Supporting Information for

Global prediction of soil saturated hydraulic conductivity using random forest in a Covariate-based Geo Transfer Functions (CoGTF) framework

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

[1] Soil and Terrestrial Environmental Physics, Department of Environmental Systems Science, ETH Zurich, Zurich, Switzerland,

[2] Envirometrix Ltd., Wageningen, the Netherlands

[3] OpenGeoHub, Wageningen, the Netherlands

[4] Bartlett School of Environment, Energy and Resources, University College London, London, UK

[5] Division of Hydrologic Sciences, Desert Research Institute, Reno, NV, USA

* Correspondence to:

Surya Gupta (surya.gupta@usys.ethz.ch)

Contents of this file

Figures S1 to S8

Table S1

Introduction

This supplementary information provides the figures of block-wise cross-validation (Figure S1), Ksat values at different depth computed with CoGTF (Figures S2 and S3), spatial distributions of Ksat and different environmental covariates in India (Figure S4), effect of clustering of Ksat samples on global map (Figure S5), global Ksat maps predicted with remote sensing or soil covariates (Figure S6) and a comparison between CoGTF and Dai et al., 2019 (Figures S7 and S8). At the end, a table lists the environmental covariates used in this study.
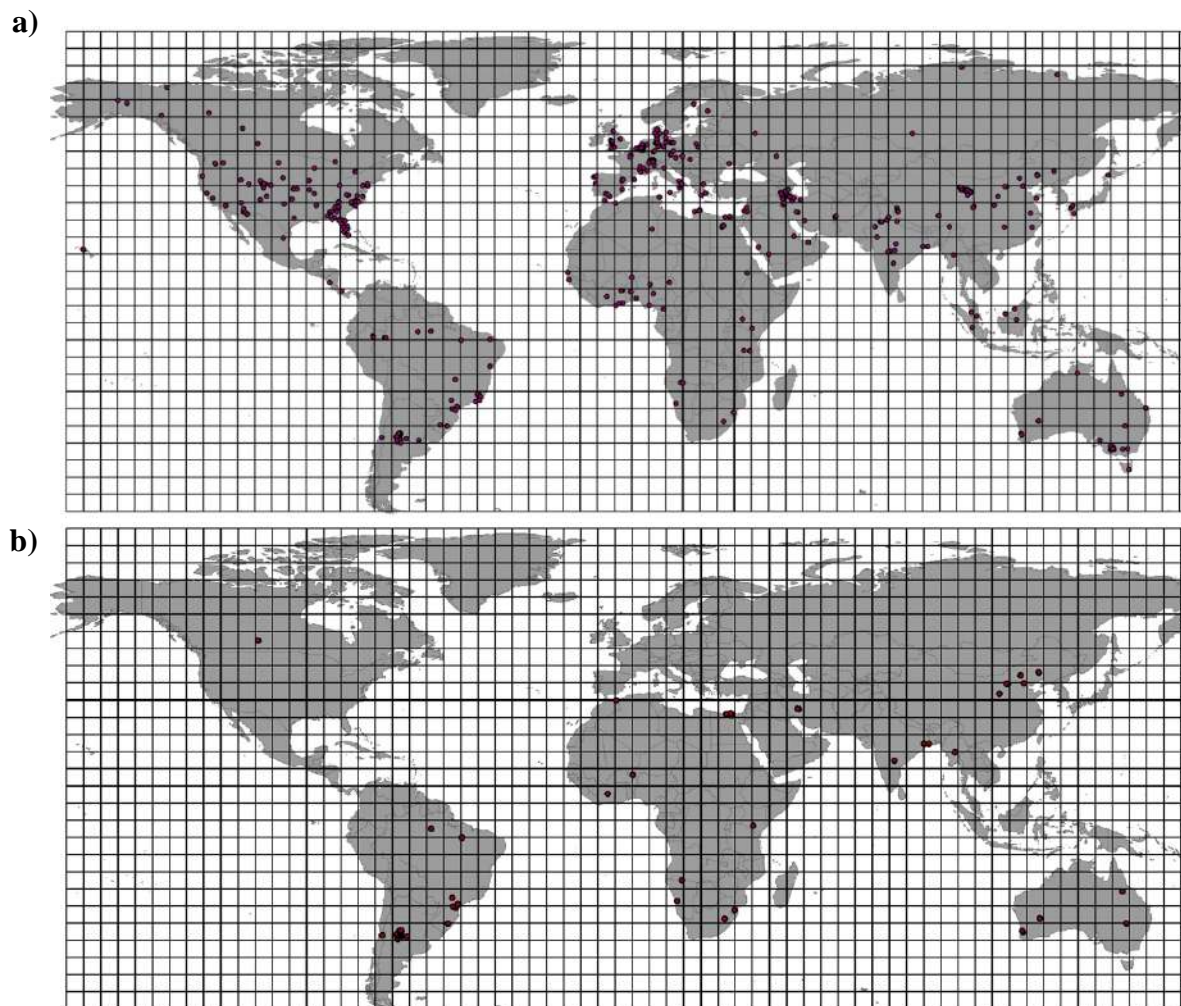
a)

b)

