



OPEN

DATA DESCRIPTOR

# SOIL-WATERGRIDS, mapping dynamic changes in soil moisture and depth of water table from 1970 to 2014

Magda Guglielmo, Fiona H. M. Tang , Chiara Pasut & Federico Maggi 

We introduce here SOIL-WATERGRIDS, a new dataset of dynamic changes in soil moisture and depth of water table over 45 years from 1970 to 2014 globally resolved at  $0.25 \times 0.25$  degree resolution (about  $30 \times 30$  km at the equator) along a 56 m deep soil profile. SOIL-WATERGRIDS estimates were obtained using the BRTsim model instructed 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. We validate our estimates against independent observations and re-analyses of the soil moisture, water table depth, wetland occurrence, and runoff. SOIL-WATERGRIDS brings into a single product the monthly mean water saturation at three depths in the root zone and the depth of the highest and lowest water tables throughout the reference period, their long-term monthly averages, and data quality. SOIL-WATERGRIDS can therefore be used to analyse trends in water availability for agricultural abstraction, assess the water balance under historical weather patterns, and identify water stress in sensitive managed and unmanaged ecosystems.

## Background & Summary

Groundwater provides about 33% of total water withdrawals worldwide<sup>1</sup>; this fulfils about 35% of the global drinking water demand<sup>2</sup> and about 40% of irrigation water in agricultural areas<sup>3,4</sup>. However, continued groundwater abstraction can decrease the storage and threaten ecosystems and communities that depend on below-surface water, especially in arid and semiarid regions<sup>4-6</sup>. For instance, near surface water tables can regulate the soil water content at the surface through capillary rise<sup>7</sup>, supply water to the root zone supporting plant transpiration, and alleviate soil evaporation not only in arid and semiarid regions<sup>8-10</sup> but also where seasonal rainfalls alternate with long dry periods<sup>11,12</sup>. In addition to anthropogenic factors, climate variability is a concurrent threat to the groundwater availability because it can alter the long-term aquifer and vadose zone water budget and can influence, as a consequence, the making of groundwater use policies<sup>13,14</sup>. Knowing the state of groundwater globally is essential to identify where groundwater provides vital support to the ecosystems and their environmental services<sup>15</sup>. Near-surface soil water measurements are usually conducted through borehole sampling. Ideally, the greater the density of monitoring stations per surface area, the greater the data reliability and usability. However, the monitoring infrastructure is expensive and requires continuing maintenance and manual data logging in the majority of cases. Despite a large number of boreholes worldwide (more than 1.6 million<sup>15</sup>), the density per surface area is not enough to produce spatially uniform assessments of the actual groundwater storage over wide geographical regions.

Currently, several georeferenced gridded datasets are available for global water resources. The most extensive catalogue of the groundwater table depth is provided by<sup>10</sup>, which combines global observations from government archives and the literature with estimates by hydrological models. Although this dataset distributes the equilibrium (long-term mean) water table depth and its long-term monthly averages from 2004 to 2013, it does not provide information about the surface and root-zone soil water content, which is fundamentally dependent on the

Laboratory for Advanced Environmental Engineering Research, School of Civil Engineering, The University of Sydney, Bld. J05, 2006, Sydney, NSW, Australia. e-mail: [magda.guglielmo@sydney.edu.au](mailto:magda.guglielmo@sydney.edu.au); [fiona.tang@sydney.edu.au](mailto:fiona.tang@sydney.edu.au); [c.pasut@sydney.edu.au](mailto:c.pasut@sydney.edu.au); [federico.maggi@sydney.edu.au](mailto:federico.maggi@sydney.edu.au)

proximity to the water table. On the other hand, existing satellite data provide distributed records of the soil water content at different depths near-surface (generally within the top 30 cm<sup>16,17</sup>), while soil moisture sensor measurements provide recordings to deeper depths<sup>18</sup> but their spatial coverage is often limited to more or less extended regions. Advanced land surface models combined with data assimilation techniques (e.g., GLDAS model in<sup>19</sup>) have contributed to expanding soil moisture data by filling in spatial and temporal gaps.

The datasets mentioned above provide invaluable information on the global water asset. However, the existing estimates of water table depth and near-surface soil moisture are independent from each other and independently validated. Also, these have a different level of detail (e.g., resolution, depth) and report different target variables (e.g., volumetric soil water content, soil water saturation, water table depth). Most importantly, no set of data currently exists for the time- and depth-resolved soil water content or saturation in the topsoil, root zone, and vadose zone and the water table depth (one or multiple) over the same geographic grid. The lack of knowledge of the water connectivity between near-surface and groundwater compartments hampers a full-scale assessment of the global surface and groundwater assets and introduces uncertainty in the understanding of how the water resources in the deep aquifer depend on the surface water availability, and *vice versa*. To improve assessments of the way surface water and groundwater are ultimately coupled and how they influence each other, datasets should provide both the soil moisture and depths of the water tables including those that form temporarily. Cross-validations against multiple independent datasets can provide reliable references to make interpretations, especially where observations are limited.

To overcome these limitations, we produced the SOIL-WATERGRIDS dataset, which was constructed by re-analysis of existing global scale datasets and by coupling these datasets with the BRTSim (BioReactive Transport Simulator<sup>20</sup>) mechanistic model for the geophysical flow of water. BRTSim was fed with globally gridded soil physical and hydraulic properties, land cover and use characteristics, and hydrometeorological data harmonised to a resolution of  $0.25 \times 0.25$  degree per grid cell (about 30 km at the equator). BRTSim resolves the variably-saturated water flow over a vertical profile from land surface to 56 m depth and covers a time frame of 45 years from 1970 to 2014. We validated the SOIL-WATERGRIDS modelling estimates of volumetric soil water content against globally-gridded satellite data (ESA/CCI<sup>21–23</sup>), modeling reconstructions and re-analyses (GLDAS<sup>19,24,25</sup>), and distributed measurements (ISMN network<sup>18,26</sup>). We validated the water table depth estimates against selected groundwater data in<sup>10</sup>. These validations are also used to determine the data quality index of SOIL-WATERGRIDS. In addition to those validations, we benchmarked the location of ponding occurrence against the SWAMPS dataset for wetlands<sup>27</sup> and our estimates of surface runoff against re-analyses in the GRUN dataset<sup>28</sup>.

## Methods

**Seeding databases and data pre-processing.** For the development of SOIL-WATERGRIDS, we deployed the BRTSim model<sup>20</sup> on a global grid constructed using existing georeferenced data. Details of all the datasets are summarised in Online-only Table 1 and described below.

The seeding data for the soil physical properties include the soil texture (sand, silt, and clay fractions) and bulk density from SoilGridsv2.0<sup>29</sup>, and porosity from SoilGridsv1.0<sup>30</sup>. The soil hydraulic properties include the permeability, the pore volume distribution index, and the air-entry suction of the Brooks and Corey model<sup>31,32</sup>, and the residual soil water content<sup>33</sup>. The land cover type in year 2014 was taken from MODIS/IGBP 2019<sup>34,35</sup>, while the ecosystem-specific root depth profiles for crops and native ecosystems are from<sup>36,37</sup>, respectively. The water security indicators and crop calendars<sup>38–40</sup> were used to estimate the irrigation volumes as in<sup>36</sup>. The seeding hydrometeorological data include the historical global daily precipitation from the Climatic Research Unit gridded Time Series dataset from 1970 to 2018 (CRU/TS<sup>41</sup>), the daily actual evapotranspiration from the Global Land Evaporation Amsterdam Model from 1980 to 2017 (GLEAM<sup>24,25</sup>), the monthly potential evapotranspiration from the CRU/TS, and the monthly snowmelt data from Global Land Data Assimilation System (NOAH/GLDAS<sup>19</sup>). We also used the soil water content and saturation of the NOAH/GLDAS Version 2, satellite data from the European Space Agency, Climate Change Initiative for Soil Moisture (hereafter ESA/CCI<sup>21–23</sup>), data-assimilation soil moisture from the Global Land Evaporation Amsterdam Model (GLEAM, Martens *et al.*, 2017<sup>25</sup>), the ground-based soil moisture measurements from the International Soil Moisture Network (ISMN<sup>18,26</sup>), and the long-term average and long-term monthly averages of the water table depth in<sup>10</sup>. Finally, we used the SWAMPS dataset for wetlands<sup>27</sup> and the monthly total surface runoff from the GRUN dataset for additional benchmarking<sup>28</sup>.

The workflow used to generate the SOIL-WATERGRIDS is represented in the Supplementary Information, Figure S1. Heterogeneous seeding datasets were first harmonised to the chosen resolution of  $0.25 \times 0.25$  degree per grid cell, which corresponds to about  $30 \times 30$  km at the equator. The resolution harmonisation was conducted with a conservative (phycnophylactic) linear interpolation for the quantities explicitly representing mass or mass fluxes (e.g., precipitation, evapotranspiration, and others), or with a simple interpolation otherwise (e.g., soil physical properties, and others). Conservation of the quantity of interest was achieved by first calculating the areal integral at the original resolution and by rescaling the resized data to the areal integral. Data resizing was conducted for either type of variables by linear interpolation with the *imresize* function in @Matlab2019b. This method achieved conservation of mass and average values in all climatic regions with a percent relative error smaller than 0.01% (see sample validation in Supplementary Information, Figure S2). Harmonised data were used in intermediate steps for data reconstruction and calculation of the surface runoff and water budget in each grid cell.

Time series reconstruction was implemented for the GLEAM actual evapotranspiration (*ETA*) in years 1970 to 1979 because data in those years are not available. To this end, we used the potential evapotranspiration (*PET*) of the CRU/TS datasets. Because *ETA* and *PET* are from two independent sources, we tested data consistency by counting the number of times *ETA* is greater than 1.2 *PET*, where 1.2 is the excess evapotranspiration in cropping

systems identified by the FAO method<sup>38</sup>. We found inconsistencies in 5.8 to 7.4% of grid cells within the computational domain across years from 1980 to 2014 (Supplementary Information, Figure S3), where both  $ETA$  and  $PET$  exist, but we have not applied adjustments to the original or harmonized data. The actual monthly evapotranspiration,  $ETA_{m,i}$  (mm month<sup>-1</sup>), in month  $m$  and grid cell  $i$  was reconstructed as

$$ETA_{m,i} = r_{m,i} \times PET_{m,i} \quad (1)$$

where  $PET_{m,i}$  is the monthly potential evapotranspiration in grid cell  $i$  available from CRU/TS and  $r_{m,i}$  is the ratio of the long-term monthly average actual evapotranspiration  $\overline{ETA}_{m,i}$  to the long-term monthly average potential evapotranspiration  $\overline{PET}_{m,i}$  in the same month  $m$  and grid cell  $i$  throughout the years where both datasets exist. Comparison of  $ETA$  reconstructed using Eq. (1) and the long-term monthly means available from the GLEAM dataset returned  $R = 0.99$  and  $p < 0.01$  (Supplementary Information, Figure S4). The actual evapotranspiration used as a boundary condition in BRTSim can be interpreted as a potential evapotranspiration flux because water can only be removed by the plant root system if water is available at a saturation above residual.

