Specifications

Tiered[1] *GlobalSoilMap* products

Release 2.4
[07/12/2015][2]

Science Committee

[1] Tier 1 – point predictions at centre of a global3 arc-second grid; Tier 2 - additional 100mx100m block predictions centred on 3 arc-second grid cells

[2] These specifications have been modified as a result of an Uncertainty Workshop held at USDA NRCS, Lincoln, Nebraska, August 27-30, 2012 and an ad hoc meeting in Jeju, Korea, June 11, 2014.

These specifications are expected to remain valid for a period of 2 years from the date of endorsement

Date of Endorsement: 07 12 2015

Table of Contents

...

List of Figures

List of Tables

Specifications for *GlobalSoilMap* products[3][4]

Context and objectives

There is a need for accurate, up-to-date and spatially referenced soil information. This need has been expressed by the modeling community, land managers, policy developers and decision makers. The need coincides with an enormous leap in technologies that allow for improvements in more accurately collecting and predicting soil properties.

A global consortium has been formed to make a new digital soil map of the world using state-of-the-art and emerging technologies. This new *GlobalSoilMap* will include prediction of soil properties at fine spatial resolution (~100 m). These data will be supplemented by interpretation and functionality options to support improved decisions for a range of global issues such as food production and hunger eradication, climate change, and environmental degradation. This is an initiative of the Digital Soil Mapping Working Group of the International Union of Soil Sciences (IUSS).

For more information see www.globalsoilmap.net

This document sets out the Specifications for *GlobalSoilMap* project data products (Tiers 1 to 4). The Specifications do not prescribe how the products must be made; only what they need to conform to in order to permit global collation and presentation of consistent standardized data.

Specifications Summary

The Specifications focus on five aspects (Table 0):
1. The spatial entities
2. The soil properties to be predicted (and the date associated with their prediction)
3. The uncertainties for each soil property prediction
4. The age of the data or information used to estimate the predicted property values
5. The validation measures to be used and reported.

[3] Agreement on the initial specifications was achieved at the *GlobalSoilMap* node meeting in Seoul, Korea on October 25-26, 2009.

[4] Agreement on version (2.1) of the specifications was achieved in June, 2011. Agreement on version (2.2) of the specifications was achieved in February 2012.

Table 0 – Summary of *GlobalSoilMap* Tiered data products

Tier	**Spatial entity**	**Grid**	**Properties**	**Uncertainty**	**Date stamping**	**Validation measure**
1	'point' with specified X, Y coordinates	Located at center of cells of 3 arc second by 3 arc second grid	Point estimate for all properties in Tables 3, 4 & 5 by standard depths	Upper and lower 90% PI for all properties at all depths	N/A	N/A
2 (includes all Tier 1 products)	100 m by 100m block	Centred on cells of 3 arc second by 3 arc second grid	Block average for all properties in Tables 3, 4 & 5 by standard depths	Upper and lower 90% PI of block average for all properties at all depths	Year or period of field data collection	N/A
3 (includes all Tier 2 products)				Marginal probability distribution for each and every xyz-point		National/Regional RMSE, etc. for point and block predictions for all properties at all depths, by independent (probability) sampling
4 [A, AA, AAA][5] (includes all Tier 3 products)				Complete spatial-multivariate probability distribution (joint probability distribution for all soil properties and xyz-points)		Thresholds of RMSE and % coverage of true values are met for each prediction

[5] accuracy thresholds for each attribute (and each depth) will be specified in tabular form, linking average prediction interval width, and the percentage of independent validation samples that fall within the prediction intervals (for A, AA, AAA etc)

GlobalSoilMap will deliver estimated values (and a measure of uncertainty) for an agreed set of functional soil properties representing two globally defined spatial entities.

The Tier 1 spatial entity is a point location with defined X, Y coordinates. *GlobalSoilMap* points are located at the cell centres of a global 3 arc second grid, which is defined to exactly match the NASA Shuttle Radar Digital Elevation Model grid (extended north and south to the poles).

The Tiers 2 to 4 spatial entity is a volumetric grid cell (a **voxel**) which represents an area of 100 m by 100 m horizontal dimensions centred at the points defined by the primary spatial entity.

In the vertical dimension, predictions of soil property values and their associated uncertainties will be made to 2 m (if total profile depth is 2 m or more)) with data reported for 6 standard depth intervals of 0-5 cm, 5-15 cm, 15-30 cm, 30-60 cm, 60-100 cm and 100-200 cm.

Twelve soil properties will be predicted for each spatial entity (Tables 3, 4 & 5). These are:

(1) total profile depth (cm)
(2) plant exploitable (effective) soil depth (cm)
(3) organic carbon (g/kg)
(4) pH (x10)
(5) sand (g/kg)
(6) silt (g/kg)
(7) clay (g/kg)
(8) gravel (m^3 m^{-3})
(9) ECEC ($cmol_c/kg$)
(10) bulk density of the fine earth (< 2 mm) fraction (excludes gravel) (Mg/m^3)
(11) bulk density of the whole soil in situ (includes gravel) (Mg/m^3) and
(12) available water capacity (mm).

Additional soil properties including, for example, EC (dS/m) may be predicted at the discretion of the Nodes but these are not mandatory.

Each soil property will have an estimate of the uncertainty associated with the prediction for each depth (for properties reported by depth) for each spatial entity. For Tier 1 and 2 uncertainty is defined here as the 90% prediction interval (PI), which is the range in values within which the true value at any prediction location is expected to be found 9 times out of 10 (90%). Methods of estimating uncertainty are not specified here but are outlined in appendices.

Each data provider will be responsible for fully documenting the inputs and procedures used to generate all products they submit. This documentation will be specified as metadata attached to each product. A template for recording metadata about how all outputs are produced will be provided as an appendix to these Specifications.

1. Spatial entities

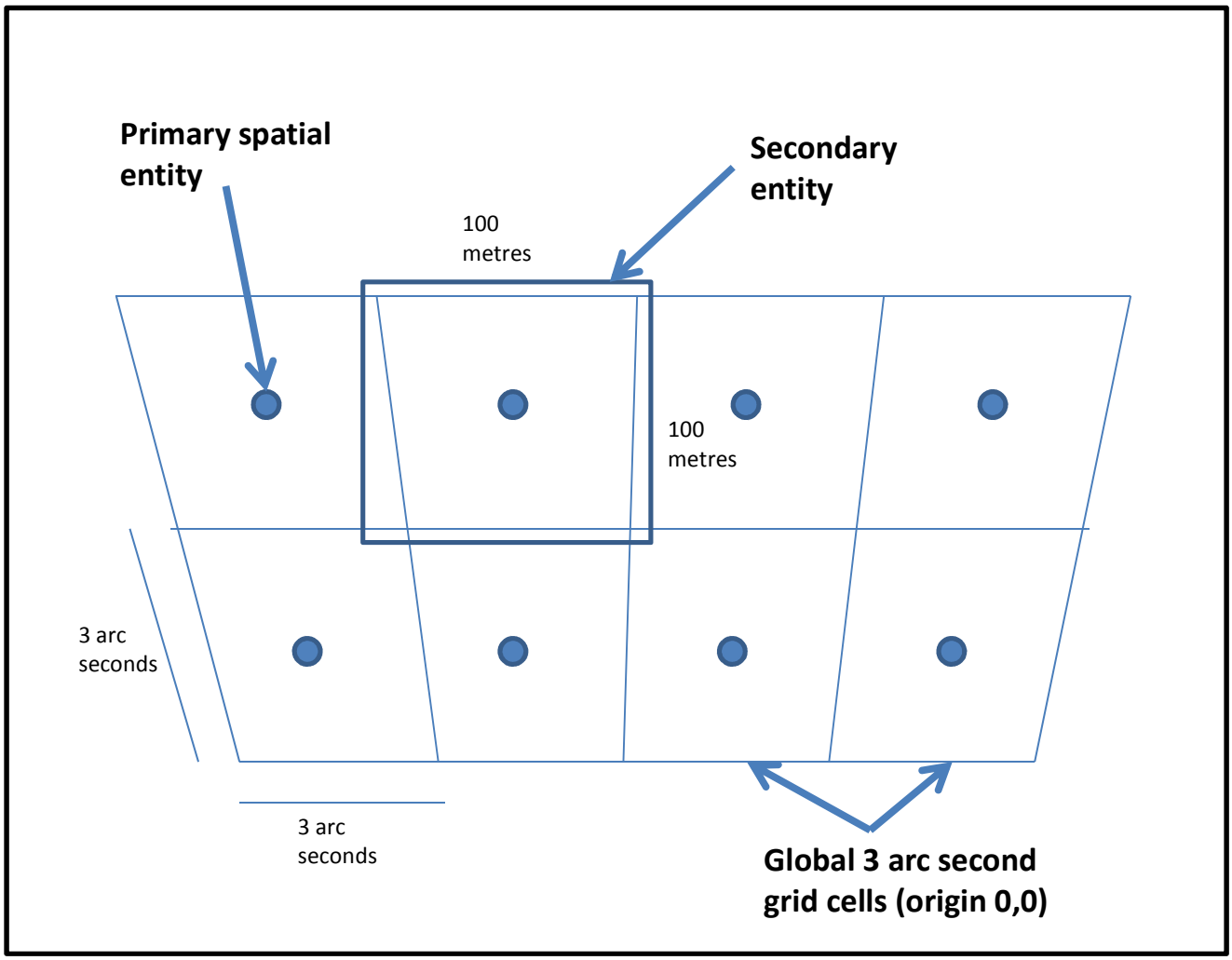
1.1 Definition of entity location and dimensions

Two spatial entities are defined (Figure 1).

The primary spatial entity for Tier 1 products is a point location with defined X, Y coordinates. *GlobalSoilMap* points are located at the cell centres of a global 3 arc-second grid, which exactly matches the NASA Shuttle Radar Digital Elevation Model grid (extended north and south to the poles). These points represent a volume of soil to a depth of 2 m (or depth to bedrock if less than 2 m) for a small irregular area with horizontal dimensions of less than 2 m by 2 m.

The secondary spatial entity for delivery in subsequent Tiers (2-4) is defined as a volume of soil to a depth of 2 m (or depth to bedrock if less than 2 m) for an area with regular, fixed horizontal dimensions of 100 m by 100 m located at the centre of the defined global 3 arc-seconds by 3 arc-seconds grid (Note: as the global 3 arc-second cells are approximately only 93 m x 93 m at the equator, the *GlobalSoilMap* blocks will overlap by increasing amounts towards the poles).

Figure 1: *GlobalSoilMap* spatial entities and the global 3 arc-second grid



1.2 Vertical Dimension- Depth

Depth is measured from the soil surface. For mineral soils, the soil surface is the top of the mineral soil. For organic soils (or mineral soils with an O horizon), the top of any surface horizon identified as an O horizon is considered the soil surface. The soil surface is the top of the part of the O horizon that is at least slightly decomposed. Fresh leaf or needle fall that has not undergone observable decomposition is excluded when determining soil depth. For soils with a cover of 80 percent or more rock fragments on the surface, the depth is measured from the surface of the rock fragments (Soil Survey Division Staff, 1993: Chapter 3 page 4).

A value will be predicted for 12 soil properties, and for the uncertainty associated with this prediction, for six fixed depth intervals (Table 1). The values reported for these six depth

intervals act as coefficients for a spline function that will provide a mechanism for reporting continuous variation with depth for all properties at all grid cells.

In addition, a value, and associated uncertainty, will be predicted for each of depth to bedrock or consolidated material and for plant extractable (effective) soil depth (depth to restricting layer).

Table 1. Depth intervals for which soil property values and uncertainty will be provided

No.	Depth Interval	Lower 5 Percentile of mean	Estimated Value of Soil Property	Upper and lower 95 Percentile of mean
1	0 - 5 cm			
2	5 - 15 cm			
3	15 - 30 cm			
4	30 - 60 cm			
5	60 - 100 cm			
6	100 - 200 cm			
7	Total profile Depth		Depth to rock in cm	
8	Plant exploitable Depth		Effective Depth in cm	

Total profile depth is depth to a lithic or paralithic contact in cm as defined below. ***Total profile depth*** refers to the depth to fixed rock. Hard and soft bedrock are distinguished. Hard bedrock is usually indurated but may be strongly cemented, and excavation difficulty would be very high or higher. Soft bedrock meets the consistence requirements for paralithic contact (Soil Survey Division Staff, 1993, Chapter 6 page 13).

Plant Exploitable (Effective) Depth is defined as: “The lower limit of biologic activity, which generally coincides with the common rooting depth of native perennial plants" (Soil Survey Staff, 1975; Soil Survey Division Staff, 1993, Chapter 1 page 5). Plant exploitable depth is the depth to a physical /chemical barrier at less than 2 m depth, or if a barrier is not present to the depth of inferred bioactivity or perennial plants.

We can define **plant exploitable soil depth** by either the evidence of the roots themselves, or on the presence of barriers to root extension. The first option requires rules for root abundance to define the lower limit, or inferences on the depth of native roots from soil morphology. Depths may differ between biomes as given in Table 2. Although logical, the approach is complex. *GlobalSoilMap* soil specifications need to be capable of consistent application across the globe, and not reliant on complex accessory data. **The second option defines the depth of a relatively-easy-rooting zone from the soil surface to a root boundary**. The boundary is defined by one or more morphological barriers. In addition to hard and soft bedrocks cited previously, these barriers include: clean sand, pan, high-density material (bulk density >1.85), extremely gravelly or densely packed gravel, permanent water table and chemical toxicity.

