https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

SoilKsatDB: global soil saturated hydraulic conductivity measurements for geoscience applications

Surya Gupta[1], Tomislav Hengl[2], Peter Lehmann[1], Sara Bonetti[1], and Dani Or[1]

[1]Soil and Terrestrial Environmental Physics, Department of Environmental Systems Science, ETH, Zürich, Switzerland
[2]OpenGeoHub foundation / EnvirometriX, Wageningen, the Netherlands

Correspondence: Gupta S.
surya.gupta@usys.ethz.ch

Abstract. Saturated soil hydraulic conductivity (Ksat) is a key parameter in many hydrological and climatic modeling applications, as it controls the partitioning between precipitation, infiltration and runoff. Ksat values are primarily determined from soil textural properties and soil forming processes, and may vary over several orders of magnitude. Despite availability of Ksat datasets at catchment or regional scale, significant efforts are required to import and bind the data before it ~~could~~ be used for modeling. In this work, a total of ~~1,910 sites with~~ 13,267 Ksat measurements were assembled from published literature and other sources, standardized, and quality-checked in order to provide a global database of soil saturated hydraulic conductivity (SoilKsatDB). The SoilKsatDB covers most ~~global~~ regions, with the highest data density from the USA, followed by Europe, Asia, South America, Africa, and Australia. In addition to Ksat, other soil variables such as soil texture (11,667 measurements), bulk density (11,151 measurements), soil organic carbon (9,787 measurements), field capacity (7,389) and wilting point (7,418) are also included in the dataset. The results of using the SoilKsatDB to fit Ksat pedotransfer functions (PTFs) for temperate climatic regions and laboratory based ~~soil samples based on~~ soil properties (sand and clay content, bulk density) show that reasonably accurate models can be fitted using Random Forest (best CCC = 0.70 and CCC = 0.73 for temperate and lab based measurements, respectively). However when temperate and laboratory based Ksat PTFs are applied to soil samples from tropical climates and field measurements, respectively, the model performance is significantly lower (CCC = 0.51 for tropical and CCC = 0.13 for field samples). PTFs derived for temperate soils and laboratory measurements might not be suitable for estimating Ksat for tropical regions or field measurements, respectively. The SoilKsatDB dataset is available at https://doi.org/10.5281/zenodo.3752721 (Gupta et al., 2020) and the code used to produce the compilation is publicly available under an open data license.

1 Introduction

Soil saturated hydraulic conductivity (Ksat) describes the water movement through water saturated soils and is defined as ratio between water flux and hydraulic gradient (Amoozegar and Warrick, 1986). It is a key variable in a number of hydrological, geomorphological, and climatological applications, such as rainfall partitioning into infiltration and runoff (Vereecken et al., 2010), optimal irrigation design (Hu et al., 2015), as well as the prediction of natural hazards including catastrophic floods and

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

landslides (Batjes, 1996; Gliński et al., 2000; Zhang et al., 2018). Accurate measurements of Ksat in the laboratory and field are laborious and time consuming and most samples are taken from agricultural soils (Romano and Palladino, 2002).

Efforts to produce reliable and spatially refined datasets of hydraulic properties date back to the 1970's with the proliferation of distributed hydrologic and climatic modeling. Some of these early notable works also provided some of the basic databases (some of which are used in this study) for Australia (McKenzie et al., 2008; Forrest et al., 1985), Belgium (Vereecken et al., 2017; Cornelis et al., 2001), Brazil (Tomasella et al., 2000, 2003; Ottoni et al., 2018), France (Bruand et al., 2004), Germany (Horn et al., 1991; Krahmer et al., 1995), Hungary (Nemes, 2002), the Netherlands (Wösten et al., 2001), Poland (Glinski et al., 1991), and USA (Rawls et al., 1982). Nemes (2011) discussed the available datasets on Ksat and hydro-physical properties in detail. Collaborative efforts have resulted in the compilation of multiple databases, including the Unsaturated Soil Hydraulic Database (UNSODA) (Nemes et al., 2001), the Grenoble Catalogue of Soils (GRIZZLY) (Haverkamp et al., 1998), and the Mualem catalog (Mualem, 1976) - these however focused on soil types and not on spatially context mapping of Ksat. In an effort to provide spatial context, Jarvis et al. (2013), Rahmati et al. (2018) and Schindler and Müller (2017) published global databases for soil hydraulic and soil physical properties. Likewise, the European soil data center also started projects for generating spatially referenced databases for several countries such as SPADE (Hiederer et al., 2006) and HYPRES (Wösten et al., 2000). Since HYPRES represents only western European countries, Weynants et al. (2013) gathered the data from 18 countries and developed the European Hydropedological Data Inventory (EU-HYDI) database - this dataset is, however, not publicly available and was not included in this compilation. The datasets mentioned above cover almost all climatic zones except tropical regions, where Ksat values could be significantly different due to the strong local weathering processes (Hodnett and Tomasella, 2002). Recently, Ottoni et al. (2018) published a dataset named HYBRAS (Hydrophysical Database for Brazilian Soils) improving the coverage of South American tropical regions. In addition, Rahmati et al. (2018) recently published the Soil Water Infiltration Global database (SWIG) collecting information on Ksat for the whole globe as deduced from infiltration experiments.

The increased observation of various surface properties using satellite based imaging capability as well as the ever increasing demand for highly resolved description of surface processes require commensurate advances in Ksat representation for modern Earth System Model (ESM) applications. Despite availability of datasets at catchment or regional scale, to be able to use the various soil datasets listed above for global modeling, a significant amount of time is required to import and bind data. In addition, several existing Ksat datasets miss either coordinates of points or these have been recorded with unknown accuracy thus limiting their applications for spatial modeling. For example the SWIG dataset misses information on soil depth and assigns a single coordinate for entire watersheds. Similarly, UNSODA dataset does not provide coordinates and soil texture information for all samples. For a few locations, HYBRAS uses a different coordinate system. Taken together, these limitations highlight that, to prepare spatially referenced global Ksat datasets for large scale applications, a serious effort to compile, standardize and quality check all literature (available publicly) is often required.

The objective of the work here is to provide a new global standardized Ksat database (SoilKsatDB) that can be used for geoscience applications. To do so, a total of 13,267 Ksat measurements have been collected, standardized, and cross-checked to produce a harmonized compilation which is analysis-ready (i.e., it can directly be used for model fitting and spatial analy-

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

sis). We collected data from existing datasets and, to improve the spatial coverage in regions with sparse data, we have further conducted a literature search to include Ksat measurements in geographic areas that were not yet covered in other existing databases. In the manuscript, we first describe the data collection process and then describe methodological steps used to spatially reference, filter, and standardize existing datasets. As an illustrative application of the dataset we derive pedotransfer functions (PTFs) for different regions and measurement methods and discuss their transferability to other regions and measurement methodologies. We fully document all importing, standardization and binding steps using R environment for statistical computing (R Core Team, 2013), so that we can collect feedback from other researchers and increase the speed of further updates and improvements. The newly created data set (SoilKsatDB) can be accessed via https://doi.org/10.5281/zenodo.3752721 and directly used to test various Machine Learning algorithms (Casalicchio et al., 2017).

2 Methods and materials

2.1 Data sources

To locate and obtain all compatible datasets for compilation, a literature search was conducted using different search engines, including Science Direct (https://www.sciencedirect.com/), Google Scholar (https://scholar.google.com/) and Scopus (https://www.scopus.com). We searched soil hydraulic conductivity datasets using keywords such as *"saturated hydraulic conductivity database"*, *"Ksat"*, and similar. The collected datasets are listed in Table 1 together with number of Ksat observations for each study, and can be classified into three main categories, namely: i) Existing datasets (in forms of tables) published and archived with a DOI in a peer-review publication; ii) legacy datasets in paper/document format (e.g.,legacy reports, PhD theses, and scientific studies), iii) on-line materials.

Existing datasets include published datasets such as HYBRAS (Ottoni et al., 2018), UNSODA (Nemes et al., 2001), SWIG (Rahmati et al., 2018), and the soil hydraulic properties over the Tibetan Plateau (Zhao et al., 2018), from which we extracted the required information as described in Table 2a. The major challenge with making the existing datasets compatible for binding (standardization, removing redundancy), was to obtain the locations for a particular sample as well as the corresponding measurement depths. For instance, the UNSODA database completely lacks geographical locations. To fill the gaps and make the data suitable also for spatial analysis, we used Google Earth to find the coordinates based on the given location (generally an address or a location name). We separated the data based on laboratory and field measurements and we computed sand, silt and clay contents based on the algorithm described in Nemes et al. (2001). We further note that, in some datasets, the coordinates were missing or reported in diverse coordinate systems. For example, in the HYBRAS database, the locations needed to be converted from UTM to a decimal degrees. In the SWIG database, the information related to location (coordinates for each point), soil depth and measurement method (laboratory or field) was completely missing, so we went through each publication referenced in Rahmati et al. (2018) (except the unpublished literature) and added coordinates and applied the necessary conversions.

In the case of legacy datasets (paper or document format, data from journals, theses, and legacy reports with and without peer-reviewed publications), we invested a significant effort to digitize tabular data, clean it and make it analysis-ready. In some

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

Table 1. List of reference articles and digitized Ksat datasets, and number of points (N) per data set used to generate the new SoilKsatDB product.