The daily snowmelt rate  $SM_i$  (mm day<sup>-1</sup>) in each grid cell  $i$  was calculated from the available monthly  $SM_{m,i}$  estimates in NOAA/GLDAS and was next combined with the daily precipitation  $P_i$  to calculate the daily surface runoff  $Q_i$  (mm day<sup>-1</sup>) in each grid cell  $i$  following the Curve Number method (NRCS-CN<sup>42</sup>) as

$$Q_i = \frac{(P_i + SM_i - I_{a,i})^2}{P_i + SM_i - I_{a,i} + S_i} \quad (2)$$

$$S_i = \frac{25,400}{CN_i} - 254 \quad (3)$$

where  $I_{a,i} = 0.2S_i$  is the initial abstraction,  $S_i$  is the maximum water retention capacity of the soil before releasing water, and  $CN_i$  is the curve number at fair moisture conditions tabulated in the NRCS-CN technical report as a function of soil type, hydrological characteristics, and land use. In addition to these classifications, the normalized antecedent precipitation index ( $NAPI^{43}$ ) was introduced to calculate the soil moisture based on the precipitation in the previous days as

$$NAPI = \frac{\sum_{t=-1}^{T_p} P(t)k_{CN}^{-t}}{P_{avg} \sum_{t=-1}^{T_p} k_{CN}^{-t}} \quad (4)$$

where  $T_p = 5$  is the number of antecedent days recommended in the NRCS-CN,  $k_{CN} = 0.85$  is the decay constant of the soil moisture over time<sup>44</sup>,  $P(t)$  is the precipitation at day  $t$ , and  $P_{avg}$  is the annual average precipitation.  $NAPI$  changes over time and in each grid cell. Soil in a grid cell is considered “dry” if  $NAPI < 0.33$ , “wet” if  $NAPI > 3$ , and at “fair condition” otherwise. Finally, the precipitation dependent curve number  $CN_i$  in Eq. (3) was written as

$$CN_i^d = \frac{CN_i}{2.281 - 0.0128CN_i}; \quad CN_i^w = \frac{CN_i}{0.427 - 0.00573CN_i} \quad (5)$$

where superscripts  $d$  and  $w$  indicate dry and wet conditions, respectively. The runoff flow direction was derived from the steepest down-slope gradient calculated from the digital elevation map<sup>45</sup>. Finally, the travel time of  $Q_i$  from the source grid cell  $i$  to the recipient grid cell was calculated as the time of concentration in days  $T_{c,i}$  using the watershed lag method of the NRCS-CN as

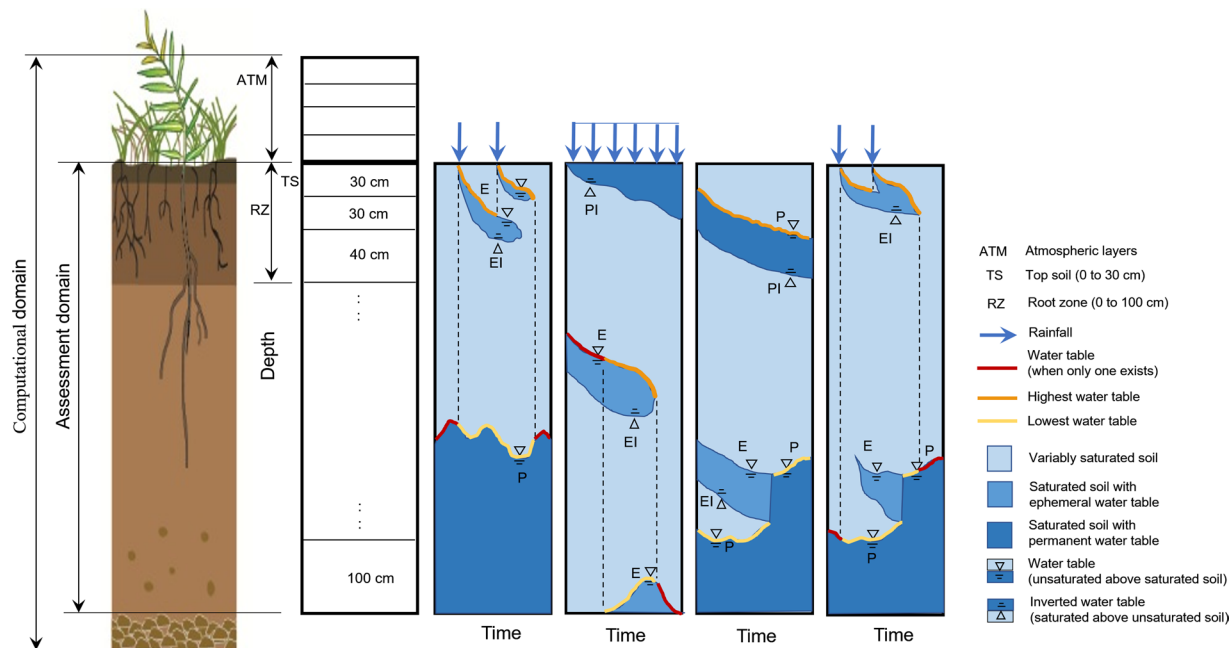
$$T_{c,i} = \frac{1}{24} \frac{L_i^{0.8} (S_i + 1)^{0.7}}{1, 140 Y_i^{0.5}} \quad (6)$$

where  $L_i$  is the distance in feet between the centroids of the source grid cell and its recipient, and  $Y_i$  is the average slope between the source and recipient grid cells. In our implementation,  $T_{c,i}$  is specific to each grid cell and is used to allocate the fraction of generated runoff,  $f_Q = \min\{1, 1/T_{c,i}\}$ , eventually leaving the source grid cell in a day. The remaining generated runoff,  $(1-f_Q)Q$ , is applied on the subsequent day using again  $f_Q$ ; this procedure is repeated until 99.9% of the initial  $Q_i$  is allocated through time. The daily runoff rate  $Q_i$  is not subtracted from  $P_i$  and  $SM_i$ , but is applied as an outlet flux from the topsoil of grid cell  $i$ . We recall that the NRCS-CN method has limitations on the accounting of snowmelt, the aspect that we address in “Technical Validation”.

Finally, the calculation of the daily irrigation rate consisted in identifying the grid cells where irrigation is applied using the crop water security indicators in<sup>39</sup>, while the crop calendars in<sup>40</sup> were re-interpreted to calculate the daily gridded normalised crop calendar and identify the crop irrigation period<sup>36</sup>. In an irrigated grid cell, the daily irrigation rate is calculated as the gap  $I_i = f(1.5ETA_i - P_i)$  between precipitation  $P_i$  and actual evapotranspiration  $ETA_i$  during the crop irrigation period with  $f = 0.8$  or  $f = 0.2$  depending on whether irrigation is major or minor, respectively, according to the classification in<sup>39</sup>.

The hydrometeorological data were used to estimate the decadal water balance,  $W_p$ , in grid cell  $i$  from 1970 to 2014 as

$$W_i = \sum P_i - \sum ETA_i + \sum SM_i + \sum Q_i + \sum I_i \quad (7)$$



**Fig. 1** Computational and assessment domains (a), and their discretisation in the BRTSim modelling (b) with examples of water table types (c to f). Letters P and E indicate “permanent” and “ephemeral” water tables, the former existing for the entire assessment period and the latter existing for a shorter period. The additional letter I associated to E and P (i.e., PI and EI) indicates inverted water tables. Water tables marked in yellow and orange are the lowest and highest, respectively, while those in red indicate cases when only one water table exists. The water tables marked with these colors are distributed in SOIL-WATERGRIDS. The assessment period is from 1970 to 2014. Drawing of the soil profile on the left is elaborated from <https://www.qld.gov.au/environment/land/management/soil/soil-explained/forms>.

where the terms on the right-hand side are the cumulative precipitation, actual evapotranspiration, snowmelt, net surface runoff (input minus output), and irrigation in each decade and in each grid cell. The irrigation term,  $I$ , is only accounted for in grid cells where the water table depth is below the depth of our computational domain of 56 m. This is because we assumed irrigation abstractions to occur from within the computational domain in grid cells with a water table above 56 m depth, and therefore the water abstraction from the aquifer and applied at land surface was not accounted for in the water budget of Eq. (7) as water is recirculated. Equation (7) is used to condition the boundary flows in each grid cell with an additional flow of water lasting for each decade of the reference period (positive or negative depending on the grid cell) to prevent the soil column defined by finite volumes in BRTSim to become fully saturated or fully depleted. Inclusion of this decade-specific water balance flow does not affect the intra day and intra seasonal fluctuations in soil water content and water table depth, and compensates for the horizontal groundwater flow not explicitly included in our modeling (see also Section “BRTSim modelling and outputs”). Estimates in SOIL-WATERGRIDS do not explicitly account for river flow discharge within a watershed and only include an approximation to the snowmelt flow into runoff. An analysis and benchmarking of our approach and estimates against existing runoff data is expanded in Section “Technical Validation”.

**Grid cell selection, design, and initialisation.** The bounding box of SOIL-WATERGRIDS extends between longitudes  $-180^{\circ}\text{W}$  and  $180^{\circ}\text{E}$  and between latitudes  $-55.5^{\circ}\text{S}$  and  $70.5^{\circ}\text{N}$ . A total of about 168,000 grid cells of interest were selected according to the following criteria: (1) we excluded grid cells with missing data using the information provided in the seeding datasets (see Section “Seeding databases and data pre-processing”); (2) we excluded grid cells above the permanent snow line at 3,500 m elevation using the digital elevation map in<sup>45</sup> and grid cells with more than 10% permafrost area using the NEO/NASA data in<sup>46</sup> to reduce biases caused by frozen soil, which is not explicitly described in our modelling and for which no data exist for validating the water table depth and surface soil moisture; and (3) we excluded grid cells with more than 10% wetland area based on the MODIS/IGBP land cover classification because we did not explicitly model wetlands hydraulics in this work even if we track the occurrence of ponding in our computational grid (more detail on ponding is provided in Section “Technical Validation”). Runoff water from and to grid cells neighbouring the computational domain was explicitly accounted for to minimise biases on water budget at the boundary of the computational domain. Grid cells included in the computational domain are represented in Figure S5 of Supplementary Information.

