ELSEVIER

GEODERMA

Geoderma 91 (1999) 27–45

Modelling soil attribute depth functions with equal-area quadratic smoothing splines

T.F.A. Bishop [a,*], A.B. McBratney [a], G.M. Laslett [b]

[a] Department of Agricultural Chemistry and Soil Science, Ross St Building, AO3, University of Sydney, Sydney NSW 2006, Australia

[b] CSIRO Mathematical and Information Sciences, Private Bag 10, Clayton South MDC, Vic 3169, Australia

Received 25 June 1998; accepted 21 December 1998

Abstract

The objective of this paper is to test the ability of equal-area quadratic splines to predict soil depth functions based on bulk horizon data. In addition, the possibility of improving the prediction quality by the use of additional samples from the top and/or bottom of soil profiles along with horizon data is examined. The predictive performance of the splines is compared with that of exponential decay functions, and 1st and 2nd degree polynomials. In addition, the predictive quality of the conventional horizon data is examined. The measure of predictive performance used is the root mean square error values calculated from differences between the 'true' depth function and the fitted depth function. The 'true' depth functions were derived from the intensive sampling and laboratory analysis of soil profiles. Three soil profiles were sampled; a Red Podzolic Soil (Red Kurosol), Podzol (Aeric Podosol) and Krasnozem (Red Ferrosol). The soil attributes that were measured included; pH, electrical conductivity (EC), clay %, sand %, organic carbon %, gravimetric water content at -33 kPa and air dry. The results clearly indicated the superiority of equal-area quadratric splines in predicting depth functions. Such splines depend on a parameter, λ that controls goodness-of-fit vs. roughness. Their quality of fit varied with the λ value used and it was found that a λ value of 0.1 was the best overall predictor of the depth functions. The results also showed that using additional samples from the top and/or bottom of the soil profiles improved the prediction quality of the spline functions.

Keywords: splines; equal-area quadratic splines; depth functions; modelling

* Corresponding author. Fax: +61-2351-3706; E-mail: t.bishop@agec.usyd.edu.au

0016-7061/99/$ - see front matter
PII: S0016-7061(99)00003-8

1. Introduction

Soil attributes in general vary continuously with depth in a soil profile (Russell and Moore, 1968). In contrast to this, the traditional method of sampling soil involves dividing a soil profile into horizons. The number of horizons and the position of each is generally based on attributes easily observed in the field, such as morphological soil properties. From each horizon, a bulk sample is taken and it is assumed to represent the average value for a soil attribute over the depth interval from which it is sampled. Consequently, the analysis of well mixed bulk horizon samples when presented as a depth function is stepped rather than varying continuously with depth (Colwell, 1970).

The discontinuous nature of depth functions derived from bulk horizon data may cause inaccuracies when attempting to predict the value of an attribute at specific depths within a soil profile, in addition the minima and maxima may be damped (Ponce-Hernandez et al., 1986). This is a source of concern as, in many different fields of soil science, accurate soil-depth data are becoming increasingly important, especially for use in computer simulation models for environmental and agricultural purposes.

Due to the possible inadequacies of horizon data in accurately representing depth functions, soil scientists have in the past attempted to use various methods to modify horizon data in order to make it more continuous. The earliest attempts to do this involved drawing freehand curves (see Jenny, 1941) between data points, where it was assumed that the value obtained for each soil attribute from the measurement of bulk horizon samples corresponded to the mid-point depth for the horizon concerned. More sophisticated methods evolved, examples being the fitting of exponential decay functions (Brewer, 1968; Russell and Moore, 1968; Moore et al., 1972), linear regression and polynomials of the 2nd-degree to 5th-degree to soil depth-data (Campbell et al., 1970; Colwell, 1970).

A disadvantage of using polynomials (and also exponential decay functions) is that any local variation in the soil profile affects the quality of fit everywhere else in the profile (Webster, 1978). Consequently they lack flexibility in fitting depth functions and the quality of fit may be quite varied (Webster, 1978). This problem may be solved by the use of spline functions which can fit a smooth curve through any set of data points by fitting piece wise a series of local independent functions over small intervals of a soil profile (Jauregui and Quirino, 1985). In spite of this, all of the functions described previously (including spline functions) fit curves through horizon averages, so that they are based on data that has already been ‘damped’ by averaging. Therefore, they tend to produce smoother depth functions than the real ones (Ponce-Hernandez et al., 1986).

Ponce-Hernandez et al. (1986) proposed a variation of the spline function, called an equal-area spline which they believed would negate the damping effects of using horizon data to model soil attribute depth functions. The key characteristics of the equal-area spline are as follows.

(1) It consists of a series of local quadratic polynomials with the 'knots' or positions of joins being located at horizon boundaries (Fig. 1).

(2) For each horizon, the area to the left of the fitted spline curve above the horizon average (X) is equal to the area to the right of the fitted spline curve below the horizon average (Y) thus ensuring the mean value of the horizon is maintained (Fig. 1).

Ponce-Hernandez et al. (1986) did not use 'real' depth functions to compare the prediction quality of the equal-area splines, but instead used fictitious depth functions derived from freehand drawn curves through horizon data. They then used the horizon data to fit the equal-area splines and compared the results to the fictitious depth functions. However, this does not seem a severe enough test of the equal-area spline approach. Ideally, 'real' depth functions should have been derived from the intensive sampling and analysis of soil profiles. The 'real' depth functions could then have been compared to the depth functions derived from the modelling of the horizon data obtained from the same soil profiles.

The study of Ponce-Hernandez et al. (1986) could also have been improved by modelling the soil depth functions with polynomials or exponential decay functions. This would have enabled direct comparison of the prediction quality of the equal-area spline with other mathematical functions that have been previously used to model depth functions. While the authors believe that the polynomials and exponential decay functions will perform poorly in comparison to the equal-area splines, we believe some comparison should be performed.

