Article

A Comparison, Validation, and Evaluation of the S-world Global Soil Property Database

Jetse J. Stoorvogel * and **Vera L. Mulder**

Soil Geography and Landscape Group, Wageningen University, 6708 PB Wageningen, The Netherlands; titia.mulder@wur.nl
* Correspondence: jetse.stoorvogel@wur.nl; Tel.: +31-3174-84043

Abstract: Despite the increased usage of global soil property maps, a proper review of the maps rarely takes place. This study aims to explore the options for such a review with an application for the S-World global soil property database. Global soil organic carbon (SOC) and clay content maps from S-World were studied at two spatial resolutions in three steps. First, a comparative analysis with an ensemble of seven datasets derived from five other global soil databases was done. Second, a validation of S-World was done with independent soil observations from the WoSIS soil profile database. Third, a methodological evaluation of S-world took place by looking at the variation of soil properties per soil type and short distance variability. In the comparative analysis, S-World and the ensemble of other maps show similar spatial patterns. However, the ensemble locally shows large discrepancies (e.g., in boreal regions where typically SOC contents are high and the sampling density is low). Overall, the results show that S-World is not deviating strongly from the model ensemble (91% of the area falls within a 1.5% SOC range in the topsoil). The validation with the WoSIS database showed that S-World was able to capture a large part of the variation (with, e.g., a root mean square difference of 1.7% for SOC in the topsoil and a mean difference of 1.2%). Finally, the methodological evaluation revealed that estimates of the ranges of soil properties for the different soil types can be improved by using the larger WoSIS database. It is concluded that the review through the comparison, validation, and evaluation provides a good overview of the strengths and the weaknesses of S-World. The three approaches to review the database each provide specific insights regarding the quality of the database. Specific evaluation criteria for an application will determine whether S-World is a suitable soil database for use in global environmental studies.

Keywords: soil carbon; soil texture; global soil map; ensemble analysis

1. Introduction

There is an increasing call for global assessments of a wide range of environmental impacts. A good example is the range of initiatives around the Global Sustainability Goals [1]. Concurrently, there is an increased need for global databases describing the status of our natural resources. Over the years, various databases on past, current, and future climatic conditions have been developed. These climatic databases are based on actual observations [2] and simulation models for past and future conditions [3]. Less attention has been paid to global soil resources although they are increasingly being placed in the spotlight through, for example, the sustainable development goals and the obvious role of soils for, e.g., food production and carbon storage [4]. Various global soil databases have been developed over the past decades. Existing soil surveys were compiled and published as the FAO Digital Soil Map of the World (DSMW [5]) at a scale of 1:5 million. Later, parts of the DSMW were replaced by more detailed information (up to a certain level of accuracy, as they were aiming at a resolution of 30 arc-seconds) if more information was available. These updates resulted in the Harmonized World Soil Database (HWSD [6]). The DSMW and the HWSD were both based on soil surveys from around the world. The DSMW

aimed to provide consistent global soil information. The HWSD improved the DSMW where possible. As a result, the level of detail of the HWSD differs between countries. It is likely that the quality of the HWSD exceeds that of the DSMW, but data in the latter are more consistent. As soon as modelers started to use the data in all kind of modelling efforts, they required insight in soil properties rather than the soil types presented in the original surveys. Soil properties were derived either from the representative soil profiles in the original surveys, or through linking databases of soil profiles [7] to the survey. To deal with complex map units that are described by multiple soil types in, e.g., the HWSD, either the dominant soil type is used or a weighted average from the soil associations within the map unit is calculated, based on the relative importance of the different soil types. Recently, an automated soil property mapping approach has been developed, denominated SoilGrids [8,9]. This approach mainly makes use of a large database of soil profiles with chemical and physical descriptions (WoSIS, [10]) and an extensive array of covariates to model the soil properties on a particular location based on local conditions using digital soil mapping.

To evaluate global land degradation and support the analysis of scenarios for the UNCCD Global Land Outlook, the Netherlands Environmental Assessment Agency required soil data for both current and natural soil conditions [11,12]. As these data were not yet available, a new approach was developed in which the HWSD was disaggregated using auxiliary data and the WISE3.1 soil profile database [7]. The resulting S-World global soil property database [13,14] provided the required information for current conditions as well as for the natural, pristine situation.

The number of different global soil property databases is gradually increasing. However, derived products like the estimated extent of soil degradation and global carbon stocks differ considerably [12,15,16]. This obviously calls for a better insight in the quality of the databases and insights in these differences. Studies provide insight in their own specific way. The DSMW and the HWSD provide general estimates on the purity of the map units. SoilGrids provides insight in terms of: i) the variance explained by the models and ii) the average overall error of estimated prediction error. S-World carried out a validation on a limited dataset of 1690 independent soil observations, but only does so for the soil organic matter contents in the topsoil. The different approaches limit the comparison of the different global soil maps [17]. What is hampering the more in-depth review of the global soil property databases?

- Independent datasets with a proper global coverage are absent.
- The support of the global soil property maps (available at different resolutions like 30 arc-seconds or 0.5 degree) differs from the point observations in the soil profile databases.
- Single quality measures like the root mean square difference (RMSD) are insufficient as it is likely that the quality of the maps differs geographically due to the natural inherent variation in soil properties, the scale of underlying soil maps, and sampling intensity.

This study attempts to explore options to provide better insight into the quality of soil property maps. This will be done with a focus on S-World. Since insight in the methodology is required for the proper interpretation of the results, we will first briefly describe the procedures applied in S-World and how it was established. Subsequently, we describe the methodology that involved three different steps: a comparison to other global soil property maps, a validation using the point data from the WoSIS dataset, and a model evaluation of some of the key assumptions underlying the methodology. Finally, the results of the individual three steps are presented and the overall evaluation is discussed.

2. Materials and Methods

2.1. S-World and Its Use in the Global Land Outlook

S-World was developed specifically for the Netherlands Environmental Assessment Agency to explore future changes in land use and land condition and the impacts on food, water, climate change and biodiversity [12,18]. The methodology estimated the current

soil conditions through a disaggregation of the HWSD. The disaggregation took place in two steps. In a first step, the complex map units of the HWSD were disaggregated into simple map units using a global digital elevation model and logical sequences of soil types according to altitude (using the concept of toposequences). This step resulted in a new map where each map unit is described by a single soil type. In a second step, the variation of soil properties within the different soil types was derived from the WISE3.1 soil profile database and described by the range defined by the 1st and 9th decile of each soil property and each soil type. Based on a meta-analysis of a broad literature survey a simple model was developed that uses landscape properties (climate, topography, vegetative cover, and land use) to estimate soil properties for each location on the earth surface on the basis of its soil type, the variation in soil properties within these soil types, and the landscape properties. The procedure underlying this disaggregation is described in detail in Stoorvogel et al. [13]. The current soil conditions are to a certain (but unknown) extent the result of human land use. In a subsequent study, the natural vegetative cover was modelled, allowing to derive the natural soil conditions using the S-World methodology. The procedure for the derivation of the natural conditions is described in detail in Stoorvogel et al. [14].

For the scenarios for the UNCCD Global Land Outlook [12], a number of soil properties from the S-World database were used. Core variables that will further be evaluated in this study are the soil organic carbon (SOC) content in topsoil (0–30 cm) and subsoil (30–100 cm), and the average clay content in the soil profile (0–100 cm, or over the entire soil profile if the depth did not reach 100 cm). These variables will also be used in this study at two different resolutions: at the original resolution of the S-World methodology (30 arc-seconds) and at a more generalized resolution (5 arc-minutes) that was used for different modelling studies [12,19,20].

2.2. A Comparative Analysis

2.2.1. Global Soil Databases

A range of different global datasets is available for environmental modelling and thus also for the comparison with the S-World database. An inventory of datasets that were at a resolution of at least 0.5 degree and that contained data on the most relevant soil properties: organic carbon content and soil texture was carried out. The following datasets were identified:

- GSDE: Global Soil Dataset for use in Earth System Models [21]. The GSDE provides a range of soil property maps at a resolution of 30 arc-second for eight layers up to a depth of 2.3 m (i.e., 0–4.5, 4.5–9.1, 9.1–16.6, 16.6–28.9, 28.9–49.3, 49.3–82.9, 82.9–138.3 and 138.3–229.6 cm). The GSDE is based on the DSMW and various regional and national soil databases. The GSDE first harmonized the databases and subsequently linked the results to the map units of the DSMW.
- HWSD: Harmonized World Soil Database [6]. The HWSD is principally a soil map with soil types at a 30 arc-seconds resolution. The 16,022 different soil map units combine data from the DSMW with regional and national updates of soil information worldwide. The majority of the soil map units are described by multiple, so-called, soil components. In addition, soil property data are provided for the topsoil (0–30 cm) and the subsoil (30–100 cm) for each of the soil components in the map units.
- IGBP-DIS: Global Gridded Surfaces of Selected Soil Characteristics [22]. The data includes 7 soil properties at a resolution of 5 arc-minutes. The data were developed by the Global Soil Data Task Group of the International Geosphere-Biosphere Programme (IGBP) Data and Information System (DIS). The, so-called, Soil Data System uses a statistical bootstrapping approach to link the pedon records in the Global Pedon Database [23] to the DSMW.
- SoilGrids. The global 3D soil information system [8] provides a wide range of soil properties at 7.5 arc-seconds resolution at various depths: 0, 5, 15, 30, 60, 100, and 200 cm. SoilGrids uses the Global Soil Information Facilities, a framework and platform to support widespread, open collaboration in the assembly, collation and production

of global soil information. The prediction models were calibrated using approximately 150,000 soil profiles. The majority of the profiles (excluding data sets with specific restricting terms of use) are also published as WoSIS Web Feature Service [10]. Even though the input training point data are extensive and cover most continents and climatic zones, some large areas that have extreme climatic conditions and/or have very restricted access, are under sampled. Therefore, expert-based pseudo-observations were inserted and gaps in the feature space were filled in the following areas: (1) semi-arid, arid lands, deserts and sand dunes, (2) mountain tops, steep slopes of mountains and similar inaccessible areas, (3) areas covered by ice and/or snow, and (4) inaccessible tropical forest.

