Geoderma Regional 2–3 (2014) 110–118

Contents lists available at ScienceDirect

Geoderma Regional

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

Predicting and mapping the soil available water capacity of Australian wheatbelt

J. Padarian [a,*], B. Minasny [a], A.B. McBratney [a], N. Dalgliesh [b]

[a] Faculty of Agriculture and Environment, Department of Environmental Sciences, The University of Sydney, New South Wales, Australia
[b] CSIRO Ecosystem Sciences, Toowoomba, Queensland, Australia

ARTICLE INFO

Article history:
Received 30 June 2014
Received in revised form 17 September 2014
Accepted 18 September 2014
Available online 16 October 2014

Keywords:
Ensemble model
Genetic programming
Field capacity

ABSTRACT

Soil available water capacity (AWC) is the main source of water for vegetation and it is the potential amount of water available for atmospheric exchange. Studying its spatial distribution is crucial for agricultural planning and management and for use in biophysical modelling. The aim of this work is to obtain a continuous spatial prediction of AWC over Australia's wheatbelt (about 1.75 million km^2), using digital soil mapping techniques. We used a data set of 806 soil profiles which have field measurements of drainage upper limit (DUL) and crop lower limit (CLL). We mapped AWC at five depth intervals (0–5, 5–15, 15–30, 30–60, and 60–100 cm) with the help of different combinations of environmental information (topographic, climatic, soils, landsat imagery, gamma-ray spectrometry) as covariates. The modelling techniques used were symbolic regression (GP), Cubist, and support vector machines (SVM). We also tried two averaging methods to generate an ensemble model. We observed decreasing RMSE values with the addition of extra covariates and also an expected decreasing soil depth. In general, SVM produced the best accuracy. We were able to improve the predictions using one of the ensemble techniques, based on a weighted average of GP, Cubist and SVM model. The map generated with the optimal ensemble model was an unrealistic representation of AWC therefore we decided to present a sub-optimal model as the final map. We stress the need to not only focus on the numerical performance in order to obtain a flexible and stable model, but also a coherent visual representation without anomalies.

1. Introduction

Soil available water capacity (AWC) is defined as the amount of water soil can store between field capacity or drainage upper limit (DUL) and wilting point or crop lower limit (CLL). It is the main source of water for vegetation development and is related to the potential amount of water a soil could make available for the atmosphere through evapotranspiration (Dunne and Willmott, 1996). Information about its distribution in space is crucial for planning and management in agriculture, and for ecological modelling.

To model the spatial distribution of AWC, digital soil mapping has been proposed (McBratney et al., 2003). The scorpan model describes that soil properties can be predicted from its predicting factors in the form of empirical regression equations. The general steps in the modelling process involve: collection of a dataset of soil observations over the chosen area of interest; compilation of relevant covariates for the area; calibration or training of a spatial prediction function based on the observed dataset; interpolation and/or extrapolation of the prediction function over the whole area of interest; calculation of uncertainty; and finally validation using existing or independent datasets.

Despite the importance of AWC, not many studies present a mapping methodology at national scale. Hong et al. (2013) successfully predicted AWC for Korea based on detailed soil series maps and modal profiles, also recognising the shortcomings due to variability within mapping units. Poggio et al. (2010) used morphological features as covariates, obtaining an optimal model selecting covariates using generalised additive mixed models, to map AWC in Scotland. Ugbaje and Reuter (2013) used two different covariate combinations (remote sensing data; terrain, climate, and vegetation attributes) and pedotransfer functions (PTFs) to map AWC in Nigeria, not finding a clear effect of number of covariates on model accuracy. Most of these studies used PTFs to predict the AWC. Thus the uncertainty of the map depends also on the accuracy of the PTFs.

In digital soil mapping, the visual representation of the product (map) depends on the covariates and the models used. Several studies that looked at the selection and parsimony of the covariates, and also studies have compared different data mining predictions. However no work has looked at the effect of both covariates and models on the visual representation of the map.

