# 19th World Congress of Soil Science

# Congress Symposium 1
# The GlobalSoilmap.net project

**Soil Solutions for a Changing World,**

**Brisbane, Australia**

**1 – 6 August 2010**

# Table of Contents

# Aligning New Zealand digital soil mapping with the global soil mapping project

Allan Hewitt[A], James Barringer[A], Guy Forrester[A], Ian Lynn[A], Thomas Mayr[B], Stephen McNeil[A] and Trevor Webb[A]

[A]Landcare Research, Lincoln, New Zealand, Email hewitta@landcareresearch.co.nz
[B]National Soil Resources Institute, Cranfield University, UK, Email t.mayr@cranfield.ac.uk

**Abstract**
New Zealand is undertaking a new phase of soil mapping (S-map) with the goal of national soil map coverage at 1:50 000 scale. One S-map objective is to develop a national capability in digital soil mapping (DSM). Concurrent development of the global soil map (GSM) project provides an opportunity for the S-map project to align its work with the GSM initiative both to achieve an enhanced national soil map and contribute New Zealand data to the GSM project. Consequently we have applied GSM method guidelines nationally. We matched the generic GSM methods to 193 New Zealand soilscapes. Soilscapes with similar levels of legacy soil information were grouped into eight soilscape groups, and DSM methods applied to them. Four soilscape groups were in low relief land where intensification of land use generates high demand for new soil information. The other four method classes were in rolling, hilly and mountainous land. The most extensive method in low relief land requires establishment of better environmental covariate layers followed by soil map reference area extrapolation, and the most extensive method in high relief land involves sampling programmes and scorpan analysis. The proposed inference engines are Bayesian belief networks, SoLIM, and Random Forests.

**Key Words**
Digital soil mapping, soilscapes, soil mapping strategy, global soil map.

## Introduction

New Zealand (NZ) is undertaking a new phase of soil mapping (S-map project, Lilburne *et al.* 2004) to meet demands for soil information and especially provision of more accurate input data for a new generation of models for application in environmental policy development and land management. The S-map goal is for national soil map coverage at 1:50 000 scale. One of the S-map objectives is to develop a national capability in digital soil mapping (DSM). Concurrent development of the global soil map project provides an opportunity for the S-map project to align its work with the global soil map initiative to achieve an enhanced national soil map, to contribute NZ data to the global project, and to develop national capability in DSM. Consequently we have applied global soil map DSM method guidelines nationally. The purpose of this paper is to outline an adaption of the Minasny and McBratney (in press) global soil map method guidelines in developing a national DSM strategy.  The global goal is to generate soil information as 90-m-resolution rasters. The NZ goal is to provide finer, 25-m-resolution rasters. The intention is to provide national DSM coverage using techniques and formats that will be compatible with global DSM coverage and from which global-resolution data for NZ may be easily extracted.

## Methods

*Soilscape definition and mapping*

The definition of NZ national soilscapes was explored by Hewitt *et al.* (in press) and completed in the work described in this paper (and available through www.landcareresearch.co.nz). We followed the soilscape definition of Lagacherie *et al.* (2001) where a soilscape is “a landscape unit including a limited number of soil classes that are geographically distributed according to an identifiable pattern”. The quality of our national soilscape coverage will improve incrementally as DSM proceeds. First-approximation soilscapes are being used to plan DSM operations as in this paper. Subsequently it will be modified and improved using DSM results to derive a second-approximation coverage suitable for use as a more generalised representation of national soils. Hewitt *et al.* (in press) explored digital methods for generating soilscapes. They found that where they are available, legacy data and expert knowledge provided an efficient basis for generating a first approximation. Accordingly soilscapes for the South Island were based on the earlier map of “soil sets” for the South Island similar in concept to “land systems” (Soil Survey Staff 1968). Soilscapes for the North Island were based on a map of “erosion terrains” derived from the NZ Land Resource Inventory (NWASCO 1979). Although originally intended for erosion and sediment yield studies, the erosion terrains efficiently



1 – 6 August 2010, Brisbane, Australia.  Published on DVD.

stratified soil patterns and rock types relevant to soilscapes mapping. For both islands, the soilscapes were arranged in a hierarchy of six levels: level 1, land province – major climate, geologic terrains and landscape units; level 2, land region – major physiographic units; level 3, lithology – major rock and cover material types; level 4, climate; level 5, altitude; and level 6, slope and landforms. For the national-scale DSM strategy we used soilscapes at level 5 for the South Island and soilscapes at level 3 for the North Island. Climate (level 4) and altitude (level 5) were not used for the North Island because these factors were less variable and were of less significance than in the South Island. Level 6 slope and landform attributes were not used because they stratified finer scale variations considered more relevant for local rather than national planning. This provided 193 soilscapes for analysis nationally (52 for the North Island and 141 for the South Island).

*Data available for DSM in NZ*
Point data are mainly limited to analysed pedons of the National Soils Database (NSD). Data quality is high but the number of sampled sites is less than 3000. These tend to be clustered in former study areas. There are few areas where soil survey auger observations are available in digital form. Although point data are sparse, there is good remaining pedologist expertise. DSM methods that are able to incorporate expert knowledge are therefore favoured.

Environmental covariates of national extent include climate layers (Leathwick *et al.* 2002), a national digital elevation model at 25-m and 15-m resolution based on national-extent 1:50 000 scale topographic elevation maps (Barringer *et al.* 2008), and nearly completed 1:250 000 scale of geological coverage update (Nathan 1993). National-extent remote sensing imagery has been collated by the EcoStat project (Dymond and Shepherd 2004) and projects in support of Kyoto Protocol compliance.

*Allocation of DSM methods to soilscapes*
A major distinction is made between "lowlands", comprising plains or basins with flat to easy rolling slopes, and "uplands", comprising rolling, hilly, plateau and steep land. Uplands have sufficient relief to make effective use of the national 25-m-resolution DEM for soil–landscape modelling. Lowlands relief is insufficient for effective use of the DEM. There is also a lack in the lowlands of good environmental covariates of sufficient resolution to support DSM techniques.

Based on Minasny and McBratney (in press) we assigned the following land and soil information attributes to all soilscapes: (1) lowland or upland, (2) area of legacy soil maps, (3) soil survey quality classification assignment for all legacy soil maps in five classes based on age and map scale, (4) number of national soil database analysed profile sites, and (5) available expert knowledge either in the form of reports, soil–landscape models, or living people. We grouped the soilscapes with similar attributes into eight soilscape groups, and for each of these we assigned generic DSM methods.

*Port Hills trials for operational DSM methods*
A study area in the Port Hills adjacent to Christchurch City was chosen to test three possible inference engines; a Bayesian belief network, SoLIM (Zhu *et al.* 2001), and a classification using the Random Forests method (Breiman 2001). Belief networks enable excellent depiction and exploration of soil–landscape models and are therefore good for capturing expert knowledge. It is time-consuming to set up the networks, and they require categorical training data and can only predict categorical data. SoLIM is also well able to capture expert data. It can accept both categorical and point data but is limited to categorical data outputs. Random Forests is very fast and able to predict both categorical (i.e. classification) and numerical data (i.e. regression). It is more able than the other methods to handle large learning and processing databases, often with little or no built-in knowledge of the data under study or the relationship between data elements. Its main disadvantage, common with all general machine-learning methods (Hastie *et al.* 2001), is that it is a black box with no direct opportunity for expert knowledge, and it is often difficult to extract the reasoning for the derived relationships.

The legacy 1:25 000 scale Port Hills soil map and report is a high quality soil survey of hilly and steep land. The area is complex and presents a challenging test area for DSM and has been used as a test-bed for developing concepts of soil–landscape models (Webb 1994). Environmental covariates were derived from geology and climate DEM layers. The three inference engines were evaluated by comparing output soil classes to the original soil survey using confusion matrices. Judgement focused on the mapping of soil classes that were least likely to be wrongly mapped in the legacy soil survey, for example, the contrast between shallow soils derived from volcanic rock on shoulder landforms verses deep colluvium soils on steep talus cone aprons.

## Results

*Soilscape groups*

Table 1 summarises the soilscape groups, their definitions, and derived DSM methods modified from Minasny and McBratney (in press).

**Table 1. Soilscape groups and proposed DSM methods based on criteria modified from Minasny and McBratney (in press). Low = lowland, Up = upland, H = high, L = low, qual = quality, pt = point, NI = North Island, SI = South Island, ECovar = environmental covariates, Extrap = extrapolation.**

| Soilscape group | Soilscape group definitions | DSM Methods | Area ($km^2$) NI | Area ($km^2$) SI |
|---|---|---|---|---|
| Group 1 | Low, H qual full cover map, good pt data | ECovar + Spatial disaggregation | - | 14788 |
| Group 2 | Low, H qual ref. area maps, sparse pt data | ECovar + Map Extrap | 24 559 | - |
| Group 3 | Low, L qual, ref. area maps, sparse pt data | ECovar + Training map + Extrap | 7 249 | 11647 |
| Group 4 | Low, L qual surveys, no pt data, | ECovar + Training map + Extrap | 57 | 1106 |
| Group 5 | Up, H qual ref. area maps, sparse pt data, | Map Extrap | 5 755 | 4383 |
| Group 6 | Up, L qual ref. area maps, sparse pt data, | Sampling + Scorpan | 23 673 | 46435 |
| Group 7 | Up, L qual ref. area maps, sparse pt data, | Sampling + Scorpan | 11 912 | |
| Group 8 | Up, no useful surveys no pt data | Sampling + Scorpan | 40 677 | 3264 |

Group 1 includes the alluvial outwash and loess-blanketed lowlands of the Southland and Canterbury plains. Soil map polygons incorporate associations and complexes of well-defined soil classes. Development of good quality environmental covariates is needed to spatially disaggregate these polygons. The NSD point data sites are not likely to be sufficient for application of scorpan kriging (McBratney *et al.* 2003) but they do provide good datasets for development of pedotransfer functions. Group 2 is the dominant lowland area that includes a wide range of alluvium, ash, loess, and sand dune soilscapes. High quality soil surveys have patchy distribution but provide good reference areas for spatial extrapolation. As for group 1, the development of spatial covariates is necessary to enable map extrapolation and spatial disaggregation of polygons. Groups 3 and 4 include land similar to group 2 but reference soil maps are of low quality. The groups have either sparse or no point data. A practical approach for both groups would be to choose and map reference training areas and then extrapolate to the full soilscape areas by the methods of group 2. If the primary application of the NSD point data is for the development of pedotransfer functions then the distinction between point data coverage between groups is not important because all point data would be pooled across all groups to develop functions. The areas of validation for these functions would most likely be independent of the soilscape groups.

In upland areas existing environmental covariates are generally suitable for DSM. Development of 5-m-resolution DEMs by ALOS Prism and radar imagery is proceeding to provide better land element discrimination in lower relief rolling and hill land. DEM derivatives are powerful soil predictors of soils in NZ because the NZ landscape and soil cover is predominantly of late-Pleistocene or Holocene in age and soils are closely related to landform position. Group 5 includes soilscapes with high coverage of quality reference area soil surveys that will provide a good basis for map extrapolation. Groups 6, 7 and 8 have either poor or unsuitable legacy soil surveys and sparse or no data. Much of the land is of low priority and a significant proportion has poor accessibility due to rugged terrain. Consequently mapping must be based on sampling programmes and scorpan modelling. Sampling strategies must take into account access.

Most of the soilscape classes are mapped as several delineations. Because of this a legacy soil map reference area may need to be extrapolated into non-contiguous unmapped areas. Feature space analysis may assist in confirming the integrity of soilscape delineations. Extrapolations will need to be validated.

*Port Hills trial*

Map outputs from the three inference engines tested on the Port Hills data were compared with the legacy soil map. The three methods had comparable performance in recognising major soil contrasts. However, Random Forests had superior performance overall. A likely explanation for this is that the Random Forests method involved data mining of soil map units. The other two methods, however, were more influenced by expert knowledge of the relationships of soil taxonomic units to what were considered key environmental covariates. Use of the legacy soil map as the standard may not then be a fair basis for comparison and suggests the need for an independent field sample based approach for a more accurate assessment.

**Conclusions**

- The global soil map has potential to provide opportunities for linkage with the international DSM community that that will help national as well as international goals.
- Development of better environmental covariates, particularly in the lowlands, is of high priority.
- As development of DSM capability is important there is need for the mapping team to gain experience by testing methods with each of the soilscape group areas.
- Because DSM techniques continue to advance it is likely that current techniques chosen will be superseded. Our choice therefore has to be provisional.
- The results of this analysis provide a basis for costing the DSM effort required in NZ.

**References**

Barringer JRF, Hewitt AE, Lynn IH, Schmidt J (2008) National mapping of landform elements in support of S-Map, a New Zealand soils database. In 'Advances in Digital Terrain Analysis'. (Eds Q Zhou, B Lees, G Tang) pp. 443–458. (Springer).

