Jurnal Ilmu Tanah dan Lingkungan Vol. 7 No.1 (2007) p: 54-67

PREDICTING SOIL PROPERTIES

Budiman Minasny
The University of Sydney, email: b.minasny.at.usyd.edu.au

Abstract

There is a global need for quantitative soil information for environmental monitoring and modelling. Traditional soil survey only has the objective of gathering information on the inventory and distribution of soil classes, and to provide a soil map. However, the increasing demand for the establishment of sustainable agricultural and improvement of land quality has requested soil survey to provide quantitative spatial distribution of soil properties. For example, for sustainable farm planning, a map of available soil water capacity is more useful than a map of soil classes.

Process-simulation models are useful in predicting the outcome of agricultural management on soil quality. The most difficult and expensive step towards the process of modelling is collection of data. Soil properties can be highly variable spatially and temporally and measuring these properties is time consuming and expensive. The data that are mostly available come from soil survey, but survey data only contain basic soil properties such as field morphology, texture, structure, and pH. Therefore it is essential to derive relationships that link the basic soil properties to the functional soil properties that are more difficult to measure. Soil information is expensive to obtain in the field, and also laboratory analysis of soil physical and chemical; properties are expensive. This paper will review and discuss various ways of obtaining and translating soil information.

Key words: pedotransfer, soil properties, soil quality, soil data base

1. Introduction

A useful way of generating soil information based on what we have is called *pedotransfer function* (PTF). The term was coined by Johan Bouma (1989) as *translating data we have into what we need*. Pedotransfer functions allow basic soil information from soil surveys or geographic information system (GIS) data layers to be translated into other more laborious and expensively determined soil properties. Pedotransfer functions are predictive functions of certain soil properties from other easily-, routinely-, or cheaply-measured soil properties. They bridge the gap between the available soil data and the required data.

A new PTF is established by first characterizing basic and specific soil properties. Using these data, hich potentially contribute to a soil database, statistical relationships relating basic to specific soil properties are derived. Once the PTF is formulated, an existing soil survey map can be used to extrapolate the specific soil properties. This framework was proposed by Wösten and Nugroho (1993) for the establishment of a regional soil hydraulic map in Indonesia.

In soil mapping the use of pedotransfer functions (PTFs) is to provide more useful information in terms of soil attributes, soil quality (Reuter, 1998) and soil functioning. There are two approaches on the use pedotransfer functions (see Figure 1). The first is a static one, where pedotransfer functions are used to estimate soil properties. The second, a dynamic approach, predicts those soil properties which will be used as inputs into a simulation or decision-support model. Such models can be used to run scenarios on the effects of different agricultural management on soil functioning.

2. A brief history of pedotransfer functions

Although not formally named until 1989, the concept of the pedotransfer function has long been applied to estimate soil properties that are difficult to determine. In the earliest stage, various ‘rule of thumb’ were formulated to estimate various soil properties. Probably because of the particular difficulty and cost of measurement, the most comprehensive research in developing PTFs has been for the estimation of water retention. The first attempt to use such predictions came from the study of Briggs and McLane (1907), which established a relationship between mechanical composition and the moisture equivalent (amount of water retained after applying centrifugation of 3000 times gravity) of soil based upon data covering 104 soil types.

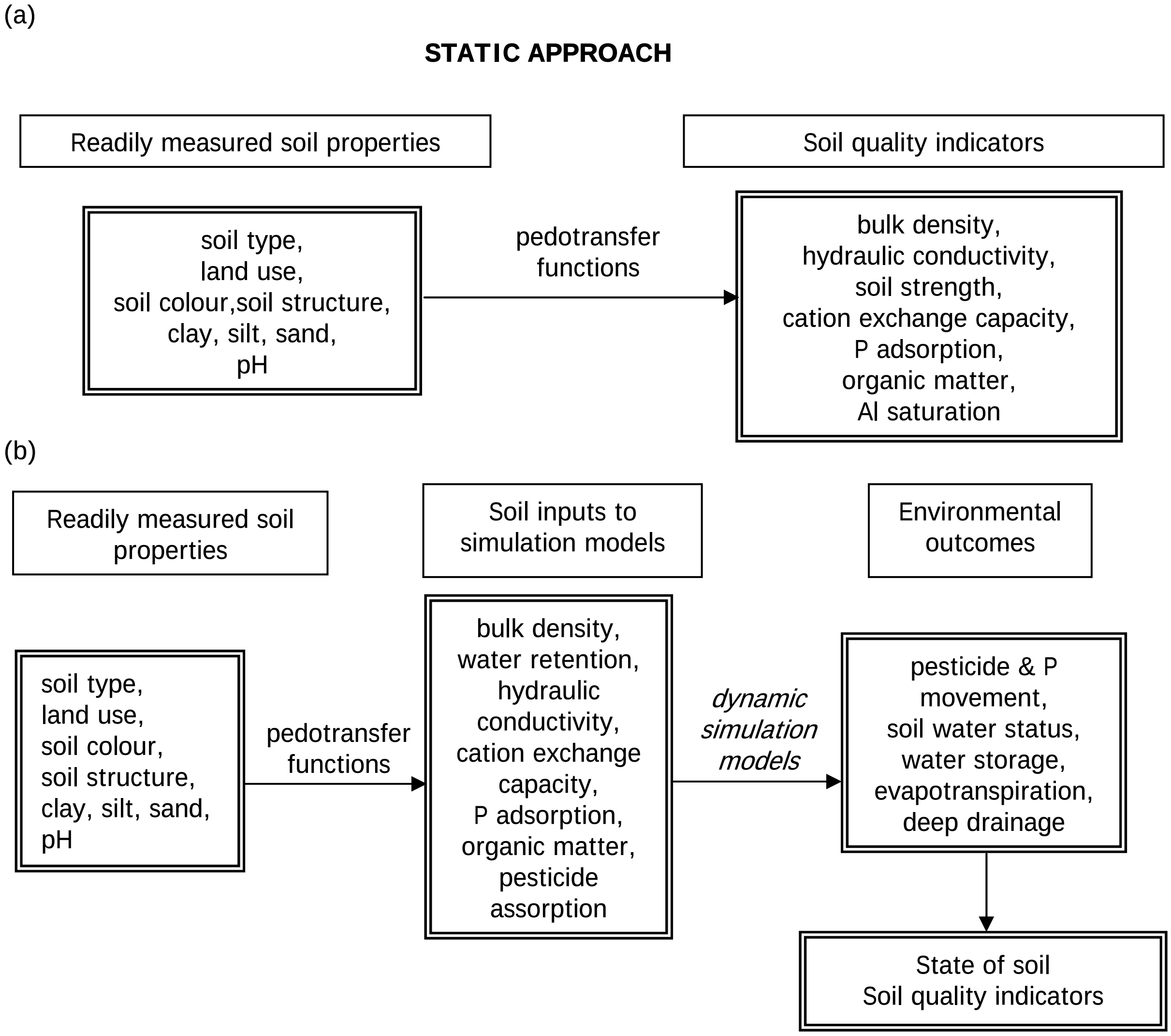

Figure 1. (a) Static vs. (b) dynamic approaches using pedotransfer functions for predicting soil quality

