

Advancing catchment-scale environmental modelling in Europe: Integrating open-source soil datasets and pedotransfer functions

Brigitta Szabó, Piroska Kassai, Svajūnas Plungė, Attila Nemes, Péter Braun,
Michael Strauch, Felix Witing, János Mészáros, Natalja Čerkasova



Aim of the study

- 1) Catalogue open-source soil datasets and pedotransfer functions (PTFs) applicable in simulation studies across European catchments.
- 2) Evaluate the performance of selected PTFs.
- 3) 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.

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

	Basic data	Soil hydraulic or physical data	Soil basic and hydraulic data
Point data	<ul style="list-style-type: none">– LUCAS Topsoil dataset– SPADE 2	<ul style="list-style-type: none">– EU-HYDI	<ul style="list-style-type: none">– WOSIS
Map	<ul style="list-style-type: none">– Topsoil chem. and physical properties for Europe– SoilGrids– OpenLandMap	<ul style="list-style-type: none">– EU-SoilHydroGrids– Motzka et al. (2017)– Zhang and Schaap (2018)– Zhang et al. (2020)– Gupta et al. (2022)– Gupta et al. (2021)	<ul style="list-style-type: none">– HWSD v 2.0– DSOLMap– GSDE (2014)– Dai et al. (2019)

The following data sites include most of the updates:

- European Soil Data Centre - soil datasets from Europe (<https://esdac.jrc.ec.europa.eu/>),
- ISRIC Soil Data Hub - 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 - open geographical and geoscientific data (<https://openlandmap.org>).

Analysed soil properties

Physical parameters

- bulk density
- porosity
- moist soil albedo
- soil erodibility factor

Hydraulic parameters

- water retention curve
- saturated hydraulic conductivity

Chemical parameters

- soil phosphorus content

Database with measured values – prediction performance

EU-HYDI

- 18,682 samples from 6,014 profiles
- used to derive euptfs

LUCAS Topsoil dataset

- 2009
- 2018: 5821 samples

Locally measured data on soil phosphorus content

- 2009: 34 agricultural parcels

Database used for comparison

MCD43A3 database

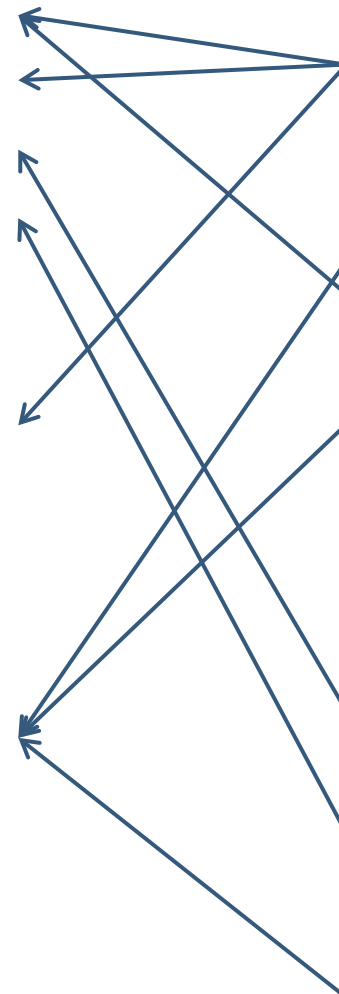
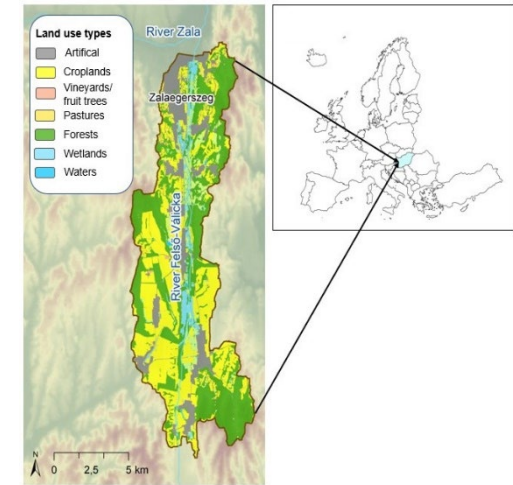
- 2022

European soil erodibility map

(Panagos et al., 2014)

European topsoil P content map

(Ballabio et al., 2019)



Bulk density

Name of the PTF	Equation	Reference
BD_Rawls	$BD = \frac{100}{\left(\frac{OM}{0.224}\right) + \frac{100 - OM}{1.27}}$	(Rawls, 1983)
BD_Alexander_A	$BD = 1.72 - 0.294 \cdot OC^{0.5}$	(Alexander, 1980)
BD_Alexander_B	$BD = 1.66 - 0.308 \cdot OC^{0.5}$	(Alexander, 1980)
BD_MAn_J_A	$BD = 1.510 - 0.113 \cdot OC$	(Manrique and Jones, 1991)
BD_MAn_J_B	$BD = 1.66 - 0.318 \cdot OC^{0.5}$	(Manrique and Jones, 1991)
BD_Hollis	<p>-for cultivated topsoils:</p> $BD = 0.80806 + (0.823844 \cdot (\exp(-0.27993 \cdot OC))) + 0.0014065 \cdot sand - 0.0010299 \cdot clay$ <p>- for mineral subsoils:</p> $BD = 0.69794 + (0.750636 \cdot (\exp(-0.230355 \cdot OC))) + 0.0008687 \cdot sand - 0.0005164 \cdot clay$ <p>- for organic horizons*:</p> $BD = 1.4903 + 0.33293 \cdot \log(OC)$	(Hollis et al., 2012)
BD_Bernoux	$BD = 1.398 - 0.042 \cdot OC - 0.0047 \cdot clay$	(Bernoux et al., 1998)
BD_Hossain	$BD = 0.074 + 2.632 \cdot \exp(-0.076 \cdot OC)$	(Hossain et al., 2015)

Porosity

Name of the PTF	Equation	Reference
POR_Schjonning_etal	$PD_{OM} = 1.241 + 0.173 \cdot \left(\frac{OM}{100}\right)$ $PD_{SMS} = 2.663 + 0.107 \cdot \left(\frac{clay}{100}\right)$ $PD = \left(\frac{\left(1 - \frac{OM}{100}\right)}{PD_{SMS}} + \frac{OM}{100} \right)^{-1}$ $POR = \left(1 - \left(\frac{BD}{PD}\right) \right) \cdot 100$	(Schjønning et al., 2017)
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)
POR_2_65	$POR = \left(1 - \left(\frac{BD}{2.65}\right) \right) \cdot 100$	(Lal and Shukla, 2004)

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 %).

