# Predicting soil water holding capacity from climate and crop yield

Enli Wang [1], Di He [1], and Zhongkui Luo[1]

[1] CSIRO Agriculture and Food, GPO Box 1700, Canberra ACT 2601, ACT, Australia, enil.wang@csrio.au

**Abstract**
Soil water holding capacity is a key soil property affecting dryland crop yield, and is therefore important for crop management in semi-arid climates like Australia. This paper explores two approaches: one developed using process-based modelling to inversely predict plant available water capacity (PAWC) of soils from crop yield, another one built with machine learning to predict soil available water capacity (AWC) spatially based on bio-climate variables. Our results indicate that soil PAWC can be skilfully predicted with water-limited crop yield ($R^2$ of 0.84~0.98 and RMSE of 14.5mm~30.2mm across 10 sites) and that the bio-climate variables together with a machine learning approach could explain up to 50% of the variance in soil AWC across sites. These results demonstrate the potential to use climate and crop yield data to predict soil water holding capacity.

**Key Words**
Soil-plant interaction, LAI, machine learning, random forest, APSIM

## Introduction

Soil water holding capacity is a key soil property that impacts on crop yield, particularly for dryland crops in a semi-arid climate. Available water capacity (AWC) of soil for growing plants in general is the amount of water held between drained upper limit (DUL) and water content at 15 bar (LL15). Plant available water capacity (PAWC) further takes the plant type into account, is defined as the amount of water held between DUL and the crop-specific lower limit (CLL), with CLL changing also with rooting depth, root density and crop water demand. While AWC is a soil property, PAWC is a combined property of soil and plant.

Spatial variability in PAWC can be high even at paddock scale. It has been well recognized that spatial variations in PAWC cause a large part of the variability in crop yield across sites (Wang et al., 2009) and within a paddock (Lawes et al., 2009; Wong and Asseng, 2006), which in many cases warrant spatially explicit management practices. Despite its importance, accurate PAWC data at the required spatial resolution are not available due to the difficulties to do soil sampling. The recently developed Soil and Landscape Grid of Australia (SLGA) (Grundy et al., 2015) contains spatial predictions of AWC at 90 m resolution, however, there are significant areas in Australia's agricultural regions that could benefit from refined and improved predictions for management applications at paddock and sub-paddock scales.

Climate directly influences the weathering processes in soil formation, and it also impacts on vegetation/crop growth. In spite of the different timeframes, soil moisture retention properties like DUL, LL15 and AWC may spatially correlate with long-term climate variables. Under a given climate, if spatial variations in vegetation dynamics and crop yield can reflect variations in soil PAWC, they can potentially be used to inversely estimate soil PAWC. Such spatial and inverse modelling approaches have the potential to predict soil PAWC at the resolution required for spatially explicit management because climate and vegetation data are much easier to obtain at high spatial resolution. This paper explores the potential of these approaches.

## Methods

*The inverse modelling approach*

While PAWC is a relatively static soil property, crop yield varies with spatial, intra- and inter-annual variations of climate. The relationship between PAWC and crop yield is therefore climate-dependent. In order to quantify the relationship, ten sites (Ballarat, Griffith, Ardlethan, Temora, Yanco, Yong, Emerald, Miles, Narrabri and Merredin) were selected to cover different rainfall patterns with an annual rainfall range of 382-647mm (Fig 1a). We used the APSIM model to simulate wheat potential yield (no nitrogen stress) for 120 years from 1889 to 2017 on 48 soils with a PAWC range of 15-286mm at each site. These soil profile data were created using pedotransfer functions and the 6 soil texture classes in the Australian Soil Resource Information System (ASRIS). A negative exponential model was developed based on the simulation results to link average crop yield ($y$) to PAWC: $y = y_m \times (1 - e^{-k(PAWC-5)/y_m})$ for all the sites together, with the parameter $y_m$ related to mean total annual rainfall and $k$ related to mean fraction of rainfall during the wheat growing season of each site. The model explained >80% of variation in simulated wheat yields.

The derived *y-PAWC* was then used to inversely estimate PAWC using simulated wheat yield. For the purpose of PAWC prediction, we used simulated wheat yield with 41 of the 48 soils to derive the negative exponential model, and the remaining wheat yield of 7 soils to test the model. We tested the model using simulated wheat yield from a single year, average of the simulated yields from 5 and 10 consecutive years, to mimic situations where only crop yield from one year, 5 and 10 consecutive years are available.

