Forecasting soil water with Deep Learning for improved dryland agriculture decision-making

Qazi Muqeet Amir, Thomas Bishop
Sydney Institute of Agriculture, School of Life and Environmental Sciences, The University of Sydney, NSW 2015

Abstract
Soil water availability is a critical constraint to agricultural productivity. The ability to predict current and future soil water accurately is key in informing decisions relating to irrigation, fertiliser use, and yield. While soil water forecasting has been conducted in literature previously for irrigated fields and near-surface soil (0-5 cm), there is limited understanding of deeper soil water forecasting for non-irrigated fields in Australia. This study aimed to identify the best performing models for forecasting topsoil soil water (0-30 cm) and thus advance our understanding of soil water behaviour in Australian dryland fields.

This study presents various conventional machine learning and novel deep learning methods of forecasting topsoil soil water. The methods applied in this study are Random Forest, XGBoost, Multi-Layer Perceptron (MLP), and Long Short-Term Memory (LSTM). Soil water probe data is sourced from the CosmOz probe network, which spans across Australia, representing its diverse soil types and climatic conditions. Various in-situ and remotely sensed meteorological data were used as features to forecast 0-30 cm topsoil soil water up to 14 days ahead, at 10 sites across Australia.

The results showed that deep learning methods outperform conventional machine learning methods in every location. The LSTM model performs the best, due to the persistence of information over the feature space. In most instances XGBoost outperforms Random Forest.

Introduction

An essential aspect of dryland agricultural practice is determining inputs and managing risks to optimise yield productivity. Soil water availability is a critical constraint in yield productivity as dryland crops are highly dependent on water available to them in the root zone. Crop dependency on soil water availability means that forecasts of soil water can inform growers' decisions relating to irrigation requirements, sowing, and fertiliser rates for a growing season. With rapidly varying weather trends across Australia, soil water forecasting is essential to efficiently plan a growing season.

Soil water forecasting has largely been conducted for irrigated fields (Togneri et al., 2022; Dubois et al., 2021; Yu et al., 2020; Cai et al., 2019; Brinkhoff et al., 2019; Adeyemi et al., 2018). In irrigated fields short-term forecasts are of use to a grower as they serve as a guide for irrigation timing. Hence, a lead time of seven days is aimed for, without motivation to predict beyond this. Irrigated fields differ from dryland fields in their forecastability as there are regular inputs of water to the soil in irrigated fields that add short-term seasonality, making them more predictable.

In an Australian context, *Brinkhoff et al., (2019*), used a variety of conventional machine learning methods (Lasso Regression, Linear Regression, Support Vector Machine, Decision Tree, and Random Forest) to predict soil water across five irrigated fields at a site in Whitton, New South Wales, Australia. They found that Random Forest produced the most accurate results across their five fields. The study is limited in its spatial and temporal domain as it only forecasts soil water at one site for a lead time of seven days. *Ahmed et al., (2021)* forecasted surface soil water (2 cm) with lead times up to 30 days using a deep learning model, Gated Recurrent Unit Model, in four dryland sites across the Murray-Darling basin. This study provides a novel approach to feature selection, input noise reduction, and deep learning architecture while producing high performance results. Surface soil water is of limited use to growers as temperate crop rooting depths grow as deep as 2 m (Fan et al., 2016). The current literature features a limited spatial domain and depth, which does not inform soil water forecastability patterns across the various soil types and climatic conditions of Australia.

The aim of this study is to identify the best performing models for forecasting topsoil (0–30 cm) soil water in Australian dryland fields, and to highlight the relationships between soil water, and various meteorological and geophysical variables. This is achieved through the development of conventional machine learning and novel deep learning models to forecast 0-30 cm deep soil water at 10 sites spanning Australia. The comparison of forecast accuracies from these four distinct machine learning and deep learning models informs the mechanisms that underlie soil water behaviour in dryland soils.

