New pedotransfer approaches to predict soil bulk density using WoSIS soil data and environmental covariates in Mediterranean agro-ecosystems

Calogero Schillaci [1], Alessia Perego [1]*, Elena Valkama [2], Michael Märker [3], Sergio Saia [4], Fabio Veronesi [5], Aldo Lipani [6], Luigi Lombardo [7], Tommaso Tadiello [1], Hannes A. Gamper [8], Luigi Tedone [9], Cami Moss [10], Elena Pareja-Serrano [11], Gabriele Amato [12], Kersten Kühl [13], Claudia Damatirca [14], Alessia Cogato [15], Nada Mzid [16], Rasu Eeswaran [17], Marya Rebelo [18], Giorgio Sperandio [19], Alberto Bosino [3], Margherita Bufalini [20], Tülay Tunçay [21], Jianqi Ding [22], Marco Fiorentini [23], Guadalupe Tiscornia [24], Sarah Conradt [25], Marco Botta [1], Marco Acutis [1]

1 Department of Agricultural and Environmental Science, University of Milan, Via Celoria 2 Milan, Italy
2 Natural Resources Institute Finland (Luke), Bioeconomy and Environment, FI-31600, Tietotie 4, Jokioinen, Finland
3 Department of Earth and Environmental Sciences, University of Pavia, Via Ferrata, 1, 27100 Pavia, Italy
4 Department of Veterinary Sciences, University of Pisa, Via delle Piagge 2, Pisa 56129, Italy
5 Water Research Centre Limited, Frankland Road, Blagrove, Swindon, Wiltshire, SN56 8YF, England
6 Department of Civil, Environmental and Geomatic Engineering, University College London (UCL), Gower Street, London, UK
7 Faculty of Geo-Information Science and Earth Observation (ITC), University of Twente, PO Box 217, Enschede, AE 7500, The Netherlands
8 Faculty of Science and Technology, Free University of Bozen, Piazza Università, 5 39100 Bolzano, Italy
9 Department of Agricultural and Environmental Science, University of Bari Aldo Moro, Via Amendola 165/A-, 70126 Bari, Italy
10 Department of Population Health, London School of Hygiene & Tropical Medicine, London WC1E 7HT, UK
11 Remote Sensing and GIS Group, Instituto de Desarrollo Regional, University of Castilla-La Mancha, 02071 Campus of Albacete, Spain
12 IFAC CNR, Via Madonna del Piano, 10, 50019 Sesto Fiorentino, Florence, Italy
13 Department of Geography Ludwig-Maximilians-University of Munich, Germany
14 Department of Agricultural, Forest and Food Sciences, University of Torino, largo Braccini 2, 10095 Grugliasco, Italy
15 Department of Land, Environmental, Agriculture and Forestry, University of Padova,35020 Legnaro, Italy
16 Department of Sciences and Technology for Agriculture, Forest, Environment and Energy, DAFNE, Università della Tuscia, Viterbo, Italy
17 Department of Plant, Soil and Microbial Sciences, Michigan State University, East Lansing, MI48824, USA
18 Department of Agriculture, Food and Environment, University of Pisa, via del Borghetto 80, 56124 Pisa, Italy
19 Dipartimento di Medicina Molecolare e Traslazionale, Università di Brescia, Viale Europa, 11, 25123 Brescia, Italy
20 University of Camerino, School of Science and Technology-Geology Division, Via Gentile III da Varano, Camerino 62032, Italy
21 Soil Fertilizer and Water Resources Central Research Institute, Ankara-Turkey
22 Department of Biological and Ecological Sciences DEB, Università della Tuscia, Viterbo, Italy
23 Department of Agricultural, Food and Environmental Sciences (D3A), Marche Polytechnic University, Ancona, Italy
24 Instituto Nacional de Investigación Agropecuaria (INIA). Unidad Agroclima y Sistemas de Información (GRAS). Ruta 48 KM10, Canelones 90200, Uruguay.
25 SCOR SE, Zurich Branch, Switzerland

***corresponding author**

Keywords: Agriculture, bulk density, pedotransfer functions, PTFs, soil carbon, soil texture

Highlights

- Three PTFs were developed to calculate bulk density of arable top- and subsoil
- WoSIS, WorldClim, and topographic data of the Mediterranean Basin were used
- Model transferability of the three new PTFs was validated with external dataset
- Topsoil ANN-PTF had R^2 of 0.89 in training and 0.45 in model transferability
- ANN-PTF outperformed the commonly employed PTF by Manrique and Jones

Abstract

For the estimation of the soil organic carbon stocks, bulk density (BD) is a fundamental parameter but measured data are usually not available especially when dealing with legacy soil data. It is possible to estimate BD by applying pedotransfer function (PTF). We applied different estimation methods with the aim to define a suitable PTF for BD of arable land for the Mediterranean Basin, which has peculiar climate features that may influence the soil carbon sequestration. To improve the existing BD estimation methods, we used a set of public climatic and topographic data along with the soil texture and organic carbon data. The present work consisted of the following steps: i) development of three PTFs models separately for top (0-0.4 m) and subsoil (0.4-1.2 m), ii) a 10-fold cross-validation, iii) model transferability using an external dataset derived from published data.

The development of the new PTFs was based on the training dataset consisting of World Soil Information Service (WoSIS) soil profile data, climatic data from WorldClim at 1 km spatial resolution and Shuttle Radar Topography Mission (SRTM) digital elevation model at 30 m spatial resolution.

The three PTFs models were developed using: Multiple Linear Regression stepwise (MLR-S), Multiple Linear Regression backward stepwise (MLR-BS), and Artificial Neural Network (ANN).

The predictions of the newly developed PTFs were compared with the BD calculated using the PTF proposed by Manrique and Jones (MJ) and the modelled BD derived from the global SoilGrids dataset.

For the topsoil training dataset (N=129), MLR-S, MLR-BS and ANN had a R^2 0.35, 0.58 and 0.86, respectively. For the model transferability, the three PTFs applied to the external topsoil dataset (N=59), achieved R^2 values of 0.06, 0.03 and 0.41. For the subsoil training dataset (N=180), MLR-S, MLR-BS and ANN the R^2 values were 0.36, 0.46 and 0.83, respectively. When applied to the external subsoil dataset (N=29), the R^2 values were 0.05, 0.06 and 0.41. The cross-validation for both top and subsoil dataset, resulted in an intermediate performance compared to calibration and validation with the external dataset. The new ANN PTF outperformed MLR-S, MLR-BS, MJ and SoilGrids approaches for estimating BD. Further improvements may be achieved by additionally considering the time of sampling, agricultural soil management and cultivation practices in predictive models.

1. Introduction

Soil bulk density (BD) is directly linked to soil functionality including mechanical support of crop plants, circulation of soil solution, and soil aeration (Håkansson and Lipiec, 2000). Relatively high values of BD indicate soil compaction which may lead to reduced water infiltration especially in agricultural land, where it can hamper the growth of crop root systems (Colombi et al., 2018). Soil BD is calculated as the dry weight of soil divided by its volume. Volumes include soil particle volume and pore space between soil particles. Soil BD is typically expressed in g cm^{-3} or Mg m^{-3} (SI). Along with soil organic carbon (SOC) concentrations, soil BD is necessary to calculate SOC stocks (Minasny et al., 2013) and to assess carbon sequestration (Tao et al., 2019). Many soil physical and chemical properties are expressed on a volumetric basis and in particular the estimation of soil biological properties depend on BD estimates (Tejada et al., 2009). In arable lands, tillage and other management practices cause high variation of BD during the year. Scientists have tried to infer BD from soil properties that are routinely measured such as textural information and organic carbon content (Acutis and Donatelli, 2003; Alvarez-Acosta et al., 2012; Pachepsky et al., 1996; Van Looy et al., 2017). The functions enabling the estimation of a given soil property (e.g. BD) from other variables, routinely obtained through laboratory measurement, are called pedotransfer functions (PTF) (Bouma, 1989; Patil and Singh, 2016). PTFs have been used at global scale to estimate the soil water retention, soil particle size, soil BD and SOC stock (Batjes and Dijkshoorn, 1999; Rawls, 1983; Rawls and Pachepsky, 2002; Reynolds et al., 2000; Saxton et al., 1986). At this scale, soil BD models had limited predictive ability (Rawls, 1983; Tietje and Tapkenhinrichs, 1993). Unfortunately, PTFs are not able to fully replace direct measurements, as highlighted in a recent publication which compared >50 PTFs using high resolution geodata in at district scale (Nasta et al., 2020; Xiangsheng et al., 2016). PTF are also frequently chosen at district scales after a sensitivity analysis (Basile et al., 2019).

Accurate models are of high interest for land management and policy-making especially where sparse data are available.

Today, BD estimates are used to quantify and model the SOC stocks in top- and subsoil at regional and global scales (Valkama et al., 2020). For example, Sun et al. (2020) recently used PTF in a meta-analysis to assess the effect of conservation agriculture on carbon stocks but did not provide an assessment of the PTF function performance.

One of the first attempts to estimate BD was made by Manrique and Jones (1991) who proposed a PTF based on SOC alone ($BD=1.660-0.318 \cdot SOC^{0.5}$) for all soil types. Since then, other PTFs for BD estimation have been developed based on the fine earth fractions and SOC, which is important to BD due to its effect on the ratio between soil macro- and micropores (Martín et al., 2017; Throop et al., 2012). Furthermore, many other functions have been proposed to describe regional (Akpa et al. 2016; Chagas et al., 2016; Chen et al., 2018; Makovníková et al., 2017; Montzka et al., 2017; Ramcharan et al., 2017; Román Dobarco et al., 2019; Wösten et al., 2013, 1999) and local conditions (Benites et al., 2007; De Vos et al., 2005; Picciafuoco et al., 2019; Sevastas et al., 2018) and to predict BD in different soil horizons (Hollis et al., 2012; Reidy et al., 2016; Sequeira et al., 2014).

In the absence of measured soil data, the availability of new topographic data such as digital elevation models and morphometric indices has also improved soil assessment (Lombardo et al. 2018, Schillaci et al., 2017a, b, 2019; Veronesi and Schillaci, 2019) and in particular to develop PTF (Leij et al., 2004; Romano and Chirico, 2004). Other geodata (e.g., climate, satellite-derived, land cover) correlated with BD have also been used to improve estimates (Aitkenhead and Coull, 2020). Various researchers have recently developed new methods to estimate BD.

Bondi et al., (2018) estimated BD for peat soils using soil visual assessment, and decision trees achieving similar performances, with around 0.6 explained variance. Premrov et al., (2018) achieved similar performances (R^2 from 0.4 to 0.6) using optimal power-transformation of measured physical and chemical soil parameters.

Chen et al. (2018) formalized an analytical protocol to test the PTF prediction at regional scales in France by building a Boosted Regression Tree (BRT) model to obtain reliable predictions (R^2 0.7) , and also applied the advanced deep learning modelling framework for the evaluation of in situ spectral measurement of SOC with in situ vis-NIR spectroscopy in southeastern Tibet (Chen et al., 2020) achieving ($R^2 = 0.92$). Rodríguez-Lado et al. (2015) used a dataset consisting of 115 topsoil observations in a catchment of approximately 100 km^2 to map soil BD and compared three methods: Stepwise Multiple Linear Regression (MLR-S), Random Forest (RF) and Artificial Neural Networks (ANN). In this procedure, RF and ANN appeared the most suitable approaches to predict the measured data, producing R^2 of 0.90 and 0.86, respectively. These results suggest that soil samples remain essential to obtain good estimates, and that PTFs derived from data collected in given locations can fail to give accurate estimates when applied elsewhere (Akpa et al., 2016). PTFs modelling is a relatively new subject and many important steps have been carried out recently (Chen et al., 2018; Sevastas et al., 2018). To extract all the contributions on soil BD, simple query can be used to gather publications from SCOPUS and Web of Knowledge (Schillaci et al., 2018). Out of this search the most used approach for the BD estimation with PTFs is multiple linear regression (60%) followed by ANN (20 %), therefore these two approaches are investigated here.

