Dear Diana Vieira,

Thank you for the time and effort you have put into the detailed review and suggestions on how to improve our manuscript. Please find below how we would address the raised issues. First, we cite your comment, then provide our response. The changes that would be applied to the manuscript text are highlighted in blue and are visible in the uploaded PDF version of this text.

GENERAL COMMENT:
This study compiles a series of datasets of soil properties and pedotransfer functions available for catchment scale modeling, evaluate them and provide an R open script for soil data derivation. I believe this is a valuable piece of work, however, I have several major concerns about its suitability for publication in its current shape but I believe it could be significantly improved. The main concerns is the fact that there is no data openness in this work, the main data used is not available to assess reproducibility and another smaller dataset (independent dataset provided by agricultural company) is not even described. If the authors are willing to share the data and correct the identified problems, I believe this could be an excellent contribution.

ANSWER FOR GENERAL COMMENT:
1) The reason for using EU-HYDI for this study is that it is the most representative soil hydraulic dataset for Europe that we could use for this study. The internal use and no external openness of the dataset has been requested by the data providers during the establishment of EU-HYDI, which was initiated and coordinated by the EC Joint Research Centre in 2013. Some contributors have given the JRC a licence to make their raw data publicly available on the European Soil Data Centre. Based on information from JRC it will soon be available from ZENODO. Information about data availability is provided here: https://github.com/melwey/euhydi_public . We will add this information to the "Data availability" section of the manuscript as soon as the link will be available from JRC:
"6,583 samples of 1999 soil profiles, summing up to 35 % of the EU-HYDI dataset, are available from ZENODO DOI LINK. The entire dataset cannot be made publicly available due to its legal restrictions."

2) Thank you for pointing out the missing description of the dataset on topsoil phosphorus content. We will add the following text on it under section 2.1 "Evaluation of the methods" in line 143:
"The LUCAS Topsoil dataset (Orgiazzi et al., 2018; Tóth et al., 2013) of 2009 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."

SPECIFIC COMMENTS
COMMENT 1:
Lines 39 -58 – The introduction could approach more deeply the importance/availability of all the soil properties mentioned in the abstract (Lines 23-25), however the examples provided seemed disconnected between each other. Alternatively, further explanation could be provided to justify giving those examples and not other.

ANSWER 1:
1.1 We will modify lines 39-44 in the following way:

"The basic soil properties, i.e., soil organic carbon content, particle size distribution, 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 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 (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."

1.2 Lines 50-58 will be moved before lines 45-58:
"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)."

1.3 Under lines 45-58 those soil properties are included which are more challenging to retrieve. Thanks for the suggestion, we will add the following explanation to the text:
"Often those soil properties are required as model input data as well, which are rarely available. One example is the data on soil cracking. 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."

COMMENT 2:
Lines 64-67 – What would this allow the scientific community to do? Support modeling studies? Assist researchers in the decision for methodological approaches? Please provide a statement.

ANSWER 2:
We will modify lines 64-67 accordingly:
"Therefore, in this study we support soil data retrieval for environmental modelling across Europe by i) systemizing information on open access datasets and PTFs applicable for Europe, ii) demonstrating 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) providing a comprehensive workflow and accompanying open-source R script and library for the derivation of missing soil data."

COMMENT 3:
Materials and Methods
I was hoping to read here how you determine the compilation in table 1. How to ensure this dataset is complete?

ANSWER 3:

We will add the following text in line 111 to highlight the continuous improvement of datasets and most important sites where information on new dataset or updates is expected to be available in the future:

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).

However, these sources do not include products from specific institutes, such as http://globalchange.bnu.edu.cn/research. The datasets included in Table 1 might be appropriate for regional and continental modelling.

COMMENT 4:
Lines 92 -109 – These are the "soil properties most frequently required by environmental models" based on what? How did you determined this list?

