Soil distribution relationships as revealed by a global soil database

Jonathan Gray[A], Geoff Humphreys[B] and Jozef Deckers[C]

[A]NSW Department of Environment, Climate Change and Water, PO Box 3720, Parramatta (Sydney), NSW 2124 Australia, Email jonathan.gray@environment.nsw.gov.au
[B]Department of Physical Geography, Macquarie University (Sydney), NSW 2109 Australia (deceased)
[C]Catholic University of Leuven, Vital Decosterstraat 102, 3000 Leuven, Belgium.

Abstract
Broad quantitative relationships between the environmental factors of climate, parent material and topography and a range of important soil properties and World Reference Base (WRB) soil types were developed using the ISRIC WISE Global soil database Three different analytical approaches were used in the analysis, involving (i) multiple linear regressions, (ii) fitted decision trees and (iii) categorical analysis with median values. Despite the relatively low to moderate strengths of the derived relationships, they can provide useful first approximations of soil character under different environmental conditions and could be applied in broad quantitative soil modelling and mapping programs. They have the potential for widespread application as they should be universally applicable, are based on readily available data and do not require sophisticated quantitative modelling techniques. Most relationships are in accord with accepted pedological thinking and support the state factor model of soil formation, but some anomalies are observed and deserve further examination. The results reveal the dominant influence of climate and parent material in controlling the distribution of many soil properties, with the influence of topography being less evident, at least at the global scale. The WRB soil classification scheme is shown to be least partly guided by genetic factors.

Key Words
Digital soil survey, soil prediction, environmental correlation, modelling, soil genesis.

Introduction
Soil data is increasingly being relied upon in ecological, hydrologic, climatic and other environmental applications in addition to its traditional role in agriculture and natural resource management. There is a need for precise quantitative relationships between soil and key environmental factors to aid soil data collection and soil modelling around the globe. Such relationships form the basis of digital soil mapping techniques, which are widely considered to be where the future of soil survey lies (Lagacherie and McBratney 2007, Hartemink 2006; Grunwald 2006). However, most published information on soil-environment relationships is of a qualitative form, and even these are often not clearly enunciated. Few universally applicable quantitative relationships have been published and widely accepted to date. Most that have been published relate to specific regions rather than to continental or global scales. Heuvelink (2005) laments that most of our knowledge on soil forming processes is available only in a conceptual or descriptive form and that so far we have not succeeded in building a generic, quantitative, reproducible model that predicts the soil from its controlling factors in a satisfactory way.

Soil-environment relationships are generally based on the principles inherent in the now widely accepted State Factor model of soil formation and the associated fundamental soil equation. This model, sometimes referred to as the *clorpt* model, maintains that the soil or soil property is a function of the factors of climate (*cl*), organism or biota (*o*), relief or topography (*r*), parent material (*p*) and time (*t*) (Jenny 1941). The equation has been described as “a general statement implying that soils are natural bodies that are distributed in a predictable way in response to a systematic interaction of environmental factors” (Hudson 1992).

Methods
The ISRIC WISE Global Datasets (2002 and 2008 versions) were examined to extract relationships that could be used to model soil behaviour at the global level. The datasets contain several thousand geo-referenced soil profiles from up to 149 countries (Batjes 2002, 2008). It was established by the International Soil Reference and Information Centre using the framework of the World Inventory of Soil Emission Potential (WISE) project. Climatic data was derived from a world climatic dataset contained in FAO/IIASA (2002).

Multiple queries were created in the Access databases to extract key soil properties, WRB soil types and the associated environmental variables relating to topography, parent material and climate. These variables were divided into classes as outlined below. Age and biotic activity could not be included in the query process due to their lack of meaningful indicators.