In a further study Briggs and Shantz (1912) determined the wilting coefficient (*percentage water content of a soil when the plants growing in that soil are first reduced to a wilted condition from which they cannot recover in an approximately saturated atmosphere without the addition of water to the soil*) as a function of mechanical composition:

$$\text{Wilting coefficient} = \frac{0.01\,\text{sand} + 0.12\,\text{silt} + 0.57\,\text{clay}}{1 \pm 0.025}$$

With the introduction of the field capacity (FC) and permanent wilting point (PWP) concepts by Veihmeyer and Hendricksen (1927), research during the period 1950-1980 attempted to correlate particle size distribution, bulk density and organic matter content with water content at field capacity (FC, θ at -33 kPa), permanent wilting point (PWP, θ at -1500 kPa), and available water content (AWC = FC - PWP). Nielsen and Shaw (1958) presented a parabolic relationship between clay content and PWP from 730 Iowa soils. A typical example (Burrows and Kirkham, 1958) is estimating field capacity as a function of clay fraction and bulk density (ρ_b Mg m^{-3}):

$$\mathbf{FC = 6.69 + 0.637\,clay + 1.67\,\rho_b\ (R = 0.932)} \qquad [1.1.3]$$

In the 1960s various papers dealt with the estimation of FC, PWP, and AWC, notably in a series of papers by Salter and Williams (1965a, 1965b, 1966, 1967, 1969). They explored relationships between texture classes and available water capacity, which are now known as *class PTFs*. They also developed functions relating the particle-size distribution to AWC, now known as *continuous PTFs*. They asserted that their functions could predict AWC to a mean accuracy of 16 %. Collection of soil data from soil survey allows establishment of national soil database, and development of empirical relationships. De Leenheer and Van Ruymbeke (1960) called their paper “Is it possible to predict some physical soil characteristics, knowing the soil components ?” In the 1970s more comprehensive research using large databases was developed. A particularly good example is the study by Hall *et al.* (1977) from soil in England and Wales; they established field capacity, permanent wilting point, available water content, and air capacity as a function of textural class, and as well as deriving continuous functions estimating these soil-water properties.

Table 1. Example of pedotransfer functions

Predicted soil properties	Predictor variables	Authors
Physical properties		
infiltration rate after certain period	initial water content, moisture deficit, total porosity, non capillary porosity, hydraulic conductivity	Canarache *et al.* (1968)
soil thermal conductivity	texture, organic matter content, water content	De Vries (1966) Hubrechts and Feyen (1996)
bulk density	particle size distribution	Rawls (1983)
Gas diffusivity	Air-filled porosity at -10 kPa	Moldrup et al. (2000)
Mechanical properties		
soil mechanical resistance	organic carbon content, clay content, bulk density	Mirreh and Ketcheson (1972) da Silva and Kay (1997)
Soil shrinkage curve	Clay content	Crescimanno and Provenzano (1999)
volumetric shrinkage, liquid limit, plastic limit, plasticity index	organic matter content, clay content, CEC	Mbagwu and Abeh (1998)
degree of overconsolidation	bulk density, void ratio	McBride and Joose (1996)
rate of structural change	organic matter content, clay content, pH	Rasiah and Kay (1994)
soil erodibility factor	geometric mean particle-size, clay and organic matter content	Torri *et al.* (1997)
Chemical properties		
cation exchange capacity (CEC)	clay content, organic matter content	Bell and van Keulen (1995); Curtin and Rostad (1997)
critical P level, P buffer coefficient	clay content	Cox (1994); Chen *et al.* (1997)
soil organic matter	soil colour	Fernandez *et al.* (1988)
P sorption	pH in NaF	Gilkes and Hughes (1994)
pH buffering capacity	organic matter content, clay content	Helyar *et al.* (1990); Noble *et al.* (1997); Curtin and Rostad (1997)
Al saturation	base saturation, organic carbon content, pH	Jones (1984)
P saturation	extractable P, Al	Kleinman *et al.* (1999)
K/Ca exchange	clay content, extractable K	Scheinost *et al.* (1997a)
nitrogen-mineralization parameters	CEC, total N, organic carbon content, silt and clay content	Rasiah (1995)
As and Cd sorption	clay content, pH , organic carbon content, dithionite extractable Fe	Schug *et al.* (1999)
Phosphorous (P) adsorption	clay content, pH, soil colour	Sheinost and Schwertmann (1995)
Cd sorption coefficient	clay content, organic carbon content, pH	Springob *et al.* (1998)
Haematite content	soil colour	Torrent *et al.* (1983)

In the USA, Gupta and Larson (1979) developed 12 functions relating particle-size distribution and organic matter content to water content at potentials ranging from -4 kPa to -1500 kPa.

With the flourishing development of hydraulic models (van Genuchten, 1980) and computer modelling of soil-water and solute transport (De Wit and van Keulen, 1975), the need for hydraulic properties as input to these models becomes more evident. Clapp and Hornberger (1978) derived average values for the parameters of a power function water retention curve, sorptivity and K_s for different texture classes. In probably the first research of its kind, Bloemen (1980) derived the relationships between parameters of the Brooks and Corey hydraulic model and particle-size distribution.

Lamp and Kneib (1981) formally introduced the term *pedofunction*, while Bouma and van Lanen (1986) used the term *transfer function*. To avoid confusion with the terminology transfer function used in soil physics, and in many other disciplines with many other meanings, Bouma (1989) later called it a *pedotransfer function*.

Since then, the development of hydraulic PTFs has become a boom industry, mostly in the US and Europe. Results of such research have been reported widely, including in the USA (Rawls *et al.*, 1982), the UK (Mayr and Jarvis, 1999), the Netherlands (Wösten *et al.*, 1995), and Germany (Scheinost *et al.*, 1997b).

