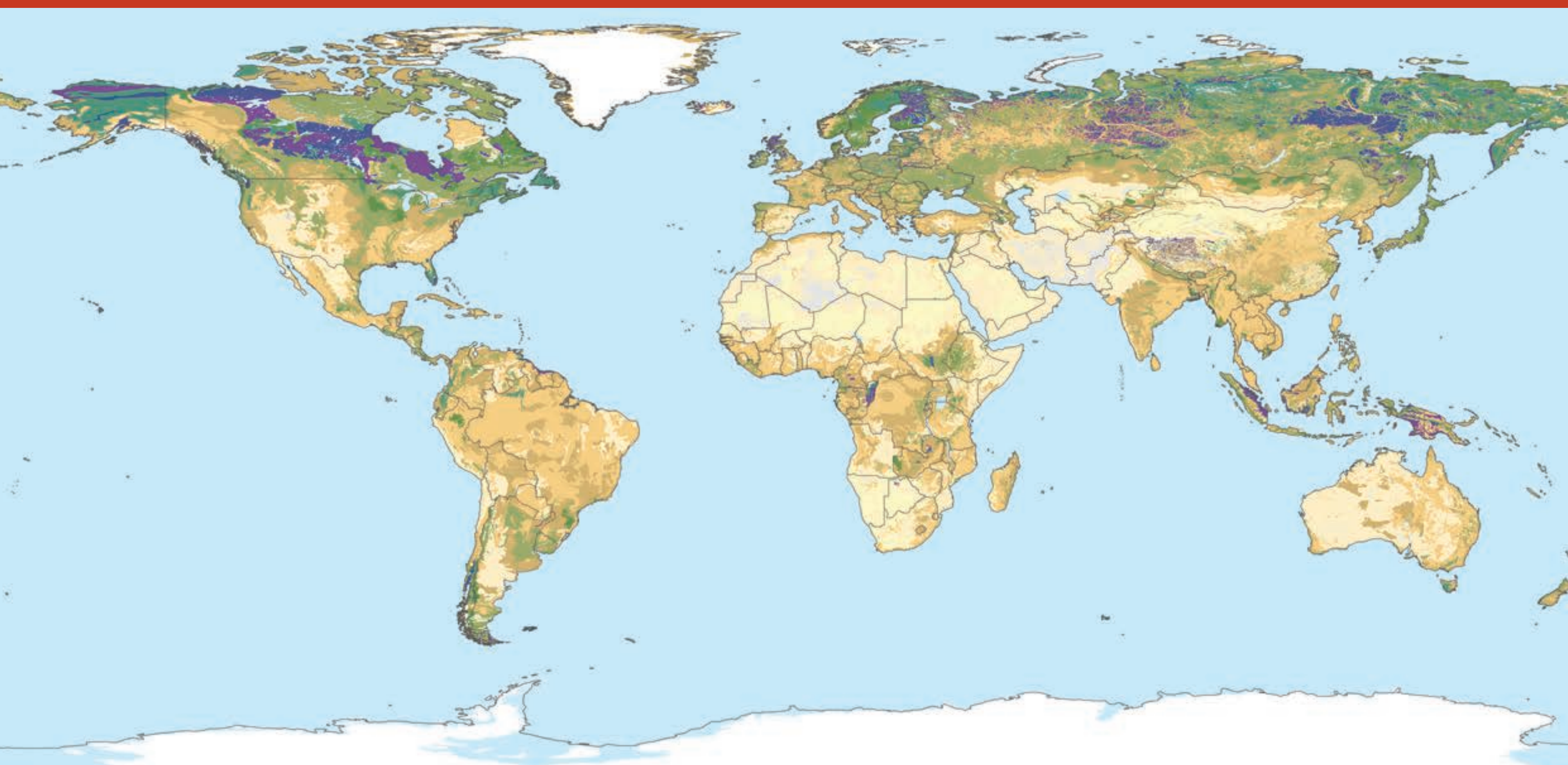
World soil property estimates for broad-scale modelling (WISE30sec)

World Soil Information

ISRIC Report 2015/01



Niels H. Batjes

2015 International Year of Soils

World soil property estimates for broad-scale modelling (WISE30sec)

Niels H. Batjes

ISRIC Report 2015/01

Wageningen, July 2015

Citation
Batjes NH 2015. World soil property estimates for broad-scale modelling (WISE30sec). Report 2015/01, ISRIC - World Soil Information, Wageningen (with data set, available at www.isric.org)

ISRIC Report 2015/01

Contents

List of tables

List of figures

Preface

ISRIC - World Soil Information has the mission to create and increase the awareness and understanding of the role of soils in major global issues. As an international institution, we inform a wide audience about the multiple roles of soils in our daily lives; this requires scientific analysis of sound soil information.

This study presents soil property estimates for the world for application at a broad scale. The GIS dataset was compiled using traditional mapping approaches. It is comprised of a soil-geographical and a soil attribute component. The former was derived from a GIS overlay of the Harmonised World Soil Database (HWSD) and Köppen-Geiger climate class map as a co-variate, while soil property estimates attribute data for these compound map units were derived using taxotransfer (TTR) procedures. The TTR scheme draws heavily on statistical analyses of analytical data managed in the ISRIC-WISE soil profile database.

The 30 by 30 arcsec resolution GIS dataset may be used for exploratory assessments at the global level keeping in mind the inherent generalisations, assumptions and possible limitations for use that are described in the report. Databases at finer spatial resolution will be needed for studies at a sub-national and regional level.

ISRIC, in its capacity of World Data Centre for Soils, is seeking collaboration with national institutes with a mission for soil resource inventories in order to further develop its world soil information services for the benefit of the international community.

ir. Rik van den Bosch
Director, ISRIC — World Soil Information

Summary

This report describes a harmonized dataset of derived soil properties for the world that is comprised of a soil-geographical and a soil attribute component. The GIS dataset was created using the soil map unit delineations of the broad scale Harmonised World Soil Database, version 1.21, with minor corrections, overlaid by a climate zones map (Köppen-Geiger) as co-variate, and soil property estimates derived from analyses of the ISRIC-WISE soil profile database for the respective mapped 'soil/climate' combinations. The dataset considers 20 soil properties that are commonly required for global agro-ecological zoning, land evaluation, crop growth simulation, modelling of soil gaseous emissions, and analyses of global environmental change. It presents 'best' estimates for: organic carbon content, total nitrogen, C/N ratio, pH(H_2O), CEC_{soil}, CEC_{clay}, effective CEC, total exchangeable bases (TEB), base saturation, aluminium saturation, calcium carbonate content, gypsum content, exchangeable sodium percentage (ESP), electrical conductivity, particle size distribution (content of sand, silt and clay), proportion of coarse fragments (> 2 mm), bulk density, and available water capacity (-33 to -1500 kPa); also the dominant soil drainage class. These estimates are presented for fixed depth intervals of 20 cm up to a depth of 100 cm, respectively of 50 cm between 100 cm to 200 cm (or less when appropriate) for so-called 'synthetic' profiles' (as defined by their 'soil/climate' class). The respective soil property estimates were derived from statistical analyses of data for some 21,000 soil profiles managed in a working copy of the ISRIC-WISE database; this was done using an elaborate scheme of taxonomy-based transfer rules complemented with expert-rules that consider the 'in-pedon' consistency of the predictions. The type of rules used was flagged to provide an indication of the possible confidence (i.e. lineage) in the derived data.

Best estimates for each attribute are given as means and standard deviations (STD), as calculated for the sample populations that remained upon application of a robust data outlier detection scheme. Results of the analyses can be linked to the spatial data through the unique map unit (grid cell) identifier, which is a combination of the soil unit and climate class code. Most map units are comprised of up to ten different components; each of these with their own range of derived soil properties and associated statistical uncertainties.

The present soil property values are 'best estimates' based on the current selection of soil profiles in WISE, and criteria and procedures for clustering the measured data in the taxotransfer scheme. They may be used for assessments at a continental and global level (scale < 1:1M) upon due consideration of the underlying generalisations and assumptions. These assessments should consider the full map unit composition, i.e. calculate first for each 'soil/climate cluster, depth layer and attribute', and then aggregate the results. Estimates of global soil organic carbon (SOC) stocks are presented as an example. According to this study, some 30% (607 ± 87 Pg C) of the total SOC stock (2060 ± 215 Pg C) to 2 m depth is held in the Northern Circumpolar Region, which is considered most sensitive to climate change.

Studies at (sub)national level should be based on regionally more detailed soil data sets.

Keywords: soil units, derived soil properties, uncertainty estimates, environmental modelling, ISRIC-WISE database, Harmonised World Soil Database, climate zones

1 Introduction

Soils play a key role in providing a range of ecosystem services (Bouma and McBratney 2013; IPCC 2006; MEA 2005). Soil information needs in support of studies of environmental, societal and economic sustainability, however, will vary with scale and the user's perspective and demand (Batjes *et al.* 2013; Finke 2006; Omuto *et al.* 2012). Despite rapid and significant progress in digital soil mapping techniques (Arrouays *et al.* 2014; Hengl *et al.* 2014), many user groups operating at a global level still require datasets that are derived from so-called 'traditional' mapping methods (FAO-GSP 2014a; Omuto *et al.* 2012).

The 'Harmonised World Soil Database' (HWSD, FAO *et al.* 2012) is probably the best traditional soil map for the world, but it has several limitations (Hengl *et al.* 2014; Omuto *et al.* 2012). Some of these relate to the spatial data and use of a two-layer model while others concern the derived attribute data, in particular their so far unquantified uncertainty. By its nature, however, a full revision of the HWSD product is beyond the scope of the present study. Such an activity would require commitment and collaboration from national soil resources institutes worldwide, as foreseen within the framework of the Global Soil Partnership (FAO-GSP 2014a).

Each map unit on HWSD is characterised by up to ten component soil units; these are either characterised by the original (FAO-Unesco 1974) or revised (FAO 1988) Legend depending on the source databases used to represent various parts of the world (FAO *et al.* 2012). Derived properties for each of these soil units have been determined using taxotransfer procedures that consider the FAO soil unit name, depth zone and soil textural class *albeit* using criteria that vary with the source materials (e.g. 1974 or 1988 FAO Legend). Further, no quantitative information has been provided with HWSD concerning the uncertainty of the different estimates, which is 'assumed to be large'.

The aim of this study is to develop a globally consistent taxotransfer procedure (TTR), for application with the soil-geographical data of the HWSD, that also provides uncertainty estimates for the derived (numerical) data. The procedure builds on earlier work of ISRIC, FAO, IIASA and other partners (Batjes *et al.* 1997; Batjes *et al.* 2007; FAO 1995; FAO *et al.* 2012), recognising that regional differences in soil type are useful 'carriers of soil information' (Bouma *et al.* 1998; Bouma *et al.* 2014). Similarly, soil taxonomic units derived from the HWSD map were shown to be an important predictor in digital soil mapping at a global scale (Hengl *et al.* 2014).

The current analyses consider the soil-geographical data of the HWSD (FAO *et al.* 2012), climate classes of the Köppen-Geiger system (Peel *et al.* 2007) as an important co-variate, as well as analyses of the full complement of soil profile (attribute) data held in the ISRIC-WISE database. Quantitative measures for the uncertainty/variability of the derived soil properties by 'soil/climate' cluster and soil layer are provided as: mean, standard deviation (STD) and standard error (which takes into account the sample size, *n*); median and median of absolute deviation; percentiles (10% steps); and, minimum and maximum. The TTR procedures uses 'best estimates' for mean and STD for the respective strata. As an initial application, the database (WISE30sec) has been used to calculate global soil organic carbon stocks to depth of 2m.

Materials and methods are described in Section 2, results of the TTR-procedure are presented and discussed in Section 3, while conclusions are drawn in Section 4. The structure of the various output tables, GIS files, and installation procedure are presented in the Appendices.

2 Materials and Methods

2.1 Soil profile data

The present study draws heavily on statistical analyses of harmonised soil profile data. For this, the ISRIC-WISE soil profile database (Batjes 2009, 2011) was complemented with some 8,000 'new' profiles, originating mainly from North America (ISCN 2014) and 'High Latitude' regions (Harden *et al.* 2012; Hugelius *et al.* 2014b; Michaelson *et al.* 2013). This included re-classification of the original USDA Soil Taxonomy names to the FAO (1988) Legend; inherently, this correlation encompassed generalisations (see Spaargaren and Batjes 1995).

Ultimately, ~21,000 profiles were available for the present study, up from ~10,200 for an earlier global mapping exercise at 5 by 5 arc minute resolution (Batjes 2012) that still drew on the spatial data of the Digital Soil Map of the World (DSMW, see FAO 1995). Nonetheless, there are still several soil geographic and taxonomic gaps in the data set and the spatial distribution of the profiles remains uneven. Further, the full complement of soil analytical attributes considered in this study (Table 1) is seldom available for many profiles.

Descriptive information on essential site data, such as climate, parent material and land use or natural vegetation, as well as detailed geo-location, is not provided in many source materials. As a consequence, stratification of the available profile data according to say main soil forming factors is not possible yet (see Gray *et al.* 2009). Profile coordinates are used to infer the broad climate at each site using GIS overlay with the Köppen-Geiger map (Peel *et al.* 2007) for subsequent consideration in the analyses.

Almost every country has its own analytical methods and these methods may vary from one laboratory to the next within one country; this is partly so because soil analyses are often soil type specific. Consequently, issues of quality and comparability of soil analytical data, collated from disparate sources, are critical in any analysis of soil profile data. Yet, there are no straightforward solutions for harmonising the data (Batjes 1999; Dobos *et al.* 2006; GlobalSoilMap.net 2013; Pleijsier 1989; van Reeuwijk 1998). At the present broad scale, these issues have been addressed pragmatically similar to what has been the case for earlier studies (Batjes 2002b; ESB 2001; FAO *et al.* 2012). Correlation of soil analytical data, however, should be done more rigorously when more detailed scientific work, at a finer spatial resolution, is considered. At the global level, the scope for concerted harmonisation efforts is being considered within the framework of Pillar 5 of the Global Soil Partnership (FAO-GSP 2014b).

2.2 Soil geographical data

HWSD combines regional and national updates of soil information worldwide (European Soil Database, Soil Map of China, SOTER and WISE-derived databases), mapped at a scale of 1:1 to 1:5 million, with the information contained within the 1:5 million scale DSMW for those countries that still have to contribute updated spatial and attribute information to the ongoing HWSD effort (see FAO *et al.* 2012; Omuto *et al.* 2012). Reliability of the information presented on the HWSD is therefore variable; it is considered lowest for those sections of the world that still draw on the DSMW, including North America, Australia, as well as large sections of South East Asia and West Africa (Nachtergaele *et al.* 2012). Updates of the information for those regions is foreseen in the framework of the Global Soil Partnership (GSP), under its Pillar 4 (FAO-GSP 2014a).

Each map unit of the HWSD is defined as being comprised of single soil units or associations thereof. Compound map units can include one dominant soil unit and up to nine associated soils. Each soil unit has been characterised (named) according to the Revised FAO Legend (FAO 1988), except for those areas that still draw on the DSMW which uses the original Legend (FAO-Unesco 1974). For the latter part, the soil unit names were correlated to those of the Revised Legend (Appendix I). This involved some generalisations as the original FAO Legend is less detailed that the Revised Legend; see FAO (1988) for a discussion. Similar types of broad scale correlation have been undertaken for application at a supra-national level by other groups (Jones *et al.* 2005; Jones *et al.* 2013).