At present the available PTF models offer wide predictive ranges and none are specifically developed for the Mediterranean area. The aim of this study was to develop new regionally-specific BD prediction models using data gathered from the literature on soil texture, SOC, topography and climate in Mediterranean agro-ecosystems. As well as providing a modelling framework that can be applied in each environmental setting. In the Mediterranean basin area, soil organic matter mineralization is boosted by high-temperature conditions (Álvaro-Fuentes and Paustian, 2011), in which rainfall has a peculiar pattern (availability during a short season vs long dry period). Moreover, the agricultural systems are conventionally plough-based (Mazzoncini et al., 2011) causing soil compaction and reduced SOC stocks.

2. Material and Methods

The study was conceptualized during the first annual summer school module "Statistical Analysis of Spatial Data in Agro-Environmental Research", organized in cooperation with Lake Como Advanced School (https://sdae.lakecomoschool.org/), and held from August 26-30, 2019. As a practical teaching activity, soil legacy data and topographic datasets were compiled to develop a PTF. The school participants were mainly PhD students and early career researchers. The present work was carried out after the school as a collaboration between students and teachers.

2.1 Operational procedures

Study work streams included PTF development using training datasets from public databases, and PTF validation using an independent validation dataset compiled from systematic review of the literature (Table 1 and Fig. 1). In the training step, we defined three PTFs – two based on statistical approaches and one based on ANN. In the validation step, we applied the three newly-defined PTFs to an external dataset. We then compared the performances of the three PTFs and benchmarked them against those of the MJ PTF and SoilGrids estimates (see below). The training and validation datasets were each split into topsoil and subsoil to infer separate PTFs.

Table 1. Study overview and workflow to develop pedotransfer functions (PTF) to infer soil bulk density (BD) in top- (0-0.4m) and subsoil (0.4-1.2) of arable fields in the Mediterranean

	New PTFs to estimate BD			**Reference PTF and BD data**	
Study stage	**MLR-S**: stepwise regression	**MLR-BS**: backward + stepwise regression	**ANN**: artificial neural network	**Manrique-Jones (1991):** PTF function for estimating BD	**SoilGrids**: estimated BD values derived from WoSIS data
Training	Developed using WoSIS database* + topographic + climatic data				
Validation and Benchmarking	Applied on external database** + topographic + climatic data			Applied on external database**	Available at 250 m grid

* WoSIS database: measured data of bulk density (BD, Mg m^{-3}), soil organic carbon (%), sand

** newly compiled database of soil bulk density, organic carbon, sand, silt and clay measurements of studies from the Mediterranean

Training dataset

Source
WoSIS open database of geo-referenced soil profiles sampling with soil properties:
Bulk density, Texture, SOC, Rock fragment content, sample depth

Extraction
WoSIS shapefile with geo-referenced study locations with soil physical and chemical soil data

Validation dataset

Source
Soil properties of studies carried out in the Mediterranean:
Bulk density, Texture, SOC, Rock fragment content, sampling depth

Extraction
Manual compilation in Excel datasheet from original publications

Criteria for inclusion in the newly compiled reference dataset:
-Köppen classification with a buffer of 250 km in the Mediterranean basin
-Availability of:

- Bulk density
- Soil organic carbon % (SOC)
- Texture (Clay, Silt, Sand)

-Land use: agricultural soil (CORINE database*)
-Rock fragments < 5%
-Depth < 1.2 meters

Quality Checking
0.9 < Bulk Density <2

Predictors based on soil properties
Mean depth, texture, SOC. Computation of power terms and interaction between variables: $Clay^2$, $Sand^2$, $SOC^{0.5}$, SOC^2, Clay x SOC, Clay x Sand, $Clay^2$x SOC^2

Additional data, compiled from WorldClim Bioclimatic data (Annual Mean T°C (BIO1), Mean Diurnal T°C Range (BIO2), Isothermality (BIO3), Temperature Seasonality (BIO4), Max T°C of the Warmest Month (BIO5), Min T°C of the Coldest Month (BIO6), Annual T°C Range (BIO7), Mean T°C of the Wettest Quarter (BIO8), Mean T°C of the Driest Quarter (BIO9), Mean T°C of the Warmest Quarter (BIO10), Mean T°C of the Coldest Quarter (BIO11), Annual Precipitation (BIO12), Precipitation of the Wettest Month (BIO13), Precipitation of the Driest Month (BIO14), Precipitation Seasonality (BIO15), Precipitation of the Wettest Quarter (BIO16), Precipitation of the Driest Quarter (BIO17), Precipitation of the Warmest Quarter (BIO18), Precipitation of the Coldest Quarter (BIO19))

Topographic data (Elevation, Slope, Northness, Profile curvature, Plan curvature**)**

Extraction per location using GIS

Figure 1. Features of the datasets used to train and validate (training and model transferability) three new pedotransfer functions (PTF) for arable soils in the Mediterranean. For a description of the WorldClim Bioclimatic data, please see (Fick and Hijmans, 2017).

2.2 Data used in the training and validation stages

2.2.1. Soil datasets

Training dataset used for PTFs model development: The World Soil Information Service WoSIS (https://www.isric.org/explore/wosis) was used to retrieve soil textural values, SOC content and bulk density. WoSIS is a world scale database containing 196,000 geo-referenced, standardized soil profile entries for soil data from multiple origins. Approximately 40 different organizations around the world provide free access to the data via WoSIS and the Soil Profile (https://www.isric.org/explore/wosis/wosis-contributing-institutions-and-experts). More information on WoSIS inclusion criteria, quality assurance, and standardization procedures are available in Batjes et al. (2017). We note that for Europe, one of the main providers of WoSIS data is the Joint Research Center of the European Community, which has made available the entire collection of soil profiles included within the Soil Profile Analytical Database (SPADE-2) (de Souza et al., 2016; Hiederer et al., 2006; Panagos et al., 2013). Using ARCGIS, we selected all the profiles of the WoSIS database belonging to the Mediterranean basin (and defined surrounding areas) with geographic coordinates in metric resolution as well as attributes including sand, silt, clay, organic carbon and bulk density data in at least one soil horizon.

External dataset used to test model transferability: To assess the model transferability, validation of the three developed PTFs was required. Accordingly, we conducted a systematic literature analysis to collate information on soil textures, SOC, and BD from studies of field crops cultivated on mineral soils in Mediterranean basin and close surrounding areas. The search was carried out in SCOPUS and Web of Science (WoS). The selection criterion was the same as that applied during the extraction of the WoSIS data: required data were BD, SOC, texture and geo-localization. It is needed to remark that systematic queries did not result in a adequate number of suitable articles, so that we used different approaches such as searching for soil dataset within agronomic journals. To compare the performances of the PTF models developed in this study with well-known approaches, in the validation phase we applied the MJ PTF (1991; $BD = 1.660 - 0.318 \cdot SOC^{0.5}$) and we fitted the available SoilGrids BD values (Hengl et al., 2017) with the data of the external validation database constructed as above. SoilGrids is a system for digital soil mapping that uses machine learning methods to map the spatial distribution of soil properties across the globe using WoSIS data and environmental predictors.

For both training and validation datasets, the analysis focused on samples that alternatively fall within the 0-0.4 m layer (i.e. topsoil) or the 0.4-1.2 m layer (i.e. subsoil). Due to the presence of multiple horizons inside the topsoil and subsoil, single observations which are part of the training dataset were not averaged. The soil sampling depth were considered as predictor. Furthermore, the inclusion of predictors such as soil properties (soil particle size fractions and SOC stock) allows to describe the soil sample at the given profile depth (e.g., SOC and clay content tend to decrease along the soil profile).

The data points used in the training phase which were derived from WoSIS were 129 and 180 for topsoil and subsoil, respectively.

As SoilGrids data are provided for six soil layers at the fixed depths (0-5, 5-15, 15-30, 30-60, 60-100, 100-200 cm), we computed a weighted average of SoilGrids BD for the comparison with the BD from the external dataset. For example, if BD is measured for the 10-25 layer then 5 cm belongs to the 5-15 cm SoilGrids layer and 10 cm to the 15-30 cm layer. Consequently, to obtain the sample value, we computed a weighted mean between the SoilGrids BD values given for the 5-15 and the 15-30 layers, using a weighting factor of 5 and 10 for the two layers, respectively. We excluded the BD values lower than 0.9 Mg m^{-3} because they were not representative of mineral soils in semiarid regions and, when present, they were likely due to tillage operations occurred close to the sampling moment. We also excluded BD values greater than 2 Mg m^{-3} because they are not representative for agricultural land. Textural plots were prepared using the ggtern software (Hamilton and Ferry, 2018).

2.2.2. Geodata

For the terrain analysis, the Shuttle radar topography mission SRTM 30 m DEM (Farr et al., 2007) was used to obtain topographic data with a resampling at 90 m. The digital elevation model was downloaded in ten tiles from the open topography website (https://opentopography.org/).The topographic indices were obtained for the whole study area using the geo-processing terrain analysis tool in SAGA (Conrad et al., 2015). Data pre-processing and maps were prepared using ArcGIS. The WorldClim climatic data (Fick and Hijmans, 2017) was used to obtain climatic data (e.g., mean annual rainfall, average annual temperature). For EU countries, CORINE land cover (Bossard et al., 2000) was used to select agricultural land use. To assign the target land cover (Agriculture) CLC was check for all the available periods, 2000, 2006, 2012, 2018. For non-EU countries – except for Turkey, which was included in CORINE land cover data – we selected soil profiles belonging to agricultural areas by observing satellite and aerial imagery available in ArcMap and Google Earth-Pro.

2.3. Study area

The study focused on the Mediterranean Basin, which covers the territory between 30° and 45° latitudes and, according to the Köppen climate classification system, belongs to the three main climate groups: *B* (dry), *C* (temperate), and *D* (continental) (Francaviglia et al., 2020) (see Fig. 2). The influence of the sea plays a key role in shaping the environment including relief characteristics, which determine the characteristic Mediterranean climate at basin scale (Lionello et al., 2006). Mediterranean soils are the result of a complex genesis (Lagacherie et al., 2018). Carbonatic and limestone parent materials are the most widespread minerals in the Mediterranean (Verheye and De La Rosa, 2005; Zdruli et al., 2011). Long-term agricultural use has altered soil structure and degraded carbon content. Soil characteristics indicate different ages of soil development and depths and there is evidence of clay particle translocation within the soil profile (Zdruli et al., 2011).

According to the World Reference Base for Soil Resources (WRB 2006), approximately a dozen soil orders can be found in the Mediterranenan basin: histosols, anthrosols, leptosols, vertisols, fluvisols, gleysols, andosols, kastanozems and phaeozems, umbrisols, gypsisols, durisols, calcisols, luvisols, arenosols, cambisols, and regosols. A brief description of these soil orders can be found in Zdruli et al., (2011). Figure 2 shows the locations of the sites included in the training (WoSIS data) and validation (extracted from the literature) datasets.

