



HAL
open science

Dataset of soil hydraulic parameters in the Yellow River Basin based on in situ deep sampling

Yongping Tong, Yunqiang Wang, Jingxiong Zhou, Xiangyu Guo, Ting Wang, Yuting Xu, Hui Sun, Pingping Zhang, Zimin Li, Ronny Lauerwald

► To cite this version:

Yongping Tong, Yunqiang Wang, Jingxiong Zhou, Xiangyu Guo, Ting Wang, et al.. Dataset of soil hydraulic parameters in the Yellow River Basin based on in situ deep sampling. *Scientific Data*, 2024, 11 (1), pp.740. 10.1038/s41597-024-03576-7. hal-04690108

HAL Id: hal-04690108

<https://hal.inrae.fr/hal-04690108v1>

Submitted on 6 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



OPEN

DATA DESCRIPTOR

Dataset of soil hydraulic parameters in the Yellow River Basin based on *in situ* deep sampling

Yongping Tong^{1,2,3,4}, Yunqiang Wang^{1,2,4,5}✉, Jingxiong Zhou^{1,2,4}, Xiangyu Guo^{1,4,6}, Ting Wang^{1,4,6}, Yuting Xu^{1,4,6}, Hui Sun^{1,4,6}, Pingping Zhang^{1,2,4}, Zimin Li^{1,2,4} & Ronny Lauerwald³

Soil hydraulic parameters are vital for precisely characterizing soil hydrological processes, which are critical indicators for regulating climate change effects on terrestrial ecosystems and governing feedbacks between water, energy, and carbon–nitrogen cycles. Although many studies have integrated comprehensive soil datasets, data quality and cost challenges result in data completeness deficiencies, especially for deep soil information. These gaps not only impede methodological endeavours but also constrain soil parameter-based ecosystem process studies spanning from local profiles to global earth system models. We established a soil dataset across the entire Yellow River Basin (YRB) (795,000 km²) using high-density *in situ* field sampling. This observation-based dataset contains records of soil texture (2924), bulk density (2798), saturated hydraulic conductivity (2782), and water retention curve parameters (1035) down to a maximum depth of 5 m. This dataset, which extends the recorded data range for deep soil hydraulic parameters, is valuable as a direct data resource for environmental, agronomical and hydrological studies in the YRB and regions with similar pedological and geological backgrounds around the world.

Background & Summary

Soils serve as a crucial interface between atmosphere, biosphere, hydrosphere and lithosphere^{1,2}, profoundly influencing matter and energy cycling within terrestrial ecosystems^{3–5}. In particular, soil hydrological processes play a pivotal role in regulating the impact of climate change on terrestrial ecosystems and feedback mechanisms between water, energy, and carbon–nitrogen cycles^{6–10}. Soil hydraulic parameters, which are in turn largely determined by soil texture and structure, serve as key factors in accurately depicting soil hydrological processes^{11–13}. For instance, the saturated conductivity (Ks) is a major control of moisture movement, distribution, and fluctuations within the soil profile^{14,15}. The matrix potential, which describes the strength of adhesive forces between soil moisture and solid components of the soil, determines the plant-availability of soil moisture. The soil water retention curve (SWRC), which defines the relationship between soil moisture content and matrix potential^{16,17}, affects a range of processes including evaporation¹⁸. Thus, it is one of the fundamental attributes that characterise soil hydraulics¹⁹.

The main methods for acquiring the aforementioned parameters encompass *in situ* sampling²⁰ and the use of pedotransfer function (PTF)^{21–23}. The considerable costs of *in situ* sampling has led to a growing interest in the establishment and use of PTFs^{9,24}. However, most PTFs are developed for specific regions, and their applicability to areas with different soil and climatic conditions is limited, necessitating re-calibration based on field measurements¹². The challenges associated with obtaining soil parameters not only impede methodological endeavours,

¹State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an, 710061, China. ²University of Chinese Academy of Sciences, Beijing, 100049, China. ³Université Paris-Saclay, INRAE, AgroParisTech, UMR Ecosys, Palaiseau, 91120, France. ⁴National Observation and Research Station of Earth Critical Zone on the Loess Plateau, 710061, Xi'an, China. ⁵Department of Earth and Environmental Sciences, Xi'an Jiaotong University, Xi'an, 710049, China. ⁶Xi'an Institute for Innovative Earth Environment Research, Xi'an, 710061, China. ✉e-mail: wangyq@ieecas.cn

such as up-scaling based on PTFs, but also impose limitations on ecosystem process studies that rely on soil parameters²⁵. These limitations affect research at a broad range of scales from site-level profile investigations^{26,27} to calibration and parametrization of comprehensive ecosystem models². Therefore, the accurate measurement of soil parameters is highly beneficial for assessing soil hydrological processes not just at local scale, but also for up-scaling to regional scales, and thus facilitating multiscale ecohydrological process studies²⁸.

Currently, a multitude of datasets, including Florida Soil Characterization Data²⁹, WoSIS³⁰, and UNSODA³¹, are dedicated to aggregating a diverse range of soil parameters derived from field measurements, comprising, in particular, the essential SWRC parameters. However, a considerable proportion of these data exhibits vague sample point coordinates and insufficient data pairs for establishing the SWRC, often lacking the wet end of the SWRC (water head ≤ 0.2 m)²⁴. To address the limitations of field measurements, some studies have employed integrated PTFs to derive soil hydraulic parameters at national and global scales^{32,33}. As illustrated by Gupta, Papritz²⁴, integrated field measurements with PTFs to effectively globally extend the applicability of soil hydraulic parameters by supplementing missing measurement data. However, the extensive datasets mentioned above still contain limited information regarding deep soil profiles, particularly regarding the scarcity of soil information below a depth of 3 m³⁴. Deep soil water, which is largely mediated by vegetation³⁵, acts an important role in enabling vegetation to withstand drought stress^{36,37} and water is also a key factor affecting the soil's ability to sequester carbon³⁸. Hence, deep soil hydraulic processes play an important role in terrestrial hydrology and soil carbon budgets³⁹. Given the potential impact of soil profile heterogeneity on hydraulic parameters^{9,28}, which constrains the applicability of shallow soil data, it becomes necessary to broaden the depth of investigation for soil hydraulic parameters. The compilation of deep soil profile information and incorporation of detailed field records would serve as a valuable complement to existing soil datasets.

Given the limitations of the current datasets outlined above, the objective of this study was to utilise geographically precise field measurements from deep soil profiles to extend existing soil datasets with reliable deep soil property records. Furthermore, we sought to provide a quantitative foundation to facilitate the development of PTFs that rely on original data. We conducted *in situ* sampling across the entire Yellow River Basin (YRB). The YRB is extensive (795,000 km²), irrigating over 15% of China's cultivated land and sustaining more than 12% of China's population⁴⁰. Furthermore, this basin encompasses most of China's important ecological barrier belt⁴¹, including the Loess Plateau (LP), the world's largest loess deposition region. Historically, severe soil erosion in this region has led to substantial loss of soil carbon to the ocean via the Yellow River, profoundly impacting the land carbon budget^{42–44}. Over the past two decades, China has been one of the leading contributors to the land greening observed around the globe⁴⁵, with the LP taking a prominent role through the "Grain for Green" program for ecological restoration^{12,46}. Given the significance of the YRB for global carbon cycling, climate change, food security, and ecological stability, the investigation of soil parameters in this region does not only hold the value for regional environmental and agronomic studies, but also provide some valuable supplementation to the current global pool of soil hydraulic datasets. Moreover, our dataset offers more possibilities for ecohydrological studies including observation and modelling that focus on deep profiles by providing soil hydraulic parameters down to a profile depth of 5 m.

