



# Bioclimatic variables as important spatial predictors of soil hydraulic properties across Australia's agricultural region

B.P. Malone <sup>a,\*</sup>, Z. Luo <sup>b</sup>, D. He <sup>a</sup>, R.A. Viscarra Rossel <sup>c</sup>, E. Wang <sup>a</sup>

<sup>a</sup> CSIRO Agriculture and Food, Black Mountain, Canberra, ACT, Australia

<sup>b</sup> College of Environmental and Resource Sciences, Zhejiang University, Hangzhou, China

<sup>c</sup> Soil and Landscape Science, School of Molecular & Life Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia

## ARTICLE INFO

### Article history:

Received 24 April 2020

Received in revised form 25 September 2020

Accepted 27 September 2020

### Keywords:

Digital soil mapping  
Available water capacity  
Drained upper limit  
Soil hydraulic variables  
Bio-climatic variables  
Australian agriculture  
Cambisols  
Ferralsols  
Lixisols  
Luvisols  
Solonchaks  
Solonetz  
Vertisols

## ABSTRACT

Soil water directly or indirectly affects almost all ecological processes. Soil available water capacity (AWC), the difference between field capacity, or drained upper limit (DUL), and wilting point, or lower limit (LL15), and saturated water content (SAT) are among the most important soil hydraulic properties controlling soil water dynamics. These properties vary across space and are expensive to measure directly. It is difficult to obtain reliable estimates of soil hydraulic properties at an appropriate scale for water and land management. Here we modelled LL15, DUL, SAT and AWC measurements from 1127 whole-soil profiles across Australian agricultural areas with the Random Forest machine learning model using 19 bioclimatic and 15 topographical covariates. The amount of variance explained by the model reached up to  $R^2 = 0.69$  depending on the property and soil depth assessed. For all soil hydraulic properties, the bioclimatic variables alone contributed to more than 90% of the explained variance. Particularly, temperature of driest and wettest quarter, and precipitation of warmest month were the three most influential variables. Using the derived models, we also mapped the four hydraulic properties across Australian agricultural areas in six sequential depths down to 2 m at a spatial resolution of 90 m. Moreover, we combined our mapping of AWC with existing products via an ensemble model averaging approach which proved to be more accurate than each of the three contributing products. Our results uncover the significant role of bioclimatic variables in regulating soil hydraulic properties, providing a benchmark assessment of soil hydraulic properties in agricultural regions for efficient water-related land management.

© 2020 Elsevier B.V. All rights reserved.

## 1. Introduction

Soil water influences almost every aspect of ecosystem behaviour, from water quality to soil fertility, from organic matter decomposition to plant growth (Weil and Brady, 2016). Accurate information about soil hydraulic properties is fundamental for the assessment of soil water dynamics and their interactions with ecological processes as well as for sustainable water and land management. Due to the high spatial variability in soil hydraulic properties even at the scale of several meters (Nielsen et al., 1973; Russo and Bresler, 1981), direct measurements via soil sampling are difficult, time-consuming, and therefore expensive, particularly across large extents.

A common approach for the prediction of soil hydraulic properties is via pedotransfer functions (Bouma, 1989; Wösten et al., 2001), which use relatively easier-to-measure soil properties as predictors such as soil bulk density, particle size distribution and organic matter (Gupta and Larson, 1979; Arya and Paris, 1981; Saxton et al., 1986; Schaap

et al., 1998; Reynolds et al., 2000; Merdun et al., 2006; McNeill et al., 2018). One issue, however, is that pedotransfer functions developed from a specific soil database may produce poor results when the function is applied to other soils which the database cannot represent or without measurements of predictors, i.e., they often do not generalise very well. An alternative approach proposes that vegetation dynamics can signal soil hydraulic properties (Mohanty and Skaggs, 2001; Araya et al., 2016). But such vegetation signals are hard to detect due to confounding effects of various environmental constraints (e.g., soil nutrient, natural and anthropogenic disturbances) and species-specific responses of plants to soil water dynamics (Martínez and Gilabert, 2009; Verbesselt et al., 2010). Due to the shortcomings of these approaches, a more reliable approach independent of other soil properties and/or using readily available data resources is required for prediction of soil hydraulic properties at a resolution that is appropriate for soil management.

Soil hydraulic properties are closely associated with soil physical properties such as particle size distribution (i.e., texture), which in turn are influenced by five factors controlling soil formation (Jenny, 1941); namely, parent materials, climate, biota, topography and age

\* Corresponding author.

E-mail address: [brendan.malone@csiro.au](mailto:brendan.malone@csiro.au) (B.P. Malone).

(Weil and Brady, 2016). These factors influence soil properties in concert. Among these factors, climate and topography are fundamental controls of soil hydraulic properties. Covariates representing these two factors are plentiful and most easily obtainable (thanks to the climate station networks across the globe and the advancement of climate models and satellite-based observations). Other variables such as gamma radiometrics – used as a proxy variable for parent material and to some extent, age – is also commonly used when it is of good quality, in terms of information content, as is generally the case across Australia (Minty et al., 2009). These data together with associated derivatives have been used in digital soil mapping studies within Australia, and specifically for the mapping of soil physical and hydraulic properties too. Key examples include Viscarra Rossel et al. (2015) as the basis for the Soil and Landscape Grid of Australia (Grundy et al., 2015) where multiple soil variables including AWC were mapped across Australia, consistent with the Global Soil Map specifications (Arrouays et al., 2014) in terms of spatial and depth support and the quantification of uncertainties. Padarian et al. (2014) used a suite of variables including topography, climate and gamma radiometric data to estimate and map the spatial variability of crop lower limit (CLL), DUL and AWC across Australia's agricultural zone. The spatial and depth support of these products were consistent with those of Viscarra Rossel et al. (2015).

While keeping these existing studies of the digital mapping of AWC in mind, a lot of our focus in this study is the nature of the climatic information used in digital soil mapping (DSM) where we believe more nuanced information regarding this soil forming factor may confer improved prediction of soil hydraulic properties. Climate covariates have been used in continental scale spatial modelling and DSM (Viscarra Rossel et al., 2015; Brevik et al., 2016) but only few key climatic variables such as annual mean precipitation and temperature,

mean annual maximum and minimum temperatures and solar radiation have been considered. As climate shows strong intra- and inter-annual variability, the role of climate would be largely underestimated by climate variables (e.g., mean annual temperature or precipitation) which cannot holistically reflect such variability.

Given the direct effect of climate as well as its indirect effect via regulating biota (e.g., the distribution of biomes) and weathering of parental materials, the proposition of this study is that models explicitly accounting for inter-annual and seasonal variability of climate may enhance the predictability of soil hydraulic properties. This hypothesis is borne from the generalisation that the Australian land surface is very flat and geologically very old and has been subject to intensive and prolonged weathering. Consequently, Australian soils are likely to be less affected by other confounding factors such as modern geologic activities-induced transportation and translocation (Johnson, 2009). By focusing on only agricultural regions, taking into account the potential influence of intra- and inter-annual variability of climate might largely improve the prediction.

