SOIL-WATERGRIDS

Technical Documentation

Authors: Federico Maggi, Magda Guglielmo, Fiona H. M. Tang , Chiara Pasut
Contact: federico.maggi@sydney.edu.au

Release chronology

SOIL-WATERGRIDS v1	June 2021	First release

Recommended citation for SOIL-WATERGRIDS
Guglielmo, M., Tang, F. H. M., Pasut, C., Maggi, F., (2021). Mapping dynamic changes in soil moisture and depth of water table from 1970 to 2014. Scientific Data.

1. GENERAL INFORMATION

1.1 Open access CC BY 4.0

1.2 Package content
The SOIL-WATERGRIDS product distributes the globally gridded soil water saturation at three depths within the root zone (0 to 100 cm) and the depth of up to two water tables, that is, the closest to the surface and the deepest. The data have an annual coverage from 1970 to 2014 and a monthly time frequency. Long-term monthly means are distributed as well. The SOIL-WATERGRIDS product also includes additional globally gridded maps of soil hydraulic parameters that can be used in conjunction with the soil water saturation to determine other hydraulic features such as the water potential and the volumetric water content.

This package also includes the full-size global modelling for use by a third party under the Open Access CC BY4.0 license. The package contains compressed (.ZIP and .RAR) folders that include the executable and bash files for running BRTSim v4-1a massively on a supercomputer, and all input files that instruct BRTSim on the solving domain (vertical discretization and soil properties, Param_*.inp files) and hydrometeorological boundary conditions (precipitation, evapotranspiration, irrigation, runoff, snowmelt, and water balance correction, Table_*.txt, RNF_*.txt, and WB_*.txt).

Features embedded in the BRTSim v4-1a solver are not reported here but are detailed in the User Manual and Technical Guide (Maggi, 2020) distributed in this package and accessible also at the archive repository of the BRTSim project home page https://sites.google.com/site/thebrtsimproject/download.

2. THE SOIL-WATERGRIDS DATASET DISTRIBUTION

2.1 SOIL-WATERGRIDS region of interest

Grid cells included in SOIL-WATERGRIDS were selected according to criteria described in Guglielmo et al., (2021), and resulted in about 168,000 grid cells in the region of interest for which SOIL-WATERGRIDS estimates are distributed (**Figure 1**). Details on the data distribution are:

Resolution: 0.25 x 0.25 degree (about 30 km at the equator)
Pixel density: 720 x 1440
Coordinates: standard WGS84
Bounding box: 180ºE-180ºW; 90ºS-90ºN
Years: 1970 to 2014
Frequecy: monthly and long-term monthly means

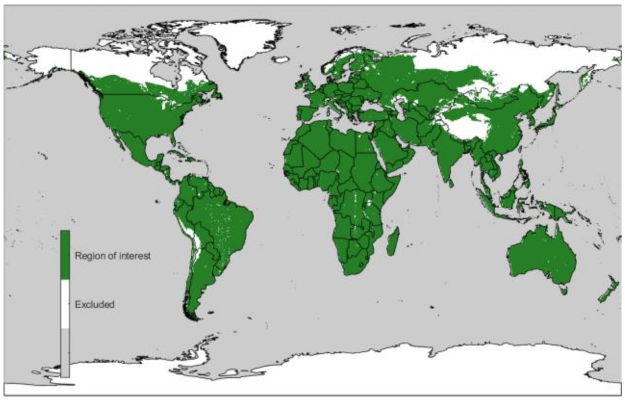

Figure 1. Region of interest of SOIL-WATERGRIDS.

2.2 Methods of estimation

SOIL-WATERGRIDS is compiled using the BRTSim v4.1a model coupled with globally gridded soil physical and hydraulic properties, land cover and use characteristics, and hydrometeorological variables to account for precipitation, ecosystem-specific evapotranspiration, snowmelt, surface runoff, and irrigation (**Figure 2**). All source data and methods of computation are described in Guglielmo et al. (2021).

To use prior information of the volumetric soil water content and water table depth, the hydraulic parameters b and ψ_S of the Brooks and Corey model (Brooks & Corey, 1964) were calculated in each grid cell *i* assuming hydraulic equilibrium, that is, the total hydraulic head is $H = \psi(S_\ell) + z$ = const. Under this assumption, the known pair of values ($S_\ell = 1$, z_{WTD}) for the water table in Fan et al. (2017) and (S_ℓ, z = 30 cm) for the saturation near surface from existing data were used to calculate *b* and ψ_S. These were next averaged with the globally-gridded values in Dai et al. (2019). Although the probability distribution of *b*, ψ_S , and ϕ values within the computational domain of SOIL-WATERGRIDS are within reported ranges, a minor fraction of grid cells (0.33%) have resulted in very high *b* values likely because of a high soil water saturation S_ℓ at the surface and a deep water table.

