INTRODUCTION

The 21st century is faced with challenges caused by the spiraling population increase and demand for food, energy, and water; indiscriminate land use conversions; and the catastrophic impacts of abrupt climate change partially attributed to soil degradation.[1] Addressing these challenges necessitate concrete availability of scientifically credible information backed by proactive and strategically targeted management. The judicious management of soil resources holds the key to sustainability of earth’s resources and thus socioeconomic viability.[2] Thus, models that synthesize soil property information into simplified formats can play a critical role in the conceptual understanding of soil systems and the identification of key sustainable practices. Soil quality is defined as the capacity of the soil to support ecosystem functions such as nutrient cycling and water purification and encompasses the fitness of the soil for specific uses (e.g., road construction and agricultural production).[3]

Soil quality can be determined qualitatively (e.g., visually where darker soils are considered to have a high organic matter, thus of high quality) or quantitatively by measuring the soil physical, chemical, and biological properties (Table 1). Soil biological properties are indirectly inferred from other soil properties [e.g., soil organic carbon (SOC)], because of 1) inaccuracies in earthworm counts (i.e., by hand); 2) difficulty in accounting microbial species diversity; and 3) difficulty in interpreting the soil respiration tests. The SOC, a proxy of soil quality, influences the soil ecosystem functions and productivity.[4] High SOC reduces risks of eutrophication from agricultural runoff laden with nitrate (NO3–) or pesticide leaching. Stocks of SOC are computed by multiplying the SOC content (%) with the dry bulk density (ρb), for each soil layer, minus the fraction of fragments >2 mm or by the use of pedotransfer functions (PTFs).

Characterizing soil quality is challenging because of the landscape spatial heterogeneity and the fuzzy definition of soil quality. Despite numerous challenges, the laboratory determination of soil quality has its own problems. For example, the determination of SOC, by chromate oxidation or “wet combustion” method, produces toxic wastes that must be properly disposed and is inaccurate because of the incomplete oxidation of soil organic matter, whereas the dry combustion is expensive and slow. Another technique, loss on ignition, though relatively cost-effective, is inaccurate because certain mineral fractions are also decomposed when heating to the high temperatures required by this method.[5] Laboratory methods can be prohibitively expensive, when the number of samples to be analyzed is large.

Advances in automation, digital computation, or machine learning have ignited new prospects for characterizing and mapping complex features.[6] Maps describe earth surface features, either on hard copy or digitally. Digital soil mapping entails the “computer-assisted” production of digital representations of soil type or soil properties, over a spatial domain.[6] Machine learning methods interpolate and make predictions based on spatially exhaustive environmental variables and are therefore useful for enumerating the interrelationship between soil properties and other covariates; a practical example being the scorpan paradigm[7] that comprises the following: 1) s: soil, other or previously measured attributes of the soil at a point; 2) c: climate, climatic properties of the environment at a point; 3) o: organisms, including land cover and natural vegetation; 4) r: topography, including terrain attributes and classes (e.g., slope, aspect, area, and direction); 5) p: parent material, including lithology; 6) a: age, the time factor; and 7) n: space, spatial or geographic position.

Generally, indicators need to be prioritized, ranked, or weighed based on importance or effectiveness in

Soil Quality Assessment by Interpolation Techniques

Vincent Obade Dr.
Ohio State University, Columbus, Ohio, U.S.A.

Abstract

Environmental hazards such as pollution, flooding, and other repercussions linked to abrupt climate change cause social and economic hardships and are exacerbated by soil degradation. Thus, detailed and accurate soil quality status maps in digital and analog format are required to identify soils susceptible to degradation. This entry explains the scientific basis of soil quality interpolation and mapping and techniques of assessment. Indeed, the accurate determination of soil quality is constrained with problems that range from data availability to the requirement of complex models to capture the spatial heterogeneity of soil properties. Thus, the integration of different approaches and tools to handle data scarcity and minimize errors to generate simplified soil quality maps is discussed.

Copyright © 2016 by Taylor & Francis. All rights reserved.
accomplishing or measuring selected tasks. The simpler or less detailed the indicator, the easier to identify, model, understand, and address the specific concerns that may include evaluating management efficacy or policy intervention. Geographical information system (GIS) generates maps from overlays of different data sets that can be used for prioritizing during decision-making and for evaluating alternative strategies. However, GIS is no panacea and has its own challenges. For example, the data used in overlays are derived from multiple sources, in different formats, time spans, and resolutions. Therefore, it is challenging to merge or standardize these data sets. Furthermore, mapping accuracy varies with the level of detail and generalization. Thus, it should not be construed that measurements made on the paper or digital map products reflect the exact location of all these features on the ground.