During three years (2008, 2018, 2019) of fieldwork, we collected a total of 2925 disturbed soil samples and 2800 undisturbed soil samples throughout the whole YRB. This extensive, and high-density observation grid contains measurements of soil hydraulic properties down to a maximum depth of 5 m. The profiles were analyzed in the laboratory, and measurements were subjected to comprehensive data quality control and cleansing processes. Furthermore, we employed the "soilhyfit" package⁴⁷ in R (4.2.3 version) to fit the SWRC via the van Genuchten (VG) model. It should be noted that all SWRC records were derived from 10 pairs of corresponding soil matrix potential and moisture content data, covering a broad range of matrix potentials from 0.1 bar to 10 bars. For our dataset, we finally retained 2924 records of soil texture, 2798 records of soil bulk density (BD), 2782 records of Ks, and 1035 SWRC records. All records were consolidated into a unified dataset. This dataset further provides detailed meta-information for each sample, including sampling time, coordinates, elevation, depth, and land use type. We opted to preserve as much of the observed data as possible, but assigned categories of data quality which may help users to balance between quantity and quality of data depending on their research objectives and requirements. This dataset will be of value as a direct resource for environmental, agronomical and hydrological studies, as well as for calibrating PTFs. Although the spatial coverage of this dataset is limited, it covers the extensive YRB, filling the data gaps in this region and will also provide a useful data resource for studying other regions with comparable environmental setting worldwide. Finally, this dataset effectively extends the range of recorded data for deep soil hydraulic parameters around the world.

Methods

Study area and sampling site layout. The study area comprised the whole YRB (Fig. 1), which covers an approximate area of 795,000 km² (95°53'–119°5'E and 32°10'–42°50'N)^{2,48}. The Yellow River spans a length of 5464 km⁴⁹, ranking as the fifth longest river in the world. We acquired disturbed and undisturbed soil samples by conducting large-scale *in situ* sampling in two phases. The first phase involved high-density shallow-profile sampling from April to November 2008. The second phase comprised medium-density deep-profile sampling conducted from September to December 2018 and from October to November 2019. We selected the sampling sites by overlaying digital maps of the sampling area by a high-density sampling grid. This grid ensured uniform partitioning of the entire basin, with the centre of each grid serving as the initial choice of the sampling site. Subsequently, the sampling locations were adjusted based on topography, soil depth, and vegetation type to increase their representativeness. Ultimately, 382 sampling sites were established in the first phase and 93 in the second phase (Fig. 1).

Field methods. In the first phase, we excavated a 40 cm deep profile at each sampling point and collected disturbed and undisturbed soil samples from two layers (0–5 cm and 20–25 cm). During this stage, 764 disturbed

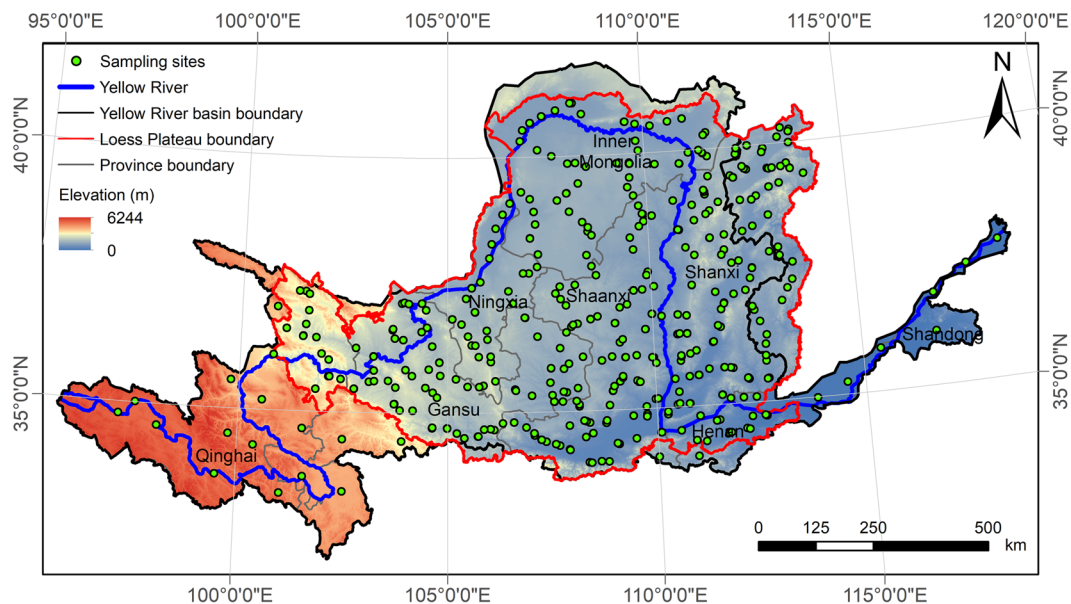


Fig. 1 Spatial distribution of soil sampling sites in the Yellow River Basin.

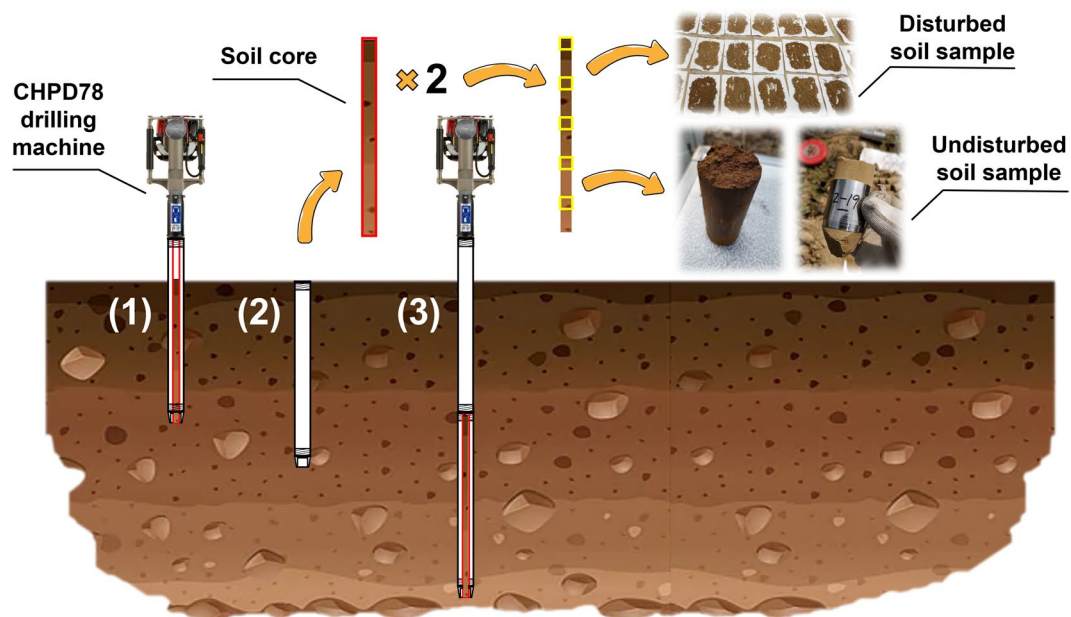


Fig. 2 Schematic of *in situ* soil sampling using a handheld drilling machine.

and 764 undisturbed soil samples were collected. Undisturbed soil cores were placed into metal cylinders after collection to facilitate the subsequent measurement of soil hydraulic parameters⁵⁰.