The computational grid was designed to describe an archetype 56 m deep soil column with a sequence of 64 finite volumes with varying thickness vertically stratified in each grid cell (Fig. 1a,b). The top soil (TS, 0 to 30 cm) consists of the first soil layer of 30 cm thickness; the root zone (RZ, 0 to 100 cm) is defined by the first three soil layers of 30, 30, and 40 cm thickness; the layers from 1 to 3 m depth have 40 cm thickness, from 3 to 6 m depth

have 50 cm thickness, and from 6 to 56 m depth have 1 m thickness. We included four atmospheric layers of 30 cm thickness above the soil to allow for occasional water ponding.

The soil textural properties (sand, silt, and clay fractions), bulk density, and porosity for each soil layer from the surface to 2 m depth were assigned using data in SoilGrids<sup>30</sup> interpolated at the depths describing the soil column in our computational domain. The soil properties below 2 m depth were assigned using the values of the closest upper known depth. The soil permeability,  $k$ , pore volume distribution index,  $b$ , and air-entry suction,  $\psi_s$ , of the Brooks and Corey model were assigned using data in<sup>32</sup> for all soil layers from the surface down to 2 m depth after interpolation at the depths of our computational domain. Below 2 m depth, the permeability,  $k$ , was calculated with a negative exponential function of the depth,  $z$ , similar to the approach in<sup>47</sup>

$$k = k_{2m} e^{-z/\beta} \quad (8)$$

where  $k_{2m}$  is the permeability at 2 m depth and  $\beta$  is an empirical parameter calculated assuming that the permeability  $k_{55.5m}$  at 55.5 m depth (i.e., the center of mass of the deepest layer in our discretisation) is one-tenth of  $k_{2m}$  after trial and error testing on randomly selected grid cells, hence,  $\beta = z_{55.5m}/\ln(10) = 24.1$  m. The hydraulic parameters  $b$  and  $\psi_s$  were calculated in each grid cell  $i$  assuming hydraulic equilibrium (i.e., the total hydraulic head is  $H = \psi(S_\ell) + z = \text{const}$ ) at the depth  $z_{WTD}$  of the known water table and at the soil surface for the known soil saturation  $S_\ell$ . The known  $z_{WTD,i}$  in grid cell  $i$  was taken from<sup>10</sup> while  $S_\ell$  at the soil surface was initialised to the average value of data in GLEAM, NOAA/GLDAS, and ESA/CCI. The values of  $b$  and  $\psi_s$  calculated in this way were next averaged with the globally-gridded values in<sup>32</sup> at the depths between the soil surface and 2 m depth. Hence, the initial soil water saturation profile produced values at the surface close to available soil moisture datasets, and water table depths close to those in<sup>10</sup>. The liquid residual saturation,  $S_{lr}$ , from<sup>47</sup> was therefore adjusted in some grid cells to satisfy the condition  $S_{lr} < S_\ell$ , while the gas residual saturation  $S_{gr} = 0.1$  was used in all soil layers and grid cells. An analysis of the actual values of  $b$  and  $\psi_s$  used in our modeling show that corrections introduced here do not substantially skew their frequency distributions (Supplementary Information, Figure S6).

**BRTSim modelling and outputs.** BRTSim-v4.1a is a general-purpose solver of reaction-advection-dispersion processes in variably-saturated soil systems. BRTSim uses a hybrid explicit-implicit finite volume scheme to solve the Richards equation<sup>48</sup> and describes the variably-saturated water flow along the vertical soil profile with heterogeneous soil physical and hydraulic properties (details available in the User Manual and Technical Guide<sup>20</sup>). For SOIL-WATERGRIDS, the relative permeability-water potential-water saturation relationships of the Brooks-Corey model were used with parameters assigned as described in Section “Grid cell selection, design, and initialisation”. The core solvers in BRTSim are fundamentally one dimensional; however, the parallelisation implemented in this work allowed us to use it over the global grid as a pseudo three-dimensional solver accounting for a one-way horizontal water flow representing surface runoff.

The BRTSim modelling outputs were post-processed to calculate the monthly mean soil water saturation,  $S_\ell$ , and volumetric water content,  $\theta$ , in the three layers of the root zone, and the water table depth, WTD, ranging from land surface to a depth of 50.5 m (corresponding to the center of mass of that soil layer), that is, the bottom five layers down to 56 m depth were excluded from assessment to reduce biases from boundary effects (Fig. 1b).

Relative to the water table, the solution of the variably-saturated Richards equation allowed us to describe dynamic changes along the vertical profile, including the formation of multiple water tables stratified at different depths. Our modelling outputs showed that up to 6 water tables occurred in a minor fraction of grid cells for short periods of time. After *a posteriori* inspection, we found that all water tables not known from prior information at the initialisation of our modelling assessment appeared after intense precipitation events and were generally located above the known water table. Hence, our assessment can include permanent water tables with variable depth but lasting for the entire duration of our assessment (marked with P in Fig. 1) and ephemeral water tables with variable depth (marked with E in Fig. 1). Inverted water tables, characterised by the saturated soil above the unsaturated soil, are associated with all instances in which ephemeral water tables develop (marked with EI, Fig. 1). Permanent inverted water tables (marked with PI in Fig. 1d) are highly unlikely over the 45 years of assessment within the geographic region of interest and we have not found occurrence of this type of water table. Ephemeral water tables can also raise if the bottom unsaturated soil layer becomes saturated for a short period of time. Both ephemeral and permanent water tables can co-exist at any time of our assessment and can also combine with each other or separate into multiple saturated lenses (Fig. 1e,f). Because our modelling can describe a wide variety of forming, persisting, and fading water tables, we identified and counted all water tables (P and E types) but, for simplicity, we distribute in SOIL-WATERGRIDS only the highest (closest to the surface) and lowest (deepest from the surface) when multiple water tables co-exist. SOIL-WATERGRIDS also identifies grid cells where ponding occurs. To the best of our knowledge, SOIL-WATERGRIDS is the only data product that releases these assessments of the water table dynamics.

## Data Records

The SOIL-WATERGRIDS distribution dataset includes NetCDF formatted files available for download from the Zenodo repository<sup>49</sup>. These files contain the variables described in detail in Table 1 with legends expanded in Table 2. In addition to the distribution dataset, we make available the full global-scale modelling inputs (deployable on a UNIX computer cluster) from the Zenodo repository<sup>49</sup> as described in Section “Code Availability”. Additional information on both distributed dataset and modelling is available in the Technical Documentation in the Zenodo repository<sup>49</sup>.

	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. <sup>(a)</sup> When only one water table exists, these report the same value of water table depth.	SL_0_30	Soil water saturation, 0 to 30 cm	[-]
			SL_30_60	Soil water saturation, 30 to 60 cm	[-]
			SL_60_100	Soil water saturation, 60 to 100 cm	[-]
			<sup>(a)</sup> WTDH	Depth of highest water table	[m]
			<sup>(a)</sup> 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 <i>QI</i> calculated as described in Section “Technical Validation”	QI	Quality index, 0 (worse) to 1 (best)	[-]
	SOIL-WATERGRIDS_ext.nc	Contains the soil porosity ( $\phi$ ), air-entry suction ( $\psi$ ), pore volume distribution index ( $b$ ), and water residual saturation ( $Slr$ ) in the three layers of the root zone. Data can be used in combination with SOIL-WATERGRIDS_YYYY.nc and SOIL-WATERGRIDS_ltm.nc to calculate the volumetric soil water content and the soil water potential $\psi$ 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_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 1.** SOIL-WATERGRIDS data distribution files and variables. All files included in the data distribution are compressed and contained in file SOIL-WATERGRIDS\_NC\_Distributed.zip available from<sup>81</sup> and<sup>49</sup>.

## Technical Validation

The quality of the distributed dataset is measured at two levels: the first, called validation, refers to the direct comparison of the target variables distributed in SOIL-WATERGRIDS against other existing data. The second, called benchmarking, refers to the comparison against other existing data of intermediate variables not distributed in the SOIL-WATERGRIDS dataset but instrumental to the estimates in SOIL-WATERGRIDS.

**Methods of validation and benchmarking.** The volumetric soil water content  $\theta$  in TS (0 to 30 cm) and RZ (0 to 100 cm), and water table depth WTD estimated in SOIL-WATERGRIDS were validated against existing independent data listed in Table 3 for the same variables. For both SOIL-WATERGRIDS and the validation datasets, we retrieved or calculated the long-term mean (LTM) and long-term monthly mean (mLTM) depending on data frequency. Validation against some datasets was limited to the available geographic regions such as for ESA/CCI, ISMN, and data in<sup>10</sup>. The metrics to measure the quality of SOIL-WATERGRIDS estimates consisted in the anomaly  $\Delta X_Y = X - Y$  of the SOIL-WATERGRIDS variables  $X$  against the corresponding validation set  $Y$ , the normalised root mean square deviation NRMSD, the spatial and temporal Pearson's correlations  $R$ , and the Duvellier coefficient  $\lambda$  accounting for data biases<sup>50</sup>. These metrics are defined in Table 4 and were calculated for either LTM, or mLTM, or both data depending on data availability. Of these metrics,  $\Delta X_Y$  and  $R$  were used to calculate the data quality index  $QI$  as described later in Section “Data quality”.

We benchmarked intermediate data not directly distributed with SOIL-WATERGRIDS. Specifically, the geographic occurrence of ponding was benchmarked against the SWAMPS wetlands dataset, while the calculated runoff and water balance were benchmarked against the GRUNv1 dataset. These benchmarking were not included in the data quality index as ponding and runoff are not the target variables distributed in SOIL-WATERGRIDS, but further supported our working approach in a quantitative way.

**Validation of the soil water content.** We compared the long-term mean volumetric water content  $\theta$  in TS against available re-analyses from NOAA/GLDAS and GLEAM, satellites data provided by ESA/CCI, and ground-logged readings from 39 networks and a total of 1,766 data points from 28 countries of the ISMN covering a maximum time span from 1952 to 2014 (Table 5). A similar comparison was repeated for  $\theta$  in RZ against the corresponding quantity in the GLEAM and NOAA/GLDAS datasets.

