# scientific data

Check for updates

## **OPEN** A global dataset of remote sensingbased soil critical point and DATA DESCRIPTOR permanent wilting point

Yawei Xu<sup>1</sup>, Qing He<sup>2</sup>, Hui Lu<sup>1,3,4</sup>, Kun Yang<sup>1</sup>, Dara Entekhabi<sup>5</sup> & Daniel J. Short Gianotti<sup>™</sup>

The critical point (CP) and permanent wilting point (PWP) are key soil hydraulic characteristics that control the land surface energy budget and water balance. There is a lack of available data for these parameters on the global scale. This study extracts CP and PWP through soil moisture drydown (SMD) and provides global yearly soil hydraulic properties from a long-term (2002–2023) remote-sensing soil moisture product (Neural Network-based Soil Moisture, NNsm). Validated against 1334 stations from the International Soil Moisture Network (ISMN), the results show that the global medians of CP and PWP based on the NNsm are robust over time, and outperform the Soil Moisture Active and Passive (SMAP) dataset in accuracy due to the advantage of daily temporal resolution. Furthermore, this dataset holds an advantage over existing products, as it is derived from a multi-year climatological mean state and solely from satellite-based soil moisture observation. The derived dataset is useful for those who wish to connect land-atmosphere characteristics with their interests, as well as calibrate land surface models.

#### **Background & Summary**

The critical point (CP) refers to the volumetric soil water content at which evaporation transitions from being energy-limited to being water-limited. It separates the Stage-I evapotranspiration regime and the Stage-II evapotranspiration regime<sup>1-3</sup>. As evaporation proceeds, soil moisture decreases, eventually reaching a level at which it can no longer be absorbed or utilised by vegetation. This moisture level is termed the permanent wilting point (PWP)<sup>4-6</sup>. The PWP and CP are important soil hydraulic properties for global hydrological models<sup>7,8</sup>, biophysical models<sup>9,10</sup>, and agricultural irrigation<sup>11</sup>. However, due to the lack of large-scale estimates of CP, and PWP, these soil parameters often exist as unified constants in global hydrological models<sup>12</sup>. The hydraulic properties of soil affect vegetation growth and are also proven to play a role in intensifying extreme events such as floods<sup>13</sup> droughts<sup>14</sup>, and heatwaves<sup>15</sup>. With the continuously increasing population<sup>16</sup>, food production demand, and water resource shortages, accurately estimating global soil hydraulic characteristics - thereby enhancing model accuracy – is of great significance in addressing the above issues<sup>17</sup>.

It is difficult to obtain soil hydraulic properties globally, as most of the existing data comes from field observations<sup>18</sup>. One traditional method for estimating PWP on a large scale is based on pedo-transfer functions (PTFs)<sup>9,12,16-21</sup>. This method requires a large number of field samples, but the accuracy of the results is still limited<sup>22,23</sup>. A more recent method uses machine learning (ML) to predict PWP. But these algorithms are based on input soil parameters such as soil porosity, sand content, clay content, etc<sup>24-27</sup> meaning that they still rely on a large number of field samples for validation, which is time-consuming and labour-intensive<sup>28,29</sup>.

Recent advancements in satellite remote sensing techniques provide an efficient way to capture the global soil moisture drydown (SMD) curve and use this to determine information related to soil hydraulic characteristics. The underlying theory posits that there is a close relationship between SMD and the two soil hydraulic properties: The SMD is defined as a process with consistent soil moisture decrement following a precipitation event<sup>30,31</sup>.

<sup>1</sup>Department of Earth System Science, Institute for Global Change Studies, Tsinghua University, Beijing, 100084, China. <sup>2</sup>Department of Civil Engineering, The University of Tokyo, Tokyo, 113-8654, Japan. <sup>3</sup>State Key Laboratory of Hydroscience and Engineering, Tsinghua University, Beijing, 100084, China. <sup>4</sup>Tsinghua University (Department of Earth System Science)- Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Beijing, 100084, China. <sup>5</sup>Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, MA02139, Cambridge, USA. <sup>™</sup>e-mail: luhui@tsinghua.edu.cn

Initially, during the SMD process, the soil moisture decrease may be due to excess water in the soil being allocated to runoff and drainage until the soil becomes unsaturated. Subsequently, the soil moisture decrease will be due to evapotranspiration controlled by energy availability. When the soil moisture reaches the CP, evapotranspiration will instead be controlled by the water supply until soil moisture reaches PWP<sup>32</sup>. The critical range of SMD can, therefore, be used to infer CP and PWP.

The method of using remote sensing soil moisture data to estimate soil hydraulic properties has already been applied in previous studies. Field capacity, a commonly used soil hydraulic parameter, was estimated in Yang *et al.*<sup>33</sup> using the Advanced Microwave Scanning Radiometer for EOS (AMSR-E) soil moisture data, where soil moisture values observed within 1-2 days following rainfall peaks were assumed to approximate field capacity. However, there are few existing products that represent the CP, a direct measure of water stress. In some previous studies<sup>34</sup>, CP has been considered to be the same as field capacity because the two points appeared very close together in time. However, in reality, CP may be slightly smaller than field capacity<sup>31,35,36</sup>. Recently, Fu *et al.*<sup>37</sup> used the evaporative fraction-soil moisture (EF-SM) method to calculate global values of CP by combining satellite and ERA5-land reanalysis data, but their results may still be dependent on the model's parameterization schemes. PWP is represented in products like Shangguan *et al.*'s Global Soil Dataset for use in Earth System Models (GSDE) dataset<sup>38,39</sup> and Tomislav and Gupta's (TG) global soil water content dataset<sup>40</sup> by soil moisture values at 1500 kPa suction, since PWP is traditionally considered to be the water content measured at a matric potential of  $\leq$ -1500 kPa. While numerous PWP datasets, such as those mentioned above, do exist, most are derived from PTFs rather than from direct observational data. What is required is a global soil CP and PWP dataset derived solely from satellite soil moisture observations with robust data quality over time.

