EstSoil-EH: An eco-hydrological modelling parameters dataset derived from the Soil Map of Estonia

Alexander Kmoch[1], Arno Kanal[1,†], Alar Astover[2], Ain Kull[1], Holger Virro[1], Aveliina Helm[3], Meelis Pärtel[3], Ivika Ostonen[1] and Evelyn Uuemaa[1]

[1]Department of Geography, Institute of Ecology and Earth Sciences, University of Tartu, Vanemuise 46, Tartu, 51003, Estonia
[2]Chair of Soil Science, Institute of Agricultural and Environmental Sciences, Estonian University of Life Sciences, Fr.R. Kreutzwaldi 5, Tartu, 51014, Estonia
[3]Department of Botany, Institute of Ecology and Earth Sciences, University of Tartu, Lai 40, Tartu, 51005, Estonia
[†] deceased, 07th of May, 2019

Correspondence to: Alexander Kmoch (alexander.kmoch@ut.ee)

Abstract. The Soil Map of Estonia is a vector dataset that maps more than 750 000 soil units throughout Estonia at a scale of 1:10 000. It is the most detailed and information-rich dataset for soils in Estonia, a Baltic country with an area of approximately 45 000 km^2. For each soil unit, it describes the soil type (i.e. soil reference group), soil texture, and layer information with a composite text code, that comprises not only of the actual texture class, but also of the classifiers for the rock content, peat soils, its distinct compositional layers and their depths. However, to use it as an input for numerical modelling using process-based physical models, these text codes must be translated into numbers. Various generalisations and aggregations for agricultural soils for less-detailed versions of the map have been made at a scale of 1:100 000 and 1:200 000, but not for the original scale of 1:10 000.

In this study, we create an extended eco-hydrological dataset for Estonia, the EstSoil-EH (Kmoch et al., 2019a; doi:10.5281/zenodo.3473289), containing derived numerical values for the following data in all of the mapped soil units in the 1:10 000 soil map: soil profiles (e.g., layers, depths), texture (clay, silt, sand components), coarse fragments and rock content, and physical variables related to water and carbon (bulk density, hydraulic conductivity, organic carbon content). Ultimately, our objective was to develop a reproducible method for deriving numerical values to support modelling and prediction of eco-hydrological processes in Estonia with numerical models such as the popular Soil and Water Assessment Tool.

The developed methodology and dataset will be an important resource for the Baltic region. Countries like Lithuania and Latvia have similar historical soil records from the Soviet era that could be turned into value-added datasets such as the one we developed for Estonia.

1 Introduction

Eco-hydrological numerical models like the Soil and Water Assessment Tool (SWAT; https://swat.tamu.edu/) or the Regional Hydro-Ecologic Simulation System (RHESSys) have been developed and applied during the past 30 years to evaluate the effects of alternative management decisions on water resources and non-point-source pollution in river basins through the simulation of physical processes (Arnold et al., 1998; Douglas-Mankin et al., 2010). SWAT is widely used internationally and is increasingly applied in Northern European and Baltic watersheds to better assess the hydrological state of the environment based on modelling of the most relevant physical processes (Piniewski et al., 2018; Tamm et al., 2016, 2018). However, a main input factor for many of these models is detailed soil data, which does not exist for many countries on national scale or which exists with insufficiently fine spatial resolution. In addition, it is complicated to derive the values of the model parameters.

The objective of the present study was to develop a numerical soil database for modelling and for predicting processes in Estonia. In this study, we derived numerical values for the key characteristics for the whole Soil Map of Estonia at a 1:10 000 scale for soil profiles (e.g., layers, depths), textures (clay, silt, and sand contents), coarse fragments and rock content, and physical variables related to the water and carbon cycle. There is no countrywide spatial dataset of BD, SOC, hydraulic conductivity, available water capacity) for Estonia. Thus, it was needed to derive predictions for these soil properties in order to map them.

Existing national-scale soil datasets that have been developed to be used by SWAT currently only exist for the United States. Cordeiro et al. (2018) developed an official soil dataset for SWAT for Canada. Apart from these efforts, no consistent methodology has been used to develop soil datasets at national, continental, or global scales so that the data is immediately applicable for the use in SWAT (Batjes, 1997; Dobos et al., 2005).

At the global level, two main soil databases are available. The first was made available by the United Nations Food and Agricultural Organisation (FAO) through its Soils Portal: the Harmonized World Soil Database (HWSD) v1.2 (Fischer et al., 2008; http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-world-soil-database-v12/en/). The dataset resulted from a collaboration between FAO and Austria’s International Institute for Applied Systems Analysis (http://www.iiasa.ac.at/), ISRIC–World Soil Information (https://www.isric.org/), the Institute of Soil Science of the Chinese Academy of Sciences (http://english.issas.cas.cn/), and the Joint Research Centre of the European Commission (https://ec.europa.eu/info/departments/joint-research-centre_en). HWSD is a 30-arc-second raster database with more than 15000 different soil mapping units. It combines existing regional and national updates of soil information from around the world, including key properties of the Soil and Terrain database (SOTER, https://www.isric.org/explore/soter), the Ecological Site Description system of the U.S. Department for Agriculture (ESD, https://esis.sc.egov.usda.gov/Welcome/pgReportLocation.aspx?type=ESD), the Soil Map of China (https://esdac.jrc.ec.europa.eu/content/soil-map-china), and homogenized sets of soil property estimates in the World Inventory of Soil Emission Potentials (WISE, https://www.isric.org/explore/wise-databases). It also contains information from the 1:5 000 000-scale FAO-UNESCO Soil Map of the World (FAO, 1990; http://www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/faounesco-soil-map-of-the-world/en/).

The other global-level soil dataset is SoilGrids250m, which provides global gridded soil information derived with machine learning methods (Hengl et al., 2017) and is made accessible via an interactive Web interface (https://soilgrids.org/) with sophisticated standards-based data access via the OGC Web Coverage Services (https://www.opengeospatial.org/standards/wcs). SoilGrids250m provides values for sand, silt, clay, and rock fractions, and organic carbon and carbon stocks at several depths, which can be used as inputs for SWAT. SoilGrids also provides a harmonized soil database for Europe. At the regional level, the European Soil Database v2.0 (Panagos et al., 2012; https://esdac.jrc.ec.europa.eu/content/european-soil-database-v20-vector-and-attribute-data) is aharmonized soil database for Europe. It contains the soil geographical database SGDBE (vector data), which includes a number of essential soil attributes, and an associated database (PTRDB), with attribute values that have been derived through pedotransfer rules. The European database also includes the Soil Profile Analytical Database, which contains measured and predicted soil profiles for Europe as well as soil organic carbon (SOC) projections for Europe that include 26 European countries at a resolution of 1 km. Further European databases that collate soil-hydraulic data and pedotransfer functions (PTFs) include the European Hydropedological Data Inventory (EU-HYDI, https://esdac.jrc.ec.europa.eu/content/european-hydropedological-data-inventory-eu-hydi) the database of HYdraulic PRoperties of European Soils (HYPRES Wösten et al. (1999)).

In Estonia, systematic large-scale soil mapping was launched in 1949, with agronomy students assisting (Estonian Landboard, 2017; "mullakaardi_seletuskiri.pdf"). Starting in 1954, a special survey was carried out under the supervision of the Ministry of Agriculture. Aerial photographs were used as the basis for this activity. By 1992, Estonia’s soil cover had been mapped by the Soil Survey Department of the former Institute of Estonian Agroprojects at a scale of 1:10 000. In addition to inspecting arable land, forests, and other land types between 1989 and 1991, the remaining former Soviet military areas were also mapped. Between 1997 and 2001, the soil map was digitized and attribute data was inserted into the database, resulting in the official National Soil Map of Estonia as a vector dataset that mapped 750 000 soil units at a scale of 1:10 000 (Estonian Landboard, 2017; https://geoportaal.maaamet.ee/est/Andmed-ja-kaardid/Mullastiku-kaart-p33.html). It is a very detailed and information-rich dataset for soils in Estonia. For each soil unit, it describes the soil type, quality, texture, and layer information using a series of complex text codes. However, to use it as an input for process-based models such as SWAT, these text codes must be translated into numbers. Processing a class-based soil dataset into the required numerical variables is a time-consuming process because not all data are readily available (Bossa et al., 2012; Rahman et al., 2012). Various datasets have been created that generalise values for agricultural soils in Estonia to produce less detailed versions at scales of 1:100 000 and 1:200 000 (Kõlli et al., 2009; Tamm et al., 2018). However, there is no national scale dataset of measurements or predictions of SOC or bulk density (BD) for Estonia and no large-scale high-resolution soil database is currently available with numerical data for a range of typical eco-hydrological process-based models.

Prévost (2004) described predictions of soil properties from the SOC content, and found that SOC was closely related to soil bulk density (BD) and porosity. Suuster et al. (2011) emphasized the importance of BD as an indicator of soil quality, site productivity, and soil compaction and proposed a PTF for the organic horizon in arable soils. Van Looy et al. (2017) reviewed existing PTFs and documented the new generation of PTFs that have been developed by different disciplines of Earth system science. They emphasized that PTF development must go hand in hand with suitable extrapolation and upscaling techniques to ensure that the PTFs correctly represent the spatial heterogeneity of the soils. Abdelbaki (2018) evaluated the predictive accuracy of 48 published PTFs for predicting BD using State Soil Geographic (STATSGO) and Soil Survey Geographic (SSURGO) soil databases from the United States. They also proposed and validated a new PTF for predicting BD using SOC inputs.

However, reliable estimates for SOC have been difficult to obtain due to a lack of global data on the SOC content of each soil type (Eswaran et al., 1993). Very few SOC datasets are available for countries or regions. For example, the Northern Circumpolar Soil Carbon Database (Tarnocai et al., 2009; https://bolin.su.se/data/ncscd/) was developed to describe the SOC pools in soils of the northern circumpolar permafrost region. SOC stocks were also predicted under future climate and land cover change scenarios using a geostatistical model for predicting current and future SOC in Europe (Yigini and Panagos, 2016). Ramcharan et al. (2018) assimilate more than 200 datasets with SSURGO in order to predict various soil properties, SOC and BD, at 100-m spatial resolution for the conterminous United States using statistical machine learning. Also, SOC and soil-hydraulic predictions for Estonia need to consider that Estonia is located relatively far north and hosts large areas of peatlands.

In summary, other datasets are available for use in Estonia-based modelling contexts. However, vector-based European or even global soil datasets are very coarse and excessively generalise large parts of the diverse Estonian landscape. High-detail datasets such as Soilgrids250m are predicted data on a grid 1km/250m, based on much less data points for Estonia. In this study, we derived numerical values for the following data in all of the mapped soil units in the 1:10 000 soil map: soil type (i.e. soil reference group), texture class, soil profiles (e.g., layers, depths), texture (clay, silt, sand components, and coarse fragments), rock content, and physical variables related to the water and carbon cycle (organic carbon content, bulk density, hydraulic conductivity, available water capacity). We also describe the development of a reproducible method for deriving numerical values from the Soil Map of Estonia to support modelling and prediction of eco-hydrological processes with the popular Soil and Water Assessment Tool and we provide an extended ready-to-use dataset containing additional parameters.