Using a comprehensive data set of measured soil hydraulic properties in 1127 whole-soil profiles across Australian agricultural soils, we tested our hypothesis on the role of climate variability in improving spatial prediction of soil hydraulic properties. Combining this data set with 19 biologically meaningful climatic variables (i.e., bioclimatic variables, Table 1) calculated based on historical climate records as well as topographic attributes (Table 1), we trained a machine learning model to: 1) assess the relative importance of climate and topography, 2) identify the most influential variables assessed, and 3) map the four hydraulic properties in the whole soil profile down to 2 m across the Australian cropping region at 90 m grid cell resolution. From these outcomes we used a model averaging approach to combine

**Table 1**  
Bioclimatic and topographic variables used in the modelling.

Theme	Variable	Definition	Source
Temperature	T1	Annual mean temperature	SILO
	T2	Mean diurnal range calculated as the mean of monthly difference of max and min temperature	
	T3	Isothermality calculated as $100 \times T2/T7$	
	T4	Temperature seasonality calculated as standard deviation $\times 100$	
	T5	Max temperature of warmest month	
	T6	Min temperature of coldest month	
	T7	Temperature annual range calculated as the difference between T5 and T6	
	T8	Mean temperature of wettest season	
	T9	Mean temperature of driest season	
	T10	Mean temperature of warmest season (i.e., summer)	
	T11	Mean temperature of coldest season (i.e., winter)	
Precipitation	P1	Annual precipitation	Gallant and Austin (2015)
	P2	Precipitation of wettest month	
	P3	Precipitation of driest month	
	P4	Precipitation seasonality (coefficient of variation)	
	P5	Precipitation of wettest season	
	P6	Precipitation of driest season	
	P7	Precipitation of warmest season (i.e., summer)	
	P8	Precipitation of coldest season (i.e., winter)	
Topography	Aspect	The direction in which a land surface slope faces expressed in degrees from north	Gallant and Austin (2015)
	fm	300 m focal median of percent slope	
	fr1000	1000 m elevation range	
	fr300	300 m elevation range	
	mrRTF	Multi-resolution ridge top flatness	
	mrVBF	Multi-resolution valley bottom flatness index	
	PI	The Prescott Index	
	planC	Plan curvature	
	profileC	Profile curvature	
	SD	Mean monthly solar radiation	
	SP	Percent slope	
SR	Slope relief		
TPlc	Topographic position index, a topographic position classification identifying upper, middle and lower parts of the landscape		
TPlm	A mask of TPlc identifies where topographic position cannot be reliably derived in low relief areas.		
TWI	Topographic wetness index calculated as $\ln(\text{specific catchment area/slope})$ and estimates the relative wetness within a catchment		

our own predictions with AWC outputs created in the [Viscarra Rossel et al. \(2015\)](#) and [Padarian et al. \(2014\)](#) to generate what would be considered the best possible estimate of this variable given to date.

## 2. Materials and methods

### 2.1. Soil profiles and harmonization

This study used 1127 whole-soil profile data from across Australian agricultural regions in the APSOIL database (<http://www.apsim.info/Products/APSoil.aspx>). This database covers a broad range of climate zones and terrain/landscape attributes ([Fig. 1](#)) and includes measurements of three volumetric hydraulic properties ( $\text{m}^3$  water per  $\text{m}^3$  soil). They are wilting point, or lower limit (LL15), field capacity, or drained upper limit (DUL), and saturated water content (SAT). We also calculated soil available water capacity (AWC) as the difference between DUL and LL15. These soil profiles are all located on agricultural land and have been accumulated since 1980s until now under enormous efforts of field work. As the soil hydraulic properties in the APSOIL database were not measured systematically across the same depth intervals, we harmonized them to six standard depths (i.e., 0–5 cm, 5–15 cm, 15–30 cm, 30–60 cm, 60–100 cm and 100–200 cm) using mass-preserving splines ([Bishop et al., 1999](#); [Malone et al., 2009](#)) specified by the GlobalSoilMap project ([Arrouays et al., 2014](#)). As AWC is more commonly expressed to farmers in terms of mm of water (gravimetrically) rather than volumetrically, we reported this variable in mm units, which is estimated by multiplying the soil layer thickness by the soil water volume.

### 2.2. Bioclimatic and terrain covariates

We obtained daily weather data (including rainfall, maximum and minimum temperature) in the period 1957–2017 from 8022 SILO climate stations across Australia (<https://legacy.longpaddock.qld.gov.au/silo/>). Based on the climate data, we calculated 19 biologically meaningful variables (i.e., bioclimatic variables) representing the intra- and inter-annual climate variability ([Table 1](#)). Eleven of the 19 bioclimatic variables are temperature-related, and eight are precipitation-related. Then, we mapped these bioclimatic variables by interpolating the calculations across Australia at the resolution of 3 arc-second (~90 m) using a hybrid regression kriging model that entailed generalised additive modelling (GAM; [Hastie and Tibshirani, 1990](#)) and residual kriging. For the GAM component, each bioclimatic variable was modelled as a

function of terrain and topo-climatic variables derived from the Australian 3 arc-second digital elevation model (Geoscience Australia, 2010). Specifically, these predictor variables included: elevation, slope, topographic wetness index and aspect. Through a series of cross-validation steps for each bioclimatic variable we determined the optimal level of wiggleness ([Wood, 2011](#)) to attribute to the smoothing spline terms for each predictor variable in the model. In this cross-validation we also assessed whether the relationship could simply be determined by simple linear regression together with the smoothing spline functions to come to the simplest or most parsimonious model. Once fitted, the models were used to predict each bioclimatic variable across the mapping extent given the grids of the predictor variables as inputs. After the GAM fitting, model residuals were assessed for auto-correlation via globally fitted variograms (here we considered either exponential, spherical or Matern model parametrisations). The fitted variograms were then used as input to kriging model residuals onto the same grid configuration as the GAM predicted bioclimatic variables. The final step in this process was to add the GAM predictions with the kriged residuals to derive a final map of each bioclimatic variable.

The mapping of the bioclimatic variables was completely performed using R ([R Core Team, 2018](#)) with the following packages: *gam* ([Hastie, 2018](#)) and *mgcv* ([Wood, 2011](#)) for the GAMs and *gstat* ([Pebesma, 2004](#)) for the variogram fitting and kriging. The raster ([Hijmans, 2019](#)) and *rdgal* ([Bivand et al., 2019](#)) packages were used for preparation and handling of raster data layers and in the spatial application of both GAM and residual kriging models. Finally, we extracted the 19 bioclimatic variables for the 1127 soil profiles from the maps.

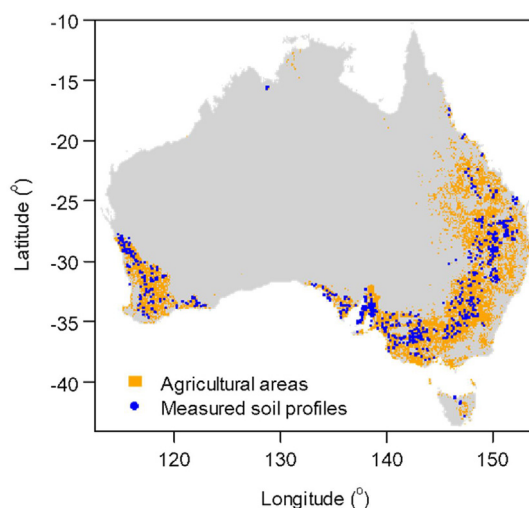
For each of the soil profiles, we also extracted 15 topographic attributes ([Table 1](#)) from maps estimated by the Shuttle Radar Topographic Mission (SRTM) digital elevation model (DEM) at the resolution of ~3 arc-second ([Gallant and Austin, 2015](#)). The details of these topographic attributes and maps were described by [Gallant and Austin \(2015\)](#).

