

S-WORLD: A GLOBAL SOIL MAP FOR ENVIRONMENTAL MODELLING

Jetse J. Stoorvogel^{1*}, Michel Bakkenes², Arnaud J. A. M. Temme¹, Niels H. Batjes³, Ben J. E. ten Brink²¹Soil Geography and Landscape Group, Wageningen University, Wageningen, The Netherlands²PBL Netherlands Environmental Assessment Agency, The Hague, The Netherlands³ISRIC – World Soil Information, Wageningen, The Netherlands

Received 19 September 2016; Revised 21 October 2016; Accepted 21 October 2016

ABSTRACT

The research community increasingly analyses global environmental problems like climate change and desertification with models. These global environmental modelling studies require global, high resolution, spatially exhaustive, and quantitative data describing the soil profile. This study aimed to develop a pedological approach that takes stock of available legacy and auxiliary data to create a global, 30 arc second soil property database for modelling. The methodology uses the Harmonized World Soil Database and the ISRIC-WISE 3.1 soil profile database. Auxiliary information at 30 arc second resolution for various landscape properties is used to describe the variation in landscape properties (temperature and precipitation, topography, elevation, land use, and land cover). Complex mapping units of the HWSD were first disaggregated using a digital elevation model and toposquences to generate delineated areas described by a single soil type. Secondly, ranges of soil properties per soil type were determined using the soil profile data. Then a meta-analysis on a broad literature survey was used to develop a simple model that, based on landscape properties at a particular location, determines the position within these ranges and thus provides an estimation of the local soil properties. Finally, the model was implemented at the global level to determine the distribution of soil properties. The methodology, denominated S-World (Soils of the world) resulted in readily available, high resolution, global soil property maps that are now available for environmental modelling. © 2016 The Authors Journal of Behavioral Decision Making Published by John Wiley & Sons Ltd.

KEY WORDS: GlobalSoilMap; digital soil mapping; Rwanda; soil property maps

INTRODUCTION

A range of international bodies like the United Nations and the Rio Conventions set international targets like the Sustainable Development Goals and specific targets to combat poverty, climate change and desertification. Many of these targets are directly related to our natural environment, including agricultural production, soil conservation, carbon sequestration, and the preservation of biodiversity. Global land use models like IMAGE (Integrated Model to Assess the Global Environment; Stehfest *et al.*, 2014) and LPJ (the Lund-Potsdam-Jena Dynamic Global Model; Bondeau *et al.*, 2007) are currently being used to assess environmental impacts and to study environmental change. In addition, regional and local models are being applied at the global level to assess specific processes like soil erosion (Arnold *et al.*, 1998; Borrelli *et al.*, 2016), soil conservation (Panagos *et al.*, 2016), soil restoration (Fava *et al.*, 2016), and crop growth (Bindraban *et al.*, 2012; Asseng *et al.*, 2013). All these studies require adequate environmental data (Brevik *et al.*, 2015; Bouma & Montanarella, 2016) but are seriously hampered by the lack of coherent, global, high resolution soil data (Sanchez *et al.*, 2009; Brevik *et al.*, 2016). This study focuses on closing a part of this data gap.

Global environmental modelling studies put a number of specific requirements on soil data. Although data requirements vary between studies, generally soil data are required that are i) spatially exhaustive at the global level, ii) at resolutions varying between 30 arc minute down to 30 arc second, iii) quantitative, iv) a description of the entire soil profile, and v) readily available or derivatives from available data.

In the nineteenth century the first soil surveys were made and Vasily Dokuchaev and his school started to look in a structured way at soil geography (Kellogg, 1974). Since then, soil surveys have become common good (Hartemink *et al.*, 2013). Many countries have, at different resolutions, carried out national soil surveys. These national surveys form the basis of many global studies. However, despite the plethora of soil surveys, some 70% of the land surface has yet to be mapped at scales larger than 1:1 million (Nachtergaele & Van Ranst, 2003; Omuto *et al.*, 2013).

Almost half a century ago, FAO initiated the development of a global soil database through the publication of the Soil Map of the World (FAO & Unesco, 1971-1981) in which a large number of national and regional soil surveys were combined into a single printed soil map, and ultimately in the Digital Soil Map of the World (DSMW; FAO, 1995), using a common legend. Recently, a consortium published an update of the DSMW as the Harmonized World Soil Database (HWSD; FAO *et al.*, 2012). Other initiatives aimed at safeguarding and disclosing soil legacy data like: i) the Joint Research Centre and partners made many

*Correspondence to: Jetse J. Stoorvogel, Soil Geography and Landscape Group, Wageningen University, P.O. Box 47, 6700 AA Wageningen, The Netherlands.
E-mail: jetse.stoorvogel@wur.nl

national soil maps accessible through the European Digital Archive on Soil Maps of the World (Panagos *et al.*, 2011), ii) ISRIC - World Soil Information developed standardized procedures to improve national soil maps in the framework of the SOTER program (Igue *et al.*, 2004; Goyens *et al.* 2007); this also generated a large global database with harmonised soil profile information for a World Inventory of Soil Emission Potentials (WISE; Batjes, 2009), and iii) the Web Soil Survey and the Soil Data Access Web Service of the National Resources Conservation Service of the US Department of Agriculture (www.nrcs.usda.org).

Various derived products were developed in response to the call from global modelling groups for quantitative and global soil property estimates. An early example is the land surface data for global circulation models provided by Webb *et al.* (1993), which drew on the old Soil Map of the World and a still limited set of soil profile data for the world. More recently, representative values for key soil parameters have been presented for the major soil types of the FAO/Unesco Soil Map of the World respectively the HWSD (Dunne & Willmott, 1996; Batjes, 1997; FAO *et al.*, 2012; Batjes, 2016). Other groups have attempted to circumvent the need for time consuming correlation to a common (FAO) legend using data assimilation (Shangguan *et al.*, 2014), albeit with varying levels of success.

In the last decade new quantitative techniques for soil survey have been developed where mapping is based on an analysis of relationships between soil properties and a series of co-variables (often related to the soil forming factors). This analysis is often referred to as Digital Soil Mapping (DSM) and uses a range of techniques including generalised linear models, classification and regression trees, neural networks, fuzzy systems and geostatistics (McBratney *et al.*, 2003). Although DSM was originally developed and implemented at detailed scale levels to support *e.g.*, precision agriculture as illustrated by López-Granados *et al.* (2005), it has also been implemented at the landscape level (Mora-Vallejo *et al.*, 2008; Stoorvogel *et al.*, 2009; Holleran *et al.*, 2015; Miller *et al.*, 2015). Recently the potential of DSM has been illustrated, for example, for Africa (<http://africasoils.net>) and Australia (Viscarra Rossel *et al.*, 2014), partly in support of the GlobalSoilMap project (Sanchez *et al.*, 2009; Arrouays *et al.*, 2014).

DSM techniques rely to a large extent on a good, detailed coverage of point-based soil profile data in combination with auxiliary data at the appropriate (spatial and temporal) resolution. Unfortunately, it is realistic to expect that it may take some time to get a proper global coverage of profile data. In addition, the necessary auxiliary data, particularly those describing the variation in parent material are often not yet available. Recently, Hengl *et al.* (2014) have developed an approach that uses available 'shared' point data and geo-statistical models to create surfaces of soil properties for the world at 1 km resolution, and subsequently at 250 m (<https://soilgrids.org>). The methodology, denominated SoilGrids, allows for automatic updates at increased resolutions of these surfaces as new point data

and co-variate data become available or alternative geostatistical models are introduced.

From a pedological point of view, the HWSD provides the most complete global soil description in which many regional or national soil maps have been brought together. Nevertheless, the database as it stands presents three major limitations. Firstly, many mapping units are described as soil complexes or soil associations, whereas many models require a single soil profile description for a particular grid cell or location. Secondly, like many broad scale soil maps, the HWSD uses soil types to describe regional soil variability. Soil classification has proven to be a useful concept in the past when the general purpose soil survey data was used for a variety of applications often with a qualitative, descriptive character. However, the qualitative soil classification is not very suitable for environmental modelling. The third limitation is that soil types explain only a small part of the actual variation in soil properties. Soil classification systems like Soil Taxonomy (Soil Survey Staff, 2014) and the World Reference Base (IUSS Working Group WRB, 2015) typically emphasize pedogenesis, placing specific soil forming processes, like clay illuviation, high in the hierarchy and placing less emphasis on the key, functional soil properties.