Erodibility

Name of the PTF	Equation	Reference
K_Sharpley_Williams	$K_{USLE} = \left(0.2 + 0.3 \cdot \exp \left(0.0256 \cdot \text{sand} \cdot \left(1 - \frac{\text{silt}}{100} \right) \right) \right) \cdot \left(\left(\frac{\text{silt}}{\text{clay} + \text{silt}} \right)^{0.3} \right) \cdot \left(1 - \left(\frac{0.25 \cdot \text{OC}}{(\text{OC} + \exp(3.72 - 2.95 \cdot \text{OC}))} \right) \right) \cdot \left(1 - \left(\frac{0.7 \cdot \left(1 - \frac{\text{sand}}{100} \right)}{\left(\left(1 - \frac{\text{sand}}{100} \right) + \exp \left(-5.51 + 22.9 \cdot \left(1 - \frac{\text{sand}}{100} \right) \right) \right)} \right) \right)$	Sharpley and Williams (1990)
K_Renard	$K_{RUSLE} = 7.594 \left(0.0034 + 0.0405 \cdot \exp \left(-0.5 \cdot \left(\frac{\log(D_g) + 1659}{0.7101} \right)^2 \right) \right)$ <p style="text-align: center;">with $D_g = \exp(0.01 \cdot \sum f_i \cdot \ln m_i)$</p>	Renard et al. (1997)

Albedo

Name of the PTF	Equation	Reference
ALB_Gascoin	$ALB = 0.31 \cdot \exp(-12.7 \cdot \theta) + 0.15$	Gascoin et al. (2009)

Soil hydraulic parameters

Available water content (AWC)

1. predict parameters of the van Genuchten model that describe the full moisture retention curve

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

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

2. compute FC and WP

$$\text{Assouline \& Or (2014): FC} = \theta_r + (\theta_s - \theta_r) \left\{ 1 + \left[\frac{n-1}{n} \right]^{(1-2n)} \right\}^{\left(\frac{1-n}{n} \right)}$$

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

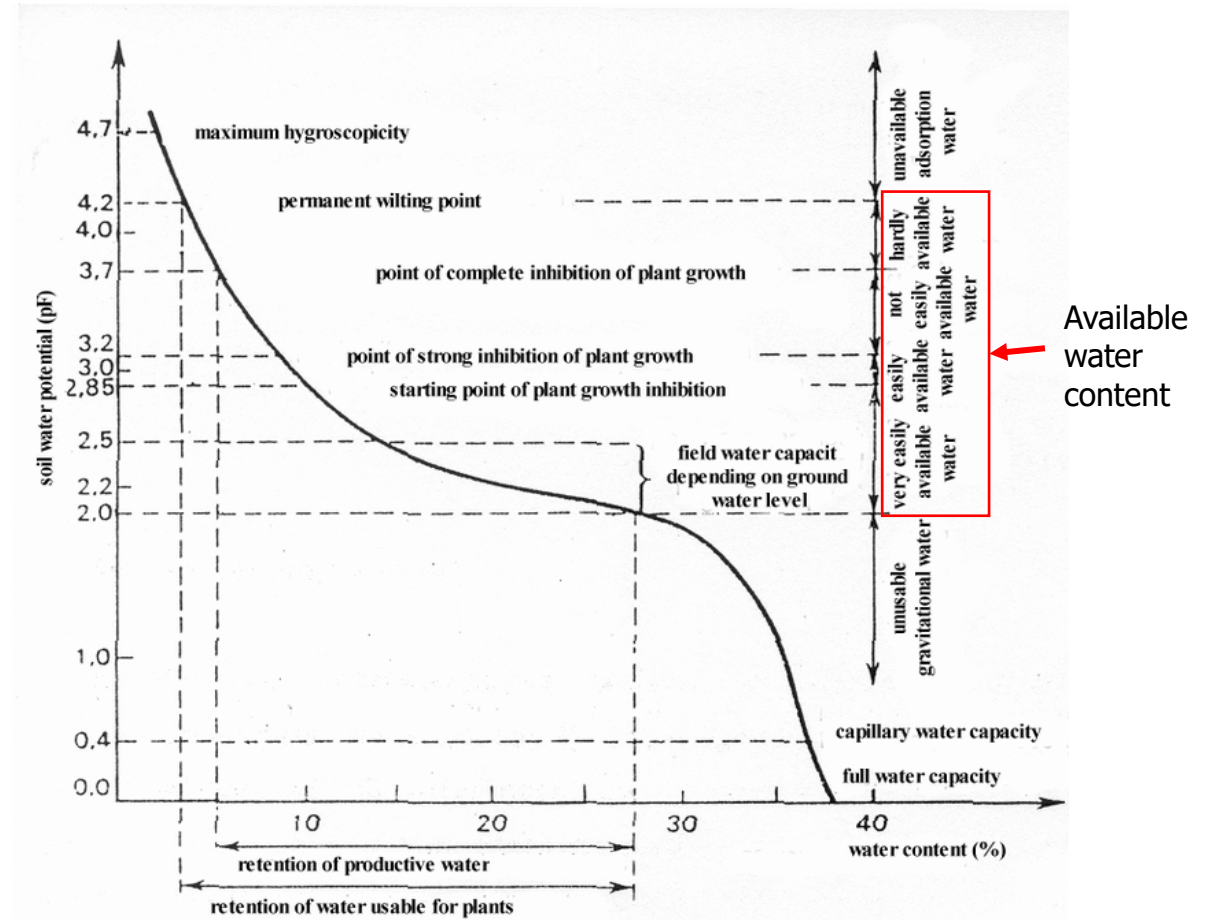
3. compute AWC based on FC and WP

$$\text{AWC} = \text{FC} - \text{WP}$$

Saturated hydraulic conductivity (KS)

compute KS from VG parameters

$$\text{Gascoin (2007): } K_S = 4.65 \cdot 10^4 \theta_s \alpha^2$$



Soil moisture retention curve after Turski R., Dom Œał H., Borowiec J., Flis-Bujak M., Misztal M., 1997. Gleboznawstwo – ćwiczania dla studentów wydziałów rolniczych. AR, Lublin.