Breiman L (2001) Random Forests. *Machine Learning* **45**, 5–32.

Dymond JR, Shepherd, JD (2004) The spatial distribution of indigenous forest and it composition in the Wellington Region, New Zealand, from ETM+ satellite imagery. *Remote Sensing of Environment* **90**, 116–125.

Hastie T, Tibshirani R, Friendman J (2001) 'The Elements of Statistical Learning.' (Springer: New York).

Hewitt AE, Barringer JRF, Forrester GJ, McNeill SJ (in press) Soilscapes basis for digital soil mapping in New Zealand. In 'Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Third Global Workshop on Digital Soil Mapping in Logan, Utah, USA (DSM 2008)'. (Eds AE Hartemink, A McBratney, J Botenger) (Springer).

Lagacherie P, Robbez-Masson JM, Nguyen-The N, Barthès JP (2001) Mapping of reference area representativity using a mathematical soilscape distance. *Geoderma* **101**, 105–118.

Leathwick J, Morgan F, Wilson G, Rutledge D, McLeod M, Johnstone K (2002) 'Land Environments of New Zealand: A Technical Guide.' (Ministry for the Environment: Wellington).

Lilburne L, Hewitt AE, Webb TH, Carrick S (2004) S-map: a new soil database for New Zealand. In 'SuperSoil 2004 : 3rd Australian New Zealand Soils Conference" University of Sydney, Australia Sydney. P [unpaged]. Available: www.regional.org.au/au/asssi

McBratney AB, Minasny B, Mendonca Santos ML (2003) On digital soil mapping. *Geoderma* **117**, 3–52.

Minasny B, McBratney AB. (in press) Methodologies for global soil mapping. In: 'Digital Soil Mapping: Bridging Research, Environmental Application, and Operation. Third Global Workshop on Digital Soil Mapping in Logan, Utah, USA (DSM 2008)' (Eds AE Hartemink, A McBratney, J Botenger) (Springer).

Nathan S (1993) Revising the 1:250 000 Geological Map of New Zealand: a discussion paper. *Institute of Geological & Nuclear Sciences Scientific Report 93/*26.

NWASCO (1979) 'Our Land Resources.' Bulletin accompanying NZLRI worksheets. (National Water and Soil Conservation Organisation: Wellington).

Soil Survey Staff (1968) General survey of the soils of the South Island, New Zealand. *Soil Bureau Bulletin* **27**. (DSIR: Wellington). 404 p., including maps.

Webb TH (Ed.) (1994) Soil-landscape modelling in New Zealand : proceedings of a workshop held at Aokautere, New Zealand, 8–9 February 1993. *Landcare Research Science Series* **5**. (Manaaki Whenua Press: Lincoln).

Zhu AX, Hudson B, Burt J, Lubich K, Simonson D (2001) Soil mapping using GIS, expert knowledge, and fuzzy logic. *Soil Science Society of America Journal* **65**, 1463–1472.

# Digital soil mapping using legacy soil data in Korea

S. Young Hong[A], Budiman Minasny[B], Yong Seon Zhang[A], Yi Hyun Kim[A] and Kang Ho Jung[C]

[A]Department of Agricultural Environment, National Academy of Agricultural Science(NAAS), Rural Development Administration(RDA), 150 Suin-ro, Suwon 441-707, Korea, Email syhong67@korea.kr; zhang@korea.kr, yhkim75@korea.kr
[B]Faculty of Agriculture, Food & Natural Resources, The University of Sydney, NSW 2006, Australia, Email b.minasny@usyd.edu.au
[C]Faculty of Agriculture, Life & Environmental Sciences, University of Alberta, Edmonton, Alberta T6G 2E3, Canada, Email stealea@hanmail.net

**Abstract**
Soil carbon storage and available water capacity are important properties for land management, plant production and environment and ecosystem management. This paper will apply the digital soil mapping concept for mapping these two properties in South Korea. A Korean soil database was compiled, which includes chemical and physical properties such as particle size, moisture retention, organic matter, cation exchange capacity, and a limited number of bulk density data based on 380 soil series. The first step is to estimate bulk density for estimation of both C storage and available water capacity. Bulk density at different depths of soils was predicted by deriving a pedotransfer function model with sand, depth, and organic matter, based on Adams' model (1973). Organic C distribution with depth was first derived by converting from mass basis C (kg/kg) to volume basis C (kg/m$^3$). C storage (kg/m$^2$) was first calculated by multiplying C on the volume basis to the thickness of each soil layer (m), and finally integrated to a depth of 1 m for each soil series. Mapping available water capacity was more challenging as only half of the database contains measurement of water retention at -33 and -1500 kPa. Field capacity was calculated from clay content and predicted bulk density and adjusted by taking into account porosity. Wilting point was calculated from clay content and adjusted for any discrepancy with predicted field capacity and porosity. Available water capacity (mm) to a depth of 1 m was estimated by multiplying the amount of water stored between field capacity and wilting point and the thickness of the layer. The carbon storage and available water capacity from surface to a depth of 1 m for the south part of whole Korean peninsula were mapped using the estimated parameters in a soil series map unit (1:25,000). Mean value of carbon density of Korea is approximately 5 kg/m$^2$ and available water capacity is approximately 154 mm. Total soil carbon storage of agricultural land in Korea is approximately 174 Gg.

**Key Words**
Soil information, carbon storage, available water capacity, Korea, digital soil mapping.

**Introduction**
The need for accurate, up-to-date, and spatially referenced soil information, which is important for land management, food production, and ecosystem management, has been identified by policy and decision makers, land users, farmers, and researchers. "This need coincides with an enormous leap in technologies that allow accurate collecting and predicting soil properties. Accordingly, there is a need for making a new digital soil map of the world using state-of-the-art and emerging technologies for soil mapping and predicting soil properties at fine resolution" (www.globalsoilmap.net). Minasny *et al.* (2006) showed the application of digital soil mapping for mapping the depth functions of soil carbon in Australia. Hong *et al.* (2009a, 2009b) introduced Korean soils and information systems and also mapped soil carbon storage and water capacity using soil profile and soil series information.

An Asian soil information working group was also organized during the workshop on "A new approach to soil information systems for natural resources management in Asian countries" which was held in Japan in 2008 initiated by Food & Fertilizer Technology Center (FFTC). The objectives of the workshop were to review the current status of SIS (or available soil data such as soil maps) in participating Asian countries, to exchange relevant information on appropriate SIS for the participating countries, to share technological know-how relevant for establishing a SIS for sustainable crop production among the participant countries, and to discuss the possibility of establishment of an appropriate regional SIS for sustainable crop production in the Asian and Pacific Council (ASPAC).

The objectives of this study were to estimate and map soil carbon storage and available water capacity of Korea using the digital soil mapping approach which has been defined as, "the creation and population of spatial soil information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables" (Lagacherie and McBratney 2007).

## Methods

*Soil database*

A Korean database used in this study was compiled based on the "Taxonomical Classification of Korean Soils" (NIAST, RDA, 2000), which was mostly collected in the 1970s for soil profile description. It includes soil chemical and physical properties of each horizon (n=1,559) such as particle size, moisture retention, organic matter, cation exchange capacity, and a limited number of bulk density data (n=108) based on 380 soil series. When described using the Soil Taxonomy of the USDA, soils in Korea are classified into seven Soil Orders which are then further divided into 14 Sub-Orders according to moisture regimes. Among those seven Soil Orders, the younger soils, Entisols and Inceptisols, are dominant. Entisols are the youngest soils, followed by Inceptisols. Alfisols and Ultisols . The working unit of soil classification is the Soil Series. So far 390 Soil Series have been identified in the country. Table 1 is a summary of the areal extent of the different Soil Orders and the number of Soil Series within them. Table 1 clearly shows that abundance of younger soils (Entisols and Inceptisols). This is a result of the influences of both Korea's unique climate, with concentrated rainfall in summer, and rugged topography as characterized by the wide occurrence of highly-sloped mountains. This strongly suggests that, if the soil resources are to be adequately conserved, serious attention must be paid to development of measures to minimize the soil erosion in hilly lands.

**Table 1. Major soil orders/sub-orders, number of soil series occurring within them and the areal extent of soil orders in Korea.**

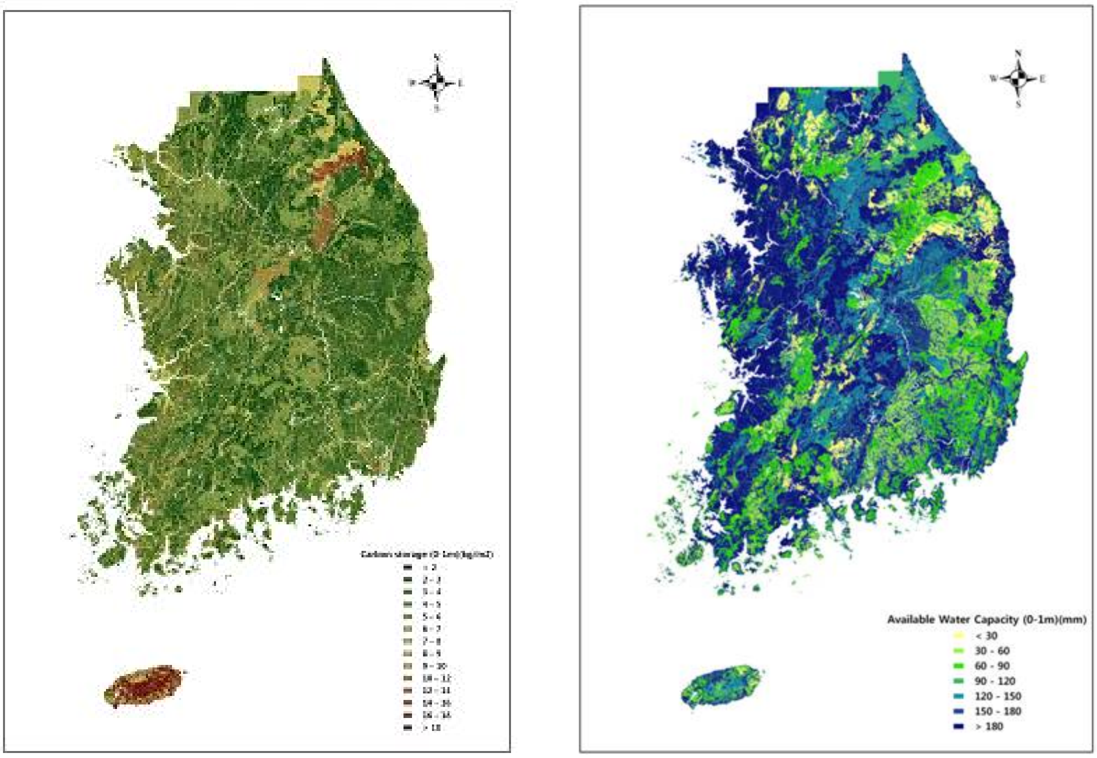
| Soil Orders | Sub orders | No. of Soil Series | Area-$10^3$ ha (%) |
|---|---|---|---|
| Inceptisols | Aquepts | 77 | 6,668 (69.2) |
| | Udepts | 133 | |
| Entisols | Aquents | 14 | 1,315 (13.7) |
| | Fluvents | 13 | |
| | Orthents | 17 | |
| | Psamments | 20 | |
| Ultisols | Udults | 28 | 398 (4.2) |
| Alfisols | Aqualfs | 7 | 276 (2.9) |
| | Udalfs | 37 | |
| Andisols | Udands | 39 | 129 (1.3) |
| | Vitrands | 1 | |
| Mollisols | Udolls | 2 | 5 (0.1) |
| Histosols | Saprists | 1 | 0.4 (0) |
| | Hemists | 1 | |
| | | 390 | |

*Digital mapping of soil C storage and available water capacity*

The first step was to estimate bulk density for estimation of both C storage and available water capacity. Bulk density at different depths of soils was predicted by deriving a pedotransfer function model with sand and depth. Adjustment for organic matter content was based on Adams' model (1973). Organic C distribution with depth was first derived by converting from mass basis C (kg/kg) to volume basis C (kg/m$^3$). C storage (kg/m$^2$) was first calculated by multiplying C on the volume basis to the thickness of each soil layer (m), and finally integrated to a depth of 1 m for each soil series. Mapping available water capacity was more challenging as only half of the database contains measurement of water retention at -33 and -1500 kPa. Furthermore measurement of water retention is in mass basis and based on disturbed soil samples. Pedotransfer functions were derived for volumetric water content at field capacity (-33 kPa) and wilting point (-1500 kPa). Further adjustments based on total soil porosity are required as the field capacity values were derived from disturbed soil samples. Field capacity was calculated from clay content and predicted bulk density and adjusted by taking into account porosity. Wilting point was calculated from clay content and adjusted for any discrepancy with predicted field capacity and porosity. Available water capacity (mm) to a depth of 1 m was estimated by multiplying the amount of water stored between field capacity and wilting point and the thickness of the layer. The carbon storage and available water capacity from surface to a depth of 1 m for the south part of whole Korean peninsula were mapped using the estimated parameters in a soil series map unit (1:25,000). Total carbon storage and available water capacity were summarized by land use type using land cover map provided by Ministry of Environment.

