Article

Global Mapping of Soil Water Characteristics Parameters—Fusing Curated Data with Machine Learning and Environmental Covariates

Surya Gupta [1,*], Andreas Papritz [1], Peter Lehmann [1], Tomislav Hengl [2,3], Sara Bonetti [4] and Dani Or [1,5]

1   Soil and Terrestrial Environmental Physics, Department of Environmental Systems Science, ETH Zürich, 8092 Zürich, Switzerland; andreas.papritz@env.ethz.ch (A.P.); peter.lehmann@env.ethz.ch (P.L.); dani.or@env.ethz.ch (D.O.)
2   OpenGeoHub Foundation, Agro Business Park 10, 6708 PW Wageningen, The Netherlands; tom.hengl@opengeohub.org
3   EnvirometriX, Agro Business Park 10, 6708 PW Wageningen, The Netherlands
4   Soil Physics and Land Management Group, Wageningen University, 6708 PB Wageningen, The Netherlands; sara.bonetti@wur.nl
5   Division of Hydrologic Sciences, Desert Research Institute, Reno, NV 89512, USA
*   Correspondence: surya.gupta@usys.ethz.ch

Abstract: Hydrological and climatic modeling of near-surface water and energy fluxes is critically dependent on the availability of soil hydraulic parameters. Key among these parameters is the soil water characteristic curve (SWCC), a function relating soil water content (θ) to matric potential (ψ). The direct measurement of SWCC is laborious, hence, reported values of SWCC are spatially sparse and usually have only a small number of data pairs (θ, ψ) per sample. Pedotransfer function (PTF) models have been used to correlate SWCC with basic soil properties, but evidence suggests that SWCC is also shaped by vegetation-promoted soil structure and climate-modified clay minerals. To capture these effects in their spatial context, a machine learning framework (denoted as Covariate-based GeoTransfer Functions, CoGTFs) was trained using (a) a novel and comprehensive global dataset of SWCC parameters and (b) global maps of environmental covariates and soil properties at 1 km spatial resolution. Two CoGTF models were developed: one model (CoGTF-1) was based on predicted soil covariates because measured soil data are not generally available, and the other (CoGTF-2) used measured soil properties to model SWCC parameters. The spatial cross-validation of CoGTF-1 resulted, for the predicted van Genuchten SWCC parameters, in concordance correlation coefficients (CCC) of 0.321–0.565. To validate the resulting global maps of SWCC parameters and to compare the CoGTF framework to two pedotransfer functions from the literature, the predicted water contents at 0.1 m, 3.3 m, and 150 m matric potential were evaluated. The accuracy metrics for CoGTF were considerably better than PTF-based maps.

Keywords: soil hydraulic properties; remote sensing; CoGTF; van Genuchten parameters

1. Introduction

The quantification of hydrological processes in the soil unsaturated zone (infiltration, runoff, drainage, evaporation, and water storage) is critically dependent on the quality of parameters of the soil water characteristics curve (SWCC) describing the relationship between soil water content (θ) and matric potential (ψ) [1–3]. Computing climatic predictions by contemporary land surface models (LSMs) requires highly resolved maps of SWCC parameters for the terrestrial surface of the Earth.

This rapidly expanding need for spatially exhaustive and highly resolved maps of SWCC parameters has prompted the use of readily available soil information such as soil texture to estimate parameters of SWCCs by pedotransfer functions (PTFs) [4,5]. PTFs

typically use soil texture, bulk density, and organic content to correlate with SWCC parameter values. A major shortcoming of these PTFs is the omission of the role that soil structure (aggregation of soil particles, biopores formed by plant roots and earthworms) exerts on surface fluxes [6,7]. PTFs are often trained using data on agricultural soils and are tailored for a specific region or country, and cannot be transferred to regions with different land-cover and soil-forming processes [8,9].

To overcome some of these inherent limitations in the estimation of soil hydraulic properties, new approaches that harness local information on soil-forming processes have been recently proposed to map the saturated hydraulic conductivity (Ksat) by linking Ksat measurements to environmental covariates on vegetation, climate, and topography using a machine learning algorithm [10]. Likewise, Chaney et al. [11] developed the maps of soil hydraulic properties (POLARIS dataset) at 30 m resolution using remote sensing covariates for the United States using the SSUGRO (Soil Survey Geographic) database. This study uses a similar approach and improves the spatial maps of SWCC parameters by injecting information on vegetation, topography, and climate along with spatial information on soil properties into SWCC mapping using a machine learning algorithm. To differentiate this approach from previous PTFs, we use the term Covariate-based GeoTransfer Functions (CoGTFs) [10].

The objectives of this study are:

1. To link an extensive dataset of SWCC parameters to environmental covariates using the CoGTF framework to develop global maps of van Genuchten (vG) parameters.
2. To compare the CoGTF derived parameter maps with earlier published maps based on PTFs (Rosetta 3 and HiHydroSoil v2.0).
3. To highlight the limitations related to the application of a model based on measured soil properties for global mapping when only predicted soil information is available.

2. Materials and Methods

2.1. Covariate-Based GeoTransfer Functions (CoGTFs) Framework

We used here the CoGTF framework proposed by Gupta et al. [10] where soil response variables were combined with environmental covariates using machine learning (specifically, we used here a random forest approach). As described in Gupta et al. [12], we prepared a geo-referenced dataset of the van Genuchten (vG) parameters that included also values of basic soil properties. Then, we overlaid the observation sites with maps of the environmental covariates to produce a regression matrix and fitted a random forest (RF) model to the data. After evaluating the performance of the RF model, we produced spatial predictions of the vG parameters for a global 1 km mesh of locations. The R environment was used for the whole analysis [13].

2.2. Training Data—Expanding the Global SWCC Dataset

We compiled and curated a dataset of 15,259 SWCCs (Figure 1) called the Global Soil Hydraulic Properties (GSHP) database [12]. This dataset is much larger than the training datasets used for two frequently applied PTFs, i.e., Rosetta 3 [14] and HiHydroSoil v2.0 [15] that is based on the PTF of Tóth et al. [16]. Rosetta 3 used 2134 SWCCs (mainly from North America and Europe) whereas HiHydroSoil v2.0 used 3773 SWCCs from Europe to develop the respective PTFs. Hence, maps of vG parameters deduced from Rosetta 3 or related to HiHydoSoil v2.0 have been calibrated by a rather limited dataset of SWCCs, and this motivated us to collect a more comprehensive dataset. We collected SWCC data from 2702 locations that represent all continents and climatic regions. The geographic and climatic distribution of locations with SWCC data was uneven, with most SWCCs measured in North America, followed by Africa, Europe, Asia, South America, and Australia/Oceania. Additionally, the majority of SWCCs (10,048 out of 15,259 curves) were obtained in temperate regions, and 2344, 1422, 1411, and 34 SWCCs were obtained from boreal, tropical, arid, and polar regions, respectively.
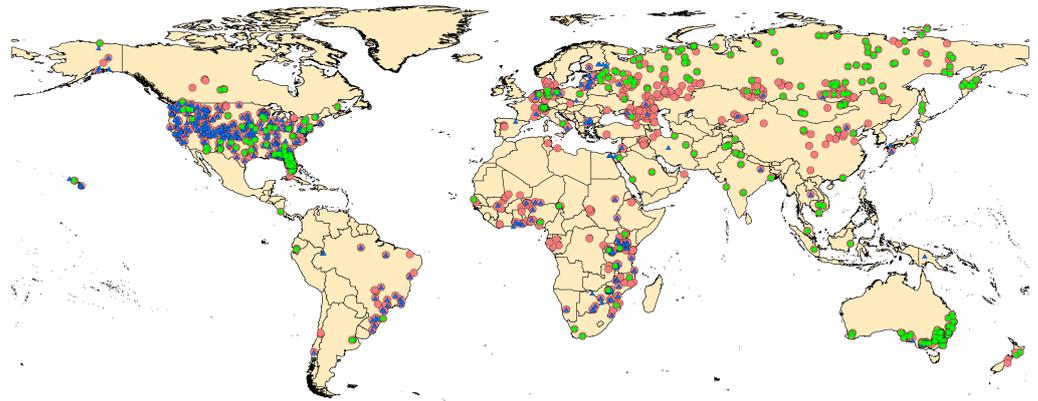
Figure 1. Collection of SWCC samples to produce global maps of van Genuchten parameters. The spatial distribution of SWCC data was shown with different categories. The colors are assigned to different quality criteria as explained in Figure A1.

We estimated the parameters of the simplified van Genuchten model with the restriction $m = 1 - 1/n$ (Equation (1)) for all 15,259 SWCCs using the *'soihypfit'* R package [17]:

$$\theta(\psi) = \theta_r + (\theta_s - \theta_r)[1 + (\alpha|\psi|)^n]^{-(1-1/n)} \tag{1}$$

where $\theta(\psi)$ is the volumetric water content (m^3/m^3) measured at matric potential ψ (m), θ_s and θ_r are the saturated and residual water contents, respectively (m^3/m^3), and α (m^{-1}) and n (dimensionless) are the inverse air entry pressure and shape parameters. To estimate the SWCC parameters, we restricted the parameter range for α to 0–100 m^{-1} and for n to 1–7. SWCCs, where the estimated values were equal to the upper limits of the parameter ranges were discarded. We further computed profile likelihood-based confidence intervals for the α and n parameters (see Chapter 4 in Uusipaikka [18]). The limits of the confidence intervals guided us to remove SWCCs that had a flat upper loglikelihood profile for α and n (i.e., SWCCs that had a high or indeterminate estimated standard error). The details of this analysis are explained in Gupta et al. [12].

2.3. Soil and Environmental Covariates

We combined spatial information on four soil properties (sand and clay content, soil depth, and bulk density) with 21 environmental covariates for modeling vG parameters, all globally available from https://www.openlandmap.org/ (accessed on 22 December 2021) [19]. The environmental covariates were selected to represent ecological conditions important for soil formation [20], including topography, precipitation patterns, and vegetation cover. The 25 spatial covariates are classified into five groups: (a) climate, (b) terrain, (c) surface reflectance, (d) vegetation, and (e) soil properties. The covariates were selected as described in Table A1 with respect to their hypothesized relevance for modeling vG parameter values.

