Using machine learning to predict soil bulk density on the basis of visual parameters: tools for in-field and post-field evaluation

Giulia Bondi[a*], Rachel Creamer[b], Alessio Ferrari[c], Owen Fenton[a], David Wall[a]

[a]Teagasc Crops, Environment and Land-Use Research Centre, Wexford, Ireland;

[b]Soil Biology and Biological Soil Quality, Wageningen University, Wageningen, The Netherlands;

[c]Consiglio Nazionale delle Ricerche, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (CNR-ISTI), Pisa, Italy.

***Corresponding author**: G. Bondi (Email: Giulia.Bondi@teagasc.ie)*

R. Creamer (Email: rachel.creamer@wur.nl); A. Ferrari (Email: alessio.ferrari@isti.cnr.it); O. Fenton (Email: Owen.Fenton@teagasc.ie); D. Wall (Email: David.Wall@teagasc.ie).

Abstract

Soil structure is a key factor that supports all soil functions. Extracting intact soil cores and horizon specific samples for determination of soil physical parameters (e.g. bulk density (B_d) or particle size distribution) is a common practice for assessing indicators of soil structure. However, these are often difficult to measure, since they require expensive and time consuming laboratory analyses. Our aim was to provide tools, through the use of machine learning techniques, to estimate the value of B_d based solely on soil visual assessment, observed by operators directly in the field. The first tool was a decision tree model, derived through a decision tree learning algorithm, which allows discrimination amongst three B_d ranges. The second tool was a linear equation model, derived through a linear regression algorithm, which predicts the numerical value of soil B_d. These tools were validated on a dataset of 471 soil horizons, belonging to 201 soil profile pits surveyed in Ireland. Overall, the decision tree model showed an accuracy of ~60%, while the linear equation model has a correlation coefficient of about 0.65 compared to the measured B_d values. For both models, the most relevant property affecting soil structural quality appears to be the humic characteristics of the soil, followed by soil porosity and pedogenic formation. The two tools are parsimonious and can be used by soil surveyors and analysts who need to have an approximate in-situ estimate of the structural quality for various soil functional applications.

Keywords: soil bulk density, soil structure, soil quality, machine learning

1. Introduction

The importance of soil structure in relation to soil quality is well known (Mueller et al., 2009; Karlen, 2004; Kay et al., 2006). A commonly used soil physical measurement to characterize soil structural quality is soil bulk density (B_d) (Armindo and Wendroth, 2016; Dam et al., 2005; Håkansson and Lipiec, 2000; Logsdon and Karlen, 2004; Moncada et al., 2015), which is defined as the oven-dry mass per unit volume of soil (IUSS Working Group, 2006; Mueller et al., 2009). Measurement of soil B_d is useful as it describes both the packing structure of the soil and its permeability (Dexter, 1988), whereby drainage characteristics can be inferred (Reidy et al., 2016). B_d measurement is often used in agronomic studies as it indicates the presence of compacted layers resulting from machinery or animal traffic (Reidy et al., 2016; Saffih-Hdadi, 2009), which may affect crop production. It is commonly considered an efficient measurement of soil carbon and nutrient stocks (Ellert and Bettany, 1995; Reidy et al., 2016).

However, the process of measuring B_d is often time consuming and open to human bias in the field and requires accurate laboratory analyses using trained personnel. Furthermore, soil texture has an important influence on the assessment of B_d e.g. in soils with high clay or sand content, or very humic soils, it may be difficult to obtain a representative sample and large variability between replicate samples can represent a problem. Also, in some soils the presence of stones can make sampling almost unmanageable. For such reasons, or constraints of budget or laboratory facilities, B_d measurements are commonly missing from soil databases (Reidy et al., 2016).

The main methods employed for the prediction of B_d are pedotransfer functions (PTF) methods, based on measurable soil attributes, such as organic carbon (OC) and clay content (Kaur et al., 2002; Leonavičiutė, 2000; Reidy et al., 2016). However, many of these methods ignore horizonation and depth variances for soil B_d prediction (Reidy et al., 2016). Furthermore, the nature of these methods, based on chemical/physical or landscape parameters, do not capture the intrinsic nature of the soil structural properties.

Our experience with respect to soil descriptions and classification has shown that the visual observations collected in the field at horizon level are often very important for the evaluation of soil quality (Fenton et al., 2017; 2015) and they become essential during the interpretation of the trend of some analytical parameters used as indicators of soil structure status e.g. B_d.

Soil structural quality has been assessed visually for millennia (Batey, 2000) e.g. soil survey manuals used in the field such as the Soils Survey Division Staff Manual (1993) or the WRB for soil resources (FAO, ISRIC and ISSS, 1998) include soil structure visual observations. However, soil scientists, for a long time, have presented repeatable procedures for the examination of soil structural form, stability and resilience (see latest review by Emmet-Booth et al., 2016 with examples from 1940's to present; Ball and Munkholm, 2015).

Taking this into account, in the present work we investigated whether, and to what extent those visual observations, called descriptors, can be used to predict soil B_d, which is considered one of the most efficient indicators in the assessment of soil structure quality (Moncada et al., 2015).

In order to achieve this objective, machine learning techniques were used. The potential of machine learning techniques have been rediscovered in the last few years through various applications in environmental sciences.

Worldwide, decision tree approaches have been used for different purposes: identifying sources of soil pollution (Xue et al., 2015); describing the extension of different forms of soil erosion in Mexico (Geissen et al., 2007); predicting chemical soil properties at national level in Australia (Henderson et al., 2005); classifying the surface soil freeze/thaw status in China (Jin et al., 2009) and even studying of soil structure through the prediction of soil hydraulic properties (Pachepsky and Rawls, 2003). However, limited literature has been found on the use of these powerful tools for environmental science in Europe.

The decision tree model output applied in this paper is based on a series of rules generated by the software, which can be visualised as paths starting from the root of the decision tree and ending at one of the leaves (Bhargava et al., 2013; Xue et al., 2013; Xue et al., 2015). Each of those paths corresponds to one or more soil *descriptors*, which are related to an internal *node* (Henderson el al., 2005). The model is able to examine all possible descriptors and then to select the most decisive *splitting attribute* (Xue et al., 2013). This operation occurs several times until all the instances are correctly classified in a set of *rules*. Each descriptor included in the model corresponds to a more defined *level* of classification.

The linear regression model applied in this paper is a classical statistical technique used to predict numerical data. It is based on the modelling of the relationship between a scalar dependent variable and one or more explanatory variables.

With our work we want to:

(i) Provide an operational strategy to estimate a range of B_d values, based on the visual soil parameters by means of a decision tree approach. This model can be used as a field tool to predict a general class of B_d (Low, Medium and High). It is an instrument able to discriminate between macro classes and has to be considered as a descriptive tool for *qualitative* estimation.

(ii) Propose an algorithm that can predict a numerical estimate of B_d. This second model should discriminate better between smaller increments. This instrument has to be considered as a more refined tool for *quantitative* estimation.

2. Methods

2.1 Primary data source and descriptors

Two pedological surveys, where full soil profile descriptions and supporting laboratory analyses, were carried out in Ireland with the aim of defining a coherent and homogeneous way to study soil formation, functions and quality:

1. The Irish Soil Information System (Irish SIS) project was established in 2008. It aimed to conduct a programme of structured research into the national distribution of soil types and construct a soil map, at 1:250,000 scale, able to identify and describe the soils according to a harmonised national legend. Irish SIS included more than 225 sites distributed around Ireland (Creamer et al., 2014).

2. The Soil Quality and Research project (SQUARE) started in 2013. The aim was to establish a baseline of soil quality in Ireland. The SQUARE soil survey included 38 grassland sites distributed within the five major agro climatic regions of Ireland defined by Holden and Bereton (2004) and classified into two drainage classes on the basis of the Irish Soil classification System.

During both (1) and (2) profile pits approximately 1 m deep, were observed and described by different operators. For the present study data from 201 profiles (168 Irish SIS, 33 SQUARE) was extracted from the larger database to cover a wide variety of Irish soil types with a specific focus on mineral soils. This data represents 471 horizons (http://gis.teagasc.ie/soils/map.php).

Although different surveyors worked across the projects mentioned, a systematic procedure was applied to describe the nature of the soil profiles, which included each of the soil horizons. Training was given to field operators. Using knowledge of soil structure and quality, the operators followed a widely understood schema of observation (developed by FAO through the Guidelines for Soil Description in 2006) which was able to investigate and finally characterize soil structure through visual parameters (FAO, 2006; FAO, ISRIC and ISS, 1998). Herein we have selected eleven descriptors presented in Table 1 (justifications are provided in Table 1), which may be considered the most important for the qualitative judgment of soil structure. Each descriptor was described and recorded on the basis of a set of pre-defined categories, reported in Table 1 in the Supplementary Material.

2.2 Soil analysis

The procedure to determine B_d of intact cores is a version of the ISO 11272:1998 – Soil Quality Part 5: Physical methods Sect. 5.6 – Determination of dry bulk density. The primary difference between the ISO and the applied methodology is that the ISO does not account for stone mass and volume in its core method, whereas the methodology applied in this study includes the following equation to calculate B_d (stone free):

$$B_d\ (g\ cm^{-3}) = (Md - Ms)/(V - Vs) \qquad \text{Eq 1}$$

where; Md: oven dry soil material weight (g), Ms: oven dry stone weight (g), V: volume of soil core (cm^{-3}), Vs: volume of stones (mL). Soil B_d values reported in this paper correspond to the mean of the three values obtained for each horizon sampled.

2.3 Model frameworks

Two models were built by means of the modelling tool WEKA (Waikato Environment for Knowledge Analysis). WEKA 3.8 is open source software for machine learning and data mining under the General public license developed at the University of Waikato in New Zealand (http://www.cs.waikato.ac.nz/ml/weka, Bhargava et al., 2013). This software includes different implementations of several machine learning algorithms. In our context, we used two specific algorithms that are made available by the tool, namely:

- the j48 algorithm, which corresponds to the WEKA's implementation of the C4.5 decision tree learner (Quinlan, 1993; Xue et al., 2015) which was used to build Model (1);
- a linear regression algorithm, used to build the Model (2). The M5 Method was used as attribute selection method for the linear model presented.

Two models were produced to achieve the objectives of our work:

- Model (1) is based on a classic decision tree model, developed to be used in the field in order to predict a B_d class using only visual descriptors as in Table 1.
- Model (2) is a linear regression model that uses the same in-field descriptors as above, and it is able to predict a numerical value of B_d with a relatively small error.

The proposed two models are both descriptive and predictive, but the decision tree is better at exploring in a descriptive way the relationship between B_d and visual parameters, as it allows further analysis of the soil pertaining to that soils own chemical and physical characteristics. On the other hand, the linear equation algorithm is stronger as a predictive tool and it offers a more precise estimate of B_d.

2.3.1 Data treatment

The entire database consists in 201 sampling points (profile pits) for a total of 471 horizons. For each horizon eleven descriptors and B_d data were used to train Model (1) and Model (2). The treatment of data can be summarized as follows:

- Data cleaning: to produce a full dataset, time was invested to ensure the data homogeneity between Irish SIS (649 horizons) and SQUARE datasets (125 horizons). In particular, descriptor rating options were double checked to reaffirm consistency across projects. To achieve uniformity within the dataset some data conversions were necessary. The final dataset consisted of 471 horizons i.e. 346 from Irish SIS and 125 from SQUARE.
- Missing values imputation: an initial analysis of the dataset highlighted the presence of some missing values for part of the considered descriptors, namely: “Fissure size”, “Void size”, “Void abundance” and “Soil consistency”. To avoid further reductions of the dataset, an IMRI imputation (performed by the WEKA 3.8 software described above) was selected as the means to predict missing values (Templ et al., 2011).

2.3.2 Model 1: Decision tree model, validation and outputs

Model (1) has been designed to predict classes of B_d data through the combination of the 11 visual descriptors outlined in Table 1. The model has been trained using B_d data at horizon level. The predicted classes of B_d are:

(i) Low B_d class: < 1.0 g cm^{-3} (n=137 cases)

(ii) Medium B_d class: between 1.0 and 1.4 g cm^{-3}(n=178 cases)

(iii) High B_d class: > 1.4 g cm^{-3} (n=156 cases)

Class ranges were selected on the basis of their homogeneity in terms of class population. The B_d measured in the majority of mineral soils under agricultural management in Ireland occur typically within the 1.0 and 1.4 g cm^{-3} range. Values of < 1.0 g cm^{-3} are usually related to Ombrotrophic or Mineratrophic Peat Soils (which correlates with the Histosol reference soil group of the WRB (IUSS Working Group WRB, 2006) or mineral soils having a Histic horizon (Reidy et al., 2016). Therefore 1.0 g cm^{-3} was selected as the lower B_d threshold e.g. herein 16 cases out of 137 belonged to the Low B_d class as Oh, Op, Of or Omf. The higher threshold (1.4 g cm^{-3}) was empirically chosen to best fit these data. In particular, multiple decision tree models were trained by varying the threshold between 1.1 g cm^{-3} and 1.8 g cm^{-3}, in 0.1 intervals. The model trained with the 1.4 g cm^{-3} threshold outperformed, other model runs in terms of accuracy.

The decision tree produced herein can be easily converted into classification rules. Each path in the tree that goes from the root to one of the leaves defines one classification rule. In our case each rule categorises the data in B_d Low, Medium and High classes. The knowledge represented in a decision tree can be extracted and represented in the form of the classification rule IF-THEN as follows:

If {condition A} AND {condition B} AND {condition C} AND {…} **then** categorization.

In our case:

If

the horizon is described as HUMOSE

then

the horizon fits into the bulk density category “Low”= <1 g cm^{-3}

A pruning technique is automatically performed by the WEKA software. This allows the identification and the removal of the outliers reducing the risk of overfitting to the training data (Bhargava et al., 2013). When decision trees are built, many of the ramifications can represent noise or outliers in the training data. The pruning process tries to identify and remove these branches with the aim of improving the accuracy of classification of future data. The next step was to prune the dataset to identify and remove branches which do not improve prediction with the aim of improving the accuracy of classification of future data.

A10-fold cross validation method was adopted, which randomly partitions the dataset into 10 parts and is used to validate the model. Then nine parts of the dataset were used to train the model, with the last part used for model testing (see Xue et al., 2015 for a similar approach).

Two measures were applied to evaluate the model performance; precision and recall values were calculated. Precision indicates how many of the instances were classified within a certain class that actually belong to that class. Whereas recall indicates how many of the instances that belong to a certain class are correctly classified by the model. To explain these measures further, it is useful to introduce the concepts of true positive, false positive and false negative.

Given a class C, we define true positives *tp* as the number of instances labelled as C in the original dataset, and classified as C by the decision tree; false positives *fp* as the number of instances not labelled as C in the original dataset, and incorrectly classified as C by the decision tree; false negatives *fn* as the number of instances labelled as C in the original dataset, and not classified as C by the decision tree. Given these definitions, precision p_C and recall r_C for the class C are defined as follows:

$$p_C = \frac{tp}{tp + fp}; \qquad \text{Eq 2}$$

$$r_C = \frac{tp}{tp + fn} \qquad \text{Eq 3}$$

Precision is negatively influenced by the number of false positive cases. Whereas, recall is negatively influenced by the number of false negative cases. High scores for precision and recall show that the classifier is returning accurate results (high precision, related to low false positive rates), as well as returning a majority of all positive results (high recall, related to a low false negative rates).

As no baseline algorithms were available, or proposed in other publications, against which to evaluate the performance of our Model (1), we resort to comparing it with a random predictor baseline, i.e. a fictional algorithm that randomly predicts the class of an instance (Alvaretz, 2002).

Let *n* be the total number of instances, and let *c* be the number of instances of class C, the precision and recall of a random predictor for the class C is given by:

$$p_C = r_C = c\,/\,n \qquad \text{Eq 4}$$

Besides precision and recall for each class, we also evaluated the average value of these measures. In addition, we evaluated the harmonic means of precision and recall, which is normally called F-measure, and is defined as follows:

$$F\text{-}measure = 2 * \frac{p_c * r_c}{p_c + r_c} \qquad \text{Eq 5}$$

Finally, as an overall indicator of the ability of the decision tree to correctly classify the instances, we evaluated the overall accuracy, which is defined as follows:

$$Accuracy\ (\%) = \frac{tp + tn}{tp + tn + fp + fn} \qquad \text{Eq 6}$$

2.3.3 Model 2: Linear regression model, validation and outputs

Using data from the visual parameters (Table 1), a linear equation that predicted an exact B_d value was developed.

As for Model (1), Model (2) was learned using B_d data at horizon level and produced by means of the WEKA software, using a linear regression algorithm. For this experiment these data were converted into numerical binary data, since the linear regression algorithm takes numerical data as input. In particular, for each value taken by each descriptor, a binary variable was created that takes either 0 (False) or 1 (True) as values. The variable was 1 if the descriptor had the value associated to the variable, and 0 otherwise.

The model builds a linear equation based on a weighted combination of the possible values taken by the 11 descriptors. In particular, the linear model has the following form:

$$B_d\ (g\ cm^{-3}) = \sum_{i=0}^{n} C_i * V_i \qquad \text{Eq 7}$$

where *Ci* are the coefficients computed by the linear regression algorithm. *Vi* are the binary variables. The linear regression algorithm is designed to select only those variables that have an influence on the final B_d value. Hence, the final model will not include all the possible variables, but only a subset. A 10-fold cross validation was performed to validate the model as described in par 2.3.2.

