Geoderma 344 (2019) 1–13

Contents lists available at ScienceDirect

Geoderma

journal homepage: www.elsevier.com/locate/geoderma

A space-time observation system for soil moisture in agricultural landscapes

N.S. Wimalathunge*, T.F.A. Bishop

Sydney Institute of Agriculture, School of Life & Environmental Sciences, The University of Sydney, Eveleigh, NSW 2015, Australia

ARTICLE INFO

Handling Editor: Jan Willem Van Groenigen

Keywords:
Agriculture
Water balance
Space time
Random Forest
Monitoring

ABSTRACT

A confluence of scientific and technological developments in geospatial data have made it possible to parameterise the components of the soil water balance equation in space and time at spatial and temporal resolutions useful for agriculture. In this work, we present results on the development of an approach that takes advantage of this opportunity to predict soil moisture at a spatial resolution of 90 m on a daily time step at multiple depths in the profile.

Three types of water balance model were examined: (i) single layer model with saturated flow (ii) multi-layer model with saturated flow and (iii) multi-layer model with unsaturated flow. Five layers were considered: 0–5, 5–15, 15–30, 30–60, and 60–100 cm, which coincide with the layers of the Soil Landscape Grid of Australia which is available at ~90 m spatial resolution. Pedotransfer functions were used to predict the bucket size for each soil layer. Precipitation and evapotranspiration are estimated by gridded SILO precipitation data (5 km, 1 day) and the MODIS 16 ET product (1 km, 8 day), respectively. Soil moisture predictions were tested against four publicly and privately owned soil moisture networks.

The multi-layer model incorporating unsaturated flow performed the best in terms of predicting soil moisture for the whole profile (0–1 m) with a median correlation coefficient (r) of just over 0.7 across all sites. When classifying the sites according to the land use; cropping sites showed better median correlation (~0.8) than grazing sites (~0.7). However, grazing sites seem to have more consistent results for all the layers. To understand the relative importance of the water balance model predictions as compared to other environmental properties, a Random Forest model was fitted to a suite of variables *e.g.* soil order, month, temperature and *etc.*, that vary in space, time or both space and time. For the analysis, only calibrated soil moisture network sites (OzNet) were considered. Soil moisture which was derived from unsaturated soil water balance model was the 9th most important variable. To assess the quality of the predictive model, leave-one-out-site cross validation (LOOSCV) was performed and across all sites the prediction quality was reasonable (Concordance = 0.66, Accuracy = 0.06 cm 3 cm^{-3}). These results highlight the potential of this approach and since it is based on readily available data it is scalable to large spatial domains.

1. Introduction

Water availability is a major determinant of food production across the globe and a key part of this is the amount of the water held by the soil. Given the potential increase in variability in climate in the future it is expected that for many agricultural production systems across the world that adapting management to soil moisture levels will become increasingly important. Knowledge of soil moisture can be used to inform a range of management decisions with the most obvious being to determine the amount and timing of irrigation (Pardossi et al., 2009). In a dryland cropping system, it can be used in conjunction with seasonal forecasts to determine target yields based on water availability for the coming season which can then inform fertiliser rates (Zhao et al., 2016). In dryland grazing systems it can be used to determine stocking rates based on the expected feed availability determined by stored soil moisture. In addition to its importance to agriculture, soil moisture is a significant part of many processes in a range of disciplines, *e.g.* hydrology, meteorology, ecology (Rodríguez-Iturbe et al., 2006).

Soil moisture is a function of a range of processes *i.e.* precipitation, evapotranspiration and is influenced by various factors *i.e.* topography, landuse, soil texture and it therefore highly varied in space and time, across different scales (Garnaud et al., 2017). The complexity and magnitude of its variation make soil moisture monitoring difficult over large areas. One way to assess the usefulness of soil moisture monitoring approaches is to consider the spatial and temporal support of the estimates of soil moisture. The spatial support is the vertical and

* Corresponding author.
E-mail address: niranjan.manikkuacharige@sydney.edu.au (N.S. Wimalathunge).

https://doi.org/10.1016/j.geoderma.2019.03.002
Received 29 August 2018; Received in revised form 6 January 2019; Accepted 1 March 2019
Available online 06 March 2019

horizontal area over which the prediction is made (Bishop et al., 2015). In terms of the horizontal spatial support, we need estimates at the within-paddock resolution (~hectare resolution) for precision agriculture applications, to the whole-paddock resolution. In the vertical spatial support, it is needed at multiple depths in the soil profile. For examples, an estimate of subsoil moisture storage is particularly important in dryland cropping systems as it greatly affects the biomass production and water use efficiency of crops (Wang et al., 2014) and to assess their capability to develop deep roots system by accessing subsoil moisture. The temporal support is the frequency of the estimates, *e.g.* daily, weekly, and the period over which they are averaged. In terms of the temporal support, soil moisture requirements depend on the agricultural system. In irrigated systems daily estimates may be needed but in dryland cropping a grower may only need estimates a few times per year, *e.g.* when they are deciding whether to sow a crop and deciding how much fertiliser to apply. In summary, if soil moisture monitoring approaches could be used to provide estimates at the hectare resolution, for multiple depths in the soil profile and on a daily time step then this would be satisfactory for most agricultural applications.

Three main approaches exist for monitoring soil moisture: remote sensing, field-based measurements such as probes, and the use of water balance models. There is an increasing range of satellite platforms available to estimate soil moisture based on microwave and radar remote sensing. For example, the European Space Agency (ESA) has a global satellite soil moisture product as part of its Climate Change Initiative (CCI) program. This product named ESA CCI SM provides global volumetric soil moisture estimates of the upper few centimetres of the soil profile with an accuracy of 0.04 $m^3 m^{-3}$ at a spatial resolution of 25 km, on a daily time step (Dorigo et al., 2015). More recently a new 1 km and 3 km resolution soil moisture product has been released, which is based on a combination of SMAP (Soil Moisture Active Passive) radiometer and Sentinel-1 SAR imagery (Das et al., 2018). The time that the SMAP overlaps the Sentinel-1 scenes determines the temporal resolution of this product, which varies across the earth. The horizontal spatial resolution is improving but this is countered by the decrease in the temporal resolution to achieve this. However, a spatial resolution of 1 km is still too coarse for most agricultural applications. The main issue for the foreseeable future for this technology is the vertical support, in that only the surface layer is measured (0 to ~5 cm depth).

In terms of field-based estimates of soil moisture, there is a classical probe, which can be placed at any depth in the profile. It provides a good way to understand temporal soil moisture patterns and when a few are placed vertically within the soil profile it also provides good estimates in terms of the vertical support. However, in terms of the horizontal spatial support they only provide point estimates and hence, large numbers of probes are required to be strategically installed across the landscape to represent horizontal spatial variation. Although probes will be getting cheaper in the future, they are unlikely to have the measurement density required to represent the landscape. Also, probes require regular maintenance and this is another limitation to their use in broad scale modelling. An alternative to static moisture probes is the use of mobile systems, which have a measurement technology mounted on a vehicle. Examples include the use of electromagnetic induction (EMI) sensors as a proxy for soil moisture (Huang et al., 2016) and mobile cosmic ray probes, which utilise naturally occurring cosmic ray neutrons as a proxy for soil moisture (Dong et al., 2014). Both approaches require extensive calibration and only provide temporally irregular measurements; however, they do provide detailed information on the spatial variation of soil moisture. Perhaps the greatest current limitation is the time it takes to perform a survey. For practical use, they would need to be installed on farm machinery, so soil moisture estimates are provided with normal field operations, *e.g.* sowing, tillage, spraying *etc.*

As an alternative to measurement in the field and through remote sensing, there are a number of monitoring approaches, which use soil water balance models that are run on spatial grids. The horizontal and vertical spatial resolution depends on the model inputs and the way the soil profile is represented, *i.e.* single layer *versus* multiple layers. Examples of implemented models include the Leaky Bucket model, which offers monthly global soil moisture at ~50 km to a depth of 1.6 m, the North American Data Assimilation System (NLDAS), which offers soil moisture at ~14 km spatial resolution to a depth of 2 m (Mitchell, 2004), and VegET offers daily Soil Water Index (SWI) at ~10 km for the crop root zone (Arndt, 2016). The Leaky Bucket is a one-layer hydrological model with a static maximum soil moisture capacity of 760 mm and uniform porosity of 0.47. It takes observed precipitation and temperature and calculates soil moisture, evaporation and runoff. NLDAS is a hydrological and meteorological product and calculates soil moisture at three depths (0–10, 10–40, 40–200 cm) with no consideration of soil texture or landuse, which vary within the landscape. The Bureau of Meteorology of Australia has a water balance model (The Australian Water Resources Assessment - AWRA) at ~5 km grid scale for three soil layers (0–10, 10–100, 100–600 cm). There is no crop representation (only grass and trees) in the model plus ET is estimated over 5 km pixels, meaning it is does not representing spatially complex crop and management patterns found in agricultural landscapes. Further AWRA reports the actual soil moisture as the percentage of available water content rather than total soil water volume. Overall, the major drawback of all of these approaches is the horizontal spatial resolution, which is 5 km or more.