In the second phase, to facilitate deep undisturbed soil sampling (the targeted depth is 5 m), we employed a hand-held drilling machine (CHPD78, Christie Engineering Pty Ltd., Australia). To prevent compression in the soil core, a dual-tube setup was used within the drilling pipe, with an inner retrievable tube designed to accommodate the soil cores. The core diameter was 37 mm, and the inner tube was replaced every 1 m during drilling to ensure the sample integrity. To ensure sample correspondence, two boreholes (with 0.5 m distance) were drilled at each sampling point to retrieve the disturbed and undisturbed soil samples (Fig. 2). For the surface layer, the disturbed and undisturbed soil samples were obtained from the depth of 0.05 m. Subsequently, the sampling was carried out every 20 cm starting from the depth of 0.2 m. During this phase, 2161 disturbed and 2036 undisturbed soil samples were collected. Owing to constraints related to soil depth and the structure in certain layers, the number of undisturbed samples was lower than that of disturbed samples. As in the first phase, the undisturbed soil samples were placed into metal cylinders after collection. To prevent samples inside the metal cylinders from disturbance, we preserved them in a shockproof foam box after sampling and promptly

No.	1	2	3	4	5	6	7	8	9	10
ψ (bar)	0.01	0.10	0.30	0.60	0.80	1.00	3.00	6.00	8.00	10.00
h (mH ₂ O)	0.10	1.02	3.06	6.12	8.16	10.20	30.59	61.18	81.58	101.97

Table 1. Soil matrix potential and corresponding water head range when measuring the soil water retention curve based on centrifugation.

returned them to the laboratory for the storage. Ultimately, a total of 2925 disturbed soil samples and 2800 undisturbed soil samples were collected in the two phases.

Laboratory methods. All disturbed soil samples underwent preprocessing, including weed removal, air-drying, grinding, and sieving (using a 1 mm mesh), before particle size distribution was analyzed. Mastersizer laser particle size analysers (Mastersizer 3000, Malvern Panalytical, UK) were used to determine soil particle size distribution. Subsequently, the soil particle sizes were categorised according to the United States Department of Agriculture (USDA) standards into clay particles (< 0.02 mm), silt particles (0.02–0.5 mm), and sand particles (> 0.5 mm), leading to the classification of soil texture following USDA standards⁵¹.

The undisturbed soil samples were initially immersed for 24 h to achieve full saturation. Subsequently, we performed the determination of Ks using the constant-head method⁵², which involves in maintaining a constant water head infiltration through the Mariotte bottle until a stable infiltration rate is reached. Then, the amount of water passing through the sample within a fixed time were measured to calculate the Ks. Each sample was measured three times to ensure the accuracy. Centrifuge and pressure plate instrument methods are the most widely used methods for SWRC in the laboratory⁵³. The distinction between the two methods is as follows: In the low-suction range, the pressure plate method yields fewer data points and leads to a lower precision, whereas the centrifuge method provides relatively higher precision. However, the centrifuge method can be notably affected by density changes in soils with coarser textures, potentially resulting in lower precision. In the high-suction range, the pressure plate method may yield less accurate results for soils with high clay and silt contents because of inadequate drainage during the measurement process. In this case, the centrifuge method is more suitable. Considering the high silt content of most samples in this study and the time costs of pressure plate instrument method, the centrifuge method is more suitable for determining the SWRC. Utilising a centrifuge (CR21N, Hitachi, Japan), we set a series of different rotate speeds to correspond to different suction conditions (as being outlined in Table 1). After implementing each centrifugation process corresponding to different rotate speeds under a constant temperature of 20 °C, we removed the metal cylinders from the rotor, weighed, and recorded the total mass of the metal cylinders and the internal soil sample. Then, using the final measurement of the metal cylinders and dry soil mass, we calculated the gravimetric soil water contents corresponding to different suctions. Prior to measuring the SWRC, the soil saturation water content (θ_s) was initially tested. Subsequently, the BD was assessed after oven-drying (at 105 °C for 10 h), enabling the conversion of gravimetric water content to volumetric water content.

SWRC fitting and parameter acquisition based on the VG model. Upon obtaining soil water suction and volumetric moisture content data for each sampling point, we employed the “soilhypfit” package in R for fitting the SWRC using the “fit_wrc_hcc” function, in line with existing research²⁴. “soilhypfit” is an R package designed for the parametric modelling of soil water retention and hydraulic conductivity data. This function allows the estimation of SWRC parameters based on the van Genuchten (VG) model¹⁷, with the constraint $m = 1 - 1/n$. The VG equation (Eq. 1) is as follows:

$$\theta(\psi) = \frac{\theta_s - \theta_r}{[1 + (\alpha|\psi|)^n]^m} + \theta_r \quad (1)$$

where $\theta(\psi)$ (m³/m³) denotes the volumetric soil water content at matric potential ψ , and θ_s (m³/m³) and θ_r (m³/m³) represent the saturated and residual water contents, respectively. The α (m⁻¹) is a parameter related to the inverse of air entry pressure, and n is a dimensionless shape parameter of the VG model. During the prediction process, the “fit_wrc_hcc” function estimates parameters of the SWRC from respective measurements using the maximum likelihood method, optionally subject to physical constraints on the estimated parameters, and utilises the optimisation algorithm from the NLOpt library⁵⁴ or the Stochastic Complex Evolution (SCE) algorithm⁵⁵. According to existing research²⁴, we constrained n within the range from 1.0 to 7.0 and α within the range from 0 to 100 (m⁻¹) during the fitting process. Field capacity (FC) and permanent wilting point (PWP) are two key parameters that determine the soil water availability for plants and the maximum soil water-holding capacity⁵⁶. Hence, utilising the SWRC curves derived from the fitted VG models at each point, we projected the volumetric water content corresponding to FC (−1/3 bar, −3.37 mH₂O) and PWP (−15 bar, −152.96 mH₂O) for these sampling locations⁵⁷.

Data Records

After collating the measured and predicted soil parameters, a comprehensive soil hydraulic parameter dataset for the YRB was established. This dataset encompasses sampling points spanning the entire basin in terms of horizontal spatial distribution, with 382 shallow profile points at a resolution of 40 × 40 km and 93 deep profile points at a resolution of 100 × 100 km. The dataset has been uploaded and can be accessed via the link of <https://doi.pangaea.de/10.1594/PANGAEA.965004>⁵⁸.

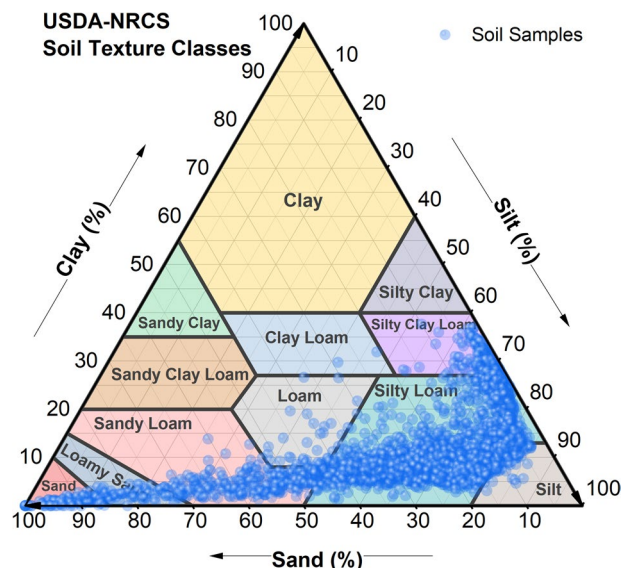


Fig. 3 Soil texture classification of samples from the Yellow River Basin. Classification was based on USDA soil texture classification standards.

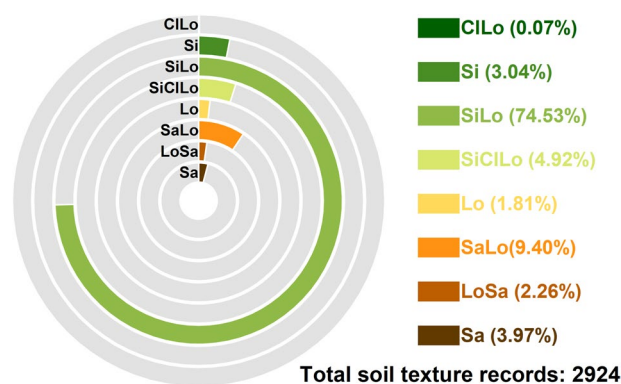


Fig. 4 Soil texture proportion of samples from the Yellow River Basin. Sa, Sand; LoSa, Loamy Sand; SaLo, Sandy Loam; Lo, Loam; SiClLo, Silt Clay Loam; SiLo, Silt Loam; Si, Silt; and CILO, Clay Loam.