2.4. CoGTF Models Based on Measured and Predicted Soil Covariates

As mentioned above, we trained the CoGTF model using environmental covariates and soil properties. However, related to soil covariates, we faced the problem that we knew measured soil properties for most SWCC locations but not for the locations of the prediction mesh. For global mapping, we completely relied on predicted soil covariates. On the one hand, models based on measured soil properties likely better reveal the 'true' relations between soil properties, environmental covariates, and SWCCs. On the other hand, the model should be trained with the same type of information that is available on a global scale otherwise different types of information will be incoherently mixed. To explore the effect of using different types of soil information on the predictions of the VG parameters, we built two models within the CoGTF framework. The first model CoGTF-1 was developed using

predicted soil properties as covariates, extracted from https://www.openlandmap.org/ (accessed on 22 December 2021) and environmental covariates with 11,705 SWCCs parameter sets. The second model, CoGTF-2, was created using the same environmental covariates along with measured soil property data available only for 9958 SWCC parameters. A total of 1747 SWCCs were discarded due to a lack of measured soil properties data (Figures 1 and A1).

2.5. Computational Aspects of the Random Forest Model

Following the estimation of the vG parameters and the formation of regression matrix for both CoGTF-1, and CoGTF-2, the R package *'ranger'* [21] was used to fit the RF model. Partial dependence plots (see Section 10.13.2 in Hastie et al. [22]) were generated by the R packages *'hexbin'* [23], *'lattice'* [24], and *'viridis'* [25]. The optimal value of the most sensitive hyper-parameter *mtry* of RF was determined by spatial five-fold cross-validation (see Section 11.4 in Lovelace et al. [26]), and default values of *ranger* were used for the remaining hyper-parameters, e.g., number of trees, minimal node size, maximum tree depth, and splitting rule. For spatial cross-validation (SCV) the Earth's surface was divided into 1-degree by 1-degree blocks. We then formed the SCV partitions by randomly assigning the blocks to 5 subsets such that each contained about 20% of the data. The RF model was then fitted 5 times, leaving a subset out at a time and using it as a validation set. This SCV procedure was repeated 3 times, and predictions of the 3 SCV repetitions were averaged. The outcome of SCV analysis will be shown in the results section for the optimal *mtry* in a hexbin plot with a LOWESS (Locally Weighted Scatterplot Smoothing) line [27] to reveal the eventual conditional bias of the vG parameters' predictions.

The relative importance of the various covariates for modeling vG parameters was assessed by the node impurity, which is computed for RF regression problems as the decrease of residual sum of squares (RSS) when a particular covariate splits the data at the nodes of a tree (see Sections 10.13.1 and 15.3.2 in Hastie et al. [22]). The variable that provided maximum decline in RSS (and consequently increase in node purity) was considered the most important variable, the variable with the second-largest RSS decrease was considered the second most important, and so on.

The SCV and the list of important covariates were estimated for both CoGTF-1 and CoGTF-2. For the SCV of CoGTF-1, the predicted soil covariates were used in calibration and validation whereas, for CoGTF-2, the measured soil covariates were used to calibrate and validate the model. To show the effect of 'mixing' soil information in model formulation and application, we conducted another SCV using measured soil properties for the calibration but predicted soil properties for the validation. The results are presented in Figure A6.

2.6. Criteria to Assess Predictive Accuracy of SWCC Parameters

The accuracy of the SCV predictions was evaluated using BIAS, root mean square error (RMSE), coefficient of determination (R^2), and concordance correlation coefficient (CCC).

BIAS and RMSE are defined as

$$\text{BIAS} = \sum_{i=1}^{N} \frac{(\hat{y}_i - y_i)}{N} \tag{2}$$

$$\text{RMSE} = \sqrt{\frac{\text{SSE}}{N}} \tag{3}$$

where

$$\text{SSE} = \sum_{i=1}^{N} (\hat{y}_i - y_i)^2$$

is the sum of squared errors between the predictions $\hat{y}_i$ (predicted vG parameters and water content θ at matric potential ψ) and the measurements y_i (vG parameter and water content θ measured at matric potential ψ), and N is the total number of observations.

R^2 is defined as

$$R^2 = \left[1 - \frac{\text{SSE}}{\text{SST}} \right] \qquad (4)$$

where

$$\text{SST} = \sum_{i=1}^{N} (y_i - \bar{y})^2$$

is the total sum of squares and $\bar{y}$ the arithmetic mean of the measurements.

The concordance correlation coefficient (CCC) [28], a further measure of the agreement between observed and predicted vG parameter values and for validating predictions of water contents θ at prescribed matrix potentials ψ in CoGTF framework, is given by

$$\text{CCC} = \frac{2 \cdot \rho \cdot s_{\hat{y}} \cdot s_y}{s_{\hat{y}}^2 + s_y^2 + \text{BIAS}^2} \qquad (5)$$

where ρ is the Pearson correlation coefficient between $\hat{y}_i$ and y_i, and $s_{\hat{y}}$ and s_y are the respective standard deviations. CCC is equal to 1 for a perfect model with $\hat{y}_i = y_i$ in which case we have $\text{BIAS} = 0$, $\rho = 1$ and $s_{\hat{y}} = s_y$.

2.7. Validation of Van Genuchten Parameter Maps

The Earth's surface was again divided into 1-degree $\times$ 1-degree blocks. For a fair comparison of the CoGTF vG parameter maps with Rosetta 3 and HiHydroSoil v2.0, we used a validation dataset from all continents except North America and Europe because Rosetta 3 and HiHydroSoil v2.0 PTFs were calibrated with the data from North America and Europe. To compute the validation predictions by the CoGTF model, we split the validation data into five parts and predicted the parameters of each part in turn by an RF model that was calibrated with the data of the remaining four parts complemented by the data from North America and Europe. The predictions by Rosetta 3 and HiHydoSoil v2.0 were obtained by overlaying the locations of the validation data with the respective maps. To compute CCC, R^2, RMSE, and BIAS, we compared the predicted water content with measured water content for three matric potentials (0.1 m, 3.3 m, and 150 m). Predicted water content at these matric potentials was calculated by inserting the predicted vG parameters into Equation (1).

As in the case of SCV, we quantified the model performance of both CoGTF models (CoGTF-1, and CoGTF2). The totals of 2621 SWCCs for COGTF-1 and 1817 SWCCs for CoGTF-2 were used as the validation datasets. Only for a subset of these SWCCs, the water content had been measured at 0.1, 3.3, and 150 m matric potential: 1572 (0.1 m), 719 (3.3 m), and 1184 (150 m) SWCCs for CoGTF-1 and 856 (0.1 m), 611 (3.3 m), and 732 (150 m) SWCCs for CoGTF-2, respectively. We evaluated the accuracy criteria for all of these subsets for either CoGTF-1 or CoGTF-2, Rosetta 2, and the HiHydoSoil v2.0 maps.

3. Results

3.1. Covariate Importance and Model Performance for the CoGTF Approach

Figure 2 lists the most important covariates for vG parameter modeling. The figure shows that the covariate importance differed between the various vG parameters (some parameters were dominated by climate covariates, others by soil properties) but was rather similar for CoGTF-1, and CoGTF-2. Climate covariates were the most important factors for the inverse air entry pressure parameter α (Figure 2a,e) whereas sand and clay were the key covariates for the shape parameter n (Figure 2b,f). In addition, for residual (Figure 2c,g) and saturated water content (Figure 2d,h), soil covariates were the dominant factors (soil depth and clay content for residual and sand and bulk density for saturated water content).
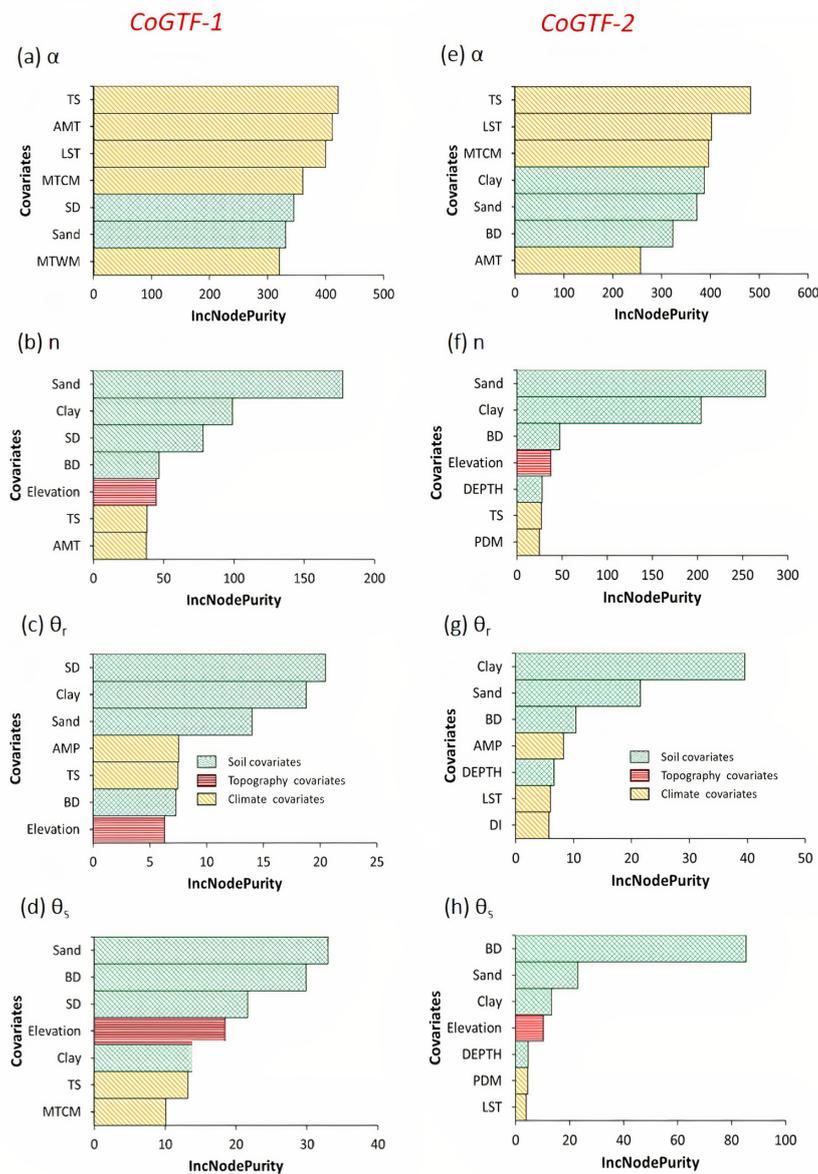
Figure 2. Importance of the covariates for modeling vG parameters by a random forest model. The x-axis displays the average increase in node purity (the larger the value, the more important is a covariate). The 7 most important covariates are shown for both models (CoGTF-1 with predicted soil covariates and CoGTF-2 with measured soil covariates). The plots (**a,e**) show the importance for the common logarithm $log_{10} \alpha$ of inverse air entry pressure parameter (unit of α m^{-1}), (**b,f**) for the shape parameter n, (**c,g**) for the residual water content θ_r, and (**d,h**) for the saturated water content θ_s. Sand content, bulk density (BD), soil depth (SD), and clay content belong to soil covariates. Elevation is one of the terrain covariates. Temperature seasonality (TS), minimum temperature of warmest month (MTCM), annual average land surface temperature (LST), minimum temperature of coldest month (MTWM), precipitation of driest month (PDM), mean annual temperature (AMT), diffuse irradiation (DI), and mean annual precipitation (AMP) belong to the climate category.

