Addressing soil data needs and data-gaps in catchment scale environmental modelling: the European perspective

Brigitta Szabó[1,2], Piroska Kassai[1,2], Svajunas Plunge[3], Attila Nemes[4], Péter Braun[1,2,5], Michael Strauch[6], Felix Witing[6], János Mészáros[1,2], Natalja Čerkasova[5,7]

[1]Institute for Soil Sciences, HUN-REN Centre for Agricultural Research, Budapest, 1022, Hungary
[2]National Laboratory for Water Science and Water Security, Budapest, 1022, Hungary
[3]Department of Hydrology, Meteorology, and Water Management, Institute of Environmental Engineering, Warsaw University of Life Sciences, Warsaw, 0-653, Poland
[4]Norwegian Institute of Bioeconomy Research, Ås, 1431, Norway
[5]Marine Research Institute, Klaipeda University, Klaipeda, 92294, Lithuania
[6]Helmholtz Centre for Environmental Research GmbH - UFZ, Department of Computational Landscape Ecology, Leipzig, 04318, Germany
[7]Texas A&M AgriLife, Blackland Research and Extension Center, Temple, TX 76502, USA

Correspondence to: Piroska Kassai (kassai.piroska@atk.hun-ren.hu)

Abstract. To effectively guide agricultural management planning strategies and policy, it is important to simulate water quantity and quality patterns and to quantify the impact of land use and climate change on soil functions, soil health, hydrological, and other underlying processes. Environmental models that depict alterations in surface and groundwater quality and quantity at a catchment scale require substantial input, particularly concerning movement and retention in the unsaturated zone. Over the past few decades, numerous soil information sources, containing structured data on diverse basic and advanced soil parameters, alongside innovative solutions to estimate missing soil data, have become increasingly available. This study aims to: i) catalogue open-source soil datasets and pedotransfer functions (PTFs) applicable in simulation studies across European catchments, ii) evaluate the performance of selected PTFs and iii) present compiled R scripts proposing estimation solutions to address soil physical, hydraulic, and chemical soil data needs and gaps in catchment-scale environmental modelling in Europe. Our focus encompassed basic soil properties, bulk density, porosity, albedo, soil erodibility factor, field capacity, wilting point, available water capacity, saturated hydraulic conductivity, and phosphorus content. We aim to recommend widely supported data sources and pioneering prediction methods that maintain physical consistency, and present them through streamlined workflows.

1 Introduction

The availability of raw and derived soil datasets, specifically soil hydraulic data, has increased significantly in Europe over the last 10 years as a results of the Soil Strategy and Soil Monitoring Law proposed by the EU Commission. Both the collection and harmonisation of soil datasets and the preparation of soil maps have intensified. Further to these improvements, the derivation of prediction algorithms, which can compute specific soil properties from easily available soil or other

environmental variables (the pedotransfer functions (PTFs)) has continued to be refined since the 1980s. The growing amount of spatiotemporal environmental data opens up possibilities for different prediction approaches, which is reflected in the terminology of the transfer functions, e.g. i) the classical PTFs mostly use only soil properties as input (Bouma, 1989), ii) those PTFs that consider not only soil properties but other environmental variables as well, are called covariate-based geo transfer functions (Gupta et al., 2021a), iii) spectral transfer functions predict non easily available soil properties from spectral data (Babaeian et al., 2015), while machine-learning-based (ML-based) soil mapping fuses prediction algorithms with geostatistical methods (Romano et al., 2023). All these improvements resulted in the emergent availability of soil maps at global, regional, and local scales.

~~Most of t~~The basic soil properties, ~~e.g~~i.e., soil organic carbon content, particle size distribution, ~~etc.,~~ in most cases are locally available at high resolution (< 100 m), but information on bulk density, albedo, soil erodibility factor, soil hydraulic properties, and soil nutrient content ~~soil hydraulic properties~~ is often lacking. There are many PTFs available in the literature that can be used to calculate soil physical (Abbaspour et al., 2019) and hydrological~~physical parameters~~ (Bouma and van Lanen, 1987; Van Looy et al., 2017) parameters from basic soil properties, but determining the most suitable one might not be obvious. Parameter estimations derive the parameters of a model that describes either water retention, hydraulic conductivity, or both across the entire matric potential range. These estimations aim to ensure a cohesive physical relationship between the computed soil hydraulic properties.

Information on soil nutrient properties often essential for environmental modelling, such as plant-available soil phosphorus or soil nitrate content, is seldom accessible at a catchment or regional scale. In the absence of measured data on nutrient content, estimating highly mobile nutrients like nitrate poses a challenge due to seasonal fluctuations influenced by factors such as fertilizer application, rainfall, plant nutrient uptake, and microbial activity. Regarding plant-available phosphorus, its levels typically exhibit minimal variation throughout a year. Therefore, approximating its quantity could rely on land use type and area-specific phosphorus fertilization loads (Ballabio et al., 2019). Nevertheless, multiple methods are employed across Europe to measure plant-available soil phosphorus content, potentially requiring conversions between these methods for broader-scale applications. A comprehensive review on conversion equations is available specifically for European studies in Steinfurth et al.(2021).

Often those soil properties are required as model input data as well, which are rarely available. One~~Another~~ example is the data on soil cracking, which is rarely readily available. Cracking intensity and number of cracks are determined by i) soil mineralogy, specifically the amount and type of clay minerals, ii) type of strength that forms soil structure (Lal and Shukla, 2004) and iii) human activity, e.g. tillage, plant spacing. The aperture and closure of cracks can be dynamically related to soil water content (Xing et al., 2023). The data that could describe the variability of cracking is also not easily available, therefore prediction of this parameter is limited at catchment scale.

~~Information on soil nutrient properties often essential for environmental modelling, such as plant-available soil phosphorus or soil nitrate content, is seldom accessible at a catchment or regional scale. In the absence of measured data on nutrient content, estimating highly mobile nutrients like nitrate poses a challenge due to seasonal fluctuations influenced by factors such as~~

fertilizer application, rainfall, plant nutrient uptake, and microbial activity. Regarding plant-available phosphorus, its levels typically exhibit minimal variation throughout a year. Therefore, approximating its quantity could rely on land use type and area-specific phosphorus fertilization loads (Ballabio et al., 2019). Nevertheless, multiple methods are employed across Europe to measure plant-available soil phosphorus content, potentially requiring conversions between these methods for broader-scale applications. A comprehensive review on conversion equations is available specifically for European studies in Steinfurth et al.(2021).

Dai et al. (2019b) provides an extensive review on global soil property maps applicable for Earth system models. Abbaspour et al. (2019) collected both soil datasets and pedotransfer functions for global Soil and Water Assessment Tool (SWAT) applications. From these global comprehensive review studies and a variety of soil datasets available among others from the

Table 1. Open access soil data that could be applied for environmental modelling in Europe.

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
Soil basic data							
LUCAS Topsoil dataset (Orgiazzi et al., 2018; Tóth et al., 2013)	topsoil point data	-	0-20	coarse, Sa, Si, Cl, pH_H2O, pH_CaCl2, OC, CaCO3, P, N, K_ext, CEC, MSP, HM	member states of the European Union and some other countries	https://esdac.jrc.ec.europa.eu/projects/lucas	from 2018 BD, soil biodiversity indicators, visual assessment of soil erosion and depth of organic soil are available, too (Fernandez-Ugalde et al., 2022)
SoilGrids (Poggio et al., 2021)	map	250 m	0-5, 5-15, 15-30, 30-60,	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O,	global	https://soilgrids.org/	-

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
			60-100, 100-200	CEC, N, with information on uncertainty.			
OpenLandMap	map	250 m	0, 10, 30, 60, 100, 200	ST, Sa, Si, Cl, OC, BD, pH_H2O, other	global	https://openlandmap.org	-
SPADE 2 (Hannam et al., 2009)	soil profile point data	-	actual soil depth	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O, pH_CaCl2, pH_KCl	member states of the European Union	https://esdac.jrc.ec.europa.eu/content/soil-profile-analytical-database-2	includes information on water regime class and water management type
Soil hydraulic or physical data							
EU-SoilHydroGrids (Tóth et al., 2017)	map	250 m	0, 5, 15, 30, 60, 100, 200	THS, FC, WP, KS, VG, MVG	Europe	https://elkh-taki.hu/en/eu_soilhydrogrids_3d , https://esdac.jrc.ec.europa.eu/content/3d-soil-hydraulic-database-europe-1-km-and-250-m-resolution	can be used with soil basic data of the SoilGrids 2017 (Hengl et al., 2017) dataset

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
Panagos et al. (2014)	map	500 m	-	K_fact	Europe	https://esdac.jrc.ec.europa.eu/content/soil-erodibility-k-factor-high-resolution-dataset-europe	can be used with the LUCAS Topsoil dataset
Montzka et al. (2017)	map	0.25°	0, 5, 15, 30, 60, 100, 200	VG, KS and scaling parameters based on 1 km SoilGrids 2017	global	https://doi.org/10.5194/essd-9-529-2017	can be used with soil basic data of the SoilGrids 2017 (Hengl et al., 2017) dataset
Zhang and Schaap (2018)	map	1 km	0, 5, 15, 30, 60, 100, 200	KO, FC, AWC with standard deviations	global	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UI5LCE	can be used with soil basic data of the SoilGrids 2017 (Hengl et al., 2017) dataset
Zhang et al. (2020)	map	10 km	0, 10, 30, 60, 100, 200	THS, FC, WP with coefficient of variation	global	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VPIN2B	can be used with the soil basic data of the OpenLandMap dataset

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
Gupta et al. (2022)	map	1 km	0, 30, 60, 100	VG	global	https://doi.org/10.5281/zenodo.6348799	environmental covariates further to soil properties were considered as input for the mapping
Gupta et al. (2021b)	map	1 km	0, 30, 60, 100	KS	global	https://doi.org/10.5281/zenodo.3935359	environmental covariates further to soil properties were considered as input for the mapping
Surface albedo, MCD43A3	map	500 m	-	ALB	global	https://doi.org/10.5067/MODIS/MCD43A3.061	16-daily data
Surface albedo (Copernicus Climate Change Service, 2018)	map	300 m, 1 km, 4 km	-	ALB	global	https://doi.org/10.24381/cds.ea87ed30	10-daily data
HYSOGs250m (Ross et al., 2018)	map	250 m	0, 5, 15, 30, 60, 100	HSG	global	https://doi.org/10.3334/ORNLDAAC/1566	can be used with soil basic data of the SoilGrids 2017 (Hengl et al., 2017) dataset

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
Soil basic and hydraulic data							
WOSIS (Batjes et al., 2020)	soil profile point data	-	actual soil depth	coarse, Sa, Si, Cl, BD, OC, pH_H2O, CaCO3, CEC, TOC, P, FC, WP, other	global	https://soilgrids.org/	-
HWSD v 2.0 (FAO & IIASA, 2023)	map	1 km	0-20, 20-40, 40-60, 60-80, 80-100, 100-150, 150-200	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O, CaCO3, CEC, EC, D_C, AWC_C, IL, RSD, SWR, other	global	https://gaez.fao.org/pages/hwsd	-
GSDE (2014) (Shangguan et al., 2014)	map	30'' (~1km)	0-4.5 4.5-9.1 9.1-16.6 16.6-28.9 28.9-49.3 49.3-82.9	ST, coarse, Sa, Si, Cl, BD, OC, TOC, pH_H2O, CaCO3, CEC, EC, N, P	global	http://globalchange.bnu.edu.cn/research/soilw	dataset is based on the Soil Map of the World and various regional and national soil databases

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes
			82.9-138.3 138.3-229.6	D_C, AWC_C, IL, RSD, SWR, FC, WP, other			
Dai et al. (2019a)	map	30″ (~1km)	0-5 5-15 15-30 30-60 60-100 100-200 and layering of four land surface models(Dai et al., 2019a)	coarse, OM, Sa, Si, Cl, Qa CB, MVG SHC, STC	global	http://globalchange.bnu.edu.cn/research/soil5.jsp	can be used with GSDE and SoilGrids 2017

*ST: soil type; coarse: coarse fragments; Sa, Si, Cl: sand, silt and clay content; Qa: quartz content of the mineral soil; BD: bulk density; OM: organic matter content; OC: organic carbon content; TOC: total organic carbon content; pH_H2O: pH in water; pH_CaCl2: pH in calcium chloride; pH_KCl: pH in potassium chloride; CaCO3: calcium carbonate content; CEC: cation exchange capacity; EC: electrical conductivity; N: total nitrogen content; K_ext: extractable potassium content; P: phosphorus content extracted by Olsen method; MSP: multispectral properties; HM: heavy metals; K_fact: soil erodibility (K-factor) based on the Revised Universal Soil Loss Equation; ALB: surface albedo; HSG: hydrological soil groups based on the National Engineering Handbook(U.S. Department of Agriculture Natural Resources Conservation Service, 2009) of the U.S. Department of Agriculture Natural Resources Conservation Service; D_C: drainage class, AWC_C: available water capacity class in the rootable depth; IL: depth class to impermeable layer; RSD: rootable soil depth class; SWR: soil water regime class; THS: saturated water content; FC: water content at field capacity; WP: water content at wilting point; KS: saturated hydraulic conductivity; VG: parameters of the van Genuchten model (θ_s, θ_r, α, n, m) to describe the soil water retention curve; MVG:

parameters of the Mualem-van Genuchten model to describe the soil water retention and hydraulic conductivity curve (θ_s, θ_r, α, n, m, K_θ, L); CB: parameters of the Campbell model (θ_s, ψ, λ, K_s) to describe the soil water retention and hydraulic conductivity curve; KO parameters of the Kosugi model to describe the soil water retention and hydraulic conductivity curve (θ_s, θ_r, h_m, σ, K_s); SHC: heat capacity of soil solids; STC: soil thermal conductivity of unfrozen saturated soil, frozen saturated soil and dry soil; other: other soil properties, which are less frequently required by environmental models.

Table 1. Open access soil data that could be applied for environmental modelling in Europe.

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes	References
Soil basic data								
LUCAS Topsoil dataset	topsoil point data	-	0-20	coarse, Sa, Si, Cl, pH_H2O, pH_CaCl2, OC, CaCO3, P, N, K_ext, CEC, MSP, HM	European Union and some other countries	https://esdac.jrc.ec.europa.eu/projects/lucas	from 2018 BD, soil biodiversity indicators, visual assessment of soil erosion and depth of organic soil are available, too (Fernandez-Ugalde et al., 2022)	Orgiazzi et al. (2018), Tóth et al. (2013)
SPADE 2	soil profile point data	-	actual soil depth	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O, pH_CaCl2, pH_KCl	European Union	https://esdac.jrc.ec.europa.eu/content/soil-profile-analytical-database-2	includes information on water regime class and water management type	Hannam et al. (2009)
SoilGrids	map	250 m	0-5, 5-15, 15-30, 30-60, 60-100, 100-200	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O, CEC, N, with information on uncertainty.	global	https://soilgrids.org/	-	Poggio et al. (2021)
OpenLand-Map	map	250 m	0, 10, 30, 60, 100, 200	ST, Sa, Si, Cl, OC, BD, pH_H2O, other	global	https://openlandmap.org	-	https://openlandmap.org
Topsoil chemical properties for Europe	map	500 m	0-20	pH_H2O, pH_CaCl2, CEC, CaCO3, C:N ratio, P, N, K_ext	European Union	https://esdac.jrc.ec.europa.eu/content/chemical-properties-european-scale-based-lucas-topsoil-data	can be used with topsoil physical properties for Europe dataset (Ballabio et al., 2016)	Ballabio et al. (2019)

Table 1. cont.

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes	References
Soil hydraulic or physical data								
EU-SoilHydroGrids	map	250 m	0, 5, 15, 30, 60, 100, 200	THS, FC, WP, KS, VG, MVG	Europe	https://elkh-taki.hu/en/eu_soilhydrogrids_3d , https://esdac.jrc.ec.europa.eu/content/3d-soil-hydraulic-database-europe-1-km-and-250-m-resolution	can be used with the SoilGrids 2017 (Hengl et al., 2017) dataset	Tóth et al. (2017)
Montzka et al. (2017)	map	0.25°	0, 5, 15, 30, 60, 100, 200	VG, KS and scaling parameters based on 1 km SoilGrids 2017	global	https://doi.org/10.5194/essd-9-529-2017	can be used with the SoilGrids 2017 (Hengl et al., 2017) dataset	Montzka et al. (2017)
Zhang and Schaap (2018)	map	1 km	0, 5, 15, 30, 60, 100, 200	KO, FC, AWC with standard deviations	global	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/UI5LCE	can be used with the SoilGrids 2017 (Hengl et al., 2017) dataset	Zhang and Schaap (2018)
Zhang et al. (2020)	map	10 km	0, 5, 15, 30, 60, 100, 200	THS, FC, WP with coefficient of variation	global	https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VPIN2B	can be used with the soil basic data of the OpenLandMap dataset	Zhang et al. (2020)
Gupta et al. (2022)	map	1 km	0, 30, 60, 100	VG	global	https://doi.org/10.5281/zenodo.6348799	-	Gupta et al. (2022)
Gupta et al. (2021b)	map	1 km	0, 30, 60, 100	KS	global	https://doi.org/10.5281/zenodo.3935359	-	Gupta et al. (2021b)
Topsoil physical properties for Europe	map	500 m	0-20	coarse, Sa, Si, Cl, BD, TEX, AWC	European Union	https://esdac.jrc.ec.europa.eu/content/topsoil-physical-properties-europe-based-lucas-topsoil-data	can be used with topsoil chemical properties for Europe dataset (Ballabio et al., 2019)	Ballabio et al. (2016)

Table 1. cont.

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes	References
Soil bulk density in Europe	map	100 m	0-20, 0-10, 10-20	BD	European Union	https://esdac.jrc.ec.europa.eu/content/soil-bulk-density-europe	-	Panagos et al. (2024)
Panagos et al. (2014)	map	500 m	-	K_fact	Europe	https://esdac.jrc.ec.europa.eu/content/soil-erodibility-k-factor-high-resolution-dataset-europe	can be used with the LUCAS Topsoil dataset	Panagos et al. (2014)
Global soil erodibility	map	1 km	-	K_fact	global	https://esdac.jrc.ec.europa.eu/content/global-soil-erodibility	can be used with the SoilGrids 2021 (Poggio et al., 2021) dataset	Gupta et al. (2024)
Surface albedo, MCD43A3	map	500 m	-	ALB	global	https://doi.org/10.5067/MODIS/MCD43A3.061	16-daily data	https://doi.org/10.5067/MODIS/MCD43A3.061
Surface albedo (Copernicus Climate Change	map	300 m,	-	ALB	global	https://doi.org/10.24381/cds.ea87ed30	10-daily data	Copernicus Climate Change Service (2018)
HYSOGs250m	map	250 m	0, 5, 15, 30, 60, 100	HSG	global	https://doi.org/10.3334/ORNLDAAC/1566	can be used with the SoilGrids 2017 (Hengl et al., 2017) dataset	Ross et al. (2018)
Soil basic and hydraulic data								
WOSIS	soil profile point data	-	actual soil depth	coarse, Sa, Si, Cl, BD, OC, pH_H2O, CaCO3, CEC, TOC, P, FC, WP, other	global	https://soilgrids.org/	-	Batjes et al. (2020)
HWSD v 2.0	map	1 km	0-20, 20-40, 40-60, 60-80, 80-100, 100-150, 150-200	ST, coarse, Sa, Si, Cl, BD, OC, pH_H2O, CaCO3, CEC, EC, D_C, AWC_C, IL, RSD, SWR, other	global	https://gaez.fao.org/pages/hwsd	-	FAO & IIASA (2023)

Table 1. cont.

