scientific **data**

OPEN

DATA DESCRIPTOR

Global Soil Hydraulic Properties dataset based on legacy site observations and robust parameterization

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

The representation of land surface processes in hydrological and climatic models critically depends on the soil water characteristics curve (SWCC) that defines the plant availability and water storage in the vadose zone. Despite the availability of SWCC datasets in the literature, significant efforts are required to harmonize reported data before SWCC parameters can be determined and implemented in modeling applications. In this work, a total of 15,259 SWCCs from 2,702 sites were assembled from published literature, harmonized, and quality-checked. The assembled SWCC data provide a global soil hydraulic properties (GSHP) database. Parameters of the van Genuchten (vG) SWCC model were estimated from the data using the R package 'soilhypfit'. In many cases, information on the wet- or dry-end of the SWCC measurements were missing, and we used pedotransfer functions (PTFs) to estimate saturated and residual water contents. The new database quantifies the differences of SWCCs across climatic regions and can be used to create global maps of soil hydraulic properties.

Background & Summary

The soil water characteristics curves (SWCCs) describes the relationship between soil water content (in gravimetric or volumetric form) and matric potential[1,2] and is a fundamental soil hydraulic property that characterizes soil water dynamics and key hydrological processes[3]. The amount of water retained in the soil pores at given matric potential is dominantly controlled by soil texture, structure, and organic matter[4]. The measurement of the SWCCs in the laboratory or field is laborious and time-consuming[5]. This limitation hinders empirical characterization of SWCC parameters over large areas as required in land surface and catchment-scale modeling. An alternative to excessive sampling to obtain spatially distributed SWCC parameters is the application of pedotransfer functions (PTFs)[6]. However, agronomic legacy of soil mapping influences PTF-derived SWCCs, which tend to be region-specific and focus on homogeneous agricultural soil[7–9]. The growing demand for hydrological land surface parameterization beyond agricultural lands and the related refinement of spatial representation require more definitive SWCC information[10]. Motivated by the growing need for more comprehensive and spatially resolved SWCCs, a key objective of this study was to pool published datasets and supplement these with anecdotal literature values for regions with poor coverage. A cursory inspection of major datasets such as WOSIS[11], UNSODA[12] and measurements by Holtan[13] shows critical deficiencies ranging from lack of spatial referencing to partial information with datasets omitting the wet-end (water content measured for matric potential ≤ 0.2 m) and dry-end (water content measured at matric potential ≥ 150 m) that is often defined as wilting point of SWCCs. For example, WOSIS[11] and Holtan[13] datasets provide SWCCs starting at 0.6 and 1.0 m matric potential, respectively, and missing the wet-end measurements. In contrast, the ZALF dataset[14] provides wet-end information but dry-end information is missing. In addition, Batjes et al.[11] and Holtan[13] used bulk density measured at 3.3 m matric potential (defined as density at field capacity) to convert gravimetric to volumetric water content, ignoring the water mass retained in the pore space at field capacity. Therefore, a critical aspect of

[1]Soil and Terrestrial Environmental Physics, Department of Environmental Systems Science, ETH Zürich, Zürich, Switzerland. [2]OpenGeoHub Foundation, Wageningen, the Netherlands. [3]EnvirometriX, Wageningen, the Netherlands. [4]Soil Physics and Land Management Group, Wageningen University, Wageningen, The Netherlands. [5]Division of Hydrologic Sciences, Desert Research Institute, Reno, NV, USA. ✉e-mail: surya.gupta@usys.ethz.ch

compiling a global SWCCs dataset for hydro-climatic applications is to harmonize the data and devise strategies for imputing missing information as we further elaborate in the section 'Materials and methods'.

The objectives of this study are to address the limitations of currently available datasets (such as lack of wet- and dry-end measurements of SWCCs, poor estimates of dry bulk density, etc.) and to provide a systematic approach to infer more reliable and geo-referenced SWCC parameters. We put an effort to collect a global soil hydraulic properties (GSHP) database with 15,259 SWCCs by importing, quality controlling, and standardizing tabular data from existing datasets and scientific literature and estimating van Genuchten (vG) model parameters from the measured SWCCs. For the quality control, firstly, we excluded SWCCs using likelihood-based confidence intervals, yielding 11,705 as 'good quality estimate'. Along with SWCCs, GSHP also provides information regarding soil texture, bulk density, organic carbon, and porosity. The database also contains 8,675 data of Ksat that allow the quantification of the unsaturated hydraulic conductivity function. The GSHP database covers most countries, climatic regions, and continents, including tropical regions with more intense soil-forming processes and inactive clay minerals. The data and codes are made publicly available to promote data-driven analysis and to collect additional data to increase the accuracy of global modeling of unsaturated soil properties in land surface and Earth system models.

Methods

Data sources. The GSHP database was assembled in two steps. Firstly, we combined well-known datasets such as UNSODA[12], HYBRAS[15], WOSIS[11], and AFSPDB[16]. Secondly, we used different search engines, including Science Direct (https://www.sciencedirect.com/), Google Scholar (https://scholar.google.com/) and Scopus (https://www.scopus.com) to collect additional data. We searched for SWCC datasets using "soil water retention curves", "moisture retention curves", and "soil water characteristics curves" as keywords. Moreover, the country names were also used with these keywords to conduct more specific searches for areas with few or no data points (this was the case, for example, for Cambodia, Vietnam, and Thailand). The sources of the collected datasets are listed in Table 1 together with the number of SWCCs for each dataset.

The next task was to check the availability of spatial coordinates for the measurements in each dataset. We assigned each SWCC to one of eight 'accuracy classes' ranging from highest (0–100 m) to lowest (more than 10,000 m or non-available information (NA)) accuracy. For example, Forrest et al.[17], and Ottoni et al.[15] provided exact coordinates of the locations, thus we assigned a location accuracy of 0–100 m (i.e., highly accurate; see Table S1 for more details). For other references, we digitized provided maps or sketches with locations of the measurements. We first georeferenced these maps using ESRI ArcGIS software (v10.3) and then digitized the coordinates from georeferenced images. Some of the documents we digitized (e.g. Nemes et al.[12]) provided the names of specific locations, and hence we used Google Earth to obtain the coordinates. Coordinates of SWCCs for which only a location name was provided, were estimated using Google Earth, while SWCCs with neither coordinates nor location name were discarded. A detailed description of the extraction of coordinates is provided in Gupta et al.[18]. Note that Batjes et al.[11], and Leenaars et al.[16] used another definition of accuracy classes in their collection of datasets. Therefore, we merged the various accuracy definitions into broader classes as shown in Table S1. For example, samples that have minimum and maximum location accuracy less than 100 m (likely 0–10 m, 0–40 m, etc) are assigned to the 0–100 m class. Likewise, 0–1000 m values are assigned to the 500–1000 m class. Furthermore, datasets were cross-checked to avoid redundancy. For example, the WOSIS dataset[11] includes the AFSPDB dataset[16], so we removed the redundant data from WOSIS and used the original dataset (AFSPDB) in the final compilation. The Florida dataset[19] was directly obtained from a project website.

Database cleaning and harmonization. The harmonization of the collected datasets was done by first converting all data to the same units. The units are described in Table 2. Note that we used potential heads (unit of length) and not pressure values for the characterization of the water potential. Furthermore, the data was cleaned as follows: a) SWCCs with maximum volumetric water content more than 1.0 m³/m³ were removed, b) SWCCs that had less than four data pairs were removed, c) SWCCs in which the water content increased more than 10% for increasing (absolute) matric potential were removed (see Figure S1), and d) SWCCs without wet-end information that did not have bulk density data were removed because it was impossible to impute the saturated water content without such information.

After data extraction from literature, geo-referencing, and harmonization, all information was collected in tabular form in the new GSHP database. The database consists of 54 variables (see Table 2) and 136,989 records with water content measurements (and complementary information) at given matric potential recorded for 15,259 SWCCs. A list of the variables of the database along with their units is given in Table 2.

Conversion of gravimetric to volumetric water content. Some datasets such as Holten[13] and Batjes et al.[11] provided the gravimetric water content, and the dry bulk density (ratio of solid mass to total volume), before potential shrinkage was required to convert to volumetric water content. However, for these samples, the dry bulk density of a soil clod had been determined by measuring the volume (and mass) of a dry 'clod' of soil and not by measuring the soil sample volume at wet state (for shrinking soils, the dry bulk density will be overestimated by measuring the sample volume at dry state). For these datasets, the 'wet' bulk density at 3.0 or 3.3 m matric potential had been measured as well, and this value had been used in the original analysis to convert gravimetric to volumetric water content. However, this 'wet' bulk density (including water mass) is higher than the dry bulk density. Here we followed a different strategy for conversion to volumetric water content and assumed that the volume of the 'clod' of soil at 3.3 m matric potential is the same as at saturation. This is a simplification because soils with structural pores can shrink by application of −3.3 m matric potential as observed by Assi et al.[20] and Bonvin et al.[21]. However, the reported bulk density changes between 0.1 and 5.2% (average 3%) were small compared to the error when using bulk density measured at 3.3 m matric potential to convert gravimetric to volumetric water