The computational domain along the vertical direction is extended to a depth of 56 m. However, the SOIL-WATERGRIDS distribution dataset includes only the soil water saturation at three depths of the root zone (0 to 100 cm) and the depth of the highest and lowest water tables down to a maximum of 50.5 m depth. Estimates at different depth can be calculated by third party users by deploying the modelling distribution described in Section 3 of this document.

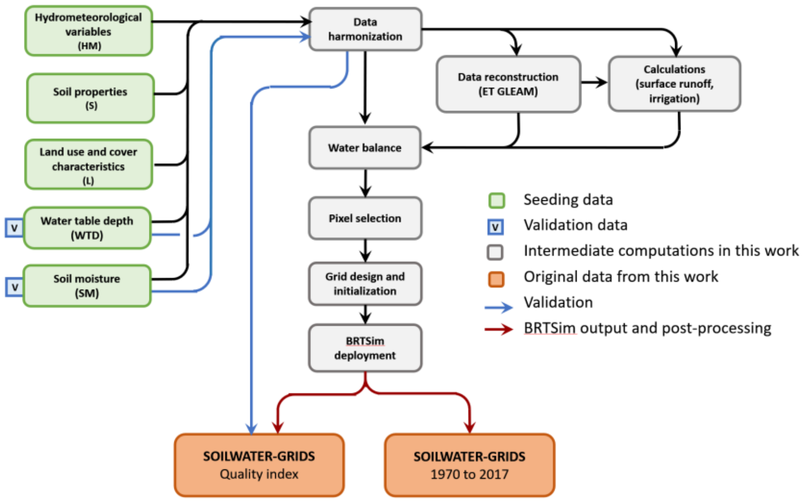

Figure 2. Conceptual workflow adopted to generate SOIL-WATERGRIDS. Acronyms in seeding data redirect to datasets in Table 1 in Guglielmo et al., (2021): HM, hydrometeorological variables; S, soil properties; L, land use and cover characteristics; WTD, water table depth; and SM, soil moisture.

2.3 SOIL-WATERGRIDS dataset structure

SOIL-WATERGRIDS includes a total of 45 files of the annual monthly soil water saturation, annual monthly depth of the water tables (highest and lowest), and the number of water tables within each grid cell of the computational domain. SOIL-WATERGRIDS also includes one file for the long-term monthly means of the same variables. SOIL-WATERGRIDS also distributes the data quality index *QI*, and extra data for the soil porosity (ϕ), air-entry suction (ψ_S), pore volume distribution index (*b*), and residual water saturation (S_{lr}) in the three layers of the root zone. These allow to calculate the volumetric soil water content and the soil water potential using the Brooks and Corey model (Brooks & Corey, 1964). The content of each distributed file is detailed in **Table 1**, while legends provided in **Table 2**.

Table 1. SOIL-WATERGRIDS data distribution files and variables.

	File Name	**Description**	**Variable**		
			Name	**Description**	**Unit**
SOIL-WATERGRIDS_NC_Distributed.zip	SOIL-WATERGRIDS_**YYYY**.nc	Contains the globally gridded data in year **YYYY** (from 1970 to 2014) of the monthly mean soil water saturation in three layers of the root zone and the monthly mean depth of the highest and lowest water tables. [(a)]When only one water table exists, these report the same value of water table depth.	Sl_0_30	Soil water saturation, 0 and 30 cm	[-]
			Sl_30_60	Soil water saturation, 30 and 60 cm	[-]
			Sl_60_100	Soil water saturation, 60 and 100 cm	[-]
			[(a)]WTDH	Depth of highest water table	[m]
			[(a)]WTDL	Depth of lowest water table	[m]
			WTDN	Number of water tables	[-]
			time	Month of the year	[-]
	SOIL-WATERGRIDS_ltm.nc	Contains the globally gridded long-term monthly mean soil water saturation in three layers of the root zone and the long-term monthly mean depth of the highest and lowest water tables	Sl_0_30_ltm	Soil saturation, 0 to 30 cm	[-]
			Sl_30_60_ltm	Soil saturation , 30 to 60 cm	[-]
			Sl_60_100_ltm	Soil saturation , 60 to 100 cm	[-]
			WTDH_ltm	Depth of highest water table	[m]
			WTDL_ltm	Depth of lowest water table	[m]
	SOIL-WATERGRIDS_qi.nc	Contains the globally gridded data quality index *QI* calculated as described in Section "Technical Validation"	QI	Quality index, 0 (worse) to 1 (best)	[-]
	SOIL-WATERGRIDS_ext.nc	Contains the soil porosity (ϕ), air-entry suction (ψ_s), pore volume distribution index (b), and and water residual saturation (*Slr*) in the three layers of the root zone. Data can be used use in combination with SOIL-WATERGRIDS_yyyy.nc and SOIL-WATERGRIDS_ltm.nc to calculate the volumetric soil water content and the soil water potential ψ using the Brooks and Corey model.	phi_0_30	0 to 30 cm	[-]
			phi_30_60	30 to 60 cm	[-]
			phi_60_100	60 to 100 cm	[-]
			psis_0_30	0 to 30 cm	[m]
			psis _30_60	30 to 60 cm	[m]
			psis_60_100	60 to 100 cm	[m]
			b_0_30	0 to 30 cm	[-]
			b_30_60	30 to 60 cm	[-]
			b_60_100	60 to 100 cm	[-]
			Slr_0_30	0 to 30 cm	[-]
			Slr_30_60	30 to 60 cm	[-]
			Slr_60_100	60 to 100 cm	[-]
	Read_NC.m	Editable script written in Matlab 2019b to read and represent .NC files (example).			