Although most PTFs have been developed to predict soil hydraulic properties, they are not restricted to hydraulic properties. PTFs for estimating soil physical, mechanical, chemical and biological properties have also been developed (Table 1).

3. Principles

As seen in Table 1, many pedotransfer functions have been developed over the past few decades. At this point, it is useful to define two principles of PTFs, so as to avoid misuse and abuse of the concept. The principles relate to effort and uncertainty.

3.1 Principle 1 - Efficiency

Our first principle of PTFs is: *Do not predict something that is easier to measure than the predictor.*

Since the objective of pedotransfer functions is to predict properties that are difficult or expensive to measure, the predictor should be more easily or cheaper to measured. The cost and effort to obtain the information on the predictor should be much less than that to obtain information on the predicted. In other words, if we define efficiency (Minasny and McBratney, 2002a) as:

Efficiency 1 = quality of information / effort
Efficiency 2 = quality of information / cost of information

the ratio between the efficiency of the predicted over the efficiency of the predictor should be > 1 for justification as an efficient PTF. This also implies that the quality of information gotten out of a PTF should be higher (or the information should be more useful) than the predictor.

This principle extends to the use of existing data to predict values that are missing (not measured). A typical example is prediction of bulk density, although it is known that density is useful for calculation of soil attributes in relation to mass and volume, and is a key predictor of water retention, it is seldom measured. This is mainly because soil survey's main purpose in the past has been to produce a soil class map and most variables analysed are for classification purposes. Although it is far more expensive to determine clay and organic matter content, model predicting bulk density from clay and organic matter is considered as an efficient PTFs. This is because it uses existing soil data in order to predict a missing variable.

However, to predict the saturated hydraulic conductivity (K_s) of a soil from its structural features, as measured by image analysis, would not constitute an efficient PTF. Although there is a good relationship between the image analysis parameters and K_s, it takes more effort to use the image analysis technique, unless the technology improves dramatically. Prediction from field morphology is an example of an efficient PTF however.

3.2 Principle 2 – Uncertainty

Do not use PTFs unless you can evaluate the uncertainty, and for a given problem, if a set of alternative PTFs is available, use the one with minimum variance.

Principle 2 implies two sub-principles, namely,

- *Uncertainty of PTFs should be quantified, and*
- *if a set of alternative PTFs is available, use the one with minimum variance.*

Many PTFs have been developed to predict the same or similar soil properties. For example, in Australia at least 10 functions are available for the prediction of water retention, while worldwide there are more than 100 functions for this property. Therefore, it is wise to choose the function that has the smallest error variance or fit the soil type. Alternatively some Bayesian analysis could be used to provide the most probable estimate, or all competing estimates could be combined with weights inversely proportional to their uncertainty.

The uncertainty of a PTF can be due to the uncertainty of the model, and uncertainty in the input data. The uncertainty associated with the model can be calculated from the non-parametric bootstrap method, or first-order analysis if the PTFs are generated using the least-squares method. The uncertainty of input data can be easily computed using Monte Carlo simulation (see Chapter 25 on Uncertainty). We should also minimise extrapolation of the soil properties.

4. Source of soil information

There are several sources of information that can be used to predict soil properties and considered as pedotransfer function. We will broaden our view on the use of PTFs in accordance with our principles related to effort. The potential predictors can come from sources such as: laboratory, field description and soil morphology, and the soil electromagnetic spectrum.

4.1 Laboratory data

Laboratory analysis in soil survey usually conducted to allocate the soil profile into an existing soil class. The high cost of laboratory analysis incite the development of empirical relationship relating more easily or routinely measured properties to other properties that are more useful. One of the well-known example is the estimation of available water capacity from particle size distribution. The development in pedotransfer functions is boosted by the availability of large soil databases, which allows the use of data mining tools. The most useful variable in predicting soil physical properties is perhaps clay content, as it affects moisture retention, soil strength and many physical and chemical processes. Routine analysis usually lack of physical data. Research is still mainly predicting hydraulic properties, such as water retention and saturated hydraulic conductivity. Some simpler analysis has been utilised to estimate a more difficult to measure properties, such as pH in NaF is an indication of phosphorous sorption capacity (Gilkes and Hughes, 1994).

4.2 Field description and soil morphology

Most research has been focused on correlating laboratory-determined soil properties with more-difficult-to measure properties, mainly because of the availability of comprehensive soil survey databases and the presumption that these properties are most appropriate for predictive purposes. However, it has also been recognised for some time that soil morphological description could be used as predictors (O'Neal, 1949, 1952; McKeague *et al.*, 1984; McKenzie and McLeod, 1989; McKenzie and Jacquier, 1997). This is critical in Australia because most soil survey do not contain detailed laboratory analysis data.

Calhoun *et al.* (2001) contended that soil morphology and field description have been under utilised in the development of pedotransfer functions. They presented the representation of Jenny's state factors through variables: physiography, parent materials, horizon, field texture and structure as collected in soil survey) in predicting bulk density. They demonstrated that morphology and field descriptors account for more variability in predicting bulk density than laboratory measurement of particle size and organic carbon.

The usefulness of physiographic description and soil morphological characterisation was illustrated by Rawls and Pachepsky (2002). They used slope gradient, position of the slope and horizon classes as collected from soil survey data to predict water retention.

Many studies have been devoted to find correlation between soil morphology and hydraulic properties. Several studies have been successful in predicting hydraulic conductivity by using soil morphological features (e.g. O'Neal, 1952; McKeague *et al.*, 1982). However, the descriptive systems and interpretative guidelines in conventional soil survey have been largely qualitative and only appropriate for a given range of soils. McKenzie *et al.* (1991) found that several published descriptive systems for inferring hydraulic properties provided poor predictions for a limited range of soils from South Australia. McKenzie and Jacquier (1997) reasoned that good predictive relationships should only be expected when the field criteria used have a logical physical connection with hydraulic properties. They further postulated that predictive systems that develop direct relationships between hydraulic properties and field criteria of physical significance should be superior to systems that rely on classified entities such as horizons or soil series. They devised a simple visual estimate of areal porosity and found that saturated conductivity can be estimated from field texture, grade of structure, areal porosity, bulk density, dispersion index, and horizon type. Lin *et al*. (1999) also proposed similar system to predict hydraulic properties from field morphology. They converted morphological properties to scores which is related to water flow. From these studies, it was concluded that additional morphological descriptors to those routinely surveyed may be needed to improve the predictive capacity.