Reference	N	Reference	N	Reference	N
Al-Darby and El-Shafei[42]	1	Li et al.[43]	3	Al Majou et al.[44]	10
Alghamdi et al.[45]	1	Abid and Lal[46].	4	Asghari et al.[47]	12
Are et al.[48]	1	Bescansa et al.[49]	4	MacVicar et al.[50]	12
Babaeian et al.[51]	1	Dlapa et al.[52]	4	Noguchi et al.[53]	12
Bhushan and Sharma[54]	1	Hoshino et al.[55]	4	Tobón et al.[56]	12
de Oliveira et al.[57]	1	Kumar et al.[58]	4	Tyagi et al.[59]	14
Garba et al.[60]	1	McBeath et al.[61]	4	AL-Kayssi[62]	15
Glab et al.[63]	1	Mondal et al.[64]	4	Karup et al.[65]	16
Kakeh et al.[66]	1	Ng et al.[67]	4	Simmons[68]	16
Lowe et al.[69]	1	Smettem and Gregory[70]	4	Wang et al.[71]	16
Medina et al.[72]	1	Xia et al.[73]	4	Cooper et al.[74]	18
Nyamangara et al.[75]	1	Chari and Vahidi[76]	5	Quang and Jansson[77]	20
Sulaeman et al.[78]	1	Eden et al.[79]	5	Pan et al.[80]	22
Thakur et al.[81]	1	Moazeni-Noghondar et al.[82]	5	Bambra[83]	23
Wickland et al.[84]	1	Nano et al.[85]	5	Marui et al.[86]	25
Zebarth et al.[87]	1	Toriyama et al.[88]	5	Vereecken and Van Looy[89]	145
Zhang et al.[90]	1	Xing et al.[91]	5	Jauhiainen et al.[92]	108
El-Asswad et al.[93]	2	Jha et al.[94]	6	Richard and Lüscher[36]	111
Ismail[95]	2	Konyai et al.[96]	6	Forrest et al.[17]	115
Khdair et al.[97]	2	Li et al.[98]	6	Vereecken and Van Looy[89]	145
Lozano et al.[99]	2	Li et al.[100]	6	Kool et al.[101]	217
Macinnis-Ng et al.[102]	2	Manyame et al.[103]	6	Nemes et al.[12]	218
Mosquera et al.[104]	2	Talat et al.[105]	6	CSIRO[106]	652
Mujdeci et al.[107]	2	Ullah et al.[108]	6	Leenaars et al.[16]	729
Abedi-koupai et al.[109]	3	Werisch et al.[110]	6	Ottoni et al.[15]	814
Basile and D'Urso[111]	3	Elliott and Price[112]	7	Stolbovoy et al.[113]	1,129
Cuenca et al.[114]	3	Ismail[115]	8	Holtan[13]	1,864
De Boever et al.[116]	3	Novak[117]	8	Batjes et al.[11]	2,541
Guzman et al.[118]	3	Saha and Kukal[119]	8	Grunwald[19]	6,008
Kassaye et al.[120]	3	Seki et al.[121]	8		

Table 1. List of sources for SWCC data and number of SWCCs (N) per dataset assembled in the GSHP database.

mass (typically, about 20% of sample mass is water at a matric potential of -3.3 m). The adapted expression for dry bulk density is then equal to:

$$\rho_{bulk} = \frac{m_{solid}}{V_{total}} = \frac{\rho_{3.3m}}{(1 + \theta_{grav3.3m})} \quad (1)$$

where ρ_{bulk} is the dry bulk density, m_{solid} is the mass of solid particles, V_{total} is the volume of the soil clod at 3.3 m matric potential), $\rho_{3.3\,m}$ is the bulk density and $\theta_{grav3.3\,m}$ is the gravimetric water content at 3.3 m matric potential. Note that Holten[13] provided the gravimetric water content at 3 m matric potential, but we assumed for simplicity that the water contents at 3 and 3.3 m matric potential are equal.

PTFs for constraining saturated (θ_s) and residual (θ_r) water contents for SWCCs without wet- and dry-end measurements. In this study, SWCCs were modeled using van Genuchten (vG) Eq. 2.

The vG model provides the volumetric water content, θ (m³/m³) at matric potential ψ (m) as

$$\theta(\psi) = \theta_r + (\theta_s - \theta_r)[1 + (\alpha|\psi|)^n]^{-(1-1/n)} \quad (2)$$

where θ_s and θ_r are the saturated and residual water contents, respectively (m³/m³), and α (m⁻¹) and n (dimensionless) are SWCC shape parameters.

To estimate the vG parameters, measurements close to full and residual saturation are needed. The GSHP database contains some SWCCs that lack the wet- and/or dry-end information (i.e., no measurements are available at matric potential ≤ 0.2 m or ≥ 150 m, respectively). Fitting the vG model to SWCCs without wet-end information leads to unreliable parameter estimates as shown in Figure S2 and Table S2. Therefore, PTFs were used to impute the wet- and/or dry-end information in this study.

PTF for θ_s. A PTF for θ_s was developed and tested based on those SWCCs that had water content information at matric potentials ranging from 0.00 or 0.01 m (saturated water content) to 150 m (permanent wilting point).

Header	Description	Units
layer_id	Unique ID of each SWCC	—
disturbed_undisturbed	Sample soil structure disturbed or undisturbed during the analysis	—
climate_classes	Climate information (temperate, boreal etc.)	—
profile_id	Unique ID of each profile	—
reference	Data reference	—
DOIs_URLs	Data DOIs or URLs	—
method	Method used to measure the SWCC	—
method_keywords	Comments on the methods if applicable	—
latitude_decimal_degrees	Ranges up to $+90$ degrees down to -90 degrees	Decimal degree
longitude_decimal_degrees	Ranges up to $+180$ degrees down to -180 degrees	Decimal degree
hzn_desgn	Soil horizon designation	—
hzn_top	Upper depth of soil sample	cm
hzn_bot	Lower depth of soil sample	cm
db_33	Bulk density at 3.3 m matric potential	g/cm^3
db_od	Dry bulk density	g/cm^3
oc	Soil organic carbon content	%
tex_psda	Soil texture classes based on USDA	—
sand_tot_psa	Mass of soil particle 2 mm for fine earth	%
silt_tot_psa	Mass of soil particle >0.05 and <2 mm for fine earth	%
clay_tot_psa	Mass of soil particles <0.002 mm for fine earth	%
ph_h2o	Soil reaction	—
ksat_field	Soil saturated hydraulic conductivity from field	cm/day
ksat_lab	Soil saturated hydraulic conductivity from lab	cm/day
porosity	Porosity	m^3/m^3
WG_33kpa	Gravimetric water content at 3.3 m matric potential	kg/kg
lab_head_m	Lab measured matric potential	m
lab_wrc	Lab measured volumetric water content	m^3/m^3
field_head_m	Field measured matric potential	m
field_wrc	Field measured volumetric water content	m^3/m^3
keywords_total_porosity	Extra information regarding porosity	—
SWCC_classes	SWCC classes (indicators for presence of wet- and dry-end information)	—
source_db	Source of the data	—
location_accuracy_min	Minimum value of location accuracy	m
location_accuracy_max	Maximum value of location accuracy	m
broad_accuracy_classes	Classes for location accuracy	—
α	vG shape parameter	m^{-1}
se_α	Standard error of α vG shape parameter	m^{-1}
n	vG shape parameter	—
se_n	Standard error of n vG shape parameter	—
θ_r	Residual water content	m^3/m^3
θ_s	Saturated water content	m^3/m^3
q2.5_α	2.5^{th} percentile of α	m^{-1}
q97.5_α	97.5^{th} percentile of α	m^{-1}
q10_α	10th percentile of α	m^{-1}
q90_α	90th percentile of α	m^{-1}
q25_α	25th percentile of α	m^{-1}
q75_α	75th percentile of α	m^{-1}
q2.5_n	2.5^{th} percentile of n	—
q97.5_n	97.5^{th} percentile of n	—
q10_n	10th percentile of n	—
q90_n	90th percentile of n	—
q25_n	25th percentile of n	—
q75_n	75th percentile of n	—
data_flag	Classes that defines the quality of the vG parameters	—

Table 2. List of all 54 variables in the GSHP database and their units.

In total, 2,287 SWCCs had both wet- and dry-end measurements and from this set we selected those 1,947 SWCCs with soil texture and bulk density data to build the PTF. After fitting the vG model to these SWCCs see below, a robust linear regression PTF was developed for θ_s using the R package 'robustbase'[22]. This package is useful when the residual errors of the regression model have long-tailed distributions.

The linear regression PTF for θ_s was developed using, as covariates, clay and sand contents, bulk density, and a categorical variable that distinguished between tropical and other climatic regions (to account for unique soil-formation conditions in tropical climate). Two PTFs of θ_s were developed depending on the available covariate information: The first PTF (Model1) was developed for the SWCCs when both soil texture and bulk density were available and the second PTF (Model2) for the SWCCs when only bulk density information was available. To account for the intense weathering processes in the wet and warm climate of the tropical regions[23], we distinguished between SWCCs from tropical and other climate regions (arid, boreal, temperate, and polar). Tropical regions have often Oxisols that are dominated by inactive (non-swelling) clay minerals (kaolinite) as shown by Ito and Wagai[24] and are characterized by intense soil-formation processes affecting soil hydraulic properties[25]. The climate region information was extracted from the Köppen-Geiger climate zone map[26,27].

The accuracy of the PTF was assessed by 20-fold cross-validation using coefficient of determination (R^2), Root Mean Square Error ($RMSE$) and $BIAS$ as accuracy measures. $BIAS$ and $RMSE$ are defined as: $BIAS = \sum_{i=1}^{N_1} \frac{(\hat{\theta}_{s,i} - \theta_{s,i})}{N_1}$ and $RMSE = \sqrt{\frac{SSE}{N_1}}$ where $SSE = \sum_{i=1}^{N_1} (\hat{\theta}_{s,i} - \theta_{s,i})^2$. SSE is the sum of squared errors between the cross-validation predictions $\hat{\theta}_{s,i}$ and the measurements $\theta_{s,i}$, and N_1 is the total number of SWCCs. The coefficient of determination R^2 is defined as: $R^2 = \left[1 - \frac{SSE}{SST} \right]$ where $SST = \sum_{i=1}^{N_1} (\theta_{s,i} - \bar{\theta}_{s,i})^2$. SST is the total sum of squares and $\bar{\theta}_{s,i}$ is the arithmetic mean of the θ_s deduced from fitting the measured SWCCs. We calculated the prediction interval for θ_s at 95% confidence and used their upper and lower limits as box constraints for θ_s when estimating the vG parameters by the 'soilhypfit'[28] (https://cran.r-project.org/web/packages/soilhypfit/index.html) R package for the SWCCs lacking wet-end measurements.