- WISE30sec: Harmonized soil property values for broad-scale modelling [24]. The dataset considers 20 soil properties. These estimates are presented for fixed depth intervals of 20 cm up to a depth of 100 cm, and intervals of 50 cm between 100 cm to 200 cm (or less when appropriate). A harmonized dataset of derived soil properties for the world that is comprised of a soil geographical and a soil attribute component. The dataset was created using the HWSD, 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.

It should be noted that the above datasets are not independent. Most datasets used the HWSD as a starting point which was made using the DSMW. All databases, except SoilGrids, used one of these databases as their starting point.

2.2.2. Data Preprocessing

An ensemble of different soil property maps was created in three different steps including data acquisition, conversion, and aggregation and/or resampling:

1. Data acquisition: GSDE data on sand, clay and SOC content were downloaded from the Land-Atmoshere Interaction Research Group at Sun Yat-sen University at 30 arc-seconds. The vertical variation in soil properties was captured by eight layers of differing depths. HWSD data were downloaded from IIASA at 30 arc-seconds resolution. The map units of the HWSD are described by 1–10 different soil types. The relative importance of the soil types and the SOC contents (for the topsoil (0–30 cm) and subsoil (30–100 cm)) and clay contents were included in the database. IGBP-DIS data were downloaded from the Distributed Active Archive Centre for Biochemical Dynamics and included soil-carbon density and bulk density data at 5 arc-minutes resolution. Data from SoilGrids were directly obtained from the developer at ISRIC and included grids of clay and SOC contents at 250 m resolution at seven different depth intervals. WISE30sec data were downloaded from the ISRIC website, including estimates of sand, silt and SOC content for fixed depth intervals of 20 cm up to a depth of 100 cm. Data were obtained in October 2017 except for the GSDE database that was downloaded on 8 January 2018.

2. Conversion: All the obtained datasets had to be converted into the proper matching variables (i.e., depth interval and measurement units). The GSDE and SoilGrids included complete data at various depth intervals. Weighted averages were calculated for the 0–30 cm topsoil and 30–100 cm subsoil for SOC, and over the 0–100 cm profile for the clay content. The HWSD already provided SOC contents for topsoil and subsoil. For the clay content a weighted average was calculated over the soil profile. If, due to a limited soil depth, no data were provided for the subsoil, clay contents of the topsoil were used. Typically, two methods are used to deal with the complex map units containing soil associations (84% of all map units): either the dominant soil type is used (which on average covers 65% of the map unit) (HWSDd) or the area weighted average of the soil properties is calculated based on the share of each soil type (HWSDw). Both methods were used to derive the required grids of soil properties. The IGBP-DIS database included soil carbon density and bulk density but lacked data on soil texture. We derived data on SOC contents for topsoil and

subsoil from the soil carbon density and bulk density. The derivation required two main assumptions: (i) differences in soil depth are not considered and (ii) carbon stocks in the 0–30 cm topsoil roughly equal the carbon stocks in the 30–100 cm subsoil (following Hiederer and Köchy [25]). Similar to the HWSD, the WISE30 database included information for the topsoil and subsoil of the different soil components. The WISE30 was used in a similar way as the HWSD: soil properties of the dominant soil component (WISE30d) and the weighted average (WISE30w) were derived.

3. Aggregation/resampling: Finally, maps at 5 arc-minutes resolution were resampled to additionally obtain maps at 30 arc-seconds resolution. Maps at 30 arc-seconds resolution or more detailed were aggregated to a 5 arc-minutes resolution. During aggregation, the mean of all grid cells was calculated while ignoring possible cells with no data values.

2.2.3. Consistency Analysis

The inventory of global soil property databases and the preprocessing provided an ensemble of soil property maps with information on soil organic carbon (in the topsoil and subsoil) and soil clay content at a 30 arc-seconds and 5 arc-minutes resolution. The variation within the ensemble was first analyzed by deriving the mean, standard deviation, and the coefficient of variation, for each of the three soil properties, and at the two spatial resolutions. To better understand the variation in the ensemble and the comparison with S-World, the analysis was stratified by the 14 terrestrial ecoregions as defined by [26]. The terrestrial ecoregions are presented in Figure 1. The consistency between the datasets is evaluated at 5 arc-minutes resolution. Ensemble mean and standard deviation for the four soil properties were derived for each individual grid cell. Subsequently, the spatially aggregated mean and standard deviation of the ensemble mean and standard deviation for each of the terrestrial ecoregions are derived, each with its own specific interpretation:

- The spatially aggregated mean of the ensemble mean ($\mu_{E\mu}$): average soil property within an ecoregion.
- The spatially aggregated standard deviation of the ensemble mean ($\sigma_{E\mu}$): the spatial variation of a soil property within an ecoregion.
- The spatially aggregated mean of the ensemble standard deviation ($\mu_{E\sigma}$): the average consistency of the estimates of a soil property within an ecoregion.
- The spatially aggregated standard deviation of the ensemble standard deviation ($\sigma_{E\sigma}$): The variation in the consistency of the estimates within an ecoregion.
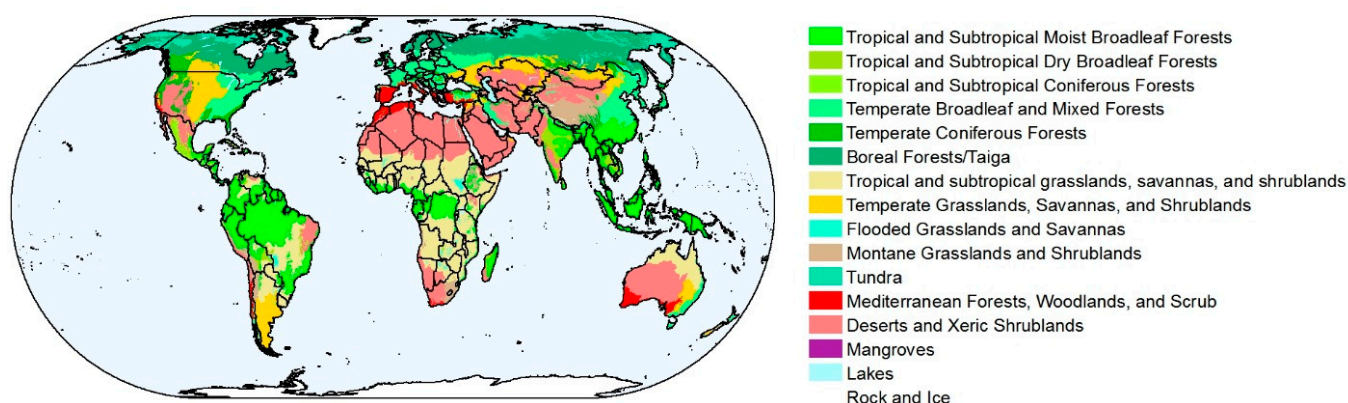

Figure 1. Terrestrial ecosystems as defined by Olson et al. [26].

The results allow us to evaluate where the various datasets are in agreement with each other. Differences are likely to occur in areas with considerable short distance soil variation, relatively high values, or with few underlying soil studies. While evaluating a soil property database, it is good to know whether the map differs significantly from the other maps in the ensemble. S-World estimates were compared to the ensemble by

calculating the difference between the ensemble mean and the S-World data. Although there are advanced statistical ways to do so, they require an ensemble of independent estimates [27]. This is certainly not the case in the ensemble of global soil property maps. Therefore, simply the difference between the S-World estimates and the mean values of the ensemble were derived. In addition, the S-World estimates are interpreted together with the map of the standard deviation in the ensemble. In this work, S-World estimates were considered representative when having an (admittedly arbitrary) deviation from the ensemble maps of less than 1.5% for SOC and 5% for clay contents. Generic thresholds are difficult to assess as the thresholds depend on the application.

2.3. Validation

WoSIS [10] is one of the most complete soil profile databases that is publicly available and with a reasonable global coverage. WoSIS aims to provide quality-assessed, georeferenced soil data to the international community upon their standardization and harmonization. So far, the focus has been on developing procedures for legacy point data with special attention to the selection of soil analytical and physical properties considered in the GlobalSoilMap specifications [28]. Profile data in WoSIS were contributed by a wide range of providers. The data have been described, sampled, and analyzed according to methods and standards in use in the originating countries. Special attention was paid to measures for data quality and the standardization of soil property definitions, and soil analytical methods. The WoSIS database contained some 118,400 soil profiles, of which some 96,000 are georeferenced. The number of measured data for each property varies between profiles and with depth, generally depending on the purpose of the initial studies. For the S-World validation only part of the database was useful:

- Soil profiles had to be independent from the S-World calculations. Therefore, all soil profiles that were also included in the WISE3.1 database were removed from the database.
- Only soil profiles that were georeferenced with a 30 arc-seconds accuracy (or higher) were included in the validation.
- Only soil profiles that included either or both SOC and soil texture data were useful.