Here, we propose a methodology denominated S-World (Soils of the World) that aims to resolve specific issues like complex mapping units and qualitative soil types. The methodology makes use of available legacy and auxiliary data to disaggregate large, compound mapping units with a view to provide insight on the spatial variability of soil properties within these units. As the methodology is rooted in the traditional soil survey, we will refer to it as a pedological method rather than the more statistically based techniques in DSM. We will start with a conceptual framework in which we explain the concepts of the S-World methodology. Subsequently, we will describe its implementation at the global level based on the HWSD and WISE databases. Although the calculations are carried out at the global level, results for Rwanda are presented here as an example as this small country exhibits significant soil variation (Imerzekene & van Ranst, 2001).

CONCEPTUAL FRAMEWORK

The conceptual framework for the development of S-World is described in Figure 1. It aims to disaggregate an exploratory soil map into numerous soil property maps. For the disaggregation, soil profile data, auxiliary data (*e.g.*, a digital elevation model), literature and pedological insight are used. The analysis is done in 5 steps: i) disaggregation of complex mapping units (*i.e.*, mapping units described by more than one soil type) into mapping units that are described by a single soil type, ii) determination of the ranges of soil properties per soil type based on soil profile data, iii) determination of the ranges of landscape properties per soil type based on auxiliary data, iv) development of a global model to assess soil properties based on the soil types and landscape properties, and v) implementation of the model to create

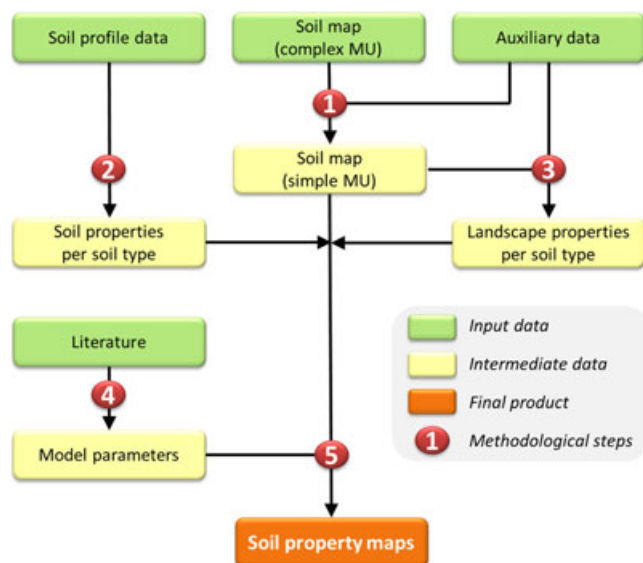


Figure 1. The conceptual model of S-World with its 5 main steps. This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

global maps of derived soil properties. Each of these steps will be described in the following sections.

Disaggregation of Complex Mapping Units

The development of exploratory soil surveys automatically leads to the creation of mapping units (MUs) described by several components. Particularly in the case of soil associations (*i.e.*, a MU with multiple soil types that could have been mapped individually at more detailed map scales) these soil types are not randomly distributed throughout the mapping units but, in many cases, along typical toposequences (Soil Survey Division Staff, 1993). This implies that the soil types are predominantly located, or formed, on specific geographic positions which can be characterised by *e.g.*, elevation and slope. The wide availability of global digital elevation models provides new opportunities to disaggregate soil associations. Soil types are first ordered on the basis of their expected relative topographic position in terms of elevation. For the ordering first the key soil properties (*e.g.*, soil drainage class) that are related to topography and included in the soil classification are identified. The soil forming factors are subsequently ordered resulting in a hierarchy that relates them to elevation. Conceptually, the ordering can be seen as the development of a toposequence. This toposequence can be used to disaggregate the delineated areas after overlaying the delineated areas of the soil map with a digital elevation model (Figure 2). The delineated areas are filled from the low areas up to the higher areas; the process starts with the soil type (from the complex mapping unit) that is located lowest in the toposequence until the area percentage of that soil type (as defined in the description of the complex mapping unit) is reached. Filling continues sequentially with the other soil types following the toposequence until the entire delineated area is represented. The disaggregation results in a new soil map with simple mapping units that are described by a single soil type. The efficacy of the

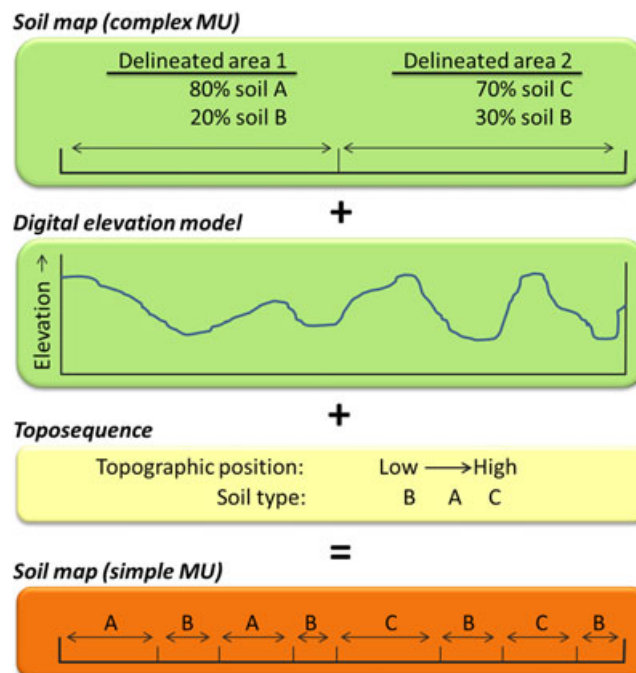


Figure 2. The process of disaggregation of complex mapping units in S-World. This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

methodology for a delineated area depends on the differences between soil types, the relation with altitude, and topography. The advantage of the methodology is that it maintains the proper areas of the soil types.

Determination of the Ranges of Soil Properties Per Soil Type

Given the pedogenetic character of many soil classification systems like the Soil Taxonomy and the revised legend of the Soil Map of the World, the variation in functional properties such as soil organic matter within a soil type can be large. Large soil profile databases are required to describe the distribution of soil properties per soil type in statistical terms. In some cases, these databases are available for soil properties that can be described morphometrically or by field estimations (*e.g.*, topsoil thickness). However, many chemical and physical soil properties require specific laboratory analyses and are often determined for a limited number of representative soil profiles only. Besides the limited number of observations, and regional differences in soil analytical methods used, many soil properties are not normally distributed. This hampers the description of the variability of soil properties per soil type. One option is to derive the minimum and maximum values reported for a given soil property. However, soil observations or measurements with a point support are expected to show more variation than larger grid cells in global soil property maps. In addition, outliers may occur in the database. To avoid an overestimation of variation, we therefore pragmatically use the majority of available values as defined here by the 1st and 9th decile. In other words, the possible range for variable *i* for soil type *s* is defined as $[v_{is1}, v_{is9}]$ where v_{is1} corresponds to the 1st decile and v_{is9} to the 9th decile. Prior to the determination

of the deciles, we verified whether the variables have a skewed distribution in which case the variables can be transformed (e.g., log-transformed) to arrive at a more normal distribution.

Determination of the Ranges of Landscape Properties Per Soil Type

Just like digital soil mapping (McBratney *et al.*, 2003), S-World uses auxiliary data related to soil forming factors as defined by Jenny (1941): parent material, time, climate, topography, and organisms. The soil classification systems provide information on parent material and time through the soil types. For the other three soil forming factors auxiliary information is available from climate databases, digital elevation models and satellite imagery. The landscape properties described by the auxiliary information (e.g., temperature, rainfall, slope, land cover) are used to explain the spatial variation in soil properties within a particular soil type. For this, the ranges in landscape properties need to be derived per soil type. S-World characterizes the variation in landscape conditions for each soil type by assessing the 1st and 9th decile for each landscape property. The choice of the 1st and 9th decile is arguably subjective. However, there are good reasons to limit the range between the minimum and maximum value reported for the landscape properties. Small positional errors in the databases may easily lead to some extreme values. Similar to the variability in soil properties, landscape properties may present a range of different distributions prohibiting the use of a mean and the standard deviation. Using the percentiles seems to be a logical, but pragmatic, choice. The deciles are derived through an overlay of the disaggregated soil map with the simple mapping units and the maps with auxiliary information. The observed range for landscape property j on soil s is defined as $[p_{js1}..p_{js9}]$ in which p_{js1} corresponds to the 1st decile and p_{js9} to the 9th decile.