All the data in the dataset, excluding the SWRC curve parameters, were derived from direct measurements. According to USDA classification, the soil texture in the dataset falls into two major categories: loamy and sandy soils (Fig. 3).

The loamy soil category includes sandy loam, loam, silt, silty loam, silty clay loam, and clay loam. The sandy soil category includes sand and loamy sand. Among them, silty loam constituted the highest proportion (74.53%), followed by sandy loam (9.40%). The remaining soil texture classes constitute less than 5% of sampling point (Fig. 4).

The original data for the SWRC curve of each sampling point was also derived from direct measurements by fitting a VG model to derive the θ_s , θ_p , and shape parameters (α and n). According to the kernel density plot, it can be observed that the fitted VG model parameters are generally distributed within a reasonable range (Fig. 5). We applied the necessary data cleaning and quality control procedures (see Technical Validation). To preserve the integrity of the original measurement data, we introduced the relative error range information (θ_s _RE_range) into the dataset to describe the quality of the SWRC parameter fitting.

Ultimately, 2925 disturbed field soil samples and 2800 undisturbed field soil samples were collected, and most profiles covered a depth down to 5 m. This dataset comprises 31 variables and 2925 records, and ultimately contains 2924 records for soil texture, 2798 records for soil BD, 2782 records for Ks, and 1035 records for SWRC parameters after data quality control and cleaning. A detailed description of each variable is provided in Table 2. Furthermore, a graphical representation (Venn diagram) illustrates the overlap among the different measurement indicators (Fig. 6).

Technical Validation

Data verification and cleaning. Prior to analysis, all field-collected samples underwent a preliminary inspection to ensure the integrity and non-mixing of the disturbed samples, and the undisturbed samples in the metal cylinders were free from vibration-induced cracking or any damage. Subsequently, the original

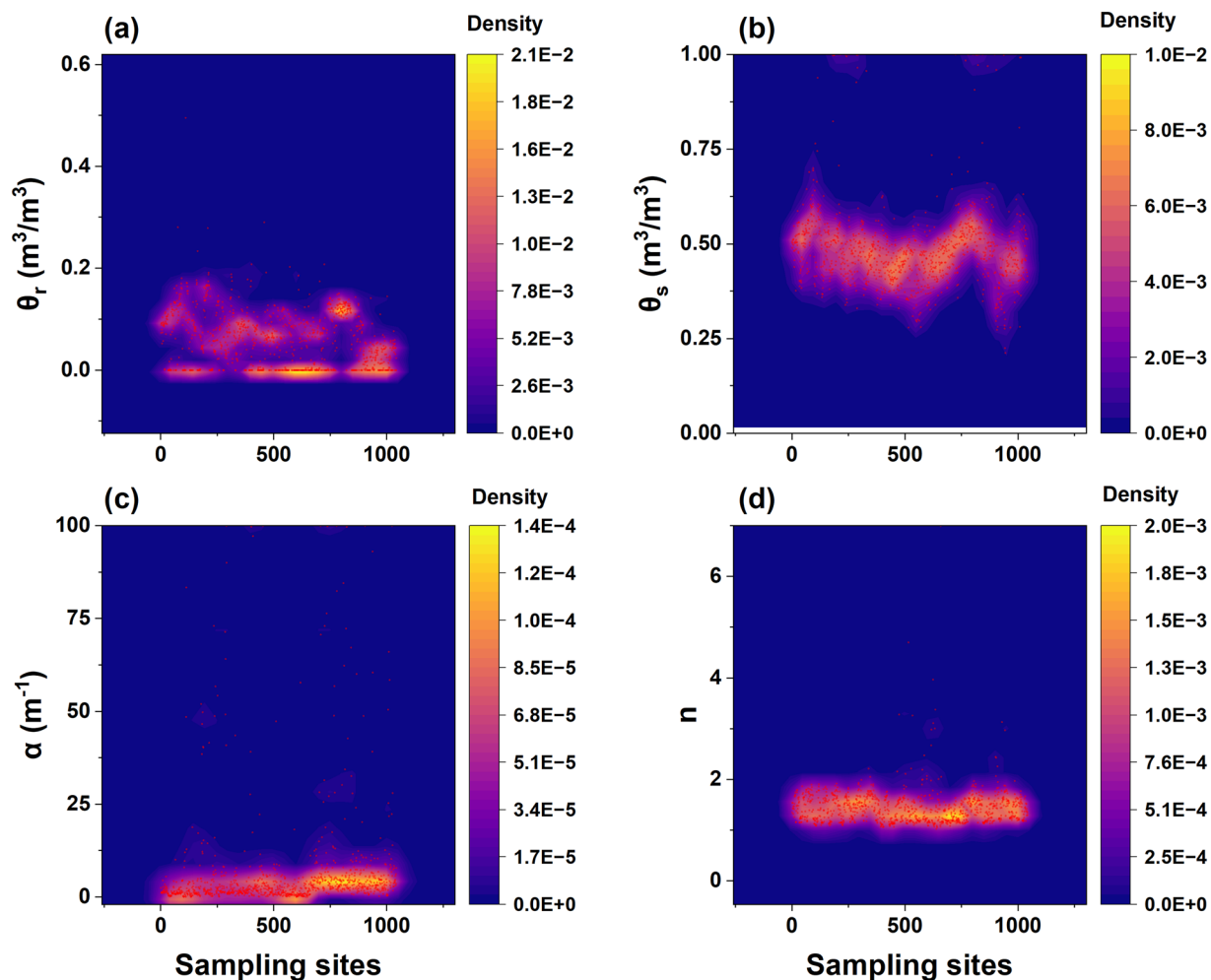


Fig. 5 Kernel density plots of van Genuchten (VG) model parameters distribution. θ_s represents the saturated water content (a); θ_r represents residual water content (b); α (c) and n (d) are the shape parameters of the VG model.

measurement data were subjected to thorough validation and data cleansing procedures. Regarding the BD, we eliminated the sample results with $BD > 2.65 \text{ g/cm}^3$ during the quality control process⁵⁹. Regarding the particle size distribution data, by comprehensively referring to the existed national measurement standards and literatures²⁴, we directly excluded the samples (11 records, denoted as “Error” in dataset) when the sum of particle size class contributions (clay + silt + sand) was not within $100 \pm 3\%$. Subsequently, samples were classified based on the absolute difference between the sum of particle size class fractions and 100% as follows: Level A ($< 1\%$), Level B ($1\% \leq \text{difference} < 2\%$), and Level C ($2\% \leq \text{difference} < 3\%$).

Constraints on VG fitting. In order to assess the quality of SWRC fitting based on the VG model using the “soilhyfit” package, we computed the coefficient of determination (R^2) for each model fit. As the “soilhyfit” package lacks a built-in function for directly calculating R^2 , we employed the following approach (Eq. 2):

$$R^2 = 1 - \frac{SSE}{SST} \quad (2)$$

where SSE represents the sum of squared errors obtained from the “ssq_wc” output of the “fit_wrc_hcc” function, while SST represents the total sum of squares total calculated based on the variances of the measured data at each point. All VG model fits yielded R^2 values above 0.9, indicating very high fitting performance.

To further assess the fitted data quality, we attempted to retrieve field surveys of soil hydraulic parameters from the same research area for the comparison. After filtering, we selected and plotted the spatial distribution of the mean hydraulic parameters within 0–5 m of the LP, which is the main body of the YRB (Fig. 7). The results show that the spatial distribution of θ_s , θ_r , α , and n all have zonal characteristics, exhibiting obvious spatial heterogeneity. Moreover, θ_s , α , and n have similar spatial distribution characteristics with existed investigation results⁶⁰ in most areas within the LP, which validates the reliability of our survey to some extents.

