Derivation and evaluation of a set of pedogenically-based empirical algorithms for predicting bulk density in British soils

S.H. Hallett, J.M. Hollis and C.A. Keay
National Soil Resources Institute, School of Applied Science, Cranfield University, Cranfield, Bedfordshire, MK43 0PL, Great Britain
Email: s.hallett@cranfield.ac.uk Website: http://www.landis.org.uk

ABSTRACT

A set of pedogenically-based empirical algorithms for predicting the variation of bulk density across all soils within England and Wales has been developed from a dataset of 1,568 measured values. The algorithms are derived from multiple regression analysis of sub-sampled datasets differentiated according to soil horizons and lithological groupings of soil substrate material. Parameters required for the algorithms are the percentage clay (<2μm), silt (2-60μm), sand (60-2000μm) and organic carbon in soil horizons. Statistical analysis shows the methodology to give significantly better prediction of the 1,568 measured values than does the method proposed by Rawls (1983), based on a large set of measured data from the USA. For the proposed method, the adjusted r^2, scaled root mean square error and model efficiency values are 0.77, 0.1 and 0.727 respectively, whereas equivalent values for the Rawls method are 0.56, 0.131 and 0.534. The improved prediction of the newly developed algorithms is attributed largely to their pedogenic component. As they are based on measured data from a variety of soil types, the algorithms are considered to be applicable to a similar range of soil types within Europe and parts of North America. Their continuous nature makes them ideally suited for incorporation within computer-based Geographical Information Systems designed to support modelling of the agricultural and semi-natural environment at a variety of scales.

INTRODUCTION

Soil bulk density, defined as the apparent density of field soil and calculated from the oven-dry mass per unit volume of field soil, is an important soil property that summarises general soil structural characteristics. It is a fundamental input requirement for virtually all mathematical models describing the transfer and interaction of soil chemical constituents within the ecosphere.

Bulk density is a relatively straight-forward property to measure and a number of extensive datasets have been compiled (Hall et al., 1977; Rawls et al., 1981). Because of this, few attempts have been made to develop methodologies for its prediction from other basic soil properties. However, the increasing interest in developing comprehensive national datasets of soil physical properties for use in spatially- or stochastically-based environmental modelling (King et al., 1995; Bruand et al., 1996) has inevitably highlighted discontinuities in the existing measured datasets. This in turn, has now focused attention on the need to develop algorithmic methodologies that can predict variation in bulk density according to the continuous variation of soil properties such as particle-size and organic matter content.

In England and Wales, comprehensive data on the soil particle-size distribution, organic carbon content and pH is available from two principal sources.

Firstly, analytical data characterising the horizons of soil profiles chosen as typical representatives of soil series under specific land uses has been collected over a period of some 50 years. Soil series form the basic unit of classification and mapping in England and Wales (Clayden and Hollis, 1984; Hollis and Avery, 1997). To date, the basic properties of some 7,000 soil horizons representing approximately 2,500 soil profiles have been analysed and the data incorporated within the SSLRC Land Information System, 'LandIS' (Hallett et al., 1996). Secondly, in addition to this data, some 5,600 sets of topsoil analyses have been undertaken as part of the National Soil Inventory, 'NSI', dataset of England and Wales (McGrath and Loveland, 1992), a statistically-valid, 5km resolution, grid-based sampling of the two countries carried out during field survey for the national 1:250 000 scale soil maps (Soil Survey Staff, 1983). Based on these two data sources, the mean value and standard deviations of pH, sand, silt, clay and organic carbon content of each soil horizon of all 412 main soil series in England and Wales under each of four different land uses: arable; short-term rotational (ley) grassland; long-term grassland and other, semi-natural vegetation, has been calculated.

These datasets represent the most comprehensive information on the continuous variation of basic soil properties within England and Wales and provide an ideal base for predicting the variation in bulk density across the two countries. This paper describes the development and evaluation of a pedogenically-based set of empirical algorithms for predicting bulk density using the national soil profile datasets described above.

MATERIALS

In order to characterise the physical properties of soil series recognised during systematic soil mapping of England and Wales at the 1:25 000 and 1:50 000 scales, triplicate undisturbed soil cores of 222 cm^3 were field-collected from individual horizons of representative soil profiles (Fig. 1). Sampling sites were selected to represent typical profiles of soil series under specific land uses. The range of soil types represented in the dataset includes Alisols, Arenosols, Cambisols, Fluvisols, Gleysols, Histosols, Leptosols, Luvisols, Phaeozems, Podzoluvisols, Podzols and Regosols (FAO, 1990a).

Most of the samples were collected over a period between 1970 and 1987, using uniform methods described by Hodgson (1976). Bulk density (g cm^{-3}), and the percentage volume of water retained at -5, -10, -40, -200 and -1500 kPa tension were measured on each sample. Retained water volumes were measured using either sand tension baths or pressure membrane apparatus, depending on the soil water tension to be measured. Bulk density was measured from the oven dry (105^{o}) mass of each undisturbed core sample. Average values of the triplicate sets of measurements were used represent the individual properties of each of the horizons. The analytical methodology is fully described by Avery and Bascomb (1982).