The results of SCV for both models CoGTF-1 and CoGTF-2 are presented by hexbin density plots in Figure 3. For parameters α, the line of LOWESS fell close to the 1:1 line, hence predictions were approximately conditionally unbiased for both models. However, for θ_s and θ_r the LOWESS lines were unbiased up to approximately 0.15 and 0.60 m^3/m^3, respectively, but slightly biased for higher water content. SCV of CoGTF-1 showed concordance correlation coefficients (CCC) for $log_{10} \alpha$ (unit of α m^{-1}), n, θ_r, and θ_s of 0.432, 0.525,

0.321, and 0.565, respectively, and coefficients of determination (R^2) of 0.273, 0.325, 0.172, and 0.412 , respectively. The root mean square errors (RMSE) for the four parameters were equal to 0.442, 0.888, 0.076, and 0.082, respectively, and the BIASes equal to 0.008, 0.050, −0.001, and 0.001, respectively.
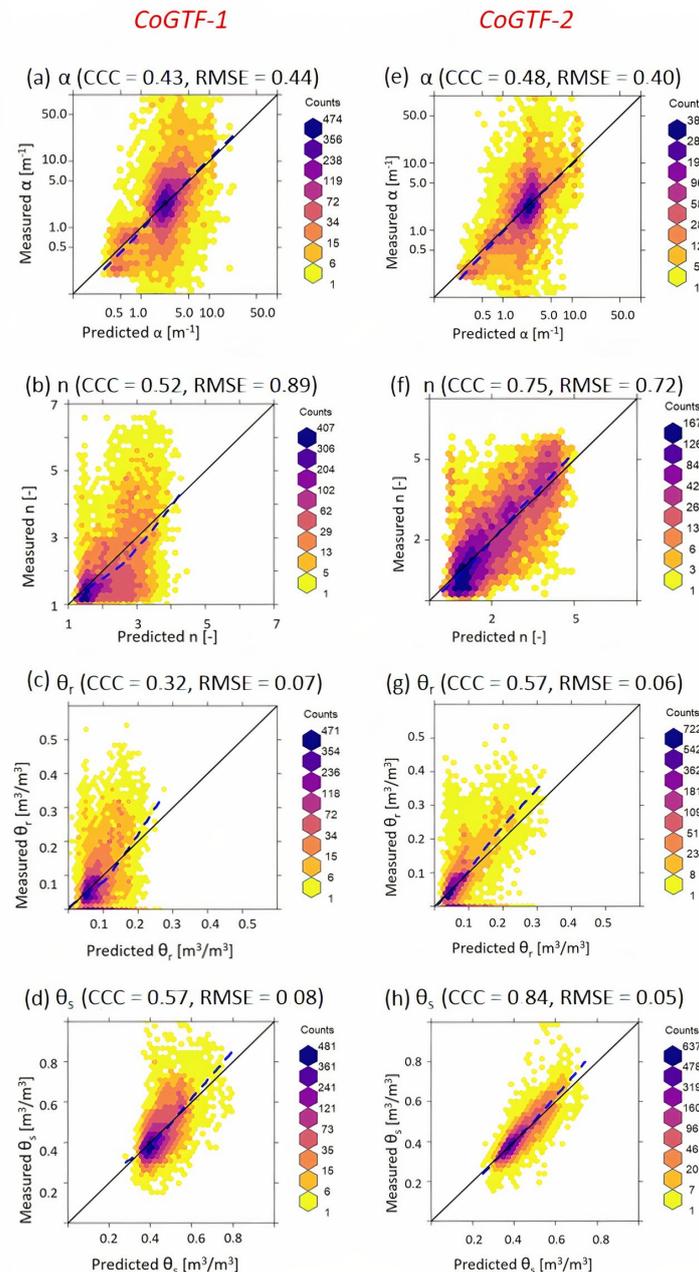

Figure 3. Correlation between observations and SCV predictions of van Genuchten parameters for the CoGTF-1 model using predicted soil covariates on the left and for the CoGTF-2 model using measured soil properties on the right. The plots (**a**,**e**) show inverse air entry pressure parameter log_{10} α (unit of α m^{-1}), (**b**,**f**) shape parameter n, (**c**,**g**) the residual water content θ_r, and (**d**,**h**) the saturated water content θ_s. The term 'measured' on the axes of the figures relates to parameter estimates obtained by fitting the vG model to measured SWCCs. The color code represented the number of observations in each hexagonal bin. The solid black line is the 1:1 line, and the blue dashed line the LOWESS (locally weighted scatter plot smoothing) curve.

Likewise, SCV for CoGTF-2 showed concordance correlation coefficients (CCC) for log_{10} α (unit of α m^{-1}), n, θ_r, and θ_s of 0.481, 0.753, 0.573, and 0.835, and coefficients of

determination (R^2) of 0.323, 0.605, 0.421, and 0.735, respectively. The root mean square errors (RMSE) for the four parameters were equal to 0.400, 0.723, 0.060, and 0.050, and the BIASes equal to -0.009, 0.004, 0.005, and 0.002, respectively. The model performance of CoGTF-2 was seemingly better than for CoGTF-1, however, the SCV criteria of CoGTF-2 do not characterize the predictive quality honestly for locations where measured soil property data are lacking, i.e., for the locations of the global prediction mesh.

3.2. Global Maps of vG Parameters Based on CoGTF-1

Figure 4 shows the global maps of the vG parameters at the soil surface (0.00 m depth) as calculated by the CoGTF-1 model. Note that the global maps with CoGTF-2 could be created as well, but are not shown here because they were built with 'mixed' soil information using measured soil covariates to calibrate the model but substituted predicted soil covariates for the missing soil property measurements when computing the predictions. CoGTF-1 predictions of the inverse air entry pressure parameter α (Figure 4a) ranged from 0.11 to 48 (m^{-1}). High α values (corresponding to soils with dominantly large pores and small air entry pressure) were predicted for the Sahara, the Middle East (regions with high sand content), and tropical regions of northern South America and Central Africa. Low α values were found in the higher northern latitudes with a colder climate. We can conclude that predicted α values are large for coarse textures and for warm regions with high rainfall rates (more intense structure formation). High values of the shape parameter n (Figure 4b) were predicted for Sahara, the Middle East, Central Asia, and the Gobi Desert (regions with high sand content) but also in the higher northern latitudes where soils with high organic matter content and high porosity (low bulk density) are common. In tropical regions both saturated (Figure 4c) and residual water contents (Figure 4d) were high. As expected, the saturated water content was strongly related to bulk density with high values in the northern regions with low bulk density soils. Figure A2 shows the cumulative distribution function (CDF) for global maps of vG parameters at different depths (0, 30, 60, and 100 cm) for CoGTF-1. There is no clear change in vG parameters values with depth except for θ_s with much higher values at the surface compared to the deeper soil layer. This is very likely due to low bulk density in topsoil horizons with high organic content.