4.3 Electromagnetic spectrum, proximal sensing & remote sensing

Electromagnetic spectrum

Developments in spectroscopy, have resulted in an increase in the potential for soil analysis. Electromagnetic radiation involves the interaction of energy with matter, radiation of different energy or frequency is used to incite specific types of electronic excitations, leading to different types of spectroscopy. Electromagnetic radiation has both particle and wave properties, and thus light of a particular wavelength also corresponds to a spatial scale of detection.

Infrared spectroscopy in both the near and mid infrared ranges allows rapid acquisition of soil information. Spectral signatures of soil materials are characterised by their reflectance to a particular wavelength in the electromagnetic spectrum. Investigations by Ben-Dor and Banin (1995), Janik and Skjemstad (1995) and others have demonstrated the ability of the reflectance to provide estimate of several soil physical, chemical and biological properties at a time. Dalal and Henry (1986) evaluated the use of near-infrared diffuse reflectance spectroscopy to predict soil moisture, organic carbon and total nitrogen content. They developed linear regression model by selecting three wavelength values that predict the soil attributes best. Janik *et al.* (2000) offered the mid-infrared spectroscopy for simultaneous estimation of lime requirement, organic carbon, exchangeable cations, air-dry moisture, clay content and biological indicators. Infrared methodology appears to have advantages in facilitating some soil analyses that are time-consuming or expensive. Ludwig *et al.* (2002) assessed the application of near-infrared spectroscopy technique to predict various chemical and biological properties of soil samples collected from two mountain ash sites in Victoria, 10 years after forest harvesting. They found that promising results obtained for C, N, Olsen P, microbial C, cumulative N and C mineralised, and potentially mineralisable C.

Digital soil spectra usually contain hundreds or thousands of reflectance values as a function of wavelength. Since there are more predictor variables than the observation and predicted variables, methods that reduce the dimension of the predictors are required. Principal component regression and partial least squares (PLS) method were commonly utilised. Principal component regression reduces the dimension of the predictors via principal component analysis, and then forming linear regression between the principal components and soil attributes (Martens and Naes, 1989; Chang *et al*., 2001). Partial Least Squares (PLS) (Martens and Naes, 1989) extracts successive linear combinations of the predictors, which optimally address the combined goals of explaining response variation and explaining predictor variation. PLS therefore balances the two objectives of explaining response variation and explaining predictor variation.

Proximal sensing

The developments of the spectroscopy method can be potentially linked to the deployment of 'on-the-go' proximal soil sensing systems or scanners. These sensing systems or scanners can overcome current problems of high cost, labour, time, imprecision of soil sampling to more efficiently and accurately represent the spatial variability of the measured properties (Viscarra Rossel and McBratney, 2002). Sudduth and Hummel (1993) devised a portable near-infrared spectrophotometer to predict soil organic matter content, moisture content and CEC.

Electromagnetic induction and soil resistivity meter attached to a vehicle provides rapid spatially referenced soil electrical conductivity. Soil bulk electrical conductivity reflects a combination of soil mineralogy, salts, moisture and texture, hence it is a good compound measure of soil. Such proximal sensing offer the possibility of producing high resolution maps of soil properties. Regression equations have been developed to predict moisture content, topsoil thickness, and clay content, however the values of ECa reading is a combination of these variables and little research has been done to unscramble these factors.

In the future NIR spectrometers for field use on soil hand samples and profile and core faces. (Such devices are in use by field geologists.) With proper calibration by the use of PTFs field estimates of many soil properties will be improved and estimates of other soil properties *e.g.*, CEC organic C will be able to be made in the field for the first time. Field observation, proximal sensing and PTFs will be fused into a new more powerful approach.

Remote sensing

The value of remote sensing over proximal sensing is that large spatial extents can be covered quickly with many estimates. The inferred value of remotely sensed data either airborne and satellite (Chapters have been shown to be an efficient means of assessing resource condition at reasonably broad scales. The remotely sensed data can include spectral, radar, thermal and radiometric signals. These data reflects the environmental and soil condition and are known to be associated with soil properties. Ben-Dor (2002) gave a review on the application of remote sensing in quantitative assessment of soil properties.

Table 2. Various methods developed for data mining purposes

Multiple regression
The general purpose of multiple regression is to analyse the relationship between several independent or predictor variables and a dependent or predicted variable. Multiple regression analysis fits a straight line (or plane in an n-dimensional space, where n is the number of independent variables) to the data.

Generalised linear models (GLM)
A class of models that arise for a natural generalisation of ordinary linear models. The transformed dependent variable values are predicted from (is linked to) a linear combination of predictor variables; the transformation is referred to as the link function; also, different distributions can be assumed for the dependent variable values.

Generalised additive models (GAM)
Models that use smoothing techniques, such as splines to identify and represent possible nonlinear relationship between the predictor and predicted variables. GAM is a generalisation of GLM where the linear function of the predictor is replaced by an unspecified (non-parametric) function, obtained by applying a scatterplot smoother to the scatterplot of partial residuals (for the transformed dependent variable values).

Partial least squares (PLS)
Alternative to multiple linear regression to deal with data having more independent variables than observation points. PLS constructs a new set of components as regressor variables which are linear combination of the original variables. The components in partial least squares are determined by both the response variable(s) and the predictor variables.

Artificial neural networks
A mathematical structure modelled after the functioning of the nervous system. The essential feature is a network of simple processing elements joined together by weights.

Regression tree
Alternative to multiple regression, rather than fitting a model to the data, a tree structure is generated by dividing the sample recursively into a number of groups, each division being chosen so as to maximise some measure difference in the predicted variable in the resulting two groups. The resulting structure provides easy interpretation as variables most important for prediction can be identified quickly.

Compiled from: Everitt (2002) The Cambridge Dictionary of Statistics, Statistics Glossary http://www.statsoftinc.com/textbook/glosfra.html.