## Results

Figure 1 shows the distribution of soil carbon storage (kg/m$^2$) and available water capacity (mm) of Korea. The distribution map of the soil properties was made by calculating mean soil organic carbon values and available water capacity to 1 meter depth for each soil series from Korean database and allotted the mean value calculated of soil properties to each of the detailed Korean soil map (1:25,000). The mean value (to 1 m depth) of carbon density of Korea is approximately 5 kg/m$^2$ and available water capacity is approximately 154 mm. Soil C density in grass and agricultural land were higher as 8.82 and 6.77 kg/m$^2$, respectively, than other land use types and available water capacity of soil was the highest as 203 mm in agricultural land as shown in Table 2. Total carbon storage and available water capacity were summarized by land use type using the land cover map provided by Ministry of Environment. Total soil carbon storage of agricultural land in Korea is approximately 174 Gg.




**a) Soil carbon storage** **b) Available water capacity**

**Figure 1. Soil carbon storage and available water capacity map of Korea.**

**Table 2. Amount of soil carbon storage and available water capacity by land use type in Korea.**

| C storage | Forest | Agr. Field | Grass | Wetland | Barren |
|---|---|---|---|---|---|
| Land use (km$^2$) | 61,394 | 25,648 | 1,858 | 1,780 | 1,439 |
| Available Water Capacity (AWC, mm) | 123.6 | 203.1 | 142.9 | 47.8 | 137.7 |
| C density (kg/m$^2$) | 4.05 | 6.77 | 8.82 | 1.10 | 4.24 |
| Total C storage (Gg) | 249 | 174 | 16 | 2 | 6 |

## Conclusion

Soil carbon storage and available water capacity in Korea, which are important for land management, food production and environment and ecosystem management, were predicted and mapped based on a Korean database for soil profile description that mostly collected in the 1970s. Mean value (1 m depth) of carbon density of Korea is approximately 5 kg/m$^2$ and available water capacity is approximately 154 mm. Total carbon storage and available water capacity were summarized by land use type using land cover map provided by Ministry of Environment. Total soil carbon storage of agricultural land in Korea is approximately 174 Gg. Further work is required to verify this amount with recent soil data.

## References

Adams WA (1973) The effect of organic matter on the bulk and true densities of some uncultivated podzolic soils. *Journal of Soil Science* **24,** 10-17.

GlobalSoilMap.net Consortium (2010) www.globalsoilmap.net

Hong SY, Minasny B, Zhang YS, Kim YH, Song KC (2009a) Digital soil mapping of soil carbon storage and available water capacity in Korea. In 'Proceedings of the 9th International Conference of The East and Southeast Asia Federation of Soil Sciences Societies, Seoul'. pp. 406-408. (Korean Society of Soil Science and Fertilizer: Suwon).

Hong SY, Zhang YS, Kim YH, Sonn YK, Hyun BK, Park CW, Song KC, Jang BC, Choe EY, Lee YJ, Ha SK, Kim MS, Lee JS, Jung GB, Ko BG, Kim GY, Jung KY (2009b) Korean soils and information system. In 'Proceedings of the 9th International Conference of The East and Southeast Asia Federation of Soil Sciences Societies, Seoul'. pp. 153-158. (Korean Society of Soil Science and Fertilizer: Suwon).

Jung YT, Park ME, Jo IS, Jo YK, Park CS, Noh DC, Lee HS, Hyeon GS, Jung SJ, Bark BS, Song KC, Song YS, Jang YS, Han SS, Cho GH, Choi JW, Park KB, Son IS, Yun ES (2000) 'Taxonomical classification of Korean soils'. (National Institute of Agricultural Science and Technology: Rural Development Administration).

Lagacherie P, McBratney, AB, Voltz M (2007) 'Digital Soil Mapping: An Introductory Perspective. Developments in Soil Science 31'. (Elsevier: Amsterdam).

Minasny B, McBratney AB, Mendoca-Santos ML, Odeh IOA, Guyon B (2006) Prediction and digital mapping of soil carbon storage in the lower Namoi valley. *Australian Journal of Soil Research* **44,** 233-244.

# *GlobalSoilMap.net* - from planning, development and proof of concept to full-scale production mapping

R.A. MacMillan[A], A. E. Hartemink[A] and A. B. McBratney[B]

[A]ISRIC - World Soil Information, PO Box 59, 6700 WB Wageningen, The Netherlands, Email Bob.MacMillan@wur.nl, alfred.hartemink@wur.nl
[B]Faculty of Agriculture, Food and Natural Resources, The University of Sydney, Sydney, Australia, Email alex.mcbratney@sydney.edu.au

**Abstract**
The *GlobalSoilMap.net* project aims to produce predictions of nine key soil properties at continuous depth intervals at a spatial resolution of 90 m for the entire world. These maps of soil properties will be produced by a participants working under the coordination of regional node leaders with responsibility for organizing and delivering results for eight defined geographic regions of the world. This paper identifies and discusses the technical impediments to moving towards commencement of operational production mapping. These are: i) agreement on specifications for all products, ii) location, digital capture and harmonization of legacy soil data, iii) assembly of covariate databases, iv) documentation of prediction methods, v) specification of data model(s) to use to capture, store and disseminate maps and data, vi) selection of cyber-infrastructure to support map production and dissemination vii) end user surveys assessment and verification, and vii) identification of methods for assessing the uncertainty and accuracy of predictions. Actions undertaken to date to address these challenges are presented and progress is evaluated. There are no significant technical reasons for not moving towards planning and implementing operational production mapping.

**Introduction**
The *GlobalSoilMap.net* project has an ambitious goal of producing predictions of continuous horizontal and vertical variation for ten key soil properties at a spatial resolution of 90 m for the entire world (Sanchez *et al.* 2009). Production of these grid maps will be achieved through the combined efforts of diverse participants whose contributions are being coordinated by regional node leaders for seven continent-sized geographic regions of the world. The wide diversity of contributors, and of soils and landscapes for which predictions will be made, results in a need to identify and address the main challenges to moving forward towards operational production mapping of soil properties.

**Eight challenges to moving forward towards operational production mapping**
Eight technical challenges to moving towards rapid commencement of operational production mapping have been identified and actions to address these have been initiated. There may well be other technical or organizational issues that conspire to delay initiation of operational mapping, but it is felt that the challenges identified here represent key impediments that must be addressed in order to move forward towards operational production mapping.

*Specification of output products*
It is not possible to begin production mapping until the specifics of what is to be produced have been agreed to by all project participants. Specifications spell out explicitly which soil properties will be predicted, at what spatial resolution and depth intervals they will be predicted and in what spatial framework (coordinate system, projection, datum) they will be stored and distributed.

An initial set of project specifications was discussed and agreed upon at a *GobalSoilMap.net* consortium meeting held in Seoul, South Korea in October, 2009. These specifications call for prediction of a mean value for each grid cell of 90 x 90 m horizontal dimensions (3 arc-seconds) for ten critical soil properties (organic carbon, sand, silt, clay, coarse fragments, pH, depth to bedrock, effective soil depth, bulk density and available water holding capacity) at six specified depth increments (0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm). A spline function will be fitted to the soil property values for each depth increment to permit depiction of continuous variation in soil properties with depth (Malone *et al.* 2009). An estimate of the uncertainty associated with each prediction at each depth will accompany each property value. The predictions from diverse sources will be delivered for assembly and redistribution in geographic coordinates (lat/long) using WGS80 datum.

*Location, capture and harmonization of legacy soil profile and map data*
The *GlobalSoilMap.net* project has a vision to capitalize on the major investments that have been made in collecting soils information and producing soils maps, locally and regionally, over the last 50 to 80 years. This archive of legacy soils data has been under-used and under-appreciated, because it has not been collated and harmonized into a uniform and easy to access or interpret whole. Using pre-existing sources of soil information for predicting a specific set of individual soil properties at a consistent resolution aims at coaxing a minimum level of consistency and uniformity from this diversity of legacy data. This project will simplify the complex presentation of soils information, often associated with conventional soil maps, into a model of continuous variation in the values of single soil properties both horizontally and with depth.

Before doing so, it must first locate and obtain the majority of existing data and then find ways to harmonize or standardize data of different age, quality, information content and density so that outputs of consistent content and appearance can be generated. A first objective of the *GlobalSoilMap.net* project is therefore to identify, locate and obtain, or rescue, as wide a selection of existing information about soils as is feasible and practical. This legacy data is viewed as having two main forms, namely legacy profile descriptions and accompanying analytical data, which mostly describe point locations, and legacy map data, that describe the horizontal variation of soil classes in space.

In the context of capturing legacy point data, procedures and data entry protocols have been developed to facilitate the capture and storage of a consistent subset of soil profile attributes from a wide diversity of sources of soil information. Metadata are recorded for each soil profile to enable identification of the source of each piece of soil information and association of it with a defined analytical method or data dictionary. This capability will be used to identify whether the values for specific soil properties reported by data from different sources are comparable and equivalent or whether there are systematic differences that need to be identified and resolved.

In the context of capturing legacy map data initial efforts are focussing on simply identifying, obtaining and scanning those existing legacy maps that have not yet been captured, even as digital images. In the longer term, the project will investigate how, or if, these maps can be used to contribute to procedures for predicting the spatial variation in the ten selected soil properties. Ultimately, it will be necessary to devise procedures for standardizing and harmonizing the content of existing legacy soil maps, regardless of whether they are already topologically structured or exist only as simple scans. One option is to use existing maps to inform the creation of uniform, country-wide or continent-wide, harmonised soil-landscape maps that are then used as one of the main covariates in prediction of individual soil properties on a continuous basis.

*Assembly and pre-processing of global and regional databases of environmental covariates*
Most methods for predicting the spatial variation of soil properties analyse relationships between evidence, in the form of point soil profile data or soil maps, and explanatory variables that represent environmental conditions believed or expected to influence the spatial distribution of soils and soil properties. Environmental covariates are selected to represent the ***scorpan*** factors as outlined by McBratney *et al.* (2003). Many environmental covariates represent the influence of the land surface on variation in soil properties. These are computed as derivatives of digital elevation models (DEMs) and are the key, but not the sole, predictors used in many soil prediction models. At the time of conception of the *GlobalSoilMap.net* project, the viable digital elevation model available for most of the world was the 3 arc-second (90 m) SRTM DEM. For most portions of the world, this 90 m SRTM DEM will represent the finest resolution DEM that is consistently available across entire continents of interest. So this SRTM DEM, and derivatives computed from it, will provide a significant contribution to the covariates available to predict individual soil properties. The cost benefit of trying to obtain and use derivatives computed from finer resolution 1 arc-second (30 m) SRTM DEM data is also being evaluated.

While many key covariates will be extracted from the best available DEMs, other sources of digital data will also be necessary to capture and represent the influence of other ***scorpan*** environmental variables, such as climate, organisms (vegetation), parent material, age and spatial context. A task has been defined to identify all major digital databases of environmental covariates that are currently produced and available in digital format at global to continental extents. We wish to avoid having different partners, working in different nodes, needing to identify and obtain these same data sets separately. The job is being be done once and the resulting global to continental scale databases can be accessed and used by all project participants.

In addition to simply collecting and collating existing global scale data sets of environmental covariates, these data sets may be used to devise and apply a global stratification into ecological regions. This definition of ecological domains or pedological provinces will be coordinated with other, existing efforts to define an agreed-upon global framework of ecological strata.

*Identification, evaluation and documentation of suitable prediction methods*

One of the main questions posed to (and by) project participants has been "what prediction methods will be used to produce these maps?" This is a key issue for which answers need to be provided before operational mapping can begin. A general conceptual design has been proposed which envisages using different methods in different areas depending upon the type and amount of soil evidence data available (point and map), the kind of landscapes and the type and strength of soil-landscape relationships, and the availability of suitable environmental covariate data sets (Minasny and McBratney 2010). Four "proof of concept" areas have been identified for areas in the USA and Canada, Australia, Europe and Africa. These areas range in size from about 50 km by 100 km for the USA/Canada pilot area up to 850 km by 250 km for the entire country of Malawi. Data for these sites have been obtained and are being used to produce digital maps of soil properties using different methods. These specific examples illustrate general concepts of mapping using each main combination of available data and landscape attributes. This exercise will help users in different areas identify and select the prediction method or methods that are best suited to the conditions with which they have to deal. The resulting documentation and accompanying data sets can be used as tutorials and training manuals to help participants learn how to apply selected methods in their own areas

*Specification of conceptual and physical data model(s) for GlobalSoilMap.net data*

The *GlobalSoilMap.net* project offers a unique opportunity to redefine the structure and content of global soil databases using the most recent database concepts and modelling tools. This opportunity has been seized upon to develop a proposal to define a new global standard for the storage and exchange of soils information. This global standard will encompass the full range of types of soil information and will not just be specific to the grid maps of soil properties being prepared by the *GlobalSoilMap.net* project. It will set standards for point profile data, area soil samples, conventional polygonal soil maps and continuous raster soil property maps. A task group has been proposed to work under the auspices of the International Union of Soil Sciences (IUSS) to debate and design a new global SoilML based upon UML and XML standards.