2 Materials and Methods

2.1 Pre-processing and screening of the initial soil database

The source dataset – the original Soil Map of Estonia - as described in the article, is not based on modelled but on fully observed data (e.g. texture, soil profile depth, rockiness, presence of organic layer etc). Systematic mapping of Estonian soils to produce soil map in scale 1:5 000 and 1:10 000 was started in 1954 (Reintam et al., 2005), with most intensive field studies in period 1965-1969, for the main purposes of land evaluation and assessing potential for agricultural use. Generally field mapping was carried out in scale 1:10 000 but in hilly or undulating areas with higher soil diversity in scale 1:5000. In 1982-1988 older mapping data was updated and new areas were included with full-area soil quality (primarily fertility, rockiness, water regime, texture, erodability) assessment. In 1988-1990 soil field studies were performed in non-arable land

and new mapping of ameliorated land. Forest soils were mapped in period 1976-1989. During large-scale field mapping of soils, the texture was determined in situ based on organoleptic methods (feel methods) and for reference profiles laboratory analyses were performed. This enabled calibration between texture defined by organoleptic method by each researcher participating in field survey and texture determined in laboratory (Estonian Land Board, Explanation to the soil map, https://geoportaal.maaamet.ee/docs/muld/mullakaardi_seletuskiri.pdf?t=20091211092214). As a result of large-scale soil mapping, 119 soil varieties in Estonian national classification system have been distinguished and more than 500 combinations of textural status have been described. About 10,000 profiles (1 profile per 330ha) have been sampled and analysed for characterisation of mineral soils (Reintam et al., 2003, 2005). Thus, the texture codes and soil types assigned to the ca. 750000 mapped soil units (polygons) are based on many decades of in-situ land surveying practices.

The original Soil Map of Estonia is available vector layer for geographical information system software that can be downloaded from the Republic of Estonia Land Board Web site (https://www.maaamet.ee/en) in several formats under a permissive open data license. A copy with the original shapefile dataset, the related required documentation and checksums has been archived for reference (Estonian Landboard, 2017; https://datadoi.ut.ee/handle/33/103). The soil map contains the following used attribute fields:

- Soil type: a designation of the soil name, the Estonian analogue to the WRB soil reference groups
- Texture: texture classes defined for fine and coarse fragments, and to which depth the same texture and coarse fragments are observed (layer)

These attributes are encoded as “string” values, which include both letters and numbers. The important fields soil type and texture, are not just stored as standardised class values, but are instead a coded description based on abbreviations that are then combined with numbers for example depths and indictors for level of erosion, and are grouped together for different depths within the same attribute field. These description-based attribute values make it difficult to derive the foundational numerical values for sand, silt, clay and coarse fragments from the codes and to make them more consistent and usable in calculations and statistical analyses. Therefore, we performed extensive database standardisation on the original Soil Map of Estonia as the working basis and derive all further variables based from the standardised dataset sequentially. Figure 1 illustrates the four major working packages to derive the desired eco-hydrological parameters. The subsequent four sections are structured accordingly: Section 2.2 represents step A (Textural properties); step B (Additional topographic variables as predictor variables) in section 2.3, section 2.4 describes step C (SOC RF model and BD); and step D (Hydrological parameters) is described in section 2.5.

2.2 Deriving texture classes soil reference groups, and extraction of basic physical and textural values

2.2.1 Deriving sand, silt, and clay fractions and rock content

The Soil Map of Estonia’s “texture” field encodes the texture and general soil layer structure for each mapped soil unit in a structured, rule-based format (based on old Soviet-era paper maps). To derive meaningful numerical values for texture and other soil variables from the soil map, we developed a computer program that encodes these rules into a computer-readable grammar. In addition, we provided a lookup table for wrongly encoded texture codes and historical data-entry errors. The program provides a complete internal data structure that represents the analysed grammatical representation, which can be evaluated and used to generate numerical values for a variety of variables. The complete parser Python module including an extended technical description is provided as supplemental material (Kmoch et al., 2019b; "soil_lib/LoimisGrammarV2.py").

There are detailed studies on reference soil profiles in Estonia, Latvia and Lithuania that relate original soil texture, so called Katchinsky texture system (Kachinsky, 1965) to USDA soil system (Calhoun et al., 1998) and erosion modelling case studies where based on laboratory analyses transfer functions from Katchinsky to USDA texture classes were developed (Laas and Kull, 2003). The relationship between the Katchinsky and Atterberg systems were provided by R. Kask (2001).

The USDA soil taxonomy and World Reference Base soil classification systems use 12 textural classes, which are defined based on the sand, silt, and clay fractions (Ditzler et al., 2017). However, the USDA system defines fine particles as having a diameter ≤ 2 mm, whereas the Soviet-era maps use a diameter of ≤1 mm. The Soviet soil classification also mostly ignores the silt fractions, and focuses on the clay fraction (Ø ≤ 0.001 mm). Based on the available analysis data and its structure, we derived meaningful numerical texture values using a lookup table (Table 2) that represents our best efforts to account for the size difference between the USDA and Soviet systems and lack of silt data in the Soviet system. The original observations were classified into the Estonian texture code system based on Katchinsky (1965) soil particle size standards at the time of observation (not by us). We translated the texture codes back into numbers. The foundational numerical values for fine earth and coarse fragments fractions of soil are solely derived from the extracted processed Estonian texture classes as demonstrated in Table 2.

We defined the variables data for the extracted numerical texture input parameters for each layer as follows:

- SOL_CLAY# (layer 1-4): clay content (% soil weight)
- SOL_SILT# (layer 1-4): silt content (% soil weight)
- SOL_SAND# (layer 1-4): sand content (% soil weight)
- SOL_ROCK# (layer 1-4): coarse fragments content (% volumetric)

Regarding uncertainties related to that process - as we take these as observed data - achieving 5% accuracy in organoleptic determination of clay content for lower value classes while possible error increased in case of heavy texture classes. There are many finely scaled texture classes in the Estonian system. We assigned USDA texture classes based on the now defined numerical values for the sand, silt and clay fractions. Table 2 shows examples of the rules. For each texture code, the table provides the combination of sand, silt, and clay contents. These texture transfer rules to select USDA particle size distributions from the Estonian texture classes were composed by the authors according to Estonian guideline “Field Soil Survey – Muldade väliuurimine” (http://pk.emu.ee/yldinfo/uudised/uudis/2013/02/20/muldade-valiuurimine) where matches of Estonian/Soviet and USDA/FAO classes for field survey is provided. It is not possible to retrospectively redefine minor differences in boundaries between different classes between texture systems, but we consider natural variation of texture within the soil mapping unit in scale 1:10 000 more significant than that of different texture systems. In additional we introduced two more classes beyond the well-known USDA textures classes, i.e. “PEAT” and “GRAVELS”. The former states that this soil unit is a peatland, where the peat layer thickness is at least 30 cm. For hydrological modelling reasons we decided to still assign sand, silt and clay fractions to these units in order to provide a continuous hydrological soil surface. To soil units with the class “PEAT” a high clay content was assigned in order to represent the low vertical conductivity at the bottom theses peat bogs. However, for applications that critically evaluate clay content for soil units, the additional “PEAT” texture class (in the LX_TYPE1-4 variable) can be used to apply additional rules to mask these soil units accordingly. The latter class “GRAVELS” is intended to demark soil units or discrete layers therein, where only a coarse fragment type but no fine textures have been coded in the original texture codes. In these cases, depending on the type of the coarse fragment the layer can consist of gravels, large rocks or massive rock. We stored the extracted Estonian fine earth type (texture class) per layer in the variable EST_TXT# (layer 1-4). For each Estonian texture class we assigned the related USDA texture class to the parameter LX_TYPE# (layer 1-4).

Similar to the Estonian texture classes there exist Estonian stoniness classes that describe a certain type of coarse fragments within the soil profile. An additional number in connection with this rock type identifier indicates the amount/volume of these rocks in 1kg of soil. We used this indicator number to designate numerical values for the coarse fragments. Table 3 shows how we derived the rock content from the coarse fragments indicator that we obtained from the soil map encoding.

This first and fundamental step concluded with a set of variables for each mapped soil unit that include now separate standardised Estonian and English/USDA texture classes per soil layer, number and depths of layers of the mapped soil unit and numerical values for fine earth and coarse fragments fractions per layer. In addition to the evaluation of layer and depth values we assign the extracted Estonian fine earth type and the related USDA texture class per layer in the variable EST_TXT# (layer 1-4, Estonian texture class) and LX_TYPE# (layer 1-4, USDA texture class). The complete workflow is coded in the supplemental materials (Kmoch et al., 2019b; "01_soilmap_soiltypes_textures_layers.ipynb").

2.2.2 Standardising soil types and assigning WRB soil reference groups

We used the main soil types from the soil type reference sheet that accompanies the Soil Map of Estonia to derive the soil type fields in the spatial dataset and added several widely-used soil types that were not in the original reference list. We developed a short algorithm that compares the Estonian “Soil type” field and tries to find the name in a in the database (Kmoch et al., 2019a; “soil_types_error_rules_lookup.xlsx”) that captures more than 300 entries that provide a soil type substitute code from the extended 135 soil types from the soil reference list. The soil types and the Estonian soil names were then related to the FAO World Reference Base (WRB) soil reference groups (FAO, 2015) after the data have been corrected and standardised for each map unit in the extended soil dataset based on expert input (BioSoils, EUR 24729 EN, https://www.icpforests.org/pdf/manual/BioSoil/Soil_sampling_BioSoil.pdf). An exemplary except is demonstrated in Table 1. The finalised table of the standardised 135 most used main soil types is provided as supplemental spreadsheet (Kmoch et al., 2019a; “soil_types_legend.csv”).

2.2.3 Deriving depths and layers

The mapped soil units also encode variations in the soil profile within a given soil unit. In Subsection 2.2.2, we processed the textual code descriptions to compile the exact Estonian texture types in a standardised and readily available data structure. A base assumption is that most soils in Estonia were sampled to a depth of 1 m, as this is the case for a default soil profile. If larger or smaller depth information was encoded in the original soil texture code, then this would be used for the overall depth of that soil sample. For each of the layers, we can also read the analysed depth from the soil surface to the bottom of each layer. We defined parameters SOL_Z# (layer 1-4) for each layer. We defined NLAYERS (the number described of layers in the profile) and depth per layer (SOL_Z#), and we derived the maximum soil depth (SOL_ZMX), which represents the maximum depth of the soil profile (mm). We eliminated additional soil parts from the dataset if their resulting layer thickness would be zero. An additional pragmatic decision was made to exclude cumulative vertical soil parts if their depth could not be reliably inferred. For example, “*sand/loamy sand”* indicates two layers, separated by “/”. The base assumption is that the profiles have been sampled to the depth of one metre when no additional depth information is available. Therefore, for the given example, no depth information is available for the second layer (“*loamy sand*”). In these cases, we decide to drop the second layer and assign the full depth of 1m for the first layer “*sand*”. Another example is, where the first layer depth would be indicated as a range of 70-110 cm. In this case to derive a single number, the average will be taken resulting in 90 cm for the first layer. The remaining 10 cm filling up to 1m can be assigned to second layer. The Soil Map of Estonia holds depths in centimetres, and for widely used conventions we convert the depths from cm to mm in this step.

2.2.4 Evaluation and validation of extracted texture values and soil reference groups

We used two other sources of cross-validation to confirm the accuracy of the derived values. First, we used a manually “decoded” part of the Estonian Soil Map for Tartu county. Tartu County covers about 10% of Estonia and offers a representative subset of the data, as it includes many different soil types, peat bogs, forest, and arable land. It contains 83 364 records. Several members of our research group cleaned and standardised the data on soil types, textures, and depth ranges over the course of several months and collated the results in a spreadsheet. We then compared the software’s results with the manual classification results. Each soil unit in question was interpreted by at least two experts, and when their classifications differed, they discussed the difference until they achieved consensus about the correct classification.

Second, we used the SoilGrids250m.For that, we loaded and averaged the raster layers for the provided seven depths from the SoilGrids250m data for the sand, silt, and clay, coarse fragments, bulk density and soil organic carbon contents and saturated hydraulic conductivity from EU-HydroSoilGrids into the layers of the EstSoil-EH datasets.. For each parameter, we calculated descriptive statistics and plotted value distribution and an overview of the spatial pattern of the EstSoil-EH parameters against the SoilGrids250m/EU-HydroSoilGrids values to assess the differences (Table 5 and Figure 9). We observed strong similarity in the general patterns. However, the variances ranged from 10 to 30%. One main cause for this high variation is the definition of discrete values for the Estonian texture classes in our data and the more continuous distribution of raster values in the SoilGrids250m dataset.