Reference	*N*	Reference	*N*	Reference	*N*
Rycroft et al. (1975)	1	Abagandura et al. (2017)	3	Jabro (1992)	18
Waddington and Roulet (1997)	1	Habel (2013)	3	Greenwood and Buttle (2014)	18
Takahashi (1997)	1	Nyman et al. (2011)	3	Wang et al. (2008)	19
Katimon and Hassan (1997)	1	Habel (2013)	3	Deshmukh et al. (2014)	19
El-Shafei et al. (1994)	1	Bhattacharyya et al. (2006)	4	Price et al. (2010)	20
Lopez et al. (2015)	1	Lopes et al. (2020)	4	Bonsu and Masopeh (1996)	24
Kramarenko et al. (2019)	1	Yasin and Yulnafatmawita (2018)	4	Bambra (2016)	24
Zakaria (1992)	1	Daniel et al. (2017)	6	Verburg et al. (2001)	26
Ramli (1999)	1	Anapalli et al. (2005)	7	Southard and Buol (1988)	27
Singh et al. (2011)	1	Arend (1941)	7	Chang (2010)	30
Campbell et al. (1977)	1	Helbig et al. (2013)	7	Yao et al. (2013)	33
Chief et al. (2008)	1	Gwenzi et al. (2011)	7	Becker et al. (2018)	34
Conedera et al. (2003)	1	Päivänen et al. (1973)	9	Baird et al. (2017)	50
Ebel et al. (2012)	1	Mahapatra and Jha (2019)	9	Keisling (1974)	56
Ferreira et al. (2005)	1	Amer et al. (2009)	9	Rahimy (2011)	56
Imeson et al. (1992)	1	Vogeler et al. (2019)	10	Hao et al. (2019)	57
Johansen et al. (2001)	1	Singh et al. (2006)	10	Kanemasu (1994)	60
Lamara and Derriche (2008)	1	Kelly et al. (2014)	10	Tete-Mensah (1993)	60
Parks and Cundy (1989)	1	Elnaggar (2017)	11	Zhao et al. (2018)	65
Ravi et al. (2017)	1	Ganiyu et al. (2018)	12	Hinton (2016)	77
Smettem and Ross (1992)	1	Cisneros et al. (1999)	12	Vieira and Fernandes (2004)	86
Helbig et al. (2013)	2	Niemeyer et al. (2014)	12	Houghton (2011)	88
Boike et al. (1998)	2	Sharratt (1990)	14	Tian et al. (2017)	91
Andrade (1971)	2	Habecker et al. (1990)	14	Li et al. (2017)	108
Beyer et al. (2015)	2	Nielsen et al. (1973)	14	Forrest et al. (1985)	120
Blake et al. (2010)	2	Robbins (1977)	15	Richard and Luescher (1987)	121
Bonell and Williams (1986)	2	Sonneveld et al. (2005)	15	Sanzeni et al. (2013)	127
Kutiel et al. (1995)	2	Quinton et al. (2008)	16	Vereecken et al. (2017)	145
Martin and Moody (2001)	2	Simmons (2014)	16	Coelho (1974)	177
Mott et al. (1979)	2	Ouattara (1977)	17	Kool et al. (1986)	240
Rab (1996)	2	Hardie et al. (2011)	17	Nemes et al. (2001)	283
Soracco et al. (2010)	2	Baird (1997)	17	Ottoni et al. (2018)	326
Varela et al. (2015)	2	Kirby et al. (2001)	17	Rahmati et al. (2018)	3637
Sayok et al. (2007)	3	Yoon (2009)	18	Grunwald (2020)	6532

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

cases we had to convert PDF documents to Microsoft Word files, after that to tabular data. Some documents had to be digitized manually due to the low resolution of PDFs. After the digitization process, all data values were cross-checked one more time with the original PDFs to avoid any artifacts or gross error in the final database.

Two datasets were also collected directly from project websites that might be peer reviewed such as the NASA project based on hydraulic and thermal conductivity (retrieved from https://daac.ornl.gov/FIFE/guides/Soil_Hydraulic_Conductivity_Data.html and described in Kanemasu (1994)) and the Florida database from Grunwald (2020).

Besides these, there are many locations, such as desert dunes, peatlands, frozen soils, ~~and similar, in the world~~, where very few data of Ksat were available publicly. Because it is essential for global modeling to provide some values or range to reduce the uncertainty in the spatial maps, we ~~have~~ also intensively searched for these areas and found several ~~minor~~ studies providing Ksat values in these locations. We then digitized the Ksat values from these studies (shown either in bar charts and line plots), georeferenced the maps where necessary, and then converted the data into tabular form. All these datasets are also listed in Table 1.

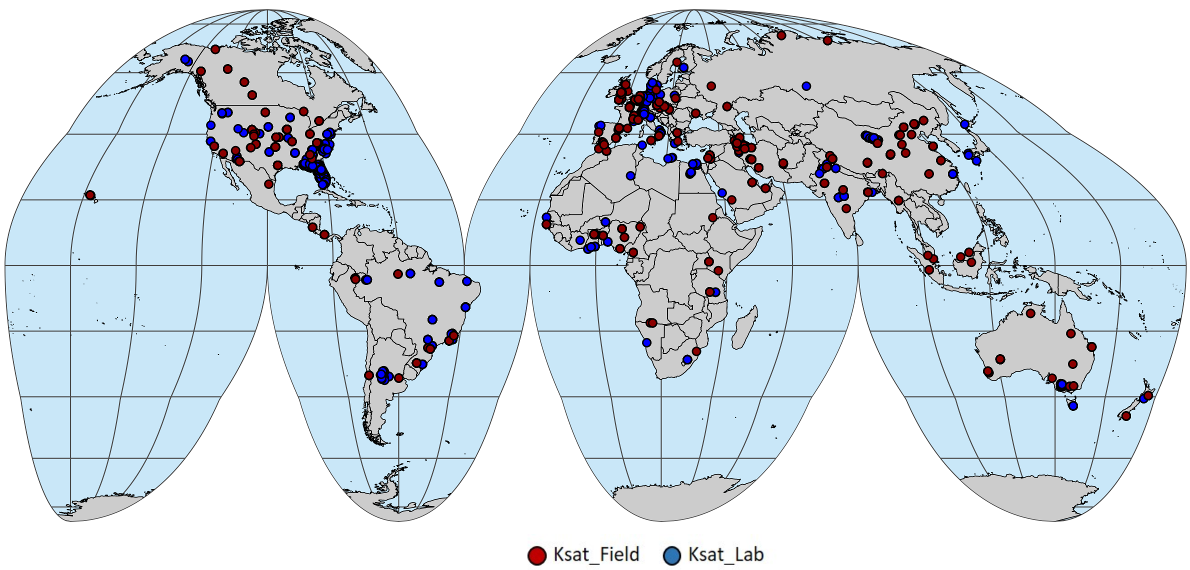

Figure 1. Spatial distribution of Ksat points (red and blue for field and laboratoratory measurements, respectively) in the SoilKsatDB. A total of 1,910 spatial locations are on this map.

2.2 Georeferencing Ksat values

Georeferencing of Ksat measurements is important for using data for local, regional or global spatial modeling. Once georeferenced, points can be directly used in hydrological and land surface models. Although many studies provided the information of spatial locations, ~~however,~~ the studies conducted in the 70's and 80's only provided the name of the locations and approximate distance from the exact location. Therefore, we extracted the latitude and longitude of the location using Google maps for some

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

datasets (which did not provide the spatial locations). Most of the studies we digitized provide maps or sketches with locations of the points. 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. (2001)) provided the names of the places, and hence we used Google Earth to obtain the coordinates. We estimate that the spatial location accuracy of these points is roughly between 0 to 5 km. Similarly, spatial maps in jpg format (e.g. Becker et al. (2018)) were geo-referenced with 100–500 m location accuracy. In contrast, few studies (e.g. Yoon (2009)) provided the extract location of the sampling with assumed location accuracy of 10–20 m.

Table 2a. Description and units of some key variables listed in the database. The complete list can be found in the link to the data base (https://doi.org/10.5281/zenodo.3752721) in the readme-file. We used the same codes adopted in the National Cooperative Soil Survey (NCSS) Soil Characterization Database (National Cooperative Soil Survey, 2016).

Headers	Description	Dimension
`site_key`	Data set identifier	—
`longitude_decimal_degrees`	Ranges up to +180 degrees down to -180 degrees	Decimal degree
`latitude_decimal_degrees`	Ranges up to +90 degrees down to -90 degrees	Decimal degree
`hzn_top`	Top of soil sample	cm
`hzn_bot`	Bottom of soil sample	cm
`db_od`	Bulk density	g cm^{-3}
`w6clod`	Soil water content at 6 kPa	vol %
`w10clod`	Soil water content at 10 kPa	vol %
`w3cld`	Soil water content at 33 kPa (field capacity)	vol %
`w15l2`	Soil water content at 1500 kPa (wilting point)	vol %
`tex_psda`	Soil texture classes based on USDA	—
`clay_tot_psa`	Mass of soil particles, < 0.002 mm	%
`silt_tot_psa`	Mass of soil particles, > 0.002 and < 0.05 mm	%
`sand_tot_psa`	Mass of soil particle, > 0.05 and < 2 mm	%
`oc`	Soil organic carbon content	%
`ph_h2o`	Soil acidity	—
`Ksat_lab`	Soil saturated hydraulic conductivity from lab	cm day^{-1}
`Ksat_field`	Soil saturated hydraulic conductivity from field	cm day^{-1}
`source_db`	Sources of the datasets	—
`confidence_degree`	Reliability on the data set based on spatial locations	—
`location_id`	Combination of latitude and logitude	—

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Table 2b. Example of Ksat database structure with key variables (from left to right: reference, longitudinal and latitudinal coordinates (decimal degree), top and bottom of soil sample (cm), bulk density (g cm^{-3}), soil textural class, clay, silt and sand content (%) and saturated hydraulic conductivity measured in lab or field (cm day^{-1}). NA is 'no value'). Note that the titles of the columns are explained in Table 2a.

site_key	longitude_decimal_degrees	latitude_decimal_degrees	hzn_top	hzn_bot	db_od	tex_psda	clay_tot_psa	silt_tot_psa	sand_tot_psa	ksat_lab	ksat_field
Saseendran_2005	-103.15	40.15	15	30	1.33	Loam	23.4	44.3	32.3	232.08	NA
Saseendran_2005	-103.15	40.15	30	60	1.32	Loam	22.3	40.7	37.0	232.08	NA
Saseendran_2005	-103.15	40.15	60	90	1.36	Loam	17.6	36.7	45.7	337.92	NA
Saseendran_2005	-103.15	40.15	90	120	1.40	Loam	12.0	42.3	45.7	284.88	NA
Saseendran_2005	-103.15	40.15	120	150	1.42	Loam	10.0	41.7	48.3	259.20	NA
Saseendran_2005	-103.15	40.15	150	180	1.42	Loam	10.0	41.7	48.3	259.20	NA
Becker_2018	-110.13	31.73	0	15	NA	Sandy loam	NA	NA	NA	NA	26.40
Becker_2018	-110.09	31.72	0	15	NA	Sandy loam	NA	NA	NA	NA	27.84
Becker_2018	-110.09	31.69	0	15	NA	Sandy loam	NA	NA	NA	NA	21.60
Becker_2018	-110.05	31.74	0	15	NA	Loam	NA	NA	NA	NA	23.76
Becker_2018	-110.04	31.72	0	15	NA	Sandy loam	NA	NA	NA	NA	39.12
Becker_2018	-110.04	31.69	0	15	NA	Sand	NA	NA	NA	NA	102.96

Table 3. Confidence weights provided to each sample based on location accuracy and method used: LM = laboratory method, FM = field method.

