

Chapter 7

Pedotransfer Functions and Soil Inference Systems

José Padarian, Jason Morris, Budiman Minasny, and Alex. B. McBratney

“You can’t make a silk purse out of a sow’s ear”.

Jonathan Swift
Anglo-Irish essayist (1667–1745)

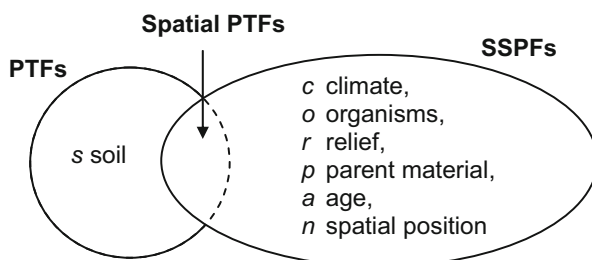
7.1 Introduction

The term *pedotransfer function* (PTF) was coined by Bouma (1989) as ‘translating data we have into what we need’. Pedotransfer functions are regression functions used to predict soil properties that would be otherwise infeasible to obtain. Typical reasons for this infeasibility include, but are not limited to, the cost, time, difficulty or hazard involved in procuring direct measurements. Each PTF is developed around some insight into a soil’s physical, chemical or biological properties that relates a set of input parameters (predictor properties) to an output parameter (a predicted property).

Pedotransfer functions (PTFs) have multiples uses. They are essential, for example, in soil carbon stock assessment (Chap. 23) based on legacy soil data, where bulk density is usually not measured. PTFs can also be used to estimate soil organic carbon pools required in soil carbon evolution models. In digital soil mapping (Chap. 12), the use of pedotransfer functions is to provide more useful information in relation to soil attributes or soil functions. Pedotransfer functions can further be used to estimate the soil’s condition or capability (e.g. available water capacity). The predicted properties resulting from PTFs can be used as inputs into process-based simulation models to run scenarios on the effects of different agricultural management on soil functioning, drainage, evapotranspiration and biomass yields.

J. Padarian (✉) • J. Morris • B. Minasny • A.B. McBratney
Sydney Institute of Agriculture & School of Life and Environmental Sciences,
The University of Sydney, Sydney, NSW 2006, Australia
e-mail: jose.padarian@sydney.edu.au; jason.morris@sydney.edu.au;
budiman.minasny@sydney.edu.au; alex.mcbratney@sydney.edu.au

Fig. 7.1 A Venn diagram showing the relationship between pedotransfer functions (PTFs), soil spatial prediction functions (SSPFs) and, their intersection, spatial PTFs



Some consider prediction of soil attributes from environmental variables (e.g. climate and topographic indices) in a spatial context as *pedotransfer functions*, but we would caution against that. We called this particular case the soil spatial prediction functions (SSPFs) (see Chap. 12 for more details). Pedotransfer functions *sensu stricto* are when we are predicting soil attributes from other soil attributes, or $S = f(s)$. There is a possible intersection or area of overlap, between PTFs and SSPFs (e.g. the spatial component), which results in what we call *spatial pedotransfer functions*. Figure 7.1 illustrates the differences and possible overlap between PTFs and SSPFs. Pachepsky et al. (2001) and Romano and Palladino (2002) illustrate examples of spatial or contextual pedotransfer functions, but they are examples of $S = f(s,r)$.

7.2 A Brief History of Pedotransfer Functions

Reviews on the development and the use of PTFs can be found in Pachepsky et al. (1999, 2015) and Wosten et al. (2001). Most of these reviews, however, are limited to the prediction of soil hydraulic properties, which regulate the retention and movement of water and chemicals in soils.

The concept of using empirical relations to predict soil properties can be traced to Briggs and McLane (1907) and Briggs and Shantz (1912) in their work on determining the wilting coefficient. Furthermore, various ‘rule of thumbs’ were formulated to estimate various soil properties. Probably because of its particular difficulty and cost of measurement, the most comprehensive research in developing PTFs has been for the estimation of water retention. With the introduction of the concepts of field capacity (FC) and permanent wilting point (PWP) by Veihmeyer and Hendricksen (1927), research during the period of 1950–1980 attempted to correlate particle-size distribution, bulk density and organic matter content with water content at field capacity (FC, θ at -33 kPa), permanent wilting point (PWP, θ at -1500 kPa) and available water content ($AWC = FC - PWP$). Nielsen and Shaw (1958), for example, presented a parabolic relationship between clay content and PWP from 730 Iowa soils.

In the 1960s various papers dealt with the estimation of FC, PWP and AWC, notably in a series of papers by Salter and Williams (1965a, b, 1966, 1967, 1969).

These explored relationships between texture classes and available water capacity, which are now known as *class PTFs*. Salter and Williams also developed functions relating the particle-size distribution to AWC, now known as *continuous PTFs*.

In the 1970s more comprehensive research using large databases was developed. A particularly good example is the study by Hall et al. (1977) who used soil samples from England and Wales. Hall et al. (1977) established field capacity, permanent wilting point, available water content and air capacity as a function of textural class, as well as deriving continuous functions estimating these soil water properties. In the USA, Gupta and Larson (1979) developed 12 functions relating particle-size distribution and organic matter content to water content at water potentials ranging from -4 to -1500 kPa.

With the flourishing development of hydraulic models (van Genuchten 1980) and computer modelling of soil water and solute transport (de Wit and van Keulen 1972), the need for hydraulic properties as input to these models became more and more evident. Clapp and Hornberger (1978) derived average values for the parameters of a power-function water retention curve, sorptivity and saturated hydraulic conductivity for different texture classes. In probably the first research of its kind, Bloemen (1980) derived the relationships between parameters of the Brooks and Corey hydraulic model and particle-size distribution.

Lamp and Kneib (1981) introduced the term *pedofunction*, while Bouma and van Lanen (1986) used the term *transfer function*. To avoid confusion with the terminology, transfer function which is used in other disciplines with many different meanings, Bouma (1989) later termed the *pedotransfer function*.

From the 1990s to the early 2000s, the development of hydraulic PTFs became a popular topic of research. Results of such research have been reported widely from various countries globally, including the UK (Mayr and Jarvis 1999), Australia (Minasny and McBratney 2000), the Netherlands (Wösten et al. 1995), Germany (Scheinost et al. 1997b) and Iran (Ghorbani and Homaei 2002).

Since the late 2000s, the popularity of developing hydraulic PTFs continued (Santra and Das 2008; Twarakavi et al. 2009; Haghverdi et al. 2012). Here, the development of PTFs for special conditions is worth noting, such as saline and saline-alkali soils of Iran (Abbasi et al. 2011), permafrost soils of China (Yi et al. 2013) and volcanic ash soils of Japan (Nanko et al. 2014), and the use and development of PTFs for continental or global extent such as the work presented by Dai et al. (2013) for China, Hollis et al. (2012) and Tóth et al. (2015) for Europe and Glendining et al. (2011) for the world.

In addition, some PTFs consider adjustments because of the differences in criteria and measurements from existing pedotransfer functions. For example, as outlined in the previous chapter of this book (Fig. 5.3), sand fractions are different according to the IUSS/Australian classification (particle diameter 20–2000 μm) and the FAO/USDA criteria (particle diameter 50–2000 μm). Padarian et al. (2012) give equations for converting between these two classification systems. On the other hand, Henderson and Bui (2002) established relationships between pH measured in water and pH measured in CaCl_2 .

Although most PTFs have been developed to predict soil hydraulic properties, they are not restricted to hydraulic properties only. PTFs for estimating soil physical, mechanical, chemical and biological properties have also been developed (Table 7.1). In addition, PTFs were developed to also predict processes such as deep

Table 7.1 Some examples of pedotransfer functions

Predicted soil properties	Predictor variables	Authors
Physical properties		
Infiltration rate after a certain period	Initial water content, moisture deficit, total porosity, non-capillary porosity, hydraulic conductivity	Canarache et al. (1968)
Soil thermal conductivity	Texture, organic matter content, water content	De Vries (1966), Hubrechts and Feyen (1996)
Bulk density	Particle-size distribution	Rawls (1983) Hollis et al. (2012)
Infiltration parameters	Particle-size distribution, bulk density, organic C content, initial water content, root content	van de Genachte et al. (1996)
Gas diffusivity	Air-filled porosity at -10 kPa	Moldrup et al. (2000)
Mechanical properties		
Soil mechanical resistance	Organic carbon content, clay content, bulk density	Mirreh and Ketcheson (1972), da Silva and Kay (1997)
Soil shrinkage curve	Clay content	Crescimanno and Provenzano (1999)
Volumetric shrinkage, liquid limit, plastic limit, plasticity index	Organic matter content, clay content, CEC	Mbagwu and Abekh (1998)
Degree of overconsolidation	Bulk density, void ratio	McBride and Joose (1996)
Rate of structural change	Organic matter content, clay content	Rasiah and Kay (1994)
Soil erodibility factor	Geometric mean particle-size, clay and organic matter content	Torri et al. (1997)
Chemical properties		
Cation exchange capacity (CEC)	Clay content, organic matter content, pH	Bell and van Keulen (1995), Curtin and Rostad (1997)
Critical P level, P buffer coefficient	Clay content	Cox (1994), Chen et al. (1997)
Soil organic matter	Soil colour	Fernandez et al. (1988)
P sorption	pH in NaF	Gilkes and Hughes (1994)
pH buffering capacity	Organic matter content, clay content	Wong et al. (2013)