Table 2. Legend of variables in SOIL-WATERGRIDS data distribution.		
Variable	**Legend**	
Sl_0_30, Sl_30_60, Sl_60_100 Sl_0_30_ltm, Sl_30_60_ltm, Sl_60_100_ltm Slr_0_30, Slr_30_60, Slr_60_100	-2 -1 [0, 1]	Ocean No data Range of values
WTDL, WTDH, WTDL_ltm, WTDH_ltm	-2 -1 0 [0.15, 50.5] 100	Ocean No data Ponding Range of values Below 50.5 m depth
WTDN	-2 -1 [0, 6]	Ocean No data Range of values
QI	-2 -1 [0, 1]	Ocean No data Range of values, 1 is best
b_0_30, b_30_60, b_60_100	-2 -1 [1.2, 112]	Ocean No data Range of values
phi_0_30, phi_30_60, phi_60_100	-2 -1 [0.1, 0.80]	Ocean No data Range of values
psis_0_30, psis _30_60, psis_60_100	2 1 [-1.60,0]	Ocean No data Range of values

2.4 Visualization of SOIL-WATERGRIDS data

There are a number of tools to visualize and elaborate on the distributed .NC files including Panoply, which is distributed free of charge from the NASA/GISS at https://www.giss.nasa.gov/tools/panoply/download/. Alternatively, the free and open source QGIS software can be used to open and process .NC files and can be downloaded at https://www.qgis.org/en/site/. In addition to third party software, we provide a simple editable script written in @Matlab 2019b that reads and represent a variable from a chosen .NC file.

3. THE SOIL-WATERGRIDS MODELLING DISTRIBUTION

3.1 SOIL-WATERGRIDS modelling data structure

The modelling package is contained in 7 .ZIP files named BRTSim_SOIL-WATERGRIDS_0p25x0p25_*.zip. Each of these files is associated to a continental region and is container of additional levels of folders. Specifically, they contain folders HPC and RNF, the former used to run the model in all grid cells of a region, and the latter containing the tables of runoff from and to adjacent grid cells (**Table 3**).

Inside the HPC folders are a number of subfolders that include the BRTSim executable, bash file, and BRTSim license (BRTSim_v41a_GLNXA64_R2019b, run_BRTSim_v41a_GLNXA64_R2019b.sh, license.txt), the Param_*.inp files to instruct BRTSim to run each grid cells allocated to the subfolder, and the tables of boundary conditions (Table_*.txt, and WB_*.txt). These subfolders also contain the bash file to launch in a sequential way all grid cells allocated to a subfolder (run_BRTSim_R*.sh). The license file is valid until the end of 2021. A new license can be obtained with no charge from the BRTSim download page (https://sites.google.com/site/thebrtsimproject/download) after expiry. The number of grid cells in the subfolders is variable.

Inside the RNF folders are a number subfolders that include tables of the runoff applied to each grid cell. These tables are aggregated by hydrologic units identified by the number appearing in the file name.

This package provides all data required to model about 168,000 grid cells globally.

Table 3. SOIL-WATERGRIDS modelling distribution folders and file tree.

Region	File Name	Folders	Number of subfolders
Africa	BRTSim_SOIL-WATERGRIDS_0p25x0p25_AF.zip	HPC.rar	119
		RNF.rar	2030
Asia Insular	BRTSim_SOIL-WATERGRIDS_0p25x0p25_AI.zip	HPC.rar	119
		RNF.rar	1325
Asia	BRTSim_SOIL-WATERGRIDS_0p25x0p25_AS.zip	HPC.rar	120
		RNF.rar	3309
Europe	BRTSim_SOIL-WATERGRIDS_0p25x0p25_EU.zip	HPC.rar	95
		RNF.rar	2223
North America	BRTSim_SOIL-WATERGRIDS_0p25x0p25_NA.zip	HPC.rar	120
		RNF.rar	2531
Oceania	BRTSim_SOIL-WATERGRIDS_0p25x0p25_OC.zip	HPC.rar	120
		RNF.rar	985
South America	BRTSim_SOIL-WATERGRIDS_0p25x0p25_SA.zip	HPC.rar	120
		RNF.rar	1680

