*Soil Use and Management*, December 2007, **23,** 437–443 doi: 10.1111/j.1475-2743.2007.00092.x

# Building and testing conceptual and empirical models for predicting soil bulk density

G. Tranter[1], B. Minasny[1], A. B. Mcbratney[1], B. Murphy[2], N. J. Mckenzie[3], M. Grundy[4] & D. Brough[4]
[1]*Faculty of Agriculture, Food and Natural Resources, University of Sydney, Sydney, NSW, Australia,* [2]*New South Wales Department of Infrastructure, Planning & Natural Resources, Cowra, NSW, Australia,* [3]*CSIRO Land and Water, Canberra, ACT, Australia, and* [4]*Department of Natural Resources and Mines, Indooroopilly, Qld, Australia*

**Abstract**

The development of pedotransfer functions offers a potential means of alleviating cost and labour burdens associated with bulk-density determinations. As a means of incorporating *a priori* knowledge into the model-building process, we propose a conceptual model for predicting soil bulk density from other more regularly measured properties. The model considers soil bulk density to be a function of soil mineral packing structures ($\rho_m$) and soil structure ($\Delta\rho$). Bulk-density maxima were found for soils with approximately 80% sand. Bulk densities were also observed to increase with depth, suggesting the influence of over-burden pressure. Residuals from the $\rho_m$ model, hereby known as $\Delta\rho$, correlated with organic carbon. All models were trained using Australian soil data, with limits set at bulk densities between 0.7 and 1.8 g cm$^{-3}$ and containing organic carbon levels below 12%. Performance of the conceptual model ($r^2 = 0.49$) was found to be comparable with a multiple linear regression model ($r^2 = 0.49$) and outperformed models developed using an artificial neural network ($r^2 = 0.47$) and a regression tree ($r^2 = 0.43$). Further development of the conceptual model should allow the inclusion of soil morphological data to improve bulk-density predictions.

**Keywords:** Bulk density, pedotransfer function, soil structure, soil compaction, data mining

## Introduction

The consequences of land pressure and land degradation in Australia have led to an inexorable increase in the need for land-resource information. Data on land resources are required for numerous biophysical models as used by many governmental agencies. However, often the financial and labour demands have made the collection of soil data, in particular hydrologic and chemical properties, unviable. In particular, soil bulk density is regularly overlooked in traditional soil surveys despite its importance in relation to plant growth and as a key parameter in numerous biophysical models. This presents a significant problem in such a large country as Australia given its dependence on natural resources through agriculture, mining and tourism.

In recent years, numerous efforts have been made to alleviate the financial and labour costs associated with bulk-density data determinations. While some have focused on lowering the cost of data capture (Grunwald *et al.*, 2001), others have investigated the use of existing or legacy data through predictive functions. These predictive functions, termed pedotransfer functions (PTFs), use existing or easily acquired data to predict a property that is otherwise expensive or difficult to measure, the main premise being the characterization of dependence structures between response variables and predictor variables (Vereecken & Herbst, 2004). We propose a simple conceptual model for bulk-density prediction as a means of incorporating *a priori* knowledge into model development and allow a posterior analysis to isolate the dominant factors affecting soil bulk density. Whilst the proposed model does not promise dramatic improvements in accuracy compared with more conventionally derived bulk-density PTFs, it does provide a more interpretable and rational form, allowing functional processes to be discerned. For comparison, functions derived from data-mining techniques were also developed.

*A priori* knowledge has enabled the development of models for predicting bulk density from organic matter measures in largely undisturbed forest soils, high in organic matter (Curtis & Post, 1964; Jeffrey, 1970; Stewart *et al.*, 1970; Adams, 1973). To a lesser extent Williams (1970) found that

Correspondence: G. Tranter. E-mail: gtra5750@mail.usyd.edu.au
Received October 2006; accepted after revision March 2007

 

up to 50% of variation in soil bulk density could be accounted for by variation in organic carbon content on arable and grasslands soils. In soils of low organic matter content Heinonen (1977) established a relationship between bulk density and clay and silt-sized fractions. Using input data on organic carbon, organic matter and texture, Rawls (1983), Bernoux *et al.* (1998) and De Vos *et al.* (2005) built more sophisticated models for bulk-density prediction.