Tuller and Or[29] model for θ_r. A physically-based model by Tuller and Or[29] was used to constrain estimates of θ_r for the SWCCs where dry-end measurements were missing. The model describes the dry-end gravimetric water content (θ_m) as a function of specific surface area (SA, in m²/kg) and the thickness of water film (h, in meters) adsorbed on mineral surfaces,

$$\theta_m = h \cdot SA \cdot \rho_w \tag{3}$$

where ρ_w is the density of water. According to Or and Tuller[30], capillary contribution becomes negligible for matric potential above 1000 m for a wide range of soil textures. For water adsorbed on planar surfaces by van der Waals forces, Iwamatsu and Horii[31] expressed the equilibrium thickness (h) of an adsorbed thin water film as a function of a matric potential (ψ) as shown below in Eq. 4:

$$h = \sqrt[3]{\frac{A_{svl}}{6\pi g \rho_w \psi}} \tag{4}$$

where A_{svl} (J) is the Hamaker constant for solid-vapor interactions (we used A_{svl} equal to 6×10^{-20} J as proposed by Tuller and Or[29]), ψ (m) is the matric potential, ρ_w (kg/m³) is the density of the liquid, and g (m/s²) is acceleration due to gravity. According to Or and Tuller[32], the water film thickness of an adsorbed water layer is equal to $3.5 \cdot 10^{-10}$ m. At 150 m matric potential, the water thickness according to Eq. 4 is approximately $7.0 \cdot 10^{-10}$ m (two layers of water molecules). For SWCCs without dry-end information, we constrained the possible range of residual water content using Eq. 3 with thickness h between one and two monolayers of water ($3.5 \cdot 10^{-10}$–$7.0 \cdot 10^{-10}$ m).

Constraints on vG shape parameters. We constrained the possible values of n from 1.0 to 7.0. These limits were assigned considering literature, expert knowledge, and values obtained from fitting the SWCCs without constraints. We used minimum and maximum α values of 0 and 100 (m^{-1}) for all textural classes. Because the inverse of α characterizes the capillary pressure in the largest soil pores, the maximum value imposed for α corresponds to a maximum pore diameter of 3 mm. Detailed Information on employed constraints when estimating the vG parameters is provided in Table S3.

Fitting and quality check of the vG parameters using soilhypfit R package. SWCCs parameters were estimated, with box-constraints described in the previous sections, using the 'soilhypfit' R package[28]. 'soilhypfit' was designed for parametric modeling of SWCCs and/or unsaturated hydraulic conductivity datasets. The main function to estimate vG parameters is called 'fit_wrc_hcc' and uses maximum likelihood (ML) to estimate with the restriction $m = 1 - 1/n$. 'fit_wrc_hcc' uses optimisation algorithms of the NLopt library[33] or the Shuffled Complex Evolution (SCE) algorithm[34]. Firstly, we used the default global optimisation algorithms to fit the SWCCs. Then, we refined the parameter estimates by the default unconstrained local algorithm, again fitting the untransformed parameters (α, n) with the same parameter range as in global optimization and using as initial values the parameters estimated by the global optimization algorithm. The reason for this sequential procedure is that the parameters returned by the global algorithm are sometimes slightly away from the parameters of the true maximum of the objective function and the computed standard errors of α and n, computed from the Hessian of loglikelihood function (=the observed Fisher information matrix), are then not accurate.
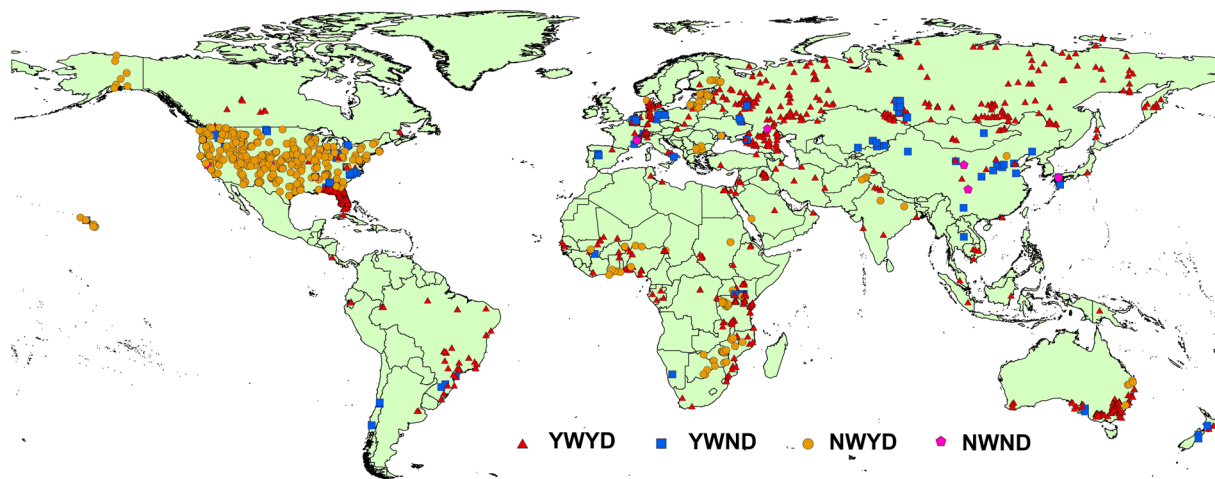
Fig. 1 Spatial distribution of SWCCs in the GSHP database. A total of 2,702 locations are shown on the map. The locations are grouped into four classes: YW and NW stand for SWCCs with and without wet-end information (water content measured for matric potential $\leq 0.2\,\mathrm{m}$), respectively, while YD and ND stand for SWCCs with and without dry-end information (water content measured at matric potential $\geq 150\,\mathrm{m}$), respectively.

Source db	$N=4$	$N=5$	$N>5$	Total number of SWCCs
AfSPDB[16]	186	255	288	**729**
Australian dataset[17,106]	1	119	648	**768**
ETH Literature	10	1884	747	**2641**
UNSODA[12]	2	2	214	**218**
WOSIS[11]	1159	374	1008	**2541**
HYBRAS[15]	—	11	803	**814**
Russia EGRPR[113]	—	1129	—	**1129**
Swiss dataset[36]	—	10	101	**111**
Belgium dataset[89]	—	—	145	**145**
Florida dataset[19]	—	—	6008	**6008**
ZALF dataset[14]	—	—	155	**155**
Total	**1,358**	**3,784**	**10,117**	**15,259**

Table 3. Number of SWCCs with 4, 5, and more than 5 data pairs (matric potential, water content) per sample along with a total number of SWCCs. The ETH literature dataset designates the SWCCs data that we collected by our own literature search, and it includes all other references shown in Table 1.

We checked the quality of estimates of the van Genuchten parameters n and α using loglikelihood profiles. From the profiles, we computed likelihood-based confidence intervals of α and n for 3 confidence levels (0.5, 0.8, and 0.95) (see chapter 4 in Uusipaikka[35]). Furthermore, we identified those SWCCs where the estimates of α and n coincided with any of the limits of the constraining intervals. We classified the parameter estimates in five classes based on the loglikelihood profiles and the standard errors of estimated parameter. The first two classes contained SWCCs with estimated parameters equal to the upper limits ($\alpha=100$, $n=7$). For these SWCCs no standard error for the estimated parameters could be specified (classes denoted as 'upper limit for α' and 'upper limit for n'). The third and fourth classes had flat likelihood profiles such that no quantiles of 75% probability or higher could be determined (classes denoted as 'flat upper profile for α' and 'flat upper profile for n'). The fifth class or final class of the SWCCs was assigned to 'good quality estimate'. After fitting the vG model to the SWCCs, standard deviations of the modelling errors, say $\hat{\sigma}_\theta$ were computed and and SWCCs with $\hat{\sigma}_\theta$ larger than 0.1 m³/m³ water content are discarded.

Data Records

The database consists of 54 variables and 136,989 records with water content measurements at given matric potential (and complementary information) recorded for 15,259 SWCCs from 2,702 locations. A list of the variables of the database along with their units is given in Table 2. The global distribution of SWCCs is shown in Fig. 1, and the source of each dataset is given in Table 1. Table 3 shows the total number of SWCCs with the corresponding number N of θ-ψ data pairs from the various sources. The SWCCs with more than 5 data pairs dominate the database with 10,117 SWCCs followed by 3,784 and 1,359 SWCCs with 5 and 4 data pairs, respectively.

Coefficients	Model1 (BD + clay + sand)		Model2 (BD)	
	Other climates	Tropical climate	Other climates	Tropical climate
a	0.917	0.932	0.987	1.011
b	−0.353	−0.353	−0.389	−0.389
c	0.00087	0.00087	—	—
d	−0.00004	−0.00004	—	—