2.3 Adding topographic variables as predictor variables

For the subsequent step of SOC prediction via the Random Forest machine-learning model, we calculated mean, median and standard deviation of several topographic and environmental variables as additional predictor variables. Topographic variables slope, Topographic Wetness Index (TWI), Terrain Ruggedness Index (TRI), and LS-factor were all calculated by using SAGA-GIS software based on a digital elevation model (Conrad et al., 2015). The LiDAR-based Digital Elevation Model with resolution 1 m was obtained from Estonian Land Board.

2.3.1 Topographic Wetness Index (TWI)

The TWI is a topo-hydrological factor proposed by Beven and Kirkby (Beven and Kirkby, 1979) and is often used to quantify topographic control on hydrological processes (Michielsen et al., 2016; Uuemaa et al., 2018) which also are relevant in the soil evolution. TWI controls the spatial pattern of saturated areas which directly affect hydrological processes at the watershed scale. Manual mapping of soil moisture patterns is often labor-intensive, costly, and not feasible at large scales. TWI provides an alternative for understanding the spatial pattern of wetness of the soil (Mokarram et al., 2015). It is a function of both the slope and the upstream contributing area:

$$TWI = \ln\left({}^{a}\!/_{\tan b}\right) \qquad (1)$$

where a is the specific upslope area draining through a certain point per unit contour length (m^2 m^{-1}), and b is the slope gradient (in degrees).

2.3.2 Terrain Ruggedness Index (TRI)

TRI reflects the soil erosion processes and surface storage capacity which again is relevant from a soil evolution perspective. The TRI expresses the amount of elevation difference between neighbouring cells, where the differences between the focal cell and eight neighbouring cells are calculated:

$$TRI = Y\left[\sum\left(x_{ij} - x_{00}\right)^2\right]^{1/2} \qquad (2)$$

where x_{ij} is the elevation of each neighbour cell to cell (0,0). Flat areas have a value of zero, while mountain areas with steep ridges have positive values.

2.3.3 LS-factor

The potential erosion in catchments can be evaluated using LS factor as used by the Universal Soil Loss Equation (USLE). The LS factor is length-slope factor that accounts for the effects of topography on erosion and is based on slope and specific catchment area (as substitute for slope length). In SAGA-GIS the calculation is based on (Moore et al., 1991):

$$LS = (n + 1)({}^{A_s}/_{22.13})^n({}^{\sin\beta}/_{0.0896})^m \qquad (3)$$

where n=0.4 and m=1.3.

2.3.4 Drainage area per mapped soil unit

In addition, we calculated the area per mapped soil unit in m^2 (area_drain) and in percent of area, which is under drainage (drain_pct). The drainage regimen considered both underground tile drainage and ditch based drainage systems. The drainage information used for this was compiled based on the Estonian Topographic Data Set (ETAK) and the official register of drainage systems (https://portaal.agri.ee/avalik/#/maaparandus/msr/systeemi-otsing) managed by the Agricultural Board of Ministry of Rural Affairs of Estonia. All the variables were calculated using the GIS software packages QGIS and SAGA.

2.4 Predicting Soil Organic Carbon (SOC) and Bulk Density (BD)

The main information described in the Soil Map of Estonia is the soil type and the soil texture. However, soil hydraulic properties and SOC data are needed for many different applications in soil hydrology. Pedotransfer functions (PTFs) have proven to be useful to indirectly estimate these parameters from more easily obtainable soil data (Van Looy et al., 2017). Therefore, several soil parameters like soil organic carbon, bulk density and saturated hydraulic conductivity must be derived via PTFs and other data assimilation methods. To apply PTFs and other data-assimilation methods, third-party datasets can be used as secondary sources. In the previous steps we have prepared a wide set of input variables, including the numerical fractions for the textural properties, standardised classes for soil type and soil textures, and additional topographic variables, which we can apply as predictor variables to model the value distribution for SOC and BD. We develop these two extended soil physical input parameters as organic carbon content in % soil weight (SOL_CBN# layer 1-4), and dry bulk density in Mg/m^3 or g/cm^3 (SOL_BD# layer 1-4).

In order to map the spatial distribution of SOC in Estonia a machine learning model random forest (RF) was used to predict SOC based on parameters derived from the soil map. RF was preferred to more advanced ML algorithms (e.g., neural networks) because it has shown to be relatively resilient towards data noise and not require preliminary hyperparameter tuning (Breiman, 2001; Caruana and Niculescu-Mizil, 2006). In addition, feature importances can be extracted from the model to determine the most influential predictor variables.

For training, we used measurements of soil organic matter (SOM) or soil organic carbon (SOC) from forest areas (samples sizes: n=100), 4 datasets of samples from Estonian open and overgrown alvars and grasslands (n: 94, 137, 146, 69), peatlands (n=175) and from arable soils transects (n=8964) resulting in 3373 distinct point locations (Kriiska et al., 2019; Noreika et al., 2019; Suuster et al., 2011). Where necessary, the SOM values were translated into SOC via: *SOC* = *SOM* / 1.724. Many samples from peatlands and arable fields were often sampled within the same mapped soil unit. For these soil units (polygons) the respective soil measurement data was averaged and joined to the respective soil units to reduce the bias of the prediction. After joining the sample size reduced to the 397 distinct training samples for machine learning (Figure 2).

This data was then randomly split into training (60%) and test (40%) sets and the model was evaluated by predicting SOC based on the predictor variables of the test set. Finally, the model was applied to soil map polygons without available SOC measurements to predict SOC content in Estonian soils.

Subsequently, we calculated soil bulk density based on predicted soil organic carbon for each layer in each mapped soil unit polygon, with following PTF (Adams, 1973; Kauer et al., 2019), which has been successfully applied in Estonia:

$$\mathrm{BD} = 1 / (0.03476 \times SOM + 0.6098) \qquad (4)$$

where: $SOM = SOC \times 1.724$

The conversion factor 1.72 is a widely used universal value. However, we acknowledge that the real value varies slightly between soils.

2.5 Assimilation of additional hydrological variables

In order for this dataset to be more useful in eco-hydrological modelling we developed and added two additional hydrological variables. Saturated hydraulic conductivity (K_{sat}) relates soil texture to a hydraulic gradient and is quantitative measure of water movement through a saturated soil. In addition to the ability of transmitting water along a hydraulic gradient we also add available water capacity (AWC) as a variable. AWC describes the soil’s ability to hold water and quantifies how much of that water is available for plants to grow. We develop two variables saturated hydraulic conductivity (mm/hr, SOL_K# layer 1-4), and available water capacity of the soil layer (mm H_2O/mm soil, SOL_AWC# layer 1-4). We calculated K_{sat} using the improved Rosetta3 software, which implements a pedotransfer model with improved estimates of hydraulic parameter distributions (Zhang and Schaap, 2017). It is based on an artificial neural network (ANN) for the estimation of water retention parameters, saturated hydraulic conductivity, and their uncertainties. For each standardised texture class from Table 2, we used the numerical fine earth fractions for sand, silt and clay as inputs for the Rosetta3 software and calculated K_{sat} for each layer in each mapped soil unit polygon. We provide a copy of the Rosetta3 program in the supplemental materials.

In order to calculate available water capacity**,** we summarized the field capacity (FC, at −330 cm matric potential −0.03 MPa) and wilting point (WP, at −15,848 cm matric potential −1.5 MPa) variables of the 7 soil depths of the EU-SoilHydroGrids 250m resolution raster datasets (Tóth et al., 2017) for each mapped soil unit for the provided depths of 0, 5, 15, 30, 60, 100, and 200 cm. The available water capacity is then calculated for each of the 7 depths by a raster calculation: AWC = FC - WP (Dipak and Abhijit, 2005). The resulting 7 AWC raster layers are then averaged into the respective depth ranges for each of the discrete layers of the Estonian mapped soil units.

3 Results

In this study, we developed a Python module that is capable of analysing, extracting, and standardising the soil type and soil texture data from the official Soil Map of Estonia into a reusable, reproducible soil dataset that uses World Reference Base and FAO soil classes and texture descriptions. Figure 4 shows a map of the classified topsoil texture classes derived from the original Estonian texture codes. In addition, it shows the peat soils that cover up to 20% of Estonia, and are an important soil type in such northern countries.

To make such soil information usable in an eco-hydrological modelling context, we derived numerical values for each of the soil units. These values include the number of discretized soil layers (NLAYERS) - a maximum of 4 separate vertical distinct soil layers where described in the original texture codes –the depth of each layer (SOL_Z1-4), and the maximum depth of the sampled profile for each mapped soil unit (SOL_ZMX). Based on the layer information and the extract texture classes we could define the percent fractions per volume of sand (SOL_SAND1-4), silt (SOL_SILT1-4), clay (SOL_CLAY1-4), and coarse fragments (SOL_ROCK1-4) per layer. Figure 5 shows the percent fractions for sand, silt, clay and coarse fragments for the top layer. Table 6 contains the full list of variable and parameters per mapped soil unit contained in the EstSoil-EH dataset.

3.1 Validation of soil type and texture classes extraction and standardisation

For the main soil types, we achieved 97.7% agreement between the software's result and the manual classification. The manual verification of the validation revealed several re-labelling issues from the error lookup table. A visual assessment by two soil sciences senior research staff asserted that the level of similarity of the soil types that were selected by the automated process were closely related. However, the mismatches (1943 records, equivalent to 2.3% of the total records) indicated that the soil experts tended to interpret "errors" based on personal knowledge that may not be reproducible in a strictly automated fashion. For example, some landforms (e.g. eroded material filling low slopes or collapsed cliffs) were originally classified as exceptions to the general classification rule based on the local knowledge of the landscape. When standardising these expert interpretations with the same more general soil type, we reduced the number of mismatched soil type identifiers to 0. Furthermore, it should be emphasised that humans tend to make mistakes when performing repetitive procedures. Therefore, we consider the high accuracy (97.7%) to be a very good result.

For the validation of textures, we used several steps. First, given the high agreement between the software-generated codes and the human-generated codes, we accepted the software's texture codes for use in our subsequent evaluations. Next, we compared the extracted main texture for each layer with the manually coded value:

- 77 870 of 83 364 records (93.4%) showed identical parsing of the full texture code
- 71 635 of the records (85.9%) showed identical interpretation of the first layer's texture type (10 312 records were differently coded, and 1417 produced "no value" errors, in which either the source or validation dataset contained no value, preventing a comparison with the other dataset's value)

- 65 000 of the records (78.0%) showed identical interpretation of the second layer's texture (with 2325 differently coded textures, and 16 038 "no value" errors, of which 15 461 occurred in the automated processed new dataset, and only 577 occurred in the validation dataset)
- 82 507 of the records (99.0%) showed identical interpretation of the third layer's texture (with most errors caused by a non-existent third layer, 334 differently coded, and 523 with a "no value" error)

For sand, silt and clay fractions we could obtain laboratory analysis **only for forest soil samples**. We calculated the root mean squared error (RMSE) and chose the Normalized Median Absolute Deviation (nMAD) as an additional measure of dispersion of error for non-gaussian distributed data:

- RMSE for sand: 13.1 %
- nMAD for sand: 9.68
- RMSE for silt: 10.7 %
- nMAD for silt: 7.0
- RMSE for clay: 6.5 %
- nMAD for clay: 3.9

Our manual assessment of the mismatches indicated the same problem that occurred with the soil types. The expert assessments aimed to keep as much information as possible available in their decoded classification, and this did not always agree with the automated processing rules. Furthermore, the complexity of the Estonian texture rules and the reliance on human judgement creates high uncertainty in some cases, even for human interpretation. In addition, to derive the grammar rules, we added a few simplifying elements, such as omitting some rarely used additional information in the soil texture descriptions. For example, the Estonian rules allow specification of several soil parts, but as a horizontal distribution within the same mapped soil unit rather than as vertical layers. This is understandably complex, making it difficult to classify this variable soil as a single soil unit. Consequently, it is inevitable that some of these descriptions will not agree with the software's classification.