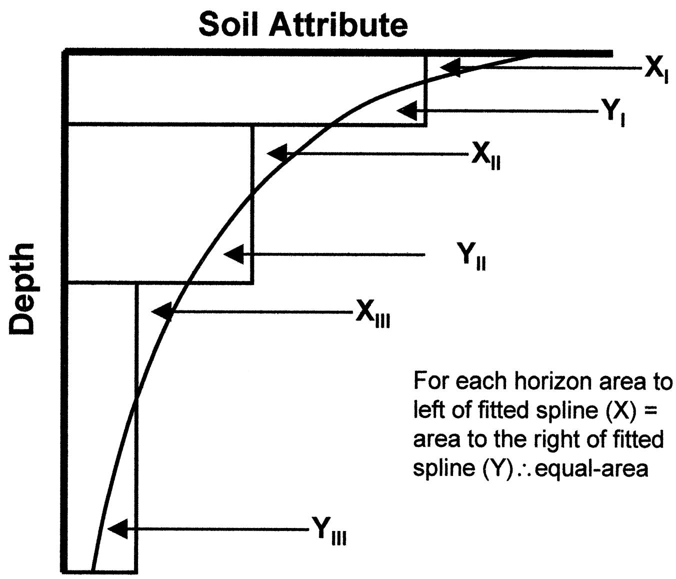

Fig. 1. An equal-area quadratic spline from Ponce-Hernandez et al. (1986).

Ponce-Hernandez et al. (1986) also attempted to see if any improvements could be made to the prediction quality of the equal-area spline by the use of soil samples from the top and bottom of soil profiles in addition to horizon samples. They believed that the additional samples would impose boundary conditions on the fitted function, thus improving the prediction quality at the bottom and top of the profile. Furthermore, for some soil attributes, the greatest variation is near the soil surface (e.g., organic carbon), therefore an additional sample at the top of the profile could further improve the quality of the fit. Their work, while not conclusive, did suggest that the use of the additional samples aided in improving the prediction quality of the equal-area spline.

The equal-area splines developed by Ponce-Hernandez et al. (1986) made no allowance for measurement errors in the data. They also require ad hoc boundary conditions. We overcame both of these deficiencies by defining equal-area quadratic smoothing splines (EAQSS), and we show that they solve a natural variational problem.

As yet, no published work has directly compared the quality of the different mathematical functions to predict soil attribute depth functions. This comparison should involve a range of different soil attributes to decide whether there is a universal mathematical function which fits a wide range of soil attributes or a series of specialist mathematical functions for individual soil attributes.

Our aims are:

1. to the examine the efficacy of the equal-area quadratic splines in predicting soil attribute depth functions based on bulk horizon data,
2. to test whether any improvements can be made to the prediction quality of the mathematical functions by using top and/or bottom profile samples in addition to bulk horizon data, and
3. to compare the results with alternative methods for modelling soil attribute depth functions.

2. Materials and methods

2.1. Sampling

Three different soil groups were sampled, according to the Australian Great Soil Group classification system (Stace et al., 1968). They were: Red Podzolic Soil, Podzol and Krasnozem. In the most recent Australian soil classification system (Isbell, 1996) they may be allocated to the classes Red Kurosol, Aeric Podosol and Red Ferrosol, respectively.

One soil profile was sampled for each of the chosen soil types. To obtain an estimate of the ‘true’ soil attribute depth functions for each profile, soil samples were taken at 2-cm depth intervals to a depth of 1 m (50 in total). In the construction of the 'true' soil attribute depth functions, we assumed that the value of each sample obtained from laboratory analysis represented the value of the soil attribute at the mid-point of the depth interval sampled.

In addition to sampling at 2-cm intervals, the number and position of horizons were chosen as is normally done in the field, that is, based on field morphological characteristics such as colour, structure, and easily measured field attributes such as field texture. From each horizon, a bulk sample was taken for laboratory analysis. The analysis of the bulk horizon samples was used to estimate, using mathematical functions, the true depth function for any particular soil attribute that had been measured.

2.2. Laboratory analyses

The following attributes were measured in the laboratory for modelling of their depth functions in each of the three soil profiles sampled; pH (1:5 soil:0.01 M $CaCl_2$ extracts), EC (1:5 soil:water extracts), clay and sand content (hydrometer method), organic carbon content (colometrically), air-dry w (oven equilibration) and w at -33 kPa (pressure plate method).

2.3. Mathematical functions

The mathematical functions used to estimate the true nature of the soil attribute depth functions were; 1st-degree (Lin), 2nd-degree (2P) and exponential decay functions (EDF) and the equal-area quadratic smoothing splines (EAQSS).

The JMP computer program (Lehman and Sall, 1995) was used to fit all of the functions (with the exception of the equal-area quadratic spline) using least squares regression.

For the exponential decay function, the following equation was used:

$$y = c\mathrm{e}^{-kx} \tag{1}$$

where y = value of the soil attribute, x = depth below the soil surface and c and k are constants.

2.4. Equal-area quadratic splines

It is assumed that the true soil attribute values vary smoothly with depth. This needs to be translated into mathematical terms. We denote depth by x, and the depth function describing the true attribute values by $f(x)$; by smoothness, we mean that $f(x)$ and its first derivative $f'(x)$ are both continuous, and that $f'(x)$ is square integrable.