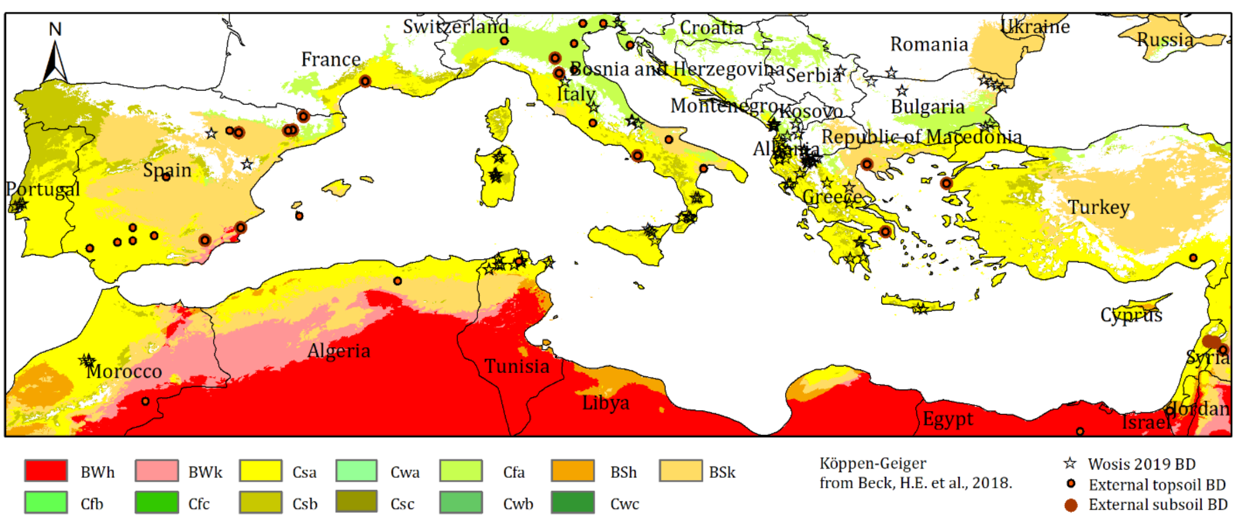

Figure 2. Study area. The location of the sampling sites (WoSIS data for training and external data for model transferability).

2.4 Development of PTFs to estimate BD

In this study, we evaluated three methods to estimate BD, namely Multiple Linear Regression (MLR) models each using two-variable selection criteria, and Artificial Neural Networks (ANN). These methods were chosen in our analyses because they are suitable when data are sparse and no spatial structure can be defined. A 10-fold cross-validation frame was used to assess the prediction accuracy (Veronesi and Schillaci, 2019). The three models were defined using a wide set of predictors, i.e. independent variables (soil properties, bioclimatic and topographic indicators). These predictors were derived from the soil and additional database (Fig. 1).

2.4.1 Multiple linear regression (MLR)

The first method (MLR-S) used was a stepwise multiple linear regression starting from no dependent variables (a constant-only model); the first dependent variable that will be included in the model is the variable that produces the maximum increase in R^2; if the increase in the explicated variance is significant (partial F test) at a given P(F), called inclusion threshold, the variable is retained in the model (forward step). The same procedure is done to evaluate the possibility to include a second independent variable and so on. At each inclusion step, there is an exclusion step too, where, among the variables included in the model, the variable that is excluded causes the lower reduction in explicated variance. If the decrease of explained variability is not significant at a given P(F), called exclusion threshold and higher than the inclusion threshold, the variable is excluded from the model. The process stops when no more dependent variables are included or excluded (Noryani et al., 2019).. In the MLR-S, a predictor is included in the model if its regression coefficient is significant at $P \leq 0.05$ and excluded if the partial F test has a $P > 0.1$ (Draper and Smith, 1998).

The second approach was a stepwise variable selection, which started by including all independent variables, then excluded non-significant variables one by one using a backward stepwise approach (MLR-BS). Variables were excluded when their contribution did not affect the model explication capability (i.e., when the partial F test have a $P>0.1$) (Ghani and Ahmad, 2010).

For both methods (MLR-S and MLR-BS), the normality test of Kolmogorov-Smirnov and the Breush-Pagan test for the homogeneity of variances (Breusch et al., 1979) were applied to the residuals of the regression models.

2.4.2 Artificial Neural Network

An ANN is part of a computing system, which is developed to mimic the way the human brain processes information. ANN allows finding non-linear behavior of the system that cannot be discovered with traditional regression-based methods. To develop a PTF, the ANN is generally made by three layers of neurons, i.e. an input layer, a hidden layer and an output layer (Ebrahimi et al., 2019; Minasny and McBratney, 2002; Schaap et al., 1998). This kind of ANN architecture is known as Multi-Layer Perceptron (MLP). ANN imposes minimal requirements for model structure or assumptions because the shape of the relationship is determined during the learning process (Haykin, 2008). We used an MLP implementation in the IBM-SPSS 26.0.0.1. One hidden layer was used with three neurons according to the default settings, using the hyperbolic tangent activation function and the identity function for the output layer. This is an identity function because this task is a regression problem. The weighted connections feed forward from the input layer to the output layer. The training algorithm works by back-propagating the prediction error, through the parameters of the neural network. In this study, the MLP had 18 input predictors and one output variable, i.e. BD. The independent variables used as predictors in the three statistical models for the BD estimation are shown in Fig. 1. The optimal fit was reached in cross-validation by using 1 hidden layer, combined with three neurons. The Hyperparameters tuning was iteratively tested by applying an ANN with one hidden layer with 2 to 10 neurons and, alternatively, an ANN with two hidden layers with 2 to 5 neurons in the first hidden layer combined with 2 or 3 neurons in the second hidden layer. The use of one single hidden layer resulted to be more effective. This result agreed with the automatic parameterization proposed by the software: (ftp://public.dhe.ibm.com/software/analytics/spss/documentation/statistics/27.0/en/client/Manuals/IBM_SPSS_Statistics_Algorithms.pdf). Regarding computation time, the model training phase takes few second.

2.5. Analysis of models' performance

The following evaluation indices were calculated to test the model performance in estimating BD:

i) R^2 coefficient of determination of the scatter plot of the predicted against the observed values;

ii) Bias and %Bias (Addiscott and Whitmore, 1987), optimal value is 0, range is from $+\infty$ to $-\infty$; when the Bias% is < 10% it may be considered very favorable (Moriasi et al., 2007);

iii) Root Mean Square Error (RMSE) and %RMSE (RMSE/(Observed Mean) *100) (Fox, 1981), optimal value is 0, range is from 0 to $+\infty$; %RMSE value lower than 10% is considered to be favorable (Bellocchi et al., 2002);

iv) The Pearson correlation coefficient, optimal value is 1, range is from +1 to –1;

v)The slope of the regression of observed data to the estimated ones, optimal value is 1, range is from $+\infty$ to $-\infty$ (Piñeiro et al., 2008).

Note that Bias is always equal to 0 when the ordinary least square (OLS) method is applied, which was the case in the two regression training sets. Moreover, in OLS analysis the slope of observed values against the estimated values is equal to 1. All indices were computed using Irene-DLL (Fila et al., 2003).

3. Results

3.1 Descriptive statistics

3.1.1 Soil properties

The highest average BD value was observed in the subsoil training dataset (1.51±0.17 Mg m^{-3}). The lowest average BD value was observed in the topsoil validation dataset (1.38±0.12 Mg m^{-3}). The SOC was higher in the topsoil testing (1.28±1 %), and lower in the subsoil testing dataset (0.61±0.35 %) (Table 3). The most variable soil property was the sand content with a coefficient of variation ranging from 48 to 85%, while BD was less variable with a coefficient of variation ranging from 9 to 14%. The references of the independent dataset for validation are listed in Table 2. The independent external dataset for validation comprised 59 observations for the topsoil and 29 for the subsoil, Table 3. Textural plots of the training and validation datasets are shown in Table 3.

Table 2. Independent dataset for validation with country, climate and reference.

Country	Köppen climate classification	Author
Algeria	BSk	Chennafi et al., 2006
Croatia	Cfa	Bogunovic et al., 2018
Egypt	BWh	Mahmoud et al., 2019
	BWh	Salem et al., 2015
	BWh	Zohry et al., 2017
France	Csa	Cardinael et al., 2017
Greece	CSa	Antonopoulos et al., 2013
Israel	BSh	Stavi et al., 2008
Italy	Cfa	Pezzuolo et al., 2017
	Cfa	Carozzi et al., 2013
	Csa	Francaviglia et al., 2015
	Cfa	Valboa et al., 2015
	Cfa	Perego et al., 2019
	Cfa	Ceotto et al., 2018
	Cfa	Diacono et al., 2018
	Csa	Vitale et al., 2017
Lebanon	Csa	Karam et al., 2007
Morocco	Csa	Ichir et al., 2003
Spain	BSk	Pareja-Sánchez et al., 2017
	Cfa	Bescansa et al., 2006
	BSk	Pardo et al., 2020
	BSk	Tolon-Becerra et al., 2011
	Bsk-Cfa	Álvaro-Fuentes et al., 2008
	BWh	Visconti et al., 2019
	BSk	Recio et al., 2018
	Csa	Marquez-Garcia et al., 2013
Syria	Bsk	Abou Zakhem et al., 2019
Tunisia	Csa	Jemai et al., 2013
Turkey	Csb	Çelik et al., 2019

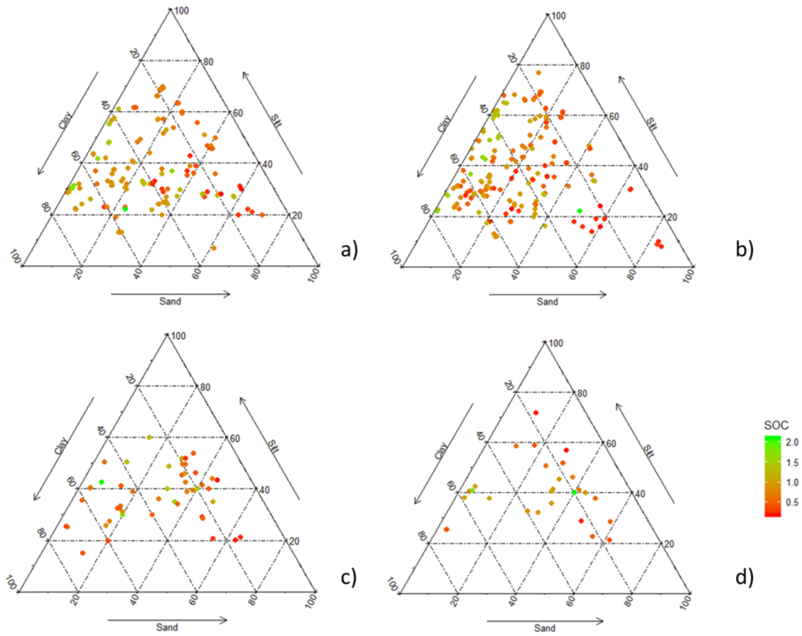

Figure 3. Textural plots, a) topsoil validation dataset, b) subsoil validation dataset, c) topsoil test dataset d) subsoil test dataset.