Domain

This study aims to forecast soil water at the locations of the soil water probes available through the CosmOz network (Hawdon et al., 2014). The Australian cosmic-ray soil water monitoring network (CosmOz) consists of 27 cosmic ray sensors that provide hourly estimates of soil water over an area of approximately 30 ha. For this study, only sites with more than five years of data were selected to maintain a suitable number of observations to train the models on. The 10 selected sites are illustrated in Figure 1. Each site had hourly data available between 2013-2023. The CosmOz soil water observations at 30 cm were used to validate our modelled soil water.

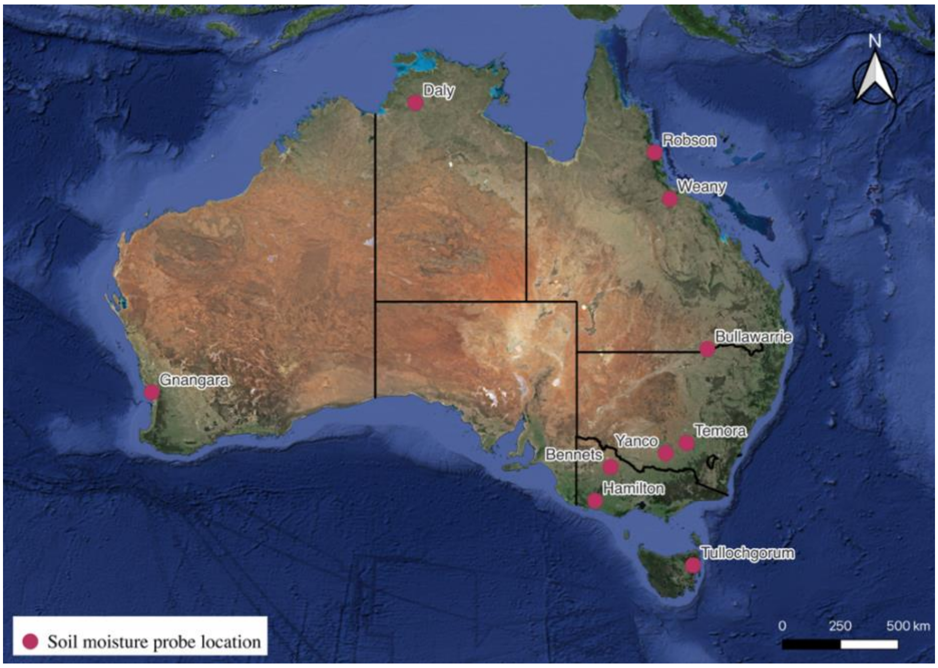

Figure 1: Locations of selected CosmOz soil water probes

Methods

Feature selection is key in developing an accurate forecasting model as the response variable is entirely dependent on the interaction of the model with the features. Eight meteorological variables were selected as features for this model based on their theoretical relationship with soil water. The features are sourced from various remotely sensed and modelled sources summarised in Table 1.

Table 1: Features

Feature	Spatial resolution (km)	Temporal Resolution	Source
Rainfall (mm)	5	Daily	SILO
Maximum temperature (°C)	5	Daily	SILO
Solar radiation (MJ/m^2)	5	Daily	SILO
Relative humidity at maximum temperature (%)	5	Daily	SILO
Evapotranspiration (mm)	0.5	8-day	NASA MODIS
Enhanced Vegetation Index	0.25	As available	ESA SENTINEL-2
Southern Oscillation Index	-	Daily	BOM

Indian Ocean Dipole	-	Daily	BOM

A Null model is employed as a benchmark model. The Null model takes the predicted soil water on a given day to equal the observed soil water from the previous n^{th} day's observation as a prediction at each time step (where, *n* is the lead time). The conventional machine learning methods selected were Random Forest and XGBoost. The Random Forest and XGBoost models are popular ensemble machine learning models that are commonly used as a baseline model due to their ease of applicability and general high performance in forecasting. Random Forest models create an ensemble of randomly selected decision trees, while XGBoost makes use of a weighted loss function that updates a decision tree to account for imbalanced and noisy data.