Despite significant development of pedotransfer functions, a bulk-density function for Australian conditions is yet to be developed. Improved access to soil data through government soil databases means that there is now the opportunity for the development of functions specific to Australian soils. The aims of this study are to: (1) develop pedotransfer functions to predict soil bulk density for Australian soils; (2) introduce and evaluate a proposed soil bulk-density conceptual model; and (3) identify and explore the functional processes affecting soil bulk density.

## A conceptual model for soil bulk density

The development of sophisticated statistical procedures and automated input selection methods has arguably led to increasingly complex biophysical models. Data-driven selection methods, in particular, have received criticism due to the perception that such methods overlook *a priori* knowledge for purely empirical relationships. Despite criticisms of data-driven selection methods (Judd & McClelland, 1989; Roecker, 1991), stepwise selection techniques do provide users with valuable data-mining power if coupled with expert knowledge or as stated by Judd & McClelland (1989, p. 204), 'It is our experience and strong belief that better models and a better understanding of one's data result from focussed data analysis, guided by substantive theory'. As a method of incorporating expert knowledge into model development, we propose the use of a conceptual model for the prediction of soil bulk density. The complex and multiplicative nature of bulk-density prediction makes the use of a conceptual model logical. Furthermore, the use of the conceptual model aims to present a model that is accurate and interpretable, delivering both predictions and information about the system.

Heinonen (1977) found that bulk density for unstructured soil could be predicted ($r^2 = 0.62$) using the clay and silt contents. Similarly, studies by Koltermann & Gorelick (1995) used variates of particle-packing models to accurately predict porosity of unconsolidated clastic sediments. Gupta & Larson (1979) found that the accuracy of packing models for natural soils based on particle size fractions was limited due to the various factors influencing packing arrangements. Childs (1969) considered soil porosity to be the combined effect of textural and structural pore spaces. Childs described textural pore space as those pores existing between individual soil particles as reflected in particle size distribution, whilst structural pore space is inter-aggregate space and related to aggregate position and orientation. On the basis of previous work, we consider soil bulk density to result from particle packing and soil structure. This forms the basis for the proposed conceptual model of the form

$$\rho_b = \rho_m + \Delta\rho + \varepsilon \quad (1)$$

where $\rho_b$ is the soil bulk density (g cm$^{-3}$), $\rho_m$ the typical bulk density of mineral matter $\approx f$ (particle size distribution, depth) (g cm$^{-3}$), $\Delta\rho$ the bulk-density variation associated with the structural component $\approx f$ (e.g. organic carbon, exchangeable sodium percentage and tillage) (g cm$^{-3}$) and $\varepsilon$ is the residual.

Not to be confused with particle density, the component $\rho_m$ responds to variation in soil bulk density associated with depth and particle size distribution and can be as a typical bulk density for a soil with a given particle size distribution, depth and average structural features. Similarly, Rawls (1983) observed a relationship between soil texture and mineral matter bulk density. The effect of particle size fractions on soil bulk density is well known, with numerous functions developed using differing components (Manrique & Jones, 1991; Bruand *et al.*, 2003; De Vos *et al.*, 2005) although mostly used in conjunction with organic matter variables. The $\rho_m$ model can be developed using standard data-mining procedures such as linear regression and neural networks.

The $\Delta\rho$ component is introduced to account for bulk-density variation as a function of soil structural change. Although quantifiable relationships between soil bulk density and soil structure remain elusive, soil properties associated with structural development or degradation offer potential as proxy structure predictors. With reference to well-accepted pedological relationships, a number of properties can be identified as prospective predictors, namely organic carbon, exchangeable sodium, calcium:magnesium ratio amongst others. Notably, organic carbon and organic matter have been used extensively in previous bulk-density PTFs, demonstrating a strong negative correlation (Curtis & Post, 1964; Adams, 1973; Alexander, 1980; Rawls, 1983).

In this study we consider the lack of fit observed in the $\rho_m$ model as a function of soil structure. For example, the $\rho_m$ residuals observed in this study were found to have a strong negative correlation with organic carbon, a widely accepted factor in soil structural development. Using standard PTF techniques the model aims to predict $\rho_m$ residuals using predictors associated with soil structure. The addition of the two components provides the final bulk-density prediction.