For Model (2) we could not evaluate the performance through precision and recall, since numerical values are involved instead of categorical ones, but we evaluated it in terms of correlation coefficient, root mean squared error, and mean absolute error:

The Root Mean Squared Error (RMSE) gives an estimation of the standard deviation of the error (Henderson et al., 2005). The lower is RMSE the higher is the predictive ability. Where *n* is the size of the dataset and $\hat{y}_t$ is the predicted value, the formula is defined as follows:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)}^{\,2} \qquad \text{Eq 8}$$

The Mean Absolute Error (MAE) is a quantity used to measure how close predictions are to the eventual outcomes. The mean absolute error is also known as the mean absolute deviation (Henderson et al., 2005). The lower the MAE value the higher is the predictive ability.

$$MAE=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t| \qquad \text{Eq 9}$$

3. Results and discussion

3.1 Model 1: Decision tree approach to assess B_d classes

3.1.1 Model performances

The number of rules generated by the decision tree algorithm was 41 in total; 11, 13 and 17 for Low, Medium and High respectively (all the rules are reported in Table 2 in the Supplementary Material). The overall tree is reported in ~~was unpacked into three sub-trees, one for each B_d class, to increase output readability (~~Figure 1~~s 1-3)~~. The decision tree hierarchy consists of six levels (level 1 has the highest classification power) of depth as follows:

Level (1): Humose

Level (2): Structure type

Level (3): Macropores Size/Void Size/Void Abundance/Plasticity

Level (4): Structure Grade/Stickiness

Level (5): Fissures

Level (6): Structure size

The descriptor “soil consistency” was excluded by the tree hierarchy, showing no influence on the prediction of B_d ranges. To discuss these results, it is useful to associate the different levels and hence, the corresponding descriptors, to specific soil quality properties. Broadly the descriptors that remained in the analysis fell into 4 main groups in order of importance as follows:

(i) soil humic characteristics; explained by level (1);

(ii) pedogenic formation; explained by level (2);

(iii) soil porosity; explained by level (3);

(iv) soil cohesive properties; explained by level (3-4).

The majority of the 41 rules were classified within the first three hierarchical levels highlighting the importance of these soil structural descriptors in quantifying B_d class. In general the parallelism between soil aggregation mechanisms, soil intrinsic characteristics related to soil forming factors, and the arrangement of solid and voids with their capacity to retain and transmit fluids, resulted in the main factors being capable of explaining soil B_d.

The overall accuracy shows that the model correctly classified 71% of the training data (Table 2). However, after the 10-fold cross validation step the model was able to correctly predict 60% of the cases (Table 2). The confusion matrix shows that mis-categorisation does not occur from Low to High B_d classes and vice versa, but can occur from Low to Medium (46 cases) and from High to medium (52 cases) (Table 3). The cause of lower model performance can be clarified by looking at the branches of the decision tree that miss-classify the highest number of instances (for discussion see par. 3.2.1).

The decision tree model was evaluated using the random predictor method (see par. 2.3.2). Performances are the following:

Low: $p_L = r_L = 137/471 = 0.29$.

Medium: $p_M = r_M = 178/471 = 0.37$.

High: $p_H = r_H = 156/471 = 0.33$.

For the Low B_d class, the decision tree outperforms a random predictor by 40% in terms of precision and by 25% in terms of recall; for the Medium B_d class, by 16% in terms of precision and by 26% in terms of recall; for the High B_d class by 30% in terms of precision and by 29% in terms of recall. These figures indicate that, although the accuracy indicates poor performance for the Low B_d class, the prediction for this class in terms of precision is actually the best amongst all 3 classes. This is due to the lower number of Low B_d instances, which are harder to detect for a random classifier, and that our algorithm predicts with a substantially higher degree of precision. Overall, the classification produced by the model is considerably better than a random classification.

3.1.2 Rules analysis

Figure~~s~~ 1~~-3~~ shows the decision tree generated for each B_d class with the corresponding instances classified for each rule and the percentage of accuracy, which refers to the percentage of true positive cases found by the model for the rule discussed. Rules classifying less than two instances are not reported in these figures.

- **Level 1: “Humose”**

The **“Humose”** descriptor emerges as the most dominant discriminator. All the horizons that highlight the presence of the “humose” feature following the rule “Humose: *yes*”, fit in the category Low B_d. The model correctly classifies 85.9% of n=64 instances for this rule with relatively few false positive cases.

The humose feature refers to an estimation of the level of humification of the organic material, so it is indirectly related to the C content. In fact, soil OC content gives an indication of decomposition rates which have a direct effect on soil aggregation (Bronick and Lal, 2005; Schulten and Leinweber, 2000). The presence of humic substances, as well decomposed organic matter (OM), contributes to the stability of soil aggregates and pores through the bonding or adhesion properties of organic materials, such as bacterial waste products, organic gels, fungal hyphae and worm secretions and casts. Moreover, OM intimately mixed with mineral soil materials induces increased moisture holding capacity and air exchange with the atmosphere (Stevenson, 1994). This mechanism is reflected by lower values of B_d, resulting in enhancement of soil porosity producing a good soil structure.

When the humose feature is recorded as absent, the model switches to level 2 for further refinement.

- **Level 2: “Structure Type”: Granular**

“Structure Type” was the most discriminating descriptor at the second level of the tree hierarchy. Here the tree branches out to account for the effect of different types of structure.

For the rule “Humose: *no*, AND Structure type: *Granular*” the model classified 23 cases as belonging to the Low B_d class, with an accuracy of 56.5%. Despite the high number of false positives, the amplitude of the group is among the highest, showing its reliability.

Granular aggregates are spheroids or polyhedrons aggregates of soil, having curved or irregular surfaces (FAO, 2006), they are normally associated with highest air capacity, and are an indicator of good soil structure (Mueller et al., 2009). Such cases are mostly A horizons, not deeper than 30 cm. Root mass, density, distribution and turnover, can positively influence the soil particle aggregation by releasing a variety of compounds, (such as root exudates) and can contribute to increased soil porosity through mechanical action (Bronick and Lal, 2005; Caravaca et al., 2002). This is reflected in the B_d, resulting in lower values.

- **Level 2: “Structure Type”: Subangular Blocky → Level 3, “Macropores Size” → Level 4 ”Structure Grade”**

In presence of the “Subangular Blocky” structure type the tree branches split again, going deeper into the third hierarchical level of tree structure, showing the **“Macropores size”** descriptor as the next level of description. For the rules “Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *(A), (B), (C), (D)*”.:

(A) Coarse (C 5.0-20.0 mm): 50% accuracy when classifying Low B_d class.

(B) Fine (F 0.5-2.0 mm): for subangular blocky aggregates, 66.1% accuracy when classifying Medium B_d class. This is a rule that has a higher number of

instances (n=92), which balances the presence of a relatively high number of false positives. For angular blocky aggregates, 72.7% accuracy when classifying Medium B_d class (n=11).

(C) Very Fine (VF< 0.5 mm): 66.7% accuracy when classifying High B_d class.

(D) Medium (M 2.0-5.0 mm): 100% accuracy when classifying Low and Medium B_d class.

Subangular blocky structure type is described as cube like with a flat surface and rounded corners (FAO, 2006; Schoeneberger et al., 2012) and is considered an indicator of an intermediate degree of soils structure quality, between the granular/crumbly aggregates and the blocky/sharp angular ones. Results for this structure type show that the size of the macropores is a crucial variable to discriminate in terms of B_d classes, as B_d is very sensitive to both alteration of macroporosity abundance and size of pores (Mueller et al., 2009). In particular the higher the macropore size, the lower the B_d class predicted by the model.

For the rule *(A)*, horizons fitting the Low B_d class are mostly Ap horizons, corresponding to the upper soil layers (maximum depth of 22 cm). Earthworm activity is mainly concentrated in the top soil (Haynes and Naidu, 1998; Lee and Foster, 1991), which promotes macropores (C 5.0-20.0 mm) which in turn can alter soil porosity thereby affecting the movement of air, water and solutes (Shipitalo and Le Bayon, 2004). For the rule *(B)* and *(C)* medium and high B_d are associated with macropore sizes from 0.5 mm to 2.0 mm. The cases following these two rules are mainly classified as A or B horizons for the medium B_d class (average maximum depth of ~40 cm) and as B horizons for the higher B_d class (average maximum depth of ~60 cm). At such depths roots become thinner resulting in reduced porosity and higher B_d. Furthermore, B_d often increases ~~exponentially~~ in the deeper horizons due to ~~compaction resulting from~~ an increase in clay material or as result of management operations.

The medium class of macropores (M 2.0-5.0 mm) insufficiently defines the B_d range and must be assisted by another splitting attribute which is triggered at level four. The **"Structure grade"** is identified by the model as an important variable to discriminate between the Low and Medium B_d class for this macropores size. The rule (*D*) is consequently developed as follows:

- "Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *M 2.0-5.0 mm* AND Structure grade: *Moderate*"; which predict Low B_d class (100% correct instances).
- "Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *M 2.0-5.0 mm* AND Structure grade: *Weak*"; which predict Medium B_d class (100% correct instances).

FAO (2006) defines soils by describing the lack (apedal) or the presence (pedal) of a defined structure. A moderate structure grade, showing the presence of a nicely structured soil, was associated with a Low B_d class. Medium B_d classes are associated with a weak structure grade, synonymous of a lower structure quality. However, herein cases classified for these two rules were similar in terms of actual B_d values. In particular, cases classified as Low B_d had actual values very close to the upper limit of this category (values between 0.9 and 1.0 g cm^{-3}) and the cases classified as Medium B_d had actual values close to the lower limit of this category (values between 1.0 and 1.16 g cm^{-3}). Considering the narrow range of values for these cases across the Low and Medium B_d categories the model correctly discriminates between categories.

- **Level 2: "Structure Type": Angular Blocky → Level 3, "Void Size" → Level 4 "Stickiness"**

The "Angular Blocky" aggregates belong to the blocky category as for the subangular blocky aggregates, but differ as they have faces intersecting at relatively sharp angles (FAO, 2006; Schoeneberger et al., 2012). In the presence of this structure type i.e. a typical indicator of a poorly structured soil, the model returns **“Void size”** as the next, most important, descriptor. This descriptor indicates the total volume of pores discernible with a *10 hand-lens (FAO, 2006). It differs from macroporosity as it is a much wider term that includes soil fissures and plane (FAO, 2006). Macropores are mostly determined by plant roots through the twisting activity within and around aggregates, and by zoological exploration, through burrowing activity, so they are usually larger pores having a size higher than 75 µm (Russell, 1975). However, here macropores and voids are described in terms of size, so a high void size can correspond to a high macropores size. Therefore, it is important to highlight here that although different from a semantic point of view, the combination of these two descriptors was very effective in explaining soil porosity.

For the rule “Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *C 5.0-20.0 mm*” the model returned 75% of correct instances for the class Low B_d. Although the structure type indicated was often that of a very poor structured soil, characterised by large, angular and sharp aggregates, this effect was mitigated by the presence of a very high overall porosity, resulting in a Low B_d (Pagliai and Vignozzi, 2002).

The rule “Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *VF <0.5 mm*”, returned High B_d class with 76.9 % accuracy. Also this rule was quite wide in terms of population (n=26). In this case the poorly structured horizons were associated with the smallest size of pores. The cases are mainly classified as Bg, BCg, Cg or Eg horizons, with an average depth ranging from 42 to 72 cm. For such cases, gleying is caused by surface water which has been held in a poorly permeable horizon, (mainly belonging to surface water gley soil type (Stagnosols reference soil group of the WRB). These poorly permeable layers showed evidence of compaction due to one or more of the following characteristics: (i) presence of a pan, resulting in severe compaction imposed by management, (ii) poor structural development, (iii) very heavy textured horizons dominated by silt and clay and (iv) natural impeded horizons due to soil formation and profile development mechanisms.

Furthermore, the presence of some argillic horizons classified as Btg following this rule, as well as having gleyic features, showed the presence of clay-sized material in coatings or as intrapedal concentrations (FAO, 2006). Clay content physically affects particle aggregation through swelling and dispersion, resulting in contiguous peds which fit together eliminating space (Attou et al., 1998; Bronick et Lal, 2005; Kay, 1998; Russel, 1975).

When the void size is classified as belonging to the middle category (Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *F 0.5-2.0 mm*), **"Stickiness"** was selected by the model as the next most informative descriptor, associating Medium B_d class with the "*Non Sticky OR Slightly Sticky*" feature (50% and 63.6% accuracy, respectively), and the High B_d class with the "*Sticky*" feature (accuracy of 75%). The rules' population size was n=44, n=22 and n=12, respectively and therefore must be considered an important branch of the overall tree. As the stickiness property is considered an indirect indicator of the clay content in soil, the results confirmed that the presence of clay material was one of the characteristics which makes the soil prone to compaction, or at least significantly decreases the level of porosity. Even in the presence of a relatively higher porosity, clay content was crucial to force a split in B_d classes, thereby attributing higher B_d values to heavier textured soils.

"Plasticity" appears at the third depth level of the decision tree for the structure type Angular Blocky to Granular. Like stickiness, plasticity is directly related to the clay content (FAO, 2006). Horizons classified as "*Plastic* OR *Slightly plastic* AND Macropores size: *Very Fine (VF< 0.5 mm)*", fit into the

High B_d class, highlighting the link between high values of B_d and clay content (accuracy respectively of 100% and 66.7%).

- **Level 2: "Structure Type": Prismatic → Level 3, "Void Abundance" → Level 4 ”Structure Grade" → Level 5: "Fissures"→ Level 6: "Structure Size"**

The "Prismatic" structure type is described by FAO, (2006) and by Schoeneberger et al. (2012) as having vertical elongated units with limited faces in the horizontal plane. Horizons having prismatic to angular blocky aggregates are classified by the model within the High B_d class (accuracy, 62.5%). The average depth ranges within 53-95 cm, classified across a number of soil types e.g. typical surface water gleys (Stagnosols reference soil group of the WRB), typical luvisols (Luvisols), typical brown podzolic (Podzols) and typical calcareous brown earths (Calcaric Cambisols). Different soil types are subjected to different mechanisms of soil particle aggregation which can exert a degree of compaction in a specific horizon, due to their intrinsic nature or to a combination of factors, such as: (i) clay concentrations in the impeded horizon, as for luvisols; (ii) translocation of Al and Fe in the spodic horizon, as for brown podzolic (Bronick and Lal, 2005; Collins, 2004), (iii) presence of carbonates, as for the calcareous brown earths (Boix-Fayos, et al., 2001) and (iv) presence of a poorly permeable horizon, as for the gleyic horizons in surface water gleys (Collins et al., 2004).

Furthermore, for the prismatic structure type, **"Voids abundance"** is selected as the main descriptor to split the data between High and Medium B_d for this structure type.

For the rules "Humose: *no*, AND Structure type: *Prismatic* AND Void abundance: *A, B, C*:

(A) High (15-40%): the model returns 66.7% of correctly classified instances as belonging to the Medium B_d class,

(B) Very Low (<2%): the model classifies for the High B_d class with an accuracy of 75%.

(C) Medium (5-15%): the model classifies for Medium and High B_d with an accuracy of 83.3 and 89%, respectively.

The Low B_d category was not defined by the model for this structure type, indicating the prismatic arrangement as having overall higher B_d values. The abundance of voids is important, as the void size for this type of structure, poorer than the angular blocky structure type. Rules *(A)* and *(B)* associated higher void abundance with higher structure quality. On the other hand the rule *(C)*, which took into account the medium void abundance category, requires further descriptors to categorise within the High and the Medium B_d classes, such as **"Structure grade"**, **"Fissures"** and **"Structure size"**.

The presence of smaller sized prismatic aggregates (Structure size: *Fine 10-20 mm*) directs the model to predict High B_d class with 100% correct instances. As found in the present study, the rapid wetting of dry soil which comes in contact with free water can cause micro-cracking. This increases ped friability, causing the production of smaller sized aggregates, which does not always result in lower values of B_d (Dexter, 2002).

Vertical fissures are also an important feature of this structure type. Although hard clods are devoid of microstructure, fissures enable the percolation of the surface water in the deeper layers (Russel et al., 1975). Such a drainage advantage is not valid for the "Platy" structure type. In this case the model returns two different outputs. Platy aggregates, probably formed by intensive mechanical intervention, indicate compaction which affects water percolation, resulting in high B_d values (accuracy, 66.7%). On the other hand, some platy aggregates were attributed to the histic horizons (Of, Omf) which showed no developed structure, associated with very low B_d (accuracy 75%).

- **Level 2: “Structure Type”: Massive**

Structure types defined as “Massive” and “Single grain” were categorised by Schoeneberger et al. (2012) as structure less soils. Massive soil material normally has stronger consistence as the soil particles are arranged in a coherent mass and are very difficult to break (FAO, 2006). Horizons with “Massive”, “Massive to Angular blocky” and “Massive to Single Grain” structure types were classified by the model as belonging to the High B_d class with an accuracy of 77.8%, (n=63; very influential rule in the overall tree output), 83.4% and 100%, respectively. Further analysis of the data showed that most of the horizons belonging to the rule: “Humose: *no*, AND Structure type: *Massive*” were described as having a hard or very hard consistence dry, fine or very fine macropores and very low void abundance. The decrease in soil macroporosity and the firm nature of the aggregates suggests a high level of compaction (Epron et al., 2016), resulting in high B_d values which ranged from 1.4 g cm^{-3} to 1.9 g cm^{-3}. In the case of the Massive to single grain structure type, **“Plasticity”** is a key attribute to assess the range of B_d (rule: “Humose: *no*, AND Structure type: *Massive to Single grain* AND Plasticity: *Non plastic*”).