In previous studies<sup>31,36</sup>, the satellite SMD analyses were mainly conducted over short time spans but a recently developed soil moisture dataset, Neural Network-based Soil Moisture (NNsm), makes it possible to extend SMD analyses to a long temporal span so that temporal variability can be investigated. This dataset is generated using a soil moisture neural network algorithm developed by Yao *et al.*<sup>41</sup>. The training target is the Soil Moisture Active and Passive (SMAP) soil moisture product, with the brightness temperature from AMSR-E/2 serving as the input to produce long-term soil moisture data. This method is able to reproduce the spatiotemporal distribution of SMAP soil moisture with an accuracy comparable to SMAP soil moisture products, and superior to the input soil moisture products of AMSR-E and AMSR2. It has been verified that NNsm matches SMAP well with similar accuracy (Unbiased Root Mean Square Error, ubRMSE  $\leq 0.04 \text{ m}^3/\text{m}^3$ )<sup>42</sup>. Another major advantage of NNsm is that the use of daily resolution data means that the estimation of the short-term drydown process will be "truer" to the real situation<sup>36</sup>.

This study derived long-term satellite-based soil CP and PWP from SMD characteristics and analyzed their annual variability. The use of multi-year data may help to reveal differences in soil hydraulic parameters under different climatic conditions, which can lead to increased errors in land surface models. The CP and PWP obtained in this way are completely based on observational data, a unique advantage over the results of previous ML models, PTF analysis, or the EF-SM method. To demonstrate the applicability of NNsm in capturing SMD characteristics, the NNsm-based SMD events are compared with those estimated from SMAP during 2015–2020 and validated against 1334 *in-situ* stations from the International Soil Moisture Networks (ISMN). Additionally, the global median CP and PWP values are compared with Fu *et al*.'s global CP products<sup>37</sup> (hereafter Fu), GSDE dataset<sup>38</sup>, and TG products<sup>40</sup> to explore the advantages of this dataset. For future applications, this dataset can provide global scale estimates of CP and PWP, which can help with agricultural yield estimation<sup>20,43,44</sup>, irrigation strategy regulation<sup>45,46</sup>, soil health monitoring<sup>23,47,48</sup>, and providing important input parameters for ecohydrological models<sup>37,49,50</sup>. Additionally it can be expected that increasing understanding of soil moisture characteristics on a global scale will help to reduce the labor cost of field sampling.

#### Methods

**Fitting soil moisture drydown.** SMD is usually divided into three processes. The first of these, in which soil moisture decreases rapidly, is dominated by drainage and runoff. This process occurs on an hourly scale making it difficult for exponential decay models, suitable for monthly scale input data, to capture. Jahlvand *et al.*<sup>51</sup> developed a method named "drainage from drydown" to estimate the coefficients of drainage, but it is only suitable for ground-based soil moisture observations. The second process of SMD is the first stage of evapotranspiration (ET-I) when soil moisture exceeds the field capacity but the rate of drydown is limited by energy. The third process is the second stage of evapotranspiration (ET-II) when the rate of SMD is limited by available water. Since the first process is far quicker than can be caught by remote sensing observation and of far shorter duration than the latter two processes, it is assumed that the first process can be neglected. McColl *et al.*<sup>36</sup> developed a drying characterization method suitable for daily scale input data, which has been proven to be highly suitable for studying SMD when the input data is satellite soil moisture data. Therefore, McColl's<sup>36</sup> method is used to calculate drydown timescales in this study (Eq. (1)):

$$\frac{d\theta(t)}{dt} = \begin{cases} -\frac{\theta(t) - \theta_w}{\tau_L}, \text{ if } P = 0 \text{ in the interval}[t - \Delta t, t] \\ -\frac{\theta(t) - \overline{\theta}}{\tau_S} + \varepsilon(t), \text{ if } P > 0 \text{ in the interval}[t - \Delta t, t] \end{cases}$$
(1)

where  $\theta(t)$  is soil moisture (with units of m<sup>3</sup>/m<sup>3</sup>), t is time (day) and  $\theta_w$  (m<sup>3</sup>/m<sup>3</sup>) is a minimum soil moisture value.  $\varepsilon(t)$  is the random error (m<sup>3</sup>/m<sup>3</sup> day),  $\overline{\theta}$  is the mean soil moisture (m<sup>3</sup>/m<sup>3</sup>) and P is precipitation (mm/ day).  $\tau_s$  is the short-term timescale of drydown events (day), which reflects energy-limited soil hydrological

Drydown regimes	Parameters [unit]	Meaning	
Infiltration & ET-I	$ au_{\mathcal{S}}$ [day]	Short-term timescale of drydown events, which reflects energy-limited soil hydrological processes.	
ET-II	$\tau_L$ [day]	Long-term timescale of drydown events, which reflects water-limited soil hydrological processe	
	CP [m <sup>3</sup> /m <sup>3</sup> ]	Initial soil moisture value at the beginning of drydown events - long-term median/mean may serve as the critical point.	
	PWP [m <sup>3</sup> /m <sup>3</sup> ]	Minimum soil moisture value of drydown events - long-term median/mean may serve as the permanent wilting point.	

Table 1. Parameters of drydown characteristics.

processes (e.g., drainage and potential ET process).  $\tau_L$  is the long-term timescale of drydown events (day), which reflects water-limited processes (e.g., the water-limited stage of ET). After simplification,  $\tau_S$  and  $\tau_L$  can be calculated from Eqs. (2) and (3), respectively.

$$\tau_{S} = -\frac{\frac{\Delta t}{2}}{\log\left(\frac{\Delta z \overline{\Delta \theta_{+}}}{\alpha}\right)} = -\frac{\frac{\Delta t}{2}}{\log(F_{p})}$$
(2)

$$\theta(t) = \Delta \theta \, \exp\left(-\frac{t - \Delta t_p}{\tau_L}\right) + \theta_w \tag{3}$$

where  $\Delta \theta_+ = \theta(t) - \theta(t - \Delta t)$ ,  $\Delta t$  is the temporal resolution of model outputs or satellite observations (day) and  $\Delta \theta$  is the drydown amplitude (m<sup>3</sup>/m<sup>3</sup>).  $t - \Delta t_p$  is defined as the time of the precipitation event (day), which is also the start of an individual drydown.  $\alpha$  is mean event precipitation (mm/day),  $F_p$  is the stored precipitation fraction developed by McColl *et al.*<sup>36</sup> and  $\Delta z$  is the depth of soil (mm).