3.3. Comparison with Alternate Global vG Parameters Maps

We compared the new parameter maps from CoGTF-1 with the maps generated with the PTFs Rosetta 3 [14] and HiHydroSoil v2.0 [15]. In Figure 5 we show the comparison for the shape parameter n and provide the other maps in Appendix A. Note that in the comparison, only CoGTF-1 maps are shown. In contrast to the CoGTF maps, Rosetta 3 predicted large values of n only for desert regions with high sand content. CoGTF predicted large values as well for arid regions in general and northern latitudes with low bulk densities in particular. The HiHydroSoil v2.0 map predicted low n values for the entire globe. The differences between the three maps are clearly visible in the probability density functions (PDFs) shown in Figure 6b. The medians of globally predicted n values were equal to 1.67 (CoGTF-1), 1.50 (Rosetta 3), and 1.34 (HiHydroSoil v2.0), respectively. CoGTF-based α and n maps showed a wider distribution compared to maps based on PTFs (see PDFs in Figure 6a,b). Both Rosetta 3 and HiHydroSoil v2.0 predicted lower α and θ_r values in most regions of the world compared to CoGTF (see PDFs in Figure 6a,c and maps in Figures A3 and A4). The CoGTF-based θ_s map appeared similar to the other two maps (see PDF in Figure 6d and global maps in Figure A5).
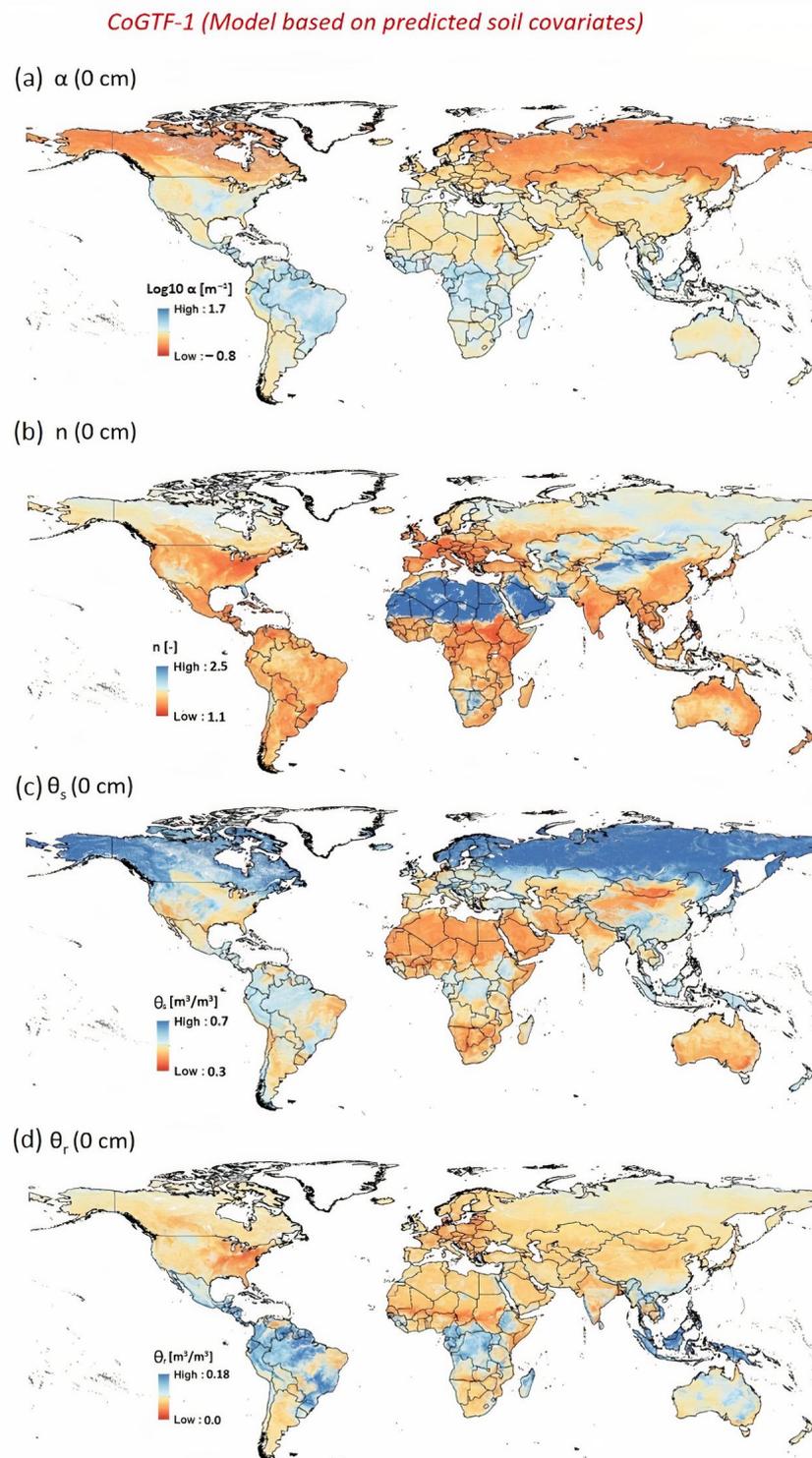
Figure 4. Global maps of van Genuchten parameter at the soil surface (marked as '0 cm') calculated with CoGTF-1 model based on predicted soil covariates. The maps (**a**) show inverse air entry pressure parameter $log_{10} \alpha$ (unit of α m^{-1}), (**b**) shape parameter n, (**c**) the residual water content θ_r, and (**d**) the saturated water content θ_s. The parameter n is high for sandy soils. Predictably, θ_s is affected by soil bulk density map (high θ_s values in high latitudes dominated by organic soils with low bulk density). In contrast, the parameter α shows low values in these northern latitudes (large pores and low capillarity).
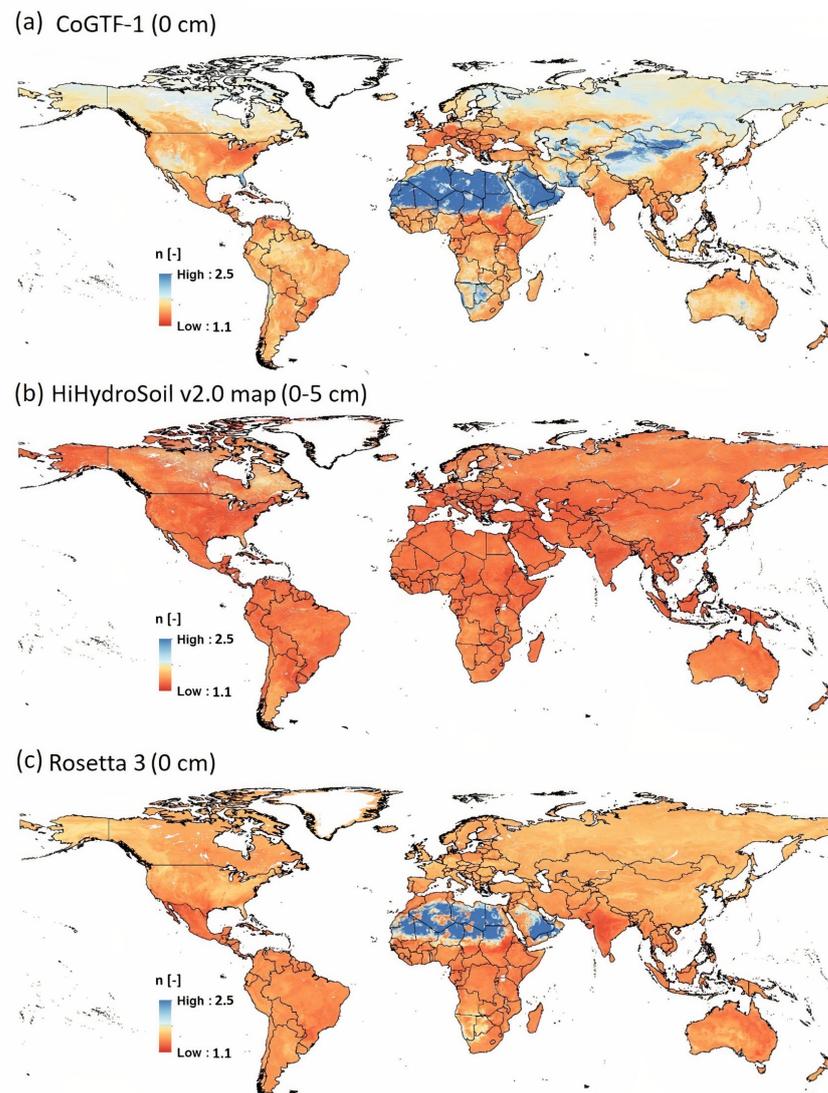
(a) CoGTF-1 (0 cm)

n [-]
High : 2.5
Low : 1.1

(b) HiHydroSoil v2.0 map (0-5 cm)

n [-]
High : 2.5
Low : 1.1

(c) Rosetta 3 (0 cm)

n [-]
High : 2.5
Low : 1.1

Figure 5. Visual comparison of global maps of shape parameter *n* between (**a**) CoGTF-1 (based on predicted soil covariates), (**b**) HiHydroSoil v2.0, and (**c**) Rosetta 3. Large values are shown for regions with high sand content for CoGTF-1 and Rosetta 3. In contrast to Rosetta 3, large *n* values were predicted as well by CoGTF-1 for cold regions with high bulk densities. The values from HiHydroSoil v2.0 map were consistently small in all regions.

3.4. Validation of CoGTF and Other PTF Based Maps

Table 1 shows the results of the validation performed for CoGTF-1 (with predicted soil covariates), Rosetta 3, and HiHydroSoil v2.0 maps, where measured water content was compared with the predicted water content at 0.1 m, 3.3 m, and 150 m matric potentials. All the maps suffered from some negative bias. The bias was small and its contribution to RMSE was negligible for water content at 0.1 m and 150 m matric potentials, however for water content at 3.3 m matrix potential BIAS contributed substantially to the overall error (HiHydroSoil v2.0: $BIAS^2/RMSE^2$ = 0.57). On the whole, the CoGTF predictions had the smallest RMSE and largest CCC and R^2 for water contents measured at the three matric potential values. Note that the R^2 was negative for the Rosetta and HiHydroSoil v2.0 predictions of the water content at 3.3 m matric potential, which means that the arithmetic mean of the validation data yielded more precise predictions than the respective maps. Likewise, the comparison of CoGTF-2 with other PTF-based maps is shown in Table A2. For the water content at higher absolute potential (3.3 and 150 m), the model performance was

similar; however, for the water content close to saturation the model based on predicted soil covariates had higher CCC and R^2 values.
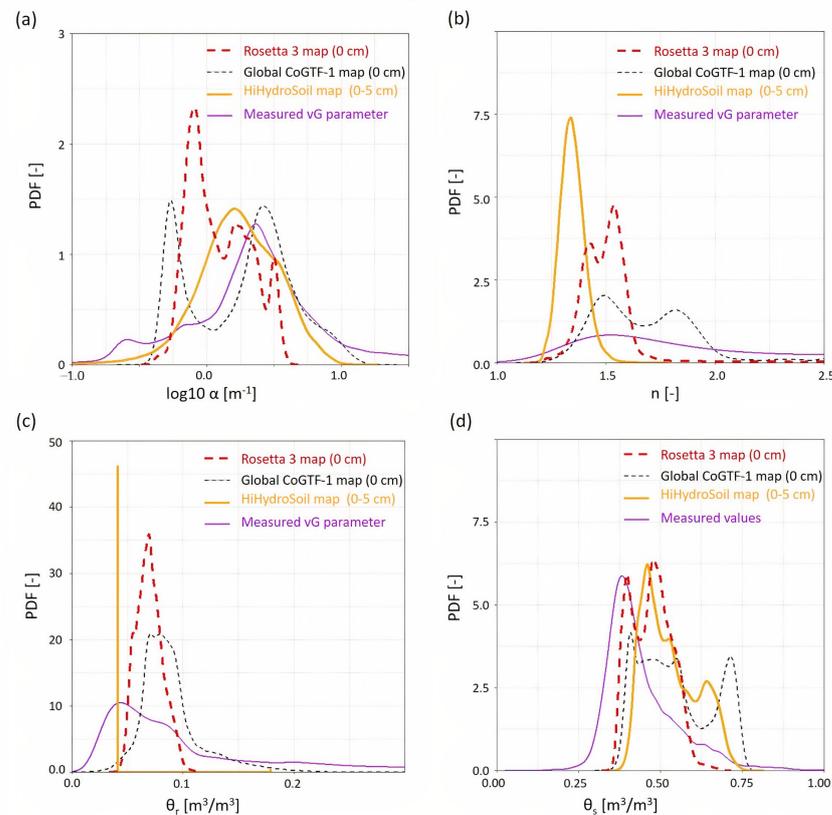

Figure 6. Probability density functions (PDF) of vG parameters for the CoGTF-1 map (black), Rosetta 3 map (red), HiHydroSoil v2.0 map (orange), and of the 'measured' values of 11,705 SWCC from GSHP dataset. The plots (**a**) show inverse air entry pressure parameter $log_{10}\ \alpha$ (unit of α m^{-1}), (**b**) shape parameter n, (**c**) the residual water content θ_r, and (**d**) the saturated water content θ_s. Note that the HiHydroSoil v02.0 map reported only 2 distinct values for θ_r with 0.04 m^3/m^3 if sand content was larger than 2% and 0.17 m^3/m^3 if sand content was below 2% (only very few locations).

Table 1. Concordance correlation coefficient (CCC), coefficient of determination (R^2), root mean square error (RMSE), and BIAS of predicted water content (WC) at 0.1 m, 3.3 m, and 150 m matric potential using CoGTF-1 (based on predicted soil covariates), Rosetta 3, and HiHydroSoil v2.0. The criteria were computed for the validation data (SWCC curves outside from Europe and North America).