## Materials and methods

Functions were developed from available data sourced from the Soil and Land Information (SALI) database provided by Natural Resources and Mining, Qld, Soil and Land Information System (SALIS) supported by the Department of

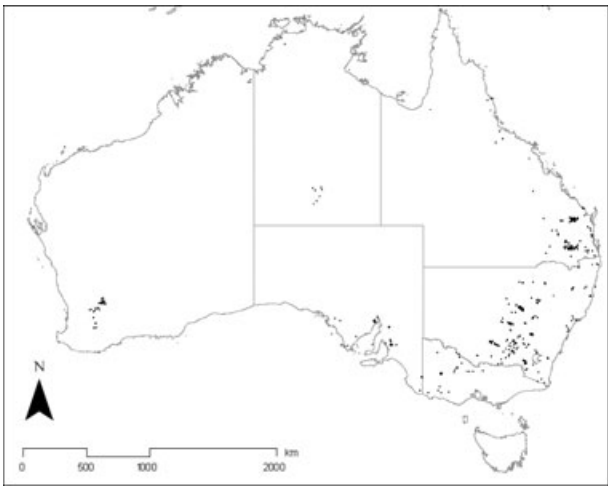


**Figure 1** Location of sampling units.

Natural Resources, NSW, in addition to data supplied by the Australian Collaborative Land Evaluation Program (ACLEP). As shown in Figure 1, the data represent soil from mainly agricultural land from Australia's eastern states of Queensland, New South Wales and Victoria.

### *Data preparation*

Based on *a priori* knowledge and an exploratory analysis, such as scatterplots, a number of potential input parameters were identified. Preliminary analysis identified organic carbon, particle size distribution and depth as candidates for further study. Organic matter contents were also required as an input for a model adapted from Stewart *et al*. (1970) and calculated indirectly using the widely accepted coefficient of 1.72. An upper limit of 12% organic carbon was set, so that samples with organic carbon levels greater than this were excluded as they represent dubious data or soil conditions rarely found in Australia. Bulk density lower and upper limits were placed at 0.7 and 1.8 g $cm^{-3}$, as samples outside this range were considered either dubious or exceptional and consequently were excluded from the analysis. Where reliable data on depth, particle size distribution and bulk density were available, a data set of 1896 samples was isolated, henceforth referred to as set A. The data were split to produce a training set of 926 samples for model construction and a validation set of 481 samples for model evaluation and validation. As shown in Table 1, the data represent a large range of soil textures and soil bulk-density values, generally low in organic carbon.

A second data set of 546 soil sampling units was isolated, henceforth referred to as set B on the basis of data availability for particle size distribution, organic carbon, depth and soil bulk density. This data set was isolated to model the effect of organic carbon on soil bulk density. The subset was divided randomly into two parts, a training set (357 samples) and a validation set (189 samples).

### *Development of bulk-density pedotransfer functions*

Despite the promise of improved model flexibility yielding superior model performance, artificial neural network (ANN) or regression tree (RT)-based models have yet to be developed for bulk-density prediction. The development of ANN- and RT-sourced models in this study is the basis for assessing the effect of model complexity on model accuracy and efficiency.

Reviews of previous work (Minasny *et al*., 2006) show that apart from Heinonen (1977), no functions solely dependent

**Table 1** Descriptive statistics for training and validation data sets

| | Training set ($N = 926$) | | | | Validation set ($N = 481$) | | | |
|---|---|---|---|---|---|---|---|---|
| Property | Mean | Min | Max | SD | Mean | Min | Max | SD |
| Set A | | | | | | | | |
| $\rho_b$ (g $cm^{-3}$) | 1.43 | 0.72 | 1.80 | 0.19 | 1.43 | 0.70 | 1.80 | 0.19 |
| $P_{<2\mu m}$ (dag $kg^{-1}$) | 38.5 | 1.0 | 85.0 | 18.9 | 40.1 | 0.5 | 89.0 | 19.5 |
| $P_{2-20\mu m}$ (dag $kg^{-1}$) | 15.7 | 0.5 | 74.0 | 9.5 | 15.6 | 0.1 | 53.7 | 9.3 |
| $P_{20-2000\mu m}$ (dag $kg^{-1}$) | 45.0 | 2.0 | 97.9 | 21.4 | 43.0 | 2.0 | 99.4 | 21.2 |
| | Training set ($N = 357$) | | | | Validation set ($N = 189$) | | | |
| Set B | | | | | | | | |
| $\rho_b$ (g $cm^{-3}$) | 1.37 | 0.72 | 1.80 | 0.23 | 1.40 | 0.75 | 1.80 | 0.21 |
| Organic carbon (% (g $g^{-1}$)) | 1.37 | 0.04 | 11.80 | 1.68 | 1.21 | 0.04 | 11.10 | 1.46 |
| $P_{<2\mu m}$ (dag $kg^{-1}$) | 34.5 | 1.0 | 79.0 | 20.8 | 36.0 | 0.5 | 86.0 | 20.3 |
| $P_{2-20\mu m}$ (dag $kg^{-1}$) | 15.9 | 0.5 | 74.0 | 9.7 | 15.9 | 0.1 | 45.7 | 9.6 |
| $P_{20-2000\mu m}$ (dag $kg^{-1}$) | 48.0 | 5.7 | 97.9 | 25.2 | 45.8 | 2.7 | 99.4 | 23.8 |