The spatial Pearson's correlation between SOIL-WATERGRIDS estimates and the validation datasets is  $R > 0.87$  in TS (Fig. 2a–c) except in relation to ISMN, which scores  $R = 0.49$  (Fig. 2d). In RZ,  $R > 0.81$  (Fig. 2e,f). Figure 2 also highlights how SOIL-WATERGRIDS aligns with the validation datasets in dry (red markers),

Variable	Legend	
Sl_0_30, Sl_30_60, Sl_60_100	−2	Ocean
Sl_0_30_ltm, Sl_30_60_ltm, Sl_60_100_ltm	−1	No data
Slr_0_30, Slr_30_60, Slr_60_100	[0, 1]	Range of values
WTDL, WTDH, WTDL_ltm, WTDH_ltm	−2	Ocean
	−1	No data
	0	Ponding
	[0.15, 50.5]	Range of values Below 50.5 m depth
WTDN	−2	Ocean
	−1	No data
	[0, 6]	Range of values
QI	−2	Ocean
	−1	No data
	[0, 1]	Range of values, 1 is best
b_0_30, b_30_60, b_60_100	−2	Ocean
	−1	No data
	[1.2, 112]	Range of values
phi_0_30, phi_30_60, phi_60_100	−2	Ocean
	−1	No data
	[0.1, 0.80]	Range of values
psis_0_30, psis_30_60, psis_60_100	2	Ocean
	1	No data
	[−1.60, 0]	Range of values

**Table 2.** Legend of variables in SOIL-WATERGRIDS data distribution.

Variable	SOIL-WATERGRIDS	GLEAM	NOAH/GLDAS	<sup>(a)</sup> ESA/CCI	<sup>(a)</sup> ISMN	<sup>(a)</sup> Fan <i>et al.</i> , (2017)
$\theta$ in TS	LTM, mLTM	LTM, mLTM	LTM, mLTM	LTM, mLTM	mLTM	
<sup>(b)</sup> $\theta$ in RZ	LTM, mLTM	LTM, mLTM	LTM, mLTM			
WTD	LTM, mLTM					LTM, mLTM

**Table 3.** Target variables in SOIL-WATERGRIDS data distribution used for validation against existing datasets. LTM and mLTM stand for long-term mean and long-term monthly mean either accessible in or calculated from the original datasets. <sup>(a)</sup> Grid cells with missing values in the source data were not paired to SOIL-WATERGRIDS. <sup>(b)</sup> The values of  $\theta$  in RZ in SOIL-WATERGRIDS were calculated as averages of  $\theta$  at the three soil depths defined in our computational domain weighted by the corresponding soil layer thicknesses.

intermediate (gray markers) and humid (blue markers) regions, where dry and humid regions are identified by  $\theta < 0.8 \times \theta_{FC}$  and  $\theta > \theta_{FC}$ , respectively, for 75% of the time within the 45 years of assessment, with  $\theta_{FC}$  the water content at field capacity corresponding to  $\psi = -33$  kPa suction (map of identified dry and humid regions is shown in Supplementary Information, Figure S7). Nonetheless, we found a few grid cells identified as humid regions that showed relatively low  $\theta$  values in TS, which can result from a combination of hydraulic parameters  $b$  and  $\psi_s$  falling in the tails of their probability distribution in those climatic areas (see also Supplementary Information, Figure S6). We note also that the spatial correlation in humid regions was generally lower in both TS and RZ than in intermediate and dry regions. This may be caused by soil water dynamics difficult to fully capture in the equatorial regions in the Amazon and South east Asia, where most wet regions were identified, or in boreal forests that experience soil frosting even if only grid cells where permafrost is less than 10% of their area were included in our assessment.

The long-term mean  $\theta$  in TS estimated in SOIL-WATERGRIDS is well in agreement with the ESA/CCI, NOAH/GLDAS and GLEAM, with 95% of the grid cells having an anomaly ranging between  $-0.06$  and  $0.08$   $\text{m}^3 \text{m}^{-3}$  (Supplementary Information, Figure S8). SOIL-WATERGRIDS generally overestimates the ISMN observations with 95% of the grids cells having anomalies between  $-0.06$  and  $0.17$   $\text{m}^3 \text{m}^{-3}$ . In RZ, 53% of grids cells in SOIL-WATERGRIDS underestimates the NOAH/GLDAS and GLEAM datasets by an average of  $-0.11$   $\text{m}^3 \text{m}^{-3}$  while the remaining cells overestimated them only slightly ( $0.004$   $\text{m}^3 \text{m}^{-3}$ ). These anomalies are distributed heterogeneously, but with geographic patterns such as lower  $\theta$  in the northern hemisphere and slightly higher  $\theta$  in some arid regions as compared to the validation datasets (Supplementary Information, Figure S8). Those anomaly patterns and their variability across datasets obtained by data reconstruction, modeling, reanalysis or their combination, suggest that SOIL-WATERGRIDS and those datasets may suffer from some biases relative to each other, but with the additional note that, among all datasets, SOIL-WATERGRIDS is the only one to be constrained to both the volumetric water content in TS and RZ, and the water table depth (see Section “Validation of the water table depth”)

Quantity	Metrics used with the long-term mean (LTM)	Metrics used with the long-term monthly means (mLTM)
Anomaly	<sup>(a)</sup> $\Delta X_Y(i) = X(i) - Y(i)$	$\Delta X_Y(i, m) = X(i, m) - Y(i, m)$
Normalised root mean square deviation		$\text{NRMSD}(m) = \frac{\sqrt{\frac{1}{N} \sum_i \Delta X_Y(i, m)^2}}{[\max\{Y(m)\}_S - \min\{Y(m)\}_S]}$
Spatial Pearson's correlation coefficient	$R_S$	$R_S(m)$
Temporal Pearson's correlation coefficient		<sup>(a)</sup> $R_T(i)$
Duvellier coefficient		$\lambda(m) = \alpha R_S(m)$ with $\alpha = 2 \left[ \frac{\sigma_X(m)_S}{\sigma_Y(m)_S} + \frac{\sigma_Y(m)_S}{\sigma_X(m)_S} + \frac{(\bar{X}(m)_S - \bar{Y}(m)_S)^2}{\sigma_X(m)_S \sigma_Y(m)_S} \right]^{-1}$

**Table 4.** Definition of metrics measuring the quality of SOIL-WATERGRIDS data distribution. Estimates  $X$  of SOIL-WATERGRIDS are compared to the corresponding independent datasets  $Y$  referred to as in Table 1. The indices  $i$  and  $m$  indicate the grid cells and the month, respectively; the subscripts  $S$  and  $T$  indicate spatial and temporal operators, respectively;  $\sigma$  indicates the variance; the bar above symbols indicates the average. <sup>(a)</sup> validation metrics used for calculation of the data quality index  $QI$  described in Section “Data quality”.

Network	Number of stations	Reference	Network	Number of stations	Reference
AACES	49 (49)	<sup>54</sup>	ORACLE	6 (6)	N/A
AMMA-CATCH	7 (7)	<sup>55</sup>	OZNET	38 (38)	<sup>56</sup>
ARM	35 (22)	N/A	PBO_H2O	159 (146)	<sup>57</sup>
AWDN	50 (50)	N/A	PTSMN	20 (20)	<sup>58</sup>
BNZ-LTER	12 (12)	<sup>59</sup>	REMEDHUS	24 (23)	N/A
CARBOAFRICA	1 (1)	<sup>60</sup>	RISMA	24 (23)	<sup>81</sup>
CHINA	57 (40)	<sup>61</sup>	RSMN	20 (19)	N/A
COSMOS	109 (11)	<sup>62</sup>	RUSWET-AGRO	212 (53)	<sup>63</sup>
CTP_SMTMN	57 (57)	<sup>64</sup>	RUSWET-GRASS	122 (76)	<sup>63</sup>
DAHRA	1 (1)	<sup>65</sup>	SASMAS	14 (14)	<sup>66</sup>
FLUXNET	2 (2)	N/A	SCAN	239 (229)	N/A
FMI	27 (27)	N/A	SMOSMANIA	23 (22)	<sup>67</sup>
FR_Aqui	5 (5)	N/A	SNOTEL	441 (437)	<sup>68</sup>
HOBE	32 (2)	<sup>69</sup>	SOILSCAPE	171 (165)	<sup>70</sup>
ICN	19 (18)	<sup>71</sup>	TERENO	5 (5)	<sup>72</sup>
IPE	2 (1)	N/A	UMBRIA	13 (13)	<sup>73</sup>
iRON	9 (9)	<sup>74</sup>	USCRN	115 (115)	<sup>75</sup>
LAB-net	3 (1)	<sup>76</sup>	USDA-ARS	4 (4)	<sup>77</sup>
MONGOLIA	44 (43)	<sup>63</sup>			

**Table 5.** ISMN networks and corresponding total and selected (in parenthesis) number of stations used in this work. Of the original 65 networks and approximately 2678 stations, selection was based on (i) at least 4 consecutive years were available for a station and (ii) data quality was reported in the original dataset as “good”. Stations located in the same grid cell of our computational domain were averaged.

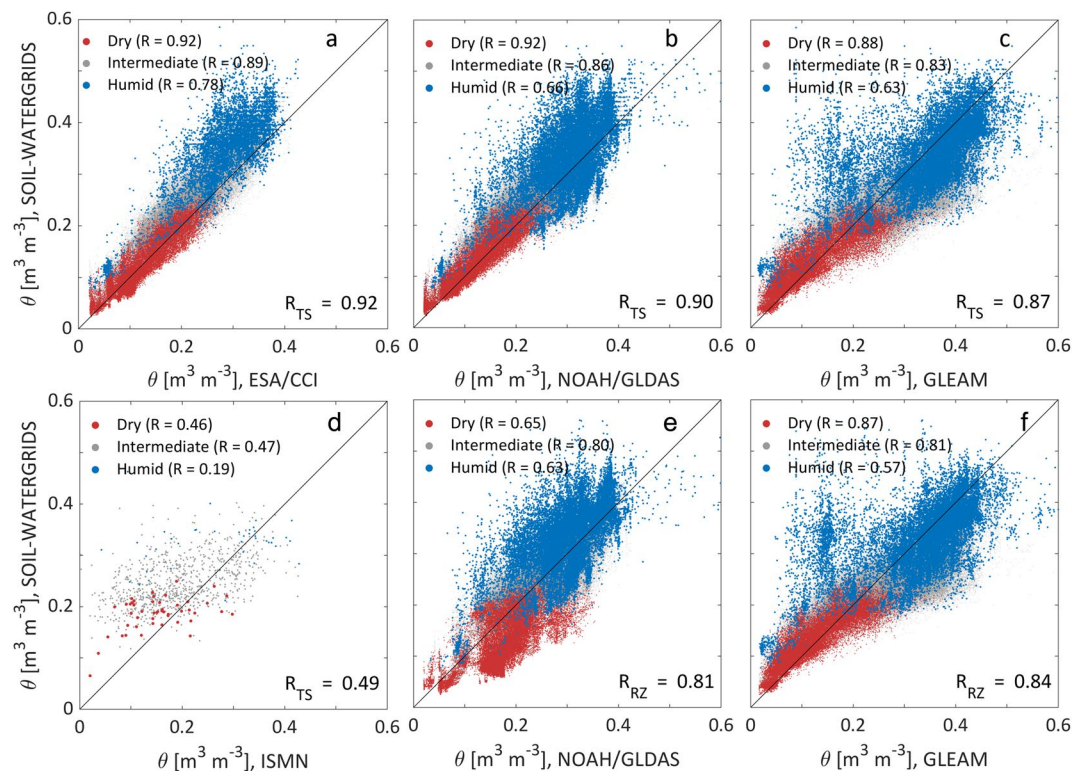
The seasonality in the long-term monthly mean  $\theta$  in TS of SOIL-WATERGRIDS captured relatively well the seasonality in ESA/CCI, NOAA/GLDAS, GLEAM, and ISMN (see 18 randomly sampled grid cells in Fig. 3). Globally, the long-term monthly means in TS and RZ do not present excessive biases, with  $\lambda > 0.7$  (best value is  $\lambda = 1$ ) and  $\text{NRMSD} < 0.15$  except in relation to ISMN data that score lower  $\lambda$  and  $R$  values (Fig. 3a–d). However, some caveats regarding the comparison with the ISMN dataset influenced our validation: in fact, the SOIL-WATERGRIDS resolution does not allow for a direct comparison of single stations and, therefore, the average of the *in-situ* measurements within the corresponding grid cell in SOIL-WATERGRIDS may not represent the grid cell-scale average. Also, the data recorded in the ISMN have different temporal coverage, thus suggesting that the long-term means in SOIL-WATERGRIDS may not fully reproduce the actual temporal variation in those specific regions.