Figure S1 Regionalization of global map for validation of CoGTF model. a) 5 degrees by 5 degrees grids plotted with positions of Ksat dataset (6,814 samples). A total of 168 grid cells contains the data points. b) 30 blocks of data were removed randomly (i.e 20% of 2,525 Ksat dataset) for validation. The 2,525 Ksat samples are a subset of the totally 6,814 samples because samples from Europe and North America were excluded (they were used to train Rosetta 3 model and could not be considered for model validation).
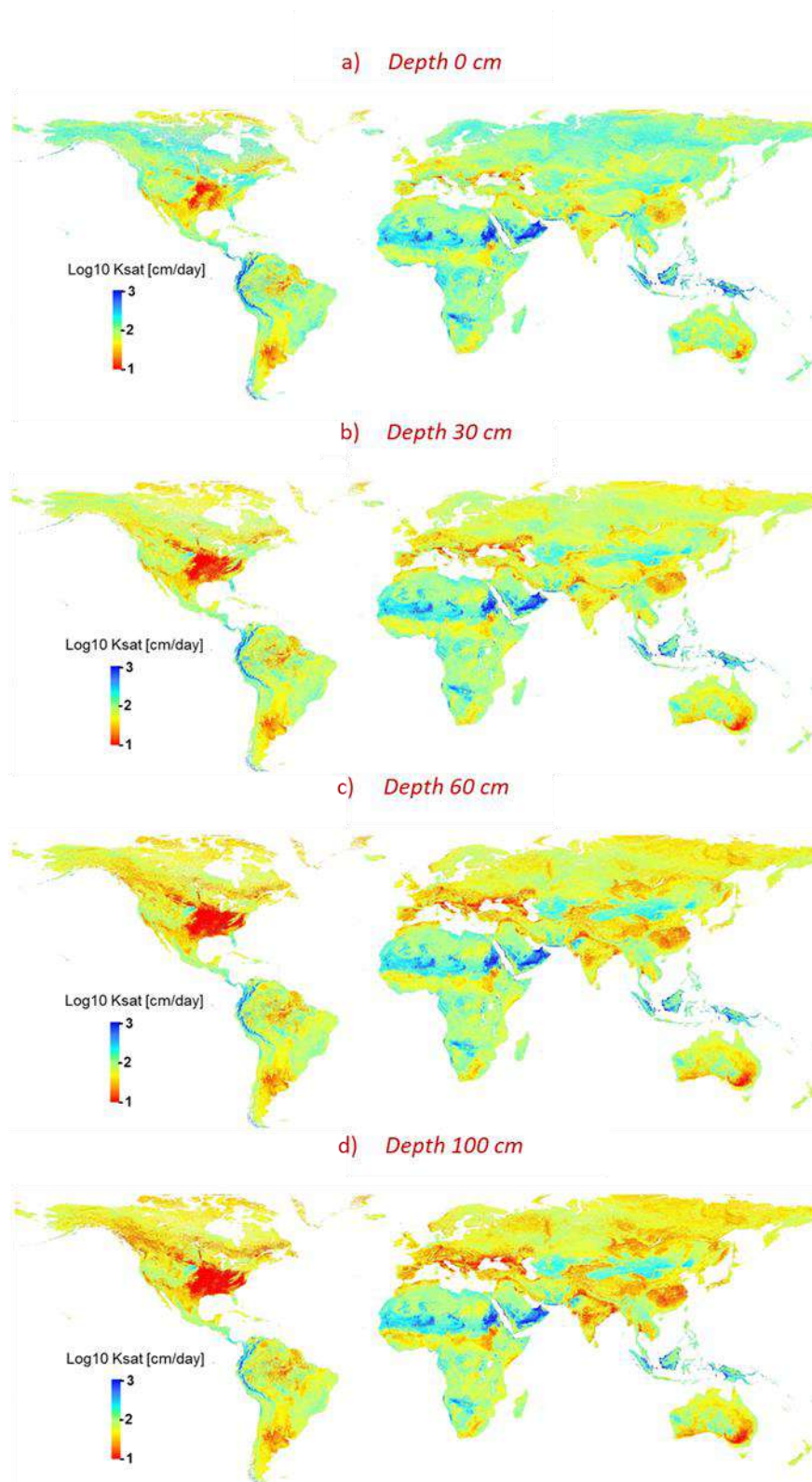
a) *Depth 0 cm*