ANSWER 4:
We will modify lines 92-93 in the following way:
"Soil properties most frequently required as static parameters 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:"

COMMENT 5:
Lines 112 – Regarding the statement "Local and national datasets provide more accurate input information", I would say depends. If is a spatially explicit modeled database I would say definitely yes, if is point data with precise coordinates of the source of the data (e.g. LUCAS) I would say no. Of course, in this condition is not possible to address quantity of information (i.e. density of points) but one would expect local data to present higher data availability, but this is not mentioned in the text. Please be more precise on this matter.

ANSWER 5:
Thank you for the suggestion, we will modify lines 111-112 accordingly:
"However, for catchment scale and national studies, local and national spatially explicit modelled datasets provide more accurate input information."

COMMENT 6:
Line 114-119 – I read this text several times, I understand you want to make the case regarding the previous sentence, but is not entirely clear please rephrase it.

ANSWER 6:

We will modify lines 112-119 lines to improve clarity:
When a certain local dataset is selected to be used as basic soil information, it is more consistent to compute the missing soil properties from this local data source rather than using other data sources. This allows to maintain consistency between the different soil properties. For example, it is not recommended to combine a local soil property map at 100 m resolution with soil hydraulic properties retrieved from EU-SoilHydroGrids at 250 m resolution (Hengl et al., 2017). Where local soil maps with soil layering, organic carbon content, clay, silt and sand content are available, it is suggested that missing soil properties such as bulk density, soil hydraulic properties, and albedo are estimated from the locally available basic soil properties to ensure consistency.

COMMENT 7:
Line 119 – 121 – Perhaps there is an expert on soil cracking in this team, but I have to admit I never used such variable. I had to search SWAT documentation, and in the EU only a small amount of soils are classified as Vertisoils. Therefore, I question the authors to justify the importance of soil cracking in comparison with all the other properties listed.

ANSWER 7:
We agree that the importance of soil cracking is significantly lower. It is included because the SWAT model requires this input and we decided to follow the structure of the SWAT+ usersoil table, but we will exclude this text because consideration of soil cracks is optional in SWAT+.

COMMENT 8:
Lines 122 – 126 – These are suggestions and not materials and methods, please move this to an appropriate section, or rephrase it if means you took those considerations when analyzing data.

ANSWER 8:
Thank you for the suggestion, we will move these recommendations under “4 Conclusions” section in line 661 in a separate paragraph, before the sentence starting with “When retrieving or deriving …”.

COMMENT 9:
Lines 127 -134 – same as before.

ANSWER 9:
We will delete the first sentence of that section and move the rest to line 67.

COMMENT 10:
Lines 139 – 141 – Not clear. I understand there is partial data availability for comparison, but I don’t understand what is the consequence of such approach. Please clarify the text.

ANSWER 10:
We will add the following clarification in line 142:
“This approach aimed to facilitate a more accurate and fair comparison among different PTFs, but decreased the number of samples used for the analysis.”

COMMENT 11:
Line 138 -139 – All right so this EU-HYDI is you data for validation right? It is reasonable to say this dataset hasn't been updated since 2013, and considering the report is actual, should I ask if you only considered measured values? Because I see in the report that part of this dataset contains estimated values. In addition to that, could you point the readers to the data itself? The possibility to reproduce the same analysis is necessary.

ANSWER 11:
Yes, we considered only measured values. The total number of samples in EU-HYDI is 18,682. For our analysis, the number of samples varied between 1,591 and 11,287 depending on which soil property was analysed. Please find information regarding the link to the data in the "ANSWER FOR GENERAL COMMENT".

COMMENT 12:
Line 152 -333– I honestly don't understand why you wrote this as a protocol format, please reformulate to describe the data analysis that you produced. This is written as a textbook that is not the purpose right? Moreover, out of the sudden I realized that there are datasets that have not been used (e.g depth to water table) and others that have been, could you synthesize that information?