The extremely deep rooting ability of some tree species in arid land, where roots penetrate to great depth in jointed rock is noted and would not be recognised in option two. The choice of option needs to consider the application of the data. If it is to map ecosystem behaviour across different biomes then option one is favoured. If it is to explore opportunities for regional or global food production then the agronomic depth provided on option 2 is favoured. ***The second***

option is proposed for these specifications because it is more likely to provide a consistent global soil information product and is likely to be more widely used*. Where option one is required it is suggested that it be mapped as an additional layer by Nodes*.

Table 2. Summary of maximum rooting depth by biome (after Canadell *et al*., 1996)

Biome	*N*	Mean maximum rooting depth (m)	Highest value for rooting depth (m)
Boreal Forest	6	2.0 ± 0.3	3.3
Cropland	17	2.1 ± 0.2	3.7
Desert	22	9.5 ± 2.4	53
Sclerophyllous shrubland and forest	57	5.2 ± 0.8	40
Temperate coniferous forest	17	3.9 ± 0.4	7.5
Temperate deciduous forest	19	2.9 ± 0.2	4.4
Temperate grassland	82	2.6 ± 0.2	6.3
Tropical deciduous forest	5	3.7± 0.5	4.7
Tropical evergreen forest	5	7.3 ± 2.8	18
Tropical savanna	15	15.0 ± 5.4	68
Tundra	8	0.5 ± 0.1	0.9

1.3 Excluded Non-soil Areas

Predictions of soil properties will not be made for spatial entities that are considered to be occupied wholly or dominantly (> 50%) by non-soil materials, including permanent water and ice, bare rock and permanently sealed surfaces (urban areas and pavements). No attempt will be made to specify the types or proportions of non-soil materials in a grid cell. Excluded grid cells values of soil properties should be identified as no data (i.e. -9999).

2. Soil Properties

2.1 Depth of Soil

In order to estimate soil properties at specific depth intervals, there is first a need to provide an estimate of the total depth of the soil within each grid cell. The project will estimate the following important depths for each grid cell.

Table 3. Specifications for properties related to reporting depth of soil

No.	Property	Units	Precision[6]	Reference	Description of Method
1	Depth to Rock	cm	N3.0	Soil Survey Division Staff, 1993 Chapter 1 page 5	Depth in cm to a lithic or paralithic contact as defined in USDA Soil Survey Manual. If depth is < 200 cm record actual depth in cm. If depth is > 200 cm record actual depth if known. If not known exactly, record depth as 999 cm
2	Plant Exploitable (Effective) Depth	cm	N3.0	Soil Survey Division Staff, 1993 Chapter 3 page 60	Effective depth in cm as defined in the USDA Soil Survey Manual. The lower limit of soil is normally the lower limit of biologic activity, which generally coincides with the common rooting depth of native perennial plants. This depth is where root penetration is strongly inhibited because of physical (including soil moisture or temperature) and/or chemical characteristics.

2.2 Primary Soil Properties – see Appendix A

GlobalSoilMap will produce estimates of soil property values, their uncertainty and their date of prediction at each of six specified depth increments for the following soil properties.

Definitions and methods of analysis for most of the soil properties are according to ISO standards as defined in FAO (2006) Annex 1: Methods for Soil Analysis (see Appendix D). Particle size distribution is defined according to the USDA Soil Survey Laboratory Methods Manual (Burt, 2004). The USDA definition of particle size classes has been recommended by FAO for use in the Soil Map of the World. Units for properties are reported in g/Kg or cm (instead of % or m) to reduce data storage and transmission costs by storing integer numbers.

[6] The notation used to describe precision (e.g. N3.0) is interpreted as N = number, 3 = length of number, 0 = number of decimal digits. Wherever possible values are reported in integer format to avoid the extra overhead associated with storing and transmitting real numbers.

Table 4. Specifications for primary soil properties

No.	Property	Unit	Precision[4]	Reference	Description of Method
3	Organic Carbon	g/kg	N4.0	ISO 10694	mass fraction of carbon by weight in the < 2 mm soil material as determined by dry combustion at 900° C
4	pHx10		N3.0	ISO 10390	1:5 soil/water (divide by 10 to get correct pH)
5	Clay	g/kg	N3.0	Burt, 2004 Page 347	< 2 µm mass fraction of the < 2 mm soil material determined using the pipette method
6	Silt	g/kg	N3.0	Burt, 2004 Page 347	2-50 µm mass fraction of the < 2 mm soil material determined using the pipette method
7	Sand	g/kg	N3.0	Burt, 2004 Page 347	50 µm - 2 mm mass fraction of the < 2 mm soil material determined using the pipette method
8	Coarse Fragments	$m^3\ m^{-3}$	N3.0	Burt, 2004 page 36	mass fraction of the soil material > 2 mm
9	ECEC	$mmol_c/kg$	N4.0	ISO 11260	Cations extracted using Barium Chloride ($BaCl_2$) plus exchangeable H + Al

2.3 Derived Soil Properties

Where no sufficient measured data exist, properties will be predicted using pedotransfer functions. For instance, the following properties may be predicted using pedo-transfer functions that will be developed and specified by the data provider:

Table 5. Specifications for derived soil properties

No.	Property	Units	Precision[4]	Reference	Description of Method
10	Bulk Density	Mg/m^3	N3.1	ISO 11272	Bulk Density of the whole soil (including coarse fragments) in mass per unit volume by a method equivalent to the core method using a pedotransfer function
11	Bulk Density	Mg/m^3	N3.1	ISO 11272	Bulk Density of the fine earth fraction of the soil (< 2 mm) in mass per unit volume by a method equivalent to the core method using a pedotransfer function
12	Available Water Capacity	mm (total over the depth range)	N4.0	Burt, 2004 Page 137	Available water capacity computed for each of the specified depth increments using a specified pedotransfer function that references the values estimated above for organic carbon, sand, silt, clay and bulk density.

NOTE: AWC = *f* (total carbon, sand, silt, clay, % coarse fragments, bulk density) for the 6 depths. Profile-AWC is AWC summed over the effective depth.

2.4 Additional Soil Properties – see also Appendix C

The soil properties identified above represent the minimum data set agreed upon by the *GlobalSoilMap* consortium. This list in no way restricts individual countries or nodes from producing a longer list of predicted soil properties for their area of interest. For example, the following secondary variable (Table 6) is considered by some nodes to be important, desirable and feasible to predict. These nodes have indicated an intention to predict this additional soil property but it is considered optional, from the point of view of these specifications.

Table 6. Specifications for an additional soil property

No.	Property	Units	Precision[4]	Reference	Description of Method
13	Electrical Conductivity	mS/m	N4.1		Electrical conductivity in 1:1 saturated paste

2.5 Time (Year)

The date of the actual or estimated time of sampling of the legacy soil data will be attached to each of the estimated soil properties at each grid cell. The date reported will reflect the year of publication for a map or the year of analysis for a sampled soil profile.

The maps of soil properties created in the *GlobalSoilMap* project will initially be based on making maximum use of legacy soils data collected and reported over many decades of field work. Data for any point or any map reflect the state of the soil at the time the point was sampled and analysed, or the map was produced. A gridded date map will be made to indicate the date (in years) that the soil property value most closely reflects.[7]

[7] It may be possible, in future Tiers (beyond Tier 2) of the *GlobalSoilMap* products to attempt to reconcile differences in soil property values reported for different times and under different land uses to one or more standardized reference dates (e.g. harmonized decadal values at 1970, 1980, 1990, 2000, 2010, etc.) and under the land use conditions current at each date. This will first require that regional legacy soil data sets be analyzed to detect and quantify directions and rates of change in soil property values under known land use and land management regimes. These regional values for rates of change under different land uses could be applied to the original predictions of soil property values, in combination with information on land use history at each grid cell, to harmonize soil property values to common reference years for each major regional land use type. This is a potential future product and is not part of the current specifications.

3. Uncertainty, Date & Accuracy

An important aspect of the *GlobalSoilMap* project is its estimate and reporting of the uncertainty associated with all soil property predictions.

3.1 Uncertainty Definition

For the purposes of these specifications, uncertainty is defined for each location and depth increment at increasing levels of sophistication as in Table 0. For Tier 1, it is the 90% Prediction Interval (PI) which reports the range of values within which the true value is expected to occur 9 times out of 10 (or 90% of the time). Preferably the PI limits are symmetric, meaning that there is a 5% probability of exceeding the upper limit and a 5% probability of being below the lower limit. If asymmetric intervals are used, then these probabilities must be specified.

3.2 Date Stamping

It would be valuable to provide a raster surface to indicate the age of observational data used in the spatial predictions. The practical reason is to show how old or out of date the data are or can be. This is of clear interest to users, and can prioritise future investments in sampling.

With each prediction location the (integer) year of the data related to the prediction can be given. The method of producing this is not specified but will be described as metadata. Methods might include the year of the nearest data point (or weighted average year of the nearest 8 data points for point-based predictions).

3.3 Validation of GlobalSoilMap predictions of soil property values

For Tier 1 and 2 products for any contiguous region no validation is required (Table 0). For Tier 1 as an addendum, but not a requirement, an appropriate measure for each property at each depth increment may be the mean error and root mean square error of the point predictions. This can be achieved by cross-validation. It is suggested, on average, one observation point per 10 000 square kilometres may be required and at least 50 points are required to obtain an estimate deemed sufficiently reliable.

In subsequent tiers beyond Tier 2 it is anticipated a richer set of validation criteria will be used, including the percentage of the map area that falls within the realised uncertainty limits (Tier 3), and ultimately should meet pre-specified accuracy measures (Tier 4).

It is essential that validation data are truly independent from the calibration data predictions and hence have not been used in any way to help make the prediction maps. Furthermore, it is encouraged that validation data are collected using probability sampling, because this allows to calculate confidence limits associated with the validation measures and to test whether a more elaborate or novel method produces more accurate results than an existing approach.

3.4 Reproducibility of GlobalSoilMap predictions of soil property values

The *GlobalSoilMap* project aims for reproducible research, meaning that it is possible to replicate or reproduce all output products given access to the inputs used to produce them. In

the first instance, for Tier 1 products, reproducibility will be enabled by requiring that for each output value reported for any soil property, associated date and associated uncertainty there will be full documentation of the methods used to produce those values. Documentation of methods used to produce each output will be reported using the standard template described in Appendix A, Table 7.

Future versions of the specifications will be expanded to require contributors to guarantee complete reproducibility by specifying all inputs, processes and outputs and by making all inputs and processes available for access and use by others.

3.5 Policy for release of interim GlobalSoilMap predictions of soil property values

It is desirable to upload and release for general use only productsthat fully meet all specifications by providing predictions of both predictions and associated uncertainties for all soil properties at all depths. However, preliminary or incomplete *GlobalSoilMap* products may be released or made available under the following circumstances.

- Where the data are made available for testing delivery processes or client use (i.e. not a full / official release).
- Where an agreed minimum area (i.e. the node and *GlobalSoilMap* agree on the size and utility of the area covered) has at least some properties available to full specification for those properties including estimates of uncertainty.

References

Brown, J.D., 2004. Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Trans. Inst. British Geographers 29, 367–381.

Calabretta, M. R. 2008. Mapping on the HEALPix grid. arXiv:astro-ph/0412607v1

Cooke, R.M., 1991. Experts in Uncertainty: Opinion and Subjective Probability in Science. Oxford University Press, Oxford.

Cazemier, D.R., Lagacherie, P., and Martin-Clouaire, R. 2001. A possibility theory approach for estimating available water capacity from imprecise information contained in soil databases. Geoderma 103, 113–132.

FAO/IIASA/ISRIC/ISS-CAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria

GlobCover Land Cover v2 2008 database. European Space Agency, European Space Agency GlobCover Project, led by MEDIAS-France. 2008. http://ionia1.esrin.esa.int/index.asp

Heuvelink, G.B.M. and Brown, J. D. 2007. Towards a soil information system for uncertain soil data. Developments in Soil Science, volume 31. P. Lagacherie, A.B. McBratney and M. Voltz (Editors). Elsevier B.V. pp. 97-106.

Heuvelink, G.B.M., Bierkens, M.F.P., 1992. Combining soil maps with interpolations from point observations to predict quantitative soil properties. Geoderma 55, 1–15.

Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modelling. Geoderma 89, 47–65.

Lilburne, L., Hewitt, A. and Ferriss, S., 2009. Progress with the design of a soil uncertainty database, and associated tools for simulating spatial realisations of soil properties. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho. pp. 5510-519.

Ma, T., Zhou, C., Xie, Y., Qin, B., Ou, Y. A discrete square global grid system based on the parallels plane projection. International Journal of Geographical Information Science. 23(10): 1297–1313

NRCS, 2004, Soil Survey Laboratory Methods Manual. Soil Survey Investigations Report No. 42, Version 4.0. Rebecca Burt (editor). United States Department of Agriculture, Natural Resources Conservation Service. November, 2004. 700 pp.

Malone, B. P., McBratney, A. B., and Minasny, B., 2011. Emperical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160, 614-626.