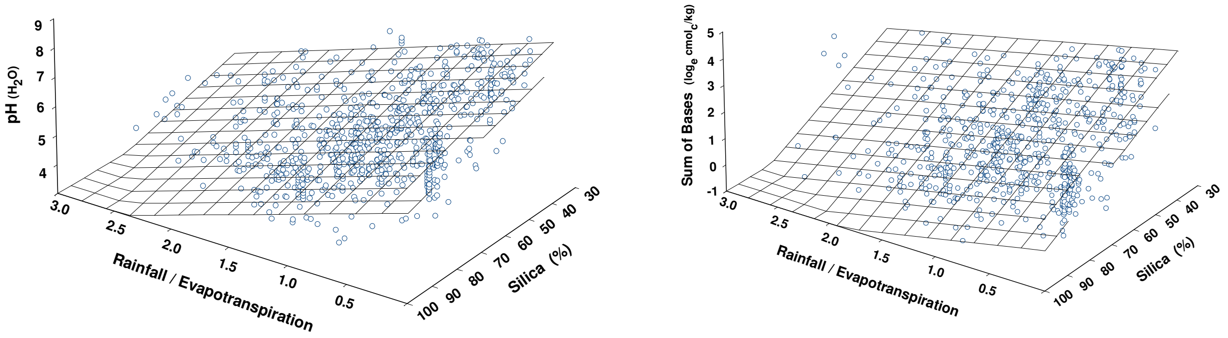
Topography – slope percent in conjunction with landform features, two classes
Parent material – broad silica (SiO_2) content; five classes with one based on calcareous character
Climate – annual rainfall (R) and rainfall/ potential evapotranspiration (R/EV); eight classes.
Only 1646 profiles within the datasets matched the final query requirements, comprised of 1108 in near level sites and 538 in sloping sites. The combination of two topographic, five parent material and eight climatic classes resulted in the creation of 80 different environmental regimes, which provided a basis for presenting categorical analysis results.

Exploratory statistical analysis was undertaken using S-Plus (2003) software, including: (i) multi-variate linear regressions (ii) fitted decision trees, and (iii) a categorical approach giving soil types and median soil property values in each of 80 different environmental regimes. Standard statistical techniques were applied to test the performance of the relationships and the relative contributions of each variable. Graphical plots were prepared for all relationships. Predictions of soil properties derived using the three approaches were tested for accuracy against actual values from 100 random samples drawn from the ISRIC database prior to the development of the relationships.

Results

Soil-environment relationships were derived using the different approaches for a wide range of soil properties (pH, sum of bases, CEC, base saturation, organic carbon, ESP, EC_e, clay and sand content) for both topsoil and subsoil horizons, and for WRB soil types. Full results are reported in Gray *et al.* (2009) and Gray *et al.* (in press).

Multiple linear regression approach
Regression relationships between the numerous soil properties and the three environmental variables were established. Results for pH and sum of bases are presented in Figure 1.



pH = 8.69 – 1.22(R/EV) – 0.025(silica%) – 0.0072(slope%)
N = 490, R^2 = 0.40, F model = 106.0, p = <0.0000

ln(SumBase) = 5.57 – 0.81(R/EV) – 0.051(silica%)
N = 781, R^2 = 0.30, F model = 88.3, p = <0.0000

Figure 1. Regression planes of key topsoil properties v R/EV and silica%.

Coefficient of determination (R^2) values are generally in the low to moderate range for most properties. For pH they are moderate (~0.30-0.40), indicating a moderate correlation and that these regression relationships are at least moderately effective in accounting for the variation of this property in the database. R^2 values are low to moderate for sum of bases and clay (~0.30-0.35) and only low for organic carbon (~0.10-0.25), indicating that these relationships are of only low to moderate effectiveness in accounting for the variation of these properties.

Decision tree approach
Decision trees were fitted to the data for each of the key soil properties. For the purposes of simplicity and ease of presentation, the sensitivity of the trees was set so as to finish with approximately 10-12 leaves at the

base. The tree for topsoil pH is presented in Figure 2. It can be seen that for a hypothetical site with very wet climate (G), highly siliceous parent material (HS) and sloping topography (S), the topsoil may be expected to have a pH of 5.2.