Models	Water Content	CCC	R^2	RMSE	BIAS
Rosetta 3	WC_0.1 m	0.172	0.079	0.122	−0.026
	WC_3.3 m	0.086	−1.066	0.129	−0.097
	WC_150 m	0.318	0.191	0.064	−0.019
HiHydroSoil v2.0	WC_0.1 m	0.184	0.099	0.121	−0.004
	WC_3.3 m	−0.044	−2.266	0.162	−0.114
	WC_150 m	0.175	−0.133	0.076	−0.036
CoGTF-1	WC_0.1 m	0.710	0.568	0.084	−0.003
	WC_3.3 m	0.489	0.018	0.088	−0.060
	WC_150 m	0.495	0.336	0.058	−0.017

4. Discussion

4.1. Characteristics of the CoGTF Global vG Parameters Maps

The study presents new global maps of vG parameters at a 1 km resolution based on the most comprehensive available dataset of SWCC measurements and harnessing local terrain, climate, and vegetation information. The effect of environmental covariates on SWCC parameters was most prominent for the inverse air entry pressure α. As shown in Figure 2a, climatic variables were more important for α than soil texture and bulk density. The size and continuity of the larger pores are controlled by soil structure formation and cannot be captured from basic soil properties alone. We can therefore conclude that for this parameter soil structure was more relevant than soil texture. A similar finding was reported for the saturated hydraulic conductivity Ksat in Gupta et al. [10], showing that large Ksat values were found not only in sandy textures but in tropical regions as well. In tropical regions, the wet and warm conditions result in different mineralogy of the clay-sized soil particle fraction [29,30]. We noticed evidence for the role of the soil-forming processes in the other vG parameters as well with higher saturated and residual water content values and lower n values for tropical regions (see maps in Figure 4). For the other three vG parameters, soil properties were the most important covariates. For example, the model of the shape parameter n was dominated by sand content, and this confirmed the findings of Haverkamp et al. [31] who showed that the shape parameter was closely correlated to soil texture. Similarly, bulk density and clay were the most important covariates for θ_s and θ_r, respectively. The same results were presented by Oleszczuk and Truba [32] and Feng and Zhang [33]. Remarkable were the high values of θ_r in tropical regions. The high values of θ_r are likely due to a large proportion of very fine pores in the Ferralsol (oxisols) soils. These pores remain water-filled at potentials as low as -4 MPa [34].

4.2. Prediction Accuracy Improves with Covariates and Larger and Well-Curated SWCC Dataset

The CoGTF-based vG parameter maps differed from PTF-based maps in the visual comparisons. As shown in Figure 6, vG parameter values for the CoGTF framework had a wider distribution than calculated with PTFs because only the CoGTF could take into account the variations of environmental covariates and their effect on soil properties. The high variations in vG parameter values were also related to the large dataset (11,705 for CoGTF-1 compared to 3773 for HiHydroSoil v2.0 and 2134 for Rosetta 3) used for training the model. The training data included samples from all soil textural classes and climatic regions. Moreover, the validation results clearly demonstrated that CoGTF performs better than the PTF based maps (Tables 1 and A2). Therefore, CoGTF overcomes the limitations of the Rosetta 3 PTF where the training data was dominated by loamy and sandy soil texture classes (see Figure 1 in [35]) from temperate climatic regions. Furthermore, Table A3 shows SCV results for a model CoGTF-A that was trained using all 15,259 SWCCs. Results showed that the CoGTF-1 model using only curated data predicts α more accurately than CoGTF-A. For the other vG parameters, the model performances of both models were similar. Likewise, Table A4 shows that the CCC and R^2 of CoGTF-A were slightly better than for CoGTF-1 except for water content at 0.1 m. This might be due to the fact that during testing most of the curves that had no wet-end (water content measured for matric potential $\leq$ 0.2 m) information and could lead to fallacious parameter values were removed.

4.3. Comparison of CoGTF Models Based on Measured and Predicted Soil Covariates

In this study, we compared two CoGTF based models (CoGTF-1 where predicted soil properties were used as covariates for model calibration and computing predictions, and CoGTF-2 where measured properties were used to calibrate the model but predicted soil covariates were used to compute the predictions). The reason for using CoGTF-1 was the lack of measured soil properties at the locations of the global prediction mesh. As stated in Section 2.4, the application of the CoGTF-2 model for global mapping was problematic because different information was used for model calibration and for computing the predictions. As expected, the CoGTF-2 models with measured soil properties (Figure 3e–h)

predicted the vG parameter in the SCV exercise more precisely than the CoGTF-1 models (Figure 3a–d) because measured soil texture and bulk density are more directly related to SWCCs than the predicted soil covariates. However, these SCV results of CoGTF-2 did not allow us to assess how well CoGTF-2 predicted the vG parameters for the global prediction mesh, where measured soil property data was unavailable. Therefore, to test the loss of accuracy when applying a model based on measured soil covariates on sites with predicted soil information, we changed the SCV scheme as follows: we calibrated the CoGTF-2 model still using measured soil properties as covariates but predicted the vG parameter of the SCV subsets using predicted soil properties as covariates. Figure A6 showed reduced model performance compared to the SCV results shown in Figure 3e–h. The performance was also slightly reduced compared to the model calibrated with predicted soil covariates (Figure 3a–d). Due to the reduced model performance and the inherent inconsistency of mixing measured and predicted soil covariate data we only showed global maps created with CoGTF-1.

4.4. Use of the Global CoGTF vG Parameters Maps and Future Developments

The new maps of vG parameters presented in this study can be used in Earth system and land surface models that need information on spatially distributed soil hydraulic parameters. These maps can also be used more directly to compute other important soil physical properties like ponding time, characteristic length of evaporation, sorptivity, and soil strength. In addition, the plant's available water is typically deduced from water contents at specific matric potential values (3.3 m for field capacity and 150 m for wilting point) and can be computed globally with the new maps. The CoGTF maps presented here have a spatial resolution of 1 km. This resolution can likely be improved in the near future, considering various initiatives to estimate soil and environmental covariate information with higher spatial resolution. In addition, the maps can be improved if more comprehensive SWCCs data from Canada and Western and Northern Australia become available.

5. Conclusions

Machine learning algorithms have been used to produce global maps of soil hydraulic properties based on well-curated spatially distributed measurements and environmental covariates. In this study, we have demonstrated the benefit of (i) improving coverage of SWCC and the quality of the estimated parameters, (ii) better coverage of information from all climatic regions for ML training, and (iii) including local covariates such as climate, terrain, and vegetation metrics in the prediction. We have shown quantitatively (based on CCC, R^2, RMSE, and BIAS values) that our new parameter maps perform better in predicting soil water retention compared to previous maps based on PTFs. We highlight that despite the possibility of calibrating a model with measured covariate values (e.g., soil texture), predicted values should be the default for the calibration step if measurements are not available at a global scale (i.e., for each pixel). One of the primary contributors to the improved predictions is the step increase in the number of SWCC data sets and the rigor by which the parameters have been derived. We conclude that including information based on local covariates implicitly injects the effects of soil formation processes on soil hydraulic properties (an easy to implement alternative to the explicit soil structure modeling reported in Fatichi et al. [6] or Bonetti et al. [7]).

Author Contributions: S.G. processed the data and created the vG maps. S.G. performed the data analyses. S.G. wrote the first draft of the manuscript. P.L., S.B., A.P., T.H. and D.O. edited the manuscript. S.G. revised the manuscript and S.B., P.L., A.P. and D.O. made significant contributions to the data interpretation. P.L. and D.O. supervised the study. All authors have read and agreed to the published version of the manuscript.
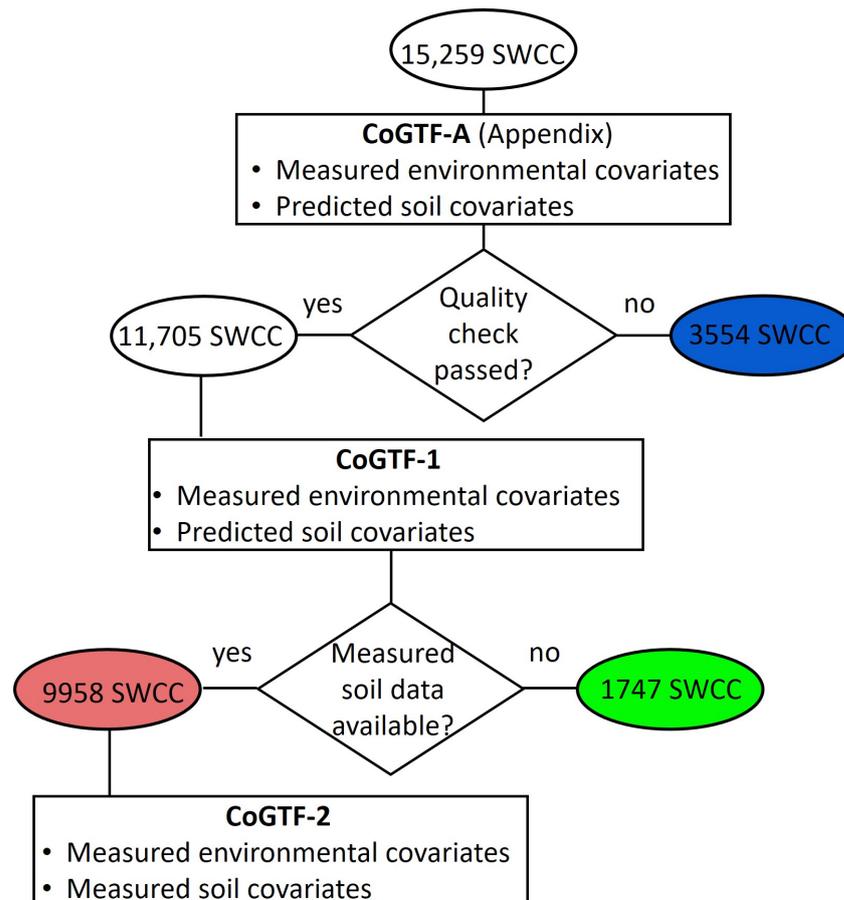
Appendix A