Thus, science-based techniques are required for establishing the least amount of data from an entire data set that mimic changes in soil quality vis-à-vis land management or geographic location. Due to the uncertainty with regard to soil quality benchmark selection, soil quality status may preferentially be determined on a relative basis. To date, there is no universal equation or soil quality prediction model that fits all ecoregions. Furthermore, different measurement and modeling techniques have different assumptions, which may create errors, exacerbating problems of interpretation and hampering efforts toward the judicious management of agroecosystems. However, it should be envisaged that measurements integrated with theories, hypotheses, or models can lead to discovery or insights on earth processes. Therefore, this entry expounds on the conceptual framework (Fig. 1) for interpolation and mapping of soil quality, limitations, and future research prospects.

Table 1  Soil quality assessment methods.

<table>
<thead>
<tr>
<th>Qualitative (e.g., by visual observation)</th>
<th>Quantitative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil color: (i.e., darker soils perceived to be of relatively higher quality)</td>
<td>Soil physical, chemical, and biological properties (e.g., SOC, texture, $\rho_b$, water-holding capacity, pH, aggregate stability, electrical conductivity, earthworm count, and soil depth; <a href="http://soilquality.org/indicators.html">http://soilquality.org/indicators.html</a>)</td>
</tr>
<tr>
<td>Soil tilth: (good-quality soils have clods easily broken on tillage)</td>
<td>Soil erosion, pollution</td>
</tr>
<tr>
<td>Compaction (good-quality soils have no hard pan and have little resistance to plough on tillage; <a href="http://pubs.cas.psu.edu/FreePubs/pdfs/ue170.pdf">http://pubs.cas.psu.edu/FreePubs/pdfs/ue170.pdf</a>)</td>
<td>Soil tensile strength/stability for construction or civil works</td>
</tr>
<tr>
<td>Water infiltration and drainage (e.g., good-quality soils drain well, compared with poor-quality soils that exhibit slaking and surface sealing, reduced water infiltration, and increased runoff and erosion)</td>
<td>Agricultural productivity (i.e., yields)</td>
</tr>
<tr>
<td>Hybrid methods: USDA Soil Quality Test Kit and Interpretive Guide; Cornell Soil Health Testing (<a href="http://soilhealth.cals.cornell.edu/">http://soilhealth.cals.cornell.edu/</a>).</td>
<td>Mathematical and statistical models (e.g., GIS, PTFs, and other regression models)</td>
</tr>
</tbody>
</table>

Fig. 1  Compartmentalized framework of soil quality spatial interpolation and mapping.
FRAMEWORK FOR SPATIAL INTERPOLATION

Sampling

Sampling entails the selection of a subset of data or observations to estimate the characteristics of the whole data set or to predict values at unsampled locations.[9] Selecting a suitable sampling design can reduce analysis costs and time, enhance precision, and support repeatability in experiments. Examples of sampling designs include the simple random sampling (SRS), stratified random sampling, and systematic random sampling. The SRS considered as the reference method, selects calibration sites at random with all sites having the same probability of being sampled, and has no restrictions on their spatial locations. Although SRS is a simple method, some of the distribution and modalities of parameters may be missing or large gaps may occur in the sample. Alternately, stratified methods generate a set of samples that more precisely reflect the shape of a multidimensional distribution for a collection of chosen ancillary variables. To fulfill the scientific requirement of replicability in metrics, an unbiased estimate with the lowest errors is desirable.

Decision Trees

Decision trees are machine learning, data mining, and rule induction algorithms that categorize data by inferring the interconnectivity between a dependent variable and a set of predictors. Decision trees 1) handle non-parametric data where the predictors are not characterized as having a specific distribution; 2) are insensitive to missing data, to inclusion of irrelevant predictors or to the presence of outliers; 3) effectively operate using numerical, ordinal, binary, and categorical classes; and 4) identify complex hierarchical relationships between predictors and response variables.[10] Decision trees consist of nodes and leaves with each node representing an if-then statement.[10] The classification tree provides a categorical outcome, whereas the regression tree provides a continuous numerical outcome.

Random Forest (RF) is an ensemble of the Classification and Regression Tree algorithm and operates by aggregating multiple trees to enhance the model accuracy.[11] RF has the advantage of incorporating “randomness” into its predictions through iterative bootstrap sampling and is less susceptible to overfitting.[9] Comparatively, “bagging” aggregates the results of many trees, and boosting considers errors from previous classifier steps when sampling data for the next iteration.