$P_{<2\mu m}$ clay, $P_{2-20\mu m}$ silt, $P_{20-2000\mu m}$ sand.

on depth and particle size have been published. Most work in recent times has used the addition of organic carbon or organic matter data as both have proved to have strong correlation with soil bulk density. However, the relative abundance of data on particle size and depth makes the development of models based on these data practical.

Set A training data were used to develop PTFs (Table 2) based on ANNs, RTs and multiple linear regression (MLR). Regression trees were derived using Cubist 2.0 software (Rulequest Research, http://www.rulequest.com/cubist-info.html). Set B training data were used to develop models based on particle size, depth and organic carbon using ANN, RT and MLR data-mining techniques. The proposed conceptual model was developed and compared with traditionally derived models (Table 2). In addition, a variation of the model proposed by Stewart *et al.* (1970) was included. The original model was developed for unstructured soils experiencing large organic matter contents and is of the form:

$$\rho_b = \frac{100}{(\mathrm{OM}\%/K_1) + [(100 - \mathrm{OM}\%)/K_2]}$$

where OM% is the organic matter percentage (g $g^{-1}$), $K_1$ the bulk density of organic matter (g $cm^{-3}$), $K_2$ the bulk density of mineral matter (g $cm^{-3}$).

Gupta & Larson (1979) proposed an adaptation to the model so that a packing model was used to predict $K_2$. Similarly, Rawls *et al.* (2004) developed a contour map of mineral bulk densities overlain on the USDA texture triangle to obtain $K_2$ estimates. In this study we adapt this model so that a function dependent on particle size distribution and depth, previously introduced as $\rho_m$, replaces the mineral matter bulk-density constant. With this adaptation the model is of the form:

$$\rho_b = \frac{100}{(\mathrm{OM}\%/\rho_{OM}) + [(100 - \mathrm{OM}\%)/\rho_m]}$$

where OM% is the organic matter percentage (g $g^{-1}$), $\rho_{OM}$ the organic matter bulk density = 0.224 g $cm^{-3}$ (Rawls *et al.*, 2004) and $\rho_m$ the mineral matter bulk density (g $cm^{-3}$).

### *Model validation and performance comparison*

Depth and particle size-dependent models were validated using set A data. Likewise, those models derived using the set B training data were validated using set B data.

The accuracy of all models was evaluated using $r^2$ and root mean-squared error (RMSE) performance criteria.

## Results and discussion

### *Regression, regression trees and artificial neural networks*

The performances of all models are listed in Table 3. A stepwise regression analysis could not prove significant