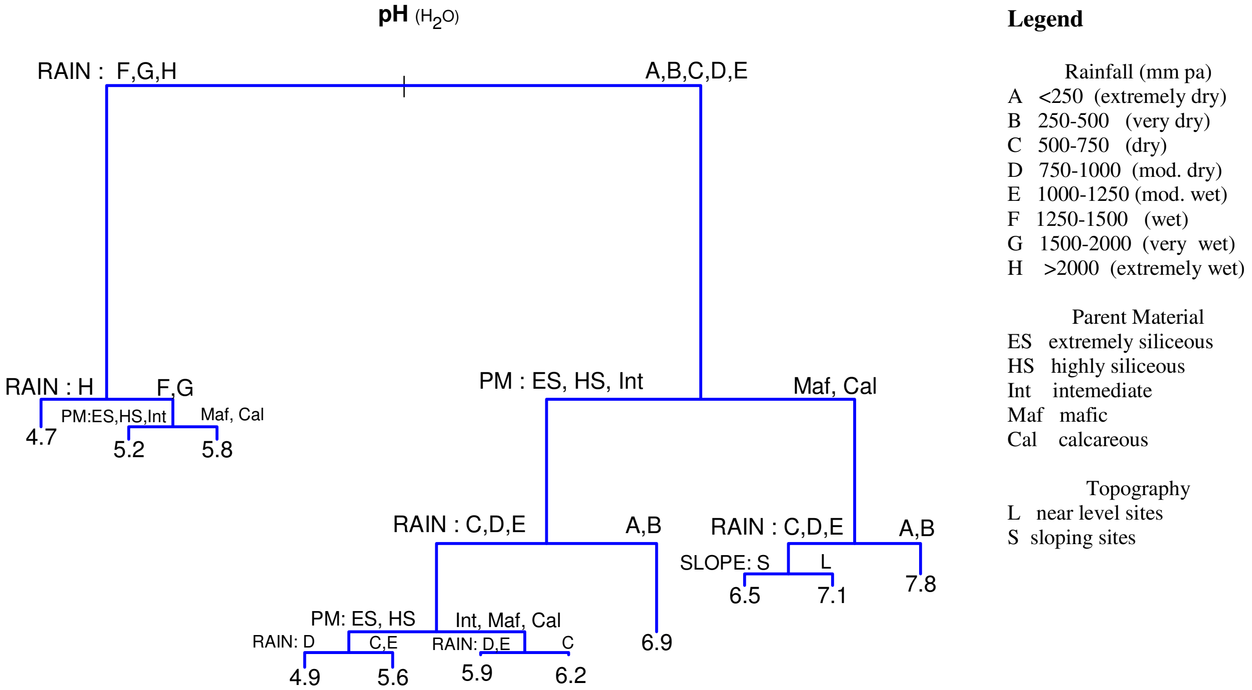


Figure 2. Fitted decision trees for topsoil pH.

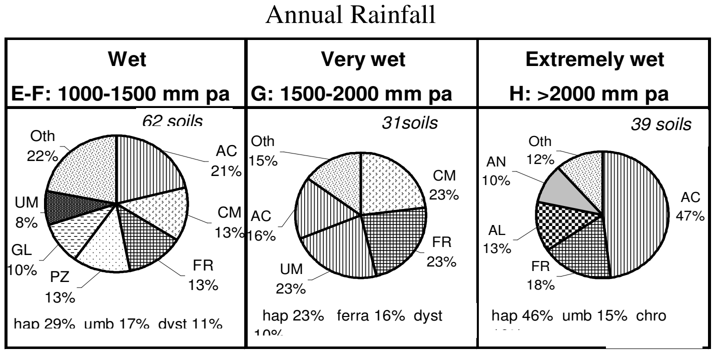
The models all reveal a generally consistent change in soil properties as the environmental factors vary, eg, decreasing pH with increasing wetness and increasing siliceous character. However, some minor anomalies occur, which may be due to low sample numbers or other sources of error as discussed later.

Categorical median value approach

The median values of the various soil properties in each of the 80 different environmental regimes were derived. These were presented in a series of tables and box plots for key properties in Gray *et al.* (2009). The plots and tables reveal the broad trends in behaviour of the soil properties with changing conditions.