*The spatial modelling approach*
We used soil profile measurements of AWC at 1,127 locations collated in the APSOIL database (http://www.apsim.info/Products/APSoil.aspx) and 60 years of weather data (1957 to 2017) at the nearest weather station (https://silo.longpaddock.qld.gov.au/) to each of the soil locations to analyse the relationship between soil AWC and climate variables. The soil and climate sites are spread across the main agricultural areas of Australia (Fig 1b). All soil profiles are harmonized to six depth intervals (i.e., 0–5, 5–15, 15–30, 30–60, 60–100 and 100–200 cm) using mass-preserving splines (Malone et al., 2009). Eleven temperature-based and eight rainfall-based bio-climate (biologically meaningful variables) variables were derived using the daily climate data to describe the mean,intra- and inter-annual variability of temperature and precipitation as well as their seasonality synchrony. Machine learning-based (ML) modelling with random forest approach was conducted to identify the most important climatic variables and develop a model for AWC prediction. The ML models were evaluated using a 10-fold cross-validation repeated 10 times in R 3.3.1 (R Core Team 2016) using the algorithms implemented in the R package *ranger* (Wright and Ziegler, 2017).

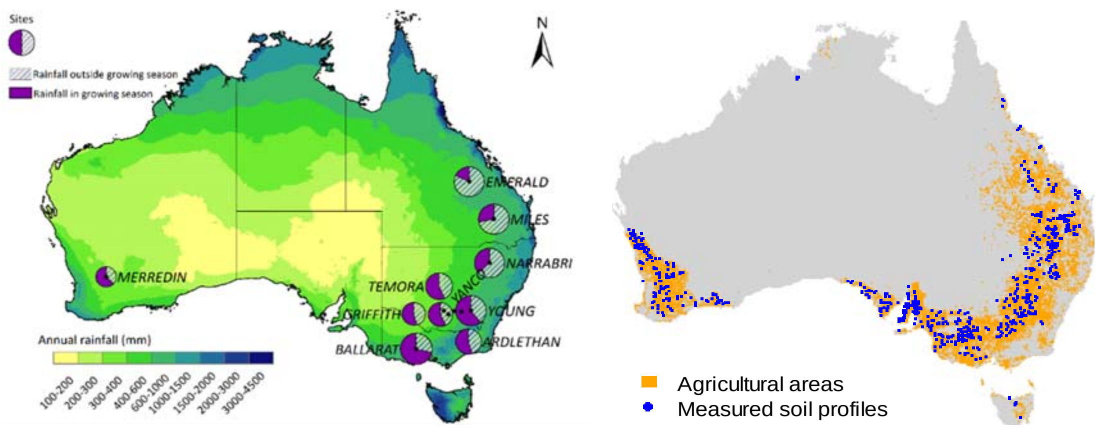


**Figure 1 Sites selected for APSIM inverse modelling of PAWC (a) and for spatial modelling of soil AWC (b).**

## Results and discussion

*The inverse modelling approach*
The negative exponential model enabled us to derive a critical PAWC value ($PAWC_C$) for each site, i.e., the PAWC above which crop yield stops increasing with PAWC. With the inverse modelling approach, useful skills with $R^2$ of 0.84~0.98 and RMSE of 14.5mm~30.2mm were achieved in the prediction of PAWC using crop yield across 10 sites when PAWC is below the $PAWC_C$ (Fig 2). Improved predictions were noticeable for: 1) low PAWC in contrast to high PAWC soils; 2) summer rainfall sites (Emerald) in comparison to uniform or winter rainfall sites (Young, Merredin), and 3) wetter than drier sites. These differences are due to the facts that when PAWC is high enough to store all rainfall crop yield is no more affected by PAWC, particularly at dry sites, and that summer rainfall sites rely more on high PAWC to store water in summer for wheat growth in the following season thus the yield-PAWC relationship becomes more stable.

The number of years of yield data has an additional significant impact on prediction outcomes, due to the fact that the *y*-PAWC model was developed based on long-term averages of simulation results. However, the results also indicate that consecutive 5-10 years of yield data would enable a much accurate estimation of PAWC (Fig 1).

While the results imply a potential to further develop the crop/vegetation-based approach for predicting PAWC, it should be emphasised that the deployed method of the inverse modelling was derived based on water-limited wheat yield under no nutrient deficit. In reality, crops are likely grown under nitrogen stressed conditions (Hochman et al., 2009) and different crops may respond differently to PAWC and climate thus complicating the inverse modelling approach. Nonetheless, our excise focusing on wheat demonstrates the potential to estimate soil PAWC through carefully measuring crops, instead of time- and labour-intensive soil sampling.