Name of dataset	Data type	Horizontal resolution	Vertical resolution (cm)	Available soil properties*	Coverage	Availability	Notes	References
DSOLMap	map	250 m	0-5, 5-10, 10-30, 30-60, 60-100, 100-200	coarse, Sa, Si, Cl, BD, OC, AWC, KS, K_fact, ALB_m	global	https://www.wateritech.com/data	data is based on OpenLandMap, and is in format of the SWAT model	(López-Ballesteros et al., (2023)
GSDE (2014)	map	30″ (~1km)	0-4.5, 4.5-9.1, 9.1-16.6, 16.6-28.9, 28.9-49.3, 49.3-82.9, 82.9-138.3, 138.3- 229.6	ST, coarse, Sa, Si, Cl, BD, OC, TOC, pH_H2O, CaCO3, CEC, EC, N, P, D_C, AWC_C, IL, RSD, SWR, other	global	http://globalchange.bnu.edu.cn/research/soilw	dataset is based on the Soil Map of the World and various regional and national soil databases	Shangguan et al. (2014)
Dai et al. (2019)	map	30″ (~1km)	0-5, 5-15, 15-30, 30-60, 60-100, 100-200 and specific layering	coarse, OM, Sa, Si, Cl, Qa, CB, MVG, SHC, STC	global	http://globalchange.bnu.edu.cn/research/soil5.jsp	can be used with GSDE and SoilGrids 2017	Dai et al. (2019a)

*ST: soil type; coarse: coarse fragments; Sa, Si, Cl: sand, silt and clay content; Qa: quartz content of the mineral soil; BD: bulk density; OM: organic matter content; OC: organic carbon content; TOC: total organic carbon content; pH_H2O: pH in water; pH_CaCl2: pH in calcium chloride; pH_KCl: pH in potassium chloride; CaCO3: calcium carbonate content; CEC: cation exchange capacity; EC: electrical conductivity; N: total nitrogen content; K_ext: extractable potassium content; P: phosphorus content extracted by Olsen method; MSP: multispectral properties; HM: heavy metals; K_fact: soil erodibility (K-factor) based on the Revised Universal Soil Loss Equation; ALB: surface albedo; ALB_m: moist soil albedo; HSG: hydrological soil groups based on the National Engineering Handbook(U.S. Department of Agriculture Natural Resources Conservation Service, 2009) of the U.S. Department of Agriculture-Natural Resources Conservation Service; D_C: drainage class, AWC_C: available water capacity class in the rootable depth; IL: depth class to impermeable layer; RSD: rootable soil depth class; SWR: soil water regime class; THS: saturated water content; FC: water content at field capacity; WP: water content at wilting point; KS: saturated hydraulic conductivity; VG: parameters of the van Genuchten model (θs, θr , α, n, m) to describe the soil water retention curve; MVG: parameters of the Mualem-van Genuchten model to describe the soil water retention and hydraulic conductivity curve (θs, θr , α, n, m, K0, L); CB: parameters of the Campbell model (θs, ψ, λ, KS) to describe the soil water retention and hydraulic conductivity curve; KO parameters of the Kosugi model to describe the soil water retention and hydraulic conductivity curve (θs, θr , hm, σ, KS); SHC: heat capacity of soil solids; STC: soil thermal conductivity of unfrozen saturated soil, frozen saturated soil and dry soil; other: other soil properties, which are less frequently required by environmental models, AWC: available water capacity; TEX: USDA soil texture class.

European Soil Data Centre (Panagos et al., 2022) (https://esdac.jrc.ec.europa.eu/) or ISRIC – World Soil Information (https://www.isric.org/), it is not straightforward which data and/or pedotransfer functions could be used for the environmental modelling in European case studies. Therefore, in this study we support soil data retrieval for environmental modelling across Europe by i) systemizeing information on open access datasets and PTFs applicable for Europe, ii) demonstrateing and quantifying the difference between some PTFs and prediction approaches to cover missing soil properties based on the point data of EU-HYDI, and iii) provideing a comprehensive workflow and accompanying open-source R script and library for the derivation of missing soil data. For the selection of the prediction approaches, three requirements had to be fulfilled: 1) the prediction algorithm had to be trained on temperate soils and should not be specific to a particular soil reference group, 2) the required predictors had to be available in most of the open access soil datasets, and 3) its ease of application. Hence, despite certain published prediction methods specifying soil depth, texture, and organic matter as requirements, those reliant on, for instance, artificial neural networks, lacking a user-friendly interface, or integration into accessible tools like R packages or Python modules, were excluded from testing due to their challenging application. For ease of reference, all the equations needed to calculate the most often required soil input parameters are given below.

2 Materials and methods

We distinguish and list soil physical and chemical parameters similarly to the terminology used by the Soil and Water Assessment Tool model documentation (Neitsch et al., 2009). We include the prediction of soil porosity since this parameter is frequently used in environmental models, e.g. MIKE SHE (DHI, 2023), HEC RAS (US Army Crops of Engineers, 2023), PIHM (Li and Duffy, 2011). Noteworthy that some models and accompanying model setup tools have an internal built-in PTF to compute porosity, e.g. SWAT+. The codes to compute the soil parameters were built based on the structure and terminology used by the SWAT+ *usersoil* table (Arnold et al., 2012). Soil properties most frequently required by the environmental models – e.g. (Abbaspour et al., 2019; Dam et al., 2008; Dang et al., 2022; DHI, 2023; Hansen et al., 2012; Šimůnek et al., 2012; Yu et al., 2020) – are:

- soil layering,
- maximum rooting depth,
- ~~information on soil cracking,~~
- effective bulk density,
- field capacity,
- wilting point,
- available water capacity,
- porosity,
- saturated hydraulic conductivity,
- organic carbon content,

- sand, silt, and clay content,
- rock fragment content,
- moist soil albedo,
- Universal Soil Loss Equation (USLE) soil erodibility factor,
- hydrologic soil group, and
- nutrient content of the surface soil layer.

We summarised ~~the~~ information about ~~possible~~potential open access sources for soil information applicable in Europe in Table 1, ~~to~~ covering most of the above listed soil properties. The availability of datasets is continuously improving. The following data sites include most of the updates:

- European Soil Data Centre, which includes soil datasets from Europe and information on EU Soil Observatory (https://esdac.jrc.ec.europa.eu/),
- ISRIC Soil Data Hub, which hosts soil data from around the world (https://data.isric.org/geonetwork/srv/eng/catalog.search#/home),
- soil related layers of the GAEZ Data Portal developed by the Food and Agriculture Organization of the United Nations (FAO) and the International Institute for Applied Systems Analysis (IIASA) (https://data.apps.fao.org),
- soil related layers of the OpenLandMap, which shares open geographical and geoscientific data (https://openlandmap.org).

Nevertheless, these sources do not include products from specific institutes, such as http://globalchange.bnu.edu.cn/research. The~~se~~ datasets included in Table 1 might be appropriate for regional and continental modelling. However, for catchment scale and national studies, local and national spatially explicit modelled datasets provide more accurate input information. When a certain local dataset is selected to be used as basic soil information, it is more consistent to compute the ~~non-easily available~~missing soil properties from this local data source rather than~~, instead of~~ using other data sources ~~of data~~. This allows to maintain ~~the correlation~~consistency between the different soil properties. For example, ~~we do~~it is not recommended to combin~~inge~~ ~~100 m resolution~~ local soil ~~basic~~ property maps at 100 m resolution with soil hydraulic properties ~~derived~~ retrieved from ~~250 m resolution~~ EU-SoilHydroGrids at 250 m resolution~~which was computed based on SoilGrids 2017 (Hengl et al., 2017)~~. ~~If~~ Where local soil maps with soil layering, organic carbon content, clay, silt, and sand content are~~is~~ available, it is suggested that ~~the~~ missing soil properties, ~~e.g.,~~such as bulk density, soil hydraulic properties, and albedo~~, etc.,~~ are estimated from ~~those~~ the locally available basic soil properties to ensure consistency. The predictions are subject to uncertainty, which depends on the similarity between the training data used for the selected prediction method and the target area in terms of soil physical and chemical characteristics (Román Dobarco et al., 2019; Tranter et al., 2009). ~~Retrieving or computing information on soil cracking is more complex than the other soil properties listed. Up to now, no approach is available to predict soil cracking volume at a catchment or regional scale, this parameter is defined for the environmental models based on local expert knowledge.~~

~~When using any available soil dataset, it is important to check the detailed description (metadata) of the dataset to avoid misinterpretation and errors in the models. Considerations such as consistent size limits for clay, silt, and sand content classification as per the model's requirements, distinction between organic carbon and organic matter, the need for dry or moist bulk density, and similar details are vital. Understanding whether the model derives soil properties that already exist in the dataset is essential, aiding in selecting the most precise parameters for the model's application.~~

~~In this paper we describe possible ways to fill data gaps with prediction algorithms that can be applied on locally available or open access soil basic data. For the selection of the prediction approaches, three requirements had to be fulfilled: i) the prediction algorithm had to be trained on temperate soils, ii) the required predictors had to be available in most of the open access soil datasets, and iii) its ease of application. Hence, despite certain published prediction methods specifying soil depth, texture, and organic matter as requirements, those reliant on, for instance, artificial neural networks, lacking a user-friendly interface, or integration into accessible tools like R packages or Python modules, were excluded from testing due to their challenging application. For ease of reference, all the equations needed to calculate the most often required soil input parameters are given below.~~

2.1 Evaluation of methods

For soil physical and hydrological properties, the performance of the prediction algorithms was assessed using the European Hydropedological Data Inventory (EU-HYDI), specifically focusing on soil parameters with available measured values in the dataset. The EU-HYDI is one of the most comprehensive European soil hydraulic datasets, which has soil data of 18,682 samples from 6,014 profiles (Weynants et al., 2013). The number of measured values varies by soil properties. EU-HYDI dataset was used to derive hydraulic PTFs, called euptfs. When comparing the performance of euptf with other methods found in the literature, only the test sets from the EU-HYDI dataset, which were not utilized in euptf's training, were included. This approach aimed to facilitate a more accurate and fair comparison among different PTFs, but decreased the number of samples used for the analysis. The analysis of bulk density prediction was performed on both the EU-HYDI and the LUCAS Topsoil dataset (Orgiazzi et al., 2018; Tóth et al., 2013) of 2018. The LUCAS Topsoil dataset of 2009 ~~(Tóth et al., 2013)~~ was used for the computation of nutrient content of the surface soil layer. For the assessment of the topsoil phosphorus maps, we used locally measured data obtained from an agricultural company. This dataset includes soil phosphorus content measured at a depth of 30 cm using the acid ammonium acetate lactate extraction (AL-P) method (Egnér et al., 1960) for 34 agricultural parcels in the year 2009. As the phosphorus content was required according to the Olsen method (Olsen-P) (Olsen et al., 1954), we applied the equation of Sárdi et al. (2009) for converting AL-P into Olsen-P. Table 2 shows the descriptive statistics of this database.

Table 2. Descriptive statistics of the locally measured phosphorus content, converted to Olsen-P, from 34 agricultural parcels.

Min	Max	Range	Mean	Median	Standard deviation

8.39	65.02	56.63	27.54	25.73	13.47

We compared the algorithms using the mean error (ME), mean absolute error (MAE), root mean squared error (RMSE), Nash-Sutcliffe efficiency (NSE), and coefficient of determination (R^2). The non-parametric Kruskal-Wallis test of the R package agricolae (De Mendiburu, 2017) at the 5% significance level was applied on the squared error values to asses if there were significant difference in performance. For soil parameters without measured data in the EU-HYDI dataset, descriptive statistics and histograms of the computed parameters were compared with studies from peer-reviewed literature focused on European applications. The statistical analysis was performed using R statistics library (R Core Team, 2022).

2.2 Analysed soil properties

We analysed soil physical, hydraulic, and chemical parameters. Under soil physical parameters, we addressed bulk density, porosity, albedo, and soil erodibility factor. For soil hydraulic parameters, we examined water retention, saturated hydraulic conductivity and hydrological soil groups. Regarding soil nutrient content, we focused on topsoil phosphorus content and described the challenges of retrieving soil nitrate content. Hereinafter information about the analysis by soil properties is provided.

2.2.1 Soil physical parameters

~~**2.2.1 Basic properties**~~

~~Initially, the data source of the most relevant soil basic properties, such as soil layering rooting depth, organic carbon content, clay, silt, and sand content, must be selected. Local data can describe the spatial variability of soil properties the best. Even if only soil basic properties are available locally, this data source could have priority against the more inclusive continental or global datasets, i.e. containing information on both soil physical, chemical, and hydraulic properties, because local datasets aim to capture the area-specific variability of soil properties as accurately as possible. If no local or national soil basic data is available with the resolution required to study a target environmental process, possible input source for soil profile data or 3D soil dataset can be found in Table 1.~~

~~Different countries and institutions measure sand, silt, and clay content using different methods by recognizing different cutoff limits and classification standards. It is important to check which particle size limits are required by the environmental model. As an example, in the widely used SWAT/SWAT+ model, the sand, silt and clay content are assumed to be classified according to the USDA system, which defines particle size limit < 0.002 mm for clay, 0.002-0.05 mm for silt and 0.05-2 mm for sand fraction. When conversions between different classifications are required to bring the local dataset to the appropriate format, it is advised to apply the k-nearest neighbour interpolation (formerly called: 'similarity technique'), which results in less uncertainty, smaller bias and shrinkage of resulting texture range compared to the simpler loglinear interpolation (Nemes et al., 1999).~~

~~In other cases, such as soil organic material, it is important to distinguish if soil organic carbon or soil organic matter is required by the model, and which of the two is available from the data source. The following most frequently used equation describes the relationship between those:~~

~~$OM = OC \cdot 1.724$~~ ~~(1)~~

~~where OM is the organic matter content (mass %) and OC is organic carbon content (mass %). The 1.724 conversion factor was defined by Van Bemmelen (1890), but can vary between 1.4 and 2.5 depending on the method used to measure organic carbon content, composition of organic matter, degree of decomposition and clay content (Minasny et al., 2020; Pribyl, 2010). Pribyl (2010) recommends using the value 2 as a general conversion factor if no specific value is available.~~

~~2.2.2~~ Bulk density

formázott: Címsor 4

~~When specifying bulk density, it is important to clarify whether the dry or effective value is required. If a measured value of neither is available, the dry bulk density can be computed from organic carbon content and particle size distribution. Further predictors, such as taxonomical information, soil structure, soil management parameters, environmental covariates are important as well (Hollis et al., 2012; Ramcharan et al., 2017) and can significantly improve the prediction performance. However, PTFs including these variables are not always possible to apply to a data scarce region. We selected the bulk density PTFs – derived on soils of the temperate region – based on previous works (Casanova et al., 2016; Hossain et al., 2015; Palladino et al., 2022; Xiangsheng et al., 2016) that tested the prediction performance of several methods.~~ Table ~~3~~2 lists the PTFs that were tested on point data in EU-HYDI and 2018 LUCAS Topsoil dataset. We selected the bulk density PTFs – derived on soils of the temperate region – based on previous works (Casanova et al., 2016; Hossain et al., 2015; Palladino et al., 2022; Xiangsheng et al., 2016) that tested the prediction performance of several methods.

Table ~~3~~2. List of pedotransfer functions tested on point data in EU-HYDI for the prediction of bulk density.

Name of the PTF	Equation	Reference	Eq.
BD_Rawls	$BD = \frac{100}{\left(\left(\frac{OM}{0.224}\right) + \frac{100 - OM}{1.27}\right)}$	(Rawls, 1983)	(1~~2~~)
BD_Alexander_A	$BD = 1.72 - 0.294 \cdot OC^{0.5}$	(Alexander, 1980)	(2~~3~~)
BD_Alexander_B	$BD = 1.66 - 0.308 \cdot OC^{0.5}$	(Alexander, 1980)	(3~~4~~)
BD_MAn_J_A	$BD = 1.510 - 0.113 \cdot OC$	(Manrique and Jones, 1991)	(4~~5~~)
BD_MAn_J_B	$BD = 1.66 - 0.318 \cdot OC^{0.5}$	(Manrique and Jones, 1991)	(5~~6~~)
BD_Hollis	-for cultivated topsoils: $BD = 0.80806 + (0.8238444 \cdot (\exp(-0.279938 \cdot OC))) + 0.0014065 \cdot sand - 0.0010299 \cdot clay$ - for mineral subsoils: $BD = 0.69794 + (0.750636 \cdot (\exp(-0.230355 \cdot OC))) + 0.0008687 \cdot sand - 0.0005164 \cdot clay$ - for organic horizons*:	(Hollis et al., 2012)	(6~~7~~)

	$BD = 1.4903 + 0.33293 \cdot \log(OC)$		
BD_Bernoux	$BD = 1.398 - 0.042 \cdot OC - 0.0047 \cdot clay$	(Bernoux et al., 1998)	(~~7~~8)
BD_Hossain **	$BD = 0.074 + 2.632 \cdot \exp(-0.076 \cdot OC)$	(Hossain et al., 2015)	(~~8~~9)

* For histic and follic horizons, which have organic carbon content equal to or greater than 20 % (IUSS Working Group WRB, 2022). **Applied only for organic soils with organic carbon content equal to or greater than 12 %. OM: organic matter content (mass %); OC: organic carbon content (mass %); sand: sand content (0.05-2 mm fraction) (mass %); clay: clay content (<0.002 mm fraction) (mass %).

~~If effective bulk density is required, it can be derived from the dry bulk density with the method of Wessolek et al. (2009):~~

~~= for soils with organic carbon content higher than 0.58 mass %:~~

~~$= +0.009 \cdot clay$ (10)~~

~~= for soils with organic carbon content less than or equal to 0.58 mass %:~~

~~$= +0.005 \cdot clay + 0.001 \cdot silt$ (11)~~

~~where BD_{eff} (g cm^{-3}) is effective bulk density, BD_{dry} (g cm^{-3}) is the dry bulk density, clay is clay content (< 0.002 mm, mass %), silt is silt content (0.002-0.063 mm, mass %). It is important to note that Eq. (11) requires the silt content with 0.002-0.063 mm limit. It can be predicted from the clay (< 0.002 mm), silt (0.002 – 0.05 mm) and sand (0.05 – 2 mm) content with the TT.text.trans function of the soiltexture R package (Moeys, 2018). This method meets the accuracy required for computing BD_{eff}, however, for other applications the transformation methods discussed by Nemes et. al. (1999) should be considered.~~

~~2.2.3~~ Porosity

~~Porosity is seldom provided in soil maps and is not commonly included as a directly measured value in soil profile datasets.~~

Porosity can be computed based on the bulk density and particle density with the following equation:

$$POR = \left(1 - \left(\frac{BD}{PD}\right)\right) \cdot 100 \quad (9\text{~~12~~})$$

where POR is porosity (volume %), BD is dry bulk density (g cm^{-3}), PD is particle density (g cm^{-3}).

As seen in literature and in SWAT+ model default assumptions (Neitsch et al., 2009), the particle density is usually set as equal to 2.65 g cm^{-3} (Lal and Shukla, 2004). However, there are PTFs that calculate the porosity based on the particle size distribution (sand, silt, clay content) and organic matter content. We selected the PTFs (Table 4~~3~~) based on the findings of Ruehlmann (2020) and analysed their prediction performance on the EU-HYDI dataset.

Table 4~~3~~. List of pedotransfer functions tested on point data in EU-HYDI for the prediction of porosity.

Name of the PTF	Equation	Reference	Eq.
POR_Schjonning_etal	$PD_{OM} = 1.241 + 0.173 \cdot \left(\frac{OM}{100}\right)$ $PD_{SMS} = 2.663 + 0.107 \cdot \left(\frac{clay}{100}\right)$	(Schjønning et al., 2017)	(10~~3~~)

formázott: Címsor 4

	$PD = \left(\frac{\left(1 - \frac{OM}{100}\right)}{PD_{SMS}} + \frac{\frac{OM}{100}}{PD_{OM}} \right)^{-1}$ $POR = \left(1 - \left(\frac{BD}{PD}\right)\right) \cdot 100$		
POR_Schjonning_etal_recal	$PD = 2.654 + 0.216 \cdot \frac{clay}{100} - 2.237 \cdot \frac{OM}{100}$ $POR = \left(1 - \left(\frac{BD}{PD}\right)\right) \cdot 100$	(Ruehlmann, 2020)	(114)
POR_2_65	$POR = \left(1 - \left(\frac{BD}{2.65}\right)\right) \cdot 100$	(Lal and Shukla, 2004)	(152)

PD_{OM}: particle density of the soil mineral substance; PD_{MS}: particle density of the soil organic matter; OM: organic matter content (mass %); sand: sand content (0.05-2 mm fraction) (mass %); clay: clay content (<0.002 mm fraction) (mass %).