Header	Description	Unit
site_id	Number of the sampling site	—
sample_id	Number of samples at each sampling site	—
longitude	Longitude coordinates based on the WGS84 system	—
latitude	Latitude coordinates based on the WGS84 system	—
elevation	Elevation of the surface of the sampling site	m
land_use	Land-use type of the sampling site	—
sampling_year	The year in which the soil sample was collected	—
sampling_depth	The depth at which the soil sample was collected	m
clay	Clay (< 0.02 mm) content in soil samples	%
silt	Silt (0.02–0.5 mm) content in soil samples	%
sand	Sand (> 0.5 mm) content in soil samples	%
method_particle	Method for measuring soil particle composition	—
soil_texture_quality	Quality level of soil particle composition measuring, the levels of “A”, “B” and “C” represent high, medium and low data quality, respectively	—
soil_texture_class1	Soil texture classification based on USDA (broad categories)	—
soil_texture_class2	Soil texture classification based on USDA (subclasses)	—
BD	Bulk density of soil	g/cm ³
method_BD	Method for measuring soil bulk density	—
Ks	Soil saturated water conductivity	cm/min
method_Ks	Method for measuring soil saturated water conductivity	—
method_SWRC	Method for measuring soil water retention curves	—
measured_θ _s	Measured saturated soil water content from soil samples	m ³ /m ³
fit_θ _s	Fitted saturated soil water content based on “soilhyfit” package in R	m ³ /m ³
fit_θ _r	Fitted residual soil water content based on “soilhyfit” package in R	m ³ /m ³
fit_α	Fitted shape parameter of van Genuchten model based on “soilhyfit” package in R	m ⁻¹
fit_n	Fitted shape parameter of van Genuchten model based on “soilhyfit” package in R	—
fit_m	Fitted shape parameter of van Genuchten model based on “soilhyfit” package in R	—
fit_FC	Field capacity predicted by fitted SWRC curve based on “soilhyfit” package in R	m ³ /m ³
fit_PWP	Permanent wilting point predicted by fitted SWRC curve based on “soilhyfit” package in R	m ³ /m ³
fit_r ²	Coefficient of determination when fitting SWRC based on “soilhyfit” package in R	—
θ _s _RE	Absolute relative error of fitted and measured saturated soil water content	%
θ _s _RE_range	Quantile range of absolute relative error of fitted and measured saturated soil water content. “Q1”, “Q3”, “Min”, and “Max” represent the “first quartile”, “third quartile”, “Q1 minus 1.5IQR”, and “Q3 plus 1.5IQR”, respectively. “-” is a symbol used to represent the range of values.	—

Table 2. List of 31 variables in the Yellow River Basin soil hydraulic parameter dataset and their descriptions and units.

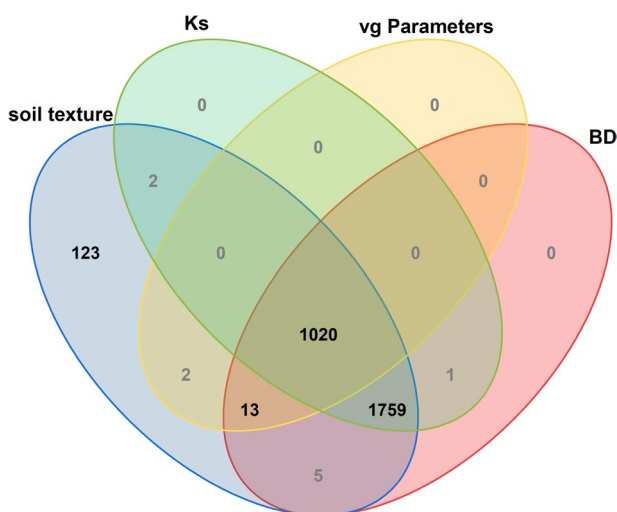


Fig. 6 Venn diagram illustrating the number of various measurement indicators derived from soil samples in the Yellow River Basin. BD represents the soil bulk density, Ks represents the soil saturated hydraulic conductivity, and vg Parameters represents the van Genuchten model parameters.

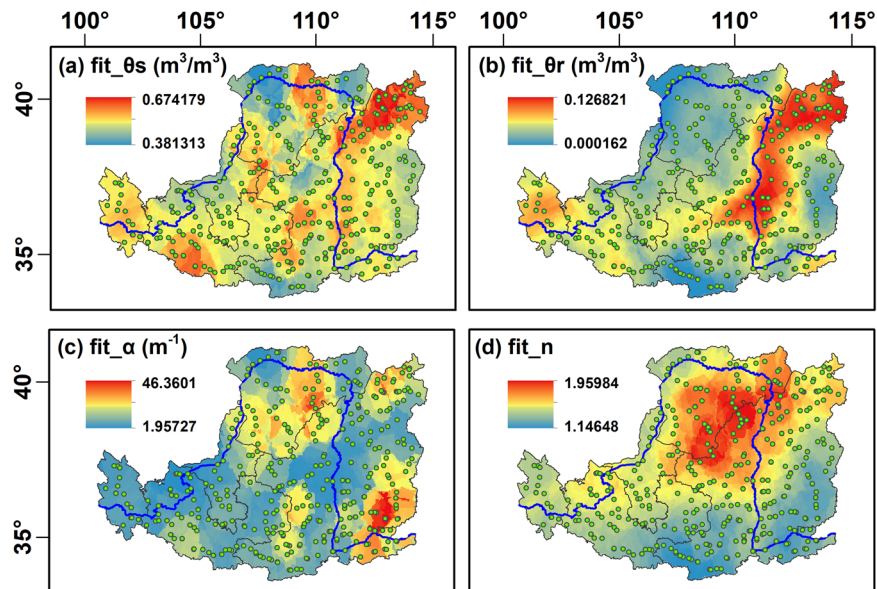


Fig. 7 Spatial distribution of the averaged soil hydraulic parameters within 0–5 m in the Chinese Loess Plateau. θ_s represents the saturated water content (**a**); θ_r represents residual water content (**b**); α (**c**) and n (**d**) are the shape parameters of the van Genuchten model. The green dots represent our sampling sites.

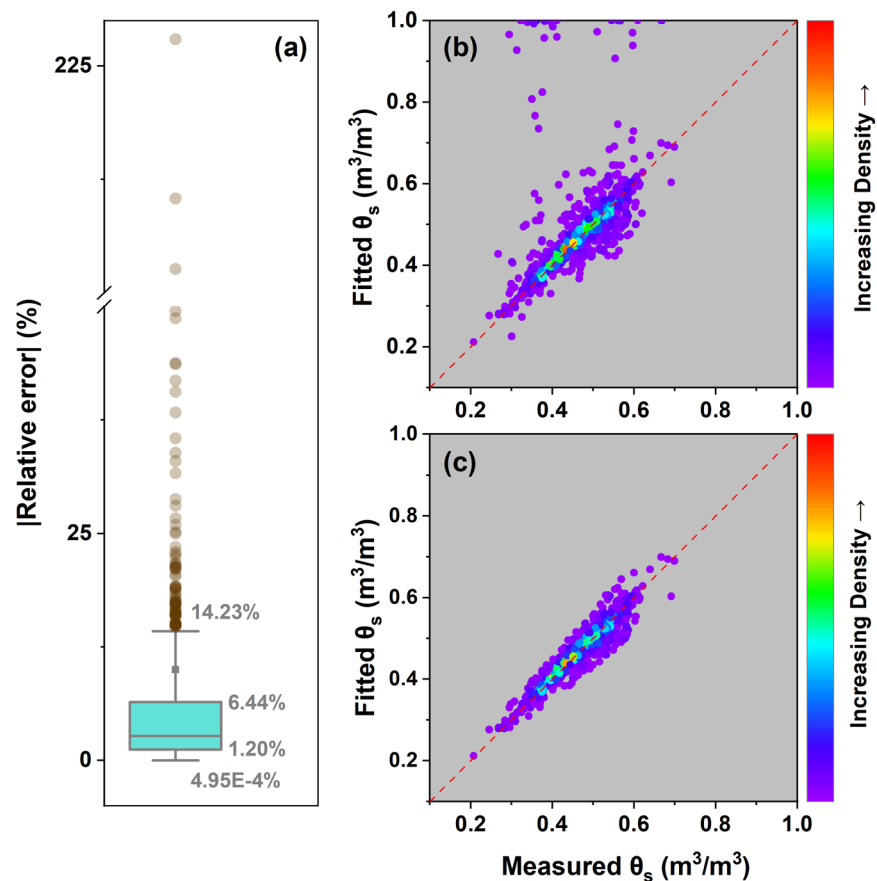


Fig. 8 Boxplot of the absolute relative error ($|RE|$) between the saturated water content (θ_s) fitted using the “soilhyfit” R package and measured θ_s (**a**); the distribution of measured and fitted θ_s for all data records (**b**); and the distribution of measured and fitted θ_s after removal of outliers of $|RE|$ (**c**). The red line in the figures represents a 1:1 ratio.