Table 4. Coefficients of linear regression PTFs for saturated water content θ_s for tropical and other climates.
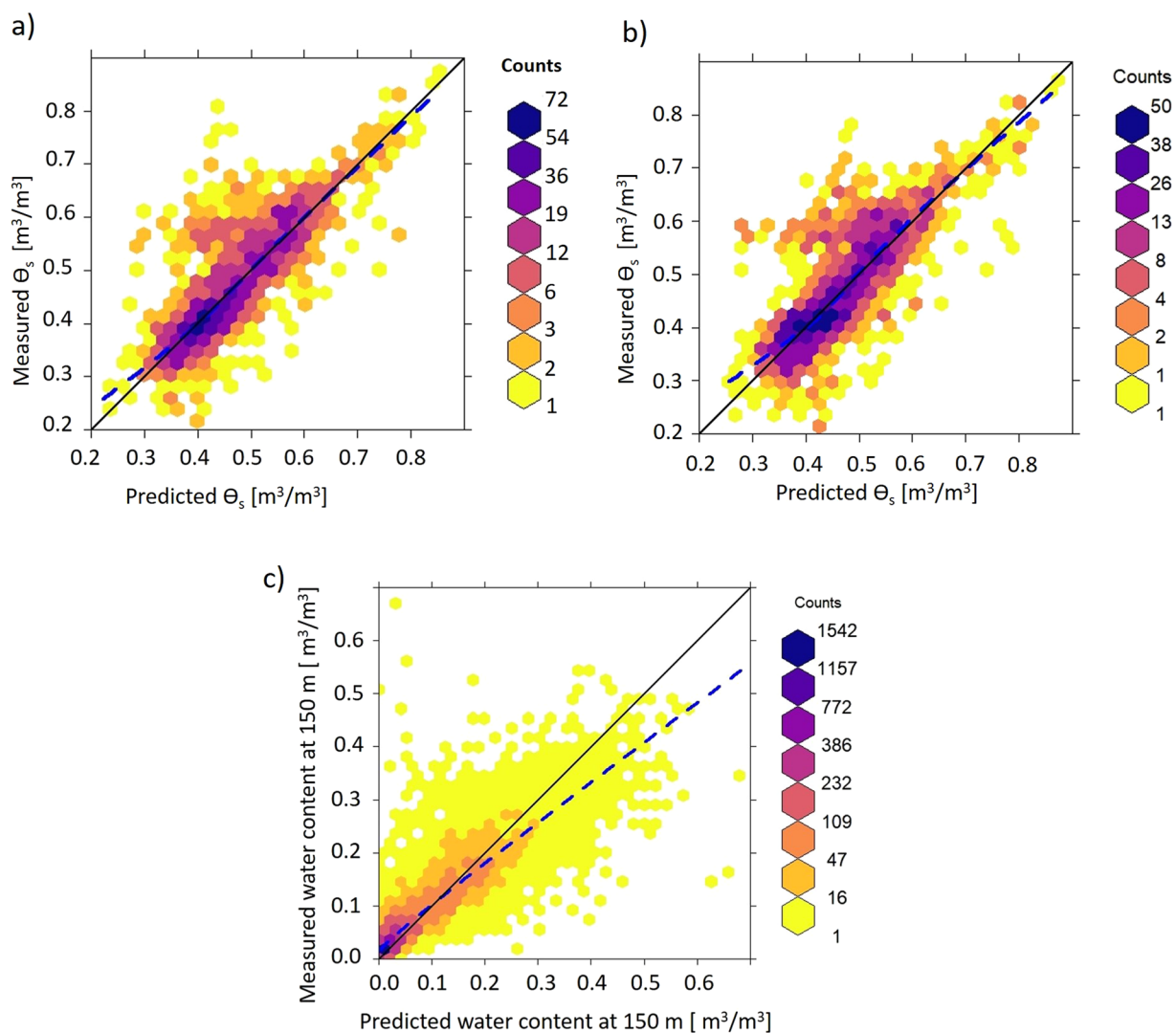

Fig. 2 Performance of linear regression PTF to estimate θ_s and of the Tuller and Or[29] model to estimate θ_r for SWCCs without wet- and dry-end information, respectively. (**a**) Relationship between measured values and cross-validation predictions of θ_s (measured θ_s are the values deduced from fitting the vG model to measured SWCCs). The solid black line is the 1:1 line, and the blue dashed line is the LOWESS (locally weighted scatter plot smoothing) curve. Cross-validation resulted in $R^2 = 0.645$ and $RMSE = 0.061$ m³/m³ with $BIAS = -0.009$ m³/m³. (**b**) Performance of PTF that uses only bulk density to estimate θ_s for SWCCs without wet-end information measurements. Cross-validation resulted in $R^2 = 0.611$ and $RMSE = 0.066$ m³/m³ with $BIAS = -0.008$ m³/m³. (**c**) Relation between measured and predicted water content at 150 m matric potential. Quantitative validation yielded the $R^2 = 0.752$, $RMSE = 0.053$ m³/m³ with $BIAS = -0.002$ m³/m³.

The dataset with highest and lowest number of SWCCs were the Florida[19] and the Swiss[36] datasets, with 6,008 and 111 SWCCs, respectively. The ETH literature dataset designates the SWCC data that we collected by our own literature search. The database and readme file is uploded on the Zenodo platform and can be accessed using this link https://doi.org/10.5281/zenodo.5547338[37].
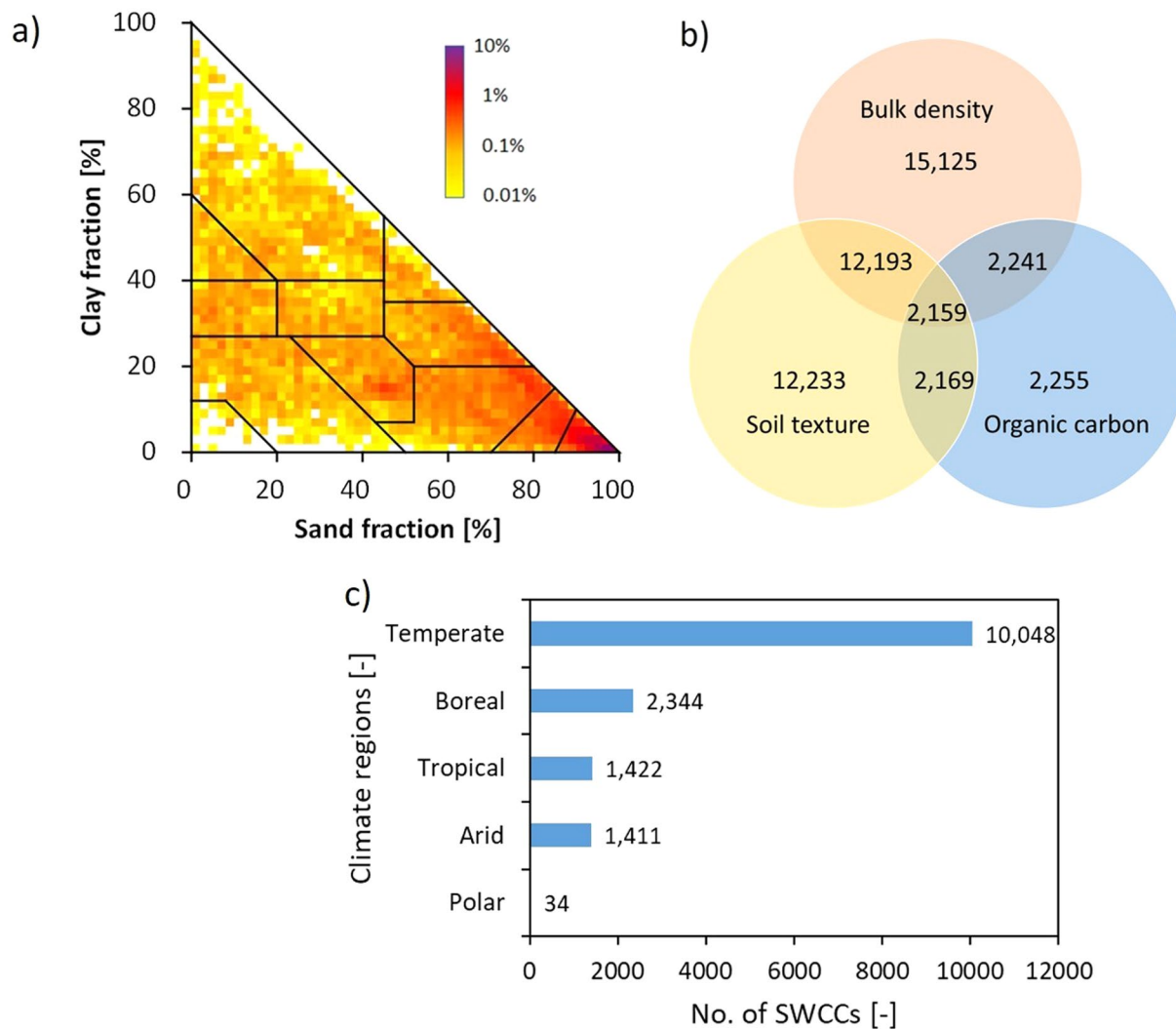
Fig. 3 Overview of SWCC data. (**a**) Distribution of soil textures of samples in the GSHP database on the USDA soil texture triangle. (**b**) Venn diagram illustrating the number of SWCCs in the GSHP database for which bulk density, soil texture, and soil organic carbon data were also available. (**c**) Number of SWCCs per climatic regions.

Technical Validation

Quality of PTFs for θ_s and θ_r. Two PTFs for θ_s were developed for SWCCs without wet-end information (Eqs. 5–6). For 30% of all SWCCs, wet-end information was missing and had to be inferred from PTFs.

$$\theta_s = a + b \cdot BD + c \cdot Clay + d \cdot Sand \quad (5)$$

$$\theta_s = a + b \cdot BD \quad (6)$$

where θ_s (m³/m³) is the saturated water content, BD the bulk density (g/cm³) and *Sand* and *Clay* fractions are given in %. Note that we differentiated between tropical and other climatic regions. Only the values of the intercepts differ for tropical climate; coefficients b, c, and d are the same for all climatic regions. The coefficients are provided in Table 4. Moreover, the variance-covariance matrices and residual standard errors are provided for both models (Eqs. 5 and 6) in the supplementary document (Tables S4 and S5). The model was calibrated with 1,947 SWCCs, and 20-fold cross-validation was applied to validate the model. The results of 20-fold cross-validation for the model with bulk density and soil texture and differentiating between tropical and other climate regions are shown in Fig. 2a. Likewise, Fig. 2b shows the results for the model using only bulk density and climatic region. Moreover, the results for θ_r were validated by comparing the predicted water content at 150 m matric potential using the Tuller and Or model with the measured water content at 150 m as shown in Fig. 2c.