The arrangement of the aggregates which are weak and tend to disintegrate when sampling in the field, suggested the presence of sandy material held together in big, hard and massive aggregates. In some soils, sand grains have a film of orientated clay particles on the surface, not enough to be detected by feel, but that are able to strongly hold the sand particles packing them together in massive aggregates (Russel, 1975). Sandy soils are prone to compaction of surface layers, due to intensive agricultural operations (Ampoorter et al., 2007; Deconchat, 2001; Teepe et al., 2004).

Low B_d class was attributed by the model to the horizons responding to the rule “Humose: *no*, AND Structure type: *Single grain*”, which is a feature typical of sandy soils not subjected to compaction phenomena, thereby conserving a large amount of wide pores (Ampoorter et al., 2007). The model correctly classified 66.7% of the instances following this rule.

3.2 Model 2: Linear regression approach to predict a numerical estimate of B_d

3.2.1 Model and performance

After several trials with different machine learning methods for numerical prediction, a linear regression model was selected based on better overall performance. This conclusion asserts that a strong linear relationship exists between the visual descriptors and B_d values. As a result, since the model produced by a linear regression algorithm is a linear equation, the predicted value of B_d can be easily computed in seconds without the need for time consuming and costly laboratory analyses.

Model (2) considers only those descriptors that influence the final B_d value. For this quantitative estimation seven descriptors out of eleven were selected by the linear regression algorithm:

(i) Humose
(ii) Structure Grade
(iii) Structure Type
(iv) Structure Size
(v) Macropores
(vi) Void Size
(vii) Void Abundance

The linear equation produced by Model 2 to predict B_d is as follows:

B_d (g cm $^{-3}$) =
0.4575 * *(NO **Humose**: =true)* +
-0.0517 * *(MODERATE **Structure Grade**=true)* +
-0.2128 * *(GRANULAR **Structure Type**=true)* +
0.093 * *(MASSIVE **Structure Type**=true)* +
-0.134 * *(SUBANGULAR BLOCKY **Structure Type**=true)* +
-0.1501 * *(SUBANGULAR BLOCKY TO GRANULAR **Structure Type**=true)* +
-0.1003 * *(ANGULAR BLOCKY TO GRANULAR **Structure Type**=true)* +
-1.1619 * *(VERY COARSE > 10mm **Structure Size**=true)* +
0.1471 * *(VERY FINE < 0.5mm **Macropores**=true)* +
-0.1791 * *(COARSE 5-20 mm **Void Size**=true)* +
0.0802 * *(VERY LOW **Void Abundance**=true)* +
-0.0784 * *(HIGH **Void Abundance**=true)* +
0.8866 Eq 10

The equation produced follows a simple framework:
(i) All the descriptors associated with a *positive coefficient* caused a significant incremental increase in B_d. Basically, as the associated coefficients are positive, a soil classified having these descriptors (*Vi=1=true*, see description in paragraph 2.3.3), will result in an increased B_d final value.

The variables *"NOHumose=true"*, *"VERY FINE<0.5mm Macropores=true", "VERY LOW Void Abundance=true"* and *"MASSIVE Structure Type=true",* all cause an increase in B_d, with 0.4575; 0.1471; 0.0802; and 0.093 explanatory power, respectively.

In line with Model (1) results, the humification degree of OM has the greatest influence on B_d prediction. This characteristic, as well as readily defining B_d class, also informed small differences in soil B_d and generally indicated pedological features that were consistent with lower B_d values.

Following the humic feature, macropores size was the second more discriminating feature, with the lower size able to distinguish a wide increment of prediction (multiplier of 0.1471). Furthermore, very low void abundance triggered the model as the fourth main attribute (multiplier of 0.0802). This was an expected result, as porosity was already investigated by the decision tree approach and was one of the most critical soil characteristics which took into account an evaluation of soil structure. In particular, size of pores was highlighted as a stronger predictor for an increase of B_d with respect to pore abundance. As seen from the decision tree output, the massive structure type has a role in increasing soil B_d. In general both models although operating at different scales produce the same descriptors for the prediction of B_d.

(ii) All the descriptors associated with a *negative coefficient* have a role decreasing B_d. Basically, as the associated coefficients are negative, a soil classified as having these descriptors, resulted in a decreased B_d, and therefore have a better soil physical quality.

In Model (2), structure size appears to be a stronger descriptor which influences a decrease in B_d when evaluated as *"VERY COARSE>10 mm Structure Size=true"* (multiplier of -1.1619). This was surprising considering that in Model (1) the size of aggregates was not particularly important (sixth level) in the tree hierarchy as a splitting attribute. This highlights that this feature is better at identifying small incremental reductions of B_d, but is less informative when splitting into wider B_d ranges.

However, in Model (2) Structure type still has to be considered one of the most informative variables, since it appears in four out of eight factors having a negative coefficient in the equation. In particular, looking at the equation, we have:

- *"GRANULAR Structure Type=true"* ,
- *"SUBANGULAR BLOCKY TO GRANULAR Structure Type=true"*,
- *"SUBANGULAR BLOCKY Structure Type=true"*,
- *"ANGULAR BLOCKY TO GRANULAR Structure Type=true"*,

as coefficients -0.2128; -0.1501; -0.134; -0.1003, respectively.
Granular structure type is responsible of a higher decrease of B_d, confirming what was found for Model (1), indicating good soil structure.

The model showed sufficient sensitivity allowing the identification of differences related to soil structure type. This is despite the relatively crude measurement of soil structure at field level in tandem with other attributes. In particular while the soil structure quality, associated with a change of structure type, gradually decreases, with diminishing negative effect on B_d, indicated by the coefficients for individual structure types. Hence, corresponding to their respective coefficients, a granular structure type will have a higher negative increment, resulting in a reduction of the B_d final value, while an angular blocky to granular structure type will have a lower negative reduction in the final predicted value, resulting in a higher B_d final value.

Structure size, Void size, Void abundance and Structure grade were, in order of importance, the next most informative features for the linear regression model.
If both the features relating to porosity appear in the Model (1) at a high level in the tree hierarchy (level 3), for this model only void size with the variable *"COARSE 5-20 mm* ***Void Size****=true"* appears to have higher negative impact, with a relatively high coefficient of -0.1791, while void abundance (*HIGH* ***Void***

Abundance*=true*) resulted in having less negative impact on the definition of a final B_d value, being associated with a smaller coefficient, -0.0784. Probably, as per the decision tree model, the shape of aggregates is again a critical point which drives the selection of the second decisive node into pore abundance or size, depending on the original structure type.

3.2.2 Overall model evaluation

The overall correlation coefficient on training set for the linear regression model is 0.71 with Root Mean Squared Error (RMSE): 0.25 and Mean Absolute Error (MAE): 0.20; (Table 4). After a 10-fold cross validation the correlation coefficient slightly dropped, to 0.65, with similar error ranges (RMSE: 0.27 and MAE: 0.21); (Table 4). The errors reported for this model may be considered quite high in relation to a standard lab-based B_d measure. However, it is important to highlight that the model has been fed using only soil visual parameters. Considering the nature of these data inputs and the influence that different operators can have during the classification phase, this range of error is low.

Figure ~~4~~2 shows the prediction performances of Model (2). The distribution of predicted values results more coherent with the real values for a middle range of B_d values. In particular we identified a range that goes between 0.8 to 1.6 g cm^{-3}, which falls within the typical range of B_d found in Irish grassland soils, where the model returns B_d values close to real values. In these cases the model is more robust, predicting a numerical estimate with a quite low standard error. Furthermore, neither overestimation nor underestimation prevails for a middle range of B_d.

The model shows the higher errors for the extreme B_d classes, namely (i) very low B_d, that we identified as values lower than 0.8 g cm^{-3}, or (ii) very high B_d values, identified as values higher than 1.6 g cm^{-3}. The algorithm appears to a have higher prediction power on medium B_d values~~, which fall within the~~

~~typical range of B_d found in Irish grassland soils~~~~, and hence, also~~ which also receive higher representation in our dataset. In general, machine learning algorithms are improved where greater input data is provided. Therefore, in our case, the algorithm is inherently biased towards the correct prediction of medium ranges.

3.3 Models choice considerations

We have chosen to use decision trees and linear regression because these simple types of models allow users to identify the B_d class (for decision trees), and the B_d value (for linear regression) without the need to rely on additional software. Furthermore, besides the predictive ability, decision trees also provide descriptive power, in that they make explicit the relationships among different characteristics of the soils and allow the user to have greater insight to these relationships. Other algorithms that were considered, i.e., support vector machines and multi-layer perceptron, although leading to similar performance, do not allow this type of insight. A full comparison between different algorithms, from the performance point of view, can be conducted in future. While the quality of the interrogated database was good, we believe that further improvement of model performance can be achieved by increasing the extent of the sample dataset, especially for horizons with low and high B_d values, which are less represented in our data. Finally, the utility of these models to assess critical thresholds for compaction should be evaluated for descriptive soil datasets where attributes such as "Compact degree" (FAO, 2006) are included.

4. Conclusion

A decision tree and linear equation model were developed to predict soil bulk density on the basis of visual descriptors. The visual soil descriptors identified, as being more informative by both models, are associated with specific soil properties. This allows the user to rank to these properties in terms of their impact on soil structural quality. For both models the most relevant properties that affect B_d appears to be soil humic characteristics, followed by soil porosity and pedogenic formation.

Overall, the decision tree model shows an accuracy of about 60%, while the linear equation model had a correlation coefficient of about 0.65 with respect to the measured B_d values. The two models are parsimonious and can be used by soil surveyors and analysts who need to have a quick and approximate *in-situ* estimate of the structural quality for various soil functional applications. Furthermore they have an enormous potential to retrofit B_d data (i.e. gap fill) to existing data sets were laboratory data are missing. Future work is required to refine these models for use on soils with very low and very high B_d classes which fall outside those typically found in Ireland. Finally, our goal is to encode the decision tree and the linear equation into a mobile application, in order to enable multiple user types to perform B_d prediction more quickly, on site, and in a user friendly manner.

Acknowledgements

The authors thank Irene Marongiu for her help with the physical and chemical soil analyses. SQUARE project has received founding from the Irish Department of Agriculture (DAFM) Research Stimulus Fund (RSF). Task 3 output.

References

Alvarez, S.A., 2002. An exact analytical relation among recall, precision, and classification accuracy in information retrieval. Tech. Rep. BCCS-02-01, Computer Science Department, Boston College.

Ampoorter, E., Goris, R., Cornelis, W. M., Verheyen, K., 2007. Impact of mechanized logging on compaction status of sandy forest soils. Forest Ecol. Manag. 241, 162–174.

Armindo, R.A., Wendroth, O. 2016. Physical soil structure evaluation based on hydraulic energy functions. Soil Sci. Am. J. 80, 1167–1180.

Attou, F., Bruand, A., Bissonnais, Y. L., 1998. Effect of clay content and silt-clay fabric on stability of artificial aggregates. Eur. J. Soil Sci. 49, 569–577.

Ball, B. C., Munkholm, L. J., 2015. Visual Soil Evaluation: Realizing Potential Crop Production with Minimum Environmental Impact. CAB International. London.

Batey, T., 2000. Soil profile description and evaluation. In: Smith, K., Mullins, C. (Eds.), Soil and environmental analysis, physical methods, 2nd edition. Marcel Dekker, New York. pp. 595–628

Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering. 3, 1114–1119.

Boix-Fayos, C., Calvo-Cases, A., Imeson, A.C., 2001. Influence of soil properties on the aggregation of some Mediterranean soils and the use of aggregate size and stability as land degradation indicators. Catena 44, 47–67.

Bronick, C. J., Lal, R. 2005. Soil structure and management: a review. Geoderma. 124, 3–22.

Caravaca, F., Hernandez, T., Garcia, C., Roldan, A., 2002. Improvement of rhizosphere aggregate stability of afforested semiarid plant species subjected to mycorrhizal inoculation and compost addition. Geoderma. 108, 133–144.

Collins, J. F., Larney, F. J., Morgan, M. A., 2004. Climate and soil management, in: Keane, T., Collins, J. F. (Eds.), Climate, Weather and Irish Agriculture. Working group on Applied Agricultural Meteorology (AGMET), Dublin, pp. 119–160.

Creamer, R., Simo, I., Reidy, Carvalho, J., Fealy, R., Hallett, S., Jones, R., Holden, A., Holden, N., Hannam, J., Massey, P., Mayr, T., McDonald, E., O'Rourke, S., Sills, P., Truckell, I., Zawadzka, J., Schulte, R., 2014. Irish Soil Information System. (2007-S-CD-1-S1) EPA STRIVE Programme 2007–2013, Synthesis Report.

Dam, R.F., Mehdi, B.B., Burgess, M.S.E., Madramootoo, C.A., Mehuys, G.R., Callum, I.R., 2005. Soil bulk density and crop yield under eleven consecutive years of corn with different tillage and residue practices in a sandy loam soil in central Canada. Soil Till. Res. 84, 41–53.

Deconchat, M., 2001. Effets des techniques d'exploitation forestie`res sur l'e´tatde surface du sol. Ann. For. Sci. 58, 653–661.

Dexter, A. R., 1988. Advances in characterization of soil structure, Soil Till. Res. 11, 199–238.

Dexter, A. R., 2002. Soil structure: the key to soil function, in: Pagliai, M., Jones, R., (Eds.), Sustainable land Management – Environmental Protection, a soil Physical Approach. Advances in Geoecology 35, Clearance Center, Inc., Reiskirchen, pp. 57–69.

Ellert, B. H., Bettany, J. R., 1995. Calculation of organic matter and nutrient stored in soils under constrain management regimes. Can. J. Soil Sci. 75, 529–538.

Emmet-Booth, J. P., Forristal, P. D., Fenton, O., Ball, B. C., Holden, N. M., 2016. A review of visual soil evaluation techniques for soil structure. Soil Use Manage. 32, 623–634.

Epron, D., Plain, C., Ndiaye, F. K., Bonnaud, P., Pasquier, C., Ranger, J., 2016. Effects of compaction by heavy machine traffic on soil fluxes of methane and carbon dioxide in a temperate broadleaved forest. Forest Ecol. Manag. 382, 1–9.

FAO, 2006. Guidelines for Soil Description, Fourth Edition. FAO, Rome.

FAO, ISRIC, ISSS, 1998. World Reference Base for Soil Resources. World Soil Resources Reports 80, FAO Rome.

Fenton, O., Vero, S., Ibrahim, T. G., Murphy, P. N. C., Sherriff, S. C., ÓHuallacháin, D., 2015. Consequences of using different soil texture determination methodologies for soil physical quality and unsaturated zone time lag estimates. J Contam. Hydrol. 182, 16–24.

Fenton, O., Vero, S., Schulte, R.P.O., O'Sullivan, L., Bondi, G., Creamer, R.E., 2017. Application of Dexter's soil physical quality index: an Irish case study Irish J. Agr. Food. 56, 45–53.

Geissen, V., Kampichler, C., López-de Llergo-Juárez, J. J., Galindo-Acántara, A., 2007. Superficial and subterranean soil erosion in Tabasco, tropical Mexico: development of a decision tree modeling approach. Geoderma. 139, 277–287.

Håkansson, I., Lipiec, J., 2000. A review of the usefulness of relative bulk densityvalues in studies of soil structure and compaction. Soil Till. Res. 53, 71–85.

Haynes, R.J., Naidu, R., 1998. Influence of lime, fertilizer and manure applications on soil organic matter content and soil physical conditions: a review. Nutr. Cycl. Agroecosyst. 51, 123–137.

Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma. 124, 383–398.

Holden, N., Brereton, A.J., 2004. Definition of agroclimatic regions in Ireland using hydro-thermal and crop yield data. Agr. Forest Meteorol. 122, 175–191.

IUSS Working Group WRB, 2006. World Reference Base for Soil Resources – a framework for international classification, correlation and communication. World Soil Resources Reports 103. FAO, Rome.

Jin, R., Li, X., Che, T., 2009. A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature. Remote Sens. Environ. 113, 2651–2660.

Karlen, D.L., 2004. Soil quality as an indicator of sustainable tillage practices. Soil Till. Res. 78, 129–130.

Kaur, R., Sanjeev, K., Gurung, H., 2002. Apedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Aust. J. Soil Res. 40, 847–857.

Kay, B.D., 1998. Soil structure and organic carbon: a review. In: Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.A. (Eds.), Soil Processes and the Carbon Cycle. CRC Press, Boca Raton, pp. 169-197.

Kay, B.D., Hajabbasi, M.A., Ying, J., Tollenaar, M., 2006. Optimum versus non-limiting water contents for root growth, biomass accumulation, gas exchange and the rate of development of maize (Zea mays L.). Soil Till. Res. 88, 42–54.

Lee, K. E., Foster, R. C., 1991. Soil fauna and soil structure. Aust. J. Soil Res. 29, 745–775.

Leonavičiutė, N., 2000. Predicting soil bulk and particle densities by pedotransfer functions from existing soil data in Lithuania, Geografijosmetraštis. 33, 317–330.