In recent times there is increasing availability of geospatial products that can be used to parameterise components of the soil water balance equation at spatial resolutions useful for agriculture. In other words much finer than the 5 km that existing modelling approaches offer. Two crucial resources are digital soil maps, which when combined with pedotransfer functions (PTFs) can be used to define the water holding capacity of the soil profile and the remote sensing products, which represent variation in evapotranspiration (ET) in space and time. The Soil and Landscape Grid of Australia (SLGA) has created soil attribute maps representing the soil attributes, such as clay, silt and sand at six depths (0–5, 5–15, 15–30, 30–60, 60–100 and 100–200 cm). These maps are in raster format at a resolution of 3 arcsec (~90 m) and available across Australia (Viscarra Rossel et al., 2015). Similar products are being developed across the world as part of the GlobalSoilMap project (Stoorvogel et al., 2016). A crucial part of the soil water balance is the representation of ET. A global coverage product is available from the MODIS Global Evapotranspiration Project (MOD16). The MOD16 product was derived based on the Penman-Monteith method, which uses daily meteorological data from Global Modelling and Assimilation Office, and 8-day remotely sensed vegetation property dynamics from the MODIS optical sensors as inputs (Mu et al., 2011). It has been found to adequately represent the expected average geographical patterns and seasonality of ET (Miralles et al., 2015). In agricultural landscapes, it is valuable as reflectance measurements from MODIS are used to estimate ET, so crop rotations are somewhat represented. As a comparison, many existing approaches use air temperature in combination with fixed land use maps to estimate ET (Li et al., 2009).

Given the lack of existing soil moisture monitoring approaches that are used for agriculture and are scalable to large regions, the aim of this paper is to present an approach for predicting soil moisture for different layers within the soil profile at a 90 m spatial resolution on a daily time step. The basis of this is using recently available geospatial data products that represent inputs and components of a water balance model with an emphasis on the use of the MODIS16 product and recently developed digital soil mapping products and PTFs to define the water holding capacity of the soil. In particular, we explore the effect of different types of water balance models: (i) single layer model with saturated flow; (ii) multi-layer model with saturated flow, and (iii) multi-layer model with unsaturated flow. We test our modelling approach on networks of soil moisture probes in south-eastern Australia situated in cropping and grazing landuses and identify pathways for further

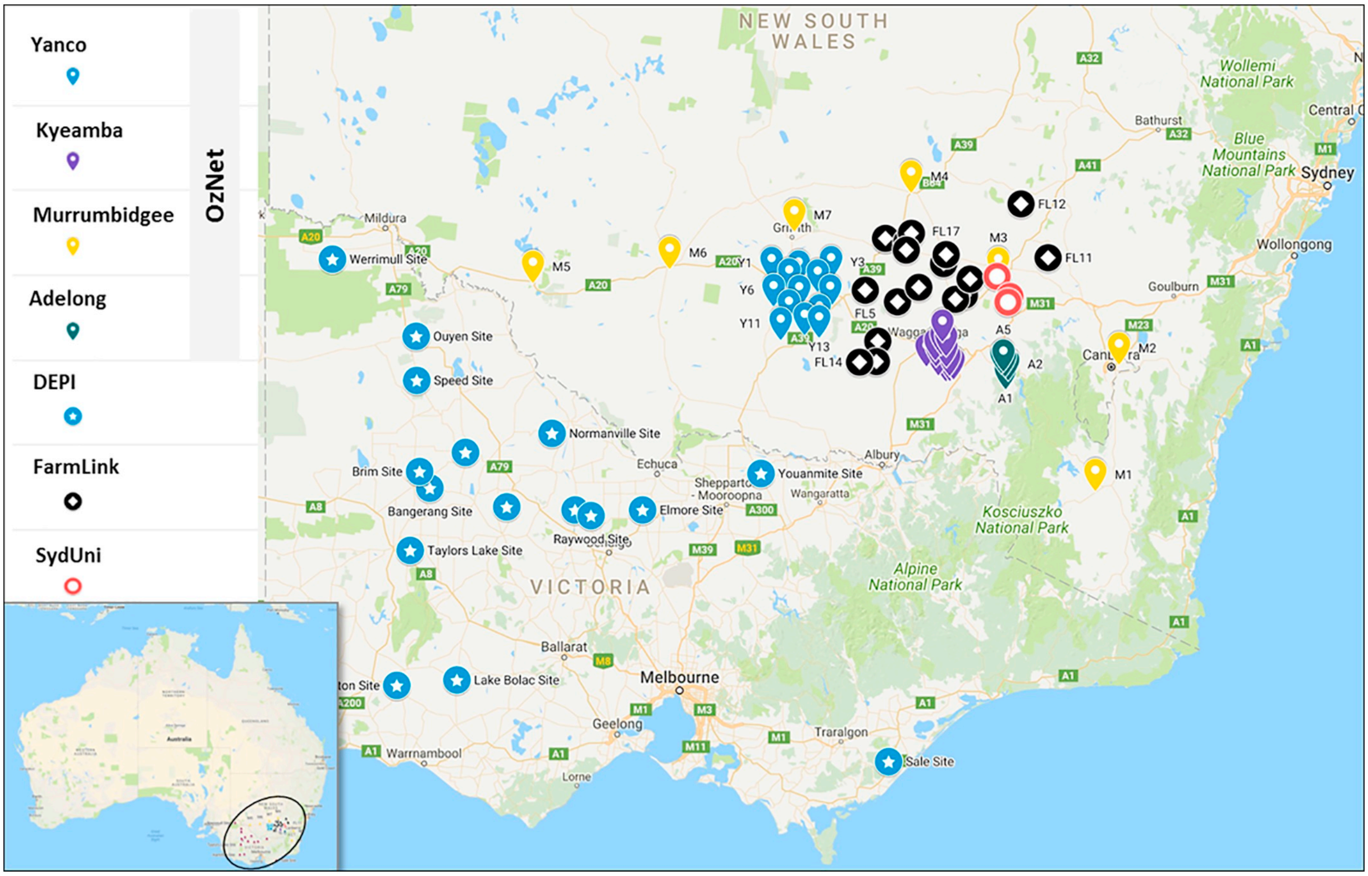

Fig. 1. Map of soil moisture probe networks overlaying a Google map image (Google, 2018).

improvements to the modelling approach.

2. Methods

2.1. Study area and soil moisture probe networks

In this work we test our modelling approach on four soil moisture probe networks; OzNet (OzNet, 2018), FarmLink (Farmlink, 2015), Victoria Department of Environment and Primary Industries (DEPI) (Agriculture Victoria, 2018); and SydUni, which are located in the agricultural landscapes of south-eastern Australia (Fig. 1).

The OzNet hydrological monitoring network was established in 2001 and encompasses 38 sites (Smith et al., 2012). It has been formed by combining four previously separate moisture probe networks; Yanco, Kyeamba, Murrumbidgee, and Adelong. The dataset consists of moisture probes installed at depth intervals of 0–5 (or 0–8), 0–30, 30–60 and 60–90 cm. Most importantly, the OzNet soil moisture probe network is field-calibrated (Smith et al., 2012). The FarmLink, DEPI and SydUni used in this work are uncalibrated. The FarmLink NSW network has 18 sites and was established in 2011. It measures soil moisture at 3–5 depths per site to a maximum depth of 112 cm (FarmLink, 2015). The SydUni soil moisture network is operated by The University of Sydney in the catchment of Muttama creek. Three sites were selected from the SydUni network, which was established in 2014. Each site has probes installed at multiple depths to a maximum depth of 120 cm. The OzNet, FarmLink and SydUni networks are located within the Murrumbidgee River Catchment (~82,000 km^2), which is a sub-catchment within the Murray-Darling Basin. Established in 2014, the Victoria Department of Environment and Primary Industries (DEPI) soil moisture probe network has 16 sites, which are spread across the state of Victoria. The DEPI probes measure soil moisture at eight depths, every 10 cm from the depth of 30 cm down to 1 m.

The annual precipitation varies across the study area from 268 to 1021 mm per year. The probes are located at elevations ranging from 60 to 930 m above sea level. In terms of land use, 59% of sites have cropping land use, 32% are grazed and 9% are located in other land uses such as native vegetation (Young et al., 2008). Fig. 2 summarises the key features of the soil moisture networks in terms of precipitation, slope, and clay content; all of which control soil moisture. The data sources for the generation of Fig. 2 can be found in Table 1.

The typical precipitation is around 450 mm with the majority of sites found in the range of 300–600 mm. The median slope is 0.5 degrees though it should be noted that some probes are located at much larger slopes meaning that lateral flow will impact on the observed soil moisture variation. Finally, clay varied in both the topsoil and subsoil from lightly textured sands and loams to medium-heavy clays, which means the inherent water storage capacity of the sites will vary significantly.

Given the variability in factors controlling soil moisture, we would expect a wide variation in soil moisture values in terms of spatial distribution and trajectory in time. This means the probe networks used in this study will provide a robust assessment of our modelling approach under a range of conditions.

2.2. The soil water balance model

The water balance for a soil profile can be expressed as:

$$\Delta SM = P + Ir - ET - DD - RO \tag{1}$$

where ΔSM is the change in moisture in the soil profile, P is precipitation, Ir is irrigation, ET is evapotranspiration, DD is deep drainage, and RO is the surface runoff. The change in the soil moisture is the difference between the amount of water added to the soil volume

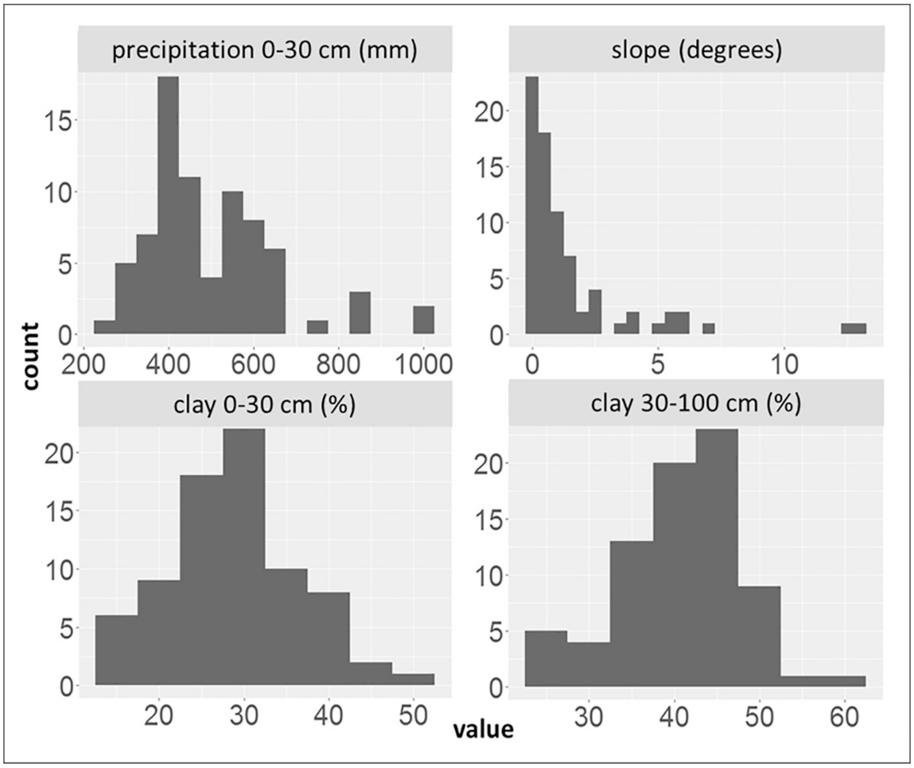

Fig. 2. Histograms of precipitation, slope, topsoil clay and subsoil clay of 75 sites.

and the amount of water withdrawn or leaving it.