In this algorithm, any event for which the decrease in soil moisture is less than 0.04 m<sup>3</sup>/m<sup>3</sup> is not considered a drydown event, as such fluctuations may occur even in ground measurements.

Through this method, four parameters (see Table 1) reflecting either soil hydraulic properties (eg. the CP and PWP) or land-atmosphere feedback variables ( $\tau_s$  and  $\tau_l$ ) are provided for both stages.

**Soil moisture data.** NNsm data ranging from 2002 to 2023 with daily resolution is used as soil moisture input for the primary analyses in this study. NNsm is a continuous and consistent global surface soil moisture dataset from 2002, with a resolution of 36 km on the daily scale. It is provided by the National Tibetan Plateau Data Center (http://data.tpdc.ac.cn)<sup>42</sup>. This dataset is a neural network-based product that utilizes multi-frequency brightness temperature data from AMSR2 and AMSRE, with the SMAP Level 3 passive daily surface soil moisture product as the training target. It does not directly use soil texture information but can accurately reproduce the SMAP surface soil moisture.

The soil moisture product from SMAP is used in this study for comparison with the SMD estimations from NNsm. Here, the SMAP Level-3 soil moisture product (SPL3SMP) was chosen since it has been evaluated intensively against ground observations and proven to have a high quality performance in characterizing surface hydrometeorological processes<sup>52</sup>. It can provide global soil moisture at a daily timescale with 36 km spatial resolution. Studies have shown that out of several available soil moisture satellite products (i.e., AMSR2/-E and The Soil Moisture and Ocean Salinity, SMOS), SMAP has shown the lowest ubRMSE against *in-situ* networks<sup>53</sup>. The SMAP data, which can be accessed at https://nsidc.org/data/spl3smp/versions/9/<sup>54</sup>, is available from 2015.04.01. Here, 5 annual cycles, with timespans overlapping the NNsm data, were chosen from within the period of 2015.04.01 – 2020.03.31. Areas with vegetation water content (VWC) greater than 5 kg/m<sup>2</sup> were masked as the dense vegetation as these areas would interfere with soil moisture observations. Previous literature shows that in Earth System Models CP may be underestimated in arid regions, while PWP may be overestimated in humid regions<sup>33,37,50</sup>. Therefore, an experiment was designed to compare the mean square error (MSE) of PWP and CP with ISMN site results at different percentiles, and ultimately select the 10<sup>th</sup> percentile (MSE = 0.010) of CP to represent CP in hyper-arid and arid areas.

*In-situ* soil moisture data from 1334 stations in the ISMN (https://ismn.earth/en/)<sup>55,56</sup> dataset are used for verification. This dataset provides open soil moisture data from across the world to enable users to easily validate and improve global satellite products as well as land surface, climate, and hydrological models<sup>57</sup>. The original sampling frequency (SF) of this dataset is 1 hour<sup>-1</sup>. The data have been aggregated to 1 day<sup>-1</sup> and 1/3 day<sup>-1</sup> to be consistent with NNsm and SMAP, respectively. To diminish the scale mismatch between *in-situ* stations and satellite pixels to the maximum degree, at each specific grid square, averaged soil moisture from all stations within the grid square is regarded as a representative value to compare with the remote sensing products. For quantifying the bias between other CP (i.e., Fu) and PWP (i.e., GSDE and TG) products, only grid squares containing more than two stations were selected.

Data quality control procedures were followed to ensure the effectiveness of the analyses. *In-situ* stations with data records shorter than 1 year were excluded. Additionally, although ISMN provides *in-situ* soil moisture at



Fig. 1 Distribution of the selected ISMN stations. Red triangles indicate the selected ISMN stations. The background map shows the Köppen-Geiger (KG) Climate Classification.

Classification	Aridity Index (AI)
Hyper-arid	AI < 0.05
Arid	$0.05 \le AI < 0.2$
Semi-arid	$0.2 \le AI < 0.5$
Dry Subhumid	$0.5 \le AI < 0.65$
Humid	$AI \ge 0.65$

 Table 2.
 Climate classification and dryland subtypes based on the Aridity Index.

------

depths from 0 to 100 cm, only data from the depth of 5 cm are used in this study to ensure consistency with the vertical depths of the NNsm and SMAP data, which are 3 cm and 5 cm, respectively. The locations of all stations and the global climate zones<sup>58</sup> are displayed in Fig. 1.

**Ancillary data.** Since the analyses of short-term drydown requires precipitation information, the contemporaneous hourly Global Precipitation Mission (GPM) product<sup>59</sup> is used. The GPM data are aggregated into the same temporal and spatial resolutions as the NNsm data. In detail, GPM Level 3 IMERG Final Daily 0.1 degree × 0.1 degree (GPM\_3IMERGDF) derived from the half-hourly GPM\_3IMERGHH was adopted in this study. The derived result represents the final estimate of the daily accumulated precipitation. The dataset is produced at the NASA Goddard Earth Sciences (GES) Data and Information Services Center (DISC) by accumulating the valid precipitation retrievals for the day in GPM\_3IMERGHH and giving the result in the unit of mm.

To filter for the PWP and CP in humid, hyper-arid, and arid regions, ERA5 reanalysis data<sup>60</sup> were used to calculate the Aridity Index (AI). According to the classification from the World Atlas of Desertification (https://wad. jrc.ec.europa.eu/patternsaridity), the global land is divided by AI into 5 categories: hyper-arid, arid, semi-arid, dry subhumid, and humid. Here the humid area (AI  $\ge$  0.65), hyper-arid, and arid area (AI < 0.2) are defined in Table 2.

To compare CP and PWP across different vegetation types and soil textures, land use data from MCD12Q1 (https://lpdaac.usgs.gov/products/mcd12q1v061/)<sup>61</sup> spanning the period 2002 to 2023 and soil texture data from SoilGrids (https://www.isric.org/explore/soilgrids)<sup>62</sup> were resampled to a 36 km resolution to align with the NNsm grid. Land cover types were classified according to the International Geosphere-Biosphere Programme (IGBP) classification scheme. Only non-waterbody and non-permafrost land types with more than 20% of their area remaining after excluding VWC values greater than 5 kg/m<sup>2</sup> were retained for further analysis. Soil texture data, representing the mean values for the following parameters: clay content (g/100g, %), sand content (g/100g, %), were derived from the 0–5 cm depth layer. All the datasets mentioned above are summarized in Table 3.