The deep learning models selected for this study are Multi-Layer Perceptron (MLP) and Long Short-Term Memory (LSTM). The MLP model is a feedforward artificial neural network consisting of a number of fully connected perceptron layers and non-linear activation functions. The MLP model is the simplest artificial neural network as they only propagate information in one direction and are trained through backpropagation. For this reason, they are a suitable baseline deep learning model. The MLP model used in this study is illustrated in Figure 2. Unlike MLPs, LSTM models do not only feed information forward but recurrently too. This allows information to persist over time and account for patterns in the feature space. The LSTM model employed in this study takes a 7-day feature series to forecast.

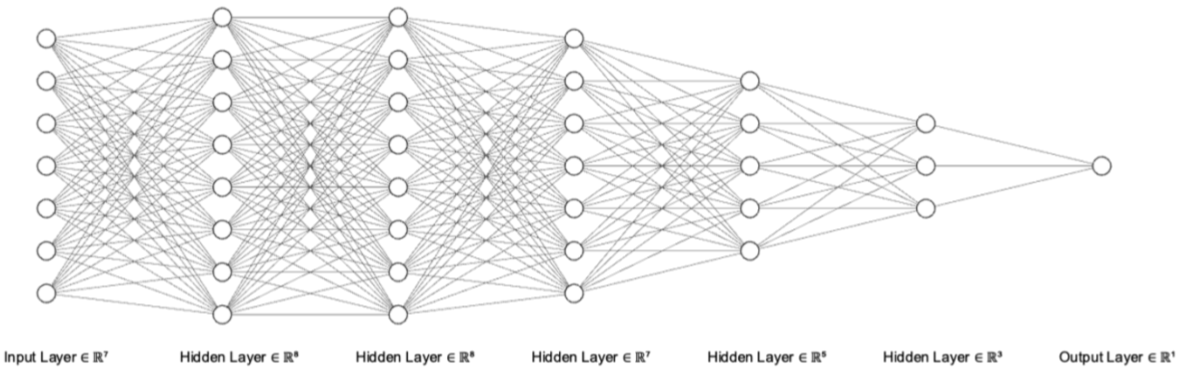

Figure 2: Multi-Layer Perceptron model architecture

Lead times of 1-14 days were selected to understand the short term forecastability of soil water. Prediction performance was assessed with the Nash-Sutcliffe Efficiency (NSE). A NSE value of 1 indicates a perfect prediction, while a NSE value of less than 0 indicates the mean of the data is a better predictor of the response variable than the modelled values.

Results and Discussion

Deep learning methods perform better than the conventional machine learning methods at every site, as illustrated in
Figure *3*. Overall, the LSTM model performs the best due to its persistence of information over the feature space. This can be expected due to LSTM models accounting for the autoregressive nature of soil water, as water penetration in soil is non-instantaneous. In most instances XGBoost outperforms Random Forest. This can be attributed to XGBoost models updating the weights of more difficult examples, thus training them better. The Null model relies entirely on the autoregressive nature of soil water, and in most instances (Robson, Hamilton, Bennetts, Temora, Bullawarrie, and Tullochgoram), outperforms Random Forest for low lead times.