A good digital soil map should have a balance of model parsimony (number of covariates), accuracy (numerical performance) and realism of the visual representations (maps). The aim of this work is to obtain a

* Corresponding author.
E-mail addresses: jose.padarian@sydney.edu.au (J. Padarian), budiman.minasny@sydney.edu.au (B. Minasny), alex.mcbratney@sydney.edu.au (A.B. McBratney), neal.dalgliesh@csiro.au (N. Dalgliesh).

http://dx.doi.org/10.1016/j.geodrs.2014.09.005
2352-0094/

Fig. 1. Location of soil profiles from APSRU database. Grey area represents the bioregion subset where predictions were made.

continuous spatial prediction of AWC over Australia, based on field measured data that reconcile these three aspects, exploring the use of different covariate combinations and modelling techniques, and visually inspecting the generated maps.

2. Materials and methods

2.1. Data sets and study area

The data set used correspond to a CSIRO Ecosystem Sciences (APSRU) compilation of 806 soil profiles that includes field measurements of DUL and CLL for the most commonly grown crops of Australia (Dalgliesh et al., 2012). Procedures for determination of these properties are described in the accessory publication of the article by Dalgliesh et al. (2009), "Procedures for determination of soil properties and states relevant to crop simulation and farmer crop management decision making". The method is a modification of the techniques described by Ratliff et al. (1983). Briefly, an area covering about 16 m^2 of soil was wetted using a trickle system. The water content and drainage were monitored using a neutron moisture meter at the access tube at the centre of the site down to a depth of 180 cm. Once the soil was judged to be thoroughly wet, it was allowed to drain until moisture monitoring indicated a minimal change in profile water status. Samples for gravimetric moisture content and bulk density were taken. For CLL, crops were grown in the field, and a rain-exclusion tent of 9 m^2 was installed. At crop maturity, soil moisture was determined at different depths.

The soil orders represented by this database, according to the Australian Soil Classification System, correspond to Calcarosol (4.22%), Chromosol (4.96%), Dermosol (2.23%), Ferrosol (0.99%), Kandosol (2.23%), Podosol (0.12%), Sodosol (5.21%), Tenosol (0.87%), Vertosol (22.08%), and 57.07% of unclassified soils. Based on the location of the unclassified soils and the dominant soil order map of Australia (ASRIS), they correspond to Dermosol (10%), Ferrosol (0.87%), Hydrosol (1.09%), Kandosol (32.61%), Kurosol (6.3%), Organosol (27.17%), Podosol (1.52%), Rudosol (0.87%), Sodosol (11.09%), Tenosol (0.43%), and Vertosol (8.04%).

A bioregion classification by Thackway and Cresswell (1995) was used to limit the study area, selecting the bioregions which contained observations of the APSRU data set. This selection, usually referred as "wheatbelt", is represented as the grey area in Fig. 1 and it is equivalent to about 1.75 million km^2.

2.2. Digital soil mapping model

In this study we used the scorpan approach (McBratney et al., 2003) as an empirical quantitative description of relationships between soil and other spatially referenced factors. It is represented as $S = f(s, c, o, r, p, a, n) + \epsilon$, where S: is the variable of interest (DUL and CLL), s: stands for soil (other properties of the soil at a point), c: climate (climatic properties of the environment at a point), o: organisms (vegetation or fauna or human activity), r: topography (landscape attributes), p: parent material (lithology); a: age (the time factor); n: space (spatial position); and ϵ correspond to the spatially modelled residuals (usually by kriging).

2.2.1. Soil attribute: S

We predicted soil properties related with water holding capacity of a soil. DUL represents the volumetric water content an initially saturated soil holds after draining for 2–3 days (Veihmeyer and Hendrickson, 1949). On the other hand, CLL corresponds to the volumetric soil water remaining in the soil after a healthy crop, with uninterrupted root development, has reached maturity under soil water-limited conditions (Hochman et al., 2001). Both properties are measured in the field independently and were governed by different processes, hence different sources of error, thus we decided to model them separately.

Table 1
Statistics of soil samples used for model generation.

	Mean	S.D.	Min.	Median	Max.
DUL (%)	30.20	11.54	3.00	32.00	56.00
CLL (%)	16.90	8.61	0.40	18.00	53.00