~~As seen in literature and in SWAT+ model default assumptions (Neitsch et al., 2009), the particle density is usually set as equal to 2.65 g cm^{-3} (Lal and Shukla, 2004). However, there are PTFs that calculate the porosity based on the particle size distribution (sand, silt, clay content) and organic matter content. We selected the PTFs (Table 3) based on the findings of Ruehlmann (2020) and analysed their prediction performance on the EU-HYDI dataset.~~

~~2.2.4~~ Albedo

formázott: Címsor 4

Bare soil albedo mostly depends on soil moisture variations, surface roughness, soil texture and organic matter content (Carrer et al., 2014). Time series of soil surface albedo could be retrieved e.g. from the MCD43A3 database or Copernicus Climate Change Service (2018) (Table 1). If a single characteristic value of soil surface albedo is required for the entire modelling period, such as e.g. in the case of the SWAT+ model, the study of Abbaspour et al. (2019) provides several formulas for its computation and suggests to substitute the actual volumetric water content with field capacity. For European applications the equation of Gascoin et al. (2009) could be used:

$$ALB = 0.31 \cdot \exp(-12.7 \cdot \theta) + 0.15 \quad (163)$$

where ALB is soil albedo and θ is volumetric water content (cm^3 cm^{-3}), which could be set to the value of field capacity.

We computed the soil albedo with Eq. (163) for the EU-HYDI topsoil samples with setting the water content to saturation, field capacity and wilting point. The EU-HYDI dataset does not include the measured soil albedo values at a certain moisture content, therefore we extracted the median surface albedo for year 2022 from the MCD43A3 database (https://doi.org/10.5067/MODIS/MCD43A3.061) for two cases: i) any surfaces in the entire year and ii) only dry bare soils. We compared the descriptive statistics of computed and mapped values. We considered the visible broadband black-sky albedo for the analysis. Dry bare soil pixels were selected using MOD09GA.061 dataset based on Normalized Difference Vegetation Index (NDVI) and Normalized Burn Ratio 2 (NBR2) indices (Safanelli et al., 2020) in Google Earth Engine platform (Gorelick et al., 2017) when NDVI values fell in the range of -0.05 and 0.30, and NBR2 values between -0.15 and 0.15. Pixels for dry

bare soils were selected to mask and compare the remote sensed soil albedo values to the albedo computed at different moisture states.

~~2.2.5~~ Soil erodibility factor

The soil erodibility factor (K-factor) required for modelling soil erosion can be computed with several methods described e.g. in Kinnell (2010) or Panagos et al. (2014). The most widely used equation that can be readily applied to the most frequently available soil properties was published by Sharpley and Williams (1990) (Eq. 14) and Renard et al. (~~Renard et al.,~~ 1997) (Eq. 15).~~,~~ ~~t~~The advantage of ~~which~~ these methods is that ~~they~~it requires only the sand, silt, clay, and organic carbon content of the soil.~~:~~

$$K_{USLE} = \left(0.2 + 0.3 \cdot \exp\left(0.0256 \cdot \text{sand} \cdot \left(1 - \frac{\text{silt}}{100}\right)\right)\right) \cdot \left(\left(\frac{\text{silt}}{\text{clay} + \text{silt}}\right)^{0.3}\right) \cdot \left(1 - \left(\frac{0.25 \cdot \text{OC}}{(\text{OC} + \exp(3.72 - 2.95 \cdot \text{OC}))}\right)\right) \cdot \left(1 - \left(\frac{0.7 \cdot \left(1 - \frac{\text{sand}}{100}\right)}{\left(\left(1 - \frac{\text{sand}}{100}\right) + \exp\left(-5.51 + 22.9 \cdot \left(1 - \frac{\text{sand}}{100}\right)\right)\right)}\right)\right) \quad (1\text{~~7~~}4)$$

$$K_{RUSLE} = 7.594\left(0.0034 + 0.0405 \cdot exp\left(-0.5 \cdot \left(\frac{log(D_g)+1659}{0.7101}\right)^2\right)\right) \quad \text{with } D_g = \exp(0.01 \cdot \sum f_i \cdot \ln m_i) \quad (15)$$

where K_{USLE} is the Universal Soil Loss Equation (USLE), K_{RUSLE} is the Revised Universal Soil Loss Equation (RUSLE) soil erodibility factor $\left(\frac{t \cdot arce \cdot h}{hundreds\ of\ acre \cdot foot - tonf \cdot inch}\right)$, silt is silt content (mass%, 0.002-0.05 mm), sand is sand content (mass %, 0.05-2 mm), OC is organic carbon content (mass %), D_g is the geometric mean particle diameter (mm), f_i is the particle size fraction (mass%), m_i is the arithmetic mean of the particle size limits of the f_i particle size fraction (mm). If the unit is required in $\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$, the ~~$K_{USLE}$~~ value of the soil erodibility factor computed with Eq. (1~~7~~4) or Eq. (15) has to be multiplied with 0.1317 (Foster et al., 1981).

We computed the soil erodibility~~K_{USLE}~~ factor for the EU-HYDI dataset. Similarly to the above mentioned albedo, there is no measured soil erodibility ~~K_{USLE}~~ value in the EU-HYDI dataset, thus we compared the values computed for the topsoils of EU-HYDI with the values extracted from the European map of Panagos et al. (2014).

~~2.3~~ 2.2.2 Soil hydraulic parameters

~~2.3.1~~ Water retention and saturated hydraulic conductivity

~~It is recommended to compute the s~~Soil water retention and hydraulic conductivity can be computed from the parameters of the widely used van Genuchten model (VG) (van Genuchten, 1980):

$$\theta(\psi) = \theta_r + \frac{\theta_s - \theta_r}{[1 + (\alpha\psi^n)]^m} \quad \text{with} \quad m = 1 - 1/n \quad (1\text{~~8~~}6)$$

formázott: Címsor 4

formázott: Címsor 3

where θ_r (cm^3 cm^{-3}) and θ_s (cm^3 cm^{-3}) are the residual and saturated soil water contents, respectively, α (cm^{-1}) is a scale parameter, m (-) and n (-) are shape parameters.

Alternative models, like the Kosugi model (Kosugi, 1996) exist for characterizing the water retention curve. However, the availability of predictive tools for their parameters and equations to derive specific soil hydraulic properties (such as saturated hydraulic conductivity and field capacity based on internal drainage dynamics) from these parameters is either limited or non-existent (Zhang et al., 2018). Utilizing the VG model to compute all necessary soil hydraulic properties ensures self-consistency of parameters and relies on a dynamic criterion based on soil internal drainage dynamics (Assouline and Or, 2014; Nasta et al., 2021).

~~The most frequently used pedotransfer functions, which can be used to predict soil water content and hydraulic conductivity from easily available soil information, were tested by Nasta et al. (2021) on European datasets: GRIZZLY, HYPRES and EU-HYDI. Based on their results we selected the approaches that performed well on the European datasets. Using the selected approaches, we then computed the field capacity (FC), wilting point (WP), plant available water capacity (AWC), and saturated hydraulic conductivity (KS) for the EU-HYDI dataset. The selected approaches are listed under Table 4. We considered only those approaches, which required the mean soil depth, sand, silt, clay content, organic carbon content, and bulk density as input. When FC, WP, AWC and KS is computed from the measured or predicted parameters of the VG model, it secures that all required soil hydraulic properties are derived from a physically based model, resulting in a physically plausible soil hydraulic property combination.~~

Usually, the FC is considered as water content at a static soil matric potential. The -330 cm matric potential is widely used for this approximation (Kutílek and Nielsen, 1994). Assouline and Or (2014) derived a physically-based analytical equation with self-consistent static and dynamic criteria for the prediction of FC. ~~This approach is recommended for the computation of FC.~~ It requires the parameters of the van Genuchten model:

$$FC = \theta_r + (\theta_s - \theta_r)\left\{1 + \left[\frac{n-1}{n}\right]^{(1-2n)}\right\}^{\left(\frac{1-n}{n}\right)} \quad (19\underline{7})$$

where FC (cm^3 cm^{-3}) is water content at field capacity, θ_r (cm^3 cm^{-3}) and θ_s (cm^3 cm^{-3}) are the residual and saturated soil water contents, respectively, α (cm^{-1}) is a scale parameter, and n (-) is the shape parameter of the van Genuchten model (van Genuchten, 1980).

Computation of WP could be performed based on the VG parameters, using Eq. (18):

$$WP = \theta_r + \frac{\theta_s - \theta_r}{[1+(\alpha \cdot 15000^n)]^{1-1/n}} \quad (\underline{20}18)$$

AWC is defined by FC and WP with the following equation:

$$AWC = FC - WP \quad (\underline{21}19)$$

Physically, it is impossible to have AWC < 0, therefore its minimum value has to be set to 0.001 cm^3 cm^{-3}.

Computation of KS from parameters of the van Genuchten model can be performed by e.g. the equation of Guarracino (2007):

$$KS = 4.65 \cdot 10^4 \theta_s \alpha^2 \quad (2\underline{2}0)$$

where KS is expressed in units of cm d^{-1}. If a unit in mm h^{-1} is required, the Eq. (2~~2~~0) has to be multiplied by 0.41667.

The most frequently used pedotransfer functions, which can be used to predict soil water content and hydraulic conductivity from easily available soil information, were tested by Nasta et al. (2021) on European datasets: GRIZZLY, HYPRES and EU-HYDI. Based on their results we selected the approaches that performed well on the European datasets. Using the selected approaches, we then computed the field capacity (FC), wilting point (WP), plant available water capacity (AWC), and saturated hydraulic conductivity (KS) for the EU-HYDI dataset. The selected approaches are listed under Table 5~~4~~. We considered only those approaches, which required the mean soil depth, sand, silt, clay content, organic carbon content, and bulk density as input. When FC, WP, AWC and KS is computed from the measured or predicted parameters of the VG model, it secures that all required soil hydraulic properties are derived from a physically based model, resulting in a physically plausible soil hydraulic property combination.

Table 5~~4~~. Approaches tested in the EU-HYDI for the prediction of field capacity (FC), wilting point (WP), available water capacity (AWC) and saturated hydraulic conductivity (KS)

Soil hydraulic property	Type of the prediction	Description	Abbreviation of the prediction	Reference
FC	direct	FC at -100 cm matric potential with PTF03 of euptfv2	pred_FC_100	(Szabó et al., 2021)
	direct	FC at -330 cm matric potential with PTF02 of euptfv2	pred_FC_330	(Szabó et al., 2021)
	from VG parameters	VG parameters predicted with PTF07 of euptfv2 for mineral soils and PTF18 of euptfv1 for organic soils, matric potential set to -100 cm	pred_FC_VG_100	(van Genuchten, 1980; Szabó et al., 2021; Tóth et al., 2015)
	from VG parameters	VG parameters predicted with PTF07 of euptfv2 for mineral soils and PTF18 of euptfv1 for organic soils, matric potential set to -330 cm	pred_FC_VG_330	(van Genuchten, 1980; Szabó et al., 2021; Tóth et al., 2015)
	from VG parameters	VG parameters predicted with PTF07 of euptfv2 for mineral soils and PTF18 of euptfv1 for organic soils + equation of Assouline and Or (2014) based on θ_s, θ_r and α	pred_FC_VG_AO	(Assouline and Or, 2014; Szabó et al., 2021; Tóth et al., 2015)
WP	direct	WP at -1500 kPa with PTF02 of euptfv2	pred_WP	(Szabó et al., 2021)
	direct	SWAT approach	pred_WP_SWAT	(Neitsch et al., 2009)
	from VG parameters	VG parameters predicted with PTF07 of euptfv2 for mineral soils and PTF18 of euptfv1 for organic soils + van Genuchten function	pred_WP_VG	(van Genuchten, 1980; Szabó et al., 2021; Tóth et al., 2015)
AWC	from VG parameters	AWC from pred_FC_VG_100 and pred_WP_VG	pred_AWC_VG_100	(van Genuchten, 1980; Szabó et al., 2021)
	from VG parameters	AWC from pred_FC_VG_330 and pred_WP_VG	pred_AWC_VG_330	(van Genuchten, 1980; Szabó et al., 2021)

	from VG parameters	AWC from pred_FC_VG_AO and pred_WP_VG	pred_AWC_VG_AO	(Assouline and Or, 2014; van Genuchten, 1980; Szabó et al., 2021)
KS	from VG parameters	VG parameters predicted with PTF07 of euptfv2 + equation of Guarracino (2007) based on θ_s and α	pred_KS_VG	(Guarracino, 2007; Szabó et al., 2021)

formázott: Címsor 4

2.3.2 Hydrological soil groups

The hydrologic soil groups (HSG) are based on the infiltration characteristic of the soil and include four groups having similar runoff potential. The groups are defined based on the saturated hydraulic conductivity, depth to high water table and depth to water impermeable layer (Table 5). More details can be found in U.S. Department of Agriculture Natural Resources Conservation Service (2009).

Table 5. Definition of soil hydrologic groups based on U.S. Department of Agriculture Natural Resources Conservation Service (2009). KS: saturated hydraulic conductivity (μm s^{-1}).

Depth to water impermeable layer*	Depth to high water table**	KS of least transmissive layer in depth range (μm s^{-1})	KS depth range	HSG***
<50 cm	—	—	—	D
50 to 100 cm	<60 cm	>40.0	0 to 60 cm	A/D
		>10.0 to ≤40.0	0 to 60 cm	B/D
		>1.0 to ≤10.0	0 to 60 cm	C/D
		≤1.0	0 to 60 cm	D
	≥60 cm	>40.0	0 to 50 cm	A
		>10.0 to ≤40.0	0 to 50 cm	B
		>1.0 to ≤10.0	0 to 50 cm	C
		≤1.0	0 to 50 cm	D
>100 cm	<60 cm	>10.0	0 to 100 cm	A/D
		>4.0 to ≤10.0	0 to 100 cm	B/D
		>0.40 to ≤4.0	0 to 100 cm	C/D
		≤0.40	0 to 100 cm	D
	60 to 100 cm	>40.0	0 to 50 cm	A
		>10.0 to ≤40.0	0 to 50 cm	B
		>1.0 to ≤10.0	0 to 50 cm	C
		≤1.0	0 to 50 cm	D
	>100 cm	>10.0	0 to 100 cm	A
		>4.0 to ≤ 10.0	0 to 100 cm	B
		>0.40 to ≤4.0	0 to 100 cm	C
		≤0.40	0 to 100 cm	D

*An impermeable layer has a KS less than 0.01 μm s^{-1} [0.0014 in h^{-1}] or a component restriction of fragipan; duripan; petrocalcic; orstein; petrogypsic; cemented horizon; densic material; placic; bedrock, paralithic; bedrock, lithic; bedrock, densic; or permafrost. **High water table during any month during the year. ***Dual HSG classes are applied only for wet soils (water table less than 60 cm [24 in]). If these soils can be drained, a less restrictive HSG can be assigned, depending on the KS.

~~For modelling purposes, it is important if tile drainage is present in the modelled area, because tile drainage systems influence the soil infiltration rate and runoff potential. The EU-HYDI dataset does not include measured data on the depth to water table and maximum rooting depth, therefore this method is not applied for this dataset.~~

~~Derivation of HSG requires local input data. If local datasets are not available, and SoilGrids 2017 (Hengl et al., 2017) was chosen as the source for the basic soil data, HSG can be retrieved from the global HYSOGs250m (Ross et al., 2018) database.~~

~~2.4~~ 2.2.3 Soil chemical parameters

formázott: Címsor 3

For mapping soil phosphorus (P) content of the topsoil we present a simple approach based on mean statistics, which is suitable for areas where data is scarce. Land use has the strongest influence on soil P content, with most agricultural areas exhibiting higher P levels compared to regions with natural land cover (Ballabio et al., 2019). The available P level in agricultural soils is influenced by the P inputs – fertilizers, manure, atmospheric deposition, chemical weathering – and outputs – plant uptake and erosion. ~~primarily influenced by~~The agricultural management practices~~fertilization practices~~ (Tóth et al., 2014)~~, which is~~ are determined by factors such as the country's economy, climate, tillage practices, and crop production characteristics. Based on the relationships mentioned above, the geometric mean of soil P is computed by land use categories and assigned to the local land use map using the mean statistics-based method. This approach comprises three main steps:

1) Selection of LUCAS Topsoil samples (EUROSTAT, 2015; Orgiazzi et al., 2018) from the adequate year and an agroclimatic zone (Ceglar et al., 2019) similar to the target area, preferably in the same country (NUTS region). Additional criteria for the data selection could be comparable soil types and fertilization systems.~~and with comparable fertilization schemes.~~ If this information is not known, the NUTS2 phosphorus map of the European cropland areas might be useful in the data selection (Tóth et al., 2014).
2) Computation of the geometric mean of soil P for each land use category.
3) Assigning the mean values to the local land use map.

Further details about the mapping can be found in Szabó and Kassai (2022) .

We prepared a soil P content map by applying the proposed method for a case study called Felső-Válicka, located in Hungary (Figure 1). The resulting map was then compared to i) the European topsoil phosphorus content map (Ballabio et al., 2019) and ii) a locally measured independent dataset provided by an agricultural company. Limited availability of soil nutrient data hampered the wider scale of comparison.

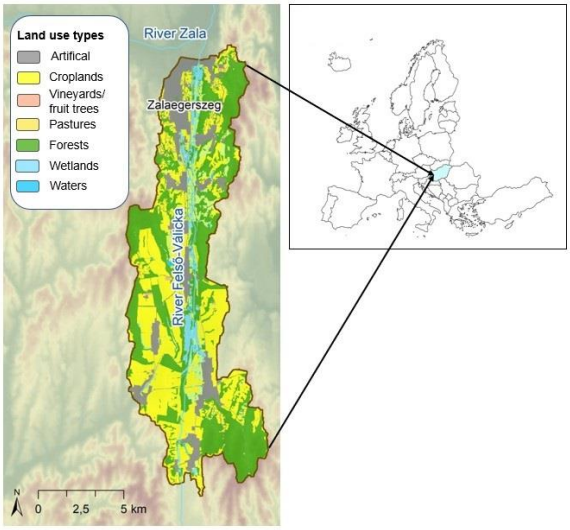


Figure 1. Local land use map of the Felső-Válicka case study in Hungary.

Organic nitrogen can be estimated from soil organic carbon content (Amorim et al., 2022; Liu et al., 2016; Pu et al., 2012; Zhai et al., 2019) if measured data are not available. The concentration of inorganic nitrogen~~ate~~ in soil is highly variable in space and time and the dynamic of its amount is significantly influenced by leaching, denitrification, volatilization, and nitrogen fertilization (Zhu et al., 2021). Therefore, no general method is available for its prediction so far. However, when simulating nitrogen uptake and losses on catchment-scale, information on the amount and timing of nitrogen fertilization is often more crucial, than knowledge of the initial nitrate content of the soil (Krevh et al., 2023). The mineral and relatively dynamic N pools are often considered to be initialized during the warmup period of the models (Yuan and Chiang, 2015). It is especially important to have a proper parameterization of the agricultural management (e.g. fertilization, residue management) setup in the model application with an appropriate length of the warmup period, where we recommend it to be no less than 4 years. Furthermore, it is beneficial to initialise the SOM levels accurately to define the large and rather slow pool of organic nitrogen (Liang et al., 2023).

3 Results and discussion

3.1 Bulk density

Table 6 shows the prediction performance of the selected PTFs. The performance varies depending on the texture classes, e.g., it is lower for clayey soils, sandy clay loams, ~~silts,~~ and ~~silty~~ loams in the EU-HYDI dataset (Figure S1 a). For the LUCAS topsoil samples, the performance of all PTFs is lower compared to their performance on EU-HYDI in terms of RMSE. Additionally, all analysed methods tend to overpredict bulk density. ~~On average, the BD_Alexander_A PTF (E.q. 3) had the lowest RMSE value when used with the EU-HYDI dataset and better performance overall.~~The BD_Alexander_A PTF (E.q. 3) ranks highest based on the sample-number-weighted average results of the Kruskal-Wallis test, analysed on both the EU-HYDI and LUCAS dataset (Table 6, weighted rank). ~~To properly derive the missing parameters in soils with high organic matter content (soils with organic carbon content equal to or higher than 12 %) it is suggested to use the PTF of Hossain (E.q. 9). The BD_Alexander_A_Hossain shows the prediction performance of the combined use of the BD_Alexander_A (for soils with organic carbon content less than 12 %) and BD_Hossain (for soils with organic carbon content equal to or higher than 12 %) PTFs.~~The BD_Alexander_A_Hossain PTF shows the ~~prediction~~ performance of the combined use of the BD_Alexander_A (for soils with organic carbon content less than 12 %) and BD_Hossain (for soils with organic carbon content equal to or higher than 12 %) PTFs. This combined PTF performs similarly to the simple BD_Alexander_A method but helps to properly derive bulk density for soils with high organic matter content. Figure 2 shows the scatterplot of measured versus predicted bulk density values of the best performing PTF, where the predefined bulk density is capped at 1.72 g cm^{-3} as product of the models constraints.