3.2 SOC prediction and validation of Random Forest model

We also calculated several extended soil properties, i.e. *SOC* content and *BD*. The RF regression model was implemented with the RandomForestRegressor function from the Scikit-learn Python library. The model was evaluated by predicting SOC based on the predictor variables of the test set for the 60:40 split. Figure 3 illustrates the cross-validation scatterplots of observed vs. predicted SOC values for the test/validation sample splits. Following characteristics are reported for the chosen RF model:

- coefficient of determination (R^2) score: 0.69
- score of the training dataset with out-of-bag estimate (oob score): 0.58
- Pearsons *r correlation coefficient*, training: 0.90, validation: 0.83

RF feature importances, top 6:

- Clay content (SOL_CLAY1): 0.65

- Terrain Roughness Index, standard deviation (tri_stdev): 0.04
- Sand content (SOL_SAND1): 0.03
- LS-factor, median (ls_median): 0.028
- Area under drainage in percent (drain_prct): 0.027
- Coarse fragments rock content (SOL_ROCK1): 0.024

Figure 6 shows the predicted values of SOC and BD for the top layer. On visual inspection the spatial distribution for the SOC content matches comparatively well with known agricultural areas, where low carbon content prevails, as well as with the peat land areas, which have a very high carbon content.

For further description and guidance on errors in the predictions for SOC and BD we calculated the RMSE and nMAD as an additional measure of dispersion of error for non-gaussian distributed data. BD observed data was only available for arable lands and forest soil samples, and should be treated accordingly.

- RMSE for SOC predictions: 2.95 %
- nMAD for SOC: 1.44
- RMSE for the subsequent BD predictions with PTF: 0.33 g/cm³
- Normalized Median Absolute Deviation (nMAD) for BD: 0.15

However, due to the small number and distribution of input samples over four distinct landscapes, namely arable lands, wetlands, forests and open/grass lands, we broke down the error distribution for these for land forms in Table 6. The prediction error characteristics differ, with the smallest errors for arable lands, then wetlands and the largest errors for open grasslands and forest.

3.3 Hydrological variables results

Based on the variables derived in previous steps, we could calculate saturated hydraulic conductivity (K_{sat}) based on the sand, silt and clay content. Available water capacity (AWC) was calculated solely by aggregating EU-SoilHydroGrids data of field capacity and wilting point. Figure 7 shows the spatial distribution of K_{sat} andAWC for the first layer. Rosetta reports the standard deviation for its internal prediction process, which draws many samples for the same input of sand, silt and clay content and then provides the mean as the predicted value for K. The summary of the predicted K_{sat} values and the standard deviation is summarized in Table 7. For peat areas and wetlands the predicted values also corresponds with ranges reported in the literature for the sand-silt-clay ratios provided (Gafni et al., 2011). We compiled all parameters into a dataset that can now be easily used with SWAT or other eco-hydrological and land-use-change models. As we are not changing the general geometry or underlying spatial data model of the original soil map, all parameters are only added to the existing mapped soil units and thus, all original soil polygons remain discernible.

4 Discussion and Future Work

The Soil Map of Estonian is a valuable resource for hydrological, ecological and agricultural studies. It is widely used in Estonia. But before our analysis, a large amount of the dataset's information was not readily usable beyond the field or farm-scale because of the need to manually interpret the specialised soil types and the complexity of the rules that describe the texture or other characteristics of the soil units. The developed dataset is of very high spatial detail based on the original Estonian national soil map, which was created from directly surveying all of Estonia. Thus, our presented dataset holds to potential to further improve our understanding of eco-hydrological processes in the landscape through the use of advanced numerical and process-based models. The derived information is much more spatially related to the landform/landuse observed there than any other dataset covering soil information for Estonia. Furthermore, the textures and SOC/BD values are directly derived from reliable observed data samples from Estonia, with a reproducible workflow which is unique in the case of Estonian soil datasets, whereas this is not true for many other reported soil datasets that cover the area of Estonia. Furthermore, the open access availability and transparency of measurement data can provide a reliable building block for advancing studying soil and hydrological processes in Estonia into temporal aspects. Especially, properties such as SOC and BD will vary extensively depending on the land use and land cover. In combination which developments that capture also the dynamics of land use change and adaptations under climate change, the evolution of the soils in Estonia could more readily be investigated.

The method created to translate original hard copy soil map (with traditional textual codes) to a digitally readable numerical GIS-based map can be used by several other countries (e.g. Latvia, Lithuania, Ukraine etc) which enables spatially more explicit modelling of ecosystems. We used a multi-step approach to derive a generalised and standardised numerical dataset that will have many potential applications for users of Estonian soil information, including support of economic, agricultural, or environmental management and input for decision-making and to support more reproducible scientific research based on Estonian soil information. Until the present analysis, numerical and process-based eco-hydrological modelling with tools and models like SWAT or the Regional Hydro-Ecologic Simulation System (RHESSys) were greatly hampered by the need to manually derive useful values from the Soil Map of Estonia to support the modelling.

One challenge in terms of validation is that the datasets we used for validation, e.g. SoilGrids and EU-SoilHydroGrids are informed to some extent by samples of Estonian soil characteristics that are not necessarily more accurate than the results of our classification. Although we accounted for this problem by providing additional comparisons, the scale mismatch between continuous raster datasets and polygon-based data inevitably introduced errors and trade-offs into the comparison. One solution to these problems would be to perform supplemental field sampling to ground-truth the source data and confirm the accuracy of our model's classification based on the field data.

From the point of end-user, the first layer is not a default 30 cm deep top soil layer. A direct interpretation of the derived discrete layer information as soil horizons should not be generalized but checked on per case basis. All physical, chemical and hydraulic properties are based on the analysis of the original texture code per mapped soil units and the resulting discrete layers per unit. This is an important usage constraint, for example in sense of biological activity, as 30 cm soil layer is most active, but for each soil unit it needs to be checked which layers extend into which actual depths. Also the *SOC* content and *BD* are not modelled in a vertical continuum but per discrete values per unit and layer. However, fertile soils, like Luvisols contain a lot of *SOC* also in deeper layers. But such additional expert knowledge is not encoded in original Soilmap of Estonia, nor in the processing algorithms that derived the extended parameters for this newly generated dataset. However, such additional knowledge, as well as more appropriate models for peatland areas, could be included as additional rules in a subsequent improvement of this dataset.

Kõlli et al. (2009) published estimates of the SOC stocks for forests, arable lands, and grasslands and for all of Estonia. Nevertheless, they constrained their finding by noting that their estimates were calculated based on the mean SOC stock for each soil type and the corresponding area in which the soil type was distributed. Putku (2016) used the large-scale Soil Map of Estonia at the polygon level for SOC stock modelling for mineral soils in arable land of Tartu county. Carbon content calculations in Estonia have historically been predominantly made for soils in agricultural areas. Exisiting literature and our results in summary are in line with SOC distribution per soil type in mineral soils in arable lands (E. Suuster et al. 2012).

The original purpose of this dataset was to derive values for hydrological modelling purposes and at the same time to stay as close to the original data as possible. From that perspective peat soil units are currently modelled with assumptions to have a similar behaviour to clay hydrologically. Therefore, the spatial distribution of clay percentage in particular, but also the concurrent physical fractions of sand and silt do not make scientific sense for these areas where peat is prevalent. In order to make the dataset as useful as possible and to identify peatland areas, we introduced the additional class “PEAT” into the USDA classification. While sand, silt, clay and rock content are directly derived values from the original texture codes, *SOC* and K_{sat} are modelled via statistical machine-learning algorithms, which include additional uncertainty. This should be considered when evaluating *BD* , which are calculated using *SOC* as an input variable.

The only variable which we did not model based in dependence of already modelled parameters is *AWC*. Here we summarised the EU-SoilHydroGrids 250m (Tóth et al., 2017) raster datasets for *FC* and *WP* as inputs an external data integration. This is not ideal and can be considered a trade-off between introducing too much uncertainty and an external un-related data source.

In the future, we foresee step-wise improvement of our software by developing better PTFs to estimate parameters and to better integrate the presence of peat soils and other specific landscapes and environments in Estonia. Furthermore, statistical machine-learning or neural network and deep learning methods could be tested in order to improve soil classifications and express more complex relationships between soil types and textures. Currently, one specificity of the newly created EstSoil-EH dataset is its discrete nature, as we are only adding derived numerical variables to the existing mapped soil units (polygons). We do not predict a continuous surface in this study, thus, comparisons with continuous surface parameters predicitons such as in SoilGrids (Hengl et al., 2017) or EU-SoilHydroGrids (Tóth et al., 2017), are not directly possible. However, the workflow could potentially be extended also for creating continuous surface. With appropriate modification (e.g., to use the soil characteristic codes more consistently for a different country), our methodology could also be applied in other countries such as Lithuania or Latvia that share similar historical land- and soil surveying practices.

Code and data availability.

The described "EstSoil-EH v1.0" dataset including all supplemental tables and figures is deposited on Zenodo, doi:10.5281/ZENODO. 3473289 (Kmoch et al., 2019a). Supplemental software and codes that were used, e.g. the texture-code parsing scripts, the machine learning model and the parameter calculation Jupyter notebooks are maintained on GitHub (https://github.com/LandscapeGeoinformatics/EstSoil-EH_sw_supplement/releases) and were also deposited on Zenodo, doi: 10.5281/zenodo.3473209 (Kmoch et al., 2019b). The original National Soil Map of Estonia (https://geoportaal.maaamet.ee/est/Andmed-ja-kaardid/Mullastiku-kaart-p33.html) was archived for reference on the DataCite- and OpenAire-enabled repository of the University of Tartu, DataDOI, doi:10.15155/re-72 (Estonian Landboard, 2017).

Author contributions.

A. Kmoch designed the experiments and code. A. Kmoch, E. Uuemaa, A. Astover, A. Kanal validated soil types and texture values. E. Uuemaa completed terrain analysis and statistics. A. Kull, A. Helm, M. Pärtel, I. Ostonen cleaned and organized the input SOC datasets. H. Virro developed the SOC RF machine-learning code and experiment. A. Kmoch and E. Uuemaa drafted the manuscript and created the figures, and all authors contributed to the paper writing.

Competing interests.

The authors declare that they have no conflict of interest.

Special issue statement.

This article is part of the special issue "Linking landscape organisation and hydrological functioning: from hypotheses and observations to concepts, models and understanding". It is not associated with a conference.

Acknowledgements.

Financial support.

This research has been supported by the Marie Skłodowska-Curie Actions individual fellowships under the Horizon 2020 Programme grant agreement number 795625, the Mobilitas Pluss Postdoctoral Researcher Grant numbers MOBJD233 and

PRG352 of the Estonian Research Council (ETAG), the European Regional Development Fund (Centre of Excellence EcolChange), and by the Estonian Environmental Investment Centre.

References

Abdelbaki, A. M.: Evaluation of pedotransfer functions for predicting soil bulk density for U.S. soils, Ain Shams Eng. J., 9(4), 1611–1619, doi:10.1016/j.asej.2016.12.002, 2018.

Adams, W. A.: The Effect of Organic Matter on the bulk and true Densities of some Uncultivated Podzolic Soils, J. Soil Sci., 24(1), 10–17, doi:10.1111/j.1365-2389.1973.tb00737.x, 1973.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. and Williams, J. R.: Large area hydrologic modeling and assessment part I: model development, J. Am. Water Resour. Assoc., 34(1), 73–89, 1998.