5. Approaches

Most survey agencies have their own 'rule of thumb' for predicting soil properties. Another form is a look-up table, which usually relates field texture class to properties such as clay content, available water capacity, etc. These rules or tables are usually derived from experience, expert knowledge, or from means of properties for particular class in a soil database. For the continuous predicted variables, a plethora of mathematical models can be used to derive PTFs, finding relationship between the predictor and predicted variables. Various methods have been developed for data mining purposes, and there is a benefit in using these tools. Many of the modern techniques are described in Hastie *et al.* (2001). The methods ranges from linear regression, generalised linear models (GLIM), generalised additive models (GAM), regression trees, neural networks and fuzzy systems. Many statistical packages now allow the use of these tools in a user-friendly environment, such as S-Plus (Insightful, 2002), JMP (SAS Institute, 2002). Programming language based software such as Matlab (Mathworks, 2002), R *(http://www.r-project.org)* offer toolboxes that contain many advanced mathematical tools.There are also specific software developed for data-mining purposes, they are usually more powerful and can handle large data sets, however they are usually more expensive than general statistical packages. The predictive power and interpretability varies between models depending on their complexity. Table 3 provides a guideline for various models. The more complex the model, the more parameters it will have, user need to be aware on the principle of parsimony (a general principle that any models, all of which provide an adequate fit for a set of data, the one with the fewest parameters is to be preferred) (Lark, 2000). There is a limit for predictive model; users should choose the simplest model that can adequately account for the variation in prediction. Model with high complexity will appear to fit the data very well, however it may also cause overfitting, or too many parameters in the model, thus the model will fit the noise of the data. It is recommended to split the data into a calibration and validation set, using the calibration data for fitting and then testing or validating the model with a validation set (see Hastie *et al.* (2001) for more detail). Wosten *et al.* (2001) compared that the performance of three models to predict water content at -33 kPa from basic soil properties using the same data set. They revealed that the accuracy of all three methods is similar, and suggested that the improvement of fit may not be expected from the use of different models, but from a better data.

Table 3. Comparison of different mathematical predictive models, ☺ = good, ☹ = poor 😐 = fair (Adapted from Table 10.1 of Hastie *et al.*, 2001).

Feature	Linear models	GLIM	GAM	Regression Tree	Neural Net
Ease of use	☺	😐	😐	☺	☹
Parsimony	☺	😐	☹	☺	☹
Interpretability	☺	😐	☹	☺	☹
Nonlinearity	☹	☹	☺	☺	☺
Handling of mixed data type (Qualitative & quantitative)	😐	☺	☺	☺	☹
Computational efficiency (large data)	☺	☺	😐	☺	☹
Predictive power	☹	😐	😐	😐	☺

6. Using pedotransfer functions

Developing new PTFs is an arduous task, as it requires a large soil database containing many soil measurements. It will be wise, in the first instance, to utilise the functions that have already been developed. But the validity of a given PTF should not be interpolated or extrapolated beyond the pedological origin or soil type on which they are developed. The distinct properties of Australian soil (Williams, 1983) means that PTFs developed elsewhere cannot be directly applied without testing. Testing is required so that the most suitable PTFs to use can be identified. Stratification and calibration of PTFs are also essential. Wosten *et al.* (2001) discussed on the issues of accuracy and reliability of PTFs. Accuracy is defined as the performance of the PTF on the training data while reliability refers to the performance outside the training data.

Some suggestions for using published PTFs can be made as follows: stratification, calibration, and testing. Stratification is needed to establish separate PTFs based on soil type and input information. Stratification has been made variously according to soil horizons (Hall *et al.*, 1977); FAO soil classes (Batjes, 1996), textural classes (Tietje and Hennings, 1996); hydraulic-functional horizons (Wösten *et al.*, 1986); great soil groups, temperature regime, moisture regime (Pachepsky and Rawls, 1999); parent material and horizon morphology (Franzmeier, 1991); numerical soil classification (Williams *et al.*, 1983); management units (Droogers and Bouma, 1997)

Calibration is required to adjust properties measured to the one required as input. Calibration is needed because of the differences in criteria and measurements from existing pedotransfer functions. For example, sand fractions are different according to the ISSS/Australian classification (particle diameter 20-2000 m) and the FAO/USDA criteria (particle diameter 50-2000 m). Minasny and McBratney (2001) gave equations for converting these two types of silt sized fractions. Little (1992), and Bui and Henderson (2003) established relationships between pH measured in water and pH measured in $CaCl_2$.

Testing of published PTFs can be made from available data. Many exotic PTFs are only applicable to certain types of soil where they were calibrated. Papers that evaluate the application of different PTFs (e.g. Tietje and Tapkenhinrichs, 1993; Minasny and McBratney, 2000) may offer a guide to which function to use.

We recommend using a consistent function when applying PTFs onto a profile or across a field. For example in the case where particle size distribution was available for each layer in a soil profile but bulk density was only available in some layers. While there are PTFs that use particle size analysis alone and particle size plus bulk density as inputs, we should consistently use the function that takes particle size plus bulk density as inputs for each layer. For layers with no measurement of bulk density, its value can be predicted from interpolation between layers or predicted from particle size distribution. Using different PTFs across a field or within layers in a profile can cause anomalous values that are due to the dissimilar structure of the functions.

7. Soil Inference Systems

While there are many similar pedotransfer functions generated using new or existing datasets there seems to be much less effort in gathering and using the available PTFs. McBratney *et al.* (2002) proposed the concept of *soil inference system* (SINFERS), where pedotransfer functions are the knowledge rules for soil inference engines. *A soil inference system takes measurements we more-or-less know with a given level of (un)certainty, and infers data that we don't know with minimal inaccuracy, by means of properly and logically conjoined pedotransfer functions.*

Dale *et al.* (1989) discussed the role of expert systems in soil classification, and similar principles can be applied to the proposed inference system. This is illustrated in Figure 2, where the system has a source, an organiser and a predictor. The sources of knowledge to predict soil properties are collections of pedotransfer functions and soil databases. The organiser arranges and categorizes the pedotransfer functions with respect to their required inputs and the soil types from which they were generated. The inference engine is a collection of logical rules selecting the pedotransfer functions with the minimum variance. The rules can simply be a collection of "if-then" statements, or based on probabilistic Bayesian inference. Uncertainty of the prediction can be assessed using Monte-Carlo simulations. The inference system operates through a user interface that will return the predictions of soil physical and chemical properties with their uncertainties based on the information provided.

McBratney *et al.* (2002) demonstrated the first approach towards building a soil inference system is to create a very rudimentary system in the form of a specially-adapted spreadsheet. It has two essentially new features; firstly it contains a suite of published pedotransfer functions. The output of one PTF can act as the input to other functions (if no measured data are available). Secondly, the uncertainties in estimates are inputs and the uncertainties of subsequent calculations are performed. The input consists of the essential soil properties. The inference engine will work in the following manner:

1. Predict all the soil properties using all possible combinations of inputs and PTFs.
2. Select the combination that leads to a prediction with the minimum variance.

We envisioned that a complete soil inference system will be built and can be adapted in a GIS context. The system should have a database of mean soil properties, such as particle-size distribution and organic matter content, for different soil types. The inference engine has knowledge rules which will determine what functions to use realizing their uncertainties. The output will be the predicted physical and chemical properties along with their uncertainties. This can be incorporated into a spatial framework, where a point in space can be predicted from the neighbourhood basic soil properties or soil class.