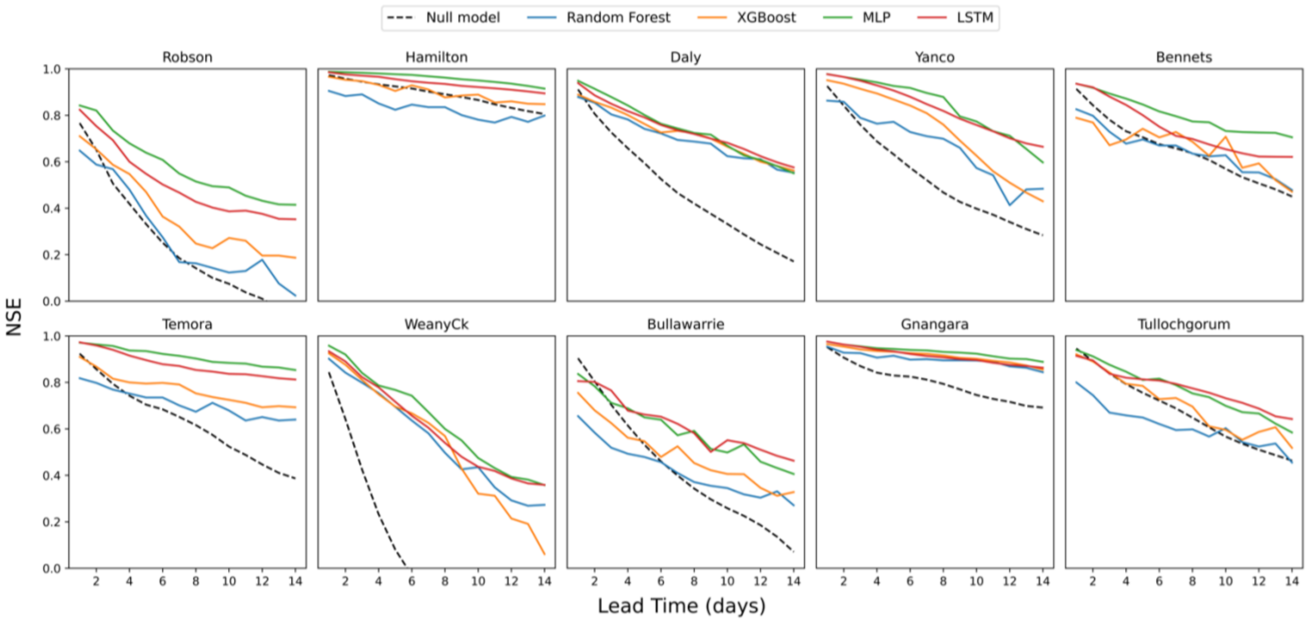

Figure 3: NSE over different lead times for each site

Conclusions and future work

Our study demonstrates the potential to forecast soil water with deep learning algorithms. While conventional machine learning methods remain a norm in forecasting soil water, a shift towards employing deep learning algorithms produces higher accuracy forecasts. These models enable the ability of growers to make decisions about sowing dates and fertiliser rates with high certainty, up to two weeks in advance. Further research can enable lead times up to 30 days to be achieved. To further improve on the deep learning forecasting capabilities, a Transformer neural network can be developed, which introduces Self-Attention capabilities that improve on long term information retention across the neural network.

References

Fan, J., McConkey, B., Wang, H., Janzen, H., 2016. Root distribution by depth for temperate agricultural crops. Field Crops Research 189, 68–74. https://doi.org/10.1016/j.fcr.2016.02.013

Togneri, R., Felipe dos Santos, D., Camponogara, G., Nagano, H., Custódio, G., Prati, R., Fernandes, S., Kamienski, C., 2022. Soil water forecast for smart irrigation: The primetime for machine learning. Expert Systems with Applications 207, 117653. https://doi.org/10.1016/j.eswa.2022.117653

Adeyemi, O., Grove, I., Peets, S., Domun, Y., Norton, T., 2018. Dynamic Neural Network Modelling of Soil Water Content for Predictive Irrigation Scheduling. Sensors 18, 3408. https://doi.org/10.3390/s18103408

Brinkhoff, J., Hornbuckle, J., Ballester, C., 2019. Soil water forecasting for irrigation recommendation. IFAC Proceedings Volumes (IFAC Papers-OnLine) 52, 385–390. https://doi.org/10.1016/j.ifacol.2019.12.586

Dubois, A., Teytaud, F., Verel, S., 2021. Short term soil water forecasts for potato crop farming: A machine learning approach. Computers and Electronics in Agriculture 180, 105902. https://doi.org/10.1016/j.compag.2020.105902

Ahmed, A.A.M., Deo, R.C., Raj, N., Ghahramani, A., Feng, Q., Yin, Z., Yang, L., 2021. Deep Learning Forecasts of Soil Water: Convolutional Neural Network and Gated Recurrent Unit Models Coupled with Satellite-Derived MODIS, Observations and Synoptic-Scale Climate Index Data. Remote Sensing 13, 554. https://doi.org/10.3390/rs13040554

Yu, J., Xin, Z., Xu, L., Dong, J., Zhangzhong, L., 2020. A hybrid CNN-GRU model for predicting soil water in maize root zone. Agricultural Water Management 245. https://doi.org/10.1016/j.agwat.2020.106649