Location errors (LM)	Confidence index	Location errors (FM)	Confidence index
0 – 100 m	1	0 – 100 m	3
100 – 250 m	3	100 – 250 m	6
250 – 500 m	5	250 – 500 m	9
0.5 – 1 km	7	0.5 – 1 km	12
1 – 5 km	9	1 – 5 km	15
5 – 10 km	20	5 – 10 km	30
>10 km	40	>10 km	40

2.3 Standardization and quality assignment

The database was cleaned on the basis of highest and lowest values of saturated hydraulic conductivity. In SWIG database, some values of Ksat were less than 10^{-14} m/day, ~~that~~ seem unreasonable, so they were not included in the database. All datasets were cross-checked to avoid redundancy. For example, UNSODA data consist of Vereecken et al. (2017) and Richard

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

Table 4. Mean values of soil hydro-physical properties for each soil texture class. The number of samples (N) is given in parenthesis under each soil variable for each soil texture classes. N values marked with * correspond to undefined soil texture classes. BD = bulk density (g/cm^3), OC = organic carbon (%), FC = field capacity (% vol), WP = wilting point (% vol), $Ksat_l$, $Ksat_f$ = laboratory and field Ksat (cm/day). For Ksat the geometric mean is reported (due to the sensitivity on few extreme values). For all other properties the arithmetic mean is provided.

Texture Classes	Clay (N)	Silt (N)	Sand (N)	BD (N)	OC (N)	FC (N)	WP (N)	$Ksat_l$ (N)	$Ksat_f$ (N)
Clay	56.3	23.8	19.9	1.27	1.98	45.0	30.9	8.17	110.33
	(835)	(835)	(835)	(609)	(454)	(452)	(454)	(507)	(331)
Clay Loam	31.4	38.6	30.0	1.27	2.49	39.7	24.1	12.25	59.96
	(543)	(543)	(543)	(382)	(360)	(76)	(76)	(139)	(423)
Loam	19.1	39.3	41.6	1.28	2.16	32.6	14.0	43.49	35.59
	(699)	(699)	(699)	(607)	(561)	(102)	(106)	(206)	(504)
Loamy Sand	7.5	8.5	84.0	1.55	1.14	17.5	6.6	96.49	127.06
	(742)	(742)	(742)	(712)	(680)	(558)	(592)	(633)	(100)
Sand	2.2	3.1	94.7	1.51	0.62	8.2	2.5	501.08	252.31
	(4526)	(4526)	(4526)	(4450)	(4193)	(4077)	(4074)	(4218)	(320)
Sandy Clay	39.3	8.1	52.6	1.53	0.23	34.7	23.4	14.02	—
	(179)	(179)	(179)	(166)	(143)	(161)	(161)	(175)	(4)
Sandy Clay Loam	26.3	12.2	61.5	1.54	1.25	28.9	17.3	19.28	14.23
	(1149)	(1149)	(1149)	(941)	(959)	(806)	(760)	(869)	(288)
Sandy Loam	13.5	16.7	69.8	1.50	1.33	24.2	11.0	34.53	85.31
	(1610)	(1610)	(1610)	(1488)	(1352)	(815)	(801)	(999)	(636)
Silt	7.5	84.7	7.8	1.17	1.65	51.43	7.5	13.27	—
	(25)	(25)	(25)	(19)	(11)	(11)	(751)	(25)	
Silt Loam	15.2	67.0	17.8	1.34	3.65	35.3	15.6	5.76	43.64
	(813)	(813)	(813)	(633)	(500)	(148)	(138)	(444)	(383)
Silty Clay	45.5	45.5	10.0	1.18	3.83	49.9	30.2	1.22	217.60
	(181)	(181)	(181)	(175)	(116)	(46)	(46)	(69)	[illegible]
Silty Clay loam	33.1	57.2	9.7	1.24	2.67	46.2	23.9	1.45	49.10
	(333)	(333)	(333)	(282)	(226)	(57)	(56)	(110)	(232)
Total	**11,635**	**11,635**	**11,635**	**10,464**	**9,555**	**7,340**	**7,275**	**8,394**	**3,333**
	(32*)	(32*)	(32*)	(687*)	(232*)	(49*)	(143*)	(413*)	(1,154*)

and Luescher (1987) datasets and SWIG database used Zhao et al. (2018). Hence we removed these datasets from UNSODA

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

and SWIG database and used the original source datasets. Moreover, in the SWIG database, soil depth information was not available, so we assumed that data were obtained from field measurements and assumed it was obtained at a depth of 0–20 cm.

To describe the accuracy and reliability of each dataset, a quality flag (or confidence degree) was assigned to each data set based on (a) positional accuracy of the site, and (b) methodology used (i.e. only differentiating between field and laboratory measurements, not accounting for different laboratory and field methods) for measuring Ksat. Here, we separated each study based on the measurement of Ksat and subjectively selected a range from 1 to 50 (i.e., 1 = highly accurate, 50 = least accurate) to describe the level of accuracy of each dataset. Table 3 shows the allocation of different weights for laboratory and field methods. Here, we assigned a slightly higher confidence to laboratory methods (compared to field ones) because the analyzed soil depth is well defined in lab samples but unclear in field infiltration measurements. In contrast, field methods are representative of larger areas. The other main difference is the entrance of atmospheric air into the soil. It is, in fact, more difficult in field methods to reach a saturated state because of the interference of atmospheric air and fast infiltration velocities at beginning of the process (Faybishenko, 1997). In addition, a higher confidence was assigned to measurements with higher spatial accuracy. For example, laboratory measurements at high spatial accuracy were given the highest confidence degree. Among these, Forrest et al. (1985) and/or Ottoni et al. (2018) measured Ksat in the laboratory and provided detailed site coordinates, thus we assigned a confidence degree of 1 (i.e., highly accurate). Zhao et al. (2018) measured Ksat using field methods and provided the exact locations of the field sites thus we assigned 3 as a confidence degree. If the spatial accuracy was be between 100–250 m, then we would have assigned a value of 6 (see Table 3 for more details). After data extraction from literature (and data bases), geo-referencing and standardization, all information was collected in tabulated form in the new data base SoilKsatDB (https://doi.org/10.5281/zenodo.3752721). The database consists of a 38 columns (various sample properties) and 13,268 rows (for column titles and 13,267 samples). An excerpt of the data base with some key properties is shown in Table 2b.

2.4 Statistical modeling

The PTF models were fitted using multivariate polynomial regression (MPR) and random forest (RF) in the R environment for statistical computing (R Core Team, 2013). We tested fitting the MPR model for Ksat values as function of primary soil properties. For 15% of samples with information on bulk density and soil texture, the value of organic content (OC) was not reported. Therefore, we expressed the PTF for Ksat as function of bulk density, clay and sand content (without OC). To test if PTFs for different climatic regions or measurement types are different, we have split fitting the PTF using (1) temperate-climate soil samples (including both laboratory and field measurements), and (2) laboratory based measured samples (including all climates). To develop PTFs with temperate climate soil samples, the dataset (total 13,267 points) was divided based on climatic regions (temperate, tropical, boreal, and arid) to account for differences in climate and related weathering processes (Hodnett and Tomasella, 2002). A total of 8,333 temperate-climate soil samples were used that contain information on sand, clay, and bulk density. The data set was randomly divided into training (6,666 samples, 80%) and testing dataset (1,667 samples, 20%). Likewise, MPR was also applied to develop a PTF for laboratory measurements. In a second application, the dataset (total 13,267) was divided into laboratory and field based soil Ksat samples. The laboratory dataset (8,055 soil samples) was

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

used for training (6,444) and testing (1,611) following the same method as used for the temperate climate PTF (i.e., 80% for training and 20% for testing).

The following equation was fitted using MPR:

$$\log(\text{Ksat}) = b_0 + b_1 \cdot \text{BD} + b_2 \cdot \text{BD}^2 + b_3 \cdot \text{CL} + b_4 \cdot \text{BD} \cdot \text{CL} + b_5 \cdot \text{CL}^2 + b_6 \cdot \text{SA} + b_7 \cdot \text{BD} \cdot \text{SA} + b_8 \cdot \text{CL} \cdot \text{SA} + b_9 \cdot \text{SA}^2 \quad (1)$$

where Ksat is in cm/day, clay (CL) and sand (SA) are expressed in % and bulk density (BD) is in g/cm^3.

Likewise, PTFs were also developed using a RF algorithm both for temperate-climate and laboratory based soil samples. The same soil variables (sand, clay and, bulk density) were fitted with Ksat values and we used the same number of points as for MPR for training and testing the models (i.e., 80% and 20%, respectively). The *'ranger'* package (Wright and Ziegler, 2015) was implemented to process the large data. The PTFs developed for temperate regions and for laboratory data were then applied to estimate Ksat in tropical climate (1,122 samples) or field measurements (2,396 samples), respectively. Root mean square error (RMSE) and concordance correlation coefficient (CCC) (Lawrence and Lin, 1989) were computed to assess the accuracy of the models.

3 Results

3.1 Data coverage

Based on the intensive literature search and data collection, we have assembled a total of 13,267 values of Ksat from 1,910 sites across the globe. Figure 1 shows the global distribution of the sites locations used in this study. Most data originate from the USA, followed by Europe, Asia, South America, Africa, and Australia. The points are often spatially clustered with the biggest cluster of points (1,103 site locations with 6,532 Ksat values) in Florida (Grunwald, 2020). Ksat data include 4,460 values from field measurement and 8,807 values from laboratory measurements. In particular, different types of infiltrometers were used for Ksat field measurements, whereas constant or falling head methods were predominantly used in laboratory analyses.

Out of the 13,267 Ksat measurements, 11,667, 11,151, 9,787, 7,389 and 7,418 points had information on soil texture, bulk density, organic carbon, field capacity and wilting point, respectively, and 8,947 samples had information for all soil basic properties (bulk density, soil texture and organic carbon) as shown in Figure 2.