O'Hagan A, Buck C, Daneshkhah A, Eiser J, Garthwaite P, Jenkinson D, Oakley J, Rakow T, 2006. Uncertain Judgements : Eliciting Experts' Probabilities (John Wiley & Sons, Ltd., Chichester).

Seong, Jeong Chang , Mulcahy, Karen A. and Usery, E. Lynn(2002) 'The Sinusoidal Projection: A New Importance in Relation to Global Image Data', The Professional Geographer, 54: 2, 218-225.

Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

van der Keur, P. and Iversen, B. V. 2006. Uncertainty in soil physical data at river basin scale. Hydrol. Earth Syst. Sci. Discuss., 3, 1281–1313.

Appendix A: Minimum data set for each 100 m grid cell

Table 7 below illustrates the information that must be provided for each grid cell and suggests a possible structure for providing that information in a data base format. This structure is not part of the current *GlobalSoilMap* specifications. A formal process is underway to develop a UML model for describing the delivery products of the *GlobalSoilMap* project. This UML model will be implemented in XML once it is defined, in order to facilitate data exchange and inter-operability. It is assumed that all nodes will submit data that has been harmonized to the methods of analysis specified for each of the soil properties in these specifications (e.g. pH, organic carbon, sand, silt, clay, gravel, etc.)

Table 7. Illustration of the minimum data required for each grid cell

Long (X)	Lat (Y)	Tile_ID	1 Deg Cell ID	AttributeAnalyzed	yearOfAnalysis	Top_Depth	Bottom_Depth
				For Example			
				pH			
				organic carbon			
				sand			
				silt			
				clay			
				gravel			
				BD			
				AWC			
				EC			
				ECEC			
				Profile depth			
				Effective Depth			
				Spline Smoothing Value			

AttributeValue	LowerUncertainty (5%)	Upper Uncertainty (95%)	Prediction Model	Pedo transfer Function

Appendix B: Some possible approaches for computing weighted mean value by grid cell.

This Appendix presents and discusses some alternative approaches for computing or assigning a bulked mean value to each grid cell for each soil property of interest. It is necessary to be more specific about how to calculate a bulked mean value for each soil property at each of the 6 depths within each square grid cell.

Below are some ideas or options to consider.

1. If using map based estimates, the bulked mean value for the soil property at a given depth for all soils listed for the polygon can be assumed to represent an areal average already and so should satisfy the requirement that the value for the cell represent a bulked area average.
2. Work at a grid resolution that is finer than the final reporting resolution (25, 30 or 50 m grid cells) and produce estimates of the soil property value at each depth for each of the finer resolution grid cells. Then compute an average value for the 100 m x 100 m grid cell as the mean of all values for the finer resolution cells.
3. Produce point-centred estimates for the centre of each grid cell at the working resolution (100 m or 90 m) and then compute a bulked mean value for each grid cell as the average value within a 3x3 or 5x5 window centred on each grid cell. This way the bulked area average reflects the average value within a larger window centred at each grid cell.
4. Don't worry about it and just assume that any point centred prediction represents the bulked mean value for the entire grid cell.

Reporting a bulked mean value has the advantage of removing the short range variability in the value of a soil property within the extent of a grid cell. The uncertainty associated with estimation of a mean value for each reference depth within the full extent of a grid cell will also be lower than the uncertainty associated with estimation of a single value for each depth at a single point at the centre of a grid cell. The values reported for each grid cell should therefore be an estimate of the mean value of that property at each of the six specified depths within the extent of the cell occupied by soil materials (excluding non-soils).

It is necessary to be aware of, and to clearly acknowledge, that in some instances of strongly contrasting soils the reported bulked average value may not exist at any single physical location within the grid cell. Consider the case of a cell that is 50% organic peat soils and 50% sandy upland soils with no or very low organic matter content. The bulked mean value for organic carbon for the cell would represent a mean value between the high value for the peat soil and the low value for the sandy soil. This value is not likely to occur anywhere within the grid cell but it is representative of the mean value within the grid cell. It will be necessary to live with this dichotomy and acknowledge it.

Appendix C: Correlations of soil properties derived from different soil analytical methods

This Appendix identifies and discusses the need for pedotransfer functions to convert soil property values from their original method of analysis to the standard *GlobalSoilMap* reference method of analysis. For discussion purposes, examples are provided to illustrate conversion of data from several widely used non-reference methods into the specified reference methods.

C1.0. Rationale

A well-known issue with using legacy soils data is the inconsistency that arises from use of many different methods for analysing soils in the laboratory or describing them in the field. These different methods yield different values that are not exactly equivalent or comparable. This creates a need to harmonize values produced using different methods in order to make them roughly equivalent and comparable. Harmonization can be challenging.

In order to make use of legacy soils data in the *GobalSoilMap* project, it will be necessary to convert measurements made using different laboratory methods into an equivalent value in the specified standard reference method. For example, values reported for organic carbon determined by non-reference methods will need to be converted into equivalent values in the reference method of dry combustion. Similarly, values for pH in 1:1 or 1:2 water will need to be converted the equivalent value in the standard reference method of pH in 1:5 water. Harmonization of values reported for sand, silt and clay computed using methods of textural analysis that use definitions for particle size fractions different from the reference method will also have to be converted to the standard particle size definitions adopted for these specifications.

Default pedotransfer functions could potentially be identified for each of the methods of analysis for each of the soil properties selected for inclusion in the project. However, locally specific pedotransfer functions have consistently proven to be more effective than global ones and there is widespread agreement that there is generally no universal equation for converting from one method to another in all instances (Konen et al., 2002; Meersmans et al., 2009; Jankauskas et al., 2006; Jolivet et al., 1998; de Vos et al., 2007).

Consequently, there will be a need to develop locally relevant pedotransfer functions at the node level that apply to restricted soil-landscape domains. Examples of conversion of values from non-reference to reference methods are presented below for the primary soil properties of organic carbon, pH, sand, silt and clay.

C1.1. Organic Carbon

The standard reference method for reporting soil organic carbon for the *GlobalSoilMap* project is by dry combustion to at least 900°C (ISO 10694). Values of organic carbon will be reported in g/Kg with integer precision (N4.0) Because of its accuracy and completeness, the dry combustion method (Leco at 1000°C) has been used in many studies as a reference method against which to calibrate other methods (Grewal et al., 1991;. Meersmans et al., 2009)

The dry combustion method is based on thermal oxidation of the OC and thermal decomposition of IC minerals by means of a furnace. It is a rapid, reliable method for the determination of the OC when IC is removed prior to combustion. In fact, dry combustion is considered to ensure oxidation of all OC so it is considered the most accurate method. It can be used as a reference to calibrate other methods against it (Biscutti et al., 2004).

In the dry combustion method, the carbon present in the soil is oxidised to carbon dioxide (CO_2) by heating the soil to at least 900°C in a flow of oxygen-containing gas that is free from carbon dioxide. The amount of carbon dioxide released is then measured by titrimetry, gravimetry, conductometry, gas chromatography or using an infrared detection method, depending on the apparatus used. When the soil is heated to a temperature of at least 900 °C, in addition to organic carbon any inorganic carbon present as carbonate is also completely decomposed.

Total organic carbon can be determined directly or indirectly. Direct determination consists of previous removal of any carbonates present by treating the soil with hydrochloric acid. Indirect determination consists of a correction of the total carbon content for the carbonates present.

Examples of studies that have used dry combustion for calibrating other methods of analyzing organic carbon include Bisutti et al., 2004; Byre and Slaton, 2003; de Vos et al., 2007; Grewal et al., 1991; Kalembasa and Jenkinson, 1973; Jankauskas et al., 2006; Jolivet et al., 1998; Konen et al., 2002; Meersmans et al., 2009; Mikhailova et al., 2003; Sleutel et al., 2007; Soon and Abboud, 1991 and Wang et al, 1996.

A review of several studies (Table 9) illustrates that it is possible to produce regression equations to permit conversion of results produced by one method into equivalent values in a specified reference method (usually dry combustion). However, the studies also highlight the fact that local calibration equations that reflect differences in soils on a regional basis are usually needed.

It has not proven possible to provide a single universal equation to convert organic carbon values produced using other methods of analysis to equivalent values in the reference method of dry combustion. Each node will need to develop and apply node-specific conversions.

Table 8. Regression equations for harmonizing values of organic carbon to a reference standard

No.	Target Method Y =	Source Method X	* Slope	+ Intercept	R2	Reference
1	Dry Combustion	Spectro-photonic	0.9800	0.0000	0.98	Soon and Abboud (1991)
2	Dry Combustion	Walkley-Black	1.0500	0.0000	0.98	Soon and Abboud (1991)
3	Dry Combustion	modified Tinsley	1.0400	0.0000	0.98	Soon and Abboud (1991)
4	Dry Combustion	modified Mebius	1.4000	0.0000	0.99	Soon and Abboud (1991)
5	Dry Combustion	Loss on Ignition (LOI)	0.6330	-9.3600	0.98	Soon and Abboud (1991)
6	Tinsley (1950)	LOI at 850 C	0.4620	-1.3600	0.99	Ball, 1964
7	Tinsley (1950)	LOI at 850 C	0.4600	-1.8700	0.99	Ball, 1964
8	Tinsley (1950)	LOI at 375 C	0.4580	-0.4000	0.99	Ball, 1964
9	DC (Leico at 875 C)	LOI at 360 C MLRA 65NE	1.1414	-0.6791	0.94	Konen et al., 2002
10	DC (Leico at 875 C)	LOI at 360 C MLRA 75NE	0.0672	-4.5359	0.94	Konen et al., 2002
11	DC (Leico at 875 C)	LOI at 360 C MLRA 95B	0.5743	0.1025	0.98	Konen et al., 2002
12	DC (Leico at 875 C)	LOI at 360 C MLRA 103 IA	0.6824	-2.8696	0.97	Konen et al., 2002
13	DC (Leico at 875 C)	LOI at 360 C MLRA 108 IL	0.6094	0.1949	0.98	Konen et al., 2002
14	DC (Dumas at 1000)	Walkley-Black	1.2500	0.1260	0.99	Grewal et al., 1991
15	LOI at 550	DC (Dumas at 1000)	1.6700	2.5100	0.76	Grewal et al., 1991
16	LOI at 550	LOI at 450	0.9970	0.5000	0.98	Grewal et al., 1991
18	DC (at 680 C)	Wet combustion	0.9920	0.0000		Kalembasa & Jenkinson, 1973
19	DC (at 680 C)	Tinsley I	0.9500	0.0000		Kalembasa & Jenkinson, 1973
20	DC (at 680 C)	Tinsley II	0.9530	0.0000		Kalembasa & Jenkinson, 1973
21	DC (at 680 C)	Tinsley III	0.9680	0.0000		Kalembasa & Jenkinson, 1973
22	DC (at 680 C)	Anne	0.9330	0.0000		Kalembasa & Jenkinson, 1973
23	DC (at 680 C)	Mebius	0.9530	0.4300		Kalembasa & Jenkinson, 1973
24	DC (at 680 C)	Walkley-Black	0.7690	-0.0800		Kalembasa & Jenkinson, 1973
25	DC (at 680 C)	Tyurin	0.9330	0.0000		Kalembasa & Jenkinson, 1973
26	DC (Leico CNS 2000)	Walkley-Black	1.3350	0.5730	0.88	Mikhailova et al., 2003
27	DC Robo-prep	Walkley-Black	1.4490	0.4110	0.90	Mikhailova et al., 2003
28	DC (Leico at 1000 C)	Walkley-Black (classic)	1.4700	0.0000	0.84	Meersmans et al., 2009
29	DC (Leico at 1000 C)	Walkley-Black (modified)	1.2000	0.0000	0.87	Meersmans et al., 2009
30	Walkley-Black (mod)	Walkley-Black (classic)	0.8200	0.6800	0.53	Brye and Slaton, 2003
31	DC (Leico at 1000 C)	DC (Carlo–Erba at 1020 C)	1.1300	-0.0600	0.99	Brye and Slaton, 2003
32	Walkley-Black (modified)	DC (Leico at 1000 C)	0.7200	0.6300	0.73	Brye and Slaton, 2003
33	Walkley-Black (modified)	DC (Carlo–Erba at 1020 C)	0.8100	0.5800	0.73	Brye and Slaton, 2003
34	Walkley-Black (classic)	DC (Leico at 1000 C)	0.8900	-0.0900	0.99	Brye and Slaton, 2003
35	Walkley-Black (classic)	DC (Carlo–Erba at 1020 C)	1.0200	0.1500	0.99	Brye and Slaton, 2003
36	Walkley-Black (classic)	LOI at 360	0.4300	-0.0900	0.88	Brye and Slaton, 2003
37	Walkley-Black (modified)	LOI at 360	0.3400	0.6300	0.44	Brye and Slaton, 2003
38	DC (Carlo–Erba at 1020 C)	LOI at 360	0.4300	0.6500	0.98	Brye and Slaton, 2003
39	DC (Leico at 1000 C)	LOI at 360	0.4800	-0.0030	0.89	Brye and Slaton, 2003

Table 9. Regression equations for harmonizing values of organic carbon to a reference standard