**Table 2** Pedotransfer function details, parameter number indicator of model complexity

| Model Ref. | Model (form) | Predictor Variables | Coefficients | Parameter No. |
|---|---|---|---|---|
| Set A | | | | |
| MLR-A | Multiple linear regression $\rho_b = a + b \times \mathrm{sand} + (c - \mathrm{sand})^2 d + e \log\ \mathrm{depth}$ | | $a = 1.35\ b = 0.0045\ c = 44.7\ d = -6 \times 10^{-5}\ e = 0.060$ | 6 |
| ANN-A | Artificial neural network (3-node) | Sandsiltclaylog depth | N/A | 19 |
| RT-A | Regression tree (Cubist 2.0) | Sandsiltclaylog depth | N/A | 11 |
| Set B | | | | |
| MLR-B1 | Multiple linear regression $\rho_b = a + b \times \mathrm{sand} + c \times \mathrm{OC} + (d + \mathrm{sand})^2 + e + f \log\ \mathrm{depth}$ | | $a = 1.20\ b = 0.0021\ c = -0.143\ d = -47.95\ e = 6 \times 10^{-5}\ f = -0.043$ | 7 |
| ANN-B1 | Artificial neural network (3-node) | Sandorganic carbonlog depth | N/A | 16 |
| ANN-B2 | Artificial neural network (6-node) | Claysiltsandorganic carbonlog depth | N/A | 23 |
| RT-B1 | Regression tree (Cubist 2.0) | Sandorganic carbonlog depth | N/A | 19 |
| Conceptual models | | | | |
| AS-B2 | Adapted Adams–Stewart model | Organic mattersanddepth | $\rho_{OM} = 0.224$ | 6 |
| $\rho_m$ | Multiple linear regression $\rho_b = a + b \times \mathrm{sand} + (c - \mathrm{sand})^2 + d + e \log\ \mathrm{depth}$ | | $a = 1.35\ b = 0.0045\ c = 44.7\ d = 6 \times 10^{-5}\ e = 0.060$ | 6 |
| $\Delta\rho$ | Multiple linear regression $\Delta\rho = a + b \log\ \mathrm{OC} + c \log\ \mathrm{depth}$ where $\Delta\rho = \rho_b - \rho_m$ | | $a = -0.217\ b = -0.114\ c = -0.077$ | 3 |
| CM-B | Conceptual model $\hat{\rho}_b = \rho_m + \Delta\rho$ | | | 9 |

 

**Table 3** Performance summary for developed pedotransfer functions

| Model | Data set | No. parameters | RMSE (g cm$^{-3}$) | $r^2$ |
|---|---|---|---|---|
| MLR-A | A | 6 | 0.153 | 0.405 |
| ANN-A | A | 19 | 0.153 | 0.394 |
| RT-A | A | 11 | 0.195 | 0.197 |
| MLR-B1 | B | 6 | 0.153 | 0.498 |
| CM-B | B | 9 | 0.155 | 0.497 |
| ANN-B1 | B | 22 | 0.155 | 0.480 |
| RT-B1 | B | 19 | 0.163 | 0.429 |
| ANN-B2 | B | 25 | 0.158 | 0.478 |
| AS-B | B | 7 | 0.176 | 0.549 |

correlations between soil bulk density and clay, and silt-sized fractions. As such, input variables for MLR-A were restricted to depth and sand fractions. This contrasts with work by Heinonen (1977) who reported that better empirical relationships were achieved using clay and silt. The improved performance of depth data was achieved using log-transformed data, implying an exponential increase with depth. The poor performance of RT-A across all performance criteria suggests the ineffectiveness of particle size distributions and depth-dependent RTs in predicting soil bulk density in this study. Despite its relative simplicity, MLR-A performed far better than RT-A and comparable with ANN-A. In terms of goodness-of-fit, MLR-A ($r^2 = 0.405$) was found to perform marginally better than ANN-A ($r^2 = 0.394$) whilst both achieved an RMSE of 0.153 g cm$^{-3}$. As expected, the addition of organic carbon data improved bulk-density prediction compared with those models based solely on particle size distribution and depth.

Results from stepwise regression analysis indicated that log-transformed organic carbon data to be the best predictor of soil bulk density ($r^2 = 0.34$) whilst no significant correlations for clay and silt-sized fraction could be identified. This concurs with the findings by a study using British soils (Williams, 1970). With the exception of RT-B, all models developed for set B showed a similar performance. MLR-B exhibited marginal improvements in relation to RMSE and $r^2$, but generally model accuracies were comparable. When provided with the same input variables, ANN-B1 (RMSE = 0.155 g cm$^{-3}$) returned a larger RMSE to that of MLR-B (RMSE = 0.153 g cm$^{-3}$) and a slightly poorer goodness-of-fit ($r^2 = 0.480$) despite the use of a greater number of model parameters. Even with additional parameters and inputs, ANN-B2 could only achieve comparable results to that of ANN-B1. As shown by stepwise regression, silt and clay-sized fractions appear to have little or no predictive power in relation to soil bulk density. Indeed, the inclusion of silt and clay data in ANN-B2 slightly reduced model accuracy, suggesting that overfitting may have occurred. In a similar way to results from set A, RTs returned inferior and inefficient bulk-density predictions compared with regression and ANN models. Whilst other studies have shown RTs to be effective data-mining tools for other PTF applications (McKenzie & Jacquier, 1997), results in this study indicate their ineffectiveness for soil bulk-density prediction on Australian soils.