3.2 Statistical properties

The distribution of soil samples based on soil texture classes is shown on the USDA soil texture triangle in Figure 3a. The database covers all textural classes, with a high clustering in sandy soils due to the numerous samples from Florida. The violin distribution plot in Figure 3 shows the range of Ksat values for the different databases. Most of the datasets showed Ksat values between $\approx 10^{-2}$ and $10^{2.5}$ cm/day, with a wider range of Ksat values observed in measurements from theses and reports

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

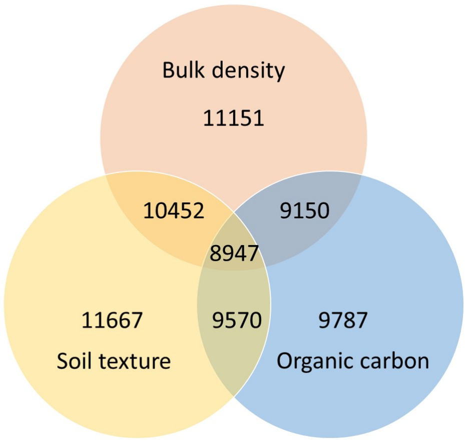

Figure 2. Venn diagram illustrating the number of samples containing information on bulk density, soil texture, and organic carbon. Out of 13,267 samples, 11,151, 11,667 and 9,787 samples have values of bulk density, soil texture and organic carbon, respectively. Furthermore, 10,452, 9,150 and 9,570 samples have information of bulk density and soil texture, bulk density and organic carbon and soil texture and organic carbon, respectively. 8,947 samples have information of all three soil properties

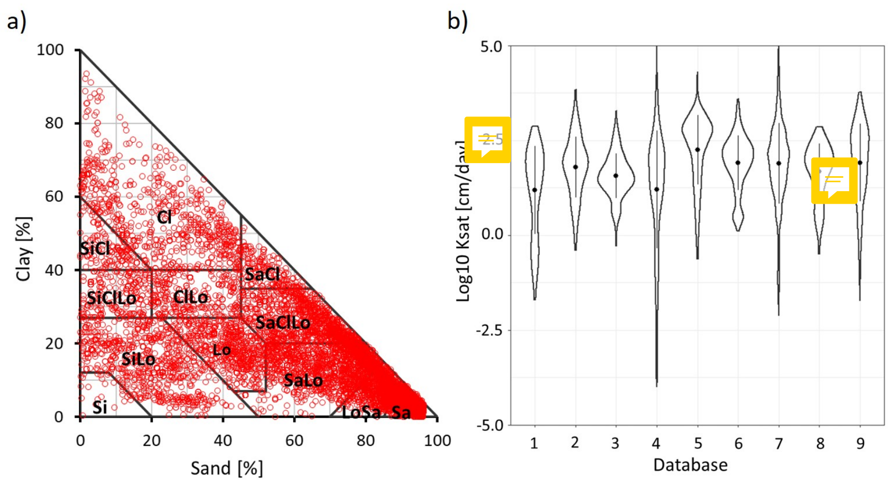

Figure 3. Distribution of collected Ksat values: (a) distribution of soil samples on the USDA soil texture triangle. The bulk of the samples were from Florida (cluster of sandy soil samples). The Ksat values covers all soil textural classes and only few samples belong to the silt textural class. The histogram plot (b) represents the range of Ksat values spanned by each data source. The dot represents the mean value, and the line represents the standard deviation for each data set. The numbers 1–9 refer to different sources and databases: 1 = Australia (Forrest et al., 1985), 2 = Belgium (Vereecken et al., 2017), 3 = China (Tian et al., 2017; Li et al., 2017), 4 = extracted from thesis and reports (see Table 1), 5 = Florida (Grunwald, 2020), 6 = HYBRAS (Ottoni et al., 2018), 7 = SWIG (Rahmati et al., 2018), 8 = Tibetan Plateau (Zhao et al., 2018), 9 = UNSODA (Nemes et al., 2001).

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

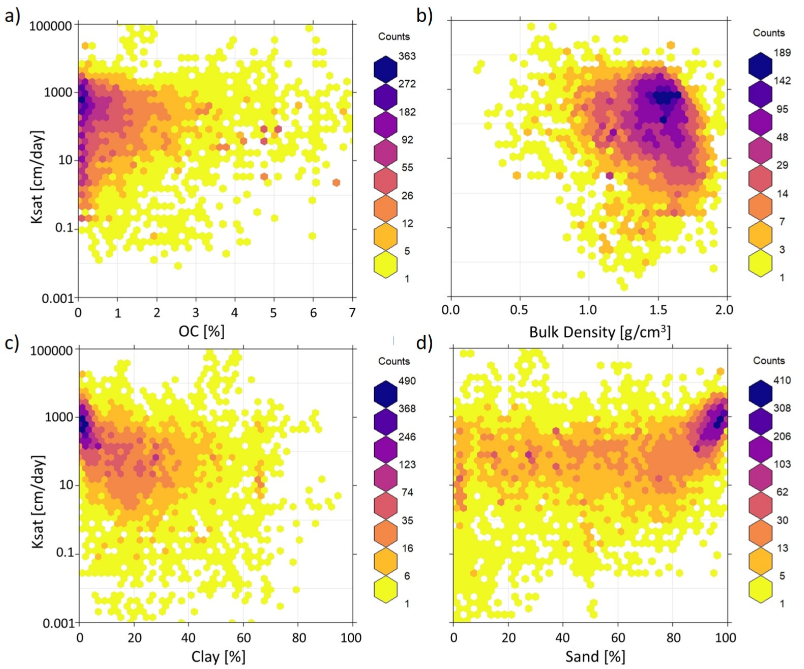

Figure 4. Partial correlation between Ksat and a) organic carbon (%), b) bulk density (g/cm^3), c) clay (%) and d) sand (%).

(including studies with extreme values from sandy desert soils and low conductive clay soils) and from the SWIG database (databases 4 and 7 in Figure 3b, respectively).

Average values of Ksat and other hydro-physical properties are shown in Table 4. Higher average organic carbon and bulk density values were observed in clayey and loamy soils compared to sandy soils. Ksat values obtained from field measurements were on average higher (depending on the type of instrument used) than those obtained from laboratory samples. Particularly, for the clay texture class much lower Ksat values were observed for laboratory (mean Ksat $\approx$ 8 cm/day) compared to field (mean Ksat $\approx$ 110 cm/day) measurements.

3.3 PTFs derivation

As a test application of SoilKsatDB, PTFs were derived for temperate climate region and laboratory based samples using basic soil properties as covariates. Such basic soil properties (i.e., clay and sand fraction, organic carbon, and bulk density) are plotted against Ksat in Figure 4, showing that Ksat decreases with increasing clay content and bulk density, and increases with sand content. The observed correlation between these soil properties and Ksat motivates their use as key variables for the estimation of PTFs. Due to limiting data availability (15% of samples without OC information) and the poor correlation between OC and Ksat (Figure 4), we built the PTF for Ksat using bulk density, clay and sand content (without OC).

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Table 5. Pedotransfer function (Eq. (1)) coefficients obtained for temperate and laboratory soil measurements.

Coefficient	Value (temp.)	Value (lab.)
b_0	2.17	1.44
b_1	0.9387	2.053
b_2	−0.8026	−1.256
b_3	0.0037	−0.0533
b_4	−0.017	−0.000051
b_5	0.000015	0.00055
b_6	0.0025	0.0079
b_7	0.00086	−0.00080
b_8	−0.00025	0.000043
b_9	0.000073	0.000052

Coefficients of Eq. (1) were fitted to values obtained from i) temperate sites and from ii) laboratory measurements. The fitted model coefficients are listed in Table 5. The fitting procedure provided R^2 of 0.47 and 0.53 for temperate and laboratory values, respectively. Validation of the fitted equations against the testing data set provided CCC and RMSE for the temperate and laboratory based predictions equal to 0.64 (CCC, temperate), 0.71 (RMSE, temperate) and 0.70 (CCC, lab), and 0.67 (RMSE, lab), respectively.

Results obtained from RF modeling using the same number of data points and the same independent variables (sand, clay, and bulk density) show a better accuracy. Specifically, the RF model performance based on CCC and RMSE was 0.69 (CCC, temperate region) and 0.70 (RMSE, temperate region), 0.73 (CCC, lab measurements), and 0.66 (RMSE, lab measurements), respectively.

Figure 5 (b and d) and Figure 6 (b and d) indicates that both models underestimated Ksat for both tropical and field measured soil samples. In fact, for the RF model we obtained CCC and RMSE values equal to 0.51 and 0.90 for tropical and 0.13 and 1.1 for field measured samples, whereas CCC and RMSE values obtained from MPR were equal to 0.53 and 0.83, and 0.16 and 1.0 for tropical and field measurements, respectively.

4 Discussion

4.1 Laboratory vs field estimated Ksat: effect of soil structure

Results showed that Ksat values were, on average, higher for samples measured using field methods compared to laboratory methods for most soil texture classes (Table 4). Figure 7 further illustrates the higher range of Ksat values obtained for finer texture soils (clay and loam) compared to coarser soils (sand). The difference in laboratory and field based Ksat values and higher range of Ksat values in fine textured soil is probably related to the effect of biologically-induced soil structure that might be neglected in laboratory measurements. In other words, variability in the Ksat values depends on the consideration of