No.	Target Method Y =	Source Method X	* Slope	+ Intercept	R2	Reference
40	A-I colorimetric	Walkley-Black (classic)	0.5410	-0.0330	0.96	Chacón et al., 2002
41	A-I colorimetric	Walkley-Black (classic)	0.4590	-0.0580	0.94	Chacón et al., 2002
42	A-I colorimetric	Walkley-Black (classic)	0.4920	0.0000	0.99	Chacón et al., 2002
43	DC (Shimadzu at 900 C)	Walkley-Black not corrected	1.5800	0.0000	0.96	De Vos et al., 2007
44	DC (Shimadzu at 900 C)	Walkley-Black corrected	1.2000	0.0000	0.96	De Vos et al., 2007
45	LOI at 375 (Lab K)	DC (Vario EL at 1150 C)	1.2530	0.5030	0.87	Jankauskas et al., 2006
46	LOI at 375 (Lab W)	DC (Vario EL at 1150 C)	1.2790	0.2380	0.89	Jankauskas et al., 2006
47	Walkley-Black NRCS 1995	DC (Vario EL at 1150 C)	1.0200	0.1680	0.97	Jankauskas et al., 2006
48	Tyurin photometric	DC (Vario EL at 1150 C)	0.8700	0.3690	0.98	Jankauskas et al., 2006
49	Tyurin titrametric classic	DC (Vario EL at 1150 C)	0.8690	0.1620	0.91	Jankauskas et al., 2006
50	LOI at 375 (Lab W)	LOI at 375 (Lab K)	0.8750	0.1500	0.88	Jankauskas et al., 2006
51	Walkley-Black NRCS 1995	LOI at 375 (Lab K)	0.6100	0.3570	0.83	Jankauskas et al., 2006
52	Tyurin photometric	LOI at 375 (Lab K)	0.5220	0.5250	0.84	Jankauskas et al., 2006
53	Tyurin titrametric classic	LOI at 375 (Lab K)	0.5280	0.1390	0.87	Jankauskas et al., 2006
54	Walkley-Black NRCS 1995	LOI at 375 (Lab W)	0.6350	0.4200	0.86	Jankauskas et al., 2006
55	Tyurin photometric	LOI at 375 (Lab W)	0.5510	0.5570	0.89	Jankauskas et al., 2006
56	Tyurin titrametric classic	LOI at 375 (Lab W)	0.5670	0.3060	0.85	Jankauskas et al., 2006
57	Tyurin photometric	Walkley-Black NRCS 1995	0.8130	0.3110	0.97	Jankauskas et al., 2006
58	Tyurin titrametric classic	Walkley-Black NRCS 1995	0.8240	0.0810	0.91	Jankauskas et al., 2006
59	Tyurin titrametric classic	Tyurin photometric	0.9540	-0.1120	0.89	Jankauskas et al., 2006
60	Walkley-Black NRCS 1995	DC (Leico at 875 C)	0.9180	1.0000	0.99	Jolivet et al., 1998
61	Walkley-Black NRCS 1995	DC (Leico at 875 C)	0.9470	0.0000	0.99	Jolivet et al., 1998
62	DC (Leico at 875 C)	LOI at 550 C	0.6130	0.6000	0.99	Jolivet et al., 1998
63	DC (Leico at 875 C)	LOI at 550 C	0.6240	0.0000	0.99	Jolivet et al., 1998
64	DC (Shimadzu at 900 C)	Walkley-Black NRCS 1995	1.5060	0.0000	0.99	Lettens et al., 2007
65	DC (Shimadzu at 900 C)	Walkley-Black NRCS 1995	1.5940	0.0000	0.99	Lettens et al., 2007
66	DC (Shimadzu at 900 C)	Walkley-Black NRCS 1995	1.7740	0.0000	0.98	Lettens et al., 2007
67	Walkley-Black 6A1	DC (Leico at 1000 C)	0.9700	0.0000	0.99	Wang et al., 1996
68	DC (Leico at 1000 C)	LOI at 375 C siltstone	0.7320	-1.6100	0.95	Wang et al., 1996
69	DC (Leico at 1000 C)	LOI at 375 C sandstone	0.5620	-0.9950	0.95	Wang et al., 1996
70	DC (Leico at 1000 C)	LOI at 375 C basalt	0.4690	-0.9410	0.95	Wang et al., 1996
71	DC (Leico at 1000 C)	LOI at 375 C combined	0.7260	-1.5980	0.96	Wang et al., 1996
72	DC (Leico at 1000 C)	LOI at 375 C basalt	0.4690	-0.9410	0.95	Wang et al., 1996
73	Walkley-Black 6A1	DC (Leico at 1000 C) other	0.7390	-1.7590	0.95	Wang et al., 1996
74	Walkley-Black 6A1	DC (Leico at 1000 C) basalt	0.4520	-0.8910	0.95	Wang et al., 1996
75	LOI at 375 C basalt	DC (Leico at 1000 C)	0.4692	-0.9410	0.95	Wang et al., 1996
76	Walkley-Black 6A1	LOI at 375 C combined	0.4880	-2.3360	0.91	Wang et al., 1996
77	Walkley-Black 1934	DC (Variomax CNS)	1.0340	0.0160	0.99	Sleutel et al., 2007
78	Walkley-Black 1934	DC (Variomax CNS)	1.0130	0.0000	0.99	Sleutel et al., 2007
79	Springer-Klee, 1954	DC (Variomax CNS)	1.0020	0.0000	0.98	Sleutel et al., 2007
80	DC (Shimadzu at 900 C)	DC (Variomax CNS)	0.9430	0.0000	0.99	Sleutel et al., 2007

C1.2. pH

As a single measurement, pH describes more than relative acidity or alkalinity. It also provides information on nutrient availability, metal dissolution chemistry, and the activity of microorganisms (Miller and Kissel, 2010).

The standard reference method for reporting pH for the *GlobalSoilMap* project is ISO 10390:2005.

This standard specifies an instrumental method for the routine determination of pH using a glass electrode in a 1:5 (volume fraction) suspension of soil in water (pH in H2O), in 1 mol/l potassium chloride solution (pH in KCl) or in 0.01 mol L^{-1} calcium chloride solution (pH in $CaCl_2$).

Values for pH for the *GlobalSoilMap* project will be reported for a 1:5 suspension of soil in water. Values will be reported in byte format as pH x 10 with a precision of (N3.0) (value range of 0-149). These values will need to be divided by 10 to produce a correct pH value with a precision of 1 decimal place.

ISO 10390:2005 is applicable to all types of air-dried soil samples, for example pre-treated in accordance with ISO 11464. The most common method for analyzing pH in North America is a 1:1 soil/water suspension (Miller and Kissel, 2010). Adopting ISO 10390:2005 as a standard with its specification of pH measured in a 1:5 suspension of soil in water will require many values to be converted from 1:1 soil/water to 1:5 soil/ water equivalent values.

The ratio of soil to water in a suspension has a net effect of increasing the pH with a decrease in the soil/water ratio. Keaton (1938) and Davis (1943) have shown that decreasing the soil/water ratio from 10:1 to 1:10 resulted in an increase of 0.40 pH units. Values for pH computed using methods with a lower ratio of soil to water (e.g. 1:1 or 1:2.5) will generally be lower than equivalent values for pH in 1:5 $CaCl_2$ solution and will need to be adjusted higher. Several authors have demonstrated that fitting quadratic or curvilinear functions to soil pH data produces regression equations with higher coefficients of determination that those obtained from a linear fit (Aitken and Moody, 1991; Miller and Kissel, 2010).

Soil pH varies with season and soil moisture content with higher pH values associated with wetter soils and winter conditions and lower pH values with drier soils and summer conditions (Miller and Kissel, 2010). The effects of both temporal variation in pH and variation due to different methods means that small differences in pH may not be meaningful in the context of predictions made for the *GlobalSoilMap* project using legacy soils data.

Table 10. Regression equations for converting values of pH between different methods

No.	Target Method (Y)	Source Method (X)	Equation	R2	Reference
1	pH (1:1 0.01 m CaCl2)	pH (1:1 water)	$y = 1.08(x) - 0.973$	0.98	Miller and Kissel (2010)
2	pH (1:1 0.01 m CaCl2)	pH (saturated paste)	$y = 1.10 (x) - 0.923$	0.98	Miller and Kissel (2010)
3	pH (1:1 0.01 m CaCl2)	pH (1:2 water)	$y = 1.05 (x) - 0.950$	0.97	Miller and Kissel (2010)
4	pH (1:1 water)	pH (1:1 0.01 m CaCl2)	$y = x + 0.267 (EC\ 1:1\ water)^{-0.445}$	0.99	Miller and Kissel (2010)
5	pH (1:2 water)	pH (1:1 0.01 m CaCl2)	$y = x + 0.239 (EC\ 1:1\ water)^{-0.505}$	0.98	Miller and Kissel (2010)
6	pH (1:5 0.01 m CaCl2)	pH (1:5 water)	$y = 1.012 (x) - 0.76$	0.99	Conyers and Davey (1988)
7	pH (1:5 0.01 m CaCl2)	pH (1:5 water)	$y = 0.979 (x) - 0.71$	0.68	Bruce et al., (1989)
8	pH (1:5 0.01 m CaCl2)	pH (1:5 water)	$y = 0.887 (x) - 0.199$	0.88	Aitken and Moody (1991)
9	pH (1:5 0.01 m CaCl2)	pH (1:5 water)	$y = 0.197 (x)^2 - 1.21 (x) + 5.78$	0.92	Aitken and Moody (1991)
10	pH (1:5 0.002 m CaCl2)	pH (1:5 water)	$y = 0.948 (x) - 0.308$	0.90	Aitken and Moody (1991)
11	pH (1:5 0.002 m CaCl2)	pH (1:5 water)	$y = 0.178 (x)^2 - 1.043 (x) + 5.10$	0.94	Aitken and Moody (1991)
12	pH (1:5 1 m KCl)	pH (1:5 water)	$y = 0.803 (x) + 0.077$	0.81	Aitken and Moody (1991)
13	pH (1:5 1 m KCl)	pH (1:5 water)	$y = 0.233 (x)^2 - 1.797 (x) + 7.143$	0.98	Aitken and Moody (1991)
14	pH (soil solution)	pH (1:5 water)	$y = 1.28 (x) - 0.613$	0.78	Aitken and Moody (1991)
15	pH (soil solution)	pH (1:5 0.01 m CaCl2)	$y = 1.105 (x) - 0.140$	0.79	Aitken and Moody (1991)
16	pH (soil solution)	pH (1:5 0.002 m CaCl2)	$y = 1.050 (x) - 0.112$	0.80	Aitken and Moody (1991)
18	pH (soil solution)	pH (1:5 1 m KCl)	$y = 1.175 (x) - 0.262$	0.80	Aitken and Moody (1991)

C1.3. Particle Size Distribution (sand, silt and clay)

Soil texture represents the relative composition of sand, silt, and clay in soil. The particle-size distribution is usually represented in a texture diagram, relating the percentages of sand, silt, and clay to a texture class (Minasny and McBratney, 2001). The standard reference method adopted by the *GlobalSoilMap* project for reporting particle size classes of sand, silt and clay (g/Kg), is as per the USDA Soil Survey Laboratory Methods Manual (3A1a) (Burt, 2004 page 34). The Kilmer and Alexander (1949) pipet method was chosen by the USDA Soil Conservation Service because it is reproducible in a wide range of soils.

The current standard for particle size classes adopted by FAO for use in the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISSCAS/JRC, 2009) is ISO 10390:2005. This standard differs from the USDA definition in defining the size range for silt as 2-63 µm instead of 2-50 µm and sand as 63-2000 µm instead of 50-2000 µm. This is a relatively new standard for FAO which previously adopted the USDA definitions for the digital soil map of the world (FAO, 1990).

Differences in values reported for soil particle size fractions can arise because of differences in method of analysis (e.g. hydrometer, pipette, laser diffraction) or differences classification of particle size fractions. Most literature on harmonization of soil texture data deals with harmonizing differences in reported particle size fractions (Figure 2).

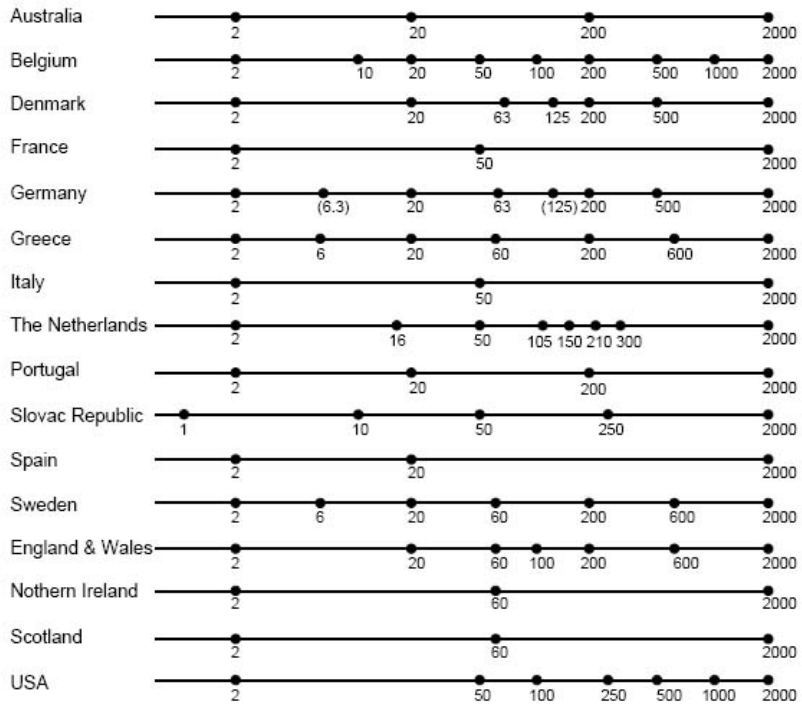


Figure 2. Particle size limits used in European countries, Australia and America (Adapted from Nemes *et al.*, 1999a and Minasny and McBratney, 2001)

Minasny and McBratney (2001) identified two major textural classifications in the world as the International and USDA/FAO systems (Table 11). The significant difference between these two was the choice of a threshold value for differentiating silt from clay of 20 µm for the International and 50 µm for the USDA. The new ISO/FAO standard adds an additional difference by changing the threshold value between silt and sand from 50 µm to 63 µm. This is a relatively minor difference but it still needs to be addressed.