Table 3. Soil properties of the training and testing data for topsoil (0-0.4m) and subsoil (0.4-1.2 m): Bulk Density (BD), Soil Organic Carbon (SOC), Fine earth fractions,

		BD (Mg m^{-3})	**SOC (%)**	**Sand (%)**	**Silt (%)**	**Clay (%)**
Topsoil Training	mean	1.44	1.26	24.4	36.5	39.1
(N=129)	Stdv	0.20	0.64	17.1	14.1	18.1
Topsoil Testing	mean	1.41	1.28	31.39	40.68	28.21
(N=59)	Stdv	0.11	1.0	13.82	9.09	14.25
Subsoil Training	mean	1.51	1.15	20.1	38.2	41.7
(N=180)	Stdv	0.17	0.67	17.0	15.7	17.4
Subsoil Testing	mean	1.48	0.61	29.04	39.49	31.56
(N=29)	Stdv	0.16	0.35	19.38	12.58	18.4

3.1.2 Environmental variables

Average precipitation reported in the training dataset was highly variable in the study area with a minimum value of 426 and a maximum of 1693 mm yr^{-1}. The validation dataset showed a minimum annual rainfall of 189 and a maximum of 1155 mm yr^{-1}. Mean annual temperature, Elevation (m), Slope (%) are reported in Table 4.

Table 4. Descriptive statistics of the selected environmental variables

		Annual Average Precipitation (mm yr^{-1})	Mean annual temperature (° C)	Elevation (m)	Slope (%)
Training (N=77 sites)	mean	774.4	10.5	321	4.3
	stdv	294.4	1.5	332	5.4
Testing (N=36 sites)	mean	495.7	16	318	4
	stdv	300	3.1	379	4.9

3.2 Model performance and transferability

Homogeneity of variance and normality tests for the MLR models were conducted using the Breush-Pagan test and Kolmogorov-Smirnov test (Table 5).

Table. 5 Homogeneity of variance and normality tests for Multiple Linear Regression (MLR) models.

	MLR-S		**MLR-BS**	
	Topsoil	**Subsoil**	**Topsoil**	**Subsoil**
Homogenity of variance of residuals*	0.056	0.051	0.065	0.051
Normality of residuals**	>0.2	>0.2	>0.2	>0.2

*Breush-Pagan test; ** Kolmogorow-Smirnov test

Topsoil model metrics are shown in Table 6. The RMSE of the topsoil training dataset (Table 6a) ranged from 0.07 (ANN) to 0.17 (MLR-S), and similar performances were obtained with the MLR-BS models. The Bias of the ANN was close to zero. The ANN model showed the highest R^2 (0.89), whereas the MLR-S model showed the lowest R^2 (0.24).

Table 6. Performance of the newly developed pedotransfer function (PTF) as developed with the topsoil training and cross validation (a) and tested with the independent external datasets for model transferability (b). Indices values reported in brackets refer to the cross-validation results.

a)	MLR-S - MLR-S CV	MLR-BS - MLR-BS CV	ANN - ANN CV
RMSE	0.17 (0.16)	0.14 (0.15)	0.07 (0.16)
rRMSE %	11.91 (11.53)	9.68 (10.81)	4.56 (11.4)
Bias	(0.0007)	(0.0047)	0.00 (0.01)
Bias %	(0.144)	(0.37)	0.10 (1.11)
r	0.51 (0.49)	0.72 (0.57)	0.94 (0.67)
R^2	0.26 (0.33)	0.51 (0.37)	0.89 (0.48)
Slope b	(0.84)	(0.71)	1.00 (0.78)
Estimated Max	1.61 (1.6)	1.80 (1.67)	1.90 7 (1.75)
Estimated Min	1.16 (1.22)	1.02 (1.14)	0.88 (1.13)
N	129		

b)	MLR-S	MLR-BS	ANN	MJ	SoilGrids
RMSE	0.14	0.32	0.16	0.17	0.13
rRMSE %	9.28	22.26	11.53	11.93	9.12
Bias	0.06	0.13	0.07	-0.11	0.008
Bias %	1.1	11.2	1.4	-6.657	0.04
r	0.34	0.05	0.64	0.24	0.09
R^2	0.12	0.00	0.41	0.06	0.01
Slope b	0.29	-0.12	1.11	0.23	0.05
Estimated Max	1.7	2.75	1.94	1.44	1.53
Estimated Min	1.24	1.03	0.92	0.76	1.28
N	59				

The RMSE in the topsoil validation dataset (Table 6b) range from 0.13 (SoilGrids) to 0.32 (MLR-BS). All the Bias values were ≤0.5. The R^2 ranged between 0.09 to 0.41, in SoilGrids and ANN, respectively.

Table 7. Performance of the newly developed pedotransfer function (PTF) as developed with the subsoil training and cross validation (a) and tested with the independent external datasets for model transferability (b). Indices values reported in brackets refer to the cross-validation results.

a)	**MLR-S**	**MLR-BS**	**ANN**
RMSE	0.14 (0.13)	0.12 (0.13)	0.07 (0.11)
rRMSE %	9.04 (9.24)	8.04 (8.67)	4.53 (7.79)
Bias	(-0.0003)	(-0.0004)	0.00 (0.003)
Bias %	(0.0056)	(0.0008)	-0.16 (0.21)
r	0.49 (0.47)	0.70 (0.58)	0.92 (0.67)
R^2	0.24 (0.21)	0.48 (0.38)	0.84 (0.48)
Slope b	(0.90)	(0.90)	0.98 (0.84)
Estimated Max	1.77 (1.68)	1.79 (1.71)	1.93 (1.76)
Estimated Min	1.12 (1.32)	1.35 (1.28)	1.10 (1.26)
N	180	180	180

b)	**MLR-S**	**MLR-BS**	**ANN**	**MJ**	**SoilGrids**
RMSE	0.17	0.39	0.21	0.14	0.17
rRMSE %	11.78	26.26	13.93	9.54	11.39
Bias	0.09	-0.04	0.15	-0.07	0.07
Bias %	2.8	-2.347	1.7	-4.66	0.708
r	0.38	0.35	0.67	0.26	-0.42
R^2	0.15	0.13	0.45	0.07	0.18
Slope b	0.37	1.06	0.94	0.001	-0.11
Estimated Max	1.99	1.95	1.88	1.53	1.64
Estimated Min	1.38	1.01	1.29	1.20	1.47
N	30				

Subsoil model metrics are shown in Table 7. The RMSE in the subsoil training dataset (Table 7a) ranged from 0.07 (ANN) to 0.14 (MLR-S). The Bias of the ANN was close to zero. The ANN model showed the highest R^2 of 0.84, whereas the MLR-S model showed the lowest R^2 of 0.24. The RMSE in the subsoil external dataset (Table 7b) are very similar and ranged from 0.14 to 0.39. The Bias % values ranging from -4.6 (MJ) to 2.8% (MLR-S). The R^2 ranged between 0.07 (MLR-S) to 0.45 (ANN), respectively.

Since the best performance was achieved with the ANN, we provide a .xlm spreadsheet file that can be used to execute the PTF developed with the ANN using the soil data, topography and

WorldClim. Furthermore, to allow users to apply the PTF based on the ANN in different statistical packages a Predictive Model Markup Language file (PMML), which is an XML-based predictive model interchange format, is available in the supplemental materials.

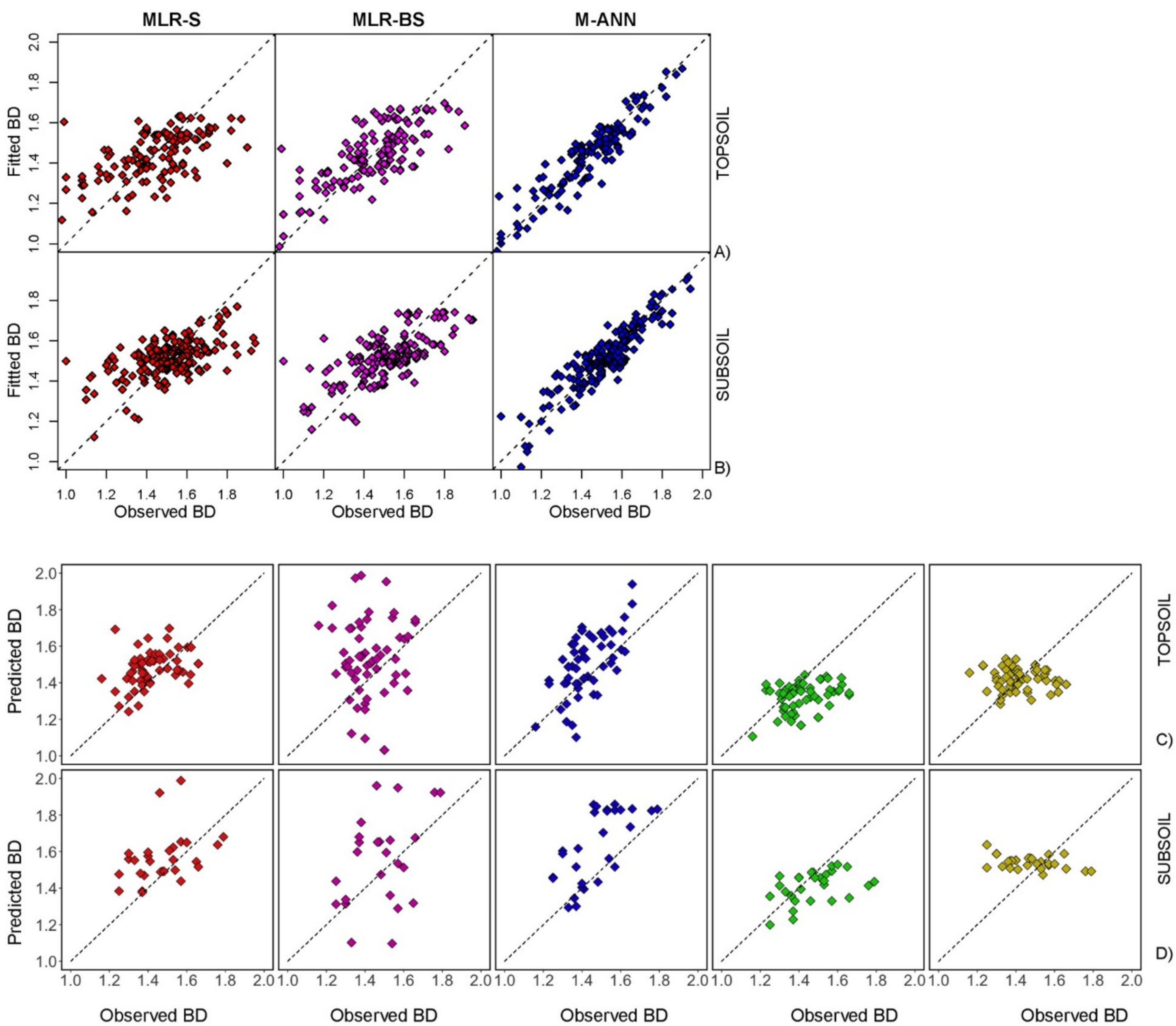

Figure 4. Predicted vs observed data (training topsoil and subsoil A and B, validation topsoil and subsoil C and D), MLR-S model, MLR-BS model, ANN neural network model, MJ PTF SoilGrids.

3.3 Variable importance

Table 8 shows the absolute standardized regression coefficient for each MLR model, considering 100% the highest beta value, to obtain a comparable result to the ANN model. Clay was the most important predictor in the topsoil MLR-S model. In the topsoil, SOC contributed approximately

25% of BD in the MLR-BS PTF, but it was not present in the MLR-S models. Similarly, $Clay^2$ was not present in the MLR-S models, slope and SOC^2 were the most important predictors in the subsoil using MLR-BS. Bioclimatic predictors such as BIO1 (Annual Mean Temperature), BIO2 (Mean Diurnal) and BIO7 (Annual T°C Range) were the most influential predictors in both topsoil and subsoil using MLR models. In topsoils, predictors included BIO7 (Annual T°C Range) and BIO14 (Precipitation of the Driest Month), heavily contributed to BD estimates within the MLR-BS and MLR-S of BD. BIO7 (Annual T°C Range) was more important than BIO14 (Precipitation of the Driest Month) in any model. Among the topographic predictors, the elevation was important in subsoil MLR-S models (contributing 24%), whereas it was not important in the topsoil or subsoil MLR-BS models. In subsoils, BIO3 (Isothermality) contributed 8% and 7% of subsoil BD in MLR-S and MLR-BS.