For pragmatic reasons, as explained in the HWSD report, the spatial data were rasterized at 30 by 30 arc-seconds, which corresponds with a grid cell size of about 0.93 km x 0.93 km (0.86 km^2) at the equator. By implication, most individual soil map units provided on the original vector maps will consist of multiple grid cells. On the HWSD, these grid cells are assumed to have the same soil unit composition, hence derived attributes, irrespective of their location in the original polygons which is a recognised simplification (FAO *et al.* 2012, p. 2). As such, the presently used '1km grid cell' size should not be confused with the actual '1 km resolution' of the SoilGrids1km product as derived from digital soil mapping (Hengl *et al.* 2014).

Thirteen map units on HWSD have been mapped as having no information (coded 'NI'), four of which are fairly large. For the larger 'polygons', namely map unit (MU-Global) number 794 (Siberia), 16512 (Sinai), and 18690 (Namibia), proxies for the grid cell composition were derived from the Soil Map of Europe (Jones *et al.* 2005) respectively the Soil Map of Africa (Jones *et al.* 2013), and the delineations of these map units where updated.

Subsequent to the above 'corrections' and the conversion/correlation to the FAO Revised Legend, the map unit composition was re-assessed and the sequential numbering for the component soil units updated (i.e. SEQ 1 for dominant soil unit, etc.). The original map unit codes were maintained to preserve consistency with the original codes (MU_Global) used with HWSD, except for the newly updated areas mentioned above for which new codes (MUGLB_NEW) were created (Appendix 4).

2.3 Climate zones

During compilation of the DSMW, it was recognised that morphologically similar soils that occur in different climate zones should be separated. Climatic conditions, in particular available moisture and air temperature are important determinants of production. Some major-climate related soil features are considered on the HWSD, either through the Revised Legend code (e.g. presence of permafrost) or through the 1974 Legend (e.g. arid conditions for Xerosols and Yermosols). Possible effects of regional variation in climate, relief, parent material, and land use and management practices on specific soil properties, however, cannot be considered explicitly on broad scale maps such as the DSMW and HWSD. By implication, so far, the same parameter estimates for soil organic carbon content for say Haplic Cambisols thus had to be used irrespective of whether these soil units are located, for example, in cool and humid parts of Western Europe as opposed to hot and arid regions in the Middle East. To account for this known source of variation, broad climate zones were introduced as a categorical covariate in the analyses.

Following initial analyses, the 0.1 degree grid 'updated world map of the Köppen-Geiger climate classification' (Peel *et al.* 2007) was selected as being most appropriate for the present exploratory study (see 2.5). The classification considers precipitation effectiveness for plant growth as the major classification factor, and uses the appropriate seasonal values of temperature and precipitation to determine the limits of climatic groupings. The Köppen system figures a shorthand code of letters designating major climate groups, subgroups within these major groups, with further subdivisions to distinguish particular seasonal characteristics of temperature and precipitation.

Each soil profile was allocated to a Köppen-Geiger climate class, using GIS overlays, based on its given location. Consideration of all possible combinations up to the 3rd level (19 classes, for example BSh: 'Arid, Steppe, hot') or 2nd level (13 classes, such as BS: 'Arid, Steppe') of the classification system, however, resulted in rather small sample populations for many of the corresponding strata. Therefore, profiles were allocated to the '1st level climate' of the grid cell in which they occur. There are five possibilities for this: 'A: Tropical', 'B: Desert', 'C: Temperate', 'D: Cold' and 'E: Polar'; detailed definitions are given in Peel *et al.* (2007). For example, a Calcaric Cambisol mapped in an Arid zone would be coded as belonging to the 'CMc/B' soil-climate cluster, whereas similar soil units mapped in the Temperate zone would be coded as 'CMc/C'.

GIS-overlay of the HWSD and main climate classes, using majority sampling (ESRI 2015), yielded over 16,000 unique 'soil/climate' map units for the world. These provided the spatial basis for application of taxotransfer rules aimed at deriving 'best estimates' for twenty soil properties to a depth of 200 cm (see Section 2.5).

2.4 Land use and management

Land use and management, thus human intervention, are an important soil forming factor (Jenny 1941). For many profiles managed in the WISE, however, there is no information on land use and management (history) as this information has generally not been documented in the underpinning source materials (Batjes 2011). Hence, it was not feasible to group the available soil profiles according to some broad land use categories, such as 'Forest Land', 'Cropland', 'Grassland', 'Wetlands', 'Settlements' and 'Other land' as used for IPCC Tier I type greenhouse gas inventories at the (supra)national level (IPCC 2006; Ravindranath and Ostwald 2008). The alternative of deriving such categories for individual (geo-referenced) profiles, sampled over a period of say 40 years, from remotely sensed land cover data using GIS overlays still poses many challenges — the associated uncertainties are large (De Paul Obade and Lal 2013; Giri *et al.* 2013; Verburg *et al.* 2011).

2.5 Taxotransfer scheme

The TTR scheme considers twenty soil properties (Table 1). First, derived values or 'best estimates' for each of these were derived from statistical analyses of the 'clustered', measured analytical data (after outlier detection, see below). The overall clustering is according to: 'soil unit/climate' cluster, depth range (i.e. five 20 cm thick layers up to 100 cm depth, and two 50 cm thick layers between 100 cm and 200 cm depth), and textural class (five following SOTER conventions, see van Engelen and Dijkshoorn 2013, p. 53).

Table 1.
List of soil variables for which property estimates are presented

Abbreviation[e]	Soil variable
ALSAT	Aluminium saturation (as % of ECEC) [a]
BSAT	Base saturation (as % of CEC_{soil}) [a]
BULK	Bulk density
CECC	Cation exchange capacity of clay size fraction (CEC_{clay}) [ac]
CECS	Cation exchange capacity (CEC_{soil})
CFRAG	Coarse fragments (> 2 mm; volume %)
CLPC	Clay (mass %)
CNrt	C/N ratio[a]
DRAIN	Soil drainage class (observed, according to FAO (2006)
ECEC	Effective cation exchange capacity [ab]
ELCO	Electrical conductivity
ESP	Exchangeable sodium percentage [a]
GYPS	Gypsum content
ORGC	Organic carbon[e]
PHAQ	Soil reaction (pH_{H2O})
SDTO	Sand (mass %)
STPC	Silt (mass %)
TAWC	Available water capacity (from -33 to -1500 kPa; cm m^{-1}) [ad]
TCEQ	Total carbonate equivalent
TEB	Total exchangeable bases
TOTN	Total nitrogen

[a] Calculated from other measured soil properties.

[b] ECEC is defined as exchangeable (Ca^{++}+Mg^{++}+K^{+}+Na^{+}) plus exchangeable (H^{+}+Al^{+++}) (van Reeuwijk 2002).

[c] CEC_{clay} is calculated from CEC_{soil} by assuming a mean contribution of 350 $cmol_c$ kg^{-1} OC, the common range being from 150 to over 750 $cmol_c$ kg^{-1} (Klamt and Sombroek 1988).

[d] Soil water potential limits for AWC conform to USDA standards (Soil Survey Staff 1983); these values have not *yet* been corrected for the presence of fragments > 2 mm.

[e] Units of measurement for all soil properties are given in Appendix 7.

Soil properties often show a skewed distribution. Various transformations can be used to correct for skewing to a greater or lesser extent; the recommended transformation to use will depend on both the direction and extent of skew (Hodge and Austin 2004; McKenzie *et al.* 2008; Snedecor and Cochran 1980). Based on several trials, that took into consideration the large range of soil properties under consideration as well as possible inconsistencies resulting from the soil/climate GIS-overlay (i.e. possible mis-representation of stratum), the following procedure was adopted here to reduce skewness: '10%-Winsorising' (Wilcox 2009, p. 27-29), which provides resistance to outliers, with subsequent detection and rejection of possible outliers based on 'box-whisker' analyses using k=1.5 (Frigge *et al.* 1989). Upon application of this robust procedure, the data distributions were found to be approximately symmetric for most soil properties.

The number of cases rejected for each cluster has been documented in the 'statistical output files', see Appendix 5. For clusters or data combinations at the FAO subunit level undifferentiated by climate, e.g. cases like 'ACfCECc/Y/Diu', of which there are 33,589 cases in total (see files like STAT_ACVR1 in Appendix 5), this information has been summarised in Table 2, corresponding to an average rejection percentage of 2% per cluster. When analysed by individual soil property, these values ranged from 0% for texture (SDTO, CLPC and STPC) up to 6% for ESP and GYPSUM.

Table 2.
Average proportion of cases rejected upon possible outlier-detection by soil property (for data clustered at FAO subunit level)

Soil property[a]	n[b]	Avg	Std
ALSA	339	2.16	4.72
BSAT	1908	2.41	5.43
BULK	1712	1.39	3.89
TCEQ	813	3.40	5.68
CECc	2009	2.15	4.50
CECS	2275	1.78	4.09
CLPC	2285	0.00	0.00
CNrt	1879	2.87	5.44
ECE	1855	2.61	4.91
ESP	1725	5.99	6.80
CFRAG	1768	2.98	6.12
GYPS	182	6.04	6.87
ORGC	2266	3.42	5.34
PHAQ	2264	1.32	3.74
SDTO	2285	0.00	0.00
STPC	2285	0.00	0.00
TAWC	581	0.96	3.24
TEB	1914	2.80	5.22
TOTN	1904	2.73	4.91

[a] Soil properties are listed in alphabetical order; see Table 1 resp. Appendix 8 for details.
[b] Results (%) are for combinations at the FAO subunit level undifferentiated by climate that is cases like 'ACfCECc/Y/Diu', corresponding with 33,589 cases in total (see files like STAT_ACVR1 in Appendix 5 for actual numbers of deleted cases). Typically, some soil properties such as ALSAT or GYPSUM are only reported for selected soil types leading to lower values for *n*.

The following statistics were calculated for the 'normalised' sample populations: sample size (*n*); mean, standard deviation (STD) and standard error (SE); median and median of absolute deviation (MAD); upper and lower quartiles as well as 10% percentiles (where $n > 10$); and minimum and maximum recorded for the given sample population. Values for SE, i.e. STD/SQRT(n), provide a quantitative measure for the uncertainty of the estimated mean(s). The initial number of observations (Num_0) before application of the outlier detection/rejection scheme was also recorded (see Appendix 5).

Finally, mean and STD values as calculated for each 'soil/climate' cluster, depth zone and soil property provided the basis for the actual taxotransfer procedure (see 3.2).

3 Results and Discussion

3.1 General

The compilation of the HWSD, based on source materials at a scale of 1:1M to 1:5M, encompassed a marked degree of data integration. The aim of such broad scale maps being to simplify the geographical distribution of soil types to a regionally representative pattern. Inherently, all soil map units will include a number of impurities, often in excess of 15% (see Landon 1991), which cannot be mapped at the given scale. Further, the exact location of the component soil units within a given map unit —and the corresponding grid cells— is not known, only their estimated proportion (see FAO 1995; FAO *et al.* 2012). Additional uncertainty is associated with the delineation of the original map unit boundaries but, as indicated, this situation cannot be remedied here. Updates of the soil-geographical data for at least those sections of HWSD that are considered to be 'less reliable' are foreseen in the context of GSP, Pillar 4 (FAO-GSP 2014a; Omuto *et al.* 2012).

Substantial changes have occurred in defining soils according to the FAO (1988) Legend and its successor the World Reference Base for Soil Resources (IUSS Working Group WRB 2014). For instance, the re-defined Arenosols, Fluvisols and Regosols, key out at a much lower hierarchical level in the WRB 2014 system than in the Revised Legend, that is below WRB 2014 Cambisols. Definitions for several diagnostic criteria have also changed and three new diagnostic horizons have been defined (IUSS Working Group WRB 2014, p. 2-3) as a result of which correlation with the FAO (1988) system is not straightforward. Strictly speaking, conversion to the WRB 2014 system (legend) may require both remapping of soil boundaries as well as reclassifying the underpinning soil profiles by national/regional soil experts. This type of comprehensive updates may be seen as the responsibility of GSP regional nodes, who may contact the national soil experts for their continents/regions.

This study involved the harmonisation of soil variables that may show a marked spatial and temporal variation; as discussed in section 2.1, these variables have been determined in a range of laboratories, according to various analytical methods, and over a range of years. Other sources and types of uncertainty are associated with the spatial data (map unit boundaries and map unit composition) and aggregation procedures; at present, these cannot be quantified. Possible implications for modelling have been discussed elsewhere (Batjes 1999; Bouwman *et al.* 1999; Cramer and Fischer 1997; Dobos *et al.* 2006; Hengl *et al.* 2014; Middelburg *et al.* 1999).

3.2 Methodological differences with HWSD

Main methodological differences with the HWSD approach are summarised in Table 2. These relate to the data model, number of soil properties under consideration, and the fact that WISE30sec presents quantified estimates of the uncertainty in the predictions as well as a qualitative measure for the 'inferred confidence' in the derived data (Appendix 8).

Table 3.
Comparison of HWSD and WISE30sec procedures

Feature	HWSD [a]	WISE30sec
Map unit composition	HWSD, v1.21	HWSD v1.21, plus some regional updates (see 2.2); also climate (see 2.3)
Grid size[b]	30 by 30 arc second	30 by 30 arc second
Soil depth	Up to 1 m[a]	Up to 2 m
Number of layers	2 (0-0.3, 0.3-1 m) [a]	7 (0-0.2, 0.2-04, 0.4-0.6, 0.6-0.8, 0.8-1, 1-1.5, 1.5-2 m)
Number of textural classes	3 (FAO 1988 conventions)	5 (SOTER conventions)
Soil attributes (per layer)	16[a]	20
Taxotransfer procedure	Procedures vary with the source materials used	Globally consistent procedure; nature of TTR-rules flagged (Appendix 8)
Uncertainty	Not given / mentioned [a]	Mean ± STD; statistical uncertainties discussed
Applications [b]	Continental and global scale	Continental and global scale
Co-variates:		
- Climate	*No* [c]	*Yes (Köppen-Geiger)*
- Land use (management)	*No* [a]	*No (not enough 'measured' data)*

[a] Considered as key constraints of HWSD according to FAO-GSP (2014a, p. 13).

[b] Many applications now require soil data at resolutions that match digital elevation models and remotely sensed imagery for application at subnational levels. Typically 250m or finer, as currently being developed for GlobalSoilMap and SoilGrids (Arrouays *et al.* 2014; Hengl *et al.* 2014).

[c] Considered only in some recent SOTWIS databases represented in HWSD v1.21 (e.g. Batjes 2010).

3.3 Taxotransfer-derived soil properties

Derived soil properties for the components of each 'soil/climate' map unit were derived using taxotransfer rules. Such rules allow to 'estimate soil properties based on modal soil characteristics of soil units, as derived from a combination of their classification name or taxon (which by definition often implies a certain range for a number of properties), expert knowledge and empirical rules, and a statistical analysis of a large number of soil profiles belonging to the same taxon' (see Batjes *et al.* 1997).

Results of the statistical analyses (i.e. mean and STD), after outlier detection and possible rejection per cluster (Appendix 5), provided the quantitative basis for the actual taxotransfer procedure (Appendix 7), which consists of four stages. These are carried out sequentially depending on the size of the original sample population(s). The overall procedure is illustrated below using an example 'FLtALSA/B/D3u': estimated Aluminium saturation (ALSA) at 40-60 cm depth (D3) for Thionic Fluvisols (FLt) mapped in coastal mangroves in an Arid climate (B). As there is no specific information about the soil textural class on the HWSD map, the 'undifferentiated' (u) or modal class is used by default.