Texture Classes	BD	OC	Porosity	Ksat	α	n	θ_r	θ_s
Clay	1.23, 0.22	1.34, 1.24	0.53, 0.08	4.00, 38.58	5.36, 8.98	1.59, 1.00	0.17, 0.13	0.55, 0.09
	(1,053)	(469)	(318)	(396)	(1,054)	(1,054)	(1,054)	(1,054)
Silty Clay	1.17, 0.25	1.79, 1.62	0.55, 0.08	1.25, 32.12	5.44, 12.43	1.47, 0.78	0.11, 0.11	0.52, 0.09
	(252)	(93)	(38)	(47)	(252)	(252)	(252)	(252)
Sandy Clay	1.44, 0.18	0.58, 0.80	0.43, 0.08	1.95, 19.05	4.97, 5.91	1.69, 0.93	0.17, 0.12	0.44, 0.08
	(214)	(61)	(40)	(132)	(214)	(214)	(214)	(214)
Clay Loam	1.27, 0.28	1.69, 2.23	0.56, 0.14	11.70, 28.96	3.10, 6.45	1.43, 0.59	0.09, 0.09	0.50, 0.12
	(427)	(104)	(44)	(70)	(428)	(428)	(428)	(428)
Silty Clay Loam	1.18, 0.24	1.69, 1.95	0.58, 0.12	2.9, 19.15	2.79, 6.49	1.49, 0.78	0.08, 0.08	0.52, 0.10
	(423)	(97)	(10)	(50)	(424)	(424)	(424)	(424)
Sandy Clay Loam	1.52, 0.20	0.60, 0.58	0.42, 0.07	1.09, 12.75	2.91, 4.29	1.63, 0.70	0.14, 0.10	0.41, 0.07
	(1061)	(200)	(80)	(703)	(1,062)	(1,062)	(1,062)	(1,062)
Silt	1.22, 0.21	1.44, 1.57	—	9.10, 7.81	0.65, 3.57	1.68, 0.46	0.03, 0.03	0.50, 0.07
	(36)	(18)	(0)	(10)	(36)	(36)	(36)	(36)
Silt Loam	1.24, 0.27	1.64, 2.48	0.51, 0.09	25.9, 9.40	1.30, 4.15	1.60, 0.58	0.06, 0.06	0.48, 0.10
	(849)	(259)	(19)	(193)	(857)	(857)	(857)	(857)
Loam	1.35, 0.29	1.36, 1.80	0.37, 0.08	15.54, 12.38	2.50, 4.65	1.50, 0.54	0.08, 0.07	0.46, 0.10
	(798)	(319)	(215)	(101)	(811)	(811)	(811)	(811)
Sandy Loam	1.46, 0.28	0.95, 1.42	0.42, 0.09	1.98, 10.34	1.88, 4.06	1.71, 0.65	0.08, 0.06	0.41, 0.09
	(1,858)	(316)	(56)	(789)	(1,865)	(1,865)	(1,865)	(1,865)
Loamy Sand	1.50, 0.21	0.55, 0.83	0.47, 0.06	5.11, 6.41	2.63, 3.44	1.90, 0.77	0.06, 0.04	0.39, 0.08
	(996)	(113)	(9)	(584)	(996)	(996)	(996)	(996)
Sand	1.50, 0.14	0.71, 1.00	0.43, 0.04	22.04, 3.34	2.66, 2.09	3.17, 1.34	0.04, 0.02	0.39, 0.06
	(4,226)	(120)	(17)	(3,884)	(4,234)	(4,234)	(4,234)	(4,234)
Total number of SWCCs respective data	**12,193**	**2,169**	**846**	**6,985**	**12,223**	**12,223**	**12,223**	**12,223**

Table 5. Means and standard deviations (format: mean, standard deviation) of basic soil properties, hydraulic conductivity, and vG SWCCs parameters per soil textural class. The number of SWCCs is given in parenthesis. BD bulk density (g/cm³), OC soil organic carbon content (%), Porosity (m³/m³), Ksat saturated hydraulic conductivity (cm/day) measured in laboratory, α (m⁻¹) and n (dimensionless) vG shape parameters, θ_r residual water content (m³/m³), θ_s saturated water content (m³/m³). For Ksat and α the geometric mean is reported (because Ksat and α are approximately lognormally distributed), while for all other properties, the arithmetic mean is provided.

Spatial coverage and auxiliary soil properties. The GSHP database contains SWCCs from all USDA soil textural classes (see Fig. 3a). Concerning the geographical distribution of the data, most of the SWCCs are from North America followed by Europe, Africa, Asia, South America, Australia/Oceania as shown in Table S6. 10,048 SWCCs belong to the temperate region, and 2,344, 1,422, 1,411, and 34 SWCCs to boreal, tropical, arid, and polar regions, respectively (Fig. 3c). Regarding the measurement method of SWCCs, 99% of the SWCCs were measured in the laboratory and only 1% stem from the field (most field SWCCs are from the UNSODA dataset). Note that different laboratory methods were used to estimate SWCCs. Most commonly, pressure plate and sand-box apparatus for wet-end and pressure chamber for the dry-end measurements were used.

Along with SWCCs, in GSHP we also collected information on soil texture, bulk density, organic carbon, porosity, and saturated hydraulic conductivity (Ksat). Out of 15,259 SWCCs, 12,233, 15,125, 2,255, 2,754 SWCCs have information on soil texture, bulk density, organic carbon, porosity, respectively. In addition, 12,193 SWCCs have both soil texture and bulk density information, and 2,159 SWCCs have information on soil texture, bulk density, and organic carbon, as shown in Fig. 3b. Note that in addition to 12,233 soil texture values, 82 measurements have soil texture information with a total (sand + silt + clay) less than 98% or greater than 102%. We did not use these SWCCs in the development of the PTF for θ_s, but kept them in the database and assigned the value "Error" to the soil texture class variable. The database also contains Ksat data for 8,675 soil samples that allow the quantification of the unsaturated hydraulic conductivity function using the Mualem-van Genuchten parameterization (see Figure S3). Regarding the number of soil profiles per location, in the GSHP database, 214 locations are without profile information whereas 2,390, 60, 17, and 21 locations have 1, 2–5, 6–10, and >10 soil profiles, respectively. Similarly, 259, 1,418, 916, and 109 locations have 1, 2–5, 6–10, and >10 soil samples, respectively. Moreover, 32 locations are without soil depth information whereas 287, 1,447, 876, 60 locations have 1, 2–5, 6–10, and >10 soil depths, respectively as shown in Table S7.

After the quality check of SWCCs, 10,373 SWCCs have been assigned as most accurate with a 0–100 m location accuracy. 1,890 SWCCs lack location accuracy information. Furthermore, after calculating the likelihood-based confidence intervals, a total of 11,705 SWCCs out of 15,259 emerges as 'good quality estimate' whereas 1,857, 633, 947, and 117 were classified as 'flat upper profile for α', 'flat upper profile for n', 'upper limit for α', and 'upper limit for n', respectively.
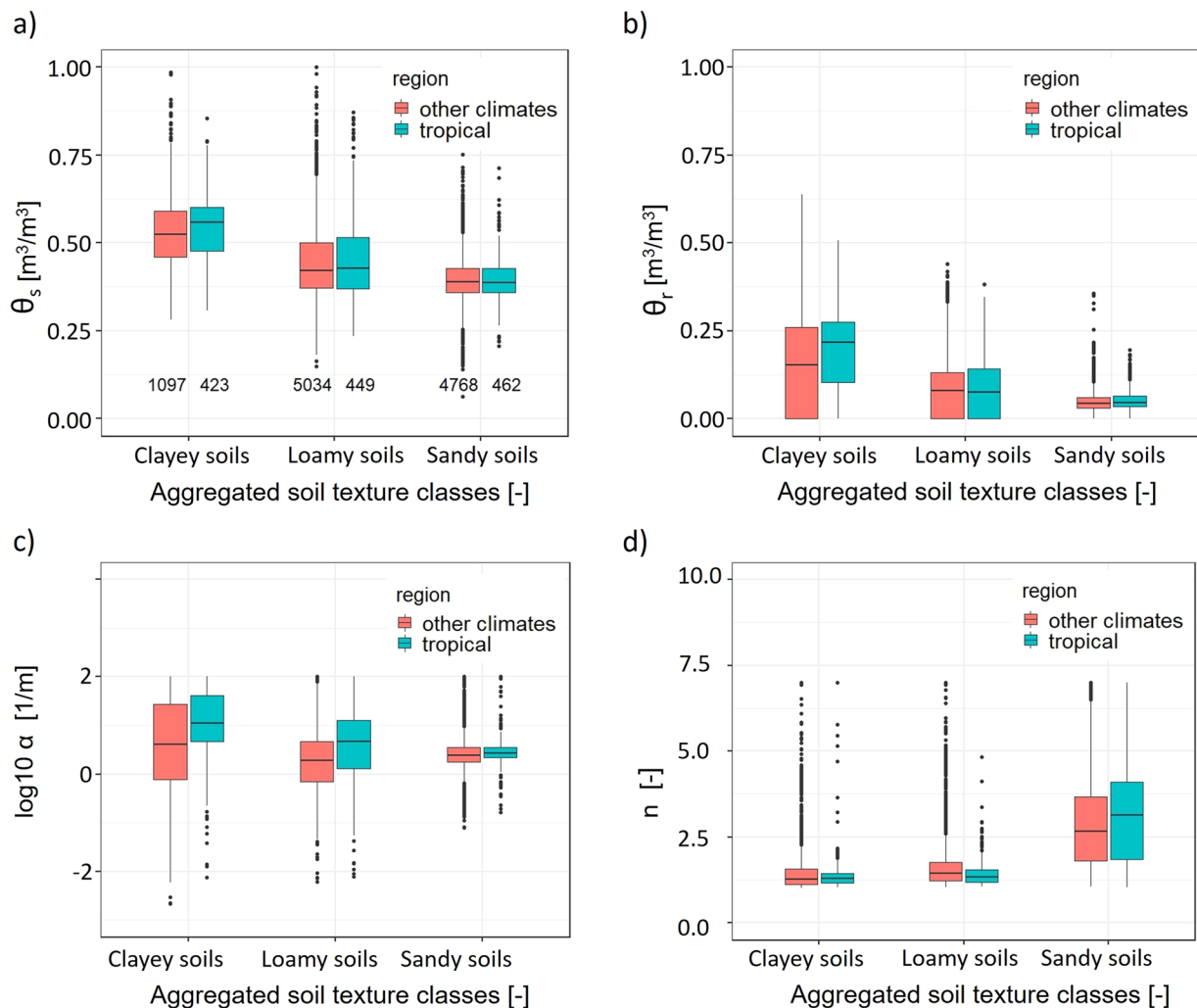
Fig. 4 Distribution of vG parameters using aggregated soil texture classes (sandy soils: sand and loamy sand; loamy soils: sandy loam, loam, silt loam, silt, silty clay loam, clay loam, and sandy clay loam; clayey soils: sandy clay, silty clay, and clay) for SWCCs from tropical and other climatic regions. Note that soil textures are estimated using the USDA-Natural Resources Conservation Service soil texture triangle. The numbers in panel (**a**) show the number of vG parameters for different aggregated soil texture classes used to make this plots.