Table 8 Normalized variable importance in the MLR-S and MLR-BS (standardized regression coefficient in %). Conditional formatting is applied, Red color marks the minimum, green color the maximum and the yellow marks the middle values.

	MLR-S TOP	MLR-BS TOP	MLR-S SUB	MLR-BS SUB
Clay				4.62
Sand		5.83		
Silt		2.97	14.67	
SOC		19.15		4.69
MeanDepth			11.73	
Elevation		2.34		4.53
Slope			24.97	4.35
Profile Curvature			13.91	
BIO1				5.70
BIO2		11.23		5.91
BIO3		8.73		6.01
BIO4	36.46			
BIO5		4.59		5.20
BIO7	46.88	7.56	14.75	4.90
BIO12				6.09
BIO13				5.26
BIO14	16.67	1.21		5.02
BIO15		1.90		
BIO17				5.08
BIO18				4.85

BIO19			4.74
$Clay^2$	3.17		4.57
$Sand^2$	2.14		
SOC^2	6.10	19.96	4.81
SOC^2*Clay^2	3.98		4.76
$SOC^{\wedge 0.5}$	14.22		4.15
Clay*SOC	4.87		4.78

Table 9 shows the importance of the predictors included in the ANN models, based on sensitivity analyses using the default option in the MLP tool in IBM SPSS (independent variable importance analysis). The models included all the predictors except interactions between soil properties which were calculated within the ANN procedure but hidden to the user. Among the main physical and chemical properties, Sand was the most important in topsoil (5.6) and SOC in subsoil (5.5) Within the ANN, the most important predictors were BIO7 (Annual T°C Range) and Profile Curvature for topsoil and subsoil, respectively. Other important predictors were MeanDepth and the BIO12 (Annual Rainfall) in topsoil. SOC, BIO1 (Annual Mean Temperature) and BIO16 (Precipitation of the Wettest Quarter) were important in topsoil and subsoil. Soil properties predicted the BD of subsoil. Clay predicted BD in subsoil (3.2) and topsoil (2.8). Among the topographic predictors, Profile and plan curvature played a stronger role predicting the BD of the topsoil (5.3) compared to the subsoil (5.6).

Table 9. Normalized variable importance for predicting the bulk density of the top- and subsoil by means of the pedotransfer function developed by the Artificial Neural Network optimization approach. *variable description is available in figure 1, Conditional formatting is applied, Red color marks the minimum, green color the maximum and the yellow marks the middle values.*

	Topsoil	Subsoil
Clay	3.2	2.8
Sand	5.6	1.2
Silt	3.4	3.7
SOC	3.8	5.5
MeanDepth	1.4	1.5

Elevation	3.8	2.7
Slope	2.6	5.3
SIN Aspect	3.3	1.7
Profile Curvature	5.3	4.5
Plan Curvature	3.8	5.6
BIO1	2.5	2.8
BIO2	5.3	5.0
BIO3	2.0	2.7
BIO4	3.7	4.4
BIO5	4.7	1.6
BIO6	3.4	1.8
BIO7	3.9	5.4
BIO8	3.5	4.0
BIO9	3.8	2.9
BIO10	1.6	2.6
BIO11	2.9	1.8
BIO12	1.4	2.6
BIO13	3.0	2.3
BIO14	3.2	3.2
BIO15	5.6	2.3
BIO16	3.3	1.8
BIO17	3.9	3.5
BIO18	4.4	3.3
BIO19	1.7	2.7

4 Discussion

Collaborative work among researchers from different branches of geosciences facilitated a systematic literature review and compilation of an up-to-date, regionally-relevant geo-dataset; and this permitted the development and validation of new pedotransfer functions (PTF) to predict the BD of top- and subsoil in the Mediterranean.

4.1 Performance of pedotransfer functions

In this study, the performances of three new PTFs to estimate BD (MLR-S, MLR-BS and ANN) were defined used WoSIS soil data in combination with environmental data. The model transferability of the new PTFs was carried out using external dataset from Mediterranean locations derived from literature. Results were also benchmarked against the widely used MJ-PTF, which uses only soil organic carbon to predict BD, and the global SoilGrids data which are based on topographic and remote-sensed estimates of BD. Among the MLR approaches, MLR-BS performed slightly better than MLR-S. The ANN model outperformed the MLR models.

Our PTF development strategy, which implies the use of topographic and climatic variables along with soil properties, agreed with the approach of Wang et al., (2014) and Akpa et al., (2016). However, we also validated our new PTFs by estimating the BD in top- and subsoil using an external and independent dataset. This resulted in less accurate predictions than those made by the training datasets as already remarked by (Khaledian and Miller, 2020; Morin and Davis, 2017; Thompson, 2006). Nevertheless, these authors suggested that the use of external dataset rather than internal validation methods provides direct evidence about whether study results will replicate (Thompson, 2006).

Here we initially used MLR because it is a hands-on tool that provides direct quantitative and easily interpretable results. By contrast, the ANN has provided an alternative machine learning approach used in relatively recent analyses (Alvarez-Acosta et al., 2012; Ballabio et al., 2016; Chen et al., 2018; Ghehi et al., 2012; Nussbaum et al., 2018). In this study, the MJ PTF was used as a simple comparator because it is independent of other physical soil parameters except SOC and is the most widely used PTF. In our MLR-BS topsoil, MLR-S subsoil, and MLR-BS subsoil models, we considered inclusion of the key determinant used (i.e., SOC square root), but it was not included in the final model because it did not significantly improve the predictions. Our MLR-S and MLR-BS performed better than the MJ because our dataset included additional factors that directly determine BD over the long-term (such as those related to the climate or topography), thus raising the prediction capability. the SoilGrids database yielded a lower prediction ability in comparison to the other models. SoilGrids is considered as an interesting solution because it is a gridded multiple depth dataset at a 250m spatial resolution and it is available worldwide. However, the present results suggest that SoilGrids BD estimations may not adequately match the observed BD values (i.e., external dataset) which were measured in specific sites located in the Mediterranean area.

4.2 Data groupings and reliability of PTFs

The fit of the MLR subsoil models was more satisfactory than for subsoil linear model MLR-S with an R^2 0.12. Subsoil models were also more satisfactory for the ANN ($R^2 = 0.45$). This is in contrast to previous publications in which grouping input data by soil depths did not improve the prediction of BD in tropical soils (De Vos et al., 2005), which might have been attributable to different level of disturbance of the soil in the study areas (Hollis et al., 2012), or differences in the additional factors analyzed.

Arable soils undergo significant changes over time due to tillage and cultivation. Therefore, physical soil properties such as BD are more stable in the subsoil than in the topsoil. Statistics for soil texture and SOC agreed with data reported for other Mediterranean countries (Çelik et al., 2019; Evrendilek et al., 2004). The MJ and SoilGrids models yielded a similar result (Table 6 and 7). Notably, MLR-BS showed an R^2 close to zero for the validation datasets. This hampers discussion of model variability comparing the training and validation datasets. Generally, negative Bias is observed in the subsoil external dataset. As for the external dataset, slightly positive Bias indicates that MLR-S, ANN, SoilGrids overestimate the average BD of 2.8%, 1.7 % and 0.7%, respectively; MJ and MLR-BS underestimate the BD of -4.6% and -2.3% which is not preferable especially for the subsoil, an exception has been encountered with the topsoil MLR-BS that has predicted in few cases values very far from the true BD. Mediterranean soils are diverse, their hydraulic properties reflect pedogenetic factors as well as recent changes in management and climate (Yaalon, 1997).

4.3 Importance of predictor variables

Previous attempts to estimate soil BD by PTF (Çelik et al., 2019; Gozubuyuk et al., 2014; Tranter et al., 2007) did not include climate parameters because they were often not readily available or not immediately obvious as a determinant of BD. Many factors related to climate, such as bioclimatic indices, affect BD (e.g., rainfall intensity or pattern, high soil temperature in summer) (Basile et al., 2019; Chen et al., 2018). Bioclimatic indices and topographic predictors contributed greatly to the performance of the MLR and comprised 100% of the variables in the MLR-S for the topsoil, and about the 33% in the MLR-BS. The regression models (MLR-S and MLR-BS) included soil textural data (MLR-S topsoil) or SOC related (MLR-S subsoil and MLR-BS topsoil and subsoil). Our results showed that important predictors of BD in the MLR models were slope, clay, SOC^2, and bioclimatic variables such as BIO1 (Annual Mean Temperature), BIO2 (Mean Diurnal Range) and BIO7 (Annual T°C Range). This is consistent with previous reports (Akpa et al., 2016). In fact, part of the BD variability is due to the diverse bioclimatic zones within the Mediterranean Basin (Beck et al., 2018).

In our study, the inclusion of the climatic and topographic data increased the model reliability. Indeed, models without topographic and climatic predictors had very low performance at the training stage (data not shown). However, the lack of field management information, which strongly affects arable soils (e.g., crop type, tillage methods, irrigation, input of organic matter), hampers the ability to infer a relationship with factors of soil formation and processes (Wadoux et al., 2019) which would potentially improve the model prediction. In the Mediterranean Basin, significant effects of cropping systems and field managements on BD have been demonstrated in field studies (Álvaro-Fuentes et al., 2008; Bogunovic et al., 2020; Çelik et al., 2019; Perego et al., 2019; Pezzuolo et al., 2017).

5. Conclusions

Arable soils are widely distributed and the estimation of their fertility and carbon sequestration ability is a prerequisite for their management at wide scale. Reliable PTF to estimate BD are thus a needed instrument for arable soils management at the regional or higher levels. In the present study, we developed a robust PTF for BD estimation by exploiting the WoSIS resource, and it was the first time that such a broad set of data are valorized for PTF development. Moreover, we considered relevant predictors such as climatic and topographic parameters, which are fully and freely available and responsible for remarkably improving the predictive capability of the PTF models.

One of the three developed PTF (i.e., ANN) showed a better capability of estimating BD data than the well-known function Manrique Jones and the SoilGrids estimation approach; this outcome proved that the work hypothesis was correct and then developing the PTF with climate-specific set of data and adding topographic and climate predictors leads to a better predictive capability.

A relevant result of the present work is a ready to be used PTF model (i.e., ANN) for to separate soil layers (i.e., topsoil and subsoil) for the arable soils in the Mediterranean basin. The potential users of this result are public authorities interested in estimating soil carbon stock by exploiting legacy soil data in which bulk density is an often-missing parameter in the large monitoring campaigns. Researchers can be also interested in a more robust method of BD estimation when elaborating sets of soil data, especially when the aim is to estimate spatial and temporal variation.

The robustness of the ANN PTF is ensured by the use of an independent external dataset compiled from the literature for the validation of the PTF models transferability.

Results from the present work provide a reproducible and externally tested tool that can be applied to obtain a BD estimation at a regional level more reliable than the presently used PTF or gridded benchmarks. Thus, the present results are an option for policy making and management at a regional level.