Soil hydraulic parameters

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)

Soil chemical parameters

Soil phosphorus: mean statistics-based approach

- 1) Selection of LUCAS Topsoil samples from the adequate year and agroclimatic zone, preferably in the same country (NUTS region), with soil types and fertilization systems similar to the target area.
- 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.

Soil nitrogen:

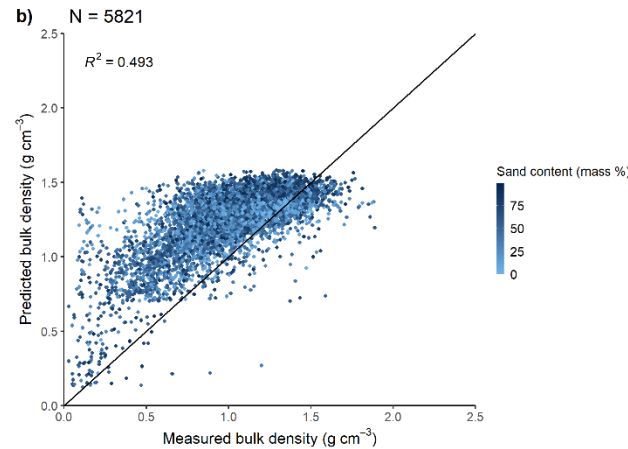
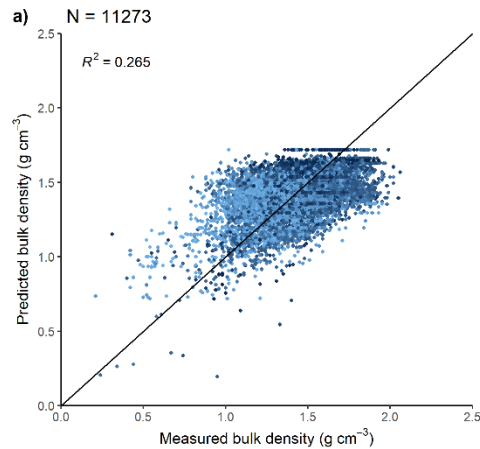
- organic nitrogen can be estimated from soil organic carbon content;
- inorganic nitrogen is highly variable in space and time and the dynamic of its amount is significantly influenced by leaching, denitrification, volatilization, and nitrogen fertilization → no general method is available for its prediction so far → information on the amount and timing of nitrogen fertilization is important + warmup period ≥ 4 years + initialise the SOM levels accurately to define the organic nitrogen pool.

Results

Bulk density

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.

PTF	EU-HYDI (N=11273)							LUCAS (N = 5821)							Weighted rank
	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



Prediction of dry bulk density could be performed with:

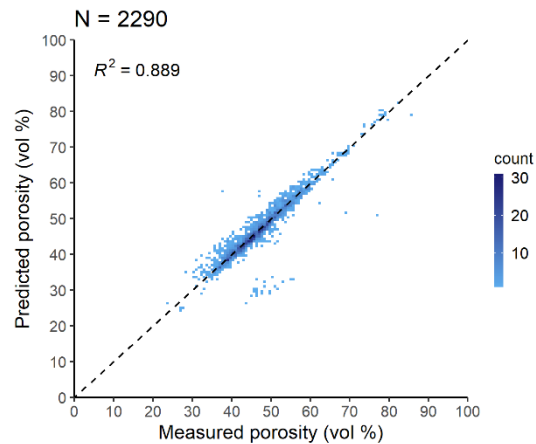
- i) BD_Alexander_A for soils with OC < 12% and
- ii) BD_Hossain for soils with OC >= 12%.

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.

Porosity

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	N	ME	MAE	RMSE	NSE	R^2	Sign. diff.
POR_Schjonning_etal	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