The basis of our methodology is to use nationally and in some cases globally available geospatial data products to represent components of the water balance equation. Precipitation (P) was obtained from Scientific Information for Land Owners (SILO) (SILO, 2018) and ET was obtained from MODIS. SILO daily precipitation data is available at a 5 km spatial resolution and MODIS16 data was available at a spatial resolution of 1 km and a temporal resolution of 8-day totals. In order to obtain daily estimates, the ET from the MOD16 product was assumed to be uniformly distributed for each 8 day period. The focus of this study is dryland agriculture, so the irrigation (Ir) term was set to zero. Deep drainage (DD) is lost soil moisture due to transport beyond the root zone. Runoff (RO) occurs by two methods, infiltration excess runoff when precipitation exceeds infiltration rate and saturation excess runoff when precipitation falls onto saturated surfaces (Liang and Xie, 2001). The following sections explain how RO and DD are dealt with in the model.

2.2.1. Representation of soil layers, water storage, and flow

In the development of a water balance model three key issues needed to be addressed: (a) how to represent the flow down the profile (saturated or unsaturated); (b) how to represent the soil profile (single layer or multiple layers); and (c) how to estimate the amount of water held within each layer.

To address a) and b) we implemented and compared three different structures for a soil water balance model:

1. Single layer soil water balance with saturated flow
2. Multi-layer soil water balance with saturated flow
3. Multi-layer soil water balance with unsaturated flow

To address c), soil layer thickness is multiplied with saturated volumetric moisture content. SLGA depths (0–5, 5–15, 15–30, 30–60 and 60–100 cm) intervals are taken as our layer thickness of the multi-layer soil water balance model. The corresponding clay, sand and bulk density values were used to calculate the saturated volumetric moisture content (θs) using a pedotransfer function (PTF). This PTF has been developed for Australia's wheat belt, which encompasses most of the sites used in this study (Padarian et al., 2014).

All models were run on a daily time step for the period 1st Jan 2000 to the last date for each site or availability of MODIS ET data. Furthermore, it is assumed that the soil is uniform within each horizontal layer and the water flow through the soil is only in the vertical dimensions. The modelling depth for all sites is 1 m. When there is no

Table 1
Covariates to describe soil moisture dynamics.

	Covariate	Source	Resolution
Spatial	Slope, aspect and solar radiation	DEM from Geoscience Australia	90 m raster
	Land use	ABARES	1:100,000 polygons
	Soil order	ASRIS	90 m raster
	Plant available water capacity (PAWC) for top 1 m soil	ASRIS	90 m raster
Spatial & temporal	Soil water balance	SILO & MODIS	1 km, daily
	Evapotranspiration (ET) and discounted ET	MODIS	1 km, daily
	precipitation & discounted precipitation	SILO	5 km, daily
	Temperature (Temp) and discounted temperature	SILO	5 km, daily
Temporal	Month number (1 – 12)	N/A	Monthly

further rain and no soil water to take by ET, the soil moisture is assumed to be equal to the residual soil moisture content (θ_r), which was estimated from a PTF generated by Padarian et al. (2014). When we started to run the models, the initial profile soil moisture was set to 10 mm as the model runs started on January 1 in the middle of summer in the southern hemisphere and the area has winter-dominant precipitation. Therefore, the profile wetness is likely to be at its driest. It should be also noted that the earliest probes were established in 2001 so the model ran for at least a year before comparison so the initial conditions would have minimal effect. The effective spatial resolution of the model predictions is 90 m as determined by the SLGA data as the model is run on each SLGA raster cell with the corresponding value for precipitation and ET. For example, the model uses the same ET value for each 90 m raster cell within the 1 km ET raster cell.

2.2.2. Single layer soil water balance model (single bucket)

Water balance models can be constructed at any level of complexity (Zhang et al., 2002). In a single layer model, only the most important processes or components of the water balance model are represented. When appropriately used, a single layer model can provide useful insights into the functional behaviour of a system (Zhang et al., 2002). The bucket (layer) is filled up by precipitation (P) and emptied by ET. When the layer is saturated, extra water is assumed to become deep drainage or runoff. The bucket size, which is the soil moisture storage capacity (*S*) is defined by the Eq. (2)

$$S\,(\mathrm{mm}) = \sum_{i=1}^{5} \theta_s\,(\mathrm{cm^3 cm^{-3}})_i \times D\,(\mathrm{mm})_i \tag{2}$$

where θ_s is the saturated soil water content, and D_i is the soil depth of each of the i^{th} SLGA soil layers up to a depth of 1 m. The mean profile bucket size for the FarmLink, DEPI, Yanco, Kyeamba, Adelong and SydUni networks were ~410 mm (*SD*: *6*), ~420 mm (*SD*: *23*), ~425 mm (*SD*: *7*), ~394 mm (*SD*: *5*), ~463 mm (*SD*: *23*) and, ~428 mm (*SD*: *22*) respectively.

2.2.3. The multi-layers soil water balance models

Soil profiles in most cases are multi-layered where each layer's water holding capacity varies according to its properties. Multi-layer models account for vertical variation in soil water holding capacity, and therefore, better represent soil moisture variation. Our second and third approaches are saturated and unsaturated multi-layer water balance models. In the saturated model, rain (or irrigation) water infiltrates to a lower layer if the current layer is in a saturated state. In the unsaturated model, water infiltrates to a lower layer freely and continuously according to the properties of the soil layers. Neither of these model structures takes into account upward (capillary motion) or lateral flow.

Fig. 3 shows the multi-layer unsaturated flow model. In the construction of these models, ET is assumed to be an equal (50:50) contribution of evaporation and transpiration. For the multi-layer saturated flow model the 0–5 cm layer (the top layer) is filling with precipitation, emptying with surface evaporation and when it is saturated, excess soil water flows to the 5–15 cm layer. Similarly, when the 5–15 cm layer is saturated, excess soil water flows to the 15–30 cm layer. This process continues for all depth intervals defined by the SLGA data product. When there is excess soil water beyond the 60–100 cm layer, it was deemed to be deep drainage. Transpiration is generally more sensitive to the moisture content of the densely rooted soil layer; the top 30 cm of the soil profile (Feddes et al., 2001). In the case of field crops, *e.g.* wheat, most of the root mass is located in the top layers, and root density decreases with depth. Using data from 250 studies, Jackson et al. (1996) fitted an exponential decrease in root density with depth for each of 11 biome classes where they found that crops have 70% of their root biomass in the upper 30 cm. Therefore, we assumed that from Layer 5–15 cm to Layer 15–30 cm soil moisture is emptied by the transpiration fractions of 0.2 and 0.15 (70% of 0.5 ET) and from Layer 30–60 cm to Layer 60–10 cm, it is 0.1 and 0.05 (30% of 0.5 ET). Since the Layer 0–5 cm is a very thin layer (5 cm topsoil), it was assumed that runoff only occurs when both Layer 0–5 cm and Layer 5–15 cm are saturated. This assumption was applied to both saturated and unsaturated models. Continuous drainage differentiates the multi-layer with unsaturated flow model from the multi-layer with saturated flow model. In the multi-layer with the saturated flow model, downward flow only occurs when the layer is saturated. However, excess moisture drains freely according to the soil properties and the force of gravity for the multi-layer with the unsaturated flow model. The relative infiltration rates were defined through expert knowledge *i.e.* infiltration rates higher in topsoil due to tillage of soil or a greater soil organic matter. The bottom layers hold more soil moisture due to increasing clay content deeper in the soil column which means in most cases infiltration declines with depth (Ameli et al., 2016).

3. Evaluation of model performance

The model performance was assessed using three approaches: (i) correlation with the probe networks; (ii) variable importance of the water balance predictions in a Random Forest model; and (iii) the prediction quality of the Random Forest model (Fig. 4). The details and rationale are described below.

3.1. Correlation between model predictions and in-situ measurements

The purpose here was to assess how well the model predictions matched the *in-situ* probe measurements and compare this between the different model structures being considered for the water balance model. In addition, the correlation with R and ET was considered so see how well the water balance (WB) predictions improve on simply using the two main components of the WB model. The correlation was calculated separately for each site through time. Since many of the probes are uncalibrated, we only assess the strength of the linear relationship between the predictions and observation rather than a direct comparison through measures of accuracy (root-mean-square error) and bias (mean error).

Since the depth of installation varied within and between networks we compared the water balance predictions to the observations based on reference depth intervals from the OzNet Network; 0–30, 30–60 and 60–100 cm, as well as whole profile interval of 0–1 m. For each site, we calculated a weighted average of soil moisture for each interval based on the depth of installation of each probe. The model structure where the predictions had the best correlation with the probe measurements was used for the next stage of model assessment based on a Random Forest model.