6. Acknowledgments

All authors acknowledge the funding for the Lake Como Advanced School provided by the grant for the SDAE Summer School 2019 from the Italian Society of Agronomy, University of Milan and University of Pavia. WLC. Landsupport H2020, grant number ID: 774234. Project "Carbon Market - Innovative cropping systems for carbon market" funded by Natural Resources Institute Finland (Luke). We express our gratitude to Cami Moss for editing the manuscript.

Views, thoughts, and opinions expressed in the text belong solely to the author and do not reflect the opinion of the author's employer, organization, committee members, or that of other course participants.

This paper is dedicated to the dear memory of Prof: Dario Sacco, who was a teacher of the SDAE School.

References

Acutis, M., Donatelli, M., 2003. SOILPAR 2.00: Software to estimate soil hydrological parameters and functions, in: European Journal of Agronomy. pp. 373–377. https://doi.org/10.1016/S1161-0301(02)00128-4

Addiscott, T.M., Whitmore, A.P., 1987. Computer simulation of changes in soil mineral nitrogen and crop nitrogen during autumn, winter and spring. J. Agr. Sci. 109, 141–157.

Aitkenhead, M., Coull, M., 2020. Mapping soil profile depth, bulk density and carbon stock in Scotland using remote sensing and spatial covariates. Eur. J. Soil Sci. https://doi.org/10.1111/ejss.12916

Akpa, S.I.C., Ugbaje, S.U., Bishop, T.F.A., Odeh, I.O.A., 2016. Enhancing pedotransfer functions with environmental data for estimating bulk density and effective cation exchange capacity in a data-sparse situation. Soil Use Manag. 32, 644–658. https://doi.org/10.1111/sum.12310

Alvarez-Acosta, C., Lascano, R.J., Stroosnijder, L., 2012. Test of the Rosetta Pedotransfer Function for Saturated Hydraulic Conductivity. Open J. Soil Sci. 02, 203–212. https://doi.org/10.4236/ojss.2012.23025

Álvaro-Fuentes, J., López, M. V., Cantero-Martinez, C., Arrúe, J.L., 2008. Tillage Effects on Soil Organic Carbon Fractions in Mediterranean Dryland Agroecosystems. Soil Sci. Soc. Am. J. 72, 541–547. https://doi.org/10.2136/sssaj2007.0164

Álvaro-Fuentes, J., Paustian, K., 2011. Potential soil carbon sequestration in a semiarid Mediterranean agroecosystem under climate change: Quantifying management and climate effects. Plant Soil 338, 261–272. https://doi.org/10.1007/s11104-010-0304-7

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

Basile, A., Bonfante, A., Coppola, A., De Mascellis, R., Falanga Bolognesi, S., Terribile, F., Manna, P., 2019. How does PTF Interpret Soil Heterogeneity? A Stochastic Approach Applied to a Case Study on Maize in Northern Italy. Water 11, 275.

https://doi.org/10.3390/w11020275
Batjes, N.H., Dijkshoorn, J.A., 1999. Carbon and nitrogen stocks in the soils of the Amazon Region. Geoderma 89, 273–286. https://doi.org/10.1016/S0016-7061(98)00086-X
Batjes, N.H., Ribeiro, E., van Oostrum, A., Leenaars, J., Hengl, T., Mendes de Jesus, J., 2017. WoSIS: providing standardised soil profile data for the world. Earth Syst. Sci. Data 9, 1–14. https://doi.org/10.5194/essd-9-1-2017
Beck, H.E., Zimmermann, N.E., McVicar, T.R., Vergopolan, N., Berg, A., Wood, E.F., 2018. Present and future köppen-geiger climate classification maps at 1-km resolution. Sci. Data 5, 1–12. https://doi.org/10.1038/sdata.2018.214
Bellocchi, G., Acutis, M., Fila, G., Donatelli, M., 2002. An indicator of solar radiation model performance based on a fuzzy expert system. Agron. J. 94, 1222–1233. https://doi.org/10.2134/agronj2002.1222
Benites, V.M., Machado, P.L.O.A., Fidalgo, E.C.C., Coelho, M.R., Madari, B.E., 2007. Pedotransfer functions for estimating soil bulk density from existing soil survey reports in Brazil. Geoderma 139, 90–97. https://doi.org/10.1016/j.geoderma.2007.01.005
Bogunovic, I., Pereira, P., Galic, M., Bilandzija, D., Kisic, I., 2020. Tillage system and farmyard manure impact on soil physical properties, CO2 emissions, and crop yield in an organic farm located in a Mediterranean environment (Croatia). Environ. Earth Sci. 79, 1–11. https://doi.org/10.1007/s12665-020-8813-z
Bondi, G., Creamer, R., Ferrari, A., Fenton, O., Wall, D., 2018. Using machine learning to predict soil bulk density on the basis of visual parameters: Tools for in-field and post-field evaluation. Geoderma 318, 137–147. https://doi.org/10.1016/j.geoderma.2017.11.035
Bossard, M., Feranec, J., Otahel, J., Steenmans, C., 2000. CORINE land cover technical guide-

Addendum 2000.

Bouma, J., 1989. Using Soil Survey Data for Quantitative Land Evaluation. pp. 177–213. https://doi.org/10.1007/978-1-4612-3532-3_4

Breusch, T., Pagan, A.R., Breusch, T., Pagan, A., 1979. A Simple Test for Heteroscedasticity and Random Coefficient Variation. Econometrica 47, 1287–94.

Çelik, İ., Günal, H., Acar, M., Acir, N., Bereket Barut, Z., Budak, M., 2019. Strategic tillage may sustain the benefits of long-term no-till in a Vertisol under Mediterranean climate. Soil Tillage Res. 185, 17–28. https://doi.org/10.1016/j.still.2018.08.015

Chagas, C. da S., de Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena 139, 232–240. https://doi.org/10.1016/j.catena.2016.01.001

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

Chen, S., Xu, D., Li, S., Ji, W., Yang, M., Zhou, Y., Hu, B., Xu, H., Shi, Z., 2020. Monitoring soil organic carbon in alpine soils using in situ vis-NIR spectroscopy and a multilayer perceptron. L. Degrad. Dev. 31, 1026–1038. https://doi.org/10.1002/ldr.3497

Colombi, T., Torres, L.C., Walter, A., Keller, T., 2018. Feedbacks between soil penetration resistance, root architecture and water uptake limit water accessibility and crop growth – A vicious circle. Sci. Total Environ. 626, 1026–1035. https://doi.org/10.1016/j.scitotenv.2018.01.129

de Souza, E., Fernandes Filho, E.I., Schaefer, C.E.G.R., Batjes, N.H., dos Santos, G.R., Pontes, L.M., 2016. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. Sci. Agric. 73, 525–534. https://doi.org/10.1590/0103-9016-2015-0485

De Vos, B., Van Meirvenne, M., Quataert, P., Deckers, J., Muys, B., 2005. Predictive quality of pedotransfer functions for estimating bulk density of forest soils. Soil Sci. Soc. Am. J. 69, 500–510. https://doi.org/10.2136/sssaj2005.0500

Draper, N.R., Smith, H., 1998. Applied Regression Analysis, Third Edition, in: DOI:10.1002/9781118625590 1998 John Wiley & Sons, Inc. pp. 1–704.

Ebrahimi, M., Sarikhani, M.R., Safari Sinegani, A.A., Ahmadi, A., Keesstra, S., 2019. Estimating the soil respiration under different land uses using artificial neural network and linear regression models. Catena 174, 371–382. https://doi.org/10.1016/j.catena.2018.11.035

Evrendilek, F., Celik, I., Kilic, S., 2004. Changes in soil organic carbon and other physical soil properties along adjacent Mediterranean forest, grassland, and cropland ecosystems in Turkey. J. Arid Environ. 59, 743–752. https://doi.org/10.1016/j.jaridenv.2004.03.002

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., Alsdorf, D.E., 2007. The shuttle radar topography mission. Rev. Geophys. 45. https://doi.org/10.1029/2005RG000183

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086

Fila, G., Bellocchi, G., Donatelli, M., Acutis, M., 2003. IRENE_DLL: A Class Library for Evaluating Numerical Estimates. Agron. J. 95, 1330–1333.

https://doi.org/10.2134/agronj2003.1330

Fox, D.G., 1981. Judging Air Quality Model Performance. https://doi.org/10.1175/1520-0477(1981)062<0599:JAQMP>2.0.CO;2

Francaviglia, R., Álvaro-Fuentes, J., Di Bene, C., Gai, L., Regina, K., Turtola, E., 2020. Diversification and management practices in selected European regions. A data analysis of arable crops production. Agronomy 10, 297. https://doi.org/10.3390/agronomy10020297

Ghani, I.M.M., Ahmad, S., 2010. Stepwise multiple regression method to forecast fish landing, in: Procedia - Social and Behavioral Sciences. Elsevier Ltd, pp. 549–554. https://doi.org/10.1016/j.sbspro.2010.12.076

Ghehi, N.G., Nemes, A., Verdoodt, A., Van Ranst, E., Cornelis, W.M., Boeckx, P., 2012. Nonparametric Techniques for Predicting Soil Bulk Density of Tropical Rainforest Topsoils in Rwanda. Soil Sci. Soc. Am. J. 76, 1172–1183. https://doi.org/10.2136/sssaj2011.0330

Gozubuyuk, Z., Sahin, U., Ozturk, I., Celik, A., Adiguzel, M.C., 2014. Tillage effects on certain physical and hydraulic properties of a loamy soil under a crop rotation in a semi-arid region with a cool climate. Catena 118, 195–205. https://doi.org/10.1016/j.catena.2014.01.006

Håkansson, I., Lipiec, J., 2000. A review of the usefulness of relative bulk density values in studies of soil structure and compaction. Soil Tillage Res. https://doi.org/10.1016/S0167-1987(99)00095-1

Hamilton, N.E., Ferry, M., 2018. Ggtern: Ternary diagrams using ggplot2. J. Stat. Softw. 87, 1–17. https://doi.org/10.18637/jss.v087.c03

Haykin, S., 2008. Neural Networks and Learning Machines, Pearson Prentice Hall New Jersey USA 936. https://doi.org/978-0131471399

Hengl, T., De Jesus, J.M., Heuvelink, G.B.M., Gonzalez, M.R., 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., Kempen, B., 2017. SoilGrids250m: Global gridded soil information based on machine learning. PLoS One 12. https://doi.org/10.1371/journal.pone.0169748

Hiederer, R., Jones, R.J.A., Daroussin, J., 2006. Soil Profile Analytical Database for Europe (SPADE): Reconstruction and validation of the measured data (SPADE/M). Geogr. Tidsskr. 106, 71–85. https://doi.org/10.1080/00167223.2006.10649546

Hollis, J.M., Hannam, J., Bellamy, P.H., 2012. Empirically-derived pedotransfer functions for predicting bulk density in European soils. Eur. J. Soil Sci. 63, 96–109. https://doi.org/10.1111/j.1365-2389.2011.01412.x

Khaledian, Y., Miller, B.A., 2020. Selecting appropriate machine learning methods for digital soil mapping. Appl. Math. Model. 81, 401–418. https://doi.org/10.1016/j.apm.2019.12.016

Lagacherie, P., Álvaro-Fuentes, J., Annabi, M., Bernoux, M., Bouarfa, S., Douaoui, A., Grünberger, O., Hammani, A., Montanarella, L., Mrabet, R., Sabir, M., Raclot, D., 2018. Managing Mediterranean soil resources under global change: expected trends and mitigation strategies. Reg. Environ. Chang. 18, 663–675. https://doi.org/10.1007/s10113-017-1239-9