The screening of the database resulted in 83,393 soil profiles that could be used in a validation (Figure 2). For a statistical analysis, it is important to consider that these observations are not a random sample. Certain regions of the world (like the Russian Federation) lack observations, while others (like Mexico) have a relatively dense observation density. The actual validation took place on a smaller number of observations. Only 53% of the observations had information on the three soil properties. 63% of the observations included topsoil SOC, 58% included subsoil SOC, and 83% included clay content. All points were overlain with the S-World database and the mean difference (MD) and the root mean square difference (RMSD) were derived. Although these data were independent from the WISE database, it was not possible to check whether they were independent of the HWSD. Some of the observations may have actually been used for the soil surveys that provide the basis for the HWSD. This indirect dependency is ignored in the analysis as it is impossible to verify. Another limitation of the validation dataset is that it is far from a random sample. To evaluate the distribution of the validation points, the average distance from each grid cell to the nearest validation point was compared to the average distance from each WISE observation to the nearest validation point. If the distance of the WISE datapoints is smaller, the validation points are located in areas where there is also a relatively high quality of the S-World database resulting in relatively positive validation results. If the distance of the WISE datapoints is larger, apparently observations in new areas were added and the validation took place in areas with a relatively lower quality, resulting in relatively negative validation results.
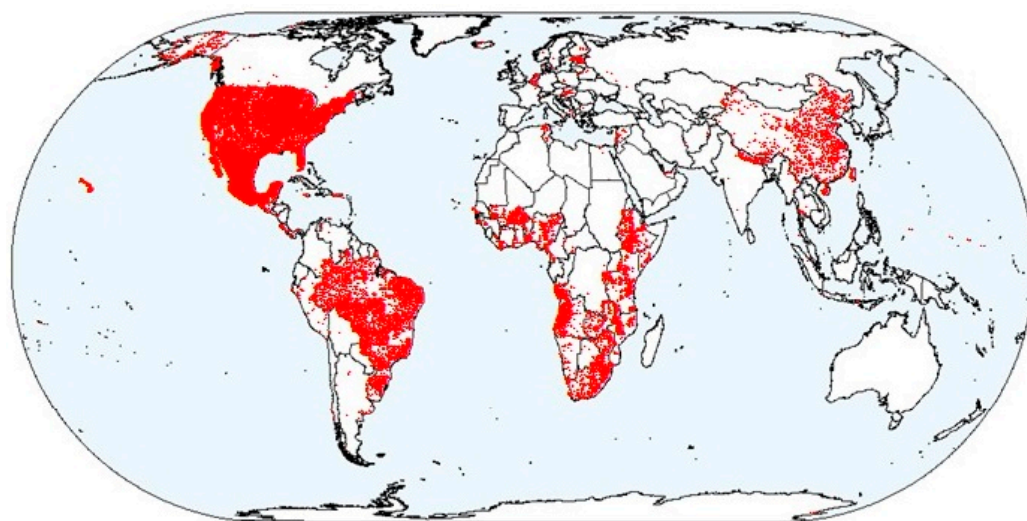
Figure 2. Selected soil profiles (*n* = 83,393) from the WoSIS database for the validation of the S-world soil property database.

2.4. Methodological Evaluation of S-World

The S-World methodology includes a large number of different steps. If it is relatively difficult to carry out a proper validation, an alternative approach would be to look at specific methodological elements of the S-World methodology. Two specific elements of the methodology were evaluated, related to the ranges of the soil properties and the support of the observations.

First, in the original S-World evaluation, ranges of soil properties within each of the soil types were derived on the basis of the WISE3.1 database [7]. The database 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 6399 soil profiles was used to estimate these ranges. Since then, the profile database at ISRIC has been expanded resulting in the much larger WoSIS database with 118,400 soil profiles. This poses the question whether the results of S-World would have been different if the larger WoSIS database would have been available. The resulting ranges in the soil properties *v* derived from WISE3.1 (*wi*) and WoSIS (*wo*) were compared in terms of the area weighted average of the overlap (*O* (in %)) as:

$$O_v = \frac{100}{A_{tot}} \sum_{s=1}^{s=181} \left(\frac{Min(D9_{wo,v,s}, D9_{wi,v,s}) - Max(D1_{wo,v,s}, D1_{wi,v,s})}{Max(D9_{wo,v,s}, D9_{wi,v,s}) - Min(D1_{wo,v,s}, D1_{wi,v,s})} * A_s \right)$$

where *D*1 is the 1st decile and *D*9 is the 9th decile of the range, and A_s is the area of soil type *s*.

Second, the support of the grid cells of 30 arc-seconds makes it likely that the variation in soil properties is less, compared to point observations. S-World therefore takes the 1st and 9th decile of the actual variation in soil properties per soil type. The choice of the 1st and 9th decile is a rather arbitrary one. Within the WoSIS database many observations are from local experiments or soil surveys at short distances. It allowed to estimate for the within-grid cell variation and to assess whether indeed the choice for the use of the deciles is legitimate. S-world assumes that the variation for the grid cells is 80% of the variation in individual observations and therefore truncates the ranges at the 1st and 9th decile. Soil profiles in the WoSIS database are often clustered as they are derived from specific detailed research projects or detailed soil surveys. In 1429 grid cells of the S-World database multiple soil profiles are located. These soil profiles were used to obtain insights regarding the variation within grid cells. To evaluate the assumptions on the discrepancy between the support of the grid cells and the soil profile observations, the variation within

the grid cells was studied in terms of the variation explained in an analysis of variation ($100\%*SSQ_{within}/SSQ_{between}$).

3. Results and Discussion

3.1. Comparative Analysis

The ensembles of soil property maps for the different properties are presented in Figures 3–5. Although, the patterns of the various maps do correspond, there are also some clear differences. Roughly above 50° northern latitude, S-world, SoilGrids and WISE30d clearly present higher SOC contents in the topsoil in comparison to the other 4 databases. Looking at the subsoil, SoilGrids is still presenting higher SOC contents in this region whereas the other databases appear to be more consistent. In the case of the clay contents, the patterns do vary locally (e.g., the arid areas around the Sahara Desert). The results on the consistency of the ensemble are summarized in Table 1 for each of the terrestrial ecoregions.

Table 1. Spatially aggregated mean and standard deviation of 4 soil properties as represented by an ensemble of global soil property maps stratified by terrestrial ecoregions.

	Topsoil SOC%				Subsoil SOC%				Clay%			
	$\mu_{E\mu}$	$\sigma_{E\mu}$	$\mu_{E\sigma}$	$\sigma_{E\sigma}$	$\mu_{E\mu}$	$\sigma_{E\mu}$	$\mu_{E\sigma}$	$\sigma_{E\sigma}$	$\mu_{E\mu}$	$\sigma_{E\mu}$	$\mu_{E\sigma}$	$\sigma_{E\sigma}$
(Sub)tropical moist broadleaf forests	2.2	2.6	1.2	1.7	1.1	2.3	0.6	1.5	35.8	8.2	7.4	3.2
(Sub)tropical dry broadleaf forests	1.5	1.0	0.7	1.0	0.5	0.3	0.2	0.4	35.5	9.5	6.9	2.7
(Sub)tropical coniferous forests	2.2	1.2	1.3	1.2	0.8	0.4	0.4	0.4	27.3	7.2	7.5	2.2
Temperate broadleaf/mixed forests	3.0	3.8	1.8	2.6	1.5	3.1	1.1	2.2	23.4	6.9	4.8	2.4
Temperate coniferous forests	3.2	3.3	2.1	2.5	1.4	2.8	1.1	2.2	20.2	6.5	4.7	2.3
Boreal forests/taiga	8.2	7.9	5.1	3.9	5.4	7.4	4.0	4.0	16.6	5.7	5.5	2.7
(Sub)tropical grasslands/savannas/shrublands	1.1	1.0	0.5	0.8	0.4	0.7	0.2	0.5	28.5	11.7	6.9	3.2
Temperate grasslands/savannas/shrublands	1.7	1.5	0.8	1.5	0.7	1.1	0.4	1.1	26.2	6.1	4.9	2.5
Flooded grasslands/savannas	2.8	4.8	2.0	3.6	1.7	3.9	1.3	2.8	30.2	10.0	6.6	3.1
Montane grasslands/shrublands	1.8	1.2	1.1	1.4	0.7	0.9	0.5	1.1	20.1	7.0	4.7	2.0
Tundra	4.5	3.7	3.6	3.2	2.4	2.9	2.3	2.3	16.0	3.6	6.2	3.1
Mediterranean forests/woodlands/scrub	1.2	0.7	0.5	0.5	0.4	0.4	0.2	0.3	25.3	6.2	4.9	1.9
Deserts and Xeric shrublands	0.7	0.4	0.4	0.6	0.3	0.2	0.1	0.3	20.5	7.2	5.8	2.4
Mangroves	4.7	6.0	2.7	3.2	3.4	5.3	2.2	3.1	36.6	8.4	8.8	3.2

The spatially aggregated mean of the ensemble mean presents large differences in SOC contents between the different ecoregions whereas the differences in clay contents are much smaller. This is probably caused by the direct relationship between SOC and the ecological conditions used to define the terrestrial ecoregions (climate and vegetation), whereas soil textural differences are stronger related to other soil forming factors like parent material and the age of the deposits. The boreal regions and to a lesser extent the tundras jump out, having very high SOC contents in the topsoil but compared to other ecosystems even more in the subsoil.