Figure A1. Conceptual workflow used to generate the different CoGTF models using the SWCCs shown in Figure 1. The CoGTF-A model was generated using all 15,259 SWCCs whereas to generate the CoGTF-1 model only 11,705 SWCCs were used with predicted soil properties. Furthermore, the CoGTF-2 model was produced using 9958 SWCCs that had measured soil properties.
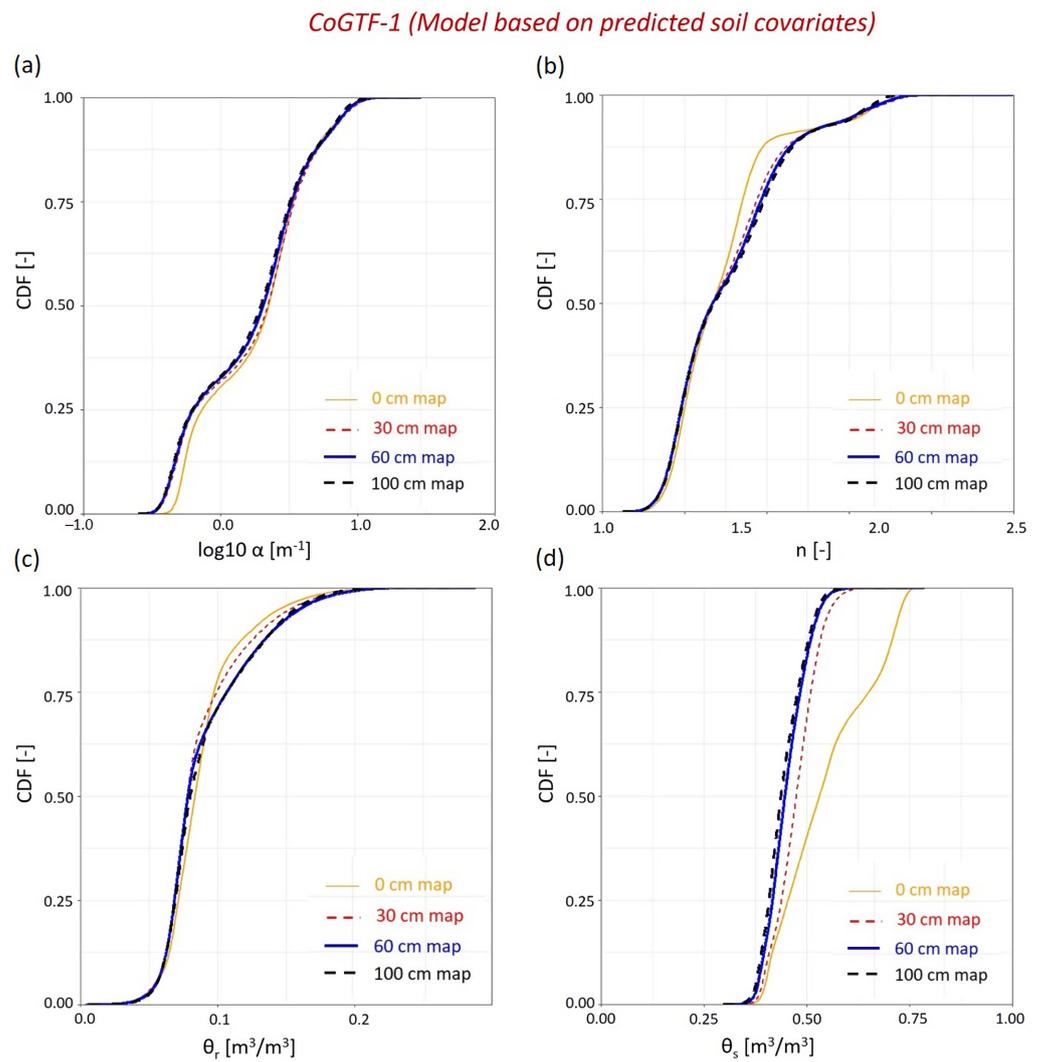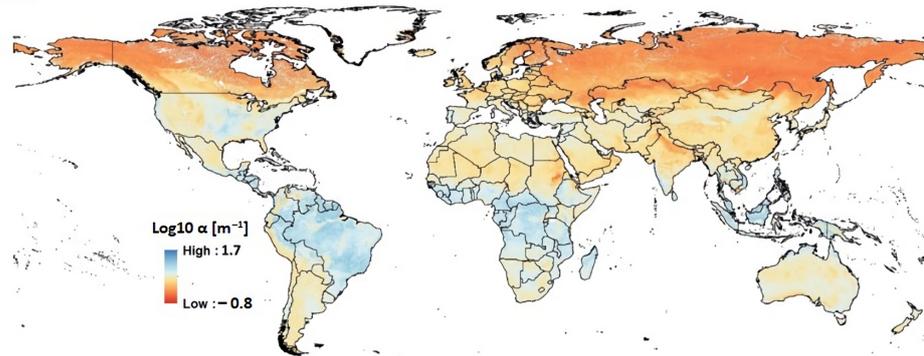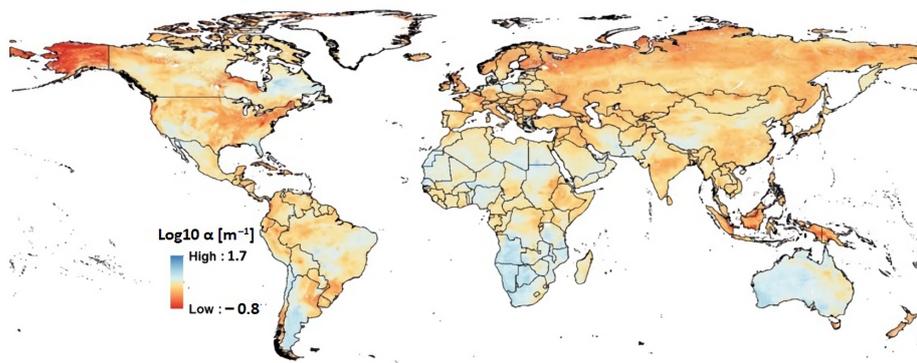
CoGTF-1 (Model based on predicted soil covariates)

Figure A2. Cumulative distribution function (CDF) for global maps of vG parameters at different depths (0, 30, 60, and 100 cm) predicted by CoGTF-1. (**a**) shows the inverse air entry pressure parameter $\log_{10} \alpha$ (unit of α m^{-1}), (**b**) shape parameter n, (**c**) the residual water content θ_r, and (**d**) the saturated water content θ_s.

(a) CoGTF-1 (0 cm)

(b) HiHydroSoil v2.0 map (0-5 cm)

(c) Rosetta 3 (0 cm)

Figure A3. Visual comparison between (**a**) CoGTF-1 (model based on predicted soil covariates), (**b**) HiHydroSoil v2.0, and (**c**) Rosetta 3 map of shape parameter $\log_{10} \alpha$ (unit of α m^{-1}).

(a) CoGTF-1 (0 cm)

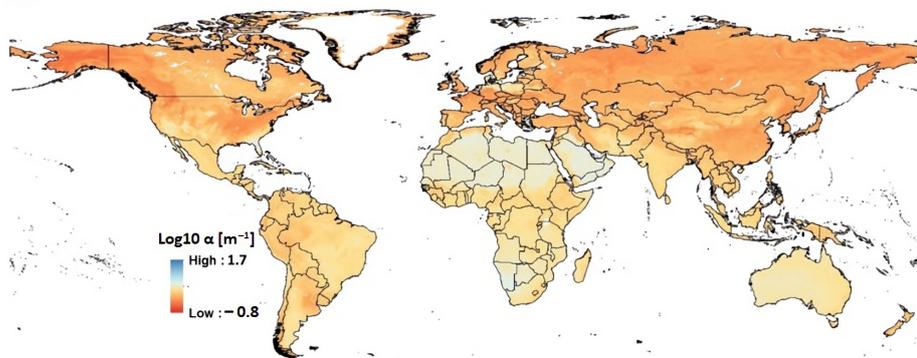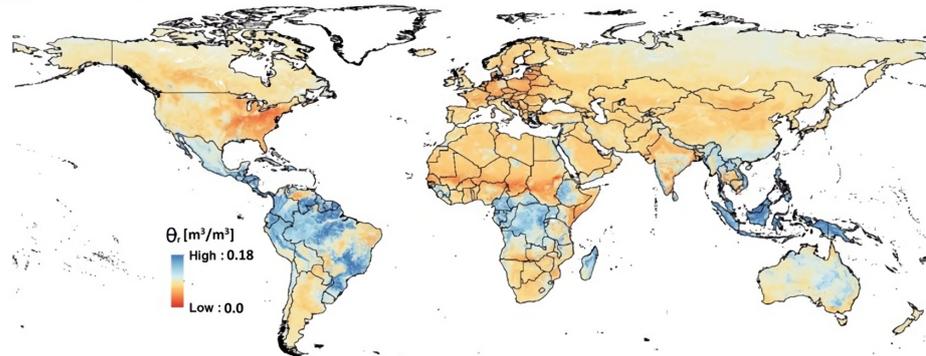(b) HiHydroSoil v2.0 map (0-5 cm)

(c) Rosetta 3 (0 cm)

Figure A4. Visual comparison between (**a**) CoGTF-1 (model based on predicted soil covariates), (**b**) HiHydroSoil v2.0, and (**c**) Rosetta 3 map of residual water content θ_r.

(a) CoGTF-1 (0 cm)

(b) HiHydroSoil v2.0 map (0-5 cm)

(c) Rosetta 3 (0 cm)