Denote the depths of the boundaries of the n horizons by $x_0 < x_1, \ldots < x_n$. Often x_0 is the soil surface, so that $x_0 = 0$, but this is not essential. The measurement of the bulk sample from horizon i is assumed to reflect the mean attribute level, apart from measurement error. Mathematically, the measurements y_i $(i = 1, \ldots, n)$ are modelled as

$$y_i = \bar{f}_i + e_i, \tag{2}$$

where $\bar{f}_i = \int_{x_{i-1}}^{x_i} f(x)\,\mathrm{d}x/(x_i x_{i-1})$ is the mean value of $f(x)$ over the interval (x_{i-1}, x_i). The errors e_i are assumed independent, with mean 0 and common variance σ^2.

The function $f(x)$ is unknown, and must be estimated from the horizon data. Various approaches to a similar problem, that of estimating $f(x)$ from point data rather than horizon data, have been suggested by mathematicians over the last few decades. One of the most popular approaches, splines, may be readily adapted to horizon data, and then consists of choosing the $f(x)$ that minimises

$$\frac{1}{n}\sum_{i=1}^{n}\left(y_i - \bar{f}_i\right)^2 + \lambda\int_{x_0}^{x_n} f'(x)^2\,\mathrm{d}x. \tag{3}$$

The first term represents fidelity to the data. The second term measures roughness of the function $f(x)$: the rougher the function, as expressed by the magnitude of $f'(x)$, the larger this term becomes. The parameter λ controls the trade-off between the fidelity term and the roughness penalty. It may be proven using the calculus of variations or otherwise, that the minimiser of Eq. (3) is a quadratic spline, which we define below. We note in passing that the roughness penalty could involve higher derivatives of $f(x)$; however, for simplicity we confine our attention to first derivatives in this paper. The choice of λ is itself a non-trivial problem, which we address later; but for the moment we assume λ is given.

We denote the quadratic spline by $s(x)$. Within each horizon, $s(x)$ conforms to a quadratic polynomial $p(x)$. The polynomials $p_i(x)$ and $p_{i+1}(x)$ for two adjacent horizons meet smoothly at the boundary between the horizons. Thus, the overall soil attribute curve is given by

$$s(x) = p_i(x) \text{ for } x_{i-1} \le x \le x_i,\ i = 1,2,\ldots,n. \tag{4}$$

The smoothness conditions are

$$\begin{aligned} p_i(x_i) &= p_{i+1}(x_i) \\ p_i'(x_i) &= p_{i+1}'(x_i) \end{aligned} \qquad \text{for } i = 1, 2, \ldots\ n-1, \tag{5}$$

and

$$\begin{aligned} p_1'(x_0) &= 0 \\ p_n'(x_n) &= 0. \end{aligned} \tag{6}$$

These latter two conditions mean that $s(x)$ is a so-called natural spline. It has $3n$ unknown parameters, and there are $2n$ smoothness conditions. If, in addition, $\bar{s}_1, \ldots, \bar{s}_n$ are given, the number of equations equals the number of parameters, so the spline is fully determined.

The n equations to determine $\bar{s}_1, \ldots, \bar{s}_n$ are obtained by minimising Eq. (3). Let R be the $(n-1)\times(n-1)$ symmetric tridiagonal matrix with diagonal elements $R_{ii}=2(x_{i+1}-x_{i-1})$ and off-diagonal elements $R_{i+1,i}=R_{i,i+1}=x_{i+1}-x_i$. Also, let Q be the $(n-1)\times n$ matrix with $Q_{ii}=-1$, $Q_{i,i+1}=1$ and $Q_{ij}=0$ otherwise. It is readily proved that the n additional equations are

$$\left[I+6n\lambda Q^t R^{-1} Q\right]\bar{s}=y, \tag{7}$$

where I is the identity matrix, $\bar{s}^t=(\bar{s}_1, \ldots, \bar{s}_n)$, $y^t=(y_1, \ldots, y_n)$ and t denotes the transpose.

The equal-area quadratic splines were fitted using code written in S-PLUS (Venables and Ripley, 1994), a statistical computer language. In each case, the λ values 10, 1, 0.1, 0.01, 0.001, 0.0001 and 0.00001 were tried, and the 'best' value selected, as we describe later.

2.5. Sampling used for predicting soil attribute depth functions

The mathematical functions used to predict the soil attribute depth functions were fitted to four different combinations of samples:

1. bulk horizon data;
2. bulk horizon data and a sample from the top of the soil profile (0–2 cm);
3. bulk horizon data and a sample from the bottom of the soil profile (98–100 cm);
4. bulk horizon data and samples from both the top (0–2 cm) and bottom of the soil profile (98–100 cm).

The samples at the top and bottom of the soil profiles, were simply treated as another, albeit thin, horizon when fitting the mathematical functions. Using the polynomials as an example, the functions were fitted through the mid-point depth of the interval concerned, so for the top profile sample the mid-point was at a depth of 1 cm and for the bottom profile sample the mid-point was at a depth of 99 cm.

2.6. Assessment of the prediction quality of the mathematical functions

The prediction quality of the mathematical functions was determined by comparing the predicted depth function with the 'true' soil attribute depth function obtained from measuring the soil profiles at 2 cm depth intervals. Each combination of mathematical function and sampling scheme was fitted to each attribute for each profile. In addition the bulk horizon data were treated exactly like any other mathematical function in that they were also compared to the 'true' depth function to determine its prediction quality.

Estimates of the prediction quality were obtained by the calculation of root mean square error values (RMSE) (Eq. 8) which measure the difference between the true and fitted function. The smaller the RMSE value, the better is the prediction quality of the fitting/sampling scheme concerned.

$$\mathrm{RMSE} = \sqrt{1/n \sum_{i=1}^{n} \left(x_{\mathrm{actual}} - x_{\mathrm{predicted}} \right)^2} \tag{8}$$

2.7. Choosing λ for the quadratic splines

The best λ value was chosen for each combination of sampling scheme, soil attribute and soil profile by plotting the λ vs. RMSE, with the lowest RMSE corresponding to the best λ value. In some cases, extra λ values outside the range described earlier were needed to minimise the RMSE.