The added flexibility of RT and ANN models does not appear to improve model performance compared with simpler techniques, such as MLR. It seems from this study that relationships are more generalized. Calhoun *et al.* (2001) suggest that organic carbon and particle size distribution can only explain between 50 and 60% of variation in bulk density, with the remaining variation attributed to interactions in management history, particle density and hydrology amongst others. As shown in Table 4, previous studies have returned performances within the range stated by Calhoun *et al.* and comparable with those achieved in this study. Improved predictions have been achieved through partitioning; however, such approaches are dependent upon and should be limited to large data sets. The apparent limitations to the predictive accuracy of organic carbon and particle size distribution in regression equations suggest the need for the inclusion of better predictors or development of improved modelling methods.

### *Adapted Adams–Stewart model*

The introduction of the $\rho_m$ component intends to describe the variation in soil bulk density associated with depth and mineral content. As shown in Figure 2, the $\rho_m$ model demonstrates a bulk-density maximum at approximately 80% sand. Interestingly, these results closely correspond to the findings by Koltermann & Gorelick (1995) where bulk-density maxima were observed at sand percentages around 70–75% on structureless unconsolidated sediments. This suggests that despite pedological organization, such as aggregation, soil is still influenced by simple packing structures more commonly associated with single-grained unstructured sediments. As suggested by Koltermann & Gorelick (1995) the bulk-density maxima occur when small particles (silt or clay-sized particles) completely fill voids created between larger supporting particles (sand sized). The study also observed that porosity decreased with increased confining pressure due to compaction. The generalized relationship between depth and overburden pressure supports the inclusion of depth, in part as a proxy measure of confining pressure. The organic matter bulk density at field condition (0.224 g cm$^{-3}$) was retained as functions and optimized values did not improve model accuracy. As shown in Table 3, the Adams–Stewart model showed superior goodness of fit ($r^2 = 0.549$), whilst achieving the worst RMSE (0.176 g cm$^{-3}$) compared with the other set B models. As $\rho_m$ was the only trained component of the model, it is expected that fully trained models achieve better RMSE. In the light of the results, the Adams–Stewart model shows promise for soil bulk-density estimation on small data sets where organic matter data is insufficient for reliable bulk-density relationships to be found.

**Table 4** Previously developed pedotransfer functions and associated accuracies on the respective data sets

| PTF | Inputs | Parameters | *N* (training/ validation) | $r^2$ |
|---|---|---|---|---|
| Manrique & Jones (1991) | OC, Clay, WC15 | 4 | 19 226/NA | 0.58 |
| Bernoux *et al.* (1998) | OC, Clay, pH | 5 | 309/NA | 0.56 |
| Calhoun *et al.* (2001) | OC, Clay, Silt, Sand, Depth | 7 | 937/NA | 0.56 |
| Heuscher *et al.* (2005) | OC, WC15, Clay, Silt, Depth | 6 | 47 015/NA | 0.45 |

WC15, water content at −1500 kPa; OC, organic carbon (g $g^{-1}$); NA, not applicable.

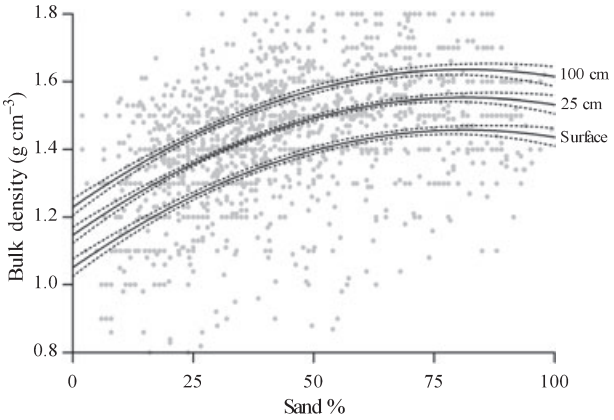


**Figure 2** Observed bulk density (dots) versus sand (200–2000 $\mu$m) percentage. Predicted bulk density (—) and associated 95% confidence limits (- - -) for $\rho_m$ at specified depths.