### 2.3. Machine learning

To identify the most important bioclimatic and topographic variables, we trained a machine learning-based statistical model - random forest (RF) - to predict LL15, DUL, SAT, and AWC at each of the six standard depths. Fundamentally, the RF model ([Breiman, 2001](#)) predicts a typical soil hydraulic property (*HP*) in each depth as a function of bioclimatic and topographic covariates:

$$HP_k = f(\text{topography, climate}) + \varepsilon, \quad (1)$$

where  $HP_k$  is the target variable of either LL15, DUL, SAT, or AWC and  $\varepsilon$  is model residuals. In brief, RF predictions are the average of predictions of individual decision trees determined via bootstrap resampling. Before training the RF model, we selected a subset of variables optimally suitable for predicting *HP* using a variable selection (which is also called feature selection) algorithm. Variable selection is an important component for finding relevant variables since it achieves parsimony (which in turn reduces the computational effort) but potentially also improves the accuracy of the machine learning algorithms ([Nilsson et al., 2007](#)). We used the Boruta algorithm ([Kursa and Rudnicki, 2010](#)), which uses a wrapper approach built around the random forest algorithm ([Breiman, 2001](#)) to eliminate irrelevant variables. We executed Boruta selection runs for each soil depth and each soil hydraulic property independently using the package *Boruta* in R. The RF model was evaluated using a 10-fold cross-validation repeated 10 times in R 3.3.1 ([R Core Team, 2018](#)) using the algorithms implemented in the R package *ranger* ([Wright and Ziegler, 2017](#)). The best model tuning parameters were targeted by running the model under a series of parameter combinations. The model performance was assessed by the root mean squared error (RMSE) which quantifies prediction accuracy, the coefficient of determination ( $R^2$ ) which quantify the amount of variance explained in the relationship between observed target variables and their



**Fig. 1.** Study region. Location of soil profiles and Australian agricultural regions (including cropping and pasture areas).

corresponding modelled predictions. The final model was selected with the minimum RMSE.

To derive the controls and the relative importance of individual predictors, we calculated the overall variable importance for each soil hydraulic characteristic in each soil depth, with the *varImp* function in the *caret* library of the R software (Kuhn et al., 2019). The function computes a linear combination of all predictor variables used in the RF algorithms and reports the percent variable importance. Focusing on the identified most important predictor variable, we also conducted a linear regression analysis to detect its first-order relationship with the corresponding soil hydraulic characteristic.

#### 2.4. Mapping soil hydraulic properties and prediction uncertainty

Using the random forest model, we mapped the four soil hydraulic properties in the six standard layers across the Australian agricultural regions at the resolution of 3 arc-second (~90 m) grid using the compiled terrain and bioclimatic raster data sources. Prediction uncertainty was quantified using a Monte Carlo approach proposed for digital soil mapping with machine learning by Viscarra Rossel et al. (2014a, 2014b, 2015) and Coulston et al. (2016). First, 200 bootstrap samples of soil profiles were randomly drawn with replacement to construct 200 random forest models. For each bootstrap resample, 90% of the data (randomly selected without replacement) was used to train a random forest model with 200 individual trees, while the leftover 10% was held-out for model testing. For each derived random forest model, then, a scaling factor  $\tau$  for each observation in the hold-out testing data was estimated to approximate prediction uncertainty (Coulston et al., 2016):

$$\tau = \sqrt{\frac{(y - \hat{y})^2}{\text{var}(\hat{y})}} \quad (2)$$

where  $y$  is the observed value,  $\text{var}(\hat{y})$  the variance of the predictions of individual trees (i.e., 200 trees) in the random forest model,  $\hat{y}$  the average of  $\hat{y}$ , i.e., the prediction of the random forest model. As such, we obtained a vector of  $\tau$  with a length of the number of observations in the hold-out data multiplied by bootstrapping times 200, which represents the distribution of model prediction errors. Finally, a new random forest model based on all available data was trained for mapping. In each grid, soil hydraulic property  $HP$  was estimated as:

$$HP = \hat{y} \pm \hat{\tau} \cdot sd(\hat{y}) \quad (3)$$

where  $\hat{y}$  is the prediction of the new random forest model,  $sd(\hat{y})$  the standard deviation of the predictions of individual trees in the new random forest model,  $\hat{\tau}$  the scaling factor for desired prediction interval width estimated based on the derived distribution of model prediction errors. Using a Monte Carlo approach, we took the 95% percentile of the distribution as  $\hat{\tau}$  to quantify the 95% prediction interval (Coulston et al., 2016). In each grid, uncertainty was standardized as the half range of the 95% prediction interval divided by the mean (i.e.,  $\hat{y}$ ).

#### 2.5. Model averaging

Acknowledging prior efforts that have already mapped AWC either across Australia (Viscarra Rossel et al., 2015) or Australia's agricultural region only (Padarian et al., 2014), we sought to integrate these with our own efforts into a model averaging framework. The main idea behind model averaging is to deliver an outcome, in our case, maps of AWC, that are either better than or equivalent to, (in terms of accuracy) the best contributing map (Malone et al., 2014). Diks and Vrugt (2010) provided extensive discussion and examples of different model averaging frameworks. Each has a fundamental basis in that a combined outcome is the result of competing contributor outcomes, each weighted

according to some measure of accuracy or prediction certainty. Note that in some cases, preferential weighting is not given and instead, each contributing model may be given equal weighting. However, probably the simplest known preferential weighting method is the variance weighted model average approach (Bates and Granger, 1969). In the work by Padarian et al. (2014) they experimented with equal weight averaging and Granger-Ramanathan averaging (GRA; Granger and Ramanathan, 1984). GRA is a special case where instead of all the contributing model weights summing to 1, the weights correspond to ordinary least square (OLS) estimates of a multiple linear regression between a target variable and predictor variables. In our case the target variable is spline processed AWC values, and predictor variables are the predicted outcomes of each of the contributing models at each corresponding point. Given the relative simplicity of the GRA approach and its relatively good performance in past studies (Diks and Vrugt, 2010; Malone et al., 2014), we use it in this study also where:

$$Y = W_0 + (AWC_{mal} \cdot W_{mal}) + (AWC_{pad} \cdot W_{pad}) + (AWC_{slga} \cdot W_{slga}) \quad (4)$$

Here  $Y$  is observed AWC data (for a given depth layer) and  $AWC_{mal}$ ,  $AWC_{pad}$ , and  $AWC_{slga}$  are corresponding modelled predictions from this present study, Padarian et al., 2014, and Viscarra Rossel et al. (2015) respectively. Eq. (4) is used to solve for the parameters:  $W_0$ ,  $W_{mal}$ ,  $W_{pad}$ , and  $W_{slga}$ .  $W_0$ , the intercept term is an 'in-built' bias correction term between the observed values and the individual model predictions. Once this model is fitted, given the availability of the Viscarra Rossel et al. (2015) data from the CSIRO Data Access Portal (<https://data.csiro.au/dap/search?q=TERN+Soil>) and from Padarian et al. (2014) via personal communication, it is used as a 'global' model to derive the combined digital soil map of AWC for each of the prescribed soil depth intervals. Note that as the digital mapping of all three products were more-or-less identical in terms of spatial and depth support, minimal processing of the raster data was required to fulfill the last step of applying the GRA model spatially. The spatial extent of the mapping for the different products were somewhat different due to methodological differences in terms of delineating Australia's agricultural areas. While this is a non-issue for the Viscarra Rossel et al. (2015) where the entire continent is mapped completely, the Padarian et al. (2014) mapping encompasses agricultural land and nearby surrounding areas that may not be classified as Agricultural land according to ABARES (2011), which is what we used in the present study to delineate the mapping extent. Therefore, the spatial application of the GRA model is extended to the areas where there were modelled predictions for each of the three contributing sources only.