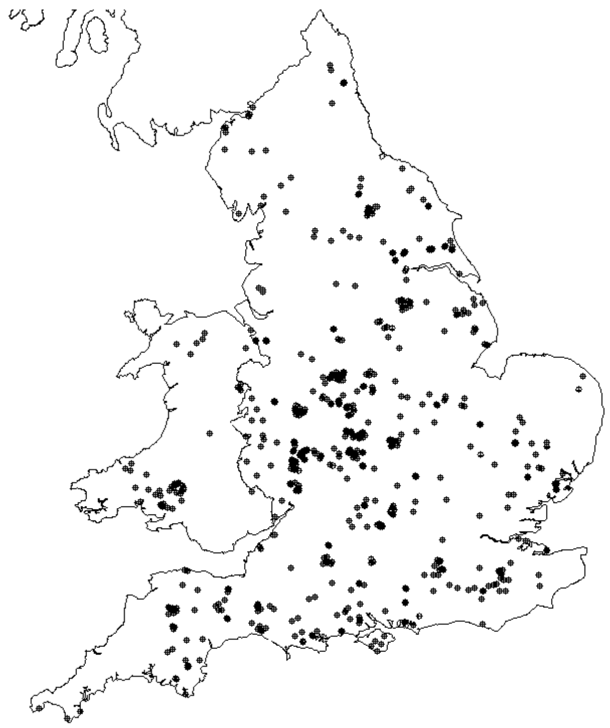


Fig. 1. Sites sampled for water retention, porosity and density studies.

In addition to the undisturbed core samples, bulk samples were taken from each soil horizon and used to determine the percentage of organic carbon present and the percentage of mineral particles with an average diameter <0.002mm, 0.002-0.06mm; 0.06-0.1mm; 0.1-0.2mm; 0.2-0.6mm and 0.6-2mm, although for some peat soil horizons, the percentage of mineral particles was not analysed. Organic carbon % was determined using the method of Tinsley (1950) and particle-size analysis determined using wet sieving or the pipette method, depending on the size fractions analysed. At each site the characteristics of each horizon in the soil profile were described using the terminology given by Hodgson (1976) and the horizon assigned a notation based on its morphological and pedogenic characteristics (Avery, 1980). The soil profile was also classified to soil series level according to the criteria described by

Avery (1980) and Clayden and Hollis (1984). At the soil series level important differentiating criteria include the presence or absence of distinctive mineralogy and the nature and lithology of the soil parent material.

The sampling programme resulted in a dataset comprising 1,606 soil horizons, of which 1,568 horizons had a complete set of measured bulk density, organic carbon content and particle-size fractions and 38 horizons had measured bulk density and organic carbon content only. Characteristics of the dataset are summarised in Table I.

TABLE I

Summary of observed soil data used in analysis

	Sand (%) PSF*[1] 60-2000μm	Clay (%) PSF < 2μm	Silt (%) PSF 2-60μm	Organic Carbon (%)		Bulk density g/cm^3	
Sample size	1568 *[2]	1568 *[2]	1568 *[2]	1606	1568	1606	1568
Maximum	98.99	89.00	87.00	74.0	50.0	1.85	1.85
Minimum	0.01	0.01	1.00	0.03	0.03	0.08	0.13
Mean	35.50	26.04	38.46	2.87	2.12	1.26	1.28
Standard Deviation	26.67	17.03	18.75	6.79	3.55	0.28	0.25

*1 PSF, Partical-Soil Fraction
*2 Only 1568 horizons have mineral PSF data

METHODS

Two methods for predicting soil bulk density based on particle-size distribution and organic matter content were tested using the dataset described above.

Rawls method

The procedure proposed by Rawls (1983) was used to predict bulk density based on the measured sand, silt and clay particle-size fractions and organic carbon percentage of the 1,568 horizons for which such data was available (see Table I). The Rawls procedure was chosen because it is the only published methodology developed from a large dataset of measured values and also because it has been used for environmental modelling where measured bulk density data are lacking (Mullins et al., 1993).

The procedure involves identifying a value of bulk density for the mineral soil fraction by interpolating between 'contour lines' of density values drawn over the standard textural triangle used by the United States Department of Agriculture (Rawls, 1983). Derived mineral density values are then corrected for organic matter content using the equation proposed by Adams (1973) which assumes an average organic matter bulk density of 0.224 g cm^{-3} (eq. 1).

$$\rho = \frac{100}{(\%OM / 0.224) + (100 - \%OM / \rho m)} \qquad (1)$$

Where ρ = soil bulk density (g cm^{-3}), %OM = percent by weight organic matter, ρ_m = bulk density of soil mineral material.

The first part of this procedure is laborious, even when partially automated using a computer-based approach. It can also be subjective as bulk density values for points falling between 'contour lines' require visual interpolation. In addition, because the separation between sand and silt fractions in the USDA system is set at 0.05mm esd, as opposed to 0.06mm esd in the England and Wales system, it was necessary to estimate the equivalent USDA sand and silt contents for each measured sand and silt content in the dataset (eq. 2).

$$\%(0.002 - 0.05mm) = \%(0.002 - 0.06mm) - [\%(0.06 - 0.1mm)x0.26] \qquad (2)$$

Although this method of conversion gives small errors in the resulting USDA sand and silt percentages, these are likely to be insignificant compared to those resulting from the visual interpolation necessary for the Rawls estimation method.