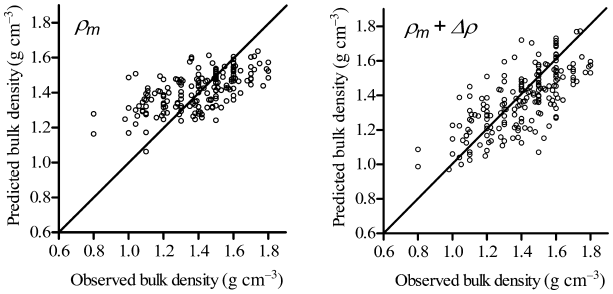


**Figure 3** Predicted vs observed soil bulk densities for the proposed conceptual model.

### *Conceptual model*

As discussed, MLR-A was found to be the best function based on particle size fraction and depth data and thus was used for the $\rho_m$ component of the conceptual model. Although the $\rho_m$ component does reflect those trends observed by Koltermann & Gorelick (1995), the large variance observed suggests the influence of another factor, namely soil structure.

This study identified organic carbon as a potential indicator of structural development and a predictor for $\Delta\rho$, the premise being that organic carbon enhances soil structure through aggregation and improved aggregate stability (Bronick & Lal, 2005). As anticipated, organic carbon data were found to be correlated with $\rho_m$ residuals. Log transformation of organic carbon coupled with log depth data further improved $\Delta\rho$ predictions. The $\Delta\rho$ component was shown to add the greatest improvement in model prediction where bulk density is $< 1.2$ and $> 1.6$ g cm$^{-3}$. As shown in Figure 3, the $\rho_m$ model tends to overestimate at small bulk densities and underestimate at large values. With the addition of the $\Delta\rho$ component these systematic errors are reduced delivering an improved goodness of fit.

As shown in Table 3, the conceptual model proved to perform comparably with other more traditional methods. The conceptual model outperforms the more complex ANN and RT methods despite using fewer input variables and model parameters. The small discrepancy between MLR-B and CM-B may be the result of MLR-B identifying interactions between sand, organic carbon and soil bulk density.

In this study we propose that the observed negative correlation between soil bulk density and organic carbon is the result of soil aggregation. Stewart *et al.* (1970) suggest that even at small organic matter contents there is an alteration to the packing of the mineral skeleton in unstructured soils. In structured soils, however, we propose that the promotion of soil aggregation by soil organic components has a greater effect on soil bulk density, primarily through an increase in inter-aggregate pore spaces. As mentioned, Childs (1969) referred to this as the structural pore space and recognized soil organic matter as a powerful aggregator.

Whilst generalized relationships between certain soil properties and soil structure are well known, the complex manner in which these properties interact makes precise quantification of relationships unrealistic. As such, a better understanding of the soil structure may be achieved through a direct description of the structure through nominal morphological data. In this study, morphological data may provide a better description of $\Delta\rho$. Calhoun *et al.* (2001) found that nominal morphological data can improve model bulk-density model predictions compared with laboratory-based models. Pachepsky *et al.* (2006) suggest that such improvements are due to morphological data providing information on the structure of the soil pore space. Although soil structural descriptions may be too coarse to describe fine-pore structure

(Pachepsky *et al*., 2006), soil morphological data may provide information on the soil structural pore space.

## Conclusions

The results indicate that increased model complexity does not necessarily provide improved model accuracy for bulk-density prediction. For both data sets, MLR was found to provide better predictions with greater efficiency than both ANNs and RTs. Results indicate that while general trends are identifiable, intricate relationships remain elusive. Whilst ANN and RT may fare better on larger, more comprehensive data sets, the ability to glean reliable models from small data sets is part of the appeal of pedotransfer functions.

The adapted Adams–Stewart model was found to provide the best goodness of fit, however, as only one component ($\rho_m$) is fitted to the data it is understandable that it yielded the poorest RMSE. Nevertheless, this method remains valuable where organic matter data are inadequate for dependable bulk-density relationships to be found. The performance of the conceptual model indicates its potential as a method for bulk-density prediction. Whilst the $\rho_m$ component is unlikely to experience significant improvement, the inclusion of data directly pertaining to soil structure should improve our estimate of $\Delta\rho$ and ultimately our bulk-density prediction.

Childs (1969) forecasted that at some time soil structure will focus on the quantitative description of the structural pore space with precise definition. Although such developments have yet to be incorporated into standard procedures, the development of superior structural descriptions would improve PTF accuracy for numerous applications.