The spatially aggregated standard deviation of the ensemble mean shows that there is considerable spatial variation in soil conditions within the ecoregions especially in the SOC contents where $\sigma_{E\mu}$ is often the same order of magnitude as $\mu_{E\mu}$. In the case of the clay content, $\sigma_{E\mu}$ is generally much smaller than the $\mu_{E\mu}$. This can be interpreted such that the ecoregions are a relatively poor predictor of the various soil properties.

The spatially aggregated mean of the ensemble standard deviation shows how much variation there is between the different maps in the ensemble. The SOC contents vary considerable within the ensemble with the $\mu_{E\sigma}$ often around 50% of the $\mu_{E\mu}$. The ensemble is much more consistent in clay contents for almost all ecoregions with coefficients of variance around 25% compared to 50% for SOC. The ecosystems with limited consistency within the ensemble also present considerable spatial variation and it appears that there is considerable variation in the consistency of the ensemble. The tundras and taigas that presented relatively large differences between the ensemble logically also present a relatively large $\mu_{E\sigma}$ for SOC of the topsoil.
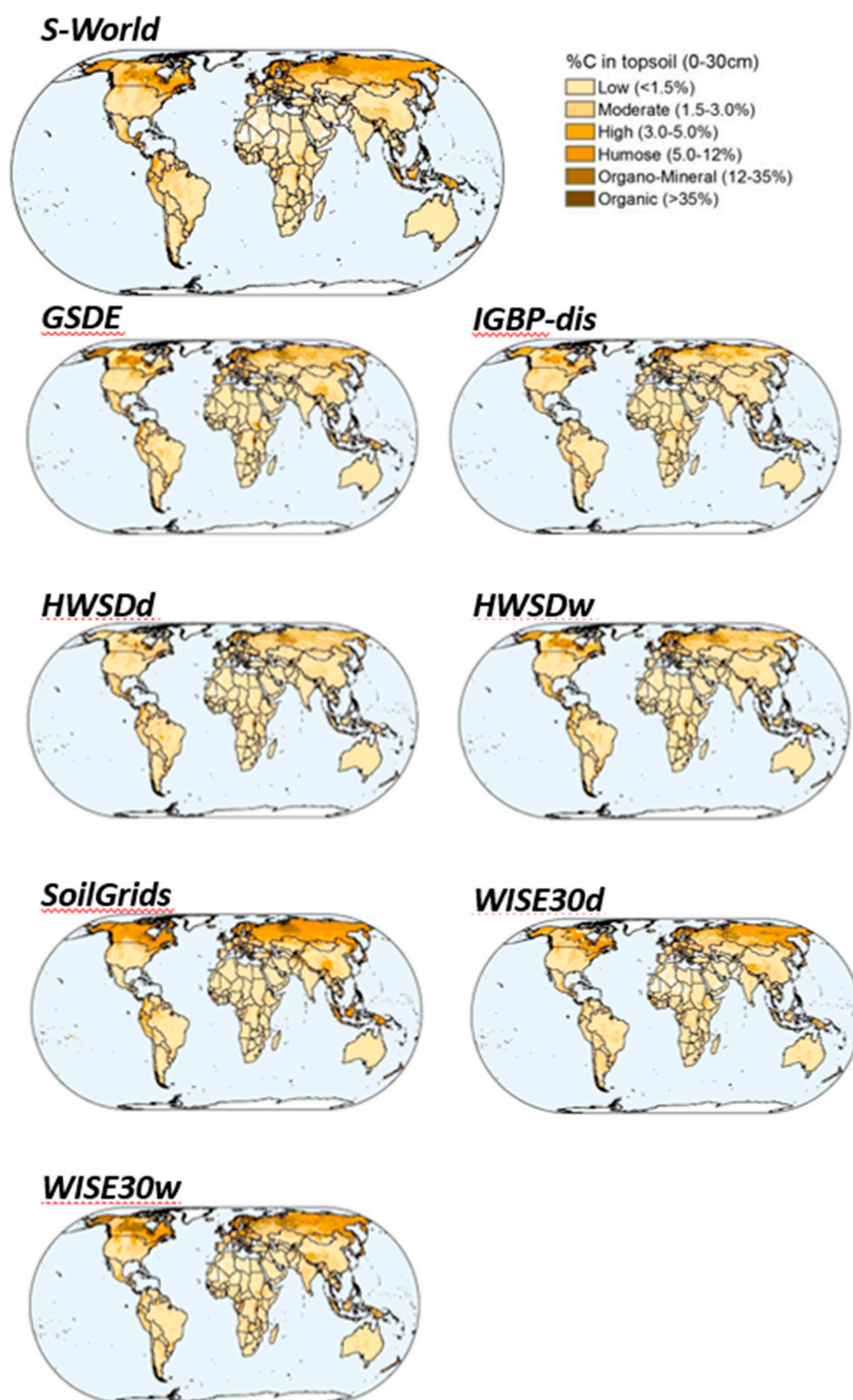
Figure 3. Global maps of soil organic carbon contents in the topsoil from different sources.
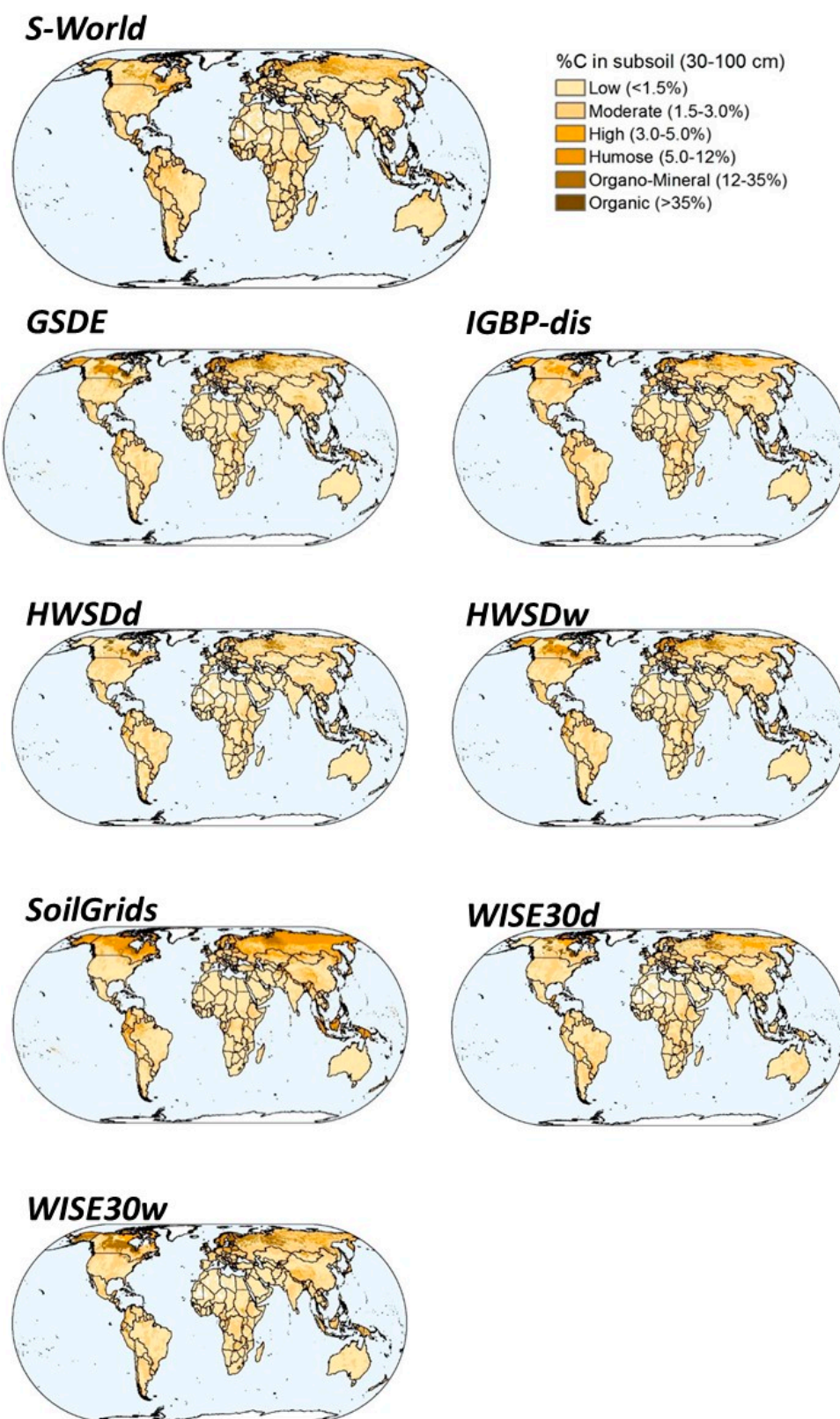
Figure 4. Global maps of soil organic carbon contents in the subsoil from different sources.
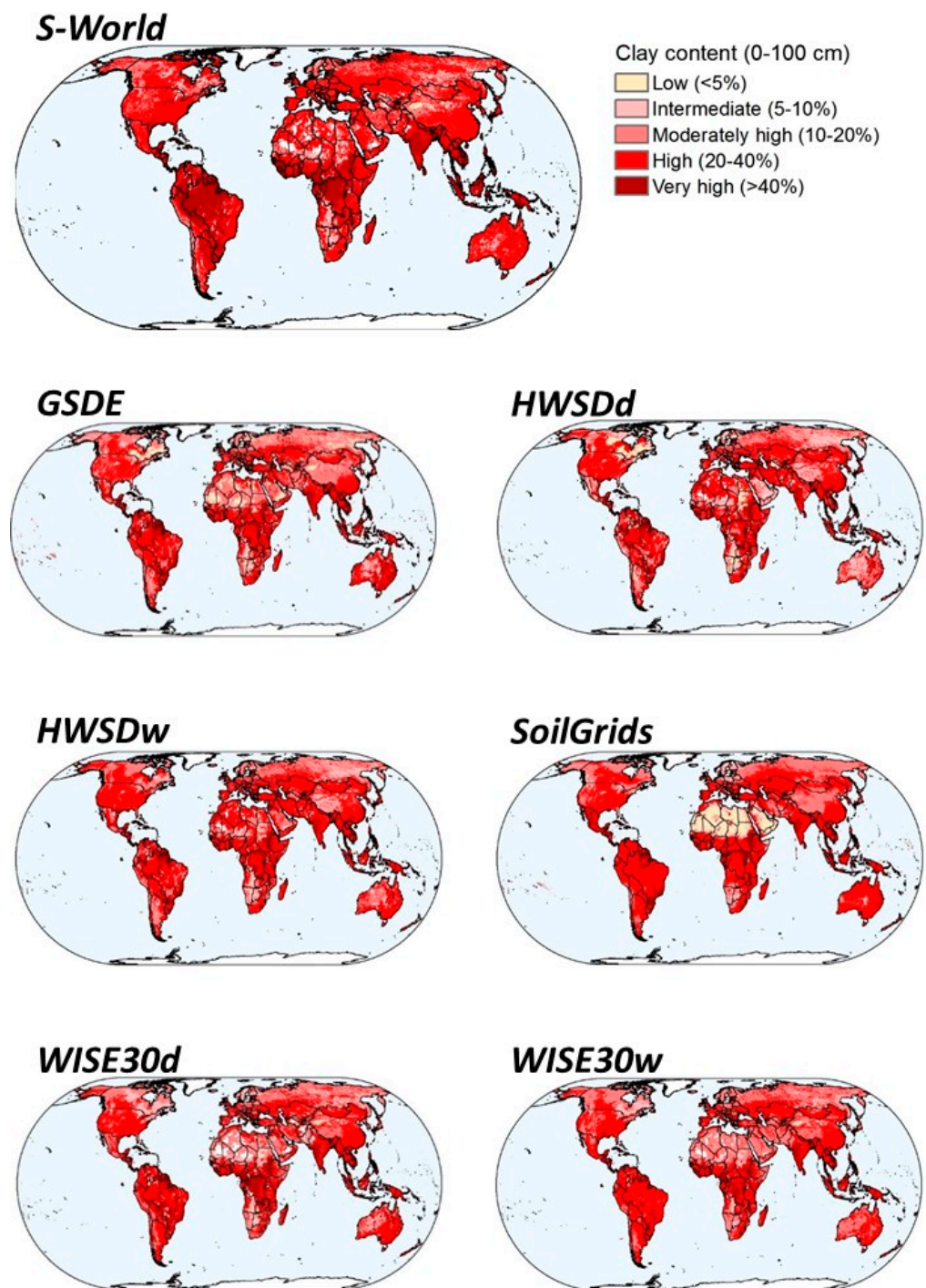
Figure 5. Global maps of clay contents in the soil profile from different sources.