World Reference Base soil pie charts

Data on WRB soil types were extracted for each of the topographic-parent material-climate regimes and presented in a series of pie charts. Each chart presents the dominant WRB soil types (up to seven) and the three most dominant qualifiers, together with their percentages and sample numbers. The associated percentages give an indication of the likely probability of occurrence of each soil type or qualifier in that regime. Three example charts are shown in Figure 3. Full results are presented in Gray *et al.* (in press).



Topography: sloping
Parent material: highly siliceous

Figure 3. WRB pie charts for three environmental regimes.

Discussion

Despite the relatively low to moderate strengths of the predictive relationships developed, they may assist in the collection and modelling of soil data as required to meet the growing demands for this information in many earth, environmental, climatic and other scientific modelling programs. More specifically, the relatively simple and pragmatic relationships may:

- provide useful first approximations of soil character under different environmental conditions that could be applied in broad quantitative soil modelling and mapping programs.
- have the potential for widespread application as they should be universally applicable, are based on readily available data and do not require sophisticated quantitative modelling techniques. They may thus facilitate an increased application of digital techniques in addition to conventional techniques in soil survey organisations and in the broader soil science community.
- assist in our understanding of soil formation and soil distribution. Most relationships revealed are in accord with accepted pedological thinking and support the state factor model of soil formation, but some anomalies are observed and deserve further examination.

The relatively low to moderate effectiveness of the relationships reflects the inherent complexity of soil property–environmental factor relationships. A high variability in soil properties is demonstrated even in samples within generally similar environmental condition. Factors that may contribute to this high variability include: the relatively small number of usable samples in the database, influence of other environmental factors not considered here including biotic and age factors; the simplification of climatic, parent material and topographic factors, the polygenetic nature of many soils with influence by past climates; and the influence of "intrinsic" rather than traditionally accepted extrinsic" factors (Phillips 2001). The WRB relationship results suggest that this classification scheme is at least moderately guided by soil forming factors and that it may represent a scheme with an appropriate balance between genetic factors and purely diagnostic features in its underlying principles.

The project has demonstrated the potential value of large world soil databases for the understanding, modelling and mapping of soil distribution around the globe. Further analyses of the ISRIC and other large soil databases with more sophisticated data mining techniques would undoubtedly help to refine the predictive tools and shed further light on factors influencing soil distribution.

References

Batjes NH (2002) 'A homogenised soil profile dataset for global and regional environmental research' (WISE, vers. 1.1), Report 2002/01. (International Soil Reference and Information Centre: Wageningen).

Batjes NH (2008) 'ISRIC-WISE harmonized Global Soil Profile Dataset (version 3.1), Report 2008/02'. (International Soil Reference and Information Centre: Wageningen).

Gray JM, Humphreys GS, Deckers JA (2009) Relationships in soil distribution as revealed by a global soil database, *Geoderma* **150**, 309-323.

Gray JM, Humphreys GS, Deckers JA (2010) Distribution patterns of World Reference Base soil as revealed by a global soil database. (in press)

Grunwald S (2006) What Do We Really Know about the Space-Time Continuum of Soil-Landscapes? In 'Environmental Soil-Landscape Modelling - Geographic Information Technologies and Pedometrics'. (Ed S Grunwald) pp. 3-36. (Taylor and Francis)

Hartemink AE (2006) 'The Future of Soil Science'. (International Union of Soil Sciences (IUSS): Wageningen).

Heuvelink GBM (2005). Foreword. In 'Environmental Soil-Landscape Modelling - Geographic Information Technologies and Pedometrics'. (Ed S Grunwald) (Taylor and Francis).

Hudson BD (1992). The soil survey as paradigm based science. *Soil Science Society of America Journal* **56**, 836-841.

Jenny H (1941) 'Factors of Soil Formation.' (McGraw-Hill Book Company Inc.: New York).

Lagacherie P, McBratney AB (2007) Spatial Soil Information Systems and Spatial Soil Inference Systems: Perspectives for Digital Soil Mapping. In 'Digital Soil Mapping'. (Eds P Lagacherie, AB McBratney, M Voltz) pp. 3-22. (Elsevier: Amsterdam).

Phillips JD (2001) The Relative Importance of Intrinsic and Extrinsic Factors in Pedodiversity. *Annals of the Association of American Geographers* **94**, 609-621.