Geoderma 259–260 (2015) 81–92

Contents lists available at ScienceDirect

Geoderma

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

Modelling the abundance of soil calcium carbonate across Australia using geochemical survey data and environmental predictors

John Wilford [a,*], Patrice de Caritat [a,b], Elisabeth Bui [c]

[a] Geoscience Australia, GPO Box 378, Canberra, ACT 2601, Australia
[b] Research School of Earth Sciences, The Australian National University, Canberra, ACT 2601, Australia
[c] CSIRO Land and Water, GPO Box 1666, Canberra, ACT 2601, Australia

ARTICLE INFO

Article history:
Received 23 February 2015
Received in revised form 7 May 2015
Accepted 10 May 2015
Available online 26 May 2015

Keywords:
Geochemical modelling
Calcium carbonate
Alteration and weathering processes
Regolith
Data mining
Australia

ABSTRACT

Soil calcium carbonate is a key component of the regolith, particularly in arid and semi-arid regions. It influences soil properties, is an important terrestrial carbon store, and is used in mineral exploration. Previous national overviews of soil carbonate distribution have been compiled from regional soil, regolith or geological maps with varying degrees of confidence and consistency. Here we develop a decision tree approach based on a piecewise linear regression model to map soil calcium carbonate abundance at the continental scale. The model is based on relationships established from 1311 field sites sampled across Australia at two depths (0–10 and ~60–80 cm) and 49 national environmental predictor datasets reflecting fundamental soil formation processes (e.g., climate, parent material, relief).

The resulting map of soil calcium carbonate distribution is based on the median results of twenty model runs generated by randomly selecting 90% training and 10% validation splits of the input data. Results present an average coefficient of determination (R^2) of 0.5 on the validation dataset. The predictors used in the modelling are consistent with our understanding of the controls on the sources (inputs), formation, and preservation of soil calcium carbonate across Australian landscapes. The model produces a continuous, internally consistent, quantitative prediction of soil calcium carbonate abundance in surficial regolith at a resolution of 90 m with associated uncertainty estimates. This type of analysis provides a framework for better understanding environmental controls on regolith processes and has the potential to also be used to predict other soil geochemical properties. Furthermore, the methodology has utility globally where geochemical survey data are available.

1. Introduction

In contrast to primary carbonate, which is associated with carbonate-bearing rocks (e.g., limestone, marble), secondary carbonate accumulates in the regolith (i.e., all weathered material above fresh bedrock, including soil) through a process of calcification (Lal and Kimble, 2000; Schaetzl and Anderson, 2005). Caliche, travertine, kunkar and calcrete are all terms used to describe secondary carbonate accumulations in the weathered zone. Soil carbonate consists mainly of calcium carbonate ($CaCO_3$) with variable degrees of Mg-substitution (Milnes and Hutton, 1983; Bui et al., 1990; Hill et al., 1999; Eggleton, 2001). Carbonate abundance influences the pH and a range of other soil properties including texture, porosity, permeability, hydraulic conductivity, structure, and cation exchange capacity, which are important attributes for the agriculture sector (Peverill et al., 2001). In terms of global carbon accounting and modelling, the amount of inorganic carbon stored in soil (950 Gt) is similar to atmospheric carbon (800 Gt) (Ontl and Schulte, 2012). In arid and semi-arid regions soil inorganic carbon stocks can be much larger than organic stocks and, depending on the land management, they could be a potentially major contributor to the global carbon flux (Sanderman, 2012). The stable isotopic composition of secondary carbonates in soils has been used to reconstruct paleoclimatic fluctuations and changes in C3/C4 vegetation patterns (e.g., Cerling, 1984; Schlesinger and Pilmanis, 1998). Secondary carbonate also plays an important role in mineral exploration as it can be a useful geochemical sampling medium for locating areas of enhanced gold (McQueen et al., 1999; Lintern et al., 2006) and copper, zinc, nickel, and cobalt (McQueen, 2006) potential. Groundwater-related carbonate deposits are also known to act as traps for uranium mineralisation (e.g., U-calcrete deposits; Butt et al., 1977).

Globally, carbonate-rich soils cover up to 13% of the Earth's land surface (Summerfield, 1991); this estimate is based on maps that are empirically based, qualitative and rarely spatially detailed. For example in Australia, the driest inhabited continent, soil and to a lesser extent surface geology maps are the principal datasets showing the distribution of soil carbonate. National coverages are produced only at low resolution (>1:1,000,000 scale), whereas more accurate information associated with 1:100,000 and 1:50,000 scale mapping has limited spatial coverage. Soil carbonate covers an estimated 21% to 50% of Australia, as

* Corresponding author.
E-mail address: john.wilford@ga.gov.au (J. Wilford).

http://dx.doi.org/10.1016/j.geoderma.2015.05.003
0016-7061/

inferred by the distribution of regolith carbonate (Chen et al., 2002: Fig. 1) and soil alkalinity (Fitzpatrick and Merry, 2000), respectively.