The seasonality in the monthly estimates of  $\theta$  in TS of SOIL-WATERGRIDS throughout the entire period of assessment from 1970 to 2014 is generally highly and positively correlated with  $\theta$  values averaged across the ESA/CCI, NOAA/GLDAS, and GLEAM datasets, with only minor regions of weak anti-correlation in very humid areas in the Amazon and very arid areas in the Sahara and Gobi deserts (Supplementary Information, Figure S9).

Finally, maps of the long-term average volumetric soil water content in TS and RZ throughout the assessment period from 1970 to 2014 are represented in Supplementary Information, Figure S10.

**Validation of the water table depth.** The long-term mean and long-term monthly mean of the water table depth (WTD) estimated in SOIL-WATERGRIDS were compared with data in<sup>10</sup> using the same metrics



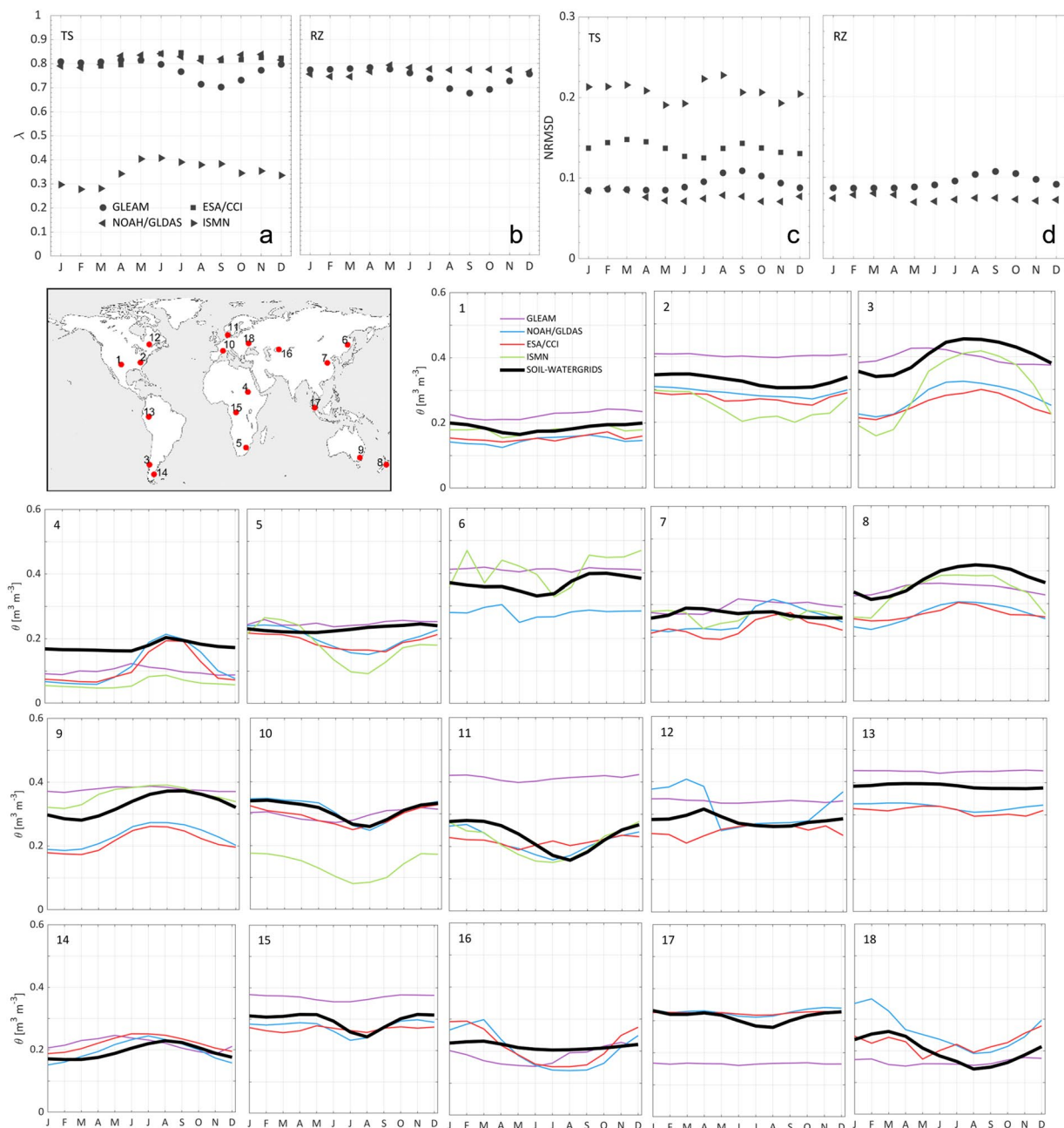


**Fig. 2** (a) to (d) Long-term mean estimates in volumetric water content  $\theta$  in the top soil (TS, 0 to 30 cm) and (e) to (f) the root zone (RZ, 0 to 100 cm) against datasets from ESA/CCI, NOAH/GLDAS, GLEAM, and ISMN, respectively. The long-term mean values of SOIL-WATERGRIDS are calculated over the assessment period 1970–2014. Dry and humid regions were identified by grid cells where  $\theta$  in the top soil is below  $0.8 \times \theta_{FC}$  and above  $\theta_{FC}$ , respectively, for 75% of the time within the 45 years of assessment, with  $\theta_{FC}$  the water content at field capacity corresponding to a suction  $\psi = -33$  kPa. A map of the geographic distribution of dry and humid regions is available in Supplementary Information, Figure S7.

described earlier. We excluded from validation about 1% of grid cells where values in<sup>10</sup> are outside of our computational depth and where WTD in SOIL-WATERGRIDS are beyond the bottom boundary zone (below 50.5 m depth) for more than 10% of the assessment period. For validation of grid cells where more than one water table existed in SOIL-WATERGRIDS, we used the water tables closest to those in<sup>10</sup>. The SOIL-WATERGRIDS estimates of the long-term mean depth of water table matched those in<sup>10</sup> reasonably well, with a high spatial correlation between the two datasets when both single and multiple water tables were present ( $R = 0.99$  and  $R = 0.74$ , respectively, Fig. 4a,b). However, data dispersion was generally higher in grid cells with multiple water tables than all other grid cells (Fig. 4b), likely because of interactions between water tables such as anticipated in Section “BRTSim modelling and outputs”. We did not find substantial biases relative to dry and humid regions (represented in Supplementary Information, Figure S11a) with the exception of some regions where multiple water tables exist, which SOIL-WATERGRIDS tends to estimate closer to surface (Fig. 4b). The goodness of fit of the long-term monthly mean was characterized by  $\lambda > 0.92$  and  $\text{NRMSD} < 0.08$  (Fig. 4c). The average anomaly ranged between  $-5.0$  and  $4.4$  m, which is slightly above our grid resolution in the vertical direction (Supplementary Information, Figure S11b).

The maps of the maximum number of water tables and the long-term average water table depth closest to data in<sup>10</sup> throughout the assessment period from 1970 to 2014 are represented in Supplementary Information, Figure S12.

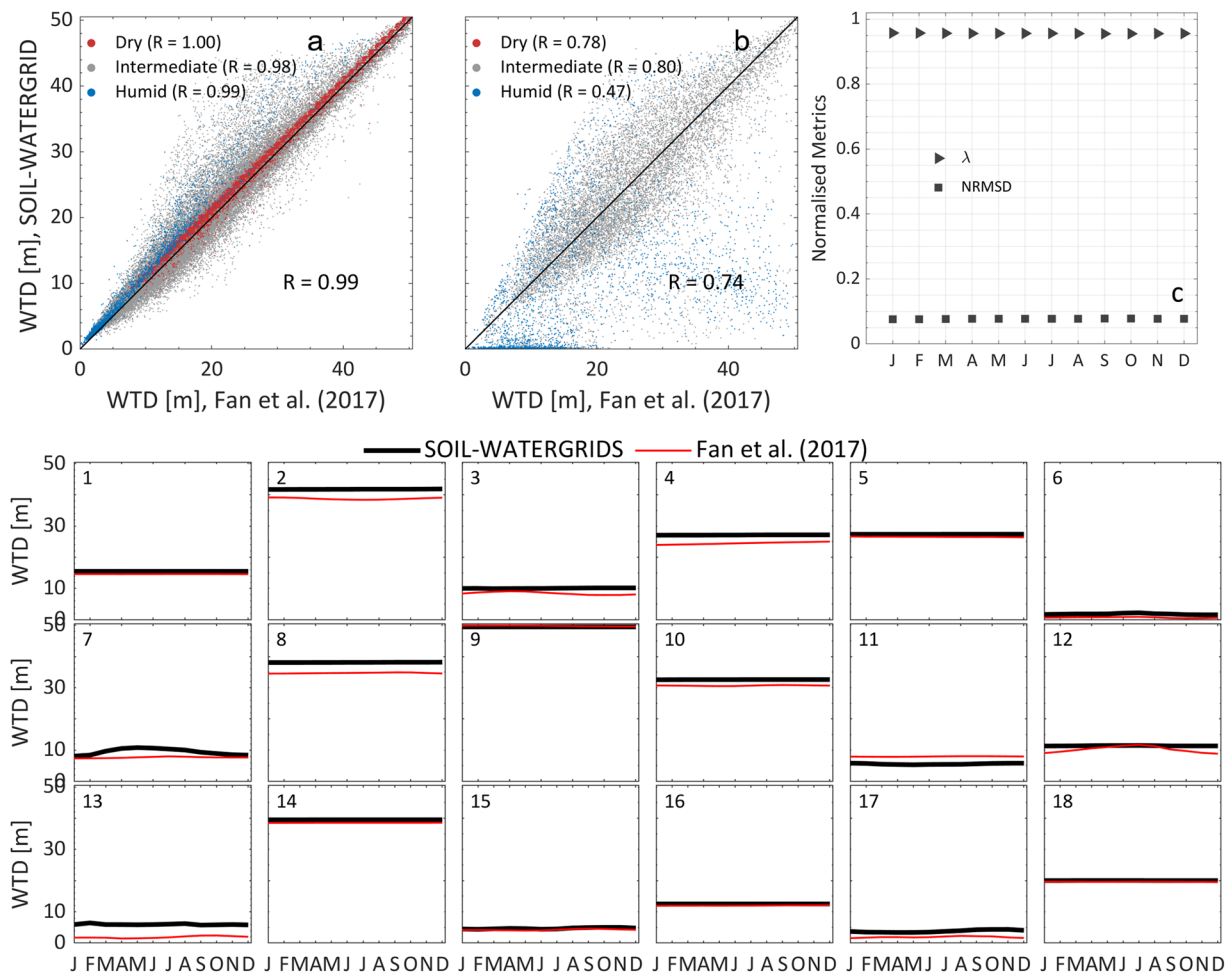
**Benchmarking the geographic distribution of ponding.** The geographic location of the long-term mean WTD estimated in SOIL-WATERGRIDS was compared against the location of the long-term mean wetland area fraction in the SWAMPS dataset to determine if the grid cells where we identified ponding coincide with those of wetlands. We considered a grid cell to contain wetlands only if the maximum wetland area fraction reported in SWAMP from 2000 to 2014 exceeded a monthly mean of 5%. About 11% of the grid cells in SOIL-WATERGRIDS show occasional ponding and 73% of those overlap with those in SWAMPS (Supplementary Information, Figure S13). Ponding in SOIL-WATERGRIDS underestimated 36% of the wetlands in SWAMPS particularly across the Australian drylands, sub Saharan region, central east Europe, and Northwest Canada. However, arid regions in the Australian and sub Saharan drylands can flood during the rainfall season only occasionally when the soil dryness does not allow fast infiltration of rainwater, and the corresponding grid cells in SOIL-WATERGRIDS may have been excluded from comparison with SWAMPS. In fact, the frequency of occurrence of wetlands in those areas is 2 or 3 times per year and covers 20% of the total area at the most. Similarly,