(continued)

Table 7.1 (continued)

Predicted soil properties	Predictor variables	Authors
Al saturation	Base saturation, organic carbon content, pH	Jones (1984)
P saturation	Extractable P, Al	Kleinman et al. (1999)
K/Ca exchange	Clay content, extractable K	Scheinost et al. (1997a)
Total nitrogen	Organic carbon content, pH, sand and clay content	Glendining et al. (2011)
As and Cd sorption	Clay content, pH, organic carbon content, dithionite extractable Fe	Schug et al. (1999)
Phosphorous (P) adsorption	Clay content, pH, soil colour	Sheinost and Schwertmann (1995)
Cd sorption coefficient	Clay content, organic carbon content, pH	Springob et al. (1998)
Haematite content	Soil colour	Torrent et al. (1983)
Biological properties		
Microbial phylotype richness and diversity	pH	Fierer and Jackson (2006)
Respiration rate	Water content	Wildung et al. (1975)
Nitrogen mineralisation parameters	CEC, total N, organic carbon content, silt and clay content	Rasiah (1995)
OC pools	Total OC, clay	Weihermueller et al. 2013

percolation. For example, Selle and Huwe (2005) used a regression tree approach to simplify process-based models to identify key soil and environmental variables which govern percolation. Wessolek et al. (2008) called these hydro-pedotransfer functions, as soil and hydrological variables are used to predict other soil processes. Wessolek et al. (2008) developed empirical functions that predict deep percolation and evapotranspiration from soil conditions, vegetation and land uses.

Pachepsky et al. (2015) reviewed more recent developments in PTFs and identified research gaps that require future work:

- The need for sufficient upscaling of PTFs. PTFs were mainly generated on point observations, and many applications require simulations on regional or continental extent. An example is saturated hydraulic conductivity which is highly dependent on the measurement support.
- The need for more regional or specific PTFs for saline soils, calcareous and gypsiferous soils, peat soils, paddy soils, soils with well-expressed shrink-swell behaviour, and soils affected by freeze-thaw cycles.
- The need for parameters governing biogeochemical processes, such as in soil carbon and nitrogen evolution models, where parameters are related to organic matter pools (e.g. Weihermueller et al. 2013). For these cases, soil heat transfer and water availability inputs can be improved.

- The need to expand work on the spatial and temporal structure of PTFs which is not well known.
- The use of PTFs in large-scale projects, where soil information is usually not represented properly. In soil carbon stock assessment studies, bulk density is usually not measured, so PTFs for bulk density are required, which can be a main source of uncertainty (Hollis et al. 2012).

7.3 Developing Pedotransfer Functions

The basic steps for developing PTFs are simple – in theory, $S = f(s)$, and therefore:

1. Collect a sufficient data set of soil properties (S and s) that are suspected to having empirical relationships to each other.
2. Set aside a certain fraction of the data for developing the PTFs, and use the remaining data for testing the performance of the PTFs (e.g. an 80:20% split of the data set).
3. Choose a modelling method f for analysing the data (e.g. linear regression, neural networks or other machine learning algorithms), and subsequently develop the empirical equations.
4. Test the empirical equations on the testing data to show their validity.
5. Calculate the output uncertainty.

7.4 Predictors

There are several sources of information that can be used to predict soil properties and that can be considered as input for pedotransfer functions. Here, we will present the use of PTFs and their potential predictors which are sourced from the laboratory, field description (including soil morphology) as well as the soil electromagnetic spectrum.

7.4.1 Laboratory Data

Laboratory analysis of soil samples is usually conducted to allocate a particular soil profile to an existing soil class. The high cost of laboratory analysis, however, drove the development of empirical relationships by relating more easily or routinely measured soil properties to other attributes that are, for example, more useful for soil management purposes. One of the well-known examples is the estimation of available water capacity from particle-size distribution. The development in

pedotransfer functions is boosted by the availability of large national or regional soil databases, which allows the use of machine learning tools. The most useful variable in predicting soil physical properties is perhaps clay content, as it affects moisture retention, soil strength and many physical and chemical processes. Routine analysis usually lacks of physical data. Research is still mainly focusing on improving the prediction of hydraulic properties, such as water retention and saturated hydraulic conductivity. Some simpler analysis, however, has also been utilised to estimate more difficult-to-measure properties, such as pH in sodium fluoride which is an indication of phosphorous sorption capacity (Gilkes and Hughes 1994).

7.4.2 Field Description and Soil Morphology

Most research has been focused on correlating laboratory-determined soil properties with more difficult-to-measure properties, mainly because of the availability of comprehensive soil survey databases and the presumption that these properties are most appropriate for predictive purposes. However, it has also been recognised for some time that soil morphological description could be used as predictor (O'Neal 1949, 1952; McKeague et al. 1984; McKenzie and McLeod 1989; McKenzie and Jacquier 1997).

Calhoun et al. (2001) contended that soil morphology and field description have been underutilised in the development of pedotransfer functions. They presented the representation of Jenny's state factors through the variables' physiography, parent material, horizon, field texture and structure as collected in soil surveys for predicting bulk density. They demonstrated that morphology and field descriptors account for more variability in predicting bulk density than laboratory measurement of particle size and organic carbon. Physiographic description and soil morphological characterisation (slope gradient, position of the slope and horizon classes) were also found as useful predictors of water retention (Rawls and Pachepsky 2002).

Several studies have been successful in predicting hydraulic conductivity by using soil morphological features (e.g. O'Neal 1952; McKeague et al. 1982). However, the descriptive systems and interpretative guidelines in conventional soil survey have been largely qualitative and only appropriate for a given range of soils. McKenzie et al. (1991) found that several published descriptive systems for inferring hydraulic properties provided poor predictions for a limited range of soils from South Australia. McKenzie and Jacquier (1997) reasoned that good predictive relationships should only be expected when the field criteria used have a logical physical connection with hydraulic properties. They further postulated that predictive systems that develop direct relationships between hydraulic properties and field criteria of physical significance should be superior to systems that rely on classified entities such as horizons or soil series. They devised a simple visual estimate of areal porosity and found that saturated conductivity can be estimated from field texture, grade of structure, areal porosity, bulk density, dispersion index and horizon type. A similar idea was performed by Lin et al. (1999), who

converted morphological properties to scores which are related to water flow. From these studies, it was concluded that additional morphological descriptors to those routinely surveyed may be needed to improve the predictive capacity.

7.4.3 Handheld and On-the-Go Proximal Soil Sensing and Remote Sensing

7.4.3.1 Handheld, Stationary, Proximal Soil Sensing

As outlined in Chap. 5, in traditional soil surveys, soil scientists used the visible light spectrum through the Munsell soil colour chart to determine soil colour and the presence of pedological features like mottles or concretions. Furthermore, it was discussed in the previous chapter that developments in spectroscopy have resulted in an increase in the potential for soil analysis, and we will include a short summary of its capability here (Fig. 7.2). Diffuse reflectance infrared spectroscopy in both the visible-near (400–700–2500 nm) and mid-infrared ranges (2500–25,000 nm) allows rapid acquisition of soil information in the field or in the laboratory. Diffuse reflectance infrared spectroscopy is based on the fact that molecules have specific frequencies at which they rotate or vibrate corresponding to discrete energy levels. Absorption spectra of compounds are a unique reflection of their molecular structure. Spectral signatures of soil materials are characterised by their reflectance to a particular wavelength in the electromagnetic spectrum. Soil spectra in the vis-NIR and MIR ranges can be used to estimate a range of soil physical, chemical and biological properties simultaneously. Good results were reported for measurement of total C, total N, clay and sand content, CEC and microbial activity (Soriano-Disla et al. 2014).

Mid-infrared (MIR) spectroscopy usually produces better predictions than vis-NIR. The use of MIR also enables estimation of various soil organic carbon pools derived from tedious and time-consuming physical fractionation procedures. These pools can be used as inputs in soil carbon evolution models. Vis-NIR spectrometers particularly are used extensively and gained popularity in soil science because they are also available in a portable format and easy and ready to use in the field and require minimal or even no sample preparation. Reviews on the use of vis-NIR for predicting soil properties can be found in Stenberg et al. (2010) and Soriano-Disla et al. (2014).