Logsdon, S.D., Karlen, D.L., 2004. Bulk density as a soil quality indicator during conversion to no-tillage. Soil Till. Res. 78, 143–149.

Moncada, M.P., Ball, B.C., Gabriels, D., Lobo, D. and Cornelis, W.M., 2015. Evaluation of soil physical quality index S for some tropical and temperate medium-textured soils. Soil Sci. Soc. America J. 79, 9–19

Mueller, L., Kay, B. D., Deen, B., Hu, C., Zhang, Y., Wolff, M., Eulenstein, F., Schindler, U., 2009. Visual assessment of soil structure: Part II. Implications of tillage, rotation and traffic on sites in Canada, China and Germany. Soil Till. Res. 103, 188–196.

Pachepsky, Y. A., Rawls, W. J., 2003. Soil structure and pedotransfer functions. Eur. J. Soil Sci. 54, 443–452.

Pagliai, M., Vignozzi, N., 2002. The soil pore system as an indicator of soil quality, in: Pagliai, M., Jones, R., (Eds.), Sustainable land Management – Environmental Protection, a soil Physical Approach. Advances in Geoecology 35, Clearance Center, Inc., Reiskirchen, pp. 71–82.

Quinlan, J. R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

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.

Russell, E. W., 1975. Soil Conditions and Plant Growth 10th edition. Ed. Longman, London.

Saffih-Hdadi, K., Défossez, P., Richard, G., Cui, Y. J., Tang, A. M., Chaplain, V., 2009. A method for predicting soil susceptibility to the compaction of surface layers as a function of water content and bulk density, Soil Till. Res. 105, 96–103.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.

Schulten, H.R., Leinweber, P., 2000. New insights into organic– mineral particles: composition, properties and models of molecular structure. Biol. Fertil. Soils. 30, 399– 432.

Shipitalo, M. J., Le Bayon, R. C., 2004. 10 Quantifying the Effects of Earthworms on Soil Aggregation and Porosity, in: Edwards, C.A. (Eds.) Earthworm ecology 2nd edition, Boca Raton, pp 183–199.

Soil Survey Division Staff, 1993. Soil Survey Manual. Agricultural Handbook N 18. USDA Natural Resources Conservation Service, Washington D.C.

Stevenson, F.J. 1994. Humus chemistry. genesis, composition, reactions. 2nd edition. Wiley Interscience, New York.

Teepe, R., Brumme, R., Beese, F., Ludig, B., 2004. Nitrous oxide emission and methane consumption following compaction of forest soils. Soil Sci. Soc. Am. J. 68, 605–611.

Templ, M., Kowarik, A., & Filzmoser, P., 2011. Iterative stepwise regression imputation using standard and robust methods. Comput. Stat. Data An. 55, 2793–2806.

Xue, D., De Baets, B., Van Cleemput, O., Hennessy, C., Berglund, M., Boeckx, P., 2013. Classification of nitrate polluting activities through clustering of isotope mixing model outputs. J. Environ. Qual. 42, 1486–1497.

Xue, D., Pang, F., Meng, F., Wang, Z., Wu, W., 2015. Decision-tree-model identification of nitrate pollution activities in groundwater: A combination of a dual isotope approach and chemical ions. J. Contam. Hydrol. 180, 25–33.

Table 1. Selection of soil structure field descriptors described by FAO, Guidelines for Soil Description, 2006.

Descriptor	Title	Description
1	Humose	This is an estimation of the degree of humification of the organic material. Surveyor must provide a positive or affirmative answer to being humose (this descriptor was recorded as a presence/absence in the database).
2	Soil Consistency	The strength with which soil materials are held together. It provides a means of describing the degree of cohesion and adhesion between the soil particles as related to the resistance of the soil to deform or rupture. It includes soil properties such as friability, plasticity, stickiness and resistance to compression. It changes with soil moisture and is highly related to the percentage of clay and OM in the soil.
3	Stickiness	It is the capacity of the soil to adhere to an object. It is evaluated pressing a small amount of wet soil between thumb and forefinger to see if it will stick to fingers.
4	Plasticity	The ability of soil material to retain a shape after pressure deformation. It is evaluated by rolling a small amount of wet soil between the hand palms until it forms a long, round strip like a wire about 3 mm thick.
Soil structure* is described as the combination of (5, 6, 7)		
5	Structure Grade	It describes the level of development of soil structure. It is expressed as the differential between cohesion within aggregates and adhesion between aggregates. It is evaluated in relation to the arrangement of the aggregates and to the strength necessary to break them.
6	Structure Type	It describes the form or shape of individual aggregates and is directly correlated with the pedogenic formation.
7	Structure Size	It describes the average size of individual aggregates. Different classes may be recognized in relation to the type of soil structure from which they come.
Voids** is described as the combination of (8, 9)		
8	Voids Abundance	An indication of the total volume of voids measured by area and was recorded as the percentage of the surface occupied by pores.
9	Voids Size	The diameter of voids and was recorded in mm.
10	Fissures size	The diameter of fissures and was recorded in mm.
11	Macropores size	The diameter of macropores, which are described as bigger void, mostly determined by plant roots, and by zoological exploration. Macropores were recorded in mm.
*Soil Structure: It refers to the spatial disposition of aggregates which are the result of the aggregation of single particles such us sand, silt and clay. Size, shape and arrangement of these solids and voids, determining the porosity and the capacity to retain fluids and inorganic and organic substances can occur in different patterns, resulting in different soil structures (Bronick et Lal, 2005).** Voids: Include all the pore space present in the soil. It is closely related to the porosity and is a good indicator of soil compactness. It is evaluated as presence/absence data. Voids were described in terms of size and abundance.		

Table 2. Decision tree model (Model 1); performances. RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

	Performance with cross validation				**Performance on training set**			
	N. of instances	Accuracy	RMSE	MAE	N. of instances	Accuracy	RMSE	MAE
Correctly Classified Instances	283	60.08 %	0.44	0.32	335	71.12%	0.37	0.27
Incorrectly Classified Instances	188	39.91 %			136	28.87 %		
	N. of instances	Precision	Recall	F-measure	N. of instances	Precision	Recall	F-measure
Low B$_d$ class	137	0.70	0.54	0.60	137	0.75	0.65	0.69
Medium B$_d$ class	178	0.53	0.63	0.58	178	0.64	0.73	0.68
High B$_d$ class	156	0.62	0.62	0.62	156	0.76	0.74	0.75
Weighted Average		0.61	0.60	0.60		0.72	0.71	0.71

Table 3. Decision tree model (Model 1); confusion matrix.

	Classes classified by decision tree model (N. of instances=471)		
	a	b	c
Low B$_d$ class (a)	**74**	46	17
Medium B$_d$ class (b)	25	**112**	41
High B$_d$ class (c)	7	52	**97**

Table 4. Linear regression model (Model 2); performances. RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

	Performance with cross validation				**Performance on training set**			
	N. of instances	Correlation coefficient	RMSE	MAE	N. of instances	Correlation coefficient	RMSE	MAE
Instances	471	0.65	0.27	0.21	471	0.71	0.25	0.20

Highlights:

- **Two models to estimate B_d values based on soil visual descriptors, are proposed.**
- **Machine learning techniques were used to build the models.**
- **Estimation of B_d by these models can replace complex laboratory analysis.**
- **Relevant properties affecting soil B_d are humic feature, porosity and pedogenesis.**

Using machine learning to predict soil bulk density on the basis of visual parameters: tools for in-field and post-field evaluation

Giulia Bondi[a]*, Rachel Creamer[b], Alessio Ferrari[c], Owen Fenton[a], David Wall[a]

[a]*Teagasc Crops, Environment and Land-Use Research Centre, Wexford, Ireland;*

[b]*Soil Biology and Biological Soil Quality, Wageningen University, Wageningen, The Netherlands;*

[c]*Consiglio Nazionale delle Ricerche, Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo" (CNR-ISTI), Pisa, Italy.*

***Corresponding author**: G. Bondi (Email: Giulia.Bondi@teagasc.ie)*

R. Creamer (Email: rachel.creamer@wur.nl); A. Ferrari (Email: alessio.ferrari@isti.cnr.it); O. Fenton (Email: Owen.Fenton@teagasc.ie); D. Wall (Email: David.Wall@teagasc.ie).

Abstract

Soil structure is a key factor that supports all soil functions. Extracting intact soil cores and horizon specific samples for determination of soil physical parameters (e.g. bulk density (B_d) or particle size distribution) is a common practice for assessing indicators of soil structure. However, these are often difficult to measure, since they require expensive and time consuming laboratory analyses. Our aim was to provide tools, through the use of machine learning techniques, to estimate the value of B_d based solely on soil visual assessment, observed by operators directly in the field. The first tool was a decision tree model, derived through a decision tree learning algorithm, which allows discrimination amongst three B_d ranges. The second tool was a linear equation model, derived through a linear regression algorithm, which predicts the numerical value of soil B_d. These tools were validated on a dataset of 471 soil horizons, belonging to 201 soil profile pits surveyed in Ireland. Overall, the decision tree model showed an accuracy of ~60%, while the linear equation model has a correlation coefficient of about 0.65 compared to the measured B_d values. For both models, the most relevant property affecting soil structural quality appears to be the humic characteristics of the soil, followed by soil porosity and pedogenic formation. The two tools are parsimonious and can be used by soil surveyors and analysts who need to have an approximate in-situ estimate of the structural quality for various soil functional applications.

Keywords: soil bulk density, soil structure, soil quality, machine learning

1. Introduction

The importance of soil structure in relation to soil quality is well known (Mueller et al., 2009; Karlen, 2004; Kay et al., 2006). A commonly used soil physical measurement to characterize soil structural quality is soil bulk density (B_d) (Armindo and Wendroth, 2016; Dam et al., 2005; Håkansson and Lipiec, 2000; Logsdon and Karlen, 2004; Moncada et al., 2015), which is defined as the oven-dry mass per unit volume of soil (IUSS Working Group, 2006; Mueller et al., 2009). Measurement of soil B_d is useful as it describes both the packing structure of the soil and its permeability (Dexter, 1988), whereby drainage characteristics can be inferred (Reidy et al., 2016). B_d measurement is often used in agronomic studies as it indicates the presence of compacted layers resulting from machinery or animal traffic (Reidy et al., 2016; Saffih-Hdadi, 2009), which may affect crop production. It is commonly considered an efficient measurement of soil carbon and nutrient stocks (Ellert and Bettany, 1995; Reidy et al., 2016).

However, the process of measuring B_d is often time consuming and open to human bias in the field and requires accurate laboratory analyses using trained personnel. Furthermore, soil texture has an important influence on the assessment of B_d e.g. in soils with high clay or sand content, or very humic soils, it may be difficult to obtain a representative sample and large variability between replicate samples can represent a problem. Also, in some soils the presence of stones can make sampling almost unmanageable. For such reasons, or constraints of budget or laboratory facilities, B_d measurements are commonly missing from soil databases (Reidy et al., 2016).

The main methods employed for the prediction of B_d are pedotransfer functions (PTF) methods, based on measurable soil attributes, such as organic carbon (OC) and clay content (Kaur et al., 2002; Leonavičiutė, 2000; Reidy et al., 2016). However, many of these methods ignore horizonation and depth variances for soil B_d prediction (Reidy et al., 2016). Furthermore, the nature of these methods, based on chemical/physical or landscape parameters, do not capture the intrinsic nature of the soil structural properties.

Our experience with respect to soil descriptions and classification has shown that the visual observations collected in the field at horizon level are often very important for the evaluation of soil quality (Fenton et al., 2017; 2015) and they become essential during the interpretation of the trend of some analytical parameters used as indicators of soil structure status e.g. B_d.

Soil structural quality has been assessed visually for millennia (Batey, 2000) e.g. soil survey manuals used in the field such as the Soils Survey Division Staff Manual (1993) or the WRB for soil resources (FAO, ISRIC and ISSS, 1998) include soil structure visual observations. However, soil scientists, for a long time, have presented repeatable procedures for the examination of soil structural form, stability and resilience (see latest review by Emmet-Booth et al., 2016 with examples from 1940's to present; Ball and Munkholm, 2015).

Taking this into account, in the present work we investigated whether, and to what extent those visual observations, called descriptors, can be used to predict soil B_d, which is considered one of the most efficient indicators in the assessment of soil structure quality (Moncada et al., 2015).

In order to achieve this objective, machine learning techniques were used. The potential of machine learning techniques have been rediscovered in the last few years through various applications in environmental sciences.

Worldwide, decision tree approaches have been used for different purposes: identifying sources of soil pollution (Xue et al., 2015); describing the extension of different forms of soil erosion in Mexico (Geissen et al., 2007); predicting chemical soil properties at national level in Australia (Henderson et al., 2005); classifying the surface soil freeze/thaw status in China (Jin et al., 2009) and even studying of soil structure through the prediction of soil hydraulic properties (Pachepsky and Rawls, 2003). However, limited literature has been found on the use of these powerful tools for environmental science in Europe.

The decision tree model output applied in this paper is based on a series of rules generated by the software, which can be visualised as paths starting from the root of the decision tree and ending at one of the leaves (Bhargava et al., 2013; Xue et al., 2013; Xue et al., 2015). Each of those paths corresponds to one or more soil *descriptors*, which are related to an internal *node* (Henderson el al., 2005). The model is able to examine all possible descriptors and then to select the most decisive *splitting attribute* (Xue et al., 2013). This operation occurs several times until all the instances are correctly classified in a set of *rules*. Each descriptor included in the model corresponds to a more defined *level* of classification.

The linear regression model applied in this paper is a classical statistical technique used to predict numerical data. It is based on the modelling of the relationship between a scalar dependent variable and one or more explanatory variables.

With our work we want to:

(i) Provide an operational strategy to estimate a range of B_d values, based on the visual soil parameters by means of a decision tree approach. This model can be used as a field tool to predict a general class of B_d (Low, Medium and High). It is an instrument able to discriminate between macro classes and has to be considered as a descriptive tool for *qualitative* estimation.

(ii) Propose an algorithm that can predict a numerical estimate of B_d. This second model should discriminate better between smaller increments. This instrument has to be considered as a more refined tool for *quantitative* estimation.

2. Methods

2.1 Primary data source and descriptors

Two pedological surveys, where full soil profile descriptions and supporting laboratory analyses, were carried out in Ireland with the aim of defining a coherent and homogeneous way to study soil formation, functions and quality:

1. The Irish Soil Information System (Irish SIS) project was established in 2008. It aimed to conduct a programme of structured research into the national distribution of soil types and construct a soil map, at 1:250,000 scale, able to identify and describe the soils according to a harmonised national legend. Irish SIS included more than 225 sites distributed around Ireland (Creamer et al., 2014).

2. The Soil Quality and Research project (SQUARE) started in 2013. The aim was to establish a baseline of soil quality in Ireland. The SQUARE soil survey included 38 grassland sites distributed within the five major agro climatic regions of Ireland defined by Holden and Bereton (2004) and classified into two drainage classes on the basis of the Irish Soil classification System.

During both (1) and (2) profile pits approximately 1 m deep, were observed and described by different operators. For the present study data from 201 profiles (168 Irish SIS, 33 SQUARE) was extracted from the larger database to cover a wide variety of Irish soil types with a specific focus on mineral soils. This data represents 471 horizons (http://gis.teagasc.ie/soils/map.php).

Although different surveyors worked across the projects mentioned, a systematic procedure was applied to describe the nature of the soil profiles, which included each of the soil horizons. Training was given to field operators. Using knowledge of soil structure and quality, the operators followed a widely understood schema of observation (developed by FAO through the Guidelines for Soil Description in 2006) which was able to investigate and finally characterize soil structure through visual parameters (FAO, 2006; FAO, ISRIC and ISS, 1998). Herein we have selected eleven descriptors presented in Table 1 (justifications are provided in Table 1), which may be considered the most important for the qualitative judgment of soil structure. Each descriptor was described and recorded on the basis of a set of pre-defined categories, reported in Table 1 in the Supplementary Material.

2.2 Soil analysis

The procedure to determine B_d of intact cores is a version of the ISO 11272:1998 – Soil Quality Part 5: Physical methods Sect. 5.6 – Determination of dry bulk density. The primary difference between the ISO and the applied methodology is that the ISO does not account for stone mass and volume in its core method, whereas the methodology applied in this study includes the following equation to calculate B_d (stone free):

$$B_d\ (g\ cm^{-3}) = (Md - Ms)/(V - Vs) \qquad \text{Eq 1}$$

where; Md: oven dry soil material weight (g), Ms: oven dry stone weight (g), V: volume of soil core (cm^{-3}), Vs: volume of stones (mL). Soil B_d values reported in this paper correspond to the mean of the three values obtained for each horizon sampled.

2.3 Model frameworks

Two models were built by means of the modelling tool WEKA (Waikato Environment for Knowledge Analysis). WEKA 3.8 is open source software for machine learning and data mining under the General public license developed at the University of Waikato in New Zealand (http://www.cs.waikato.ac.nz/ml/weka, Bhargava et al., 2013). This software includes different implementations of several machine learning algorithms. In our context, we used two specific algorithms that are made available by the tool, namely:

- the j48 algorithm, which corresponds to the WEKA's implementation of the C4.5 decision tree learner (Quinlan, 1993; Xue et al., 2015) which was used to build Model (1);
- a linear regression algorithm, used to build the Model (2). The M5 Method was used as attribute selection method for the linear model presented.