**Fig. 3** Seasonality assessment. (a) to (d) Duvelier coefficient  $\lambda$  and normalized root mean square deviation (NRMSD) calculated for the long-term monthly mean volumetric water content  $\theta$  in the top soil (TS, 0 to 30 cm) and root zone (RZ, 0 to 100 cm) of SOIL-WATERGRIDS relative to the GLEAM, ESA/CCI, NOAH/GLDAS, and ISMN validation data.  $\lambda$  and NRMSD are calculated as described in Section “Technical Validation”, Tables 3 and 4. (map) geographic location of 18 randomly selected grid cells and (numbered panels) corresponding long-term monthly mean volumetric water content  $\theta$  in the top soil (TS, 0 to 30 cm) estimated in SOIL-WATERGRIDS as compared to the validation datasets throughout the entire assessment period from 1970 to 2014.

northern latitudes are characterized by widespread peatland areas with the water table mostly below the soil surface<sup>51</sup> even if they are accounted for as wetlands in SWAMPS, and this may explain why SOIL-WATERGRIDS does not detect ponding in those regions.

**Benchmarking the runoff flow rate.** We compared our estimates of surface runoff  $Q_i$  calculated using the NRCS-CN method against the monthly total runoff from the surface, groundwater, and river discharge from 1970 to 2014 available in the GRUNv1 dataset. Because GRUNv1 accounts for fluxes excluded in SOIL-WATERGRIDS (i.e., river flow discharge and groundwater flow), we first verified whether GRUNv1 data are the upper bound for the surface runoff in SOIL-WATERGRIDS (i.e.,  $Q_i < Q_{i,GRUN}$ ) and we found that runoff  $Q_i$  estimated in this work



**Fig. 4** (a) and (b) scatterplot of SOIL-WATERGRIDS estimated WTD against data in<sup>10</sup> when single and multiple water tables exist, respectively. When more than one water table existed in SOIL-WATERGRIDS, we used the water table closest to those in<sup>10</sup>. (c) Duvelier coefficient  $\lambda$  and normalized root mean square deviation (NRMSD) calculated for the long-term monthly mean WTD of SOIL-WATERGRIDS estimates against data in<sup>10</sup>.  $\lambda$  and NRMSD are calculated as described in Section “Technical Validation” and Table 4. Dry and humid regions were identified by grid cells where  $\theta$  in the top soil is below  $0.8 \times \theta_{FC}$  and above  $\theta_{FC}$ , respectively, for 75% of the time within the 45 years of assessment, with  $\theta_{FC}$  the water content at field capacity corresponding to a suction  $\psi = -33$  kPa. A map of the geographic distribution of dry and humid regions is available in Supplementary Information, Figure S7.

is less than in GRUN in about 83% of grid cells ( $R = 0.30$ , Supplementary Information, Figure S14a). We next verified whether our hypothesis that the water balance  $W_i$  in Eq. (7) eventually compensated for unaccounted water fluxes at the surface or within the aquifer (i.e.,  $Q_i + W_i \approx Q_{i,GRUN}$ ) and we found SOIL-WATERGRIDS estimates are nearly equally distributed around GRUNv1 data ( $R = 0.65$ , Supplementary Information, Figure S14b). We concluded that known limitations in the application of the NRCS-CN method in relation to snowmelt and the time of concentration, and the unaccounted river flow discharge and groundwater flow were fairly moderated by the decadal water balance  $W_i$  in our approach.

**Data quality.** To calculate the data quality index  $QI$  of SOIL-WATERGRIDS, we considered two quality factors,  $QF_{\Delta}$  and  $QF_R$ , which account for the anomaly and temporal Pearson’s correlation against validation data, respectively.

$QF_{\Delta}$  describes the quality relative to the anomalies of SOIL-WATERGRIDS estimates in the validation of long-term mean soil water content  $\theta$  in TS and RZ, and the long-term mean depth of water table.  $QF_{\Delta}(i)$  is calculated in each grid cell  $i$  as

$$QF_{\Delta}(i) = \frac{1}{n_i} \sum_{n_i} [1 - \delta X_Y^{n_i}(i)] \quad (9)$$

where  $\delta X_Y^{n_i}(i) = |\Delta X_Y^{n_i}(i)| / \max\{\Delta X_Y^{n_i}(i)\}$  is the normalised anomaly for variable  $X^{n_i}$  with  $\Delta X_Y^{n_i}(i)$  the anomaly defined as in Table 4 in grid cell  $i$  against the dataset  $Y^{n_i}$ , with the number  $n_i$  of assessable variables in grid cell  $i$  ranging between 4 and 7 depending on data availability (Table 3).

$QF_R$  describes the quality relative to the temporal Pearson's correlation of the long-term monthly mean soil water content  $\theta$  in TS and RZ, and WTD of SOIL-WATERGRIDS estimates against the corresponding validation datasets.  $QF_R(i)$  in grid cell  $i$  is calculated as

$$QF_R(i) = \frac{1}{n_i} \sum_{n_i} \frac{R_T^{n_i}(i) + 1}{2} \quad (10)$$

where the argument in the sum on the right-hand side is the normalised Pearson's coefficient  $R_T^{n_i}(i)$  defined in Table 4 and ranging between 0 and 1.

Equations (9) and (10) were combined to calculate the quality index  $QI$  in each grid cell  $i$  as

$$QI(i) = \frac{QF_\Delta(i) + QF_R(i)}{2} \quad (11)$$

where  $QI=0$  means "worse" and  $QI=1$  means "best" quality. Relative to the distributed variables, 69% of grid cells in SOIL-WATERGRIDS estimates have a  $QI$  between 0.6 and 0.8, and 4% have  $QI$  values between 0.8 and 1 (Fig. 5). The  $QI$  georeferenced data are distributed with the SOIL-WATERGRIDS distribution dataset in<sup>49</sup>.

**Known limitations.** We acknowledge some limitations in this first version of SOIL-WATERGRIDS that may be addressed in future updates of the dataset.

The heterogeneity of source datasets in hydrometeorological conditions and land cover can result in some inconsistencies within and across interrelated data. For example, some inconsistencies observed between the potential and actual evapotranspiration from the CRU/TS and GLEAM datasets are quantified and highlighted in the Supplementary Information, but those inconsistencies may reduce in future releases of the original datasets or through implementation of numerical corrections across various datasets when used in combinations. Likewise, we have not explicitly accounted for the land cover change in the 45 year period of assessment, and we used data relative to 2014. Incorporation of land cover change is challenging in modelling at these spatial and temporal scales because it implies changes in runoff but also rooting systems in soil, changes in atmospheric conditions such as temperature and relative humidity, albedo/reflectance, heat exchange fluxes, urbanization, and possibly many other aspects, part of which are not yet fully characterized. We do not have yet tools to fully account for land cover changes at the current stage of progression in our modelling capabilities; we presume that land changes are accounted for in some implicit way in the datasets we have used (e.g., precipitation, evapotranspiration) because those data are re-analyses based also on satellite observations. Hence, some biases observed in dry and humid regions may be caused by these uncertainties.

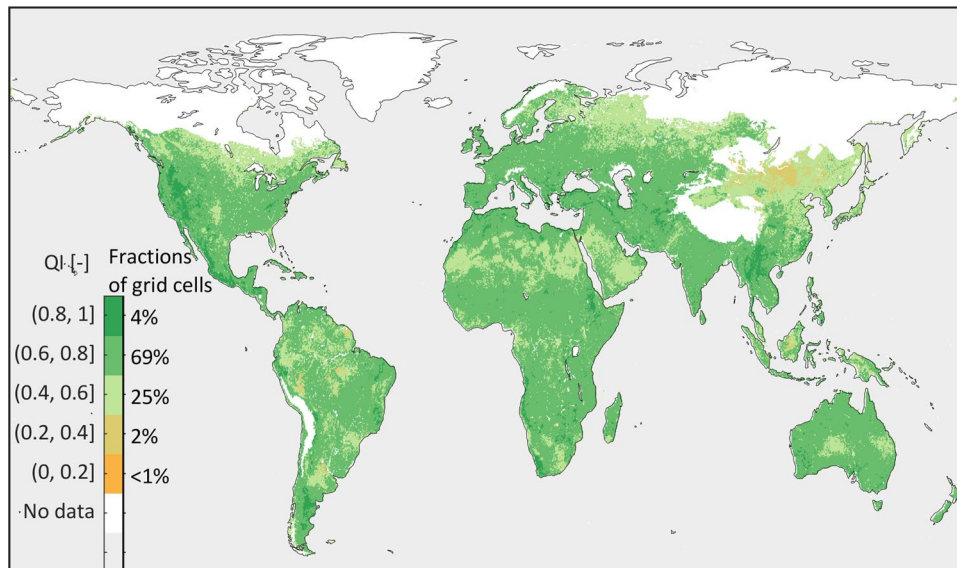
The calculation of surface runoff using the NRCS-CN method has advantages because it follows a uniform protocol to forecast the daily surface runoff that does not require additional calibration of parameters, hence it is relatively simple and stable in many applications<sup>52</sup> but is also mostly empirical as compared to other mechanistic approaches such as TOPMODEL<sup>53</sup>. We introduced an element of process accounting via infiltration into soil using the Richards equations, but a better process description should explicitly describe the land surface runoff, stream flow, and groundwater flow. For these reasons, we conditioned the water balance in each grid cells of SOIL-WATERGRIDS for the surface runoff calculated using the NRCS-CN method and we tested against the GRUNv1 dataset whether this approach was able to represent surface processes. Nonetheless, the NRCS-CN method may lack in representation of snow, ice, and soil thawing, hence the inclusion of a water balance such as in our approach still limits the capability to describe soil water in regions that undergo freezing during part of the year, thus potentially affecting also downstream grid cells.