3. Results

3.1. Profile descriptions

A profile description containing the results of the laboratory analysis of the horizons, top (0–2 cm) profile samples and bottom (98–100 cm) profile samples is presented in Tables 1–3 for each of the soil profiles used in the experiment.

In determinations of the efficacy of the various fitting/sampling schemes, versatility was important. To ensure the versatility was tested, the determinations of prediction quality were made across all soil profiles for the various aspects tested, i.e., for best function, sampling scheme etc.

3.2. Choosing a λ value

In a normal situation where no prior information is available about a soil profile it would not be possible to predict the best λ value as described in the

Table 1
Profile description of Red Podzolic Soil from Camden, NSW

Horizon	Depth interval (cm)	pH	EC (mS m^{-1})	Clay content (dag kg^{-1})	Sand content (dag kg^{-1})	Organic carbon (dag kg^{-1})	−33 kPa *w* (kg kg^{-1})	AD *w* (kg kg^{-1})
Top	0–2	4.99	10.86	20.9	48.1	4.33	0.271	0.033
A1	0–8	4.74	8.00	21.6	48.5	2.56	0.230	0.024
A2	8–28	4.39	3.28	39.8	47.3	1.42	0.207	0.027
B21	28–68	3.99	8.22	91.4	8.20	1.00	0.395	0.053
B22	68–100	3.85	8.45	87.5	10.6	0.22	0.372	0.056
Bottom	98–100	3.77	8.00	78.6	18.3	0.24	0.375	0.063

Table 2
Profile description of Podzol from Botany, NSW

Horizon	Depth interval (cm)	pH	EC ($mS\ m^{-1}$)	Clay content ($dag\ kg^{-1}$)	Sand content ($dag\ kg^{-1}$)	Organic carbon ($dag\ kg^{-1}$)	-33 kPa w ($kg\ kg^{-1}$)	AD w ($kg\ kg^{-1}$)
Top	0 − 2	4.65	5.65	1.46	97.07	1.09	0.044	0.007
A1	0–30	4.32	4.06	1.51	97.99	0.89	0.026	0.005
A2	30–68	4.54	0.68	0.50	99.01	0.11	0.001	0.001
B2	68–100	4.21	0.24	3.34	96.66	0.48	0.032	0.007
Bottom	98–100	3.96	0.21	1.51	96.63	0.22	0.028	0.005

Section 2. Therefore a general pattern was looked for in the data for λ values which were most commonly among the best and could therefore be used as guidelines for further examination. Two examples are presented in Figs. 2 and 3. From this cursory examination it was found that λ values of 0.01, 0.1, and 1 were typically among the best. Therefore, the following examination of the prediction quality concentrates on the prediction quality of each combination of sampling, soil attribute and profile for the quadratic splines fitted with λ values of 0.01, 0.1 and 1. In addition the λ values with the best and worst RMSE were used in the following comparisons, along with the mean RMSE value across all values of λ. This was done to examine the variation in prediction quality when the chosen λ value may be unknowingly among the worst or best.

3.3. Examples of modelled soil attribute depth functions

Due to the large number of soil attribute depth functions that were modelled, it is impossible to present individually a significant portion of the fitted depth functions. Therefore, only a small example set of modelled soil attribute depth functions will be shown. These figures show the improvement in the prediction quality of the equal-area quadratic spline functions when additional samples

Table 3
Profile description of Krasnozem from Robertson, NSW

Horizon	Depth interval (cm)	pH	EC ($mS\ m^{-1}$)	Clay content ($dag\ kg^{-1}$)	Sand content ($dag\ kg^{-1}$)	Organic carbon ($dag\ kg^{-1}$)	-33 kPa w ($kg\ kg^{-1}$)	AD w ($kg\ kg^{-1}$)
Top	0–2	4.02	3.92	73.13	12.6	0.29	0.331	0.059
A11	0–4	4.05	4.34	66.32	17.55	0.49	0.364	0.068
A12	4–44	4.22	3.30	69.03	15.72	0.66	0.369	0.067
B2	44–100	4.04	4.68	72.30	13.24	0.27	0.333	0.063
Bottom	98–100	3.92	3.43	80.31	8.98	0.21	0.359	0.064

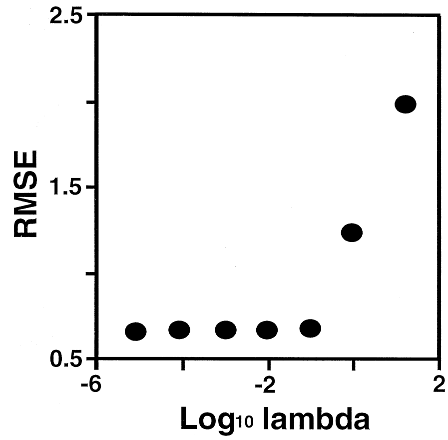

Fig. 2. RMSE value vs. $\log_{10} \lambda$ of equal-area quadratic splines fitted through EC horizon data in addition to top and bottom profile samples for the Red Podzolic Soil.

from the top and bottom of the profile are used for modelling the depth functions (Fig. 4) as compared to when only horizon data is used to model the depth function (Fig. 5). Furthermore Fig. 4 illustrates the effect of different λ values with a value of 0.1 fitting the data better than a value of 10.