Two models were produced to achieve the objectives of our work:

- Model (1) is based on a classic decision tree model, developed to be used in the field in order to predict a B_d class using only visual descriptors as in Table 1.
- Model (2) is a linear regression model that uses the same in-field descriptors as above, and it is able to predict a numerical value of B_d with a relatively small error.

The proposed two models are both descriptive and predictive, but the decision tree is better at exploring in a descriptive way the relationship between B_d and visual parameters, as it allows further analysis of the soil pertaining to that soils own chemical and physical characteristics. On the other hand, the linear equation algorithm is stronger as a predictive tool and it offers a more precise estimate of B_d.

2.3.1 Data treatment

The entire database consists in 201 sampling points (profile pits) for a total of 471 horizons. For each horizon eleven descriptors and B_d data were used to train Model (1) and Model (2). The treatment of data can be summarized as follows:

- Data cleaning: to produce a full dataset, time was invested to ensure the data homogeneity between Irish SIS (649 horizons) and SQUARE datasets (125 horizons). In particular, descriptor rating options were double checked to reaffirm consistency across projects. To achieve uniformity within the dataset some data conversions were necessary. The final dataset consisted of 471 horizons i.e. 346 from Irish SIS and 125 from SQUARE.
- Missing values imputation: an initial analysis of the dataset highlighted the presence of some missing values for part of the considered descriptors, namely: “Fissure size”, “Void size”, “Void abundance” and “Soil consistency”. To avoid further reductions of the dataset, an IMRI imputation (performed by the WEKA 3.8 software described above) was selected as the means to predict missing values (Templ et al., 2011).

2.3.2 Model 1: Decision tree model, validation and outputs

Model (1) has been designed to predict classes of B_d data through the combination of the 11 visual descriptors outlined in Table 1. The model has been trained using B_d data at horizon level. The predicted classes of B_d are:

(i) Low B_d class: < 1.0 g cm^{-3} (n=137 cases)

(ii) Medium B_d class: between 1.0 and 1.4 g cm^{-3}(n=178 cases)

(iii) High B_d class: > 1.4 g cm^{-3} (n=156 cases)

Class ranges were selected on the basis of their homogeneity in terms of class population. The B_d measured in the majority of mineral soils under agricultural management in Ireland occur typically within the 1.0 and 1.4 g cm^{-3} range. Values of < 1.0 g cm^{-3} are usually related to Ombrotrophic or Mineratrophic Peat Soils (which correlates with the Histosol reference soil group of the WRB (IUSS Working Group WRB, 2006) or mineral soils having a Histic horizon (Reidy et al., 2016). Therefore 1.0 g cm^{-3} was selected as the lower B_d threshold e.g. herein 16 cases out of 137 belonged to the Low B_d class as Oh, Op, Of or Omf. The higher threshold (1.4 g cm^{-3}) was empirically chosen to best fit these data. In particular, multiple decision tree models were trained by varying the threshold between 1.1 g cm^{-3} and 1.8 g cm^{-3}, in 0.1 intervals. The model trained with the 1.4 g cm^{-3} threshold outperformed, other model runs in terms of accuracy.

The decision tree produced herein can be easily converted into classification rules. Each path in the tree that goes from the root to one of the leaves defines one classification rule. In our case each rule categorises the data in B_d Low, Medium and High classes. The knowledge represented in a decision tree can be extracted and represented in the form of the classification rule IF-THEN as follows:

If {condition A} AND {condition B} AND {condition C} AND {…} **then** categorization.

In our case:

If

the horizon is described as HUMOSE

then

the horizon fits into the bulk density category “Low”= <1 g cm^{-3}

A pruning technique is automatically performed by the WEKA software. This allows the identification and the removal of the outliers reducing the risk of overfitting to the training data (Bhargava et al., 2013). When decision trees are built, many of the ramifications can represent noise or outliers in the training data. The pruning process tries to identify and remove these branches with the aim of improving the accuracy of classification of future data. The next step was to prune the dataset to identify and remove branches which do not improve prediction with the aim of improving the accuracy of classification of future data.

A10-fold cross validation method was adopted, which randomly partitions the dataset into 10 parts and is used to validate the model. Then nine parts of the dataset were used to train the model, with the last part used for model testing (see Xue et al., 2015 for a similar approach).

Two measures were applied to evaluate the model performance; precision and recall values were calculated. Precision indicates how many of the instances were classified within a certain class that actually belong to that class. Whereas recall indicates how many of the instances that belong to a certain class are correctly classified by the model. To explain these measures further, it is useful to introduce the concepts of true positive, false positive and false negative.

Given a class C, we define true positives *tp* as the number of instances labelled as C in the original dataset, and classified as C by the decision tree; false positives *fp* as the number of instances not labelled as C in the original dataset, and incorrectly classified as C by the decision tree; false negatives *fn* as the number of instances labelled as C in the original dataset, and not classified as C by the decision tree. Given these definitions, precision p_C and recall r_C for the class C are defined as follows:

$$p_C = \frac{tp}{tp + fp}; \qquad \text{Eq 2}$$

$$r_C = \frac{tp}{tp + fn} \qquad \text{Eq 3}$$

Precision is negatively influenced by the number of false positive cases. Whereas, recall is negatively influenced by the number of false negative cases. High scores for precision and recall show that the classifier is returning accurate results (high precision, related to low false positive rates), as well as returning a majority of all positive results (high recall, related to a low false negative rates).

As no baseline algorithms were available, or proposed in other publications, against which to evaluate the performance of our Model (1), we resort to comparing it with a random predictor baseline, i.e. a fictional algorithm that randomly predicts the class of an instance (Alvaretz, 2002).

Let *n* be the total number of instances, and let *c* be the number of instances of class C, the precision and recall of a random predictor for the class C is given by:

$$p_C = r_C = c / n \qquad \text{Eq 4}$$

Besides precision and recall for each class, we also evaluated the average value of these measures. In addition, we evaluated the harmonic means of precision and recall, which is normally called F-measure, and is defined as follows:

$$\textit{F-measure}=2 * \frac{p_c * r_c}{p_c + r_c} \qquad \text{Eq 5}$$

Finally, as an overall indicator of the ability of the decision tree to correctly classify the instances, we evaluated the overall accuracy, which is defined as follows:

$$\textit{Accuracy (\%)} = \frac{tp + tn}{tp + tn + fp + fn} \qquad \text{Eq 6}$$

2.3.3 Model 2: Linear regression model, validation and outputs

Using data from the visual parameters (Table 1), a linear equation that predicted an exact B_d value was developed.

As for Model (1), Model (2) was learned using B_d data at horizon level and produced by means of the WEKA software, using a linear regression algorithm. For this experiment these data were converted into numerical binary data, since the linear regression algorithm takes numerical data as input. In particular, for each value taken by each descriptor, a binary variable was created that takes either 0 (False) or 1 (True) as values. The variable was 1 if the descriptor had the value associated to the variable, and 0 otherwise.

The model builds a linear equation based on a weighted combination of the possible values taken by the 11 descriptors. In particular, the linear model has the following form:

$$B_d \ (g\ cm^{-3}) = \sum_{i=0}^{n} C_i * V_i \qquad \text{Eq 7}$$

where *Ci* are the coefficients computed by the linear regression algorithm. *Vi* are the binary variables. The linear regression algorithm is designed to select only those variables that have an influence on the final B_d value. Hence, the final model will not include all the possible variables, but only a subset. A 10-fold cross validation was performed to validate the model as described in par 2.3.2.

For Model (2) we could not evaluate the performance through precision and recall, since numerical values are involved instead of categorical ones, but we evaluated it in terms of correlation coefficient, root mean squared error, and mean absolute error:

The Root Mean Squared Error (RMSE) gives an estimation of the standard deviation of the error (Henderson et al., 2005). The lower is RMSE the higher is the predictive ability. Where *n* is the size of the dataset and $\hat{y}_t$ is the predicted value, the formula is defined as follows:

$$RMSE=\sqrt{\frac{1}{n}\sum_{t=1}^{n}(y_t-\hat{y}_t)}^{\,2} \qquad \text{Eq 8}$$

The Mean Absolute Error (MAE) is a quantity used to measure how close predictions are to the eventual outcomes. The mean absolute error is also known as the mean absolute deviation (Henderson et al., 2005). The lower the MAE value the higher is the predictive ability.

$$MAE=\frac{1}{n}\sum_{t=1}^{n}|y_t-\hat{y}_t| \qquad \text{Eq 9}$$

3. Results and discussion

3.1 Model 1: Decision tree approach to assess B_d classes

3.1.1 Model performances

The number of rules generated by the decision tree algorithm was 41 in total; 11, 13 and 17 for Low, Medium and High respectively (all the rules are reported in Table 2 in the Supplementary Material). The overall tree is reported in Figure 1. The decision tree hierarchy consists of six levels (level 1 has the highest classification power) of depth as follows:

Level (1): Humose

Level (2): Structure type

Level (3): Macropores Size/Void Size/Void Abundance/Plasticity

Level (4): Structure Grade/Stickiness

Level (5): Fissures

Level (6): Structure size

The descriptor "soil consistency" was excluded by the tree hierarchy, showing no influence on the prediction of B_d ranges. To discuss these results, it is useful to associate the different levels and hence, the corresponding descriptors, to specific soil quality properties. Broadly the descriptors that remained in the analysis fell into 4 main groups in order of importance as follows:

(i) soil humic characteristics; explained by level (1);

(ii) pedogenic formation; explained by level (2);

(iii) soil porosity; explained by level (3);

(iv) soil cohesive properties; explained by level (3-4).

The majority of the 41 rules were classified within the first three hierarchical levels highlighting the importance of these soil structural descriptors in quantifying B_d class. In general the parallelism between soil aggregation mechanisms, soil intrinsic characteristics related to soil forming factors, and the arrangement of solid and voids with their capacity to retain and transmit fluids, resulted in the main factors being capable of explaining soil B_d.

The overall accuracy shows that the model correctly classified 71% of the training data (Table 2). However, after the 10-fold cross validation step the model was able to correctly predict 60% of the cases (Table 2). The confusion matrix shows that mis-categorisation does not occur from Low to High B_d classes and vice versa, but can occur from Low to Medium (46 cases) and from High to medium (52 cases) (Table 3). The cause of lower model performance can be clarified by looking at the branches of the decision tree that miss-classify the highest number of instances (for discussion see par. 3.2.1).

The decision tree model was evaluated using the random predictor method (see par. 2.3.2). Performances are the following:
Low: $p_L = r_L = 137/471 = 0.29$.
Medium: $p_M = r_M = 178/471 = 0.37$.
High: $p_H = r_H = 156/471 = 0.33$.

For the Low B_d class, the decision tree outperforms a random predictor by 40% in terms of precision and by 25% in terms of recall; for the Medium B_d class, by 16% in terms of precision and by 26% in terms of recall; for the High B_d class by 30% in terms of precision and by 29% in terms of recall. These figures indicate that, although the accuracy indicates poor performance for the Low B_d class, the prediction for this class in terms of precision is actually the best amongst all 3 classes. This is due to the lower number of Low B_d instances, which are harder to detect for a random classifier, and that our algorithm predicts with a substantially higher degree of precision. Overall, the classification produced by the model is considerably better than a random classification.

3.1.2 Rules analysis

Figure 1 shows the decision tree generated for each B_d class with the corresponding instances classified for each rule and the percentage of accuracy, which refers to the percentage of true positive cases found by the model for the rule discussed. Rules classifying less than two instances are not reported in these figures.

- **Level 1: "Humose"**

The **"Humose"** descriptor emerges as the most dominant discriminator. All the horizons that highlight the presence of the "humose" feature following the rule "Humose: *yes*", fit in the category Low B_d. The model correctly classifies 85.9% of n=64 instances for this rule with relatively few false positive cases.

The humose feature refers to an estimation of the level of humification of the organic material, so it is indirectly related to the C content. In fact, soil OC content gives an indication of decomposition rates which have a direct effect on soil aggregation (Bronick and Lal, 2005; Schulten and Leinweber, 2000). The presence of humic substances, as well decomposed organic matter (OM), contributes to the stability of soil aggregates and pores through the bonding or adhesion properties of organic materials, such as bacterial waste products, organic gels, fungal hyphae and worm secretions and casts. Moreover, OM intimately mixed with mineral soil materials induces increased moisture holding capacity and air exchange with the atmosphere (Stevenson, 1994). This mechanism is reflected by lower values of B_d, resulting in enhancement of soil porosity producing a good soil structure.

When the humose feature is recorded as absent, the model switches to level 2 for further refinement.

- **Level 2: “Structure Type”: Granular**

“Structure Type” was the most discriminating descriptor at the second level of the tree hierarchy. Here the tree branches out to account for the effect of different types of structure.

For the rule “Humose: *no*, AND Structure type: *Granular*” the model classified 23 cases as belonging to the Low B_d class, with an accuracy of 56.5%. Despite the high number of false positives, the amplitude of the group is among the highest, showing its reliability.

Granular aggregates are spheroids or polyhedrons aggregates of soil, having curved or irregular surfaces (FAO, 2006), they are normally associated with highest air capacity, and are an indicator of good soil structure (Mueller et al., 2009). Such cases are mostly A horizons, not deeper than 30 cm. Root mass, density, distribution and turnover, can positively influence the soil particle aggregation by releasing a variety of compounds, (such as root exudates) and can contribute to increased soil porosity through mechanical action (Bronick and Lal, 2005; Caravaca et al., 2002). This is reflected in the B_d, resulting in lower values.

- **Level 2: “Structure Type”: Subangular Blocky → Level 3, “Macropores Size” → Level 4 ”Structure Grade”**

In presence of the “Subangular Blocky” structure type the tree branches split again, going deeper into the third hierarchical level of tree structure, showing the **“Macropores size”** descriptor as the next level of description. For the rules “Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *(A), (B), (C), (D)*”.:

(A) Coarse (C 5.0-20.0 mm): 50% accuracy when classifying Low B_d class.

(B) Fine (F 0.5-2.0 mm): for subangular blocky aggregates, 66.1% accuracy when classifying Medium B_d class. This is a rule that has a higher number of

instances (n=92), which balances the presence of a relatively high number of false positives. For angular blocky aggregates, 72.7% accuracy when classifying Medium B_d class (n=11).

(C) Very Fine (VF< 0.5 mm): 66.7% accuracy when classifying High B_d class.

(D) Medium (M 2.0-5.0 mm): 100% accuracy when classifying Low and Medium B_d class.

Subangular blocky structure type is described as cube like with a flat surface and rounded corners (FAO, 2006; Schoeneberger et al., 2012) and is considered an indicator of an intermediate degree of soils structure quality, between the granular/crumbly aggregates and the blocky/sharp angular ones. Results for this structure type show that the size of the macropores is a crucial variable to discriminate in terms of B_d classes, as B_d is very sensitive to both alteration of macroporosity abundance and size of pores (Mueller et al., 2009). In particular the higher the macropore size, the lower the B_d class predicted by the model.

For the rule *(A)*, horizons fitting the Low B_d class are mostly Ap horizons, corresponding to the upper soil layers (maximum depth of 22 cm). Earthworm activity is mainly concentrated in the top soil (Haynes and Naidu, 1998; Lee and Foster, 1991), which promotes macropores (C 5.0-20.0 mm) which in turn can alter soil porosity thereby affecting the movement of air, water and solutes (Shipitalo and Le Bayon, 2004). For the rule *(B)* and *(C)* medium and high B_d are associated with macropore sizes from 0.5 mm to 2.0 mm. The cases following these two rules are mainly classified as A or B horizons for the medium B_d class (average maximum depth of ~40 cm) and as B horizons for the higher B_d class (average maximum depth of ~60 cm). At such depths roots become thinner resulting in reduced porosity and higher B_d. Furthermore, B_d often increasesin the deeper horizons due to an increase in clay material or as result of management operations.

The medium class of macropores (M 2.0-5.0 mm) insufficiently defines the B_d range and must be assisted by another splitting attribute which is triggered at level four. The **"Structure grade"** is identified by the model as an important variable to discriminate between the Low and Medium B_d class for this macropores size. The rule (*D*) is consequently developed as follows:

- "Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *M 2.0-5.0 mm* AND Structure grade: *Moderate*"; which predict Low B_d class (100% correct instances).
- "Humose: *no*, AND Structure type: *Subangular Blocky* AND Macropores: *M 2.0-5.0 mm* AND Structure grade: *Weak*"; which predict Medium B_d class (100% correct instances).

FAO (2006) defines soils by describing the lack (apedal) or the presence (pedal) of a defined structure. A moderate structure grade, showing the presence of a nicely structured soil, was associated with a Low B_d class. Medium B_d classes are associated with a weak structure grade, synonymous of a lower structure quality. However, herein cases classified for these two rules were similar in terms of actual B_d values. In particular, cases classified as Low B_d had actual values very close to the upper limit of this category (values between 0.9 and 1.0 g cm^{-3}) and the cases classified as Medium B_d had actual values close to the lower limit of this category (values between 1.0 and 1.16 g cm^{-3}). Considering the narrow range of values for these cases across the Low and Medium B_d categories the model correctly discriminates between categories.