Because soil is a complex mixture of materials, it is difficult to assign specific features of the spectra to specific chemical components. Ultraspectral data obtained from infrared spectrometers contain thousands of reflectance values as a function of wavelength. Since there are more predictor variables than the observations and predicted soil attributes as outlined in the previous chapter, methods that reduce the dimension of the spectra are required. Principal component regression and partial least squares (PLS) methods are commonly utilised. Principal component regression reduces the dimension of the spectra via principal component analysis and then form

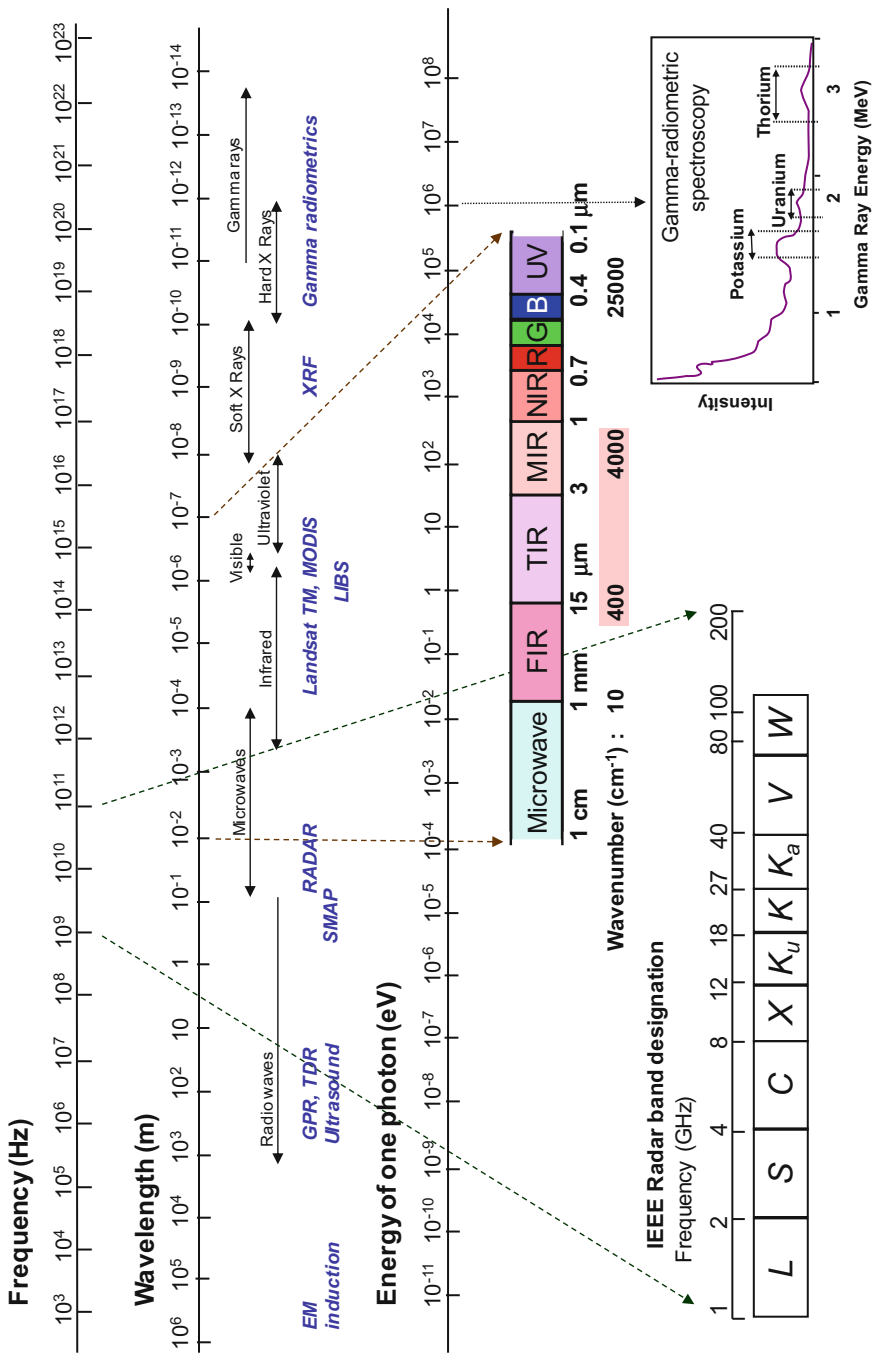


Fig. 7.2 The electromagnetic spectrum and regions useful for soil measurement

linear regression between the principal components and soil attributes (Martens and Naes 1989; Chang et al. 2001). Partial least squares (PLS) (Martens and Naes 1989) extracts successive linear combinations of the spectra, which optimally address the combined goals of explaining response variation and explaining predictor variation. Other machine learning techniques that are capable of variable (wavelength) selection have also been found useful (e.g. Minasny and McBratney 2008; Sarajith et al. 2016).

In addition to vis-NIR spectroscopy, the direct measurement of the elemental concentration of soils in the field also became possible using energy-based portable X-ray fluorescence (XRF) devices (Weindorf et al. 2012). Bulk density can also be estimated utilising photogrammetry via a digital single-lens camera or laser scanning (Bauer et al. 2014; Rossi et al. 2008).

7.4.3.2 On-the-Go Proximal Soil Sensing

While we can collect detailed soil information at limited locations using conventional methods of soil analysis and interpolate resulting values across space and time using geostatistics, in some instances it would be more beneficial if we could directly measure soil information at a fine spatial scale (e.g. measurements every 2–20 m). In this instance, proximal soil sensing offers a cost- and time-effective solution (Viscarra Rossel et al. 2010). Proximal soil sensing acquires information about soil through the use of field-based sensors that are placed in proximity to the soil (within 2 m) or within the soil body, which is in contrast to remote sensing (McBratney et al. 2011a, b). The development and use of on-the-go proximal soil sensing techniques is motivated by the need for high-resolution spatial and temporal soil information. Proximal soil sensors operate on a range of frequencies in the electromagnetic spectrum, from microwaves to gamma rays. These sensing devices either measure soil properties directly or can be used to make inferences via PTFs about specific soil properties. Often sensors are also used simultaneously to overcome the limitations of single-sensor data interpretation (Wong et al. 2010). For example, electromagnetic induction instruments (EMI) are used to measure the soil's electrical conductivity, a highly valuable soil property that is influenced by soil porosity, moisture content, salinity, temperature and the amount and composition of soil colloids.

Ground-penetrating radar, electrical resistivity as well as electrical conductivity sensors are available to monitor the spatial distribution of soil moisture (Adamchuk et al. 2004). In addition, gamma ray spectrometers have been used to measure the amount of potassium, uranium and thorium in the upper soil profile which is most likely directly related to the parent material the surveyed soil originated from (Dickson and Scott 1997). Local PTFs have been developed to estimate soil attributes (such as clay and organic carbon content) from the sensed variables (e.g. bulk electrical conductivity, gamma K).

As outlined in Chap. 5, portable sensors can now be used in the field on profile and core faces for pedological studies, which is termed digital soil morphometrics

(Hartemink and Minasny 2014). Field observation via proximal sensors and PTFs should be fused in an inference system into a powerful approach for estimating a range of soil properties for pedological studies, precision farming or contamination assessment (Horta et al. 2015).

7.4.3.3 Remote Sensing

The value of remote sensing over proximal sensing is that large spatial extents can be covered quickly with many estimates. The inferred value of remotely sensed data either airborne or satellite sourced has been shown to be an efficient means of assessing the condition of natural resources at reasonably broad scales (and this will be discussed further in Chap. 13). The remotely sensed data can include spectral, radar, thermal and radiometric signals. These reflect the environmental and soil condition and are known to be associated with soil properties. Mulder et al. (2011) reviewed the application of optical and microwave remote sensing for soil and terrain mapping. Soil properties that have been measured include mineralogy, texture, soil iron content, soil moisture content, soil organic carbon content, soil salinity and carbonate content. Its use for soil mapping is, however, hampered by vegetation cover. Nevertheless, indicators, such as plant functional groups, NDVI and productivity changes, can be used as indications of soil properties.

The application of remotely sensed infrared data for mapping soil clay content and mineralogy is demonstrated by Mulder et al. (2013) and Gomez et al. (2015). Some studies demonstrated that time series data collected from remotely sensed data can be used to derive soil hydraulic properties. Dimitrov et al. (2014) derived soil hydraulic parameters, surface roughness and soil moisture of a tilled bare soil plot using measured brightness temperatures at 1.4 GHz (L-band), rainfall and potential soil evaporation. This required a radiative transfer model and a soil hydrologic model combined with an optimisation routine.