If there are enough measured data (defined here as $n \geq 5$, see Table 3) for the given combination of 'soil-climate', textural class, depth layer and soil property, the TTR procedure will use the mean and STD for cluster 'FLtALSA/B/D3u' as 'best estimates' (coded as TTRsub). Else, if $n < 5$, best estimates for the nearest 'broader' cluster, *in casu* 'FLtALSA/Y/D3u' where 'Y' stands for undifferentiated climate, will be considered (TTRsubY). However, if n is still < 5 at this stage (or if soil types have solely been mapped at the major group level) 'best estimates' for the taxonomically closest major soil group and climate will be used (FLALSA/B/D3u; TTRmain). Finally, provided $n \geq 5$, 'best estimates' for the closest major soil group and 'undifferentiated' climate will be used (FLALSA/Y/D3u; TTRmainY).

Based on the present selection and analyses of soil analytical data a large number of TTRs could be defined: 39,108 rules for TTRsub, 18,642 rules for TTRsubY, 13,956 for TTRmain, and 4,691 for TTRmainY. Nonetheless, at the end of the TTR application there were still gaps for some soil attributes (values coded as '-8'). In such instances, where possible, 'best estimates' were allocated using expert-judgement (using results from auxiliary analyses of the whole data set). These automated expert rules (EXR) also served to check for in-pedon consistency. For example, if pH_{water} is less than 6.5 in a given layer the content of calcium carbonate should be zero. Similarly, there should be no exchangeable acidity in basic soils.

Results of the taxotransfer procedure are stored in table *HW30S_ParEst* (Appendix7) resp. *HW30S_Full* (Appendix 9); these tables include quantitative measures for the uncertainty of the predictions. The actual steps, or TTR and EXR 'lineage', are documented in a separate table, called *Log_TTRexr* (Appendix 8). Each flag consists of a sequence of letters followed by a numeral as illustrated in Figure 1; see Appendix 8 for soil property codes. The number code reflects the size of the sample population that underpins the TTR applied (Table 4). Use of a capital indicates that the substitution is based on the whole set for the corresponding soil unit/climate and depth layer, irrespective of soil texture (i.e. 'u' or undifferentiated); conventions for coding the soil attributes are listed in Appendix 8). The flags provide an overall (qualitative) measure for the possible level of confidence in the derived data, by depth layer. Overall this confidence will decrease from TTR stage 1 to stage 4 and, within of these stages, with the number of observations available for the cluster under consideration. Quantitatively, this is expressed by the standard error (SE) as listed in the tables described in Annex 5 for the respective cluster_id's.

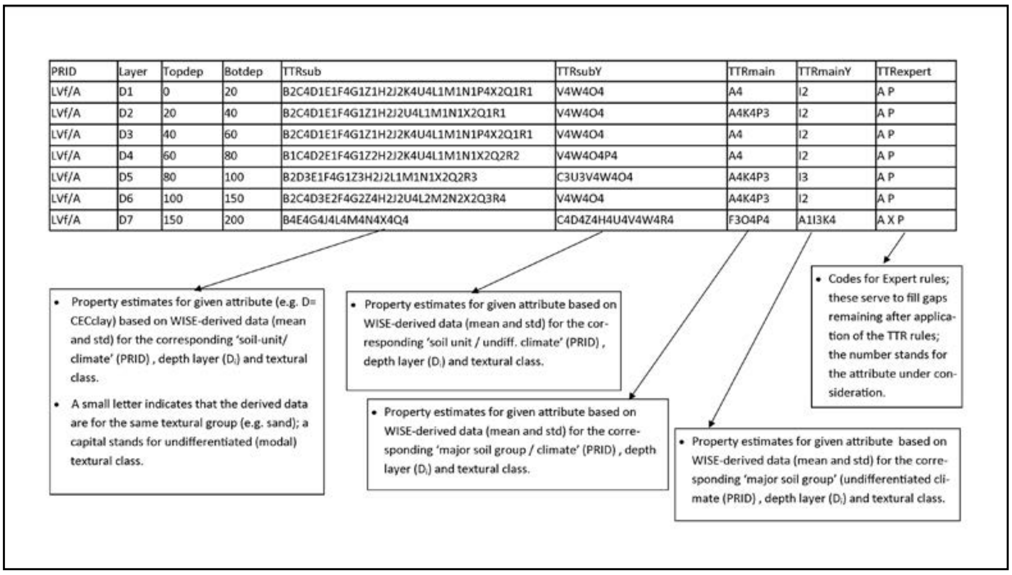


Figure 1.
Flagging of taxotransfer and expert rules for a hypothetical profile (defined by soil type and climate) by depth zone, soil textural class and soil property

Table 4.
Criteria for defining 'inferred confidence' in the derived data

Code	Confidence level	n_{WISE} [a]
1	Very high	≥ 60
2	High	30-60
3	Moderate	15-29
4	Low [b]	5-14
5	Very low	< 5

[a] n_{WISE} is the sample size after outlier detection and rejection.
[b] The cut-off point for applying any TTR is $n_{WISE} < 5$ (see Batjes *et al.* 1997).

For some cases, it was not yet possible to present 'best estimates' using the present set of TTRs and EXRs (implying the need for collation of additional soil profiles for several FAO soil units); as indicated, these have been flagged with a negative value (-8) in the database. Similarly, for miscellaneous units, such as rock outcrops or land ice, specific negative values have been inserted in the database, mainly to facilitate GIS visualisation:

- WR, Water bodies: -1
- GG, Land ice and glaciers: -2
- RK, Rock outcrops: -3
  (Note: -7 is used for the rocky sublayers of shallow soil units, such as Leptosols)
- DS, Dunes/Shifting sands: -4
- ST, Salt flats: -5
- Other miscellaneous units, including fish ponds (FP), humanly disturbed (HD), small islands (IS), and urban (UR) areas: -9

3.4 Sources of uncertainty

The calculation and mapping of derived soil properties at the global level remains fraught with uncertainty, being largely dependent on the quality and resolution of the soil-geographical data and environmental covariates, as well as the selection/distribution/quality of the underpinning soil (point) profile data and their locational accuracy. Even with geostatistical methods (Hengl *et al.* 2014), prediction accuracies at 1km resolution assessed using 5–fold cross-validation were between 23–51% (point-based). Similarly, with the present mapping approach the variation in soil properties observed within each map unit remains large (see Appendix 9). This is not unexpected; the natural range in measured chemical and physical properties can be considerable at the level of the soil unit, or even at that of the soil series, with coefficients of variation (CV) often exceeding 40% (Beckett and Webster 1971; Burrough 1993; Landon 1991; Spain *et al.* 1983). Similarly, the median of absolute deviations (MADs) —the median of the differences between each observation and the median— are often large: additional information on the data distribution for each data cluster is provided in the form of 10%-percentiles (Appendix 5).

The overall assumption in this assessment has been that the confidence in a TTR-based soil property estimate should increase with the size of the underpinning sample population. In addition, in principle, the confidence in soil property estimates derived using TTRsub type procedures should be higher than for those derived using TTRsubY, TTRmain and TTRmainY type procedures respectively. Nonetheless, a high (inferred) confidence rating does not necessarily imply that the soil property estimates shown will be representative for the data cluster under consideration. Profile selection for the WISE profile database, as for most other large historic databases (e.g. ESB 2001; Leenaars *et al.* 2014; NCSS 2010), is not probabilistic but based on available data and purposive sampling. Similarly, the soil profiles that underpin broad scale soil property maps derived from digital soil mapping, such as SoilGrids1km (Hengl *et al.* 2014), are largely based on purposive sampling.

Results may also be biased for those soil properties that were recorded as 'not observed' or 'nil' in the original surveys, for example volumetric gravel content (*CFRAG*). In such cases, derived properties computed using the TTRs may well give a biased impression of 'modal' conditions for some soil units. For example, for soil units that are generally devoid of coarse fragments, field surveyors may only have noted limiting gravel (> 2mm) contents as being 'relevant' for land suitability evaluation. Further, property estimates for the deeper soil layers (say below 80 cm) are generally based on a smaller number of samples, as such layers are less frequently sampled during routine soil survey, thus considered to be 'less reliable'.

Several soil chemical properties —such as high aluminium saturation in parts of the subsoil or a high exchangeable sodium percentage in parts of the topsoil— may be 'levelled out' during depth weighting. When occurring, this will also be reflected in derived values obtained through pedotransfer. In addition, some of the soil attributes under consideration, such as the presence of fragments > 2 mm and water holding capacity, are not diagnostic in the FAO Legend, but they should have similar values for broad soil types (e.g. Regosols versus Ferralsols).

Some soil processes and properties are readily modified by changes in land use and management. For example, soil pH and aluminium saturation upon liming; salinity and electrical conductivity upon irrigation or soil drainage; soil organic matter quantity and quality upon changes in land use, nutrient and water management, tillage practices or climate change. Such effects, however, cannot be considered explicitly when analysing the available profile data. In particular, the information on land use and management (history) for the available soil profiles is limited. Alternatively, in principle, the present data set could be used to define different *phenoforms* for main soil types (or *genoforms* as represented by their 'modal' soil properties) along the lines proposed by Droogers and Bouma (1997) and Bouma *et al*. (1998).

Further, as indicated earlier, there are uncertainties in the spatial (soil-geographical) data. For example, in the delineation and naming of the various map units, as well as in the estimated relative proportion for each component soil unit within a given map unit. Consideration of this type of errors, which are known to be significant (see Batjes 1999; Omuto *et al.* 2012), and their propagation during data processing, is beyond the scope of this study as this may require some re-mapping.

3.5 Appropriate data use

There are various options to display or spatially aggregate derived soil data when dealing with compound map units, each of these having their strengths and limitations (e.g., Batjes 2006; Carter and Scholes 2002; FAO 1995; Kern 1995). Generally, the type of research purpose will determine which soil property estimates or class intervals (for binned data sets) will be required for a specific application. The necessary data selections can best be made using tailor-made programs designed to meet the scope of these applications*; these programs should consider the full map unit (grid cell) composition* and relevant depth intervals (Appendix 9). This aspect is illustrated below using map unit 'WD50012250' in the boreal zone as an example for organic carbon content, which involves analysis of data on SOC concentration, bulk density and proportion of coarse fragments for defined soil depths (Batjes 1996b). Table 5 shows estimates of SOC content (kg C m^{-2} to 20 cm depth) that are obtained when (1) the full map unit composition, (2) area-weighted values, and (3) only the spatially dominant soil unit are considered in the calculations. For this example, these estimates range from 14.2 kg C m^{-2} for 'approach 1' to 6.8 kg C m^{-2} for 'approach 3'.

Table 5.
Calculations should consider the full map unit composition –computation of organic carbon content as an example

Approach	Soil component	Proportion (%)	CFRAG (%)	BULK (g cm^{-3})	ORGC (g kg^{-1})	C stock[a] (kg m^2)
1) Full map-unit composition	1 – LPq	50	22	1.37	31.6	6.75
	2 – CMu	30	18	1.03	59.41	10.03
	3 – HSi	20	5	0.42	292.5	23.34
						14.22
2) Area-weighted	3 – X	100	17	1.08	92.1	*16.5*
3) Dominant unit	1 – LPq	'100'	22	1.37	31.6	*6.75*

[a] For 0 to 20 cm depth (D1): CFRAG proportion of fragments > 2mm; BULK is bulk density, ORGC is organic carbon. Example is for mapping unit 'WD50012250', the corresponding data are held in file HW30S_Full, see Appendix 9.

Nonetheless, for ease of 'rapid' visualization using GIS, generalisations can be useful so that one single set of derived soil property values can be assigned to a given grid or map unit (for a given depth layer). For this purpose *only*, so-called 'summary' files holding area-weighted soil property estimates for each map unit (and soil layer) have been included in the data set (see Appendix 10). For good order, as illustrated above, such area-weighted files should not be used/combined in GIS to 'rapidly' calculate e.g. soil carbon stocks as this may lead to erroneous results.

Summarising, assessment of the accuracy and applicability of any data set for a specific purpose remains a user responsibility. The issue of scale is particularly important in this respect (e.g., Finke 2006). The soil polygon maps that underpin this study are at scale of 1:1 to 1:5 M. By implication, as indicated earlier, HWSD's gridding at 30 by 30 arc-seconds resulted in multiple grid cells with identical soil units occurring in individual map units (polygons) as presented on the original vector maps (FAO *et al.* 2012, p.2). This is a simplification since the actual location of the component soil units in a given map unit is not known, only their estimated relative proportion. Further, it should be noted that it is better to aggregate model results — for a given spatial or temporal unit — rather than to aggregate the data before modelling (Bouwman *et al.* 1999; Middelburg *et al.* 1999).

3.6 Example of application: Soil carbon stocks

Terrestrial ecosystems, and their soils, are key components in the global carbon cycle. Considering the full map unit composition and calculation procedure described in Batjes (1996b) global stocks were calculated for soil organic carbon (SOC) for the following depth intervals: 0-30 cm, 0-50 cm, 0-100 cm, 0-150 cm and 0-200 cm. Uncertainties in the estimated SOC content for these depth layers were calculated first using error propagation procedures as described in Ku (1966), necessarily assuming independency of variables (i.e. parameter estimates for SOC, CFRAG and BULK). Possible variation in the SOC stock for each map unit to a given depth (e.g. 0-30 cm, the default used by the IPCC (2006), as associated for example with possible regional differences in land use and management, was simulated next to put bounds on global SOC stocks rather than presenting a single figure. For this, 400 (randomised) simulation runs (for each NEWSUID) were used to emulate the possible variation in soil conditions within a soil map using the computed values for mean and STD (based on full map unit composition), and possible outer limits of 'mean ± 2xSTD'. This approach is recognised to be 'conservative' in the sense that is does not assume a normal probability distribution (e.g. Kempen 2011), which would lead to more stringent confidence limits. The simulated variability in SOC values within each map unit will reflect both variations in the soil type and land use/management, as well as those associated with the methods of sampling and measurement.

The information resulting from the 400 simulations was linked to the soil geographical information to arrive at *n* realizations of world SOC stocks. The resulting 95%-confidence interval for the mean SOC stock for the world, up to a depth of 2 m, is listed in Table 4. These figures are in line with earlier estimates which tend to converge to some 1400 to 1600 Pg C (1 Pg = 10^{15} g) to 1 m depth and 1990-2460 Pg C to 2 m depth (see Govers *et al.* 2012), although there is still considerable variation between studies that consider measured soil data (Scharlemann *et al.* 2014) respectively modelled data (Tian *et al.* 2015; Todd-Brown *et al.* 2013). Alternatively, important differences in the regional distribution of SOC content (Figure 2) are reported in various studies (Hugelius *et al.* 2014b; Johnson *et al.* 2011; Scharlemann *et al.* 2014; Tian *et al.* 2015). Hence the persevering need for updating the information on world soil resources at various resolutions (e.g. Anon. 2011; Arrouays *et al.* 2014; FAO-GSP 2014a; Hengl *et al.* 2015; Hugelius *et al.* 2014a).