Given differences in terms of data used for model calibration and cross-validation between each of the three studies, all the available data ( $n = 1127$ ) was used for the GRA model. However, repeated (5000 iterations) cross-validation was used to assess the sensitivity of the model parameters to different dataset configurations. We found there to be tight distributions around the means of each parameter, and consequently used these for the spatial prediction component. The predicted values from both the Padarian et al. (2014) and Viscarra Rossel et al. (2015) maps were acquired at each of the point locations using the *raster::extract* (nearest neighbour) function (Hijmans, 2019) in R.

### 3. Results and discussion

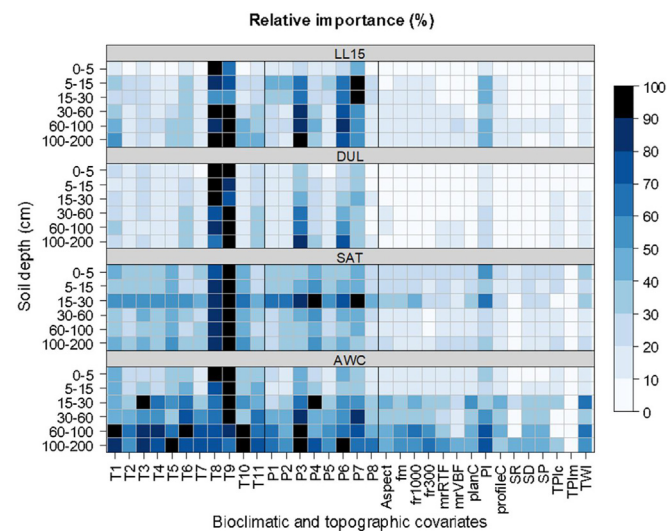
Cross-validation results suggest that the random forest model driven by 19 bioclimatic variables and 15 terrain attributes returned  $R^2$  outcomes (modelled predictions vs. observed data) of between 20 and 69% for the four soil hydraulic properties depending on soil depth interval considered (Table 2). The modelling better explains LL15 and DUL, compared with SAT and AWC (Table 2). It is generally found that soil hydraulic characteristics in deeper soil layers are more difficult to explain as evidenced by the decreased  $R^2$  with soil depth (Table 2). In particular,

**Table 2**  
Cross-validation statistics of model performance. The statistics are the coefficient of determination ( $R^2$ ) and the root mean squared error (RMSE).

Variable	Soil depth (cm)	$R^2$	RMSE
LL15 ( $m^3/m^3$ )	0–5	0.681	0.0471
	5–15	0.666	0.0473
	15–30	0.615	0.0511
	30–60	0.549	0.0551
	60–100	0.498	0.0564
	100–200	0.512	0.0561
DUL ( $m^3/m^3$ )	0–5	0.688	0.0720
	5–15	0.676	0.0713
	15–30	0.629	0.0738
	30–60	0.620	0.0706
	60–100	0.574	0.0691
	100–200	0.547	0.0676
SAT ( $m^3/m^3$ )	0–5	0.391	0.0587
	5–15	0.436	0.0559
	15–30	0.452	0.0564
	30–60	0.453	0.0533
	60–100	0.403	0.0524
	100–200	0.381	0.0517
AWC (mm)	0–5	0.410	2.66
	5–15	0.440	4.86
	15–30	0.366	7.16
	30–60	0.327	13.64
	60–100	0.284	19.27
	100–200	0.209	51.21

$R^2$  decreased from ~0.41 for AWC in the top 0–5 cm soil layer to ~0.20 in the 100–200 cm soil layer (Table 2). For model accuracy (i.e., RMSE), those results follow the similar pattern to the  $R^2$  results (Table 2).

The relative importance analyses revealed that bioclimatic variables (which contribute >90% to the explained variances) are much more important than topographic variables (Fig. 2). For all four soil hydraulic properties in all soil layers, the most important variable was climate-related. However, neither T1 (i.e., annual mean temperature) nor P1 (annual mean precipitation) were the most important climate variable. Rather, T8 (mean temperature of wettest quarter), T9 (mean temperature of driest quarter) and/or P7 (precipitation of warmest quarter) were the three most influential variables depending on soil hydraulic characteristic and soil depth (Fig. 2). Fig. 3 shows the linear relationship between soil hydraulic characteristics in each layer with their corresponding most influential predictor variables. The results indicated



**Fig. 2.** Relative importance of bioclimatic and topographic covariates. The relative importance of bioclimatic and topographic variables in predicting LL15, DUL, SAT and AWC in different soil layers. All values are normalized to the most important variable.

that LL15, for example, in the top 0–5 cm soil is positively correlated with T8, which alone can explain 50% ( $R^2 = 0.5$ ) of the variance (Fig. 3).

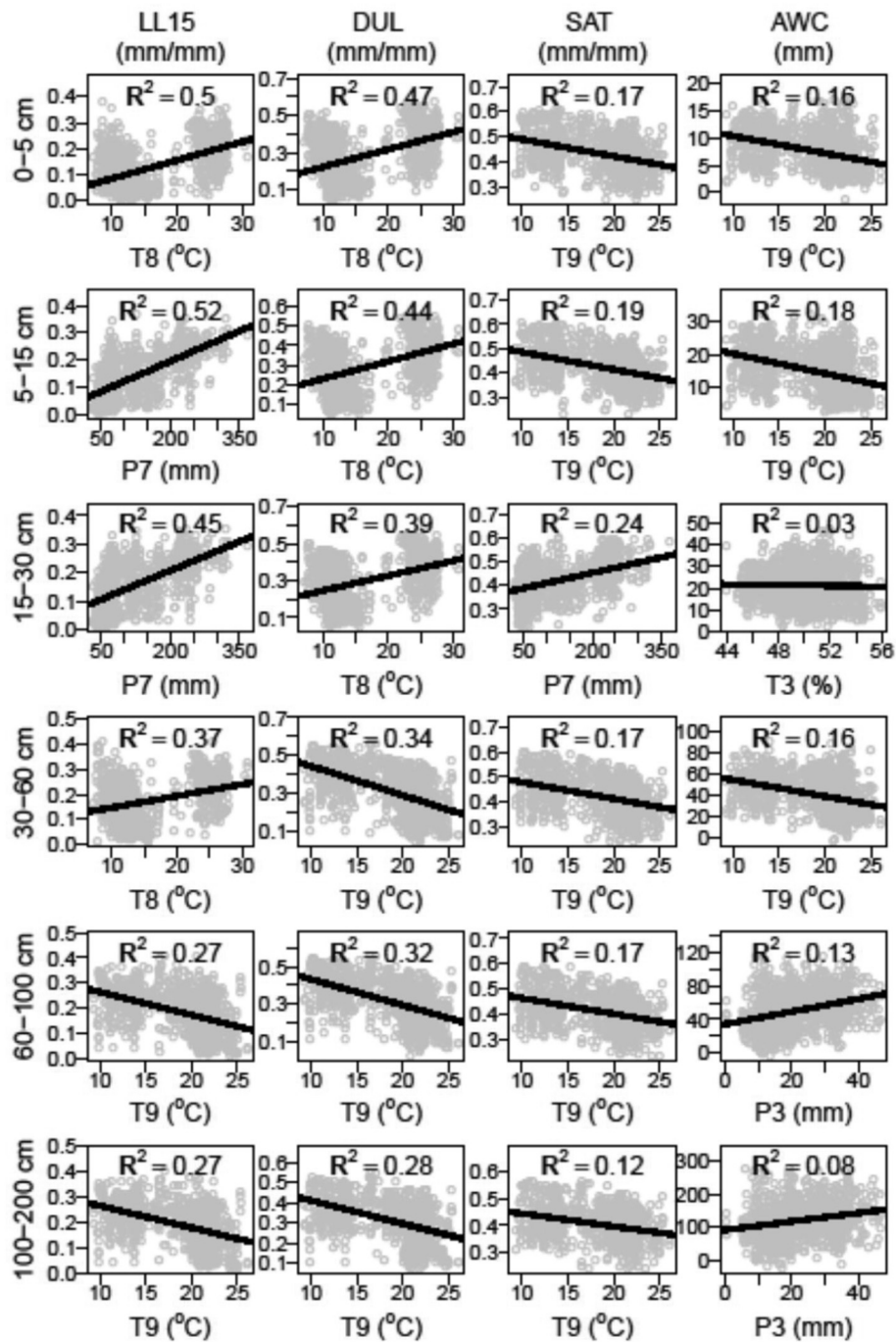
Using the derived random forest model and taking into account the uncertainty in predictions, we mapped the four soil hydraulic characteristics for the six soil layers at the resolution of 3 arc-seconds (~90 m) across the whole Australian agricultural region. Across the region, the LL15 in the top 0–5 cm soil layer ranges from  $0.0136 m^3/m^3$  to  $0.296 m^3/m^3$ , showing a clear increasing pattern from west to east and from south to north (Fig. 4). These maps also show the uncertainty of the predicted four hydraulic properties which is calculated as the half range of 95% prediction interval divided by the mean. The spatially explicit analysis provides a unique inventory to derive soil hydraulic characteristics adequately along the soil profile. Note for better visualisation of the maps, we show only the outputs corresponding to the 0–5 cm only (Fig. 4). The supplementary figures sections contain maps for all variables at each depth with their associated uncertainty estimates.