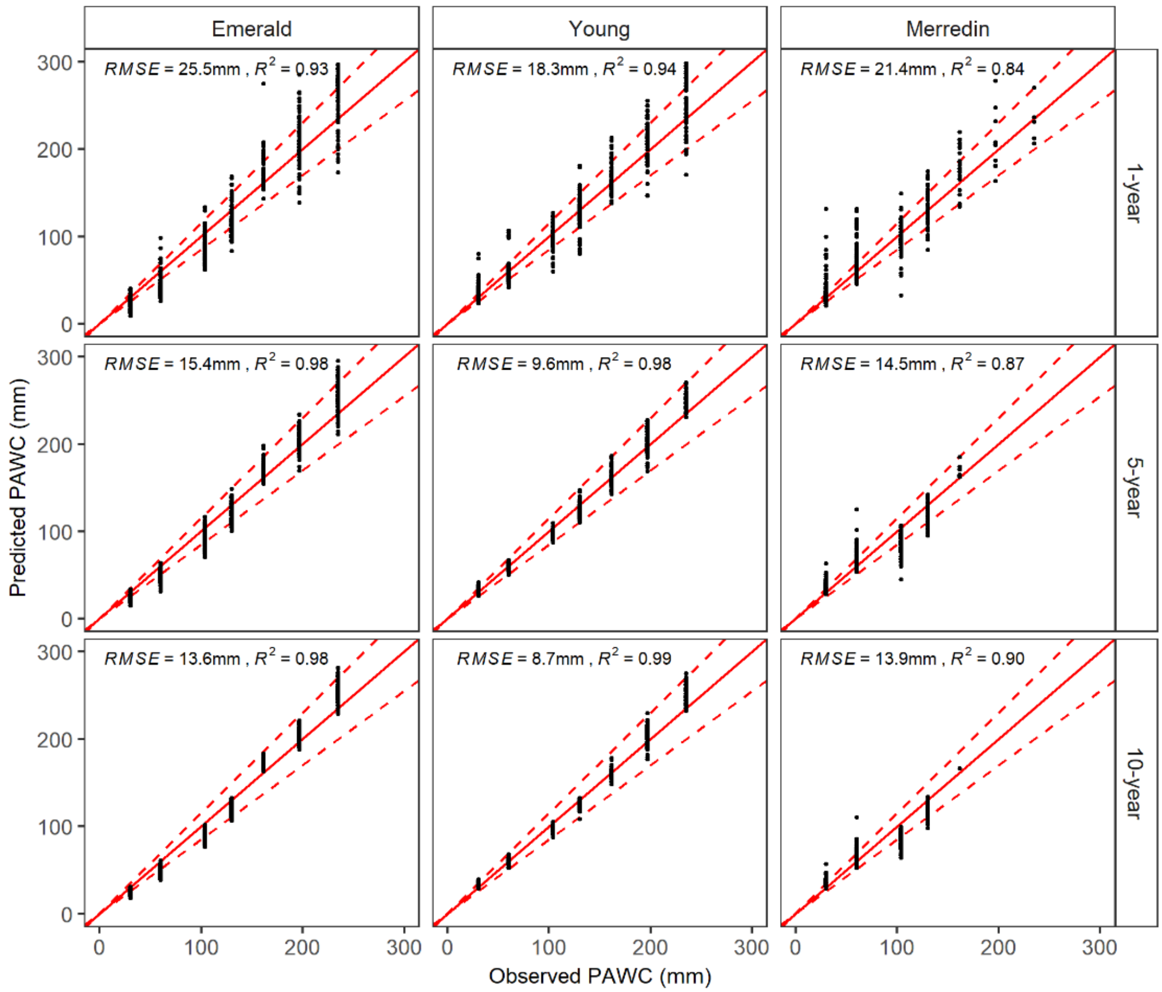


**Figure 2 Skills of PAWC prediction with APSIM inverse modelling approach at three typical sites, assuming wheat yield data from one year, 5- and 10- consecutive years are available. Solid red line is the 1:1 line. Dashed red lines indicate 15% deviation from the 1:1 line.**

*The spatial modelling approach*

The spatial modelling approach revealed significant correlations between soil AWC and the bio-climate variables (Fig 3). In the 0-5 cm and 5-15 cm depth intervals, the 19 climatic variables together can explain 50% (i.e., $R^2$ = 0.50) and 52% ($R^2$ = 0.52) of the variance in AWC, respectively.