3.4. Determination of the accuracy of the fitting / sampling schemes

The calculation of RMSE (Eq. 8) was used to determine the accuracy of the various fitting/sampling schemes used to model the soil attribute depth func-

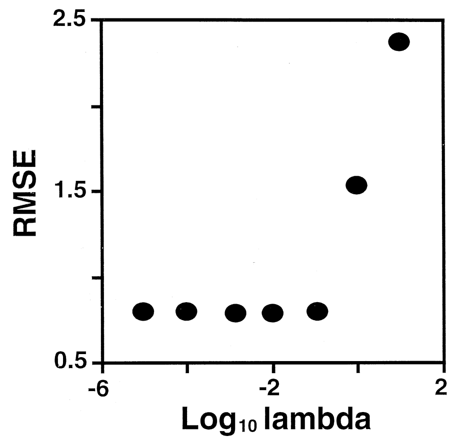

Fig. 3. RMSE value vs. $\log_{10} \lambda$ of equal-area quadratic splines fitted through EC horizon data for the Red Podzolic Soil.

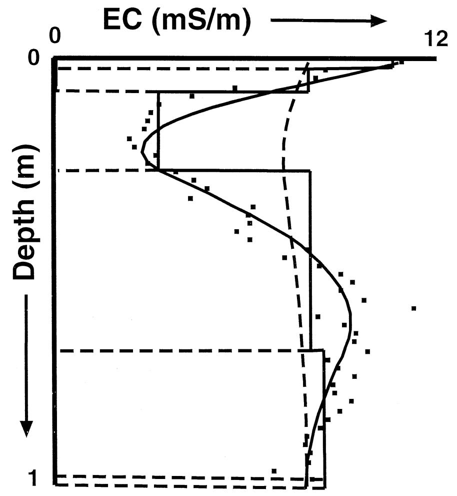

Fig. 4. Equal-area quadratic splines with λ values of 0.1 (solid line) and 10 (dashed line) fitted through horizon data in addition to top and bottom profile sample data for the Red Podzolic soil.

tions. The RMSE gives the average absolute difference between the fitting/sampling scheme and the true soil attribute depth function. The difference is in the same units as the soil attribute being predicted. The magnitude of the RMSE value is, in part, dependent on the magnitude of the soil attribute in question. As a result, RMSE values cannot be averaged across the three different soil profiles

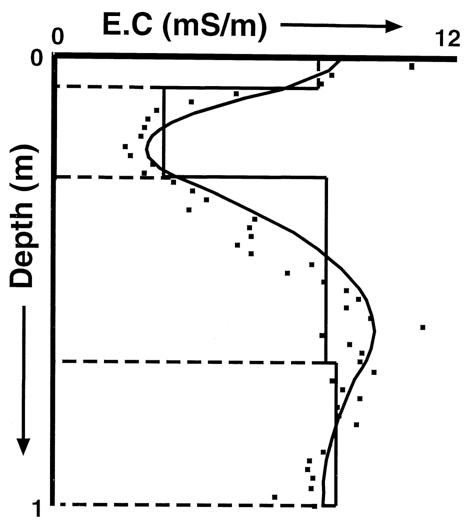

Fig. 5. Equal-area quadratic spline (λ values of 0.1) fitted through horizon data for the Red Podzolic soil.

to find the most accurate fitting/sampling scheme for predicting a particular soil attribute. For example, it would not be feasible to compare, by averaging, the RMSE of a fitting/sampling scheme in predicting the clay content of a Podzol and a Krasnozem because a Podzol has a clay content of only a few percent and the Krasnozem may have clay contents up to 60–80%. Therefore, simply averaging the RMSE values could give rise to uneven weighting between the three soil types used in this study.

A more robust method for determining the most accurate fitting/sampling scheme was needed. A simple ranking system was decided upon, where the RMSE of each fitting/sampling scheme for each soil attribute and for each profile was ranked from the smallest (1st) to largest (last).

3.5. General determination of the most accurate fitting procedure

An initial analysis of the results was performed to determine which fitting procedure was the most accurate predictor for all the soil attributes. To do this, the mean and standard deviation of rank was determined for each type of function across all the soil attributes, and for all soil types. The sampling scheme was not considered in this analysis, so the four different sampling schemes used for each function were treated as being the same. The mean rank and standard deviation of rank were then plotted for each fitting procedure where the closest to the bottom left-hand corner was the most accurate predictor (Fig. 6). The standard deviation of rank is important as it is a measure of the consistency of prediction of the fitting procedure, but in marginal cases the mean rank is deemed more important.

Fig. 6 and Table 4 shows that quadratic spline functions are generally the most accurate predictors of soil attribute depth functions. Not surprisingly, the spline fitted with the best λ value is the best predictor followed by the 0.1λ,

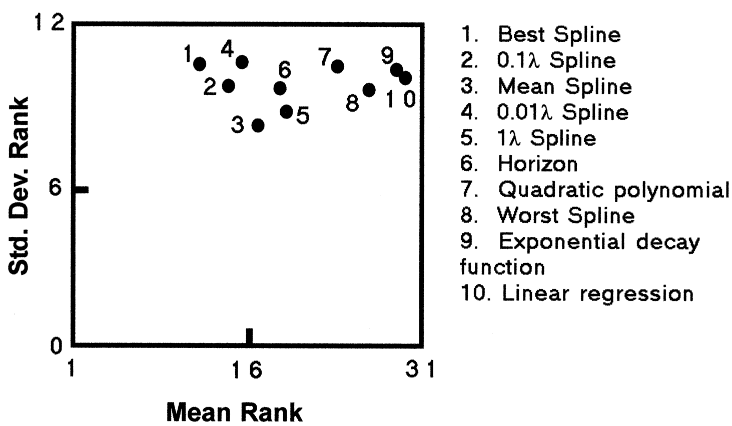

Fig. 6. Testing the accuracy of the fitting procedures across all soil attributes.