Batjes, N. H.: A world dataset of derived soil properties by FAO-UNESCO soil unit for global modelling, Soil Use Manag., 13(1), 9–16, doi:10.1111/j.1475-2743.1997.tb00550.x, 1997.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology, Hydrol. Sci. Bull., 24(1), 43–69, doi:10.1080/02626667909491834, 1979.

Bossa, A. Y., Diekkrüger, B., Igué, A. M. and Gaiser, T.: Analyzing the effects of different soil databases on modeling of hydrological processes and sediment yield in Benin (West Africa), Geoderma, 173–174, 61–74, doi:10.1016/j.geoderma.2012.01.012, 2012.

Breiman, L.: Random Forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

Calhoun, T. E., Ellermäe, O., Kõlli, R., Lemetti, I., Penu, P. and Smith, C. W.: Benchmark Soils of Estonia Researched thru Baltic –American Collaboration. Problems of Estonian Soil Classification, Trans. Est. Agric. Univ., 198, 76–114, 1998.

Caruana, R. and Niculescu-Mizil, A.: An Empirical Comparison of Supervised Learning Algorithms, in Proceedings of the 23rd International Conference on Machine Learning, pp. 161–168, ACM, New York, NY, USA., 2006.

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V. and Böhner, J.: System for Automated Geoscientific Analyses (SAGA) v. 2.1.4, Geosci. Model Dev., 8(7), 1991–2007, doi:10.5194/gmd-8-1991-2015, 2015.

Cordeiro, M. R. C., Lelyk, G., Kröbel, R., Legesse, G., Faramarzi, M., Masud, M. B. and McAllister, T.: Deriving a dataset for agriculturally relevant soils from the Soil Landscapes of Canada (SLC) database for use in Soil and Water Assessment Tool (SWAT) simulations, Earth Syst. Sci. Data, 10(3), 1673–1686, doi:10.5194/essd-10-1673-2018, 2018.

Dejanović, I., Milosavljević, G. and Vaderna, R.: Arpeggio: A flexible PEG parser for Python, Knowledge-Based Syst., 95, 71–74, doi:10.1016/j.knosys.2015.12.004, 2016.

Dipak, S. and Abhijit, H.: Physical and Chemical Methods in Soil Analysis -, New Age International Ltd., New Delhi., 2005.

Ditzler, C., Scheffe, K. and Monger, H. C.: Soil survey manual. USDA Handbook 18, Soil Science Division. Government

Printing Office, Washington, D.C., 2017.

Dobos, E., Daroussin, J. and Montanarella, L.: An SRTM-based procedure to delineate SOTER Terrain Units on 1: 1 and 1: 5 million scales, in European Commission Directorate General, Joint Research Centre. EUR 21571 EN, p. 55, Luxembourg. [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.397.1892&rep=rep1&type=pdf, 2005.

Douglas-Mankin, K. R. R., Srinivasan, R. and Arnold, J. G. J. G.: Soil and Water Assessment Tool (SWAT) model: current developments and applications, Trans. Am. Soc. Agric. Biol. Eng., 53(5), 1423–1431, doi:10.13031/2013.34915, 2010.

Estonian_Landboard: Soilmap of Estonia - Mullastiku kaart, , doi:http://dx.doi.org/10.15155/re-72, 2017.

Eswaran, H., Van Den Berg, E. and Reich, P.: Organic Carbon in Soils of the World, Soil Sci. Soc. Am. J., 57(1), 192, doi:10.2136/sssaj1993.03615995005700010034x, 1993.

FAO: World reference base for soil resources 2014 International soil classification system., 2015.

FAO, I.: Guidelines for soil description 3rd Edition, Soil Resour. Manag., 1990.

Fischer, G., Nachtergaele, F., Prieler, S., van Velthuizen, H. T., Verelst, L. and Wiberg, D.: Global Agro-ecological Zones Assessment for Agriculture (GAEZ 2008), in IIASA, Laxenburg, Austria and FAO, Rome, Italy., 2008.

Gafni, A., Malterer, T., Verry, E., Nichols, D., Boelter, D. and Päivänen, J.: Physical Properties of Organic Soils, in Peatland Biogeochemistry and Watershed Hydrology at the Marcell Experimental Forest, edited by R. Kolka, S. Sebestyen, E. S. Verry, and K. Brooks, pp. 135–176, CRC Press, Boca Raton, FL., 2011.

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.

Kachinsky, N.: Fizika potchv. Soil physics. In Russian, Vol. 1. in., Moscow University Press, Moscow., 1965.

Kask, R.: On the English Equivalents of the Estonian Terms for the Textural Classes of Estonian Soils, J. Agric. Sci., 14, 93–96 [online] Available from: http://agrt.emu.ee/pdf/proceedings/toim_2001_14_kaskr.pdf, 2001.

Kauer, K., Astover, A., Viiralt, R., Raave, H. and Kätterer, T.: Evolution of soil organic carbon in a carbonaceous glacial till as an effect of crop and fertility management over 50 years in a field experiment, Agric. Ecosyst. Environ., 283, 106562, doi:10.1016/j.agee.2019.06.001, 2019.

Kmoch, A., Kanal, A., Astover, A., Kull, A., Virro, H., Helm, A., Pärtel, M., Ostonen, I. and Uuemaa, E.: EstSoil-EH v1.0: An eco-hydrological modelling parameters dataset derived from the Soil Map of Estonia (data deposit), doi:10.5281/zenodo.3473289, 2019a.

Kmoch, A., Virro, H. and Uuemaa, E.: EstSoil-EH v1.0 software supplement release for deposit, doi:10.5281/zenodo.3473209, 2019b.

Kõlli, R., Ellermäe, O., Köster, T., Lemetti, I., Asi, E. and Kauer, K.: Stocks of organic carbon in Estonian soils, Est. J. Earth Sci., 58(2), 95, doi:10.3176/earth.2009.2.01, 2009.

Kriiska, K., Frey, J., Asi, E., Kabral, N., Uri, V., Aosaar, J., Varik, M., Napa, Ü., Apuhtin, V., Timmusk, T. and Ostonen, I.:

Variation in annual carbon fluxes affecting the SOC pool in hemiboreal coniferous forests in Estonia, For. Ecol. Manage., 433, 419–430, doi:10.1016/j.foreco.2018.11.026, 2019.

Laas, A. and Kull, A.: Sustainable Planning and Development, edited by A. G. K. E. Beriatos, C.A. Brebbia, H. Coccossis, Boston: Wessex Institute of Techonology Press, Southampton., 2003.

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.

Michielsen, A., Kalantari, Z., Lyon, S. W. and Liljegren, E.: Predicting and communicating flood risk of transport infrastructure based on watershed characteristics, J. Environ. Manage., doi:10.1016/j.jenvman.2016.07.051, 2016.

Mokarram, M., Roshan, G. and Negahban, S.: Landform classification using topography position index (case study: salt dome of Korsia-Darab plain, Iran), Model. Earth Syst. Environ., doi:10.1007/s40808-015-0055-9, 2015.

Moore, I. D., Grayson, R. B. and Ladson, A. R.: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications, Hydrol. Process., doi:10.1002/hyp.3360050103, 1991.

Noreika, N., Helm, A., Öpik, M., Jairus, T., Vasar, M., Reier, Ü., Kook, E., Riibak, K., Kasari, L., Tullus, H., Tullus, T., Lutter, R., Oja, E., Saag, A., Randlane, T. and Pärtel, M.: Forest biomass, soil and biodiversity relationships originate from biogeographic affinity and direct ecological effects, Oikos, oik.06693, doi:10.1111/oik.06693, 2019.

Panagos, P., Liedekerke, M. van, 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.

Piniewski, M., Szcześniak, M., Huang, S. and Kundzewicz, Z. W.: Projections of runoff in the Vistula and the Odra river basins with the help of the SWAT model, Hydrol. Res., 49(2), 303–317, doi:10.2166/nh.2017.280, 2018.

Prévost, M.: Predicting Soil Properties from Organic Matter Content following Mechanical Site Preparation of Forest Soils, Soil Sci. Soc. Am. J., 68(3), 943, doi:10.2136/sssaj2004.9430, 2004.

Putku, E.: Prediction models of soil organic carbon and bulk density of arable mineral soils, Estonian University of Life Sciences., 2016.

Rahman, M., Bolisetti, T. and Balachandar, R.: Hydrologic modelling to assess the climate change impacts in a Southern Ontario watershed, Can. J. Civ. Eng., 39(1), 91–103, doi:10.1139/l11-112, 2012.

Ramcharan, A., Hengl, T., Nauman, T., Brungard, C., Waltman, S., Wills, S. and Thompson, J.: Soil Property and Class Maps of the Conterminous United States at 100-Meter Spatial Resolution, Soil Sci. Soc. Am. J., 82(1), 186–201, doi:10.2136/sssaj2017.04.0122, 2018.

Reintam, L., Kull, A., Palang, H. and Rooma, I.: Large-Scale Soil Maps and a Supplementary Database for Land Use Planning in Estonia, J. Plant Nutr. Soil Sci. Fur Pflanzenernahrung Und Bodenkd., 166(2), 225−231, 2003.

Reintam, L., Rooma, I., Kull, A. and Kõlli, R.: Soil information and its application in Estonia, in Research report, vol. 9, edited by European Soil Bureau, pp. 121–132., 2005.

Suuster, E., Ritz, C., Roostalu, H., Reintam, E., Kõlli, R. and Astover, A.: Soil bulk density pedotransfer functions of the humus horizon in arable soils, Geoderma, 163(1–2), 74–82, doi:10.1016/j.geoderma.2011.04.005, 2011.

Tamm, O., Luhamaa, A. and Tamm, T.: Modeling future changes in the North-Estonian hydropower production by using SWAT, Hydrol. Res., 47(4), 835–846, doi:10.2166/nh.2015.018, 2016.

Tamm, O., Maasikamäe, S., Padari, A. and Tamm, T.: Modelling the effects of land use and climate change on the water resources in the eastern Baltic Sea region using the SWAT model, CATENA, 167, 78–89, doi:10.1016/j.catena.2018.04.029, 2018.

Tarnocai, C., Canadell, J. G., Schuur, E. A. G., Kuhry, P., Mazhitova, G. and Zimov, S.: Soil organic carbon pools in the northern circumpolar permafrost region, Global Biogeochem. Cycles, 23(2), doi:10.1029/2008GB003327, 2009.

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

Uuemaa, E., Hughes, A. O. and Tanner, C. C.: Identifying feasible locations for wetland creation or restoration in catchments by suitability modelling using light detection and ranging (LiDAR) Digital Elevation Model (DEM), , 10(4), doi:10.3390/w10040464, 2018.

Wösten, J. H. ., Lilly, A., Nemes, A. and Le Bas, C.: Development and use of a database of hydraulic properties of European soils, Geoderma, 90(3–4), 169–185, doi:10.1016/S0016-7061(98)00132-3, 1999.

Yigini, Y. and Panagos, P.: Assessment of soil organic carbon stocks under future climate and land cover changes in Europe, Sci. Total Environ., 557–558, 838–850, doi:10.1016/J.SCITOTENV.2016.03.085, 2016.

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, 39–53, doi:10.1016/j.jhydrol.2017.01.004, 2017.