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

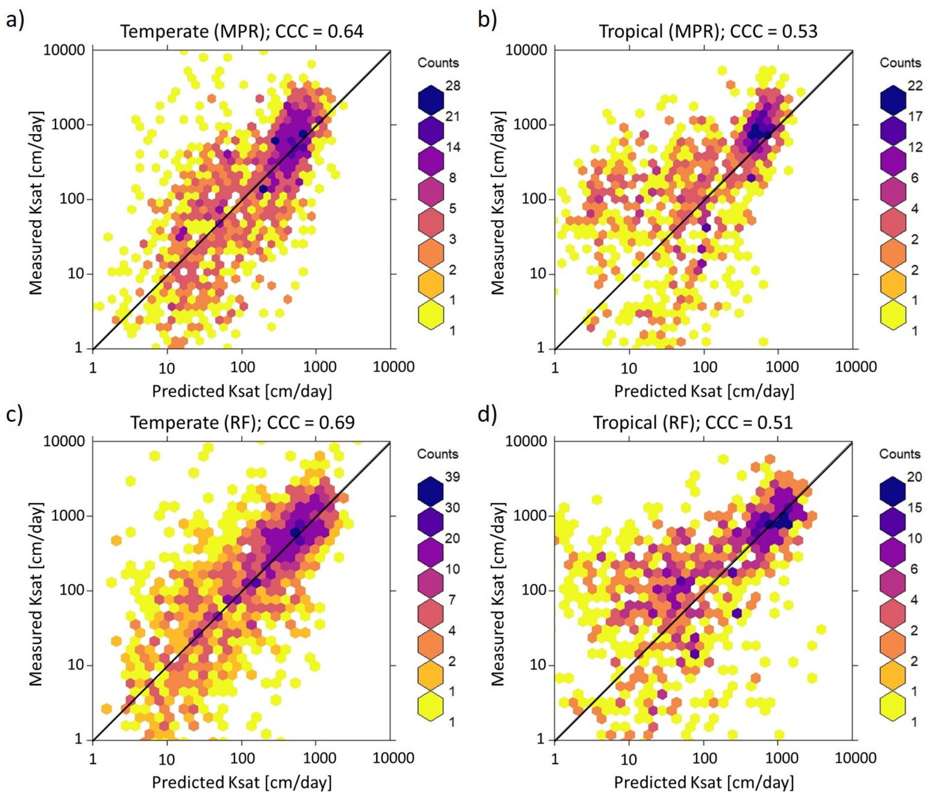

Figure 5. Correlation between observed and predicted Ksat values obtained from (a, b) multivariate polynomial regression (MPR) and (c, d) random forest (RF) models. Models were obtained by fitting 6,666 ~~temperate climate~~ training points and tested on temperate (1,667 samples, panels a, c) and tropical testing points (1,122 samples, panels b, d). The density of point pairs for Ksat is shown in logarithmic scale. CCC is the concordance correlation coefficient. PTFs showed reasonable agreement for both MPR (CCC = 0.64) and RF (CCC = 0.69) algorithms ~~with~~ temperate soil samples, ~~while~~ lower CCC values were obtained for tropical soil samples (0.53 and 0.51 for MPR and RF, respectively). PTFs determined for temperate regions cannot be easily transferred to tropical regions due to different soil forming processes.

soil macropores by the measurement methods. Soil macropores change the pore size distribution and subsequently affect Ksat values (Tuller and Or, 2002). Such an effect is likely to be neglected more in laboratory measurements compared to field ones. Mohanty et al. (1994), for example, compared the three field methods and one laboratory method and found that the sample size affects the measurement of Ksat and maximum variability observed in the Ksat values at shallow depth might be due to the presence and absence of open-ended pores. Likewise, Braud et al. (2017) used three field methods for Ksat measurements and found significant variation between these methods of measurements.

As shown in Figure 6 Ksat values measured in the field were underestimated by PTFs derived from laboratory measurements. The omission of soil structures in many laboratory samples limits the possibility to properly reproduce field observations that are likely to be more affected by the presence of biopores (Fatichi et al., 2020).

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Open Access Earth System Science Data Discussions

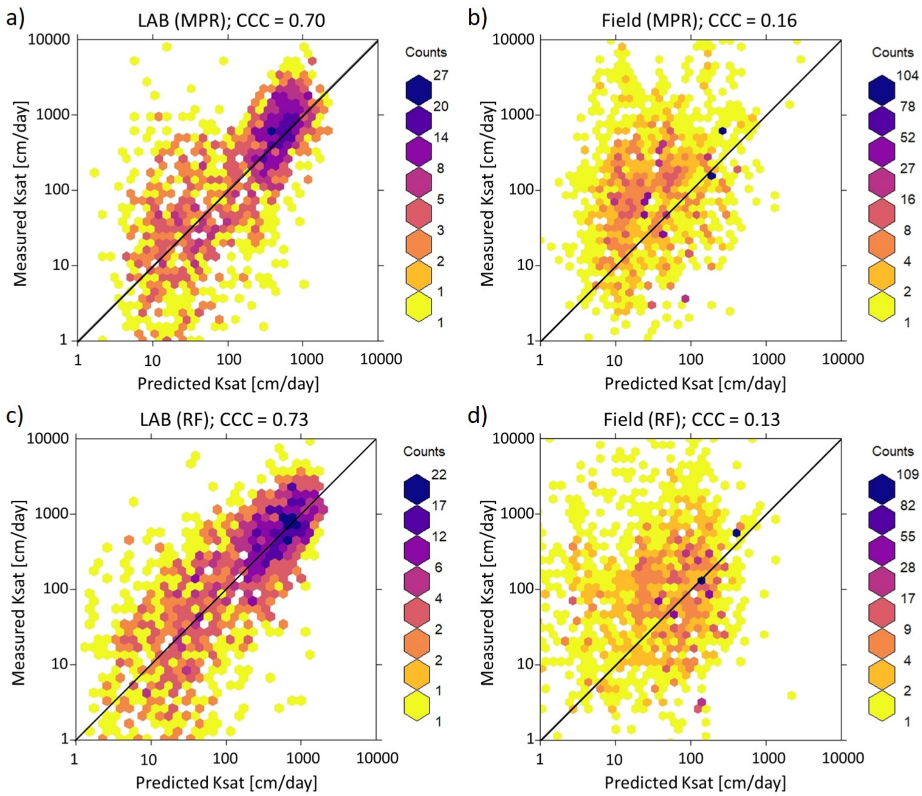

Figure 6. The correlation between observed and predicted Ksat values obtained from (a, b) multivariate polynomial regression (MPR) and (c, d) random forest (RF) models. The model was fitted using laboratory measurements and tested on both laboratory (a, c) and field (b, d) measurements. Results showed reasonable agreement (CCC = 0.70, CCC = 0.73) using both algorithms (RF and MPR) for laboratory measurements, but low CCC (0.16, 0.13) for field measurements. PTFs developed based on laboratory measurements do not provide accurate estimates of Ksat measured in the field.

4.2 Temperate vs tropical soils: effect of clay mineralogy

Results showed that PTFs obtained for temperate soils performed poorly for tropical soils (Figure 5), with Ksat being underestimated by the temperate-based PTFs. This result is in agreement with Tomasella et al. (2000) who derived PTFs using data from tropical Brazilian soils, which did not properly capture observations in temperate soils. We argue that the significant differences in the models fitted for tropical and temperate soils are due to the differences in the soil-forming processes defining the clay type and mineralogy. In fact, Oxisols (highly weathered clay minerals in tropical regions) are turned into inactive (non-swelling) clay minerals as a result of high rainfall and temperatures. On the other hand, in the temperate regions, active (smectite) and moderately active clay minerals (illite) are the dominant clay minerals. These swelling clay minerals retain the water within internal structures with very low hydraulic conductivity. Therefore, such a difference in clay mineralogy is likely responsible for the underestimation of Ksat in tropical soils from PTFs obtained in temperate ones.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

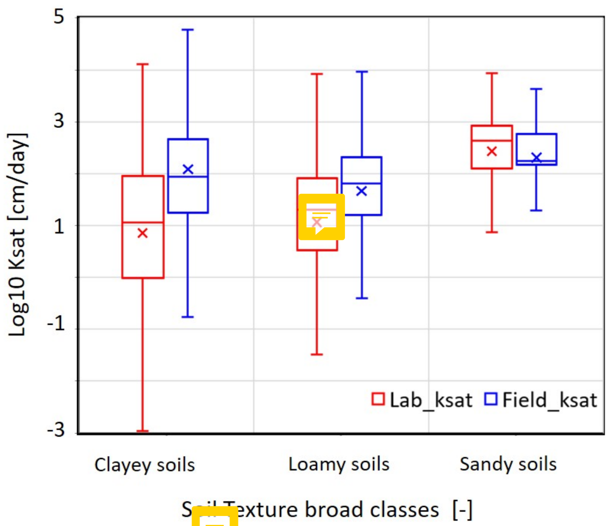

Figure 7. The distribution of Ksat values based on laboratory and field methods. Field measurements gave higher values than laboratory ones in clayey and loamy soils likely due to the effect of structure.

4.3 Limitations of SoilKsatDB

We have put an effort to collect laboratory and field data from all parts of the globe. However, we acknowledge that there are still gaps in some regions such as Russia and higher northern latitudes in general, which may produce uncertainties in Ksat estimations in such regions. The SoilKsatDB could also be of limited use for fine-resolution applications because many data points were characterized by limited spatial accuracy and missing soil depth information. Specifically, the spatial accuracy of many points is between tens of meters to several kilometers (see the methodology sections regarding the extraction of the spatial locations using Google Earth). In addition, in the SWIG database the soil depth and measurement method information were not provided, and often one location was used to represent an entire watershed. We tried to revisit each publication and extract the most accurate coordinates of assumed sampling locations and we assumed that most of the samples belonged to the field measurements as authors used different infiltrometers to compute Ksat. Hence, there might be few points in our SoilKsatDB that belong to laboratory measurements and that we have incorrectly assigned to field measurements.

For each measurement, a confidence index (1 = highest, 50 = lowest) was assigned based on the sampling location accuracy and measurement technique (laboratory or field), which can be used as a weight or probability argument in Machine Learning. We acknowledge that this was a rather subjective decision and a more objective way to assign weights would be to use the actual measurement and spatial positioning errors. Because these were not available for most of the datasets, we have opted for the definition of a confidence index estimated from the available documentation.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

4.4 Further developments

We envisage several further developments of this database. The advancement in remote sensing technology opens the doors to link the hydraulic properties with global environmental features. Using satellite-based maps of environmental properties enables ~~to~~ incorporate local information on vegetation, climate, and topography for specific areas, which are often ignored by basic PTFs. For example, Sharma et al. (2006) developed PTFs using environmental variables such as topography and vegetation and concluded that these attributes, at finer spatial scales, were useful to capture the observed variations within the soil mapping units. Likewise, Szabó et al. (2019) used the random forest machine learning algorithm for mapping soil hydraulic properties and incorporated local environmental variable information.

5 Data availability

All collected data and related soil characteristics are provided online for reference and are available at https://doi.org/10.5281/zenodo.3752721 (Gupta et al., 2020).

6 Summary and conclusions

We prepared a comprehensive global compilation of measured Ksat training point data ($N = 13,267$) by importing, quality controlling, and standardizing tabular data from existing soil profile databases and legacy reports.

The produced SoilKsatDB covers a broad range of soil types and climatic regions and hence is ~~applicable~~ for global soil modeling. A higher variation in Ksat values was observed in fine-textured soil compared to coarse-textured soils, possibly indicating the effect of soil structure on Ksat. Moreover, Ksat values obtained from field measurements were generally higher than those from laboratory measurements, likely due to impact of macropores at larger scale in field measurements.