7.5 Modelling Approaches

Approaches to develop PTFs can be purely empirical or physico-empirical. Empirical approaches attempt to find relationships between the predictor and predicted variables using regression analysis or various machine learning models. In a physico-empirical approach, the soil properties are derived based on some physical principles. For example, in water retention curve prediction, Arya and Paris (1981) translated the particle-size distribution into a water retention curve by converting solid mass fractions to water content and pore-size distribution into hydraulic potential by means of the capillary equation. Zeiliger et al. (2000) proposed an additive model for soil water retention, which assumed that water retention of a soil can be approximated by the sum of the components of water retention of its textural composition.

Considering the type of data we wish to predict, we can distinguish single point and parametric PTFs. Single point PTFs predict a soil property, while parametric PTFs predict parameters of a model.

Most survey agencies have their own ‘rule of thumb’ for predicting soil properties. One form is a look-up table, which usually relates field texture class to properties such as clay content, available water capacity, etc. These rules or tables are usually derived from experience and expert knowledge or from means of properties for a particular class in a soil database.

For the continuous predicted variables, a range of machine learning models can be used to derive PTFs, finding relationships between the predictor and predicted variables. Many of the modern regression techniques are described in Hastie et al. (2009). The methods range from linear regression, generalised linear models (GLM) and generalised additive models (GAM) to regression trees, random forests, neural networks, genetic programming and fuzzy systems. Most of these tools are available in commercial and open-source projects. R (<https://www.r-project.org>) and Python (<https://www.python.org/>) are commonly used by the scientific community, because they offer many free-of-use advanced mathematical and machine learning tools.

The predictive power and interpretability vary between models depending on their complexity. Tables 7.2 and 7.3 provide a guideline for various models. The more complex the model, the more parameters it will have, so users need to be aware of the principle of parsimony (which is a general principle that for any model, which provides an adequate fit for a set of data, the one with the fewest parameters

Table 7.2 Common machine learning algorithms used for developing PTFs

<i>Multiple regression</i>
The general purpose of multiple regression is to analyse the relationship between several independent or predictor variables and a dependent or predicted variable. Multiple regression analysis fits a straight line (or plane in an n -dimensional space, where n is the number of independent variables) to the data
<i>Generalised linear models (GLM)</i>
A class of models that arise for a natural generalisation of ordinary linear models. The transformed dependent variable values are predicted from (are linked to) a linear combination of predictor variables; the transformation is referred to as the link function; also different distributions can be assumed for the dependent variable values
<i>Generalised additive models (GAM)</i>
Models that use smoothing techniques, such as splines to identify and represent possible nonlinear relationships between the predictor and predicted variables. GAM is a generalisation of GLM where the linear function of the predictor is replaced by an unspecified (non-parametric) function, obtained by applying a scatterplot smoother to the scatterplot of partial residuals (for the transformed dependent variable values)
<i>Partial least squares (PLS)</i>
This is an alternative to multiple linear regression that can deal with data having more independent variables than observation points. PLS constructs a new set of components as regressor variables which are a linear combination of the original variables. The components in partial least squares are determined by both the response variable(s) and the predictor variables

(continued)

Table 7.2 (continued)

<i>Artificial neural networks</i>
A flexible mathematical structure modelled after the functioning of the nervous system, capable of fitting nonlinear relationships. The essential feature is a network of simple processing elements (neurons) joined together by weights
<i>Regression tree</i>
This is an alternative to multiple regression. Rather than fitting a model to the data, a tree structure is generated by dividing the sample recursively into a number of groups, each division being chosen so as to maximise some measurable difference in the predicted variable in the resulting two groups. The resulting structure provides easy interpretation as variables most important for prediction can be identified quickly
<i>Random forests</i>
Random forest is an extension of regression trees where many trees are generated, varying the number of covariates used and using a bootstrap sampling of the training data. Then, an ensemble model is generated by aggregating the individual trees
<i>Genetic programming</i>
Machine learning method for evolving computer programs, following the concepts of natural selection and genetics, to solve complex problems
<i>Support vector machine (SVM)</i>
A method that looks for an optimal separating hyperplane between two classes by maximising the margin between the closest observations of each class. In a regression case, the observations lie in between the two borders of the margin (supporting vectors), which are separated from the hyperplane by $\pm \varepsilon$ (maximum error)

Based on Hastie et al. (2009)

is to be preferred) (Lark 2001). There is a limit for predictive models; here, users should choose the simplest model that can adequately account for the variation in the prediction. Models with high complexity will appear to fit the data very well; however, these may also cause overfitting or include too many parameters in the model; thus the model will fit the noise of the data. It is recommended to split the data into a calibration and validation set, using the calibration data for fitting and then testing or validating the model with a validation set (see Hastie et al. (2009) for more detail). Wosten et al. (2001) compared the performance of three models to predict water content at -33 kPa from basic soil properties using the same data set. They reported that the accuracy of all three methods was similar and suggested that the improvement of fit may not be expected from the use of different models, but from a better set of data.

7.5.1 Ensemble Models

An alternative to selecting a single predictive model is model ensembles. This consists of creating multiple models and combining them to obtain a single final model. The advantage of this method is that, most of the time, the combined model performs better than any of the individual models in terms of lower error and

Table 7.3 Comparison of different mathematical predictive models

Feature	Linear models	GLM	GAM	Regression tree	Random forests	Neural network	Genetic programming	SVM
Parsimony	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Interpretability	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Nonlinearity	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Handling of mixed data type (qualitative and quantitative)	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Computational efficiency (large data)	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕
Predictive power	⊕	⊕	⊕	⊕	⊕	⊕	⊕	⊕

⊕ good, ⊖ poor ⊕ fair
 Adapted from Hastie et al. (2009)

unaltered bias. This method has been used for almost 200 years as pointed out in an interesting review by Clemen (1989). Baker and Ellison (2008a) discussed various aspects of implementation of ensemble methods for soil studies. In soil science, examples of its use are Baker and Ellison (2008b) who used ensemble ANN for PTFs. Kim et al. (2015) combined two microwave satellite soil moisture products, Malone et al. (2014) combined estimates of soil properties from soil maps and regression kriging prediction, and Padarian et al. (2014) generated an ensemble map of soil available water capacity in Australia.

Guber et al. (2009) suggested the use of all available PTFs in a multimodel prediction technique. They used 19 published PTFs as inputs in Richards' soil water flow equation; the output of the 19 simulations was then combined to obtain a more optimal soil water prediction. The challenge in this type of ensemble method is how to calibrate and to use appropriate weighting for each of the PTF to obtain an optimal prediction.

7.6 Characterising PTF's Performance

As with all numerical methods, there are questions concerning how well any prediction agrees with real observational data. In the literature, PTFs can be characterised by their accuracy, reliability, uncertainty and validity, as well as their ultimate utility. A brief survey of these concepts follows.

7.6.1 Accuracy

Accuracy refers to how well a PTF predicts its target property based on inputs taken from the training data. It measures the performance of a PTF on its training data (a PTF has 'seen' the data). Usually accuracy is expressed in terms of error, the difference between observed and predicted values. Weynants et al. (2009) amongst others used several common statistic measures for evaluating the accuracy of PTFs: the root-mean-square error (RMSE), mean absolute error (MAE), mean error (ME) or bias, coefficient of determination (R^2) and the model efficiency. Accuracy in PTFs can also be computed with other statistics, e.g. the concordance correlation coefficient which measures how close the model predictions fall along a 45-degree line from the origin with the measured data (or a slope of exactly 1) (Lawrence and Lin 1989).

7.6.2 *Reliability*

Reliability in PTFs refers to a PTF's performance in making predictions on data outside its original training data (data a PTF has not 'seen') (Pachepsky and Rawls 1999). A reliable PTF should produce accurate predictions for seen data (data used in the model development process), as well as unseen data (data that had not been used in the model development process) (Baker and Ellison 2008a). Pachepsky and Rawls (1999) state that the reliability of PTFs can be estimated by cross validation, or using an independent data set. In the cross validation method, the training data set is split into two subsets – a calibration set and a validation set; two-thirds of the data for calibration and one-third for testing are a common practice. However, the results from such cross validation can be biased against the data set used. If a PTF is intended for prediction over a region, the independent test data set should contain observations that are unbiased (in statistics, collected based on a random sampling approach). PTFs that lack independent validation result in potentially optimistic assumptions about the functions' predictive performance.