*Selection and implementation of appropriate cyber-infrastructure support*

Storage and on-line delivery of the large global data sets of soil properties that the project will produce represents another challenge to be addressed. The project is building a cyber-infrastructure that can support collaborative efforts to acquire and process data sets of legacy soil data (points and maps) and environmental covariates of global to continental extent. This cyber-infrastructure will provide immediate support for on-line collaborative sharing of data, inter-active sharing of prediction methods and tools, and inter-active discussion forums and communication amongst project participants. In the longer term, a task group has been set up to draw on the experience of *GlobalSoilMap.net* partner agencies (e.g. USDA-NRCS, JRC, CIESIN, CSIRO, ISRIC) in storing and delivering large volumes of soils data on-line. This task group will investigate and recommend options for facilitating the storage, discovery, visualization, analysis and downloading of the data produced by the project.

*Assessment and verification of ability to meet needs of major end users*

This task seeks to validate the underlying assumption that "if we build it they will come". Essentially, the plan is to conduct a complete and systematic identification of all of the potential users of the soil property maps that will be prepared by the project. We then propose to identify the specific needs of each major user for soils data and assess the degree to which these needs can and will be met by the products that the project proposes to deliver. A task group has been formed with a mandate to systematically identify all major potential end users, to identify the specific needs of these end users for soils data and to verify the extent to which soil information produced by the project will meet those specific needs. The results of these investigations should be available for reporting by the time of the meeting at which this paper is presented.

*Identification and implementation of methods for assessing uncertainty and accuracy of predictions*

One of the exciting and important new aspects of the project is its commitment to providing an estimate of the uncertainty attached to each soil property prediction at each depth at each grid cell location. Awareness of the high level of uncertainty for particular areas may stimulate the collection of new data. A method has been proposed for estimating the uncertainty of predictions of soil properties by depth (Malone *et al.* 2009) that makes use of analysis of all geo-referenced soil property values reported for a given area. On a second front, a task group has been set up to investigate and identify viable options for collecting independent field samples to support computation of estimates of predictive accuracy for the soil property maps at different levels of aggregation.

**Planning of operational production mapping**

Production mapping will not start unless the individual nodes agree to set specific targets for both initiating and completing production mapping for specific areas of large extent. Node leaders are being encouraged to consider how they can obtain the resources (human and financial) to support full scale operational production mapping and also need to set targets for extents to be mapped within specific time frames.

**Conclusions**

The *GlobalSoilMap.net* consortium aims to move forward from initial planning and proof of concept activities towards full scale operational production mapping. Several challenges have been identified and actions are being undertaken by specific task groups to address them. Sufficient progress has been made to encourage project participants to begin thinking about how they can move forward from planning and proof of concept to operational mapping. Initial examples from some participating nodes have demonstrated that it is indeed possible to produce predictions of soil property values by depth at 90 m for entire countries or states. Node leaders are being encouraged to set targets and timelines for operational mapping of their nodes and are actively seeking the funding and resources that will be required to succeed in these challenges..

**References**


Malone BP, McBratney AB, Minasny B, Laslett GM (2009) Mapping continuous depth functions of soil carbon storage and available water capacity. *Geoderma* **154**, 138-152

McBratney AB, Mendoca Santos ML, Minasny B (2003) On digital soil mapping. *Geoderma* **117**, 3-52.

Minasny B, McBratney AB (2010) Methodologies for global soil mapping: Digital Soil Mapping: Bridging Research, Production, and Environmental Application.In 'Proceedings of the 3rd Global Workshop on Digital Soil Mapping (DSM 2008 USA)' Progress in Soil Science, Vol. 2 (Eds JL Boettinger, D Howell, AC Moore, A Hartemink, S Kienast-Brown). (Springer)

Sanchez PA, Ahamed S, Carré F, Hartemink AE, Hempel J, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, Mendonça-Santos ML, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen TG, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang GL (2009) Digital soil map of the world. *Science* **325**, 680-681.

# *GlobalSoilMap.net:* Canada-United States digital soil mapping case

**Jon Hempel**[A], Steve DeGloria[B], Walter Fraser[C], Xioyuan Geng[D], Amand Moore[E], Scott. Smith[F], James Thompson[G], Arnie Waddell[H] and Tom D'Avello[J]

[A]USDA-NRCS-National Soil Survey Center, Lincoln, NE, Email jon.hempel@lin.usda.gov
[B]Cornell University, Ithaca, NY, Email sdd4@cornell.edu
[C]Agriculture and Agri-Food Canada, Winnipeg, MB, Email fraserw@agr.gc.ca
[D]Agriculture and Agri-Food Canada, Ottawa, ON, Email gengx@agr.gc.ca
[E]USDA-NRCS-Maryland, Annapolis, MD Email amanda.moore@md.usda.gov
[F]Agriculture and Agri-Food Canada, Summerland, BC, Email smithCAS@agr.gc.ca
[G]Division of Plant and Soil Sciences West Virginia University, Morgantown, WV, Email james.thompson@mail.wvu.edu
[H]Agriculture and Agri-Food Canada, Winnipeg, MB, Email waddella@agr.gc.ca
[J]USDA-NRCS-National Soil Survey Center, Morgantown, WV, Email tom.d'Avello@wv.usda.gov

**Abstract**
A consortium of pedologists has formulated a global soil mapping initiative. The GlobalSoilMap.net project currently consists of seven continental nodes that support the goal of mapping selected soil properties over 80% of the Earth's land surface. The North American Node, lead by the Natural Resources Conservation Service in Morgantown, WV has initiated a collaborative study to test the feasibility of this effort. Agri-Food and Agriculture Canada (AAFC), the Natural Resources Conservation Service (NRCS) and several universities have begun the study with specific objectives to map selected soil properties based on *detailed* (1:12,000-1:40,000) and *generalized* (1:250,000-1:1,000,000) soil survey information using digital soil mapping methods and environmental covariate data, and to compare soil property maps developed from detailed and generalized soil mapping. We also seek to advance international exchange and quality assessment of soil data and information that will improve management of agricultural and natural resources, especially those that transcend national boundaries. The study area is located in north-central North Dakota and southwestern Manitoba where glaciated landscapes support rain-fed small grains agriculture, oilseed and forage crops, and grazing lands. Major environmental concerns are water quality and quantity, accelerated soil erosion, soil salinization, aggregate stability, and organic matter and soil productivity maintenance. To date, the team of soil scientists have assembled spatial data, developed a work plan and have implemented appropriate DSM methods and begun to evaluate mapping outcomes for addressing local and transnational resource management needs.

**Key Words**
GlobalSoilMap.net. digital soil mapping, soil survey information.

## Introduction
We have adopted a case studies approach to digital soil mapping for land areas of mutual interest to collaborating organizations. Mapping will be conducted at spatial resolutions less than 100 m cell size (<= 1 hectare) and for a set of soil properties being considered by the global soil mapping community and relevant to our intra-continental, trans-national collaborators and users. Soil properties are being estimated and mapped from existing soil geographic databases or from model predictions based on judicious use of environmental covariates and pedo-transfer functions.
The objectives for this case study are:
1. Map selected soil properties based on *detailed* soil mapping in Canada and USA using digital soil mapping methods and spatial data of high spatial resolution.
2. Map selected soil properties based on *generalized* soil mapping in Canada and USA, and compare to soil property maps developed from spatial data of high spatial resolution.
3. Map selected soil properties based on *detailed* mapping using spatial data at highest spatial resolution available (>= 30 m), and compare to soil property maps based on spatial data of lower spatial resolution (>= 90m).

## Methods
*Study area*
The spatial extent of our case study area occupies a rectangular land area of approximately 6,224 km$^2$, or 622,400 hectares (~1,538,000 acres). The extent encompasses parts of the Northern Black Glaciated Plains Major Land Resource Area (MLRA 55A) and Aspen Parkland and Southwest Manitoba Uplands Ecoregions

in the Prairies Ecozone of Canada. The area includes portions of north central North Dakota and southwestern Manitoba bisecting the Turtle Mountains to the east and major portions of the trans-national Souris River watershed to the west (Figure 1). Landscapes in the case study area are described generally as level to undulating and hummocky glacial till plains, glacio-lacustrine deposits, sandy eolian materials, kettle holes, kames, moraines, and glacial lake plains.  Soils are dominated by Mollisols with frigid soil temperature regimes and udic or aquic soil moisture regimes (Soil Classification Working Group 1998; Soil Survey Staff 2006) and by Chernozems and Gleysols (Canadian System of Soil Classification, 3rd edition)

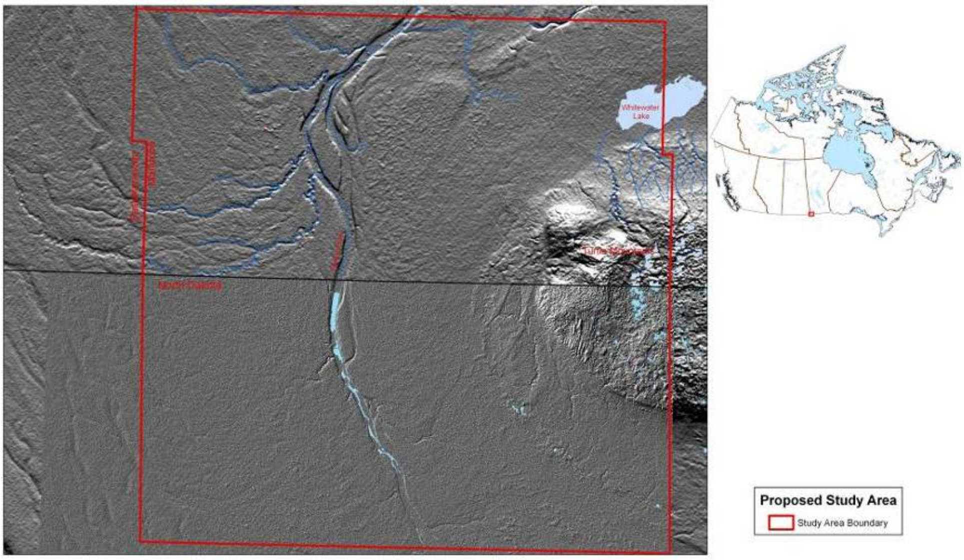



**Figure 1. Study area location in southwestern Manitoba and north-central North Dakota within the northern glaciated plains region of North America (courtesy Soil Resources Group, AAFC).**

*Digital and field data*
Data for Canada include Canadian detailed soil survey data at 1:20,000 – 1:40,000 scale in addition to generalized soil survey data at 1:1,000,000 scale (Soil Landscapes of Canada, SLC), numerous pedon observations (n ~ 500), and other geo-referenced soil inspection points are available in southern Manitoba. Data for the United States include detailed soil survey data at 1:12,000 - 1:24,000 scale (SSURGO) in addition to generalized soil survey data at 1:250,000 scale (US GSM, or STATSGO2), pedon descriptions and laboratory data, and field transects.  Numbers and locations of pedon descriptions and field transects in the study area are being compiled.  Shuttle Radar Topographic Mission (SRTM) digital elevation models (DEM) at 90 m spatial resolution are the only elevation data available for the entire proposed study area. DEM data at finer spatial resolution (<= 30 m) are available in North Dakota. In addition, SRTM data are not available for land areas more northerly than 60$^{o}$ N latitude. This excludes all of northern Canada and nearly all of the State of Alaska in the North American continent.

*Estimating and mapping soil properties*
We propose and prioritize the following minimum set of soil properties which follows the GlobalSoilMap.net consortium specifications (McMillan *et al.* 2009) for estimated properties for the case study area using digital soil mapping methods:

1. Organic Carbon (g/kg)
2. Clay (%)
3. Bulk Density (kg/m$^3$)

From these attributes, the following two properties can also be predicted using pedo-transfer functions:

4. Carbon Density (computed from Carbon % and Bulk Density; given in kg/m$^3$)
5. Available Water Capacity (given in mm/m)

The project has also identified the following “secondary” variables that are considered to be desirable and feasible to predict but which are still considered optional for delivery by nodes.