- **Level 2: "Structure Type": Angular Blocky → Level 3, "Void Size" → Level 4 "Stickiness"**

The "Angular Blocky" aggregates belong to the blocky category as for the subangular blocky aggregates, but differ as they have faces intersecting at relatively sharp angles (FAO, 2006; Schoeneberger et al., 2012). In the presence of this structure type i.e. a typical indicator of a poorly structured soil, the model returns **“Void size”** as the next, most important, descriptor. This descriptor indicates the total volume of pores discernible with a *10 hand-lens (FAO, 2006). It differs from macroporosity as it is a much wider term that includes soil fissures and plane (FAO, 2006). Macropores are mostly determined by plant roots through the twisting activity within and around aggregates, and by zoological exploration, through burrowing activity, so they are usually larger pores having a size higher than 75 µm (Russell, 1975). However, here macropores and voids are described in terms of size, so a high void size can correspond to a high macropores size. Therefore, it is important to highlight here that although different from a semantic point of view, the combination of these two descriptors was very effective in explaining soil porosity.

For the rule “Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *C 5.0-20.0 mm*” the model returned 75% of correct instances for the class Low B_d. Although the structure type indicated was often that of a very poor structured soil, characterised by large, angular and sharp aggregates, this effect was mitigated by the presence of a very high overall porosity, resulting in a Low B_d (Pagliai and Vignozzi, 2002).

The rule “Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *VF <0.5 mm*”, returned High B_d class with 76.9 % accuracy. Also this rule was quite wide in terms of population (n=26). In this case the poorly structured horizons were associated with the smallest size of pores. The cases are mainly classified as Bg, BCg, Cg or Eg horizons, with an average depth ranging from 42 to 72 cm. For such cases, gleying is caused by surface water which has been held in a poorly permeable horizon, (mainly belonging to surface water gley soil type (Stagnosols reference soil group of the WRB). These poorly permeable layers showed evidence of compaction due to one or more of the following characteristics: (i) presence of a pan, resulting in severe compaction imposed by management, (ii) poor structural development, (iii) very heavy textured horizons dominated by silt and clay and (iv) natural impeded horizons due to soil formation and profile development mechanisms.

Furthermore, the presence of some argillic horizons classified as Btg following this rule, as well as having gleyic features, showed the presence of clay-sized material in coatings or as intrapedal concentrations (FAO, 2006). Clay content physically affects particle aggregation through swelling and dispersion, resulting in contiguous peds which fit together eliminating space (Attou et al., 1998; Bronick et Lal, 2005; Kay, 1998; Russel, 1975).

When the void size is classified as belonging to the middle category (Humose: *no*, AND Structure type: *Angular Blocky* AND Void size: *F 0.5-2.0 mm*), **"Stickiness"** was selected by the model as the next most informative descriptor, associating Medium B_d class with the "*Non Sticky OR Slightly Sticky*" feature (50% and 63.6% accuracy, respectively), and the High B_d class with the "*Sticky*" feature (accuracy of 75%). The rules' population size was n=44, n=22 and n=12, respectively and therefore must be considered an important branch of the overall tree. As the stickiness property is considered an indirect indicator of the clay content in soil, the results confirmed that the presence of clay material was one of the characteristics which makes the soil prone to compaction, or at least significantly decreases the level of porosity. Even in the presence of a relatively higher porosity, clay content was crucial to force a split in B_d classes, thereby attributing higher B_d values to heavier textured soils.

"Plasticity" appears at the third depth level of the decision tree for the structure type Angular Blocky to Granular. Like stickiness, plasticity is directly related to the clay content (FAO, 2006). Horizons classified as "*Plastic* OR *Slightly plastic* AND Macropores size: *Very Fine (VF< 0.5 mm)*", fit into the

High B_d class, highlighting the link between high values of B_d and clay content (accuracy respectively of 100% and 66.7%).

- **Level 2: "Structure Type": Prismatic → Level 3, "Void Abundance" → Level 4 "Structure Grade" → Level 5: "Fissures"→ Level 6: "Structure Size"**

The "Prismatic" structure type is described by FAO, (2006) and by Schoeneberger et al. (2012) as having vertical elongated units with limited faces in the horizontal plane. Horizons having prismatic to angular blocky aggregates are classified by the model within the High B_d class (accuracy, 62.5%). The average depth ranges within 53-95 cm, classified across a number of soil types e.g. typical surface water gleys (Stagnosols reference soil group of the WRB), typical luvisols (Luvisols), typical brown podzolic (Podzols) and typical calcareous brown earths (Calcaric Cambisols). Different soil types are subjected to different mechanisms of soil particle aggregation which can exert a degree of compaction in a specific horizon, due to their intrinsic nature or to a combination of factors, such as: (i) clay concentrations in the impeded horizon, as for luvisols; (ii) translocation of Al and Fe in the spodic horizon, as for brown podzolic (Bronick and Lal, 2005; Collins, 2004), (iii) presence of carbonates, as for the calcareous brown earths (Boix-Fayos, et al., 2001) and (iv) presence of a poorly permeable horizon, as for the gleyic horizons in surface water gleys (Collins et al., 2004).

Furthermore, for the prismatic structure type, **"Voids abundance"** is selected as the main descriptor to split the data between High and Medium B_d for this structure type.

For the rules "Humose: *no*, AND Structure type: *Prismatic* AND Void abundance: *A, B, C*:

(A) High (15-40%): the model returns 66.7% of correctly classified instances as belonging to the Medium B_d class,

(B) Very Low (<2%): the model classifies for the High B_d class with an accuracy of 75%.

(C) Medium (5-15%): the model classifies for Medium and High B_d with an accuracy of 83.3 and 89%, respectively.

The Low B_d category was not defined by the model for this structure type, indicating the prismatic arrangement as having overall higher B_d values. The abundance of voids is important, as the void size for this type of structure, poorer than the angular blocky structure type. Rules *(A)* and *(B)* associated higher void abundance with higher structure quality. On the other hand the rule *(C)*, which took into account the medium void abundance category, requires further descriptors to categorise within the High and the Medium B_d classes, such as **"Structure grade"**, **"Fissures"** and **"Structure size"**.

The presence of smaller sized prismatic aggregates (Structure size: *Fine 10-20 mm*) directs the model to predict High B_d class with 100% correct instances. As found in the present study, the rapid wetting of dry soil which comes in contact with free water can cause micro-cracking. This increases ped friability, causing the production of smaller sized aggregates, which does not always result in lower values of B_d (Dexter, 2002).

Vertical fissures are also an important feature of this structure type. Although hard clods are devoid of microstructure, fissures enable the percolation of the surface water in the deeper layers (Russel et al., 1975). Such a drainage advantage is not valid for the "Platy" structure type. In this case the model returns two different outputs. Platy aggregates, probably formed by intensive mechanical intervention, indicate compaction which affects water percolation, resulting in high B_d values (accuracy, 66.7%). On the other hand, some platy aggregates were attributed to the histic horizons (Of, Omf) which showed no developed structure, associated with very low B_d (accuracy 75%).

- **Level 2: “Structure Type”: Massive**

Structure types defined as “Massive” and “Single grain” were categorised by Schoeneberger et al. (2012) as structure less soils. Massive soil material normally has stronger consistence as the soil particles are arranged in a coherent mass and are very difficult to break (FAO, 2006). Horizons with “Massive”, “Massive to Angular blocky” and “Massive to Single Grain” structure types were classified by the model as belonging to the High B_d class with an accuracy of 77.8%, (n=63; very influential rule in the overall tree output), 83.4% and 100%, respectively. Further analysis of the data showed that most of the horizons belonging to the rule: “Humose: *no*, AND Structure type: *Massive*” were described as having a hard or very hard consistence dry, fine or very fine macropores and very low void abundance. The decrease in soil macroporosity and the firm nature of the aggregates suggests a high level of compaction (Epron et al., 2016), resulting in high B_d values which ranged from 1.4 g cm^{-3} to 1.9 g cm^{-3}. In the case of the Massive to single grain structure type, **“Plasticity”** is a key attribute to assess the range of B_d (rule: “Humose: *no*, AND Structure type: *Massive to Single grain* AND Plasticity: *Non plastic*”).

The arrangement of the aggregates which are weak and tend to disintegrate when sampling in the field, suggested the presence of sandy material held together in big, hard and massive aggregates. In some soils, sand grains have a film of orientated clay particles on the surface, not enough to be detected by feel, but that are able to strongly hold the sand particles packing them together in massive aggregates (Russel, 1975). Sandy soils are prone to compaction of surface layers, due to intensive agricultural operations (Ampoorter et al., 2007; Deconchat, 2001; Teepe et al., 2004).

Low B_d class was attributed by the model to the horizons responding to the rule “Humose: *no*, AND Structure type: *Single grain*”, which is a feature typical of sandy soils not subjected to compaction phenomena, thereby conserving a large amount of wide pores (Ampoorter et al., 2007). The model correctly classified 66.7% of the instances following this rule.

3.2 Model 2: Linear regression approach to predict a numerical estimate of B_d

3.2.1 Model and performance

After several trials with different machine learning methods for numerical prediction, a linear regression model was selected based on better overall performance. This conclusion asserts that a strong linear relationship exists between the visual descriptors and B_d values. As a result, since the model produced by a linear regression algorithm is a linear equation, the predicted value of B_d can be easily computed in seconds without the need for time consuming and costly laboratory analyses.

Model (2) considers only those descriptors that influence the final B_d value. For this quantitative estimation seven descriptors out of eleven were selected by the linear regression algorithm:

(i) Humose
(ii) Structure Grade
(iii) Structure Type
(iv) Structure Size
(v) Macropores
(vi) Void Size
(vii) Void Abundance

The linear equation produced by Model 2 to predict B_d is as follows:

B_d (g cm $^{-3}$) =
0.4575 * *(NO **Humose**: =true)* +
-0.0517 * *(MODERATE **Structure Grade**=true)* +
-0.2128 * *(GRANULAR **Structure Type**=true)* +
0.093 * *(MASSIVE **Structure Type**=true)* +
-0.134 * *(SUBANGULAR BLOCKY **Structure Type**=true)* +
-0.1501 * *(SUBANGULAR BLOCKY TO GRANULAR **Structure Type**=true)* +
-0.1003 * *(ANGULAR BLOCKY TO GRANULAR **Structure Type**=true)* +
-1.1619 * *(VERY COARSE > 10mm **Structure Size**=true)* +
0.1471 * *(VERY FINE < 0.5mm **Macropores**=true)* +
-0.1791 * *(COARSE 5-20 mm **Void Size**=true)* +
0.0802 * *(VERY LOW **Void Abundance**=true)* +
-0.0784 * *(HIGH **Void Abundance**=true)* +
0.8866 Eq 10

The equation produced follows a simple framework:
(i) All the descriptors associated with a *positive coefficient* caused a significant incremental increase in B_d. Basically, as the associated coefficients are positive, a soil classified having these descriptors (*Vi=1=true*, see description in paragraph 2.3.3), will result in an increased B_d final value.

The variables *"NOHumose=true"*, *"VERY FINE<0.5mm Macropores=true", "VERY LOW Void Abundance=true"* and *"MASSIVE Structure Type=true",* all cause an increase in B_d, with 0.4575; 0.1471; 0.0802; and 0.093 explanatory power, respectively.

In line with Model (1) results, the humification degree of OM has the greatest influence on B_d prediction. This characteristic, as well as readily defining B_d class, also informed small differences in soil B_d and generally indicated pedological features that were consistent with lower B_d values.

Following the humic feature, macropores size was the second more discriminating feature, with the lower size able to distinguish a wide increment of prediction (multiplier of 0.1471). Furthermore, very low void abundance triggered the model as the fourth main attribute (multiplier of 0.0802). This was an expected result, as porosity was already investigated by the decision tree approach and was one of the most critical soil characteristics which took into account an evaluation of soil structure. In particular, size of pores was highlighted as a stronger predictor for an increase of B_d with respect to pore abundance. As seen from the decision tree output, the massive structure type has a role in increasing soil B_d. In general both models although operating at different scales produce the same descriptors for the prediction of B_d.

(ii) All the descriptors associated with a *negative coefficient* have a role decreasing B_d. Basically, as the associated coefficients are negative, a soil classified as having these descriptors, resulted in a decreased B_d, and therefore have a better soil physical quality.

In Model (2), structure size appears to be a stronger descriptor which influences a decrease in B_d when evaluated as *"VERY COARSE>10 mm Structure Size=true"* (multiplier of -1.1619). This was surprising considering that in Model (1) the size of aggregates was not particularly important (sixth level) in the tree hierarchy as a splitting attribute. This highlights that this feature is better at identifying small incremental reductions of B_d, but is less informative when splitting into wider B_d ranges.

However, in Model (2) Structure type still has to be considered one of the most informative variables, since it appears in four out of eight factors having a negative coefficient in the equation. In particular, looking at the equation, we have:

- *"GRANULAR Structure Type=true"* ,
- *"SUBANGULAR BLOCKY TO GRANULAR Structure Type=true"*,
- *"SUBANGULAR BLOCKY Structure Type=true"*,
- *"ANGULAR BLOCKY TO GRANULAR Structure Type=true"*,

as coefficients -0.2128; -0.1501; -0.134; -0.1003, respectively.
Granular structure type is responsible of a higher decrease of B_d, confirming what was found for Model (1), indicating good soil structure.

The model showed sufficient sensitivity allowing the identification of differences related to soil structure type. This is despite the relatively crude measurement of soil structure at field level in tandem with other attributes. In particular while the soil structure quality, associated with a change of structure type, gradually decreases, with diminishing negative effect on B_d, indicated by the coefficients for individual structure types. Hence, corresponding to their respective coefficients, a granular structure type will have a higher negative increment, resulting in a reduction of the B_d final value, while an angular blocky to granular structure type will have a lower negative reduction in the final predicted value, resulting in a higher B_d final value.

Structure size, Void size, Void abundance and Structure grade were, in order of importance, the next most informative features for the linear regression model.
If both the features relating to porosity appear in the Model (1) at a high level in the tree hierarchy (level 3), for this model only void size with the variable *"COARSE 5-20 mm* ***Void Size****=true"* appears to have higher negative impact, with a relatively high coefficient of -0.1791, while void abundance (*HIGH* ***Void***

Abundance=*true*) resulted in having less negative impact on the definition of a final B_d value, being associated with a smaller coefficient, -0.0784. Probably, as per the decision tree model, the shape of aggregates is again a critical point which drives the selection of the second decisive node into pore abundance or size, depending on the original structure type.

3.2.2 Overall model evaluation

The overall correlation coefficient on training set for the linear regression model is 0.71 with Root Mean Squared Error (RMSE): 0.25 and Mean Absolute Error (MAE): 0.20; (Table 4). After a 10-fold cross validation the correlation coefficient slightly dropped, to 0.65, with similar error ranges (RMSE: 0.27 and MAE: 0.21); (Table 4). The errors reported for this model may be considered quite high in relation to a standard lab-based B_d measure. However, it is important to highlight that the model has been fed using only soil visual parameters. Considering the nature of these data inputs and the influence that different operators can have during the classification phase, this range of error is low.

Figure 2 shows the prediction performances of Model (2). The distribution of predicted values results more coherent with the real values for a middle range of B_d values. In particular we identified a range that goes between 0.8 to 1.6 g cm^{-3}, which falls within the typical range of B_d found in Irish grassland soils, where the model returns B_d values close to real values. In these cases the model is more robust, predicting a numerical estimate with a quite low standard error. Furthermore, neither overestimation nor underestimation prevails for a middle range of B_d.

The model shows the higher errors for the extreme B_d classes, namely (i) very low B_d, that we identified as values lower than 0.8 g cm^{-3}, or (ii) very high B_d values, identified as values higher than 1.6 g cm^{-3}. The algorithm appears to a have higher prediction power on medium B_d values which also receive higher representation in our dataset. In general, machine learning algorithms are improved where greater input data is provided. Therefore, in our case, the algorithm is inherently biased towards the correct prediction of medium ranges.

3.3 Models choice considerations

We have chosen to use decision trees and linear regression because these simple types of models allow users to identify the B_d class (for decision trees), and the B_d value (for linear regression) without the need to rely on additional software. Furthermore, besides the predictive ability, decision trees also provide descriptive power, in that they make explicit the relationships among different characteristics of the soils and allow the user to have greater insight to these relationships. Other algorithms that were considered, i.e., support vector machines and multi-layer perceptron, although leading to similar performance, do not allow this type of insight. A full comparison between different algorithms, from the performance point of view, can be conducted in future. While the quality of the interrogated database was good, we believe that further improvement of model performance can be achieved by increasing the extent of the sample dataset, especially for horizons with low and high B_d values, which are less represented in our data. Finally, the utility of these models to assess critical thresholds for compaction should be evaluated for descriptive soil datasets where attributes such as “Compact degree” (FAO, 2006) are included.

4. Conclusion

A decision tree and linear equation model were developed to predict soil bulk density on the basis of visual descriptors. The visual soil descriptors identified, as being more informative by both models, are associated with specific soil properties. This allows the user to rank to these properties in terms of their impact on soil structural quality. For both models the most relevant properties that affect B_d appears to be soil humic characteristics, followed by soil porosity and pedogenic formation.

Overall, the decision tree model shows an accuracy of about 60%, while the linear equation model had a correlation coefficient of about 0.65 with respect to the measured B_d values. The two models are parsimonious and can be used by soil surveyors and analysts who need to have a quick and approximate *in-situ* estimate of the structural quality for various soil functional applications. Furthermore they have an enormous potential to retrofit B_d data (i.e. gap fill) to existing data sets were laboratory data are missing. Future work is required to refine these models for use on soils with very low and very high B_d classes which fall outside those typically found in Ireland. Finally, our goal is to encode the decision tree and the linear equation into a mobile application, in order to enable multiple user types to perform B_d prediction more quickly, on site, and in a user friendly manner.