Throughout the various maps $\mu_{E\sigma}$ exceeds $\sigma_{E\mu}$ which means that the variation between the maps in the ensemble is smaller than the spatial variation within the ecoregions.

The ensemble is compared to the S-World database in Figure 6 presenting the difference between the ensemble mean and the S-world estimate. The largest differences in SOC contents between ensemble and S-World are found in terrestrial ecoregions with colder and wetter conditions, including the boreal forests, taiga, and tundra. Those are also the areas where the largest discrepancies between the various maps (as indicated by $\mu_{E\sigma}$) are found but also with the highest spatial variation (as indicated by $\sigma_{E\mu}$). In the topsoil, also differences are found in the topical moist broadleaf forests of the Amazon. Differences

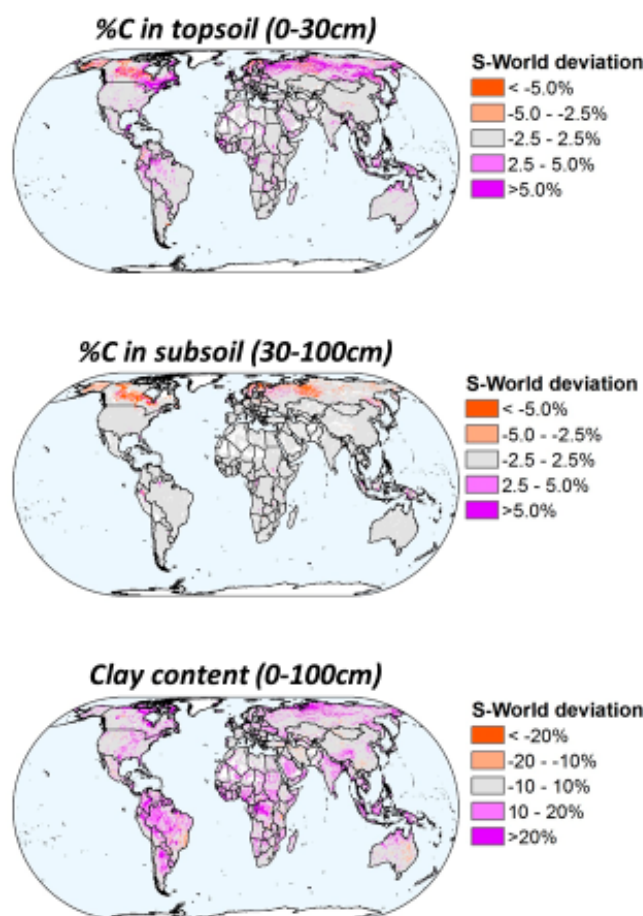between clay contents by the ensemble and S-world are much more spread out over all terrestrial ecoregions.

Figure 6. The comparative analysis of S-world against an ensemble of different soil property maps.

Figure 6 gave a comparison between the ensemble mean and S-world. An alternative approach that was followed is to derive the area where the S-World estimates differ considerably from the ensemble mean in each of the ecoregions. For this, the admittedly arbitrary thresholds of $\pm 1.5\%$ for SOC and $\pm 10\%$ for clay content were used. These thresholds depend on the specific application of the dataset. The results are presented in Table 2 and indicate that the majority of the area falls within these limits: 67.3% for topsoil SOC, 87.2% for subsoil SOC, and 86.1% for the clay contents. However, again, large differences occur between the different terrestrial ecoregions. Large differences occur: for the topsoil SOC only 32.4% of the area falls within the threshold in the boreal forests and taiga, whereas in the deserts there is correspondence in 91.9%. This clearly is the result of the use of a fixed threshold where variation tends to increase with the magnitude of SOC contents. The SOC contents in the boreal forests and taiga are simply much larger than the SOC contents in the deserts. This is also clearly presented if one compares the values of the $\mu E \sigma$ with the area that corresponds. A large variation in the ensemble almost automatically leads to a small area that falls within the threshold and vice versa.

3.2. Validation Results

The validation results are presented in Table 3. It immediately becomes apparent that some of the terrestrial ecoregions have few soil profiles. These are typically the regions that also had few observations in the WISE3.1 database, but also the regions where the level of detail in the HWSD was relatively low. As a result, it is not surprising that these regions are the ones with the largest RMSD, with the exception of temperate coniferous forests.

The latter can be partly explained by the fact that there were few samples available for validation from the WoSIS database for this ecoregion (Figure 2). These are also the ones which had the largest variability within the ensemble maps (Table 1).

Table 2. The variation in the ensemble of soil property maps in terms of the spatially aggregated mean of the ensemble standard deviation ($\mu_{E\sigma}$) and % area in correspondence (A_{corr}).

	Topsoil SOC%		Subsoil SOC%		Clay%	
	$\mu_{E\sigma}$	A_{corr}	$\mu_{E\sigma}$	A_{corr}	$\mu_{E\sigma}$	A_{corr}
(Sub)tropical moist broadleaf forests	1.2	55.3	0.6	94.6	7.4	76.1
(Sub)tropical dry broadleaf forests	0.7	81.8	0.2	99.2	6.9	88.8
(Sub)tropical coniferous forests	1.3	66.7	0.4	98.9	7.5	96.2
Temperate broadleaf/mixed forests	1.8	70.5	1.1	89.3	4.8	94.1
Temperate coniferous forests	2.1	65.4	1.1	89.7	4.7	95.0
Boreal forests/taiga	5.1	32.4	4.0	57.6	5.5	85.4
(Sub)tropical grasslands/savannas/shrublands	0.5	83.3	0.2	99.1	6.9	85.4
Temperate grasslands/savannas/shrublands	0.8	92.8	0.4	97.2	4.9	93.9
Flooded grasslands/savannas	2.0	72.8	1.3	87.9	6.6	80.9
Montane grasslands/shrublands	1.1	87.5	0.5	97.4	4.7	94.0
Tundra	3.6	49.7	2.3	77.7	6.2	64.8
Mediterranean forests/woodlands/scrub	0.5	87.8	0.2	99.7	4.9	99.1
Deserts and Xeric shrublands	0.4	91.9	0.1	99.9	5.8	95.6
Mangroves	2.7	48.3	2.2	71.7	8.8	60.4
Overall	2.0	67.3	1.3	87.2	5.9	86.1

Table 3. Validation of S-World results with WoSIS observations (excluding the WISE observations) in terms of MD and RMSD (between parentheses).