Table 6. Prediction performance of bulk density (g cm^{-3}) computed by available pedotransfer functions on the point data of EU-HYDI (N = 11,273) and LUCAS (N = 5821). ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

	EU-HYDI (N=11273)							LUCAS (N = 5821)							Weighted rank
PTF	ME	MAE	RMSE	NSE	R^2	Sign. diff.*	Rank	ME	MAE	RMSE	NSE	R^2	Sign. diff.*	Rank	
BD_Alexander_A	0.01	0.15	0.19	0.22	0.27	g	1	-0.22	0.26	0.32	-0.01	0.49	b	6	2.70
BD_Alexander_A_Hossain	0.01	0.15	0.19	0.22	0.27	g	1	-0.24	0.27	0.33	-0.06	0.49	b	6	2.70
BD_Alexander_B	0.08	0.16	0.21	0.05	0.27	e	4	-0.14	0.21	0.27	0.28	0.49	e	3	3.66
BD_MAn_J_A	0.07	0.16	0.21	-0.04	0.23	f	3	-0.10	0.27	0.44	-0.90	0.39	c	5	3.68
BD_MAn_J_B	0.09	0.17	0.21	-0.01	0.27	d	5	-0.12	0.20	0.26	0.32	0.49	f	2	3.98
BD_Rawls	0.27	0.29	0.33	-1.40	0.27	a	8	-0.03	0.18	0.23	0.47	0.51	g	1	5.62
BD_Bernoux	0.20	0.23	0.28	-0.72	0.22	b	7	-0.15	0.24	0.30	0.13	0.35	d	4	5.98
BD_Hollis	0.04	0.20	0.25	-0.45	0.10	c	6	-0.26	0.28	0.34	-0.17	0.47	a	8	6.68

~~Name of PTF~~	~~ME~~	~~MAE~~	~~RMSE~~	~~NSE~~	~~R^2~~	~~Sign. difference *~~
~~BD_Alexander_A_Hossain~~	~~0.009~~	~~0.145~~	~~0.185~~	~~0.222~~	~~0.265~~	~~g~~
~~BD_Alexander_A~~	~~0.010~~	~~0.145~~	~~0.185~~	~~0.220~~	~~0.269~~	~~g~~
~~BD_Hollis~~	~~0.000~~	~~0.149~~	~~0.189~~	~~0.188~~	~~0.277~~	~~f~~
~~BD_Alexander_B~~	~~0.084~~	~~0.162~~	~~0.205~~	~~0.045~~	~~0.269~~	~~d~~
~~BD_MAn_J_B~~	~~0.094~~	~~0.167~~	~~0.211~~	~~-0.009~~	~~0.269~~	~~e~~
~~BD_MAn_J_A~~	~~0.069~~	~~0.159~~	~~0.214~~	~~-0.043~~	~~0.229~~	~~e~~
~~BD_Bernoux~~	~~0.202~~	~~0.233~~	~~0.275~~	~~-0.718~~	~~0.215~~	~~b~~
~~BD_Rawls~~	~~0.271~~	~~0.285~~	~~0.325~~	~~-1.404~~	~~0.266~~	~~a~~

*Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example, performance indicated with the letter c is significantly better than the one noted with letters b and a.

If only the soil's organic carbon content is known, the prediction accuracy is restricted. The RMSE value of BD_Alexander_A_Hossain PTF on the EU-HYDI is comparable with the accuracy of an ML-based PTF built on a French dataset (Chen et al., 2018), when computed on independent validation sets, which reported RMSE between 0.17 and 0.22 g cm^{-3}. This performance is better than the results of a model transferability test of a PTF derived on soils from Campania, Italy, analysed on the EU-HYDI (Palladino et al., 2022), which had RMSE = 0.210 g cm^{-3}. Yi et al. (Xiangsheng et al., 2016) and De Souza et al.(Souza et al., 2016) found RMSE values higher than 0.185 g cm^{-3} when they applied PTFs trained on temperate soils, available from the literature, on a Chinese permafrost region and Brazilian catchment, respectively. This outcome underscores the significance of refraining from using a PTF that was trained on soils formed under different conditions – i.e. with different soil forming factors – , making it inapplicable to the specific target area (Chen et al., 2018; Tranter et al., 2009).

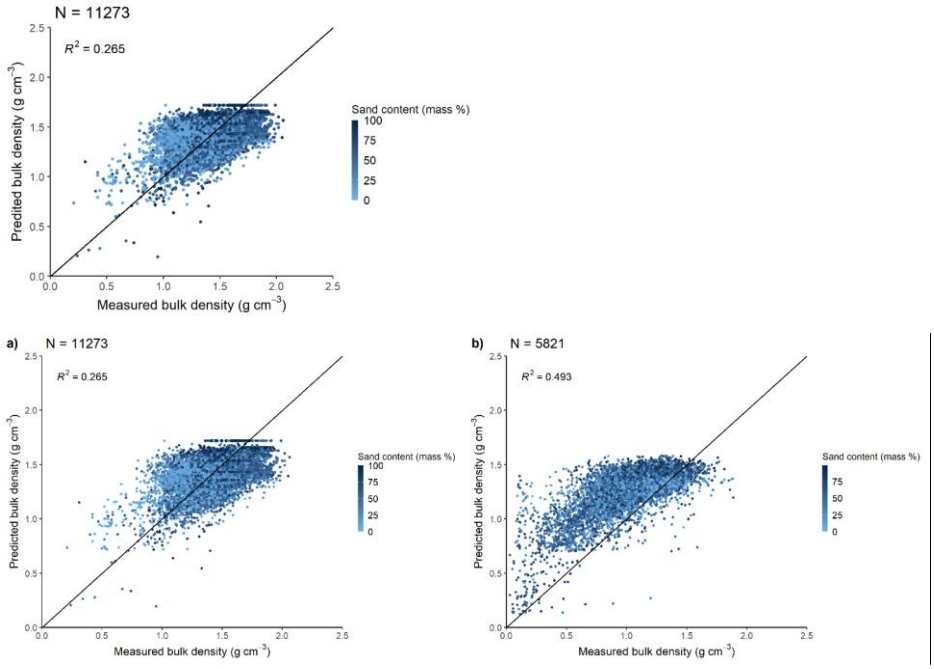


Figure 2. Scatterplot of measured versus predicted bulk density values of the best performing PTF (BD_Alexander_A_Hossain) analysed on the point data of EU-HYDI (a) and LUCAS (b) dataset.

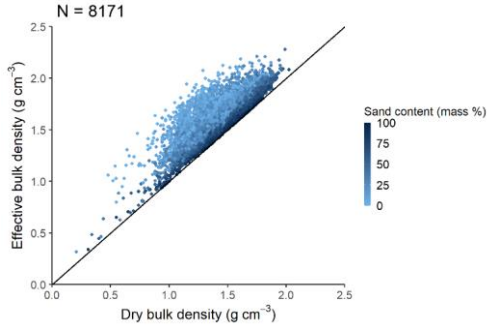


Figure 3. Scatterplot of dry versus effective bulk density analysed based on the point data of EU-HYDI.

Effective bulk density is always higher than dry bulk density. Effective bulk density value computed for the EU-HYDI dataset with Eq. (~~10~~22) and (~~11~~23) was between 0.32 and 2.17 g cm^{-3}. Figure 3 shows the scatterplot of dry bulk density versus computed effective bulk density based on the EU-HYDI dataset.

Based on the performance analysis on EU-HYDI (N = 11,273) the prediction of dry bulk density could be performed with i) Eq. (~~23~~) (BD_Alexander_A) for soils with OC < 12% and ii) Eq. (~~89~~) (BD_Hossain) for soils with OC >= 12%.

3.2 Porosity

The porosity values computed based on the particle density predicted by Schjønning et al. PTF (POR_Schjonning_etal) implemented in Eq. (10~~3~~) were significantly more accurate on those EU-HYDI samples, which considered measured particle density value for the computation of porosity (Table 7). If solely samples with low organic matter content, specifically less than 1%, were considered for analysis, no notable differences between the methods were observed. In the case of soils with organic matter content higher than 1 % the prediction of porosity significantly improved if particle density was computed based on distinction between organic matter and mineral substrates. Figure 4 displays the scatterplot of measured versus Eq. (10~~3~~) (POR_Schjonning_etal) predicted porosity values.

Table 7. Prediction performance of porosity (vol %) computed by available pedotransfer functions on the point data of EU-HYDI results are structured by organic matter content. OM: organic matter content (mass %), N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Name of PTF	OM (mass %)	N	ME	MAE	RMSE	NSE	R^2	Sign. diff.*
POR_Schjonning_etal	any	2290	0.19	1.38	2.53	0.882	0.889	c
POR_Schjonning_etal_recal		2290	1.05	1.81	2.84	0.852	0.878	a
POR_2_65		2290	0.23	1.67	2.71	0.866	0.883	b
POR_Schjonning_etal	0 =< OM < 10	2246	0.20	1.38	2.55	0.860	0.869	c
POR_Schjonning_etal_recal		2246	1.06	1.81	2.86	0.824	0.855	a
POR_2_65		2246	0.29	1.64	2.70	0.843	0.861	b
POR_Schjonning_etal	0 =< OM < 5	1943	0.23	1.34	2.48	0.841	0.849	c
POR_Schjonning_etal_recal		1943	1.01	1.76	2.78	0.801	0.834	a
POR_2_65		1943	0.52	1.57	2.61	0.824	0.840	b
POR_Schjonning_etal	0 =< OM < 1	492	-0.22	1.32	1.84	0.879	0.881	a
POR_Schjonning_etal_recal		492	-0.01	1.25	1.69	0.898	0.898	a
POR_2_65		492	0.23	1.23	1.63	0.905	0.907	a
POR_Schjonning_etal	10 =< OM	44	-0.24	1.41	1.94	0.968	0.969	b
POR_Schjonning_etal_recal		44	0.92	1.49	1.91	0.969	0.980	b
POR_2_65		44	-2.85	2.86	3.29	0.909	0.977	a

*Different letters indicate significant differences at the 0.05 level between the accuracy of the methods based on the squared error; for example, performance indicated with the letter c is significantly better than the one noted with letters b and a.

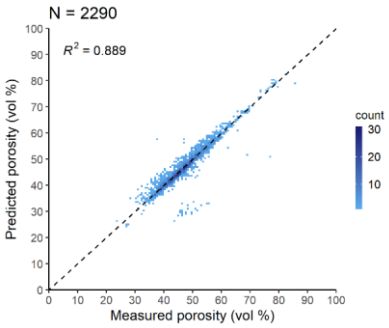


Figure 4. Scatterplot of measured versus predicted porosity values of the best performing PTF, POR_Schjonning_etal (Eq. 103) analysed based on the EU-HYDI subset with measured particle density values. Count: the number of cases in each quadrangle.

When data on porosity is missing, some studies use the saturated water content as its approximation, although based on the literature the saturated water content is usually equal or less than the total porosity (Lal and Shukla, 2004). Figure 5 shows the relationship between porosity and saturated water content for 391 EU-HYDI samples with measured values of both parameters. Among these samples, ~~in the EU-HYDI dataset.~~ 56.5% ~~of the samples with measured porosity and saturated water content~~ have a total porosity larger or equal to the saturated water content. ~~In the case of~~For the samples where the saturated water content is higher than the total porosity, the reason may~~can~~ be the uncertainties in the measurement of both parameters. It is possible that ~~uncertainty of the saturated water content measurement, i.e.~~ free water could have pounded on top of the sample when its saturated weight was measured, and errors in the measurement of particle density used to compute porosity may have also contributed (Kutílek and Nielsen, 1994; Nimmo, 2004), resulting in a lower porosity.

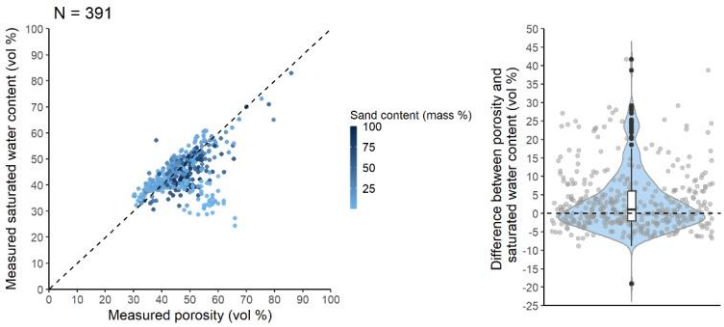


Figure 5. Scatterplot of measured porosity values versus measured saturated water content and boxplot of the difference between the two values tested on point data in EU-HYDI dataset.

Based on the study performed in EU_HYDI, prediction of porosity could be performed with the Schjønning et al. PTF Eq. (~~13~~10) instead of defining particle density as 2.65 g cm^{-3} in Eq. (12).

3.2 Albedo

The range of soil albedo computed with Eq. (1~~6~~3) for the topsoil layers of the EU-HYDI dataset with different moisture states (Table 8) is within the range of the values available from the literature, which is 0.10-0.43 in the case of ECOCLIMAP-U dataset (Carrer et al., 2014). The median dry, bare soil albedo and surface albedo values of year 2022 extracted from the MCD43A3 database to the EU-HYDI topsoil layers are significantly lower than the computed values (Figure 6). The histogram of the monthly surface albedo and dry, bare soil albedo values extracted to the EU-HYDI topsoil samples are show on Figure S2a and b. It's crucial to specify the moisture condition for which the albedo value is needed in the modelling process.

Table 8. Descriptive statistics of soil albedo values computed with the simplified Gascoin et al. (2009) equation on the topsoil samples of EU-HYDI dataset (N = 7,537) at different moisture states: based on saturation (ALB_comp_THS), field capacity (ALB_comp_FC), wilting point (ALB_comp_WP).

Albedo at different moisture state	Minimum	Maximum	Range	Mean	Median	Standard deviation
ALB_comp_THS	0.15	0.17	0.02	0.15	0.15	0.00
ALB_comp_FC	0.15	0.31	0.16	0.17	0.16	0.02
ALB_comp_WP	0.15	0.46	0.31	0.22	0.19	0.08

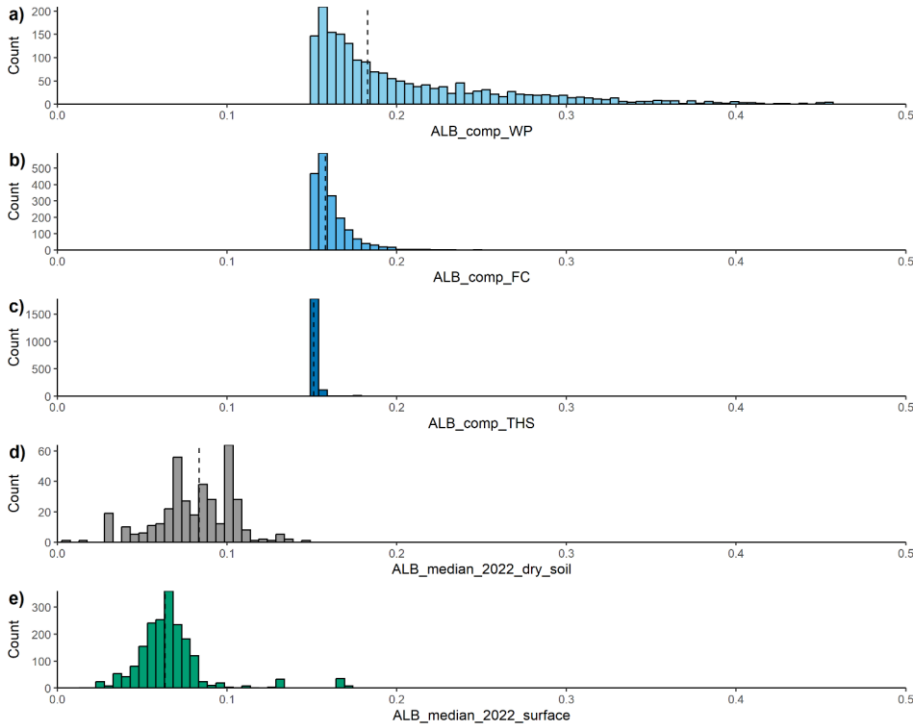


Figure 6. Histograms of the soil albedo computed with the Gascoin et al. (2009) equation for the topsoil layers of the EU-HYDI dataset in the case of three moisture states: at saturation (ALB_comp_THS) (a), internal drainage dynamics-based field capacity (ALB_comp_FC) (b) and wilting point (ALB_comp_WP) (c) (N = 2408), and median surface (d) and dry, bare soil albedo (e) of year 2022 (ALB_median_2022_dry_soil, ALB_median_2022_surface) extracted from the MCD43A3 global database for the EU-HYDI topsoil layers. Vertical dashed lines indicate the median values.

3.3 Soil erodibility factor

The soil erodibility factor (K-factor) computed on the topsoil samples of the EU-HYDI dataset with Eq. (1~~7~~4) are comparable with the values of the European 500 m resolution soil erodibility map published by Panagos et al. (Panagos et al., 2014) in terms of range, mean and density of the values (Table 9 and Figure 7), although the relationship between the computed and mapped values was weak (Figure 8). For the computation of the European map soil organic matter content, soil texture, coarse fragments content, soil structure and stoniness were considered. The Renard et al. (Eq. 15) equation resulted in a higher median

value but lower possible maximum value because the computed soil erodibility factor is capped at 0.044 $\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$ due to the constraints of the model. The relationship between the soil erodibility factors derived by different methods is strongest between the values computed using the Sharpley and Williams (1990) method and the Renard et al. (1997) method. This is logical because both methods consider the particle size distribution of the soil as input information.

Both approaches, whether directly applying the equations~~Eq.~~ (Eq. 1~~4~~7 or 15) or extracting values, generate predicted soil erodibility values. While both can be used for environmental modelling, i) European soil erodibility map could be linked with LUCAS topsoil dataset and maps, ii) employing Eq. (1~~4~~7) or (15) might offer greater consistency with the other local basic and physical soil data, aligning more seamlessly with the modelling process. Given the scarcity of measured K-factor values, our suggestion is to initially utilize these predicted values as preliminary approximations. However, we recommend fine-tuning this factor during the model calibration process.

Table 9. Descriptive statistics of soil erodibility factor values computed with the Sharpley and Williams (1990) and Renard et al. (1997) equations on the topsoil samples of the EU-HYDI dataset (N = 11,287) provided in two different units.