Bouma (1989) defined pedotransfer functions in terms of data translation. We can describe this translation function as information. This information, when properly and logically conjoined, constitutes knowledge. Knowledge can generate various data. Soil inference systems take measurements we know with a certain precision and infer properties we don't know with given precision, by means of properly and logically conjoined pedotransfer functions.

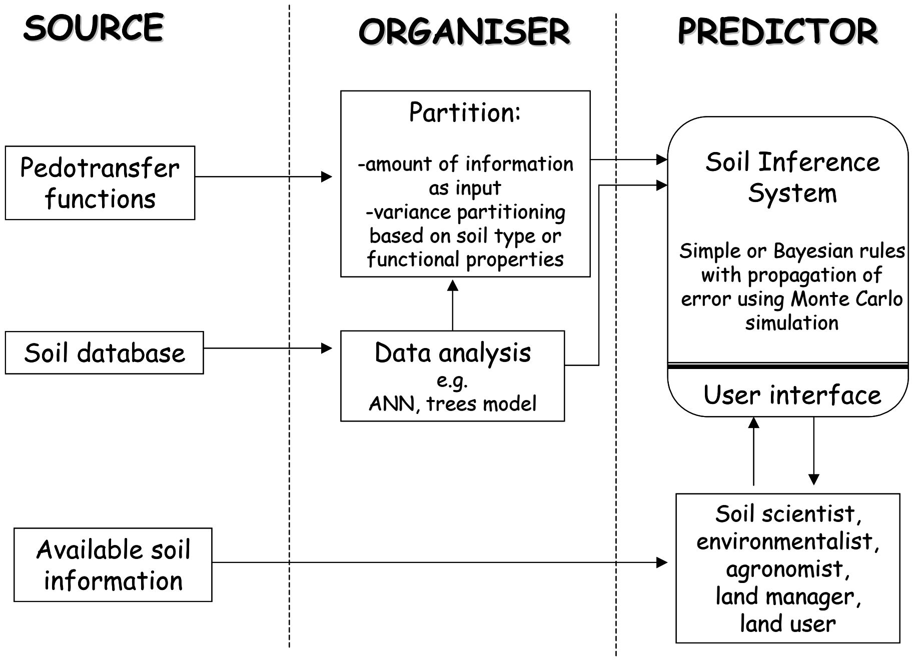

Figure 3. A soil inference system

References

Arya, L. M., Paris, J.F., 1981. A physicoempirical model to predict soil moisture characteristics from particle size distribution and bulk density data. Soil Science Society of America Journal 45, 1023 1030.

Batjes, N.H., 1996. Development of a world data set of soil water retention properties using pedotransfer rules. Geoderma 71, 31-52.

Ben-Dor, E., 2002. Quantitative remote sensing of soil properties. Advances in Agronomy 75, 173-243.

Ben-Dor, E., Banin, A. 1995. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. Soil Science Society of America Journal 59, 364–72.

Boucneau, G., 1998. Geographical Information Science Applied to Soils of West-Flanders. PhD thesis. Faculty of Agricultural and Applied Biological Sciences. University of Ghent.

Bouma, J., 1989. Using soil survey data for quantitative land evaluation. Advances in Soil Science 9, 177-213.

Bui, E., Henderson, B., Moran, C., Johnston, R., 2002. Continental-scale spatial modelling of soil properties. Paper No. 1470. 17th World Congress of Soil Science. 14-21 August 2002, Bangkok, Thailand.

Calhoun, F.G., Smeck, N.E., Slater, B.L., Bigham, J.M., Hall, G.F., 2001. Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. Soil Science Society of America Journal. 65, 811-819.

Campbell, G.S., 1974. A simple method for determining unsaturated hydraulic conductivity from moisture retention data. Soil Science 117, 311-314.

Carlile, P., Bui, E., Moran, C., Simon, D., Henderson, B., 2001. Method used to generate soil attribute surfaces for the Australian Soil Resource Information System using soil maps and look-up tables. CSIRO Land and Water Technical Report 24/01.

Chang, C-W., Laird, D.W., Mausbach, M.J., Hurburgh, C.R., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480-490.

Chen, G., Yost, R.S., Li, Z.C., Wang, X., Cox, F.R., 1997. Uncertainty analysis for knowledge-based decision aids: application to PDSS (Phosphorous Decision Support System). Agricultural Systems 55, 461-471.

Colwell, J.D., 1977. National Soil Fertility Project. CSIRO Division of Soils in collaboration with state departments of agriculture and the fertilizer industry.

Crescimanno, G., Provenzano, G., 1999. Soil shrinkage characteristic curve in clay soils. Soil Science Society of America Journal 63, 25-32.

Cresswell, H. P., Paydar, Z., 1996. Water retention in Australian soils. I. Description and prediction using parametric functions. Australian Journal of Soil Research 34, 195-212.

Cresswell, H., Pierret, C., Brebner, P., Paydar, Z., 2002. The SH-Pro V1.03 software for predicting and analysing soil hydraulic properties. CSIRO Land and Water, Canberra.

Cresswell, H.P.,.McKenzie, N.J, Paydar, Z., 1999. A strategy for determination of hydraulic properties of Australian soil using direct measurement and pedotransfer functions. In: van Genuchten, M.Th., Leij, F.J., Wu, L. (Eds). Characterization and Measurement of the Hydraulic Properties of Unsaturated Porous Media. pp. 1143-1159. University of California: Riverside, CA.

Da Silva, A., Kay, B.D., 1997. Estimating the least limiting water range of soils from properties and management. Soil Science Society of America Journal 61, 877-883.

Dalal RC, Henry RJ. 1986. Simultaneous determination of moisture, organic carbon and total nitrogen by near infrared reflectance spectrometry. Soil Science Society of America Journal 50, 120-123.

Dale, M.B., McBratney, A.B., Russell, J.S., 1989. On the role of expert systems and numerical taxonomy in soil classification. Journal of Soil Science 40, 223-234.

Dane, J.H., Topp, G.C. (Eds.) 2002. Methods of Soil Analysis Part 4 – Physical Methods. SSSA Book Seties 5. Soil Science Society of America, Madison, Wisconsin.

De Leenheer and Van Ruymbeke (1960) Is it possible to predict some physical soil characteristics, knowing the soil components? 7th International Congress of Soil Science, Madison, Wisc. 1.25 272-278.