	5 Arc Minutes			30 Arc-Seconds		
	SOC_{top}	SOC_{sub}	Clay	SOC_{top}	SOC_{sub}	Clay
(Sub)tropical moist broadleaf forests (*n* = 7824)	0.8	−0.1	7.1	0.7	−0.2	7.6
	(4.0)	(2.3)	(21.4)	(4.1)	(2.1)	(23.5)
(Sub)tropical dry broadleaf forests (*n* = 3097)	1.3	0.0	7.9	1.2	0.0	8.0
	(3.3)	(1.3)	(19.7)	(3.4)	(1.5)	(20.9)
(Sub)tropical coniferous forests (*n* = 2486)	1.5	0.2	−2.1	1.5	0.2	−2.4
	(2.9)	(1.0)	(15.4)	(3.2)	(1.1)	(16.2)
Temperate broadleaf/mixed forests (*n* = 20074)	1.7	0.0	7.7	1.6	0.0	7.8
	(5.6)	(2.6)	(16.5)	(5.9)	(2.7)	(17.1)
Temperate coniferous forests (*n* = 8478)	4.0	0.2	11.4	3.9	0.1	11.6
	(11.4)	(4.5)	(19.0)	(12.0)	(5.2)	(20.1)
Boreal forests/taiga (*n* = 200)	0.8	0.6	17.5	1.0	0.5	16.7
	(11.4)	(5.9)	(21.0)	(13.6)	(6.2)	(22.5)
(Sub)tropical grasslands/savannas/shrublands (*n* = 10141)	1.0	0.0	9.2	0.9	0.0	9.6
	(2.5)	(1.5)	(20.1)	(3.0)	(1.6)	(21.3)
Temperate grasslands/savannas/shrublands (*n* = 17205)	0.6	0.0	9.3	0.6	0.0	10.1
	(2.0)	(1.1)	(16.5)	(2.0)	(1.1)	(17.6)
Flooded grasslands/savannas (*n* = 545)	2.9	0.4	18.6	2.8	0.3	19.6
	(9.1)	(3.6)	(29.3)	(10.4)	(4.6)	(31.5)
Montane grasslands/shrublands (*n* = 2023)	0.7	−0.2	0.0	0.7	−0.2	1.0
	(2.7)	(1.6)	(15.9)	(3.1)	(1.6)	(17.0)
Tundra (*n* = 193)	−12.4	−7.6	13.0	−11.7	−7.2	14.7
	(20.5)	(15.2)	(17.2)	(21.7)	(15.5)	(20.1)
Mediterranean forests/woodlands/scrub (*n* = 935)	0.3	−0.1	3.6	0.3	−0.2	3.5
	(2.2)	(1.7)	(15.3)	(2.3)	(1.8)	(16.0)
Deserts and Xeric shrublands (*n* = 9808)	0.3	−0.1	4.7	0.3	−0.1	5.1
	(1.2)	(0.9)	(16.3)	(1.3)	(0.9)	(17.1)
Mangroves (*n* = 184)	1.9	0.3	15.7	1.9	0.3	15.3
	(5.4)	(4.9)	(26.7)	(6.4)	(5.1)	(26.8)

At 30 arc-seconds, the overall MDs and RMSDs were larger compared to the 5 arc-minutes. The difference was anticipated and can be explained by the smoothing effect of aggregating to the coarser resolution. SOC_{MD} and SOC_{RMSD} were substantially higher in the topsoil (14.1%, 22.1%) than in the subsoil (1%, 6.3%) in topsoil. The clay content was overestimated (MD = −16.9, RMSD = 25.7). At 5 arc-minutes, the trends were similar though less strong. The differences strongly deviate by terrestrial ecoregion but the MD's were found low when evaluated per ecoregion, the majority of the ecoregions had MD smaller than 2%. In general, S-World slightly overestimates the carbon content within most terrestrial ecoregions, though for tundra environments the SOC estimate was underestimated. Areas where the ensemble maps show large variability coincide with the areas where S-World is uncertain. However, S-World is within range for most of the land surface.

However, the validation dataset had a bias to those areas where the WISE observations were located. The distance from the validation points to the nearest WISE observation was 143 km (s.d. 108 km), whereas the average distance from grid cells to the validation points was 344 km (s.d. 350 km).

3.3. Methodology Evaluation

When S-World was developed the only large global soil profile database was the WISE database. The recent availability of the much larger WoSIS database now allows us to evaluate whether the WISE database gave a good indication in the ranges of soil properties per soil type. The results for the range in topsoil SOC of the 50 most important soil types are presented in Figure 7. Large differences are found between the soil types. For some soil types the ranges are almost identical but for other soil types there are large differences which will certainly lead to large differences in the estimates in the S-world database. Overall, there was a 54.2% overlap as calculated as an area weighted average overlap over all the soil types. The results for the subsoil SOC and clay showed even larger discrepancies with an overlap of 44.5% and 29.2% respectively. A more in-depth study identified a number of explanations. There was a clear relationship between the number of observations and the overlap. A larger number of observations in the WISE in general gave a better estimate. However, in some cases, the soil profiles in the WISE and WoSIS database are clustered. This means that a soil type may be described by a relatively large number of profiles, but that many of the profiles are clustered in a small region. The larger WoSIS database may introduce a number of soil profiles of the soil type from different areas with distinct agro-ecological conditions. Finally, S-world had to use ranges of the soil groups for some soil types where the number of soil profiles for the soil type was too limited. This happened in the case of 13 soil types. With the larger WoSIS databases, this only happened in the case of 6 soil types. It is clear that the larger WoSIS database provides a good opportunity to improve the estimates of the S-World database.

A second methodological evaluation was done by evaluating the arbitrary choice of considering the 1st and 9th decile to determine the ranges in soil properties. In the WoSIS database, there were many soil profiles that fell within the same grid cell. This allowed us to carry out the percentage of the variation that was explained within the grid cells. The results of the analysis showed that the variation within the grid cells was almost similar to the variation between the grid cells (43% for topsoil SOC, 23% for subsoil SOC, and 51% for clay%). Although the analysis is unable to identify whether the 1st and 9th decile should be used or, e.g., the 1st and 3rd quartile, the analysis does show that it is appropriate to decrease the range and eliminate some of the more extreme values.
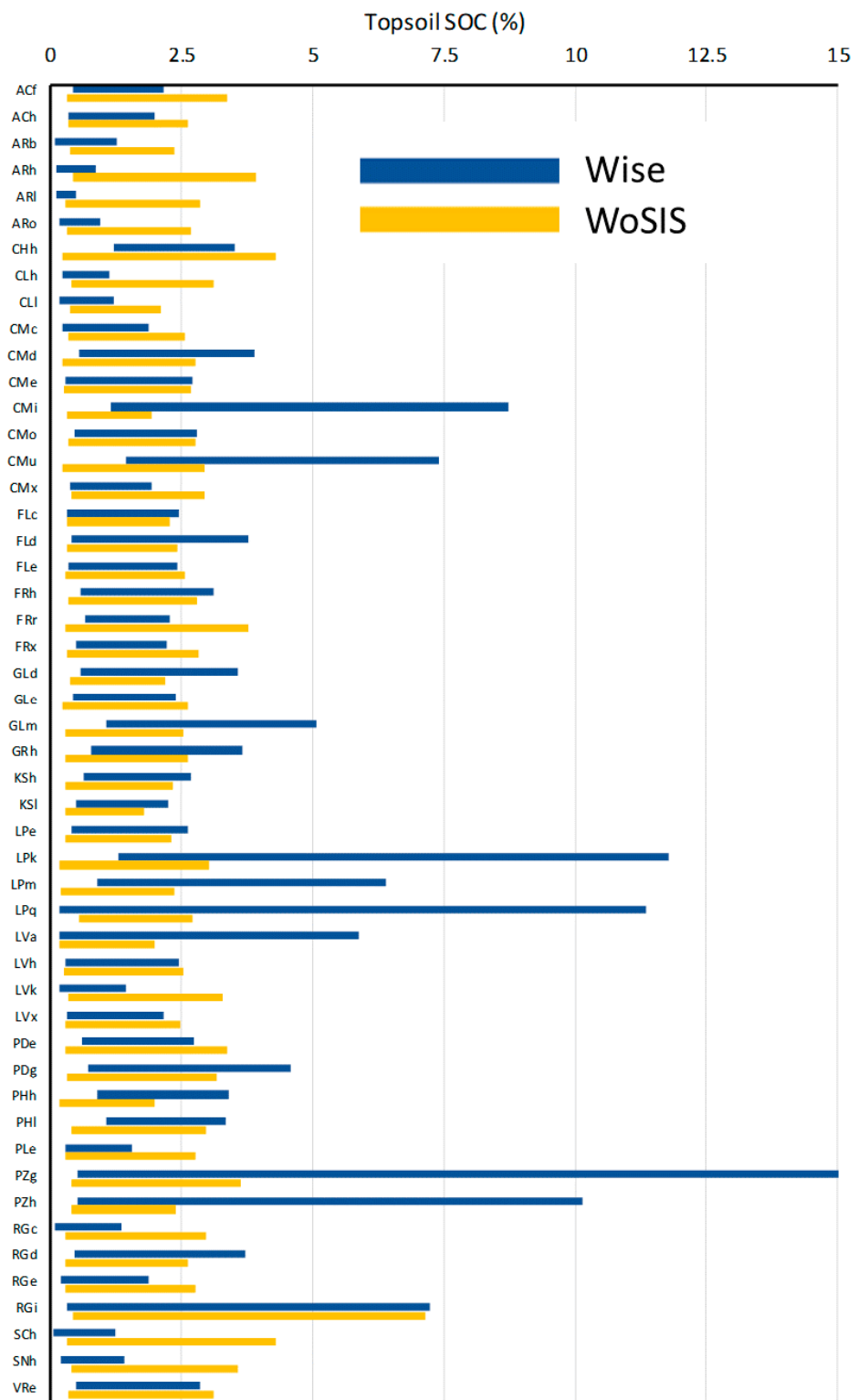
Figure 7. The ranges of topsoil SOC for the 50 most important soil types of the HWSD as estimated by the WISE (blue bars) and WoSIS (orange bars) databases.