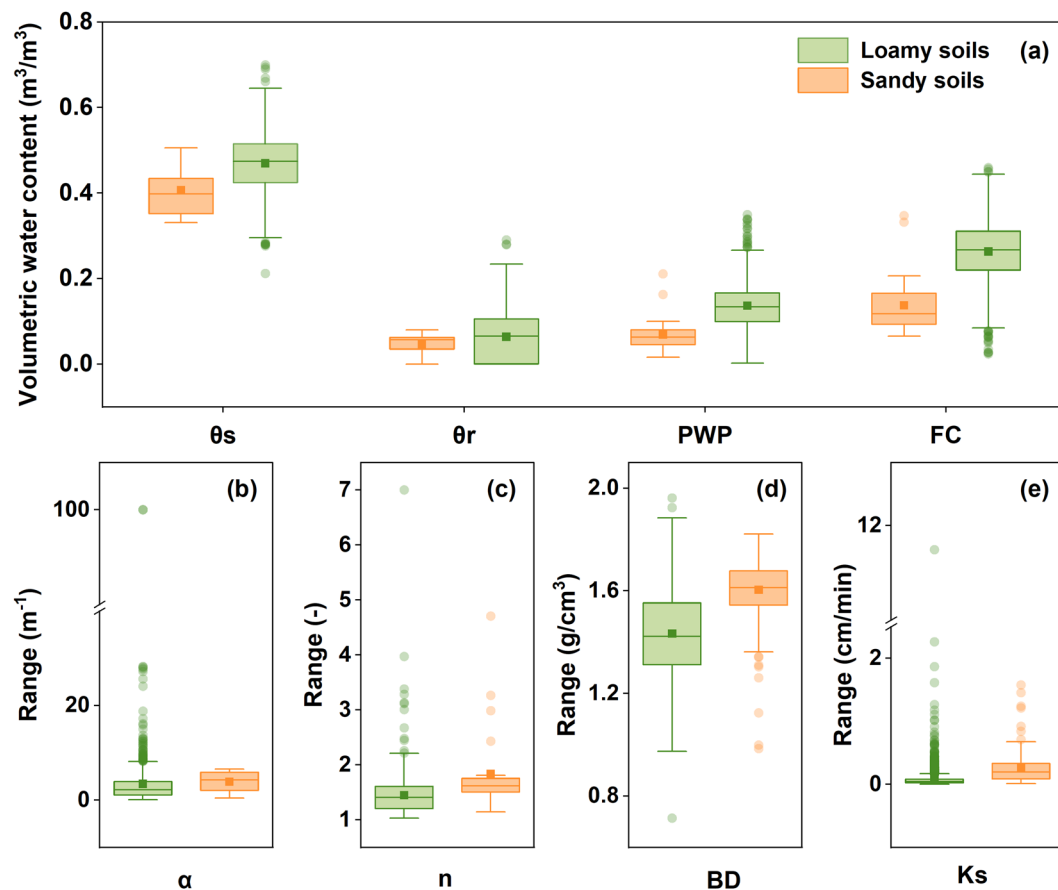


Fig. 9 Boxplots of the distribution of hydraulic parameters for major soil texture categories. θ_s represents the saturated water content, θ_r represents the residual water content, PWP represents the permanent wilting point and FC represents the field capacity (a); α (b) and n (c) represent the shape parameters of the van Genuchten model; BD represents the bulk density (d); and Ks represents the saturated hydraulic conductivity (e).

It should be noted that finding an investigation that perfectly matches our dataset in terms of survey range, depth, and number of sample sites still presents a challenge, limiting the quantitative comparison in space. Therefore, the quality of the fitted data needs to be further assessed. Besides, the inherent limitations of the VG model for fitting to soils with high sand/clay content also need to be considered. Hence, we further calculated the relative errors between fitted θ_s and measured θ_s to quantitatively evaluate the fitting quality of each sample point. The calculation method of $|RE|$ is as follows:

$$|RE| = \left| \frac{\theta_{s\text{fitted}} - \theta_{s\text{measured}}}{\theta_{s\text{measured}}} \times 100\% \right| \quad (3)$$

Subsequently, we identified the distribution characteristics of $|RE|$ by calculating the quartiles of these relative errors (Fig. 8a). The first (Q1) and third (Q3) quartiles were 1.20% and 6.44%, respectively. The $Q3 + 1.5IQR$ was 14.23%, and the $Q1 - 1.5IQR$ was 4.95E-4%. The results indicated that there were no $|RE|$ values lower than $Q1 - 1.5IQR$. Upon comparing the fitted θ_s and measured θ_s before and after outlier removal, it can be observed that the points after outlier removal are largely distributed along the 1:1 line (Fig. 8b,c). Therefore, in the final dataset, we further classified the SWRC parameters according to the quartile range of $|RE|$, and marked the sample points where $|RE|$ exceeds the range of $Q3 + 1.5IQR$ as “outliers”. To retain as much of the original data as possible, we included all the RE levels in the dataset.

By comparing the soil hydraulic parameters of the two main soil types in our datasets (loamy soil and sandy soil) after the outlier removal, we observed that for the loamy soil, all θ_s , θ_r , PWP, and FC were higher than those for sandy soil. In contrast, α , n , BD, and Ks for loamy soil were lower than those for sandy soil (Fig. 9). Furthermore, following the study by Goldberg *et al.*⁶¹, we further removed points with $FC > 48\%$ and $PWP > 36\%$.

Usage Notes

Considering that all data in this dataset originated from measurements of *in situ* samples, we strived to preserve the maximum number of sample test records and provided a grading system based on our quality assessment of soil texture measurements and SWRC fitting. Our intention was to allow researchers to freely choose which

data to use, and choose between quantity and quality of data according to their requirements. For soil texture measurements, we suggest to use the data at Level A ($< 1\%$) with confidence, while to use the data at level B ($1\% \leq \text{difference} < 2\%$) and C ($2\% \leq \text{difference} < 3\%$) selectively based on their specific requirements. For SWRC data, despite our efforts, some parameters still exceeded the predefined validity range, which included: one n parameter reached 7 and ten α parameters reached 100 m^{-1} (comprising 0.09% and 0.96% of the total valid SWRC count, respectively). Moreover, 236 θ_r were predicted as zero due to their inherently small actual values and 116 $|\text{RE}|$ of θ_s were listed as outliers (comprising 22.8% and 11.2% of the total valid SWRC count, respectively). We recommend cautious utilisation of these records. Constrained by sampling costs, the volume of the dataset remains limited. Nonetheless, we believe that this dataset, entirely based on measured data from *in situ* samples and encompassing soil hydraulic records down to a profile depth of 5 m, can effectively address the gaps in the pool of existing observational data, and the absence of deep soil information in particular.

Code availability

The code used to calculate SWRC parameters can be found on Github (<https://github.com/TONGYP1116/SoilHydraulicParameter.git>).