Table 11. Differences between the International, USDA and ISO/FAO particle size classifications

Size Fraction	International	USDA	ISO/FAO
clay	< 2 µm	< 2 µm	< 2 µm
silt	2 - 20 µm	2 - 50 µm	2 - 63 µm
sand	20-2000 µm	50-2000 µm	63-2000 µm

Both Minasny and McBratney (2001) and Nemes et al., (1999a) investigated options for harmonizing values for sand, silt and clay reported using different systems for classifying particle size fractions.

Using a compilation of four large databases consisting of a total of 1620 samples, Minasny and McBratney (2001) developed a single multiple linear regression model for converting between silt fraction based on the international standard of 2-20 µm ($P_{2\text{-}20}$) to the 2-50 µm range of the USDA standard ($P_{2\text{-}50}$) and vice versa. The equations are as follows:

$$P_{2\text{-}50} = -18.3914 + 2.0971\,(P_{2\text{-}20}) + 0.6726\,(P_{20\text{-}2000}) - 0.0142\,(P_{2\text{-}20})^2 - 0.0049\,(P_{20\text{-}2000})^2$$

($R^2 = 0.823$)

If $P_{2\text{-}50} < 0$ then $P_{2\text{-}50} = 0.8289\,(P_{2\text{-}20}) + 0.0198\,(P_{20\text{-}2000})$

and

$$P_{2\text{-}20} = -0.4070 - 0.1271\,(P_{<2}) + 0.5527\,(P_{2\text{-}50}) + 0.0017\,(P_{<2})^2 - 0.0019\,(P_{2\text{-}50})^2 + 0.0059\,(P_{<2})\,(P_{2\text{-}50})$$

($R^2 = 0.818$)

If $P_{2\text{-}20} < 0$ then $P_{2\text{-}20} = 0.1147\,(P_{<2}) + 0.2212\,(P_{2\text{-}50})$

Minasny and McBratney (2001) argued that most countries should consider adopting the particle size limits and texture classes of the USDA system. They noted that the 2 - 50 µm particle size range is usually more useful than the 2 - 20 µm range for estimating water retention in pedo transfer functions and observed that translations from one system into another were relatively easy, given improved computing power and algorithms.

There is already a package in R that supports conversion of particle size data reported in one system of classification to values in any specified other system. This package, provided by Julien Moeys with contributions by Wei Shangguan, applies a log-linear transformation of soil texture data from one particle size system into another (Moeys, 2010). Two modules exist, one that only accepts three data values as input (TT.text.transf) and the other that can translate any number of values for any number of size fractions (TT.text.transf.X). Log linear transformations have been shown to be the least reliable method for converting between different particle size classifications (Minasny and McBratney, 2001; Nemes, 1999a) but the simple fact that routines

already exist in R to support rapid and efficient conversion from different systems into the USDA reference standard is encouraging. The *GlobalSoilMap* project will look at extending the functionality of this R package provided by Moeys (2010) to include additional options for converting between particle size classification systems.

The *GlobalSoilMap* project will develop an extended library of R functions for converting from systems of particle size classification different from the USDA to the standard particle size classes of the USDA system (clay = < 2 µm, silt = 2-50 µm and sand = 50-2000 µm). We will investigate and implement three main options of a) the spline and similarity methods of Nemes *et al*., (1999a,b) b) the regression equations of Minasny and McBratney (2001), and c) the graphical PSD conversion nomograms of Shirazi *et al*., 2001.

C1.4. Bulk Density

The standard reference method for reporting bulk density for the *GlobalSoilMap* project is the core method (ISO 11272).

The dry bulk density (BD) is the ratio between the mass of oven dry soil material and the volume of the undisturbed fresh sample. The ISO standard defines dry bulk density as the ratio of the oven-dry mass of the solids to the volume (the bulk volume includes the volume of the solids and of the pore space) of the soil.

The recommended ISO method (core method) uses steel cylinders of known volume (100 mL, 400 mL) that are driven in the soil vertically or horizontally by percussion. Sampling large volumes results in smaller relative errors but requires heavy equipment. The method cannot be used if stones or large roots are present or when the soil is too dry or too hard.

For soils with a high stone or root content or when the soil is too dry or too hard, methods based on the excavation technique are used as an alternative to the core method. In the excavation method a hole on a horizontal surface is dug and then filled with a material with a known density (e.g. sand which packs to a calibrated volume or water separated from the soil material by an elastic membrane). The soil obtained from the hole, is dried to remove the water and the dry mass is weighed.

The volumetric percentage of the coarse fragments needs to be determined in order to calculate the bulk density of the fine earth.

Experience has shown that organic carbon (OC) and texture predominately determine soil bulk density. Organic carbon and texture information is often available in soil survey campaigns. Therefore many attempts have been made to estimate soil bulk densities through some pedo-transfer functions (PTFs) based on soil OC and texture data (Curtis and Post 1964; Adams 1973; Alexander 1980; Federer 1983; Rawls 1983; Huntington et al. 1989; Manrique and Jones 1991; Bernoux et al. 1998; Tomasella and Hodnett 1998).

Heuscher *et al*., (2007) applied a stepwise multiple regression procedure to predict oven-dried bulk density from soil properties using the NRCS National Soil Survey Characterization Data. The database included both subsoil and topsoil samples. An overall regression equation for predicting oven-dried bulk density from soil properties (R^2 = 0.45, P < 0.001) was developed using almost 47,000 soil samples. Partitioning the database by soil suborders improved regression relationships (R^2 =0.62, P < 0.001). Of the soil properties considered, the stepwise multiple regression indicated that organic C content was the strongest contributor to bulk density prediction. Other significant variables included clay content, water content and to a lesser extent, silt content, and depth.

Tranter et al., 2007 proposed a conceptual model that incorporated *a priori* knowledge for predicting soil bulk density from other more regularly measured properties. The model considers soil bulk density to be a function of soil mineral packing structures (ρ_m) and soil structure ($\Delta\rho$). Bulk-density maxima were found for soils with approximately 80% sand. Bulk densities were also observed to increase with depth, suggesting the influence of over-burden pressure. Residuals from the ρ_m model, referred to as $\Delta\rho$, correlated with organic carbon.

Torri et al., (2007) developed a nomogram for transforming rock fragment content from a by-mass to a by-volume basis and vice versa. This nomogram facilitates comparison of data on rock fragment content expressed in different units.

Most PTFs for predicting bulk density, except those developed by Rawls (1983), Tomasella and Hodnett (1998), and Bernoux et al. (1998), are a function only of organic matter (OM)/OC content. Although studies conducted by Saini (1966) and Jeffrey (1970) have shown that OM has a dominating effect on soil bulk density and that it can be used alone as a good predictor of soil bulk density, it has been observed (e.g. Alexander 1980; Huntington et al. 1989; Manrique and Jones 1991) that soil texture plays a major role in controlling bulk density where OM is a minor component.

McBratney *et al*., (2002) proposed the concept of a soil inference system (SINFERS) that incorporated both expert soil knowledge and statistical prediction equations. The proposed system was intended to implement two major functions, namely:

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

The SINFER approach proposed by McBratney *et al*., (2002) will be the basis for efforts to create and apply PTFs for predicting soil bulk density for the *GlobalSoilMap* project.

C1.5. Available Water Capacity

The standard reference method adopted by the *GlobalSoilMap* project for reporting available water capacity is as per the USDA Soil Survey Laboratory Methods Manual (3D5a) (Burt, 2004 page 137).

Calculation of the water retention difference (WRD) is considered the initial step in the approximation of the available water capacity (AWC). WRD is a calculated value that denotes the volume fraction for water in the whole soil that is retained between 1500-kPa suction and an upper limit of usually 33 or 10-kPa suction (Burt, 2004 page 137). The upper limit (lower suction) is selected so that the volume of water retained approximates the volume of water held at field capacity. The 10-, 33- and 1500-kPa gravimetric water contents are then converted to a whole soil volume basis by multiplying by the bulk density (Db33) and adjusting downward for the volume fraction of rock fragments, if present in the soil. The lower suctions, e.g., 10 or 5-kPa, are used for coarse materials.

Results of research to develop hydraulic PTFs have been reported widely, including in the USA (Rawls et al., 1982), the UK (Mayr and Jarvis, 1999), the Netherlands (Wösten et al., 1995), and Germany (Scheinost et al., 1997b)." This research has attempted to correlate particle size distribution, bulk density and organic matter content with water content at field capacity (FC, θ at -33 kPa), permanent wilting point (PWP, θ at -1500 kPa), and available water content (AWC = FC - PWP) (Minasny, 2007). Other examples include studies by Nielsen and Shaw (1958), Burrows and Kirkham (1958), Slater and Williams (1965a, 1965b, 1966, 1967, 1969), Hall et al., (1977)Gupta and Larson (1979) Clapp and Hornberger (1978) and Bloemen (1980).

Gijsman *et al*., (2007) reported that many PTFs for estimating soil hydraulic properties have been published (see overviews by Rawls et al. (1991), Timlin et al. (1996) and Wösten et al. (2001). Timlin et al. (1996) reported 49 methods and estimated that this covers only about 30% of the total. Gijsman et al. (2002) compared eight methods for all the soil classes that make up the texture triangle. They went through the triangle in steps of 1% sand, 1% silt and 1% clay and determined the estimated values of wilting point or lower limit of plant extractable water (LL), field capacity, also referred to as the drained upper limit (DUL) and soil saturation (SAT) . Gijsman et al. (2002) concluded that none of the methods were universally good. The best method in the comparison of Gijsman et al. (2002) was Saxton et al. (1986), closely followed by Rawls et al. (1982).

Jagtap et al. (2004) developed an approach that does not fit a mathematical equation through the data, but rather compares the soil layer for which the key soil water contents of LL, DUL and SAT have to be estimated with all layers in a database of field-measured soil–water-retention data. The layer that is most similar in texture and organic carbon concentration is considered to be the 'nearest neighbor' among all the layers in the database and its soil–water-retention values are assumed to be similar to those that need to be estimated. To avoid making estimated soil–water-retention values dependent on only one soil in the database, the six 'nearest neighbors' are used and weighted according to their degree of similarity (Jagtap et al., 2004). This is a non-parametric procedure, in the sense that it does not assume a fixed mathematical relationship between the physical properties and the water holding properties of

soils. The similarity method to convert soil particle size fraction data proposed by Nemes *et al*. (1999a,b) is a direct analogue of this similarity method of Jagtap et al., (2004).

Zacharias and Wessolek (2007) identified three different approaches for deriving the WRC from more easily available parameters as:

1. Point-based estimation methods: estimating the water content of selected matric potentials from predictors such as the percentage of sand, silt, or clay, the amount of organic matter, or the bulk density (e.g., Gupta and Larson, 1979; Rawls and Brakensiek, 1982).
2. Semiphysical approach: deriving the WRC from information on the cumulative particle size distribution (Arya and Paris, 1981); theoretically, this approach is based on the similarity between cumulative particle size distribution and water retention curves. The water contents are derived from the soil's predicted pore volume and the hydraulic potentials are derived from capillarity relationships.
3. Parameter estimation methods: using multiple regression to derive the parameters of an analytical closed-form equation for describing the WRC, using predictors such as the percentage of sand, silt, and clay, the amount of organic matter, or the bulk density (e.g., Vereecken et al., 1989; Wösten et al., 1999).

Zacharias and Wessolek (2007) concluded that approach 1 has the disadvantage that it uses a large number of regression parameters depending on the number of WRC sampling points, which makes its use in the mathematical modelling more difficult while for approach 2 very detailed information about the particle size distribution is required. They therefore preferred use of the parameter estimation methods.

Zacharias and Wessolek (2007) observed that pedotransfer functions that do not use the OM are rare and gave the following examples. Hall et al. (1977) developed point-based regression equations using soil texture and bulk density (only for subsoils) for British soils. Oosterveld and Chang (1980) developed an exponential regression equation for Canadian soils for fitting the relationship between clay and sand content, depth of soil, and moisture content. Equations to estimate the WRC from mean particle diameter and bulk density have been proposed by Campbell and Shiozawa (1989). Williams et al. (1992) analyzed Australian data sets and developed regression equations for the relationship between soil moisture and soil texture, structure information, and bulk density including variants for both the case where there is available information on OM and where the OM is unknown. Rawls and Brakensiek (1989) reported regression equations to estimate soil water retention as a function of soil texture and bulk density. Canarache (1993) developed point based regression equations using clay content and bulk density for Romanian soils. More recently, Nemes et al. (2003) developed different PTFs derived from different scales of soil data (Hungary, Europe, and international data) using artificial neural network modeling including a PTF that uses soil texture and bulk density only.