Empirical-pedogenic method

Because of difficulties in applying the procedure proposed by Rawls to predict bulk density within a large national dataset of particle-size and organic matter and because of concerns about the purely empirical basis of the approach with the resulting uncertainty as to its transference to soils in England and Wales a new methodology was developed based upon multiple regression analysis of sub-sampled datasets differentiated on a pedogenic basis. Unlike the Rawls method, this procedure takes into account structural factors by considering the degree to which pedogenesis has modified the organisation of primary particles within the soil profile. In order to do this, the horizon nomenclature adopted by the Food and Agriculture Organisation (FAO, 1990b) and the Soil Survey and Land Research Centre (Avery, 1980) was first used to group the 1,606 mineral and organic horizons for which data was available (see Table I) as shown in Table II.

TABLE II

Horizon nomenclature and associated characteristics used to group the datasets used in the regression analysis

Horizon[*1] nomenclature	Characteristics
H	Surface organic layers not resulting from waterlogging
Of	Fibrous, relatively unhumified organic layers resulting from waterlogging
Om	Semi-fibrous, partly humified organic layers resulting from waterlogging
Oh	Well humified surface organic layers originally formed as a result of waterlogging
Oh1	Well humified subsurface organic layers originally resulting from waterlogging and formed at or above 50cm depth.
Oh2	Well humified subsurface organic layers originally resulting from waterlogging and formed below 50cm depth.
A	Surface mineral layer showing distinct incorporation of organic matter
B_{podz}	Subsurface mineral 'podzolic' Bs or Bh horizon enriched with 'illuvial' organic matter, iron and/or aluminium
E	Subsurface mineral layers from which clay and or iron have been lost by the process of 'eluviation'
Eg	'Gleyed' E horizon resulting from seasonal waterlogging
Bt	Subsurface mineral layer enriched with 'illuvial' clay from overlying horizons
Bw1	Subsurface mineral layer formed at or above 50 cm depth and with distinct soil structure and colour indicating slight weathering
Bg1	'Gleyed' subsurface mineral layer formed at or above 50 cm depth, with distinct soil structure and resulting from seasonal waterlogging
Bw2	Subsurface mineral layer formed below 50 cm depth and with distinct soil structure and colour indicating slight weathering
Bg2	'Gleyed' subsurface mineral layer formed below 50 cm depth, with distinct soil structure and resulting from seasonal waterlogging
BC	Subsurface layer with structural characteristics transitional between an overlying B horizon and an underlying, unweathered C horizon
C	Subsurface mineral layer unaffected by soil forming processes, showing no soil structural or weathering characteristics

[*1] In the FAO horizon nomenclature, 'H' and 'O' horizons are designated 'O' and 'H' respectively

Within organic soil horizons designated, as 'H' or 'O', bulk density is determined mainly by organic matter content, mode of formation and degree of humification. These characteristics are used to specify the horizon nomenclature and therefore no further stratification of the identified groups is necessary. A simple linear regression analysis was used to define the relationship between bulk density and organic matter content in each of these groups.

Surface soil layers designated as 'A' represent those mineral horizons subjected to the maximum amount of pedological weathering and reorganisation, including the incorporation of organic matter. Within this group, in addition to particle-size distribution and organic matter content, bulk density is considered likely to depend upon the amount of disturbance resulting from human activity. The 'A' horizon data subset was therefore further subdivided into four categories of decreasing 'disturbance' representing mineral soil layers under 'arable' cultivation (A_ar), short-term rotational 'ley' grassland (A_le), long-term 'permanent' managed grassland (A_pg) and 'other' semi-natural vegetation (A_ot).

The 'B_{podz}' designation separates 'podzolic B' (Avery, 1980; FAO, 1990a) or 'spodic' (Soil Survey Staff, 1996) subsurface mineral horizons where the bulk density differs significantly from other types of mineral subsurface horizon because of their characteristic genesis. No further subdivision of this group is considered necessary.

Within the majority of the other mineral subsurface layers, the amount of pedological weathering and re-organisation decreases from the 'E' to the 'C' horizons as shown in Table II. Those soil horizons designated as 'C' characterise unconsolidated or weakly consolidated mineral soil layers relatively unaltered by pedogenic processes. In some soil horizons however, soil parent material factors are considered to have an overriding influence on bulk density. Thus, within all subsoil horizons formed in recently deposited colluvium, alluvium, lake marl or tufa (Avery, 1980), the mode of formation is considered to be more important in determining bulk density than any structural modification that may be reflected in the horizon type. Similarly, within all subsoil horizons formed in soft, slowly permeable or impermeable parent materials, as categorised according to the Hydrology of Soil Types (HOST) classification system (Boorman et al., 1995), bulk density is considered to depend more on the degree of consolidation of the parent material than on any structural factors related to horizon type.

Because of these considerations, subsoil horizons formed in recent colluvium, recent alluvium, lake marl, tufa, soft slowly permeable or soft impermeable materials were not stratified according to horizon nomenclature, but simply sub-divided into seven 'parent material groups', designated LITH 1 to LITH 7, depending on the mode of formation or the consolidation and permeability of their soil parent materials, as shown in Table III.