Statistical characteristics of vG parameters.   Table 5 reports the means and standard deviations of vG parameters and the number of SWCCs per soil textural class. The largest mean values of α were observed for clayey soils (classes clay and silty clay); loamy soils, and sandy soils showed smaller values of α. In contrast, the largest mean n was observed for sandy soils (classes loamy sand and sand). For the other textural classes, n ranged from 1.34 to 1.71. Similarly, highest and lowest θ_s and θ_r were observed for clayey and sandy soils, respectively. Regarding other soil properties, the largest Ksat values were obtained for sandy soils followed by clay soils, most likely due to the presence of macropores in these fine-textured soils[18]. Fig. 4 shows the distribution of vG parameters for combinations of aggregated soil texture classes and tropical and other climatic regions. In the database, there are 1,097 clayey, 5,034 loamy, and 4,768 sandy SWCCs from other climates and 423 clayey, 449 loamy, and 462 sandy SWCCs from tropical regions in the GSHP database.

Usage Notes
Challenges to determine vG parameters.   The major problem to determine the vG parameters was missing measurements of water content close to saturation. Imposing the expected range of water content close to saturation had a large effect on the estimated vG parameters as shown with three illustrative examples in Figure S2 and Table S2. To scope with SWCCs that have not enough measured data pairs to cover the full matric potential and water content range, we imposed limits for the range of parameter values based on literature. Although we provided the PTF-based values of θ_s for SWCCs that lack the wet-end information, after estimating the likelihood-based confidence intervals, we noticed that some of the SWCCs are fitted with vG parameters equal to upper limits of box constraints used in the fitting process. For example, 15% SWCCs without measured wet-end

information were fitted with α values equal to a maximum value (100 m^{-1}). Moreover, only 37% (1,714 out of 4,598 SWCCs) of SWCCs without wet-end information were assigned to the class of 'good quality estimate'. We conclude that the limited number of measured data pairs and water content range imposes a large uncertainty of vG parameter values. But due to the scarcity of more complete measurements in many regions of the world, we must use such SWCCs as well that allow some general characterization of the unsaturated soil properties.

Effect of climate on SWCCs. Soil hydraulic properties do not depend on soil texture alone but on soil structure as well, especially saturated hydraulic conductivity Ksat and the shape parameter α that is inversely related to the air entry value[10,18]. Soil-formation processes are particularly intense in tropical regions, it can be expected that SWCCs parameter values differ for tropical and other climatic regions[38]. For both clayey and loamy soils, we found larger α values for tropical regions. Similarly, Ottoni et al.[15]. also reported that soil hydraulic properties (vG parameters and Ksat) are different in the tropical compared to the temperate regions. Likewise, Gupta et al.[18] illustrate that the PTFs developed using temperate Ksat measurements could not predict successfully tropical Ksat measurements. We hypothesize that this is due to the differences in the soil-forming processes that also determine the clay type and mineralogy. Tropical soils are often Oxisols rich inactive (non-swelling) clay minerals (kaolinite). In contrast to tropical soils, active (smectite) and moderately active clay minerals (illite) are the dominant clay minerals in other climate regions. These swelling clay minerals retain water within internal structures with strong capillary forces. Recently, Lehmann et al.[25] revealed that the incorporation of clay-type informed PTFs could improve characterization of soil hydraulic and mechanical properties.

Limitations of GSHP database. Some limitations need to be taken into account when using the GSHP database, as shortly detailed here. The database still lacks data from some regions (mainly Canada, and northern and western Australia). The database, in addition to SWCCs with $\geq$5 data pairs, also includes 9% SWCCs with 4 data pairs only (see spatial distribution of θ-ψ data pairs in Figure S4). For these SWCCs, the vG parameters are not well estimated because the measurements do not allow to properly capture the shape of the SWCC. 94% of SWCCs with 4 data pairs belong to the class without wet-end information. For some samples, we extracted the coordinates using Google Earth, and 12.3% SWCCs have no location accuracy information, hence, we advise using the coordinates with caution. Some estimated parameters were equal to the limits of the higher range of vG parameters but this represents only a limited portion (1% SWCCs attained the n value = 7 and 6.2% SWCCs obtained α = 100 m^{-1}) of the total database. In the end, 11,705 SWCCs emerge as the good quality estimate after passing the profile likelihood test. Therefore, we recommend using these SWCCs with confidence, and other SWCCs should be used with care.

Future applications of the GSHP database. The new GSHP database presents SWCC data from all continents and covers regions in Russia, which were so far not represented in currently available datasets and generally not included in PTF training and global mapping of soil hydraulic properties (note that there are still some gaps in the geographical representation of the data, especially data are lacking from Canada and northern and southern Australia). This global coverage is a great asset for the development of global PTFs for vG parameters. Remote sensing technology opens the doors to link measured hydraulic properties with global environmental data. For example, Gupta et al.[39] linked Ksat with satellite-based maps of environmental covariates such as local information on vegetation, climate, and topography to generate a global map of Ksat with 1-km resolution. Likewise Chaney et al.[40] developed the maps of soil hydraulic properties (POLARIS dataset) at 30 m resolution using remote sensing covariates for the United States using SSUGRO (Soil Survey Geographic) database. We recently used the GSHP dataset to generate highly resolved global maps of vG parameters. For that purpose, we applied machine learning approaches to relate vG parameters to soil and environmental covariates and predicted vG parameter values for all locations as a function of remote sensing information[41].

Code availability

All collected data and related soil characteristics are provided online for reference and are available at https://doi.org/10.5281/zenodo.5547338[37]. The R code used to create the GSHP database is available on Github (https://github.com/ETHZ-repositories/GSHP-database).

References

1. Brooks, R. H. & Corey, A. T. Hydraulic properties of porous media and their relation to drainage design. *Trans. ASAE* **7**, 26–0028 (1964).
2. van Genuchten, M. T. A closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **44**, 892–898 (1980).
3. Jarvis, N., Bergstrom, L. & Dik, P. Modeling water and solute transport in macroporous soil. 2. Chloride breakthrough under nonsteady flow. *J. Soil Sci* **42**, 71–81 (1991).
4. Tuller, M., Or, D. & Hillel, D. Retention of water in soil and the soil water characteristic curve. *Encycl. Soils Environ* **4**, 278–289 (2004).
5. Klute, A. et al. Methods of soil analysis (1986).
6. Bouma, J. Using soil survey data for quantitative land evaluation. In *Adva. Soil Sci.*, 177–213 (Springer, 1989).
7. Weynants, M., Vereecken, H. & Javaux, M. Revisiting Vereecken pedotransfer functions: Introducing a closed-form hydraulic model. *Vadose Zone J* **8**, 86–95 (2009).
8. Jorda, H., Bechtold, M., Jarvis, N. & Koestel, J. Using boosted regression trees to explore key factors controlling saturated and near-saturated hydraulic conductivity. *Eur. J. Soil Sci.* **66**, 744–756 (2015).

11

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.* **67**, 276–284 (2016).

10. Or, D. The tyranny of small scales—On representing soil processes in global land surface models. *Water Resour. Res.* 56 (2020).

11. Batjes, N. H., Ribeiro, E. & Van Oostrum, A. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth Syst. Sci. Data* **12**, 299–320 (2020).

12. Nemes, A. D., Schaap, M., Leij, F. & Wösten, J. Description of the unsaturated soil hydraulic database UNSODA version 2.0. *J. Hydrol.* **251**, 151–162 (2001).

13. Holtan, H. N. *Moisture-tension data for selected soils on experimental watersheds*, vol. 41 (Agricultural Research Service, US Department of Agriculture, 1968).

14. Schindler, U. G. & Müller, L. Soil hydraulic functions of international soils measured with the Extended Evaporation Method (EEM) and the HYPROP device. *Open Data J. for Agric. Res.* **3** (2017).

15. Ottoni, M. V., Ottoni Filho, T. B., Schaap, M. G., Lopes-Assad, M. L. R. & Rotunno Filho, O. C. Hydrophysical database for Brazilian soils (HYBRAS) and pedotransfer functions for water retention. *Vadose Zone J.* 17 (2018).

16. Leenaars, J., Kempen, B., van Oostrum, A. & Batjes, N. Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa. *Arrouays et al.(eds.), 2014b* 51–57 (2014).

17. Forrest, J., Beatty, H., Hignett, C., Pickering, J. & Williams, R. Survey of the physical properties of wheatland soils in eastern Australia. Tech. Rep., CSIRO Division of Soils (1985).

18. Gupta, S., Hengl, T., Lehmann, P., Bonetti, S. & Or, D. SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications. *Earth Syst. Sci. Data* **13**, 1593–1612 (2021).

19. Grunwald, S. Florida soil characterization data, Soil and water science department, IFAS-Instituite of food and agriculture science. Tech. Rep., University of Florida (2020).

20. Assi, A. T., Blake, J., Mohtar, R. H. & Braudeau, E. Soil aggregates structure-based approach for quantifying the field capacity, permanent wilting point and available water capacity. *Irrigation Sci* **37**, 511–522 (2019).

21. Boivin, P., Garnier, P. & Vauclin, M. Modeling the soil shrinkage and water retention curves with the same equations. *Soil Sci. Soc. Am. journal* **70**, 1082–1093 (2006).

22. Maechler, M. *et al.* Package 'robustbase'. *Basic Robust Stat.* (2021).

23. 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* **108**, 155–180 (2002).

24. Ito, A. & Wagai, R. Global distribution of clay-size minerals on land surface for biogeochemical and climatological studies. *Sci. Data* **4**, 1–11 (2017).