The new database was applied to develop pedotransfer functions (PTFs) for Ksat using temperate ~~and~~ laboratory based soil samples using both MPR and RF algorithms. Both algorithms provided reasonable accuracy. However, PTFs developed for a certain climatic region (temperate) or measurement method (laboratory) could not be satisfactorily applied to estimate Ksat for other regions (tropical) or measurement method (field) due to the role of different soil forming processes (inactive clay minerals in tropical soils and impact of biopores in field measurements).

There are still some gaps in the geographical representation of sampling points, especially in Russia and the higher northern latitudes, that could induce uncertainty in global modeling. Therefore, the data set can be further improved by covering the missing areas and achieve better accuracy in the hydrological applications.

The SoilKsatDB was developed in R software and is available via https://www.openml.org/d/42332 and https://doi.org/10.5281/zenodo.3752721. We have made code and data publicly available to enable further developments and improvements as a collective effort.

Acknowledgements. The SoilKsatDB is a compilation of numerous existing datasets from which the most significant: SWIG dataset (Rahmati et al., 2018), UNSODA (Leij et al., 1996; Nemes et al., 2001), and HYBRAS (Ottoni et al., 2018). The study was supported by ETH Zurich (Grant ETH-18 18-1). OpenGeoHub maintains an global repository of Earth System Science datasets at www.openlandmap.org. We thank Zhongwang Wei for helping in collecting the datasets and for insightful discussions. We would also want to thank Samuel Bickel (ETH Zurich) for boosting the leading author's confidence in High Performance Computing.

References

Abagandura, G. O., Nasr, G. E.-D. M., and Moumen, N. M.: Influence of tillage practices on soil physical properties and growth and yield of maize in jabal al akhdar, Libya, Open Journal of Soil Science, 7, 118–132, 2017.

Amer, A.-M. M., Logsdon, S. D., and Davis, D.: Prediction of hydraulic conductivity as related to pore size distribution in unsaturated soils, Soil science, 174, 508–515, 2009.

Amoozegar, A. and Warrick, A.: Hydraulic conductivity of saturated soils: field methods, Methods of Soil Analysis: Part 1 Physical and Mineralogical Methods, 5, 735–770, 1986.

Anapalli, S. S., Nielsen, D. C., Ma, L., Ahuja, L. R., Vigil, M. F., and Halvorson, A. D.: Effectiveness of RZWQM for simulating alternative Great Plains cropping systems, Agronomy journal, 97, 1183–1193, 2005.

Andrade, R. B.: The influence of bulk density on the hydraulic conductivity and water content-matric suction relation of two soils, 1971.

Arend, J. L.: Infiltration rates of forest soils in the Missouri Ozarks as affected by woods burning and litter removal, J. For., 39, 726–728, 1941.

Baird, A. J.: Field estimation of macropore functioning and surface hydraulic conductivity in a fen peat, Hydrological Processes, 11, 287–295, 1997.

Baird, A. J., Low, R., Young, D., Swindles, G. T., Lopez, O. R., and Page, S.: High permeability explains the vulnerability of the carbon store in drained tropical peatlands, Geophysical Research Letters, 44, 1333–1339, 2017.

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.

Batjes, N. H.: Total carbon and nitrogen in the soils of the world, European journal of soil science, 47, 151–163, 1996.

Becker, R., Gebremichael, M., and Märker, M.: Impact of soil surface and subsurface properties on soil saturated hydraulic conductivity in the semi-arid Walnut Gulch Experimental Watershed, Arizona, USA, Geoderma, 322, 112–120, 2018.

Beyer, M., Gaj, M., Hamutoko, J. T., Koeniger, P., Wanke, H., and Himmelsbach, T.: Estimation of groundwater recharge via deuterium labelling in the semi-arid Cuvelai-Etosha Basin, Namibia, Isotopes in environmental and health studies, 51, 533–552, 2015.

Bhattacharyya, R., Prakash, V., Kundu, S., and Gupta, H.: Effect of tillage and crop rotations on pore size distribution and soil hydraulic conductivity in sandy clay loam soil of the Indian Himalayas, Soil and Tillage Research, 86, 129–140, 2006.

Blake, W. H., Theocharopoulos, S. P., Skoulikidis, N., Clark, P., Tountas, P., Hartley, R., and Amaxidis, Y.: Wildfire impacts on hillslope sediment and phosphorus yields, Journal of Soils and Sediments, 10, 671–682, 2010.

Boike, J., Roth, K., and Overduin, P. P.: Thermal and hydrologic dynamics of the active layer at a continuous permafrost site (Taymyr Peninsula, Siberia), Water Resources Research, 34, 355–363, 1998.

Bonell, M. and Williams, J.: The two parameters of the Philip infiltration equation: their properties and spatial and temporal heterogeneity in a red earth of tropical semi-arid Queensland, Journal of Hydrology, 87, 9–31, 1986.

Bonsu, M. and Masopeh, B.: Saturated hydraulic conductivity values of some forest soils of Ghana determined by a simple method, Ghana Journal of Agricultural Science, 29, 75–80, 1996.

Braud, I., Desprats, J.-F., Ayral, P.-A., Bouvier, C., and Vandervaere, J.-P.: Mapping topsoil field-saturated hydraulic conductivity from point measurements using different methods, Journal of Hydrology and Hydromechanics, 65, 264–275, 2017.

Bruand, A., Duval, O., and Cousin, I.: Estimation des propriétés de rétention en eau des sols à partir de la base de données SOLHYDRO: Une première proposition combianant le type d'horizon, sa texture et sa densité apparente., 2004.

Campbell, R. E., Baker, J., Ffolliott, P. F., Larson, F. R., and Avery, C. C.: Wildfire effects on a ponderosa pine ecosystem: an Arizona case study, USDA For. Serv. Res. Pap. RM-191. Fort Collins, CO: US Department of Agriculture, Forest Service, Rocky Mountain Forest and Range Experimental Station. 12 p., 191, 1977.

Casalicchio, G., Bossek, J., Lang, M., Kirchhoff, D., Kerschke, P., Hofner, B., Seibold, H., Vanschoren, J., and Bischl, B.: OpenML: An R package to connect to the machine learning platform OpenML, Computational Statistics, pp. 1–15, 2017.

Chang, Y.-J.: Predictions of saturated hydraulic conductivity dynamics in a midwestern agricultural watershed, Iowa, 2010.

Chief, K., Ferré, T., and Nijssen, B.: Correlation between air permeability and saturated hydraulic conductivity: Unburned and burned soils, Soil Science Society of America Journal, 72, 1501–1509, 2008.

Cisneros, J., Cantero, J., and Cantero, A.: Vegetation, soil hydrophysical properties, and grazing relationships in saline-sodic soils of Central Argentina, Canadian Journal of Soil Science, 79, 399–409, 1999.

Coelho, M. A.: Spatial variability of water related soil physical properties., 1974.

Conedera, M., Peter, L., Marxer, P., Forster, F., Rickenmann, D., and Re, L.: Consequences of forest fires on the hydrogeological response of mountain catchments: a case study of the Riale Buffaga, Ticino, Switzerland, Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group, 28, 117–129, 2003.

Cornelis, W. M., Ronsyn, J., Van Meirvenne, M., and Hartmann, R.: Evaluation of pedotransfer functions for predicting the soil moisture retention curve, Soil Science Society of America Journal, 65, 638–648, 2001.

Daniel, S., Gabiri, G., Kirimi, F., Glasner, B., Näschen, K., Leemhuis, C., Steinbach, S., and Mtei, K.: Spatial distribution of soil hydrological properties in the Kilombero floodplain, Tanzania, Hydrology, 4, 57, 2017.

Deshmukh, H., Chandran, P., Pal, D., Ray, S., Bhattacharyya, T., and Potdar, S.: A pragmatic method to estimate plant available water capacity (PAWC) of rainfed cracking clay soils (Vertisols) of Maharashtra, Central India, Clay Res, 33, 1–14, 2014.

Ebel, B. A., Moody, J. A., and Martin, D. A.: Hydrologic conditions controlling runoff generation immediately after wildfire, Water Resources Research, 48, 2012.

El-Shafei, Y., Al-Darby, A., Shalaby, A., and Al-Omran, A.: Impact of a highly swelling gel-forming conditioner (acryhope) upon water movement in uniform sandy soils, Arid Land Research and Management, 8, 33–50, 1994.

Elnaggar, A.: Spatial Variability of Soil Physiochemical Properties in Bahariya Oasis, Egypt, Egyptian J. of Soil Sci. (EJSS), 57, 313–328, https://doi.org/10.21608/EJSS.2017.4438, 2017.

Fatichi, S., Or, D., Walko, R., Vereecken, H., Young, M. H., Ghezzehei, T. A., Hengl, T., Kollet, S., Agam, N., and Avissar, R.: Soil structure is an important omission in Earth System Models, Nature Communications, 11, 2020.

Faybishenko, B.: Comparison of laboratory and field methods for determining the quasi-saturated hydraulic conductivity of soils, Tech. rep., Lawrence Berkeley National Lab.(LBNL), Berkeley, CA (United States), 1997.

Ferreira, A., Coelho, C., Boulet, A., and Lopes, F.: Temporal patterns of solute loss following wildfires in Central Portugal, International Journal of Wildland Fire, 14, 401–412, 2005.

Forrest, J., Beatty, H., Hignett, C., Pickering, J., and Williams, R.: Survey of the physical properties of wheatland soils in eastern Australia, 1985.

Ganiyu, S., Rabiu, J., and Olatoye, R.: Predicting hydraulic conductivity around septic tank systems using soil physico-chemical properties and determination of principal soil factors by multivariate analysis, Journal of King Saud University-Science, 2018.

Glinski, J., Ostrowski, J., Stepniewska, Z., and Stepniewski, W.: Soil sample bank representing mineral soils of Poland, Problemy Agrofizyki (Poland), 1991.

Gliński, J., Stępniewski, W., Stępniewska, Z., Włodarczyk, T., Brzezińska, M., et al.: Characteristics of aeration properties of selected soil profiles from central Europe., International agrophysics, 14, 17–31, 2000.

Greenwood, W. and Buttle, J.: Effects of reforestation on near-surface saturated hydraulic conductivity in a managed forest landscape, southern Ontario, Canada, Ecohydrology, 7, 45–55, 2014.