6. pH (specify method, $H_20$, $CaCl_2$, KCl)
7. CEC (Cations plus exchangeable acidity cmols/kg)
8. EC (Electrical conductivity dS/m)

For each soil property, we will determine values both horizontally across the landscape as well as vertically through the soil profile to a soil depth of at least one meter, or soil depth to restrictive layer, by taxonomic horizon. In some cases, vertical variation can be integrated with horizontal variation for some soil functions (e.g., Root Zone Available Water Holding Capacity). In other cases, modeling soil property variation at multiple depth increments, or layers, may be more appropriate (e.g., particle size distribution, pH). Digital soil mapping will be based upon detailed soil surveys in Canada and detailed soil surveys (SSURGO) in the USA for Objectives 1 and 3. Soil Landscapes of Canada (at 1:1,000,000 scale) and STATSGO2 (at 1:250,000 scale) soil geographic databases will be used for DSM applications under Objective 2.

*Inference models*

In general, the DSM methodology for the case study within the North American continental node will be related explicitly to, and incorporate to the degree possible, legacy soil survey data within each survey area including their description, location and extent, quality and relevance to soil properties being mapped. Other modeling approaches for this particular case study will follow standards as documented and recently applied by others in the digital soil mapping consortium and community (e.g., Bui *et al.* 1996). We will also build upon and complement methodologies implemented by the North American Soil Characteristics Database for Hydrological and Meteorological Modeling (NOAM-Soil) project building on the work of Miller and White (1998) and Padbury *et al.* (2002).

**Results**

The following are expected results from this project:

- Literature review of selected soil properties and appropriate DSM methods to meet land management needs in case study region.
- Development of a collaborative, intra-continental, trans-national approach to digital soil mapping over a range of spatial scales.
- An approach for predicting and mapping relevant soil properties for the region using advanced digital soil mapping (DSM) methodologies..
- Trans-national geo-spatial database development, application, and assessment for mapping selected soil properties using soil, climate, land cover, and terrain variables.
- Development of predictive soil property maps, prioritized by user application need.
- Recommendations on future trans-national collaboration for North America GSM node.

Collaborators note that both coarse (90 m) and fine (30 m) spatial resolution DEM are unlikely to adequately resolve subtle soil landscape patterns and processes in our study area (see figure 2). To meet the 90 m resolution digital soil property map standard as proposed by the GlobalSoilMap.net consortium, we plan to employ DSM methods that are less reliant on terrain model derivatives for similar landscapes throughout the North American continent. In the final analysis for this case study, however, soil property maps will be produced at 90 m resolution, which will result in very fine scale soil variation mapped at finer spatial resolutions being aggregated, or integrated, over the range of the coarser resolution data.

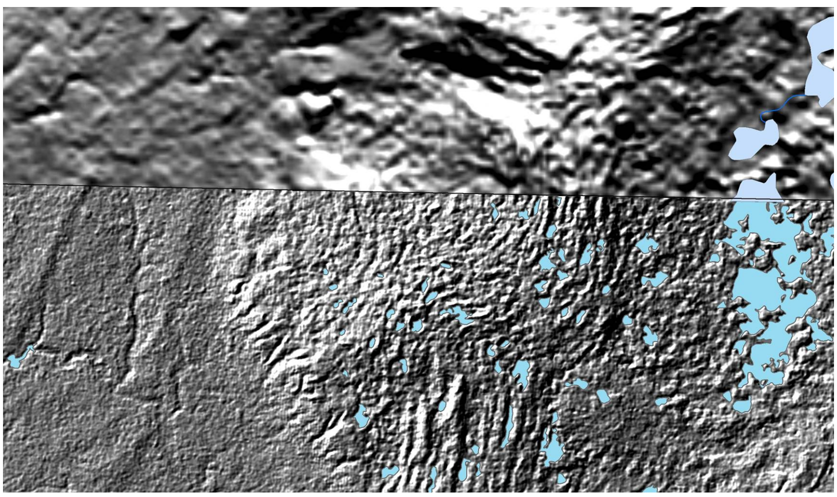


**Figure 2. Comparison of SRTM-derived hillshade maps for Manitoba at 90 m resolution (top) and for North Dakota at 30 m resolution (bottom) for a portion of the northern glaciated plains study area (courtesy Soil Resource Group, AAFC).**

## Conclusion

The outcomes anticipated from this case study will include a set of mapped soil properties of continental rather than regional importance and those that relate to the soil properties (and soil functions) defined by the global soil map consortium (i.e., carbon density, infiltration, permeability, drainage, nutrient supplying capacity, and plant available water capacity). In addition, we need to focus on land degradation and land management needs relevant to our case study area and operational work plans and policies of associated collaborating institutions

## References


Bui EN, Smettem KRJ, Moran CJ, Williams J (1996) Use of soil survey information to assess regional salinization risk using geographical information systems. *J. Env. Quality* **25,** 433-439.

McMillan RA *et al.* (2009) 'Specifications for GlobalSoilMap.net products'.

Miller DA, White RA (1998) A conterminous United States multilayer soil characteristics dataset for regional climate and hydrology modeling. *Earth Interactions* **2,** 1-26.

Padbury G, Waltman S, Caprio J, Coen G, McGinn S, Mortensen D, Nielsen G, Sinclair R (2002) Agroecosystems and land resources of the northern great plains. *Agron. J.* **94,** 251-261.

Soil Classification Working Group (1998) 'The Canadian System of Soil Classification' 3rd ed. Agriculture and Agri-Food Canada Publication 1646, (NRC Research Press: Canada).

Soil Survey Staff (2006) 'Keys to Soil Taxonomy'. 10th ed. (USDA-Natural Resources Conservation Service: Washington, DC).

# Regional approach to soil property mapping using legacy data and spatial disaggregation techniques

James A. Thompson[A], Timothy Prescott[B], Amanda C. Moore[C], James "Skip" Bell[D], Darrell Kautz[E], Jon Hempel[F], Sharon W. Waltman[G] and C. Hobie Perry[H]

[A]Division of Plant and Soil Sciences West Virginia University, Morgantown, WV, Email james.thompson@mail.wvu.edu
[B]USDA Natural Resources Conservation Service, Morgantown, WV, USA, Email timothy.prescott@wv.usda.gov
[C]USDA-Natural Resources Conservation Service, Annapolis, MD, USA, Email amanda.moore@md.usda.gov
[D]USDA-Natural Resources Conservation Service, Morgantown, WV, USA, Email skip.bell@wv.usda.gov
[E]USDA-Natural Resources Conservation Service, Morgantown, WV, Email darrell.kautz@wv.usda.gov
[F]USDA-Natural Resources Conservation Service, Lincoln, NE, Email jon.hempel@lin.usda.gov
[G]USDA-Natural Resources Conservation Service, Morgantown, WV, USA, Email sharon.waltman@wv.usda.gov
[H]USDA-Forest Service, St. Paul, MN, USA, Email charleshperry@fs.fed.gov

**Abstract**
Current regional and national estimates of soil properties for the USA, such as organic carbon (SOC) storage or root zone available water capacity (AWC), are based on analysis of soil maps developed at a small scale and using methods that have considerable uncertainty. Recent improvements in the availability of detailed digital soils data, as well as computing capacity to handle large spatial data sets and statistical approaches to incorporate existing data in various formats, provide an opportunity to develop more detailed and accurate estimates of soil properties. Our objective is to improve the accuracy and precision of regional and national soil property estimates using spatial disaggregation techniques that combine detailed soil class maps with spatial data on environmental covariates such as topography and geology to discern the spatial distribution, variability, and extent of component soils--and the associated soil properties--within soil map units. A regional approach is employed based on recognized major land resource areas (MLRA), which are expected to have relatively consistent soil-landscape relationships. Two map units of large extent in the southern portion of the Eastern Allegheny Plateau and Mountains (MLRA 127) provide an illustration of the disaggregation approach to produce raster-based, landscape-scale maps of SOC. The disaggregated data identifies locations of component soils within soil class map units, depicting the spatial distribution of soils with higher and lower SOC stocks instead of using an average SOC value for the entire extent of a soil map unit. For this example, the disaggregated data predicted 6% higher average SOC content compared to the published soil class map data.

**Key Words**
GlobalSoilMap.net. digital soil mapping, soil organic carbon, soil survey, SSURGO.

**Introduction**
The GlobalSoilMap.net project seeks to produce continental-scale maps of soil properties using a raster format. The anticipated soil property data layers are soil organic carbon (SOC) content, clay content, and bulk density, with additional properties such as carbon density and available water capacity (AWC) predicted using pedotransfer functions. These data are of interest to soil scientists and to other environmental scientists, modelers, and policy-makers.

In the United States, current regional and national estimates of soil properties, such as SOC storage or root zone AWC, are based on analysis of the USDA–NRCS State Soil Geographic Database (STATSGO2; Soil Survey Staff 2006) (e.g., Bliss *et al.* 1995). However, STATSGO2 was developed at a small scale and the methods used to create STATSGO have considerable uncertainty. The impending completion of the initial soil survey of private lands in the USA will allow the more detailed Soil Survey Geographic Database (SSURGO) to be used to estimate soil properties. Because of its larger scale and finer detail, SSURGO based estimates of soil properties will be more precise especially when coupled with land use data and estimates of management induced differences in soil properties. Yet using SSURGO data to develop estimates of SOC, AWC, and other soil properties presents its own challenges. For example, artificial boundaries in the data associated with geopolitical boundaries lead to discontinuities in map unit composition and soil property data. Within SSURGO map units, unnamed components (e.g., components designated as "Other soils") are not included in the determination of soil properties, but may represent a significant proportion of map unit composition. Even when all components are named, the individual components can vary greatly in soil properties but the location of these components within the larger map unit delineation is not represented.

The goal of this project is to improve the accuracy and precision of regional and national soil property estimates by developing models based upon SSURGO polygon data and data from USDA-NRCS and other databases. Our objective is to combine the SSURGO data with spatial data on environmental covariates such as topography and geology, to discern the spatial distribution, variability, and extent of the individual components within SSURGO map units. Digital soil mapping techniques will provide added value soil survey data to meet the needs of a wider user community. These products will include disaggregated polygon maps (soil component maps) and soil property maps at a variety of resolutions. This approach will provide for more reliable data on soil properties, and give modelers, policy-makers, and planners better data sets to develop assessments and form public policy.

**Methods**

Digital soil mapping technology allows for the production of raster-based, landscape-scale predictions of soil classes or continuous soil properties at a variety of resolutions, and SOC density is a soil property that is of great interest to modelers, policy-makers, and planners. The methods and results presented here focus on SOC, but are applicable to the estimation and mapping of other soil properties.

The SOC stock calculated for a given SSURGO map unit represents an average SOC value based on all of the component soils identified in the map unit. However, each component, while not mapped spatially, often occurs in specific landscape positions. For example, a map unit may consist of two components, with the first is found predominantly on north-facing slopes and the second on south-facing slopes. In this case, slope aspect could be used to predict the distribution of these soils within the map unit. It has been shown that soil map units can be disaggregated into individual components based on soil-landscape relationships documented in existing soil surveys (Bui *et al.* 1999; Bui and Moran 2001).

*Regional approach*

Major land resource areas (MLRA) are “geographically associated land resource units” (USDA-NRCS 2006) that have been established to aid in state, regional, and national planning. MLRA regions delineate areas with similar physiography, geology, climate, soils, and hydrology relative to agricultural productivity. For the US National Cooperative Soil Survey (NCSS), MLRA regions are the basis for all future soil survey updates and management. It is also expected that within MLRA, soil-landscape relationships and environmental covariates are mostly homogeneous, making MLRA regions useful subdivisions for development of spatial disaggregation rules and soil-landscape models based on relationships between soils and environmental variables. MLRA 127 (Eastern Allegheny Plateau and Mountains) was selected for this case study. MLRA 127 is located in the northeastern USA, including eastern West Virginia and central Pennsylvania, as well as parts of western Virginia, western Maryland, and southern New York. It covers 50,370 km$^2$, with a range in mean annual precipitation of 840 to 1,725 mm and a range in mean annual air temperature of 6 to 12°C. The steep slopes of this highly dissected plateau expose the level-bedded sandstone, shale, coal, and limestone strata that underlie this landscape (USDA-NRCS 2006). The dominant soils across MLRA 127 are Ultisols and Inceptisols.

*Environmental variables*

Terrain attributes were derived from digital elevation model (DEM) data acquired from US Geologic Survey National Elevation Dataset with a resolution of 30 m. Terrain attributes calculated from these DEM included slope gradient, slope aspect, profile (down slope) curvature, contour (cross-slope) curvature, total curvature, tangential curvature, and relative slope position. Hillslope elements, which are defined based on differences in slope steepness and slope curvature, were derived using the methods of Schmidt and Hewitt (2004).