In general, LL15 increases with soil depth. In the 100–200 cm soil layer, for example, the range of LL15 is increased to from  $0.0349 m^3/m^3$  to  $0.394 m^3/m^3$ . The DUL and SAT across the region as well as in different layers generally shown the similar pattern as LL15. For the AWC, the spatial distribution shows that the AWC is higher in eastern regions (e.g., New South Wales) than in western regions (e.g., Western Australia).

While Table 2 summarises the model diagnostics based on cross-validation outcomes, Table 3 however is based on all available data and therefore the  $R^2$  and RMSE appear to show improved model performance of the processes conducted in this study. The intention here however is to provide a side-by-side comparison between our estimates with those from Viscarra Rossel et al. (2015) and Padarian et al. (2014) and the combined model from GRA, and this is only really possible by using all of the available data given differences between studies in terms of data used in model calibration and cross-validations.

The results from this study are comparable with those from Viscarra Rossel et al. (2015) where a full suite of predictive covariates was used. The Padarian et al. (2014) results show for all depths less accuracy than what was achieved in this study and from Viscarra Rossel et al. (2015). The simple explanation for this is possibly differences in the data used for the modelling, where firstly the Padarian et al. (2014) effort, 806 soil profiles were used from the APSOIL database, whereas in the present study 1127 were used. Secondly, in the present study we used LL15 for the dry end of the soil hydraulic property whereas in Padarian et al. (2014) the CLL was used, and by definition were estimating plant AWC (PAWC) rather than AWC. Therefore, estimates of AWC are going to differ as CLL is crop dependent and measured in the field setting, while LL15 is a physical measure of the soil and measured in the laboratory. While this may be true for most cases in the APSOIL database, we have since learned (and noted for future investigations) that for many characterisations, LL15 is an estimated parameter (derived from CLL) rather than measured (pers comm. K. Verburg), and it is unknowable to determine which characterisation are measured or estimated due to issues with data entry into the database. Strictly speaking the equitable approach to combine outputs from the present study with those from Padarian et al. (2014) would be to focus on the DUL variable, but for practical reasons, AWC is a much more useful variable to investigate. The third reason is that in both the present study and Viscarra Rossel et al. (2015), AWC was modelled directly whereas in Padarian et al. (2014) it was derived as the difference between DUL and CLL. Together these similarities and differences equate to the present study predictions being comparable to those from Viscarra Rossel et al. (2015) whom used the same data, despite the present study using a lesser range of predictive covariates, but more nuanced in terms of climatic and temperature data characterisations.

Combining the three models using GRA model averaging yielded improved prediction accuracy for all depths which was to be expected and signifies the value in considering other covariates not used in this study



**Fig. 3.** The most important variable among all assessed variables. The effect of the most important variable identified by random forest model on four soil hydraulic characteristics in six soil layers. From left to right, LL15, DUL, SAT and AWC. From top to bottom, 0–5, 5–15, 15–30, 30–60, 60–100, and 100–200 cm soil layers.

such as gamma radiometrics and remote sensing data. This highlights the value of the GRA model as we don't necessarily need to explicitly incorporate these variables into the modelling procedure, but rather incorporate them indirectly via contributing models that use them through the model combination process. Regarding the relative contribution of each model to the GRA models for each depth, Table 4 summarises the OLS estimates of the weighting attributed to each contributing model. Our interpretation of these parameter values is to compare them relative to one another which after bias correction given by the  $W_0$  parameter, both the estimates for the present study and Viscarra Rossel et al. (2015) are quite similar, meaning similar contribution to the GRA estimates. On the other hand, the contribution of the Padarian

et al. (2014) predictions appears relatively smaller, the reasons for which we believe to be like the ones described above.

Fig. 5 shows the GRA model estimated AWC for the 0–5 cm depth interval. This figure also shows focused mapping centred on southern Australia's wheat-sheep region to illustrate similarities and differences between each of the contributing maps and the combined GRA model estimates. Just from these focused maps, the AWC estimates of prior studies have substantial areas of lower AWC compared with estimates made in the present study. As can be seen in the GRA model mapping, there is a subtle adjustment of our modelled AWC to reflect the predictions made in the prior studies. The contribution of the Padarian et al. (2014) estimates appear to be minimal compared with those from Viscarra Rossel

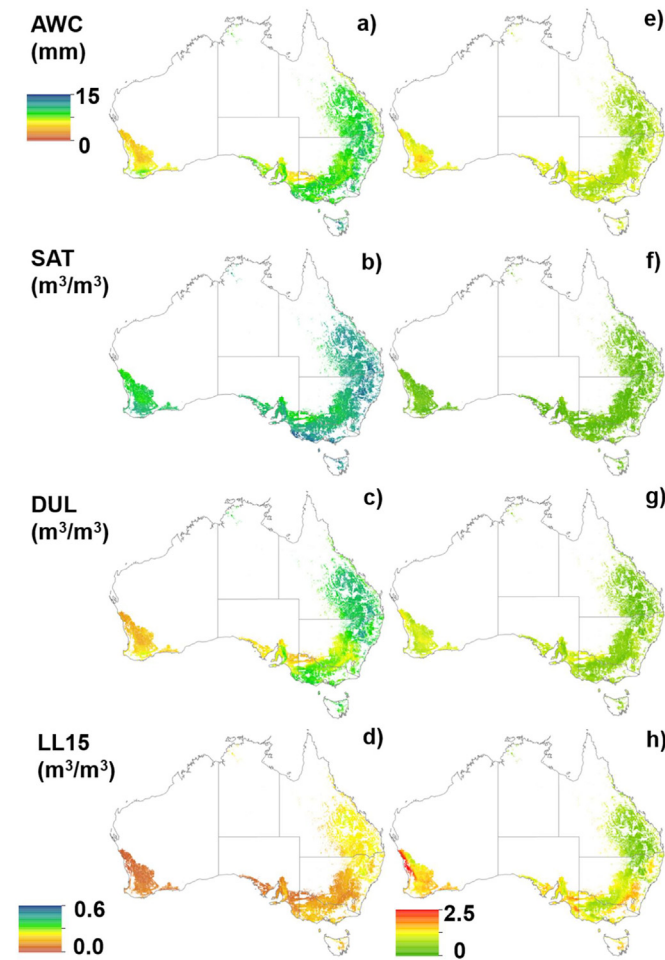


Fig. 4. Maps of predicted AWC, SAT, DUL and LL15 and associated prediction uncertainties for the 0–5 cm depth interval. Similar maps are provided for the other depth intervals in the supplementary figure section.

et al. (2015) which is to be expected given the parameter estimates established by the GRA modelling. Supplementary figures show the GRA modelled estimates for all the depths investigated in this study.