Table 4
Results of the determination of accuracy of the fitting procedures across all soil attributes (Fig. 6)

Fitting procedure	Mean rank	Standard deviation of rank
Best spline	11.86	10.62
0.1 λ spline	14.38	9.75
0.01 λ spline	15.58	10.63
Mean spline	16.89	8.19
Horizon data	18.77	9.79
1 λ spline	19.56	8.74
Quadratic polynomial	23.82	10.56
Worst spline	26.65	9.78
Exponential decay function	29.01	10.53
Linear regression	29.81	10.18

0.01λ and mean spline. Next is the 1λ spline and the horizon data which have very similar prediction qualities. Another noticeable result is that the spline fitted with worst λ value is still a better predictor than the linear regression or exponential decay functions.

3.6. General determination of the most accurate fitting / sampling combination

To find the most accurate combination of fitting and sampling procedures for all soil attributes, the same procedure was used as explained previously. The only exception being that the different sampling procedures used for each fitting procedure were considered separately enabling a determination of the most accurate fitting/sampling combination.

Fig. 7 and Table 5 show the dominance of the spline functions. Key features of the results are the spline fitted with the best λ being the best predictor for each different sampling scheme, followed by the 0.1λ which was marginally better than the 0.01λ splines. Furthermore for each predictor, in general, the sampling scheme giving the best predictions was the one using the top and bottom profile samples, followed by using only bottom or top profile samples which had similar predictive ability. The worst sampling scheme was to use only horizon data.

3.7. Determination of the most accurate fitting / sampling combination for individual soil attributes

To ascertain whether any fitting/sampling combination was particularly accurate in predicting an individual soil attribute, the mean rank and standard deviation of rank for individual soil attributes were obtained across the three soil types sampled. In each case, the standard deviation of rank was plotted against

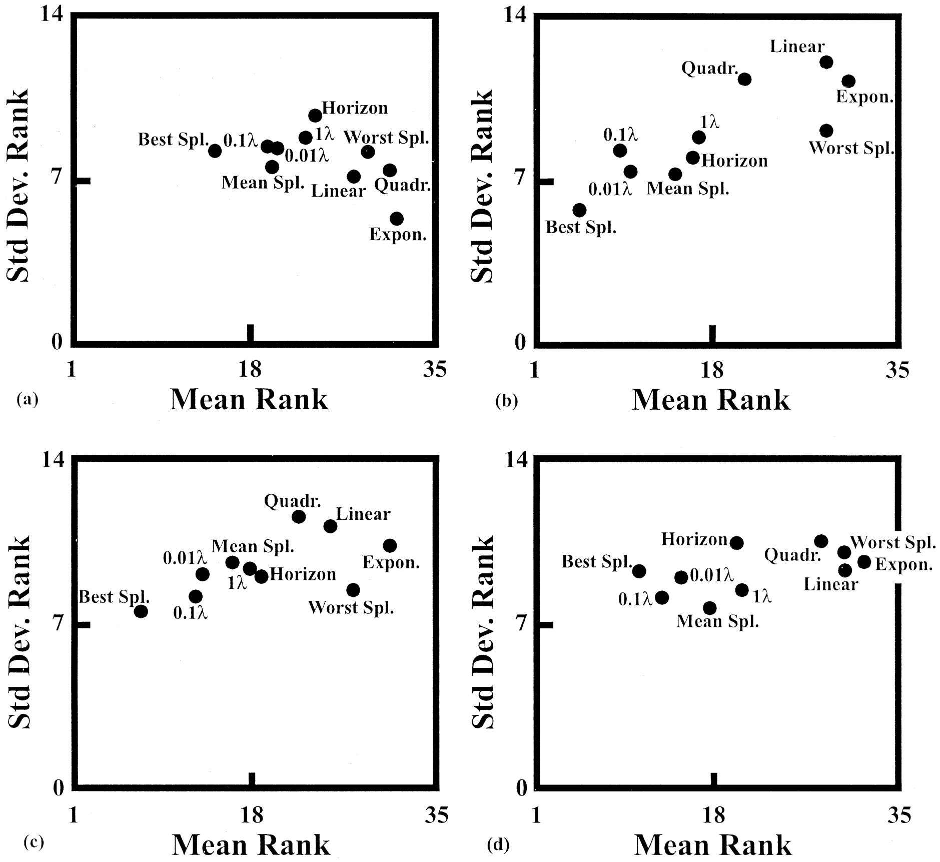

Fig. 7. Testing the accuracy of different combinations of fitting and sampling procedures (a) horizon data (b) horizon data, top and bottom profile data (c) horizon and bottom profile data (d) horizon and top profile data.

Table 5
Results of the ten most accurate fitting/sampling combinations across all soil attributes (Fig. 7)

Fitting/sampling comb.	Mean rank	Standard deviation of rank
Best Spline-horizon, top + bottom	5.00	5.94
Best Spline-horizon, bottom	7.52	7.45
0.1 λ Spline-horizon, top + bottom	9.10	8.46
0.01 λ Spline-horizon, top + bottom	9.86	7.55
Best Spline-horizon, top	10.42	9.29
0.1 λ Spline-horizon, bottom	12.86	8.13
0.1 λ Spline-horizon, top	13.09	8.12
0.01 λ Spline-horizon, bottom	13.33	9.05
Best Spline-horizon, normal	13.81	8.21
0.01 λ Spline-horizon, top	14.09	8.99

Table 6
Results of the determinations of the most accurate (top 10) fitting/sampling combinations for individual soil attributes[abc]