b) *Depth 30 cm*

c) *Depth 60 cm*

d) *Depth 100 cm*

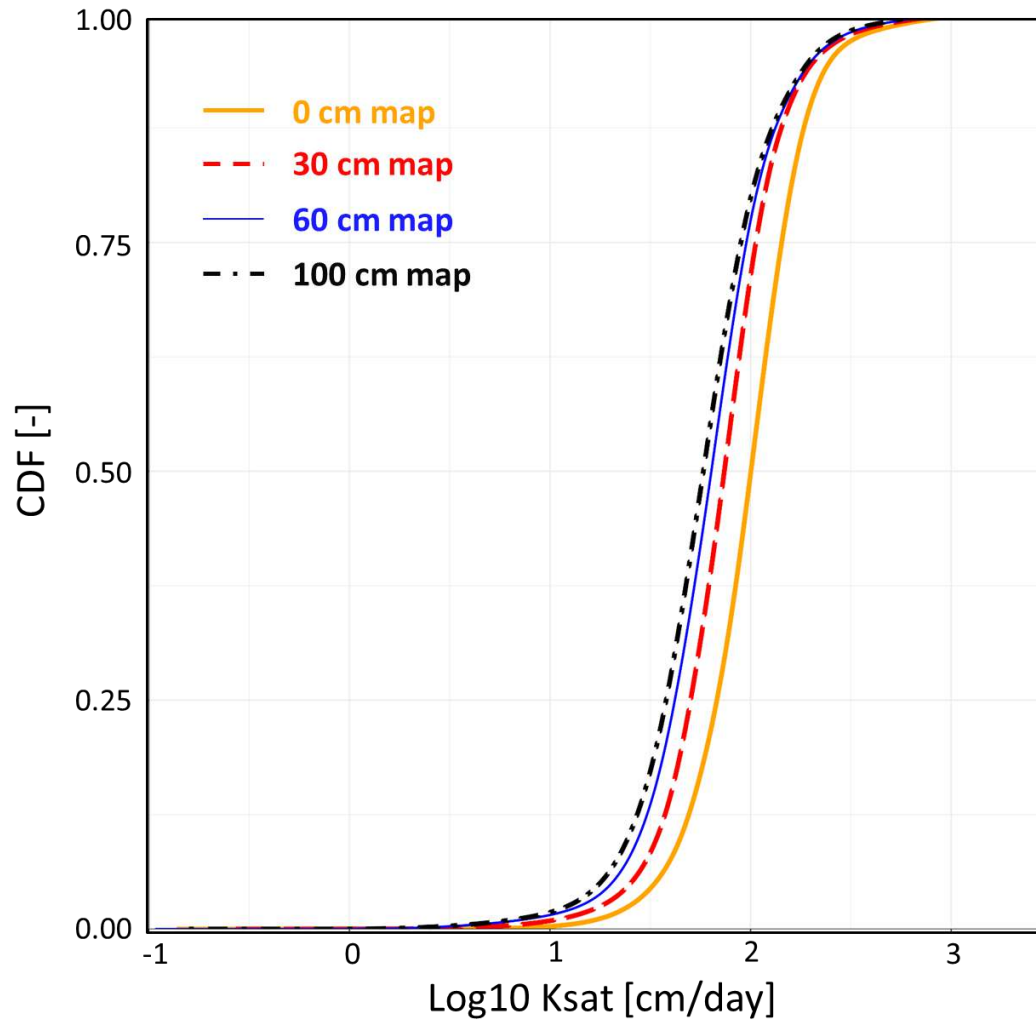**Figure S2** Ksat maps at different depths a) 0 cm b) 30 cm, b) 60 cm, c) 100 cm computed with CoGTF