Received: 24 January 2024; Accepted: 26 June 2024;

Published online: 07 July 2024

References

- Paustian, K. *et al.* Climate-smart soils. *Nature* **532**, 49–57, <https://doi.org/10.1038/nature17174> (2016).
- Fatichi, S. *et al.* Soil structure is an important omission in Earth System Models. *Nat. Commun.* **11**, 522, <https://doi.org/10.1038/s41467-020-14411-z> (2020).
- Jung, M. *et al.* Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954, <https://doi.org/10.1038/nature09396> (2010).
- Cleveland, C. C. *et al.* Patterns of new versus recycled primary production in the terrestrial biosphere. *P. Natl. Acad. Sci.* **110**, 12733–12737, <https://doi.org/10.1073/pnas.1302768110> (2013).
- Peng, S. *et al.* Simulated high-latitude soil thermal dynamics during the past 4 decades. *Cryosphere* **10**, 179–192, <https://tc.copernicus.org/articles/10/179/2016/> (2016).
- Seneviratne, S. I. *et al.* Investigating soil moisture–climate interactions in a changing climate: A review. *Earth-Sci. Rev.* **99**, 125–161, <https://doi.org/10.1016/j.earscirev.2010.02.004> (2010).
- Lohse, K. A., Brooks, P. D., McIntosh, J. C., Meixner, T. & Huxman, T. E. Interactions between biogeochemistry and hydrologic systems. *Annu. Rev. Env. Resour.* **34**, 65–96, <https://doi.org/10.1146/annurev.enviro.33.031207.111141> (2009).
- Green, J. K. *et al.* Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature* **565**, 476–479, <https://doi.org/10.1038/s41586-018-0848-x> (2019).
- Vereecken, H. *et al.* Soil hydrology in the Earth system. *Nat. Rev. Earth Env.* **3**, 573–587, <https://doi.org/10.1038/s43017-022-00324-6> (2022).
- Zhu, Q., Castellano, M. J. & Yang, G. Coupling soil water processes and the nitrogen cycle across spatial scales: Potentials, bottlenecks and solutions. *Earth-Sci. Rev.* **187**, 248–258, <https://doi.org/10.1016/j.earscirev.2018.10.005> (2018).
- Lin, H. *et al.* Hydrogeology: Synergistic integration of pedology and hydrology. *Water Resour. Res.* **42**, <https://doi.org/10.1029/2005WR004085> (2006).
- Bai, X., Shao, M. A., Jia, X. & Zhao, C. Prediction of the van Genuchten model soil hydraulic parameters for the 5-m soil profile in China's Loess Plateau. *CATENA* **210**, 105889, <https://doi.org/10.1016/j.catena.2021.105889> (2022).
- Wang, T. *et al.* Evaluating climate and soil effects on regional soil moisture spatial variability using EOFs. *Water Resour. Res.* **53**, 4022–4035, <https://doi.org/10.1002/2017WR020642> (2017).
- Zhao, C., Shao, M. A., Jia, X., Nasir, M. & Zhang, C. Using pedotransfer functions to estimate soil hydraulic conductivity in the Loess Plateau of China. *CATENA* **143**, 1–6, <https://doi.org/10.1016/j.catena.2016.03.037> (2016).
- Usovich, B. & Lipiec, J. Spatial variability of saturated hydraulic conductivity and its links with other soil properties at the regional scale. *Sci. Rep.* **11**, 8293, <https://doi.org/10.1038/s41598-021-86862-3> (2021).
- Brooks, R. H. & Corey, A. T. Hydraulic properties of porous media and their relation to drainage design. *T. ASAE* **7**, 26–0028, <https://doi.org/10.13031/2013.40684> (1964).
- van Genuchten, M. T. A Closed-form equation for predicting the hydraulic conductivity of unsaturated soils. *Soil Sci. Soc. Am. J.* **44**, 892–898, <https://doi.org/10.2136/sssaj1980.03615995004400050002x> (1980).
- Ciocca, F., Lunati, I. & Parlange, M. B. Effects of the water retention curve on evaporation from arid soils. *Geophys. Res. Lett.* **41**, 3110–3116, <https://doi.org/10.1002/2014GL059827> (2014).
- Assouline, S., Tessier, D. & Bruand, A. A conceptual model of the soil water retention curve. *Water Resour. Res.* **34**, 223–231, <https://doi.org/10.1029/97WR03039> (1998).
- Vereecken, H. *et al.* On the value of soil moisture measurements in vadose zone hydrology: A review. *Water Resour. Res.* **44**, <https://doi.org/10.1029/2008WR006829> (2008).
- Vereecken, H. *et al.* Using pedotransfer functions to estimate the van Genuchten–Mualem soil hydraulic properties: A review. *Vadose Zone J.* **9**, 795–820, <https://doi.org/10.2136/vzj2010.0045> (2010).
- Zhang, Y. & Schaap, M. G. Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3). *J. Hydrol.* **547**, 39–53, <https://doi.org/10.1016/j.jhydrol.2017.01.004> (2017).
- Van Looy, K. *et al.* Pedotransfer functions in Earth system science: Challenges and perspectives. *Rev. Geophys.* **55**, 1199–1256, <https://doi.org/10.1002/2017RG000581> (2017).
- Gupta, S. *et al.* Global soil hydraulic properties dataset based on legacy site observations and robust parameterization. *Sci. Data* **9**, 444, <https://doi.org/10.1038/s41597-022-01481-5> (2022).
- Comber, A. *et al.* A generic approach for live prediction of the risk of agricultural field runoff and delivery to watercourses: linking parsimonious soil-water-connectivity models with live weather data apis in decision tools. *Front. Sustain. Food Syst.* **3**, 42, <https://doi.org/10.3389/fsufs.2019.00042> (2019).
- Hartemink, A. E. & Minasny, B. Towards digital soil morphometrics. *Geoderma* **230–231**, 305–317, <https://doi.org/10.1016/j.geoderma.2014.03.008> (2014).
- Vereecken, H. *et al.* Infiltration from the pedon to global grid scales: An overview and outlook for land surface modeling. *Vadose Zone J.* **18**, 180191, <https://doi.org/10.2136/vzj2018.10.0191> (2019).
- Pan, T., Hou, S., Liu, Y. & Tan, Q. Comparison of three models fitting the soil water retention curves in a degraded alpine meadow region. *Sci. Rep.* **9**, 18407, <https://doi.org/10.1038/s41598-019-54449-8> (2019).