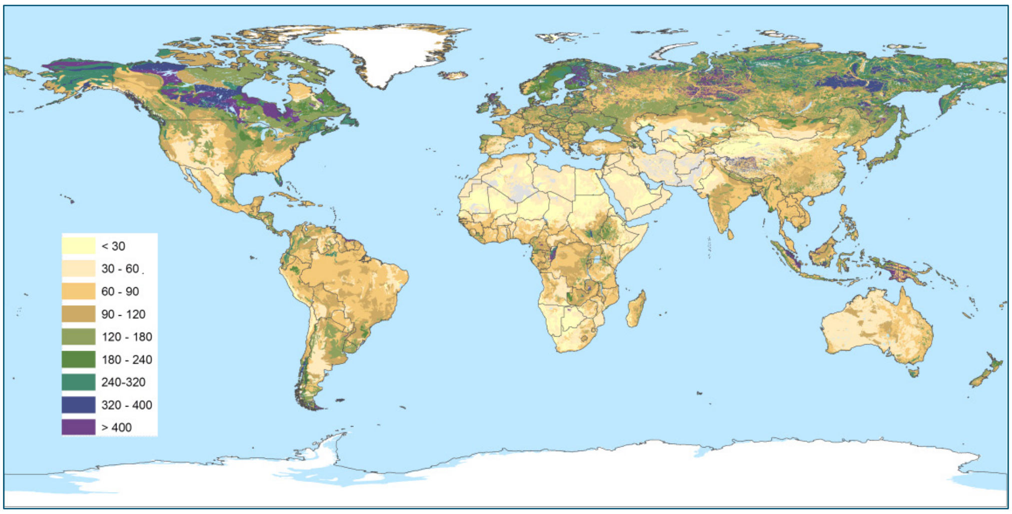


Figure 2.
Soil organic carbon content to 1m depth (Mg C ha^{-1})

Table 6.
95%-confidence limits for worldwide stocks of soil organic carbon

Variable[a]	Stocks (Pg C to specified depth)			
	95%Lo	Mean	95%Up	STD
SOC_P1	743	755	768	119
SOC_P2	979	993	1006	129
SOC_P3	1392	1408	1424	154
SOC_P4	1757	1778	1798	198
SOC_P5	2038	2060	2082	217

[a] SOC stands for soil organic carbon. Coding for depth layers: 0-30 cm (P1), 0-50 cm (P2), 0-100 cm (P3), 0-150 cm (P4) and 0-200 cm (P5). Pg, Petagram= 10^{15} g. 95%Lo and 95%Up are 95%-confidence limits for SOC stocks based on 400 simulation runs, see text for assumptions.

Based on analyses of the WISE30sec database, some 30% (601 ± 87 Pg C) of the total SOC stock (2060 ± 215 Pg C, Table 6) to 2 m depth is held in the Northern Circumpolar Region (as defined by http://bolin.su.se/data/ncscd/), the organic soil carbon reserves of which are considered most sensitive to climate change. For 0-0.3 m and 0-1 m these estimates are 223 ± 52 Pg C and 411 ± 65 Pg C respectively. These estimates are in line with those of Tarnocai *et al.* (2009) who reported estimates of 191 Pg C (0-0.3 m), 496 Pg C (0-1 m) and 1024 Pg C (0-3 m) for SOC stocks for the northern circumpolar permafrost region. Alternatively, Hugelius *et al.* (2014b), reported summarised SOC stocks (mean ± 95% confidence limits) of 827 ± 108 Pg C for 0-2 m resp. 1035 ± 150 Pg C to 3 m depth for the northern circumpolar permafrost region using different upscaling techniques; their study drew on more recent soil geographical data than are presently considered in the HWSD.

4 Conclusions

WISE30sec supersedes earlier versions of WISE-derived GIS datasets that still drew on the spatial data of the now largely outdated Digital Soil Map of the World (FAO 1995) and a more limited selection of harmonised soil profiles (see Appendix 12). Being largely automated, new releases of the WISE30sec database may be produced as updated soil-geographical and/or additional soil profiles become freely available, for example through the Global Soil Partnership and similar.

There are often gaps or omissions in the information provided in the soil literature or in auxiliary databases with respect to several of the input variables considered in the present analyses. As such, the set of derived soil property values presented here should be seen as 'best possible' estimates as derived from the current selection of measured soil profile data, scheme of taxotransfer and expert-rules, and spatial data. The overall assumption is that the inferred confidence in the soil property estimates presented for a given combination of 'soil/climate unit, soil property, depth zone, and soil textural class' should increase with the size of the underpinning sample population. Nonetheless, the statistical uncertainty attached to individual soil property estimates can be large; uncertainty estimates are provided as standard deviations, standard errors, and 10% percentiles, critical information that is not available for the original HWSD and DSMW products.

Changes in the number, spatial distribution and type of profiles analysed as well as differences in data clustering and analysis procedures, used for the attribute and spatial data, will lead to different property estimates (and binned maps) for any given soil variable. Such changes may or may not be significant depending on the soil property under consideration.

WISE30sec is considered appropriate for exploratory assessments at a broad scale (< 1:1 M; 30 by 30 arc-second or coarser), keeping in mind the generalisations, assumptions, uncertainties and recommendations for 'appropriate data use' as described in this report.

An important application of the present dataset will be to carry out 'cross checks' against predictions derived from digital soil mapping at 1 km resolution for the globe. Such comparative analyses will allow identification of possible discrepancies, and their sources, and thereby should provide a basis for further refinements to both approaches in terms of data collation and methodology development. Such complementary activities may be envisaged, for example, within the broader sponsored-framework of the Global Soil Partnership as they will require international collaboration and free sharing of source data, elements that are essential to create 'international ownership' of any derived global soil product.

Acknowledgements

This study was carried out at ISRIC - World Soil Information (WDC-Soils) in support of the International Year of Soils 2015, the aim of which is 'to be a platform for raising awareness of the importance of soils for food security and essential eco-system functions'. It builds on collaborative work with FAO and IIASA initiated in the framework of the WISE-project and draws on soil profile data shared by a wide range of institutions and experts.

Special thanks are expressed to Jennifer Harden (Scientist Emeritus, U.S. Geological Survey) and Gustaf Hugelius (Stockholm University, Sweden) for sharing soil profile data from the so far underrepresented Northern Circumpolar Region. I am also grateful to Jennifer for having initiated collaboration with the International Soil Carbon Network (ISCN). I also thank my colleagues Bas Kempen, Gerard Heuvelink and Johan Leenaars for constructive editorial comments, and Maria Gonzalez Ruiperez for GIS support.

References

Anon. 2011. Evolving the IPCC. *Nature* 1, 227-227

Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GBM, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, Mendonca-Santos MdL, Minasny B, Montanarella L, Odeh IOA, Sanchez PA, Thompson JA, Zhang G-L and Donald LS 2014. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. *Advances in Agronomy* Volume 125, 93-134

Batjes NH 1996a. *Documentation to ISRIC-WISE global data set of derived soil properties on a ½ by ½ grid*. Work. Pap. 96/05, ISRIC, Wageningen, 12 p *http://www.isric.org/isric/webdocs/docs/isric_wp96_05.pdf*

Batjes NH 1996b. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science* 47, 151-163

Batjes NH 1999. Soil vulnerability mapping in Central and Eastern Europe: Issues of data acquisition, quality control and sharing. In: Naff T (editor), *Data Sharing for International Water Resource Management: Eastern Europe, Russia and the CIS*. NATO Science Series 2: Environmental Security (Vol. 61). Kluwer Academic Publishers, Dordrecht, pp 187-206

Batjes NH 2002a. *ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid (ver. 2.0)*. Report 2002/03, International Soil Reference and Information Centre (ISRIC), Wageningen, 14 p *http://www.isric.org/isric/webdocs/docs/ISRIC_Report_2002_03.pdf*

Batjes NH 2002b. *Soil parameter estimates for the soil types of the world for use in global and regional modelling (Version 2.1)*. ISRIC Report 2002/02c, International Food Policy Research Institute (IFPRI) and International Soil Reference and Information Centre (ISRIC), Wageningen. Available at: http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2002_02c.pdf; accessed 12/2008., 58 p

Batjes NH 2005. *ISRIC-WISE global data set of derived soil properties on a 0.5 by 0.5 degree grid (ver. 3.0)*. Report 2005/08, ISRIC - World Soil Information, Wageningen 20 p *http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2005_08.pdf*

Batjes NH 2006. *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.0)*. Report 2006/02, ISRIC - World Soil Information, Wageningen, 50 p *http://www.isric.org/isric/webdocs/Docs/ISRIC_Report_2006_02.pdf*

Batjes NH 2009. Harmonized soil profile data for applications at global and continental scales: updates to the WISE database. *Soil Use and Management* 25, 124-127

Batjes NH 2010. *Soil property estimates for Tunisia derived from SOTER and WISE (SOTWIS-Tunisia, ver. 1.0)*. Report 2010/01, ISRIC - World Soil Information, Wageningen, 41 p *http://www.isric.org/isric/Webdocs/Docs/ISRIC_Report_2010_01.pdf*

Batjes NH 2011. Soil organic carbon stocks under native vegetation - revised estimates for use with the simple assessment option of the Carbon Benefits Project system. *Agriculture, Ecosystems & Environment* 142, 365-373

Batjes NH 2012. *ISRIC-WISE derived soil properties on a 5 by 5 arc-minutes global grid (ver. 1.2)*. Report 2012/01, ISRIC - World Soil Information, Wageningen, 57 p *http://www.isric.org/sites/default/files/isric_report_2012_01.pdf*

Batjes NH, Bridges EM and Nachtergaele FO 1995. World Inventory of Soil Emission Potentials: development of a global soil data base of process-controlling factors. In: Peng S, KT Ingram, HU Neue and LH Ziska (editors), *Climate Change and Rice*. Springer-Verlag, Heidelberg, pp 102-115

Batjes NH, Bindraban PS and Reuter HI 2013. *Towards improved soil information for quantification of environmental, societal and economic sustainability*. Report 2013/05, ISRIC - World Soil Information, Wageningen, 44 p *http://dx.doi.org/10.13140/2.1.4831.0560*

Batjes NH, Fischer G, Nachtergaele FO, Stolbovoy VS and van Velthuizen HT 1997. *Soil data derived from WISE for use in global and regional AEZ studies (ver. 1.0)*. Interim Report IR-97-025, FAO/ IIASA/ ISRIC, Laxenburg, 27 p *http://www.iiasa.ac.at/Admin/PUB/Documents/IR-97-025.pdf*

Batjes NH, Al-Adamat R, Bhattacharyya T, Bernoux M, Cerri CEP, Gicheru P, Kamoni P, Milne E, Pal DK and Rawajfih Z 2007. Preparation of consistent soil data sets for SOC modelling purposes: secondary SOTER data sets for four case study areas. *Agriculture, Ecosystems and Environment* 122, 26-34

Beckett PHT and Webster R 1971. Soil variability - a review. *Soils and Fertilizers* 34, 1-15

Bouma J and McBratney A 2013. Framing soils as an actor when dealing with wicked environmental problems. *Geoderma* 200-201, 130-139

Bouma J, Batjes NH and Groot JJR 1998. Exploring land quality effects on world food supply. *Geoderma* 86, 43-59

Bouma J, Batjes NH, Sonneveld † MPW and Bindraban PS 2014. Enhancing soil security for smallholder agriculture. In: Lal R and BA Stewart (editors), *Soil Management of Smallholder Agriculture*. Advances in Soil Science. CRC Press, Boca Raton (FL), pp 17-37

Bouwman AF, Derwent RG and Dentener FJ 1999. Towards reliable global bottom-up estimates of temporal and spatial patterns of emissions of trace gases and aerosols from land-use related and natural resources. In: Bouwman AF (editor), *Approaches to Scaling of Trace Gas Fluxes in Ecosystems*. Elsevier, Amsterdam, pp 3-26

Bridges EM, Batjes NH and Nachtergaele FO (editors) 1998. World Reference Base for Soil Resources: Atlas. ISSS Working Group RB, ISRIC-FAO-ISSS-Acco. Acco Press, Leuven, 79 p

Burrough PA 1993. Soil variability: a late 20th century view. *Soils and Fertilizers* 56, 529-562

Carter AJ and Scholes RJ 2002. *Generating a global database of soil properties (Available on: Global Soil Data Products, CD-ROM (IGBP-DIS), USDA, CSIR, IGBP-DIS, FAO and ISRIC)*, IGBP Data and Information Services, Potsdam

Carvalho Ribeiro E, Mendes de Jesus J, Van Oostrum A, Leenaars JGB and Batjes NH 2015 (in prep.). *ISRIC World Soil Information Service (WoSIS, version 2.0)*, ISRIC - World Soil Information, Wageningen

Cramer W and Fischer A 1997. Data requirements for global terrestrial ecosystem modelling. In: Walker B and W Steffen (editors), *Global Change and Terrestrial Ecosystems*. Cambridge University Press, Cambridge, pp 529-565

De Paul Obade V and Lal R 2013. Assessing land cover and soil quality by remote sensing and geographical information systems (GIS). *CATENA* 104, 77-92

Dobos E, F. C, Hengl T, Reuter HI and Toth G 2006. *Digital soil mapping as a support to production of functional maps*. EUR 22123 EN, Office for Official Publications of the European Communities, Luxembourg, 68 p

Droogers P and Bouma J 1997. Soil survey input in exploratory modeling of sustainable management practices. *Soil Science Society of America Journal* 61, 1704-1710

ESB 2001. *Georefenced soil database for Europe, Manual of Procedures (version 1.0)*, European Soil Bureau (on behalf of the Contributing Organisations), Office for Official Publications of the European Communities, Luxembourg (available at: http://eusoils.jrc.ec.europa.eu/esdb_archive/eusoils_docs/esb_rr/n05_ManualVer11EN4.pdf), 178 p

ESRI 2015. ARGIS Resource Centre (Desktop 10) *http://help.arcgis.com/en/arcgisdesktop/10.0/help/*

FAO-GSP 2014a. *Plan of action for Pillar Four of the Global Soil Partnership : Building an enduring and authoritative global system to monitor and forecast the condition of the Earth's soil resources* FAO Global Soil Partnership, Rome, 33 p *http://www.fao.org/fileadmin/user_upload/GSP/docs/plenary_assembly_II/pillar4.pdf*

FAO-GSP 2014b. *Plan of Action for Pillar Five of the Global Soil Partnership: Harmonization of methods, measurements and indicators for the sustainable management and protection of soil resources*, FAO Global Soil Partnership, Rome, 37 p *http://www.fao.org/fileadmin/user_upload/GSP/docs/ITPS_Pillars/annexV_pillar5.pdf*

FAO-Unesco 1971-1981. *FAO-Unesco Soil Map of the World, 1:5,000,000 (Vol. 1 to 10)*. United Nations Educational, Scientific, and Cultural Organization, Paris

FAO-Unesco 1974. *Soil Map of the World, 1:5 000000, Volume 1 - Legend*, United Nations Educational, Scientific, and Cultural Organization, Paris

FAO 1988. *FAO-Unesco Soil Map of the World, Revised Legend, with corrections and updates*. World Soil Resources Report 60, FAO, Rome; reprinted with updates as Technical Paper 20 by ISRIC, Wageningen,1997, 140 p

FAO 1995. *Digital Soil Map of the World and derived properties (ver. 3.5)*. FAO Land and Water Digital Media Series # 1, FAO, Rome, 23 p *http://www.fao.org/geonetwork/srv/en/metadata.show?id=14116*

FAO 2006. *Guidelines for soil description (Fourth ed.)*, FAO, Rome, 97 p *ftp://ftp.fao.org/agl/agll/docs/guidel_soil_descr.pdf*

FAO, IIASA, ISRIC, ISSCAS and JRC 2012. *Harmonized World Soil Database (version 1.2)*, Food and Agriculture Organization of the United Nations, International Institute for Applied Systems Analysis, ISRIC - World Soil Information, Institute of Soil Science - Chinese Academy of Sciences, Joint Research Centre of the European Commission, Laxenburg *http://www.iiasa.ac.at/Research/LUC/External-World-soil-database/HWSD_Documentation.pdf*

Finke P 2006. Quality assessment of digital soil maps: producers and users perspectives. In: Lagacherie P, A McBratney and M Voltz (editors), *Digital soil mapping: An introductory perspective*. Elsevier, Amsterdam, pp 523-541

Frigge M, Hoaglin DC and Iglewicz B 1989. Some implementations of the boxplot. *The American Statistician* 43, 50-54

Giri C, Pengra B, Long J and Loveland TR 2013. Next generation of global land cover characterization, mapping, and monitoring. *International Journal of Applied Earth Observation and Geoinformation* 25, 30-37