Development of a Global Model to Assess Soil Properties

Large soil databases to derive statistical relations between the landscape properties and the soil properties are lacking. One option is to reduce the number of mapped soil types (Kempen *et al.*, 2010) by considering higher levels in the soil classification system (e.g., ignoring qualifiers and using Soil Groups). However, the lower levels may provide essential information on key soil properties like soil organic matter (e.g., Umbric or Mollic), soil texture (e.g., Arenic or Vertic), or drainage (e.g. Gleyic). It is desirable to maintain the intrinsic information in the soil classification as far as possible and base the methodology on the limited available point data. Therefore, an alternative, expert-based model correlates landscape properties to the soil properties for each of the soil types. Conceptually, this model aims to down-scale a certain soil property i with a known range based on n landscape properties or explanatory factors $[p_1, p_2 \dots p_n]$. We estimate v_i at location x where we find soil s as $v_{isx} = v_{is1} + w_{ix} * (v_{is9} - v_{is1})$. with w_{ix} being a weight $w \in [0..1]$ that determines where v is in the range $[v_{is1}..v_{is9}]$. Different

explanatory factors represented by the landscape properties determine w . Weight w at location x for soil property i is calculated as a linear combination of weights for the individual landscape properties:

$$w_{ix} = \sum_{j=1}^n w_{ijx}.$$

The weight w_{ijx} for landscape property j is calculated as:

$$\begin{aligned} \text{Case : } c_{ij} \geq 0 : w_{ijx} &= c_{ij} * \frac{(p_{jx} - p_{js1})}{(p_{js9} - p_{js1})} \\ c_{ij} < 0 : w_{ijx} &= c_{ij} * \frac{(p_{jx} - p_{js9})}{(p_{js9} - p_{js1})} \end{aligned}$$

In which c_{ij} is a constant that indicates the relative importance of the landscape property j for soil property i . Landscape property j on location x is described by p_{jx} . The sign of c_{ij} indicates whether there is a positive or negative relationship between the landscape property and the soil property. When $\sum_{j=1}^n |c_{ij}| = 1$ and all $w \in [0..1]$ then all values

in the range $[v_{is1}..v_{is9}]$ are possible based on the landscape properties. Although in practice c is specific for each landscape property, soil type, and soil property, data are lacking to estimate c at that level of specificity. Therefore the model assumes that c is constant per soil type and landscape property, or, in other words, the relative impact of landscape properties on a specific soil property is assumed to be constant over the different soil types. The values for c_{ij} can be derived statistically or through a meta-analysis. However, currently in most cases the number of observations will not permit such an analysis. Therefore, for the moment we based the values for c_{ij} on expert knowledge and a general meta-analysis of literature.

Implementation of the Model to Create Global Surfaces of Soil Properties

In theory, the global model can be applied to any of the soil properties that are considered in the soil profile databases. After disaggregating the complex mapping units (Step 1), each grid cell is characterized by a single soil type characterized by the inferred ranges in soil properties and various landscape properties. Where the values of landscape properties fall outside the range $[p_{js1}..p_{js9}]$, they are truncated:

$$\begin{aligned} \text{case : } p_{jx} < p_{js1} : p_{jx} &= p_{js1} \\ p_{jx} > p_{js9} : p_{jx} &= p_{js9} \end{aligned}$$

Theoretically, the model can be implemented for each soil property. However, for some properties estimation of the model parameters can be difficult due to lack of appropriate literature, and for some other properties, a meaningful relation with landscape properties within soil types is absent. A categorisation for soil properties is therefore important: i) soil properties for which point data are available and that correlate to the available explanatory data, ii) soil properties

that do not correlate to the available explanatory data but that do vary between soil types, and iii) soil properties for which there are almost no measured data. For the first group the above procedures can be followed, while for the second group averages per soil type can be derived and linked to the disaggregated soil map. The last group typically includes soil physical properties like water retention and conductivity. These soil properties are often derived using pedo-transfer functions as presented by Saxton & Rawls (2006) or Romero *et al.* (2012). Combining results of the analysis for the three groups of derived soil properties will provide a complete set of soil properties.

S-World Implementation

The conceptual framework was implemented to create global surfaces of soil properties suitable for global environmental modelling. While all calculations were carried out at the global level, Rwanda will be taken as an example country to present some of the results. The 26,338 km² country is small but shows considerable variation in altitude (between 950 and 4507 meters above sea level) and soil conditions.

Legacy Data

The starting point for our analysis was the Harmonized World Soil Database (HWSD; FAO *et al.*, 2012). The HWSD is based on updated materials for large parts of the world while the remainder is still based on the DSMW (FAO, 1995). The HWSD classified soils using the original (FAO & Unesco, 1974) or revised (FAO *et al.*, 1990) legend depending on the source materials. Many of the mapping units are described by complex mapping units with 2–10 different soil types. The HWSD has a nominal resolution of 30 arc second, corresponding to cells with a surface of roughly 1 km² at the equator. However, as it had essentially been derived from soil maps at a scale of 1:1 to 1:5 million, the HWSD presented many large delineated areas which were assumed to have the same soil composition throughout.

A large global soil profile dataset is WISE3.1 (Batjes, 2009) which contains quantitative soil horizon information for 10,253 soil profiles from 149 countries. Not all the soil profiles have full analytical data as these have not always been collected during the underlying soil surveys. Therefore, a subset of soil profiles was used to estimate the ranges of soil properties for S-World. The revised legend of the Soil Map of the World was used (FAO and Unesco, 1990). HWSD soil classifications based on the DSMW were reclassified to the revised (1990) legend. Although this reclassification encompassed some generalisation or loss of detail (Bridges *et al.*, 1998), it allowed for a link between the WISE3.1 database and the HWSD using the 1990 legend. The WISE database was not developed for direct interpolation using geo-locations, rather to represent major FAO soil types. To assess the distribution of soil properties it is nonetheless important that the available profiles are well spread over the earth's land surface, which is not yet the case. For example, there are still few profiles for Europe,

Northern Asia and Canada in WISE3.1. This range will be expanded in the framework of the Global Soil Partnership (GSP, 2014) and the WoSIS initiative (Ribeiro *et al.*, 2015).

Auxiliary Data

Auxiliary data were used in S-World to describe three soil forming factors: topography, climate, and organisms. Various digital elevation models provide topographic information to spatially disaggregate the complex mapping units of HWSD. For most parts of the world (except the areas north of 60°N and south of 60°S) data from the Shuttle Radar Topography Mission (SRTM) provide elevation data at a 90 m resolution (Jarvis *et al.*, 2008). The GTOPO30 database (<http://edcdaac.usgs.gov/gtopo30/gtopo30.asp>) is available for the entire globe at a resolution of 30 arc minute. Recently, this database was updated into the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) as described by Danielson & Gesch (2011). GMTED2010 was selected as the most appropriate DEM for S-World given its global coverage and spatial resolution (GMTED2010 also provides the data at a resolution of 30 arc second). Topography was interpreted here as slope. However, slopes derived from a 30 arc second grid do not represent actual topography. Since the development of high resolution (*e.g.*, 30 – 90 m) DEMs, it is now possible to derive the variation in altitude for the 30 arc second grid cells. GMTED2010 provides the standard deviation in elevation for each of the grid cells. S-World uses this standard deviation in elevation as an indicator of topography in a grid cell.

For climate, mean annual temperature and rainfall were used from the WorldClim database (Hijmans *et al.*, 2005). The WorldClim database is the result of the interpolation of mean annual rainfall data from 47,554 weather stations and mean annual temperature from 14,835 weather stations. In WorldClim, weather data are aggregated over several decades of observations, and elevation is used as an additional spatial explanatory variable in the interpolation.