25. Lehmann, P. *et al.* Clays are not created equal: How clay mineral type affects soil parameterization. *Geophys. Res. Lett.* **48**, e2021GL095311 (2021).

26. Rubel, F. & Kottek, M. Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorol. Zeitschrift* **19**, 135–141 (2010).

27. Hamel, P. *et al.* Sediment delivery modeling in practice: Comparing the effects of watershed characteristics and data resolution across hydroclimatic regions. *Sci. Total. Environ.* **580**, 1381–1388 (2017).

28. Papritz, A. *soilhypfit: Modelling of Soil Water Retention and Hydraulic Conductivity Data*. R package version 348 0.1-5 (2022).

29. Tuller, M. & Or, D. Water films and scaling of soil characteristic curves at low water contents. *Water Resour. Res.* 41 (2005).

30. Or, D. & Tuller, M. Liquid retention and interfacial area in variably saturated porous media: Upscaling from single-pore to sample-scale model. *Water Resour. Res.* **35**, 3591–3605 (1999).

31. Iwamatsu, M. & Horii, K. Capillary condensation and adhesion of two wetter surfaces. *J. Colloid Interface Sci* **182**, 400–406 (1996).

32. Leão, T. P. & Tuller, M. Relating soil specific surface area, water film thickness, and water vapor adsorption. *Water Resour. Res.* **50**, 7873–7885 (2014).

33. Johnson, S. G. The NLopt nonlinear-optimization package (2014).

34. Duan, Q., Sorooshian, S. & Gupta, V. K. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.* **158**, 265–284 (1994).

35. Uusipaikka, E. *Confidence Intervals in Generalized Regression Models* (Chapman & Hall/CRC Press, 2008).

36. Richard, F. & Lüscher, P. *Physikalische Eigenschaften von Böden der Schweiz. Lokalformen. Eidg. Anstalt für das forstliche Versuchswesen. Sonderserie.* (Eidgenössische Technische University, 1983/87).

37. Gupta, S. *et al.* GSHP: Global database of soil hydraulic properties. *Zenodo* https://doi.org/10.5281/zenodo.5547338 (2021).

38. Tomasella, J., Hodnett, M. G. & Rossato, L. Pedotransfer functions for the estimation of soil water retention in Brazilian soils. *Soil Sci. Soc. Am. J.* **64**, 327–338 (2000).

39. 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.* **13**, e2020MS002242 (2021).

40. Chaney, N. W. *et al.* POLARIS: A 30-meter probabilistic soil series map of the contiguous United States. *Geoderma* **274**, 54–67 (2016).

41. Gupta, S. *et al.* Global Mapping of Soil Water Characteristics Parameters—Fusing Curated Data with Machine Learning and Environmental Covariates. *Remote Sensing*, **14**(8), 1947 (2022).

42. Al-Darby, A., Al-Asfoor, S. & El-Shafei, Y. Effect of Soil Gel-Conditioner on the Hydrophysical Properties Sandy Soil J. *Saudi Soc. Agr. Sci* **1**, 14–40 (2002).

43. Li, Y., Kinzelbach, W., Zhou, J., Cheng, G. & Li, X. Modelling irrigated maize with a combination of coupled-model simulation and uncertainty analysis, in the northwest of China. *Hydrol. Earth Syst. Sci.* **16**, 1465–1480 (2012).

44. Al Majou, H., Muller, F., Penhoud, P. & Bruand, A. Prediction of water retention properties of Syrian clayey soils. *Arid Land Res. Manag.* 1–20 (2021).

45. Alghamdi, A. G., Alkhasha, A. & Ibrahim, H. M. Effect of biochar particle size on water retention and availability in a sandy loam soil. *J. Saudi Chem. Soc.* **24**, 1042–1050 (2020).

46. Abid, M. & Lal, R. Tillage and drainage impact on soil quality: II. Tensile strength of aggregates, moisture retention and water infiltration. *Soil Tillage Res* **103**, 364–372 (2009).

47. Asghari, S., Ahmadnejad, S. & Keivan Behjou, F. Deforestation effects on soil quality and water retention curve parameters in eastern Ardabil, Iran. *Eurasian Soil Sci* **49**, 338–346 (2016).

48. Are, K., Oshunsanya, S. & Oluwatosin, G. Changes in soil physical health indicators of an eroded land as influenced by integrated use of narrow grass strips and mulch. *Soil Tillage Res* **184**, 269–280 (2018).

49. Bescansa, P., Imaz, M., Virto, I., Enrique, A. & Hoogmoed, W. Soil water retention as affected by tillage and residue management in semiarid Spain. *Soil Tillage Res* **87**, 19–27 (2006).

50. MacVicar, C., Loxton, R. & van der Eyk, J. South African soil series. Part II: Profile descriptions and analytical data (1965).

51. Babaeian, E. *et al.* A comparative study of multiple approaches for predicting the soil–water retention curve: Hyperspectral information vs. basic soil properties. *Soil Sci. Soc. Am. J.* **79**, 1043–1058 (2015).

52. Dlapa, P. *et al.* The Impact of land-use on the hierarchical pore size distribution and water retention properties in loamy soils. *Water* **12**, 339 (2020).

53. Noguchi, S. *et al.* Soil physical properties and preferential flow pathways in tropical rain forest, Bukit Tarek, Peninsular Malaysia. *J. For. Res* **2**, 115–120 (1997).

54. Bhushan, L. & Sharma, P. Long-term effects of lantana residue additions on water retention and transmission properties of a medium-textured soil under rice–wheat cropping in northwest India. *Soil Use Manag* **21**, 32–37 (2005).

55. Hoshino, A. *et al.* Effects of crop abandonment and grazing exclusion on available soil water and other soil properties in a semi-arid Mongolian grassland. *Soil Tillage Res* **105**, 228–235 (2009).

56. Tobón, C., Bruijnzeel, L., Frumau, K. A. & Calvo-Alvarado, J. Changes in soil physical properties after conversion of tropical montane cloud forest to pasture in northern Costa Rica. *Trop. Montane Cloud For. Sci. for Conserv. Manag.* 502–515 (2010).

57. de Oliveira, L. A. *et al.* Atrazine movement in corn cultivated soil using HYDRUS-2D: A comparison between real and simulated data. *J. Environ. Manag.* **248**, 109311 (2019).

58. Kumar, S., Sekhar, M., Reddy, D. & Mohan Kumar, M. Estimation of soil hydraulic properties and their uncertainty: comparison between laboratory and field experiment. *Hydrol. Process.* **24**, 3426–3435 (2010).

59. Tyagi, J., Qazi, N., Rai, S. & Singh, M. Analysis of soil moisture variation by forest cover structure in lower western Himalayas. *India. J. For. Res* **24**, 317–324 (2013).

60. Garba, M., Cornelis, W. M. & Steppe, K. Effect of termite mound material on the physical properties of sandy soil and on the growth characteristics of tomato (Solanum lycopersicum L.) in semi-arid Niger. *Plant Soil* **338**, 451–466 (2011).

61. McBeath, T., Grant, C., Murray, R. & Chittleborough, D. Effects of subsoil amendments on soil physical properties, crop response, and soil water quality in a dry year. *Soil Res* **48**, 140–149 (2010).

62. AL-Kayssi, A. Use of water retention data and soil physical quality index S to quantify hard-setting and degree of soil compactness indices of gypsiferous soils. *Soil Tillage Res* **206**, 104805 (2021).

63. Głąb, T. *et al.* Fertilization effects of compost produced from maize, sewage sludge and biochar on soil water retention and chemical properties. *Soil Tillage Res* **197**, 104493 (2020).

64. Mondal, S. *et al.* Effect of different rice establishment methods on soil physical properties in drought-prone, rainfed lowlands of Bihar, India. *Soil Res* **54**, 997–1006 (2016).

65. Karup, D., Moldrup, P., Tuller, M., Arthur, E. & de Jonge, L. W. Prediction of the soil water retention curve for structured soil from saturation to oven-dryness. *Eur. J. Soil Sci.* **68**, 57–65 (2017).

66. Kakeh, J. *et al.* Biological soil crusts determine soil properties and salt dynamics under arid climatic condition in Qara Qir, Iran. *Sci. Total. Environ.* **732**, 139168 (2020).

67. Ng, C. W. W., Owusu, S. T., Zhou, C. & Chiu, A. C. F. Effects of sesquioxide content on stress-dependent water retention behaviour of weathered soils. *Eng. Geol.* **266**, 105455 (2020).

68. Simmons, L. A. *Soil hydraulic and physical properties as affected by logging management.* Ph.D. thesis, [University of Missouri–Columbia] (2014).

69. Lowe, M.-A. *et al.* Bacillus subtilis and surfactant amendments for the breakdown of soil water repellency in a sandy soil. *Geoderma* **344**, 108–118 (2019).

70. Smettem, K. & Gregory, P. The relation between soil water retention and particle size distribution parameters for some predominantly sandy Western Australian soils. *Soil Res* **34**, 695–708 (1996).

71. Wang, J., Wang, P., Qin, Q. & Wang, H. The effects of land subsidence and rehabilitation on soil hydraulic properties in a mining area in the Loess Plateau of China. *Catena* **159**, 51–59 (2017).

72. Medina, H., Tarawally, M., del Valle, A. & Ruiz, M. E. Estimating soil water retention curve in rhodic ferralsols from basic soil data. *Geoderma* **108**, 277–285 (2002).

73. Xia, J., Zhao, Z. & Fang, Y. Soil hydro-physical characteristics and water retention function of typical shrubbery stands in the Yellow River Delta of China. *Catena* **156**, 315–324 (2017).

74. Cooper, M., Boschi, R. S., Silva, L. F. S. D., Toma, R. S. & Vidal-Torrado, P. Hydro-physical characterization of soils under the Restinga Forest. *Sci. Agric.* **74**, 393–400 (2017).

75. Nyamangara, J., Gotosa, J. & Mpofu, S. Cattle manure effects on structural stability and water retention capacity of a granitic sandy soil in Zimbabwe. *Soil Tillage Res* **62**, 157–162 (2001).