3.2 File naming

Running the BRTSim solver on each grid cell requires a Param_*.inp file, a table Table_*.txt, table WB_*.txt, and tables RNF_*.txt (optional) depending on the hydrologic settings relative to runoff. Naming of those files is as follows:

Param_**ID**_py**Y**-720_px**X**-1440_lat**LAT**_lon**LON**_**BEDWAT**_**RNF**.inp
Table_ **ID** _py**Y**-720_px**X**-1440_lat**LAT**_lon**LON**_**BEDWAT**_**RNF**.txt
WB_ **ID** _py**Y**-720_px**X**-1440_lat**LAT**_lon**LON**_**BEDWAT**_**RNF**.inp
RNF_ **A** _**ID**.txt (optional)
RNF_ **ID** _**B**.txt (optional)

where **ID** is the unique identifier for the grid cell, **Y** is the matrix coordinate in the N-S orientation, **X** is the matrix coordinate in the E-W orientation, **LAT** is the latitude (-90 to 90), **LON** is the longitude (-180 to 180), **BEDWAT** is either “bedrock” or “watertable” and indicates if the bottom layer of the grid cell is a bedrock or has a saturated aquifer, respectively, while **RNF** is either empty, “null”, or “ocean” to indicate if the grid cell discharges runoff water to some other grid cells, if the grid cell does not discharge runoff water, or whether it discharges runoff water to the ocean, respectively, **A** is the source identifier for runoff into **ID,** and **B** is the recipient identifier for the runoff from **ID**.

3.3 Supported operating system

The modelling and modelling workflow is designed to operate on a UNIX system (GLNXA64) high performing computing (HPC) cluster with all subfolders in folder HPC running in parallel. For operating the model, BRTSim requires that the Matlab MCR9.7 (Matlab 2019b) is installed on the operating system. The MCR is distributed free at the MathWorks home page for LINUX (https://au.mathworks.com/products/compiler/matlab-runtime.html).

3.4 Launching and workflow

Launching of the model on a UNIX HPC cluster depends on the scheduler adopted for parallelization and queue management, hence we do not distribute the job launchers because these contain confidential information and are system dependent for memory, walltime, and number of cores. The user must therefore write job launchers (typically .PBS files) and call the bash files in each subfolder (run_BRTSim_R*.sh). The workflow is embedded in the bash files inside each subfolder (run_BRTSim_R*.sh) and can be redesigned by the user depending on space availability on their systems. In total, the compressed full-size model outputs occupy about 990GB of hard drive (about 10TB when uncompressed).

3.5 Accompanying sample and scripts

We distribute along with all input files for global modelling also one sample grid cell (Sample_Grid_Cell.ZIP) and a few editable codes written in Matlab R2019b that can be modified by the user to read the model output of individual grid cells. The sample grid cell can be run using the BRTSim computational environment on WIN64 OS systems (BRTSim_v41a_WIN64_R2019b.exe) and requires the Matlab 2019b MCR be installed (can be downloaded from the Mathworks free of charge). To run the sample grid cell, the user can launch the batch file (.BAT). Scripts can be launched after the sample grid cell has finished running. Details are given in **Table 4**.

Table 4. Editable codes distributed in this package for reading modelling outputs.

Editable code	Type	Purpose
ReadTime.m	Function	Read the Time_*.out file and organizes all state variables by time and by node (space) for ease of data manipulation.
plotPro.m	Script	Plot profiles of user-selected state variables over space and different times. It calls the ReadTime.m function.
plotWTD.m	script	Calculates and represents the depth of the water table/s over time. It calls the ReadTime.m function.

References

Brooks, R. H., & Corey, A. T. (1964). Hydraulic properties of porous media and their relation to drainage design. Transactions of the ASAE, 7, 26-28.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., . . . Lu, X. (2019). A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling. *Journal of Advances in Modeling Earth Systems, 11*, 2996-3023.

Fan, Y., Miguez-Macho, G., Jobbágy, E. G., Jackson, R. B. & Otero-Casal, C. Hydrologic regulation of plant rooting depth. *PNAS*, **114**, 10572-10577 (2017).

Guglielmo, M., Tang, F. H. M., Pasut, C., Maggi, F., (2021). Soil-WATERGRIDS, mapping dynamic changes in soil moisture and depth of water table from 1970 to 2014. Scientific Data.

Maggi, F. (2020). BRTSim v4.1a, A general-purpose multiphase and multispecies computational solver for biogeochemical reaction-advection-dispersion processes in porous and non-porous media. User Manual and Technical Guide, pp 85, November 2020, accessible at https://sites.google.com/site/thebrtsimproject/home