3.2. Random Forest model and variable importance

In this part of the assessment, we model the measured soil moisture using a Random Forest model (Breiman, 2001). The response variable is 0–1 m profile soil moisture. We only apply this to the OzNet probe network as it has been calibrated (Smith et al., 2012). The Random Forest model was fitted to a data cube of variables that vary in space, time or both space and time, all of which are conceptually related to soil moisture dynamics (Table 1). Included in the data cube were the WB predictions. The purpose of this approach was to twofold, to see the variable importance of the WB predictions relative to other variables, and perhaps based on the variable importance identify processes not well represented in the water balance model structure.

Land use and soil order information for the study area was obtained from the Australian Bureau of Agricultural and Resource Economics (ABARES) (http://www.agriculture.gov.au/abares/aclump) and Australian Soil Resource Information System (http://www.asris.csiro.au) respectively. A 90 m digital elevation model (DEM) was obtained from Geoscience Australia (http://www.ga.gov.au/elvis). Slope, aspect

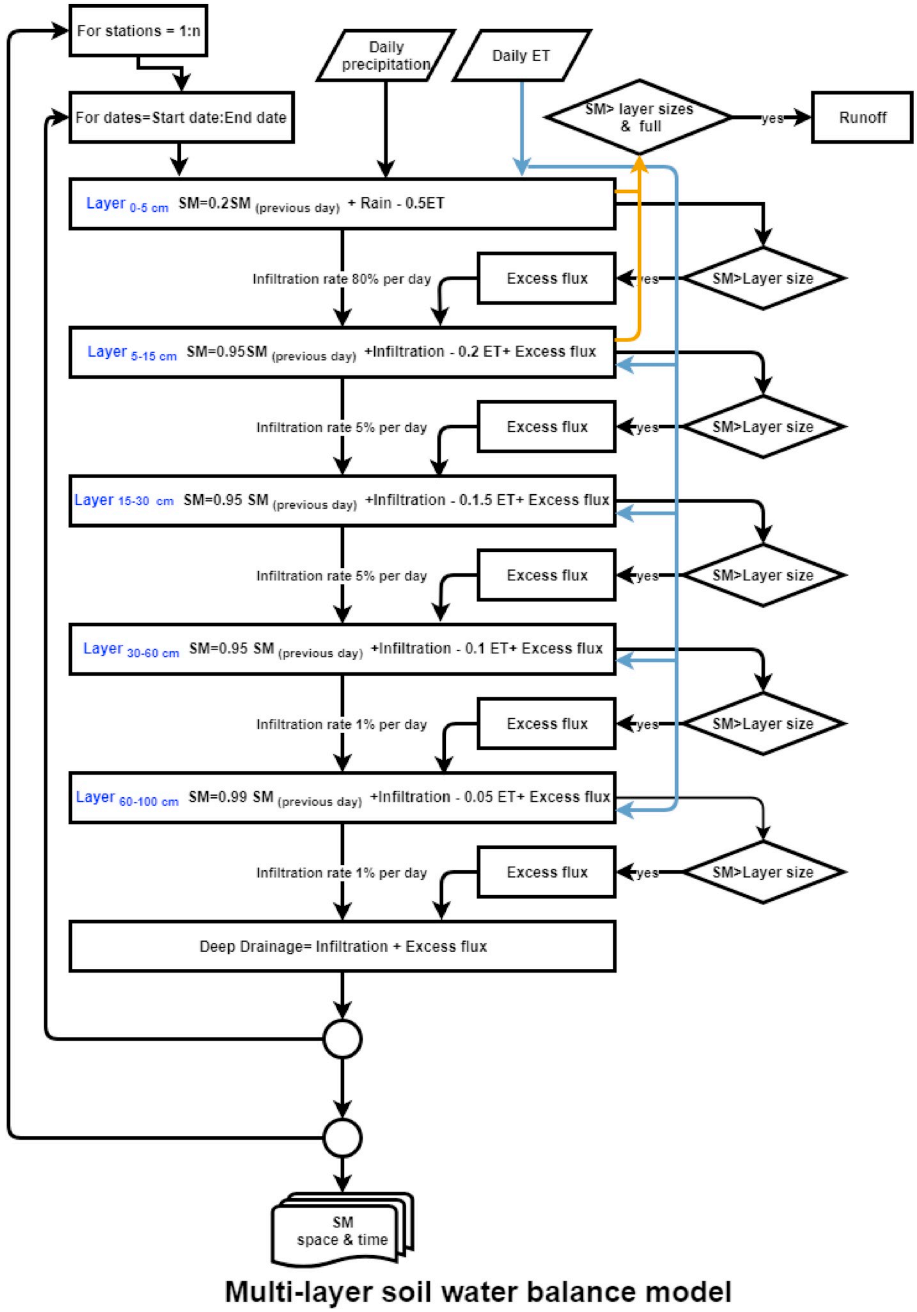

Fig. 3. Flow chart of the multi-layer soil water balance model (unsaturated). In the saturated model, there is no infiltration to the bucket below, unless the layer is saturated. SM: soil moisture, ET: evapotranspiration.

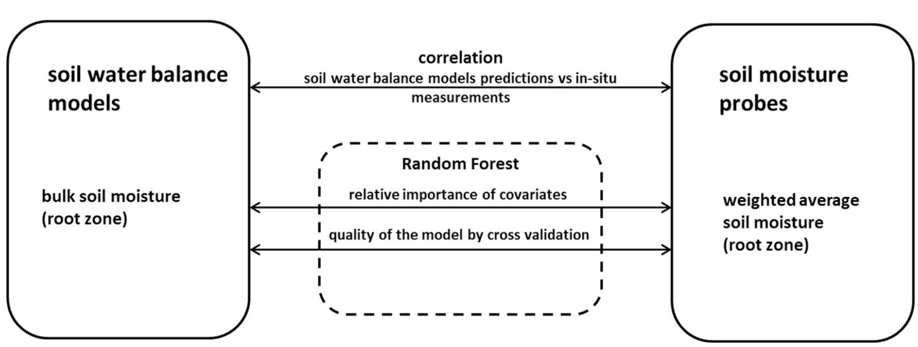

Fig. 4. Approaches used for model assessment.

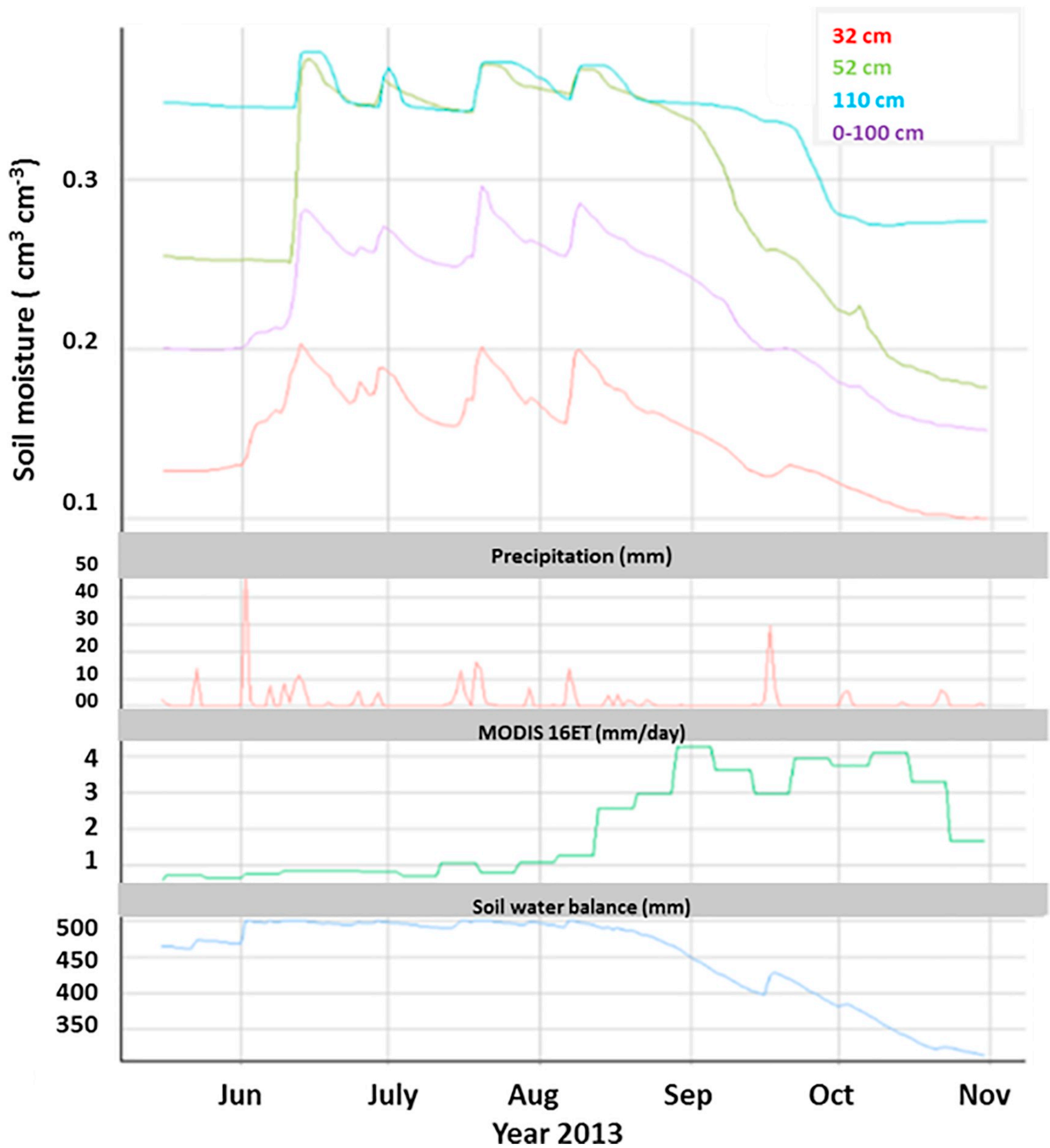

Fig. 5. FarmLink Site 12: time series plots of observed soil moisture, precipitation, MODIS16 ET, and single layer soil water balance predictions from May to November 2013.

and solar radiation were derived from the DEM using GRASS GIS (Neteler et al., 2012). In addition to these covariates, discounted ET, precipitation and temperature were added as supplementary covariates, which use a weighting function (for calculating the temporal mean) 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). Discount factors (df) that were considered include 0.5, 0.7, 0.9, 0.95, 0.99 and 0.999. The df controls the rate per day at which the weighting decays with time in the past. For example, with a df of 0.5, the precipitation of two days ago gets a weighting of $0.5^2 = 0.25$ when calculating the average precipitation (Horta and Bishop, 2013).