At the broad scale of S-World, the soil forming factor “organisms” was interpreted as land cover and land use. Two databases were used, GlobCov and NDVI. The GlobCov database (Bontemps *et al.*, 2010) includes current land use at a 10 arc second resolution. The land use classes of the GlobCov database were translated into a land use intensity value ranging from 0% in natural areas with no (or almost no) human influence to 100% in areas with very intensive management under arable farming where soil may be tilled, organic matter (crop product and, possibly, crop residues) may be removed, and the soil may be bare for part of the year. The intermediate classes include less intensive forms of land use (*e.g.*, pasture) or mixed cropping. The reclassification is provided in Table I. After the reclassification the global grid was aggregated to a resolution of 30 arc second by averaging the land use intensities in each of the larger grid cells. Besides the land use intensity the actual land cover is being used. Land cover has a dual effect. A good vegetative cover protects land against erosion. In addition,

Table I. Reclassification of the GLOBCOV database in terms of land use intensity

Description	Land use intensity (%)
Post-flooding or irrigated croplands (or aquatic)	75
Rainfed croplands	100
Mosaic cropland (50-70%) / vegetation (grassland/shrubland/forest) (20-50%)	75
Mosaic vegetation (grassland/shrubland/forest) (50-70%) / cropland (20-50%)	50
Artificial surfaces and associated areas (Urban areas >50%)	75
Natural vegetation categories	0
Water bodies	0
Permanent snow and ice	0
No data (burnt areas, clouds,...)	0

the vegetation will provide organic matter to the soil. Land cover is described by the Normalized Difference Vegetation Index (NDVI). Five years (2006-2010) of monthly 30 arc second resolution SPOT NDVI images (<http://free.vgt.vito.be/>) were downloaded from which an average vegetation index was calculated. The average NDVI was used as a proxy for the vegetative cover of the soil. Illustrations of the input data for the Rwanda example are presented in Figure 3.

Disaggregation of Complex Mapping Units

The HWSO contains slightly over 200,000 delineated areas that are classified as 16,022 different mapping units. Only 2,503 of these delineated areas are described by a single soil type. On average, the other mapping units are described by a combination of 3 soil types. The soil types of the 1990 revised legend were ordered according to their expected topographic position in the landscape based on 4 different criteria: drainage (*e.g.*, Gleysols and Gleyic soil types), salt accumulation (*e.g.*, Solonchaks, Solonetz, and Salic soil types), parent material or specific processes defined in the classification procedures (*e.g.*, Fluvisols, Lithosols), and soil fertility (*e.g.*, Eutric versus Dystric soil types). Overall, the soil classification is related to important soil forming factors as illustrated by Bockheim & Gennadiyev (2000), some of which are strongly correlated to topography. The resulting toposquence is a fictive one as many soils do not occur together in the landscape or in a particular mapping unit. This is particularly so for some of the zonal soils that are linked to specific climatic conditions. However, this does not influence the analysis since the toposquence is only used to order the soil types representing particular mapping units. An overlay of the HWSO with the DEM resulted in smaller areas with equal altitude (mostly individual grid cells). Finally, each of the delineated areas is “filled” with the appropriate soil types starting in the lower parts with the soil types that are found on the lowest topographic positions, followed by the other soil types in the appropriate order along the toposquence. When delineated areas were smaller than 5 grid cells, the major soil type was assumed to represent the entire delineated area. The result of the procedure was a

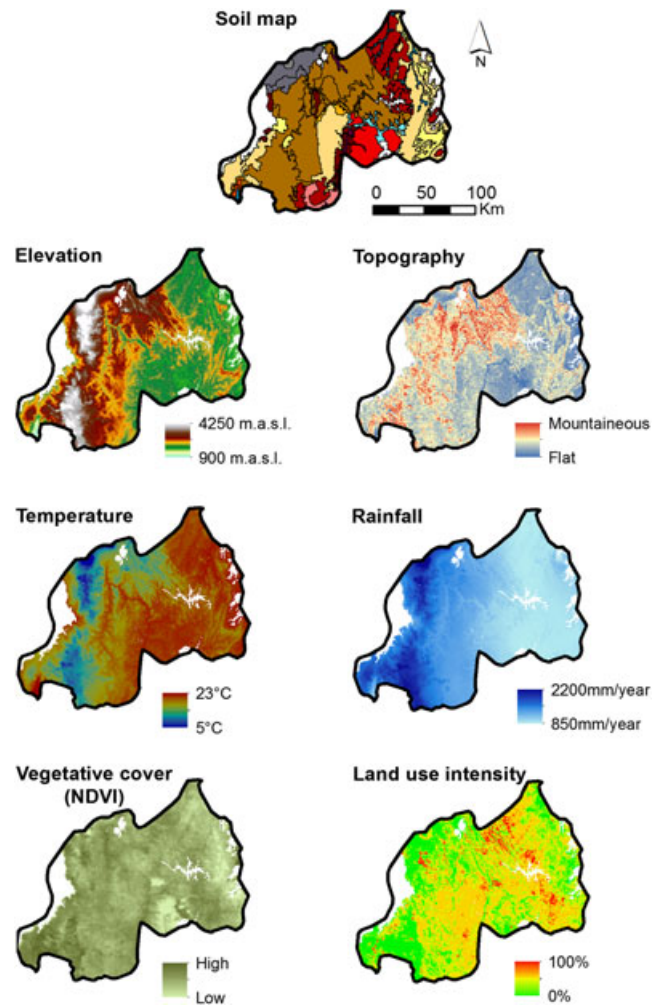


Figure 3. Legacy and auxiliary data used for S-World (sources are presented in the text) using Rwanda as an example. This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

new soil map with a single soil type for each of the delineated areas, as illustrated in Figure 4 for the Rwanda example.

Ranges of Soil Properties Per Soil Type

S-World aims to support environmental modelling. During two expert meetings six key soil properties were identified: i) topsoil thickness (cm), ii) soil depth (cm), iii) soil organic carbon (SOC) content in the top 30 cm (%), iv) SOC content in the subsoil (30 to 120 cm) (%), v) clay content in the soil profile (%), and vi) sand content in the soil profile (%). This selection was a trade-off between data requirements, data availability, and the consideration whether certain soil properties can be adequately described at the global scale. Values for the selected soil properties were taken from soil profiles in WISE3.1 that had sufficient attribute data. Being based on field observations, the profile descriptions did not provide any information about conditions below the deepest horizon sampled. As a result, the exact (physical) depth of the many profiles was unknown. It was assumed that if a soil profile was described to be less than 1.20 m deep, there was a true depth limitation. For many practical applications in

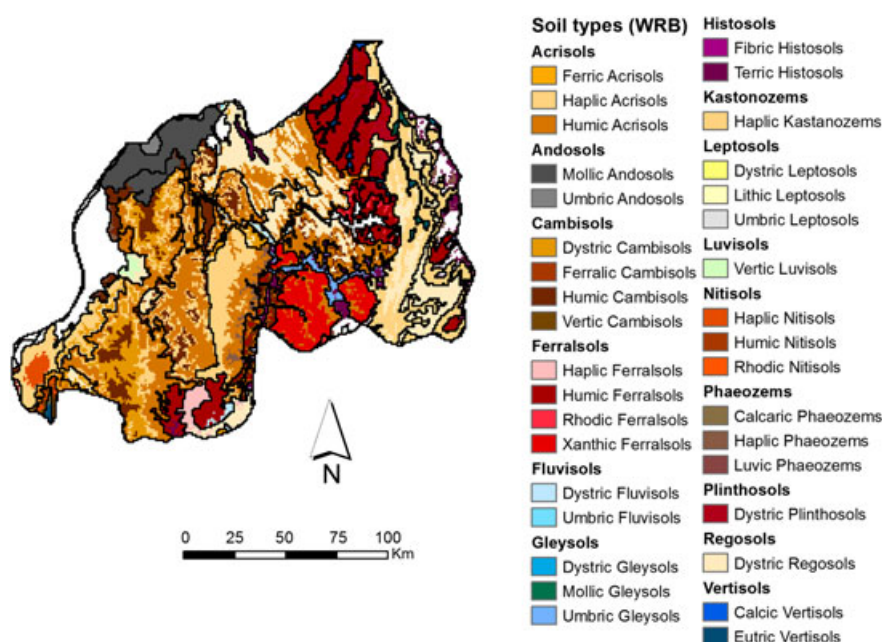


Figure 4. Disaggregated soil map for Rwanda with the lines representing the original delineated areas of the HWSD. This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

environmental modelling (*e.g.*, soil suitability) differences in soil depth beyond 1.20 m are considered of less importance. Therefore, the ‘truncation’ of the deeper soil profiles at 1.20 m was not considered a major problem. Similar challenges were encountered in the definition of the topsoil as horizon codes were sometimes not assigned or named consistently in the WISE database, which is derived from disparate sources. Therefore it was not possible to simply select the A-horizon as corresponding with the topsoil. Following initial tests, all horizons with a SOC content greater than 50% of the SOC content of the first horizon were pooled. Together, these horizons were considered to represent the topsoil. By implication, topsoil thickness could only be determined for soil profiles with complete SOC data (*i.e.*, for all the soil horizons of the soil profile). The procedure provided a proper determination of the topsoil thickness for 89% of the soil profiles for which horizon codes were provided in the ISRIC-Wise database. For clay and sand content, the weighted average was calculated for each layer represented in each truncated soil profile. In view of the above conditions, only 6399 soil profiles (out of 10,253) were available for S-World, which on average corresponds with 36 soil profiles per soil type. For 40 soil types (out of 181) there were less than 10 soil profiles, and 22 soil types even had less than 5 soil profile descriptions. We used the deciles of the corresponding soil groups (rather than soil units) as best estimates (note that the revised legend distinguishes 28 different soil groups) if less than 10 soil profiles were available.