7.6.3 *Validity*

Validity has to do with how appropriate a particular PTF is in predicting a soil property from a given soil sample. The greater the similarity of a soil sample to the soil used to develop a PTF, the greater the assumed validity of that PTF. Validity can be in terms of the geographical and pedological region over which a PTF's original training data were collected. If a PTF is used to predict soil properties outside its original data boundaries, its validity is doubtful (Wösten et al. 1999). Not surprisingly, PTFs perform best on soils having similar parent material and pedogenesis to the soils used to develop them (Bruand et al. 2003). Acutis and Donatelli (2003) stated that validity in PTFs is strictly related to the data set used to develop them. They add that when many PTFs are available to predict the same property, knowing which one to choose is a difficult task.

An important mechanism for establishing validity is stratification or the custom creation of PTFs strictly on soil-type or classification scheme basis. Stratification has been conducted according to soil horizons (Hall et al. 1977); soil classes (Batjes 1996); textural classes (Tietje and Hennings 1996); hydraulic-functional horizons (Wösten et al. 1986); great soil groups, temperature regime and moisture regime (Pachepsky and Rawls 1999); parent material and horizon morphology (Franzmeier 1991); numerical soil class (Williams et al. 1983); and management units (Droogers and Bouma 1997).

Validity can also refer to the congruence between some input data set and the original training set. Despite knowledge that the validity of a given PTF should not be interpolated or extrapolated beyond the pedological origin or soil type on which it is developed, there is still a lack of appropriate information that adequately describes

the calibration data, and, thus, we know very little about where a published PTF may be applied. There is still a lack of a mechanism that can automatically check its validity. Tranter et al. (2009) give a method for determining the valid domain of a PTF based on Mahalanobis distance of the predictor space, cautioning that it is unwise to extrapolate PTFs beyond these bounds. The uncertainty estimates of a PTF can also be a measure of its validity.

7.6.4 *Uncertainty*

Uncertainty refers to the variability in a prediction from its mean value. This occurs because PTF inputs and outputs are random variables. Therefore, they have a mean value and a variance. PTF uncertainty is typically reported as the prediction variance. Uncertainty in PTF prediction can be quantified in terms of structural uncertainty due to flaws in the PTF model, uncertainty due to sampling and measurement errors and parameter uncertainty of the PTF.

Vereecken and Herbst (2004) suggest three approaches to handling uncertainty in PTFs: (1) Compute the RMSE at 90% confidence; (2) Quantify parameter uncertainty using a covariance in PTFs during the calibration process; and (3) Use a Monte Carlo analysis to quantify parameter uncertainty associated with sampling effects in the calibration database, e.g. the bootstrap method (Efron and Tibshirani 1993).

PTF uncertainty can be computed empirically based on the calibration error using the fuzzy k-means with extragrade (FkME) method given by Tranter et al. (2010). It does not seek to disseminate sources of error but rather expresses uncertainty in the form of a prediction interval determined empirically from the calibration data. The method partitions the predictor space into classes of similar model errors, with each class represented by a prediction interval determined from the empirical distribution of the error. In addition, it also identifies those observations that exist outside the convex hull of the calibration data, thus ascertaining validity of the PTF. Those observations outside the convex hull are considered outliers of the calibration data and subsequently have their uncertainty penalised by a simple multiplier.

7.6.5 *Utility*

Wösten et al. (2001) stated that the utility of PTFs in modelling is defined as the correspondence between measured and simulated functional soil behaviour. This can be interpreted to mean that the authors advocate validating the final use (utility) of the PTFs, not just the PTF predictions. An example would be to develop some PTFs that predict water retention and conductivity and then to use those predictions in crop simulations to predict seasonal water storage. Thus, the validation occurs

at the seasonal water storage level, not at the level of the individual predictions of water retention and conductivity.

7.7 Spatial Pedotransfer Functions

Most PTFs have been calibrated from point source data and assume spatial independence. In digital soil mapping, we are interested in estimating the spatial distribution of soil properties. Pringle et al. (2007) recommended that an investigator who wishes to apply a PTF in a spatially distributed manner first has to establish the spatial scales relevant to their particular study site. Following this, the investigator must ascertain whether these spatial scales correspond to those that are adequately predicted by the available PTFs. Pringle et al. (2007) proposed three aspects of performance in the evaluation of a spatially distributed PTF: (i) the correlation of observed and predicted quantities across different spatial scales, (ii) the reproduction of observed variance across different spatial scales and (iii) the spatial pattern of the model error. For an example of predicting water retention across a 5 km transect, they showed that the tested PTFs performed quite well in reproducing a general spatial pattern of soil water retention; however, the magnitude of observed variance was underestimated. Springer and Cundy (1987) compared the parameters of the Green-Ampt infiltration equation from field measurements and those calculated from PTFs. They showed that the mean and variance of the parameters when estimated by PTFs were not preserved; the variances are always lower. The spatial trends and cross-correlations amongst the parameters were also reduced. They further used the PTFs to simulate overland flow and found that the results were significantly different when using field-measured parameters.

When measured properties are spatially limited, spatial prediction is required to generate a continuous map. Combination of spatial interpolation methods such as kriging and PTFs can generate a continuous map, and there are two possibilities to combine them. The first approach is to first interpolate related soil properties at unvisited locations using kriging and then to apply PTFs to the interpolated variables. The second approach applies PTFs to point measurements and then interpolates the predicted results. Bocneau (1998) compared these approaches to estimate CEC in West Flanders province, Belgium, and found that the performance of both methods is almost equal. Sinowski et al. (1997) compared these approaches in estimating the water retention curve and found that the first approach yields better prediction.

Heuvelink and Pebesma (1999) discussed the role of support or scale. As most PTFs were derived from point sources, they are not valid at the block support. This means that in the situation where the PTF input is available at point support and where output is required at block support, spatial aggregation should take place after the functions are calculated. It is essential to separate spatial aggregation from spatial interpolation. Interpolation should better take place before a function or model is executed because this enables a more efficient use of the spatial distribution

characteristics of individual inputs. When a model is executed with interpolated inputs, it is important to note the uncertainty of the interpolation.

7.8 Soil Inference Systems

While there are many similar pedotransfer functions generated using new or existing data sets, there seems to be much less effort in gathering and using the available PTFs. McBratney et al. (2002) proposed a soil inference system that would match the available input with the most appropriate PTF to predict properties with the lowest uncertainty. The soil inference system was proposed as a way of collecting and making better use of pedotransfer functions that have been abundantly generated. McBratney et al. (2002) demonstrated the first approach towards building a soil inference system is to create a very rudimentary system in the form of a specially adapted spreadsheet. Such a rudimentary inference system has two essentially new features. Firstly, it contains a suite of published pedotransfer functions, and the output of one PTF can act as the input to other functions (if no measured data are available). Secondly, the uncertainties in estimates are inputs, and the uncertainties of subsequent calculations are performed. The input consists of the essential soil properties.

The inference engine will work in the following manner:

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

There have been some attempts at pattern matching of PTFs using a distance metric (Tranter et al. 2009) or nearest-neighbour algorithm (Nemes et al. 2006). However, there have been no research applications that do what soil inference systems (SINFERS) aim to do, to build a system that would chain the PTF predictions together while accounting for uncertainty.

Morris et al. (2016) built an expert system software, which uses rules to select appropriate PTFs and predicts new property values and error estimates. SINFERS can use the estimated property values as new inputs, which can trigger more matching patterns and more PTFs to ‘fire’ cyclically until the knowledge base is exhausted and SINFERS has inferred everything it can about what it was originally given.

7.9 Soil Spectral Inference Systems

As discussed in Sect. 7.4.3, soil spectroscopy and proximal soil sensing research have mainly focused on spectral calibration and prediction of a range of soil properties using multivariate statistics. PTF research, on the other hand, is mainly focusing on predicting soil model parameters from other soil properties. There is no

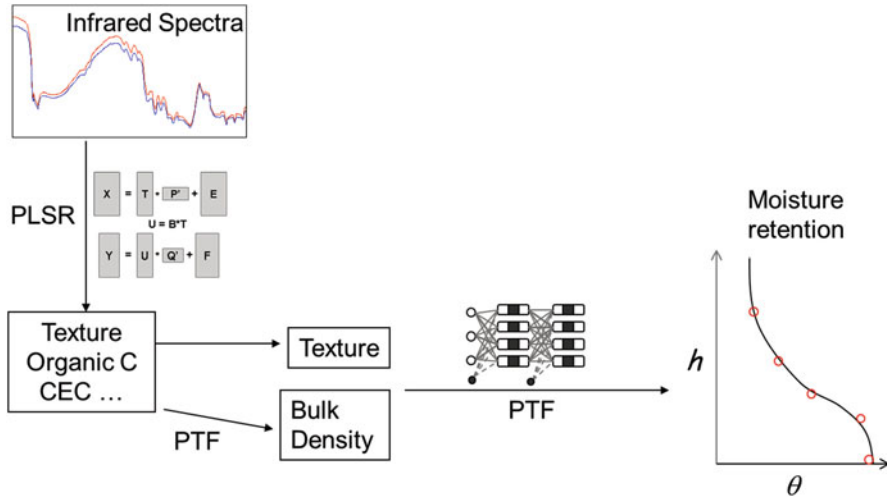


Fig. 7.3 An example of a spectral soil inference system. Soil spectra were used to predict important soil properties, and these properties were, in turn, used to predict other properties, applying well established PTFs

real connection between these research areas which have the same aim, to predict one soil property from other soil properties $S = f(s)$.