TABLE III

Parent material groups used to differentiate subsoil horizons formed in recent alluvium[*1], recent colluvium[*1], lake marl[*1], tufa[*1], soft slowly permeable[*2] or soft impermeable[*2] material

Parent material group	Definition
LITH 1	Lake marl and tufa
LITH 2	Alluvium and colluvium
LITH 3	Weakly consolidated Glaciolacustrine deposits
LITH 4	Weakly consolidated glacial till
LITH 5	Impermeable pre-quaternary clays, sandy clays and mudstones
LITH 6	Weathered igneous and metamorphic rocks
LITH 7	Pre-quaternary chalk marl, loam or siltstone

[*1] As defined by Avery, 1980
[*2] As defined by Boorman et al., 1995

Each of the stratified mineral soil groupings described above were analysed using the MINITAB statistical package (Ryan et al., 1976) to derive the best possible regression relating bulk density to the combination of organic carbon, clay, silt and sand content. The derived regression equations for both organic and mineral soil stratified groupings are shown in Table IV. They explain between 6% and 74% of the variation in measured bulk density within each stratified group and 77% of the variation in measured bulk density for the complete dataset of 1,568 soil horizons. Statistical data for the regressions are summarised in Table V.

TABLE IV

Results of the SSLRC stratified regression analysis methodology

D_b [H] = 0.8964 - 0.018 OrgC
D_b [Oh] = 0.585 - 0.007 OrgC
D_b [Of] = Mean measured value used
D_b [Om] = 0.4166 - 0.0046 OrgC
D_b [Oh1] = 0.486 - 0.00504 OrgC
D_b [Oh2] = 0.558 - 0.00803 OrgC
D_b [A_ar] = 1.46 - 0.0254 Log_eClay + 0.0279 Log_eSand - 0.261 Log_eOrgC
D_b [A_le] = 0.807 + 0.0989 Log_eClay + 0.106 Log_eSand - 0.215 Log_eOrgC
D_b [A_pg] = 0.999 + 0.0451 Log_eClay + 0.0784 Log_eSand - 0.244 Log_eOrgC
D_b [A_ot] = 0.870 + 0.0710 Log_eClay + 0.0930 Log_eSand - 0.254 Log_eOrgC
D_b [Bpodz] = 0.998 - 0.0702 Log_eSilt + 0.0798 Log_eSand - 0.131 Log_eOrgC
D_b [Eg] = 1.50 - 0.00067 Silt + 0.00262 Clay - 0.139 OrgC
D_b [E] = 1.54 - 0.000583 Silt - 0.00008 Clay - 0.162 OrgC
D_b [Bw1] = 1.55 - 0.00147 Silt - 0.00018 Clay - 0.209 OrgC
D_b [Bg1] = 1.47 - 0.00727 Silt + 0.00716 Clay - 0.082 OrgC
D_b [Bt] = 1.66 - 0.00069 Silt - 0.00827 Clay + 0.0123 OrgC
D_b [Btg] = 1.67 + 0.000751 Silt - 0.0105 Clay + 0.0316 OrgC
D_b [Bw2] = 1.54 - 0.00546 Silt + 0.00338 Clay - 0.160 OrgC
D_b [Bg2] = 1.69 + 0.00210 Silt - 0.00231 Clay - 0.505 OrgC
D_b [BC] = 1.49 - 0.00029 Silt + 0.00437 Clay - 0.314 OrgC
D_b [C] = 1.50 - 0.00059 Silt + 0.00085 Clay - 0.254 OrgC
D_b [LITH 1] = 0.7132 - 0.0336 OrgC
D_b [LITH 2] = 1.56 - 0.00124 Silt - 0.00372 Clay - 0.0668 OrgC
D_b [LITH 3] = 0.618 + 0.095 Log_eSilt + 0.100 Log_eClay + 0.0195 Log_eSand - 0.178 Log_eOrgC
D_b [LITH 4] = -0.015 + 0.119 Log_eSilt + 0.102 Log_eClay + 0.186 Log_eSand - 0.141 Log_eOrgC
D_b [LITH 5] = 1.96 - 0.0158 Log_eSilt - 0.154 Log_eClay + 0.0102 Log_eSand - 0.113 Log_eOrgC
D_b [LITH 6] = 5.01 - 0.931 Log_eSilt + 0.038 Log_eClay - 0.173 Log_eSand - 0.365 Log_eOrgC
D_b [LITH 7] = 2.37 - 0.246 Log_eSilt + 0.0266 Log_eClay - 0.0178 Log_eSand - 0.114 Log_eOrgC