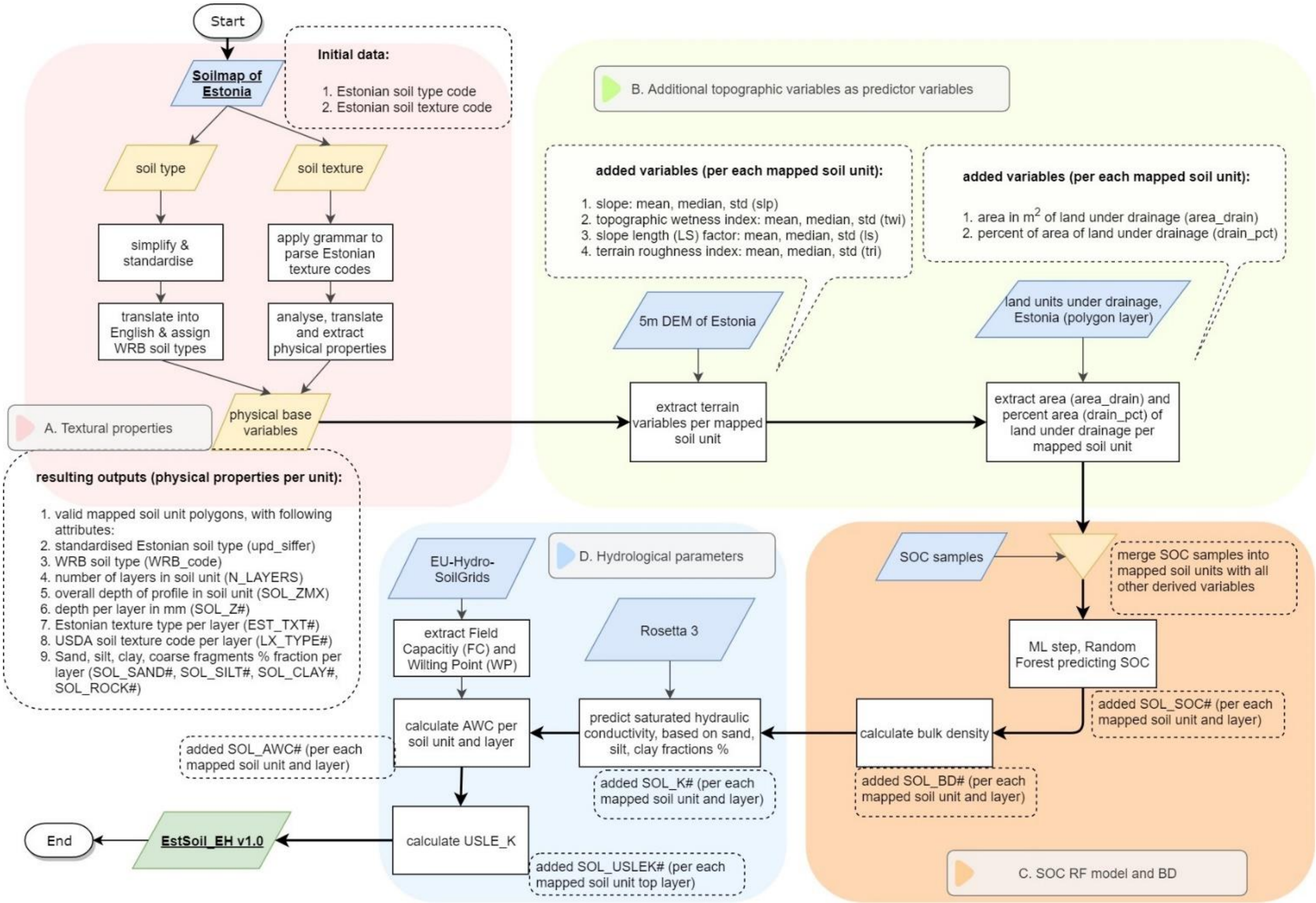

Start
Soilmap of Estonia
Initial data:
1. Estonian soil type code
2. Estonian soil texture code
soil type
soil texture
simplify & standardise
apply grammar to parse Estonian texture codes
translate into English & assign WRB soil types
analyse, translate and extract physical properties
physical base variables
A. Textural properties
resulting outputs (physical properties per unit):
1. valid mapped soil unit polygons, with following attributes:
2. standardised Estonian soil type (upd_siffer)
3. WRB soil type (WRB_code)
4. number of layers in soil unit (N_LAYERS)
5. overall depth of profile in soil unit (SOL_ZMX)
6. depth per layer in mm (SOL_Z#)
7. Estonian texture type per layer (EST_TXT#)
8. USDA soil texture code per layer (LX_TYPE#)
9. Sand, silt, clay, coarse fragments % fraction per layer (SOL_SAND#, SOL_SILT#, SOL_CLAY#, SOL_ROCK#)
B. Additional topographic variables as predictor variables
added variables (per each mapped soil unit):
1. slope: mean, median, std (slp)
2. topographic wetness index: mean, median, std (twi)
3. slope length (LS) factor: mean, median, std (ls)
4. terrain roughness index: mean, median, std (tri)
added variables (per each mapped soil unit):
1. area in m^2 of land under drainage (area_drain)
2. percent of area of land under drainage (drain_pct)
5m DEM of Estonia
land units under drainage, Estonia (polygon layer)
extract terrain variables per mapped soil unit
extract area (area_drain) and percent area (drain_pct) of land under drainage per mapped soil unit
SOC samples
merge SOC samples into mapped soil units with all other derived variables
ML step, Random Forest predicting SOC
added SOL_SOC# (per each mapped soil unit and layer)
calculate bulk density
added SOL_BD# (per each mapped soil unit and layer)
C. SOC RF model and BD
D. Hydrological parameters
EU-Hydro-SoilGrids
Rosetta 3
extract Field Capacitiy (FC) and Wilting Point (WP)
calculate AWC per soil unit and layer
predict saturated hydraulic conductivity, based on sand, silt, clay fractions %
added SOL_AWC# (per each mapped soil unit and layer)
added SOL_K# (per each mapped soil unit and layer)
calculate USLE_K
added SOL_USLEK# (per each mapped soil unit top layer)
EstSoil_EH v1.0
End

Figure 1: Flowchart for executed processing steps grouped into the four major work packages

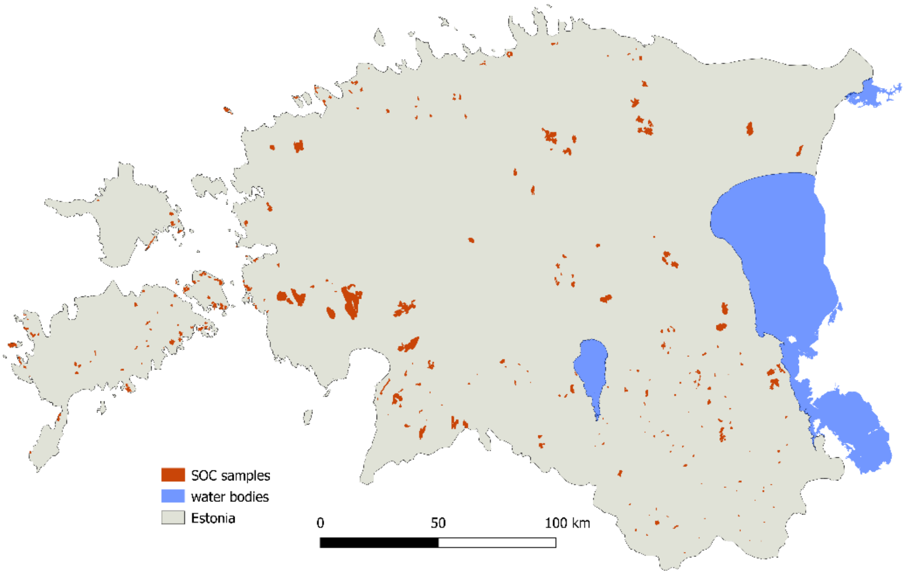


Figure 2: Distinct soil unit polygons including all sampling locations for the ML training sample.

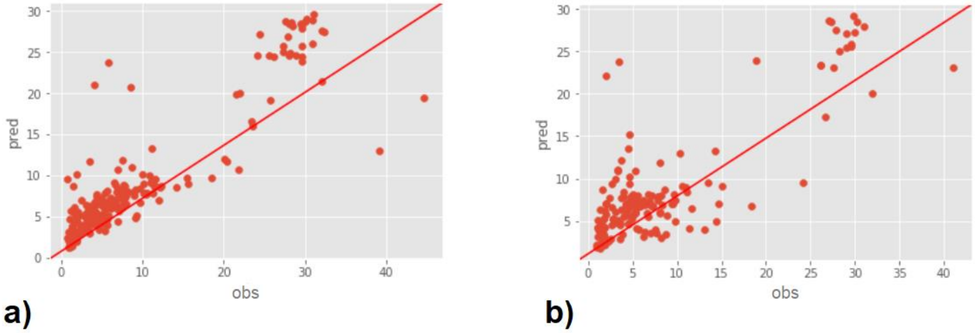


Figure 3: Random Forest model cross-validation scatterplot of observed vs. predicted SOC values for the test/validation sample splits: a) training subsample and b) validation subsample

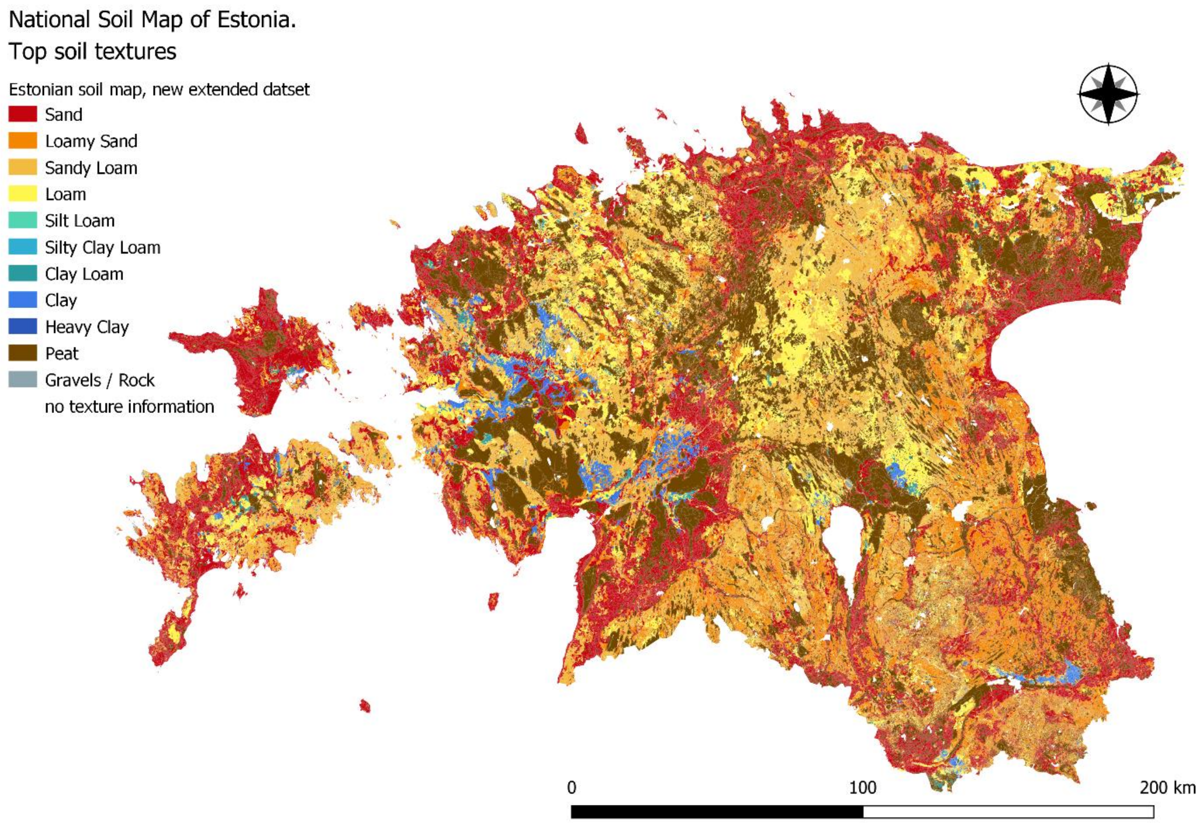


Figure 4: USDA topsoil textures derived from the original Estonian texture codes by the software developed in the present study, including additional classes "PEAT" and "GRAVELS".