## References


Adams, W.A. 1973. The effect of organic matter on the bulk and true density of some uncultivated podzolic soils. *Journal of Soil Science*, **24,** 10–17.

Alexander, E.B. 1980. Bulk density of Californian soils in relation to other soil properties. *Soil Science Society of America Journal*, **44,** 689–692.

Bernoux, M., Arrouays, D., Cerri, C., Valkoff, B. & Jolivet, C. 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. *Soil Science Society of America Journal*, **62,** 743–749.

Bronick, C.J. & Lal, R. 2005. Soil structure and management: a review. *Geoderma*, **124,** 3–22.

Bruand, A., Perez Fernandez, P. & Duval, O. 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. *Soil Use and Management*, **19,** 232–242.

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 characterisation data. *Soil Science Society of America Journal*, **65,** 811–819.

Childs, E.C. 1969. *An introduction to the physical basis of soil water phenomena*. John Wiley and Sons Ltd., London.

Curtis, R.O. & Post, B.W. 1964. Estimating bulk density from organic matter content in some Vermont forest soils. *Soil Science Society of America Proceedings*, **28,** 285–286.

De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J. & Muys, B. 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. *Soil Science Society of America Journal*, **69,** 500–510.

Grunwald, S., Rooney, D.J., McSweeney, K. & Lowery, B. 2001. Development of pedotransfer functions for a profile cone penetrometer. *Geoderma*, **100,** 25–47.

Gupta, S.C. & Larson, W.E. 1979. A model for predicting packing density of soils. *Soil Science Society of America Journal*, **43,** 758–764.

Heinonen, R. 1977. Towards "normal" soil bulk density. *Soil Science Society of America Journal*, **41,** 1214–1215.

Heuscher, S.A., Brandt, C.C. & Jardine, P.M. 2005. Using soil physical and chemical properties to estimate bulk density. *Soil Science Society of America Journal*, **69,** 51–56.

Jeffrey, D.W. 1970. A note on the use of ignition loss as a means for the approximate estimation of soil bulk density. *Journal of Ecology*, **58,** 297–299.

Judd, C.M. & McClelland, G.H. 1989. *Data analysis: a model comparison approach*. Harcourt Brace Jovanovich, New York.

Koltermann, C.E. & Gorelick, S.M. 1995. Fractional packing model for hydraulic conductivity derived from sediment mixtures. *Water Resources Research*, **31,** 3283–3297.

Manrique, L.A. & Jones, C.A. 1991. Bulk density of soils in relation to soil physical and chemical properties. *Soil Science Society of America Journal*, **55,** 476–481.

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.

Minasny, B., McBratney, A.B., Mendonca-Santos, M.L., Odeh, I.O.A. & Guyon, B. 2006. Prediction and digital mapping of soil carbon storage in the Lower Namoi Valley. *Australian Journal of Soil Research*, **44,** 233–244.

Pachepsky, Y.A., Rawls, W.J. & Lin, H.S. 2006. Hydropedology and pedotransfer functions. *Geoderma*, **131,** 308–316.

Rawls, W.J. 1983. Estimating soil bulk density from particle size analysis and organic matter content. *Journal of Soil Science*, **135,** 123–125.

Rawls, W.J., Nemes, A. & Pachepsky, Y.A. 2004. Effect of soil organic carbon on soil hydraulic properties. In: *Development of pedotransfer functions in soil hydrology* (eds Y.A. Pachepsky & W.J. Rawls), pp. 95–115. Elsevier, Amsterdam.

Roecker, E.B. 1991. Prediction error and its estimation for subset-selected models. *Technometrics*, **33,** 459–468.

Stewart, V.I., Adams, W.A. & Abdulla, H.H. 1970. Quantitative pedological studies on soils derived from Silurian mudstones. II. The relationship between stone content and apparent density of the fine earth. *Journal of Soil Science*, **21,** 248–255.

Vereecken, H. & Herbst, M. 2004. Statistical regression. In: *Developments of pedotransfer functions in soil hydrology* (eds Y.A. Pachepsky & W.J. Rawls), pp. 3–21. Elsevier B.V., Amsterdam.

Williams, R.J.B. 1970. Relationships between the composition of soils and physical measurements made on them. *Report of Rothamsted Experimental Station*, **Pt 2,** 5–35.