*SOC estimation*

Initial SOC estimates were calculated using the methods of Bliss *et al.* (1995) using published SSURGO data. The SSURGO databases report a high, a low, and a representative value of soil organic matter for each soil horizon. These values are converted to SOC values by dividing by 1.724 (Soil Survey Laboratory Staff 1996). The SOC content of each horizon (to a depth of 20 cm or 100 cm) was calculated using SOC content, bulk density, thickness, and rock fragment content data of each horizon. The SOC content of each horizon was summed over the prescribed depth to determine the SOC content of each soil in the survey area. The SOC content of each map unit was then calculated as the weighted average of all the component soils represented in each map unit.

*Spatial disaggregation of SSURGO*

Spatial disaggregation provides a process for separating soil map units into individual components. Our approach focused on the conversion of soil information encoded as unmapped entities (inclusions and minor components) within a map unit polygon to a sequence of component soils across the landscape. Soil-landscape patterns and relationships that are embedded in the soil map unit descriptions in soil survey reports or stored as a series of values within the USDA-NRCS National Soil Information System (NASIS) database were used to develop spatial disaggregation rules. These rules were applied to the SSURGO data and the ancillary digital data that are used to represent key landscape characteristics (e.g., slope gradient, slope aspect, landform elements) to map the spatial extent of individual components. Disaggregated component soils were assigned SOC values derived from SSURGO database for the same named components. These raster maps were produced with a horizontal resolution of 30 m.

**Results**

Two map units of large extent in the southern portion of MLRA 127 (Figure 1a) provide an illustration of the disaggregation approach. The Gilpin-Laidig association is mapped across 40,530 ha in MLRA 127. According to the Soil Survey of Webster County, WV (Delp 1998), this map unit consists of about 45% Gilpin soils, 35% Laidig soils, and 20% other soils. The Gilpin soils are typically found on upper backslopes, while the Laidig soils are found on the lower backslopes. The other soils included in this map unit are the Cateache and Dekalb soils on ridges and shoulders, the Guyandotte soils in north-facing hollows and footslopes, the Meckesville soils on lower backslopes, Pineville and Shouns soils in south-facing hollows and footslopes, and Craigsville soils in drainageways. The Pineville-Gilpin-Guyandotte association is mapped across 18,098 ha in MLRA 127. According to the Soil Survey of Webster County, WV (Delp 1998), this map unit consists of 35% Pineville soils, 25% Gilpin soils, 15% Guyandotte soils, and 25% other soils. The Pineville soils are typically found on lower backslopes and south-facing hollows, the Gilpin soils are found on upper backslopes, and the Guyandotte soils on north-facing upper backslopes and north-facing hollows. The other soils included in this map unit are the Dekalb soils on ridges and shoulders, the Laidig soils on footslopes, and Craigsville soils in drainageways. These descriptions were used to develop the spatial disaggregation rules to be applied to the SSURGO data (Figure 1a) and the various DEM derivatives to develop disaggregated component soil maps (Figure 1b). For example, if an area is mapped as Pineville-Gilpin-Guyandotte association and the DEM-derived hillslope element is a north-facing lower backslope, then that grid cell is designated as Guyandotte. However, if an area is mapped as Pineville-Gilpin-Guyandotte association and the DEM-derived hillslope element is shoulder, then that grid cell is designated as Dekalb.

(a)

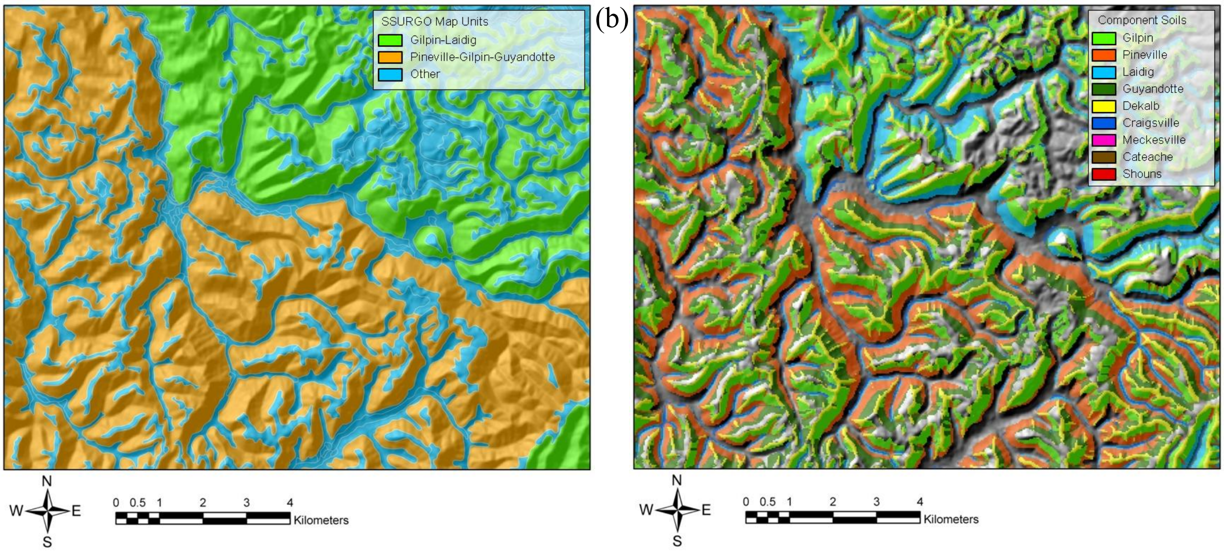



**Figure 1. An example of (a) the published SSURGO data for a portion of southern MLRA 127 showing the large extent of the two survey map units and (b) the disaggregated component soils for this same area.**

Each of the soils in the SSURGO map units have different amounts of SOC. When estimating SOC stocks from the SSURGO data the lack of spatial representation of component soils leads to a lack of spatial detail in the representation of SOC stock by the SSURGO data (Figure 2a) because an average value for the entire map unit must be used. Using the disaggregated soils map, it is possible to depict the locations of areas of soils with higher (e.g., Guyandotte) and lower (e.g., Gilpin) SOC stocks. As a result, the disaggregated data predicted a different amount of SOC for the area covered by the disaggregated map units. For example, for the Pineville-Gilpin-Guyandotte association, the disaggregated data predicted a 6% higher average SOC content compared to the published SSURGO data for the area.

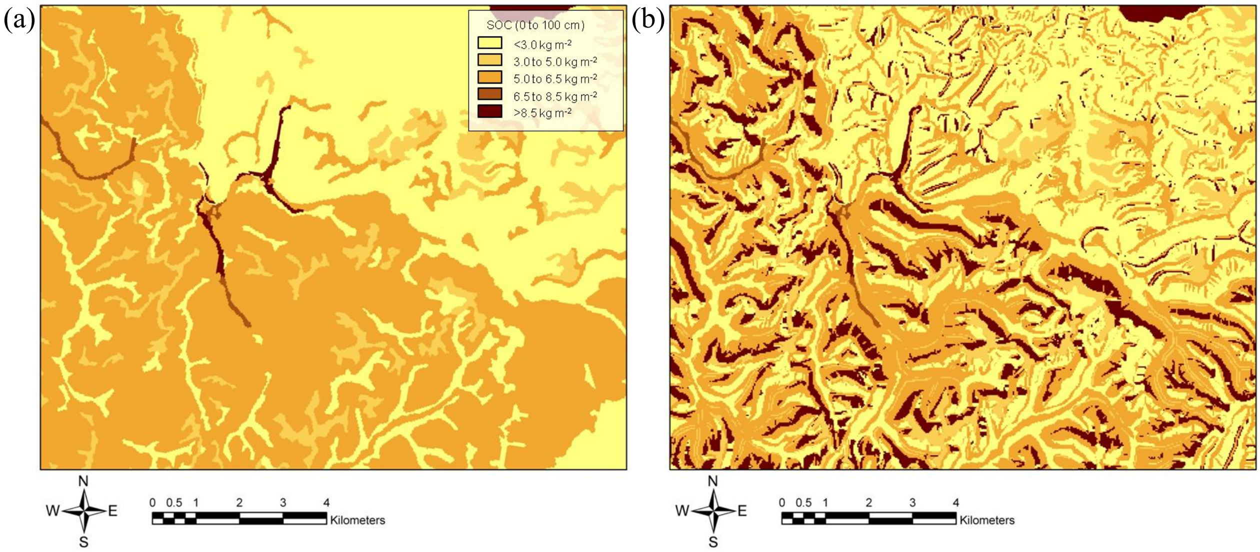



**Figure 2. Calculated SOC stock in the upper 100 cm of soil as determined from (a) the published SSURGO data for a portion of southern MLRA 127 and (b) the disaggregated component soils for this same area.**

## Conclusion

Spatial disaggregation provides a methodology for representing the spatial distribution of component soils that are known to occur within a SSURGO map unit, including both the dominant soils and the included minor soils. Furthermore, the disaggregated soil map units can be used to represent the spatial distribution of soil properties that are associated with the component soils, such as SOC stock or root zone AWC. This spatial disaggregation approach will require an MLRA-wide examination of map unit composition and component landscape properties spanning the numerous soil survey areas within the MLRA. As a result, it will be necessary to harmonize information on soil map units, component soils, and component soil properties, including (i) correlating soil map units between existing soil survey area legends, (ii) updating geomorphic properties associated with each component soil, (iii) reconciling soil property values associated with component soils, and (iv) rectifying positional displacement of SSURGO map unit delineations compared to DEM-derived landform elements.

## References


Bliss NB, Waltman SW, Petersen GW (1995) Preparing a soil carbon inventory for the United States using geographic information systems. In 'Soils and global change'. (Eds R Lal *et al.*) pp. 275-295. (CRC Press: Boca Raton, FL)

Bui EN, Loughhead A, Corner R (1999) Extracting soil–landscape rules from previous soil surveys. *Australian Journal of Soil Research* **37**, 495– 508.

Bui EN, Moran CJ (2001) Disaggregation of polygons of surficial geology and soil maps using spatial modeling and legacy data. *Geoderma* **103**, 79– 94.

Delp CH (1998) Soil Survey of Webster County, West Virginia. USDA-Natural Resources Conservation Service, WV Agric. Exp. Stn., and USDA-Forest Service (U.S. Gov. Print. Office: Washington, DC)

Schmidt J, Hewitt A (2004) Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma*, **121**, 243-256.

Soil Survey Staff (2006) U.S. General Soil Map (STATSGO2). USDA-Natural Resources Conservation Service. Available online at http://soildatamart.nrcs.usda.gov

Soil Survey Staff (2008) Soil Survey Geographic Database (SSURGO). USDA-Natural Resources Conservation Service. Available online at http://soildatamart.nrcs.usda.gov

USDA-NRCS (2006) Land Resource Regions and Major Land Resource Areas of the United States, the Caribbean, and the Pacific Basin. U.S. Department of Agriculture Handbook 296.

# Soil property mapping over large areas using sparse ad-hoc samples

A-Xing Zhu[A,B], Jing Liu[B,C], Chengzhi Qin[B], Shujie Zhang[B], Ya-ning Chen[D] and Xingwang Ma[E]

[A]Department of Geography, University of Wisconsin, Madison, Wisconsin, U.S.A.
[B]State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China.
[C] School of Geography, Beijing Normal University, Beijing 100875, China.
[D]Key Laboratory of Oasis Ecology and Desert Environment, Xinjiang Institute of Ecology and Geography, Chinese Academy of Sciences, Urumqi, Xinjiang 830011, China.
[E]Institute of Soil and Fertilizer, Xinjiang Academy of Agricultural Sciences, Urumqi, Xinjiang 830000, China.

**Abstract**
This paper presents a new approach to predict soil properties and quantify uncertainty in the derived soil property maps over large areas using sparse and ad-hoc samples. According to the soil-landscape model, each soil sample contains corresponding relationships between soil and environment conditions. Under the assumption that the more similar the environment conditions between two locations the more similar the soil property values, each sample can be considered as a representative (individual representativeness) over areas of similar environmental conditions. The level of representativeness of an individual sample to an unsampled location can be approximated by the similarity in environmental conditions between the two locations. Based on this "individual representativeness" concept and with the use of Case-based Reasoning (CBR) idea, which solves new problems by referring to similar cases, soil property values at unsampled locations can be predicted based on their environmental similarity to the individual samples. Furthermore, the uncertainty associated with each prediction is related to the similarity and can thus be quantified. A case study located in Illy Region, Xinjiang, Northwest China, has demonstrated that the predicted map of soil organic matter of top layer is of good quality and the quantified uncertainty is positively correlated with prediction residual. This suggests that the approach can be an effective alternative for predicting soil property and reporting uncertainty in the resulting soil map over large areas with sparse and ad-hoc samples.

**Key Words**
Individual representativeness, soil-landscape model, case-based reasoning, digital soil mapping, uncertainty, SoLIM.