#### 4. Discussion

Temperature of driest (T9) and wettest quarter (T8) and precipitation of warmest month (P7) appear to be the most influential bioclimatic variables on average, albeit the most influential bioclimatic

Table 3

Coefficient of determination ( $R^2$ ) and the root mean squared error (RMSE) between AWC measured at the 1127 sites and associated predictions at these sites for each depth interval from the GRA model averaging estimates, and the contributing estimates from the present study, Padarian et al. (2014) and Viscarra Rossel et al. (2015).

Soil depth interval	GRA model estimate		Our estimate		Padarian et al. (2014)		Viscarra Rossel et al. (2015)	
	R2	RMSE	R2	RMSE	R2	RMSE	R2	RMSE
0–5 cm	0.61	2.31	0.54	2.38	0.41	2.69	0.42	2.66
5–15 cm	0.64	3.89	0.54	4.45	0.44	4.91	0.49	4.69
15–30 cm	0.65	5.42	0.51	6.33	0.41	6.96	0.53	6.20
30–60 cm	0.59	10.74	0.49	11.97	0.38	13.17	0.53	11.53
60–100 cm	0.55	15.31	0.44	17.22	0.30	19.32	0.47	16.80
100–200 cm	0.57	36.89	0.42	41.68	0.21	48.67	0.40	42.45

Table 4

GRA model parameter estimates attributed to intercept term ( $W_0$ ) and each of the contributing models.

Soil depth interval	$W_0$	Our estimate	Padarian et al. (2014)	Viscarra Rossel et al. (2015)
0–5 cm	–3.00	0.96	–0.09	0.57
5–15 cm	–5.55	0.67	0.02	0.75
15–30 cm	–9.73	0.73	–0.05	0.80
30–60 cm	–16.39	0.71	–0.02	0.71
60–100 cm	–27.84	0.89	–0.08	0.73
100–200 cm	–109.67	1.09	–0.04	0.80

variable is not consistent for different soil hydraulic properties in different soil layers. It is intriguing to note that these three variables reflect some interdependence of temperature and moisture, reflecting the importance of soil weathering including physical and chemical weathering for determining soil hydraulic characteristics. This may be due to the fact that weathering is the most important process regulating soil formation and soil physical structure such as particle size distribution

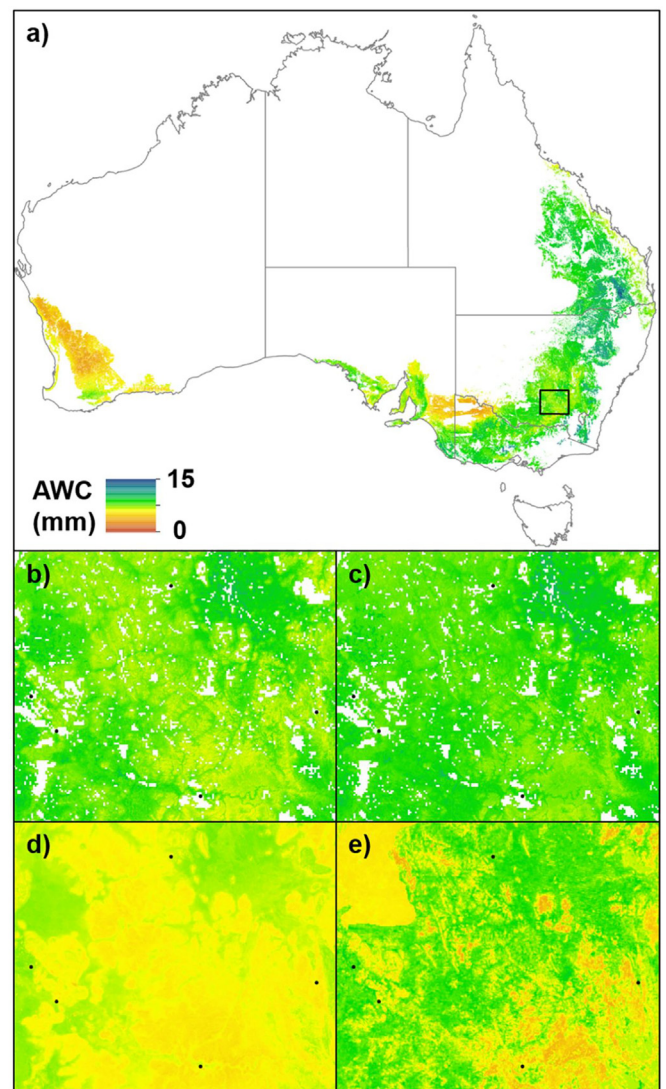


Fig. 5. GRA model estimated AWC for the 0–5 cm depth interval (a). Focussed region corresponds to a region of the southern Australian wheat-sheep belt. (b) GRA model estimates of AWC (c) AWC estimated by our own modelling using bioclimatic and topographic variables, (d) AWC estimated by Viscarra Rossel et al. (2015), (e) AWC estimated by Padarian et al. (2014).

(Weil and Brady, 2016). Physical weathering is usually accentuated in very cold and very dry environments, which is correctly captured by the importance of T9. T9 reflects the fact that temperature may experience great variability (e.g., apparent and intensive heat up and cool down cycle), causing alternate expansion and contraction of soil. As some minerals expand more than others, temperature changes set up differential stresses that eventually cause the rock to crack apart (i.e., soil formation and soil physical structure). Furthermore, as expected, T9 (temperature of driest quarter) shown opposite effects on the assessed soil hydraulic characteristics compared with the effect of T8 (temperature of wettest quarter). For chemical weathering, it is mainly chemical reactions which are most intense under wet and hot conditions, which is also correctly captured by the importance of P7 (i.e., precipitation of warmest month).

It is a common recognition that soil hydraulic characteristics are closely correlated to soil particle size distribution (Gupta and Larson, 1979; Arya and Paris, 1981; Saxton et al., 1986) – a partial reflection of soil physical structure. Indeed, most pedotransfer functions (including emerging machine learning-based approaches) used to estimate soil hydraulic characteristics are primarily driven by soil particle size (Schaap et al., 1998; Reynolds et al., 2000; Merdun et al., 2006; McNeill et al., 2018). However, the predictive power of soil particle size alone is very limited. For example, a study assessing available water capacity (AWC) across the whole of Australia using similar machine learning-based approach and a large proportion of the same dataset for the agricultural regions (the APRSU set) found that only <30% variance in AWC can be explained by a series of environmental covariates including soil mineralogy (Viscarra Rossel et al., 2015). While, in our study, >50% variance in AWC can be explained by bioclimatic variables alone (Fig. 2). We assume that the reason for the greater explained variance in our modelling here might be due to the smaller scale and the focus on only agricultural regions of Australia, where water is not likely to be as limited as elsewhere in Australia. Most application of pedotransfer functions across large scales depends on other soil mapping products (Miller and White, 1998; Reynolds et al., 2000; McBratney et al., 2003; Grunwald, 2009). The uncertainties in those mapping products are inevitably propagated into the estimation of soil hydraulic characteristics. Based on our findings, here, we suggest that bioclimatic variables are useful predictors of soil hydraulic characteristics at the large spatial scale investigated than soil particle size distribution as they reflect the dominant control of soil weathering on a series of soil physical properties therefore on soil hydraulic characteristics.