Leij, F.J., Romano, N., Palladino, M., Schaap, M.G., Coppola, A., 2004. Topographical attributes to predict soil hydraulic properties along a hillslope transect. Water Resour. Res. 40. https://doi.org/10.1029/2002WR001641

Lionello, P., Malanotte-Rizzoli, P., Boscolo, R., Alpert, P., Artale, V., Li, L., Luterbacher, J., May, W., Trigo, R., Tsimplis, M., Ulbrich, U., Xoplaki, E., 2006. The Mediterranean climate: An overview of the main characteristics and issues. Dev. Earth Environ. Sci.

https://doi.org/10.1016/S1571-9197(06)80003-0
Makovníková, J., Širáň, M., Houšková, B., Pálka, B., Jones, A., 2017. Comparison of different models for predicting soil bulk density. Case study - Slovakian agricultural soils. Int. Agrophysics 31, 491–498. https://doi.org/10.1515/intag-2016-0079
Manrique, L.A., Jones, C.A., 1991. Bulk density of soils in relation to soil physical and chemical properties. Soil Sci. Soc. Am. J. 55, 476–481. https://doi.org/10.2136/sssaj1991.03615995005500020030x
Martín, M.Á., Reyes, M., Taguas, F.J., 2017. Estimating soil bulk density with information metrics of soil texture. Geoderma 287, 66–70. https://doi.org/10.1016/j.geoderma.2016.09.008
Mazzoncini, M., Sapkota, T.B., Bàrberi, P., Antichi, D., Risaliti, R., Bahadur, T., Ba, P., 2011. Long-term effect of tillage, nitrogen fertilization and cover crops on soil organic carbon and total nitrogen content. Soil Tillage Res. 114, 165–174. https://doi.org/10.1016/j.still.2011.05.001
Minasny, B., McBratney, A.B., 2002. The neuro-m method for fitting neural network parametric pedotransfer functions. Soil Sci. Soc. Am. J. 66, 352–361. https://doi.org/10.2136/sssaj2002.3520
Minasny, B., McBratney, A.B., Malone, B.P., Wheeler, I., 2013. Chapter One – Digital Mapping of Soil Carbon, in: Advances in Agronomy. pp. 1–47. https://doi.org/10.1016/B978-0-12-405942-9.00001-3
Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., Vereecken, H., 2017. A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. Earth Syst. Sci. Data 9, 529–543. https://doi.org/10.5194/essd-9-529-

2017

Moriasi D., Arnold J. G., Van Liew M. W., Bingner R. L., Harmel R. D., Veith T. L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Trans. ASABE 50, 885–900. https://doi.org/10.13031/2013.23153

Morin, K., Davis, J.L., 2017. Cross-validation: What is it and how is it used in regression? Commun. Stat. - Theory Methods 46, 5238–5251. https://doi.org/10.1080/03610926.2015.1099672

Nasta, P., Palladino, M., Sica, B., Pizzolante, A., Trifuoggi, M., Toscanesi, M., Giarra, A., D'Auria, J., Nicodemo, F., Mazzitelli, C., Lazzaro, U., Di Fiore, P., Romano, N., 2020. Evaluating pedotransfer functions for predicting soil bulk density using hierarchical mapping information in Campania, Italy. Geoderma Reg. https://doi.org/10.1016/j.geodrs.2020.e00267

Noryani, M., Sapuan, S.M., Mastura, M.T., Zuhri, M.Y.M., Zainudin, E.S., 2019. Material selection of natural fibre using a stepwise regression model with error analysis. J. Mater. Res. Technol. 8, 2865–2879. https://doi.org/10.1016/j.jmrt.2019.02.019

Nussbaum, M., Spiess, K., Baltensweiler, A., Grob, U., Keller, A., Greiner, L., Schaepman, M.E., Papritz, A., 2018. Evaluation of digital soil mapping approaches with large sets of environmental covariates. SOIL 4, 1–22. https://doi.org/10.5194/soil-4-1-2018

Pachepsky, Y.A., Timlin, D., Varallyay, G., 1996. Artificial neural networks to estimate soil water retention from easily measurable data. Soil Sci. Soc. Am. J. 60, 727–733. https://doi.org/10.2136/sssaj1996.03615995006000030007x

Panagos, P., Hiederer, R., Van Liedekerke, M., Bampa, F., 2013. Estimating soil organic carbon in Europe based on data collected through an European network. Ecol. Indic. 24, 439–450.

https://doi.org/10.1016/j.ecolind.2012.07.020

Patil, N.G., Singh, S.K., 2016. Pedotransfer Functions for Estimating Soil Hydraulic Properties: A Review. Pedosphere 26, 417–430. https://doi.org/10.1016/S1002-0160(15)60054-6

Perego, A., Rocca, A., Cattivelli, V., Tabaglio, V., Fiorini, A., Barbieri, S., Schillaci, C., Chiodini, M.E., Brenna, S., Acutis, M., 2019. Agro-environmental aspects of conservation agriculture compared to conventional systems: A 3-year experience on 20 farms in the Po valley (Northern Italy). Agric. Syst. 168, 73–87. https://doi.org/10.1016/j.agsy.2018.10.008

Pezzuolo, A., Dumont, B., Sartori, L., Marinello, F., De Antoni Migliorati, M., Basso, B., 2017. Evaluating the impact of soil conservation measures on soil organic carbon at the farm scale. Comput. Electron. Agric. 135, 175–182. https://doi.org/10.1016/j.compag.2017.02.004

Picciafuoco, T., Morbidelli, R., Flammini, A., Saltalippi, C., Corradini, C., Strauss, P., Blöschl, G., 2019. A pedotransfer function for field-scale saturated hydraulic conductivity of a small watershed. Vadose Zo. J. 18. https://doi.org/10.2136/vzj2019.02.0018

Piñeiro, G., Perelman, S., Guerschman, J.P., Paruelo, J.M., 2008. How to evaluate models: Observed vs. predicted or predicted vs. observed? Ecol. Modell. 216, 316–322. https://doi.org/10.1016/j.ecolmodel.2008.05.006

Premrov, A., Cummins, T., Byrne, K.A., 2018. Bulk-density modelling using optimal power-transformation of measured physical and chemical soil parameters. Geoderma 314, 205–220. https://doi.org/10.1016/j.geoderma.2017.10.060

Ramcharan, A., Hengl, T., Beaudette, D., Wills, S., 2017. A soil bulk density pedotransfer function based on machine learning: A case study with the ncss soil characterization database. Soil Sci. Soc. Am. J. 81, 1279–1287. https://doi.org/10.2136/sssaj2016.12.0421

Rawls, W.J., 1983. Estimating soil bulk density from particle size analysis and organic matter content. Soil Sci. https://doi.org/10.1097/00010694-198302000-00007

Rawls, W.J., Pachepsky, Y.A., 2002. Using field topographic descriptors to estimate soil water retention. Soil Sci. 167, 423–435. https://doi.org/10.1097/00010694-200207000-00001

Reidy, B., Simo, I., Sills, P., Creamer, R.E., 2016. Pedotransfer functions for Irish soils – estimation of bulk density (ρ_b) per horizon type. SOIL 2, 25–39. https://doi.org/10.5194/soil-2-25-2016

Reynolds, C.A., Jackson, T.J., Rawls, W.J., 2000. Estimating soil water-holding capacities by linking the Food and Agriculture Organization Soil map of the world with global pedon databases and continuous pedotransfer functions. Water Resour. Res. 36, 3653–3662. https://doi.org/10.1029/2000WR900130

Rodríguez-Lado, L., Rial, M., Taboada, T., Cortizas, A.M., 2015. A Pedotransfer Function to Map Soil Bulk Density from Limited Data. Procedia Environ. Sci. 27, 45–48. https://doi.org/10.1016/j.proenv.2015.07.112

Román Dobarco, M., Cousin, I., Le Bas, C., Martin, M.P., 2019. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. Geoderma 336, 81–95. https://doi.org/10.1016/j.geoderma.2018.08.022

Romano, N., Chirico, G.B., 2004. The role of terrain analysis in using and developing pedotransfer functions. Dev. Soil Sci. https://doi.org/10.1016/S0166-2481(04)30016-4

Saxton, K.E., Rawls, W.J., Romberger, J.S., Papendick, R.I., 1986. Estimating generalized soil-water characteristics from texture. Soil Sci. Soc. Am. J. 50, 1031–1036. https://doi.org/10.2136/sssaj1986.03615995005000040039x

Schaap, M.G., Leij, F.J., Van Genuchten, M.T., 1998. Neural network analysis for hierarchical prediction of soil hydraulic properties. Soil Sci. Soc. Am. J. 62, 847–855. https://doi.org/10.2136/sssaj1998.03615995006200040001x

Schillaci, C., Acutis, M., Lombardo, L., Lipani, A., Fantappiè, M., Märker, M., Saia, S., 2017a. Spatio-temporal topsoil organic carbon mapping of a semi-arid Mediterranean region: The role of land use, soil texture, topographic indices and the influence of remote sensing data to modelling. Sci. Total Environ. 601–602, 821–832. https://doi.org/10.1016/j.scitotenv.2017.05.239

Schillaci, C., Acutis, M., Vesely, F., Saia, S., 2019. A simple pipeline for the assessment of legacy soil datasets: An example and test with soil organic carbon from a highly variable area. Catena 175, 110–122. https://doi.org/10.1016/j.catena.2018.12.015

Schillaci, C., Lombardo, L., Saia, S., Fantappiè, M., Märker, M., Acutis, M., 2017b. Modelling the topsoil carbon stock of agricultural lands with the Stochastic Gradient Treeboost in a semi-arid Mediterranean region. Geoderma 286, 35–45. https://doi.org/10.1016/j.geoderma.2016.10.019

Schillaci, C., Saia, S., Acutis, M., 2018. Modelling of Soil Organic Carbon in the Mediterranean area: a systematic map. Rend. Online della Soc. Geol. Ital. 4, 161–166.

Sequeira, C.H., Wills, S.A., Seybold, C.A., West, L.T., 2014. Predicting soil bulk density for incomplete databases. Geoderma 213, 64–73. https://doi.org/10.1016/j.geoderma.2013.07.013

Sevastas, S., Gasparatos, D., Botsis, D., Siarkos, I., Diamantaras, K.I., Bilas, G., 2018. Predicting bulk density using pedotransfer functions for soils in the Upper Anthemountas basin, Greece. Geoderma Reg. 14. https://doi.org/10.1016/j.GEODRS.2018.e00169

Tao, F., Palosuo, T., Valkama, E., Mäkipää, R., 2019. Cropland soils in China have a large potential for carbon sequestration based on literature survey. Soil Tillage Res. https://doi.org/10.1016/j.still.2018.10.009

Tejada, M., Hernandez, M.T., Garcia, C., 2009. Soil restoration using composted plant residues: Effects on soil properties. Soil Tillage Res. 102, 109–117. https://doi.org/10.1016/j.still.2008.08.004

Thompson, B., 2006. Foundations of Behavioral Statistics, pp 470. Guilford Publications.