#### **Data Records**

The data records<sup>63</sup> contain a global dataset of yearly soil critical point (CP) and permanent wilting point (PWP) and are publicly available at https://doi.org/10.6084/m9.figshare.27969276. Using McColl's<sup>36</sup> method to calculate drydown timescales, the dataset includes four other parameters that are related to SMD at two SFs: global medians of estimated parameters (i.e.,  $\tau_s$ , annual drydown times ( $n_{dd}$ ),  $\tau_1$ , and the fitting coefficient ( $R^2$ )) of the SMD

Dataset Name	Variable	Spatial resolution	Time	Link
NNsm	Soil moisture	36 km	2002.06.01-2023.12.31	https://cstr.cn/18406.11.Soil.tpdc.270960
SPL3SMP	Soil moisture	36 km	2015.04.01-2020.03.31	https://nsidc.org/data/spl3smp/versions/9/
ISMN	Soil moisture	1334 points	2015.04.01-2020.03.31	https://ismn.earth/en/
GPM_3IMERGDF	Precipitation	0.1 degree	2002.06.01-2023.12.31	https://gpm1.gesdisc.eosdis.nasa.gov/data/GPM_L3/GPM_3IMERGDF.07/
ERA5	2 m temperature, Total precipitation, Surface pressure, Surface net solar radiation, Surface net thermal radiation	0.25 degree	2002.06.01-2018.12.31	https://cds.climate.copernicus.eu/datasets/reanalysis-era5-single-levels?tab=download
Köppen- Geiger Climate Classification	Climate Classification	0.5 degree	1976-2000	http://koeppen-geiger.vu-wien.ac.at/shifts.htm
MCD12Q1	Land cover types are derived from the IGBP	250 m	2002-2023	https://lpdaac.usgs.gov/products/mcd12q1v061/
SoilGrids250m 2.0	Clay content, Sand content, Silt content	250 m	_	https://files.isric.org/soilgrids/latest/data/
Fu	СР	0.25 degree	—	https://zenodo.org/records/11183719
GSDE	PWP	0.083333 degree	_	https://cstr.cn/18406.11.Soil.tpdc.270578
TG	PWP	250 m	—	https://doi.org/10.5281/zenodo.2784001

Table 3. Summary of the datasets used in this study.

			1		
	Variable Name	Parameters [unit]	Meaning		
Parameter	tauS	$\tau_{S}$ [day]	Short-term timescale of drydown.		
	ndd	n <sub>dd</sub> [times]	Annual times of drydown event.		
	tauL	$\tau_L$ [day]	Long-term timescale of drydown.		
	thetaC	CP [m <sup>3</sup> /m <sup>3</sup> ]	Critical point.		
	thetaW	PWP [m <sup>3</sup> /m <sup>3</sup> ]	Permanent wilting point.		
	R2	R <sup>2</sup> [-]	Fitting coefficient of drydown curve. Range in [0-1]. The closer to 1, the better.		
Sample frequency	f1	1 [day <sup>-1</sup> ]	Once a day.		
	f3	1/3 [day-1]	Every three days.		
Year	From 2002 to 2023.				

Table 4. Composition of filename and corresponding meanings.

.....

curves. This dataset contains the global 36 km resolution soil characteristics from June 2002 to December 2023. There are 406 (latitude)  $\times$  964 (longitude) grid points in each file. The file format is.mat. All files can be opened in MATLAB. Each file stores data for one year. SMD data are missing from October to December 2011 and January to June 2012, due to missing NNsm data. The spatial resolution is  $36 \text{ km} \times 36 \text{ km}$ . The composition of the filename is: "Parameter"\_"Sample frequency"\_"Year".mat. Full details of these filename components are given in Table 4.

#### **Technical Validation**

**Spatial and temporal characteristics of soil CP and PWP.** Figure 2 shows the mean values of CP and PWP in different climate zones from 2002 to 2023. Due to missing data from November 2011 to June 2012, the 2012 data is replaced with the mean of 2011 and 2013 values. It can be seen that the time series of CP and PWP have only small fluctuations for most climate zones, indicating that the method of extracting CP and PWP using SMD is very robust.

Figure 3 shows the global median value and variation of CP and PWP. Areas with dense vegetation have been masked in advance. It can be seen that the global CP value (mode = 0.21) is higher than the PWP value (mode = 0.04). The range of PWP is very small (0–0.3 m<sup>3</sup>/m<sup>3</sup>). The coefficient of variation (CV) map clearly illustrates the low degree of temporal variability of CP (mean = 14%) and PWP (mean = 23%) over 22 years.

**Comparison with other CP and PWP products.** Figure 4 shows the global distribution of the other CP and PWP products. Comparison with Fig. 3 reveals that the spatial pattern of CP and PWP in our new dataset is broadly similar to that of the other products. However, the PWP of our dataset is slightly lower than that of the other products in North America, Brazil, South Africa, and Australia, while our CP is slightly higher than Fu's CP. The Fu results (mode = 0.15) are based on reanalysis data, while the GSDE results (mode = 0.11) are based on legacy site observations, and the TG results (mode = 0.13) are based on the estimation of PTFs. The variations in the data and methods used in these products from the statistical results based on satellite observations may cause the observed differences in the CP and PWP.



**Fig. 2** Mean value of (**a**) CP and (**b**) PWP based on Köppen-Geiger climate zones. The colored bars represent the mean values. Abbreviations in the legend represent different climate zones. A: equatorial, B: arid, C: warm temperate, D: snow, E: polar, W: desert, S: steppe, f: fully humid, s: summer dry, w: winter dry.





.....