29. Grunwald, S. Florida soil characterization data. Soil and water science department, IFAS-Institute of Food and Agriculture Science, Tech. Rep., University of Florida, website: <https://www.sgrunwald.org/big-data> (2020).
30. Batjes, N. H., Ribeiro, E. & van Oostrum, A. Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019). *Earth Syst. Sci. Data* **12**, 299–320, <https://doi.org/10.5194/essd-12-299-2020> (2020).
31. Nemes, A., Schaap, M. G., Leij, F. J. & Wösten, J. H. M. Description of the unsaturated soil hydraulic database UNSODA version 2.0. *J. Hydrol.* **251**, 151–162, [https://doi.org/10.1016/S0022-1694\(01\)00465-6](https://doi.org/10.1016/S0022-1694(01)00465-6) (2001).
32. Dai, Y. *et al.* Development of a China dataset of soil hydraulic parameters using pedotransfer functions for land surface modeling. *J. Hydrometeorol.* **14**, 869–887, <https://doi.org/10.1175/JHM-D-12-0149.1> (2013).
33. Zhang, Y., Schaap, M. G. & Wei, Z. Development of hierarchical ensemble model and estimates of soil water retention with global coverage. *Geophys. Res. Lett.* **47**, e2020GL088819, <https://doi.org/10.1029/2020GL088819> (2020).
34. Qiao, J., Zhu, Y., Jia, X., Huang, L. & Shao, M. Development of pedotransfer functions for soil hydraulic properties in the critical zone on the Loess Plateau, China. *Hydrol. Processes* **32**, 2915–2921, <https://doi.org/10.1002/hyp.13216> (2018).
35. Wang, Y. *et al.* Soil moisture decline in China's monsoon loess critical zone: More a result of land-use conversion than climate change. *P. Natl. Acad. Sci.* **121**, e2322127121, <https://doi.org/10.1073/pnas.2322127121> (2024).
36. Miguez-Macho, G. & Fan, Y. Spatiotemporal origin of soil water taken up by vegetation. *Nature* **598**, 624–628, <https://doi.org/10.1038/s41586-021-03958-6> (2021).
37. Gao, X. *et al.* Disentangling the impact of event- and annual-scale precipitation extremes on critical-zone hydrology in semiarid loess vegetated by apple trees. *Water Resour. Res.* **59**, e2022WR033042, <https://doi.org/10.1029/2022WR033042> (2023).
38. Heckman, K. A. *et al.* Moisture-driven divergence in mineral-associated soil carbon persistence. *P. Natl. Acad. Sci.* **120**, e2210044120, <https://doi.org/10.1073/pnas.2210044120> (2023).
39. Dawson, T. E., Hahm, W. J. & Crutchfield-Peters, K. Digging deeper: what the critical zone perspective adds to the study of plant ecophysiology. *New Phytol.* **226**, 666–671, <https://doi.org/10.1111/nph.16410> (2020).
40. Lin, M., Biswas, A. & Bennett, E. M. Spatio-temporal dynamics of groundwater storage changes in the Yellow River Basin. *J. Environ. Manage.* **235**, 84–95, <https://doi.org/10.1016/j.jenvman.2019.01.016> (2019).
41. Yin, L. *et al.* Trade-offs and synergy between ecosystem services in National Barrier Zone. *Geogr. Res.* **38**, 2162–2172, <https://doi.org/10.11821/dlyj020180578> (2019).
42. Ran, L. *et al.* Spatial and seasonal variability of organic carbon transport in the Yellow River, China. *J. Hydrol.* **498**, 76–88, <https://doi.org/10.1016/j.jhydrol.2013.06.018> (2013).
43. Yang, Y. *et al.* Estimating soil organic carbon redistribution in three major river basins of China based on erosion processes. *Soil Res.* **58**, 540–550, <https://doi.org/10.1071/SR19325> (2020).
44. Wang, X., Ma, H., Li, R., Song, Z. & Wu, J. Seasonal fluxes and source variation of organic carbon transported by two major Chinese rivers: The Yellow River and Changjiang (Yangtze) River. *Global Biogeochem. Cy.* **26**, <https://doi.org/10.1029/2011GB004130> (2012).
45. Chen, C. *et al.* China and India lead in greening of the world through land-use management. *Nat. Sustain.* **2**, 122–129, <https://doi.org/10.1038/s41893-019-0220-7> (2019).
46. Chen, Y. *et al.* Balancing green and grain trade. *Nat. Geosci.* **8**, 739–741, <https://doi.org/10.1038/ngeo2544> (2015).
47. Papritz, A. Soilhyppfit: modelling of soil water retention and hydraulic conductivity data. R package version 0.1-7. website: <https://rdrr.io/cran/soilhyppfit/> (2022).
48. Wang, W., Zhang, Y. & Tang, Q. Impact assessment of climate change and human activities on streamflow signatures in the Yellow River Basin using the Budyko hypothesis and derived differential equation. *J. Hydrol.* **591**, 125460, <https://doi.org/10.1016/j.jhydrol.2020.125460> (2020).
49. Xie, P. *et al.* Spatial-temporal variations in blue and green water resources, water footprints and water scarcities in a large river basin: A case for the Yellow River basin. *J. Hydrol.* **590**, 125222, <https://doi.org/10.1016/j.jhydrol.2020.125222> (2020).
50. Wang, Y., Shao, M. A., Liu, Z. & Horton, R. Regional-scale variation and distribution patterns of soil saturated hydraulic conductivities in surface and subsurface layers in the loessial soils of China. *J. Hydrol.* **487**, 13–23, <https://doi.org/10.1016/j.jhydrol.2013.02.006> (2013).
51. Soil Science Division Staff. *Soil survey manual*. In USDA Handbook 18 (Government Printing Office, Washington, D.C., 2017).
52. Klute, A. & Dirksen, C. in *Methods of Soil Analysis*, 687–734, <https://doi.org/10.2136/sssabookser5.1.2ed.c28> (SSSA Book Series, 1986).
53. Rahardjo, H., Nong, X. F., Lee, D., Leong, E. C. & Fong, Y. Expedited soil–water characteristic curve tests using combined centrifuge and chilled mirror techniques. *Geotech. Test. J.* **41**, 207–217, <https://doi.org/10.1520/GTJ20160275> (2018).
54. Johnson, S. G. The NLOpt Nonlinear-Optimization Package, website: <https://nlopt.readthedocs.io/en/latest/> (2014).
55. Duan, Q., Sorooshian, S. & Gupta, V. K. Optimal use of the SCE-UA global optimization method for calibrating watershed models. *J. Hydrol.* **158**, 265–284, [https://doi.org/10.1016/0022-1694\(94\)90057-4](https://doi.org/10.1016/0022-1694(94)90057-4) (1994).
56. Li, X., Shao, M. A. & Zhao, C. Estimating the field capacity and permanent wilting point at the regional scale for the Hexi Corridor in China using a state-space modeling approach. *J. Soil. Sediment.* **19**, 3805–3816, <https://doi.org/10.1007/s11368-019-02314-6> (2019).
57. Reynolds, C. A., Jackson, T. J. & Rawls, W. J. Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. *Water Resour. Res.* **36**, 3653–3662, <https://doi.org/10.1029/2000WR900130> (2000).
58. Tong, Y. *et al.* Dataset of Soil Hydraulic Parameters in the Yellow River Basin, PANGAEA, <https://doi.pangaea.de/10.1594/PANGAEA.965004> (2024).
59. Morales-Durán, N., Fuentes, S. & Chávez, C. A soil database from Queretaro, Mexico for assessment of crop and irrigation water requirements. *Sci. Data* **10**, 429, <https://doi.org/10.1038/s41597-023-02332-7> (2023).
60. Fang, S. *et al.* The distribution of Van Genuchten model parameters on soil–water characteristic curves in Chinese Loess Plateau and new predicting method on unsaturated permeability coefficient of loess. *PLoS ONE* **18**, e0278307, <https://doi.org/10.1371/journal.pone.0278307> (2023).
61. Goldberg, D., Gornat, B. & Rimon, D. *Drip Irrigation: Principles, Design and Agricultural Practices* (Drip Irrigation Scientific Publications, 1976).

Acknowledgements

This research was supported by National Key Research and Development Program of China (No. 2019YFA0607303), National Natural Science Foundation of China (No. U2243204, 42177306, and 41977422), Innovation Cross Team-Key Laboratory project of the Chinese Academy of Sciences, the Youth Innovation Promotion Association CAS, Shaanxi Province Natural Science Basic Research Program (No. 2023JC-XJ-10), and Shaanxi Province Innovation Capability Support Plan Project (No. 2024RS-CXTD- 45).

Author contributions

Yongping Tong: Investigation, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. Yunqiang Wang: Conceptualization, Project administration, Supervision, Writing – review & editing. Jingxiong Zhou: Investigation. Xiangyu Guo: Investigation. Ting Wang: Investigation. Yuting Xu: Investigation. Hui Sun: Data curation. Pingping Zhang: Writing – review & editing. Zimin Li: Writing – review & editing. Ronny Lauerwald: Conceptualization, Supervision, Writing – review & editing.

Competing interests

The authors declare that they have no competing interests. This includes, but is not limited to, financial interests, non-financial interests, patents, or any other interests that may be perceived as influencing the research or its presentation. The corresponding author affirms that this information has been discussed and agreed upon by all authors listed.

Additional information

Correspondence and requests for materials should be addressed to Y.W.

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 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 licence, and indicate if changes were made. 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/4.0/>.

© The Author(s) 2024