Figure S3 Cumulative distribution function CDF for global maps of Ksat at different depths (0, 30, 60 and 100 cm) computed with CoGTF.
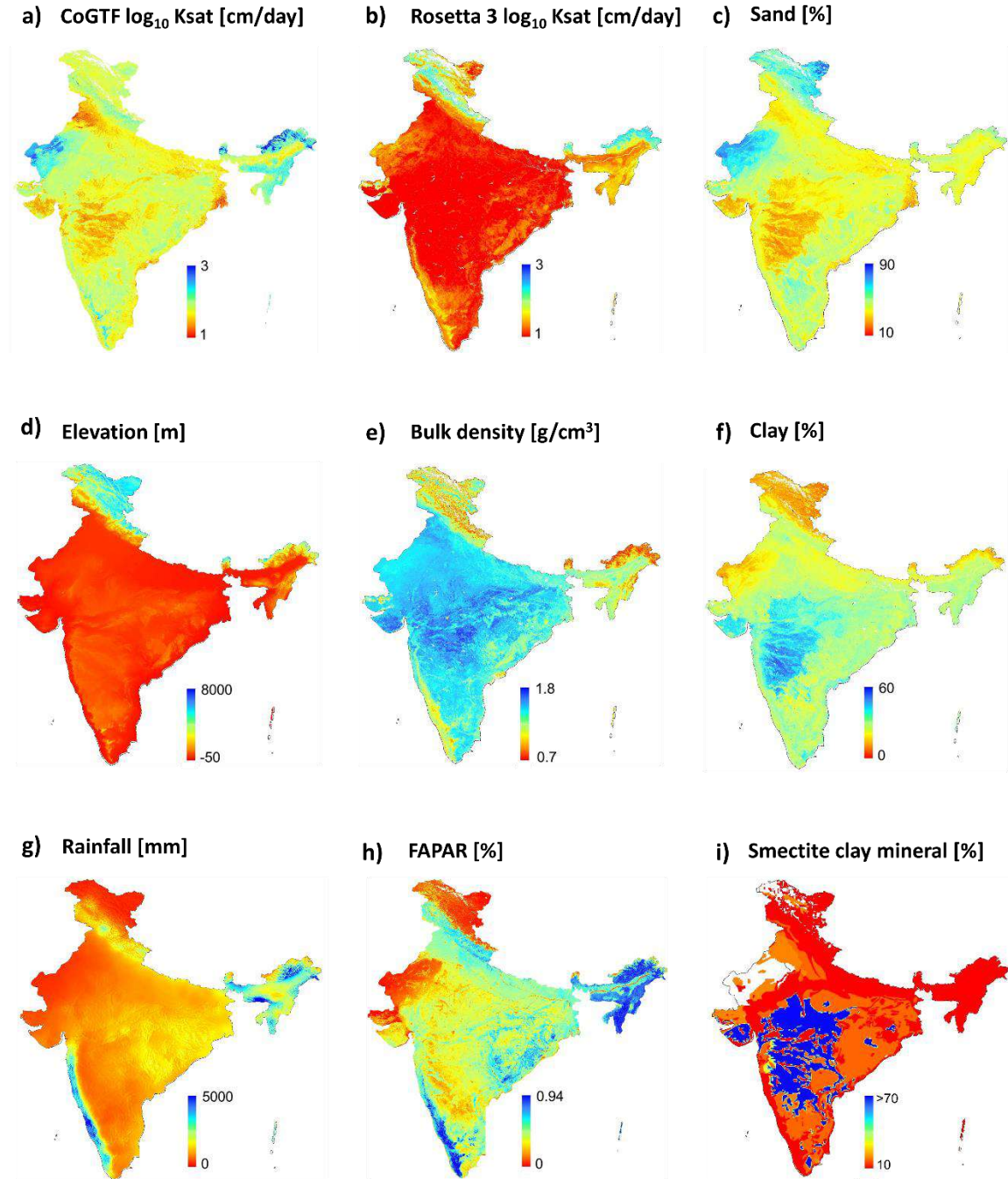
Figure S4 Ksat values for India predicted with CoGTF (a), spatial patterns of the Rosetta 3 Ksat map b), and the first four most important covariates (c-f, see figure3 in the main text): sand fraction (%), elevation (meters above sea level), bulk density (g/cm³) and clay fraction (%). Other covariates important for soil formation liked with Ksat are shown as well (g-i): mean annual rainfall (mm), fraction of absorbed photosynthetically active radiation (FAPAR, values in %) and kaolinite (in %) clay mineral.
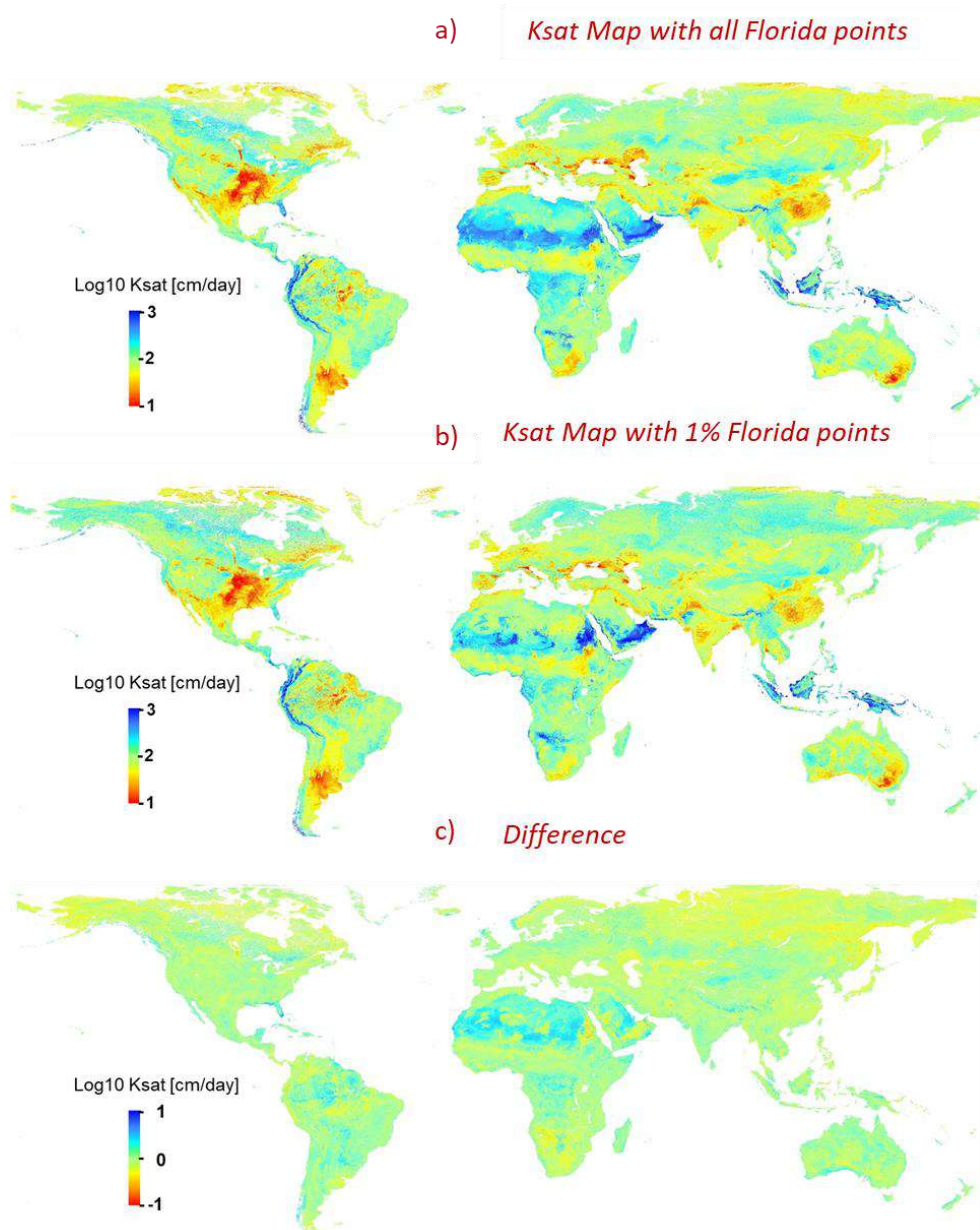
a) *Ksat Map with all Florida points*