To ensure the stability in variable selection, the Boruta feature selection algorithm was used (Kursa and Rudnicki, 2010) when fitting the Random Forest model. We use the R package "Boruta",which has a wrapper function built around the random forest classification algorithm implemented in the R package "randomForest" (Liaw and Wiener, 2002). The steps of the Boruta algorithm are: (1) add randomness to the dataset creating shuffled copies called shadow features; (2) train a Random Forest classifier and calculate the Mean Decrease in Accuracy (MDA) to evaluate the importance of each feature; (3) at every iteration it checks whether a feature has a higher importance than the best of its shadow features and removes features, which are highly unimportant; and (4) algorithm stops either when all features get confirmed or rejected or it reaches a specified limit of random forest runs (Kursa and Rudnicki, 2010). It trains a Random Forest classifier on the extended data set and applies a feature importance measure (the default is MDA) to evaluate the importance of each feature where a higher value means the feature (predictor) is more important.

3.3. Assessment of Random Forest model predictions

The final assessment of the modelling approach was to examine the quality of the predictions of the Random Forest model. This tests an approach where we use probe networks to train machine learning methods to predict soil moisture in space and time. The rationale here is that in the future with more widespread soil moisture probes we could use a data-driven approach to predict soil moisture. The purpose of the

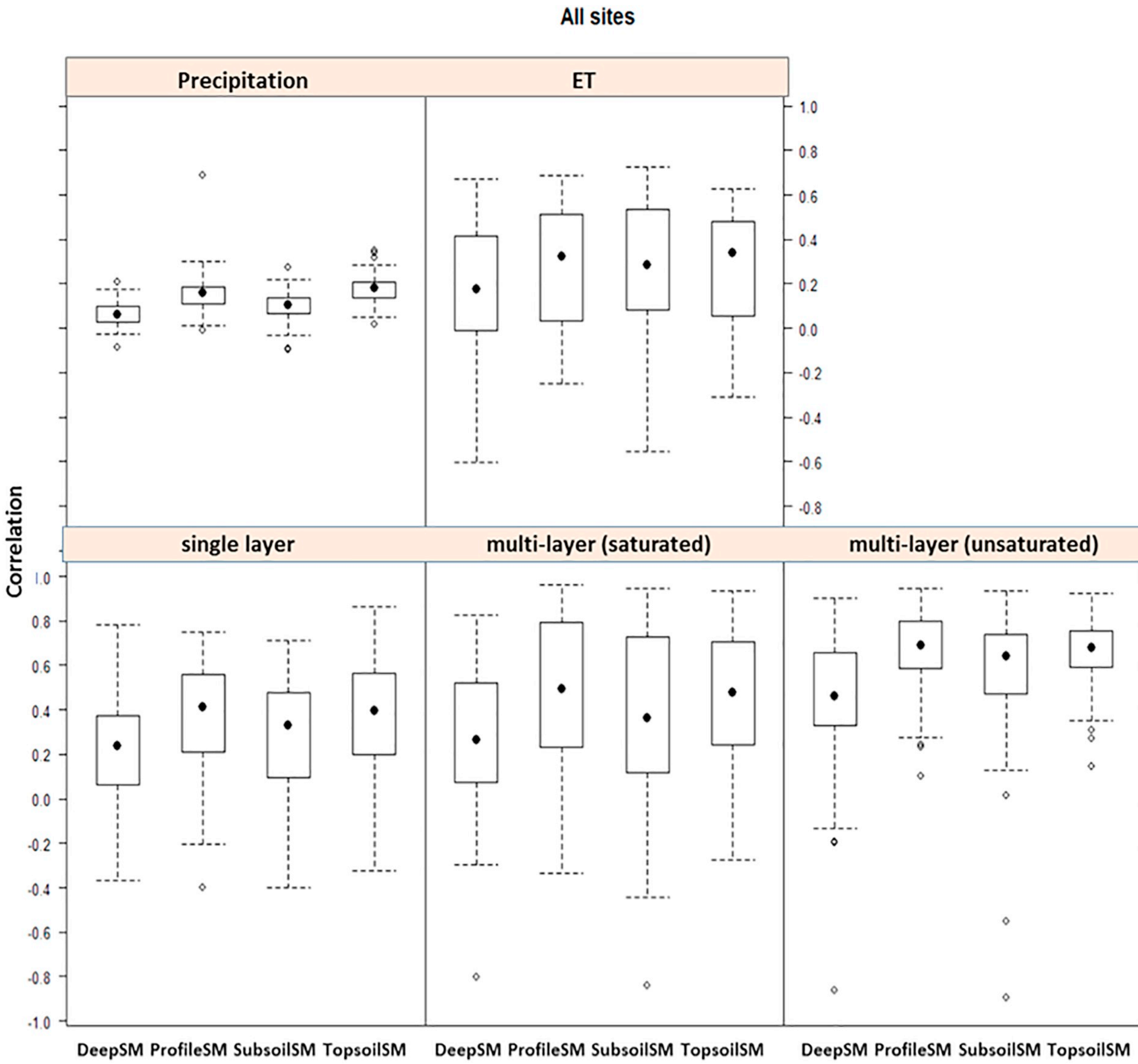

Fig. 6. Boxplots of the correlation between probe observations (all sites) and ET, precipitation and water balance model predictions.

water balance model would then be a feature extraction method for deriving better predictors of soil moisture. Random Forests has out-of-bag estimates (based on observations not selected by bootstrapping) of prediction quality but due to the serial (temporal) auto-correlation expected to be evident in the soil moisture data, this is an inadequate test of prediction quality. Therefore, we used leave-one-out-site cross validation (LOOSCV) to check the quality of the model predictions for all sites. The main aim of the LOOSCV is to characterise how the model performs at a location that is not used in fitting the model; in other words the generalisability of the model. This 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 validation metrics are calculated; Lin's concordance correlation coefficient (LCCC) and the root-mean-square error (RMSE). Lin's concordance represents how well the observations follow the 45-degree line when plotting observed *versus* predicted (Lin, 1989) and the RMSE represents the accuracy.

4. Results

4.1. An example

To illustrate the modelling approach Fig. 5 shows the observed soil moisture and the single layer model inputs/outputs for the growing season of a cereal crop. The site is part of the FarmLink Network and is situated 15 km east of Greenthorpe in southern New South Wales, Australia. From June to mid-August the observed soil moisture has increased at all depths as there are frequent precipitation events. As the precipitation continues, moisture increases at depth in the soil profile. The moisture decreases from mid-August onwards at all depths as there is less precipitation and increased evapotranspiration due to the growing cereal crop. This pattern is also shown in our soil water balance model predictions.

4.2. Correlation between soil water balance models input/outputs and observations

Fig. 6 presents boxplots of the correlations across all sites between the model predictions and observations for different depth intervals; 0–30 cm (TopsoilSM), 30–60 cm (SubsoilSM), 60–100 cm (DeepSM); and 0–1 m (ProfileSM).