Droogers, P., Bouma, J., 1997. Soil survey input in exploratory modelling of sustainable soil management practices. Soil Science Society of America Journal 61, 1704-1710.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Monographs on Statistics and Applied Probability 57. Chapman & Hall, London, UK.

Forrest, J.A., Beatty, J., Hignett, C.T., Pickering, J.H., Williams, R.G.P., 1985. A Survey of the Physical Properties of Wheatland Soils in Eastern Australia. CSIRO Australia Divison of Soils, Divisional Report No. 78.

Franzmeier, D.P., 1991. Estimation of hydraulic conductivity from effective porosity data for some Indiana soils. Soil Science Society of America Journal 55, 1801-1803.

Gilkes, R.J., Hughes, J.C., 1994. Sodium fluoride pH of south-western Australian soils as an indicator of P-sorption. Australian Journal of Soil Research 32, 755-766.

Hall, D.G., Reeve, M.J., Thomasson, A.J., Wright, V.F., 1977. Water Retention, Porosity and Density of Field Soils. Technical Monograph No. 9. Soil Survey of England and Wales, Harpenden.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning: Data Mining, Inference and Prediction. Springer Series in Statistics. Springer-Verlag, New York.

Helyar, K.R., Cregan, P.D., Godyn, D.L., 1990. Soil acidity in New south Wales - current pH values and estimates of acidification rate. Australian Journal of Soil Research 28, 523-527.

Henderson, B. L., Bui, E. N., 2003. An improved calibration curve between soil pH measured in water and CaCl2. Australian Journal of Soil research (In Press).

Heuvelink, G.B.M., 1998. Error Propagation in Environmental Modelling with GIS. Taylor & Francis, London, Bristol.

Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modeling. Geoderma 89, 47–65.

Insightful. 2001. S-Plus6 for Windows User's Guide. Insightful Coporation, Seattle, WA. http://www.insightful.com

Janik, L J., Skjemstad, J.O., Raven, M.D. 1995. Characterization and analysis of soils using mid-infrared partial least squares. I. Correlations with XRF-determined major element composition. Australian Journal of Soil Research 33, 621–636.

Janik, L., Skjemstad, J.O. 1995. Characterization and analysis of soils using mid-infrared partial least squares. II. Correlations with some laboratory data. Australian Journal of Soil Research 33, 637–650.

Lark, R.M., 2001. Some tools for parsimonious modelling and interpretation of within-field variation of soil and crop systems, Soil and Tillage Research 58, 99-111.

Lin, H.S., McInnes, K.J., Wilding, L.P., Hallmark, C.T., 1999. Effects of soil morphology on hydraulic properties: I. Quantification of soil morphology. Soil Science Society of America Journal 63, 948-954.

Little, I.P., 1992. The relationship between soil pH measurements in calcium chloride and water suspensions. Australian Journal of Soil Research 30, 587-592.

Littleboy, M. 1998. Spatial generalisation of biophysical simulation models for quantitative land evaluation: A case study for dryland wheat growing areas of Queensland. PhD Thesis. The University of Queensland.

Loch, R.J., B.K. Slater and C. Devoil. 1998. Soil erodibility (Km) values for some Australian soils. Australian Journal of Soil Research 36:1045-1055.

Ludwig, B., Khanna, P. K , Bauhus, J. Hopmans, P.2002. Near infrared spectroscopy of forest soils to determine chemical and biological properties related to soil sustainability, Forest Ecology and Management 171,121-132

Martens, G., Naes, T., 1989. Multivariate Calibration. John Wiley and Sons, New York.

Mathworks, 2002. Matlab Release 13. MathWorks Inc., Natick, MA. http://www.mathworks.com

McBratney, A.B., Minasny, B., Cattle, S.R., Vervoort, R.W., 2002. From pedotransfer function to soil inference system. Geoderma 109, 41-73.

McGarry, D., Ward, W.T., McBratney, A.B., 1989. Soil Studies in the Lower Namoi Valley: Methods and Data. The Edgeroi Data Set. CSIRO Division of Soils, Glen Osmond, South Australia.

McKeague, J.A., Wang, C., Topp, G.C. 1982. Estimating saturated hydraulic conductivity from soil morphology. Soil Science Society of America Journal 46, 1239-1244.

McKenzie N.J., Jacquier, D.W., 1997 Improving the field estimation of saturated hydraulic conductivity in soil survey. Australian Journal of Soil Research 35, 803–825.

McKenzie, N.J., Coughlan, K., Cresswell, H.P. (Eds.) 2002. Soil Physical Measurement and Interpretation for Land Evaluation. CSIRO Publisihing, Collingwood, Victoria.

McKenzie, N.J., Jacquier, D.W., Ashton L.J., Cresswell, H.P. 2000 Estimation of Soil Properties Using the Atlas of Australian Soils. CSIRO Land and Water Technical Report 11/00.

McKenzie, N.J., MacLeod, D.A. 1989. Relationships between soil morphology and soil properties relevant to irrigated and dryland agriculture. Australian Journal of Soil Research 27, 235-258.

McKenzie, N.J., Smettem, K.R.J., Ringrose-Voase, A.J., 1991. Evaluation of methods for inferring air and water properties of soils from field morphology. Australian Journal of Soil Research 29, 587-602.

Minasny B., McBratney, A.B., 2000. Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. Australian Journal of Soil Research 38, 905-926.

Minasny B., McBratney, A.B., Bristow, K.L., 1999. Comparison of different approaches to the development of pedotransfer functions for water-retention curves. Geoderma 93, 225-253.

Minasny, B., McBratney, A.B., 2001. The Australian soil texture boomerang: a comparison of the Australian and USDA/FAO soil particle-size classification systems. Australian Journal of Soil Research 39, 1443-1451.

Minasny, B., McBratney, A.B., 2002a. The efficiency of various approaches to obtaining estimates of soil hydraulic properties. Geoderma 107, 55-70.

Minasny, B., McBratney, A.B., 2002b. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Science Society of America Journal 66, 352-361.

Minasny, B., McBratney, A.B., 2002c. Neuropack: Neural networks package for fitting pedotransfer functions. Australian Centre for Precision Agriculture, the University of Sydney. Avalaible at: http://www.usyd.edu.au/su/agric/acpa

Noble, A.D., Cannon, M., Muller, D. 1997. Evidence of accelerated soil acidification under Stylosanthes-dominated pastures. Australian Journal of Soil Research 35,1309-1322.

Nugroho, K., Wösten, J.H.M., 1993. Use of soil survey data for the determination of soil hydraulic function. Indonesian Journal of Crop Science 8, 39-47.