Method	USLE K factor in different units	Minimum	Maximum	Range	Mean	Median	Standard deviation
Sharpley and Williams (1990)	$\left(\frac{t \cdot arce \cdot h}{hundreds\ of\ acre \cdot foot - tonf \cdot inch}\right)$	0.000.0 0	0.480.4 8	0.480.4 8	0.270.2 7	0.270.2 7	0.090.09
	$\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$	0.0000. 000	0.0630. 063	0.0630. 063	0.0360. 036	0.0350. 035	0.0120.012
Renard et al. (1997)	$\left(\frac{t \cdot arce \cdot h}{hundreds\ of\ acre \cdot foot - tonf \cdot inch}\right)$	0.05	0.33	0.29	0.24	0.27	0.09
	$\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$	0.006	0.044	0.038	0.032	0.035	0.012

táblázatot formázott
formázott: Balra zárt
formázott
formázott
formázott
formázott
formázott
formázott
formázott
formázott
formázott
formázott
formázott
formázott: Balra zárt
formázott
formázott: Balra zárt
formázott
formázott
formázott
formázott
formázott
formázott
formázott: Balra zárt
formázott
formázott
formázott
formázott
formázott
formázott

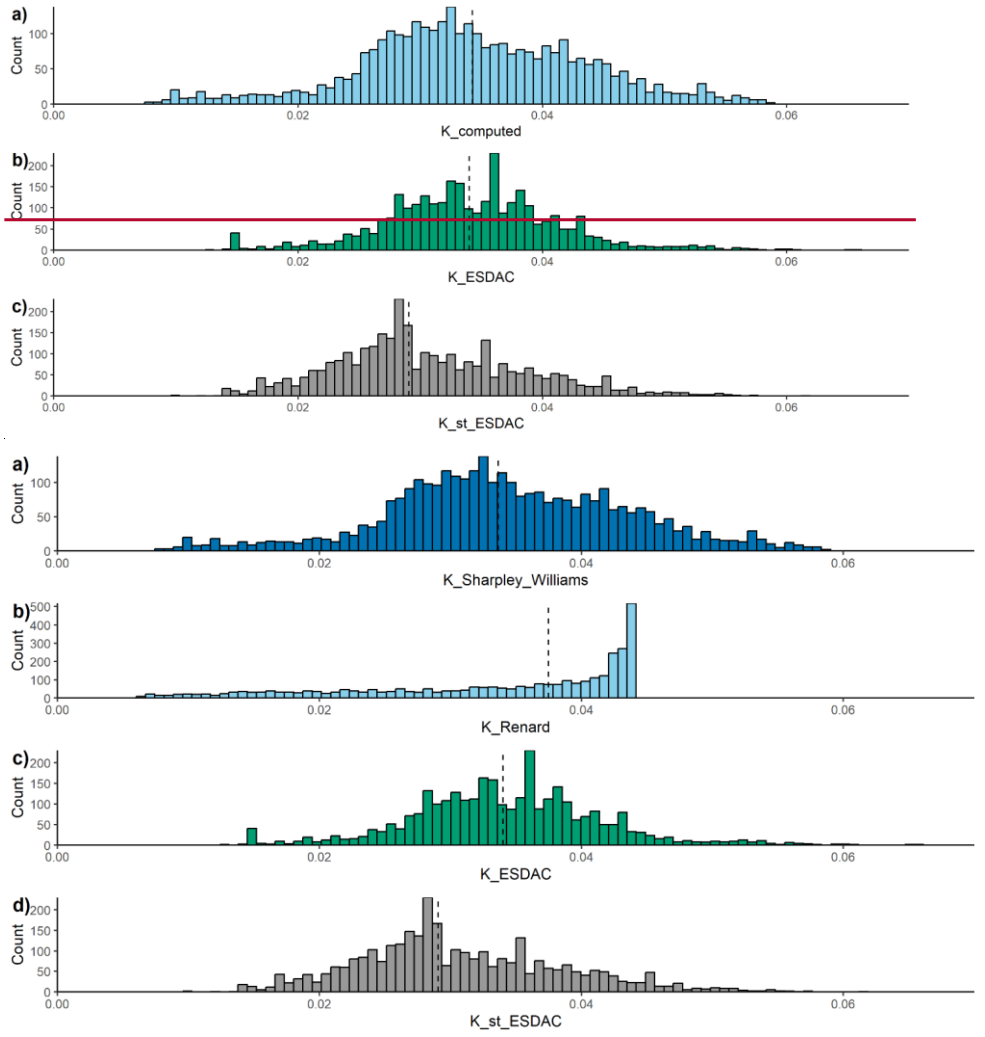
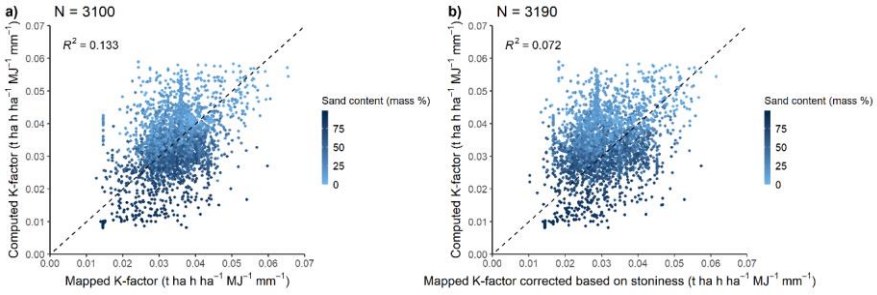
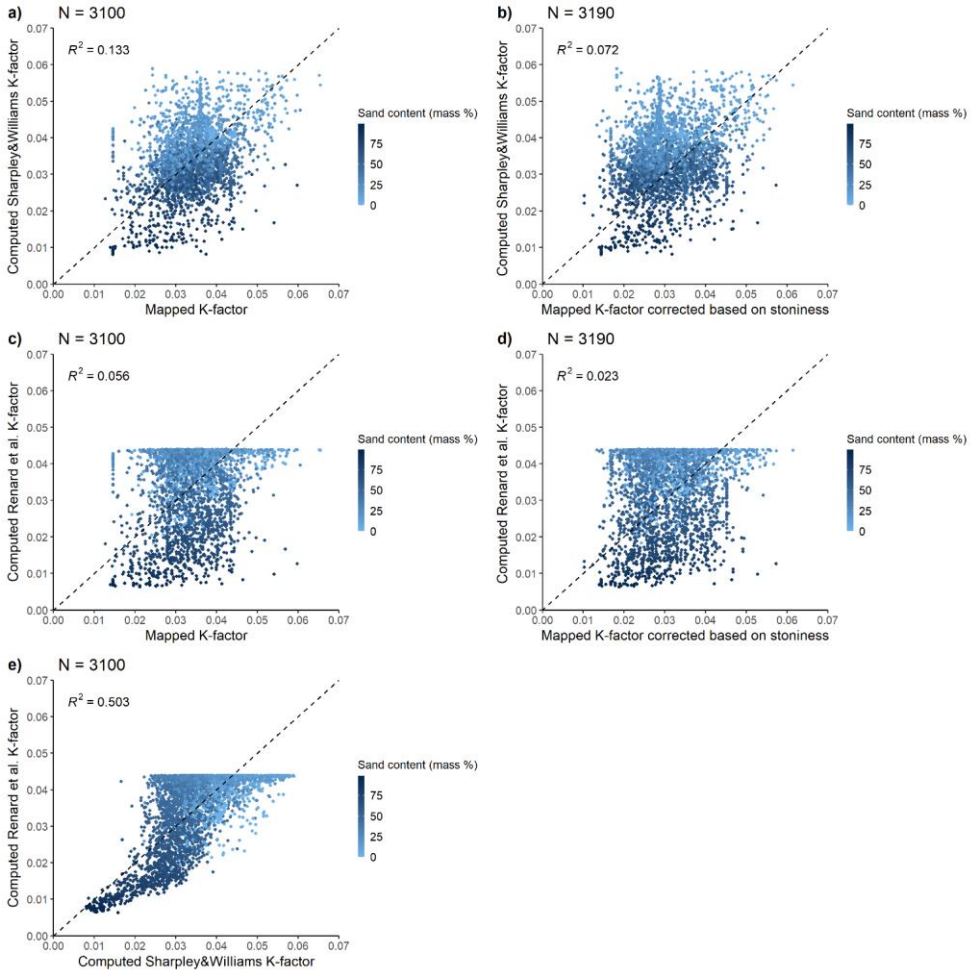


Figure 7. Histogram of the soil erodibility factor $\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$ computed with the Sharpley and Williams (1990) (K_Sharpley_Williams, N = 3276) (a) and Renard et al. (1997) (K_Renard, N = 3276) (b) equations on the topsoil samples of

the EU-HYDI dataset ~~(K_computed, N = 3276) (a),~~ and extracted from the soil erodibility map of Europe for the EU-HYDI topsoil layers without (K_ESDAC, N = 3100) (~~c~~b) and considering stoniness (K_st_ESDAC, N = 3190) (~~d~~e). Vertical dashed lines indicate the median values.


a)
N = 3100
R^2 = 0.133
Computed K-factor (t ha h ha^{-1} MJ^{-1} mm^{-1})
Mapped K-factor (t ha h ha^{-1} MJ^{-1} mm^{-1})
Sand content (mass %)
75
50
25
0
b)
N = 3190
R^2 = 0.072
Computed K-factor (t ha h ha^{-1} MJ^{-1} mm^{-1})
Mapped K-factor corrected based on stoniness (t ha h ha^{-1} MJ^{-1} mm^{-1})
Sand content (mass %)
75
50
25
0



555

Figure 8. Scatterplot of computed soil erodibility factors ~~K-factor (FC_VG_AO)~~ versus extracted from the European ~~K-factor~~ soil erodibility factor map without (a, c) and with considering stoniness (b, d) based on the topsoil samples of the EU-HYDI dataset $\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$. Plot (e) shows the relationship between the values computed by the Sharpley and Williams (1990) and Renard et al. (1997) methods.

3.4 Field capacity

The FC defined (see abbreviations in Table 5) based on soil internal drainage dynamics (FC_VG_AO) differed from the field capacity measured at -100 cm, or -330 cm matric potential (FC_100 and FC_330 respectively) or computed from VG parameters at -100 cm, or -330 cm matric potential (FC_VG_100 and FC__VG_330 respectively) (Figure 9), as was expected. The scale of difference depends on i) the predefined soil matric potential value, which we consider using as measured field capacity, and ii) characteristics soil properties that influence soil hydraulic behaviour, such as soil texture, organic matter content, bulk density, clay mineralogy, structure, etc. Figures S3 and S4 show that for soils with low sand content (< 25 %) and high silt content (> 50 %) or low bulk density (< 0.7 g cm^{-3}) the FC_VG_AO is lower than water content measured at -100 cm or -330 cm matric potential (FC_VG_AO vs. FC_100 and FC_VG_AO vs. FC_330).

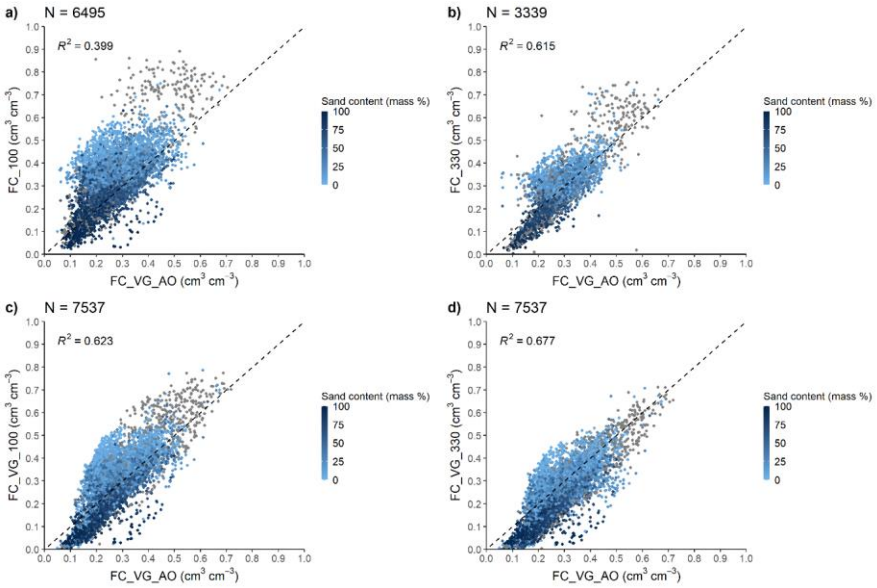


Figure 9. Scatterplot of internal drainage dynamics-based field capacity (FC_VG_AO) versus field capacity at -100 cm matric potential (a), at -330 cm matric potential (b), computed based on VG model with parameter h (head) set at -100 cm matric potential (c) and -330 cm matric potential (d).

If FC at a single matric potential value is computed from the fitted VG parameters (FC_VG_100, FC_VG_330) their Pearson correlation with the FC_VG_AO is higher than in the case of FC measured at -100 or -330 cm matric potential (Figure S5). This is logical because in the case of FC_VG_100 and FC_VG_300 the same VG parameters are used for the computation as for FC_VG_AO. In the case of EU-HYDI the FC_VG_330 is the closest to the FC_VG_AO. The only exception are sands where FC measured at -330 cm matric potential has the highest correspondence with FC_VG_AO (Figure S6).

Table 10. Prediction performance of internal drainage dynamics-based field capacity (cm^3 cm^{-3}) computed by pedotransfer functions on the FC and VG test sets of the EU-HYDI dataset. N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Approach to predict FC*	N	ME	MAE	RMSE	NSE	R^2
pred_FC_VG_AO	1591	0.005	0.043	0.058	0.514	0.519
pred_FC_100	1413	-0.071	0.083	0.106	-0.779	0.297
pred_FC_330	782	-0.010	0.047	0.061	0.210	0.395
pred_FC_VG_100	1591	-0.015	0.070	0.090	-0.184	0.320
pred_FC_VG_330	1591	0.045	0.073	0.091	-0.198	0.339

*pred_FC_VG_AO: predicted internal drainage dynamics-based field capacity based on VG parameters predicted from basic soil properties; pred_FC_100, pred_FC_330: field capacity at -100 and -330 cm matric potential directly predicted from basic soil properties; pred_FC_VG_100, pred_FC_VG_330: field capacity at -100 and -330 cm matric potential based on VG parameters predicted from basic soil properties.

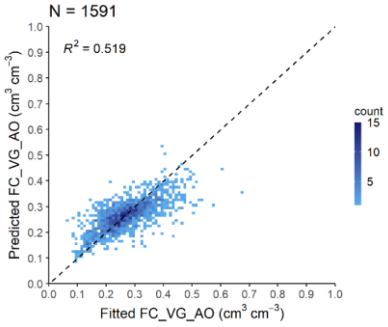


Figure 10. Scatterplot of internal drainage dynamics-based FC (FC_VG_AO) computed from fitted and predicted VG parameters analysed on the VG test set of the EU-HYDI dataset. Count: the number of cases in each quadrangle.

Table 10 illustrates the prediction performance of the FC_VG_AO for various approaches. If the FC_VG_AO was computed based on VG parameters predicted by the PTF07 of euptfv2, the RMSE value was 0.058 cm^3 cm^{-3}, which is comparable with

the literature values (Román Dobarco et al., 2019; Zhang and Schaap, 2017). Its correlation with the FC computed based on predicted VG parameters at -100 or -330 cm matric potential is weaker (with RMSE 0.090 and 0.091 $cm^3\ cm^{-3}$), aligning with the results drawn from the FC computed from fitted VG parameters (Figure 9 c) and d)).

Figure 10 shows the scatterplot of FC_VG_AO computed from fitted and predicted VG parameters analysed only on those samples of the EU-HYDI which were not used for training of the VG PTF07. Performance of VG PTF07 was published in Szabó et al. (2021) with 0.054 $cm^3\ cm^{-3}$ RMSE on the test set.

Thus FC_VG_AO could be used as FC and computed with Eq. (179) based on VG parameters predicted with i) euptfv2 (Szabó et al., 2021) for mineral soils and ii) euptfv1 (Tóth et al., 2015) class PTF (PTF18) for organic soils.

3.5 Wilting point

Calculating WP (see abbreviations in Table 5) from predicted VG parameters yields greater accuracy compared to using the equation provided by SWAT+ model (Figure 11, Table 11). Predicting WP directly from soil properties instead of deriving it from predicted VG parameters tends to yield greater accuracy (Børgesen and Schaap, 2005; Szabó et al., 2021; Tomasella et al., 2003) (Table 12). When multiple soil hydraulic parameters are needed, deriving all from a model encompassing the entire matric potential range secures the physical relationship between them (Weber et al., 2023).

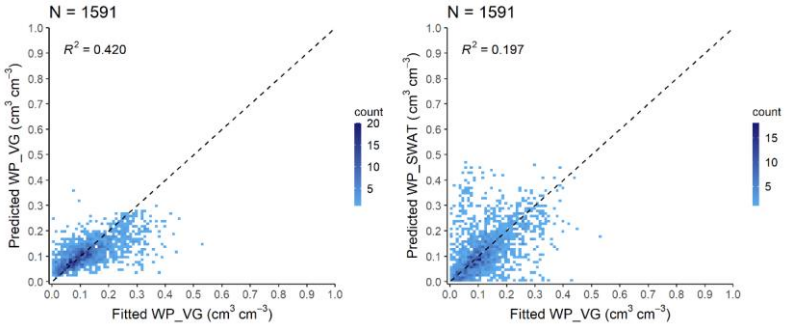


Figure 11. Scatterplot of wilting point computed from fitted VG parameters (Fitted WP_VG) versus a) wilting point computed from VG parameters predicted with euptfv2 (Predicted WP_VG) and b) wilting point predicted with the SWAT+ approach (Predicted WP_SWAT), analysed on the VG test set of the EU-HYDI dataset. Count: the number of cases in each quadrangle.

Table 11. Prediction performance of wilting point (cm^3 cm^{-3}) derived with the VG model, computed by pedotransfer functions on the VG test set of the EU-HYDI dataset. Observed variable is the WP value computed based on the fitted parameters of the VG model. N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Approach to predict WP*	N	ME	MAE	RMSE	NSE	R^2
pred_WP_VG	1591	0.016	0.045	0.065	0.382	0.420
pred_WP_SWAT	1591	-0.001	0.062	0.093	-0.239	0.197

*pred_WP_VG: wilting point computed based on VG parameters predicted from basic soil properties; pred_WP_SWAT: wilting point predicted with the equation built in the SWAT model.

Table 12. Prediction performance of wilting point (cm^3 cm^{-3}) computed by pedotransfer functions on the WP test set of the EU-HYDI dataset. Observed variable is the measured WP value. N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Approach to predict WP*	N	ME	MAE	RMSE	NSE	R^2
pred_WP_VG	2088	0.052	0.060	0.087	0.105	0.431
pred_WP_SWAT	2088	0.028	0.046	0.066	0.490	0.630
pred_WP	2088	0.000	0.033	0.046	0.755	0.755

*pred_WP_VG: wilting point computed based on VG parameters predicted from basic soil properties; pred_WP_SWAT: wilting point predicted with the equation built in the SWAT model; pred_WP: wilting point directly predicted from basic soil properties.

WP could be computed with Eq. (~~20~~18) based on VG parameters predicted with i) euptfv2 (Szabó et al., 2021) for mineral soils and ii) euptfv1 (Tóth et al., 2015) class PTF (PTF18) for organic soils.

3.6 Available water capacity

If only AWC (see abbreviations in Table 5) is required as input for a model, i.e., without FC and WP, a feasible option could involve direct prediction using a PTF like euptfv2. However, its estimation is more accurate if the internal drainage dynamics-based FC is considered for its computation (Gupta et al., 2023). Figure 12 and S9 show that coefficient of determination is low between the internal drainage dynamics-based AWC (AWC_VG_AO) and AWC based on FC at fixed matric potential (AWC_100, AWC_300, AWC_VG_100, AWC_VG_330). Which approach is the closest to the AWC_VG_AO varies based on texture classes (Figure S10).

The available water capacity based on field capacity measured at -100 cm head (AWC_100) is higher than the AWC_VG_AO, especially in the case of low sand content (< 25 %) and high silt content (> 50 %) (Figure 12c and S7). The available water capacity based on field capacity measured at -330 cm head (AWC_330) is higher than AWC_AO_VG when sand content is low (< 25 %) and silt content is high (> 50 %) and lower than AWC_AO_VG when sand content is higher than 25 % and silt content is less than 50 % (Figure 12d and S8).

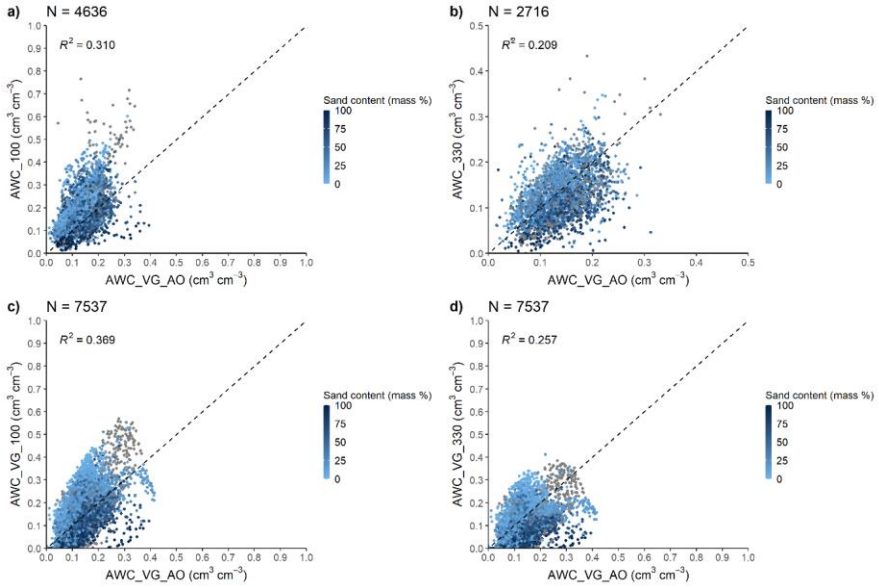


Figure 12. Scatterplot of available water capacity computed from internal drainage dynamics-based field capacity and wilting point derived based on VG parameters predicted from basic soil properties (AWC_VG_AO) versus (a, b) available water capacity computed from measured field capacity at -100 and -330 cm matric potential and wilting point, (c, d) available water capacity computed from field capacity at -100 and -330 cm matric potential and wilting point based on VG parameters predicted from basic soil properties.