Acknowledgements

The authors thank Irene Marongiu for her help with the physical and chemical soil analyses. SQUARE project has received founding from the Irish Department of Agriculture (DAFM) Research Stimulus Fund (RSF). Task 3 output.

References

Alvarez, S.A., 2002. An exact analytical relation among recall, precision, and classification accuracy in information retrieval. Tech. Rep. BCCS-02-01, Computer Science Department, Boston College.

Ampoorter, E., Goris, R., Cornelis, W. M., Verheyen, K., 2007. Impact of mechanized logging on compaction status of sandy forest soils. Forest Ecol. Manag. 241, 162–174.

Armindo, R.A., Wendroth, O. 2016. Physical soil structure evaluation based on hydraulic energy functions. Soil Sci. Am. J. 80, 1167–1180.

Attou, F., Bruand, A., Bissonnais, Y. L., 1998. Effect of clay content and silt-clay fabric on stability of artificial aggregates. Eur. J. Soil Sci. 49, 569–577.

Ball, B. C., Munkholm, L. J., 2015. Visual Soil Evaluation: Realizing Potential Crop Production with Minimum Environmental Impact. CAB International. London.

Batey, T., 2000. Soil profile description and evaluation. In: Smith, K., Mullins, C. (Eds.), Soil and environmental analysis, physical methods, 2nd edition. Marcel Dekker, New York. pp. 595–628

Bhargava, N., Sharma, G., Bhargava, R., Mathuria, M., 2013. Decision tree analysis on j48 algorithm for data mining. Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering. 3, 1114–1119.

Boix-Fayos, C., Calvo-Cases, A., Imeson, A.C., 2001. Influence of soil properties on the aggregation of some Mediterranean soils and the use of aggregate size and stability as land degradation indicators. Catena 44, 47–67.

Bronick, C. J., Lal, R. 2005. Soil structure and management: a review. Geoderma. 124, 3–22.

Caravaca, F., Hernandez, T., Garcia, C., Roldan, A., 2002. Improvement of rhizosphere aggregate stability of afforested semiarid plant species subjected to mycorrhizal inoculation and compost addition. Geoderma. 108, 133–144.

Collins, J. F., Larney, F. J., Morgan, M. A., 2004. Climate and soil management, in: Keane, T., Collins, J. F. (Eds.), Climate, Weather and Irish Agriculture. Working group on Applied Agricultural Meteorology (AGMET), Dublin, pp. 119–160.

Creamer, R., Simo, I., Reidy, Carvalho, J., Fealy, R., Hallett, S., Jones, R., Holden, A., Holden, N., Hannam, J., Massey, P., Mayr, T., McDonald, E., O'Rourke, S., Sills, P., Truckell, I., Zawadzka, J., Schulte, R., 2014. Irish Soil Information System. (2007-S-CD-1-S1) EPA STRIVE Programme 2007–2013, Synthesis Report.

Dam, R.F., Mehdi, B.B., Burgess, M.S.E., Madramootoo, C.A., Mehuys, G.R., Callum, I.R., 2005. Soil bulk density and crop yield under eleven consecutive years of corn with different tillage and residue practices in a sandy loam soil in central Canada. Soil Till. Res. 84, 41–53.

Deconchat, M., 2001. Effets des techniques d'exploitation forestie`res sur l'e´tatde surface du sol. Ann. For. Sci. 58, 653–661.

Dexter, A. R., 1988. Advances in characterization of soil structure, Soil Till. Res. 11, 199–238.

Dexter, A. R., 2002. Soil structure: the key to soil function, in: Pagliai, M., Jones, R., (Eds.), Sustainable land Management – Environmental Protection, a soil Physical Approach. Advances in Geoecology 35, Clearance Center, Inc., Reiskirchen, pp. 57–69.

Ellert, B. H., Bettany, J. R., 1995. Calculation of organic matter and nutrient stored in soils under constrain management regimes. Can. J. Soil Sci. 75, 529–538.

Emmet-Booth, J. P., Forristal, P. D., Fenton, O., Ball, B. C., Holden, N. M., 2016. A review of visual soil evaluation techniques for soil structure. Soil Use Manage. 32, 623–634.

Epron, D., Plain, C., Ndiaye, F. K., Bonnaud, P., Pasquier, C., Ranger, J., 2016. Effects of compaction by heavy machine traffic on soil fluxes of methane and carbon dioxide in a temperate broadleaved forest. Forest Ecol. Manag. 382, 1–9.

FAO, 2006. Guidelines for Soil Description, Fourth Edition. FAO, Rome.

FAO, ISRIC, ISSS, 1998. World Reference Base for Soil Resources. World Soil Resources Reports 80, FAO Rome.

Fenton, O., Vero, S., Ibrahim, T. G., Murphy, P. N. C., Sherriff, S. C., ÓHuallacháin, D., 2015. Consequences of using different soil texture determination methodologies for soil physical quality and unsaturated zone time lag estimates. J Contam. Hydrol. 182, 16–24.

Fenton, O., Vero, S., Schulte, R.P.O., O'Sullivan, L., Bondi, G., Creamer, R.E., 2017. Application of Dexter's soil physical quality index: an Irish case study Irish J. Agr. Food. 56, 45–53.

Geissen, V., Kampichler, C., López-de Llergo-Juárez, J. J., Galindo-Acántara, A., 2007. Superficial and subterranean soil erosion in Tabasco, tropical Mexico: development of a decision tree modeling approach. Geoderma. 139, 277–287.

Håkansson, I., Lipiec, J., 2000. A review of the usefulness of relative bulk densityvalues in studies of soil structure and compaction. Soil Till. Res. 53, 71–85.

Haynes, R.J., Naidu, R., 1998. Influence of lime, fertilizer and manure applications on soil organic matter content and soil physical conditions: a review. Nutr. Cycl. Agroecosyst. 51, 123–137.

Henderson, B. L., Bui, E. N., Moran, C. J., Simon, D. A. P., 2005. Australia-wide predictions of soil properties using decision trees. Geoderma. 124, 383–398.

Holden, N., Brereton, A.J., 2004. Definition of agroclimatic regions in Ireland using hydro-thermal and crop yield data. Agr. Forest Meteorol. 122, 175–191.

IUSS Working Group WRB, 2006. World Reference Base for Soil Resources – a framework for international classification, correlation and communication. World Soil Resources Reports 103. FAO, Rome.

Jin, R., Li, X., Che, T., 2009. A decision tree algorithm for surface soil freeze/thaw classification over China using SSM/I brightness temperature. Remote Sens. Environ. 113, 2651–2660.

Karlen, D.L., 2004. Soil quality as an indicator of sustainable tillage practices. Soil Till. Res. 78, 129–130.

Kaur, R., Sanjeev, K., Gurung, H., 2002. Apedo-transfer function (PTF) for estimating soil bulk density from basic soil data and its comparison with existing PTFs. Aust. J. Soil Res. 40, 847–857.

Kay, B.D., 1998. Soil structure and organic carbon: a review. In: Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.A. (Eds.), Soil Processes and the Carbon Cycle. CRC Press, Boca Raton, pp. 169-197.

Kay, B.D., Hajabbasi, M.A., Ying, J., Tollenaar, M., 2006. Optimum versus non-limiting water contents for root growth, biomass accumulation, gas exchange and the rate of development of maize (Zea mays L.). Soil Till. Res. 88, 42–54.

Lee, K. E., Foster, R. C., 1991. Soil fauna and soil structure. Aust. J. Soil Res. 29, 745–775.

Leonavičiutė, N., 2000. Predicting soil bulk and particle densities by pedotransfer functions from existing soil data in Lithuania, Geografijosmetraštis. 33, 317–330.

Logsdon, S.D., Karlen, D.L., 2004. Bulk density as a soil quality indicator during conversion to no-tillage. Soil Till. Res. 78, 143–149.

Moncada, M.P., Ball, B.C., Gabriels, D., Lobo, D. and Cornelis, W.M., 2015. Evaluation of soil physical quality index S for some tropical and temperate medium-textured soils. Soil Sci. Soc. America J. 79, 9–19

Mueller, L., Kay, B. D., Deen, B., Hu, C., Zhang, Y., Wolff, M., Eulenstein, F., Schindler, U., 2009. Visual assessment of soil structure: Part II. Implications of tillage, rotation and traffic on sites in Canada, China and Germany. Soil Till. Res. 103, 188–196.

Pachepsky, Y. A., Rawls, W. J., 2003. Soil structure and pedotransfer functions. Eur. J. Soil Sci. 54, 443–452.

Pagliai, M., Vignozzi, N., 2002. The soil pore system as an indicator of soil quality, in: Pagliai, M., Jones, R., (Eds.), Sustainable land Management – Environmental Protection, a soil Physical Approach. Advances in Geoecology 35, Clearance Center, Inc., Reiskirchen, pp. 71–82.

Quinlan, J. R., 1993. C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers.

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.

Russell, E. W., 1975. Soil Conditions and Plant Growth 10th edition. Ed. Longman, London.

Saffih-Hdadi, K., Défossez, P., Richard, G., Cui, Y. J., Tang, A. M., Chaplain, V., 2009. A method for predicting soil susceptibility to the compaction of surface layers as a function of water content and bulk density, Soil Till. Res. 105, 96–103.

Schoeneberger, P.J., D.A. Wysocki, E.C. Benham, and Soil Survey Staff. 2012. Field book for describing and sampling soils, Version 3.0. Natural Resources Conservation Service, National Soil Survey Center, Lincoln, NE.

Schulten, H.R., Leinweber, P., 2000. New insights into organic– mineral particles: composition, properties and models of molecular structure. Biol. Fertil. Soils. 30, 399– 432.

Shipitalo, M. J., Le Bayon, R. C., 2004. 10 Quantifying the Effects of Earthworms on Soil Aggregation and Porosity, in: Edwards, C.A. (Eds.) Earthworm ecology 2nd edition, Boca Raton, pp 183–199.

Soil Survey Division Staff, 1993. Soil Survey Manual. Agricultural Handbook N 18. USDA Natural Resources Conservation Service, Washington D.C.

Stevenson, F.J. 1994. Humus chemistry. genesis, composition, reactions. 2nd edition. Wiley Interscience, New York.

Teepe, R., Brumme, R., Beese, F., Ludig, B., 2004. Nitrous oxide emission and methane consumption following compaction of forest soils. Soil Sci. Soc. Am. J. 68, 605–611.

Templ, M., Kowarik, A., & Filzmoser, P., 2011. Iterative stepwise regression imputation using standard and robust methods. Comput. Stat. Data An. 55, 2793–2806.

Xue, D., De Baets, B., Van Cleemput, O., Hennessy, C., Berglund, M., Boeckx, P., 2013. Classification of nitrate polluting activities through clustering of isotope mixing model outputs. J. Environ. Qual. 42, 1486–1497.

Xue, D., Pang, F., Meng, F., Wang, Z., Wu, W., 2015. Decision-tree-model identification of nitrate pollution activities in groundwater: A combination of a dual isotope approach and chemical ions. J. Contam. Hydrol. 180, 25–33.

Table 1. Selection of soil structure field descriptors described by FAO, Guidelines for Soil Description, 2006.

Descriptor	Title	Description
1	Humose	This is an estimation of the degree of humification of the organic material. Surveyor must provide a positive or affirmative answer to being humose (this descriptor was recorded as a presence/absence in the database).
2	Soil Consistency	The strength with which soil materials are held together. It provides a means of describing the degree of cohesion and adhesion between the soil particles as related to the resistance of the soil to deform or rupture. It includes soil properties such as friability, plasticity, stickiness and resistance to compression. It changes with soil moisture and is highly related to the percentage of clay and OM in the soil.
3	Stickiness	It is the capacity of the soil to adhere to an object. It is evaluated pressing a small amount of wet soil between thumb and forefinger to see if it will stick to fingers.
4	Plasticity	The ability of soil material to retain a shape after pressure deformation. It is evaluated by rolling a small amount of wet soil between the hand palms until it forms a long, round strip like a wire about 3 mm thick.
Soil structure* is described as the combination of (5, 6, 7)		
5	Structure Grade	It describes the level of development of soil structure. It is expressed as the differential between cohesion within aggregates and adhesion between aggregates. It is evaluated in relation to the arrangement of the aggregates and to the strength necessary to break them.
6	Structure Type	It describes the form or shape of individual aggregates and is directly correlated with the pedogenic formation.
7	Structure Size	It describes the average size of individual aggregates. Different classes may be recognized in relation to the type of soil structure from which they come.
Voids** is described as the combination of (8, 9)		
8	Voids Abundance	An indication of the total volume of voids measured by area and was recorded as the percentage of the surface occupied by pores.
9	Voids Size	The diameter of voids and was recorded in mm.
10	Fissures size	The diameter of fissures and was recorded in mm.
11	Macropores size	The diameter of macropores, which are described as bigger void, mostly determined by plant roots, and by zoological exploration. Macropores were recorded in mm.
*Soil Structure: It refers to the spatial disposition of aggregates which are the result of the aggregation of single particles such us sand, silt and clay. Size, shape and arrangement of these solids and voids, determining the porosity and the capacity to retain fluids and inorganic and organic substances can occur in different patterns, resulting in different soil structures (Bronick et Lal, 2005).** Voids: Include all the pore space present in the soil. It is closely related to the porosity and is a good indicator of soil compactness. It is evaluated as presence/absence data. Voids were described in terms of size and abundance.		

Table 2. Decision tree model (Model 1); performances. RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

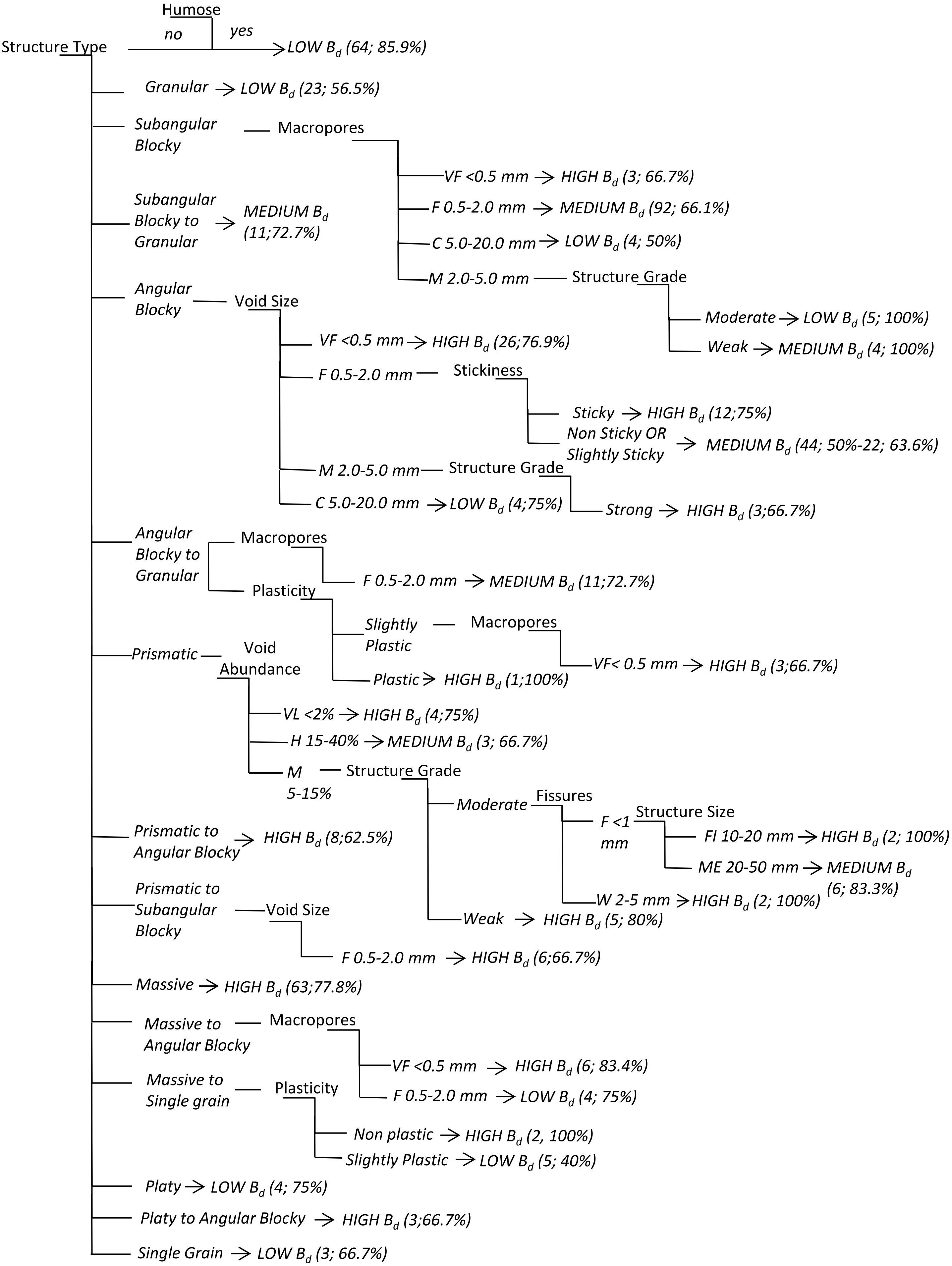
	Performance with cross validation				Performance on training set			
	N. of instances	Accuracy	RMSE	MAE	N. of instances	Accuracy	RMSE	MAE
Correctly Classified Instances	283	60.08 %	0.44	0.32	335	71.12%	0.37	0.27
Incorrectly Classified Instances	188	39.91 %			136	28.87 %		
	N. of instances	Precision	Recall	F-measure	N. of instances	Precision	Recall	F-measure
Low B_d class	137	0.70	0.54	0.60	137	0.75	0.65	0.69
Medium B_d class	178	0.53	0.63	0.58	178	0.64	0.73	0.68
High B_d class	156	0.62	0.62	0.62	156	0.76	0.74	0.75
Weighted Average		0.61	0.60	0.60		0.72	0.71	0.71