Table II presents the summary statistics per soil type for the soil properties. The explained variance was calculated through an analysis of variance. The RMSD is the Root Mean Square Difference between the actual soil property and the average soil property of the soil type. The soil types

explained a considerable part of the variation in soil properties, but at the same time variation within soil types was still large. This supports the main conceptual idea of S-World to start from the HWSD with its mapped spatial distribution of soil types. The relevance of downscaling the soil map becomes apparent when we look at the ranges in estimated soil properties by soil type. For Rwanda, our example, the most important soil types mapped on the HWSD are the humic and haplic Acrisols, which jointly cover over 30% of the country. Topsoil SOC contents (defined by the 1st and 9th decile) for humic Acrisols range between 1.3 and 7.7%, while this is 0.4 and 3.6% for haplic Acrisols. Although these numbers show clear differences between the two soil types, they also indicate that there is significant variation in SOC content within the soil types. Commonly, coefficients of variation for SOC content in excess of 30% have been reported at series level (Spain *et al.*, 1983). However, this variation is not necessarily found within Rwanda itself. The soil types were identified by the HWSD under a range of different locations all over the world (see Table III).

Table II. Summary statistics from the analysis of soil properties over the different soil types based on the WISE database

		Average	Explained variance (%)	RMSD
Thickness (cm)	Topsoil	35	13	29
	Total soil	106	38	19
Soil organic carbon (%)	Topsoil	2.2	34	2.5
	0-50 cm	1.4	54	1.3
Soil texture (%)	Subsoil	0.6	43	1.1
	Sand	43	39	21
	Clay	31	43	14

Table III. Modelled variation in landscape properties for the humic and haplic Acrisols at the global level and within Rwanda

Soil type	Topography (m)		Rainfall (mm/yr)		Temperature (°)		NDVI ^a		Land use intensity	
	Low	High	Low	High	Low	High	Low	High	Low	High
Globally (1 st and 9 th deciles)										
- humic Acrisols	0	10	1200	3264	14.9	22.6	138	200	0	100
- haplic Acrisols	0	5	807	1669	14.6	26.6	141	200	0	100
Rwanda (min max)										
- humic Acrisols	0	16	867	1690	15.2	22.1	52	200	0	100
- haplic Acrisols	0	16	859	1751	14.7	21.9	52	200	0	100

^aSPOT expresses NDVI as an index value between 0 and 255. The values were truncated to the 50-200 range since the low and large values represent little variation in vegetation.

Ranges of Landscape Properties Per Soil Type

Five landscape properties were considered in the present study to explain the variation in soil properties within the different soil types: topography, rainfall, average annual temperature, land cover, and land use intensity. The variation in these landscape properties was determined for each of the soil types. Similar to the variation in soil properties per soil type, a large variation in landscape properties per soil type was observed. As an illustration the variation in landscape properties for the two main soil types of Rwanda are presented in Table III. Humic Acrisols were found under more humid conditions than the haplic Acrisols. However, the conditions under which these soil types were found in Rwanda may differ from the overall conditions under which similar soil types were found globally. For example, in different areas the same soil types may be associated with steeper slopes or with other NDVI classes.

Towards a Global Model for Soil Formation

S-World aimed to disaggregate the variation in soil properties using the variation in landscape properties to support studies of environmental change (see Figure 1). Due to the lack of data it was difficult to develop a complex model describing the relation between landscape properties and soil properties at a broad scale. The model that was developed uses a number of constants c that indicate the explanatory power of each of the landscape properties on the soil properties. Although in practice c is specific for each landscape property, soil type, and soil property, data were lacking to estimate c at that level of specificity. Therefore, the relative impact of landscape properties on a specific soil property was assumed to be constant over the different soil types. Core of the model was the assessment of the constants c_{pv} . A literature study, similar to the one by Guo & Gifford (2002), was adopted. Basically, a large number of studies were identified in which the effect of a landscape property on a soil property was analysed. For each combination of soil and landscape property, we aimed for at least 6 different studies if there was no meta-analysis. The complete literature study is provided in Table S1.

Depth of the topsoil was defined on the basis of the soil organic matter profile. As a result it was a function of organic matter input, mineralization rates and erosion/

deposition rates. Increased temperatures generally lead to a higher net primary production but also result in higher mineralization rates. An increase in rainfall also increases net primary production but leads to a decline in mineralization rates (lower oxygen concentrations in the soil). Steep slopes will lead to erosion and thus a removal of topsoil which may be redeposited elsewhere, subject to land use type and intensity. The necessary information will have to be derived from auxiliary data sources, as depicted in Figure 1.

Topsoil soil organic carbon contents are driven by similar processes like the depth of the topsoil. However their relative impacts are very different. Increases in temperature and rainfall will generally lead to an increase in net primary production. Increased mineralization may offset the positive effect of the increased organic matter input. Steep slopes may result in topsoil removal. Depending on the SOC profile in the soil this may result in no changes in the case of well homogenized topsoils to a reduction in topsoil SOC if there is a rapid decline in SOC over depth. Land use intensity is known as a major driving force in topsoil SOC. After clearing a natural vegetation topsoil SOC often rapidly declines.

Soil depth is determined by the soil formation and the parent material. Both parameters are typically not available at the global scale, although they are to a certain extent incorporated in the soil classification. Positive conditions for soil formation and biological activity like temperature and rainfall may help to increase soil depth. On the other hand, soil erosion will lead to a reduction in soil depth. The effect of land use intensity is relatively limited although, just like in the case of topsoil depth, intensification of land on sloping lands may result in increased erosion and consequently a reduction in soil depth.

Soil texture is probably the most stable factor of the 4 soil properties under consideration. Soil texture is in most cases strongly related to sedimentation or geological formations. Although increases in temperature and rainfall may result in increased weathering rates and thus finer soil textures, the effect will be rather limited. Slope and land use intensity may result in coarser textures due to the preferential erosion of finer particle sizes.

Results of the literature survey are summarised in Table IV which lists the model parameters adopted for the analysis. A large value indicates that the landscape property

Table IV. Relative weights for different landscape properties on their impact on 7 different soil properties estimated using a literature survey and expert knowledge

		Annual temperature	Annual rainfall	Slope	Land use intensity	NDVI
Thickness (cm)	Topsoil	0.2	0.2	-0.4	-0.1	0.1
	Soil	0.2	0.2	-0.4	-0.1	0.1
Soil organic carbon (%)	Topsoil	0.1	0.3	-0.1	-0.3	0.2
	0-30 cm	0.1	0.3	-0.1	-0.3	0.2
	Subsoil	0.2	0.2	-0.4	-0.1	-0.1
Soil texture	Sand (%)	0.3	0.3	-0.3	0	0.1
	Clay	-0.3	-0.3	0.3	0	-0.1

plays an important role in explaining the variation in soil properties within a soil type, while positive values mean that there is a positive relationship between landscape property and soil property.

Application of the Model

The actual model uses all the inputs, and assumptions, as described in the previous sections including the disaggregated soil map, the 5 maps of the landscape properties, the deciles of soil properties and landscape properties per soil type, and the various model parameters. As such, the model disaggregates the original HWSD map units to grid cells where the landscape properties now determine the derived soil properties. Figures 5 and 6 show the results for mean topsoil SOC and clay contents for Rwanda and the world, as an example. The disaggregated soil units can still be recognised. Some of the soil units show little variation in their properties, which can be caused by a narrow (measured) range in soil properties for that particular soil unit or by limited variation in the landscape properties. In the future, these results should be compared with output generated using other mapping approaches as discussed earlier in this paper.