Rank	AD w	−33 kPa w	Clay	Sand	OC	pH	EC
1st	BSpTB^{4}	BSpTB$^{1.3}$	BSpB$^{8.3}$	BSpT$^{5.7}$	BSpB$^{4.3}$	BSpTB^{2}	BSpTB$^{2.3}$
2nd	BSpB$^{4.7}$	0.1λTB^{3}	0.1λT^{10}	BSpTB^{7}	0.1λB$^{5.7}$	0.1λTB$^{3.3}$	BSpB$^{4.7}$
3rd	0.1λTB^{6}	0.01λTB^{4}	2PTB$^{9.3}$	0.1λT$^{7.7}$	0.1λTB^{7}	0.01λTB$^{3.7}$	BSpN$^{7.7}$
4th	BSpN$^{6.7}$	BSpB$^{4.3}$	BSpTB$^{9.7}$	0.1λTB$^{8.7}$	BSpTB$^{6.7}$	BSpT^{4}	BSpT$^{9.3}$
5th	BSpT$^{8.7}$	0.01λB$^{10.3}$	0.01λT$^{11.3}$	1λTB$^{9.3}$	2PB$^{7.7}$	0.01λT^{5}	0.01λT^{12}
6th	0.01λTB^{9}	0.1λB$^{10.7}$	BSpN$^{12.3}$	0.1λB^{10}	0.01λTB^{8}	0.1λT$^{6.7}$	1λTB^{13}
7th	0.01λB$^{12.3}$	BSpN^{11}	1SpN$^{12.7}$	BSpB$^{8.7}$	0.01λB^{8}	BSpB^{10}	1λN$^{14.3}$
8th	MSpTB^{13}	MSpTB^{11}	0.1λTB$^{12.3}$	HorB$^{10.3}$	HorB^{10}	BSpN$^{11.3}$	1λB$^{15.3}$
9th	0.01λB$^{13.7}$	BSpT$^{11.7}$	0.01λTB$^{12.3}$	HorTB^{12}	BSpT^{13}	0.01λB^{13}	0.1λN$^{14.3}$
10th	HorTB^{14}	HorT$^{12.3}$	1λB$^{13.7}$	0.01λB$^{11.3}$	1λTB$^{15.3}$	0.1λB$^{13.3}$	0.01λTB^{16}

[a]Numbers in superscript: mean ranking.

[b]Key for table (fitting procedure used): BSp: Best spline, MSp; Mean spline, 0.01λ: 0.01λ Spline, 0.1λ: 0.1λ Spline, 1λ: 1λ Spline, Hor: Horizon sampling, 2P: Quadratic polynomial.

[c]Key for table (sampling procedure used): N: horizon data only, TB: horizon data, top and bottom profile samples used, T: horizon data and top profile sample used, B: horizon data and bottom profile sample used.

mean rank to determine the most accurate fitting/sampling procedure. A summary of the results is presented in Table 6.

4. Discussion

4.1. The predictive quality of traditional methods of modelling depth functions

The assessments of the accuracy of the various fitting procedures showed that the spline functions were the best predictors of soil attribute depth functions. As expected the polynomial and exponential decay functions performed poorly when used to model soil attribute depth functions.

When only the fitting procedures were considered, the horizon data was more accurate in reconstructing the depth functions than any of the other functions used (excepting the spline functions) (Fig. 6 and Table 4). This is a very poor result for these functions, as they attempt to improve upon the horizon data. Furthermore, when considering the overall prediction quality of the fitting/sampling combinations to predict soil attribute depth functions, Fig. 7 shows that, excluding spline functions, the horizon data with samples from the top and bottom of the profile outperforms all of the other fitting/sampling combinations.

More specifically, exponential decay functions underperformed when attempting to predict soil attribute depth functions (Fig. 7 and Fig. 8). In particular they performed poorly when attempting to predict organic carbon depth functions for which they are traditionally believed to be good predictors (Webster, 1978). This is evidenced by their absence from the top ten predictors of organic carbon depth functions (Table 6). Other studies (Brewer (1968), Russell and Moore (1968) and Moore et al. (1972)) have used exponential decay functions to model organic carbon depth functions but admittedly this was not from bulk horizon samples.

4.2. The predictive quality of the equal-area quadratic splines

In nearly every aspect, the spline functions were the best predictors of depth functions. First, when only the fitting procedures were considered, the spline functions were the best predictors (Fig. 6 and Table 4). Furthermore, when the combinations of fitting/sample procedures were considered, the top ten combinations involved splines (Fig. 7 and Table 5). When the best combinations of fitting/sampling procedures for individual attributes were considered, spline functions dominated the top 10 predictors of each soil attribute (Table 6).

The spline functions were the best predictors due to their high degree of flexibility compared to the other fitting methods (Webster, 1978). The flexibility is due to fitting quadratic polynomials through each sample interval, e.g., horizon depth, therefore local variations in the soil profile do not affect the fit elsewhere (Ponce-Hernandez et al., 1986). The other mathematical functions used were less flexible and therefore affected by local variations in soil profiles. The other reason for the spline functions being the best predictors was the equal-area criteria which minimised the effect of 'damping' caused by horizon sampling (Ponce-Hernandez et al., 1986).

4.3. The best λ value to use for predicting depth functions

While the spline functions were obviously the best predictor of depth functions, the best λ value to use has yet to be discussed. For the quadratic splines to be of any use, a standard a priori λ value must be recommended which can perform reasonably well under most circumstances without resorting to the impracticable method described earlier (Section 2) to determine the best λ value. Initially it was believed that the best λ value to use would be 0.01, 0.1 or 1. Examination of Figs. 6 and 7 and Tables 4–6 indicates that a λ values of 0.1 is marginally better than 0.01, and both are significantly better than a λ of 1.