Zacharias and Wessolek (2007) developed two different regression equations depending upon the percentage of sand in a soil as follows:

Sand content < 66.5%

$\Theta_r = 0$

$\Theta_s = 0.788 + 0.001clay – 0.263D_b$

$Ln(\alpha) = -0.648 + 0.023sand + 0.044clay – 3.168D_b$

$n = 1.392- 0.418sand^{-0.024} + 1.212clay^{-0.704}$

Sand content > 66.5%

$\Theta_r = 0$

$\Theta_s = 0.890 - 0.001clay – 0.332D_b$

$Ln(\alpha) = -4.197 + 0.013sand + 0.076clay – 0.276D_b$

$n = 2.562+ 7x10^{-9}sand + 3.750\ clay^{-0.016}$

The regression coefficients from these models were almost identical to those reported by Vereecken et al., (1989) (Vereecken $\Theta_s = 0.81 + 0.001clay – 0.283D_b$) for a different data set, adding further credibility to their general applicability. Zacharias and Wessolek (2007) recommended using the PTFs of Vereecken et al., (1989) if data on OM were available but felt that their proposed equations were suitable for use where OM data were not available.

Empirical equations developed by Williams et al. (1992) for the prediction of the constants A and B in the Campbell function have been widely used in Australia and elsewhere. These regression equations require particle size distribution, field texture and bulk density inputs as follows:

A = 1.996 + 0.136(ln C) - 0.00007(FS215) + 0.145(ln SI) + 0.382(ln TEX)

B = -0.192 + 0.0946(ln TEX) - 0.00151(FS)

C is % clay (< 0.002 mm); SI is % silt (0.002-0.02 mm); FS is % fine sand (0.02-0.20 mm), and TEX is texture group from 1-6 as defined by Northcote (1971).

Cresswell *et al*., (2006) demonstrated applicability of the Williams et al. (1992) method for French soils and confirmed that the approach of assuming a Campbell SWC model and empirically predicting the slope and air entry potential has merit. They concluded that the empirical regression equations of Campbell appeared transferable to different data sets from very different geographical locations. . They provided regression equations for all samples and stratified by horizon type that had r^2 values ranging from 0.81 to 0.91.

Cresswell *et al*., (2006) suggested a strategy for achieving adequate coverage of soil hydraulic property data for France that included an efficient sampling strategy based on the use of functional horizons (Bouma 1989) and a series of reference sites where soil hydraulic properties could be measured comprehensively. They argued that functional horizon method recognizes the soil horizon rather than the profile as the individual or building block for prediction (Wösten et al. 1985; Wösten and Bouma 1992). A significant feature of this approach is the capacity to create a complex range of different hydrologic soil classes from simple combinations of horizon type, sequence, and thickness.

It is anticipated that the SINFER approach proposed by McBratney *et al*., (2002) will be the basis for efforts to create and apply PTFs for predicting available water capacity for the *GlobalSoilMap* project. These PTFs have yet to be developed.

References for Appendix C

Adams, W.A. 1973. The effect of organic matter on the bulk and true density of some uncultivated podzolic soils. Journal of Soil Science, 24, 10-17.

Ahem, C.R., Baker, D.E., and Aitken, R.L. 1995. Models for relating pH measurements in water and calcium chloride for a wide range of pH, soil types and depths. Plant and Soil 171: 47-52.

Aitken, R. L. and Moody, P. W. 1991. Interrelations between Soil pH Measurements in Various Electrolytes and Soil Solution pH in Acidic Soils. Aust. J. Soil Res., 29, 483-91.

Alexander, E.B. 1980. Bulk density of Californian soils in relation to other soil properties. Soil Science Society of America Journal, 44, 689-692.

Arya, L.M., and J.F. Paris. 1981. A physico-empirical approach to predict the soil water moisture characteristic from particle size distribution and bulk density data. Soil Sci. Soc. Am. J. 45:1023-1030.

Ball, D. F. 1964. Loss-on-ignition as an estimate of organic matter and organic carbon in non-calcareous soils.. Journal of Soil Science, Vol. 15, No. 1. pp. 84-92.

Bernoux, M., Arrouays, D., Cerri, C., Valkoff, B. and Jolivet, C. 1998. Bulk densities of Brazilian Amazon soils related to other soil properties. Soil Science Society of America Journal, 62, 743-749.

Bisutti, Isabella, Hilke, Ines., and Raessler, Michael. 2004. Determination of total organic carbon – an overview of current methods. Trends in Analytical Chemistry, Vol. 23(10–11): 716-726.

Bouma, J. 1989. Land qualities in space and time. In: Land qualities in space and time, eds J Bouma & AK Bregt. Proceedings of a symposium organized by the International Society of Soil Science, Pudoc, Wageningen. pp. 3-13.

Bruand, A., Perez Fernandez, P. and Duval, O. 2003. Use of class pedotransfer functions based on texture and bulk density of clods to generate water retention curves. Soil Use and Management, 19, 232-242.

Bruce, R. C., Warrell, L. A., Bell, L. C., and Edwards, D. G. 1989. Chemical attributes of some Queensland acid soils. I. Solid and solution phase compositions. Aust. J. Soil Res. 27, 333-51.

Brye, K. R. and Slaton, N. A. 2003. Carbon and Nitrogen Storage in a Typic Albaqualf as Affected by Assessment Method. Communications in Soil Science and Plant Analysis, 34: 11, 1637-1655.

Calhoun, F.G., N.E. Smeck, B.L. Slater, J.M. Bigham, and G.F. Hall. 2001. Predicting bulk density of Ohio soils from morphology, genetic principles, and laboratory characterization data. Soil Sci. Soc. Am. J. 65:811-819.

Campbell, G.S., and S. Shiozawa. 1989. Prediction of hydraulic properties of soils using particle-size distribution and bulk density data. p. 317–328. In M.Th. van Genuchten et al. (ed.) Proc. Int. Worksh. on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, Riverside, CA. 11–13 Oct. 1989. Univ. of California, Riverside.

Canarache, A. 1993. Physical-technological soil maps—A possible product of soil survey for direct use in agriculture. Soil Technol. 6:3-15.

Chacón, N., Dezzeo, N., Fölster, H. and Mogollón, P. 2002. Comparison between colorimetric and titration methods for organic carbon determination in acidic soils, Communications in Soil Science and Plant Analysis, 33: 1, 203-211

Conyers, M.K., and Davey, B.G. 1988. Observations on some routine methods for soil pH determination. Soil Sci. 145, 29-36.

Cresswell, H.P., Coquet , Y., Bruand, A. and McKenzie, N.J. 2006. The transferability of Australian pedotransfer functions for predicting water retention characteristics of French soils. Soil Use and Management, 22:62-70.

Curtis, R.O. and Post, B.W. 1964. Estimating bulk density from organic matter content in some Vermont forest soils. Soil Science Society of America Proceedings, 28, 285-286.

De Vos, B., Lettens, S., Muys, B. and Deckers, J. A. 2007. Walkley–Black analysis of forest soil organic carbon: recovery, limitations and uncertainty. Soil Use and Management, 23, 221-229

De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J. and Muys, B. 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. Soil Science Society of America Journal, 69, 500-510.

FAO/IIASA/ISRIC/ISSCAS/JRC, 2009. Harmonized World Soil Database (version 1.1). FAO, Rome, Italy and IIASA, Laxenburg, Austria.

Federer, C.A., Turcotte, D.E. and Smith, C.T. 1993. The organic- bulk-density relationship and the expression of nutrient content in forest soils. Can. J. For. Res. 23:1026-1032.

Gijsman, A.J., Jagtap, S.S., Jones, J.W. 2002. Wading through a swamp of complete confusion: how to choose a method for estimating soil water retention parameters for crop models. Eur. J. Agron. 18, 75-105.

Gijsman, A.J., Thornton, P. K. and Hoogenboom, G. 2007. Using the WISE database to parameterize soil inputs for crop simulation models Computers and Electronics in Agriculture 56, 85-100

Grewal, K. S., Buchan, G. D. and Sherlock, R. R. 1991. A comparison of three methods of organic carbon determination in some New Zealand soils. Journal of Soil Science, 42, 251-257.

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

Heuscher, S.A., Brandt, C.C. and Jardine, P.M. 2005. Using Soil Physical and Chemical Properties to Estimate Bulk Density. Soil Sci. Soc. Am. J. 69:1-7.

Jankauskas B, Jankauskiene G, Slepetiene A, Fullen MA and Booth CA 2006. International Comparison of Analytical Methods of Determining the Soil Organic Matter Content of Lithuanian Eutric Albeluvisols. Communications in Soil Science and Plant Analysis 37, 707 - 7/20 (http://dx.doi.org/10.1080/00103620600563499)

Jagtap, S.S., Lal, U., Jones, J.W., Gijsman, A.J., Ritchie, J.T., 2004. A dynamic nearest-neighbor method for estimating soil water parameters. Trans. ASAE 47, 1437-1444.

Jeffrey, D.W. 1970. A note on the use of ignition loss as a means for the approximate estimation of soil bulk density. Journal of Ecology, 58, 297-299.

Jolivet, C. , Arrouays, D. and Bernoux, M. 1998. Comparison between analytical methods for organic carbon and organic matter determination in sandy Spodosols of France, Communications in Soil Science and Plant Analysis, 29: 15, 2227-2233

Kalembasa, S.J. and Jenkinson, D.S. 1973. A Comparative Study of Titrimetric and Gravimetric Methods for the Determination of Organic Carbon in Soil. J. Sci. Fd Agric., 24,1085-1090.

Kilmer, V.J., and Alexander, L.T. 1949. Methods of making mechanical analyses of soils. Soil Sci. 68:15-24.

Konen, M.E., Jacobs, P.M., Burras, C.L., Talaga, B.J. and Mason, J.A. 2002. Equations for Predicting Soil Organic Carbon Using Loss-on-Ignition for North Central U.S. Soils. Soil Sci. Soc. Am. J. 66: 1878-1881.

Manrique, L.A., and Jones, C.A. 1991. Bulk-density of soils in relation to soil physical and chemical properties. Soil Sci. Soc. Am. J. 55: 476-481.

Meersmans J., Van Wesemael, B. and Van Molle, M. 2009. Determining soil organic carbon for a agricultural soils: a comparison between the Walkley & Black and the dry combustion methods (north Belgium). Soil Use and Management 25, 346-353.

Mikhailova, E. A., Noble, R. R. P. and Post, C. J. 2003. Comparison of Soil Organic Carbon Recovery by Walkley-Black and Dry Combustion Methods in the Russian Chernozem, Communications in Soil Science and Plant Analysis, 34: 13, 1853-1860.

Miller, R.O. and Kissel, D.E. 2010. Comparison of Soil pH Methods on Soils of North America. Soil Sci. Soc. Am. J. 74:310-316

Minasny, B. 2007. Predicting soil properties. Jurnal Ilmu Tanah dan Lingkungan Vol. 7 No.1 (2007) p: 54-67.

Minasny, B. and McBratney, A.B. 2001. The Australian soil texture boomerang: a comparison of the Australian and USDA/FAO soil particle-size classification systems. Aust. J. Soil Res., 2001, 39, 1443-1451.

Moeys, J. 2010. Package – Soil texture V 2.02. Functions for soil texture plot, classification and transformation. http://soiltexture.r-forge.r-project.org/

Nemes, A., Wösten J.H.M., Lilly, A. and Oude Voshaar J.H. 1999a Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases. Geoderma 90, 187-202.

Nemes, A., Schaap M.G. and Leij, F.J. 1999b. 'The UNSODA unsaturated soil hydraulic database Version 2.0.' (US Salinity Laboratory: Riverside, CA)

Nemes, A., Schaap, M.G. and Wösten, J.H.M. 2003. Functional evaluation of pedotransfer functions derived from different scales of data collection. Soil Sci. Soc. Am. J. 67:1093-1102.

Oosterveld, M. and C. Chang, C. 1980. Empirical relationship between laboratory determinations of soil texture and moisture retention. Can. Agric. Eng. 22:149-151.

Pachepsky, Y.A., Rawls, W.J. and Lin, H.S. 2006. Hydropedology and pedotransfer functions. Geoderma, 131, 308-316.

Rawls, W.J. 1983. Estimating soil bulk-density from particle-size analysis and organic matter content. Soil Sci. 135:123–125.

Rawls, W.J., and Brakensiek, D.L. 1982. Estimating soil water retention from soil properties. J. Irrig. Drainage Div. Am. Soc. Civ. Eng. 108:166-171.

Rawls, W.J., and Brakensiek, D.L. 1989. Estimation of soil water retention and hydraulic properties. p. 275–300. In H.J. Morel-Seytoux (ed.) Unsaturated fl ow in hydrologic modeling; theory and practice. Proc. NATO Adv. Res. Worksh. Hydrology. NATO Sci. Ser. C. Kluwer Acad. Publ., Dordrecht, the Netherlands.

Rawls, W.J., Gish, T.J. and Brakensiek, D.L. 1991. Estimating soil water retention from soil physical properties and characteristics. Adv. Agron. 16, 213-234.