It is desired to develop soil spectral calibrations for a complete suite of soil physical, chemical and biological properties. However, this might not be possible, mainly for two reasons: (i) Not all soil properties show a spectral response and (ii) the development of a comprehensive soil spectral library is quite challenging. McBratney et al. (2006) proposed a spectral soil inference system (SPEC-SINFERS), where soil diffuse reflectance spectroscopy is linked with PTFs. SPEC-SINFERS uses soil spectra to estimate various basic soil properties which are then used to infer other important and functional soil properties via pedotransfer functions (Fig. 7.3). An important feature to be considered is the propagation of both input and model uncertainties. Tranter et al. (2008) demonstrated the use of the SPEC-SINFERS approach in predicting volumetric soil water retention. This is for sure a research area that requires future investigations.

References

- Abbasi Y, Ghanbarian-Alavijeh B, Liaghat AM, Shorafa M (2011) Evaluation of pedotransfer functions for estimating soil water retention curve of saline and saline-alkali soils of Iran. *Pedosphere* 21(2):230–237
- Acutis M, Donatelli M (2003) SOILPAR 2.00: software to estimate soil hydrological parameters and functions. *Eur J Agron* 18(3):373–377

- Adamchuk VI, Hummel JW, Morgan MT, Upadhyaya SK (2004) On-the-go soil sensors for precision agriculture. *Comput Electron Agric* 44(1):71–91
- Arya LM, Paris JF (1981) A physicoempirical model to predict soil moisture characteristics from particle size distribution and bulk density data. *Soil Sci Soc Am J* 45(1023):1030
- Baker L, Ellison D (2008a) The wisdom of crowds – ensembles and modules in environmental modelling. *Geoderma* 147(1):1–7
- Baker L, Ellison D (2008b) Optimisation of pedotransfer functions using an artificial neural network ensemble method. *Geoderma* 144(1):212–224
- Batjes NH (1996) Development of a world data set of soil water retention properties using pedotransfer rules. *Geoderma* 71:31–52
- Bauer T, Strauss P, Murer E (2014) A photogrammetric method for calculating soil bulk density. *J Plant Nutr Soil Sci* 177(4):496–499
- Bell MA, Van Keulen H (1995) Soil pedotransfer functions for four Mexican soils. *Soil Sci Soc Am J* 59(3):865–871
- Bloemen GW (1980) Calculation of hydraulic conductivities of soils from texture and organic matter content. *J Plant Nutr Soil Sci* 143(5):581–605
- Boucneau G (1998) Geographical information science applied to soils of West-Flanders. PhD thesis. Faculty of Agricultural and Applied Biological Sciences. University of Ghent
- Bouma J (1989) Using soil survey data for quantitative land evaluation. *Adv Soil Sci* 9:177–213
- Bouma J, Van Lanen HAJ (1987) Transfer functions and threshold values: from soil characteristics to land qualities
- Briggs LJ, McLane JW (1907) The moisture equivalent of soils. *USDA Bur Soils Bull* 45:1–23
- Briggs LJ, Shantz HL (1912) The wilting coefficient and its indirect determination. *Bot Gaz* 53(1):20–37
- Bruand A, Fernandez PP, Duval O (2003) Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. *Soil Use Manag* 19(3):232–242
- Calhoun FG, Smeck NE, Slater BL, Bigham JM, Hall GF (2001) Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. *Soil Sci Soc Am J* 65:811–819
- Canarache A, Motoc E, Dumitriu R, Rijtema P (1968) Infiltration rate as related to hydraulic conductivity, moisture deficit and other soil properties. In: *Water in the unsaturated zone. Proceedings of the Wageningen symposium, vol 1*, pp 392–401
- Chang C-W, Laird DW, Mausbach MJ, Hurburgh CR (2001) Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Sci Soc Am J* 65:480–490
- Chen G, Yost RS, Li ZC, Wang X, Cox FR (1997) Uncertainty analysis for knowledge-based decision aids: application to PDSS (Phosphorous Decision Support System). *Agric Syst* 55:461–471
- Clapp RB, Hornberger GM (1978) Empirical equations for some soil hydraulic properties. *Water Resour Res* 14(4):601–604
- Clemen RT (1989) Combining forecasts: a review and annotated bibliography. *Int J Forecast* 5(4):559–583
- Cox FR (1994) Current phosphorus availability indices: characteristics and shortcomings. In: *Soil testing: prospects for improving nutrient recommendations, (soiltestingpros)*, pp 101–113
- Crescimanno G, Provenzano G (1999) Soil shrinkage characteristic curve in clay soils. *Soil Sci Soc Am J* 63:25–32
- Curtin D, Rostad HPW (1997) Cation exchange and buffer potential of Saskatchewan soils estimated from texture, organic matter and pH. *Can J Soil Sci* 77(4):621–626
- Da Silva A, Kay BD (1997) Estimating the least limiting water range of soils from properties and management. *Soil Sci Soc Am J* 61:877–883
- Dai Y, Shangguan W, Duan Q, Liu B, Fu S, Niu G (2013) Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling. *J Hydrometeorol* 14(3):869–887

- DeVries DA (1966) Thermal properties of soils. In: van Wijk WR (ed) *Physics of plant environment*, 2nd edn, pp 231–234
- de Wit CT, van Keulen H (1972) Simulation of transport processes in soils (No. 2, p 100) Pudoc
- Dickson BL, Scott KM (1997) Interpretation of aerial gamma-ray surveys: adding the geochemical factors. *AGSO J Aust Geol Geophys* 17(2):187–200
- Dimitrov M, Vanderborcht J, Kostov KG, Jadoon KZ, Weihermüller L, Jackson TJ, Bindlish R, Pachepsky Y, Schwank M, Vereecken H (2014) Soil hydraulic parameters and surface soil moisture of a tilled bare soil plot inversely derived from L-band brightness temperatures. *Vadose Zone J* 13(1)
- Dröogers P, Bouma J (1997) Soil survey input in exploratory modelling of sustainable soil management practices. *Soil Sci Soc Am J* 61:1704–1710
- Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. In: *Monographs on statistics and applied probability* 57. Chapman & Hall, London
- Fernandez RN, Schulze DG, Coffin DL, Van Scoyoc GE (1988) Color, organic matter, and pesticide adsorption relationships in a soil landscape. *Soil Sci Soc Am J* 52(4):1023–1026
- Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* 103(3):626–631
- Franzmeier DP (1991) Estimation of hydraulic conductivity from effective porosity data for some Indiana soils. *Soil Sci Soc Am J* 55:1801–1803
- Ghorbani DS, Homaei M (2002) Derivation of the retention curve parameters using pedotransfer functions. *J Agric Eng Res* 3(12):1–16
- Gilkes RJ, Hughes JC (1994) Sodium fluoride pH of south-western Australian soils as an indicator of P-sorption. *Aust J Soil Res* 32:755–766
- Glendining MJ, Dailey AG, Powlson DS, Richter GM, Catt JA, Whitmore AP (2011) Pedotransfer functions for estimating total soil nitrogen up to the global scale. *Eur J Soil Sci* 62(1):13–22
- Gomez C, Drost APA, Roger JM (2015) Analysis of the uncertainties affecting predictions of clay contents from VNIR/SWIR hyperspectral data. *Remote Sens Environ* 156:58–70
- Guber AK, Pachepsky YA, van Genuchten MT, Simunek J, Jacques D, Nemes A, Nicholson TJ, Cady RE (2009) Multimodel simulation of water flow in a field soil using pedotransfer functions. *Vadose Zone J* 8(1):1–10
- Gupta S, Larson WE (1979) Estimating soil water retention characteristics from particle size distribution, organic matter percent, and bulk density. *Water Resour Res* 15(6):1633–1635
- Haghverdi A, Cornelis WM, Ghahraman B (2012) A pseudo-continuous neural network approach for developing water retention pedotransfer functions with limited data. *J Hydrol* 442:46–54
- Hall DG, Reeve MJ, Thomasson AJ, Wright VF (1977) Water retention, porosity and density of field soils, Technical Monograph No. 9. Soil Survey of England and Wales, Harpenden
- Hartemink AE, Minasny B (2014) Towards digital soil morphometrics. *Geoderma* 230:305–317
- Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference and prediction*, Springer Series in Statistics, 2nd edn. Springer-Verlag, New York
- Henderson BL, Bui EN (2002) An improved calibration curve between soil pH measured in water and CaCl₂. *Soil Res* 40(8):1399–1405
- Heuvelink GBM, Pebesma EJ (1999) Spatial aggregation and soil process modeling. *Geoderma* 89:47–65
- Hollis JM, Hannam J, Bellamy PH (2012) Empirically derived pedotransfer functions for predicting bulk density in European soils. *Eur J Soil Sci* 63(1):96–109
- Horta A, Malone B, Stockmann U, Minasny B, Bishop TFA, McBratney AB, Pallasser R, Pozza L (2015) Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: a prospective review. *Geoderma* 241:180–209
- Hubrechts L, Feyen J (1996) *Pedotransfer functions for thermal soil properties*. Institute for Land and Water Management. Katholieke University, Leuven
- Jones CA (1984) Estimation of percent aluminum saturation from soil chemical data 1. *Commun Soil Sci Plant Anal* 15(3):327–335