Throop, H.L., Archer, S.R., Monger, H.C., Waltman, S., 2012. When bulk density methods matter: Implications for estimating soil organic carbon pools in rocky soils. J. Arid Environ. 77, 66–71. https://doi.org/10.1016/j.jaridenv.2011.08.020

Tietje, O., Tapkenhinrichs, M., 1993. Evaluation of Pedo-Transfer Functions. Soil Sci. Soc. Am. J. 57, 1088–1095. https://doi.org/10.2136/sssaj1993.03615995005700040035x

Tranter, G., Minasny, B., Mcbratney, A.B., Murphy, B., Mckenzie, N.J., Grundy, M., Brough, D., 2007. Building and testing conceptual and empirical models for predicting soil bulk density. Soil Use Manag. 23, 437–443. https://doi.org/10.1111/j.1475-2743.2007.00092.x

Valkama, E., Kunypiyaeva, G., Zhapayev, R., Karabayev, M., Zhusupbekov, E., Perego, A., Schillaci, C., Sacco, D., Moretti, B., Grignani, C., Acutis, M., 2020. Can conservation agriculture increase soil carbon sequestration? A modelling approach. Geoderma 369, 114298. https://doi.org/10.1016/j.geoderma.2020.114298

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., Vereecken, H., 2017. Pedotransfer Functions in Earth System Science: Challenges and Perspectives. Rev. Geophys. https://doi.org/10.1002/2017RG000581

Verheye, W., De La Rosa, D., 2005. ©UNESCO-EOLSS Encyclopedia of Life Support Systems Mediterranean Soils.

Veronesi, F., Schillaci, C., 2019. Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation. Ecol. Indic. 101, 1032–1044. https://doi.org/10.1016/j.ecolind.2019.02.026

Wadoux, A.M.J.C., Samuel-Rosa, A., Poggio, L., Mulder, V.L., 2019. A note on knowledge discovery and machine learning in digital soil mapping. Eur. J. Soil Sci. https://doi.org/10.1111/ejss.12909

Wang, Y., Shao, M., Liu, Z., Zhang, C., 2014. Prediction of Bulk Density of Soils in the Loess Plateau Region of China. Surv. Geophys. https://doi.org/10.1007/s10712-013-9249-8

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

Wösten, J.H.M., Verzandvoort, S.J.E., Leenaars, J.G.B., Hoogland, T., Wesseling, J.G., 2013. Soil hydraulic information for river basin studies in semi-arid regions. Geoderma 195–196, 79–86. https://doi.org/10.1016/j.geoderma.2012.11.021

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

Yaalon, D.H., 1997. Soils in the Mediterranean region: What makes them different? Catena 28, 157–169. https://doi.org/10.1016/S0341-8162(96)00035-5

Zdruli, P., Kapur, S., Çelik, I., 2011. Soils of the mediterranean region, their characteristics, management and sustainable use, in: Sustainable Land Management: Learning from the Past for the Future. Springer Berlin Heidelberg, pp. 125–142. https://doi.org/10.1007/978-3-642-14782-1_4

--------------------- REFERENCE FOR TAB 2 --

Abou Zakhem, B., Al Ain, F., Hafez, R., 2019. Assessment of Field Water Budget Components for Increasing Water Productivity Under Drip Irrigation in Arid and Semi-Arid Areas, Syria. Irrig. Drain. 68, 452–463. https://doi.org/10.1002/ird.2286

Ali, S.A., Tedone, L., Verdini, L., Cazzato, E., De Mastro, G., 2019. Wheat response to no-tillage and nitrogen fertilization in a long-term faba bean-based rotation. Agronomy 9, 50. https://doi.org/10.3390/agronomy9020050

Álvaro-Fuentes, J., López, M. V., Cantero-Martinez, C., Arrúe, J.L., 2008. Tillage Effects on Soil Organic Carbon Fractions in Mediterranean Dryland Agroecosystems. Soil Sci. Soc. Am. J. 72, 541–547. https://doi.org/10.2136/sssaj2007.0164

Antonopoulos, V.Z., Georgiou, P.E., Kolotouros, C.A., 2013. Soil water dynamics in cropped and uncropped fields in northern Greece using a dual-permeability model. Hydrol. Sci. J. 58, 1748–1759. https://doi.org/10.1080/02626667.2013.816424

Bescansa, P., Imaz, M.J., Virto, I., Enrique, A., Hoogmoed, W.B., 2006. Soil water retention as affected by tillage and residue management in semiarid Spain. Soil Tillage Res. 87, 19–27. https://doi.org/10.1016/j.still.2005.02.028

Bogunovic, I., Pereira, P., Kisic, I., Sajko, K., Sraka, M., 2018. Tillage management impacts on soil compaction, erosion and crop yield in Stagnosols (Croatia). Catena 160, 376–384. https://doi.org/10.1016/j.catena.2017.10.009

Cardinael, R., Chevallier, T., Cambou, A., Béral, C., Barthès, B.G., Dupraz, C., Durand, C., Kouakoua, E., Chenu, C., 2017. Increased soil organic carbon stocks under agroforestry: A survey of six different sites in France. Agric. Ecosyst. Environ. 236, 243–255. https://doi.org/10.1016/j.agee.2016.12.011

Carozzi, M., Bregaglio, S., Scaglia, B., Bernardoni, E., Acutis, M., Confalonieri, R., 2013. The development of a methodology using fuzzy logic to assess the performance of cropping systems based on a case study of maize in the Po Valley. Soil Use Manag. 29, 576–585. https://doi.org/10.1111/sum.12066

Çelik, İ., Günal, H., Acar, M., Acir, N., Bereket Barut, Z., Budak, M., 2019. Strategic tillage may sustain the benefits of long-term no-till in a Vertisol under Mediterranean climate. Soil Tillage Res. 185, 17–28. https://doi.org/10.1016/j.still.2018.08.015

Ceotto, E., Marchetti, R., Castelli, F., 2018. Residual soil nitrate as affected by giant reed cultivation and cattle slurry fertilisation. Ital. J. Agron. 13, 317–323. https://doi.org/10.4081/ija.2018.1264

Chennafi, H., Aïdaoui, A., Bouzerzour, H., Saci, A., 2006. Yield response of durum wheat (Triticum durum Desf.) cultivar Waha to deficit irrigation under semi arid growth conditions.

Asian J. Plant Sci. 5, 854–860. https://doi.org/10.3923/ajps.2006.854.860

Diacono, M., Ciaccia, C., Canali, S., Fiore, A., Montemurro, F., 2018. Assessment of agro-ecological service crop managements combined with organic fertilisation strategies in organic melon crop. Ital. J. Agron. 13, 172–182. https://doi.org/10.4081/ija.2018.951

Jemai, I., Ben Aissa, N., Ben Guirat, S., Ben-Hammouda, M., Gallali, T., 2013. Impact of three and seven years of no-tillage on the soil water storage, in the plant root zone, under a dry subhumid Tunisian climate. Soil Tillage Res. 126, 26–33. https://doi.org/10.1016/j.still.2012.07.008

Kargas, G., Kerkides, P., Poulovassilis, A., 2012. Infiltration of rain water in semi-arid areas under three land surface treatments. Soil Tillage Res. 120, 15–24. https://doi.org/10.1016/j.still.2012.01.004

Mahmoud, E., Ibrahim, M., Abd El-Rahman, L., Khader, A., 2019. Effects of Biochar and Phosphorus Fertilizers on Phosphorus Fractions, Wheat Yield and Microbial Biomass Carbon in Vertic Torrifluvents. Commun. Soil Sci. Plant Anal. 50, 362–372. https://doi.org/10.1080/00103624.2018.1563103

Marquez-Garcia, F., Gonzalez-Sanchez, E.J., Castro-Garcia, S., Ordoñez-Fernandez, R., 2013. Improvement of soil carbon sink by cover crops in olive orchards under semiarid conditions. Influence of the type of soil and weed. Spanish J. Agric. Res. 11, 335–346. https://doi.org/10.5424/sjar/2013112-3558

Muñoz-Rojas, M., Doro, L., Ledda, L., Francaviglia, R., 2015. Application of CarboSOIL model to predict the effects of climate change on soil organic carbon stocks in agro-silvo-pastoral Mediterranean management systems. Agric. Ecosyst. Environ. 202, 8–16. https://doi.org/10.1016/j.agee.2014.12.014

Ozpinar, S., Ozpinar, A., Cay, A., 2018. Soil management effect on soil properties in traditional and mechanized vineyards under a semiarid Mediterranean environment. Soil Tillage Res. 178, 198–208. https://doi.org/10.1016/j.still.2018.01.004

Pardo, J.J., Martínez-Romero, A., Léllis, B.C., Tarjuelo, J.M., Domínguez, A., 2020. Effect of the optimized regulated deficit irrigation methodology on water use in barley under semiarid conditions. Agric. Water Manag. 228, 105925. https://doi.org/10.1016/j.agwat.2019.105925

Pareja-Sánchez, E., Plaza-Bonilla, D., Ramos, M.C., Lampurlanés, J., Álvaro-Fuentes, J., Cantero-Martínez, C., 2017. Long-term no-till as a means to maintain soil surface structure in an agroecosystem transformed into irrigation. Soil Tillage Res. 174, 221–230. https://doi.org/10.1016/j.still.2017.07.012

Perego, A., Rocca, A., Cattivelli, V., Tabaglio, V., Fiorini, A., Barbieri, S., Schillaci, C., Chiodini, M.E., Brenna, S., Acutis, M., 2019. Agro-environmental aspects of conservation agriculture compared to conventional systems: A 3-year experience on 20 farms in the Po valley (Northern Italy). Agric. Syst. 168, 73–87. https://doi.org/10.1016/j.agsy.2018.10.008

Pezzuolo, A., Dumont, B., Sartori, L., Marinello, F., De Antoni Migliorati, M., Basso, B., 2017. Evaluating the impact of soil conservation measures on soil organic carbon at the farm scale.

Comput. Electron. Agric. 135, 175–182. https://doi.org/10.1016/j.compag.2017.02.004
Recio, J., Vallejo, A., Le-Noë, J., Garnier, J., García-Marco, S., Álvarez, J.M., Sanz-Cobena, A., 2018. The effect of nitrification inhibitors on NH3 and N2O emissions in highly N fertilized irrigated Mediterranean cropping systems. Sci. Total Environ. 636, 427–436. https://doi.org/10.1016/j.scitotenv.2018.04.294
Stavi, I., Ungar, E.D., Lavee, H., Sarah, P., 2008. Grazing-induced spatial variability of soil bulk density and content of moisture, organic carbon and calcium carbonate in a semi-arid rangeland. Catena 75, 288–296. https://doi.org/10.1016/j.catena.2008.07.007
Tolon-Becerra, A., Tourn, M., Botta, G.F., Lastra-Bravo, X., 2011. Effects of different tillage regimes on soil compaction, maize (Zea mays L.) seedling emergence and yields in the eastern Argentinean Pampas region. Soil Tillage Res. 117, 184–190. https://doi.org/10.1016/j.still.2011.10.003
Valboa, G., Lagomarsino, A., Brandi, G., Agnelli, A.E., Simoncini, S., Papini, R., Vignozzi, N., Pellegrini, S., 2015. Long-term variations in soil organic matter under different tillage intensities. Soil Tillage Res. 154, 126–135. https://doi.org/10.1016/j.still.2015.06.017
Visconti, F., Salvador, A., Navarro, P., de Paz, J.M., 2019. Effects of three irrigation systems on 'Piel de sapo' melon yield and quality under salinity conditions. Agric. Water Manag. 226. https://doi.org/10.1016/j.agwat.2019.105829
Vitale, L., Polimeno, F., Ottaiano, L., Maglione, G., Tedeschi, A., Mori, M., De Marco, A., Di Tommasi, P., Magliulo, V., 2017. Fertilizer type influences tomato yield and soil N2O emissions. Plant, Soil Environ. 63, 105–110. https://doi.org/10.17221/678/2016-PSE
Zohry, A., Ouda, S., Hamd-alla, W., Shalaby, E.-S., 2017. Evaluation of different crop sequences for wheat and maize in sandy soil. https://doi.org/10.14720/aas.2017.109.2.21