O'Connell, D.A., Ryan, P.J., 2002. Prediction of three key hydraulic properties in a soil survey of a small forested catchment. Aust. J. Soil Res. 40, 191–206.

O'Neal, A.M. 1949. Some characteristics significant in evaluating permeability. Soil Science 67, 403-409.

O'Neal, A.M. 1952. A key for evaluating soil permeability by means of certain field clues. Soil Science Society of America Proceedings 16, 312-315.

Pachepsky, Y.A., Timlin, D.J., Ahuja, L.R., 1999. The current status of pedotransfer functions: their accuracy, reliability and utility in field- and regional-scale modeling. In: Corwin, D.L., Loage, K., Ellsworth, T.R. (Eds.) Assessment of non-point source pollution in vadose zone. Geophysical Monograph 108. pp. 223-234. American Geophysical Union, Washington, D.C.

Pachepsky, Y.A., Timlin, D.J., Rawls, W.J., 2001. Soil water retention as related to topographic variables. Soil Science Society of America Journal 65, 1787-1795.

Pachepsky, Ya. A., Rawls, W.J., 1999. Accuracy and reliability of pedotransfer functions as affected by grouping soils. Soil Science Society of America Journal 63, 1748-1756.

Paydar, Z., Cresswell, H.P., 1996. Water retention in Australian soils. II Prediction using particle-size, bulk density and other properties. Australian Journal of Soil Research 34, 679-693.

Prebble, R.E., 1970. Soil physical measurements and grass roots distributions on a red podzolic at Samford, South East Queensland. CSIRO Division of Soils Technical Memorandum 13/70. CSIRO, Australia.

Rawls, W.J., Gish, T.J., Brakensiek, D.L., 1991. Estimating soil water retention from soil physical properties and characteristics. Advances in Soil Science 16, 213-234.

Rawls, W.J., Pachepsky, Y.A., 2002. Using field topographic descriptors to estimate soil water retention. Soil Science 167, 423-435.

Reuter, D.J., 1998. Developing indicators for monitoring catchment health: the challenges. Aust. J. Experimental Agriculture 38,637-648.

Romano, N. Palladino, M. 2002. Prediction of soil water retention using soil physical data and terrain attributes. Journal of Hydrology 265, 56-75.

Romano, N., Santini, A., 1997. Effectiveness of using pedo-transfer functions to quantify the spatial variability of soil water retention characteristics. Journal of Hydrology 202, 137-157.

SAS Institute. JMP Statistics and Graphics Guide, Version 5. SAS Institute Inc., Cary, NC. http://www.jmpdiscovery.com

Schaap, M.G., Leij, F.L., van Genuchten, M.Th., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Science Society of America Journal 62, 847-855.

Scheinost, A.C., Schwertmann, U., 1995. Predicting phosphate adsorption-desorption in a soilscape. Soil Science Society of America Journal 59, 1575-1580.

Scheinost, A.C., Sinowski, W., Auerswald, K., 1997a. Regionalization of soil buffering functions: A new concept applied to K/Ca exchange curves. Advances in GeoEcology 30, 23-38.

Scheinost, A.C., Sinowski, W., Auerswald, K., 1997b. Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. Geoderma 78, 129-143.

Sinowski, W., Scheinost, A.C., Auerswald, K 1997. Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. Geoderma 78, 145-159.

Smettem, K.J.R., Gregory, P.J., 1996. The relation between soil water retention and particle size distribution parameters for some predominantly sandy Western Australian soils. Australian Journal of Soil Research 34, 695-708.

Smettem, K.R.J., Bristow, K.L., 1999. Obtaining soil hydraulic properties for water balance and leaching models from survey data 2. Hydraulic conductivity. Australian Journal of Agricultural Research 50, 1259-1262.

Smettem, K.R.J., Oliver, Y.M., Heng, L.K., Bristow, K.L., Ford, E.J., 1999. Obtaining soil hydraulic properties for water balance and leaching models from survey data 1. Water retention. Australian Journal of Agricultural Research 50, 283-289.

Springer, E.P., Cundy, T.W., 1987. Field-scale evaluation of infiltration parameters from soil texture for hydrologic analysis. Water Resources Research 23, 325-334.

Stirk, G.B., 1957. Physical properties of soils of the lower Burdekin valley, North Queensland. CSIRO Division of Soils Divisional Report 1/57. CSIRO, Australia.

Sudduth, K.A., Hummel, J.W., 1993. Soil organic matter, CEC (cation exchange capacity), and moisture sensing with a portable NIR spectrophotometer. Transactions of the ASAE. 36, 1571-1582.

Tietje, O. Hennings, V. 1996. Accuracy of the saturated hydraulic conductivity prediction by pedo-transfer functions compared to the variability within FAO textural classes. Geoderma 69,71-84.

Tietje, O., Tapkenhinrichs, M. 1993. Evaluation of pedo-transfer functions. Soil Science Society of America Journal 57, 1088-1095.

Viscarra Rossel, R.A., McBratney, A.B. 1998. Soil chemical analytical accuracy and costs: implications from precision agriculture. Australian Journal of Experimental Agriculture 38, 765–775.

Williams, J., Ross, P. J., Bristow, K. L., 1992. Prediction of the Campbell water retention function from texture, structure and organic matter. In: van Genuchten, M. Th., Leij, F. J., Lund, L. J. (Eds.) Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. pp. 427-441. University of California, Riverside, CA.

Wösten, J. H. M., Bannink, M.H., De Gruijter, J.J., Bouma, J., 1986. A procedure to identify different groups of hydraulic conductivity and moisture retention curves for soil horizons. Journal of Hydrology 86, 133-145.

Wösten, J.H.M., 1997. Pedotransfer functions to evaluate soil quality. In: Gregorich, E.G., Carter, M.R. (Eds.) Soil Quality for Crop Production and Ecosystem Health. Developments in Soil Science 25. pp. 221-245. Elsevier, Amsterdam.

Wösten, J.H.M., Finke, P.A., Jansen, M.J.W., 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66, 227-237.

Wösten, J.H.M., Pachepsky, Y.A., Rawls, W.J., 2001. Pedotransfer functions: bridging gap betwen available basic soil data and missing soil hydraulic characteristics. Journal of Hydrology 251, 123-150.

Zeiliguer, A.M., Pachepsky, Ya.A., Rawls, W.J. 2000. Estimating water retention of sandy soils using the additivity hypothesis. Soil Science 165, 373-383.

Φ