In this paper we demonstrate the use of geo-referenced chemical data from a national geochemical survey to generate a spatially explicit calcium carbonate abundance prediction. We apply a spatial correlative modelling approach to quantitatively predict the distribution of soil calcium carbonate over Australia at 3 arc second resolution (~90-m). The diverse climates and landscapes found in Australia provide a suitable analogue for understanding factors influencing soil calcium carbonate distribution globally. The aims of the paper are to: 1) address inconsistencies and gaps in the existing national maps of soil carbonate; 2) present a generic approach that can be used to model and map geochemical and mineralogical properties in the upper regolith; and 3) better understand the environmental controls on soil carbonate formation and preservation in the landscape.

2. Methods

2.1. Modelling approach

Relationships between soil calcium carbonate distribution and environmental predictors (e.g., rainfall, parent material, terrain, vegetation) are explored using the machine learning decision tree software 'Cubist' (www.rulequest.com). 'Cubist' has been used effectively in predicting soil properties at the national scale in previous studies (Henderson et al., 2005; Bui et al., 2009) and is described in Kuhn and Johnson (2013). The 'Cubist' model structure consists of a decision-based component coupled with multiple piecewise linear regression models. Continuous and categorical variables can be used in the decision tree to split the data into increasingly homogeneous sub-regions. Splits in the decision tree are based on recursive partitioning of the prediction variable to minimise the standard deviation across all potential splits (Henderson et al., 2005). Predictions within these sub-regions are then explained using linear regression models. This nested modelling approach enables local linear segments of the data to be captured from an otherwise poorly correlated dataset or one that exhibits non-linearity. The 'Cubist' model rule-set is expressed as paired conditional statements and associated linear models as shown in the example below.

(1) If (conditional statement based on decision tree splits)

Rainfall < a

and

Lithology = (p, q, r)

(2) Then (linear model)

$$\text{Property} = c_1 * \text{rainfall} + c_2 * \text{slope} + c_3 * \text{elevation} + \ldots..$$

where a, c_1, c_2, and c_3 are constants, p, q, and r are classes, and Property is the abundance of calcium carbonate.

Being explicit, the rules can be evaluated with expert knowledge and compared with conceptual models of landscape and soil development. Where more than one rule explains the prediction response the average of the rule predictions is used for the final model.

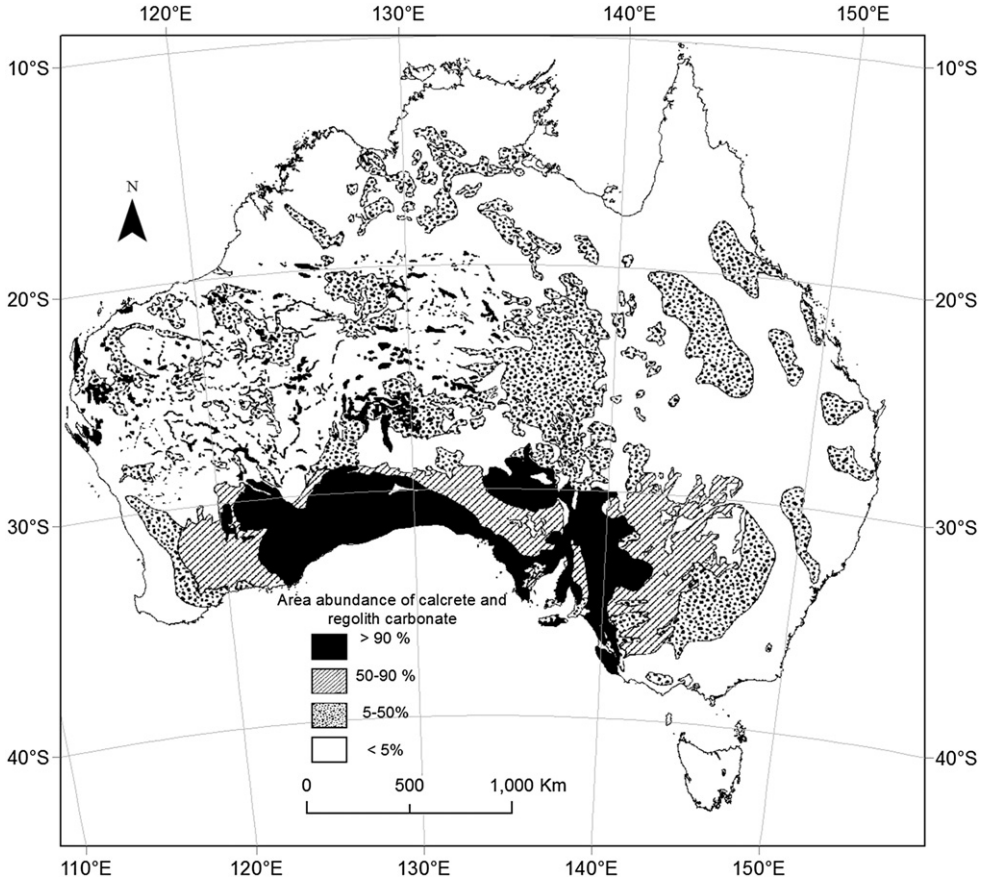

Fig. 1. Interpretative map of the distribution of Australian calcrete and regolith carbonates based largely on soil and geological mapping. From Chen et al. (2002).