Figure 5 shows scatter plots of CP and PWP from the different products against values from the ISMN sites. The correlation coefficient (r) and MSE are used to quantify the agreement between the two soil moisture datasets. The results show that, out of the three products, the PWP extracted based on NNsm is the closest to the site data. Although the r between the CP of Fu and the site data is slightly higher, it is found that the CP of our new dataset has a smaller MSE than that of Fu. Therefore, the CP and PWP based on long-term remote sensing observations of soil moisture have a higher accuracy compared to the other data products.

In 2015, 35 ISMN stations, distributed across different climate zones, were selected for calculation of their CP and PWP. These values were compared with those from Fu, TG, GSDE, and the dataset of CP and PWP created in this study. The CP and PWP values at these stations are presented in Fig. 6, along with comparisons to other products. Figure 6c shows that for 29 of the 35 stations, our CP values were closest to the ISMN station results. Due to the lower global grid point density of the Fu product compared to the CP grid points in our dataset, CP values could only be extracted from the Fu product for 21 out of 35 stations (60%). Therefore, in terms of data point validity, our dataset is more reliable.

For PWP, the station data exhibit a wide distribution range, indicating the large degree of variability in soil moisture characteristics across regions. The NNsm-derived PWP values have a closer alignment with the ISMN median in the Cfa, Cfb, and Dfc climate zones, as illustrated in Fig. 6b. In the Dfb and Dwc climate zones, the



**Fig. 4** Spatial distribution of the (**a**) PWP from GSDE (%), (**b**) PWP from TG (%), (**c**) CP from Fu ( $m^3/m^3$ ), (**d**) mean values by longitude based on PWP ( $m^3/m^3$ ) calculated in this study and other PWP products, and (**e**) mean values by longitude based on CP ( $m^3/m^3$ ) calculated in this study and CP of Fu. The subgraphs in panels a, b and c show the probability density function (PDF).



Fig. 5 CP and PWP calculated using various products compared with values from ISMN sites. (a) PWP derived from NNsm. (b) PWP derived from GSDE. (c) PWP derived from TG. (d) CP derived from NNsm. (e) CP derived from Fu. The r in the upper left corner of each panel represents the correlation coefficient (p < 0.01). MSE is the mean square error.

PWP medians from all products were relatively concentrated and closely matched the ISMN values, suggesting consistency in these regions. However, in some tropical and polar regions, there were insufficient valid data points to reliably validate the data.



**Fig. 6** CP and PWP in 2015 at 14 network sites and from different products. (**a**,**b**) show the boxplots of CP and PWP in different climate regions. Box edges are the 25th and 75th percentiles of the distribution bounding the median, and whiskers extend to extrema (maximum and minimum). (**c**,**d**) present the values of CP and PWP, respectively, as measured and derived from different products at the individual sites.

**Comparison of NNsm results with those from other soil moisture products.** To demonstrate the applicability and advantages of the NNsm-based CP and PWP, their values and those calculated based on SMAP were compared and validated with those reported by ISMN sites worldwide.

Figure 7 shows the results of the comparison of CP and PWP based on both NNsm and SMAP with the data from the ISMN sites. It is apparent that the values of CP (r = 0.42, p < 0.01) and PWP (r = 0.37, p < 0.01) calculated based on NNsm are a better match to the site-based data than are the values of CP (r = 0.24, p < 0.01) and

![](_page_8_Figure_1.jpeg)

**Fig.** 7 Scatter plots of NNsm and SMAP compared with ISMN for (**a**,**b**) CP and (**c**,**d**) PWP. The r in the upper left corner of each panel represents the correlation coefficient.

PWP (r = 0.13, p < 0.01) calculated based on SMAP. This better performance may be because NNsm's soil moisture data is closer to ground observations: the result of NNsm's ability to utilize daily soil moisture data, while SMAP's data is only provided every 2–3 days. Many studies have shown that satellite remote sensing observation data shows soil moisture decreasing at a faster rate than shown by station data<sup>64,65</sup>, a fact that indicates that there are likely to be some differences in the results produced by different soil moisture products. However, there are several reasons why both of the satellite calculation results differ from the site-based calculation results and these are described below.

First, the representativeness of the site space is limited. A site cannot capture all drydown events in a  $36 \text{ km} \times 36 \text{ km}$  grid, and the remote sensing products contain errors caused by large scale meteorological disturbances. The second reason is the heterogeneity of detection depth. There is a mismatch between the sensing depth of soil moisture based on the L-band radiometer and the detection depth of 5 cm at the sites<sup>66</sup>. In some densely vege-tated areas, the detection depth of the L-band may be less than 5 cm, but in some arid regions it may exceed this limit. There are also differences in site quality, and the variety of measurement methods used can result in differences. The effects of vegetation optical depth, climate type, surface roughness, and spatial heterogeneity need to be taken into account. According to Ma *et al.*<sup>67</sup>, the SMAP soil moisture products are similar to the site results under medium vegetation optical depth (VOD), small surface roughness, low heterogeneity conditions, and in temperate and frigid climate types. However, the product quality in high VOD, high roughness, or high terrain complexity and in tropical or desert areas needs to be improved. The above limitations are common issues for all satellite products, including SMAP and NNsm.

**Distribution of CP and PWP under different vegetation types and soil texture conditions.** As shown in Fig. 8, the CP and PWP of barren land had relatively low values. This was also the case in grassland and shrubland. In contrast, the CP and PWP of woody savannas and deciduous broadleaf forests were higher: a reasonable result as only sufficient precipitation can sustain the growth of vegetation with higher water demands. When the water requirements of the vegetation growing in a given area are higher, the CP and PWP also tend to be higher. Compared to other products, the median of CP extracted based on NNsm is closer to that extracted based on ISMN sites in woody savannas, cropland, and grassland. However, our PWP calculation performed poorly in woody savannas, exhibiting a large spread. This is possibly due to the excessive VWC of the woodland, which can interfere with L-band satellite observations of soil moisture. Despite this, the median PWP of this study is still close to that of TG in woody savannas. In other vegetation types, the PWP of this study was similar to the results of other products.