Validation

Validation of global databases like S-World is extremely difficult due to i) the global scale, ii) the support of the data, and iii) independent data. Many existing datasets have been used for the development of national soil surveys which form the basis of the HWSD. At the same time S-World predicts soil properties at the support of grid cells whereas data are typically observed for individual points. Nevertheless there is an urgent call to provide quality assessment to global

soil databases. Typically the validation requires a soil database with a proper global coverage, that is independent to regional or national surveys and that provides data at the same resolution as the data provided by S-World (30 arc seconds). Data at the 30 arc second resolution are extremely difficult to find, but one could establish an independent dataset of point observations and use that to validate the S-World database. To evaluate the improvement of the S-World database, we established a single database from independent studies from Costa Rica ($n=193$), Ecuador ($n=95$), Egypt ($n=22$), Ethiopia ($n=67$), Ivory Coast ($n=31$), Kenya ($n=256$), Morocco ($n=67$), Mozambique ($n=108$), Netherlands ($n=127$), Peru ($n=93$), Philippines ($n=124$), Russia ($n=178$), Senegal ($n=110$), Uganda ($n=54$), USA ($n=120$), and Vietnam ($n=45$) (Table S2). All 1690 georeferenced observations included data on topsoil organic carbon or organic matter. We compared the samples to the predictions of S-World and the HWSD by calculating the Root Mean Squared Difference. The HWSD provides average SOC contents for the soil types in the mapping units, but the soil types are not geo-referenced within the mapping units. Two different approaches were followed: i) the dominant soil type for the mapping unit ($\text{HWSD}_{\text{dominant}}$), and ii) the weighted average for all soil types within the mapping unit ($\text{HWSD}_{\text{weighted}}$). The RMSD for S-World was considerably smaller (2.1% SOC) than the RMSD for $\text{HWSD}_{\text{weighted}}$ (2.9% SOC) and $\text{HWSD}_{\text{dominant}}$ (3.6% SOC). However, still large deviations were found despite the fact that there were almost no organic soils in the validation dataset. Notwithstanding some of the limitations of the validation, it suggests that S-World improves the estimates of SOC contents.

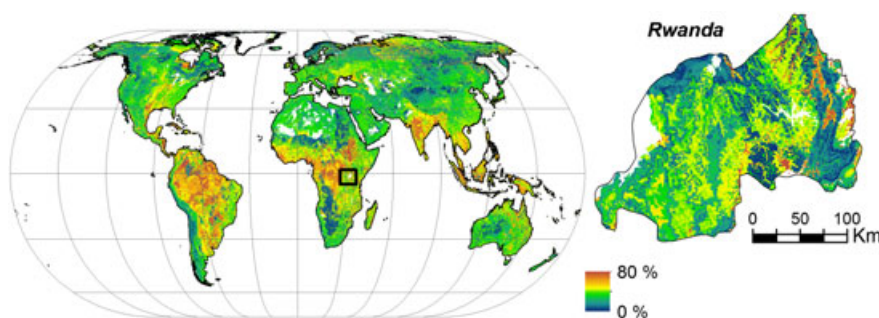


Figure 5. S-World results for modelled clay (%) for the world and Rwanda (white areas represent areas with no data as there is no soil). This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

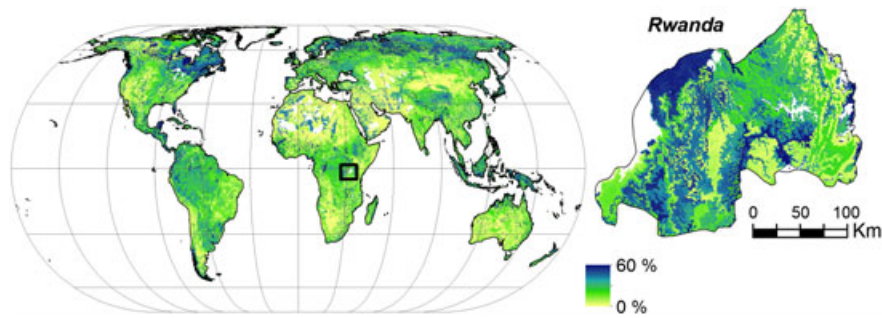


Figure 6. S-World results for modelled SOC (%) for the world and Rwanda (topsoil) (white areas represent areas with no data as there is no soil). This figure is available in colour online at wileyonlinelibrary.com/journal/ldr

DISCUSSION

S-World uses the basic concepts of digital soil mapping where extensive use of legacy and auxiliary data allows us to develop a disaggregated, soil property database for the world. The global scale level poses a number of challenges. These are mostly related to data availability. At present, the available legacy data do not, for example, allow for the derivation of statistical relationships between the landscape and soil properties for all soil types mapped in the HWSD. Such data limitations can be addressed in different ways, ranging from geostatistical interpolation techniques like in SoilGrids1km (Hengl *et al.*, 2014) to the pedological approach proposed in S-World. Each methodology has its own strengths and weaknesses (Minasny & McBratney, 2010; Padarian *et al.*, 2014; Lal *et al.*, 2015; Batjes, 2016). We therefore support the use of ensemble runs in which different datasets are being used in a similar fashion as is being done in climate change research with various models (Porter *et al.*, 2014) or different climate change projections with different earth system models (Todd-Brown *et al.*, 2013; Challinor *et al.*, 2014). These comparisons may serve to identify priority areas for further research and profile data collation, possibly within the broader framework of the Global Soil Partnership (GSP 2014), which may also help to address the pressing issue of restricted data licences.

The data do not allow for a proper validation, which limits a proper quality assessment. Nevertheless, it is important to look at the major sources of error in the S-World methodology. It is important to realize that many parts of the world (on HWSD) have not been mapped at detailed scale levels, making the proposed disaggregation less useful. The HWSD is a compilation of different soil maps of the world, generalised at scales of 1:1 million to 1:5 million. Production of HWSD has been a careful balance between the use of the best data that were made available for a particular region, aiming to be as consistent as possible in terms of resolution and the description of the mapping units. For example, soil mapping units may include 30% of inclusions (in detailed surveys, Soil Survey Division Staff, 1993), and these will not be considered in the present disaggregation procedure. Soil profiles held in large databases for the world, such as WISE and its successor WoSIS (Ribeiro *et al.*, 2015), are

largely based on purposive sampling (*i.e.*, derived from routine (legacy) soil surveys from 1940 to 2010). This is less of a problem for the pedological approach followed here where the soil profiles are used to derive variation in soil types. However, geo-statistical analysis require probabilistic sampling and are seriously constrained by the purposive sampling. Nonetheless, further development of soil profile databases as done by ISRIC (Leenaars, 2013) is often undervalued, but essential and should go hand-in-hand with further model development and intercomparison.

The S-World methodology combines the recently developed innovation in digital soil mapping with intensive use of soil legacy and auxiliary data. The model generates global soil data as it is required for a wide range of global environmental studies like the applications of the IMAGE model (Stehfest *et al.*, 2014). Such global assessments are aimed especially at restoration options, that may be of direct relevance for achieving the three Rio conventions, as well as for the Sustainable Development Goals on food, water and energy security (Bouma & Montanarella 2016; Keesstra *et al.*, 2016; Montanarella *et al.*, 2016). The S-World database adds to an increasingly large suite of different global soil databases (Hengl *et al.*, 2014; Batjes, 2016) providing environmental studies an option to select the most appropriate dataset or carry out ensemble runs with various datasets.

CONCLUSIONS

The soil science community continuously seeks for spatially exhaustive, global maps of soil properties. The rapidly increasing availability of legacy and auxiliary data opens new opportunities to generate global grids of soil properties that can be used in a wide range of environmental models. The S-World methodology provides an operational methodology in which existing general soil maps are disaggregated into 30 arc second grids with soil properties. The methodology is particularly suited to deal with complex mapping units and the variation in soil properties within soil types. The resulting maps show considerable short range variation in soil properties due to differences in topography and land use which is not shown on traditional exploratory soil maps.

ACKNOWLEDGEMENTS

Funding from the Netherlands Environmental Assessment Agency (PBL), the Agricultural Model Intercomparison and Improvement Project (AgMIP), and the Soil and Water Assessment Tool (SWAT) development team for the development of S-World is highly appreciated. The initial methodology is the result of various discussions with staff members from the Netherlands Environmental Assessment Agency, ISRIC-World Soil Information, and Future Water. We would like to thank specifically Kees Klein Goldewijk, Maurits van den Berg, Prem Bindraban, Vincent van Engelen, Zhanguo Bai, Stephan Mantel, Johan Leenaars, and Jelle Brandsma.