Grunwald, S.: Florida soil characterization data, Soil and water science department, IFAS-Instituite of food and agriculture science, University of Florida, http://soils.ifas.ufl.edu, 2020.

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S., and Or, D.: SoilKsatDB: a global compilation of soil saturated hydraulic conductivity measurements, Zenodo, https://doi.org/10.5281/zenodo.3752722, 2020.

Gwenzi, W., Hinz, C., Holmes, K., Phillips, I. R., and Mullins, I. J.: Field-scale spatial variability of saturated hydraulic conductivity on a recently constructed artificial ecosystem, Geoderma, 166, 43–56, 2011.

Habecker, M., McSweeney, K., and Madison, F.: Identification and genesis of fragipans in Ochrepts of north central Wisconsin, Soil Science Society of America Journal, 54, 139–146, 1990.

Habel, A. Y.: The role of climate on the aggregate stability and soil erodibility of selected El-Jabal Al-Akhdar soils-Libya, Alexandria Journal of Agricultural Research, 58, 261–271, 2013.

Hao, M., Zhang, J., Meng, M., Chen, H. Y., Guo, X., Liu, S., and Ye, L.: Impacts of changes in vegetation on saturated hydraulic conductivity of soil in subtropical forests, Scientific reports, 9, 8372, 2019.

Hardie, M. A., Cotching, W. E., Doyle, R. B., Holz, G., Lisson, S., and Mattern, K.: Effect of antecedent soil moisture on preferential flow in a texture-contrast soil, Journal of Hydrology, 398, 191–201, 2011.

Haverkamp, R., Zammit, C., Bouraoui, F., Rajkai, K., Arrúe, J., and Heckmann, N.: GRIZZLY: Grenoble catalogue of soils: Survey of soil field data and description of particle-size, soil water retention and hydraulic conductivity functions, Lab. d'Etude des Transferts en Hydrol. et Environ., Grenoble, France, 1998.

Helbig, M., Boike, J., Langer, M., Schreiber, P., Runkle, B. R., and Kutzbach, L.: Spatial and seasonal variability of polygonal tundra water balance: Lena River Delta, northern Siberia (Russia), Hydrogeology journal, 21, 133–147, 2013.

Hiederer, R., Jones, R. J., and Daroussin, J.: Soil Profile Analytical Database for Europe (SPADE): reconstruction and validation of the measured data (SPADE/M), Geografisk Tidsskrift-Danish Journal of Geography, 106, 71–85, 2006.

Hinton, H.: Land Management Controls on Hydraulic Conductivity of an Urban Farm in Atlanta, GA, 2016.

Hodnett, M. and 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.

Horn, A., Stumpfe, A., Kues, J., Zinner, H.-J., and Fleige, H.: Die Labordatenbank des Niedersächsischen Bodeninformationssystems (NIBIS)-. Teil: Fachinformationssystem Bodenkunde, Geologisches Jahrbuch. Reihe A, Allgemeine und regionale Geologie BR Deutschland und Nachbargebiete, Tektonik, Stratigraphie, Paläontologie, pp. 59–97, 1991.

Houghton, T. B.: Hydrogeologic characterization of an alpine glacial till, Snowy Range, Wyoming, Ph.D. thesis, Colorado State University. Libraries, 2011.

Hu, W., She, D., Shao, M., Chun, K. P., and Si, B.: Effects of initial soil water content and saturated hydraulic conductivity variability on small watershed runoff simulation using LISEM, Hydrological Sciences Journal, 60, 1137–1154, 2015.

Imeson, A., Verstraten, J., Van Mulligen, E., and Sevink, J.: The effects of fire and water repellency on infiltration and runoff under Mediterranean type forest, Catena, 19, 345–361, 1992.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Jabro, J.: Estimation of saturated hydraulic conductivity of soils from particle size distribution and bulk density data, Transactions of the ASAE, 35, 557–560, 1992.

Jarvis, N., Koestel, J., Messing, I., Moeys, J., and Lindahl, A.: Influence of soil, land use and climatic factors on the hydraulic conductivity of soil, Hydrology and Earth System Sciences, 17, 5185–5195, 2013.

Johansen, M. P., Hakonson, T. E., and Breshears, D. D.: Post-fire runoff and erosion from rainfall simulation: contrasting forests with shrublands and grasslands, Hydrological processes, 15, 2953–2965, 2001.

Kanemasu, E.: Soil Hydraulic Conductivity Data (FIFE), ORNL Distributed Active Archive Center, https://doi.org/10.3334/ORNLDAAC/107, 1994.

Katimon, A. and Hassan, A. M. M.: Field hydraulic conductivity of some Malaysian peat, Malaysian Journal of Civil Engineering, 10, 1997.

Keisling, T. C.: Precision with which selected physical properties of similar soils can be estimated, Ph.D. thesis, Oklahoma State University, 1974.

Kelly, T. J., Baird, A. J., Roucoux, K. H., Baker, T. R., Honorio Coronado, E. N., Ríos, M., and Lawson, I. T.: The high hydraulic conductivity of three wooded tropical peat swamps in northeast Peru: measurements and implications for hydrological function, Hydrological Processes, 28, 3373–3387, 2014.

Kirby, J., Kingham, R., and Cortes, M.: Texture, density and hydraulic conductivity of some soils in San Luis province, Argentina, Ciencia del suelo, 19, 20–28, 2001.

Kool, J., Albrecht, K. A., Parker, J., Baker, J., et al.: Physical and chemical characterization of the Groseclose soil mapping unit, 1986.

Krahmer, U., Hennings, V., Müller, U., and Schrey, H.-P.: Ermittlung bodenphysikalischer Kennwerte in Abhängigkeit von Bodenart, lagerungsdichte und Humusgehalt, Zeitschrift für Pflanzenernährung und Bodenkunde, 158, 323–331, 1995.

Kramarenko, V., Brakorenko, N., and Molokov, V.: Hydraulic conductivity of peat in Western Siberia, in: E3S Web of Conferences, vol. 98, p. 11003, EDP Sciences, 2019.

Kutiel, P., Lavee, H., Segev, M., and Benyamini, Y.: The effect of fire-induced surface heterogeneity on rainfall-runoff-erosion relationships in an eastern Mediterranean ecosystem, Israel, Catena, 25, 77–87, 1995.

Lamara, M. and Derriche, Z.: Prediction of unsaturated hydraulic properties of dune sand on drying and wetting paths, Electron. J. Geotech. Eng, 13, 1–19, 2008.

Lawrence, I. and Lin, K.: A concordance correlation coefficient to evaluate reproducibility, Biometrics, pp. 255–268, 1989.

Leij, F., Alves, W., Van Genuchten, M. T., and Williams, J.: The UNSODA Unsaturated Soil Hydraulic Database; User's Manual, Version 1.0, Rep. EPA/600/R-96, 95, 103, 1996.

Li, X., Liu, S., Xiao, Q., Ma, M., Jin, R., Che, T., Wang, W., Hu, X., Xu, Z., Wen, J., et al.: A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system, Scientific data, 4, 170 083, 2017.

Lopes, V. S., Cardoso, I. M., Fernandes, O. R., Rocha, G. C., Simas, F. N. B., de Melo Moura, W., Santana, F. C., Veloso, G. V., and da Luz, J. M. R.: The establishment of a secondary forest in a degraded pasture to improve hydraulic properties of the soil, Soil and Tillage Research, 198, 104 538, 2020.

Lopez, O., Jadoon, K., and Missimer, T.: Method of relating grain size distribution to hydraulic conductivity in dune sands to assist in assessing managed aquifer recharge projects: Wadi Khulays dune field, western Saudi Arabia, Water, 7, 6411–6426, 2015.

Mahapatra, S. and Jha, M. K.: On the estimation of hydraulic conductivity of layered vadose zones with limited data availability, Journal of Earth System Science, 128, 75, 2019.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Martin, D. A. and Moody, J. A.: Comparison of soil infiltration rates in burned and unburned mountainous watersheds, Hydrological Processes, 15, 2893–2903, 2001.

McKenzie, N., Jacquier, D., and Gregory, L.: Online soil information systems–recent Australian experience, in: Digital soil mapping with limited data, pp. 283–290, Springer, 2008.

Mohanty, B., Kanwar, R. S., and Everts, C.: Comparison of saturated hydraulic conductivity measurement methods for a glacial-till soil, Soil Science Society of America Journal, 58, 672–677, 1994.

Mott, J., Bridge, B., and Arndt, W.: Soil seals in tropical tall grass pastures of northern Australia, Soil Research, 17, 483–494, 1979.

Mualem, Y.: Catalogue of the hydraulic properties of unsaturated soils, Technion Israel Institute of Technology, Technion Research & Development, 1976.

National Cooperative Soil Survey: National cooperative soil survey characterization database, United States Department of Agriculture, Natural Resoucres Conservation, Lincoln, NE, 2016.

Nemes, A.: Unsaturated soil hydraulic database of Hungary: HUNSODA, Agrokémia és Talajtan, 51, 17–26, 2002.

Nemes, A.: Databases of soil physical and hydraulic properties, Encyclopedia of agrophysics, pp. 194–199, 2011.

Nemes, A. d., Schaap, M., Leij, F., and Wösten, J.: Description of the unsaturated soil hydraulic database UNSODA version 2.0, Journal of Hydrology, 251, 151–162, 2001.

Nielsen, D., Biggar, J., and Erh, K.: “Spatial variability of field-measured soil water properties. Hilgardia, 42 (7), 215–259., 1973.

Niemeyer, R., Fremier, A. K., Heinse, R., Chávez, W., and DeClerck, F. A.: Woody vegetation increases saturated hydraulic conductivity in dry tropical Nicaragua, Vadose Zone Journal, 13, 2014.

Nyman, P., Sheridan, G. J., Smith, H. G., and Lane, P. N.: Evidence of debris flow occurrence after wildfire in upland catchments of south-east Australia, Geomorphology, 125, 383–401, 2011.

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

Ouattara, M.: Variation of saturated hydraulic conductivity with depth for selected profiles of Tillman-Hollister soil, Ph.D. thesis, Oklahoma State University, 1977.

Päivänen, J. et al.: Hydraulic conductivity and water retention in peat soils., Suomen metsätieteellinen seura, 1973.