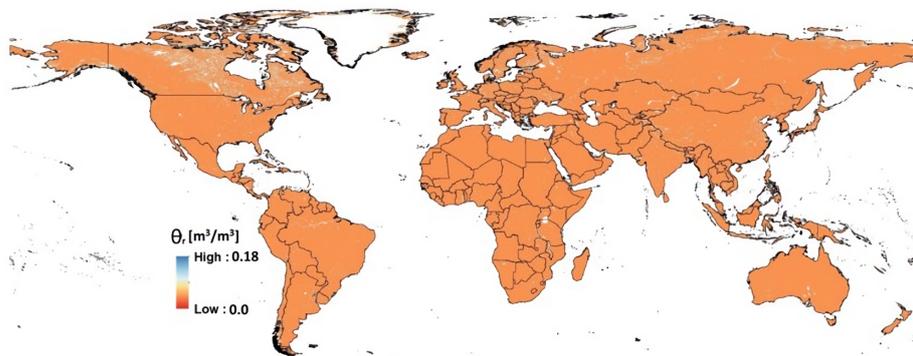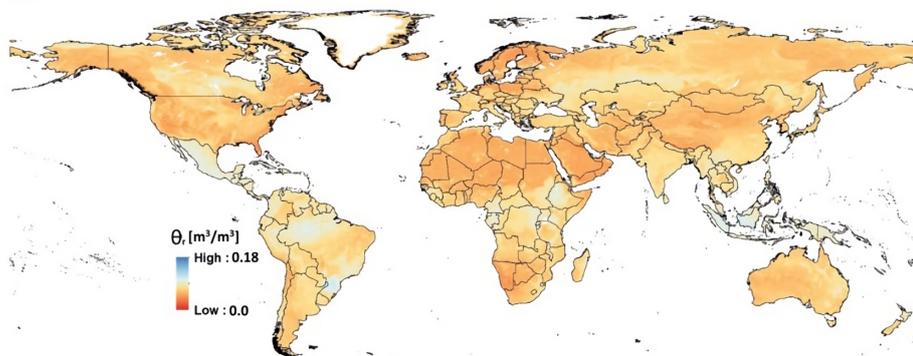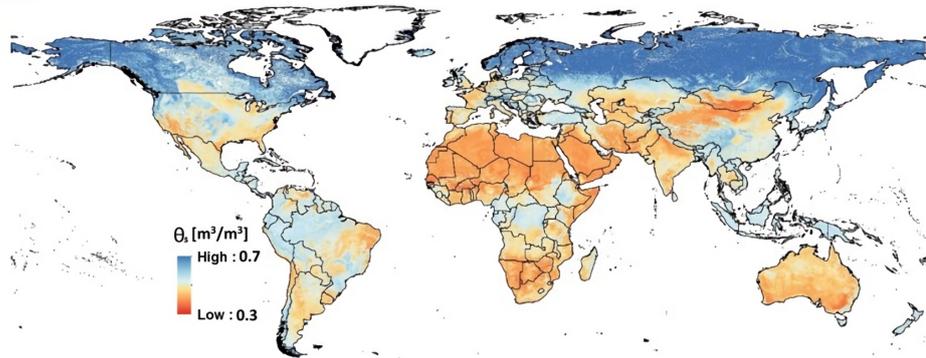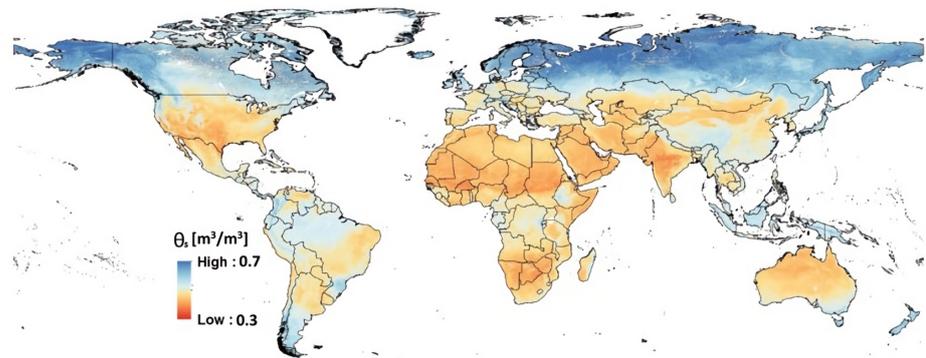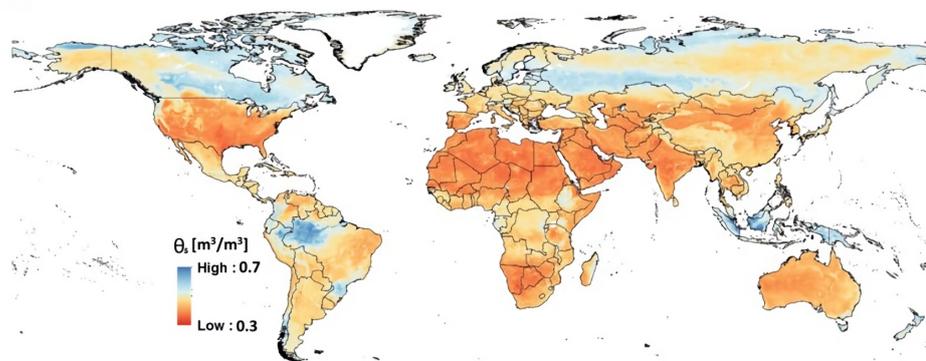**Figure A5.** Visual comparison between (**a**) CoGTF-1 (model based on predicted soil covariates), (**b**) HiHydroSoil v2.0, and (**c**) Rosetta 3 θ_s map of saturated water content θ_s.
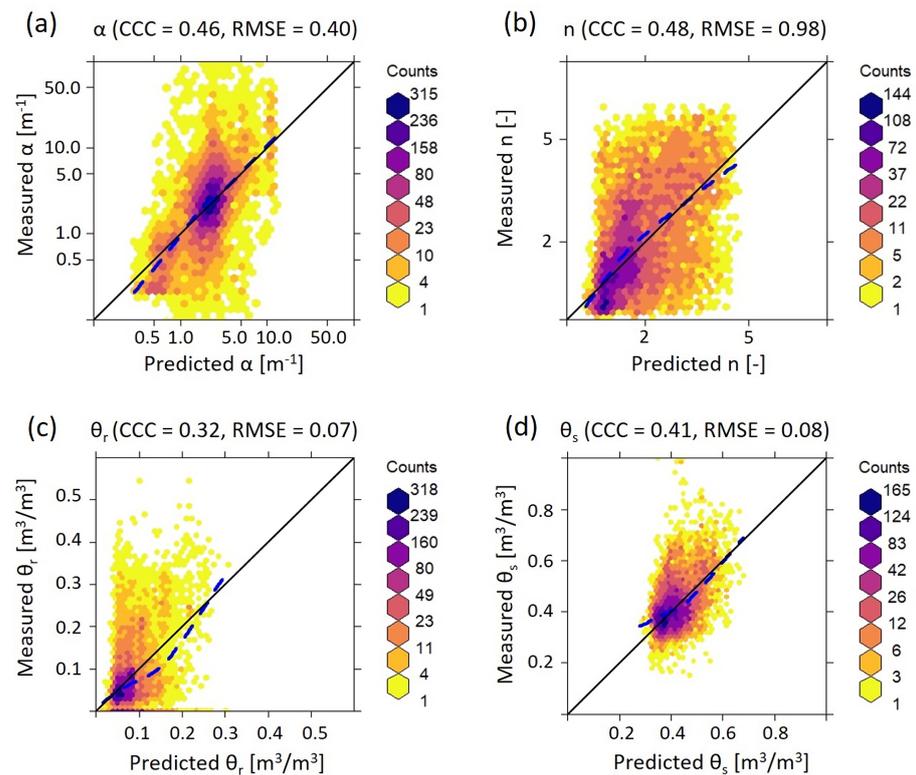
Figure A6. Results of SCV of SWCC parameter predictions to show the effect of mixing soil information for model calibration (using measured soil properties) and computing predictions (using predicted soil properties). (**a**) inverse air entry pressure parameter $log_{10}\ \alpha$ (unit of α m^{-1}), (**b**) shape parameter n, (**c**) the residual water content θ_r, and (**d**) the saturated water content θ_s. The term 'measured' on the axes of the figures relates to parameter estimates obtained by fitting the vG model to measured SWCCs.

Table A1. List of covariates used for creating the vG maps.

S.No	List of Covariates	Source
	Climate (clm)	
1	clm_annual mean temperaturebio1_m_1 km_s0..0cm_1979-2013_v1.0	http://chelsa-climate.org/bioclim/, (accessed on 20 October 2021)
2	clm_temperature seasonalitybio4_ m_1 km_s0..0cm_1979-2013_v1.0	Karger et al. [36]
3	clm_max temperature of warmest monthbio5_m_1 km_s0..0cm_1979-2013_v1.0	
4	clm_min temperature of coldest monthbio6_m_1 km_ s0..0cm_1979-2013_v1.0	
5	clm_annual precipitationbio12_m_1 km_1979_2013_v1.0	
6	clm_precipitation of wettest monthbio13_m_1 km_1979_2013	
7	clm_precipitation of driest monthbio14_m_1 km_1979_2013	
8	clm_diffuse.irradiation_solar.atlas.kwhm2.100_m_1 km_s0..0cm_2016_v1	https://globalsolaratlas.info/download/world, (accessed on 20 October 2021)
9	clm_direct.irradiation_solar.atlas.kwhm2.10_m_1 km_s0..0cm_2016_v1	
10	clm_land surface temperature_mod11a2.annual.day_m_1 km_s0..0cm_2000..2017_v1.0	https://lpdaac.usgs.gov/products/mod11a2v006/, (accessed on 20 October 2021)
11	clm_land surface temperature_mod11a2.annual.day_sd_1 km_s0..0cm_2000..2017_v1.0	

Table A1. *Cont.*

S.No	List of Covariates	Source
	Digital terrain model (dtm)	
12	dtm_topographic wetness index_merit.dem_m_1 km_s0..0cm_2017_v1.0	
13	dtm_slope_merit.dem_m_1 km_s0..0cm_2017_v1.0	
14	dtm_aspect.cosine_merit.dem_m_1 km_s0..0cm_2018_v1.0	https://zenodo.org/record/1447210#.XllTejFKhaQ, (accessed on 25 October 2021) Yamazaki et al. [37]
15	dtm_elevation_merit.dem_m_1 km_s0..0cm_2017_v1.0	
16	dtm_lithology_usgs.ecotapestry.acid.plutonics_p_1 km_s0..0cm_2014_v1.0	
	Surface reflectance (lcv)	
17	lcv_landsat.near infrared_wri.forestwatch_m_1 km_s0..0cm_2018_v1.2	
18	lcv_landsat.red_wri.forestwatch_m_1 km_s0..0cm_2018_v1.2	Hansen et al. [38]
19	lcv_landsat.short wave infrared_wri.forestwatch_m_1 km_s0..0 cm_2018_v1.2	
20	lcv_wetlands.regularly.flooded_upmc.wtd_p_1 km_b0..200cm_2010..2015_v1.0	https://doi.pangaea.de/10.1594/PANGAEA.892657 Tootchi et al. [39]
	Vegetation covariates (veg)	
21	veg_fraction of bbsorbed photosynthetically active radiation _proba.v.annnual_d_1km_s0..0cm_2014..2019_v1.0	https://land.copernicus.eu/global/products/fapar, (accessed on 25 October 2021)
	Soil properties (sol)	
22	sol_clay.wfraction_usda.3a1a1a_m_1 km_b0_10_30_60_100_200cm_ 1950..2017_v0.2	
23	sol_sand.wfraction_usda.3a1a1a_m_1 km_b0_10_30_60_100_200cm _1950..2017_v0.2	https://www.openlandmap.org/, (accessed on 22 December 2021)
24	sol_bulk_density.wfraction_usda.3a1a1a_m_1 km_b0_10_30_60_100_200cm_1950..2017_v0.2	
25	DEPTH	

Table A2. Concordance correlation coefficient (CCC), coefficient of determination (R^2), root mean square error (RMSE), and BIAS of predicted water content (WC) at 0.1 m, 3.3 m, and 150 m matric potential using CoGTF-2 based on measured soil covariates, Rosetta 3, and HiHydroSoil v2.0. These values are computed for the validation data.