76. Chari, M. M. & Vahidi, A. Effect of mean soil radius on estimation of water-soil moisture curve based on Aria-Paris model. *Iran. J. Irrigation & Drainage* **14**, 2234–2243 (2021).

77. Van Quang, P., *et al.* Soil penetration resistance and its dependence on soil moisture and age of the raised-beds in the Mekong Delta, Vietnam (2012).

78. Sulaeman, D., Sari, E. & Westhoff, T. Effects of peat fires on soil chemical and physical properties: a case study in South Sumatra. In *IOP Conference Series: Environ. Earth Sci.* vol. 648, 012146 (IOP Publishing, 2021).

79. Eden, M., Bens, O., Betz, S. & Völkel, J. Characterization of soil structure in Neuras, a Namibian desert-vineyard. *DIE ERDE–Journal Geogr. Soc. Berlin* **151**, 207–226 (2020).

80. Pan, T., Hou, S., Liu, Y. & Tan, Q. Comparison of three models fitting the soil water retention curves in a degraded alpine meadow region. *Sci. Reports* **9**, 1–12 (2019).

81. Thakur, V. K., Sreedeep, S. & Singh, D. N. Evaluation of various pedo-transfer functions for developing soil-water characteristic curve of a silty soil. *Geotech. Test. J.* **30**, 25–30 (2007).

82. Moazeni-Noghondar, S., Golkarian, A., Azari, M. & Asgari Lajayer, B. Study on soil water retention and infiltration rate: a case study in eastern Iran. *Environ. Earth Sci.* **80**, 1–18 (2021).

83. Bambra, A. *Soil loss estimation in experimental orchard at Nauni in Solan district of Himachal Pradesh.* Ph.D. thesis, Dr. Yashwant Singh Parmar, University of horticulture and forestry (2016).

84. Wickland, B., Wilson, W., Wijewickreme, D. & Fredlund, D. Mixtures of waste rock and tailings: resistance to acid rock drainage. *J. Am. Soc. Min. Reclam.* (2006).

85. Nano, C. U., Nicolardot, B., Quinche, M., Munier-Jolain, N. & Ubertosi, M. Effects of integrated weed management based cropping systems on the water retention of a silty clay loam soil. *Soil Tillage Res* **156**, 74–82 (2016).

86. Marui, A. *et al.* Soil physical properties to grow the wild licorice at semi-arid area in Mongolia. *J. Arid Land Stud* **22**, 33–36 (2012).

87. Zebarth, B., Neilsen, G., Hogue, E. & Neilsen, D. Influence of organic waste amendments on selected soil physical and chemical properties. *Can. J. Soil Sci.* **79**, 501–504 (1999).

88. Toriyama, J. *et al.* Soil pore characteristics of evergreen and deciduous forests of the tropical monsoon region in Cambodia. *Hydrol. Process.* **25**, 714–726 (2011).

89. Vereecken, H., Van Looy, K., Weynants, M. & Javaux, M. Soil retention and conductivity curve data base sDB, link to MATLAB files (2017).

90. Zhang, Y.-L., Feng, S.-Y., Wang, F.-X. & Binley, A. Simulation of soil water flow and heat transport in drip irrigated potato field with raised beds and full plastic-film mulch in a semiarid area. *Agric. Water Manag* **209**, 178–187 (2018).

91. Xing, X., Li, Y. & Ma, X. Water retention curve correction using changes in bulk density during data collection. *Eng. Geol.* **233**, 231–237 (2018).

92. Jauhiainen, M. *et al. Relationships of particle size distribution curve, soil water retention curve and unsaturated hydraulic conductivity and their implications on water balance of forested and agricultural hillslopes* (Helsinki University of Technology, 2004).

93. El-Asswad, R., Said, A. & Mornag, M. Effect of olive oil cake on water holding capacity of sandy soils in Libya. *J. Arid Environ.* **24**, 409–413 (1993).

94. Jha, P., Mohapatra, K. & Dubey, S. Impact of land use on physico-chemical and hydrological properties of ustifluvent soils in riparian zone of river Yamuna, India. *Agrofor. Syst* **80**, 437–445 (2010).

95. Ismail, S. M. Influence of effective microorganisms and green manure on soil properties and productivity of pearl millet and alfalfa grown on sandy loam in Saudi Arabia. *Afr. J. Microbiol. Res.* **7**, 375–382 (2013).
96. Konyai, S., Sriboonlue, V., Trelo-Ges, V. & Muangson, N. Hysteresis of water retention curve of saline soil. In *Unsaturated Soils* 2006, 1394–1404 (2006).
97. Khdair, A. I., Khdair, S. I. & Abu-Rumman, G. A. Dataset on some soil properties improvement by the addition of olive pomace. *Data Brief* **24**, 103878 (2019).
98. Li, L. *et al.* Soil physical qualities in an Oxic Paleustalf under different tillage and stubble management practices and application of S theory. *Soil Tillage Res* **113**, 82–88 (2011).
99. Lozano, L. A., Soracco, C. G., Buda, V. S., Sarli, G. O. & Filgueira, R. R. Stabilization of soil hydraulic properties under a long term no-till system. *Revista Brasileira de Ciência do Solo* **38**, 1281–1292 (2014).
100. Li, Z. *et al.* Quantification of soil water balance components based on continuous soil moisture measurement and the Richards equation in an irrigated agricultural field of a desert oasis. *Hydrol. Earth Syst. Sci.* **23**, 4685–4706 (2019).
101. Kool, J., *et al.* Physical and chemical characterization of the Groseclose soil mapping unit. *Tech. Rep.*, Virginia Agricultural Experiment Station (1986).
102. Macinnis-Ng, C. *et al.* Root biomass distribution and soil properties of an open woodland on a duplex soil. *Plant Soil* **327**, 377–388 (2010).
103. Manyame, C. *et al.* Modeling hydraulic properties of sandy soils of Niger using pedotransfer functions. *Geoderma* **141**, 407–415 (2007).
104. Mosquera, G. M. *et al.* A field, laboratory, and literature review evaluation of the water retention curve of volcanic ash soils: How well do standard laboratory methods reflect field conditions? *Hydrol. Process.* **35**, e14011 (2021).
105. Talat, A. E., Galal, M. E., Yeser, A. & Saad El-Dein, A. Quantifying the hydraulic properties of some Egyptian soils using RETC code. *Arab. Univ. J. Agric. Sci.* **28**, 685–694 (2020).
106. CSIRO. CSIRO National Soil Site Database, v4, CSIRO, Data Collection. (2020).
107. Mujdeci, M. *et al.* The effects of organic amendments on soil water retention characteristics under conventional tillage system. *Fresenius Environ. Bull.* **26**, 4075–4081 (2017).
108. Ullah, R. *et al.* Soil Water Release Curves: Indicator to Suit Sustainable Cropping Scheme under Sloppy Rain-Fed Climatic Conditions of Pothowar Plateau of Punjab-Pakistan. *Agritropica: J. Agric. Sci.* **1**, 9–24 (2018).
109. Abedi-Koupai, J., Sohrab, F. & Swarbrick, G. Evaluation of hydrogel application on soil water retention characteristics. *J. Plant Nutr.* **31**, 317–331 (2008).
110. Werisch, S., Grundmann, J., Al-Dhuhli, H., Algharibi, E. & Lennartz, F. Multiobjective parameter estimation of hydraulic properties for a sandy soil in Oman. *Environ. Earth Sci* **72**, 4935–4956 (2014).
111. Basile, A. & D'Urso, G. Experimental corrections of simplified methods for predicting water retention curves in clay-loamy soils from particle-size determination. *Soil Technol* **10**, 261–272 (1997).
112. Elliott, J. & Price, J. Comparison of soil hydraulic properties estimated from steady-state experiments and transient field observations through simulating soil moisture in regenerated Sphagnum moss. *J. Hydrol.* **582**, 124489 (2020).
113. Stolbovoy, V., Molchanov, E. & Sheremet, B. Morphogenetic basis of the unified state register of soil resources of Russia. *Bull. Soil Sci. Institute. VV Dokuchaeva* (2016).
114. Cuenca, R. H., Stangel, D. E. & Kelly, S. F. Soil water balance in a boreal forest. *J. Geophys. Res. Atmospheres* **102**, 29355–29365 (1997).
115. Ismail, A. Soil properties and moisture characteristics and their relationship with crop mid-day stress in the Sudan Gezira. *GeoJournal* **23**, 233–237 (1991).
116. De Boever, M., Gabriels, D., Ouessar, M. & Cornelis, W. Influence of Acacia trees on near-surface soil hydraulic properties in arid Tunisia. *Land Degrad. & Dev* **27**, 1805–1812 (2016).
117. Novak, K. D. *et al.* Investigation of soil-water properties for reclaimed oil sands, fire-disturbed, and undisturbed forested soils in Northern Alberta, Canada. Ph.D. thesis, University of Saskatchewan (2017).
118. Guzman, J. G., Ussiri, D. A. & Lal, R. Soil physical properties following conversion of a reclaimed minesoil to bioenergy crop production. *Catena* **176**, 289–295 (2019).
119. Saha, D. & Kukal, S. Soil structural stability and water retention characteristics under different land uses of degraded lower Himalayas of North-West India. *Land Degrad. & Dev* **26**, 263–271 (2015).
120. Kassaye, K. T. *et al.* Soil water content and soil temperature modeling in a vadose zone of Andosol under temperate monsoon climate. *Geoderma* **384**, 114797 (2021).
121. Seki, K. *et al.* Physical and chemical properties of soils in the fire-affected forest of East Kalimantan, Indonesia. *J. Trop. For. Sci.* 414–424 (2010).

Acknowledgements

Author contributions

S.G. processed the data and S.G. and T.H. assembled the database. S.G. and A.P. performed the data analyses. A.P. developed the 'soilhypfit' R package for fitting parametric models to SWCCs. 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. S.B., P.L., A.P, T.H. and D.O. read and approved the final version of the manuscript. P.L., A.P. and D.O. supervised the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41597-022-01481-5.

Correspondence and requests for materials should be addressed to S.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.