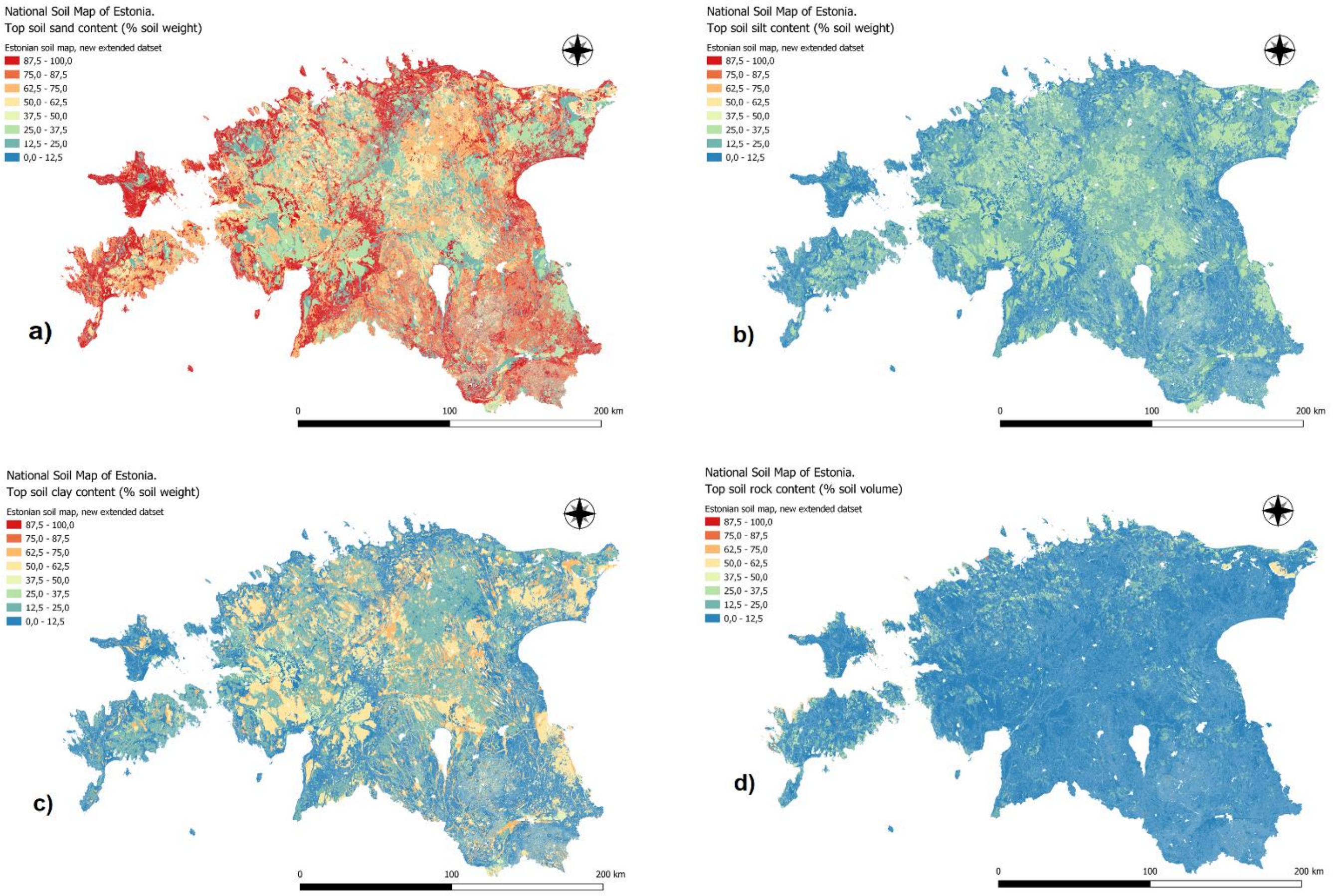

National Soil Map of Estonia.
Top soil sand content (% soil weight)
Estonian soil map, new extended datset
87,5 - 100,0
75,0 - 87,5
62,5 - 75,0
50,0 - 62,5
37,5 - 50,0
25,0 - 37,5
12,5 - 25,0
0,0 - 12,5
a)
0
100
200 km
National Soil Map of Estonia.
Top soil silt content (% soil weight)
Estonian soil map, new extended datset
87,5 - 100,0
75,0 - 87,5
62,5 - 75,0
50,0 - 62,5
37,5 - 50,0
25,0 - 37,5
12,5 - 25,0
0,0 - 12,5
b)
0
100
200 km
National Soil Map of Estonia.
Top soil clay content (% soil weight)
Estonian soil map, new extended datset
87,5 - 100,0
75,0 - 87,5
62,5 - 75,0
50,0 - 62,5
37,5 - 50,0
25,0 - 37,5
12,5 - 25,0
0,0 - 12,5
c)
0
100
200 km
National Soil Map of Estonia.
Top soil rock content (% soil volume)
Estonian soil map, new extended datset
87,5 - 100,0
75,0 - 87,5
62,5 - 75,0
50,0 - 62,5
37,5 - 50,0
25,0 - 37,5
12,5 - 25,0
0,0 - 12,5
d)
0
100
200 km

Figure 5: Physical soil properties: assigned soil texture fractions of a) sand, b) silt, c) clay, and d) coarse fragments in the first soil layer based on texture classes.

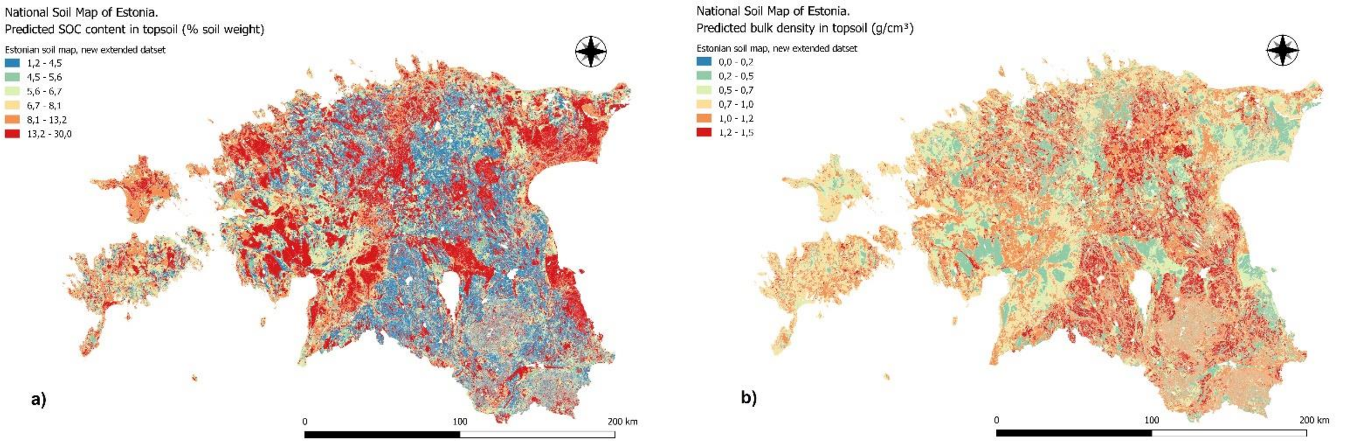

National Soil Map of Estonia.
Predicted SOC content in topsoil (% soil weight)
Estonian soil map, new extended datset
1,2 - 4,5
4,5 - 5,6
5,6 - 6,7
6,7 - 8,1
8,1 - 13,2
13,2 - 30,0
a)
0
100
200 km
National Soil Map of Estonia.
Predicted bulk density in topsoil (g/cm³)
Estonian soil map, new extended datset
0,0 - 0,2
0,2 - 0,5
0,5 - 0,7
0,7 - 1,0
1,0 - 1,2
1,2 - 1,5
b)
0
100
200 km

Figure 6: Extended physical soil parameters: a) predicted soil organic carbon (SOC) and b) bulk density (BD) of the first layer.

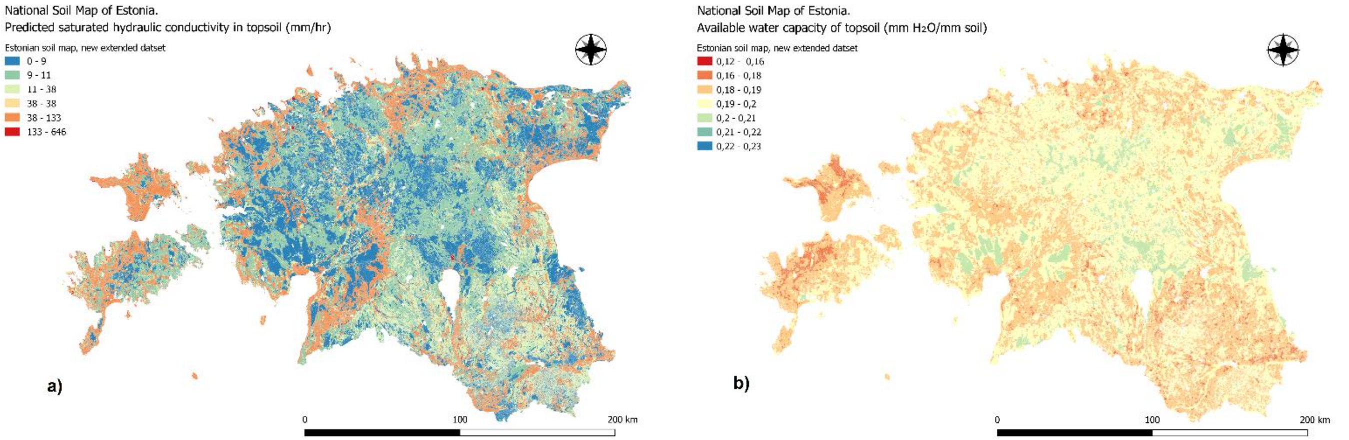
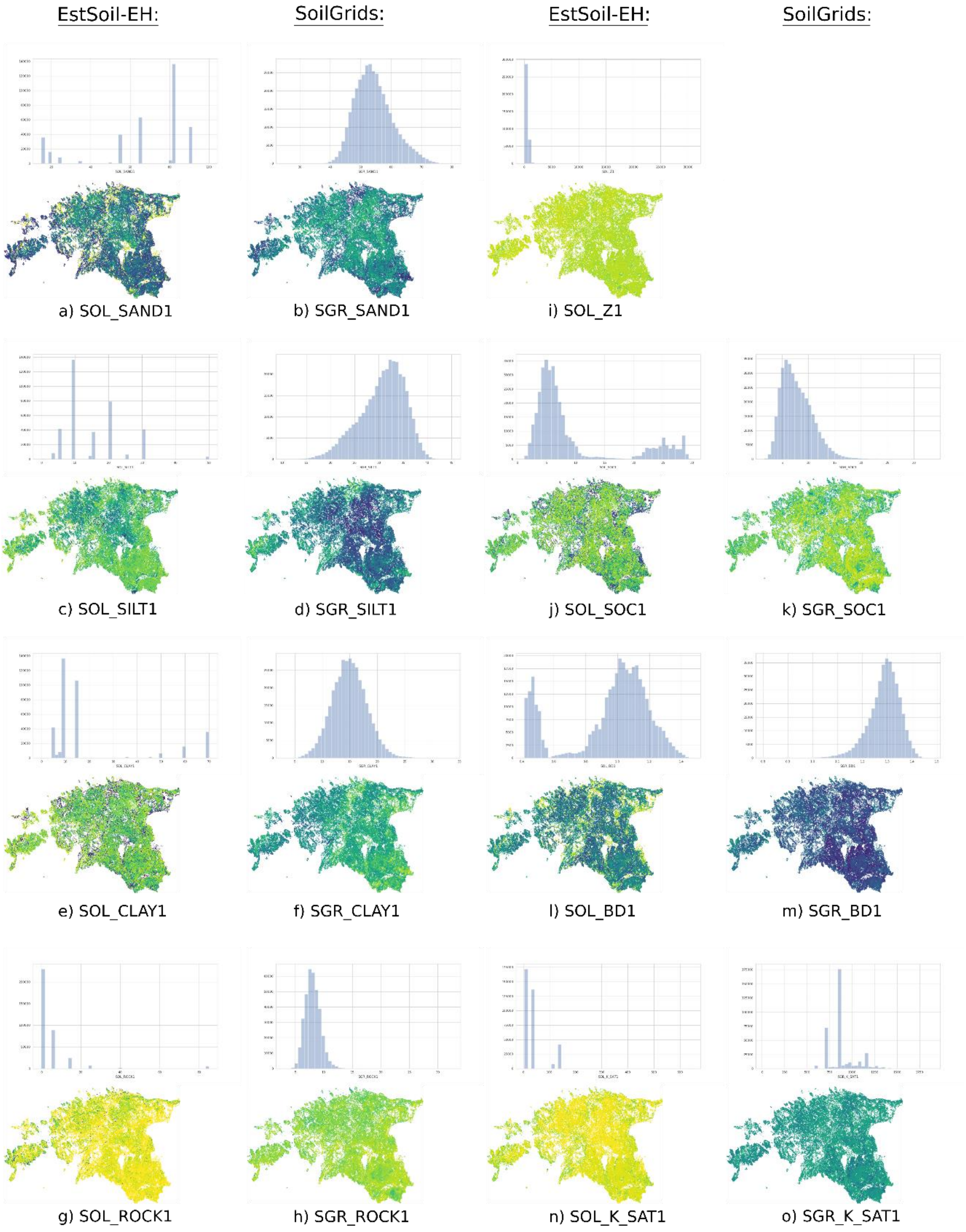