3.4. General Discussion

Given the fact that true validations are practically impossible, there is an urgent need for other ways to properly review the range of available soil property maps. The inventory in this paper showed that there are multiple global soil property maps available. In various applications, studies make use of different maps. For example, SOC stocks

are estimated on the basis of the HWSD [29–32], the DSMW [15], or SoilGrids [33]. The water holding capacity is estimated on the basis of the DSMW [34] or on S-World [35]. The differences in the soil property maps may be one of the causes underlying the differences in many global environmental studies [19]. Theoretically, the results can be caused by differences in the acquisition methods. However, at the global level, many of the soil property maps rely on re-compilations and analyses of national surveys and soil profile databases. As such the variation between the various maps is also surprising as most maps are correlated to each other. However, the intensity with which soil resources have been studies varies considerably across the globe. This can be illustrated by the WoSIS database which compiled soil profile data. The observation density differs considerably between the various ecoregions (Table 4). The boreal forests/taiga had one observation per 73,325 km^2 whereas the (sub)tropical coniferous forests had one observation per 282 km^2.

Table 4. Observation density of the WoSIS database per terrestrial ecoregion.

Ecoregion	km^2/obs	Ecoregion	km^2/obs
(Sub)tropical moist broadleaf forests	2520	Temperate grasslands/savannas/shrublands	583
(Sub)tropical dry broadleaf forests	969	Flooded grasslands/savannas	1929
(Sub)tropical coniferous forests	282	Montane grasslands/shrublands	2416
Temperate broadleaf/mixed forests	635	Tundra	40,103
Temperate coniferous forests	476	Mediterranean forests/woodlands/scrub	3435
Boreal forests/taiga	75,325	Deserts and Xeric shrublands	2402
(Sub)tropical grasslands/savannas/shrublands	1961	Mangroves	1751

In the current study, we tried to circumvent some of the problems that are limiting a proper validation. We compared S-World to the ensemble of other maps available. The current study evaluated the S-World database that made use of the WISE database, which gave us the opportunity to carry out some level of validation making use of the other soil profiles in the WoSIS data. Finally, the same soil profiles could be used for the evaluation of some specific methodological elements. Although each of the steps provided additional insights concerning the quality of the S-World, each step also had it is clear disadvantages. The comparison with the ensemble is not a 100% fair comparison as many of the databases are based on the HWSD and DSMW. The validation with the "independent" WoSIS observations had a focus on those areas where also the WISE profiles were concentrated. In addition, WoSIS observations may have been used in surveys that are part of the HWSD. The methodological evaluation concentrated on some elements of the S-World methodology, but certainly was not able to deal with every individual step.

The analysis clearly showed some of the strengths and weaknesses of the S-world database. Whether or not the S-world database is good enough strongly depends on the specific requirements of each individual application. When studies focus on agricultural production then the large variation in the ensemble in the boreal areas is less important. When studies focus on the identification of relatively differences a bias in some of the variables may not be that serious. However, if global soil carbon stocks are estimated the actual level may be really important, whereas local spatial differences are less important. Like most global databases, the S-world database applied a single methodology over the entire globe. This has the advantage that the outcomes have a certain level of consistency. However, unique conditions like permafrost in the boreal areas may require specific approaches to be applied. Using an alternative approach or re-compilations from soil surveys for those regions is likely to improve the results. Good examples include the Northern Circumpolar Soil Carbon Database [36], the Unified North American Soil Map [37], and the Soil and Landscape Grid of Australia [38]. In addition, the review of the S-world soil property databases showed considerable differences between the three soil properties under study. As a result, it really depends which soil properties are being used in the application. Therefore, the choice of database and the evaluation of the database depends

whether one is studying soil erosion [39], soil salinity [40], ammonia emissions [41], crop growth [42], or Carbon stocks [43]

The study confirmed the importance of soil legacy data rescue [44] and global soil profile databases like WoSIS, as well as the necessity of the global soil community to invest in this kind of initiative.

4. Conclusions

A true validation of global soil property maps is impossible with the datasets that are currently available. However, the review in this paper, which included a comparison, a limited validation, and a methodological evaluation, certainly allows for a structured review of a global soil property map as illustrated for the S-world database. The comparison clearly indicates regional differences where studies are in agreement and where the largest uncertainties exist. When differences between the ensemble and the soil property database under review are found, this allows for further evaluation. The validation was limited as the database had a limited number of observations in the areas where the largest uncertainties (as indicated by the comparison) were found. Nevertheless, it helped to indicate the order of magnitude. The methodological evaluation clearly indicated the potential for improvement by using the WoSIS database and the importance of using deciles to truncate the ranges of soil properties per soil type.

Author Contributions: J.J.S. and V.L.M. jointly collected the databased from a variety of resources. Concepts were developed by J.J.S. and V.L.M. Data analysis was carried out by J.J.S. Writing was done jointly by J.J.S. and V.L.M. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The S-world database will become publicly available in the second quarter of 2021. All other datasets are publicly available: WoSIS: http://www.isric.org/data/wosis (accessed on 6 May 2021); GSDE: http://globalchange.bnu.edu.cn/research/soilw (accessed on 6 May 2021); HWSD: http://webarchive.iiasa.ac.at/Research/LUC/External-World-soil-database/HTML/ (accessed on 6 May 2021); IGBP DIS: https://daac.ornl.gov/SOILS/guides/igbp-surfaces.html (accessed on 6 May 2021); SoilGrids: https://www.soilgrids.org (accessed on 6 May 2021); WISE: http://www.isric.org/explore/wise-databases (accessed on 6 May 2021).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. UN-SDSN. *An Action Agenda for Sustainable Development. Report for the UN Secretary-General*; Prepared by the Leadership Council of the Sustainable Development Solutions Network; UN-SDSN: New York, NY, USA, 2014.
2. Hijmans, R.J.; Cameron, S.E.; Parra, J.L.; Jones, P.G.; Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **2005**, *25*, 1965–1978. [CrossRef]
3. Fujiwara, M.; Wright, J.S.; Manney, G.L.; Gray, L.J.; Anstey, J.; Birner, T.; Davis, S.; Gerber, E.P.; Harvey, V.L.; Hegglin, M.I.; et al. Introduction to the SPARC Reanalysis Intercomparison Project (S-RIP) and overview of the reanalysis systems. *Atmos. Chem. Phys.* **2017**, *17*, 1417–1452. [CrossRef]
4. Bouma, J. Soil science contributions towards Sustainable Development Goals and their implementation: Linking soil functions with ecosystem services. *J. Plant Nutr. Soil Sci.* **2014**, *177*, 111–120. [CrossRef]
5. FAO. *The Digital Soil Map of the World*; Version 3.5; FAO: Rome, Italy, 1995.
6. FAO; IIASA; ISRIC; ISSCAS; JRC. *Harmonized World Soil Database (Version 1.2)*; FAO: Rome, Italy; IIASA: Laxenburg, Austria, 2012; 42p.
7. Batjes, N.H. Harmonized soil profile data for applications at global and continental scales: Updates to the WISE database. *Soil Use Manag.* **2009**, *25*, 124–127. [CrossRef]

8. Hengl, T.; Mendes de Jesus, J.; Heuvelink, G.B.; Ruiperez Gonzalez, M.; Kilibarda, M.; Blagotić, A.; Shangguan, W.; Wright, M.N.; Geng, X.; Bauer-Marschallinger, B.; et al. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **2017**, *12*, e0169748. [CrossRef]

9. de Sousa, L.M.; Poggio, L.; Batjes, N.H.; Heuvelink, G.B.M.; Kempen, B.; Riberio, E.; Rossiter, D. SoilGrids 2.0: Producing quality-assessed soil information for the globe. *Soil* **2020**. [CrossRef]

10. Batjes, N.H.; Ribeiro, E.; van Oostrum, A.; Leenaars, J.; Hengl, T.; Mendes de Jesus, J. WoSIS—Providing standardised soil profile data for the world. *Earth Syst. Sci. Data* **2017**, *9*, 1–14. [CrossRef]

11. Stehfest, E.; Van Vuuren, D.; Kram, T.; Bouwman, L.; Alkemade, R.; Bakkenes, M.; Biemans, H.; Bouwman, A.; den Elzen, M.; Janse, J.; et al. *Integrated Assessment of Global Environmental Change with IMAGE 3.0. Model Description and Policy Applications*; PBL Netherlands Environmental Assessment Agency: The Hague, The Netherlands, 2014.