We also acknowledge that the computational domain we developed for SOIL-WATERGRIDS can describe the soil profile vertically down to 56 m depth, but deeper soil profiles may be needed to increase the capability to inform on global water assets relative to the lower water storage. Similarly, the spatial resolution across the latitude and longitude may be increased to reduce aliasing and use more effectively data that are distributed at high resolution such as SoilGridsv2 amongst others.

Overall comparison against available datasets provides a picture of where estimates in SOIL-WATERGRIDS stand as compared to existing ones. In particular, anomalies show the geographic distribution of differences and detailed time sequence analyses can provide also information on when anomalies may appear. It is worth to acknowledge that all datasets used for validation, including those based on observations such as the ESA/CCI, ISMN, and water table depth in<sup>10</sup>, may suffer from uncertainty and biases. Nonetheless, efforts in elaborating estimates of global scale water assets have an extraordinary importance in face of the current and expected planetary stresses on soil, water, and food security.

## Usage Notes

The SOIL-WATERGRIDS dataset and modelling distributions contain inputs to and outputs of our model. The former are mostly standard TXT formatted files readable in any freeware text editor (e.g., Textpad or similar), while the latter are standard NetCDF4 formatted files, which can be read in different coding languages (e.g., MATLAB, Python, Julia) or used within specialised licenced software (ArcGIS) and free software (e.g., QGIS, Panoply). The data release also contains a Technical Documentation reporting details on the SOIL-WATERGRIDS dataset and modelling distribution, including file organization, and editable custom codes written in MathWorks @Matlab 2019b.



**Fig. 5** Data quality index  $QI$  combining the anomaly and temporal Pearson's correlation between the long-term mean and long-term monthly mean of SOIL-WATERGRIDS estimates and the corresponding validation datasets calculated as prescribed in Eq. (11).

### Code availability

SOIL-WATERGRIDS has been generated using the BRTSim (BioReactive Transport Simulator) computational solver version v4.1a (2020). The BRTSim software is multiplatform and can be deployed on Microsoft, Unix/Linux/Ubuntu, and Mac operating systems. The full BRTSim package, inclusive of executables, examples, basic post-processing scripts, and User Manual and Technical Guide are available for download at the BRTSim home page <https://sites.google.com/site/thebrtsimproject/home> under the CC BY4.0 license.

Seeding datasets used in our modelling are not distributed here because they are publicly accessible from the links reported in Online-only Table 1. BRTSim input files required to produce the full-size global-scale simulation of SOIL-WATERGRIDS are distributed in this data release and are available at the Zenodo repository<sup>49</sup> (about 50 GB). The BRTSim modelling distribution package for SOIL-WATERGRIDS includes 7 continental regions (Africa, Asia insular, Asia, Europe, North America, South America, Oceania) organized in compressed folders that include the BRTSim executable and license file, bash files (BRTSim\_v41a\_GLNXA64\_R2019b, run\_BRTSim\_v41a\_GLNXA64\_R2019b.sh, license.txt), the Param\_\*.inp files to instruct BRTSim to run each grid cells of the computational domain, and the tables of boundary conditions (Table\_\*.txt, WB\_\*.txt, and RNF\_\*.txt) organized by continental regions. The license file is valid until the end of 2021 and a new license can be obtained with no charge from the BRTSim home page after expiry. The distributed package allows to model about 168,000 grid cells globally. Raw outputs of the full-scale model run occupy about 10TB (uncompressed) and is not distributed. Full details on the BRTSim modelling including data structure, file naming, supported operating system, modelling launching and workflow, and accompanying scripts are available in the SOIL-WATERGRIDS Technical Documentation<sup>49</sup>.

Received: 16 February 2021; Accepted: 24 August 2021;

Published online: 06 October 2021

### References

1. Siebert, S. *et al.* Groundwater use for irrigation—a global inventory. *Hydrol. Earth Syst. Sci.* **14**, 1863–1880 (2010).
2. Clarke, R., Lawrence, A. & Foster, S. S. *Groundwater: A Threatened Resource*. (UNEP, United Nations Environment Programme, 1996).
3. Famiglietti, J. S. The global groundwater crisis. *Nat. Clim. Change* **4**, 945–948 (2014).
4. Thomas, B. F. & Famiglietti, J. S. Identifying climate-induced groundwater depletion in GRACE observations. *Sci. Rep.* **9**, 1–9 (2019).
5. Gurdak, J. J. Groundwater: Climate-induced pumping. *Nat. Geosci.* **10**, 71–71 (2017).
6. Kumm, M., Ward, P. J., Moel, H. & Varis, O. Is physical water scarcity a new phenomenon? Global assessment of water shortage over the last two millennia. *Environ. Res. Lett.* **5**, 034006 (2010).
7. Guglielmo, M. *et al.* Time- and depth-resolved mechanistic assessment of water stress in Australian ecosystems under the CMIP6 scenarios. *Adv. Wat. Res.* **148**, 103837 (2021).
8. Ridolfi, L., D'Odorico, P., Laio, F., Tamea, S. & Rodriguez-Iturbe, I. Coupled stochastic dynamics of water table and soil moisture in bare soil conditions. *Water Resour. Res.*, **44** (2008).
9. Tsyppkin, G. G. & Shargatov, V. A. Influence of capillary pressure gradient on connectivity of flow through a porous medium. *Int. J. Heat Mass Trans.* **127**, 1053–1063 (2018).
10. 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).
11. Dawson, T. E. & Ehleringer, J. R. Streamside trees that do not use stream water. *Nat.* **350**, 335–337 (1991).
12. Elliott, S., Baker, P. J. & Borchert, R. Leaf flushing during the dry season: the paradox of Asian monsoon forests. *Glob. Ecol. Biogeogr.* **15**, 248–257 (2006).