However, the quality of model predictions decline in deeper soil layers. In the 100-200 cm interval, only 18% of the variance in AWC can be explained. In terms of RMSE, it generally increases with soil depth. In the top two layers, RMSE is 7.9 mm and 7.3 mm, respectively, indicating the good predictive accuracy, but it increases to 19.5 mm in the 100-200 cm interval. Looking into the relationship between AWC and individual climatic variables, we found that the most important variable is the temperature of direst quarter ($T_{dq}$). In the top two layers, $T_{dq}$ alone can explain more than 20% of the variance in AWC.

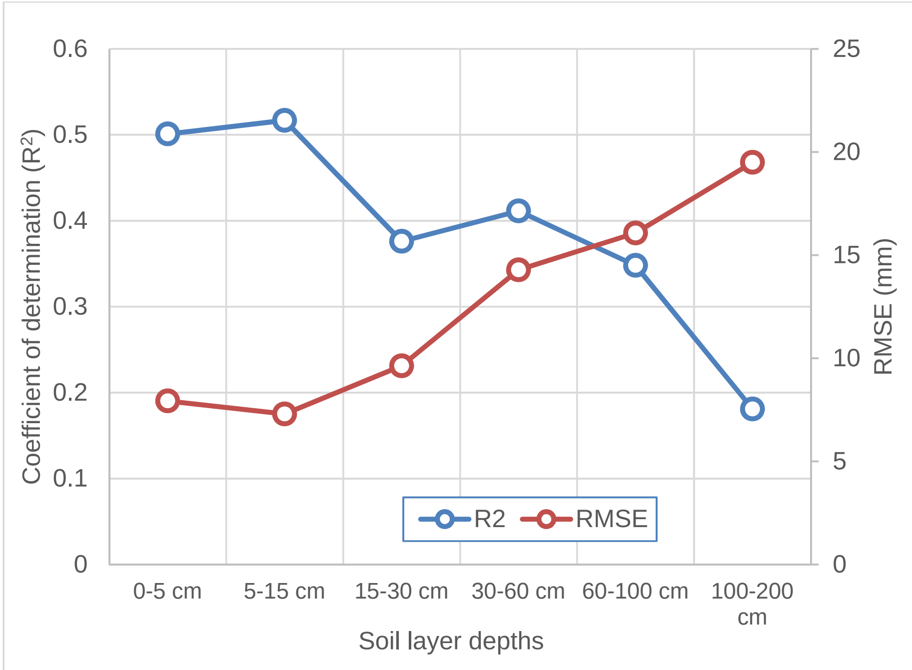


**Figure 3 Skills of the spatial modelling approach to predict AWC in different soil layers with bio-climate variables.**

## Conclusion

Our results show that PAWC can be inversely predicted with water-limited crop yield in the absence of nutrient stress. Biologically meaningful climate variables (bio-climate variables) can be calculated from historical climate data, and these bio-climate variables together with a machine learning approach could explain up to 50% of the variance in soil AWC in top soil layers across sites. These results demonstrate the potential to use climate and crop yield data to predict soil water holding capacity.

## References

Grundy, M.J. et al., 2015. Soil and Landscape Grid of Australia. Soil Research, 53(8): 835-844.

Hochman, Z., Holzworth, D. and Hunt, J.R., 2009. Potential to improve on-farm wheat yield and WUE in Australia. Crop and Pasture Science, 60(8): 708-716.

Lawes, R.A., Oliver, Y.M. and Robertson, M.J., 2009. Integrating the effects of climate and plant available soil water holding capacity on wheat yield. Field Crops Research, 113(3): 297-305.

Wang, E., Cresswell, H., Xu, J. and Jiang, Q., 2009. Capacity of soils to buffer impact of climate variability and value of seasonal forecasts. Agricultural and Forest Meteorology, 149(1): 38-50.

Wong, M.T.F. and Asseng, S., 2006. Determining the Causes of Spatial and Temporal Variability of Wheat Yields at Sub-field Scale Using a New Method of Upscaling a Crop Model. Plant and Soil, 283(1): 203-215.

Wright, M.N. and Ziegler, A., 2017. ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. Journal of Statistical Software, 77(i01).