![](_page_9_Figure_1.jpeg)

![](_page_9_Figure_2.jpeg)

Figure 9 reveals a clear relationship between soil texture and both CP and PWP. Clay and silt fractions exhibit positive correlations with PWP (r = 0.18, p < 0.01 and r = 0.17, p < 0.01, respectively) and CP (r = 0.24, p < 0.01 and r = 0.13, p < 0.01, respectively), indicating that higher clay and silt content enhances the soil's water retention capacity and the lower limit of plant-available water. In contrast, sand content shows a negative correlation with both PWP (r = -0.24, p < 0.01) and CP (r = -0.24, p < 0.01), suggesting that increased sand content reduces water retention due to its larger pore size and better drainage properties. These findings align with the results reported by Feldman *et al.*<sup>68</sup> who also identified significant relationships between soil texture, vegetation type, and CP. Overall, the results underscore the importance of soil texture in determining soil hydraulic properties and their implications for vegetation growth.

**Limitations and future applications.** Our approach to extracting CP and PWP has several limitations. First, in regions with weak land-atmosphere interactions and frozen areas, the water and energy limitation stages of SMD may not occur concurrently<sup>1</sup> and therefore only one estimated parameter may hold valid in these regions. This limitation is due to an assumption made in the current SMD detection algorithm that requires further improvement in future studies. In addition, in some areas with high VWC, the observed soil moisture may be too high, leading to an overestimation of PWP. Users should be cautious when applying the data in the above areas. Second, the spatial resolution of our dataset is still too coarse to be implemented in regional hyper-resolution (e.g., <1 km) land surface simulations. Currently, very few global hyper-resolution soil moisture datasets exist. However, this is one promising direction for future studies as the framework proposed here has flexible applicability in terms of spatial resolution. It will be possible to release updated products of the derived CP and PWP parameters with higher resolution once the global soil moisture dataset is available. Users may also apply our framework to obtain their own estimates independently by using locally available soil moisture data.

Despite these limitations, our product has the ability to provide long-term, global-scale CP and PWP data. Compared to previous products, our dataset demonstrates improved accuracy when validated against ground station measurements, achieving results that align more closely with ISMN data. Our dataset is particularly valuable for large-scale global studies, as it provides comprehensive soil CP and PWP data across extensive spatial domains. Applications of this dataset include assessing multi-year dynamic changes in variables such as the timescale of drydown, CP, and PWP. These variables can serve as stable input parameters to advance the development of land surface models<sup>69</sup>. For this purpose, one recent study has already applied the preliminary version of our product to calibrating a global soil texture map and demonstrated the benefit of our product in improving land surface simulation accuracy<sup>70</sup>. Furthermore, researchers can utilize our dataset to monitor global soil moisture variations and identify regions where soil moisture falls below the PWP or CP, aiding the prediction of future drought trends<sup>37</sup>.

![](_page_10_Figure_1.jpeg)

Fig. 9 The relationship between (a-c) CP and soil texture and (d-f) PWP and soil texture. Box edges are the 25th and 75th percentiles of the distribution bounding the median, and whiskers extend to extrema (maximum and minimum). Bin counts are all greater than 600.

.....

#### **Code availability**

The code used in this study to calculate CP and PWP from NNsm is available at https://github.com/ AimeeMoMo1998/code-of-soil-hydraulic-properties.git

Received: 9 December 2024; Accepted: 22 April 2025; Published online: 30 April 2025

#### References

- 1. Akbar, R. *et al.* Estimation of Landscape Soil Water Losses from Satellite Observations of Soil Moisture. *J. Hydrometeorol.* **19**(5), 871–889 (2018).
- Feldman, A. F., Short Gianotti, D. J., Trigo, I. F., Salvucci, G. D. & Entekhabi, D. Satellite-based assessment of land surface energy partitioning-soil moisture relationships and effects of confounding variables. *Water Resour. Res.* 55(12), 10657–10677 (2019).
- . Seneviratne, S. I. *et al.* Investigating soil moisture-climate interactions in a changing climate: A review. *Earth Sci. Rev.* **99**, 125–161 (2010).
- 4. Lopez, F. B. & Barclay, G. F. in Pharmacognosy 45-60 (2017).
- 5. Rai, R. K., Singh, V. P. & Upadhyay, A. in Planning and Evaluation of Irrigation Projects 505-523 (2017).
- 6. Kirkham, M. B. in Principles of Soil and Plant Water Relations 169-189 (2023).
- 7. Zhao R. Hydrological Simulation-Xin' an jiang model and North of Shaanxi Model (in Chinese) (Beijing: Water Conservancy and Electric Power Press, 1984).
- Zhou, M. et al. Estimating potential evapotranspiration using Shuttleworth-Wallace model and NOAA AVHRRNDVI data to feed a distributed hydrological model over the Mekong river basin. J. Hydrometeor. 327, 151–173 (2006).
- Richard, C. et al. End-user-oriented pedotransfer functions to estimate soil bulk density and available water capacity at horizon and profile scales. Soil Use Manag 39(1), 270–285 (2022).
- Rab, M. et al. Modeling and prediction of soil water contents at field capacity and permanent wilting point of dryland cropping soils. Soil Res. 49, 389–407 (2011).
- 11. Assi, A., Blake, J., Mohtar, R. H. & Braudeau, E. Soil aggregates structure-based approach for quantifying the field capacity, permanent wilting point and available water capacity. *Irrig. Sci.* 37, 511–522 (2019).
- Reynolds, C., Jackson, T. & Rawls, W. Estimating soil water-holding capacities by linking the Food and Agriculture Organization soil map of the world with global pedon data bases and continuous pedotransfer functions. *Water Resour. Res.* 36(12), 3653–3662 (2000).
- Bonan, G. & Stillwell-Soller, L. Soil water and the persistence of floods and droughts in the Mississippi River basin. Water Resour. Res. 34(10), 2693–2701 (1998).
- 14. Sehgal, V., Gaur, N. & Mohanty, B. P. Global flash drought monitoring using surface soil moisture. *Water Resour. Res.* 57(9), e2021WR029901 (2021).
- Lorenz, R., Jaeger, E. B. & Seneviratne, S. I. Persistence of heat waves and its link to soil moisture memory. *Geophys. Res. Lett.* 37, L09703 (2010).
- Bayabil, H. K., Dile, Y. T., Tebebu, T. Y., Engda, T. A. & Steenhuis, T. S. Evaluating infiltration models and pedotransfer functions: Implications for hydrologic modeling. *Geoderma*. 338, 159–169 (2019).
- Pento's, K., Pieczarka, K. & Serwata, K. The Relationship between Soil Electrical Parameters and Compaction of Sandy Clay Loam Soil. Agriculture. 11, 114 (2021).
- 18. Scott, F. & Ramon, L. Automatic in situ determination of field capacity using soil moisture. Irrig. Drain. 61(3), 416-424 (2012).
- 19. Qiao, J. *et al.* Pedotransfer functions for estimating the field capacity and permanent wilting point in the critical zone of the Loess Plateau. China. J. Soils Sediments. 19, 140–147 (2019).
- 20. Amorim, R. *et al.* Water retention and availability in Brazilian Cerrado (neotropical savanna) soils under agricultural use: Pedotransfer functions and decision trees. *Soil Till. Res.* **224**, 105485 (2022).