National Soil Map of Estonia.
Predicted saturated hydraulic conductivity in topsoil (mm/hr)
Estonian soil map, new extended datset
0 - 9
9 - 11
11 - 38
38 - 38
38 - 133
133 - 646
a)
0
100
200 km
National Soil Map of Estonia.
Available water capacity of topsoil (mm H₂O/mm soil)
Estonian soil map, new extended datset
0,12 - 0,16
0,16 - 0,18
0,18 - 0,19
0,19 - 0,2
0,2 - 0,21
0,21 - 0,22
0,22 - 0,23
b)
0
100
200 km

Figure 7: Soil hydraulic parameters: a) saturated hydraulic conductivity (K_sat) and b) available water capacity (AWC) in the first layer.



a) SOL_SAND1 b) SGR_SAND1 i) SOL_Z1

c) SOL_SILT1 d) SGR_SILT1 j) SOL_SOC1 k) SGR_SOC1

e) SOL_CLAY1 f) SGR_CLAY1 l) SOL_BD1 m) SGR_BD1

g) SOL_ROCK1 h) SGR_ROCK1 n) SOL_K_SAT1 o) SGR_K_SAT1

Figure 9: Comparative overview plots between EstSoil-EH and SoilGrids-based variables, value frequency and spatial distribution (darker is higher). Full resolution plots (histograms and spatial) for all variables are included in the data supplement.

Table 1: Examples of the final list of standardised soil types and the added English WRB classes, full list as supplemental spreadsheet ("soil_types_legend.csv")

Estonian soil code	Estonian name	Scientific English	WRB_code
Ag	Gleistunud lammimuld	Endogleyic Fluvisols	FL-gln
AG	Lammi-gleimuld	Gleyic Fluvisols	FL-gl
AG1	Lammi-turvastunud muld	Histic Fluvisols	FL-hi
AM'	Väga õhuke lammi-madalsoomuld	Rheic Sapric Histosols (fluvic)	HS-sa.rh-fv
AM"	Õhuke lammi-madalsoomuld	Rheic Sapric Histosols (fluvic)	HS-sa.rh-fv
Dg	Gleistunud deluviaalmuld	Endogleyic Umbrisols (deluvic, novic)	UM-gln-del.nv
E2I	Keskmiselt erodeeritud kahkjas leetunud ja leetunud muld	Dystric Regosols	RG-dy
E2k	Keskmiselt erodeeritud rähkmuld	Epicalcaric Regosols	RG-cap
E2o	Keskmiselt erodeeritud leostunud ja leetjas muld	Eutric Brunic Regosols	RG-br.eu
E3I	Tugevasti erodeeritud kahkjas leetunu ja leetunud muld	Dystric Regosols	RG-dy
E3k	Tugevasti erodeeritud rähkmuld	Epicalcaric Regosols	RG-cap
E3o	Tugevasti erodeeritud leostunud ja leetjas muld	Eutric Brunic Regosols	RG-br.eu

Table 2: Example of the basic rules for deriving numerical values for texture (sand, silt, and clay contents) from the Estonian texture codes and assigned new English and USDA texture classes. These rules were selected by the authors. The full table is provided as a supplemental Excel spreadsheet ("texture_rules_lookup.xlsx")

Estonian texture code	**Estonian name**	**English name**	**USDA texture code**	**Proportion (%) of total weight**		
				Sand	Silt	Clay
l	Liiv	sand	S	90	5	5
l_1	sõre liiv	coarse sand	S	95	5	0
l_2	sidus liiv	fine sand	S	90	3	7
sl	saviliiv	loamy sand	LS	82	9	9
sl_1	saviliiv	loamy sand	LS	82	9	9
ls	liivsavi	loam	L	55	30	15
ls_1	kerge liivsavi	sandy loam	SL	65	20	15
ls_2	keskmine liivsavi	loam	L	55	30	15
s	Savi	clay	C	25	30	45

Table 3: The relationship between the coarse fragments (rock content and shape) indicator from the soil map encoding and the rock content as a % of the total volume. We used the average of each defined range as the singular value required by the SWAT model

	Scale of conversion for rock content					
“Skeleton” indicator number	1	2	3	4	5	6
Inferred rock content (% of volume)	6	15	25	40	60	85

Table 4: Comparison of descriptive statistics (mean, standard deviation, minimum, 25-, 50-, and 75 percentile, and maximum value) for EstSoil-Eh and SoilGrids, based on zonal aggregation of SoilGrids into the EstSoil layer polygons (SOL*: EstSoil-EH, SGR*_ SoilGrids-corresponding variable) for the first out of maximum four layers in EstSoil-EH (full table in data deposit supplement: “estsoil_vs_soilgrids_stats.csv”)

variable	mean	std	min	25 %	50 %	75 %	max
SOL_Z1	787.67	304.97	20.00	500.00	1000.00	1000.00	30750.00
SOL_SAND1	68.70	24.16	0.00	55.00	82.00	90.00	100.00
SGR_SAND1	55.12	6.42	0.00	50.55	54.40	59.00	80.08
SOL_SILT1	14.15	9.83	0.00	5.00	9.00	20.00	50.00
SGR_SILT1	30.12	4.77	0.00	27.17	30.88	33.58	46.17
SOL_CLAY1	17.15	18.99	0.00	5.00	9.00	15.00	70.00
SGR_CLAY1	14.74	3.17	0.00	12.58	14.65	16.75	35.67
SOL_ROCK1	5.46	15.71	0.00	0.00	0.00	6.00	85.00
SGR_ROCK1	8.42	1.60	0.00	7.36	8.28	9.27	38.17
SOL_SOC1	9.37	7.18	1.20	5.05	6.69	9.38	29.98
SGR_SOC1	9.66	4.56	1.18	6.20	8.64	12.16	47.62
SOL_BD1	0.94	0.23	0.00	0.85	0.99	1.10	1.47
SGR_BD1	1.33	0.07	0.80	1.29	1.34	1.38	1.51
SOL_K_SAT1	55.27	66.43	0.00	9.20	37.54	133.21	645.68
SGR_K_SAT1	773.84	193.29	24.38	588.33	701.33	870.83	1884.17

Table 5: Description of variables and parameters available in the EstSoil-EH dataset

name of variable per mapped soil unit	data_type	description
upd_siffer	string	Estonian soil type
WRB_code	string	FAO WRB soil reference group (1st and 2nd level)
wrb_main	string	FAO WRB main soil reference group (1st level)
Loimis1	string	Estonian long texture encoding description
loimis_rec	string	reconstructed error-free interpretation of Estonian texture encoding description

nlayers	number	number of recognized layers/horizons
SOL_ZMX	float64	depth in mm: max depth of the sample analysed soil profile in the mapped soil unit
SOL_Z1-4	float64	depth in mm: the bottom of the layer
EST_TXT1-4	string	Estonian texture class
LXTYPE1-4	string	USDA texture class
EST_CRS1-4	string	Estonian coarse fragment type
SOL_SAND1-4	int64	% mass of Sand in fine earth fraction
SOL_SILT1-4	int64	% mass of Silt in fine earth fraction
SOL_CLAY1-4	int64	% mass of Clay in fine earth fraction
SOL_ROCK1-4	int64	% volumetric in kg soil
SOL_SOC1-4	float64	% soil weight
SOL_BD1-4	float64	g/cm^3
SOL_K1-4	float64	mm/hr
SOL_AWC1-4	float64	mm H_2O/mm soil
slp_mean	float64	mean slope, calculated from DEM
slp_median	float64	median of slope
slp_stdev	float64	standard deviation
twi_mean	float64	mean terrain wetness index, calculated from DEM
twi_median	float64	median of terrain wetness index
twi_stdev	float64	standard deviation
ls_mean	float64	ls-factor, calculated from DEM
ls_median	float64	median ls-factor
ls_stdev	float64	standard deviation
tri_mean	float64	terrain roughness index, calculated from DEM
tri_median	float64	median of terrain roughness index
tri_stdev	float64	standard deviation
area_drain	float64	area per unit under a (e.g. tile-)drainage regimen
drain_pct	float64	percent of the area of the soil unit under drainage
geometry	geometry	EPSG:3301 Estonian National Grid

Table 6: Table 6: Statistical description of SOC prediction error per land form

landform	mean	std	min	25%	50%	75%	95%	max	median	nMAD	kurtosis	skew	RMSE
wetland	1.74	2.73	-5.22	-0.05	1.71	3.51	6.66	8.09	1.71	2.15	-0.1	0.04	3.23
arable	-1.54	1.78	-21.2	-2.12	-1	-0.63	-0.12	6.82	-1	1.12	29.65	-4	2.35
forest	-2.08	4.46	-24.56	-3.07	-1.52	-0.09	3.44	25.65	-1.52	2.79	7.39	-0.44	4.92
grassland	1.06	4.28	-8.47	-1.79	0.52	2.92	10.94	11.78	0.52	3.21	1.16	0.59	4.38

Table 7: Predicted K sat values and reported standard deviation from Rosetta

texture class	sand	silt	clay	k_sat	k_sat_std
GRAVELS	100	0	0	645.68	1.29
S (coarse sand)	95	5	0	362.25	1.19
S (sand)	90	5	5	133.21	1.13
S (fine sand)	90	3	7	113.71	1.17
LS (Estonian 'sl1-3' classes)	82	9	9	37.54	1.15
LS (Estonian 'tsl1' class)	80	14	6	40.2	1.18
SL	65	20	15	11.02	1.18
no_info	60	20	20	7.04	1.22
L (Estonian 'ls2' class)	55	30	15	9.04	1.21
CL	50	15	35	3.67	1.3
L (Estonian 'tls1' class)	40	45	15	8.16	1.37
SiL	35	50	15	8.89	1.35
SiCL	30	40	30	3.97	1.34
PEAT (Estonian 't1' class)	25	25	50	5.09	1.53
HUMUS	25	25	50	5.09	1.53
HC	25	30	45	4.29	1.43
C	25	30	45	4.29	1.43
PEAT (Estonian 't2' class)	20	20	60	7.24	1.81
PEAT (Estonian 't3' class)	15	15	70	9.2	2.45