TABLE V

Summary of the SSLRC stratified regression analyses

Group	n [*2]	Max D_b	Min D_b	Mean D_b	St.Dev D_b	r^2 adj. %	F statistic
H	7	0.67	0.30	0.4014	0.1317	66.8	13.09
Om	10	0.32	0.11	0.1960	0.0667	21.6	3.47
Of [*1]	2	0.20	0.16	0.18	-	-	-
Oh	11	0.60	0.16	0.331	0.143	51.6	11.64
Oh1	7	0.41	0.08	0.286	0.122	44.1	5.74
Oh2	5	0.61	0.10	0.27	0.199	36.2	3.27
A_ar	247	1.76	0.60	1.3007	0.2147	66.9	167.00
A_le	108	1.71	0.71	1.2952	0.1827	50.9	38.00
A_pg	233	1.69	0.51	1.0598	0.2017	56.6	101.66
A_ot	60	1.73	0.46	1.081	0.2461	57.2	27.24
Bpodz	35	1.67	0.65	1.1551	0.2888	74.1	33.39
Eg	21	1.61	0.78	1.3257	0.2584	66.6	14.32
E	68	1.67	0.88	1.3976	0.1540	23.9	8.02
Bw1	110	1.74	0.81	1.3228	0.1928	40.0	25.19
Bg1	16	1.53	0.94	1.3662	0.1384	19.0	2.17
Bt	44	1.79	0.86	1.4123	0.21364	18.0	4.15
Btg	12	1.63	1.39	1.4675	0.0808	31.3	2.67
Bw2	22	1.53	0.91	1.3836	0.1703	27.4	3.65
Bg2	9	1.60	1.32	1.5011	0.0983	54.5	4.19
BC	41	1.76	0.94	1.4202	0.1679	47.8	13.21
C	24	1.68	1.22	1.4437	0.1137	6.2	1.50
LITH 1	5	0.88	0.45	0.5980	0.1764	55.5	2.66
LITH 2	183	1.61	0.84	1.2875	0.1638	45.4	51.49
LITH 3	16	1.63	1.12	1.4438	0.1344	63.6	7.54
LITH 4	136	1.85	0.93	1.4177	0.1830	41.8	25.29
LITH 5	145	1.71	1.03	1.3962	0.1603	45.0	30.46
LITH 6	8	1.51	1.18	1.3538	0.1082	65.0	4.26
LITH 7	21	1.70	1.27	1.4990	0.1158	26.4	2.79
All Groups	1606	1.85	0.08	1.2603	0.2870	77.1	4169.07

[*1] Not enough source data was present, so the mean value was adopted.
[*2] n = number of cases used by regression

RESULTS AND METHOD EVALUATION

Using the observed data described above, both the Rawls method and the empirical-pedogenic method (in future, referred to as the SSLRC method) were used to predict bulk density for each of the 1,568 horizons for which measured mineral and organic fraction data was available (see Table I). The two predicted datasets were then compared against the measured data using a number of statistical parameters designed to evaluate the relative goodness of fit. Table VI presents a summary of the predicted and observed bulk density data.

TABLE VI

Summary of observed and predicted bulk density

	Field Observed Db	Rawls Predicted Db	SSLRC Predicted Db
Maximum	1.85	1.80	1.73
Minimum	0.13	0.25	0.16
Mean	1.28	1.25	1.28
C.I. (95%)	± 0.012	± 0.010	± 0.010
Standard Deviation	0.25	0.20	0.21

Data Sample Size for comparison: 1568

Analysis of Variance

In comparing the two predictive methodologies, it is useful to consider observed variation within the datasets. Variation is due either to 'within-group variation' or 'between group variation'. Table VII presents a one-way analysis of variance conducted using the two predicted and the observed datasets to establish whether 'between group variation' is significantly greater than 'within group variation'.

TABLE VII

Analysis of Variance

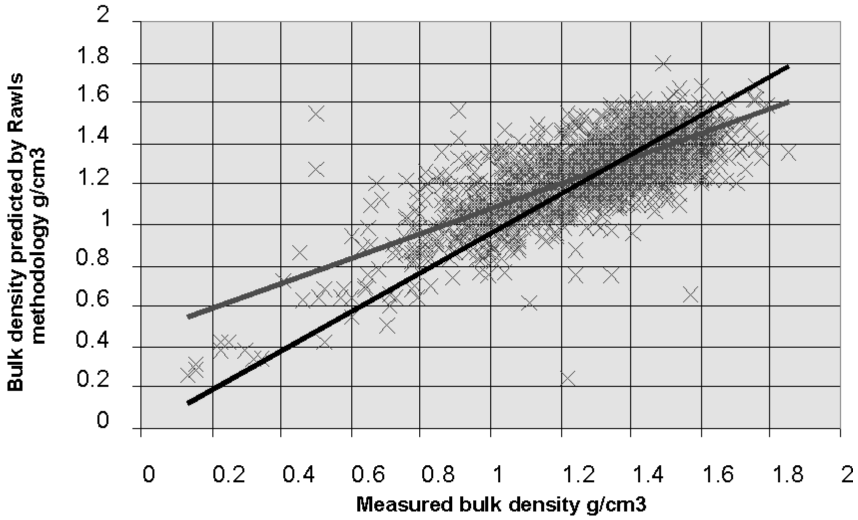
Source of variation	No. Degrees of Freedom (DF)	Sum of Squares (SS)	Mean Sum of Squares (MS)	Ratio (F)
Variation between groups	2	1.07	0.54	11.15
Variation due to random error	4701	225.77	0.05	
Total	4703	226.84		

A null hypothesis that the three datasets exhibit similar means is disproved and significant inter-group differences are observed. Given the means and confidence intervals in Table VI, it is apparent that the Rawls dataset is the cause of the discrepancy.