12. Van der Esch, S.; ten Brink, B.; Stehfest, E.; Bakkenes, M.; Sewell, A.; Bouwman, A.; Meijer, J.; Westhoek, H.; van den Berg, M. *Exploring Future Changes in Land Use and Land Condition and the Impacts on Food, Water, Climate Change and Biodiversity: Scenarios for the Global Land Outlook*; PBL Netherlands Environmental Assessment Agency: The Hague, The Netherlands, 2017.

13. Stoorvogel, J.J.; Bakkenes, M.; Temme, A.J.A.M.; Batjes, N.H.; ten Brink, B.J.E. S-World: A Global Soil Map for Environmental Modelling. *Land Degrad. Dev.* **2017**, *28*, 22–33. [CrossRef]

14. Stoorvogel, J.J.; Bakkenes, M.; ten Brink, B.J.E.; Temme, A.J.A.M. To what extent did we change our soils? A global comparison of natural and current conditions. *Land Degrad. Dev.* **2017**, *28*, 1982–1991. [CrossRef]

15. Scharlemann, J.P.W.; Tanner, E.V.J.; Hiederer, R.; Kapos, V. Global soil carbon: Understanding and managing the largest terrestrial carbon pool. *Carbon Manag.* **2014**, *5*, 81–91. [CrossRef]

16. Sterk, G.; Stoorvogel, J.J. Desertification–Scientific Versus Political Realities. *Land* **2020**, *9*, 156. [CrossRef]

17. Dai, Y.; Shangguan, W.; Wei, N.; Xin, Q.; Yuan, H.; Zhang, S.; Yan, F. A review of the global soil property maps for earth system models. *Soil* **2019**, *5*, 137–158. [CrossRef]

18. IPBES. *Summary for Policymakers of the Thematic Assessment Report on Land Degradation and Restoration of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services*; IPBES Secretariat: Bonn, Germany, 2018.

19. Hendriks, C.M.J.; Gibson, H.S.; Trett, A.; Python, A.; Weiss, D.J.; Vrieling, A.; Moyes, C.L. Mapping geospatial processes affecting the environmental fate of agricultural pesticides in Africa. *Int. J. Environ. Res. Public Health* **2019**, *16*, 3523. [CrossRef]

20. Tifafi, M.; Guenet, B.; Hatté, C. Large differences in global and regional total soil carbon stock estimates based on SoilGrids, HWSD, and NCSCD: Intercomparison and evaluation based on field data from USA, England, Wales, and France. *Glob. Biogeochem. Cycles* **2018**, *32*, 42–56. [CrossRef]

21. Shangguan, W.; Dai, Y.; Duan, Q.; Liu, B.; Yuan, H. A Global Soil Data Set for Earth System Modeling. *J. Adv. Modeling Earth Syst.* **2014**, *6*, 249–263. [CrossRef]

22. Global Soil Data Task Group. *Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS)*; ORNL DAAC: Oak Ridge, TN, USA, 2000. [CrossRef]

23. Batjes, N.H. (Ed.) *A Homogenized Soil Data File for Global Environmental Research: A Subset of FAO, ISRIC and NRCS Profiles (Version 1.0)*; Working Paper and Preprint 95/10b; International Soil Reference and Information Centre: Wageningen, The Netherlands, 1995.

24. Batjes, N.H. Harmonised soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. *Geoderma* **2016**, *269*, 61–68. [CrossRef]

25. Hiederer, R.; Köchy, M. *Global Soil Organic Carbon Estimates and the Harmonized World Soil Database*; JRC-IEC: Luxembourg, 2011.

26. Olson, D.M.; Dinerstein, E.; Wikramanayake, E.D.; Burgess, N.D.; Powell, G.V.N.; Underwood, E.C.; D'Amico, J.A.; Itoua, I.; Strand, H.E.; Morrison, J.C.; et al. Terrestrial Ecoregions of the World: A New Map of Life on Earth. *BioScience* **2001**, *51*, 933–938. [CrossRef]

27. Chen, S.; Mulder, V.L.; Heuvelink, G.B.M.; Poggio, L.; Caubet, M.; Román Dobarco, M.; Arrouays, D. Model averaging for mapping topsoil organic carbon in France. *Geoderma* **2020**, *366*. [CrossRef]

28. Arrouays, D.; McKenzie, N.J.; Hempel, J.; Richer de Forges, A.C.; McBratney, A. *GlobalSoilMap: Basis of the Global Spatial Soil Information System*; CRC Press Taylor & Francis Group: Boca Rotan, FL, USA, 2014.

29. Chen, S.; Arrouays, D. Soil carbon stocks are underestimated in mountainous regions. *Geoderma* **2018**, *320*, 146–148. [CrossRef]

30. Deng, X.; Chen, X.; Ma, W.; Ren, Z.; Zhang, M.; Grieneisen, M.L.; Long, W.; Ni, Z.; Zhan, Y.; Lv, X. Baseline map of organic carbon stock in farmland topsoil in East China. *Agric. Ecosyst. Environ.* **2018**, *254*, 213–223. [CrossRef]

31. Köchy, M.; Hiederer, R.; Freibauer, A. Global distribution of soil organic carbon—Part 1: Masses and frequency distributions of SOC stocks for the tropics, permafrost regions, wetlands, and the world. *Soil* **2015**, *1*, 351–365. [CrossRef]

32. Yigini, Y.; Montanarella, L.; Panagos, P. Chapter Thirteen—European Contribution Towards a Global Assessment of Agricultural Soil Organic Carbon Stocks. *Adv. Agron.* **2017**, *142*, 385–410.

33. Mulder, V.L.; Lacoste, M.; Richer-de-Forges, A.C.; Martin, M.P.; Arrouays, D. National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma* **2016**, *263*, 16–34. [CrossRef]

34. Reynolds, C.A.; Jackson, T.J.; Rawls, W.J. Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resour. Res.* **2000**, *36*, 3653–3662. [CrossRef]

35. Stoorvogel, J.J.; Mulder, V.L.; Hendriks, C.M.J. The effect of disaggregating soil data for estimating soil hydrological parameters at different scales. *Geoderma* **2019**, *347*, 185–193. [CrossRef]

36. Hugelius, G.; Tarnocai, C.; Broll, G.; Canadell, J.G.; Kuhry, P.; Swanson, D.K. The Northern Circumpolar Soil Carbon Database: Spatially distributed datasets of soil coverage and soil carbon storage in the northern permafrost regions. *Earth Syst. Sci. Data* **2013**, *5*, 3–13. [CrossRef]
37. Liu, S.; Wei, Y.; Post, W.M.; Cook, R.B.; Schaefer, K.; Thornton, M.M. The Unified North American Soil Map and its implication on the soil organic carbon stock in North America. *Biogeosciences* **2013**, *10*, 2915–2930. [CrossRef]
38. Grundy, M.J.; Viscarra Rossel, R.A.; Searle, R.D.; Wilson, P.L.; Chen, C.; Gregory, L.J. Soil and Landscape Grid of Australia. *Soil Res.* **2015**, *53*, 835–844. [CrossRef]
39. Borrelli, P.; Robinson, D.A.; Fleischer, L.R.; Lugato, E.; Ballabio, C.; Alewell, C.; Meusburger, K.; Modugno, S.; Schütt, B.; Ferro, V.; et al. An assessment of the global impact of 21st century land use change on soil erosion. *Nat. Commun.* **2017**, *8*, 2013. [CrossRef]
40. Hassani, A.; Azapagic, A.; Shokri, N. Predicting long-term dynamics of soil salinity and sodicity on a global scale. *Proc. Natl. Acad. Sci. USA* **2021**, *117*, 33017–33027. [CrossRef]
41. Ma, R.; Zou, J.; Han, Z.; Yu, K.; Wu, S.; Li, Z.; Zhu-Barker, X. Global soil-derived ammonia emissions from agricultural nitrogen fertilizer application: A refinement based on regional and crop-specific emission factors. *Glob. Chang. Biol.* **2021**, *27*, 855–867. [CrossRef]
42. Romero, C.C.; Hoogenboom, G.; Baigorria, G.A.; Koo, J.; Gijsman, A.J.; Wood, S. Reanalysis of a global soil database for crop and environmental modeling. *Environ. Model. Softw.* **2012**, *35*, 163–170. [CrossRef]
43. Stockmann, U.; Padarian, J.; McBratney, A.; Minasny, B.; de Brogniez, D.; Montanarella, L.; Young Hong, S.; Rawlins, B.G.; Field, D.J. Global soil organic carbon assessment. *Glob. Food Secur.* **2015**, *6*, 9–16. [CrossRef]
44. Arrouays, D.; Leenaars, J.G.; Richer-de-Forges, A.C.; Adhikari, K.; Ballabio, C.; Greve, M.; Grundy, M.; Guerrero, E.; Hempel, J.; Hengl, T.; et al. Soil legacy data rescue via GlobalSoilMap and other international and national initiatives. *GeoResJ* **2017**, *14*, 1–19. [CrossRef] [PubMed]