Although bioclimatic variables are important for all soil hydraulic characteristics in all assessed soil depths, it should be noted that their effects become weaker in deeper soil layers. This may be attributed to the general observation that deep soil is relatively younger than top soil, and its properties may be largely controlled by parental materials or something other than climate. In addition, the deep soil environment is inherently more stable than top soil. The effect of inter- and intra-annual climate variability is largely buffered by top soil layers. The effect of biota in deep soils is also very limited due to constrained biological activity in deep soils. As the close association between climate and biota, the effect of both climate and biota is thus limited in deep soils, resulting in the observed larger effect of bioclimatic variables on soil hydraulic characteristics in surface soils.

Our study focuses on the soil hydraulic characteristics in Australian agricultural soils. Some special features of the Australian continent may intensify the role of bioclimatic variables in regulating soil formation and thus soil hydraulic characteristics. First, the Australian continent, particularly the agricultural regions, is relatively flat compared with other continents. As such, the soil in general is less affected by topography (McKenzie et al., 2004) as evidenced by the relative unimportance of topographic attributes. Second, the Australian continent is old and less affected by modern geological activities such as orogeny and volcanic activity (Johnson, 2009). For this reason, Australian soils are to large extent stable and intact.

The results in this study revealed that bioclimatic variables alone can explain more than 50% of variance in some soil hydraulic characteristics such as LL15 and DUL in the top 30 cm soil layer. Soil hydraulic characteristics are important parameters to predict soil moisture dynamics (McColl et al., 2017) and therefore any ecological processes relating to soil moisture. Their reliable, affordable and fast estimation is a persistent challenge (Minasny et al., 1999; Cornelis et al., 2001; Dobarco et al., 2019) and vital for effective water management, particularly in agroecosystems (Saxton and Rawls, 2006; Dobriyal et al., 2012). For example, irrigated agriculture alone accounts for ~85% of global total human consumption of freshwater (Gleick, 2003). Our finding could be particularly valuable for the design of effective irrigation systems which requires high-quality information on soil hydraulic characteristics. Irrigation can be tuned to local soil hydraulic properties to alleviate water scarcity, which is important in this era of global water crisis (Hanjra and Qureshi, 2010; Cosgrove and Rijsberman, 2014). As bioclimatic variables can be directly calculated based on climate records which are readily obtainable for most regions across the globe, our findings mean these data could potentially be useful for other similar studies where there is need to reliably and cheaply predict soil hydraulic characteristics at local scale as well as across large scales. Digital soil mapping is important for the communication of the relevant information among land-users, scientists, and policy-makers, and is critical for the development of specific soil management recommendations (Sanchez et al., 2009). Our mapping and the relevant uncertainty estimation of soil hydraulic characteristics provide critical inputs to models predicting soil water-regulated ecosystem changes in response to global climatic and human disturbances and making local-specific soil management strategies.

## 5. Conclusions

In this study, we hypothesized that soil hydraulic properties could be skilfully estimated with a suite of environmental variables, especially bioclimatic variables because of their direct and indirect effects via regulating biota and the weathering of parental materials. Testing this hypothesis using data from Australian agricultural soils, we found that temperature of driest and wettest quarter, and precipitation of warmest month were strong predictor variables where overall the bioclimatic variables accounted for 19–69% of the variance in the four soil hydraulic properties determining soil holding capacity and water availability for plant growth. We further improved upon our estimates of AWC through an ensemble modelling approach which combined estimates from earlier studies. This approach is ideal as it facilitated indirect use of predictive covariates not considered in our own modelling experiments. Collectively, these new estimates of hydraulic properties provide an important inventory for Australian agricultural soils.

## Data availability statement

The data of soil profiles used in this study is at: <http://www.apsim.info/Products/APSsoil.aspx>. The bioclimate and terrain data, the code used to generate the maps, and, as a demonstrating example, the maps of AWC including its average and uncertainty in the top 5 cm soil layer are at: <http://10.6084/m9.figshare.9988706>.

## Author contributions

E.W conceived the project; Z.L performed the data analyses, and led interpretations and writing; B.M and D.H. contributed to the data analysis; B.M., R.A.V.R., and E.W. contributed to interpretations and writing.

## Declaration of Competing Interest

The authors declare no competing interests.



## Acknowledgement

This study is financially supported by the CSIRO Digiscape Future Science Platform via the project: Inverse modelling of soil properties. Z.L. is partially supported by funding from the National Natural Science Foundation of China.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.geodrs.2020.e00344>.