GlobalSoilMap.net 2013. *Specifications Version 1 GlobalSoilMap.net products Release 2.1 http://www.globalsoilmap.net/system/files/GlobalSoilMap_net_specifications_v2_0_edited_draft_Sept_2011_RAM_V12.pdf*

Govers G, Merckx R, van Ost L and van Wesemael B 2012. *Managing Soil Organic Carbon for Global Benefits: A STAP Technical Report.*, Global Environmental Facility, Washington, DC, 70 p *http://www.thegef.org/gef/sites/thegef.org/files/publication/STAP-SOC-Report-lowres-1.pdf*

Gray JM, Humphreys GS and Deckers JA 2009. Relationships in soil distribution as revealed by a global soil database. *Geoderma* 150, 309–323

Harden JW, Koven CD, Ping C-L, Hugelius G, David McGuire A, Camill P, Jorgenson T, Kuhry P, Michaelson GJ, O'Donnell JA, Schuur EAG, Tarnocai C, Johnson K and Grosse G 2012. Field information links permafrost carbon to physical vulnerabilities of thawing. *Geophysical Research Letters* 39, L15704

Hengl T, Mendes de Jesus J, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro EC, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG and Gonzalez MR 2014. SoilGrids1km— global soil information based on automated mapping. *PLoS ONE* 9, e105992

Hengl T, Heuvelink GMB, Kempen B, Leenaars JGB, Walsh MG, Shepherd K, Sila A, MacMillan RA, Mendes de Jesus J, Tamene L and Tondoh JE 2015. Mapping soil properties of Africa at 250 m resolution: random forests significantly improve current predictions. *PLOS One*, doi: 10.1371/journal.pone.0125814 (in press)

Hodge VJ and Austin J 2004. A survey of outlier detection methodologies. *Artificial Intelligence Review* 22, 85-126

Hugelius G, Strauss J, Zubrzycki S, Harden JW, Schuur EAG, Ping C-L, Schirrmeister L, Grosse G, Michaelson GJ, Koven CD, O'Donnell JA, Elberling B, Mishra U, Camill P, Yu Z, Palmtag J and Kuhry P 2014a. Estimated stocks of circumpolar permafrost carbon with quantified uncertainty ranges and identified data gaps. *Biogeosciences* 11, 6573-6593

Hugelius G, Strauss J, Zubrzycki S, Harden JW, Schuur EAG, Ping CL, Schirrmeister L, Grosse G, Michaelson GJ, Koven CD, O'Donnell JA, Elberling B, Mishra U, Camill P, Yu Z, Palmtag J and Kuhry P 2014b. Improved estimates show large circumpolar stocks of permafrost carbon while quantifying substantial uncertainty ranges and identifying remaining data gaps. *Biogeosciences Discuss.* 11, 4771-4822

IPCC 2006. *IPCC Guidelines for National Greenhouse Gas Inventories Volume 4: Agriculture, Forestry and other Land Use*. IPCC National Greenhouse Gas Inventories Programme, Hayama (JP)

ISCN 2014. *ISCN Database Reports*, International Soil Carbon Network *http://iscn.fluxdata.org/Pages/default.aspx*

IUSS Working Group WRB 2014. *World Reference Base for soil resources 2014 - International soil classification system for naming soils and creating legends for soil maps*. World Soil Resources Reports 106, Global Soil Partnership, International Union of Soil Sciences, and Food and Agriculture Organization of the United Nations, Rome, 182 (http://www.fao.org/3/a-i3794e.pdf) p *http://www.fao.org/3/a-i3794e.pdf*

Jenny H 1941. *Factors of Soil Formation: A System of Quantified Pedology*. McGraw-Hill, New York (Reprinted in 1994 by Dover Publications, Mineaola, NY)

Johnson KD, Harden J, McGuire AD, Bliss NB, Bockheim JG, Clark M, Nettleton-Hollingsworth T, Jorgenson MT, Kane ES, Mack M, O'Donnell J, Ping C-L, Schuur EAG, Turetsky MR and Valentine DW 2011. Soil carbon distribution in Alaska in relation to soil-forming factors. *Geoderma* 167-168, 71-84

Jones A, Montanarella L, Jones R and others 2005. *Soil Atlas of Europe*. European Soil Bureau Network, Office for Offical Publications of the Eureopean Communities, Luxembourg, 128 p

Jones A, H. B-M, Brossard M, Dampha A, Deckers J, Dewitte O, Gallali T, Hallett S, Jones R, Kilasara M, Le Roux P, Micheli E, Montanarella L, Spaargaren O, Thiombiano L, Van Ranst E, Yemefack M and R. Z (editors) 2013. Soil Atlas of Africa. European Commission, Publications Office of the European Union, European Union, Luxembourg, 176 p

Kempen B 2011. Updating soil information with digital soil mapping. PhD Thesis, Wageningen University, Wageningen, 218 pp *http://edepot.wur.nl/187198*

Kern JS 1995. Geographic patterns of soil-water holding capacity in the contiguous United States. *Soil Sci. Soc. Am. J.* 59

Klamt E and Sombroek WG 1988. Contribution of organic matter to exchange properties of Oxisols. In: Beinroth FH, MN Camargo and H Eswaran (editors), *Classification, characterization and utilization of Oxisols. Proc. of the 8th International Soil Classification Workshop (Brazil, 12 to 23 May 1986)*. Empresa Brasileira de Pesquisa Agropecuaria (EMBRAPA), Soil Management Support Services (SMSS) and University of Puerto Rico (UPR), Rio de Janeiro, pp 64-70

Ku HH 1966. Notes on the use of propagation of error formulas. *Journal of Research (of the National Bureau of Standards, Washington D.C.)* 70C, 263-273

Landon JR 1991. *Booker Tropical Soil Manual*. Longman Scientific & Technical, New York, 474 p

Leenaars JGB, Kempen B, van Oostrum AJM and Batjes NH 2014. Africa Soil Profiles Database: a compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa In: Arrouays D, N McKenzie, J Hempel, AR de Forge and A McBratney, (editors), *GlobalSoilMap - Basis of the global spatial information system*. Taylor & Francis Group, Orléans pp 51-57 *http://books.google.nl/books?hl=en&lr=&id=S5ClAgAAQBAJ&oi=fnd&pg=PA51&ots=ilrpyldHtj&sig=8KsouYF6P23GR5R1UQ2oaV6U_30&redir_esc=y#v=onepage&q&f=false*

McKenzie N, Grundy MJ, Webster R and Ringrose-Voase AJ 2008. *Guidelines for surveying soil and land resources (3nd Ed.)*. CSIRO Publishing, Collingwood (AUS), 547 p

MEA 2005. *Ecosystems and Human Well-Being: Synthesis, Millennium Ecosystem Assessment* Island Press, Washington

Michaelson GJ, Ping C-L and Clark M 2013. Soil pedon carbon and nitrogen data for Alaska: An analysis and update. *Open Journal of Soil Science* 3, 132-142

Middelburg JJ, Liss PS, Dentener FJ, Taminski T, Kroeze C, Malingreau J-P, Noväl M, Panikov NS, Plant R, Starink M and Wanninkhof R 1999. Relations between scale, model approach and model parameters. In: Bouwman AF (editor), *Approaches to Scaling of Trace Gas Fluxes in Ecosystems*. Elsevier, Amsterdam, pp 219-232

Nachtergaele FO, Van Engelen VWP and Batjes NH 2012. Qualitative and quantitative aspects of world and regional soil databases and maps. In: Pan Ming Huang, Yuncong Li and ME Sumner (editors), *Handbook of Soil Sciences: Resource management and environmental aspects (2nd ed.)*. CRC Press, Boca Raton, pp 27-1/10

NCSS 2010. *National Cooperative Soil Survey. National Cooperative Soil Characterization Database (available at: http://ssldata.nrcs.usda.gov/)*

Omuto C, Nachtergaele F and Vargas Rojas R 2012. *State of the Art Report on Global and Regional Soil Information: Where are we? Where to go?*, FAO, Italy, 69 p *http://www.fao.org/docrep/017/i3161e/i3161e.pdf*

Peel MC, Finlayson BL and McMahon TA 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci. Discuss* 11, 1633-1644

Pleijsier K 1989. Variability in soil data. In: Bouma J and AK Bregt (editors), *Land Qualities in Space and Time*. PUDOC, Wageningen, pp 89-98

Ravindranath NH and Ostwald M 2008. *Carbon inventory methods - Handbook for greenhouse gas inventory, carbon mitigation and roundwood production projects*. Advances in Global Change Research, Volume 29. Springer, Heidelberg, 304 p

Scharlemann JPW, Tanner EVJ, Hiederer R and Kapos V 2014. Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management* 5, 81-91

Snedecor GW and Cochran WG 1980. *Statistical Methods (7th. ed.)*. The Iowa State University Press, Iowa, 507 p

Soil Survey Staff 1983. *Soil Survey Manual (rev. ed.)*. United States Agriculture Handbook 18, USDA, Washington

Spaargaren OC and Batjes NH 1995. *Report on the classification into FAO-Unesco soil units of profiles selected from the NRCS pedon database for IGBP-DIS*. Work. Pap. 95/01, ISRIC, Wageningen, 15 p *http://www.isric.org/sites/all/modules/pubdlcnt/pubdlcnt.php?file=/isric/webdocs/docs/ISRIC_Report_1995-01.pdf&nid=332*

Spain AV, Isbell RF and Probert ME 1983. Soil organic matter, *Soils: An Australian Viewpoint*. CSIRO Melbourne Academic Press, London, pp 552-563

Tarnocai C, Canadell JG, Schuur EAG, Kuhry P, Mazhitova G and Zimov S 2009. Soil organic carbon pools in the northern circumpolar permafrost region. *Global Biogeochem. Cycles* 23

Tian H, Lu C, Yang J, Banger K, Huntzinger DN, Schwalm CR, Michalak AM, Cook R, Ciais P, Hayes D, Huang M, Ito A, Jain AK, Lei H, Mao J, Pan S, Post WM, Peng S, Poulter B, Ren W, Ricciuto D, Schaefer K, Shi X, Tao B, Wang W, Wei Y, Yang Q, Zhang B and Zeng NCGB 2015. Global patterns and controls of soil organic carbon dynamics as simulated by multiple terrestrial biosphere models: Current status and future directions. *Global Biogeochemical Cycles*, doi: 10.1002/2014GB005021

Todd-Brown KEO, Randerson JT, Post WM, Hoffman FM, Tarnocai C, Schuur EAG and Allison SD 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* 10, 1717-1736

van Engelen VWP and Dijkshoorn JA 2013. *Global and National Soils and Terrain Digital Databases (SOTER) - Procedures manual (Ver. 2.0)*. ISRIC Report 2013/04, IUSS, ISRIC and FAO, Wageningen, 191 p *http://www.isric.org/sites/all/modules/pubdlcnt/pubdlcnt.php?file=/sites/default/files/ISRIC_Report_2013_04.pdf&nid=331*

van Reeuwijk LP 1998. *Guidelines for quality management in soil and plant laboratories*, FAO, Rome, 143 p *http://www.fao.org/docrep/W7295E/W7295E00.htm*

van Reeuwijk LP 2002. *Procedures for soil analysis (6th ed.)*. Technical Paper 9, ISRIC, Wageningen, 1.1 to 19.5 p *http://www.isric.org/Isric/Webdocs/Docs/ISRIC_TechPap09_2002.pdf*

Verburg PH, Neumann K and Nol L 2011. Challenges in using land use and land cover data for global change studies. *Global Change Biology* 17, 974-989

Wilcox RR 2009. *Basic statistics: Understanding conventional methods and modern insights*. Oxford University Press, New York

Appendix 1
Number of profiles by FAO90 soil unit

FAO_90	Description [a]	n[b]
AC: Acrisols		1724
ACf	Ferric Acrisol	311
ACg	Gleyic Acrisol	77
ACh	Haplic Acrisol	1111
ACp	Plinthic Acrisol	101
ACu	Humic Acrisol	124
AL: Alisols		593
ALf	Ferric Alisol	67
ALg	Gleyic Alisol	97
ALh	Haplic Alisol	338
ALj	Stagnic Alisol	8
ALp	Plinthic Alisol	15
ALu	Humic Alisol	68
AN: Andosols		544
ANg	Gleyic Andosol	16
ANh	Haplic Andosol	62
ANi	Gelic Andosol	2
ANm	Mollic Andosol	66
ANu	Umbric Andosol	228
ANz	Vitric Andosol	170
AR: Arenosols		948
ARa	Albic Arenosol	37
ARb	Cambic Arenosol	92
ARc	Calcaric Arenosol	75
ARg	Gleyic Arenosol	68
ARh	Haplic Arenosol	350
ARl	Luvic Arenosol	177
ARo	Ferralic Arenosol	149
AT: Anthrosols		279
Ata	Aric Anthrosol	16
ATc	Cumulic Anthrosol	222
ATf	Fimic Anthrosol	16
ATu	Urbic Anthrosol	25
CH: Chernozems		968
CHg	Gleyic Chernozem	263
CHh	Haplic Chernozem	334
CHk	Calcic Chernozem	148
CHl	Luvic Chernozem	220
CHw	Glossic Chernozem	3
CL: Calcisols		616

FAO_90	Description [a]	n[b]
CLh	Haplic Calcisol	351
CLl	Luvic Calcisol	189
CLp	Petric Calcisol	76
CM: Cambisols		2939
CMc	Calcaric Cambisol	473
CMd	Dystric Cambisol	580
CMe	Eutric Cambisol	713
CMg	Gleyic Cambisol	308
CMi	Gelic Cambisol	257
CMo	Ferralic Cambisol	181
CMu	Humic Cambisol	252
CMv	Vertic Cambisol	86
CMx	Chromic Cambisol	89
FL: Fluvisols		969
FLc	Calcaric Fluvisol	321
FLd	Dystric Fluvisol	98
FLe	Eutric Fluvisol	362
FLm	Mollic Fluvisol	67
FLs	Salic Fluvisol	36
FLt	Thionic Fluvisol	52
FLu	Umbric Fluvisol	33
FR: Ferralsols		582
FRg	Geric Ferralsol	41
FRh	Haplic Ferralsol	247
FRp	Plinthic Ferralsol	13
FRr	Rhodic Ferralsol	106
FRu	Humic Ferralsol	64
FRx	Xanthic Ferralsol	111
FRg	Geric Ferralsol	41
GL: Gleysols		783
GLa	Andic Gleysol	1
GLd	Dystric Gleysol	121
GLe	Eutric Gleysol	318
GLi	Gelic Gleysol	31
GLk	Calcic Gleysol	33
GLm	Mollic Gleysol	150
GLt	Thionic Gleysol	8
GLu	Umbric Gleysol	121
GR: Greyzems		38
GRg	Gleyic Greyzem	3
GRh	Haplic Greyzem	35
GY: Gypsisols		115
GYh	Haplic Gypsisol	47
GYk	Calcic Gypsisol	38
GYl	Luvic Gypsisol	11
GYp	Petric Gypsisol	19
HS: Histosols		380
HSf	Fibric Histosol	103
HSi	Gelic Histosol	152
HSl	Folic Histosol	9