- Kim S, Parinussa RM, Liu YY, Johnson FM, Sharma A (2015) A framework for combining multiple soil moisture retrievals based on maximizing temporal correlation. *Geophys Res Lett* 42(16):6662–6670
- Kleinman PJ, Bryant RB, Reid WS (1999) Development of pedotransfer functions to quantify phosphorus saturation of agricultural soils. *J Environ Qual* 28(6):2026–2030
- Lamp J, Kneib W (1981) Zur quantitativen Erfassung und Bewertung von Pedofunktionen. *Mitteilungen der Deutschen Bodenkundlichen Gesellschaft* 32:695–711
- Lark RM (2001) Some tools for parsimonious modelling and interpretation of within-field variation of soil and crop systems. *Soil Tillage Res* 58:99–111
- Lawrence I, Lin K (1989) A concordance correlation coefficient to evaluate reproducibility. *Biometrics*:255–268
- Lin HS, McInnes KJ, Wilding LP, Hallmark CT (1999) Effects of soil morphology on hydraulic properties: I. Quantification of soil morphology. *Soil Sci Soc Am J* 63:948–954
- Malone BP, Minasny B, Odgers NP, McBratney AB (2014) Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* 232:34–44
- Martens G, Naes T (1989) *Multivariate calibration*. Wiley, New York
- Mayr T, Jarvis NJ (1999) Pedotransfer functions to estimate soil water retention parameters for a modified Brooks–Corey type model. *Geoderma* 91(1):1–9
- McBratney AB, Minasny B, Cattle SR, Vervoort RW (2002) From pedotransfer function to soil inference system. *Geoderma* 109:41–73
- Mbagwu JSC, Abeh OG (1998) Prediction of engineering properties of tropical soils using intrinsic pedological parameters. *Soil Sci* 163(2):93–102
- McBratney AB, Minasny B, Viscarra Rossel R (2006) Spectral soil analysis and inference systems: a powerful combination for solving the soil data crisis. *Geoderma* 136(1–2):272–278
- McBratney AB, Minasny B, Tranter G (2011a) Necessary meta-data for pedotransfer functions. *Geoderma* 160(3):627–629
- McBratney AB, Minasny B, Whelan BM (2011b) Defining proximal soil sensing. 2nd Global workshop – proximal soil sensing. Montreal, Canada
- McBride RA, Joosse PJ (1996) Overconsolidation in agricultural soils: II. Pedotransfer functions for estimating preconsolidation stress. *Soil Sci Soc Am J* 60(2):373–380
- McKeague JA, Wang C, Topp GC (1982) Estimating saturated hydraulic conductivity from soil morphology. *Soil Sci Soc Am J* 46:1239–1244
- McKeague JA, Eilers RG, Thomasson AJ, Reeve MJ, Bouma J, Grossman RB, Favrot JC, Renger M, Strebel O (1984) Tentative assessment of soil survey approaches to the characterization and interpretation of air-water properties of soils. *Geoderma* 34(1):69–100
- McKenzie NJ, Jacquier DW (1997) Improving the field estimation of saturated hydraulic conductivity in soil survey. *Aust J Soil Res* 35:803–825
- McKenzie NJ, MacLeod DA (1989) Relationships between soil morphology and soil properties relevant to irrigated and dryland agriculture. *Aust J Soil Res* 27:235–258
- McKenzie NJ, Smettem KRJ, Ringrose-Voase AJ (1991) Evaluation of methods for inferring air and water properties of soils from field morphology. *Aust J Soil Res* 29:587–602
- Minasny B, McBratney AB (2000) Evaluation and development of hydraulic conductivity pedotransfer functions for Australian soil. *Aust J Soil Res* 38:905–926
- Minasny B, McBratney AB (2008) Regression rules as a tool for predicting soil properties from infrared reflectance spectroscopy. *Chemom Intell Lab Syst* 94(1):72–79
- Mirreh HF, Ketcheson JW (1972) Influence of soil bulk density and matric pressure on soil resistance to penetration. *Can J Soil Sci* 52(3):477–483
- Moldrup P, Olesen T, Gamst J, Schjønning P, Yamaguchi T, Rolston DE (2000) Predicting the gas diffusion coefficient in repacked soil water-induced linear reduction model. *Soil Sci Soc Am J* 64(5):1588–1594
- Morris J (2016) General method for predicting soil data via pattern matching on pedotransfer functions. University of Sydney, Faculty of Agriculture and Environment. <http://hdl.handle.net/2123/14575>

- Mulder VL, De Bruin S, Schaepman ME, Mayr TR (2011) The use of remote sensing in soil and terrain mapping – a review. *Geoderma* 162(1):1–19
- Mulder VL, Plötze M, de Bruin S, Schaepman ME, Mavris C, Kokaly RF, Egli M (2013) Quantifying mineral abundances of complex mixtures by coupling spectral deconvolution of SWIR spectra (2.1–2.4 μm) and regression tree analysis. *Geoderma* 207:279–290
- Nanko K, Ugawa S, Hashimoto S, Imaya A, Kobayashi M, Sakai H, Ishizuka S, Miura S, Tanaka N, Takahashi M, Kaneko S (2014) A pedotransfer function for estimating bulk density of forest soil in Japan affected by volcanic ash. *Geoderma* 213:36–45
- Nemes A, Rawls WJ, Pachepsky YA (2006) Use of the nonparametric nearest neighbor approach to estimate soil hydraulic properties. *Soil Sci Soc Am J* 70(2):327–336
- Nielsen DR, Shaw RH (1958) Estimation of the 15-atmosphere moisture percentage from hydrometer data. *Soil Sci* 86(2):103–105
- O’Neal AM (1949) Some characteristics significant in evaluating permeability. *Soil Sci* 67:403–409
- O’Neal AM (1952) A key for evaluating soil permeability by means of certain field clues. *Soil Sci Soc Am Proc* 16:312–315
- Pachepsky YA, Rawls WJ (1999) Accuracy and reliability of pedotransfer functions as affected by grouping soils. *Soil Sci Soc Am J* 63:1748–1756
- Pachepsky YA, Timlin DJ, Ahuja LR (1999) The current status of pedotransfer functions: their accuracy, reliability and utility in field- and regional-scale modeling. In: Corwin DL, Loage K, Ellsworth TR (eds) *Assessment of non-point source pollution in vadose zone*, Geophysical Monograph, vol 108. American Geophysical Union, Washington, DC, pp 223–234
- Pachepsky YA, Timlin DJ, Rawls WJ (2001) Soil water retention as related to topographic variables. *Soil Sci Soc Am J* 65:1787–1795
- Pachepsky YA, Rajkai K, Tóth B (2015) Pedotransfer in soil physics: trends and outlook—A review—. *Agrokém Talajt* 64(2):339–360
- Padarian J, Minasny B, McBratney A (2012) Using genetic programming to transform from Australian to USDA/FAO soil particle-size classification system. *Soil Res* 50(6):443–446
- Padarian J, Minasny B, McBratney AB, Dalgliesh N (2014) Predicting and mapping the soil available water capacity of Australian wheatbelt. *Geoderma Reg* 2:110–118
- Pringle MJ, Romano N, Minasny B, Chirico GB, Lark RM (2007) Spatial evaluation of pedotransfer functions using wavelet analysis. *J Hydrol* 333(2):182–198
- Rasiah V (1995) Comparison of pedotransfer functions to predict nitrogen mineralization parameters of one and two pool models. *Commun Soil Sci Plant Anal* 26(11–12):1873–1884
- Rasiah V, Kay BD (1994) Characterizing changes in aggregate stability subsequent to introduction of forages. *Soil Sci Soc Am J* 58(3):935–942
- Rawls WJ, Pachepsky YA (2002) Using field topographic descriptors to estimate soil water retention. *Soil Sci* 167:423–435
- Romano N, Palladino M (2002) Prediction of soil water retention using soil physical data and terrain attributes. *J Hydrol* 265:56–75
- Rossi AM, Hirmas DR, Graham RC, Sternberg PD (2008) Bulk density determination by automated three-dimensional laser scanning. *Soil Sci Soc Am J* 72(6):1591–1593
- Salter PJ, Williams JB (1965a) The influence of texture on the moisture characteristics of soils: I. A critical comparison for determining the available water capacity and moisture characteristics curve of a soil. *J Soil Sci* 16:1–15
- Salter PJ, Williams JB (1965b) The influence of texture on the moisture characteristics of soils: II. Available water capacity and moisture release characteristics. *J Soil Sci* 16:310–317
- Salter PJ, Berry G, Williams JB (1966) The influence of texture on the moisture characteristics of soils: III. Quantitative relationships between particle size, composition and available water capacity. *J Soil Sci* 17:93–98
- Salter PJ, Williams JB (1967) The influence of texture on the moisture characteristics of soils: IV. A method of estimating available water capacities of profiles in the field. *J Soil Sci* 18:174–181