## References

- ABARES, 2011. Guidelines for Land Use Mapping in Australia: Principles, Procedures and Definitions, Edition 4. Australian Bureau of Agricultural and Resource Economics and Sciences, Canberra available at: [http://data.daff.gov.au/anrdl/metadatas\\_files/pe\\_abares99001806.xml](http://data.daff.gov.au/anrdl/metadatas_files/pe_abares99001806.xml).
- Araya, S., Lyle, G., Lewis, M., Ostendorf, B., 2016. Phenologic metrics derived from MODIS NDVI as indicators for plant available water-holding capacity. *Ecol. Indic.* 60, 1263–1272.
- Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., 2014. GlobalSoilMap: toward a fine-resolution global grid of soil properties. *Adv. Agron.* 93–134 (Elsevier).
- Arya, L.M., Paris, J.F., 1981. A Physicoempirical model to predict the soil moisture characteristic from particle-size distribution and bulk density data 1. *Soil Sci. Soc. Am. J.* 45, 1023–1030.
- Bates, J., Granger, C., 1969. The Combination of Forecasts. *J. Oper. Res. Soc.* 20, 451–468 <https://doi.org/10.1057/jors.1969.103>.
- Bishop, T., McBratney, A., Laslett, G., 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. *Geoderma* 91, 27–45.
- Bivand, R., Keitt, T., Rowlingson, B., 2019. rgdal: Bindings for the 'Geospatial' Data Abstraction Library. R Package Version. 1, p. 4. <https://CRAN.R-project.org/package=rgdal>.
- Bouma, J., 1989. Using soil survey data for quantitative land evaluation. In: Stewart, B.A. (Ed.), *Advances in Soil Science*. vol. 9. Springer US, New York, NY, pp. 177–213.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Brevik, E.C., Calzolari, C., Miller, B.A., Pereira, P., Kabala, C., Baumgarten, A., Jordán, A., 2016. Soil mapping, classification, and pedologic modeling: history and future directions. *Geoderma* 264, 256–274.
- Cornelis, W.M., Ronsyn, J., Van Meirvenne, M., Hartmann, R., 2001. Evaluation of pedotransfer functions for predicting the soil moisture retention curve. *Soil Sci. Soc. Am. J.* 65, 638–648.
- Cosgrove, W.J., Rijsberman, F.R., 2014. *World Water Vision: Making Water everybody's Business*. Routledge.
- Coulston, J.W., Blinn, C.E., Thomas, V.A., Wynne, R.H., 2016. Approximating prediction uncertainty for random forest regression models. *Photogramm. Eng. Remote Sens.* 82, 189–197.
- Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stoch. Env. Res. Risk A.* 24 (6), 809–820.
- Dobarco, M.R., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma* 336, 81–95.
- Dobryl, P., Qureshi, A., Badola, R., Hussain, S.A., 2012. A review of the methods available for estimating soil moisture and its implications for water resource management. *J. Hydrol.* 458, 110–117.
- Gallant, J.C., Austin, J.M., 2015. Derivation of terrain covariates for digital soil mapping in Australia. *Soil Res.* 53, 895–906.
- Geoscience Australia, 2010. Geoscience Australia, 3 second SRTM Digital Elevation Model (DEM) v01 Bioregional Assessment Source Dataset Viewed 1 April 2019. <http://data.bioregionalassessments.gov.au/dataset/12e0731d-96dd-49cc-aa21-ebfd65a3f67a>.
- Gleick, P.H., 2003. *Water Use*. *Annu. Rev. Environ. Resour.* 28, 275–314.
- Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecasts. *J. Forecast.* 3 (2), 197–204.
- Grundy, M., Rossel, R.V., Searle, R., Wilson, P., Chen, C., Gregory, L., 2015. Soil and landscape grid of Australia. *Soil Res.* 53, 835–844.
- Grunwald, S., 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma* 152, 195–207.
- Gupta, S., Larson, W., 1979. Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. *Water Resour. Res.* 15, 1633–1635.
- Hanjra, M.A., Qureshi, M.E., 2010. Global water crisis and future food security in an era of climate change. *Food Policy* 35, 365–377.
- Hastie, T., 2018. gam: Generalized Additive Models. R package version 1.16. <https://CRAN.R-project.org/package=gam>.
- Hastie, T.J., Tibshirani, R.J., 1990. *Generalized additive models*. Chapman & Hall/CRC. New York/Boca Raton.
- Hijmans, R.J., 2019. raster: Geographic Data Analysis and Modeling. R Package Version 2.9–5. <https://CRAN.R-project.org/package=raster>.
- Jenny, H., 1941. *Factors of Soil Formation, a System of Quantitative Pedology*. McGraw-Hill, New York.
- Johnson, D., 2009. *The Geology of Australia*. Cambridge University Press.
- Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, A., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., the R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2019. caret: Classification and Regression Training. R Package Version 6.0–84. <https://CRAN.R-project.org/package=caret>.
- Kursa, M.B., Rudnicki, W.R., 2010. Feature selection with the Boruta package. *J. Stat. Softw.* 36, 1–13.
- Malone, B.P., McBratney, A., Minasny, B., Laslett, G., 2009. Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* 154, 138–152.
- Malone, B.P., Minasny, B., Odgers, N.P., McBratney, A.B., 2014. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232–234, 34–44.
- Martínez, B., Gilabert, M.A., 2009. Vegetation dynamics from NDVI time series analysis using the wavelet transform. *Remote Sens. Environ.* 113, 1823–1842.
- McBratney, A.B., Santos, M.M., Minasny, B., 2003. On digital soil mapping. *Geoderma* 117, 3–52.
- McColl, K.A., Alemohammad, S.H., Akbar, R., Konings, A.G., Yueh, S., Entekhabi, D., 2017. The global distribution and dynamics of surface soil moisture. *Nat. Geosci.* 10, 100.
- McKenzie, N., Jacquier, D., Isbell, R., Brown, K., 2004. *Australian Soils and Landscapes: An Illustrated Compendium*. CSIRO Publishing.
- McNeill, S.J., Lilburne, L.R., Carrick, S., Webb, T.H., Cuthill, T., 2018. Pedotransfer functions for the soil water characteristics of New Zealand soils using S-map information. *Geoderma* 326, 96–110.
- Merdun, H., Çınar, Ö., Meral, R., Apan, M., 2006. Comparison of artificial neural network and regression pedotransfer functions for prediction of soil water retention and saturated hydraulic conductivity. *Soil Tillage Res.* 90, 108–116.
- Miller, D.A., White, R.A., 1998. A continental United States multilayer soil characteristics database for regional climate and hydrology modeling. *Earth Interact.* 2, 1–26.
- Minasny, B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. *Geoderma* 93, 225–253.
- Minty, B., Franklin, R., Milligan, P., Richardson, M., Wilford, J., 2009. The radiometric map of Australia. *Explor. Geophys.* 40 (4), 325–333.
- Mohanty, B.P., Skaggs, T., 2001. Spatio-temporal evolution and time-stable characteristics of soil moisture within remote sensing footprints with varying soil, slope, and vegetation. *Adv. Water Resour.* 24, 1051–1067.
- Nielsen, D., Biggar, J., Erh, K., 1973. Spatial variability of field-measured soil-water properties. *Hilgardia* 42, 215–259.
- Nilsson, R., Peña, J.M., Björkegren, J., Tegnér, J., 2007. Consistent feature selection for pattern recognition in polynomial time. *J. Mach. Learn. Res.* 8, 589–612.
- Padarian, J., Minasny, B., McBratney, A.B., Dalglish, N., 2014. Predicting and mapping the soil available water capacity of Australian wheatbelt. *Geoderma Reg.* 2–3, 110–118.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30, 683–691.
- R Core Team, 2018. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria URL: <https://www.R-project.org/>.
- Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the food and agriculture organization soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resour. Res.* 36, 3653–3662.
- Russo, D., Bresler, E., 1981. Soil hydraulic properties as stochastic processes: I. an analysis of field spatial variability 1. *Soil Sci. Soc. Am. J.* 45, 682–687.
- Sanchez, P.A., Ahamed, S., Carré, F., Hartemink, A.E., Hempel, J., Huisung, J., Lagacherie, P., McBratney, A.B., McKenzie, N.J., Mendonça-Santos, M.D.L., Minasny, B., Montanarella, L., Okoth, P., Palm, C.A., Sachs, J.D., Shepherd, K.D., Vágen, T.-G., Vanlauwe, B., Walsh, M.G., Winowiecki, L.A., Zhang, G.-L., 2009. Digital soil map of the world. *Science* 325, 680–681.
- Saxton, K.E., Rawls, W.J., 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Sci. Soc. Am. J.* 70, 1569–1578.
- Saxton, K., Rawls, W.J., Romberger, J., Papendick, R., 1986. Estimating generalized soil-water characteristics from texture 1. *Soil Sci. Soc. Am. J.* 50, 1031–1036.
- Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. *Soil Sci. Soc. Am. J.* 62, 847–855.
- Verbesselt, J., Hyndman, R., Newnham, G., Culvenor, D., 2010. Detecting trend and seasonal changes in satellite image time series. *Remote Sens. Environ.* 114, 106–115.
- Viscarra Rossel, R.A., Webster, R., Bui, E., Baldock, J., 2014a. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. *Glob. Chang. Biol.* 20 (9), 2953–2970.
- Viscarra Rossel, R.A., Webster, R., Kidd, D., 2014b. Mapping gamma radiation and its uncertainty from weathering products in a Tasmanian landscape with a proximal sensor and random forest kriging. *Earth Surf. Process. Landf.* 39, 735–748.
- Viscarra Rossel, R.A., Chen, C., Grundy, M., Searle, R., Clifford, D., Campbell, P., 2015. The Australian three-dimensional soil grid: Australia's contribution to the GlobalSoilMap project. *Soil Res.* 53, 845–864.
- Weil, R., Brady, N., 2016. *The Nature and Properties of Soils*. 15th ed. .
- Wood, S.N., 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J. Royal Stat. Soc. (B)* 73 (1), 3–36.
- Wösten, J., Pachepsky, Y.A., Rawls, W., 2001. Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics. *J. Hydrol.* 251, 123–150.
- Wright, M.N., Ziegler, A., 2017. ranger: a fast implementation of Random Forests for high dimensional data in C++ and R. *J. Stat. Softw.* 77.