ANSWER 12:
Thank you for the suggestion on reformulation.
12.1 We will add the following section above line 150:
"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."
12.2 We will reformulate the lines 152-303 in the following way:
1) move lines 151-174 under "3.9 Suggested workflow to derive soil input parameters" section.
2) reorganize level of subtitles: "Soil physical parameters", "Soil hydraulic parameters" and "Soil Chemical parameters" will go under "2.2 Analysed soil properties". There would be no numbered subtitles, only the name of the soil property in the case of bulk density, porosity, albedo, soil erodibility factor, and water retention and saturated hydraulic conductivity.
3) move lines 176-180 and 189-198 under "3.9 Suggested workflow to derive soil input parameters" section. We will keep the following text under "Bulk density section":
"Table 2 lists the PTFs that were tested on point data in EU-HYDI 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."
4) delete line 200 and move lines 209-212 above Table 3.
5) rephrase lines 246-247 in the following way:
"Soil water retention and hydraulic conductivity can be computed from the parameters of the widely used van Genuchten model (VG) (van Genuchten, 1980):"
6) move lines 257-265 above Table 4.
7) delete "This approach is recommended for the computation of FC." from line 268.

8) move text on Hydrological groups (lines 285-303) without subtitle to “3.9 Suggested workflow to derive soil input parameters” section above sentence starting with “Figures 18-21 …”.

COMMENT 13:
Line 312 – How can we know about local fertilization schemes?

ANSWER 13:
We will add the following information in line 312:
"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. If this information is not known, the NUTS2 phosphorus map of the European cropland areas (Tóth et al., 2014) might be useful in the data selection ."

COMMENT 14:
Line 319 - Locally independent measured dataset?? Provided by an agricultural company? How many samples? When samples were taken? Laboratorial methods? Statistical analysis? This is clearly an insufficient methods description.

ANSWER 14:
Thank you for pointing it out, we will add the text inserted under point 2) of ANSWER FOR GENERAL COMMENT in line 143.

COMMENT 15:
Porosity – Lines 396-399 – so what have you done regarding this? If 43% of the samples presented errors, was this data excluded? (explain for all parameters)

ANSWER 15:
Finding discrepancy between porosity and saturated water content is common in international soil hydraulic databases. This rests in the complexity of their measurement when it comes to small laboratory practicalities, as well as assumptions that are frequently made. In terms of measuring water content, samples often drain some of the water by the time the technician has a chance to raise it from the water bath and put the sample onto the scale. It can also happen that some water is still ponding on top of the sample when the measurement is taken. At the same time, there are known error sources in the determination of particle-density, but this property is often only assumed to be 2.65g/cm3 without measuring. The same applies to measuring bulk density: different standards and/or routines are known to exist worldwide (e.g. measuring on clods vs. ring samples, or measuring at a dry state vs. at an equilibrated moisture level), and the determination of this property is also very sensitive to sample quality. Samples can often over or underfill the rings. It is also customary that data reporters just equate porosity and saturated water content without measuring one of them.
When both properties are reported, discrepancies are often seen due to the above reasons – and beyond. The independent user is typically not equipped to judge which one to trust and which one not to. Therefore, it is a routine procedure to cross-check them and report on data quality, but to only be concerned about the value-pairs that show large discrepancies. The proportion of the suspicious samples in the further analysis is low, 39 out of 1591. Therefore, we did not decrease the number of samples used for the analysis.

We will precise lines 396-397 in the following way:
"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, 56.5% have a total porosity larger or equal to the saturated water content. For the samples where the saturated water content is higher than the total porosity (N = 170), the reason may be the uncertainties in the measurement of both parameters. It is possible that 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."

COMMENT 16:
Soil erodibility – I thought the Renard et al 1997 was the most used version of the RUSLE model (almost 5 thousand citations according to semantic scholar), but in case I am not right would you provide an information about it?
Predicting Soil Erosion by Water: A Guide to Conservation Planning With the Revised Universal Soil Loss Equation (RUSLE) (usda.gov) page 65.
Either way, why testing only one equation?