Table 13. Prediction performance of available water capacity (cm^3 cm^{-3}) computed by pedotransfer functions on the VG test set of the EU-HYDI dataset. N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Approach to predict AWC*	N	ME	MAE	RMSE	NSE	R^2
pred_AWC_VG_AO	1591	-0.011	0.034	0.048	0.339	0.372
pred_AWC_VG_100	1591	-0.031	0.071	0.090	-1.325	0.072
pred_AWC_VG_330	1591	0.029	0.061	0.078	-0.725	0.044

*pred_AWC_VG_AO: available water capacity computed from internal drainage dynamics-based field capacity and wilting point derived based on VG parameters predicted from basic soil properties; pred_AWC_VG_100, pred_AWC_VG_330: available water capacity computed from field capacity at -100 and -330 cm matric potential and wilting point based on VG parameters predicted from basic soil properties.

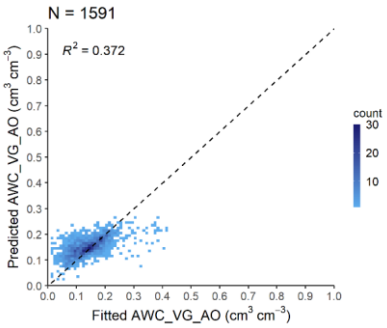


Figure 13. Scatterplot of internal drainage dynamics-based AWC (AWC_VG_AO) computed from fitted and predicted VG parameters analysed on the VG test set of the EU-HYDI dataset. Count: the number of cases in each quadrangle.

Table 13 shows the prediction performance of internal drainage dynamics-based AWC (AWC_VG_AO). As expected, the predicted internal drainage dynamics-based AWC had the lowest RMSE and highest R^2 value. The AWC computed based on the FC at 100 cm matric head derived with the predicted VG parameters (pred_AWC_VG_100) had the lowest performance. This approach yielded over-prediction of the AWC_VG_AO values when AWC_VG_AO is lower than 0.10 cm^3 cm^{-3} and under-prediction when AWC_VG_AO is higher than 0.25 cm^3 cm^{-3} (Figure 13).

Based on the findings, we recommend to compute the AWC based on the internal drainage dynamics-based FC (FC_VG_AO) and VG parameters-based WP (WP_VG) in Eq. (~~21~~19).

3.7 Saturated hydraulic conductivity

Figure 14 shows the relationship between measured KS and computed with Eq. (202) based on the fitted VG parameters (KS_VG) (see abbreviation in Table 5). The coefficient of determination between the measured and computed values is low, however fitted (not predicted) VG parameters were used for the computation. Prediction performance of KS_VG is comparable with the published widely used PTFs (Nasta et al., 2021) (Figure 15, Table 14).

Prediction of saturated hydraulic conductivity (KS) has the highest uncertainty among the soil hydraulic properties. This uncertainty originates from the differences in the measurement methods applied to measure KS, in terms of sampling volume, sample dimensions, difference between in-situ and laboratory methods (Ghanbarian et al., 2017). Due to the uncertainty of the measurements, uncertainty of the prediction is minimum one order of magnitude during the application of a PTF (Nasta et al., 2021). Estimation of KS by traditional PTFs that use basic soil properties as input is rather limited, because KS of a sample is largely determined by its structural properties and pore network characteristics, of which we lack quantitative descriptors and data (Lilly et al., 2008). There is also at least one order of magnitude difference between replicated measurements on samples

coming from the same soil horizon due to the extreme spatial variability of this particular soil property. Hence, it's important to note that while we might improve individual sample predictions for KS, the representativeness of these samples within their specific fields remains constrained. We suggest initializing this soil property using the VG parameters with Eq. (202), but keeping in mind that it should be adjusted during model calibration as a variable.

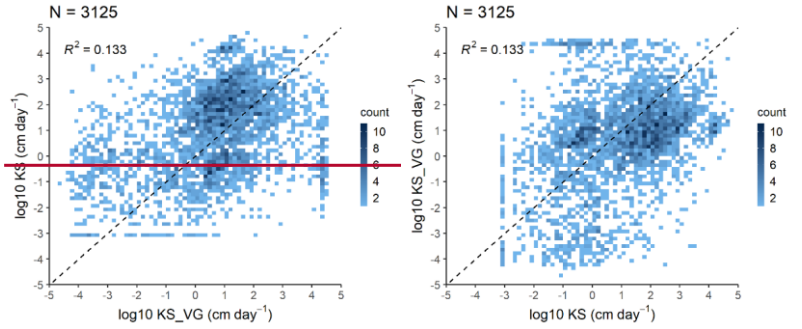


Figure 14. Scatterplot of measured saturated hydraulic conductivity (KS) versus saturated hydraulic conductivity computed from fitted VG parameters (KS_VG).

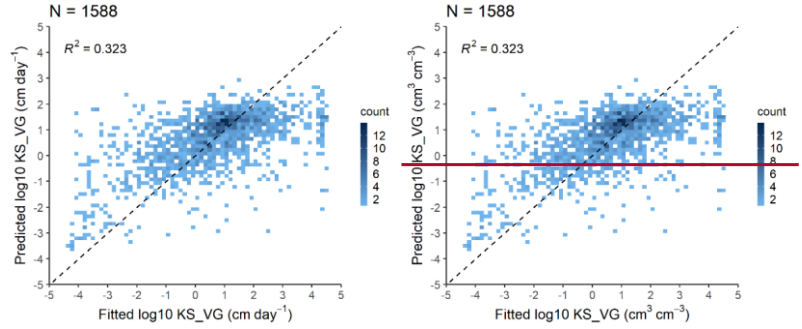


Figure 15. Scatterplot of saturated hydraulic conductivity computed from fitted and predicted VG parameters (KS_VG) analysed on the VG test set of the EU-HYDI dataset. Count: the number of cases in each quadrangle.

Table 14. Prediction performance of saturated hydraulic conductivity (cm day^{-1}) computed by pedotransfer function on the VG test set of the EU-HYDI dataset. N: number of samples, ME: mean error, MAE: mean absolute error, RMSE: root mean squared error, NSE: Nash-Sutcliffe efficiency, R^2: coefficient of determination.

Approach to predict KS*	N	ME	MAE	RMSE	NSE	R^2
log10pred_KS_VG	1591	-0.06	1.07	1.48	0.303	0.307

*log10pred_KS_VG: logarithmic 10 based saturated hydraulic conductivity computed based on VG parameters predicted from basic soil properties.

3.8 Phosphorus content of the topsoil

Figure 16 shows the European P map (Ballabio et al., 2019) clipped for the area of the Felső-Válicka study site (A) and the P map created with the mean statistics-based method using the local land use map (B) and the map of the hydrological response units (HRU) defined in the SWAT+ model (C). The spatial pattern of the two phosphorus maps is similar, but the map created with our proposed method has a higher resolution and follows the polygons of the HRU map.

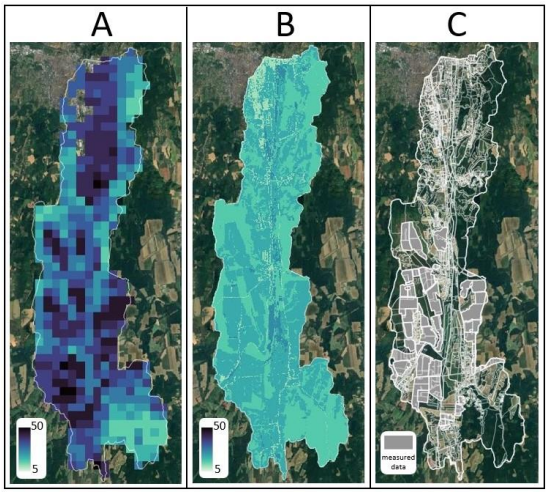


Figure 16. European topsoil P content map (Ballabio et al., 2019) (A), region-specific mean statistics-based P content map (B), hydrological response units with indication of agricultural parcels with measured P values (C) in the Felső-Válicka case study.

Figure 17 shows the geometric mean P values of the HRUs by land use categories of the European soil P map and the region-specific mean statistics-based P map in the area of Felső-Válicka. Comparing the results of the geometric mean P values, we can see that the European topsoil P map on average has a higher P concentration, with no significant differences observed between the land use categories. Based on the region specific LUCAS Topsoil dataset, artificial land use areas (urban fabric and industrial, commercial and transport units), forests and pastures are expected to have lower P concentration values. The mean statistics-based P map is more suitable at identifying differences resulting from local land use variation in the analysed case study. The P monitoring data measured on the 34 agricultural parcels, classified as arable land ~~and provided by a local agricultural company,~~ shows that the geometric mean of Olsen P in the area is 24 mg kg^{-1}, which is slightly higher than predicted by the mean statistics-based method (19.78 mg kg^{-1}).

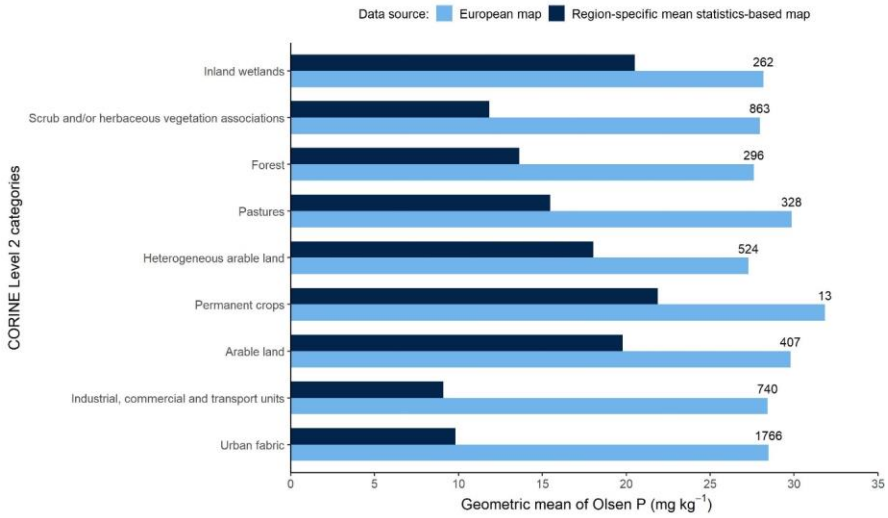


Figure 17. Geometric mean values of Olsen P across CORINE Level 2 land cover categories in the Felső-Válicka case study for both the European topsoil P content map and the region-specific mean statistics-based P content map with number of samples by categories indicated.

Ballabio et al. (2019) found that land use was the most important predictor for computing the topsoil phosphorus content map for Europe. This underscores that a soil P content map derived based on a local, fine-resolution, field-boundary-based land use map can provide more accurate results than one based on continental land use maps. ~~In summary, a P map based on local land~~

~~use information can result in more accurate estimations and higher resolution, thereby increasing its applicability in catchment-scale models, especially when defining hydrological response units based on land use data.~~

For regional or local studies, it is more plausible to use a local land use map and compute the geometric mean soil P values by land use categories based on the LUCAS Topsoil dataset, which is relevant for the target area from a fertilization point of view. ~~If~~ Where available, it is recommended to use measured data to overwrite the geometric mean values, ~~with the measured data~~creating a multi-data source solution~~,~~ that reflect~~ings~~ the spatial pattern of nutrient content within arable land areas. For continental-scale studies, the European topsoil P map (Ballabio et al., 2019) could be used.

3.9 Suggested workflow to derive soil input parameters

~~**2.2.1 Basic properties**~~

Based on the above results, we describe the most efficient workflow to retrieve the soil input parameters for European environmental modelling.

Initially, the data source of the most relevant soil basic properties, such as soil layering, rooting depth, organic carbon content, clay, silt, and sand content, must be selected. Local data can describe the spatial variability of soil properties the best. Even if only soil basic properties are available locally, this data source could have priority against the more inclusive continental or global datasets, i.e. containing information on both soil physical, chemical, and hydraulic properties, because local datasets aim to capture the area-specific variability of soil properties as accurately as possible. If no local or national soil basic data is available with the resolution required to study a target environmental process, possible input source for soil profile data or 3D soil dataset can be found in Table 1.

Different countries and institutions measure sand, silt, and clay content using different ISO protocols and methods ~~–~~by recognizing different cutoff limits and classification standards. It is important to check which particle size limits are required by the environmental model. As an example, in the widely used SWAT/SWAT+ model, the sand, silt and clay content are assumed to be classified according to the USDA system, which defines particle size limit < 0.002 mm for clay, 0.002-0.05 mm for silt and 0.05-2 mm for sand fraction. When conversions between different classifications are required to bring the local dataset to the appropriate format, it is advised to apply the k-nearest neighbour interpolation (formerly called: 'similarity technique'), which results in less uncertainty, smaller bias and shrinkage of resulting texture range compared to the simpler loglinear interpolation (Nemes et al., 1999).

In other cases, such as soil organic material, it is important to distinguish if soil organic carbon or soil organic matter is required by the model, and which of the two is available from the data source. The following most frequently used equation describes the relationship between those:

$$OM = OC \cdot 1.724 \quad (21)$$

where OM is the organic matter content (mass %) and OC is organic carbon content (mass %). The 1.724 conversion factor was defined by Van Bemmelen (1890), but can vary between 1.4 and 2.5 depending on the method used to measure organic

carbon content, composition of organic matter, degree of decomposition and clay content (Minasny et al., 2020; Pribyl, 2010). Pribyl (2010) recommends using the value 2 as a general conversion factor if no specific value is available.

When specifying bulk density, it is important to clarify whether the dry or effective value is required. If a measured value of neither is available, the dry bulk density can be computed from organic carbon content and particle size distribution. Further predictors, such as taxonomical information, soil structure, soil management parameters, environmental covariates are important as well (Hollis et al., 2012; Ramcharan et al., 2017) and can significantly improve the prediction performance. However, PTFs including these variables are not always possible to apply to a data scarce region.

If effective bulk density is required, it can be derived from the dry bulk density with the method of Wessolek et al. (2009):

- for soils with organic carbon content higher than 0.58 mass %:

$$BD_{eff} = BD_{dry} + 0.009 \cdot clay \quad (10\,22)$$

- for soils with organic carbon content less than or equal to 0.58 mass %:

$$BD_{eff} = BD_{dry} + 0.005 \cdot clay + 0.001 \cdot silt \quad (11\,23)$$

where BD_{eff} (g cm^{-3}) is effective bulk density, BD_{dry} (g cm^{-3}) is the dry bulk density, clay is clay content (< 0.002 mm, mass %), silt is silt content (0.002-0.063 mm, mass %). It is important to note that Eq. (11 23) requires the silt content with 0.002-0.063 mm limit. It can be predicted from the clay (< 0.002 mm), silt (0.002 – 0.05 mm) and sand (0.05 – 2 mm) content with the TT.text.trans function of the soiltexture R package (Moeys, 2018). This method meets the accuracy required for computing BD_{eff}, however, for other applications the transformation methods discussed by Nemes et. al. (1999) should be considered.

The hydrologic soil groups (HSG) are based on the infiltration characteristic of the soil and include four groups having similar runoff potential. The groups are defined based on the saturated hydraulic conductivity, depth to high water table and depth to water impermeable layer (Table 15). More details can be found in U.S. Department of Agriculture Natural Resources Conservation Service (2009).

For modelling purposes, it is important if tile drainage is present in the modelled area, because tile drainage systems influence the soil infiltration rate and runoff potential. Derivation of HSG requires local input data. If local datasets are not available, and SoilGrids 2017 (Hengl et al., 2017) was chosen as the source for the basic soil data, HSG can be retrieved from the global HYSOGs250m (Ross et al., 2018) database.

Table 15. Definition of soil hydrologic groups based on U.S. Department of Agriculture Natural Resources Conservation Service (2009). KS: saturated hydraulic conductivity (μm s^{-1}).

Depth to water impermeable layer*	Depth to high water table**	KS of least transmissive layer in depth range (μm s^{-1})	KS depth range	HSG***
≤50 cm	—	—	—	D
50 to 100 cm	<60 cm	>40.0	0 to 60 cm	A/D
		>10.0 to ≤40.0	0 to 60 cm	B/D
		>1.0 to ≤10.0	0 to 60 cm	C/D
		≤1.0	0 to 60 cm	D
	≥60 cm	>40.0	0 to 50 cm	A
		>10.0 to ≤40.0	0 to 50 cm	B
		>1.0 to ≤10.0	0 to 50 cm	C
		≤1.0	0 to 50 cm	D
>100 cm	<60 cm	>10.0	0 to 100 cm	A/D
		>4.0 to ≤10.0	0 to 100 cm	B/D
		>0.40 to ≤4.0	0 to 100 cm	C/D
		≤0.40	0 to 100 cm	D
	60 to 100 cm	>40.0	0 to 50 cm	A
		>10.0 to ≤40.0	0 to 50 cm	B
		>1.0 to ≤10.0	0 to 50 cm	C
		≤1.0	0 to 50 cm	D
	>100 cm	>10.0	0 to 100 cm	A
		>4.0 to ≤ 10.0	0 to 100 cm	B
		>0.40 to ≤4.0	0 to 100 cm	C
		≤0.40	0 to 100 cm	D

*An impermeable layer has a KS less than 0.01 μm s^{-1} [0.0014 in h^{-1}] or a component restriction of fragipan; duripan; petrocalcic; orstein; petrogypsic; cemented horizon; densic material; placic; bedrock, paralithic; bedrock, lithic; bedrock, densic; or permafrost. **High water table during any month during the year. ***Dual HSG classes are applied only for wet soils (water table less than 60 cm [24 in]). If these soils can be drained, a less restrictive HSG can be assigned, depending on the KS.

Figures 18-2~~1~~2 summarize the workflows to derive soil physical, hydraulic, and chemical parameters covered in this study. The workflows highlight the target soil property, necessary input, computation approach with suggested order of the computations. Indirect initialization of soil mineral N is recommended via proper management data and model warm-up period. It is important to highlight that prediction approaches trained on local data are expected to be more accurate; therefore, those could replace the indicated methods where possible.

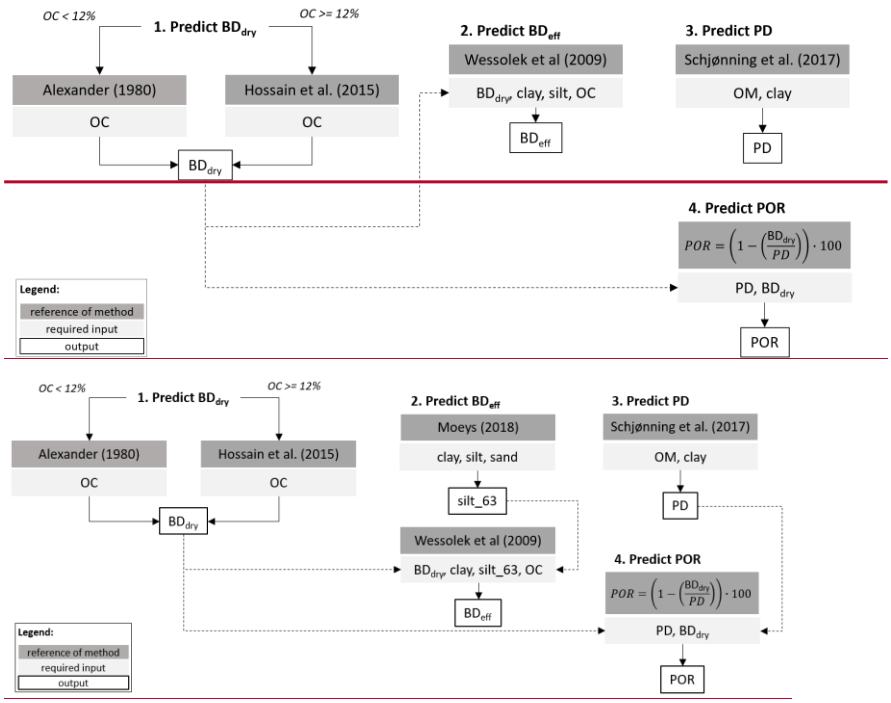


Figure 18. Prediction of soil physical properties. BD_{dry}: dry bulk density; clay: clay content (0-0.002 mm); silt: silt content (0.002-0.05 mm); sand: sand content (0.05-2 mm); silt_63: silt content (0.002-0.063 mm) OC: organic carbon content; BD_{eff}: effective bulk density; PD: particle density; POR: porosity.