## Introduction
Information on spatial variation of soil properties over large areas is a critical piece of input data for environmental modeling at the regional to continental scales (Abramopoulos *et al.* 1988; Bonan 1996; Dai and Zeng 1996; Chen and Dudhia 2001, Zhu and Mackay 2001). Yet, quality information on soil spatial variation over large areas is rather difficult to obtain due to the large number of field samples needed and the requirement of sound global representativeness imposed by the existing mapping techniques (Journel and Huijbregts 1978; Isaaks and Srivastava 1989; Cressie and Noel 1993; Goovaerts 1999; Mitas and Mitasova 1999; Schloeder *et al.* 2001; McBratney *et al.* 2003; Zhu *et al.* 2008). Due to the constraints of field conditions and project budget and the complexity of spatial variation of soil properties, field sampling can rarely meet these requirements (both the number of samples and the sound global representativeness). As a result, the collected samples are often sparse and ad-hoc (poor global representativeness) in nature. The soil property maps derived based on these samples using the existing mapping techniques are not only at low quality but also lack the information on the uncertainty introduced by samples' poor global representativeness. The lack of uncertainty information in the derived soil property maps also prevents proper uncertainty assessment of model outputs when the derived soil information is used as one of the inputs.

## Methods
The approach is based on the concept of soil-landscape model (Jenney 1941; McBratney *et al.* 2000; McBrateney *et al.* 2003) which states that each sample contains certain corresponding relationship between soil and associated environmental conditions in parameter space. With the assumption that the more similar the environment conditions between two locations the more similar the soil property values, each sample can be considered as a representative over locations (not necessarily contiguous) with similar environmental conditions, that is, each sample owns "individual representativeness". The level of representativeness of an

individual sample for an unsampled location can be approximated by the similarity in environmental conditions between the two locations. Based on this "individual representativeness" concept and with the use of Case-based Reasoning (CBR) idea, which solves new problems just by identifying existing similar cases while not requiring the presence of a global model for the entire problem domain (Aamodt and Plaza 1994; Watson and Abdullah 1994; Leake 1996; Watson 1998), soil property values at unsampled locations can be predicted based on the environmental similarities to the individual samples. Moreover, the uncertainty associated with each prediction is related to the similarity. For example, if a location is not similar or at a low degree of similarity to the current set of individual samples, the uncertainty associated with the predicted value for that location is high because none of the existing samples is a good representative of this location. Then, the uncertainty associated with the prediction at each location can be quantified by analyzing the nature of the similarity values to the individual samples (Zhu 1997).

The new approach consists of three major components: 1) The selection of environment variables (covariates) and characterization of associated environment conditions using these variables; 2) Calculation of similarity in environmental conditions; 3) Estimation of soil property value and quantify uncertainty based on the environmental similarity.

For environment characterization, the selected environment variables should be responsible for soil formation or co-varying with soil closely so that they can be used to indicate spatial variation of soil effectively. The approach uses a raster data model for spatial representation. For soil mapping over large areas the grid size is often large. The characterization of environmental conditions over large grid size depends on the variable. For variables (climate and geology) which do not vary rapidly over the area of a pixel, we use one value to represent the environmental conditions at each pixel. For variables (such as topographic variables and vegetation variables) that vary rapidly over a pixel area we use the probability density function estimated using the Kernal Density Estimation (KDE) method to characterize the environmental conditions at each pixel.

Similarity estimation was conducted at two levels: the individual environment variable level and the case (sample) level which integrates all similarities from the individual variable level. The methods for the first level depend on the data type and the characterization method of each variable. We adopted Gower distance for measuring similarity in climate variables, Boolean function for parent materials, and a consistent Measure (CM) for topographic variables and vegetation variables which are characterized using probability density functions (Zhu 1999). The methods for the second level depend on the perception of interaction of environment variables. With the knowledge that over large area climate conditions would control the general spatial distribution pattern of soil, parent material would then differentiate soils in the same climate zone, while specific topographic conditions would influence the local variation in the same parent material area, we adopted a hierarchy approach in this research to integrate the similarities from individual variables.

For uncertainty quantification and soil property prediction, similarities at each location to individual samples would form a similarity-vector characterizing the representativeness of sample cases at that location. By analyzing this similarity vector, uncertainty associated with the prediction related to samples' representativeness was quantified (Zhu 1997). Soil property value at an unsampled location was predicted using a similarity weighted average method which integrates similarities with sample attributes. The result from this approach contains two parts: a soil property map and the associated uncertainty map.

## Results

A case study located in Illy Region, Xinjiang, Northwest China, has been conducted to examine the validity of this approach. The study area is about 50,000 $km^2$ in size. The variables used are: average annual precipitation, average annual temperature, average annual relative humidity, maximum and minimum monthly precipitation, maximum and minimum monthly temperature, and maximum and minimum monthly relative humidity, parent materials; elevation, slope gradient, profile curvature, surface area ratio and land position index. A cross validation method with 73 field observation points was used to evaluate the performance of the method. The RMSE between the predicted and the observed values is 0.32 which is much smaller than 3.16, the standard deviation of these 73 field points. The correlation coefficient between the values of uncertainty and the prediction residuals at these points is 0.537 which is significant at the 0.05 level.

## Conclusions

This paper presented a new approach to predict soil property over large area based on "individual representativeness" of sparse ad-hoc samples. This approach does not require the global representativeness of the whole sample set and is able to quantify prediction uncertainty introduced by the poor global representativeness of the sparse and ad-hoc samples. The results suggest that this approach is an effective and accurate way to map soil properties over large areas and is capable of providing uncertainty associated with the derived property map. The uncertainty information is a valuable piece of information for evaluating the credibility of prediction at each location. We conclude that this approach can serve as an effective alternative for predicting soil property and reporting prediction uncertainty over large areas with sparse and ad-hoc samples.


## Acknowledgements

This study is supported by National Natural Science Foundation of China (No. 40601078), National Basic Research Program of China (No. 2007CB407207), and Key Projects in the National Science & Technology Pillar Program during the Eleventh Five-year Plan Period (2007BAC15B01-1).


## References


Aamodt A, Plaza E (1994) Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. *AI Communications* **7**, 39-59.

Abramopoulos F, Rosenzweig C, Choudhury B (1988) Improved Ground Hydrology Calculations for Global Climate Models (GCMs): Soil Water Movement and Evapotranspiration. *Journal of Climate* **1**, 921-941.

Bonan GB (1996) A land surface model (LSM version 1.0) for ecological, hydrological, and atmospheric studies: technical description and user's guide. NCAR Tech. Note NCAR/TN-417+STR, 150 p.

Chen F, Dudhia J (2001) Coupling an Advanced Land Surface–Hydrology Model with the Penn State–NCAR MM5 Modeling System. Part I: Model Implementation and Sensitivity. *American Meteorological Society* **129**, 569-585.

Cressie N, Noel AC (1993) Statistics for Spatial Data. John Wiley & Sons, Inc, New York, 900 p.

Dai Y, Zeng QC (1996) A land surface model (IAP94) for climate studies. Part I: Formulation and validation in off-line experiments. *Advances in Atmosphere Science* **14**, 433-460.

Goovaerts P (1999) Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* **89**, 1-45.

Isaaks EH, Srivastava RM (1989) Applied Geostatistics. Oxford University Press, New York, 561p.

Jenny H (1941) Factors of Soil Formation: A System of Quantitative Pedology. Dover Publ., New York, 281p.

Journel AG, Huijbregts CJ (1978) Mining Geostatistics. Academic Press, London, US, 600 p

Leake DB (1996) Case-based reasoning. *The knowledge engineering review* **9**, 61-64.

McBratney AB, Odeh IOA, Bishop T, Dunbar M, Shatar T (2000) An overview of pedometric techniques of use in soil survey. *Geoderma* **97**, 293-327.

McBratney AB, Mendonça Santos M, Minasny B 2003. On digital soil mapping. *Geoderma* **117**, 3-52.

Mitas L, Mitasova H (1999) Spatial Interpolation. In 'Geographical Information Systems: Principles, Techniques, Management and Applications'. (Eds P Longley, MF Goodchild, DJ Maguire, DW Rhind) pp. 481-492. (GeoInformation International: New York).

Schloeder CA, Zimmerman NE, Jacobs MJ (2001). Comparison of methods for interpolating soil properties using limited data. *Soil Sci. Soc. Am. J.* **65**, 470-479.

Watson ID, Abdullah S (1994) Developing Case-Based Reasoning Systems: A Case Study in Diagnosing Building Defects. In, Proc. IEEE Colloquium on Case-Based Reasoning: Prospects for Applications, Digest No: 1994/057, 1/1-1/3.

Watson I.D (1998) Applying Case-Based Reasoning: techniques for enterprise systems. Morgan Kaufmann Publishers Inc, San Francisco, CA, USA.

Zhu AX (1997) Measuring uncertainty in class assignment for natural resource maps using a similarity model. *Photogrammetric Engineering & Remote Sensing* **63**, 1195-1202.

Zhu AX (1999) A personal construct-based knowledge acquisition process for natural resource mapping using GIS. *International Journal of Geographic Information Science* **13**, 119-141.

Zhu AX, Mackay DS (2001) Effects of spatial detail of soil information on watershed modeling. *Journal of Hydrology*, **248**, 54-77.

Zhu AX, Yang L, Li BL, Qin CZ, English E, Burt JE, Zhou CH (2008). Purposive sampling for digital soil mapping for areas with limited data. In 'Digital Soil Mapping with Limited Data'. (Eds AE Hartemink, AB McBratney, ML Mendonca Santos) pp.233-245. (Springer-Verlag: New York)

# Soil spectral diagnostics – infrared, x-ray and laser diffraction spectroscopy for rapid soil characterization in the Africa Soil Information Service

Keith D Shepherd[A]

[A]World Agroforestry Centre (ICRAF), PO Box 30677-00100, Nairobi, Kenya, Email k.shepherd@cgiar.org

**Abstract**

One of the key challenges in establishing the Africa Soil Information Service (AfSIS) is how to measure soil functional properties on tens of thousands of georeferenced soil samples in a consistent way. To solve this problem, AfSIS uses spectral diagnostics – low cost, high throughput analytical techniques based on reflectance of electromagnetic radiation. This paper describes use of infrared spectroscopy (IR), total x-ray fluorescence spectroscopy (TXRF), x-ray diffraction (XRD), and laser diffraction particle size analysis (LDPSA) techniques. The data generated by these high-throughput techniques can all be treated as spectra and used as input to pedotransfer functions for prediction of soil functional properties that are expensive or time-consuming to measure. In addition, comparative LDPSA data for different dispersion treatments can provide functional indicators of soil stability. Further research should establish the added value or redundancy in pedotransfer functions when IR is complemented with TXRF, XRD and LDPSA data.

**Key Words**

Spectral diagnostics, infrared spectroscopy, x-ray fluorescence, laser diffraction, Africa soils, pedotransfer functions

## Introduction

The Africa Soil Information Service (www.africasoils.net) is being established to provide accurate, up-to-date and spatially referenced soil information to support agricultural development and scientific advancement in Africa. This need coincides with advances in technologies that allow for accurate collection and prediction of soil properties (Sanchez *et al.* 2009). The project will develop a practical, timely, cost-effective, soil health surveillance service to map soil conditions, set a baseline for monitoring changes and to provide options for improved soil management. Soil testing under AfSIS is designed to meet the diverse needs of different users: diagnosis of soil constraints for agriculture, monitoring of trends in soil health, land capability for agriculture, soil testing for engineering and stabilisation purposes, ecological and human health risk assessment; and prognostic testing to inform investment decisions (e.g. fertilizer rates, soil conditioners, soil drainage, soil conservation). Over the next four years the project will collect over 30,000 georeferenced soil samples from sub-Saharan Africa and characterize them. This paper describes how low cost high-throughput spectroscopy methods are being used both as a front line screening technique for development of pedo-transfer functions and for the direct development of indicators of soil functional properties.

## Methods

*Soil processing*

All soil samples are initially air-dried and 2-mm sieved. A 20 g sub-sample of soil is obtained by coning and quartering and hand-ground using an agate pestle and mortar to pass a 75 µm sieve. These finely ground samples can be shipped at low cost and are adequate for analysis by mid-infrared diffuse reflectance spectroscopy (MIR), TXRF, XRD, LDPSA, and total CNS analysis by combustion.