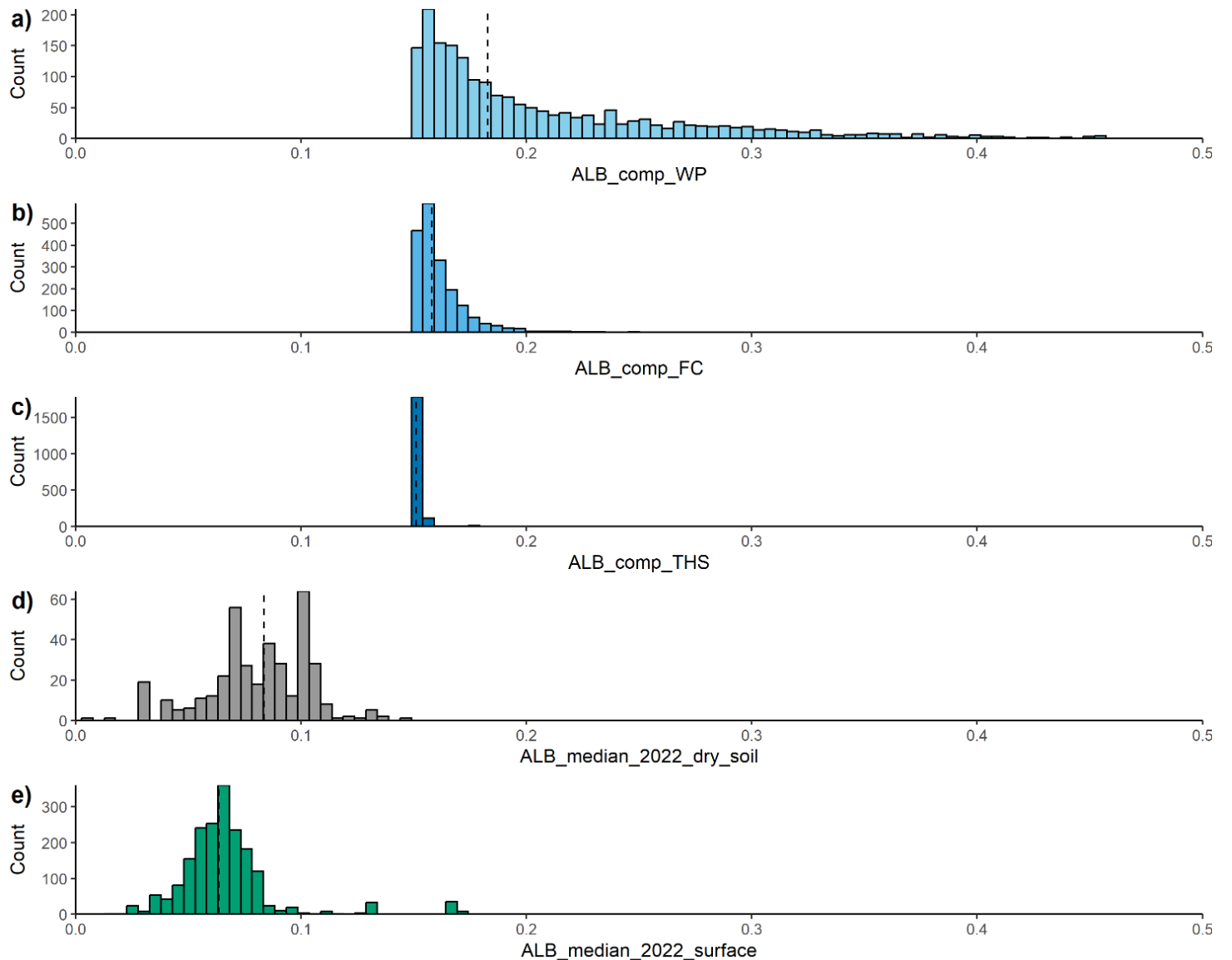


Porosity could be computed based on particle density predicted by the Schjønning et al. PTF instead of defining particle density as 2.65 g cm^{-3} .

Scatterplot of measured versus predicted porosity values of the best performing PTF, analysed based on the EU-HYDI subset with measured particle density values.

Albedo

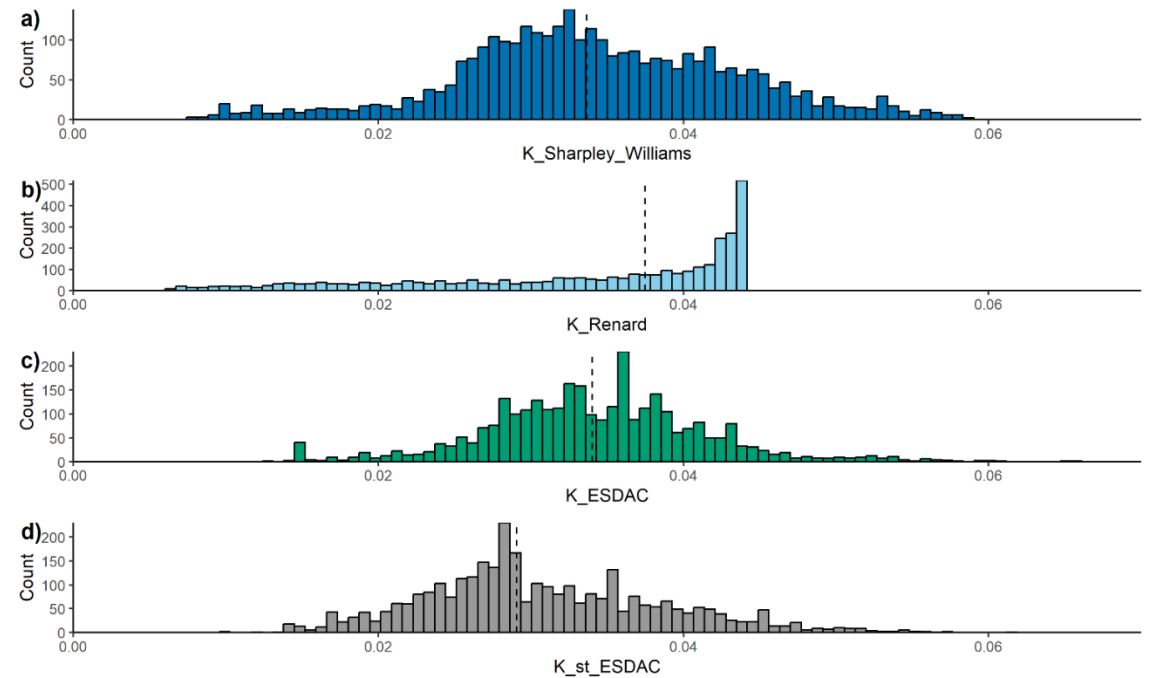
Histograms of the soil albedo computed with the Gascoïn et al. (2009) equation for the topsoil layers of the EU-HYDI dataset in the case of three moisture states: at saturation (ALB_comp_THS) (a), internal drainage dynamics-based field capacity (ALB_comp_FC) (b) and wilting point (ALB_comp_WP) (c) (N = 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.



It's crucial to specify the moisture condition for which the albedo value is needed in the modelling process.

Soil erodibility factor

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



Method	USLE K factor in different units						
	Unit	Min	Max	Range	Mean	Median	Standard deviation
Sharpley and Williams (1990)	$\left(\frac{t \cdot arce \cdot h}{hundreds\ of\ acre \cdot foot - tonf \cdot inch}\right)$	0.00	0.48	0.48	0.27	0.27	0.09
	$\left(\frac{t \cdot ha \cdot h}{ha \cdot MJ \cdot mm}\right)$	0.000	0.063	0.063	0.036	0.035	0.012
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

These predicted values could be used as preliminary approximations, but should be fine-tuned during the model calibration process.

Soil hydraulic properties

Prediction performance of internal drainage dynamics-based field capacity ($\text{cm}^3 \text{cm}^{-3}$) computed by pedotransfer functions on the FC and VG test sets of the EU-HYDI dataset.

Approach to predict	N	ME	MAE	RMSE	NSE	R ²
FC*						
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

Prediction performance of wilting point ($\text{cm}^3 \text{cm}^{-3}$) 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.

Approach to predict	N	ME	MAE	RMSE	NSE	R ²
WP*						
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

AWC could be computed based on the internal drainage dynamics-based FC and VG parameters-based WP.

Prediction performance of available water capacity ($\text{cm}^3 \text{cm}^{-3}$) computed by pedotransfer functions on the VG test set of the EU-HYDI dataset.