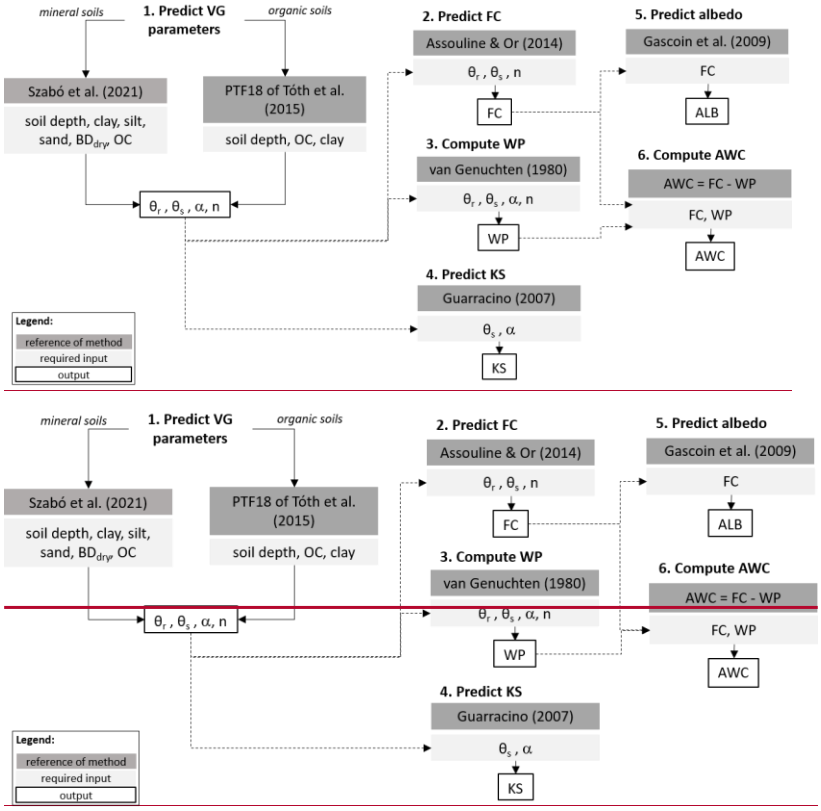


Figure 19. Prediction of soil hydraulic properties and moist soil albedo. Soil depth: mean soil depth of the soil sample; clay: clay content (0-0.002 mm); silt: silt content (0.002-0.05 mm); sand: sand content (0.05-2 mm); BD_{dry}: dry bulk density; OC: organic carbon content; θ_r : residual water content; θ_s: saturated soil water content; α : scale parameter; n: shape parameter; FC: water content at field capacity; WP: water content at wilting point; KS: saturated hydraulic conductivity; AWC: available water capacity; ALB: soil albedo.

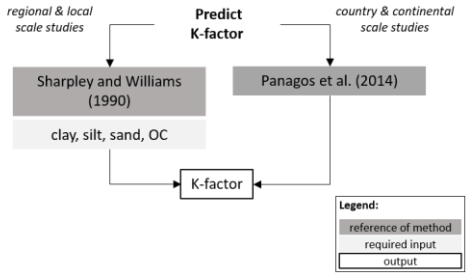


Figure 20. Prediction of soil erodibility factor (K-factor). clay content (0-0.002 mm); silt: silt content (0.002-0.05 mm); sand: sand content (0.05-2 mm); OC: organic carbon content.

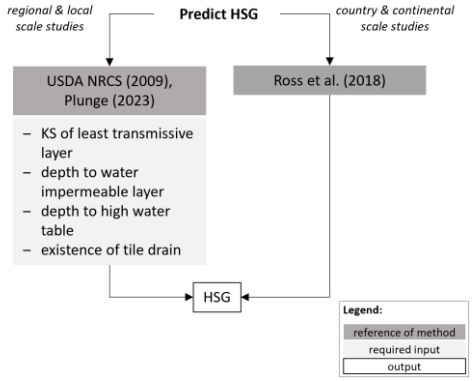


Figure 2~~0~~1. Prediction of hydraulic soil groups (HSG). KS: saturated hydraulic conductivity.

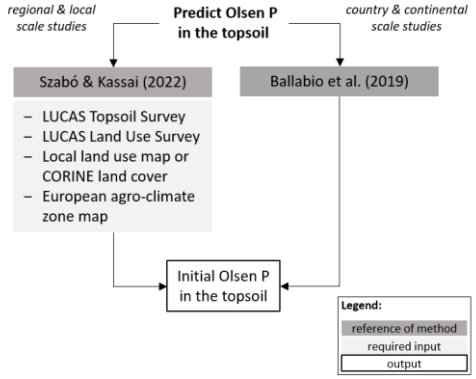


Figure 2~~1~~2. Prediction of Olsen phosphorus (P) content of the topsoil.

4 Conclusions

This study presents particular techniques and resources for extracting region-specific soil characteristics from national and global databases. While these databases might contain segments of soil information, they often lack comprehensive data required by various environmental models, such as the SWAT+ model. Through evaluation and recommendation of selected PTFs, as well as the provision of compiled R scripts for estimation solutions addressing soil data gaps, the study aims to streamline input data preparation procedures for soil physical, hydraulic, and chemical properties in environmental modelling. ~~Key findings underscored the significance of local soil data over global or large-scale datasets in environmental modelling.~~ Local data tend to retain finer soil details, hence it is recommended that users prioritise the utilisation of local (national) soil databases when it is deemed representative and reliable~~whenever possible~~. Even if these databases only offer basic soil properties, they should take precedence over broader continental or global datasets. The study demonstrated that missing soil properties could be estimated effectively from a basic set of soil parameters using appropriate PTFs developed for specific pedoclimatic regions, ensuring consistency in computed properties.

-We prepared a set of workflows to derive soil input parameters for usage in various modelling studies. In cases where this approach is not viable, we offer comprehensive guidance on alternative soil databases, outlining strategies to derive the absent soil properties effectively.

When using any available soil dataset, it is important to check the detailed description (metadata) of the dataset to avoid misinterpretation and errors in the models. Considerations such as consistent size limits for clay, silt, and sand content classification as per the model's requirements, distinction between organic carbon and organic matter, the need for dry or moist

bulk density, and similar details are vital. Understanding whether the model derives soil properties that already exist in the dataset is essential, aiding in selecting the most precise parameters for the model's application.

When retrieving or deriving missing soil input data, it is crucial to consider: i) which dataset and prediction approaches can offer physically plausible soil input data, and ii) the uncertainty associated with the derived soil input data for their appropriate use in the environmental model. For computing physically consistent soil hydraulic property values, namely FC, WP, AWC, and KS, it is plausible to use parameters of a model that describes soil water retention across the entire matric potential range. The parameters of the VG model have been widely employed to derive water retention at specific matric potential values or KS, hence can be used to derive physically plausible soil hydraulic properties. The static definition of FC could be replaced with a dynamic one that considers a soil-specific matric potential at which the continuity of soil water is reduced or disrupted. For the computation of the drainage dynamics-based AWC, the use of the VG model parameters is required for deriving both FC and WP. When computing FC, WP, AWC, and KS using the predicted VG parameters, we maintain the physical relationships among them, which is highly relevant in process-based modelling applications. Misuse of these parameters could lead to flawed model outcomes, impacting policy-making and agricultural management decisions.

It is important to note that soil parameter uncertainty encompasses not only the uncertainty of the PTF but also that stemming from the fitting of the VG model. The prediction uncertainty of soil properties varies significantly. It is essential to tailor its treatment based on the specificities of the target environmental model, particularly when it is utilized as an initial static value, in model calibration, or as a fixed input parameter.

The research emphasized the challenge of selecting suitable datasets and PTFs due to their abundance, providing quantitative performance metrics to aid potential environmental modellers. The workflows and findings presented in the study offer practical guidance for model setup and data preprocessing in various modelling endeavours across Europe, such as hydrological simulations, assessment of soil health, land evaluation, crop modelling, and analysis of soil erosion risk among others. The study's methodology can be ~~replicated~~ applied for soil databases not only in Europe but also in other regions or global datasets, highlighting its potential for broader applicability in multiple modelling contexts worldwide. We encourage the wider scientific and modelling community to use and adopt our recommended workflows to derive soil input parameters, bridging gaps in data for broader utilisation in diverse modelling studies. The presented workflows could be further improved by using a multi-model approach and applying geostatistical methods. The open-source library is available (see Code availability section) for use and adoption to meet the user-specific need.

Code availability

The `get_usersoil_table()` function in the R package SWATprepR (Plunge, 2023) was developed for this study. It facilitates the calculation of multiple soil parameters using PTF methods presented in the article. The functionality requires information on soil depth, sand, silt, clay, and organic matter content. The functions use the input information and calculates other soil parameters required for the SWAT+ model. The derivation of HSG is optional based on the KS of the least transmissive layer,

depth to water impermeable layer, depth to high water table and information on the existence of any tile drains. The entire package, source code, documentation, and installation instructions are openly accessible on the GitHub repository: https://github.com/biopsichas/SWATprepR/ .

Data availability

6,583 samples of 1999 soil profiles, summing up to 35 % of the EU-HYDI dataset, are available upon request from the European Soil Data Centre (ESDAC) at the European Commission Joint Research Centre. The entire dataset cannot be made publicly available due to its legal restrictions.~~The EU-HYDI dataset cannot be made publicly available due to its legal restrictions.~~ LUCAS TOPSOIL data can be accessed through ESDAC (European Commission Joint Research Centre, 2024; Panagos et al., 2012, 2022). Local measured topsoil phosphorus data is private, only results of analysis and derived information can be published.

Supplement link

Supplement is attached, the link will be added by Copernicus.

Author contributions

Conceptualization: BS, AN, NC, FW, MS. Data curation: BS, KP, JM**,** SP. Formal analysis: BS, KP, JM, MS, AN. Funding acquisition: FW, MS. Methodology: BS, KP, NC, AN, FW, SP, MS, BP, JM. Project administration: FW. Software: BS, SP, KP, JM. Supervision: BS. Validation: BS, KP, AN, NC, JM, SP, MS, FW. Visualization: BS, KP. BS prepared the manuscript with contributions from all co-authors. All authors reviewed the manuscript.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862756, project OPTAIN (OPtimal strategies to retAIN and re-use water and nutrients in small agricultural catchments across different soil-climatic regions in Europe) and the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA).

Financial support

This research has been supported by the Horizon 2020 Framework Programme (grant no. 862756) and Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences.

References

Abbaspour, K. C., AshrafVaghefi, S., Yang, H. and Srinivasan, R.: Global soil, landuse, evapotranspiration, historical and future weather databases for SWAT Applications, Sci. Data, 6:263, doi:https://doi.org/10.1038/s41597-019-0282-4, 2019.

Alexander, E. B.: Bulk Densities of California Soils in Relation to Other Soil Properties, Soil Sci. Soc. Am. J., 44(4), 689–692, doi:10.2136/sssaj1980.03615995004400040005x, 1980.

Amorim, H. C. S., Hurtarte, L. C. C., Souza, I. F. and Zinn, Y. L.: C:N ratios of bulk soils and particle-size fractions: Global trends and major drivers, Geoderma, 425(March), 116026, doi:10.1016/j.geoderma.2022.116026, 2022.

Arnold, J. G., Kiniry, J. R., Srinivasan, R., Williams, J. R., Haney, E. B. and Neitsch, S. L.: Soil & Water Assessment Tool, Input/Output Documentation, , 654, 2012.

Assouline, S. and Or, D.: The concept of field capacity revisited: Defining intrinsic static and dynamic criteria for soil internal drainage dynamics, Water Resour. Res., 50, 4787–4802, doi:10.1002/2014WR015475, 2014.

Babaeian, E., Homaee, M., Vereecken, H., Montzka, C., Norouzi, A. A. and van Genuchten, M. T.: A Comparative Study of Multiple Approaches for Predicting the Soil–Water Retention Curve: Hyperspectral Information vs. Basic Soil Properties, Soil Sci. Soc. Am. J., 79(4), 1043–1058, doi:10.2136/sssaj2014.09.0355, 2015.

Ballabio, C., Panagos, P. and Monatanarella, L.: Mapping topsoil physical properties at European scale using the LUCAS database, Geoderma, 261, 110–123, doi:10.1016/j.geoderma.2015.07.006, 2016.

Ballabio, C., Lugato, E., Fernández-Ugalde, O., Orgiazzi, A., Jones, A., Borrelli, P., Montanarella, L. and Panagos, P.: Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression, Geoderma, 355(July), 113912, doi:10.1016/j.geoderma.2019.113912, 2019.

Batjes, N. H., Ribeiro, E. and Van Oostrum, A.: Standardised soil profile data to support global mapping and modelling (WoSIS snapshot 2019), Earth Syst. Sci. Data, 12(1), 299–320, doi:10.5194/essd-12-299-2020, 2020.

Van Bemmelen, J. M.: Über die Bestimmung des Wassers, des Humus, des Schwefels, der in den colloïdalen Silikaten gebundenen Kieselsäure, des Mangans usw im Ackerboden., Die Landwirthschaftlichen Versuchs-Stationen, 37, 279–290, 1980.

Bernoux, M., Arrouays, D., Cerri, C. C., Volkoff, B. and Jolivet, C.: Bulk Densities of Brazilian Amazon Soils Related to Other Soil Properties., Soil Sci. Soc. Am. J., 62, 743–749, 1998.

Børgesen, C. D. and Schaap, M. G.: Point and parameter pedotransfer functions for water retention predictions for Danish soils, Geoderma, 127(1–2), 154–167, doi:10.1016/j.geoderma.2004.11.025, 2005.

Bouma, J.: Using Soil Survey Data for Quantitative Land Evaluation, pp. 177–213, Springer US., 1989.

Bouma, J. and van Lanen, H. A. J.: Transfer functions and threshold values: from soil characteristics to land qualities., in Proceedings of the International Workshop on Quantified Land Evaluation Procedures, edited by K. J. Beek, P. A. Burrough, and D. E. MacCormack, pp. 106–110, International Institute for Aerospace Survey and Earth Sciences (ITC), Washington, DC. [online] Available from: http://library.wur.nl/WebQuery/wurpubs/4195 (Accessed 27 March 2017), 1987.

Carrer, D., Meurey, C., Ceamanos, X., Roujean, J. L., Calvet, J. C. and Liu, S.: Dynamic mapping of snow-free vegetation and bare soil albedos at global 1km scale from 10-year analysis of MODIS satellite products, Remote Sens. Environ., 140, 420–432, doi:10.1016/j.rse.2013.08.041, 2014.

Casanova, M., Tapia, E., Seguel, O. and Salazar, O.: Direct measurement and prediction of bulk density on alluvial soils of central Chile, Chil. J. Agric. Res., 76(1), 105–113, doi:10.4067/S0718-58392016000100015, 2016.

Ceglar, A., Zampieri, M., Toreti, A. and Dentener, F.: Observed Northward Migration of Agro-Climate Zones in Europe Will Further Accelerate Under Climate Change, Earth's Futur., 7(9), 1088–1101, doi:10.1029/2019EF001178, 2019.

Chen, S., Richer-de-Forges, A. C., Saby, N. P. A., Martin, M. P., Walter, C. and Arrouays, D.: Building a pedotransfer function for soil bulk density on regional dataset and testing its validity over a larger area, Geoderma, 312(June 2017), 52–63, doi:10.1016/j.geoderma.2017.10.009, 2018.

Copernicus Climate Change Service, C. D. S.: Surface albedo 10-daily gridded data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)., , doi:10.24381/cds.ea87ed30, 2018.

Dai, Y., Xin, Q., Wei, N., Zhang, Y., Shangguan, W., Yuan, H., Zhang, S., Liu, S. and Lu, X.: A Global High-Resolution Data Set of Soil Hydraulic and Thermal Properties for Land Surface Modeling, J. Adv. Model. Earth Syst., doi:10.1029/2019ms001784, 2019a.

Dai, Y., Shangguan, W., Wei, N., Xin, Q., Yuan, H., Zhang, S., Liu, S., Lu, X., Wang, D. and Yan, F.: A review of the global soil property maps for Earth system models, Soil, 5(2), 137–158, doi:10.5194/soil-5-137-2019, 2019b.

Dam, J. C. van, Groenendijk, P., Hendriks, R. F. A. and Kroes, J. G.: Advances of Modeling Water Flow in Variably Saturated Soils with SWAP, Vadose Zo. J., 7(2), 640–653, doi:10.2136/VZJ2007.0060, 2008.

Dang, N. A., Jackson, B. M., Tomscha, S. A., Lilburne, L., Burkhard, K., Tran, D. D., Phi, L. H. and Benavidez, R.: Guidelines and a supporting toolbox for parameterising key soil hydraulic properties in hydrological studies and broader integrated modelling, One Ecosyst., 7, doi:10.3897/ONEECO.7.E76410, 2022.

DHI: MIKE SHE User Guide and Reference Manual, Agern Alle, 816, 2023.

Egnér, H., Riehm, H. and Domingo, W. R.: Untersuchungen über die chemische Bodenanalyse als Grundlage für die Beurteilung des Nährstoffzustandes der Böden. II. Chemische Extraktionsmethoden zur Phosphor- und Kaliumbestimmung., Lantbr. Ann., 26, 199-215., 1960.

European Commission Joint Research Centre: European Soil Data Centre (ESDAC), [online] Available from: esdac.jrc.ec.europa.eu (Accessed 17 April 2024), 2024.

EUROSTAT: LUCAS 2015 (Land Use/Cover Area Frame Survey) – Technical Reference Document: C3 Classification (Land

Use and Land Cover). Technical Report., 2015.

FAO & IIASA: Harmonized World Soil Database version 2.0., edited by F. Nachtergaele, H. van Velthuizen, L. Verelst, D. Wiberg, M. Henry, F. Chiozza, and Y. Yigini, Food and Agricultural Organization of the United Nations and International Institute for Applied Systems Analysis, Rome and Laxenburg., 2023.

Fernandez-Ugalde, O., Scarpa, S., Orgiazzi, A., Panagos, P., Van Liedekerke, M., Marechal, A. and Jones, A.: LUCAS 2018 Soil Module. Presentation of dataset and results, Luxembourg., 2022.

Foster, G. R., McCool, D. K., Renard, K. G. and Moldenhauer, W. C.: Conversion of the universal soil loss equation to SI metric units., J. Soil Water Conserv., 36(6), 355–359, 1981.

Gascoin, S., Duchame, A., Ribstein, P., Perroy, E. and Wagnon, P.: Sensitivity of bare soil albedo to surface soil moisture on the moraine of the Zongo glacier (Bolivia), Geophys. Res. Lett., 36(2), 2–6, doi:10.1029/2008GL036377, 2009.

van Genuchten, M. T.: A closed-form equation for predicting the hydraulic conductivity of unsaturated soils, Soil Sci. Soc. Am. J., 44, 892–898, 1980.

Ghanbarian, B., Taslimitehrani, V. and Pachepsky, Y. A.: Accuracy of sample dimension-dependent pedotransfer functions in estimation of soil saturated hydraulic conductivity, Catena, 149, 374–380, doi:10.1016/j.catena.2016.10.015, 2017.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D. and Moore, R.: Google Earth Engine: Planetary-scale geospatial analysis for everyone, Remote Sens. Environ., 202, 18–27, doi:10.1016/j.rse.2017.06.031, 2017.

Guarracino, L.: Estimation of saturated hydraulic conductivity Ks from the van Genuchten shape parameter α, Water Resour. Res., 43(11), 15–18, doi:10.1029/2006WR005766, 2007.

Gupta, S., Hengl, T., Lehmann, P., Bonetti, S. and Papritz, A.: Global prediction of soil saturated hydraulic conductivity using random forest in a Covariate-based Geo Transfer Functions (CoGTF) framework, J. ofAdvances Model. Earth Syst., e2020MS002242, doi:https://doi.org/10.1029/2020MS002242, 2021a.

Gupta, S., Lehmann, P., Bonetti, S., Papritz, A. and Or, D.: Global Prediction of Soil Saturated Hydraulic Conductivity Using Random Forest in a Covariate-Based GeoTransfer Function (CoGTF) Framework, J. Adv. Model. Earth Syst., 13(4), 1–15, doi:10.1029/2020MS002242, 2021b.

Gupta, S., Papritz, A., Lehmann, P., Hengl, T., Bonetti, S. and Or, D.: Global Mapping of Soil Water Characteristics Parameters— Fusing Curated Data with Machine Learning and Environmental Covariates, Remote Sens., 14(8), doi:10.3390/rs14081947, 2022.

Gupta, S., Lehmann, P., Bickel, S., Bonetti, S. and Or, D.: Global Mapping of Potential and Climatic Plant-Available Soil Water, J. Adv. Model. Earth Syst., 15(11), 1–16, doi:10.1029/2022MS003277, 2023.