b) *Ksat Map with 1% Florida points*

c) *Difference*

Figure S5 The difference between Ksat map including all Florida samples (a) and using only 1% of these Florida Ksat points (b) to build the CoGTF model. In the maps of differences (c), blue color represents higher values when all Florida points are included, yellow represents approximately the same value in both maps, and red shows locations with higher Ksat when only 1% of Florida samples are included. The Florida cluster showed a large impact on the sandy regions such as Sahara and center part of Africa and middle east as it significantly increased the Ksat values. A similar effect was observed in parts of South America and Australia. On the other hand, south of Africa and higher Nothern latitude showed high Ksat values for map that includes 1% of Florida samples.
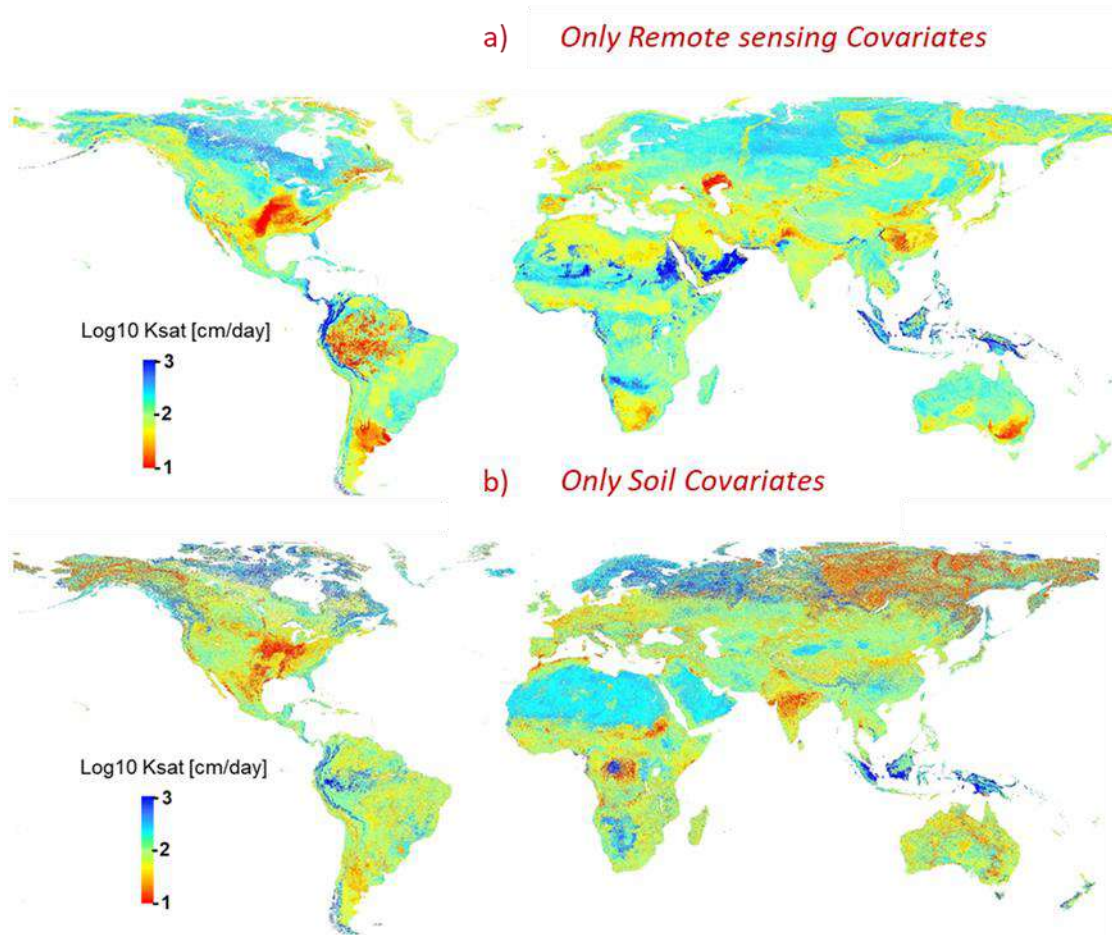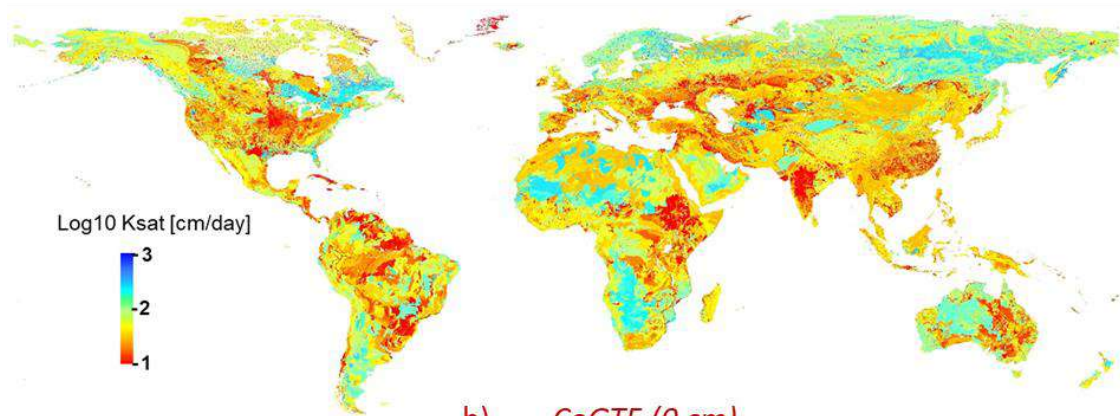
Figure S6 Ksat maps computed with Random Forest approach for soil depth of 0 cm. a) Only 24 remote sensing covariates were used to build model and to compute the map. b) Only soil properties were used (sand content, clay content and bulk density). Note that high contrast in northern latitudes in Eurasia are controlled by changes in bulk density (a dominant pattern in Rosetta 3 map).