- Filho, T. B. O., Leal, I. F., Macedo, J. R. D. & Reis, B. C. B. An Algebraic Pedotransfer Function to Calculate Standardized in situ Determined Field Capacity. J. Agric. Sci. 8, 158 (2016).
- 22. Wu, X. *et al*. An Integration Approach for Mapping Field Capacity of China Based on Multi-Source Soil Datasets. *Water.* **10**(6), 728 (2018).
- Liu, L. & Ma, X. Prediction of Soil Field Capacity and Permanent Wilting Point Using Accessible Parameters by Machine Learning. AgriEngineering 6, 2592–2611 (2024).
- Nourbakhsh, F., Afyuni, M., Abbaspour, K. C. & Schulin, R. Research note: estimation of field capacity and wilting point from basic soil physical and chemical properties. Arid Land Res. Manag. 19(1), 81–85 (2004).
- Mohanty, M. et al. Modelling Soil Water Contents at Field Capacity and Permanent Wilting Point Using Artificial Neural Network for Indian Soils. Natl. Acad. Sci. Lett. 38, 373–377 (2015).
- Tunçay, T., Alaboz, P., Dengiz, O. & Başkan, O. Application of regression kriging and machine learning methods to estimate soil moisture constants in a semi-arid terrestrial area. *Comput. Electron. Agric.* 212, 108118 (2023).
- Zotarelli, L., Dukes, M. & Morgan, K. Interpretation of Soil Moisture Content to Determine Soil Field Capacity and Avoid Over-Irrigating Sandy Soils Using Soil Moisture Sensors. AE460/AE460, 2/2010. EDIS 2010 (2010).
- Mohanty, B. & Gaur, N. Near Surface Soil Moisture Controls Beyond the Darcy Support Scale: A Remote Sensing Perspective. In Proceedings of the AGU Fall Meeting Abstracts. San Francisco, CA, USA, 15–19, H13D-1135 (2014).
- 29. Tunçay, T., Ba Skan, O., Bayramın, I., Dengız, O. & Kılıç, S. Geostatistical approach as a tool for estimation of field capacity and permanent wilting point in semi-arid terrestrial ecosystem. Arch. Agron. Soil Sci. 64, 1240–1253 (2018).
- Raoult, N., Ottlé, C., Peylin, P., Bastrikov, V. & Maugis, P. Evaluating and Optimizing Surface Soil Moisture Drydowns in the ORCHIDEE Land Surface Model at *In Situ* Locations. *J. Hydrometeorol.* 22(4), 1025–1043 (2021).
- McColl, K. A. et al. Global characterization of surface soil moisture drydowns: surface soil moisture drydown analysis. Geophys. Res. Lett. 44(8), 3682–3690 (2017).
- 32. Tso, C. M. *et al.* Multiproduct Characterization of Surface Soil Moisture Drydowns in the United Kingdom. *J. Hydrometeor.* 24, 2299–2319 (2023).
- Yang, S., Wu, B. & Yan, N. Estimation of Soil Field Capacity in North China Plain and Northeast China Based on AMSR-E Data. Chinese J. Soil Sci. 43(2), 301–305 (2012).
- 34. Bonan, G. Climate Change and Terrestrial Ecosystem Modeling (Cambridge Univ. Press, 2019).
- Van Looy, K. *et al.* Pedotransfer functions in Earth system science: Challenges and perspectives. *Rev. Geophys.* 55, 1199–1256 (2017).
   McColl, K. A., He, Q., Lu, H. & Entekhabi, D. Short-Term and Long-Term Surface Soil Moisture Memory Time Scales Are Spatially Anticorrelated at Global Scales. *J. Hydrometeorol.* 20(6), 1165–1182 (2019).
- 37. Fu, Z. et al. Global critical soil moisture thresholds of plant water stress. Nat. Commun. 15, 4826 (2024).
- Shangguan, W., Dai, Y., Duan, Q., Liu, B. & Yuan, H. A global soil data set for earth system modelling. J. Adv. Model. Earth Syst. 6(1), 249–263 (2014).
- Shangguan, W., Dai, Y. The global soil dataset for earth system modelling. National Tibetan Plateau / Third Pole Environment Data Center https://doi.org/10.11888/Soil.tpdc.270578 (2014).
- Tomislav, H. L. & Gupta, S. Soil water content (volumetric %) for 33 kPa and 1500 kPa suctions predicted at 6 standard depths (0, 10, 30, 60, 100 and 200 cm) at 250 m resolution (v0.1). Zenodo https://doi.org/10.5281/zenodo.2784001 (2019).
- Yao, P. P., Shi, J. C., Zhao, T. J., Lu, H. & Al-Yaari, A. Rebuilding Long Time Series Global Soil Moisture Products Using the Neural Network Adopting the Microwave Vegetation Index. Remote Sens. 9(8), 849 (2017).
- 42. Yao, P. et al. A long term global daily soil moisture dataset derived from AMSR-E and AMSR2 (2002–2019). Sci. Data. 8(1), 143 (2021).
- Sarmadian, F. & Keshavarzi, A. Developing pedotransfer functions for estimating some soil properties using artificial neural network and multivariate regression approaches. World Acad. Sci. Eng. Technol. 48, 427–433 (2010).
- Rodríguez-Iturbe, I. & Porporato, A. Ecohydrology of Water-Controlled Ecosystems: Soil Moisture and Plant Dynamics (*Cambridge Univ. Press*, 2007).
- Bautista, E., Clemmens, A. J., Strelkoff, T. S. & Schlegel, J. Modern analysis of surface irrigation systems with WinSRFR. Agric. Water Manag. 96(7), 1146–1154 (2009).
- Givi, J., Prasherb, S. O. & Patel, R. M. Evaluation of pedotransfer functions in predicting the soil water contents at field capacity and wilting point. Agric. Water Manag. 70, 83–96 (2004).
- 47. Morgan, C. L. Assessing Soil Health: Soil Water Cycling. Crops Soils. 53, 35–41 (2020).
- 48. Jian, J., Du, X. & Stewart, R. D. A database for global soil health assessment. Sci. Data. 7, 16 (2020).
- Bassiouni, M., Higgins, C. W., Still, C. J. & Good, S. P. Probabilistic inference of ecohydrological parameters using observations from point to satellite scales. *Hydrol. Earth Syst. Sci.* 22, 3229–3243 (2018).
- 50. Fu, Z. et al. Critical soil moisture thresholds of plant water stress in terrestrial ecosystems. Sci. Adv. 8, eabq 7827 (2022).
- Jahlvand, E. et al. Estimating the drainage rate from surface soil moisture drydowns: Application of DfD model to in situ soil moisture data. J. of Hydrol. 565, 489–501 (2018).
- 52. Entekhabi, D. et al. The Soil Moisture Active Passive (SMAP) Mission. Proc. IEEE. 98(5), 704-716 (2010).
- 53. Chan, S. K. et al. Assessment of the SMAP Passive Soil Moisture Product. IEEE Trans. Geosci. Remote Sens. 54(8), 4994–5007 (2016).
- O'Neill, P. E., Chan, S., Njoku, E. G., Jackson, T. & Bindlish, R. SMAP L3 Radiometer Global Daily 36 km EASE-Grid Soil Moisture, Version 4. NASA National Snow and Ice Data Center Distributed Active Archive Center https://doi.org/10.5067/OBBHQ5W22HME (2016).
- 55. 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**, 3 (2013).
- Dorigo, W. A. *et al.* The International Soil Moisture Network: serving Earth system science for over a decade. *Hydrol. Earth Syst. Sci.* 25, 5749–5804 (2021).
- 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).
- Rubel, F. & Kottek, M. Observed and projected climate shifts 1901–2100 depicted by world maps of the Köppen-Geiger climate classification. *Meteorol. Z.* 19, 135–141, http://koeppen-geiger.vu-wien.ac.at/shifts.htm (2010).
- Huffman, G. J., Stocker, E. F., Bolvin, D. T., Nelkin, E.J. & Tan, J. GPM IMERG Final Precipitation L3 1 day 0.1 degree x 0.1 degree V06, Goddard Earth Sciences Data and Information Services Center (GES DISC) https://doi.org/10.5067/GPM/IMERG/3B-HH/07 (2023).
- Hersbach, H. et al. ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) https://doi.org/10.24381/cds.adbb2d47 (2023).
- Friedl, M. & Sulla-Menashe, D. MODIS/Terra+Aqua Land Cover Type Yearly L3 Global 500m SIN Grid V061. NASA EOSDIS Land Processes Distributed Active Archive Center https://doi.org/10.5067/MODIS/MCD12Q1.061 (2022).
- 62. Poggio, L. et al. SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. Soil 7, 217–240 (2021).
- 63. Xu, Y., He, Q., Lu, H., Yang, K. & Entekhabi, D. A long-term dataset of global soil moisture drydowns. *figshare* https://doi. org/10.6084/m9.figshare.27969276 (2024).
- 64. Rondinelli, W. J. *et al.* Different rates of soil drying after rainfall are observed by the SMOS satellite and the South Fork *in situ* Soil Moisture Network. *J. Hydrometeorol.* **16**(2), 889–903 (2015).