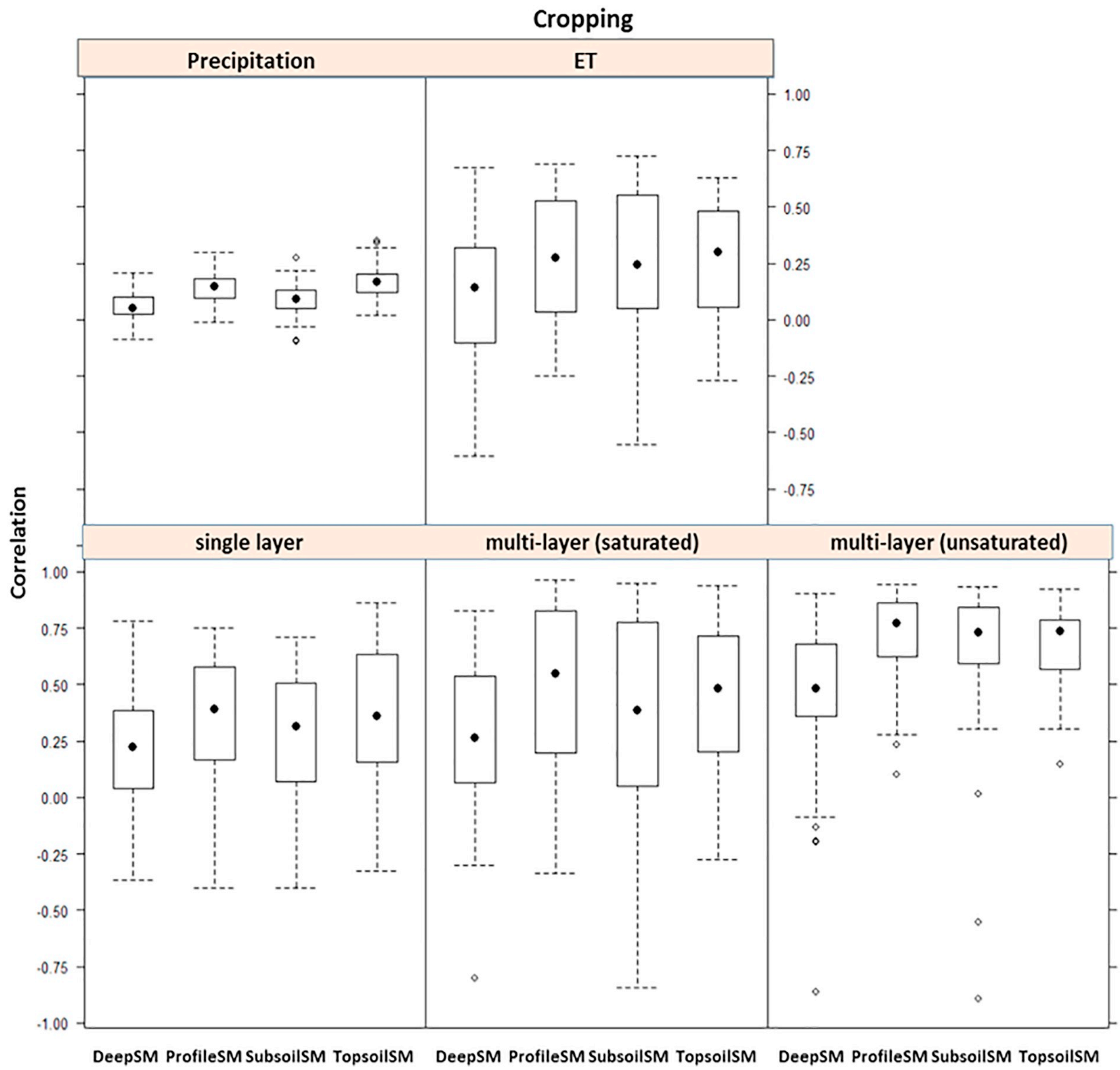

Fig. 7. Boxplots of the correlation between probe observations (cropping) and ET, precipitation and water balance model predictions.

The general trend across all depth intervals is that as the model complexity increases the correlations improve (single layer > multi-layer saturated > multi-layer unsaturated). In addition, the WB predictions are better correlated with observed soil moisture than R and ET. When we focus on the best model (multi-layer unsaturated flow), the correlations do decline with depth with a median correlation ~0.7 in the topsoil, with reduces to a median of ~0.5 in the lower part of the profile (0.6–1.0 m). This is likely due to difficulty in representing flow between the different layers and attributing where ET processes are extracting water from within the profile. An interesting result is that the multi-layer unsaturated flow model gives better predictions of the whole profile (median correlation ~0.7) than the single layer model (median correlation ~0.45). This indicates that modelling the within profile soil moisture dynamics should be attempted even when we are only interested in estimating the whole-profile soil moisture storage.

It should also be noted that some sites have a negative correlation as shown in Fig. 6. One possibility is that the water balance inputs at these sites are poorly represented with the most likely issues being with precipitation and ET. For example, the site could be situated in a region with variation in land use in the footprint of the 1 km ET pixels, such as being adjacent to an urban area or forest or have multiple paddocks at different stages of their rotation. This would mean the spectral measurements from MODIS could have a mixed signal which may not represent the reality of what is occurring at the site. However, the model that we concern about the multilayer (unsaturated), mainly has positive correlations (Figs. 6, 7 & 8).

Figs. 7–8 present the results separated by the two dominant land uses - grazing and cropping. In terms of the median correlation for all depths, there is a slight improvement in correlation for cropping as compared to grazing. As an example for the 0–1 m depth interval, the median correlation coefficient of cropping sites is ~0.8 for the multi-layer unsaturated model and for the grazing sites the median correlation coefficient is ~0.7. However, for the grazing sites, the variation in the correlation coefficient is smaller for all depth intervals. This is likely to be due to the effect of rotations in the cropping land use, which causes a less stable performance across all sites.

Overall, the multi-layer unsaturated model is the best of the three model structures and is used for further analysis with the Random Forest modelling where we focus on modelling the profile soil moisture, represented by the 0–1 m depth interval.

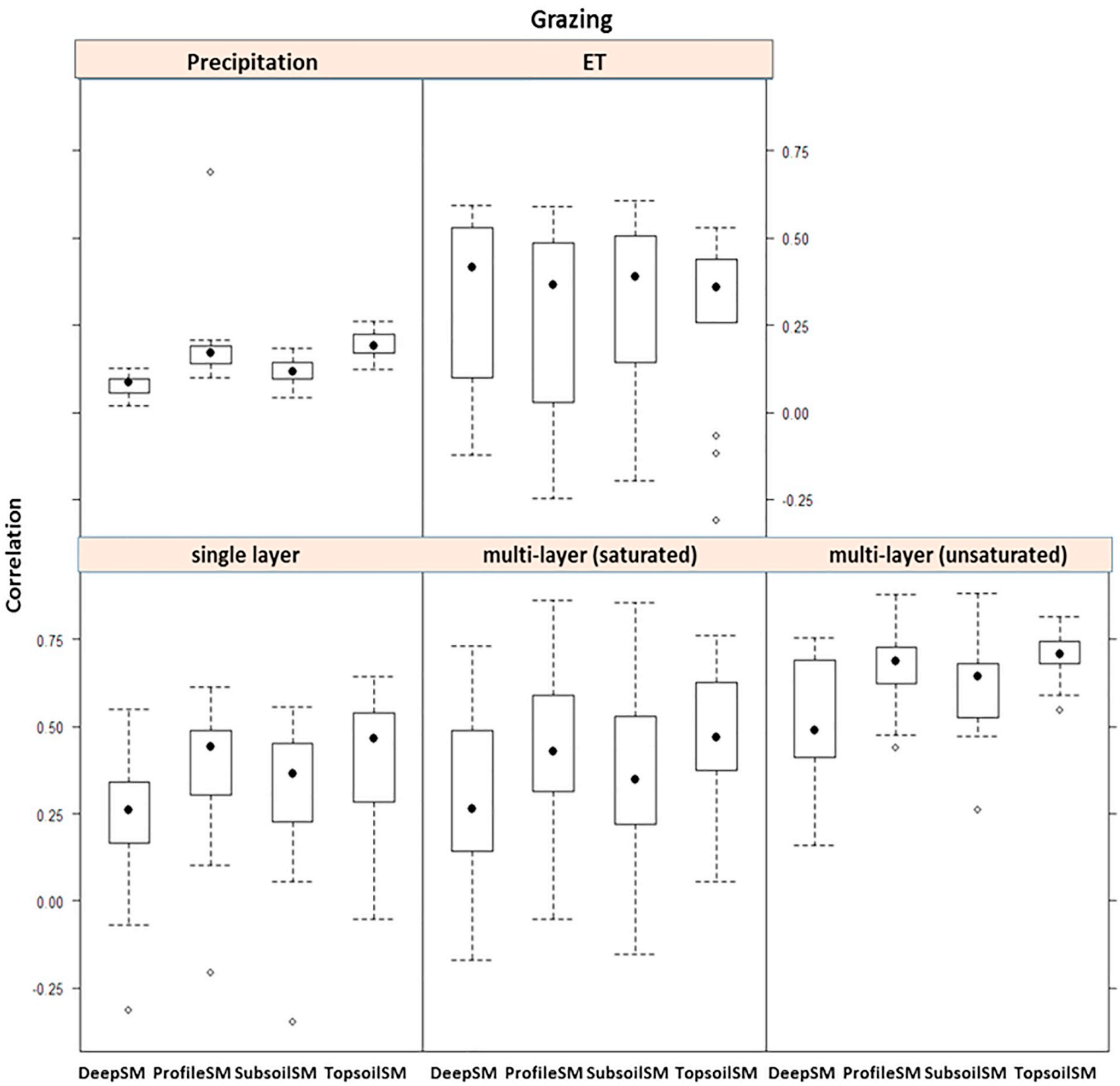

Fig. 8. Boxplots of the correlation between probe observations (grazing) and ET, precipitation and water balance model predictions.

4.3. Importance of covariates Random Forest predictive model

Fig. 9 presents the relative importance of each covariate based on Random Forest modelling of the OzNet dataset.

The water balance predictions (SM) are the 9th most important and above it, there are three general types of predictors, a temporal trend term (calendar month), longer term precipitation and ET averages as represented by discount factors, and spatial variables related to the landscape shape and orientation. The month is the most important predictor, which indicates the water balance model is not representing the temporal trend in the soil moisture. This is likely due to our representation of the rooting depth as constant through time when in reality for winter cropping, which dominates this area it would vary with the crop development, or in some cases, there may not be a root system present at all, *i.e.* fallow. Spatial variables extracted from DEM: aspect (3rd), slope (4th) and solar radiation (6th) are among the top predictors. According to Martinez et al. (2008), aspect is considered to be the main factor influencing the spatial and temporal distribution of near-surface soil moisture. This finding is supported by Western et al. (1999), who found that aspect, expressed in terms of a potential radiation index, exerts significant control over the spatial distribution of soil moisture. Moreover, the MODIS evapotranspiration has a 1 km spatial resolution, which would average out the effects of slope aspect and steepness on plant water use meaning that terrain attributes effectively downscale the ET estimates used in the WB model.