Linear regression of measured vs. predicted values

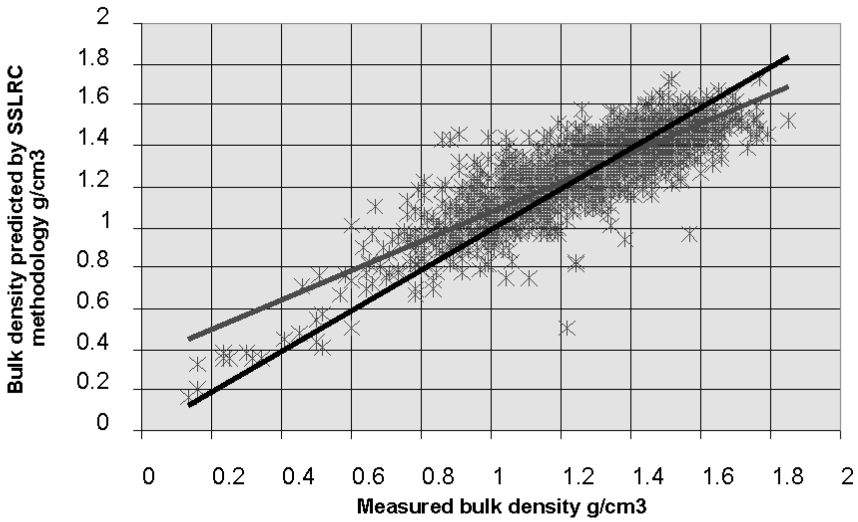
The datasets predicted by both methodologies were plotted against the field observed bulk density data, and a linear regression calculated for each dataset (Fig. 2).

The plotted regression line was also compared with the line of unity. The SSLRC method explains 17% more of the measured variation in bulk density than does the Rawls method. In addition, the regression line for the SSLRC-based predictions has a slope (0.72) that is closer to unity than is that for the Rawls-based predictions (0.61) whereas the intercept value of 0.36 is closer to 0 than that of the Rawls-based regression line (0.47).



$y = 0.61\ (\pm\ 0.03;\ 95\%)\ x + 0.47\ (\pm\ 0.04;\ 95\%)$
$r^2 = 0.56$

Fig. 2a. Bulk density predicted by the Rawls methodology using SSLRC field data



$y = 0.72\ (\pm\ 0.02;\ 95\%)\ x + 0.36\ (\pm 0.03;\ 95\%)$
$r^2 = 0.73$

Fig. 2b. Bulk density predicted by the SSLRC regression methodology using SSLRC field data

Cumulative error

The cumulative error of both predictive methodologies is shown in Fig. 3. For the SSLRC method, 95% of the predictions have an error $\geq$ 0.25 whereas for the Rawls method, the equivalent value is $\geq$ 0.32.

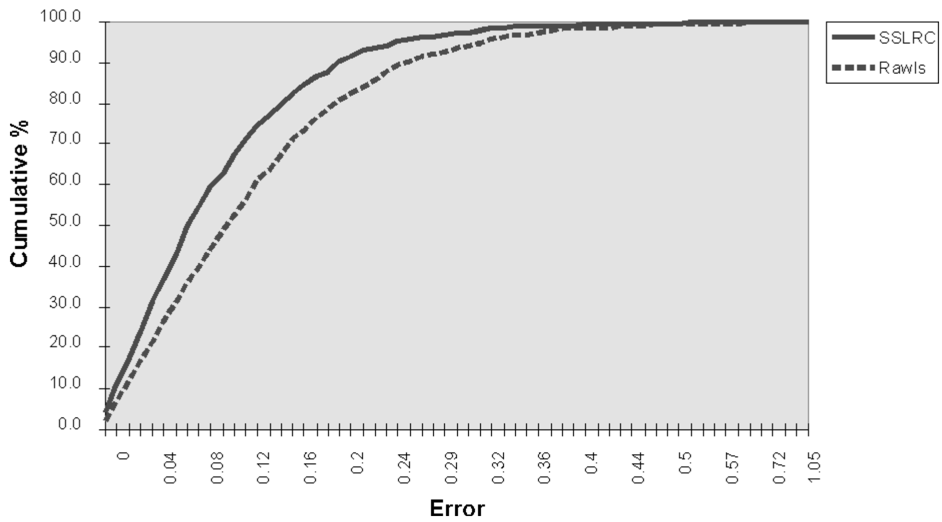


Fig. 3. Comparison of cumulative error between Rawls and SSLRC methodologies

Scaled Root Mean Square Error (SRMSE) and Model Efficiency (ME)

Both these statistical tests are suggested by Walker et al. (1995) as suitable for assessing the overall 'goodness of fit' of predicted to measured data sets. SRMSE is a measure of the 'spread' of predicted values about the line of unity between predicted and observed values. The equation (eq. 3) includes a scaling factor to take into account the number of paired values in the datasets. The closer the SRMSE is to 0, the better the performance of the model is in predicting the observed values.

$$SRMSE = \frac{1}{\overline{O}} \cdot \sqrt{\frac{\sum_{1}^{N} i(Pi - Oi)^2}{N}} \qquad (3)$$

Where Oi = Observed value in the i-th layer (i=1,…,N), Pi = Predicted value in the i-th layer (i=1,…,N)

ME gives a measure of the performance of a predictive methodology against observed values. The form used (eq. 4) is known as the 'Sutton-Rathcliffe Coefficient'. ME has no lower bound, its upper bound being 1 at a point where optimal predictive efficiency is achieved.

$$ME = 1 - \frac{\sum_{1}^{N} i(Pi - Oi)^2}{\sum_{1}^{N} i(Oi - \overline{O})^2} \qquad (4)$$

Where Oi = Observed value in the i-th layer (i=1,…,N), Pi = Predicted value in the i-th layer (i=1,…,N)

Calculated SRMSE and ME for the SSLRC and the Rawls method are given below. For both SRMSE and ME, the SSLRC method gives values closer to the optimum than does the Rawls method.