Geostatistical Analyses

Geostatistical methods are based on the Tobler’s law and predict unsampled values at point locations from observations at neighboring positions, under the assumption that the values at different locations are spatially autocorrelated.[8] Tobler’s law states that observations or measurements close to each other are more likely to be similar than those farther apart.[12] Examples of geostatistical methods include local spatial averaging, inverse distance weighting, and kriging. The local spatial average computes the value of unsampled locations from the mean of neighboring values; the problem being to define this local neighborhood. Comparatively, the inverse distance weighting computes the values for unsampled locations as the weighted mean of neighboring values, with the weights decreasing linearly from the prediction location, with the problem here being how to deal with the distances close to zero? In kriging, the linear model is fitted by ordinary least squares, and then, a variogram is estimated for the residuals. Cokriging is the multivariate modification of kriging that combines a sparsely measured primary variable (or target variable) with a denser set of ancillary data considered as secondary variable (e.g., remote sensing data) to enhance accuracy.[13] The limitation of geostatistics is the requirement of dense point data sets.

Remote Sensing

Remote sensing is the science of gathering information by recording sensed or emitted energy from the earth’s surface without physical contact. Remote sensing provides spatially continuous spectral digital data useful for modeling earth system processes and downstream science applications. However, remotely sensed data can be expensive to collect, analyze, archive, and maintain. Furthermore, remote sensors measure only surrogate variables such as reflectance, brightness temperature, and backscatter, from which earth information can be gleaned.

The diffuse reflectance spectroscopy (DRS) visible (Vis) (400–780 nm) and near-infrared reflectance (NIR) (780–2500 nm) are non-destructive, fast, precise, inexpensive, proximal remote sensing tools useful for mapping soil properties.[14] Practically, models are generated that relate soil spectra, from the Vis and NIR wavelengths with laboratory-based measurements of soil properties. The DRS has been utilized to estimate cation exchange capacity, base saturation, pH, exchangeable bases, and extractable phosphorus, clay content, extractable iron (Fe), total elements such as calcium, magnesium, Fe, manganese, potassium, and copper.[15] The challenge remains to spatially interpolate spectra of these soil properties across a spatial domain, a feat that may be attainable by classifying soil spectra.

Soil classification for mapping can be done digitally by 1) grouping data/observations that present homogeneous attributes without prior knowledge (unsupervised) or 2) training a model to provide maps based on known soil type observations (i.e., supervised). In the same vein, feature selection and separability algorithms can
be used to separate relevant features from those that are irrelevant. Spectral mixture analyses decompose spectra within pixels based on proportional cover of each pure class, or end-member, enhancing clarity of map products. Unfortunately, the full range of spectral, spatial, and temporal properties for detailed soil classification is difficult to ascertain.

Time series analyses or temporal mapping with satellite imagery can be used for monitoring. Inherent monitoring challenges, such as data gaps from atmospheric scattering, cloud-covered, or shadowed pixels, can be eliminated by normalization. Normalization algorithms identify and merge the pseudo-invariant (i.e., temporally unchanged) features on both the ground and imagery. Mapping soil characteristics remotely may require sensors that generate signals that penetrate obstacles (e.g., vegetative cover, roads, and soil depths) or algorithms that consider these variables.

ASSESSING INTERPOLATION EFFICACY

A key challenge when interpreting maps is quantifying the accuracy and explanatory power of the information. Soil mapping errors can be attributed to the spatial heterogeneity in soil properties and the equipment errors. Accuracy can be assessed by comparing predicted with observed values. For example, a sample data set can be randomly split into 70% for model calibration and development and 30% for validation. Metrics such as the coefficient of determination ($R^2$), mean error (ME), and root mean square error (RMSE) can be used, with a high $R^2$ or smallest RMSE or ME indicating higher accuracy. For classified remotely sensed products, the error matrix or contingency table is computed, with the overall accuracy being the ratio of the correctly classified pixels (the sum of diagonal number of pixels in the matrix) to the total number of classified pixels, whereas the kappa index being the probability of a pixel classified by chance.

CONCLUSION

This entry provides a synopsis of issues to consider during soil quality interpolation especially under diverse land management scenarios. For comparability and monitoring purposes, fusing data from diverse sources to produce a synthesized indicator of soil quality would suffice. However, because one field can have several soil types, a limitation is the inherent scalar uncertainty when continuously mapping or predicting soil quality. Because interpolation approaches are data driven, interpretations should be done cautiously with local expert knowledge to avoid making false assumptions, but the major challenge remains how to define the spatial dimensions (e.g., depth) for detailed analyses and interpretation of soil quality.

REFERENCES

CMS ID: **120053503**

Entry Title: **Soil Quality Assessment by Interpolation Techniques**

<table>
<thead>
<tr>
<th>Query No.</th>
<th>Page No.</th>
<th>Column No.</th>
<th>Query Detail</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Query</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Keywords**: Accuracy, digital mapping, soil quality, spatial interpolation.