- Salter PJ, Williams JB (1969) The influence of texture on the moisture characteristics of soils: V. Relationships between particle-size composition and moisture contents at the upper and lower limits of available water. *J Soil Sci* 20:126–131
- Santra P, Das BS (2008) Pedotransfer functions for soil hydraulic properties developed from a hilly watershed of Eastern India. *Geoderma* 146(3):439–448
- Sarathjith MC, Das BS, Wani SP, Sahrawat KL (2016) Variable indicators for optimum wavelength selection in diffuse reflectance spectroscopy of soils. *Geoderma* 267:1–9
- Scheinost AC, Sinowski W, Auerswald K (1997a) Regionalization of soil buffering functions: a new concept applied to K/Ca exchange curves. *Adv GeoEcol* 30:23–38
- Scheinost AC, Sinowski W, Auerswald K (1997b) Regionalization of soil water retention curves in a highly variable soilscape, I. Developing a new pedotransfer function. *Geoderma* 78:129–143
- Schug B, Hoß T, Düring RA, Gäth S (1999) Regionalization of sorption capacities for arsenic and cadmium. *Plant Soil* 213(1-2):181–187
- Selle B, Huwe B (2005) Optimising soil-hydrological predictions using effective CART models. *Adv Geosci* 5:37–41
- Sinowski W, Scheinost AC, Auerswald K (1997) Regionalization of soil water retention curves in a highly variable soilscape, II. Comparison of regionalization procedures using a pedotransfer function. *Geoderma* 78:145–159
- Soriano-Disla JM, Janik LJ, Viscarra Rossel RA, Macdonald LM, McLaughlin MJ (2014) The performance of visible, near-, and mid-infrared reflectance spectroscopy for prediction of soil physical, chemical, and biological properties. *Appl Spectrosc Rev* 49(2):139–186
- Springer EP, Cundy TW (1987) Field-scale evaluation of infiltration parameters from soil texture for hydrologic analysis. *Water Resour Res* 23:325–334
- Springob G, Böttcher J (1998) Parameterization and regionalization of Cd sorption characteristics of sandy soils. II. Regionalization: Freundlich k estimates by pedotransfer functions. *J Plant Nutr Soil Sci* 161(6):689–696
- Stenberg B, Rossel RAV, Mouazen AM, Wetterlind J (2010) Chapter five-visible and near infrared spectroscopy in soil science. *Adv Agron* 107:163–215
- Tietje O, Hennings V (1996) Accuracy of the saturated hydraulic conductivity prediction by pedotransfer functions compared to the variability within FAO textural classes. *Geoderma* 69:71–84
- Torrent J, Schwertmann U, Fechter H, Alferez F (1983) Quantitative relationships between soil color and hematite content. *Soil Sci* 136(6):354–358
- Torri D, Poesen J, Borselli L (1997) Predictability and uncertainty of the soil erodibility factor using a global dataset. *Catena* 31(1):1–22
- Tóth B, Weynants M, Nemes A, Makó A, Bilas G, Tóth G (2015) New generation of hydraulic pedotransfer functions for Europe. *Eur J Soil Sci* 66(1):226–238
- Tranter G, Minasny B, McBratney AB, Viscarra Rossel R, Murphy B (2008) Comparing spectral soil inference systems and mid-infrared spectroscopy predictions of soil volumetric moisture retention. *Soil Sci Soc Am J* 72:1394–1400
- Tranter G, McBratney AB, Minasny B (2009) Using distance metrics to determine the appropriate domain of pedotransfer function predictions. *Geoderma* 149(3):421–425
- Tranter G, Minasny B, McBratney AB (2010) Estimating pedotransfer function prediction limits using fuzzy-means with extragrades. *Soil Sci Soc Am J* 74(6):1967–1975
- Twarakavi NK, Šimůnek J, Schaap MG (2009) Development of pedotransfer functions for estimation of soil hydraulic parameters using support vector machines. *Soil Sci Soc Am J* 73(5):1443–1452
- Van de Genachte G, Mallants D, Ramos J, Deckers JA, Feyen J (1996) Estimating infiltration parameters from basic soil properties. *Hydrol Process* 10(5):687–701
- Van Genuchten MT (1980) A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci Soc Am J* 44(5):892–898
- Veihmeyer FJ, Hendrickson AH (1927) Soil-moisture conditions in relation to plant growth. *Plant Physiol* 2(1):71
- Vereecken H, Herbst M (2004) Statistical regression. *Dev Soil Sci* 30:3–19

- Viscarra Rossel RA, McBratney AB, Minasny B (2010) Proximal soil sensing. Progress in soil science. Springer, Dordrecht
- Weiermueller L, Graf A, Herbst M, Vereecken H (2013) Simple pedotransfer functions to initialize reactive carbon pools of the RothC model. *Eur J Soil Sci* 64(5):567–575
- Weindorf DC, Zhu Y, McDaniel P, Valerio M, Lynn L, Michaelson G, Clark M, Ping CL (2012) Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and albic horizons. *Geoderma* 189–190:268–277
- Wessolek G, Duijnisveld WHM, Trinks S (2008) Hydro-pedotransfer functions (HPTFs) for predicting annual percolation rate on a regional scale. *J Hydrol* 356(1):17–27
- Weynants M, Vereecken H, Javaux M (2009) Revisiting Vereecken pedotransfer functions: introducing a closed-form hydraulic model. *Vadose Zone J* 8(1):86–95
- Wildung RE, Garland TR, Buschbom RL (1975) The interdependent effects of soil temperature and water content on soil respiration rate and plant root decomposition in arid grassland soils. *Soil Biol Biochem* 7(6):373–378
- Williams J, Prebble JE, Williams WT, Hignett CT (1983) The influence of texture, structure and clay mineralogy on the soil moisture characteristic. *Aust J Soil Res* 21:15–32
- Wong MTF, Wittwer K, Oliver YM, Robertson MJ (2010) Use of EM38 and gamma ray spectrometry as complementary sensors for high-resolution soil property mapping. In: Viscarra Rossel RA, McBratney AB, Minasny B (eds) Proximal soil sensing. Springer, Dordrecht, pp 343–349
- Wong MTF, Webb MJ, Wittwer K (2013) Development of buffer methods and evaluation of pedotransfer functions to estimate pH buffer capacity of highly weathered soils. *Soil Use Manag* 29(1):30–38
- Wösten JHM, Bannink MH, De Gruijter JJ, Bouma J (1986) A procedure to identify different groups of hydraulic conductivity and moisture retention curves for soil horizons. *J Hydrol* 86:133–145
- Wösten JHM, Lilly A, Nemes A, Le Bas C (1999) Development and use of a database of hydraulic properties of European soils. *Geoderma* 90(3):169–185
- Wösten JHM, Finke PA, Jansen MJW (1995) Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. *Geoderma* 66:227–237
- Wösten JHM, Pachepsky YA, Rawls WJ (2001) Pedotransfer functions: bridging gap between available basic soil data and missing soil hydraulic characteristics. *J Hydrol* 251:123–150
- Yi X, Li G, Yin Y (2013) Comparison of three methods to develop pedotransfer functions for the saturated water content and field water capacity in permafrost region. *Cold Reg Sci Technol* 88:10–16
- Zeiliguer AM, Pachepsky YA, Rawls WJ (2000) Estimating water retention of sandy soils using the additivity hypothesis. *Soil Sci* 165:373–383