FAO_90	Description [a]	n[b]
HSs	Terric Histosol	105
HSt	Thionic Histosol	11
KS: Kastanozems		1411
KSh	Haplic Kastanozem	543
KSk	Calcic Kastanozem	173
KSl	Luvic Kastanozem	695
LP: Leptosols		431
LPd	Dystric Leptosol	87
LPe	Eutric Leptosol	103
LPi	Gelic Leptosol	31
LPk	Rendzic Leptosol	76
LPm	Mollic Leptosol	54
LPq	Lithic Leptosol	24
LV: Luvisols		3127
LPu	Umbric Leptosol	56
LVa	Albic Luvisol	116
LVf	Ferric Luvisol	123
LVg	Gleyic Luvisol	474
LVh	Haplic Luvisol	1721
LVj	Stagnic Luvisol	139
LVk	Calcic Luvisol	221
LVv	Vertic Luvisol	42
LVx	Chromic Luvisol	291
LXa	Albic Lixisol	4
LX: Lixisols		554
LXf	Ferric Lixisol	200
LXg	Gleyic Lixisol	39
LXh	Haplic Lixisol	289
LXj	Stagnic Lixisol	1
LXp	Plinthic Lixisol	21
NT: Nitisols		151
NTh	Haplic Nitisol	69
NTr	Rhodic Nitisol	47
NTu	Humic Nitisol	35
PD: Podzoluvisols		290
PDd	Dystric Podzoluvisol	92
PDe	Eutric Podzoluvisol	137
PDg	Gleyic Podzoluvisol	45
PDj	Stagnic Podzoluvisol	16
PH: Phaeozems		955
PHc	Calcaric Phaeozem	61
PHg	Gleyic Phaeozem	64
PHh	Haplic Phaeozem	339
PHj	Stagnic Phaeozem	15
PHl	Luvic Phaeozem	476
PLd	Dystric Planosol	44
PLe	Eutric Planosol	130
PLm	Mollic Planosol	20
PLu	Umbric Planosol	6

FAO_90	Description [a]	n[b]
PL: Planosols		200
PTa	Albic Plinthosol	21
PTd	Dystric Plinthosol	50
PTe	Eutric Plinthosol	29
PT: Plinthosols		103
PTa	Albic Plinthosol	21
PTd	Dystric Plinthosol	50
PTe	Eutric Plinthosol	29
PTu	Humic Plinthosol	3
PZ: Podzols		621
PZb	Cambic Podzol	25
PZc	Carbic Podzol	42
PZf	Ferric Podzol	11
PZg	Gleyic Podzol	105
PZh	Haplic Podzol	381
PZi	Gelic Podzol	57
RG: Regosols		735
RGc	Calcaric Regosol	233
RGd	Dystric Regosol	114
RGe	Eutric Regosol	304
RGi	Gelic Regosol	32
RGu	Umbric Regosol	30
RGy	Gypsic Regosol	22
SC: Solonchaks		255
SCg	Gleyic Solonchak	60
SCh	Haplic Solonchak	65
SCi	Gelic Solonchak	4
SCk	Calcic Solonchak	30
SCm	Mollic Solonchak	19
SCn	Sodic Solonchak	42
SCy	Gypsic Solonchak	35
SN: Solonetz		554
SNg	Gleyic Solonetz	86
SNh	Haplic Solonetz	220
SNj	Stagnic Solonetz	22
SNk	Calcic Solonetz	113
SNm	Mollic Solonetz	99
SNy	Gypsic Solonetz	14
VR: Vertisols		927
VRd	Dystric Vertisol	54
VRe	Eutric Vertisol	571
VRk	Calcic Vertisol	278
VRy	Gypsic Vertisol	24

[a] For detailed descriptions see (FAO 1988).

[b] Data for some 10,000 of these profiles (Batjes 2009), out of a total of 21,000, are freely available (CC BY-NC) whereas the remainder was made available by third parties (cited elsewhere) for making derived products only. Subject to such licence permissions, profiles so far managed in the ISRIC-WISE soil profile database will become gradually available through the ISRIC World Soil Information Service (Carvalho Ribeiro *et al.* 2015 (in prep.)).

Appendix 2
Number and type of soil profiles by UN Region

Continent	UN Region	Soil profile description status[a]				
		1	2	3	4	ALL
Africa						
	Africa - Eastern	206	259	507	15	987
	Africa - Middle	12	36	250	1	299
	Africa - Northern	50	76	41	0	167
	Africa - Southern	67	791	148	1	1007
	Africa - Western	85	317	1445	6	1853
America						
	America - Caribbean	26	222	69	0	317
	America - Central	94	58	4	10	166
	America - Northern	686	8348	492	3	9530
	America - South	150	1421	313	1	1885
Antarctica						
	Antarctica	4	25	1	0	30
Asia						
	Asia - Central	0	0	13	0	13
	Asia - Eastern	112	1665	49	0	1826
	Asia - South Eastern	190	276	70	1	537
	Asia - Southern	71	196	134	8	409
	Asia - Western	99	301	202	1	603
Europe						
	Europe - Eastern	78	573	220	7	878
	Europe - Northern	29	241	95	3	368
	Europe - Southern	73	135	90	9	307
	Europe - Western	45	152	103	1	301
Oceania						
	Oceania - Australia & New Zealand	8	103	116	5	232
	Oceania - Melanesia	13	4	25	0	42
	Oceania - Micronesia	13	3	0	0	16
	Oceania - Polynesia	16	3	0	0	19
ALL	World	2127	15205	4387	72	21792

[a] The number code under ‘profile description status’ refers to the completeness and apparent reliability of the soil profile descriptions and accompanying analytical data for the specified profile in the original source. The status is highest for ‘1’ and lowest for ‘4’ (see FAO 2006).

Appendix 3
Key to main output files

The whole dataset is referred to as WISE30sec, version 1.0. Compact names are used for the various data files for ease of handling in GIS.

Target	Description	Appendix
Spatial data:		
wise30sec_fin	Raster file (30x30 arcsec; about 0.9x0.9 km at the equator)	4
Descriptive statistics:		
STAT_ACCL1	Acrisols (AC) to Calcisols (CL) units, per climate region (1)	5
STAT_CMHS1	As above for Cambisol to Histosol units (1)	5
STAT_KSPH1	As above for Kastanozem to Phaeozem units (1)	5
STAT_PLVR1	As above for Planosol to Vertisol units (1)	5
STAT_ACCL2	Acrisols (AC) to Calcisols (CL) units, undifferentiated climate (2)	5
STAT_CMHS2	As above for Cambisol to Histosol units (2)	5
STAT_KSPH2	As above for Kastanozem to Phaeozem units (2)	5
STAT_PLVR2	As above for Planosol to Vertisol units (2)	5
STAT_ACCL3	Acrisols (AC) to Calcisols (CL) groups, per climate region (3)	5
STAT_CMHS3	As above for Cambisol to Histosol groups (3)	5
STAT_KSPH3	As above for Kastanozem to Phaeozem groups (3)	5
STAT_PLVR3	As above for Planosol to Vertisol groups (3)	5
STAT_ACCL4	Acrisols (AC) to Calcisols (CL) groups, undiff. climate (4)	5
STAT_CMHS4	As above for Cambisol to Histosol groups (4)	5
STAT_KSPH4	As above for Kastanozem to Phaeozem groups (4)	5
STAT_PLVR4	As above for Planosol to Vertisol groups (4)	5
Soil attribute data:		
HW30S_ParEst	Table listing soil parameter estimates (mean and std), as derived from the TTR-procedures (which consider data from the appropriate *STAT_Xi* files (Appendix 5), for all mapped 'soil/climate' combinations (rounded values)	7
Log_TTRexr	Table summarising the type of TTR and EXR rules applied for each possible 'soil/climate' cluster (as shown in *HW30s_ParEst*); log file	7
HW30S_MapUnit	Map unit composition	6
HW30S_FULL	Table with soil parameter estimates (mean and std), as derived from the TTR procedure (see *HW30s_ParEstim*), for all map units on the GIS map (see *Wise30sec_fin*) and all (7) depth layers (rounded values). This 'full map unit composition' file should be used for calculations (see Section 3.4)	9
HW30S_wDi	Area-weighted soil parameter estimates for <u>all</u> layers, D1 to D7 (0-200 cm). This file and the subsets per layer derived from it (i.e. *HW30S_wD1* to *HW30S_wD7*) are included *only* to allow for 'quick' visualisation using GIS; as indicated, calculations should consider the full map unit composition (*HW30S_FULL*)	10
HW30S_wD1	As above for layer D1 (0-20 cm)	10
HW30S_wD2	As above for layer D2 (20-40 cm)	10
HW30S_wD3	As above layer D3 (40-60 cm)	10
HW30S_wD4	As above for layer D4 (60-80 cm)	10
HW30S_wD5	As above for layer D5 (80-100 cm)	10
HW30S_wD6	As above for layer D6 (100-150 cm)	10
HW30S_wD7	As above for layer D7 (150-200 cm)	10

Appendix 4
Raster GIS files

The soil-geographical (spatial) data are provided as a raster GIS file (WGS_1984), which has been derived from the HWSD (FAO *et al.* 2012), with some regional updates (see text), and the Köppen climate zones map (Peel *et al.* 2007).

Structure of file *wise30sec_fin*

Name	Description
Rowid	GIS-generated number
Value	GIS-generated number (same as MUGLB_NEW)
MU_GLOBAL	Map unit number for the 'original' HWSD, v1.21
MUGLB_NEW	Revised map unit number for updated spatial data set
Coverage[a]	Lineage of the spatial data
Code[b]	Köppen-Geiger climate class (Peel *et al.* 2007)
Climate_CO	Code for given combination of climate code and soil mapping unit number
NEWSUID[c]	Code for map unit (recoded/derived from Climate_CO) used for joining the spatial and derived soil attribute data
Count	GIS-generated number

[a] Code for lineage of the spatial data: 0: none, 1= ESDB, 2= China, 3= SOTWIS, 4= DSMW, 5= Minor updates made for this study, see Section 2.2 for details and full references (with special thanks to Maria Ruiperez Gonzalez for GIS support).

[b] Code for Köppen climate classes: 1–A: Tropical ; 2–B: Arid; 3–C: Temperate; 4–D: Cold; 5–E: Polar.

[c] Map unit (grid cell) code, for example 'WD1000234' where 'WD' stands for World, '1' stands for Köppen climate 'A: Tropical', and '234' is the code for MUGLB_NEW; '000' are 'fillers' to arrive at a standard length of 9 characters.

Appendix 5
Statistical output files

Results of the statistical analyses of measured soil data in WISE (after screening according to broad comparability of soil analytical methods and application of an outlier detection/rejection) by soil type unit, climate class, soil attribute, depth layer, and soil textural class, as summarised in the *Cluster_id* name.

Depending on the level of detail possible (i.e. cluster sample sizes) the output (or input for the TTR procedure) is handled in four tables that have similar structure, and will be accessed in sequence as needed during the TTR procedure (see section 3.2):

1. *STAT_XXZZ1* for cases like 'ACfBSAT/A/D1u' with results for soil units for a given climate.
2. *STAT_XXZZ2* for cases like 'ACfBSAT/Y/D1u' with results for soil units for undifferentiated climate.
3. *STAT_XXZZ3* for cases like 'ACBSAT/A/D1u' with results for major soil groups only, for a given climate.
4. *STAT_XXZZ4* for jcases like 'ACBSAT/Y/D1u' with results for major soil groups, and undifferentiated climate.

Structure of statistical output tables:

Name	Type	Description
Cluster_id	Text	Code comprising abbreviation for FAO90 soil unit resp. soil group, attribute, climate class, depth layer, and soil textural class. For example, *Cluster_id* ACfBSAT/A/D1u' for the cluster of ferric Acrisols (ACf) from the Tropics (A), base saturation (BSAT) for layer D1 (0-20 cm), and modal texture(undifferentiated). [Note: This is the rightmost field in the output table]
Num0	Integer	Number of observations (before outlier detection)
Num	Integer	Number of observations (*n*, after outlier rejection)
Mean	Single	Mean
STD	Single	Standard deviation
SE	Single	Standard error (i.e. STD/SQRT(Num))
CV	Single	Coefficient of variation
Median	Single	Median (second quartile)
MAD	Single	Median of absolute deviations
Min	Single	Minimum
Max	Single	Maximum
Var	Single	Variance
FAO_90	Text	FAO90 code (this field is intentionally left blank)
P10	Single	10^{th} percentile (only given where $n \geq 10$)
P20	Single	20^{th} percentile
P30	Single	30^{th} percentile
P40	Single	40^{th} percentile
P50	Single	50^{th} percentile
P60	Single	60^{th} percentile
P70	Single	70^{th} percentile
P80	Single	80^{th} percentile

Name	Type	Description
P90	Single	90th percentile
Quart1	Single	25th percentile, first quartile
Quart3	Single	75th percentile, third quartile

Footnotes:

1) The output for *STAT_XXZZ1* etc. is subdivided in four similar tables in view of functionality and space requirements (for naming conventions see Appendix 3). For example, XX represents the first major soil group (e.g. AC, Acrisols) and YY the last (e.g. CL, Calcisols) in a given XXZZ subset.
2) Statistical output files have the same structure. However, the nature of the information presented under *cluster_id* will change depending on the table name, as explained above. During the TTR procedure, these tables are queried sequentially from STAT_XXZZ1 to STAT_XXZZ4, corresponding with increasingly 'coarse' approximations for the derived soil property under consideration.
3) Descriptive statistics are for depth-weighted data, per layer (from D1 to D7, see text). Present analyses are for mineral soil layers, except for Histosols, the top of which has been defined as 0 cm.
4) These tables list results for all statistical analyses after outlier detection and rejection. The taxotransfer scheme, however, will only consider means and standard deviations from the corresponding tables when $n \geq 5$ (see n_{WISE} in text).

Appendix 6
Soil map unit composition file

Table *HW30S_MapUnit* gives the full composition of each map unit in terms of its main soil units (FAO 1988), their relative extent, and the identifier for the corresponding synthetic soil profile, and climate class. The contents of this table can be joined to the spatial data using the *NEWSUID* field.

Structure of table HW30S_MapUnit [a]

Name	Type	Description
NEWSUID	Text	Globally unique code, e.g. 'WD1000234' (see Appendix 4)
KopPeel	Text	Code for main Köppen-Geiger climate class (Peel *et al.* 2007)
NoSoilComponent	Number	Maximum number of soil components in map unit (up to 10)
DomFAO_Name	Text	Dominant FAO90 major group in map unit (Note: This need not always be SOIL1)
DomFAO_Prop	Number	Proportion of dominant FAO major soil/climate class within in soil unit
PropSynthProf	Integer	Proportion of map unit represented by synthetic profiles (always 100% for WISE30sec)
SoilMapunit [b]	Text	Aggregated code for map unit summarizing the overall composition; final letter stands for the climate class
SOIL1	Text	Characterization of the first (main) soil unit according to the FAO90 Legend
PROP1	Integer	Proportion, as a percentage, that the main soil unit occupies within the given map unit
PRID1	Text	Unique code for the corresponding synthetic soil profile (e.g GLe/A, for a gleyic Acrisol mapped in the Tropics)
SOIL2	Text	As above but for the next soil unit
PROP2	Integer	As above
PRID2	Text	As above
SOIL3	Text	As above but for the next soil unit
PROP3	Integer	As above
PRID3	Text	As above
SOIL4	Text	As above but for the next soil unit
PROP4	Integer	As above
PRID4	Text	As above
SOIL5	Text	As above but for the next soil unit
PROP5	Integer	As above
PRID5	Text	As above
SOIL6	Text	As above but for the next soil unit
PROP6	Integer	As above
PRID6	Text	As above
SOIL7	Text	As above but for the next soil unit
PROP7	Integer	As above
PRID7	Text	As above
SOIL8	Text	As above but for the next soil unit
PROP8	Integer	As above
PRID9	Text	As above
SOIL9	Text	As above but for the next soil unit
PROP9	Integer	As above
SOIL10	Text	As above but for the next soil unit
PROP10	Integer	As above

Name	Type	Description
PRID10	Text	As above
WATER	Number	Proportion of units permanently covered by water (left blank)

[a] For the sake of consistency, table structure conventions used for secondary SOTER databases have been retained here — as such this table has the same structure as table SOTERxxx (see Batjes 2010).