Gupta, S., Borrelli, P., Panagos, P. and Alewell, C.: An advanced global soil erodibility (K) assessment including the effects of saturated hydraulic conductivity, Sci. Total Environ., 908(July 2023), 168249, doi:10.1016/j.scitotenv.2023.168249, 2024.

Hannam, J. A., Hollis, J. M., Jones, R. J. A., Bellamy, P. H., Hayes, S. E., Holden, A., Van Liedekerke, M. H. and Montanarella, L.: SPADE-2: the soil profile analytical database for Europe, version 2.0 Beta Version, Unpubl. Eur. Soil Bur. Res. Rep.,

(March), 1–27 [online] Available from: http://eusoils.jrc.ec.europa.eu/Esdb_Archive/eusoils_docs/esb_rr/SPADE-2_Beta_Report.pdf, 2009.

Hansen, S., Abrahamsen, P., Petersen, C. T. and Styczen, M.: Daisy: Model Use, Calibration, and Validation, Trans. ASABE, 55(4), 1317–1333, doi:10.13031/2013.42244, 2012.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, edited by B. Bond-Lamberty, PLoS One, 12(2), e0169748, doi:10.1371/journal.pone.0169748, 2017.

Hollis, J. M., Hannam, J. and Bellamy, P. H.: Empirically derived pedotransfer functions for predicting bulk density in European soils, Eur. J. Soil Sci., 63, 96–109, 2012.

Hossain, M. F., Chen, W. and Zhang, Y.: Bulk density of mineral and organic soils in the Canada's arctic and sub-arctic, Inf. Process. Agric., 2(3–4), 183–190, doi:10.1016/j.inpa.2015.09.001, 2015.

IUSS Working Group WRB: World Reference Base for Soil Resources. International soil classification system for naming soils and creating legends for soil maps. 4th edition., International Union of Soil Sciences (IUSS), Vienna, Austria., 2022.

Kinnell, P. I. A.: Event soil loss, runoff and the Universal Soil Loss Equation family of models: A review, J. Hydrol., 385(1–4), 384–397, doi:10.1016/j.jhydrol.2010.01.024, 2010.

Kosugi, K.: Lognormal Distribution Model for Unsaturated Soil Hydraulic Properties, Water Resour. Res., 32(9), 2697–2703, doi:10.1029/96WR01776, 1996.

Krevh, V., Filipović, L., Petošić, D., Mustać, I., Bogunović, I., Butorac, J., Kisić, I., Defterdarović, J., Nakić, Z., Kovač, Z., Pereira, P., He, H., Chen, R., Toor, G. S., Versini, A., Baumgartl, T. and Filipović, V.: Long-term analysis of soil water regime and nitrate dynamics at agricultural experimental site: Field-scale monitoring and numerical modeling using HYDRUS-1D, Agric. Water Manag., 275(September 2022), doi:10.1016/j.agwat.2022.108039, 2023.

Kutílek, M. and Nielsen, D. R.: Soil hydrology, Catena-Verlag., 1994.

Lal, R. and Shukla, M. K.: Principles of soil physics, Marcel Dekker, Inc., New York., 2004.

Li, S. and Duffy, C. J.: Fully coupled approach to modeling shallow water flow, sediment transport, and bed evolution in rivers, Water Resour. Res., 47(3), 1–20, doi:10.1029/2010WR009751, 2011.

Liang, K., Zhang, X., Liang, X.-Z., Jin, V. L., Birru, G., Schmer, M. R., Robertson, G. P., McCarty, G. W. and Moglen, G. E.: Simulating agroecosystem soil inorganic nitrogen dynamics under long-term management with an improved SWAT-C model, Sci. Total Environ., 879, 162906, doi:https://doi.org/10.1016/j.scitotenv.2023.162906, 2023.

Lilly, A., Nemes, A., Rawls, W. J. and Pachepsky, Y. a: Probabilistic approach to the identification of input variables to estimate hydraulic conductivity, Soil Sci. Soc. Am. J., 72(1), 16–24, doi:10.2136/sssaj2006.0391, 2008.

Liu, M., Ussiri, D. A. N. and Lal, R.: Soil Organic Carbon and Nitrogen Fractions under Different Land Uses and Tillage Practices, Commun. Soil Sci. Plant Anal., 47(12), 1528–1541, doi:10.1080/00103624.2016.1194993, 2016.

Van Looy, K., Bouma, J., Herbst, M., Koestel, J., Minasny, B., Mishra, U., Montzka, C., Nemes, A., Pachepsky, Y. A.,

Padarian, J., Schaap, M. G., Tóth, B., Verhoef, A., Vanderborght, J., van der Ploeg, M. J., Weihermüller, L., Zacharias, S., Zhang, Y. and Vereecken, H.: Pedotransfer Functions in Earth System Science: Challenges and Perspectives, Rev. Geophys., 55(4), 1199–1256, doi:10.1002/2017RG000581, 2017.

López-Ballesteros, A., Nielsen, A., Castellanos-Osorio, G., Trolle, D. and Senent-Aparicio, J.: DSOLMap, a novel high-resolution global digital soil property map for the SWAT + model: Development and hydrological evaluation, Catena, 231(June), doi:10.1016/j.catena.2023.107339, 2023.

Manrique, L. A. and Jones, C. A.: Bulk Density of Soils in Relation to Soil Physical and Chemical Properties, Soil Sci. Soc. Am. J., 55, 476–481, doi:https://doi.org/10.2136/sssaj1991.03615995005500020030x, 1991.

Minasny, B., McBratney, A. B., Wadoux, A. M. J. C., Akoeb, E. N. and Sabrina, T.: Precocious 19th century soil carbon science, Geoderma Reg., 22(November 1830), e00306, doi:10.1016/j.geodrs.2020.e00306, 2020.

Moeys, J.: soiltexture: Functions for Soil Texture Plot, Classification and Transformation., [online] Available from: https://cran.r-project.org/package=soiltexture, 2018.

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A. and Vereecken, H.: A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves, Earth Syst. Sci. Data, 9(2), 529–543, doi:10.5194/essd-9-529-2017, 2017.

Nasta, P., Szabó, B. and Romano, N.: Evaluation of pedotransfer functions for predicting soil hydraulic properties : A voyage from regional to field scales across Europe, J. Hydrol. Reg. Stud., 37(February), 100903, doi:10.1016/j.ejrh.2021.100903, 2021.

Neitsch, S. L., Arnold, J. G., Kiniry, J. R. and Williams, J. R.: Soil and Water Assessment Tool Theoretical Documentation—Version 2009., , 618, 2009.

Nemes, A., Wösten, J., Lilly, A. and Voshaar, J. O.: Evaluation of different procedures to interpolate particle-size distributions to achieve compatibility within soil databases, Geoderma, 90, 187–202 [online] Available from: http://www.sciencedirect.com/science/article/pii/S0016706199000142 (Accessed 7 May 2013), 1999.

Nimmo, J. R.: Porosity and Pore-Size Distribution, in Encyclopedia of Soils in the Environment, vol. 3, edited by D. Hillel, pp. 295–303, Elsevier, London., 2004.

Olsen, R., Cole, C. V., Watanabe, F. S. and Dean, L. A.: Estimation of available phosphorus in soils by extraction with sodium bicarbonate., Washington DC., 1954.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A. and Fernández-Ugalde, O.: LUCAS Soil, the largest expandable soil dataset for Europe: a review, Eur. J. Soil Sci., 69(1), 140–153, doi:10.1111/ejss.12499, 2018.

Palladino, M., Romano, N., Pasolli, E. and Nasta, P.: Developing pedotransfer functions for predicting soil bulk density in Campania, Geoderma, 412(January), 115726, doi:10.1016/j.geoderma.2022.115726, 2022.

Panagos, P., Van Liedekerke, M., Jones, A. and Montanarella, L.: European Soil Data Centre: Response to European policy support and public data requirements, Land use policy, 29(2), 329–338, doi:10.1016/j.landusepol.2011.07.003, 2012.

Panagos, P., Meusburger, K., Ballabio, C., Borrelli, P. and Alewell, C.: Soil erodibility in Europe: A high-resolution dataset

based on LUCAS, Sci. Total Environ., 479–480(1), 189–200, doi:10.1016/j.scitotenv.2014.02.010, 2014.

Panagos, P., Van Liedekerke, M., Borrelli, P., Köninger, J., Ballabio, C., Orgiazzi, A., Lugato, E., Liakos, L., Hervas, J., Jones, A. and Montanarella, L.: European Soil Data Centre 2.0: Soil data and knowledge in support of the EU policies, Eur. J. Soil Sci., 73(6), 1–18, doi:10.1111/ejss.13315, 2022.

Panagos, P., De Rosa, D., Liakos, L., Labouyrie, M., Borrelli, P. and Ballabio, C.: Soil bulk density assessment in Europe, Agric. Ecosyst. Environ., 364(October 2023), 108907, doi:10.1016/j.agee.2024.108907, 2024.

Plunge, S.: SWATprepR: SWAT+ Input Data Preparation Package, , doi:10.5281/zenodo.10167076, 2023.

Poggio, L., De Sousa, L. M., Batjes, N. H., Heuvelink, G. B. M., Kempen, B., Ribeiro, E. and Rossiter, D.: SoilGrids 2.0: Producing soil information for the globe with quantified spatial uncertainty, SOIL, 7(1), 217–240, doi:10.5194/SOIL-7-217-2021, 2021.

Pribyl, D. W.: A critical review of the conventional SOC to SOM conversion factor, Geoderma, 156(3–4), 75–83, doi:10.1016/j.geoderma.2010.02.003, 2010.

Pu, X., Cheng, H., Shan, Y., Chen, S., Ding, Z. and Hao, F.: Factor controlling soil organic carbon and total nitrogen dynamics under long-term conventional cultivation in seasonally frozen soils, Acta Agric. Scand. Sect. B Soil Plant Sci., 62(8), 749–764, doi:10.1080/09064710.2012.700318, 2012.

R Core Team: R: A Language and Environment for Statistical Computing, [online] Available from: https://www.r-project.org/, 2022.

Ramcharan, A., Hengl, T., Beaudette, D. and Wills, S.: A Soil Bulk Density Pedotransfer Function Based on Machine Learning: A Case Study with the NCSS Soil Characterization Database, Soil Sci. Soc. Am. J., 81(0), 1279–1287, doi:10.2136/sssaj2016.12.0421, 2017.

Rawls, W. J.: Estimating soil bulk density from particle size analysis and organic matter content, Soil Sci., 135(2), 123–125 [online] Available from: https://journals.lww.com/soilsci/Fulltext/1983/02000/ESTIMATING_SOIL_BULK_DENSITY_FROM_PARTICLE_SIZE.7.aspx, 1983.

Renard, K. G., Foster, G. R., Weesies, G. A., McCool, D. K. and Yoder, D. C.: Predicting soil erosion by water: a guide to conservation planning with the Revised Universal Soil Loss Equation (RUSLE), U.S. Department of Agriculture. [online] Available from: http://www.ars.usda.gov/SP2UserFiles/Place/64080530/RUSLE/AH_703.pdf, 1997.

Román Dobarco, M., Cousin, I., Le Bas, C. and Martin, M. P.: Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty, Geoderma, 336(April 2018), 81–95, doi:10.1016/J.GEODERMA.2018.08.022, 2019.

Romano, N., Szabó, B., Belmonte, A., Castrignano, A., Ben Dor, E., Francos, N. and Nasta, P.: Mapping soil properties for unmanned aerial system-based environmental monitoring, in Unmanned Aerial Systems for Monitoring Soil, Vegetation, and Riverine Environments, edited by S. Manfreda and E. Ben Dor, pp. 155–178, Elsevier., 2023.

Ross, C. W., Prihodko, L., Anchang, J., Kumar, S., Ji, W. and Hanan, N. P.: HYSOGs250m, global gridded hydrologic soil

groups for curve-number-based runoff modeling, Sci. Data, 5(150091), doi:10.1038/sdata.2018.91, 2018.

Ruehlmann, J.: Soil particle density as affected by soil texture and soil organic matter: 1. Partitioning of SOM in conceptional fractions and derivation of a variable SOC to SOM conversion factor, Geoderma, 375(June), 114542, doi:10.1016/j.geoderma.2020.114542, 2020.

Safanelli, J. L., Chabrillat, S., Ben-Dor, E. and Demattê, J. A. M.: Multispectral Models from Bare Soil Composites for Mapping Topsoil Properties over Europe., Remote Sens., 12, 1369., doi:10.3390/rs12091369, 2020.

Sárdi, K., Csathó, P. and Osztoics, E.: Evaluation of soil phosphorus contents in long-term experiments from environmental aspects., in Proceedings of the 51st Georgikon Scientific Conference, pp. 807-815., Keszthely, Hungary., 2009.

Schjønning, P., McBride, R. A., Keller, T. and Obour, P. B.: Predicting soil particle density from clay and soil organic matter contents, Geoderma, 286, 83–87, doi:10.1016/j.geoderma.2016.10.020, 2017.

Shangguan, W., Dai, Y., Duan, Q., Liu, B. and Yuan, H.: A global soil data set for earth system modeling, J. Adv. Model. Earth Syst., 6(1), 249–263, doi:10.1002/2013MS000293, 2014.

Sharpley, A. N. and Williams, J. R.: EPIC — Erosion / Productivity Impact Calculator: 1. Model Documentation., 1990.

Šimůnek, J., Van Genuchten, M. T. and Šejna, M.: Hydrus: Model use, calibration, and validation, Trans. ASABE, 55(4), 1261–1274, 2012.

Souza, E. De, Batjes, N. H. and Pontes, L. M.: Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin, Sci. Agric., 73(6), 525–534, doi:10.1590/0103-9016-2015-0485, 2016.

Steinfurth, K., Hirte, J., Morel, C. and Buczko, U.: Conversion equations between Olsen-P and other methods used to assess plant available soil phosphorus in Europe – A review, Geoderma, 401, 115339, doi:10.1016/j.geoderma.2021.115339, 2021.

Szabó, B. and Kassai, P.: Map topsoil phosphorus content (v1.0.1)., 2022.

Szabó, B., Weynants, M. and Weber, T. K.: Updated European Hydraulic Pedotransfer Functions with Communicated Uncertainties in the Predicted Variables (euptfv2), Geosci. Model Dev., 14, 151–175, doi:10.5194/gmd-14-151-2021, 2021.

Tomasella, J., Crestana, S. and Rawls, W. J.: Comparison of Two Techniques to Develop Pedotransfer Functions for Water Retention, Soil Sci. Soc. Am. J., 67, 1085–1092, 2003.

Tóth, B., Weynants, M., Nemes, A., Makó, A., Bilas, G. and Tóth, G.: New generation of hydraulic pedotransfer functions for Europe., Eur. J. Soil Sci., 66(1), 226–238, doi:10.1111/ejss.12192, 2015.

Tóth, B., Weynants, M., Pásztor, L. and Hengl, T.: 3D soil hydraulic database of Europe at 250 m resolution, Hydrol. Process., 31(14), 2662–2666, doi:10.1002/hyp.11203, 2017.

Tóth, G., Jones, A. and Montanarella, L.: LUCAS Topsoil Survey. Methodology, data and results., Publications Office of the European Union, Luxembourg:, 2013.

Tóth, G., Guicharnaud, R.-A., Tóth, B. and Hermann, T.: Phosphorus levels in croplands of the European Union with implications for P fertilizer use, Eur. J. Agron., 55, 42–52, doi:10.1016/j.eja.2013.12.008, 2014.

Tranter, G., McBratney, a. B. and Minasny, B.: Using distance metrics to determine the appropriate domain of pedotransfer function predictions, Geoderma, 149(3–4), 421–425, doi:10.1016/j.geoderma.2009.01.006, 2009.

U.S. Department of Agriculture Natural Resources Conservation Service: Part 630 Hydrology, Chapter 7 Hydrologic Soil Groups, in National Engineering Handbook., 2009.

US Army Crops of Engineers: HEC-RAS Hydraulic Reference Manual, Version 6.4, , (August), 470, 2023.

Weber, T. K. D., Weihermüller, L., Nemes, A., Bechtold, M., Degré, A., Diamantopoulos, E., Fatichi, S., Filipović, V., Gupta, S., Hohenbrink, T. L., Hirmas, D. R., Jackisch, C., de Jong van Lier, Q., Koestel, J., Lehmann, P., Marthews, T. R., Minasny, B., Pagel, H., van der Ploeg, M., Svane, S. F., Szabó, B., Vereecken, H., Verhoef, A., Young, M., Zeng, Y., Zhang, Y. and Bonetti, S.: Hydro-pedotransfer functions: A roadmap for future development. [online] Available from: https://doi.org/10.5194/egusphere-2023-1860, 2023.

Wessolek, G., Kaupenjohann, M. and Renger, M.: Bodenphysikalische Kennwerte und Berechnungsverfahren für die Praxis, Bodenökologie und Bodengenese, 40, 1–80 [online] Available from: https://www.boden.tu-berlin.de/fileadmin/fg77/_pdf/Rote_Liste/Rote_Reihe_Heft_40.pdf, 2009.

Weynants, M., Montanarella, L., Tóth, G., Arnoldussen, A., Anaya Romero, M., Bilas, G., Borresen, T., Cornelis, W., Daroussin, J., Gonçalves, M. D. C., Haugen, L.-E., Hennings, V., Houskova, B., Iovino, M., Javaux, M., Keay, C. A., Kätterer, T., Kvaerno, S., Laktinova, T., Lamorski, K., Lilly, A., Mako, A., Matula, S., Morari, F., Nemes, A., Patyka, N. V., Romano, N., Schindler, U., Shein, E., Slawinski, C., Strauss, P., Tóth, B. and Woesten, H.: European HYdropedological Data Inventory (EU-HYDI), EUR – Scientific and Technical Research series – ISSN 1831-9424, Luxembourg., 2013.

Xiangsheng, Y., Guosheng, L. and Yanyu, Y.: Pedotransfer Functions for Estimating Soil Bulk Density: A Case Study in the Three-River Headwater Region of Qinghai Province, China, Pedosphere, 26(3), 362–373, doi:10.1016/S1002-0160(15)60049-2, 2016.

Xing, X., Nie, W., Chang, K., Zhao, L., Li, Y. and Ma, X.: A numerical approach for modeling crack closure and infiltrated flow in cracked soils, Soil Tillage Res., 233(5), 105794, doi:10.1016/j.still.2023.105794, 2023.

Yu, L., Zeng, Y. and Su, Z.: Understanding the mass, momentum, and energy transfer in the frozen soil with three levels of model complexities, Hydrol. Earth Syst. Sci., 24(10), 4813–4830, doi:10.5194/hess-24-4813-2020, 2020.

Yuan, Y. and Chiang, L.-C.: Sensitivity analysis of SWAT nitrogen simulations with and without in-stream processes, Arch. Agron. Soil Sci., 61(7), 969–987, doi:10.1080/03650340.2014.965694, 2015.

Zhai, X., Liu, K., Finch, D. M., Huang, Di., Tang, S., Li, S., Liu, H. and Wang, K.: Stoichiometric characteristics of different agroecosystems under the same climatic conditions in the agropastoral ecotone of northern China, Soil Res., 57(8), 875–882, doi:10.1071/SR18355, 2019.

Zhang, Y. and Schaap, M. G.: Weighted recalibration of the Rosetta pedotransfer model with improved estimates of hydraulic parameter distributions and summary statistics (Rosetta3), J. Hydrol., 547(January 2017), 39–53, doi:10.1016/j.jhydrol.2017.01.004, 2017.

Zhang, Y., Schaap, M. G. and Zha, Y.: A High-Resolution Global Map of Soil Hydraulic Properties Produced by a Hierarchical

Parameterization of a Physically Based Water Retention Model, Water Resour. Res., 54(12), 9774–9790, doi:10.1029/2018WR023539, 2018.

Zhang, Y., Schaap, M. G. and Wei, Z.: Development of Hierarchical Ensemble Model and Estimates of Soil Water Retention With Global Coverage, Geophys. Res. Lett., 47(15), 1–12, doi:10.1029/2020GL088819, 2020.

Zhu, Y., Chen, Y., Ali, M. A., Dong, L., Wang, X., Archontoulis, S. V., Schnable, J. C. and Castellano, M. J.: Continuous in situ soil nitrate sensors: The importance of high-resolution measurements across time and a comparison with salt extraction-based methods, Soil Sci. Soc. Am. J., 85(3), 677–690, doi:10.1002/saj2.20226, 2021.