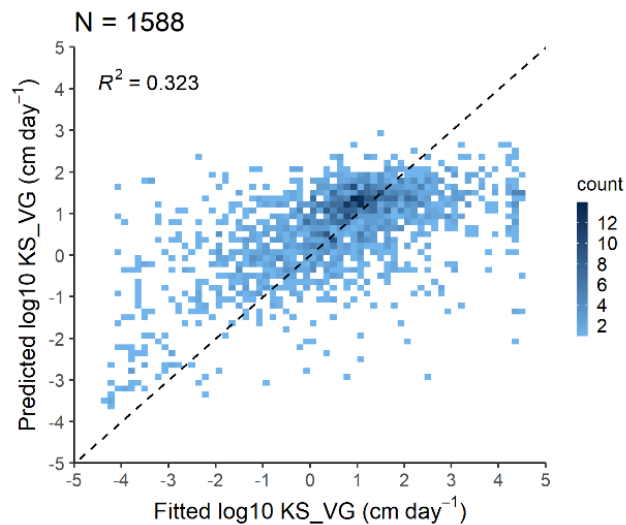
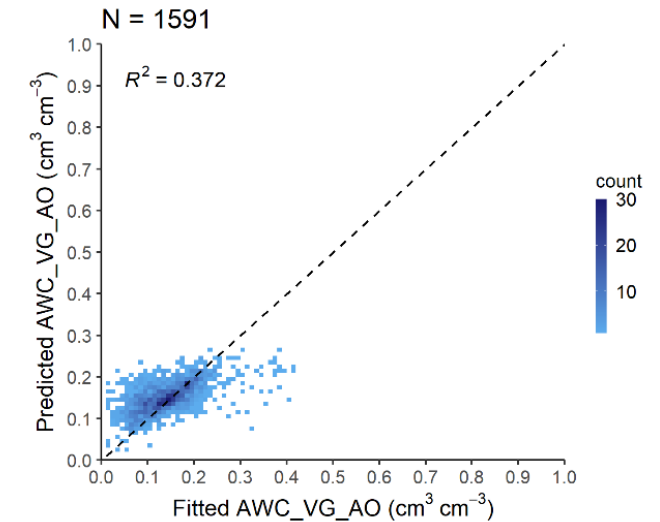
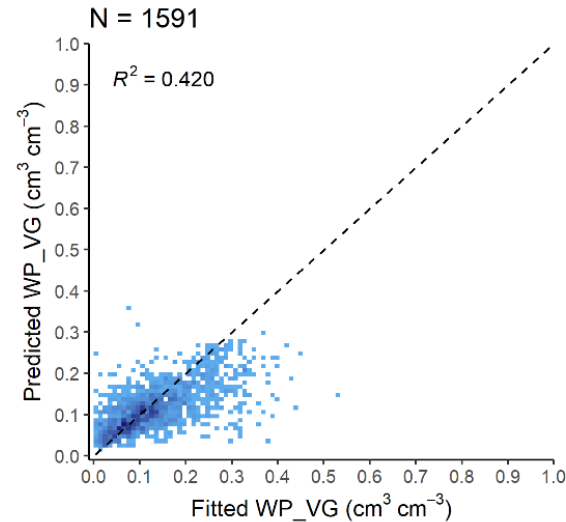
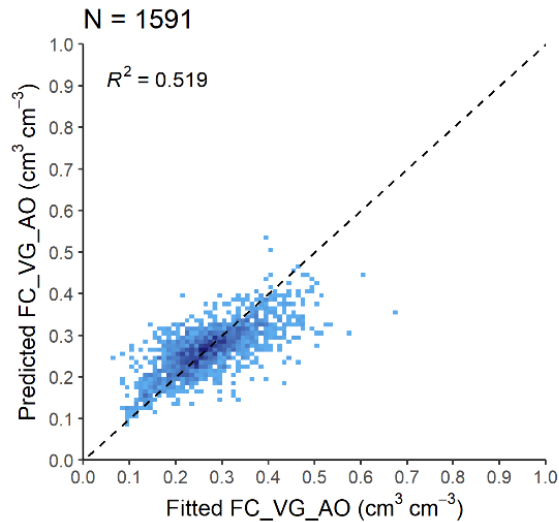
Approach to predict	N	ME	MAE	RMSE	NSE	R ²
AWC*						
pred_AWC_VG_AO	1591	-0.011	0.034	0.048	0.339	0.372
pred_AWC_VG_100	1591	-0.031	0.071	0.090	-1.325	0.072
pred_AWC_VG_330	1591	0.029	0.061	0.078	-0.725	0.044

Prediction performance of saturated hydraulic conductivity (cm day^{-1}) computed by pedotransfer function on the VG test set of the EU-HYDI dataset.

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

KS could be initialized using the VG parameters, but it should be adjusted during model calibration as a variable.

Soil hydraulic properties

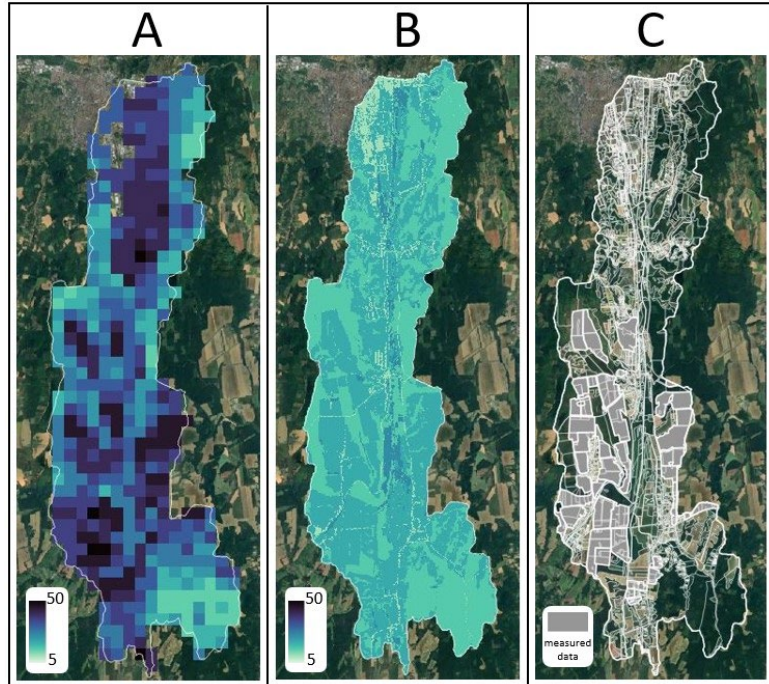


The prediction accuracy of water retention can be increased if not only soil depth, clay, silt, and sand content, organic carbon content and bulk density are used for the prediction, but pH and CEC as well.

Would it be possible to define FC, WP and porosity in the user soil table of SWAT+, and option to not compute it by the model?