[b] As indicated, each 'soil-climate' map unit may comprise up to ten component soils. For ease of legibility (or reference purposes), the relative extent of the component units of each map unit has been coded to arrive at a compact map unit code: 1 – from 80 to 100 per cent; 2 – from 60 to 80 per cent; 3 – from 40 to 60 per cent; 4 – from 20 to 40 per cent, and 5 – less than 20 per cent.

Appendix 7
Taxotransfer rule-based soil property estimates

Table *HW30S_ParEst* lists soil property estimates for all synthetic profiles (e.g. ACh/A) considered in the derived data set. This *PRID*-based information can be linked to the geographical component of the database through the unique map unit identifier, taking into account the full map unit composition (*NEWSUID*, see Appendix 4).

Table *HW30S_ParEst* should be consulted in conjunction with table *Log_TTRexr* that documents the type of taxotransfer respectively expert rules that have been applied (see Appendix 8).

Structure of table *HW30S_ParEst*

Name	Type	Description
CLAF	Text	FAO, Revised Legend code (FAO90)
PRID	Text	profile ID (as documented in table *HW30S_MapUnit*, see Appendix 6)
Drain	Text	FAO soil drainage class
Layer	Text	code for depth layer (from D1 to D7; e.g. D3 is from 40 to 60 cm)
TopDep	Integer	depth of top of layer (cm)
BotDep	Integer	depth of bottom of layer (cm)
CFRAG	Integer	coarse fragments (vol% > 2 mm)
CFRAG_STD	Single	standard deviation for above
SDTO	Integer	sand (mass %)
SDTO_STD	Single	standard deviation for above
STPC	Integer	silt (mass %)
STPC_STD	Single	standard deviation for above
CLPC	Integer	clay (mass %)
CLPC_STD	Single	standard deviation for above
PSCL	Text	texture class (SOTER conventions)
BULK	Single	bulk density (kg dm^{-3})
BULK_STD	Single	standard deviation for above
TAWC	Integer	available water capacity (cm m^{-1}, -33 to -1500 kPa conform to USDA standards)
TAWC_STD	Single	standard deviation for above
CECs	Single	cation exchange capacity ($cmol_c$ kg^{-1}) for fine earth fraction
CECs_STD	Single	standard deviation for above
TEB	Single	total exchangeable bases ($cmol_c$ kg^{-1})
TEB_STD	Single	standard deviation for above
BSAT	Integer	base saturation as percentage of $CEC_{s(oil)}$
BSAT_STD	Single	standard deviation for above
CECc	Single	CEC_{clay}, corrected for contribution of organic matter ($cmol_c$ kg^{-1})
CECc_STD	Single	standard deviation for above
ECEC	Single	effective CEC ($cmol_c$ kg^{-1})
ECEC_STD	Single	standard deviation for above
ESP	Single	exchangeable sodium (%) of CEC_{soil}

Name	Type	Description
ESP_STD	Single	standard deviation for above
ALSA	Single	exchangeable aluminium (as proportion of ECEC)
ALSA_STD	Single	standard deviation for above
PHAQ	Single	pH measured in water
PHAQ_STD	Single	standard deviation for above
TCEQ	Single	total carbonate equivalent (g C kg^{-1})
TCEQ_STD	Single	standard deviation for above
GYPS	Single	gypsum content (g kg^{-1})
GYPS_STD	Single	standard deviation for above
ELCO	Single	electrical conductivity (dS m^{-1})
ELCO_STD	Single	standard deviation for above
ORGC	Single	organic carbon content (g C kg^{-1})
ORGC_STD	Single	standard deviation for above
TOTN	Single	total nitrogen (g kg^{-1})
TOTN_STD	Single	standard deviation for above
CNrt	Single	C/N ratio
CNrt_STD	Single	standard deviation for above

Notes: A minus 8 indicates that no meaningful substitution was possible for the specified ‘soil unit/climate’ cluster and attribute using TTR based on the present selection of soil profiles, -1 is used for Oceans and inland waters, -2 for Glaciers and snow caps, -3 for rock outcrops (resp. -7 for ‘rocky’ subsoils as for Leptosols), -4 for Dunes/Shifting sands, -5 for Salt flats, and -9 for all remaining miscellaneous units mainly to facilitate visualization using GIS.

Appendix 8
Flagging of taxotransfer rules

The type of taxotransfer and expert rules that have been used when creating table *HW30S_ParEstim* (see Appendix 7) is documented in table *Log_TTRexr*. Further details on coding conventions may be found in the text (Section 3.2).

Structure of table *Log_TTRexr*

Name	Type	Description
CLAF	Text	FAO Revised Legend code
PRID	Text	Unique identifier for synthetic profile (e.g. ACh/A)
Layer	Text	Depth of soil layer (D1=0-20; D2=40-40; D3=40-60; D4=60-80; D5=80-100; D6=100-150; D7=150-200 cm)
NewTopdep	Integer	Depth of top of layer (cm)
NewBotdep	Integer	Depth of bottom of layer (cm)
TTRsub	Text	Codes for the type of taxotransfer rule used (based on derived data for similar soil units and climate; see text)
TTRsubY	Text	Codes for the type of taxotransfer rule used (based on derived data for similar soil units, undiff. climate; Y see text)
TTRmain	Text	Codes for the type of taxotransfer rule used (based on data derived data for similar major soil groups and climate)
TTRmainY	Text	Codes for the type of taxotransfer rule used (based on data derived data for similar major soil groups and undiff. climate)
TTRexpert	Text	Flag for type of expert rule(s) used (see text for details)

Note: The exchangeable aluminium percentage (ALSA) has been set at zero when pH_{water} is higher than 5.5 (EXR= 1). Similarly, the content of gypsum (GYPS) and content of carbonates (TCEQ) have been set at zero when pH_{water} is less than 6.5.

Conventions for coding soil attributes considered in table *Log_TTRexr* are listed below:

Soil variable	TTRflag[a]	EXRflag	Description[b]
ALSA	A	A	exchangeable Aluminium percentage (% of ECEC)
BSAT	B	B	base saturation (% of CECs)
BULK	C	C	bulk density
CECC	D	D	cation exchange capacity of clay fraction (corrected for organic C)
CECS	E	-	cation exchange capacity
CFRAG	F	F	coarse fragments vol% (> 2 mm)
CLPC	G	-	clay wt%
CNrt	Z	-	C/N ratio (for the measured data)
ECEC	H	-	effective CEC
ECE	I	I	electrical conductivity
ESP	J	J	exchangeable Na percentage (as % of CECs)
GYPS	K	K	gypsum content
ORGC	Q	-	organic carbon content
PHAQ	L	-	pH in water
SDTO	M	-	sand wt%
STPC	N	-	silt wt%
TAWC	O	O	volumetric water content (-33 to -1500 kPa, cm m^{-1})
TEB	X	X	total exchangeable bases
TCEQ	P	P	carbonate equivalent content
TOTN	R	-	total nitrogen content

[a] The same codes are used for flagging the TTRs (i.e. *TTRsub*, *TTRsubY, TTRmain* and *TTRmainY*) and EXRs (i.e. *TTRexpert*) for a given attribute.

[b] See Table 3 for units of measurement. C/N ratios have been calculated 'as is' from the measured data (CNrt), not as the ratio of the derived values for C and N, *ditto* for CEC_{clay}, as this would introduce additional errors.

Appendix 9
Summary files of derived soil properties (full map unit composition)

HW30S_Full

This large file holds all the derived data grouped by map unit at the highest level; as such it consists of the 'un-binned' soil properties values, derived from the TTR procedure, for all component soil units that occur in the given map unit (or grid cell).

Structure of table *HW30S_Full*

Name	Type	Description
NEWSUID	Text	Globally unique map unit code, see Appendix 4
SCID	Integer	Number of soil unit within the given map unit (ranges from 1 to 10)
PROP	Integer	Relative proportion of above in given map unit
CLAF	Text	FAO90 Legend code
PRID	Text	Profile ID (as documented in table *HW30S_MapUnit*, App. 6)
Drain	Text	FAO soil class
DrainNum	Text	Numerical value for FAO soil drainage class (see Note 4)
Layer	Text	Code for depth layer (D1 to D7, e.g. D1=0-20 cm)
TopDep	Integer	Upper depth of layer (cm)
BotDep	Integer	Lower depth of layer (cm)
CFRAG	Integer	Coarse fragments (vol. % > 2mm), mean
CFRAG_std	Single	Coarse fragments (vol. % > 2mm), standard deviation
SDTO	Integer	Sand (mass %), mean
SDTO_std	Single	Standard deviation for above
STPC	Integer	Silt (mass %)
STPC_std	Single	Standard deviation for above
CLPC	Integer	Clay (mass %)
CLPC_std	Single	Standard deviation for above
PSCL	Text	Texture class (SOTER conventions)
BULK	Single	Bulk density (kg dm^{-3}, g cm^{-3})
BULK_std	Single	Standard deviation for above
TAWC	Integer	Available water capacity (cm m^{-1}, -33 to -1500 kPa, conform USDA standards)
TAWC_std	Single	Standard deviation for above
CECS	Single	Cation exchange capacity ($cmol_c$ kg^{-1}) of fine earth fraction
CECs_std	Single	Standard deviation for above
BSAT	Integer	Base saturation as percentage of CECsoil
BSATstd	Single	Standard deviation for above
ESP	Integer	Exchangeable sodium percentage
ESP_std	Single	Standard deviation for above
CECc	Single	CECclay, corrected for contribution of organic matter ($cmol_c$ kg^{-1})
CECc_std	Single	Standard deviation for above

Name	Type	Description
PHAQ	Single	pH measured in water
PHAQ_std	Single	Standard deviation for above
TCEQ	Single	Total carbonate equivalent (g C kg^{-1})
TCEQ_std	Single	Standard deviation for above
GYPS	Single	Gypsum content (g kg^{-1})
GYPS_std	Single	Standard deviation for above
ELCO	Single	Electrical conductivity (dS m^{-1})
ELCO_std	Single	Standard deviation for above
ORGC	Single	Organic carbon content (g kg^{-1})
ORGC_std	Single	Standard deviation for above
TOTN	Single	Total nitrogen (g kg^{-1})
TOTN_std	Single	Standard deviation for above
CNrt	Single	C/N ratio
CNrt_std	Single	Standard deviation for above
ECEC	Single	Effective CEC ($cmol_c$ kg^{-1})
ECEC_std	Single	Standard deviation for above
ALSA	Integer	Aluminum saturation (as % of ECEC)
ALSA_std	Single	Standard deviation for above

Notes:

1) These are depth-weighted values, per 20 cm (D1 to D5) resp. 50 cm layer (D6 and D7).
2) Component soils within a given mapping are numbered sequentially starting with the spatially dominant one. The sum of the relative proportions of all soil units and miscellaneous units within a map unit is always 100 per cent.
3) Soil drainage classes: 1, **V**ery poorly; 2, **P**oorly drained; 3, **I**mperfectly drained; 4, **M**oderately well drained; 5, **W**ell-drained; 6, **S**omewhat excessively; and 7, **E**xcessively drained soils (Note: numerical versus class codes in bold).

In view of the map unit complexity, additional operations will often be needed before results can be visualized or analysed (meaningfully) using GIS or when binned or un-binned data (e.g. pH_{water} class 1 may correspond to pHwater < 3.5 etc.) are needed for a given soil layer of a component soil unit. Complex calculations should consider the full map unit composition, see text (3.4). However, to facilitate 'quick' visualisation using with GIS, a file with area-**w**eighted derived properties has been generated for each depth layer. These are called HW30S_**w**Di, where Di stands for D1 to D7; see Appendix 10.

Appendix 10
Summary files of derived soil properties (full map unit composition)

As discussed in section 3.4, comprehensive calculations should consider the full map unit composition (see Appendix 9). However, this compound information cannot be 'visualised' easily using GIS for which a file with area-weighted soil property estimates to a depth of 2 m (*HW30S_wDi*) has been generated. Separate files are also presented for each layer, e.g. file *HW30S_wD3* holds area-weighted data for 40-60 cm (layer D3, see text).

Structure of table *HW30S_wDi*

Name	Type	Description
NEWSUID	Text	Globally unique map unit code
NofComponents	Long Integer	Number of soil components in map unit
SoilUnits	Text	Code for resp. FAO soils units and their proportion
Layer	Text	Code for depth layer (D1 to D7, e.g. D1=0-20 cm)
PROP_aw	Integer	Relative proportion of soil units in given map unit
MiscUnits	Text	Code for resp. miscellaneous units and their proportion
Prop_misc	Integer	Proportion of miscellaneous units in map unit
Drain	Text	Dominant FAO soil drainage class
DrainProp	Integer	Proportion of above
DrainNum	Text	Number of different drainage classes in map unit
DrainMin	Integer	Lowest drainage class (numerical)
DrainMax	Integer	Highest drainage class (numerical)
TopDep	Integer	Upper depth of layer (cm)
BotDep	Integer	Lower depth of layer (cm)
CFRAG	Integer	Coarse fragments (vol. % > 2mm), mean
CFRAG_std	Single	Coarse fragments (vol. % > 2mm), standard deviation
SDTO	Integer	Sand (mass %), mean
SDTO_std	Single	Standard deviation for above
STPC	Integer	Silt (mass %)
STPC_std	Single	Standard deviation for above
CLPC	Integer	Clay (mass %)
CLPC_std	Single	Standard deviation for above
PSCL	Text	FAO texture class (see Figure 8)
BULK	Single	Bulk density (kg dm^{-3}, g cm^{-3})
BULK_std	Single	Standard deviation for above
TAWC	Integer	Available water capacity (cm m^{-1}, -33 to -1500 kPa, conform USDA standards)
TAWC_std	Single	Standard deviation for above
CECS	Single	Cation exchange capacity ($cmol_c$ kg^{-1}) of fine earth fraction
CECs_std	Single	Standard deviation for above

Name	Type	Description
BSAT	Integer	Base saturation as percentage of CECsoil
BSATstd	Single	Standard deviation for above
ESP	Integer	Exchangeable sodium percentage
ESP_std	Single	Standard deviation for above
CECc	Single	CECclay, corrected for contribution of organic matter cmol$_c$ kg^{-1})
CECc_std	Single	Standard deviation for above
PHAQ	Single	pH measured in water
PHAQ_std	Single	Standard deviation for above
TCEQ	Single	Total carbonate equivalent (g C kg^{-1})
TCEQ_std	Single	Standard deviation for above
GYPS	Single	Gypsum content (g kg^{-1})
GYPS_std	Single	Standard deviation for above
ELCO	Single	Electrical conductivity (dS m^{-1})
ELCO_std	Single	Standard deviation for above
ORGC	Single	Organic carbon content (g kg^{-1})
ORGC_std	Single	Standard deviation for above
TOTN	Single	Total nitrogen (g kg^{-1})
TOTN_std	Single	Standard deviation for above
CNrt	Single	C/N ratio
CNrt_std	Single	Standard deviation for above
ECEC	Single	Effective CEC (cmol$_c$ kg^{-1})
ECEC_std	Single	Standard deviation for above
ALSA	Integer	Aluminum saturation (as % of ECEC)
ALSA_std	Single	Standard deviation for above

Notes:

1) These are depth-weighted (and area-weighted) values, per 20 cm (D1 to D5) resp. 50 cm layer (D6 and D7).
2) The sum of the relative proportions of all soil units and miscellaneous units within a map unit is always 100 per cent.
3) Soil drainage classes: 1, **V**ery poorly; 2, **P**oorly drained; 3, **I**mperfectly drained; 4, **M**oderately well drained; 5, **W**ell-drained; 6, **S**omewhat excessively; and 7, **E**xcessively drained soils (Note: numerical versus class codes in bold).
4) Uncertainties (e.g. CFRAG_std or ORGC_std) are calculated using error propagation procedures as described in Ku (1966), assuming independency of variables (i.e. parameter estimates for the different component soils in a given map unit or grid cell).