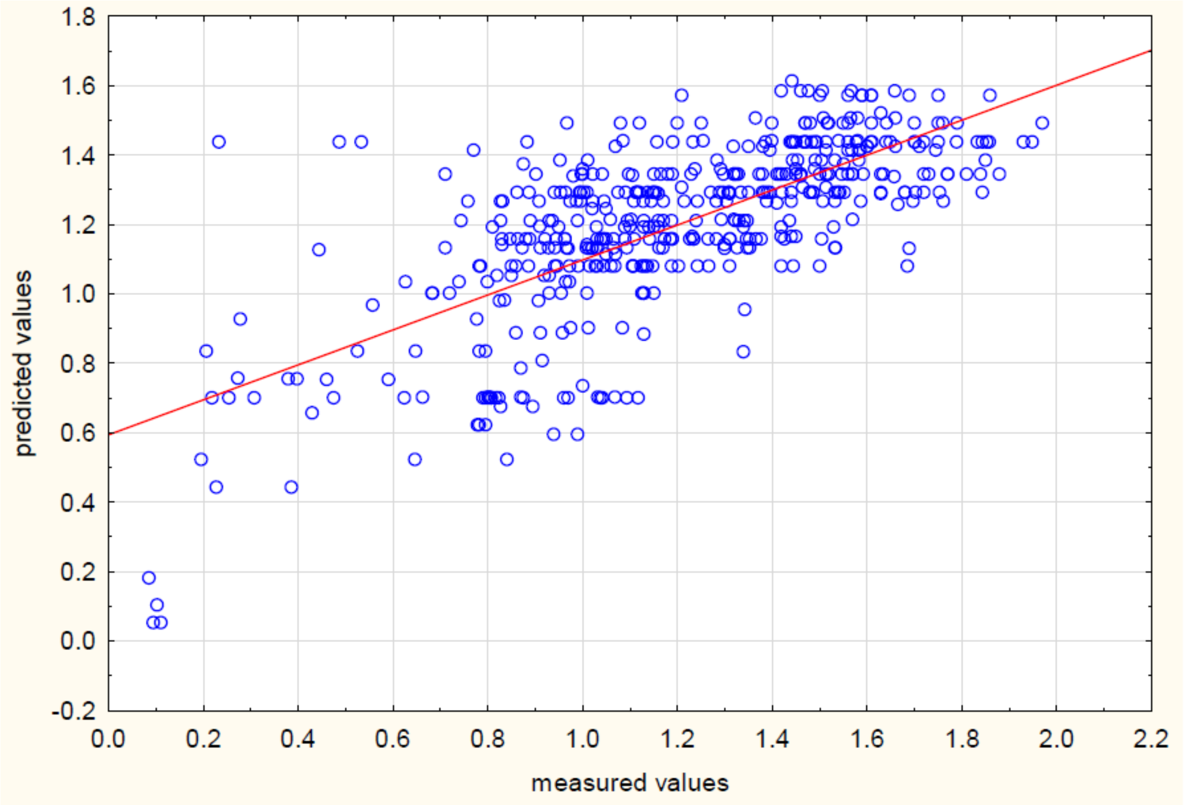
Table 3. Decision tree model (Model 1); confusion matrix.

	Classes classified by decision tree model (N. of instances=471)		
	a	b	c
Low B_d class (a)	**74**	46	17
Medium B_d class (b)	25	**112**	41
High B_d class (c)	7	52	**97**

Table 4. Linear regression model (Model 2); performances. RMSE: Root Mean Squared Error; MAE: Mean Absolute Error

	Performance with cross validation				Performance on training set			
	N. of instances	Correlation coefficient	RMSE	MAE	N. of instances	Correlation coefficient	RMSE	MAE
Instances	471	0.65	0.27	0.21	471	0.71	0.25	0.20


Structure Type
Humose
no
yes
LOW B_d (64; 85.9%)
Granular
LOW B_d (23; 56.5%)
Subangular Blocky
Macropores
VF <0.5 mm
HIGH B_d (3; 66.7%)
F 0.5-2.0 mm
MEDIUM B_d (92; 66.1%)
C 5.0-20.0 mm
LOW B_d (4; 50%)
M 2.0-5.0 mm
Structure Grade
Moderate
LOW B_d (5; 100%)
Weak
MEDIUM B_d (4; 100%)
Subangular Blocky to Granular
MEDIUM B_d (11;72.7%)
Angular Blocky
Void Size
VF <0.5 mm
HIGH B_d (26;76.9%)
F 0.5-2.0 mm
Stickiness
Sticky
HIGH B_d (12;75%)
Non Sticky OR Slightly Sticky
MEDIUM B_d (44; 50%-22; 63.6%)
M 2.0-5.0 mm
Structure Grade
Strong
HIGH B_d (3;66.7%)
C 5.0-20.0 mm
LOW B_d (4;75%)
Angular Blocky to Granular
Macropores
F 0.5-2.0 mm
MEDIUM B_d (11;72.7%)
Plasticity
Slightly Plastic
Macropores
VF< 0.5 mm
HIGH B_d (3;66.7%)
Plastic
HIGH B_d (1;100%)
Prismatic
Void Abundance
VL <2%
HIGH B_d (4;75%)
H 15-40%
MEDIUM B_d (3; 66.7%)
M 5-15%
Structure Grade
Moderate
Fissures
F <1 mm
Structure Size
FI 10-20 mm
HIGH B_d (2; 100%)
ME 20-50 mm
MEDIUM B_d (6; 83.3%)
W 2-5 mm
HIGH B_d (2; 100%)
Weak
HIGH B_d (5; 80%)
Prismatic to Angular Blocky
HIGH B_d (8;62.5%)
Prismatic to Subangular Blocky
Void Size
F 0.5-2.0 mm
HIGH B_d (6;66.7%)
Massive
HIGH B_d (63;77.8%)
Massive to Angular Blocky
Macropores
VF <0.5 mm
HIGH B_d (6; 83.4%)
F 0.5-2.0 mm
LOW B_d (4; 75%)
Massive to Single grain
Plasticity
Non plastic
HIGH B_d (2, 100%)
Slightly Plastic
LOW B_d (5; 40%)
Platy
LOW B_d (4; 75%)
Platy to Angular Blocky
HIGH B_d (3;66.7%)
Single Grain
LOW B_d (3; 66.7%)


1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
-0.2
predicted values
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
1.8
2.0
2.2
measured values

Figures Captions:

Figure 1. Decision tree model for predicting bulk density classes (Model 1): LOW B_d <1 g cm^{-3}; MEDIUM B_d 1-1.4 g cm^{-3}; HIGH B_d >1.4 g cm^{-3}. The number of cases classified for the rule and the percentage of accuracy are reported.

~~**Figure 2**. Decision tree model for predicting bulk density classes (Model 1): MEDIUM B_d 1-1.4 g cm^{-3}. The number of cases classified for the rule and the percentage of accuracy are reported.~~

~~**Figure 3**. Decision tree model for predicting bulk density classes (Model 1): HIGH B_d >1.4 g cm^{-3}. The number of cases classified for the rule and the percentage of accuracy are reported.~~

Figure ~~4~~2. Relationship between measured bulk density (B_d) and predicted bulk density values for the linear regression model (Model 2). B_d values are reported in g cm^{-3}.

Figures Captions:

Figure 1. Decision tree model for predicting bulk density classes (Model 1): LOW B_d <1 g cm^{-3}; MEDIUM B_d 1-1.4 g cm^{-3}; HIGH B_d >1.4 g cm^{-3}. The number of cases classified for the rule and the percentage of accuracy are reported.

Figure 2. Relationship between measured bulk density (B_d) and predicted bulk density values for the linear regression model (Model 2). B_d values are reported in g cm^{-3}.

Supplementary Material Table 1. Field descriptors choice options. Abbreviation and definitions.

Field Descriptor	**Full Name**	**Abbreviation/Definition**
1. Humose		
	No Humic	NH
	Humic	H
2. Soil Consistency		
	Loose	LO: Non-coherent.
	Very friable	VFR: Crushes under very gentle pressure, but coheres when pressed together.
	Friable	FR: Crushes under gentle to moderate pressure between thumb and forefinger and coheres when pressed together.
	Firm	FI: Crushes under moderate pressure between thumb and forefinger; resistance is noticeable.
	Very firm	VFI: Crushes under strong pressure; barely crushable between thumb and forefinger.
	Extremely firm	EFI: Crushes under only very strong pressure; cannot be crushed between thumb and forefinger.
3. Stickiness		
	Non-sticky	NST: No soil material adheres to thumb and finger after release of pressure.
	Slightly sticky	SST: Soil material adheres to thumb and finger after release of pressure, but it is easily removed.
	Sticky	ST: Soil material adheres to thumb and finger after release of pressure, and tends to stretch and pull apart rather than coming away from each digit.
	Very sticky	VST: Soil material adheres strongly to thumb and finger after release of pressure, and stretches when fingers are separated.
4. Plasticity		
	Non-plastic	NPL: No wire is formable.
	Slightly plastic	SPL: Wire formable but breaks immediately if bent into a ring; deformation by slight force.
	Plastic	PL: Wire formable but breaks if bent into a ring; deformation by slight to moderate force.
	Very plastic	VPL: Wire formable and can be bent into a ring; deformation by moderately strong to strong force.
5. Structure Grade		
	Weak	WE; Aggregates barely visible in situ and only weak arrangement of natural surfaces that breaks when gently disturbed.
	Moderate	MO: Aggregates are visible in situ and there is a distinct arrangement of material. When disturbed it breaks into a mixture of entire and broken aggregates.
	Strong	ST: Aggregates are clearly visible in situ and there is prominent arrangement of material. When disturbed it breaks into distinct whole aggregates.

	Weak to Moderate	WM: Show weak and moderate properties
	Moderate to strong	MS: show moderate and strong properties.
6. Structure Type		
	Granular	GR
	Granular to Single Grain	GRSG
	Angular Blocky	AB
	Angular Blocky to Granular	ABGR
	Subangular Blocky	SB
	Subangular Blocky to Granular	SBGR
	Prismatic	PR
	Prismatic to Angular Blocky	PRAB
	Prismatic to Subangular Blocky	PRSB
	Platy	PL
	Platy to Angular Blocky	PLAB
	Platy to Subangular Blocky	PLSB
	Single Grain	SG
	Massive	MA
	Massive to Single Grain	MASG
	Massive to Angular Blocky	MAAB
7. Structure Size*	*Structure size is evaluated taking into account the main categories of structure type in the following order.	
Blocky	Coarse	C : 20-50mm
Prismatic/Columnar	Coarse	C : 50-100mm
Granular/Platy	Coarse	C : 5-10mm
Prismatic/Columnar	Fine	F : 10-20mm

Granular/Platy	Fine	F : 1-2mm
Blocky	Fine	F : 5-10mm
Blocky	Medium	M : 10-20mm
Prismatic/Columnar	Medium	M : 20-50mm
Granular/Platy	Medium	M : 2-5mm
Prismatic/Columnar	Very Coarse	VC : > 100mm
Granular/Platy	Very Coarse	VC: > 10mm
Blocky	Very Coarse	VC: > 50mm
Prismatic/Columnar	Very Fine	VF: < 10mm
Granular/Platy	Very Fine	VF: < 1mm
Blocky	Very Fine	VF: < 5mm
8. Voids Abundance		
	Very Low	VL: <2%
	Low	L: 2-5%
	Medium	M: 5-15%
	High	H: 15-40%
	Very High	VH: >40%
9. Voids Size		
	Very Fine	VF: < 0.5mm
	Fine	F: 0.5-2 mm
	Medium	M: 2-5 mm
	Coarse	C : 5-20 mm
	Very Coarse	VC: 20-50mm
10. Fissures Size		
	Fine	F: < 1mm
	Medium	M: 1-2 mm
	Wide	W: 2-5 mm

	Very Wide	VW: 5-10 mm
	Extremely Wide	EW: > 10mm
11. Macropores Size		
	Very Fine	VF: < 0.5mm
	Fine	F : 0.5-2 mm
	Medium	M : 2-5 mm
	Coarse	C : 5-20 mm
	Very Coarse	VC : 20-50mm

Supplementary Material Table 2. Total list of rules for Model 1.

Rule n.	Bulk density Class	Description
1	Low	Humose=Yes
2	Low	Humose=No AND Structure type=Granular
3	Low	Humose=No AND Structure type=Subangular Blocky AND Macropores= M 2.0-5.0 mm AND Structure Grade=Moderate
4	Low	Humose=No AND Structure type=Subangular Blocky AND Macropores= C 5.0-20 mm
5	Low	Humose=No AND Structure type=Angular Blocky AND Void Size= C 5.0-20 mm
6	Low	Humose=No AND Structure type=Angular Blocky AND Void Size= F 0.5-2.0 mm AND Stickiness= Very Sticky
7	Low	Humose=No AND Structure type=Single Grain
8	Low	Humose=No AND Structure type=Massive to Angular Blocky AND Macropores= F 0.5-2.0 mm
9	Low	Humose=No AND Structure type=Prismatic to Subangular Blocky AND Void Size= M 2.0-5.0 mm
10	Low	Humose=No AND Structure type=Platy
11	Low	Humose=No AND Structure type=Massive to Single grain AND Plasticity= Slightly Plastic
12	Medium	Humose=No AND Structure type=Subangular Blocky AND Macropores F 0.5-2 mm
13	Medium	Humose=No AND Structure type=Subangular Blocky AND Macropores M 2.0-5.0 mm AND Structure Grade= Weak
14	Medium	Humose=No AND Structure type=Angular Blocky AND Void size= M 2-5 mm AND Structure Grade= Moderate
15	Medium	Humose=No AND Structure type=Angular Blocky AND Void size= F 0.5-2 mm AND Stickiness= Non Sticky
16	Medium	Humose=No AND Structure type=Angular Blocky AND Void size= F 0.5-2 mm AND Stickiness= Slightly Sticky
17	Medium	Humose=No AND Structure type=Prismatic AND Void abundance= H 15-40%
18	Medium	Humose=No AND Structure type=Prismatic AND Void abundance= M 5-15% AND Structure Grade= Moderate AND Fissures= F <1 mm AND Structure Size= ME 20-50 mm
19	Medium	Humose=No AND Structure type=Prismatic AND Void abundance= M 5-15% AND Structure Grade= Moderate AND Fissures= F <1 mm AND Structure Size= CO 50-100 mm
20	Medium	Humose=No AND Structure type=Prismatic AND Void abundance= M 5-15% AND Structure Grade= Strong

21	Medium	Humose=No AND Structure type=Prismatic to Subangular Blocky AND void size=VF <0.5 mm
22	Medium	Humose=No AND Structure type= Subangular Blocky to Granular
23	Medium	Humose=No AND Structure type= Angular Blocky to Granular AND Macropores= F 0.5-2.0 mm
24	Medium	Humose=No AND Structure type= Angular Blocky to Granular AND Macropores= M 2.0-5.0 mm
25	High	Humose=No AND Structure type= Massive
26	High	Humose=No AND Structure type= Subangular Blocky AND Macropores= VF <0.5 mm
27	High	Humose=No AND Structure type= Angular Blocky AND Void size= M 2.0-5.0 mm AND Structure grade= Strong
28	High	Humose= No AND Structure type= Angular blocky AND Void size=F 0.5-2 mm AND Stickiness= Sticky
29	High	Humose=No AND Structure type= Angular blocky AND Void size= VF <0.5 mm
30	High	Humose=No AND Structure type= Prismatic AND Void abundance= M 5-15% AND Structure grade= Moderate AND Fissures= F <1mm AND Structure size= FI 10-20 mm
31	High	Humose=No AND Structure type= Prismatic AND Void abundance= M 5-15% AND Structure grade= Moderate AND Fissures= W 2.0-5.0 mm
32	High	Humose=No AND Structure type= Prismatic AND Void abundance= M 5-15% AND Structure grade= Weak
33	High	Humose=No AND Structure type= Prismatic AND Void Abundance=VL
34	High	Humose=No AND Structure type= Prismatic to Angular Blocky
35	High	Humose= No AND Structure type= massive to Angular Blocky AND Macropores=VF < 0.5 mm
36	High	Humose=No AND Structure type= Prismatic to Subangular Blocky AND Void size= F 0.5-2 mm
37	High	Humose=No AND Structure type= Platy to Angular Blocky
38	High	Humose=No AND Structure type=Massive to Single grain AND Plasticity= Non Plastic
39	High	Humose=No AND Structure Type=Angular Blocky to Granular AND Plasticity=Slightly plastic AND Macropores= VF< 0.5 mm
40	High	Humose=No AND Structure type= Angular Blocky to Granular AND Plasticity=Non plastic
41	High	Humose=No AND Structure type= Angular Blocky to Granular AND Plasticity= Plastic

.