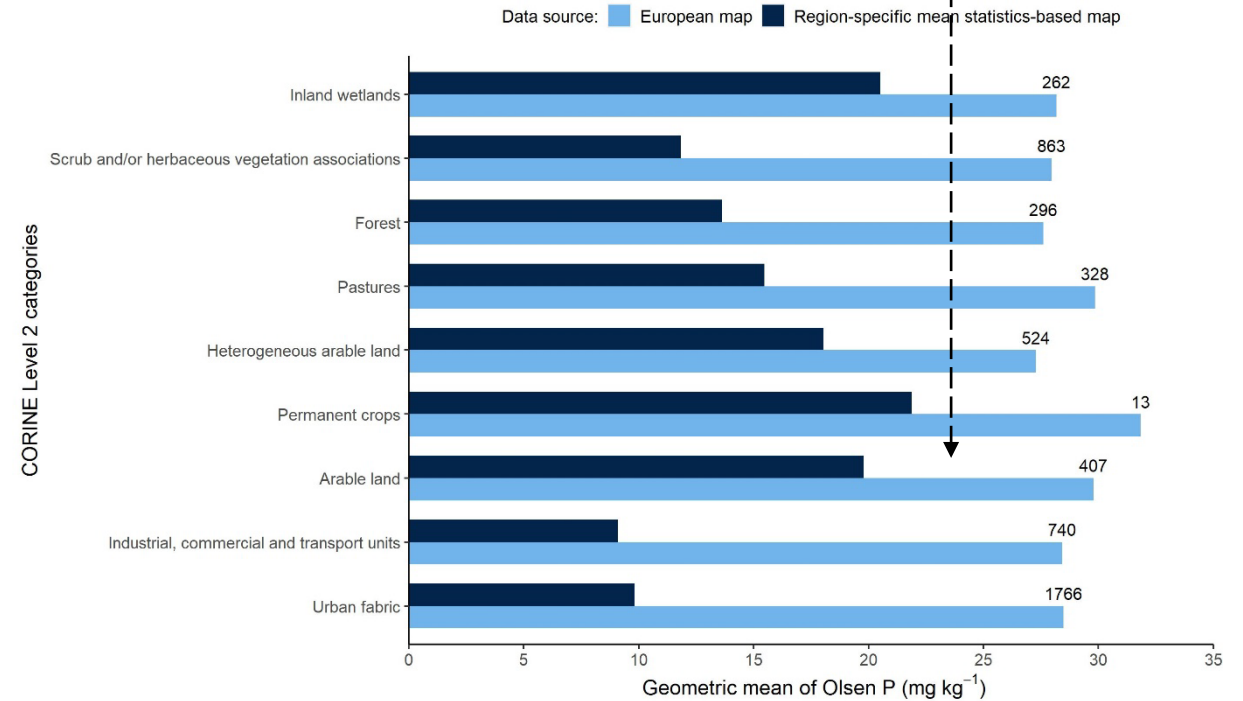
It would allow the use of any PTFs by the users.

Phosphorus content of the topsoil



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.

mean value of measured parcels: 24 mg kg⁻¹



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.

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. Where available, it is recommended to use measured data to overwrite the geometric mean values.

Suggested workflow

Soil physical properties

Organic matter or organic carbon content

$$OM = OC \cdot 1.724$$

Particle size limits required by the model vs. available in soil input data

Different countries/institutes measure soil particle-size distribution (PSD) using different methods and recognizing different classification standards (limit between silt and sand fractions \rightarrow 0.05 mm?).

Dry bulk density or effective bulk density

Effective bulk density derived from the dry bulk density by Wessolek et al. (2009) method:

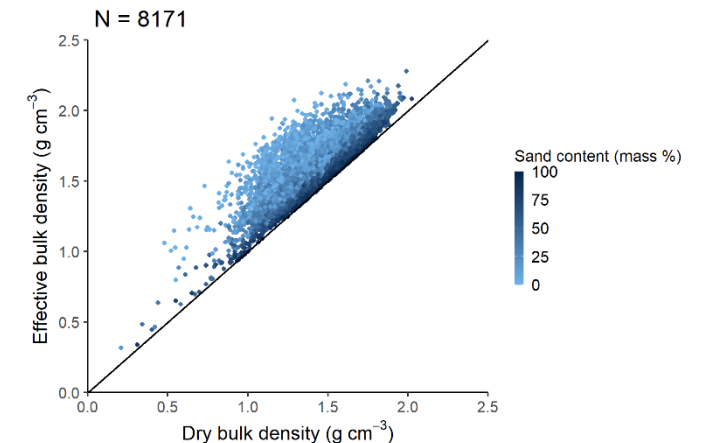
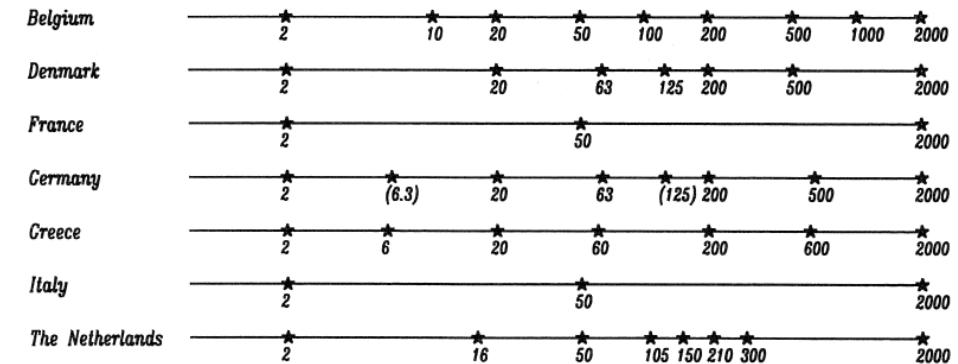
- for soils with organic carbon content higher than 0.58 %:

$$BD_{eff} = BD_{dry} + 0.009 \cdot clay$$

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

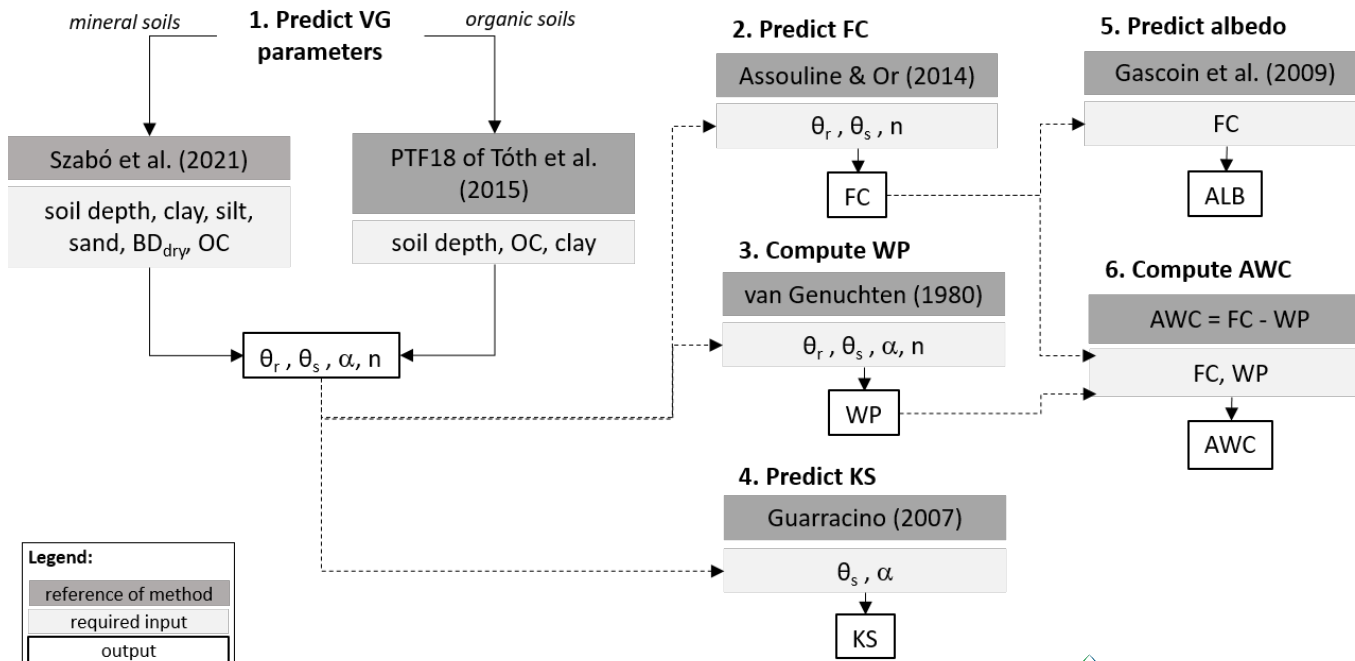
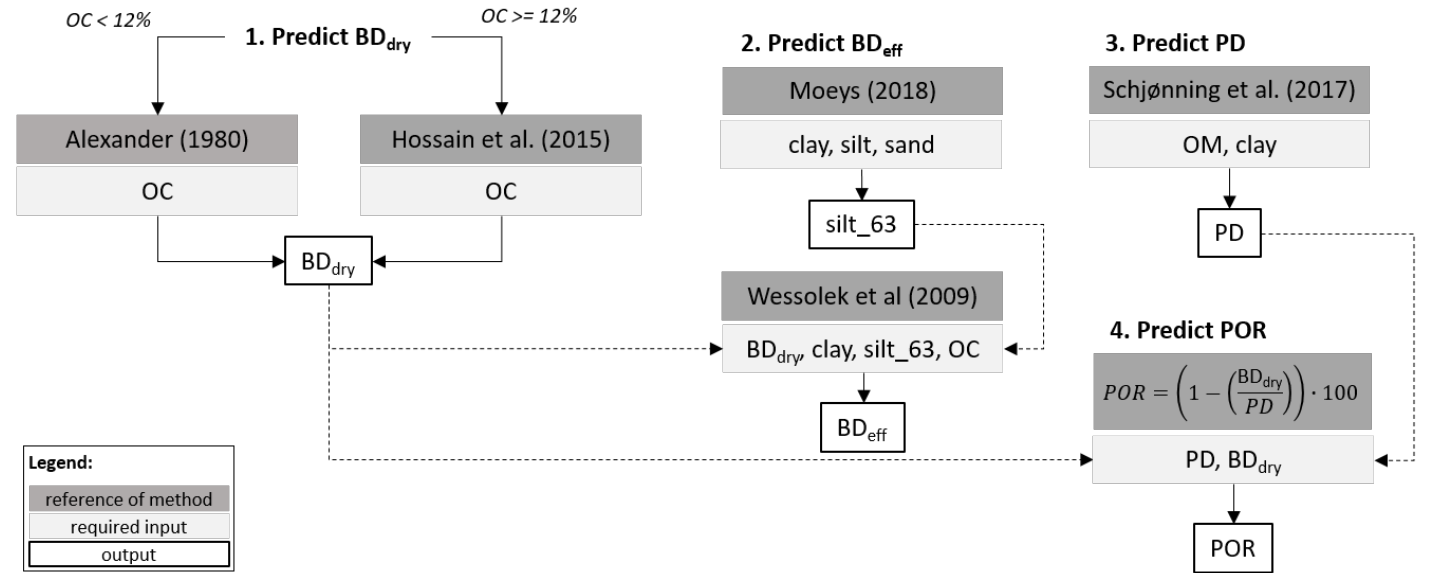
$$BD_{eff} = BD_{dry} + 0.005 \cdot clay + 0.001 \cdot silt$$