4.4. Assessment of Random Forest model predictions

Across all sites, the model predicts reasonably well with LCCC of 0.66 and RMSE of 0.06 $cm^3 cm^{-3}$. The grazing sites gave better predictions (LCCC = 0.69) as compared to cropping sites (LCCC = 0.50). The scatter plots (Fig. 10) also look visually better for the grazing with a tighter cluster around the 1:1 line than for cropping, which has more spread and some systematic under-prediction. This is likely due to cropping land uses being more dynamic in terms of fallows and rotations than grazing making them more difficult to model. In addition, some of the systematic features can be attributed to unmodeled features, for example, one site (Y4 – Yanco) had flood irrigation in 2004 (Smith et al., 2012) meaning soil moisture was under-predicted using the RF model predictions. We kept this site in the study to illustrate what could go wrong if something site-specific, *e.g.* an occasional irrigation event, is not included in the modelling approach.

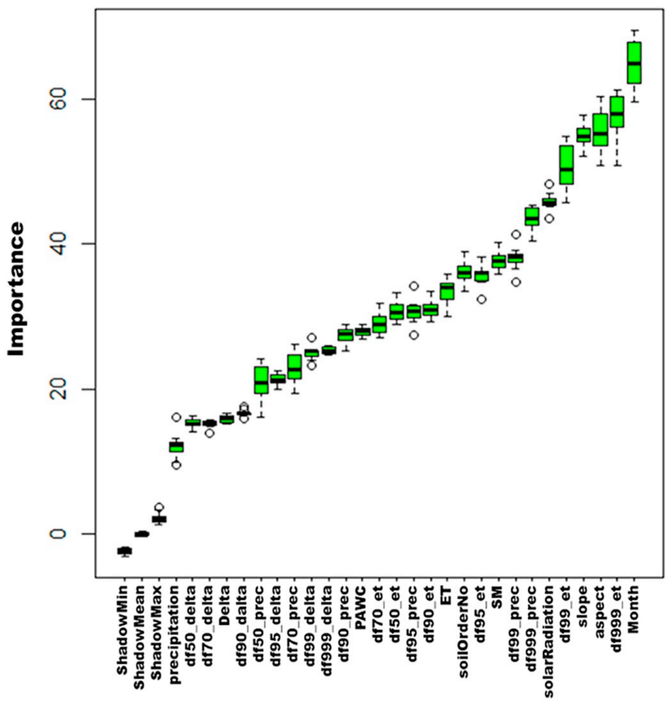

Fig. 9. Importance of covariates (OzNet). All the attributes were confirmed to be important by the Boruta algorithm. ShadowMin, shadowMax and shadowMean are Boruta shadow features as explained in Section 3.2.

5. Conclusion

This study presented an approach for predicting daily soil moisture at ~90 m resolution for different depths in the soil profile. Three structures for the water balance model were compared, and the best was the multi-layer model with unsaturated flow. The correlation between model predictions and observed soil moisture declined with depth with a median correlation of ~0.7 for the topsoil and ~0.5 for the subsoil and ~0.7 for the root-zone.

The variable importance of covariates in the Random Forest model-showed that shortcomings in the water balance model were related to not accounting for short range terrain features represented by slope and aspect, which we speculate as being due to the 1 km spatial resolution of the MODIS16 product. A further shortcoming identified was the assumption of constant rooting depth through time when estimating from which layers in the soil that the plants extracted water. Despite these shortcomings, the predictive capability of the Random Forest model was quite good with LCC of 0.66 with RMSE 0.06 $cm^3\,cm^{-3}$ with grazing being better predicted than cropping land uses.

Over time it is expected that we will have better geospatial products (precipitation and evapotranspiration) in terms of accuracy and resolution, and therefore the water balance model predictions will improve. For example, the MODIS16 ET product is quite coarse, but an alternative may be possible from Sentinel and Landsat platforms (Senay et al., 2016) which provide sub-hectare resolution predictions of ET. In terms of the model structure, future work should explore improvements in the modelling of unsaturated flow in combination with the use of PTFs for predicting hydraulic conductivity. Part of the improvements to the model structure would also be to include rooting depth as a time dependent variable based on crop type and development. This is now possible as dynamic crop type mapping using remotely sensed imagery

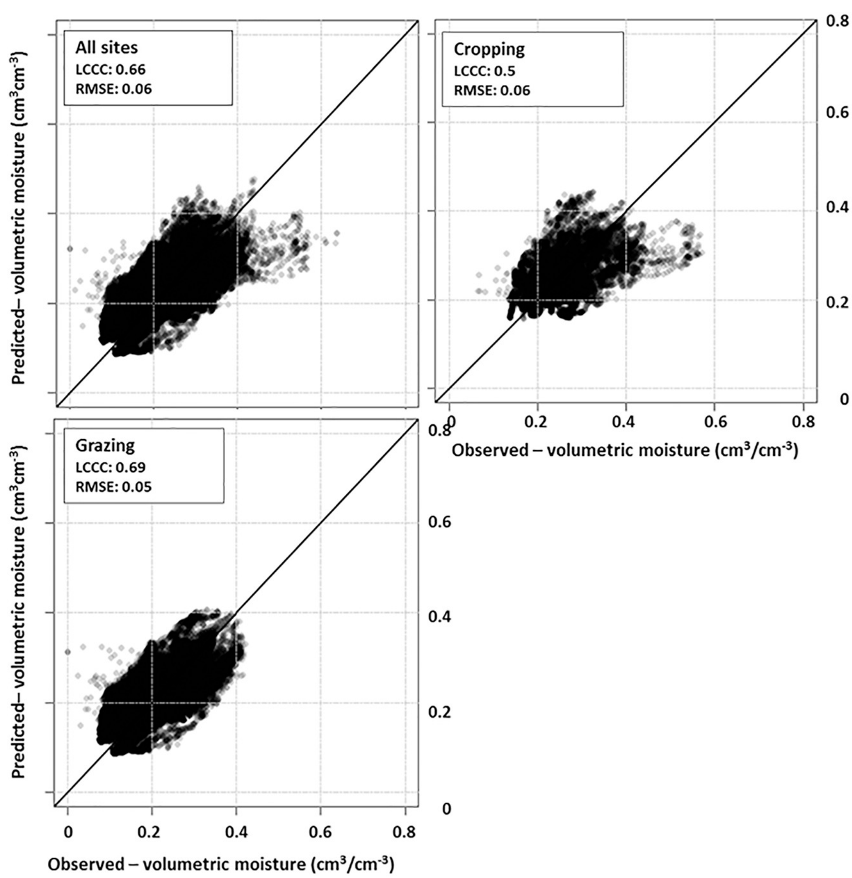

Fig. 10. Plot of observed *vs* predicted soil moisture (0–1 m depth) for OzNet.

is now becoming operational, one example being Pringle et al. (2018). A further area for improving the approach is to incorporate soil moisture observations into the approach, for example, the Sentinel/SMAP (1 km, 12–24 day, top 5 cm) soil moisture product (Das et al., 2018). These soil moisture estimates can be incorporated as a constraint to improve the model's predictions. Alternatively, we can also fine tune the model (parameters) itself, *i.e.* those related to unsaturated flow. Finally, the use of the soil moisture probe measurements as the response variable for predictive modelling and mapping was promising. Therefore, rather than using observations to test or improve water balance model predictions, an alternative is to use all soil moisture measurements in a large data-driven predictive model whether they are from *in-situ* probes, mobile field-based sensors or from remote sensing. The key will be to account for the different spatial and temporal resolutions of the observations in the predictive model. One possible approach is an extension of an area-to-point kriging approach to 3D and/or 4D with time. Typically this is performed in the horizontal dimension (Orton et al., 2015) or the vertical dimension (Orton et al., 2016) but not in both or 4D (including time).

In conclusion, we have presented a methodology for predicting soil moisture in space and time for agricultural landscapes. The approach relies on freely available geospatial data and is scalable to Australia and beyond. Future work will focus on improving the approach and creating an operational system.

Acknowledgments

The authors would like to thank OZNET for making available to soil moisture datasets publicly available. We also would like to acknowledge Dale Boyd (http://agriculture.vic.gov.au/agriculture) for giving access to DEPI datasets. The authors also acknowledge Farmlink Research, (http://www.farmlink.com.au/) for providing us with the FarmLink dataset. Also, we would like to gratefully acknowledge Ross Searle from CSIRO, and the funding support provided by the Australian Government through the National Landcare Programme for the PhD scholarship of the first author. Additionally, the author's thank Edward Jones and Senani Karunaratne for their constructive feedback on early drafts of this paper, and the anonymous reviewers for the useful comments on the submitted manuscript.

Appendix A

Table 2
Dataset sizes.

Site	FarmLInk	DEPI	Kayeamba	Yanco	Murrumbidgee	Adelong	SydUni
1	1069	546	3229	1428	3012	2998	931
2	984	806	3058	1118	3206	1628	930
3	950	1255	2814	196	1036	2458	931
4	950	489	3152	1381	3363	1447	
5	716	489	2938	1491	2133	2561	
6	830	631	2341	1342	3148	–	
7	996	543	2547	1513	3407	–	
8	754	1275	2014	1379	–	–	
9	754	1258	–	1271	–	–	
10	985	806	2179	1500	–	–	
11	1009	1270	2023	1287	–	–	
12	832	1128	2158	1410	–	–	
13	1031	1263	2552	1137	–	–	
14	357	537	2291	–	–	–	
15	204	1264	–	–	–	–	
16	223	1270	–	–	–	–	
17	356	–	–	–	–	–	
18	556	–	–	–	–	–	
Dataset size	13,556	14,830	33,296	16,453	19,305	11,092	2792

References

Agriculture Victoria, 2018. Soil Moisture Monitoring in Dryland Cropping Areas | Crop production | Grains and other crops | Agriculture | Agriculture Victoria. [online] Available at: http://agriculture.vic.gov.au/agriculture/grains-and-other-crops/crop-production/soil-moisture-monitoring-in-dryland-cropping-areas [Accessed 29 Jul. 2018]. .

Ameli, A., McDonnell, J., Bishop, K., 2016. The exponential decline in saturated hydraulic conductivity with depth: a novel method for exploring its effect on water flow paths and transit time distribution. Hydrol. Process. 30 (14), 2438–2450.