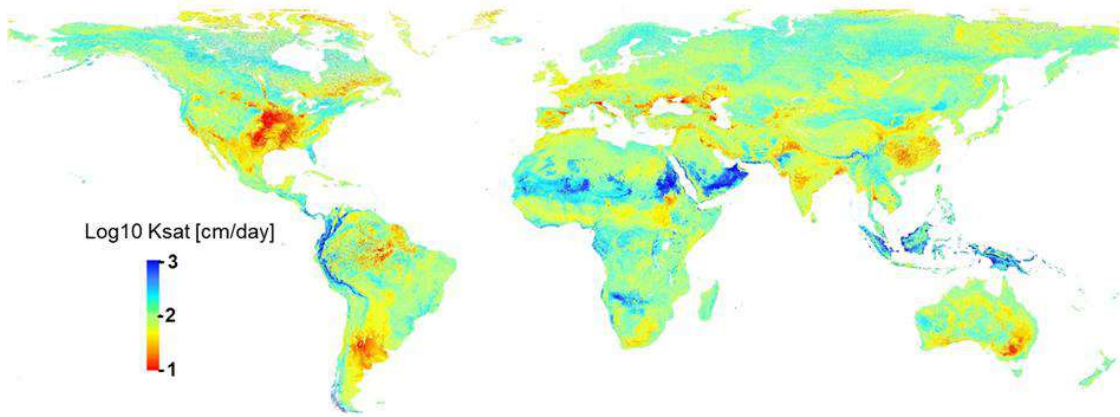a) **Dai et al.,2019 (0-5 cm)**

b) *CoGTF (0 cm)*

Figure S7 a) Ksat map at 0-5 cm depth from Dai et al. (2019) computed from an ensemble of 16 pedotransfer functions. The map used soil information from Global Soil Dataset for Earth System Models (GSDE; Shangguan et al., 2017) and SoilGrids (Hengl et al., 2017). b) CoGTF Ksat map at 0 cm.
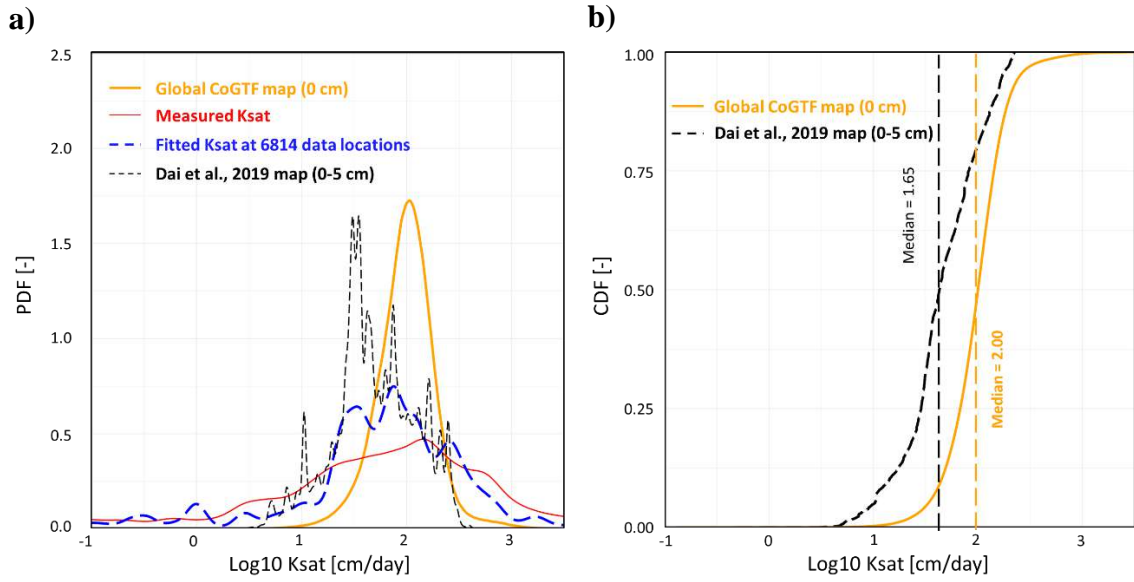
a)

b)

Figure S8 Difference in probability density functions (PDF) of Ksat values: (a) between global CoGTF map (yellow) and Dai et al. (2019) (black dotted line), measured (red) and fitted (blue) Ksat values at the sampling sites; (b) cumulative distribution functions for Dai et al. (2019) map (black dotted line) and CoGTF map (yellow) for soil depth 0 cm.