Parks, D. S. and Cundy, T. W.: Soil hydraulic characteristics of a small southwest Oregon watershed following high-intensity wildfires, in: In: Berg, Neil H. tech. coord. Proceedings of the Symposium on Fire and Watershed Management: October 26-28, 1988, Sacramento, California. Gen. Tech. Rep. PSW-109. Berkeley, Calif.: US Department of Agriculture, Forest Service, Pacific Southwest Forest and Range Experiment Station: 63-67, vol. 109, 1989.

Price, K., Jackson, C. R., and Parker, A. J.: Variation of surficial soil hydraulic properties across land uses in the southern Blue Ridge Mountains, North Carolina, USA, Journal of Hydrology, 383, 256–268, 2010.

Quinton, W. L., Hayashi, M., and Carey, S. K.: Peat hydraulic conductivity in cold regions and its relation to pore size and geometry, Hydrological Processes: An International Journal, 22, 2829–2837, 2008.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, http://www.R-project.org/, 2013.

Rab, M.: Soil physical and hydrological properties following logging and slash burning in the Eucalyptus regnans forest of southeastern Australia, Forest Ecology and Management, 84, 159–176, 1996.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Rahimy, P.: Effects of Soil Depth and Saturated Hydraulic Conductivity Spatial Variation on Runoff Simulation by the Limburg Soil Erosion Model, LISEM: A Case Study in Faucon Catchment, France, University of Twente Faculty of Geo-Information and Earth Observation (ITC), 2011.

Rahmati, M., Weihermüller, L., Vanderborght, J., Pachepsky, Y. A., Mao, L., Sadeghi, S. H., Moosavi, N., Kheirfam, H., Montzka, C., Van Looy, K., et al.: Development and analysis of the Soil Water Infiltration Global database, 2018.

Ramli, M.: Management of Groundwater Resources from Peat in Sarawak, 1999.

Ravi, S., Wang, L., Kaseke, K. F., Buynevich, I. V., and Marais, E.: Ecohydrological interactions within "fairy circles" in the Namib Desert: Revisiting the self-organization hypothesis, Journal of Geophysical Research: Biogeosciences, 122, 405–414, 2017.

Rawls, W. J., Brakensiek, D. L., and Saxtonn, K.: Estimation of soil water properties, Transactions of the ASAE, 25, 1316–1320, 1982.

Richard and Luescher: Phys. Eig. von Boeden der Schweiz, vol. 1-5, Swiss Federal Institute for Vers. CH-8093 Birmensdorf, 1987.

Robbins, C. W.: Hydraulic conductivity and moisture retention characteristics of southern Idaho's silt loam soils, 1977.

Romano, N. and Palladino, M.: Prediction of soil water retention using soil physical data and terrain attributes, Journal of Hydrology, 265, 56–75, 2002.

Rycroft, D., Williams, D., and Ingram, H.: The transmission of water through peat: I. Review, The Journal of Ecology, pp. 535–556, 1975.

Sanzeni, A., Colleselli, F., and Grazioli, D.: Specific surface and hydraulic conductivity of fine-grained soils, Journal of Geotechnical and Geoenvironmental Engineering, 139, 1828–1832, 2013.

Sayok, A., Ayob, K., Melling, L., Goh, K., Uyo, L., and Hatano, R.: Hydraulic conductivity and moisture characteristics of tropical peatland-preliminary investigation, Malaysian Society of Soil Science (MSSS), 2007.

Schindler, U. G. and Müller, L.: Soil hydraulic functions of international soils measured with the Extended Evaporation Method (EEM) and the HYPROP device, Open Data Journal for Agricultural Research, 3, 2017.

Sharma, S. K., Mohanty, B. P., and Zhu, J.: Including topography and vegetation attributes for developing pedotransfer functions, Soil Science Society of America Journal, 70, 1430–1440, 2006.

Sharratt, B. S.: Water retention, bulk density, particle size, and thermal and hydraulic conductivity of arable soils in interior Alaska, 1990.

Simmons, L. A.: Soil hydraulic and physical properties as affected by logging management, Ph.D. thesis, University of Missouri–Columbia, 2014.

Singh, I., Awasthi, O., Sharma, B., More, T., Meena, S., et al.: Soil properties, root growth, water-use efficiency in brinjal (Solanum melongena) production and economics as affected by soil water conservation practices, Indian Journal of Agricultural Sciences, 81, 760, 2011.

Singh, R., Van Dam, J., and Feddes, R. A.: Water productivity analysis of irrigated crops in Sirsa district, India, Agricultural Water Management, 82, 253–278, 2006.

Smettem, K. and Ross, P.: Measurement and prediction of water movement in a field soil: The matrix-macropore dichotomy, Hydrological processes, 6, 1–10, 1992.

Sonneveld, M., Everson, T., and Veldkamp, A.: Multi-scale analysis of soil erosion dynamics in Kwazulu-Natal, South Africa, Land Degradation & Development, 16, 287–301, 2005.

Soracco, C. G., Lozano, L. A., Sarli, G. O., Gelati, P. R., and Filgueira, R. R.: Anisotropy of saturated hydraulic conductivity in a soil under conservation and no-till treatments, Soil and Tillage Research, 109, 18–22, 2010.

Southard, R. and Buol, S.: Subsoil saturated hydraulic conductivity in relation to soil properties in the North Carolina Coastal Plain, Soil Science Society of America Journal, 52, 1091–1094, 1988.

Szabó, B., Szatmári, G., Takács, K., Laborczi, A., Makó, A., Rajkai, K., and Pásztor, L.: Mapping soil hydraulic properties using random-forest-based pedotransfer functions and geostatistics, Hydrology and Earth System Sciences, 23, 2615–2635, 2019.

Takahashi, H.: Studies on microclimate and hydrology of peat swamp forest in Central Kalimantan, Indonesia, in: Biodiversity and Sustainability of Tropical peatlands, Samara Publishing Limited, 1997.

Tete-Mensah, I.: Evaluation of Some Physical and Chemical Properties of Soils Under two Agroforestrv Practices, Ph.D. thesis, University of Ghana, 1993.

Tian, J., Zhang, B., He, C., and Yang, L.: Variability in soil hydraulic conductivity and soil hydrological response under different land covers in the mountainous area of the Heihe River Watershed, Northwest China, Land degradation & development, 28, 1437–1449, 2017.

Tomasella, J., Hodnett, M. G., and Rossato, L.: Pedotransfer functions for the estimation of soil water retention in Brazilian soils, 2000.

Tomasella, J., Pachepsky, Y., Crestana, S., and Rawls, W.: Comparison of two techniques to develop pedotransfer functions for water retention, Soil Science Society of America Journal, 67, 1085–1092, 2003.

Tuller, M. and Or, D.: Unsaturated Hydraulic Conductivity of Structured Porous MediaA Review of Liquid Configuration–Based Models, Vadose Zone Journal, 1, 14–37, 2002.

Varela, M., Benito, E., and Keizer, J.: Influence of wildfire severity on soil physical degradation in two pine forest stands of NW Spain, Catena, 133, 342–348, 2015.

Verburg, K., Bridge, B. J., Bristow, K. L., and Keating, B. A.: Properties of selected soils in the Gooburrum–Moore Park area of Bundaberg, CSIRO Land and Water Technical Report, 9, 77, 2001.

Vereecken, H., Weynants, M., Javaux, M., Pachepsky, Y., Schaap, M., Genuchten, M. T., et al.: Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review, Vadose Zone Journal, 9, 795–820, 2010.

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

Vieira, B. C. and Fernandes, N. F.: Landslides in Rio de Janeiro: the role played by variations in soil hydraulic conductivity, Hydrological Processes, 18, 791–805, 2004.

Vogeler, I., Carrick, S., Cichota, R., and Lilburne, L.: Estimation of soil subsurface hydraulic conductivity based on inverse modelling and soil morphology, Journal of Hydrology, 574, 373–382, 2019.

Waddington, J. and Roulet, N.: Groundwater flow and dissolved carbon movement in a boreal peatland, Journal of Hydrology, 191, 122–138, 1997.

Wang, T., Zlotnik, V. A., Wedin, D., and Wally, K. D.: Spatial trends in saturated hydraulic conductivity of vegetated dunes in the Nebraska Sand Hills: Effects of depth and topography, Journal of Hydrology, 349, 88–97, 2008.

Weynants, M., Montanarella, L., Toth, G., Arnoldussen, A., Anaya Romero, M., Bilas, G., Borresen, T., Cornelis, W., Daroussin, J., Gonçalves, M. D. C., et al.: European HYdropedological Data Inventory (EU-HYDI), EUR Scientific and Technical Research Series, 2013.

Wösten, J., Pachepsky, Y. A., and Rawls, W.: Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, Journal of hydrology, 251, 123–150, 2001.

Wösten, J. et al.: The HYPRES database of hydraulic properties of European soils., Advances in GeoEcology, pp. 135–143, 2000.

Wright, M. N. and Ziegler, A.: Ranger: a fast implementation of random forests for high dimensional data in C++ and R, arXiv preprint arXiv:1508.04409, 2015.

https://doi.org/10.5194/essd-2020-149
Preprint. Discussion started: 9 July 2020

Yao, S., Zhang, T., Zhao, C., and Liu, X.: Saturated hydraulic conductivity of soils in the Horqin Sand Land of Inner Mongolia, northern China, Environmental monitoring and assessment, 185, 6013–6021, 2013.

Yasin, S. and Yulnafatmawita, Y.: Effects of Slope Position on Soil Physico-chemical Characteristics Under Oil Palm Plantation in Wet Tropical Area, West Sumatra Indonesia, AGRIVITA, Journal of Agricultural Science, 40, 328–337, 2018.

Yoon, S. W.: A measure of soil structure derived from water retention properties: A kullback-Leibler distance approach, Ph.D. thesis, Rutgers University-Graduate School-New Brunswick, 2009.

Zakaria, S.: Water management in deep peat soils in Malaysia, Ph.D. thesis, Cranfield University, 1992.

Zhang, S., Xiahou, Y., Tang, H., Huang, L., Liu, X., and Wu, Q.: Study on the spatially variable saturated hydraulic conductivity and deformation behavior of accumulation reservoir landslide Based on surface nuclear magnetic resonance survey, Advances in civil engineering, 2018.

Zhao, H., Zeng, Y., Lv, S., and Su, Z.: Analysis of soil hydraulic and thermal properties for land surface modeling over the Tibetan Plateau, Earth system science data, 10, 1031, 2018.