13. Meixner, T. *et al.* Implications of projected climate change for groundwater recharge in the western United States. *J. Hydrol* **534**, 124–138 (2016).
14. Richey, A. S. *et al.* Quantifying renewable groundwater stress with GRACE. *Water Resour. Res.* **51**, 5217–5238 (2015).
15. Fan, Y., Li, H. & Miguez-Macho, G. Global patterns of groundwater table depth. *Science* **339**, 940–943 (2013).
16. Entekhabi, D. *et al.* The soil moisture active passive (SMAP) mission. *Proc. IEEE* **98**, 704–716 (2010).
17. Gruber, A., Scanlon, T., Schalie, R., Wagner, W. & Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **11**, 717–739 (2019).
18. Dorigo, W. A. *et al.* Global Automated Quality Control of *In Situ* Soil Moisture Data from the International Soil Moisture Network. *Vadose Zone J.* **12**, 1–21 (2013).
19. Rodell, M. *et al.* The Global Land Data Assimilation System. *Bull. Am. Meteorol. Soc.* **85**, 381–394 (2004).
20. Maggi, F. (2021). *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.
21. Gruber, A., Dorigo, W. A., Crow, W. & Wagner, W. Triple Collocation-Based Merging of Satellite Soil Moisture Retrievals. *IEEE Trans. Geosci. Remote Sens.* **55**, 6780–6792 (2017).
22. Gruber, A., Scanlon, T., Schalie, R., Wagner, W. & Dorigo, W. Evolution of the ESA CCI Soil Moisture climate data records and their underlying merging methodology. *Earth Syst. Sci. Data* **11**, 717–739 (2019).
23. Dorigo, W. *et al.* ESA CCI Soil Moisture for improved Earth system understanding: State-of-the art and future directions. *Remote Sens. Environ.* **203**, 185–215 (2017).
24. Miralles, D. G. *et al.* Global land-surface evaporation estimated from satellite-based observations. *Hydrol. Earth Syst. Sci.* **15**, 453–469 (2011).
25. Martens, B. *et al.* GLEAM v3: satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **10**, 1903–1925 (2017).
26. Dorigo, W. A. *et al.* The International Soil Moisture Network: a data hosting facility for global *in situ* soil moisture measurements. *Hydrol Earth Syst Sci* **15**, 1675–1698 (2011).
27. Poulter, B. *et al.* Global wetland contribution to 2000–2012 atmospheric methane growth rate dynamics. *Environ. Res. Lett.* **12**, 094013 (2017).
28. Ghiggi, G., Humphrey, V., Seneviratne, S. I. & Gudmundsson, L. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data* **11**, 1655–1674 (2019).
29. Poggio, L. *et al.* SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240 (2021).
30. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, **12** (2017).
31. Brooks, R. H. & Corey, A. T. Hydraulic properties of porous media and their relation to drainage design. *Trans ASABE* **7**, 26–28 (1964).
32. Dai, Y. *et al.* A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling. *J. Adv. Model* **11**, 2996–3023 (2019).
33. Zhang, Y., Schaap, M. G. & Zha, Y. A high-resolution global map of soil hydraulic properties produced by a hierarchical parameterization of a physically based water retention model. *Water Resour. Res.* **54**, 9774–9790 (2018).
34. Friedl, M. & Sulla-Menashe, D. MCD12Q1 MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MCD12Q1.006> (2019).
35. Sulla-Menashe, D., Gray, J. M., Abercrombie, S. P. & Friedl, M. A. Hierarchical mapping of annual global land cover 2001 to present: The MODIS Collection 6 Land Cover product. *Remote Sens Environ* **222**, 183–194 (2019).
36. Maggi, F., la Cecilia, D., Tang, F. H. & McBratney, A. The global environmental hazard of glyphosate use. *Sci. Total Environ.* **717**, 137167 (2020).
37. Canadell, J. *et al.* Maximum rooting depth of vegetation types at the global scale. *Oecologia* **108**, 583–595 (1996).
38. Allen, R. G. *et al.* Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome* **300**, D05109 (1998).
39. Thenkabail, P. S. *et al.* NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security Support Analysis Data (GFSAD) Crop Dominance 2010 Global 1 km V001. *Monograph. NASA EOSDIS Land Processes DAAC, South Dakota, USA* <http://oar.icrisat.org/id/eprint/10982> (2016).
40. Sacks, W. J., Deryng, D., Foley, J. A. & Ramankutty, N. Crop planting dates: an analysis of global patterns. *Glob. Ecol. Biogeogr.* **19**, 607–620 (2010).
41. Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. *Sci. Data* **7**, 1–18 (2020).
42. Hong, Y. & Adler, R. F. Estimation of global SCS curve numbers using satellite remote sensing and geospatial data. *Int J. Remote Sens.* **29**, 471–477 (2008).
43. Hong, Y., Adler, R. F., Hossain, F., Curtis, S. & Huffman, G. J. A first approach to global runoff simulation using satellite rainfall estimation. *Water Resour. Res.* **43**, W08502 (2007).
44. Heggen, R. J. Normalized antecedent precipitation index. *J. Hydrol. Eng.* **6**, 377–381 (2001).
45. Edwards, M. Data announcement 88-MGG-02: Digital relief of the surface of the earth. *National Oceanic and Atmospheric Administration, National Geophysical Data Center, Boulder, CO, USA*, <https://www.ngdc.noaa.gov/mgg/global/etopo5.HTML> (1988)
46. Brown, J., Ferrians, O. J. J., Heginbottom, J. A. & Melnikov, E. S. *Circum-Arctic map of permafrost and ground-ice conditions*. National Snow and Ice Data Center/World Data Center for Glaciology (1997, revised 2002).
47. Fan, Y. & Miguez-Macho, G. A simple hydrologic framework for simulating wetlands in climate and earth system models. *Clim. Dyn.* **37**, 253–278 (2011).
48. Richards, L. A. Capillary conduction of liquids through porous mediums. *Physics* **1**, 318–333 (1931).
49. Maggi, F., Guglielmo, M., Tang, F. H. M. & Pasut, C. SOIL-WATERGRIDS v1, mapping dynamic changes in soil moisture and depth of water table from 1970 to 2014, dataset and modelling. *Zenodo* <https://doi.org/10.5281/zenodo.4997453> (2021).
50. Duveiller, G., Fasbender, D. & Meroni, M. Revisiting the concept of a symmetric index of agreement for continuous datasets. *Sci Rep.* **6**, 1–14 (2016).
51. Prigent, C., Papa, F., Aires, F., Rossow, W. B. & Matthews, E. Global inundation dynamics inferred from multiple satellite observations, 1993–2000. *J. Geophys. Res. Atmos.* **112**.D12, (2007).
52. Ponce, V. M. & Hawkins, R. H. Runoff curve number: Has it reached maturity? *Journal of hydrologic engineering* **1**, 11–19 (1996).
53. Walter, M. T. *et al.* Refined conceptualization of TOPMODEL for shallow subsurface flows. *Hydrological Processes* **16**, 2041–2046 (2002).
54. Peischl, S. *et al.* The AACES field experiments: SMOS calibration and validation across the Murrumbidgee River catchment. *Hydrol Earth Syst Sci*, **16**, 1697–1708.
55. Pellarin, T. *et al.* Hydrological modelling and associated microwave emission of a semi-arid region in South-western Niger. *J. Hydrol* **375**, 262–272 (2009).
56. Smith, A. B. *et al.* The Murrumbidgee soil moisture monitoring network data set. *Water Resour. Res.* **48** (2012).
57. Larson, K. M., *et al.* Use of GPS receivers as a soil moisture network for water cycle studies. *Geophys. Res. Lett.*, **35** (2008).
58. Hajdu, I., Yule, I., Bretherton, M., Singh, R. & Hedley, C. Field performance assessment and calibration of multi-depth AquaCheck capacitance-based soil moisture probes under permanent pasture for hill country soils. *Agric. Water Manag.* **217**, 332–345 (2019).

59. Van Cleve, K., Chapin, F. S. & Ruess, R. W. Bonanza Creek LTER: Hourly Evaporation measurements at Core Sites from 1988 to Present in the Bonanza Creek Experimental Forest near Fairbanks, Alaska. *Environmental Data Initiative* <https://doi.org/10.6073/pasta/dcbc5a5649c12cfc28af9554d4c7089b> (2018).
60. Ardö, J. A 10-Year Dataset of Basic Meteorology and Soil Properties in Central Sudan. *Dataset Papers in Science* **2013**, 297973 (2013). ID.
61. Liu, S., Mo, X., Li, H., Peng, G. & Robock, A. Spatial variation of soil moisture in China: Geostatistical characterization. *J. Meteorol. Soc. Japan. Ser. II*, **79**, 555–574 (2001).
62. Zreda, M. *et al.* COSMOS: The cosmic-ray soil moisture observing system. *Hydrol Earth Syst Sci* **16**, 4079–4099 (2012).
63. Robock, A. *et al.* The global soil moisture data bank. *Bull Am Meteorol Soc* **81**, 1281–1300 (2000).
64. Yang, K. *et al.* A multiscale soil moisture and freeze–thaw monitoring network on the third pole. *Bull Am Meteorol Soc* **94**, 1907–1916 (2013).
65. Tagesson, T. *et al.* Ecosystem properties of semiarid savanna grassland in West Africa and its relationship with environmental variability. *Glob Chang Biol* **21**, 250–264 (2015).
66. Rüdiger, C. *et al.* Goulburn River experimental catchment data set. *Water Resour. Res* **43**, W10403 (2007).
67. Albergel, C. *et al.* From near-surface to root-zone soil moisture using an exponential filter: an assessment of the method based on *in-situ* observations and model simulations. *Hydrol. Earth Syst. Sci.* **12**, 1323–1337 (2008).
68. Leavesley, G. H. *et al.* A modeling framework for improved agricultural water supply forecasting. *AGU Fall Meeting Abstracts* **2008**, C21A–0497 (2008).
69. Bircher, S., Skou, N., Jensen, K. H., Walker, J. P. & Rasmussen, L. A soil moisture and temperature network for SMOS validation in Western Denmark. *Hydrol. Earth Syst. Sci.* **16**, 1445–1463 (2012).
70. Moghaddam, M. *et al.* Soil Moisture Profiles and Temperature Data from SoilSCAPE Sites. *USA ORNL DAAC* <https://doi.org/10.3334/ORNLDAAC/1339> (2016).
71. Hollinger, S. E. & Isard, S. A. A soil moisture climatology of Illinois. *J. Clim.* **7**, 822–833 (1994).
72. Zacharias, S. *et al.* A network of terrestrial environmental observatories in Germany. *Vadose Zone J.* **10**, 955–973 (2011).
73. Brocca, L. *et al.* Soil moisture estimation through ASCAT and AMSR-E sensors: An intercomparison and validation study across Europe. *Remote Sens. Environ.* **115**, 3390–3408 (2011).
74. Osenga, E. C., Arnott, J. C., Endsley, K. A. & Katzenberger, J. W. Bioclimatic and soil moisture monitoring across elevation in a mountain watershed: Opportunities for research and resource management. *Water Resour. Res* **55**, 2493–2503 (2019).
75. Bell, J. E. *et al.* US Climate Reference Network soil moisture and temperature observations. *J. Hydrometeorol.* **14**, 977–988 (2013).
76. Mattar, C., Santamaria-Artigas, A., Durán-Alarcón, C., Olivera-Guerra, L. & Fuster, R. LAB-net the first Chilean soil moisture network for remote sensing applications. *Proceedings of the IV Recent Advances in Quantitative Remote Sensing*, 22–26 (2014).
77. Jackson, T. J. *et al.* Validation of advanced microwave scanning radiometer soil moisture products. *IEEE Trans. Geosci. Remote Sens.* **48**, 4256–4272 (2010).
78. Pelletier, J. D. *et al.* Global 1-km gridded thickness of soil, regolith, and sedimentary deposit layers. *ORNL DAAC* <https://doi.org/10.3334/ORNLDAAC/1304> (2016).
79. Beaudoin, H. & Rodell, M. NASA/GSFC/HSL. GLDAS Noah Land Surface Model L4 monthly 0.25 x 0.25 degree, Version 2.0. *NASA Goddard Earth Sciences Data and Information Services Center* <https://doi.org/10.5067/9SQ1B3ZXP2C5> (2019).
80. Menne, M. J., Durre, I., Vose, R. S., Gleason, B. E. & Houston, T. G. An overview of the global historical climatology network-daily database. *J. Atmos. Ocean Technol.* **29**, 897–910 (2012).
81. Ojo, E. R. *et al.* Calibration and evaluation of a frequency domain reflectometry sensor for real-time soil moisture monitoring. *Vadose Zone J.* **14**, 1–12 (2015).

## Acknowledgements

We thank the Editorial Board and two anonymous Reviewers for the suggestions and recommendations to the earlier submissions of this paper. Magda Guglielmo and Fiona Tang are supported by the SREI2020 of the University of Sydney. Magda Guglielmo is also supported by the W H Gladstones Population and Environment Fund of the Australian Academy of Science awarded to Fiona Tang. Chiara Pasut is supported by the The University of Sydney Postgraduate Scholarship Award “Contaminant Hydrology”. The authors acknowledge the Sydney Informatics Hub and the University of Sydney’s high-performance computing cluster Artemis for providing the high-performance computing resources that have contributed to the results reported within this work. The authors acknowledge the use of the National Computational Infrastructure (NCI) which is supported by the Australian Government, and accessed through the NCMAS allocation scheme award to Maggi, 2020, “Global soil and water resource in a changing climate”, and the Sydney Informatics Hub HPC Allocation Scheme, which is supported by the Deputy Vice-Chancellor (Research), University of Sydney and the ARC-LIEF, 2019: Smith, Muller, Thornber *et al.*, Sustaining and strengthening merit-based access to National Computational 613 Infrastructure (LE190100021).

## Author contributions

Magda Guglielmo developed the dataset, curated data, performed technical validation, led data visualisation, and writing. Chiara Pasut elaborated the surface runoff dataset and aided in the modelling and figure compilation. Fiona H. Tang and Federico Maggi aided in the development of the dataset, data curation, reviewing, and editing the manuscript. All authors contributed to global scale modelling. Federico Maggi led the conceptualisation, methods development, and project administration.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41597-021-01032-4>.

**Correspondence** and requests for materials should be addressed to M.G., F.H.M.T., C.P. or F.M.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

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



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

The Creative Commons Public Domain Dedication waiver <http://creativecommons.org/publicdomain/zero/1.0/> applies to the metadata files associated with this article.

© The Author(s) 2021