Rawls, W.J., Pachepsky, Y.A., Ritchie, J.C., Sobecki, T.M. and Bloodworth, H. 2003. Effect of soil organic carbon on soil water retention. Geoderma, 116:61-67.

Rousseva, S.S. 1997. Data transformations between soil texture schemes. European Journal of Soil Science, December 1997, 48, 749-758.

Saxton, K.E., Rawls, W.J., Romberger, J.S. and Papendick, R.I. 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50, 1031-1036.

Saini, G.R. 1966. Organic matter as a measure of bulk density of soil. Nature. 210:1295

Schumacher, B.A. 2002. Methods for the determination of total organic carbon (TOC) in soils and sediments. United States Environmental Protection Agency. Environmental Sciences Division National Exposure Research Laboratory. NCEA-C- 1282. 25 pp.

Shirazi M.A. and Boersma, L. 1984. A unifying quantitative analysis of soil texture. Soil Science Society of America Journal 48, 142-147.

Shirazi, M.A., Boersma, L. and Hart, J.W. 1988. A unifying quantitative analysis of soil texture: Improvement of precision and extension of scale. Soil Science Society of America Journal, 52, 181-190.

Skaggs, T.H., Arya, L.M., Shouse, P.J. and Mohanty, B.P. 2001. Estimating Particle-Size Distribution from Limited Soil Texture Data. Soil Sci. Soc. Am. J. 65:1038-1044.

Sleutel, S., De Neve, S., Singier, B. and Hofman, G. 2007. Quantification of Organic Carbon in Soils: A Comparison of Methodologies and Assessment of the Carbon Content of Organic Matter, Communications in Soil Science and Plant Analysis, 38: 19, 2647-2657.

Soon, Y.K. and Abboud, S. 1991. A comparison of some methods for soil organic carbon determination, Communications in Soil Science and Plant Analysis, 22: 9, 943-954.

Sumner, M. E. 1994. Measurement of soil pH: Problems and solutions, Communications in Soil Science and Plant Analysis, 25: 7, 859-879.

Timlin, D.J., Pachepsky, Y.A., Acock, B. and Whisler, F. 1996. Indirect estimation of soil hydraulic properties to predict soybean yield using GLYCIM. Agric. Syst. 52, 331-353.

Torri, D., Poesen, J., Monaci , F. and Busoni, E. 1994. Rock fragment content and fine soil bulk density. Catena. 23:65-71

Tranter, G., Minasny, B., McBratney, A. B., Murphy, B., McKenzie, N. J., Grundy, M. and Brough, D. 2007. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use and Management. 1-6.

Vogel, A.W. 1994. Comparability of soil analytical data: determinations of cation exchange capacity, organic carbon, soil reaction, bulk density, and volume percentage of water at selected pF values by different methods. Work. Pap. 94/07, ISRIC, Wageningen (Available at http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_1994-07.pdf ; last accessed 12/2008)

Wang, X.J., Smethurst, P.J. and Herbed, A.M. 1996. Relationships between three measures of organic matter or carbon in soils of eucalypt plantations in Tasmania. Aust. J. Soil Res., 1996, 34, 545-53.

Williams, J., Ross, P.J. and Bristow, K.L. 1992. Prediction of the Campbell water retention function from texture, structure and organic matter. In: van Genuchten, M. Th., Leij, F. J., Lund, L. J. (Eds.) Proceedings of the International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils. Riverside, CA. 11–13 Oct. 1989. University of California, Riverside, CA. pp. 427-441.

Wösten, J.H.M., Bouma, J. and Stoffelsen G.H. 1985. Use of soil survey data for regional soil water simulation models. Soil Science Society of America Journal 49, 1238-1244.

Wösten, J.H.M., Bannink, M.H., de Gruijter J.J. & Bouma, J. 1986. A procedure to identify different groups of hydraulic conductivity and moisture retention curves for soil horizons. Journal of Hydrology 86, 133-145.

Wösten, J.H.M. and Bouma, J. 1992. Applicability of soil survey data to estimate hydraulic properties of unsaturated soils. In: Proceedings of an International Workshop on Indirect Methods for Estimating the Hydraulic Properties of Unsaturated Soils, eds M. Th. van Genuchten, F.J. Leij & L.J. Lund, University of California, Riverside, CA, USA. pp. 463-472.

Wösten, J.H.M., Finke, P.A. and Jansen, M.J.W. 1995. Comparison of class and continuous pedotransfer functions to generate soil hydraulic characteristics. Geoderma 66:227-237.

Wösten, J.H.M., Lilly, A., Nemes, A. Le Bas, C. 1999. Development and use of a database of hydraulic properties of European soils. Geoderma 90:169-185.

Wösten, J.H.M., Pachepsky, Y.A. and Rawls, W.J. 2001. Pedotransfer functions: Bridging the gap between available basic soil data and missing soil hydraulic functions. J. Hydrol. 251:123-150.

Zacharias, S. and Wessolek, G. 2007. Excluding Organic Matter Content from Pedotransfer Predictors of Soil Water Retention. Soil Sci. Soc. Am. J. 71:43-50

Appendix D: Equal Area projections for use by each of the *GlobalSoilMap* nodes

This Appendix proposes to suggest a preferred equal area projection for each node to use for collating and processing projected data sets used to predict soil properties (Table D1).

Table 12. Suggested continental scale projections and their parameters for each node

Abbr	Description	Australia	Africa	Asia	Europe	N. America	S. America
Proj	Projection	Albers EA	Lambert EA	Mercator	Lambert EA	Albers EA	Albers EA
lat_1	Latitude of 1st standard parallel	-18		0		29.5	-5
lat_2	Latitude of 2nd standard parallel	-36				45.5	-42
lat_0	Latitude of Origin	0	5		52	23	-32
lat_ts	Latitude of true scale			0			
lon_0	Central Meridian	132	20	0	10	-96	-60
x_0	False Easting - X	0	0	0	4321000	0	0
y_0	False Northing - Y	0	0	0	3210000	0	0
ellps	Ellipsoide	GRS80	WGS84		GRS80	GRS80	aust_SA
datum	Datum	toWGS84	WGS84			NAD83	
units	Units of Distance	metres	metres	metres	metres	metres	metres
a	Semimajor radius of the ellipsoid axis			6378137			
b	Semiminor radius of the ellipsoid axis			6378137			
k	Scalng factor			1			
nadgrids	Grid based datum adjustment			"@null"			
wktext							
no_defs	Don't use the defaults file						

It is expected that each node will define a single node-wide projection in which to work. It is further expected that this projection will be some type of equal area projection in which all grid cells have the same fixed resolution. Equal area projections which organize data into grid cells of fixed horizontal dimensions are required by some of the key programs used to compute terrain attributes from DEM data or to implement geostatistical procedures such as kriging.

Nodes that work at a grid resolution finer than 100 m will be able to use the finer resolution data to compute bulked mean values for a 3 arc-second by 3 arc-second grid cell by averaging the values for all grid cells that occupy a target 3 arc-second by 3 arc-second reporting grid cell.

Nodes that elect to work at a grid resolution of 100 m or greater will need to use the property values of surrounding grid cells to compute a weighted average value for each property at each depth for each target 3 arc-second by 3 arc-second reporting grid cell.

Suggested projections expressed in terms of R-code

- **1st level** - **Whole world compilation projections**
    - **world** proj4: +proj=lonlat +ellps=WGS84
    - **Googlemaps** proj4: +proj=merc +a=6378137 +b=6378137 +lat_ts=0.0 +lon_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +units=m +nadgrids=@null +wktext +no_defs
- **2nd level – Continental scale compilation projections at the node level**
    - **au** (Australia and New Zealand) proj4: +proj=aea +lat_1=-18 +lat_2=-36 +lat_0=0 +lon_0=132 +x_0=0 +y_0=0 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_defs
    - **af** (Africa) proj4: +proj=laea +lat_0=5 +lon_0=20 +x_0=0 +y_0=0 +units=m +ellps=WGS84 +datum=WGS84
    - **as** (Asia) proj4: +proj=merc +a=6378137 +b=6378137 +lat_ts=0.0 +lon_0=0.0 +x_0=0.0 +y_0=0 +k=1.0 +units=m +nadgrids=@null +wktext +no_defs
    - **eu** (Europe) proj4: +proj=laea +lat_0=52 +lon_0=10 +x_0=4321000 +y_0=3210000 +ellps=GRS80 +units=m +no_defs
    - **na** (North America) proj4: +proj=aea +lat_1=29.5 +lat_2=45.5 +lat_0=23 +lon_0=-96 +x_0=0 +y_0=0 +ellps=GRS80 +datum=NAD83 +units=m +no_defs
    - **sa** (South/Central America) proj4: +proj=aea +lat_1=-5 +lat_2=-42 +lat_0=-32 +lon_0=-60 +x_0=0 +y_0=0 +ellps=aust_SA +units=m +no_defs

Appendix E: Background on uncertainty and guidelines for uncertainty methods

Tier 1 and Tier 2 products for any contiguous region, should include some minimal statement of accuracy, although more formal validation is not required.

At Tier 3 and 4, an appropriate validation measure for each property at each depth increment is the root mean square error of the point predictions. This can be achieved for instance by cross validation for point-based methods, and true validation for soil- class map based methods. In the latter case at least, on average, one observation point per 10 000 square kilometres may be required and probably at least 50 points are required to obtain an estimate.

In higher Tiers it is anticipated a richer set of validation criteria will be used, including mean error, and the percentage of the map area that fall within the uncertainty limits. Ultimately, sampling will be required to produce such quality estimates.

A general framework for assessing and representing uncertainties in environmental data is provided by Brown (2004).

Heuvelink and Brown (2007) observed that "soil data are rarely certain or 'error free', and that these errors may be difficult to quantify in practice". Indeed, the quantification of error (defined here as a 'departure from reality') implies that the 'true' state of the environment is known. They reported that "in recent years, a distinct spectrum of methods, not altogether statistical, has emerged for dealing with situations of 'imperfect knowledge' in scientific research (see Ayyub, 2001 also)". A spectrum of methods for uncertainty analysis is indeed important for DSM (and *GlobalSoilMap*) with due consideration of the potential sources of uncertainty—namely from inputs (observed data and covariate information), model parameters and model structure. Similarly, methods of uncertainty analysis will vary on the basis of whether soil point data or existing soil maps are used for producing *GlobalSoilMap* outputs. Moreover, methods will also vary dependent on the density of point data as well.

As such, guidance on uncertainty analysis for *GlobalSoilMap* product will take the following form:

When there are sufficiently many point observations, there are two general approaches:

1. Statistical modeling (principally geostatistical models) of the soil properties directly. The uncertainty of predictions is generated from the model as a byproduct.
2. Statistical modeling (principally geostatistical models) of residuals from independent data set or resampling techniques

If there are insufficient point observations then the use of expert knowledge may be a viable option. Examples of this include:

1. Uncertainty parameters e.g. the lower and upper limits for 90% prediction intervals for soil attributes. In practice this could mean the upper and lower limits for each property at each location and depth.

2. Expert elicitation for (parametric form of) distributions of soil attributes in specific soil classes
3. For soil maps where soil class proportions and attribute ranges are quantified, these values can be combined in a conservative manner (via lowest of the low etc.) or via numerical methods.
4. Expert elicitation for variogram specification

As stated in the body of the specification document, for tier 1, our uncertainty of the reality is to be expressed as the 90% Prediction Interval (PI) which reports the range of values within which the true value is expected to occur 9 times out of 10 (or 90% of the time).

For all uncertainty methods, it is the probabilistic ones that may be most practicable for *GlobalSoilMap*. In regards to probability density functions (pdfs), Heuvelink and Brown (2007) argued that they confer a number of advantages over non-probabilistic techniques. For example, pdfs include methods for describing interdependence or correlation between uncertainties, methods for propagating uncertainties through environmental models and methods for tracing the sources of uncertainty in environmental data and models (Heuvelink, 1998). Notwithstanding these advantages, and the current popularity of stochastic methods in environmental research, there are a number of ongoing challenges for the successful application of pdfs to environmental data. In particular, there is a need to support the identification and estimation of pdfs in specific cases, as well as their storage in environmental databases.

Thus, the general pdfs need to be simplified in order to make them estimable in practice and tractable to storage within a soil database. The pdf of a numerical or categorical constant may be simplified by describing the uncertainty with a characteristic shape function, for which a small number of parameters must be estimated. Rather than specifying the entire pdf it is therefore sufficient to define the shape function and to estimate its parameters. For example, measurement error in a continuous numerical attribute is often assumed to follow a normal distribution (Heuvelink, 1998). This implies that the pdf is reduced to only two parameters, namely the mean and standard deviation, which describe the bias and average magnitude of uncertainty in the soil attribute, respectively. Similarly, it may be reasonable to assume that the number of stones in a volume of soil is Poisson distributed, for which the discrete pdf is reduced to only one parameter.

Useful simplifications must satisfy two conditions. First, the simplified pdfs must be estimable in practice, as well as tractable to storage within a soil database. Secondly, they must approximate the uncertainty in a soil variable sufficiently for their intended application. Among others, the elaboration and subsequent storage in the database must include the following aspects:

1. Uncertainty is subjective. The database must allow the opinions of different ‘experts’ to be stored.
2. Uncertainty information is very sensitive to the support size of the data items (Heuvelink and Pebesma, 1999). Here ‘support’ refers to the volume, magnitude and length of the

entity described. Support size (in time and space) should always be specified in a database.