Table S1 List of covariates used for creating the Ksat map

S. no	List of Covariates	Source
	Climate	
1	clm_annual mean temperaturebio1_m_1km_s0..0cm_1979-2013_v1.0	http://chelsa-climate.org/bioclim/ (Karger et al., 2017)
2	clm_temperature seasonalitybio4_ m_1km_ s0..0cm_1979-2013_v1.0	
3	clm_max temperature of warmest monthbio5_m_1km_ s0..0cm_1979-2013_v1.0	
4	clm_min temperature of coldest monthbio6_ m_1km_ s0..0cm_1979-2013_v1.0	
5	clm_annual precipitationbio12_ m_1km_1979_2013_v1.0	
6	clm_precipitation of wettest monthbio13_ m_1km_1979_2013	
7	clm_precipitation of driest monthbio14_ m_1km_1979_2013	
8	clm_cloud.fraction_earthenv.modis.annual_m_1km_s0..0cm_2000..2015_v1.0	http://www.earthenv.org/cloud (Wilson & Jetz, 2016).
9	clm_diffuse.irradiation_solar.atlas.kwhm2.100_m_1km_s0..0cm_2016_v1	https://globalsolaratlas.info/download/world
10	clm_direct.irradiation_solar.atlas.kwhm2.10_m_1km_s0..0cm_2016_v1	
11	clm_lst_mod11a2.annual.day_m_1km_s0..0cm_2000..2017_v1.0	https://lpdaac.usgs.gov/products/mod11a2v006/
12	clm_lst_mod11a2.annual.day_sd_1km_s0..0cm_2000..2017_v1.0	
13	clm_precipitation_sm2rain.annual_m_1km_s0..0cm_2007..2018_v0.2	https://zenodo.org/record/3405563#.XlgdNTFKhaQ (Brocca et al., 2019)
	Digital terrain model	
14	dtm_twi_merit.dem_m_1km_s0..0cm_2017_v1.0	https://zenodo.org/record/1447210#.XllTejFKhaQ
15	dtm_slope_merit.dem_m_1km_s0..0cm_2017_v1.0	(Yamazaki et al., 2017)
16	dtm_aspect.cosine_merit.dem_m_1km_s0..0cm_2018_v1.0	
17	dtm_elevation_merit.dem_m_1km_s0..0cm_2017_v1.0	
18	dtm_lithology_usgs.ecotapestry.acid.plutonics_p_1km_s0..0cm_2014_v1.0	

	Surface reflectance	
19	lcv_landsat.nir_wri.forestwatch_m_1km_s0..0cm_2018_v1.2	Hansen et al. (2013)
20	lcv_landsat.red_wri.forestwatch_m_1km_s0..0cm_2018_v1.2	
21	lcv_landsat.swir2_wri.forestwatch_m_1km_s0..0cm_2018_v1.2	
22	lcv_snow_probav.lc100_p_1km_s0..0cm_2017_v1.0	Tsendbazar et al. (2017)
23	lcv_wetlands.regularly.flooded_upmc.wtd_p_1km_b0..200cm_2010..2015_v1.0	https://doi.pangaea.de/10.1594/PANGAEA.892657 (Tootchi et al., 2019)
	Vegetation covariates	
24	veg_fapar_proba.v.annnual_d_1km_s0..0cm_2014..2019_v1.0	https://land.copernicus.eu/global/products/fapar
	Predicted soil properties	
25	sol_clay.wfraction_usda.3a1a1a_m_1km_b0_10_30_60_100_200cm_1950..2017_v0.2	
26	sol_sand.wfraction_usda.3a1a1a_m_1km_b0_10_30_60_100_200cm _1950..2017_v0.2	https://www.openlandmap.org/
27	sol_bulk_density.wfraction_usda.3a1a1a_m_1km_ b0_10_30_60_100_200cm _1950..2017_v0.2	Hengl et al. (2017)