*Infrared spectroscopy*

Diffuse reflectance infrared spectroscopy (IR) is an established technology for rapid, non-destructive characterization of the composition of materials based on the interaction of electromagnetic energy with matter. Both the visible near infrared (VNIR, 0.35-2.5 µm) and mid infrared (MIR, 2.5-25 µm) wavelength regions have been investigated for non-destructive analyses of soils and can potentially be usefully applied to predict a number of important soil properties. including: soil colour, mineral composition, organic matter and water content (hydration, hygroscopic, and free pore water), iron form and amount, carbonates, soluble salts, and aggregate and particle size distribution (Shepherd and Walsh, 2004; 2007). In AFSIS, IR is used as a frontline screening tool in regional laboratories using 2-mm sieved air-dried soil samples. The regional laboratories are equipped with fourier-transform NIR spectrometers with in-built gold reference and instrument validation routines to ensure reproducibility of results over time and among laboratories. All other measurements are centralized at the World Agroforestry Centre's Soil-Plant Spectral Diagnostics Laboratory, except for conventional extraction soil tests, which are done using ICP mass spectroscopy in an external certified laboratory. Fine ground samples are analysed with MIR using a robotic high-throughput system employing micro-titre plates (Shepherd and Walsh, 2007).

*Total x-ray fluorescence spectroscopy*

TXRF (KlocKenKämper 1997) provides for rapid simultaneous analysis of all elements from Na to U (except Mo) with minimal sample preparation time. The main principle of X-ray Fluorescence Spectroscopy is that atoms, when irradiated with X-rays, emit secondary X-rays – the fluorescence radiation. On this basis XRF analysis is possible because (i) the wavelength and energy of the fluorescence radiation is specific for each element, and (ii) the concentration of each element can be calculated using the intensity of the fluorescence radiation. Compared with conventional XRF, TXRF also has the advantages of greatly reduced background noise, and consequently much higher sensitivities, and a significant reduction of matrix effects. Standardisation is internal and only requires addition of an element that is not present in the sample for quantification purposes, and no external standardization is required in most cases. In AfSIS, TXRF is used to analyse total elements in soil (samples are suspended in detergent, pipetted onto carriers, and dried) and in soil water extracts after centrifuging the same sample. Lower detection limits are in the parts per million concentration range for suspended soil and parts per billion levels in soil water. The total element concentration profiles (essentially spectra) are used to fingerprint soils, to capture key mineralogical differences, and as an input to pedotransfer functions.

*X-ray diffraction*

Despite the critical importance of soil mineralogy in the determination of soil functional properties and as a soil forming factor, there has been relatively little work to move beyond largely descriptive studies (Dixon and Schulze, 2002) to the quantitative linking of soil function to soil mineralogy (Cornu *et al.* 2009). New instrumentation developments in benchtop high-throughput X-ray powder diffraction (XRD) and steady improvements in mineral identification databases and software have opened up new opportunities for quantitative determination of mineral phases on large sample numbers. AfSIS extends the infrared spectroscopy profiling approach (Shepherd and Walsh 2007) to include X-ray diffraction.

Finely ground (<50 µm) samples are loaded into sample holders and analysed using a X-ray diffractometer equipped with a compound silicon strip, 1-dimensional detector with Theta / Theta geometry. The angular range measured is 0 to 80° 2Theta with an accuracy of ± 0.02° throughout the measuring range. The raw XRD spectra (counts versus angle) can be used directly as input to pedotransfer functions, in the same way infrared spectra are used. Phase search, identification and semi-quantitative analysis are done using the International Centre for Diffraction libraries and fully quantitative phase analysis is proposed on subsets of samples using the Rietveld method. TXRF soil element information can also used to focus mineralogy searches.

*Laser diffraction particle size analysis*

Soil particle size distribution is a fundamental soil property that affects many soil functional properties, but its determination using conventional hydrometer or pipette methods suffers problems of poor repeatability and reproducibility and variable dispersion in many tropical soils, due to cementing actions of iron and aluminium hydroxides. There is uncertainty on what methods best reflect functional aspects of soil particle size distribution (e.g. dispersing aggregates using dispersion agents may not reflect functional effects in the field). In fact soil particle size is usually not interpreted directly to provide information on soil functions but is rather a covariate used in predicting or conditioning soil functional properties, such as nutrient retention, tillage properties, and hydraulic properties. Therefore emphasis should be on rapid and repeatable measures rather than accurate measures of particle size distribution.

Dry aggregate size distribution and dispersed and non-dispersed particle size distribution have been proposed as indicators of soil erodibility, even though erodibility may be affected by a number of variables. Various measures of dispersion have also been used to classify soil susceptibility to structural faults and piping in subsoil's (e.g. dam walls) and surface soil structural problems (e.g. hardsetting). Response of particle size distribution to different levels of ultrasonic energy can be used to derive an absolute measure of soil stability.

In AfSIS, laser diffraction particle size analysis is used to estimate particle size distribution and soil stability. The analysis can be done using small quantities of soil (<5 g). A representative cloud or ‘ensemble’ of particles passes through a broadened beam of laser light which scatters the incident light onto a Fourier lens. This lens focuses the scattered light onto a detector array and, using an inversion algorithm, a particle size distribution is inferred from the collected diffracted light data. Mie theory is used to provide a volume-based continuous distribution of particle sizes based on the correlation between the intensity and the angle of light scattered from particles.

AfSIS samples are analysed using a detectable size range of 0.01-3000 µm. The instrument allows continuous flow of a soil sample suspended in (i) a dry air stream or (ii) a water stream, to which different sonification cycles can be applied using an in-built ultrasonic probe. The protocol begins with measurement of particle size distribution of dry soil suspended in the air stream to provide a measure of micro-aggregation without wetting. Particle size distribution is then measured in water, followed by a second reading one minute later, and finally after full dispersion using Calgon and sonification. The shift in particle size distribution with these treatments is used to provide comparative indices of stability (Muggler *et al.* 1996). Destruction of organic matter and removal of soluble, salts, gypsum, carbonates, and iron and aluminium oxides is not done with this method, as comparisons of ‘functional’ particle size distribution are of primary interest, as opposed to accurate measurement of ‘absolute’ particle size distribution of primary particles. A subset of soils is also analysed using the conventional hydrometer method to provide correlations with the laser diffraction measurements.

**Results**

Total element concentration spectra for three contrasting soil types from Kenya are illustration in Figure 1. There is large variation in concentrations among soils in the range from element number 13 (Aluminium) to 40 (Zirconium), especially in levels of P, K, Ca, Mn, and Fe, indicating mineralogical differences. A key area of current research is how much redundancy there is in IR, TXRF, XRD and LDPSA data in prediction of functional properties such as soil water holding capacity and nutrient supply capacity.

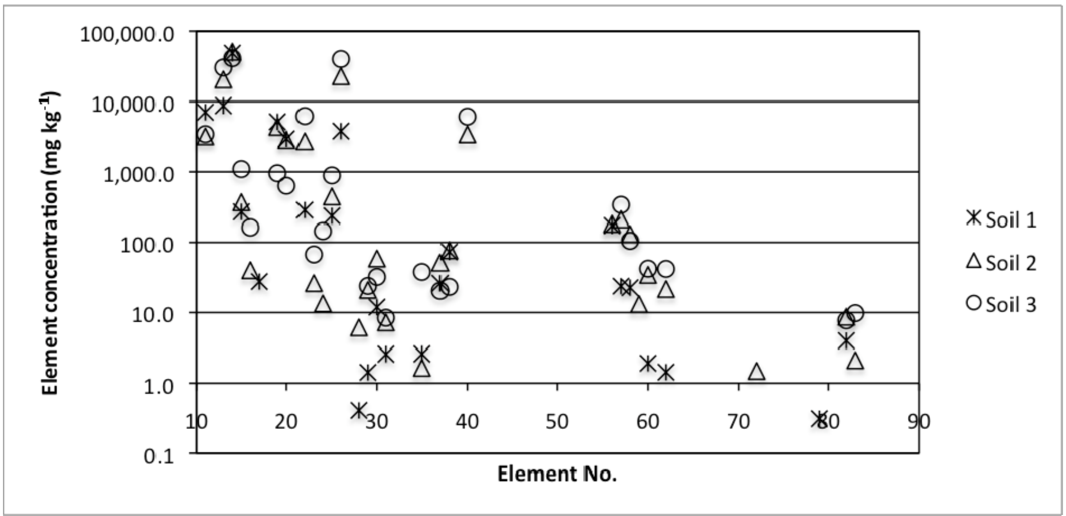



**Figure 1. Total element concentration against element number for three soils from Kenya determined using total x-ray fluorescence spectroscopy. Missing data points indicate an element was not detectable in the sample.**

The use of laser diffraction particle size analysis to assess soil stability on dispersion in water is illustrated in Figure 2. In the dry state, the unstable soil (a Fluvisol derived form lake sediments) actually had a higher proportion of aggregates >1 mm (51%) than the stable (Nitisol) soil (10%). On wetting, the stable soil showed little decrease in particle size distribution over the succession of wet treatments, whereas the unstable soil showed a successive decrease in particle sizes with the sequence of treatments. For the wet reading after one minute, the proportion of particles smaller than 10 µm was 56% in the unstable soil but only 14% in the stable soil. After the full dispersion treatment, the unstable soil had 83% of particles less than 10 µm, compared with 18% in the stable soil.

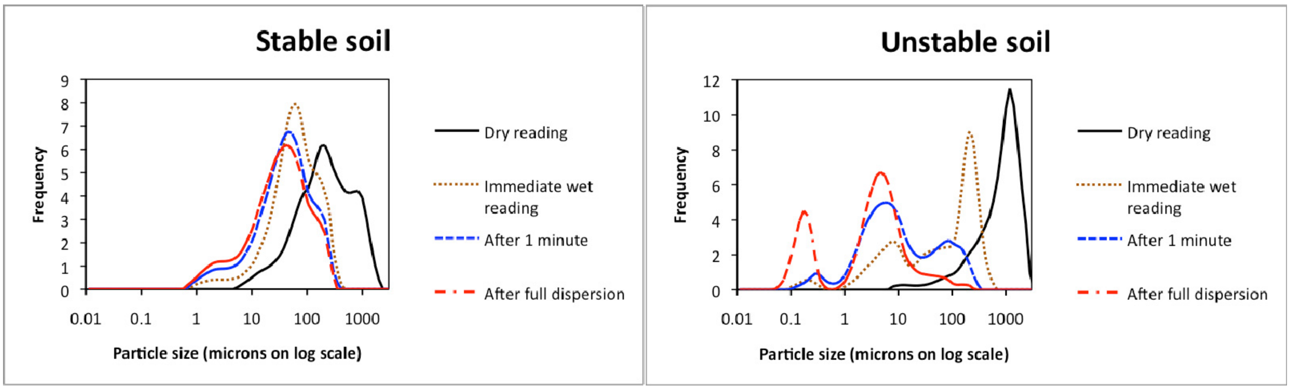



**Figure 2. Frequency distribution of particle sizes in a stable soil (Nitisol) and unstable soil (Fluvisol) from Kenya measured in (i) a dry air stream, (ii) immediately on addition to water, (iii) one minute later, and (iv) after full dispersion with calgon and sonification.**

## Conclusion

Although infrared spectral measurements of soils can predict several soil properties (such as organic carbon, exchangeable calcium, pH, and total P) calibrations need to be adjusted for different soil types. High throughput TXRF, XRD and LDPSA measurements could supplement IR as an input to pedotransfer functions and help stabilize IR calibrations across soil types with widely different mineralogy. Further research should also test whether TXRF and XRD could be useful supplements to improve prediction of properties not predicted well from IR, such as soluble or extractable nutrients. Laser diffraction particle size analysis under different dispersion treatments can serve as a rapid, functional indicator of soil stability for environmental and engineering purposes.

## References

Dixon JB, Schulze DG (2002) ‘Soil Mineralogy with Environmental Applications’. (Soil Science Society of America: Madison, WI).

Cornu S, Chevalier M, Hardy M, Bourenane H, Josiere O, Pernes M, Jolivet C, Boulonne L, Arouys D (2009) X-ray diffraction of minerals carrying trace elements in soil: application to the French soil quality monitoring network. *Comm. Soil Sci. Plant Anal*. **40**, 1138-1147.

KlocKenKämper R (1997) ‘Total-Reflection X-Ray Fluorescence Analysis’. (Wiley & Sons).

Muggler CC, Pape Th, Buurman P (1996) Laser grain-size determination in soil genetic studies. 2 Clay content, clay formation and aggregation in some Brazilian Oxisols. *Soil Sci*. **162**, 219-228.

Sanchez, PA (2009). Digital soil map of the world. *Science* **325**, 680-681.

Shepherd KD, Walsh MG (2004). Diffuse reflectance spectroscopy for rapid soil analysis. In ‘Encyclopedia of Soil Science’. (Ed R Lal) (Marcel Dekker Inc, New York. Online Published by Marcel Dekker: 04/26/200).

Shepherd KD, Walsh MG (2007) Infrared spectroscopy—enabling an evidence-based diagnostic surveillance approach to agricultural and environmental management in developing countries. *Journal of Near Infrared Spectroscopy* **15**, 1-19.