REFERENCES

- Arnold JG, Srinivasan R, Muttiah RS, Williams JR. 1998. Large Area Hydrologic Modeling and Assessment: Part I. Model Development. *JAWRA* **34**: 73–89. DOI:10.1111/j.1752-1688.1998.tb05961.x.
- Arrouays D, Grundy MG, Hartemink AE, Hempel JW, Heuvelink GBM, Hong SY, Lagacherie P, Lelyk G, McBratney AB, McKenzie NJ, Mendonca-Santos ML, Minasny B, Montanarella L, Odeh IOA, Sanchez PA, Thompson JA, Walsh MG, Zhang GL. 2014. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. *Advances in Agronomy* **125**: 93–134. DOI:10.1016/B978-0-12-800137-0.00003-0.
- Asseng S, Ewert F, Rosenzweig C, Jones JW, Hatfield JL, Ruane AC, Boote KJ, Thorburn PJ, Rötter RP, Cammarano D, Brisson N, Basso B, Martre P, Aggarwal PK, Angulo C, Bertuzzi P, Biernath C, Challinor AJ, Doltra J, Gayler S, Goldberg R, Grant R, Heng L, Hooker J, Hunt LA, Ingwersen J, Izaurralde RC, Kersebaum KC, Müller C, Naresh Kumar S, Nendel C, O'Leary G, Olesen JE, Osborne TM, Palosuo T, Priesack E, Ripoche D, Semenov MA, Shcherbak I, Steduto P, Stöckle C, Stratonovitch P, Streck T, Supit I, Tao F, Travasso M, Waha K, Wallach D, White JW, Williams JR, Wolf J. 2013. Uncertainty in simulating wheat yields under climate change. *Nature Climate Change* **3**: 827–832. DOI:10.1038/nclimate1916.
- Batjes NH. 1997. A world dataset of derived soil properties by FAO-UNESCO soil unit for global modelling. *Soil Use Management* **13**: 9–16. DOI:10.1111/j.1475-2743.1997.tb00550.x.
- 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. DOI:10.1111/j.1475-2743.2009.00202.x.
- Batjes NH. 2016. Harmonized soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **269**: 61–68. DOI:10.1016/j.geoderma.2016.01.034.
- Bindraban PS, van der Velde M, Ye L, van den Berg M, Materechera S, Kiba DI, Tamene L, Ragnarsdóttir KV, Jongschaap R, Hoogmoed M, Hoogmoed W, van Beek C, van Lynden G. 2012. Assessing the impact of soil degradation on food production. *Current Opinion in Environmental Sustainability* **4**: 478–488. DOI:10.1016/j.cosust.2012.09.015.
- Bockheim JG, Gennadiyev AN. 2000. The role of soil-forming processes in the definition of taxa in Soil Taxonomy and the World Soil Reference Base. *Geoderma* **95**: 53–72. DOI:10.1016/S0016-7061(99)00083-X.
- Bondeau A, Smith PC, Zaehle S, Schaphoff S, Lucht W, Cramer W, Gerten D, Lotze-Campen H, Müller C, Reichstein M, Smith B. 2007. Modelling the role of agriculture for the 20th century global terrestrial carbon balance. *Global Change Biology* **13**: 679–706. DOI:10.1111/j.1365-2486.2006.01305.x.
- Bontemps S, Defourny P, van Bogaert E. 2010. GLOBCOVER 2009, Products description and validation report. ESA and Université Catholique de Louvain. 53 pp. DOI:10.1594/PANGAEA.787668.
- Borrelli P, Panagos P, Ballabio C, Lugato E, Weynants M, Montanarella L. 2016. Towards a Pan-European Assessment of Land Susceptibility to Wind Erosion. *Land Degradation and Development* **27**: 1093–1105. DOI:10.1002/ldr.2318.
- Bouma J, Montanarella L. 2016. Facing policy challenges with inter- and transdisciplinary soil research focused on the UN Sustainable Development Goals. *SOIL* **2**: 135–145. DOI:10.5194/soil-2-135-2016.
- Brevik EC, Calzolari C, Miller BA, Pereira P, Kabala C, Baumgarten A, Jordán A. 2016. Soil mapping, classification, and modeling: history and future directions. *Geoderma* **264**: 256–274. DOI:10.1016/j.geoderma.2015.05.017.
- Brevik EC, Cerdà A, Mataix-Solera J, Pereg L, Quinton JN, Six J, Van Oost K. 2015. The interdisciplinary nature of soil. *SOIL* **1**: 117–129. DOI:10.5194/soil-1-117-2015.
- Bridges EM, Batjes NH, Nachtergaele FO (eds). 1998. World Reference Base for Soil Resources: Atlas. ISSS Working Group RB, ISRIC-FAO-ISSS-Acco. Acco Press: Leuven 79 pp.
- Challinor A, Martre P, Asseng S, Thornton P, Ewert F. 2014. Making the most of climate impacts ensembles. *Nature Climate Change* **4**: 77–80. DOI:10.1038/nclimate2117.
- Danielson JJ, Gesch DB. 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). U.S. Geological Survey Open-File Report 2011–1073. 26 pp.
- Dunne KA, Willmott CJ. 1996. Global distribution of plant-extractable water capacity of soil. *International Journal of Climatology* **16**: 841–859. DOI:10.1002/(SICI)1097-0088(199608)16:8 <841::AID-JOC60>3.0.CO;2-8.
- FAO, IIASA, ISRIC, ISSCAS, JRC. 2012. Harmonized World Soil Database (version 1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria. 42 pp.
- FAO, Unesco, ISRIC. 1990. Revised legend of the soil map of the world. World Soil Resources Report, FAO, Rome, Italy.
- FAO, Unesco. 1971–1981. The FAO-Unesco Soil Map of the World. Legend and 9 volumes. UNESCO: Paris.
- FAO, Unesco. 1974. Legend of the soil map of the world. FAO: Rome, Italy 59 pp.
- FAO. 1995. Digital Soil Map of the World and derived properties (ver. 3.5). FAO Land and Water Digital Media Series # 1. FAO: Rome.
- Fava F, Pulighe G, Monteiro AT. 2016. Mapping Changes in Land Cover Composition and Pattern for Comparing Mediterranean Rangeland Restoration Alternatives. *Land Degradation and Development* **27**: 671–681. DOI:10.1002/ldr.2456.
- Goyens C, Verdoodt A, Van de Waauw J, Baert G, van Engelen VWP, Dijkshoorn JA, Van Ranst E. 2007. Base de données numériques sur les sols et le terrain (SOTER) de l'Afrique Centrale (RD Congo, Rwanda et Burundi). *Etude et Gestion des Sols* **14**: 207–218.
- GSP. 2014. 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 26 pp.
- Guo LB, Gifford RM. 2002. Soil carbon stocks and land use change: a meta analysis. *Global Change Biology* **8**: 345–360. DOI:10.1046/j.1354-1013.2002.00486.x.
- Hartemink AE, Kraslininkov P, Bockheim JG. 2013. Soil maps of the world. *Geoderma* **207–208**: 256–267. DOI:10.1016/j.geoderma.2013.05.003.
- Hengl T, de Jesus JM, MacMillan RA, Batjes NH, Heuvelink GBM, Ribeiro E, Samuel-Rosa A, Kempen B, Leenaars JGB, Walsh MG, Ruiperez GA. 2014. SoilGrids1km - Global Soil Information Based on Automated Mapping. *PLoS ONE* **9**: e105992. DOI:10.1371/journal.pone.0105992.
- Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965–1978. DOI:10.1002/joc.1276.
- Holleran M, Levi M, Rasmussen C. 2015. Quantifying soil and critical zone variability in a forested catchment through digital soil mapping. *SOIL* **1**: 47–64. DOI:10.5194/soil-1-47-2015.
- Igue AM, Gaiser T, Stahr K. 2004. A soil and terrain digital database (SOTER) for improved land use planning in Central Benin. *European Journal of Agronomy* **21**: 41–52. DOI:10.1016/S1161-0301(03)00062-5.
- Imerzekene S, van Ranst E. 2001. Une banque de données pédologiques et son SIG pour une nouvelle politique agricole au Rwanda Bull. *Séanc. Acad. R. Sci. Outre-Mer* **47**: 299–329.
- IUSS Working Group WRB. 2015. World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106. FAO, Rome. 192 pp.
- Jarvis A, Reuter HI, Nelson A, Guevara E. 2008. Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). Cali: Colombia.
- Jenny H. 1941. Factors of soil formation. A system of quantitative pedology. McGraw-Hill, New York 281pp.
- Keesstra SD, Bouma J, Wallinga J, Tittoneil P, Smith P, Cerdà A, Montanarella L, Quinton JN, Pachepsky Y, van der Putten WH,