- 65. Shellito, P. J. et al. SMAP soil moisture drying more rapid than observed in situ following rainfall events. Geophys. Res. Lett. 43, 8068–8075 (2016).
- 66. Zheng, D. *et al.* Sampling depth of L-band radiometer measurements of soil moisture and freeze-thaw dynamics on the Tibetan Plateau. *Remote Sens. Environ.* **226**, 16–25 (2019).
- 67. Ma, H. *et al.* Satellite surface soil moisture from SMAP, SMOS, AMSR2 and ESA CCI: A comprehensive assessment using global ground-based observations. *Remote Sens. Environ.* 231, 111215 (2019).
- 68. Feldman, A. F. et al. Moisture pulse-reserve in the soil-plant continuum observed across biomes. Nat. Plants 4, 1026–1033 (2018).
- He, Q., Lu, H. & Yang, K. Soil moisture memory of land surface models utilized in major reanalyses differ significantly from SMAP observation. *Earth's Future* 11, e2022EF003215 (2023).
- He, Q. et al. Global Optimization of Soil Texture Maps From Satellite-Observed Soil Moisture Drydowns and Its Implementation in Noah-MP Land Surface Model. J. Adv. Model Earth Sy. 16, e2023MS004142 (2024).

#### Acknowledgements

This work was jointly supported by the National Key Research and Development Program of China (2024YFE0106700 & 2022YFC3002901) and the National Natural Science Foundation of China (42371340). Computing support was provided by the High Performance Computing Centre of Tsinghua University. Qing He acknowledges funding from International Research Fellowships of Japan Society for the Promotion of Science (JSPS).

#### **Author contributions**

Hui Lu and Yawei Xu designed the framework of this work. Yawei Xu performed the computations, conducted data analysis, and wrote the paper. Qing He, Hui Lu and Kun Yang supervised the progress of this work, provided critical suggestions, and revised the paper. Dara Entekhabi and Daniel J. Short Gianotti advised on the method and results.

#### **Competing interests**

The authors declare no competing interests.

### **Additional information**

Correspondence and requests for materials should be addressed to H.L.

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

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

**Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, 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 licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

© The Author(s) 2025