Arndt, E. (2016). Did You Know?|Monitoring References|National Centers for Environmental Information (NCEI). [online] Ncdc.noaa.gov. Available at: https://www.ncdc.noaa.gov/monitoring-references/dyk/soil-moisture-models [Accessed 14 Jul. 2016].

Bishop, T., Horta, A., Karunaratne, S., 2015. Validation of digital soil maps at different spatial supports. Geoderma 241-242, 238–249.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.

Das, N., Entekhabi, D., Kim, S., Jagdhuber, T., Dunbar, S., Yueh, S., O'Neill, P., Colliander, A., Walker, J., Jackson, T., 2018. High-resolution enhanced product based on SMAP active-passive approach using sentinel 1A and 1B SAR data. ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-5, 203–205.

Dong, J., Ochsner, T., Zreda, M., Cosh, M. and Zou, C. (2014). Calibration and validation of the COSMOS rover for surface soil moisture measurement. Vadose Zone Journal, 13(4), p.0.

Dorigo, W., Gruber, A., De Jeu, R., Wagner, W., Stacke, T., Loew, A., Albergel, C., Brocca, L., Chung, D., Parinussa, R., Kidd, R., 2015. Evaluation of the ESA CCI soil moisture product using ground-based observations. Remote Sens. Environ. 162, 380–395.

Farmlink Research, (2015). Available from: ttp://www.farmlink.com.au/project/moisture-knowledge-network. [28 June 2015].

Feddes, R., Hoff, H., Bruen, M., Dawson, T., de Rosnay, P., Dirmeyer, P., Jackson, R., Kabat, P., Kleidon, A., Lilly, A., Pitman, A., 2001. Modeling root water uptake in hydrological and climate models. Bull. Am. Meteorol. Soc. 82 (12), 2797–2809.

Garnaud, C., Bélair, S., Carrera, M., McNairn, H., Pacheco, A., 2017. Field-scale spatial variability of soil moisture and L-band brightness temperature from land surface modeling. J. Hydrometeorol. 18 (3), 573–589.

Google (2018). Sign in - Google accounts. [online] Available at: https://www.google.com.au/maps/d/edit?mid = 1QfartVlm8EOyaOo8F4QydPAJlco&ll = -34.653266706600995%2C145.13906710644528&z = 7 [Accessed 27 Aug. 2018].

Horta, A., Bishop, T., 2013. Spatio-temporal analysis to improve agricultural management. In: Oliver, M., Marchant, B. (Eds.), Precision Agriculture for Sustainability and Environmental. Protection, Abingdon, pp. 119–132. https://doi.org/10.4324/9780203128329.

Huang, J., Monteiro Santos, F., Triantafilis, J., 2016. Mapping soil water dynamics and a moving wetting front by spatiotemporal inversion of electromagnetic induction data. Water Resour. Res. 52 (11), 9131–9145.

Jackson, R., Canadell, J., Ehleringer, J., Mooney, H., Sala, O., Schulze, E., 1996. A global analysis of root distributions for terrestrial biomes. Oecologia 108 (3), 389–411.

Kursa, M. and Rudnicki, W. (2010). Feature selection with the Boruta package. Journal of Statistical Software, 36(11). p.3.

Lessels, J., Bishop, T., 2013. Estimating water quality using linear mixed models with stream discharge and turbidity. J. Hydrol. 498, 13–22.
Li, Z., Tang, R., Wan, Z., Bi, Y., Zhou, C., Tang, B., Yan, G., Zhang, X., 2009. A review of current methodologies for regional evapotranspiration estimation from remotely sensed data. Sensors 9 (5), 3801–3853.
Liang, X., Xie, Z., 2001. A new surface runoff parameterization with subgrid-scale soil heterogeneity for land surface models. Adv. Water Resour. 24 (9–10), 1173–1193.
Liaw A, Wiener M (2002). "Classification and regression by random forest." R News, 2(3), 18–22. URL http://CRAN.R-project.org/doc/Rnews/.
Lin, L., 1989. A concordance correlation coefficient to evaluate reproducibility. Biometrics 45, 255–268.
Martinez, C., Hancock, G.R., Kalma, J.D., Wells, T., 2008. Spatio-temporal distribution of near-surface and root zone soil moisture at the catchment scale. Hydrol. Process. 22 (14), 2699–2714. https://doi.org/10.1002/hyp.6869.
Miralles, D., Jiménez, C., Jung, M., Michel, D., Ershadi, A., McCabe, M., Hirschi, M., Martens, B., Dolman, A., Fisher, J., Mu, Q., Seneviratne, S., Wood, E., Fernaìndez-Prieto, D., 2015. The WACMOS-ET project – part 2: evaluation of global terrestrial evaporation data sets. Hydrol. Earth Syst. Sci. Discuss. 12 (10), 10651–10700.
Mitchell, K., 2004. The multi-institution North American Land Data Assimilation System (NLDAS): utilizing multiple GCIP products and partners in a continental distributed hydrological modeling system. J. Geophys. Res. 109 (D7).
Mu, Q., Zhao, M., Running, S., 2011. Improvements to a MODIS global terrestrial evapotranspiration algorithm. Remote Sens. Environ. 115 (8), 1781–1800.
Neteler, M., Bowman, M., Landa, M., Metz, M., 2012. GRASS GIS: a multi-purpose open source GIS. Environ. Model Softw. 31, 124–130.
Orton, T., Pringle, M., Allen, D., Dalal, R., Bishop, T., 2015. A geostatistical method to account for the number of aliquots in composite samples for normal and lognormal random variables. Eur. J. Soil Sci. 66 (6), 1023–1032.
Orton, T., Pringle, M., Bishop, T., 2016. A one-step approach for modelling and mapping soil properties based on profile data sampled over varying depth intervals. Geoderma 262, 174–186.
OzNet (2018). OzNet Hydrological Monitoring Network. [online] Available at: http://www.oznet.org.au/ [Accessed 19 Apr. 2018].
Padarian, J., Minasny, B., McBratney, A., Dalgliesh, N., 2014. Predicting and mapping the soil available water capacity of Australian wheatbelt. Geoderma Regional 2-3, 110–118.
Pardossi, A., Incrocci, L., Incrocci, G., Malorgio, F., Battista, P., Bacci, L., Rapi, B., Marzialetti, P., Hemming, J., Balendonck, J., 2009. Root zone sensors for irrigation management in intensive agriculture. Sensors 9 (4), 2809–2835.
Pringle, M., Schmidt, M., Tindall, D., 2018. Multi-decade, multi-sensor time-series modelling—based on geostatistical concepts—to predict broad groups of crops. Remote Sens. Environ. 216, 183–200.
Rodríguez-Iturbe, I., Isham, V., Cox, D., Manfreda, S., Porporato, A., 2006. Space-time modeling of soil moisture: stochastic rainfall forcing with heterogeneous vegetation. Water Resour. Res. 42 (6).
Senay, G., Friedrichs, M., Singh, R., Velpuri, N., 2016. Evaluating Landsat 8 evapotranspiration for water use mapping in the Colorado River Basin. Remote Sens. Environ. 185, 171–185.
SILO, 2018. SILO climate data (Long Paddock). [online] Available at: https://legacy.longpaddock.qld.gov.au/silo/ [Accessed 19 Apr. 2018]. .
Smith, A., Walker, J., Western, A., Young, R., Ellett, K., Pipunic, R., Grayson, R., Siriwardena, L., Chiew, F. and Richter, H. (2012). The Murrumbidgee soil moisture monitoring network data set. Water Resources Research, 48(7), p.n/a-n/a.
Stoorvogel, J., Bakkenes, M., Temme, A., Batjes, N., ten Brink, B., 2016. S-world: a global soil map for environmental modelling. Land Degrad. Dev. 28 (1), 22–33.
Viscarra Rossel, R., Chen, C., Grundy, M., Searle, R., Clifford, D., Campbell, P., 2015. The Australian three-dimensional soil grid: Australia. Soil Research.
Wang, Y., Kuhnert, P., Henderson, B., 2011. Load estimation with uncertainties from opportunistic sampling data – a semiparametric approach. J. Hydrol. 396 (1–2), 148–157.
Wang, B., Wen, F., Wu, J., Wang, X. and Hu, Y. (2014). Vertical profiles of soil water content as influenced by environmental factors in a small catchment on the hilly-gully loess plateau. PLoS ONE, 9(10), p.e109546.
Western, A.W., Grayson, R.B., Bloschl, G., Willgoose, G.R., McMahon, T.A., 1999. Observed spatial organization of soil moisture and its relation to terrain indices. Water Resour. Res. 35 (3), 797–810.
Young, R., Walker, J., Yeoh, N., Smith, A., Ellett, K., Merlin, O., Western, A., 2008. Soil Moisture and Meteorological Observations From the Murrumbidgee Catchment. The Univeristy of Melbourne, Department of Civil and Environmental Enginnering.
Zhang, L. Walker, G.R. and Dawes, W.R. (2002). Water balance modelling: concepts and applications. In: McVicar, T.R., Li Rui, Walker, J., Fitzpatrick, R.W. and Liu Changming (eds), Regional Water and Soil Assessment for Managing Sustainable Agriculture in China and Australia, ACIAR Monograph No. 84, 31–47.
Zhao, H., Wang, Z., Xue, C., Huang, D., Malhi, S., 2016. Increased dryland winter wheat yields by nitrogen fertilizer topdressing and its relationship to soil moisture, available N, P and K in northwestern China. Commun. Soil Sci. Plant Anal. 47 (5), 543–553.

A space-time observation system for soil moisture in agricultural landscapes

Wimalathunge, N. S.; Bishop, T. F.A.

01 mingxi zhang Page 4

23/5/2024 2:25

02 mingxi zhang Page 5

23/5/2024 3:26