Predictive Method	SRMSE	ME
SSLRC	0.100	0.727
Rawls	0.131	0.534

DISCUSSION

The statistical analyses conducted demonstrate that, for the large set of measured bulk density data from England and Wales, the SSLRC empirical-pedogenic method gives a significantly better overall prediction than does the Rawls method. This can be attributed in part to the semi-empirical nature of the SSLRC method, being based on data from England and Wales. However, a purely empirically-based set of regression equations for predicting bulk density, developed from an earlier analysis of the measured data, gave a much poorer overall prediction (r^2 adj. = 58.7%) which suggests that the better prediction of the final SSLRC methodology is largely the result of its pedogenic basis. This undoubtedly gives improved prediction of bulk density for a wider range of soils than does the Rawls method which does not cope well with organic or podzolic/spodic horizons, nor take into account differences in matrix consolidation inherited from different types of soft pre-Quaternary, Quaternary and Holocene materials. It also fails to predict the different bulk densities arising from pedogentically derived structural differences between cambic and argic ‘B’ horizons (FAO, 1990a) formed in materials with similar particle-size distribution and organic matter content.

Although the SSLRC method is ideally suited for application within England and Wales, its basis suggests it can be applied with some confidence to soils within a similar range of pedogenic soil groups outside these two countries, providing that a consistent identification of the 'soft slowly permeable' and 'soft impermeable' lithological groupings can be made. The methodology should thus be applicable to most of Europe, as well as much of eastern North America. However, the methodology has yet to be tested on soils outside England and Wales and for this purpose, a comprehensive set of measured data from many European countries, such as the HYPRES database (Lilly, 1996) would be ideal. Until such testing has been undertaken, use of the proposed methodology outside England and Wales should be treated with caution.

Whereas the SSLRC method gives improved prediction of bulk density for a wide range of soils, there remain a number of limitations to its use. Firstly, and most importantly, prediction of bulk density in arable topsoil (A or O) horizons is likely to give an 'average annual' value only and is unlikely to reflect a measured value taken at any one point in the annual loosening and consolidation cycle that occurs in such cultivated horizons. Secondly, some of the individual algorithms give poor predictions within pedogenic groupings and for such groupings the statistics suggest that the algorithms give no better prediction than taking the mean value of the measured data. However, rather than do this, or develop an alternative regression, it was considered more important to preserve a consistent empirical methodology for analysis within all groupings. Future work will concentrate on improving the within-group predictions.

One consequence of this research is that bulk density predictions and algorithms can now be incorporated within computer-based GIS applications designed to support modelling of natural and peri-natural environmental phenomena. One such system, developed by SSLRC, is named 'SEISMIC' (Hallett et al., 1995), an interactive database management and interrogation system that provides easy access to comprehensive national benchmark soil data for the agricultural environment of England and Wales, including the bulk density data generated using the methodologies outlined here.

CONCLUSIONS

This paper has presented a new methodology for predicting soil bulk density based upon pedogenic stratification and empirical regression analysis of observed soil data. For a large set of measured data from England and Wales, the method gives significantly better prediction than the widely used method of Rawls (1983) based on data from the USA. This is attributed, at least in part, to its pedogenic basis which takes into account more factors and embraces a wider range of soil types, for example Histosols and Podzols/Spodosols, than does the Rawls method. Due to its pedogenic basis, the SSLRC method should be applicable to the soils of other countries with a similar range of pedogenic soil types to those included in this analysis. However, the method requires testing against a large set of measured data from outside England and Wales.

The algorithmic basis of the methodology makes it well suited to incorporation into large-scale, GIS-based modelling applications where the only widely available data expressing the spatial variation of soil properties is basic particle-size distribution and organic matter content of pedogenic soil types. Its is also ideally suited to the creation of national databases of soil properties, such as those included in the SEISMIC system (Hallett et al., 1995).

There are areas of this research that may be considered as a basis for future development. Individual regressions for certain pedogenic groupings exhibited relatively poor prediction, mainly because of the scarcity of observed data. To address this, it is anticipated that additional measured data could be utilised from other sources or that aggregations be made between some groupings. Furthermore, it is also expected that predictive approaches other than regression could be adopted for some or all of the strata identified, however, for the purposes of consistency in this research, all strata were subjected to the same approach.

ACKNOWLEDGEMENTS

The authors thank Professor Peter Bullock for help and encouragement in the preparation of this paper. Much of the work described was carried out during the development of the databases included within the SEISMIC system. The authors gratefully acknowledge joint funding of that research by the Ministry of Agriculture, Fisheries and Food and the British Agrochemicals Association.