ANSWER 16:
Thanks you for the suggestion! We wanted to use only those equations which can be readily applied for the soil properties most frequently available and not use the ones that require non-easily available soil properties, such as soil structure or permeability. The Renard et al. 1997 equation fits into the logic, therefore we will add K factor computed with it (K_computed_Lenard) and compare its result with the methods already included.

COMMENT 17:
Field capacity – Would be worth to make a reference to table 4 early in the beginning. (for all parameters)

ANSWER 17:
Thank you for the suggestion, we will add the followings:
in line 452: "The FC defined (see abbreviations in Table 4) based on …"
in line 495: "Calculating WP (see abbreviations in Table 4) from …"
in line 522: "If only AWC (see abbreviations in Table 4) is required …"
in line 560-561: "Figure 14 shows the relationship between measured KS and computed with Eq. (22) based on the fitted VG parameters (KS_VG) (see abbreviation in Table 4)."

COMMENT 18:
Wilting point – I don't understand why the number of points available to assess VG (table 11) differs from the ptf (table 12)

ANSWER 18:
The number of samples differs because Table 11 shows the performances on the VG test set of the EU-HYDI, Table 12 includes performances analysed on the WP test set of the EU-HYDI. Analysis of direct WP prediction (pred_WP) was added to show the difference in accuracy between pred_WP_VG and

pred_WP. To increase clarity, we will add the following modifications in the captions of Table 11 and 12:
"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. …"
"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. …"

COMMENT 19:
Saturated hydraulic conductivity – found strange you didn't use this database ESSD - SoilKsatDB: global database of soil saturated hydraulic conductivity measurements for geoscience applications (copernicus.org) for comparisons also, any justification for that?

ANSWER 19:
Our aim was to use EU-HYDI dataset, because that is the most representative soil hydraulic dataset for Europe, therefore we would not add further datasets for the KS analysis.

COMMENT 20:
Line 625 - So these workflow are the result of your analysis, whereas you described the most efficient workflow for better data quality. Right?

ANSWER 20:
Yes, thank you for the suggestion on describing the aim of this section. We will add in line 626 the following:
"Based on the above results, we describe the most efficient workflow to retrieve the soil input parameters for European environmental modelling."
Then we will continue with the "protocol type" text that we move from the material and methods – mentioned under "ANSWER 12":
"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. …"

COMMENT 21:
Line 653 – "Key findings underscored the significance of local soil data over global or large-scale datasets in environmental modelling", I'm afraid you haven't provided hard evidence on this. Besides this Figure 17, you presented zero information about the sampling, the timing of the year, the laboratory analysis of this procedure, among many other details. Please provide the sufficient info in order to assess if such assessment is even comparable.

ANSWER 21:
Thank you for highlighting it. We agree with you, we will delete this conflicting sentence, because recommendation on density and timing of sampling, type of laboratory methods and other related topics is out of the scope of this manuscript.

COMMENT 22:

Data availability – I find this justification rather poor considering we are not talking about personal information, nor information that could reduce the value of the land. There is a report online, and this work was paid with taxpayer's money already more than 10 years ago. In the meanwhile, many things changed in science, and the open data is the new reality. Making this data available, in an open data journal as this one, would help the scientific community to overcome many obstacles in the hydrological modelling.

ANSWER 22:
Please find our answer in "ANSWER FOR GENERAL COMMENT". Please note that EC JRC covered the cost of harmonising the data structure and meetings on creating EU-HYDI, but data collection and laboratory analysis was covered by the participating institutions. EU-HYDI is not the property of any of the authors of this manuscript. Two of the authors were collaborators during the construction of EU-HYDI, therefore the dataset could be accessed for the presented analysis according to its licensing. Accommodating the opening of EU-HYDI dataset is beyond the task and reach of these authors.
We will add the information included in "ANSWER FOR GENERAL COMMENT" and the following text to "Data availability" section:
"LUCAS TOPSOIL data can be accessed through European Soil Data Centre (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."

REFERENCES ADDED IN THE DOCUMENT:

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.

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.

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.

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

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.

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., 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.

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.

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

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.

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.