4.4. The 'best' sampling scheme

Since spline functions are obviously the best predictors of depth functions and 0.1 has been chosen as the standard λ value to use when modelling depth functions, it was decided to consider the best sampling combination in conjunction with only the 0.1λ spline function. Simply by looking at Fig. 7 and Table 5, it can be seen that in order, the best sampling schemes in conjunction with spline functions are the top and bottom of the profile samples followed by the using only the bottom of the profile samples which is marginally better than using only top of the profile samples. Using only the horizon data gives the worst predictions.

These results are expected, since the more information available from which depth functions can be reconstructed, the more accurate the prediction should be. Therefore, the sampling procedure with the extra samples from the top and bottom profile samples would be expected to be the best and modelling depth functions based solely on horizon data would be excepted to be the worst.

4.5. The possibility of 'specialist' functions for modelling the depth functions of individual soil attribute

Another aspect of the results that has not been discussed, is the possibility of specialist fitting procedures for individual soil attributes rather than one universal fitting procedure for all soil attributes. Due to the limited number of soil profiles used in the experiment it is not possible to make a definitive statement about specialist fitting procedures. Examination of Table 6 shows that for five of the soil attributes measured (excluding the best λ splines), the best fitting/sampling combination had an approximate mean ranking of between five and ten. Since it would be expected that a specialist fitting/sampling procedure would have a lower mean ranking, none of the combinations seem to be potential specialists for a particular soil attribute. The exception to this is the modelling of the pH and -33 kPa which for the 0.1λ spline using top and bottom profile samples had a mean ranking of 3.3 and 3, respectively.

This results is not as significant as its seems since this fitting/sampling combination has already been identified as the best general predictor of depth functions. A more significant result would have occurred if an otherwise poor predictor, e.g. quadratic polynomial, was particularly good at predicting an individual soil attribute. Even so, due to the limited number of soil types that were sampled, there still remains a possibility that 'specialist' functions for certain soil attributes do exist.

5. Conclusions and future research

In this limited study, spline functions were clearly the best general predictors of soil attribute depth functions. In choosing a λ value for the splines, it is recommended that 0.1 or 0.01 be used as these were in general, the best predictors.

A consideration of the optimum sampling procedure to use in conjunction with the spline functions revealed that significant improvements could be made to the prediction quality by taking extra samples from the top and bottom of the soil profiles in addition to the horizon samples. The best predictions were achieved when both of the extra samples were used to model the depth functions. If only one extra sample could be taken and measured, it was inconclusive whether it was better to use a sample from either the top or bottom of the profile. The splines modelled only through horizon data were worse predictors. Nevertheless, for each possible sampling procedure, the spline function outperformed all other fitting procedures. In particular the 0.1 and 0.01 λ splines outperformed the other splines.

For individual soil attributes, no other fitting procedures were obviously better than the spline functions. Therefore, no specialist fitting procedures for individual soil attributes were evident. Due to this, the 0.01 or 0.1 λ 'equal-area' quadratic spline function for the soil types and attributes studied is considered to be the best predictor of depth functions for all possible sampling combinations.

The method presented here have as at least two direct applications. First, the method can be used to provide soil data over specified depth ranges for one-dimensional simulation models of soil processes. Secondly, in three-dimensional spatial prediction of soil properties, the method proposed here is a natural candidate for dealing with vertical non-stationarity with a change of support.

Clearly, a limited set of soil profile classes and attributes were investigated here. For an unbiased assessment of the results in this paper, ideally a separate data set should have been used to examine the quality of the 'equal-area quadratic spline', in particular the prediction quality with the recommended λ values of 0.01 and 0.1. Unfortunately the collection of further validation sets is time consuming and expensive as evidenced by Ponce-Hernandez et al. (1986) not using any real data in their paper. Therefore the results in this paper should be only be taken as a preliminary study.

References

Brewer, R., 1968. Clay illuviation as a factor in particle-size differentiation in soil profiles. Transactions of the 9th International Congress of Soil Science 4, 489–499.

Campbell, N.A., Mulcahy, M.J., McArthur, W.M., 1970. Numerical classification of soil profiles on the basis of field morphological properties. Australian Journal of Soil Research 8, 43–58.

Colwell, J.D., 1970. A statistical–chemical characterization of four great soil groups in southern New South Wales based on orthogonal polynomials. Australian Journal of Soil Research 20, 221–238.

Isbell, R.F., 1996. The Australian Soil Classification. CSIRO, Melbourne, Australia.

Jauregui, M.A., Quirino, P., 1985. Spline response functions for direct and carry over effects involving a single nutrient. Soil Science Society of America Journal 49, 140–145.

Jenny, H., 1941. Factors of Soil Formation: a System of Quantitative Pedology. McGraw-Hill, New York and London.

Lehman, A., Sall, J., 1995. JMP statistics and graphics guide. SAS Institute, Cary, NC.

Moore, A.W., Russell, J.S., Ward, W.T., 1972. Numerical analysis of soils: a comparison of three soil profile models with field classifications. Journal of Soil Science 23, 193–209.

Ponce-Hernandez, R., Marriott, F.H.C., Beckett, P.H.T., 1986. An improved method for reconstructing a soil profile from analyses of a small number of samples. Journal of Soil Science 37, 455–467.

Russell, J.S., Moore, A.W., 1968. Comparison of different depth weightings in the numerical analysis of anisotropic soil profile data. Transactions of the 9th International Congress of Soil Science 4, 205–213.

Stace, H.C.T., Hubble, G.D., Brewer, R., Northcote, K.H., Sleeman, J.R., Mulcahy, M.J., Hallsworth, E.G., 1968. A handbook of Australian soils. Rellim, South Australia.

Venables, W.N., Ripley, B.D., 1994. Modern Applied Statistics with S-Plus. Springer, New York.

Webster, R., 1978. Mathematical treatment of soil information. Transactions of the 11th International Congress of Soil Science 3, 161–190.