Appendix 11
Correlation rules FAO74 to FAO90 Legend[1]

Important sections of the HWSD, such as South-East Asia, Australia, North America and West Africa, are still based on materials from the old Digital Soil Map of the World, which uses the original FAO-Unesco (1974), while for the remainder of the world the revised legend has been used. For the former, the FAO74 names were correlated to the revised FAO (1988) Legend. Main changes in definitions and concepts are described in FAO (1988). Some of these cannot be addressed unambiguously during this type of broad scale correlation (Bridges *et al.* 1998); the table below should be seen as a first approximation (ver. 1.1) commensurate with the adopted broad scale of mapping and scope of this study.

	FAO74 Legend	FAO90 Legend		Remarks
Code	Name	Code (corr.)	Name	
A	Acrisols	AC	ACRISOLS	
Af	Ferric Acrisols	ACf	Ferric Acrisols	
Ag	Gleyic Acrisols	ACg	Gleyic Acrisols	
Ah	Humic Acrisols	ACu	Humic Acrisols	
Ao	Orthic Acrisols	ACh	Haplic Acrisols	
Ap	Plinthic Acrisols	PTd	Dystric Plinthosols	*First approximation*
B	Cambisols	CM	CAMBISOLS	
Bc	Chromic Cambisols	CMx	Chromic Cambisols	
Bd	Dystric Cambisols	CMd	Dystric Cambisols	
Be	Eutric Cambisols	CMe	Eutric Cambisols	
Bf	Ferralic Cambisols	CMo	Ferralic Cambisols	
Bg	Gleyic Cambisols	CMg	Gleyic Cambisols	
Bh	Humic Cambisols	CMu	Humic Cambisols	
Bk	Calcic Cambisols	CMc	Calcaric Cambisols	
Bv	Vertic Cambisols	CMv	Vertic Cambisols	
Bx	Gelic Cambisols	CMi	Gelic Cambisols	
C	Chernozems	CH	CHERNOZEMS	
Cg	Glossic Chernozems	CHw	Glossic Chernozems	
Ch	Haplic Chernozems	CHh	Haplic Chernozems	
Ck	Calcic Chernozems	CHk	Calcic Chernozems	
Cl	Luvic Chernozems	CHl	Luvic Chernozems	
D	Podzoluvisols	PD	PODZOLUVISOLS	
Dd	Dystric Podzoluvisol	PDd	Dystric Podzoluvisols	
De	Eutric Podzoluvisols	PDe	Eutric Podzoluvisols	
Dg	Gleyic Podzoluvisols	PDg	Gleyic Podzoluvisols	
E	Rendzinas	LPm	Mollic Leptosols	
F	Ferralsols	FR	FERRALSOLS	
Fa	Acric Ferralsols	FRg	Geric Ferralsols	

[1] Version 1.2 of conversion rules.

FAO74 Legend		FAO90 Legend		Remarks
Code	Name	Code (corr.)	Name	
Fh	Humic Ferralsols	FRu	Humic Ferralsols	
Fo	Orthic Ferralsols	FRh	Haplic Ferralsols	
Fp	Plinthic Ferralsols	PT	PLINTHOSOLS	*First approximation*
Fr	Rhodic Ferralsols	FRr	Rhodic Ferralsols	
Fx	Xanthic Ferralsols	FRx	Xanthic Ferralsols	
G	Gleysols	GL	GLEYSOLS	
Gc	Calcaric Gleysols	GLk	Calcic Gleysols	
Gd	Dystric Gleysols	GLd	Dystric Gleysols	
Ge	Eutric Gleysols	GLe	Eutric Gleysols	
Gh	Humic Gleysols	GLu	Umbric Gleysols	
Gm	Mollic Gleysols	GLm	Mollic Gleysols	
Gp	Plinthic Gleysols	PT	PLINTHOSOLS	*First approximation*
Gx	Gelic Gleysols	GLi	Gelic Gleysols	
H	Phaeozems	PH	PHAEOZEMS	
Hc	Calcaric Phaeozems	PHc	Calcaric Phaeozems	
Hg	Gleyic Phaeozems	PHg	Gleyic Phaeozems	
Hh	Haplic Phaeozems	PHh	Haplic Phaeozems	
Hl	Luvic Phaeozems	PHl	Luvic Phaeozems	
I	Lithosols	LPq	Lithic Leptosols	
J	Fluvisols	FL	FLUVISOLS	
Jc	Calcaric Fluvisols	FLc	Calcaric Fluvisols	
Jd	Dystric Fluvisols	FLd	Dystric Fluvisols	
Je	Eutric Fluvisols	FLe	Eutric Fluvisols	
Jt	Thionic Fluvisols	FLt	Thionic Fluvisols	
K	Kastanozems	KS	KASTANOZEMS	
Kh	Haplic Kastanozems	KSh	Haplic Kastanozems	
Kk	Calcic Kastanozems	KSk	Calcic Kastanozems	
Kl	Luvic Kastanozems	KSl	Luvic Kastanozems	
L	Luvisols	LV	LUVISOLS	
La	Albic Luvisols	LVa	Albic Luvsiols	
Lc	Chromic Luvisols	LVx	Chromic Luvisols	
Lf	Ferric Luvisols	LXf	Ferric Lixisols	
Lg	Gleyic Luvisols	LVg	Gleyic Luvisols	
Lk	Calcic Luvisols	LVk	Calcic Luvisols	
Lo	Orthic Luvisols	LVh	Haplic Luvisols	
Lp	Plinthic Luvisols	LXp	Plinthic Lixisols	
Lv	Vertic Luvisols	LVv	Vertic Luvisols	
M	Greyzems	GR	GREYZEMS	
Mg	Gleyic Greyzems	GRg	Gleyic Greyzems	
Mo	Orthic Greyzems	GRh	Haplic Greyzems	
N	Nitosols	NT	NITISOLS	
Nd	Dystric Nitosols	NTh	Haplic Nitisols	*First approximation*
Ne	Eutric Nitosols	NTr	Rhodic Nitisols	*First approximation*
Nh	Humic Nitosols	NTu	Humic Nitisols	
O	Histosols	HS	HISTOSOLS	
Od	Dystric Histosols	HS	HISTOSOLS	*Undiff., first approximation*
Oe	Eutric Histosols	HS	HISTOSOLS	*Undiff., first approximation*
Ox	Gelic Histosols	HSi	Gelic Histosols	
P	Podzols	PZ	PODZOLS	
Pf	Ferric Podzols	PZf	Ferric Podzols	
Pg	Gleyic Podzols	PZg	Gleyic Podzols	

	FAO74 Legend	FAO90 Legend		Remarks
Code	Name	Code (corr.)	Name	
Ph	Humic Podzols	PZc	Carbic Podzols	
Pl	Leptic Podzols	PZb	Cambic Podzols	
Po	Orthic Podzols	PZh	Haplic Podzols	
Pp	Placic Podzols	PZf	Ferric Podzols	
Q	Arenosols	AR	ARENOSOLS	
Qa	Albic Arenosols	ARa	Albic Arenosols	
Qc	Cambic Arenosols	ARb	Cambic Arenosols	
Qf	Ferralic Arenosols	ARo	Ferralic Arenosols	
Ql	Luvic Arenosols	ARl	Luvic Arenosols	
R	Regosols	RG	REGOSOLS	
Rc	Calcaric Regosols	RGc	Calcaric Regosols	
Rd	Dystric Regosols	RGd	Dystric Regosols	
Re	Eutric Regosols	RGe	Eutric Regosols	
Rx	Gelic Regosols	RGi	Gelic Regosols	
S	Solonetz	SN	SOLONETZ	
Sg	Gleyic Solonetz	SNg	Gleyic Solonetz	
Sm	Mollic Solonetz	SNm	Mollic Solonetz	
So	Orthic Solonetz	SNh	Haplic Solonetz	
T	Andosols	AN	ANDOSOLS	
Th	Humic Andosols	ANu	Umbric Andosols	
Tm	Mollic Andosols	ANm	Mollic Andosols	
To	Ochric Andosols	ANh	Haplic Andosols	
Tv	Vitric Andosols	ANz	Vitric Andosols	
U	Rankers	LPu	Umbric Leptosols	
V	Vertisols	VR	VERTISOLS	
Vc	Chromic Vertisols	VR	VERTISOLS	*undiff. (climatic clustering)*
Vp	Pellic Vertisols	VR	VERTISOLS	*undiff. (climatic clustering)*
W	Planosols	PL	PLANOSOLS	
Wd	Dystric Planosols	PLd	Dystric Planosols	
We	Eutric Planosols	PLe	Eutric Planosols	
Wh	Humic Planosols	PLu	Umbric Planosols	
Wm	Mollic Planosols	PLm	Mollic Planosols	
Ws	Sodic Planosols	PLe	Eutric Planosols	*First approximation*
Wx	Gelic Planosols	PLi	Gelic Planosols	
X	Xerosols	RG	REGOSOLS	*Several options possible*
Xh	Haplic Xerosols	RGc	Calcaric Regosols	*First approximation*
Xk	Calcic Xerosols	CLh	Haplic Calcisols	
Xl	Luvic Xerosols	LVh	Haplic Luvisols	*First approximation*
Xy	Gypsic Xerosols	GY	GYPSISOLS	
Y	Yermosols	RG	REGOSOLS	*First approx. (climatic clustering)*
Yh	Haplic Yermosols	RGc	Calcaric Regosols	
Yk	Calcic Yermosols	CLh	Haplic Calcisols	
Yl	Luvic Yermosols	LVh	Haplic Luvisols	*First approximation*
Yt	Takyric Yermosols	RGe	Eutric Regosols	*First approximation*
Yy	Gypsic Yermosols	GY	GYPSISOLS	
Z	Solonchaks	SC	SOLONCHAKS	
Zg	Gleyic Solonchaks	SCg	Gleyic Solonchaks	
Zm	Mollic Solonchaks	SCm	Mollic Solonchaks	
Zo	Orthic Solonchaks	SCh	Haplic Solonchaks	
Zt	Takyric Solonchaks	SCh	Haplic Solonchaks	

Appendix 12
Revision notes (WISE-databases)

Main methodological changes and improvements between successive versions of WISE-derived spatial databases are summarised in the overview below. Initial versions (1996-2012) were based on analyses of the soil-geographical information of the 'old' Digital Soil Map of the World (see Batjes *et al.* 1995; FAO 1995) as opposed to the present Harmonised World Soil Database (FAO *et al.* 2012), a much smaller number of harmonised soil profile data for the world (up from ~4,350 in 1996 to ~21,000 in 2015), a more limited set of soil properties (3 then versus 19 now), and a less elaborate taxotransfer procedure (e.g., going from a 2-layer to a 7-layer data model, and from 3 to 5 textural classes). The most important improvement, however, is that WISE30sec includes estimates of uncertainty for all soil properties (mean and standard deviation). Main differences with the HWSD-approach are summarised in Table 3 (Section 3.2).

Soil geo-graphical data	WISE version (Reference)	Main elements of taxotransfer procedure
HWSD, 30'x30' [a]		
	WISE30sec, v1.0 (This study)	– Soil profiles[d]: ~21,000 – Layer model: 7, i.e. 5 x 20 cm up to 100 cm depth, and 2 x 50 cm up to 200 cm depth – Textural classes: 5 (SOTER conventions: Coarse, Medium, Medium Fine, Fine, and Very Fine) – Co-variates: climate (Peel *et al.* 2007) – Soil variables: 19 – Measure of uncertainty: mean ± std by map unit; descriptive statistics per soil 'cluster' including 10% and 25% percentiles
DSMW, 5'x5' [b]		
	WISE5min, v1.2 (Batjes 2012)	– Soil profiles: ~11,000 – Layer model: 5 times 20 cm up to 100 cm depth – Textural classes: 5 (SOTER conventions) – Soil variables: 19 – Measure of uncertainty: descriptive statistics by soil 'cluster'
	WISE5min, v1.0 (Batjes 2006)	– Soil Profiles: ~10,250 – Layer model: 5 times 20 cm up to 100 cm depth – Textural classes: 3 (DSMW conventions: Coarse, Medium, Fine) – Soil variables: 19 – Measure of uncertainty: descriptive statistics by soil 'cluster'

Soil geo-graphical data	WISE version (Reference)	Main elements of taxotransfer procedure
DSMW, 0.5°x0.5° [c]		
	WISE30min v3.0 (Batjes 2005)	– Soil Profiles: ~9,600 – Layer model: 2 (0-30 and 30-100 cm) – Textural classes: 3 (DSMW conventions) – Soil variables: 22 soil variables – Measure of uncertainty: descriptive statistics by soil 'cluster' – Binned and un-binned GIS layers
	WISE30min v2.0 (Batjes 2002a)	– Soil Profiles: ~9,600 – Layer model: 2 (0-30 and 30-100 cm) – Textural classes: 3 (DSMW conventions) – Soil variables: 10 soil variables – Measure of uncertainty: descriptive statistics by soil 'cluster'
	WISE30min v1.0 (Batjes 1996a)	– Soil profiles: ~4,350 – Layer model: 2 (0-30 and 30-100 cm) – Textural classes: 3 (coarse, medium, fine; DSMW conventions) – Soil variables: soil organic carbon, carbonate carbon, and pH_{water}

[a] The HWSD was rasterised at 0.5'x0.5' (or 30x30 arc-second) from original polygon maps at scale 1:1 to 1:5 M, hence of variable detail depending on the source polygons maps (FAO *et al.* 2012). For this study, all soil units (for map units still derived from the old DSMW) were correlated to the Revised FAO Legend (1988); see Appendix 11 for first approximation.

[b] Considers soil-geographical information derived from the 5x5 arc minute DSMW (FAO 1995), as rasterised from the 1:5 M scale series of printed Soil Maps of the World (FAO-Unesco 1971-1981).

[c] Spatial data at 0.5°x0.5° were derived at FAO from the original 5'x5' raster DSMW soil-geographical data set with consideration of the full map unit composition (see Batjes *et al.* 1995). Soils units are characterised according to the original FAO-Unesco Legend (1974).

[d] Refers to soil profiles held in successive releases of the harmonised, ISRIC-WISE soil profile database (e.g. Batjes 2009).

Appendix 13
Installation procedure

The data set is provided in one single zip file, called *WISE30sec_v1.zip.* This file can be accessed through http://www.isric.org/data/data-download (with scope of dataset= global). By default, it will be de-compressed (unzipped) to folder *X:\WISE30sec,* where *X* is the actual folder*.*

The zip file includes a raster GIS file (WISE30sec) as well as a range of soil attribute data files (see Appendix 3 of the documentation). Using ArcGIS® or similar, users may join the raster data to the derived soil properties files, as managed the MSAccess® database (WISE30sec.mdb).

Linkage is through the map unit code or grid cell identifier (*NEWSUID*) of the raster set and the *NEWSUID* of the various soil attribute data files.

Depending on the proposed applications, users may select the appropriate data set(s) with due consideration for the issues raised in section 3.4 on 'appropriate use of the derived data': comprehensive studies should consider the full mapping unit composition. Methodological and technical details are provided in the documentation.

Citation:
Batjes NH 2015. World soil property estimates for broad-scale modelling (WISE30sec, ver. 1.0). Report 2015/01, ISRIC—World Soil Information, Wageningen (available via: http://www.isric.org/data/data-download)

ISRIC – World Soil Information has a mandate to serve the international community as custodian of global soil information and to increase awareness and understanding of soils in major global issues.

More information: www.isric.org

ISRIC – World Soil Information has a strategic association
with Wageningen UR (University & Research centre)