A. Nemes et al. / Geoderma 90 (1999) 187–202

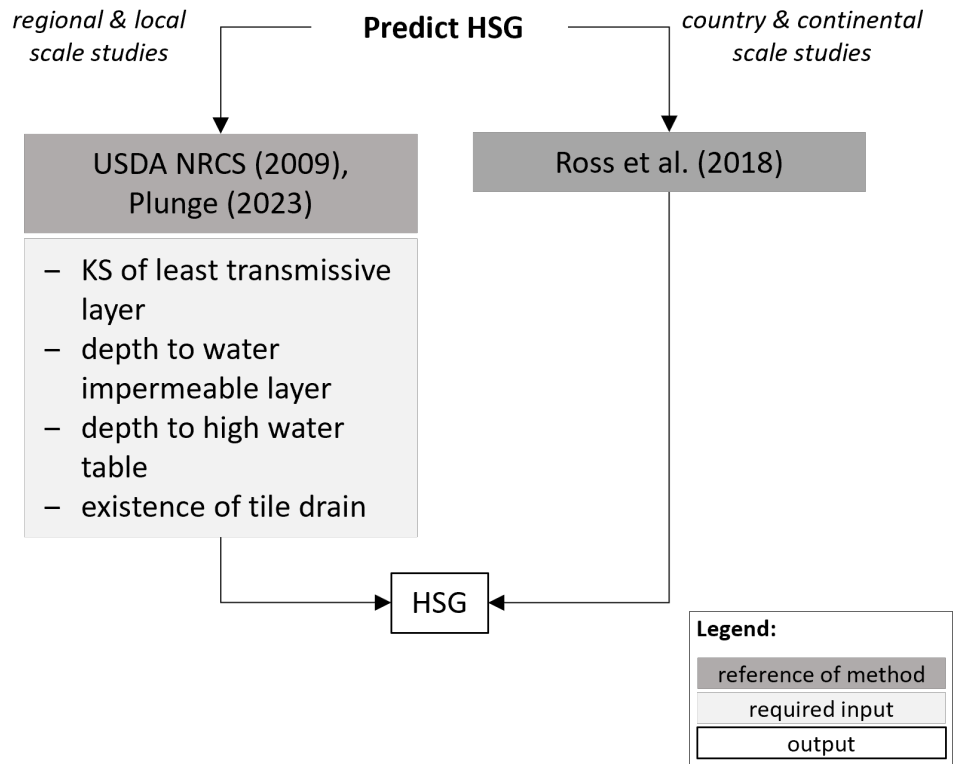


Scatterplot of dry versus effective bulk density analysed based on the point data of EU-HYDI.

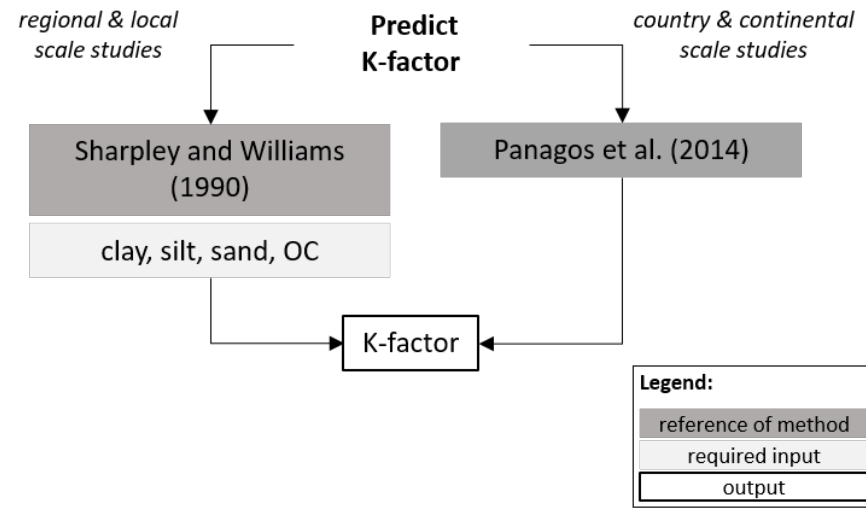
Prediction of soil physical properties.



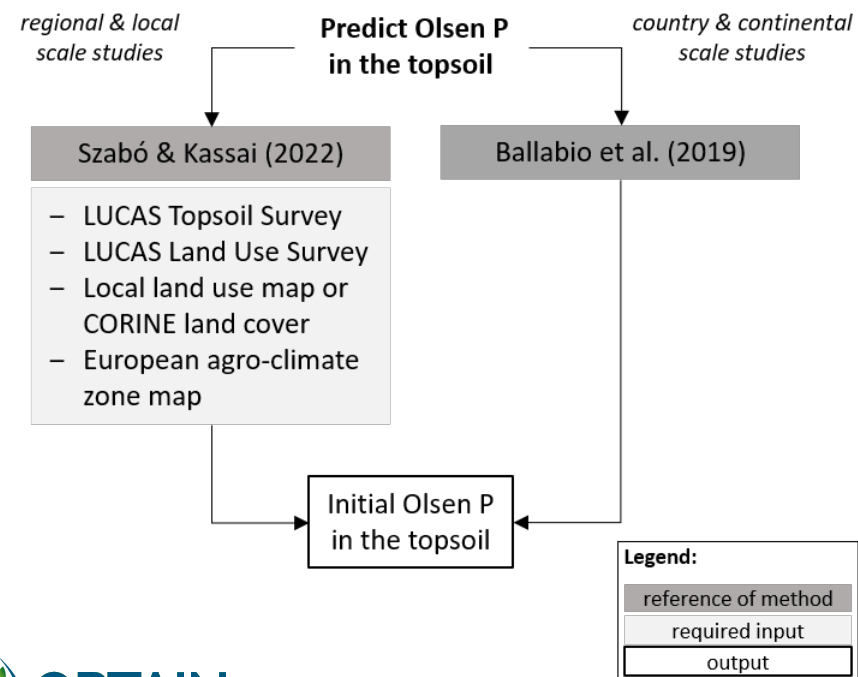
Prediction of soil hydraulic properties and moist soil albedo.



Prediction of **hydraulic soil groups (HSG)**.



Prediction of **soil erodibility factor (K-factor)**.



Prediction of **Olsen phosphorus content of the topsoil**.

Tools to derive soil properties

- derive user soil table with R package SWATprepR:
<https://github.com/biopsichas/SWATprepR>
- compute soil hydraulic properties with euptfv2:
 - user friendly web interface: <https://ptfinterface.rissac.hu>
 - R package: <https://github.com/tkdweber/euptf2>
- algorithm to harmonize soil particle size data (sand, silt and clay content) to the FAO/USDA system:
<https://doi.org/10.5281/zenodo.7353722>
- map topsoil phosphorus content:
<https://doi.org/10.5281/zenodo.6656537>

Preprint

Szabó, B., Kassai, P., Plunge, S., Nemes, A., Braun, P., Strauch, M., Witing, F., Mészáros, J., and Čerkasova, N.: Addressing soil data needs and data-gaps in catchment scale environmental modelling: the European perspective, EGU sphere [preprint], <https://doi.org/10.5194/egusphere-2023-3104>, 2024.



2nd OPTAIN webinar

Topic: Modelling of water and nutrient retention in agricultural catchments in the scope of OPTAIN project

Date and time: 3 September 2024, 15:00



Thank you

This work received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 862756, project OPTAIN and the Sustainable Development and Technologies National Programme of the Hungarian Academy of Sciences (FFT NP FTA).

Brigitta Szabó

Institute for Soil Sciences,
HUN-REN Centre for
Agricultural Research

szabo.brigitta@hun-ren.atk.hu



@H2020_OPTAIN



@H2020OPTAIN

WWW.OPTAIN.EU



Soil properties most frequently required by the environmental models

- soil layering,
- maximum rooting depth,
- 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.