Models	Water Content	CCC	R^2	RMSE	BIAS
Rosetta 3	WC_0.1 m	0.112	−0.122	0.122	−0.047
	WC_3.3 m	0.093	−1.112	0.112	−0.085
	WC_150 m	0.373	0.275	0.052	−0.008
HiHydroSoil v2.0	WC_0.1 m	0.218	0.058	0.112	−0.020
	WC_3.3 m	−0.027	−2.143	0.137	−0.088
	WC_150 m	0.233	−0.168	0.066	−0.006
CoGTF-2	WC_0.1 m	0.622	0.454	0.085	−0.019
	WC_3.3 m	0.586	0.180	0.070	−0.048
	WC_150 m	0.505	0.335	0.050	−0.022

Table A3. Comparison of the SCV predictions of the models trained with all 15,259 SWCCs (CoGTF-A with predicted soil covariates) or with the curated dataset consisting of 11,705 SWCCs (CoGTF-1 with predicted soil covariates). The curated dataset predictions are the same SCV predictions as in Figure 3a–d.

All and Curated Data	CCC	R^2	RMSE	BIAS
α_all	0.393	0.263	0.588	−0.042
α_cur	0.432	0.273	0.442	0.009
n_all	0.522	0.325	0.951	0.088
n_cur	0.525	0.328	0.888	0.050

Table A3. *Cont.*

All and Curated Data	CCC	R^2	RMSE	BIAS
θ_r_all	0.315	0.152	0.069	0.0003
θ_r_cur	0.325	0.172	0.076	−0.001
θ_s_all	0.552	0.393	0.085	−0.001
θ_s_cur	0.565	0.412	0.082	0.001

Table A4. Number of SWCCs, concordance correlation coefficient (CCC), coefficient of determination (R^2), root mean square error (RMSE), and BIAS of predicted water content (WC) at 0.1 m, 3.3 m, and 150 m matric potential using CoGTF curated (CoGTF-1) and all SWCCs (CoGTF-A). Note that a similar method (see Section 2.6) is used to estimate the accuracy parameters when all SWCCs are used.

Models	SWCCs	Water Content	CCC	R^2	RMSE	BIAS
CoGTF-1	1572	WC_0.1 m	0.710	0.568	0.084	−0.003
	719	WC_3.3 m	0.489	0.018	0.088	−0.060
	1184	WC_150 m	0.495	0.336	0.058	0.017
CoGTF-A	1671	WC_0.1 m	0.699	0.551	0.088	0.0003
	836	WC_3.3 m	0.521	0.064	0.099	−0.066
	1371	WC_150 m	0.529	0.359	0.062	−0.020

References

1. Van Genuchten, M.T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **1980**, *44*, 892–898. [CrossRef]
2. Mualem, Y. *Catalogue of the Hydraulic Properties of uNsaturated Soils*; Technion Israel Institute of Technology, Technion Research & Development: Haifa, Israel, 1976.
3. Brooks, R. Hydraulic properties of porous. *Media Hydrol. Pap.* **1964**, *3*, 1–27.
4. Bouma, J. Using soil survey data for quantitative land evaluation. In *Advances in Soil Science*; Springer: Berlin/Heidelberg, Germany, 1989; pp. 177–213.
5. Van Looy, K.; Bouma, J.; Herbst, M.; Koestel, J.; Minasny, B.; Mishra, U.; Montzka, C.; Nemes, A.; Pachepsky, Y.A.; Padarian, J.; et al. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. *Rev. Geophys.* **2017**, *55*, 1199–1256. [CrossRef]
6. Fatichi, S.; Or, D.; Walko, R.; Vereecken, H.; Young, M.H.; Ghezzehei, T.A.; Hengl, T.; Kollet, S.; Agam, N.; Avissar, R. Soil structure is an important omission in Earth System Models. *Nat. Commun.* **2020**, *11*, 522. [CrossRef] [PubMed]
7. Bonetti, S.; Wei, Z.; Or, D. A framework for quantifying hydrologic effects of soil structure across scales. *Commun. Earth Environ.* **2021**, *2*, 107. [CrossRef]
8. Or, D. The tyranny of small scales–on representing soil processes in global land surface models. *Water Resour. Res.* **2020**, *56*. [CrossRef]
9. Khlosi, M.; Alhamdoosh, M.; Douaik, A.; Gabriels, D.; Cornelis, W. Enhanced pedotransfer functions with support vector machines to predict water retention of calcareous soil. *Eur. J. Soil Sci.* **2016**, *67*, 276–284. [CrossRef]
10. Gupta, S.; Lehmann, P.; Bonetti, S.; Papritz, A.; Or, D. Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework. *J. Adv. Model. Earth Syst.* **2021**, *13*, e2020MS002242. [CrossRef]
11. Chaney, N.W.; Wood, E.F.; McBratney, A.B.; Hempel, J.W.; Nauman, T.W.; Brungard, C.W.; Odgers, N.P. POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* **2016**, *274*, 54–67. [CrossRef]
12. Gupta, S.; Papritz, A.; Lehmann, P.; Hengl, T.; Bonetti, S.; Or, D. GSHP: Global database of soil hydraulic properties. *Sci. Data under review*. [CrossRef]
13. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2013.
14. Zhang, Y.; Schaap, M.G. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *J. Hydrol.* **2017**, *547*, 39–53. [CrossRef]
15. Simons, G.; Koster, R.; Droogers, P. HiHydroSoil v2.0-High Resolution Soil Maps of Global Hydraulic Properties. 2020. Available online: https://www.futurewater.nl/wp-content/uploads/2020/10/HiHydroSoil-v2.0-High-Resolution-Soil-Maps-of-Global-Hydraulic-Properties.pdf (accessed on 28 November 2021).
16. Tóth, B.; Weynants, M.; Nemes, A.; Makó, A.; Bilas, G.; Tóth, G. New generation of hydraulic pedotransfer functions for Europe. *Eur. J. Soil Sci.* **2015**, *66*, 226–238. [CrossRef] [PubMed]

17. Papritz, A. *Soilhypfit: Modelling of Soil Water Retention and Hydraulic Conductivity Data*; R Package Version 0.1-5; R Foundation for Statistical Computing: Vienna, Austria, 2022.
18. Uusipaikka, E. *Confidence Intervals in Generalized Regression Models*; Chapman & Hall: Chapel Hill, NC, USA; CRC Press: Boca Raton, FL, USA, 2008. [CrossRef]
19. Hengl, T.; Collins, T.; Wheeler, I.; MacMillan, R. Everybody Has a Right to Know What's Happening with the Planet: Towards a Global Commons. 2019. Available online: https://zenodo.org/record/3274294#.Yl0aJ9NByUk (accessed on 28 November 2021).
20. Jenny, H. *Factors of Soil Formation: A System of Quantitative Pedology*; Dover Publications, Inc.: New York, NY, USA, 1994.
21. Wright, M.N.; Ziegler, A. ranger: A fast implementation of random forests for high dimensional data in C++ and R. *arXiv* **2015**, arXiv:1508.04409.
22. Hastie, T.; Tibshirani, R.; Friedman, J. *The Elements of Statistical Learning; Data Mining, Inference and Prediction*, 2nd ed.; Springer: New York, NY, USA, 2009.
23. Carr, D.; Nicholas, L.-K.; Maechler, M.; Deepayan, S. *Hexbin: Hexagonal Binning Routines*; R Package Version 1.28.1; R Foundation for Statistical Computing: Vienna, Austria, 2020.
24. Sarkar, D. *Lattice: Multivariate Data Visualization with R*; Springer: New York, NY, USA, 2008; ISBN 978-0-387-75968-5.
25. Garnier, S. *Viridis: Default Color Maps from 'Matplotlib'*; R Package Version 0.5.1; R Foundation for Statistical Computing: Vienna, Austria, 2018.
26. Lovelace, R.; Nowosad, J.; Muenchov, J. *Geocomputation with R*; The R Series; CRC Press: Boca Raton, FL, USA, 2019. [CrossRef]
27. Cleveland, W.S. LOWESS: A program for smoothing scatterplots by robust locally weighted regression. *Am. Stat.* **1981**, *35*, 54. [CrossRef]
28. Lawrence, I.; Lin, K. A concordance correlation coefficient to evaluate reproducibility. *Biometrics* **1989**, *45*, 255–268.
29. Hodnett, M.; Tomasella, J. Marked differences between van Genuchten soil water-retention parameters for temperate and tropical soils: A new water-retention pedo-transfer functions developed for tropical soils. *Geoderma* **2002**, *108*, 155–180. [CrossRef]
30. Lehmann, P.; Leshchinsky, B.; Gupta, S.; Mirus, B.B.; Bickel, S.; Lu, N.; Or, D. Clays are not created equal: How clay mineral type affects soil parameterization. *Geophys. Res. Lett.* **2021**, *48*, e2021GL095311. [CrossRef]
31. Haverkamp, R.; Leij, F.J.; Fuentes, C.; Sciortino, A.; Ross, P. Soil water retention: I. Introduction of a shape index. *Soil Sci. Soc. Am. J.* **2005**, *69*, 1881–1890. [CrossRef]
32. Oleszczuk, R.; Truba, M. The analysis of some physical properties of drained peat-moorsh soil Layers. *Ann. Wars. Univ. Life Sci.-SGGW Land Reclam.* **2013**, *45*, 41–48. [CrossRef]
33. Feng, J.; Zhang, G. Influence of clay fraction and mineral Composition on unsaturated characteristics of cohesive foundation soil in airports. *Geofluids* **2021**, *2021*, 1–10. [CrossRef]
34. Chauvel, A.; Grimaldi, M.; Tessier, D. Changes in soil pore-space distribution following deforestation and revegetation: An example from the Central Amazon Basin, Brazil. *For. Ecol. Manag.* **1991**, *38*, 259–271. [CrossRef]
35. Schaap, M.G.; Leij, F.J.; Van Genuchten, M.T. Rosetta: A computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *J. Hydrol.* **2001**, *251*, 163–176. [CrossRef]
36. Karger, D.N.; Conrad, O.; Böhner, J.; Kawohl, T.; Kreft, H.; Soria-Auza, R.W.; Zimmermann, N.E.; Linder, H.P.; Kessler, M. Climatologies at high resolution for the earth's land surface areas. *Sci. Data* **2017**, *4*, 170122. [CrossRef] [PubMed]
37. Yamazaki, D.; Ikeshima, D.; Tawatari, R.; Yamaguchi, T.; O'Loughlin, F.; Neal, J.C.; Sampson, C.C.; Kanae, S.; Bates, P.D. A high-accuracy map of global terrain elevations. *Geophys. Res. Lett.* **2017**, *44*, 5844–5853. [CrossRef]
38. Hansen, M.C.; Potapov, P.V.; Moore, R.; Hancher, M.; Turubanova, S.A.; Tyukavina, A.; Thau, D.; Stehman, S.V.; Goetz, S.J.; Loveland, T.R.; et al. High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **2013**, *342*, 850–853. [CrossRef]
39. Tootchi, A.; Jost, A.; Ducharne, A. Multi-source global wetland maps combining surface water imagery and groundwater constraints. *Earth Syst. Sci. Data* **2019**, *11*, 189–220. [CrossRef]