- Bardgett RD, Moolenaar S, Mol G, Jansen B, Fresco LO. 2016. The significance of soils and soil science towards realization of the United Nations Sustainable Development Goals. *SOIL* **2**: 111–128. DOI:10.5194/soil-2-111-2016.
- Kellogg CE. 1974. Soil genesis, classification, and cartography: 1924–1974. *Geoderma* **12**: 347–362. DOI:10.1016/0016-7061(74)90026-3.
- Kempen B, Heuvelink GBM, Brus DJ, Stoorvogel JJ. 2010. Pedometric mapping of soil organic matter using a soil map with quantified uncertainty. *European Journal of Soil Science* **61**: 333–347. DOI:10.1111/j.1365-2389.2010.01232.x.
- Lal R, Negassa W, Lorenz K. 2015. Carbon sequestration in soil. *Current Opinion in Environmental Sustainability* **15**: 79–86. DOI:10.1016/j.cosust.2015.09.002.
- Leenaars JGB. 2013. Africa Soil Profiles Database, Version 1.1. A compilation of geo-referenced and standardized legacy soil profile data for Sub Saharan Africa (with dataset). ISRIC report 2013/03. Africa Soil Information Service (AFSIS) project. ISRIC – World Soil Information, Wageningen, the Netherlands. 160 pp.
- López-Granados F, Jurado-Expósito M, Peña-Barragán JM, García-Torres L. 2005. Using geostatistical and remote sensing approaches for mapping soil properties. *European Journal of Agronomy* **23**: 279–289. DOI:10.1016/j.eja.2004.12.003.
- McBratney AB, Mendonça Santos ML, Minasny B. 2003. On digital soil mapping. *Geoderma* **117**: 3–52. DOI:10.1016/S0016-7061(03)00223-4.
- Miller BA, Koszinski S, Wehrhan M, Sommer M. 2015. Comparison of spatial association approaches for landscape mapping of soil organic carbon stocks. *SOIL* **1**: 217–233. DOI:10.5194/soil-1-217-2015.
- Minasny B, McBratney AB. 2010. Methodologies for global soil mapping. In *Digital Soil Mapping. Progress in Soil Science Volume 2*: Springer; 429–436.
- Montanarella L, Pennock DJ, McKenzie N, Badraoui M, Chude V, Baptista I, Mamo T, Yemefack M, Aulakh MS, Yagi K, Hong SY, Vijarnsorn P, Zhang GL, Arrouays D, Black H, Krasilnikov P, Sobocká J, Alegre J, Henriquez CR, Mendonça-Santos ML, Taboada M, Espinosa-Victoria D, AlShankiti A, AlaviPanah SK, Elsheikh EAEM, Hempel J, Arbestain MC, Nachtergaele F, Vargas R. 2016. World's soils are under threat. *SOIL* **2**: 79–82. DOI:10.5194/soil-2-79-2016.
- Mora-Vallejo A, Claessens L, Stoorvogel JJ, Heuvelink GBM. 2008. Small scale digital soil mapping in Southeastern Kenya. *Catena* **76**: 44–53. DOI:10.1016/j.catena.2008.09.008.
- Nachtergaele F, Van Ranst E. 2003. Qualitative and quantitative aspects of soil databases in tropical countries. In *Evolution of tropical soil science. Past and future*. Royal Academy of Overseas Sciences: Brussels; 107–26.
- Omuto C, Nachtergaele F, Vargas RR. 2013. State of the Art Report on Global and Regional Soil Information: Where are we? Where to go? FAO: Italy 81 pp.
- Padarian J, Minasny B, McBratney AB. 2014. The evolving methodology for global soil mapping. In: *GlobalSoilMap: Basis of the Global Spatial Soil Information System - Proceedings of the 1st GlobalSoilMap Conference*, pp. 215–220.
- Panagos P, Imeson A, Meusbürger K, Borrelli P, Poesen J, Alewell C. 2016. Soil conservation in Europe: wish or reality? *Land Degradation and Development* **27**: 1547–1551. DOI:10.1002/ldr.2538.
- Panagos P, Jones A, Bosco C, Senthil Kumar PS. 2011. European digital archive on soil maps (EuDASM): preserving important soil data for public free access. *International Journal of Digital Earth* **4**: 434–443. DOI:10.1080/17538947.2011.596580.
- Porter CH, Villalobos C, Holzworth D, Nelson R, White JW, Athanasiadis IN, Janssen S, Ripoche D, Cufi J, Raes D, Zhang M, Knapen R, Sahajpal R, Boote K, Jones JW. 2014. Harmonization and translation of crop modeling data to ensure interoperability. *Environmental Modelling and Software* **62**: 495–508. DOI:10.1016/j.envsoft.2014.09.004.
- Ribeiro E, Batjes NH, Leenaars JGB, Van Oostrum AJM, de Mendes JJ. 2015. Towards the standardization and harmonization of world soil data: Procedures Manual ISRIC World Soil Information Service (WoSIS version 2.0). ISRIC – World Soil Information: Wageningen, The Netherlands 110 pp.
- Romero CC, Hoogenboom G, Baigorria GA, Koo J, Gijsman AJ, Wood S. 2012. Reanalysis of a global soil database for crop and environmental modeling. *Environmental Modelling and Software* **35**: 163–170. DOI:10.1016/j.envsoft.2012.02.018.
- Sanchez PA, Ahamed S, Carre F, Hartemink AE, Hempel, Huising J, Lagacherie P, McBratney AB, McKenzie NJ, de Mendonça-Santos ML, Minasny B, Montanarella L, Okoth P, Palm CA, Sachs JD, Shepherd KD, Vågen TG, Vanlauwe B, Walsh MG, Winowiecki LA, Zhang GL. 2009. Digital Soil Map of the World. *Science* **325**: 680–681. DOI:10.1126/science.1175084.
- Saxton KE, Rawls WJ. 2006. Soil water characteristic estimates by texture and organic matter for hydrologic solutions. *Soil Science Society of America Journal* **70**: 1569–1578. DOI:10.2136/sssaj2005.0117.
- Shangguan W, Dai Y, Duan Q, Liu B, Yuan H. 2014. A global soil data set for earth system modeling. *Journal of Advances in Modeling Earth Systems* **6**: 249–263. DOI:10.1002/2013MS000293.
- Soil Survey Division Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18 315 pp. Washington DC.
- Soil Survey Staff. 2014. Keys to Soil Taxonomy, 12th edn. USDA-Natural Resources Conservation Service: Washington, DC 372 pp.
- Spain AV, Isbell RF, Probert ME. 1983. Soil organic matter. In CSIRO. Soils: An Australian Viewpoint. CSIRO Melbourne Academic Press: London; 552–563.
- Stehfest E, van Vuuren D, Kram T, Bouwman L, Alkemade R, Bakkenes M, Biemans H, Bouwman A, den Elzen M, Janse J, Lucas P, van Minnen J, Müller M, Prins A. 2014. Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model description and policy applications. PBL Netherlands Environmental Assessment Agency: The Hague 336 pp.
- Stoorvogel JJ, Kempen B, Heuvelink GBM, de Bruin S. 2009. Implementation and evaluation of existing knowledge for digital soil mapping in Senegal. *Geoderma* **149**: 161–170. DOI:10.1016/j.geoderma.2008.11.039.
- Todd-Brown KEO, Randerson JT, Post WM, Hoffman FM, Tamocai C, Schuur EAG, Allison SD. 2013. Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations. *Biogeosciences* **10**: 1717–1736.
- Viscarra Rossel RA, Webster R, Bui EN, Baldock JA. 2014. Baseline map of organic carbon in Australian soil to support national carbon accounting and monitoring under climate change. *Global Change Biology* **20**: 2953–2970. DOI:10.1111/gcb.12569.
- Webb RS, Rosenzweig CE, Levine ER. 1993. Specifying Land Surface Characteristics in General Circulation Models: Soil Profile Data Set and Derived Water Holding Capacities. *Global Biogeochemical Cycles* **7**: 97–108. DOI:10.1029/92GB01822.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site:

Table S1. Literature used for the development of the S-World methodology to disaggregate the Harmonized World Soil Database.

Table S2. Validation dataset for the S-world dataset.