3. The uncertainty in a particular variable may well be statistically dependent on the uncertainty in another variable. Statistical dependencies (and cross correlations) between uncertain variables can have a marked influence on how uncertainties propagate in a modelling study. These create a need to address uncertainty in spatially distributed or dynamic attributes because these are strongly affected by about dependencies and correlations.

In terms of implementation, where there is sufficient point data to define the underlying probability distribution function (pdf) for conventional statistical analysis, a 90% confidence interval will be calculated to establish the range of the prediction interval (PI) for each predicted soil property. When using statistical modeling (principally geostatistical models) of the soil properties directly. The uncertainty of predictions is generated from the model as a byproduct from which the necessary PIs can be derived. Sufficient data or computing power is often not available to achieve this.

An alternative method to estimate PI has been presented and described in Malone *et al*. 2011). Here uncertainty is treated as the probability distribution of the output model errors, which comprises all sources of uncertainty (model structure, model parameters and input data). And since it is estimated through an empirical distribution, it is not necessary to make any assumption about residuals (Solomatine and Shrestha 2009). This method is particularly useful when we are dealing with soil spatial prediction functions that include data-mining tools or neural networks (as examples) in combination with the regression-kriging approach, where it would be difficult to use other existing methods (of uncertainty analysis) to estimates of uncertainty. The approach can be summarized in the following steps:

1. Apply an unsupervised classification technique (e.g. fuzzy k-means) to the covariate data layers assembled and used to make the predictions of soil properties for a particular area or soil-landscape zone to produce functional classes (4-5).
2. Overlay all available geo-registered soil profile analytical data on the resulting 4-5 functional class map for a particular region or soil-landscape zone of interest.
3. Compute the probability distribution function (pdf) for each soil property of interest, at each depth of interest, within the 4-5 functional classes. This establishes the range and distribution of observed soil property values within each of the 4-5 functional classes.
4. Use the pdf computed for each soil property at each depth for each class to identify the values at the 5% and 95% confidence limits (the 90% prediction interval or PI).
5. Use the values at the 5% and 95% confidence limits of the pdf for each class as inputs in calculating a weighted fuzzy mean value for the upper and lower confidence intervals for each grid cell.
6. The method for computing upper and lower confidence limits for any grid cell is based on computing a weighted average of the confidence limit value for each of the 4-5 classes times the fuzzy likelihood value of that class for all n classes at each location.

7. The estimate of uncertainty at each grid cell is a weighted average of the similarity of the conditions at each cell to the conditions that define each of the N classes.

The approach of Malone *et al*., (2011) requires that there be a sufficient number and density of point observations within any given prediction area (30 per class) to support a data driven assessment of the pdf of a given soil property by class within the geographic extent of an area of interest.

If sufficient information does not exist to support conventional statistical analysis, the range will have to be assessed by appropriate local or national experts. Fuzzy logic (Cazemier *et al*., 2001) and Bayesian beliefs (O'Hagan *et al*., 2006) have been proposed as suitable frameworks for establishing estimates of uncertainty in the absence of sufficient hard field data.

See also Appendix H

As an example, Lilburne et al., (2009) presented a method based on using expert knowledge to estimate the pdf in situations where there is insufficient information to support conventional statistical analysis. This following method is adapted and presented as one example of how expert knowledge can be used to estimate uncertainty for data derived from a polygon soil map, as follows:

1. Best available expert knowledge and observed or measured data are examined for each soil map unit, or for taxonomic classes included within the map unit.
2. The variability of the property is described in the form of a probability distribution function (pdf).
3. If data are available, a normal, lognormal, or beta function can be used. An additional combination pdf termed a duplex function has also been proposed. This combines a triangular or uniform distribution with a single-valued discrete pdf for the minimum or maximum value.
4. Confidence in the base property data is indicated by an expert assigned confidence code.

For the present, uncertainty will be reported as the best feasible estimate of the range of values within which a prediction of a soil property at any depth and any location is expected with 90% confidence.

References for Appendix E

Brown, J.D., 2004. Knowledge, uncertainty and physical geography: towards the development of methodologies for questioning belief. Trans. Inst. British Geographers 29, 367–381.

Heuvelink, G.B.M., Pebesma, E.J., 1999. Spatial aggregation and soil process modelling. Geoderma 89(1-2), 47-65.

Heuvelink, G.B.M., Brown, J. D. 2007. Towards a soil information system for uncertain soil data. Developments in Soil Science, volume 31. P. Lagacherie, A.B. McBratney and M. Voltz (Editors). Elsevier B.V. pp. 97-106.

Lilburne, L., Hewitt, A. and Ferriss, S., 2009. Progress with the design of a soil uncertainty database, and associated tools for simulating spatial realisations of soil properties. 7th International Symposium on Spatial Accuracy Assessment in Natural Resources and Environmental Sciences. Edited by M. Caetano and M. Painho. pp. 5510-519.

Malone, B.P., McBratney, A.B., Minasny, B., 2011. Empirical estimates of uncertainty for mapping continuous depth functions of soil attributes. Geoderma 160(3-4), 614-626.

Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model uncertainty using machine learning techniques. Water Resources Research 45, Article Number: W00B11.

Appendix F: North American Node tiering concepts

The North America Node of *GlobalSoilMap* has developed a tiering concept where each successive tier of map products incorporates improvements over the previous tier. In the United States, initial tiers will rely heavily on the use of existing area-class maps as held in STATSGO2 and SSURGO, with the understanding that later tiers will make better use of spatial disaggregation and point observations and eventually lead to truly continuous raster property maps. For the United States, the proposed tiering system is as follows:

"Tiers 0.1–0.4". Essentially rasterised versions of the existing STATSGO2 map. Soil components have been fitted with equal-area splines (Bishop et al., 1999; Malone et al., 2009), which allows property estimates to be made at the GlobalSoilMap standard depth increments . Weighted means are then calculated within map units where more than one soil component possesses data, otherwise the single soil component's property estimate is reported. The products may have gaps where soil exists but property values were not recorded. Map units will not have been harmonized so artifacts (data discontinuities) may exist at political boundaries. The products are not truly continuous. Pre-existing scale discrepancies are inherited from STATSGO2. For initial tiers, the "upper" and "lower" values reported in STATSGO2 are assumed to be the 95% confidence limits for uncertainty estimation. Later tiers may estimate probability distribution functions and hence confidence intervals from legacy point observations in the National Soil Characterization Database.

"Tiers 0.5–0.9". Essentially as per tiers 0.1–0.4 but the higher-detail SSURGO map will be used. Tiers will begin to incorporate harmonization of map units at political boundaries, which may necessitate the use of spatial disaggregation and other predictive techniques. Techniques will need to be developed to fill gaps in SSURGO map where no SSURGO data currently exists. Disaggregation of STATSGO2 may be a potential solution, as may other predictive approaches such as the homosoil concept (Mallavan et al., 2010). Uncertainty as per tiers 0.1–0.4, or other quantitative techniques.

"Tiers 1.0+". These tiers will be produced using predictive techniques. Products will not contain gaps other than where bedrock or water exists. Products will not contain scale discrepancies. Uncertainty will be represented as 95% confidence interval estimated by quantitative techniques will be refined as more point observations become available. **Tiers <1** will be released for comment and feedback, but it is anticipated that collated international products from nodes will meet v1 standards.

Appendix G: National variants of *GlobalSoilMap* Products

Besides the concept of tiering there will be a need for various individual countries to produce products that otherwise meet *GlobalSoilMap* tiering standards in terms of resolution, depth range, uncertainty etc., but *do not correspond to* the standard set of properties. These are designated a national superset,

GLOBALSOILMAPnss ‘name of country’ ‘tier no’ ‘soil property’

e.g., GLOBALSOILMAPnss USA V1 pH in KCl(1:2),

and are intended for national use only, Common examples may be soil properties based on local analytical methods, particularly pH and carbon, cation exchange capacity. It is envisaged that in some or many instances these may be produced first and pedotransfer functions (Appendix C) used to convert these to the *GlobalSoilMap* standard.

Appendix *H*: An approach for uncertainty estimation of soil spatial predictions based on soil class (maps) with limited within-class variability information

Introduction

Often we do not have a sufficient quantity of soil property data per soil class in order to generate soil property distributions and estimate their parameters (e.g., mean and variance), which may then be used to derive confidence intervals. Techniques exist for estimating parameters of distributions based on small samples but they may be computationally expensive or specific for a particular distribution (normal, lognormal, etc.).

Although sample data may be limiting, often we have some idea of the range and typical value of soil properties for soil classes. In the presence of such data, a useful distribution to use is the triangular distribution. It may be positively or negatively skewed or symmetrical, but it must be unimodal (Johnson, 1997). It has been used in risk analysis, where it is often a proxy for the more computationally complex beta distribution often used when sample data is limiting (Williams, 1992). The probability density function of the triangular distribution is given in3.

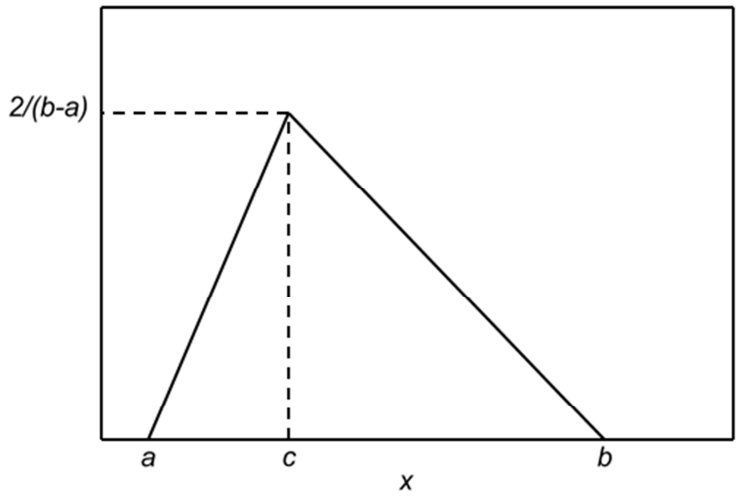


Figure 3. Probability density function of the triangular distribution.

We can use the triangular distribution in a sampling simulation approach to estimate uncertainty, as will be explained in the following section.

General approach

For some spatial entity (e.g., a map unit delineation or a grid cell), assume *s* soil classes S_i, $i = 1…s$ exist or have the potential to exist. Draw *n* times from the triangular soil property distribution of each soil class. A random draw from the triangular distribution is determined as follows:

$$\begin{cases} g = a + \sqrt{U(b-a)(c-a)} & \text{for } 0 < U < F(c) \\ g = b - \sqrt{(1-U)(b-a)(b-c)} & \text{for } F(c) \leq U < 1 \end{cases} \quad \text{(H1)}$$

where g is the soil property value drawn from the distribution, U is a random variate drawn from the uniform distribution in the interval (0,1), $F(c) = (c - a)/(b - a)$ and a, b and c are the parameters of the triangular distribution..

For each draw N = 1…n, calculate the weighted mean property value, μ_N^*, as:

$$\mu_N^* = \sum_{i=1}^{s} w_i g_{iN} \quad \text{(H2)}$$

where g_{iN} is the soil property value drawn from the triangular distribution and w_i is the weight associated with S_i. Weights are explained below.

After all draws have taken place, we will have a set of n weighted mean soil property values from which a distribution can be generated. The lower and upper bounds of the 90% confidence interval are then the 5th percentile and 95th percentile of the distribution of weighted means, respectively.

Special cases

Two special cases of this approach exist. The first case involves situations where the most feasible approach is to generate within-map unit spatially weighted mean soil property maps. In this case the spatial entity upon which calculations are carried out is the soil map unit polygon. For each polygon, s soil classes are assumed to exist in known (or assumed) proportions, but the precise spatial distribution of each soil class is unknown. The soil property map thus reports a single weighted mean soil property value for each map unit, for each depth increment. In this case, the weight w_i in equation 2 is the areal proportion of soil class S_i in the map unit (Odgers et al., 2012).

The second case involves situations where the soil polygon map has been spatially disaggregated. The result of the spatial disaggregation is a set of s raster maps where the map for S_i indicates the probability of occurrence of S_i at each grid cell. This probability varies continuously in space. The spatial entity upon which calculations are carried out is the grid cell. In this case the weight w_i in equation 2 is the probability of occurrence of S_i at the given grid cell.

Obviously these methods can be applied for any probability density function.

References

Johnson, D., 1997. The triangular distribution as a proxy for the beta distribution in risk analysis. Journal of the Royal Statistical Society, Series D (The Statistician) 46(3), 387–398. doi: 10.1111/1467-9884.00091

Odgers, N.P., Libohova, Z., Thompson, J.A., 2012. Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. Geoderma 189–190, 153–163. doi: 10.1016/j.geoderma.2012.05.026

Williams, T.M., 1992. Practical use of distributions in network analysis. The Journal of the Operational Research Society 43(3), 265–270.

Appendix I Suggested data density for point-based methods

Experience suugests that a density of between 1 and 10 observations per 1000 km^2 is required for point predictions. The larger densities are required for regions with larger instrinsic pedodiversity and where the legacy sampling is clustered (in space or with repect to the covariates).