REFERENCES

Avery, B.W., 1980. Soil Classification for England and Wales [Higher Categories], Soil Survey Technical Monograph No. 14. Rothamsted Experimental Station, Harpenden, 67 pp.

Avery, B.W. and Bascomb, C.L., 1982. Soil Survey Laboratory Methods, Soil Survey Technical Monograph No. 6. Rothamsted Experimental Station, Harpenden, 83 pp.

Adams, W.A., 1973. The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. J. Soil Sci., 24: 10-17.

Boorman, D.B., Hollis, J.M. & Lilly, A. 1995 Hydrology of Soil Types: A hydrologically-based classification of the soil in the UK. Institute of Hydrology Report No. 126, Wallingford, UK. 137 pp.

Bruand, A., Duval, O., Wösten, J.H.M. and Lilly, A., (eds.). 1996. The use of Pedotransfer in Soil Hydrology Research in Europe. Proceedings of the second workshop. INRA, Orleans, France, 10-12/10/1996. INRA, Orleans and EC/JRC Ispra. 211 pp.

Clayden, B. and Hollis, J.M., 1984. Criteria for Differentiating Soil Series, Soil Survey Technical Monograph No. 17. Rothamsted Experimental Station, Harpenden, 59 pp.

FAO, 1990a. Soil Map of the World: Revised Legend. World Soil Resources Report 60. FAO - Unesco, Rome. 119 pp.

FAO, 1990b. Guidelines for Soil Profile Description. 3rd Edition (Revised). FAO-Unesco, Rome. 70 pp.

Hall, D.G.M., Reeve, M.J., Thomasson, A.J. and Wright, V.F., 1977. Water Retention, Porosity and Density of Field Soils, Soil Survey Technical Monograph No. 9. Rothamsted Experimental Station, Harpenden, 75 pp.

Hallett, S.H., Jones, R.J.A. and Keay, C.A., 1996. Environmental Information Systems Developments for Planning Sustainable Land Use. Int. J. Geographical Information Systems, 10.1: 47-64.

Hallett, S.H., Thanigasalam, P.T. and Hollis, J.M., 1995. SEISMIC: A desktop Information System for Assessing the Fate and Behaviour of Pesticides in the Environment. Computers and Electronics in Agriculture, 13: 227-242.

Hodgson, J.M. (ed.), 1976. Soil Survey Field Handbook, Soil Survey Technical Monograph No. 5. Rothamsted Experimental Station, Harpenden, 99 pp.

Hollis, J.M. and Avery, B.W., 1997. History of Soil Survey and Development of the Soil Series Concept in the U.K. Advances in GeoEcology 29, 109-144.

King, D., Jones, R.J.A. and Thomasson, A.J. (eds.), 1995. European Land Information Systems for Agro-environmental Monitoring. CEC, JRC Ispra, Italy. 284 pp.

Lilly, A., 1996. A description of the HYPRES database (Hydraulic Properties of European Soils). In: The use of Pedotransfer in Soil Hydrology Research in Europe. Proceedings of the second workshop. INRA, Orleans, France, 10-12/10/1996. (eds. Bruand, A., Duval, O., Wösten, J.H.M. and Lilly, A.). INRA, Orleans and EC/JRC Ispra. 161-182.

McGrath, S.P. and Loveland, P.J., 1992. The Soil Geochemical Atlas of England and Wales. Blackie Academic & professional, London. 101 pp.

Mullins, J.A., Carsel, R.F., Scarborough, J.E. and Ivery, A.M., 1993. PRZM-2, a Model for Predicting Pesticide Fate in the Crop and Unsaturated Soil Zones: Users Manual for Release 2.0. Environmental Research Laboratory, US EPA, Athens, GA, USA.

Rawls, W.J., 1983. Estimating Soil Bulk Density from Particle-Size Analysis and Organic Matter Content. J. Soil Sci., 135(2): 123-125.

Rawls, W.J., Brakensiek, D.L. and Saxton, K.F., 1981. Soil Water Characteristics. Am. Soc. Agric. Eng. Paper No. 81-2510,

Ryan, T.A., Joiner, B.L. and Ryan, B.F., 1976. Minitab Student Handbook. PWS, 341 pp.

Soil Survey Staff, 1983. Soils of England and Wales: (In six map sheets at 1:250 000 scale). Ordnance Survey, Southampton.

Soil Survey Staff, 1996. Keys to Soil Taxonomy. Seventh Edition. United States Department of Agriculture, Natural Resources Conservation Service, Washington, USA. 644 pp.

Tinsley, J., 1950. The determination of organic carbon in soils by dichromate mixtures. Trans. 4th int. Congr. Soil Sci. 1, 161-4.

Walker, A., Calvet, R., Del Re, A.A.M, Pestemer, W. and Hollis, J.M., 1995. Evaluation and improvement of mathematical models of pesticide mobility in soils and assessment of their potential to predict contamination of water systems. Report of EU project IV.2.1 (Environment Research Programme) -

Risk from agricultural technologies, EV5V-CT92-0226, Blackwell Wissenschafts-Verlag, Berlin, 115 pp.