

Research article

The effects of climate and soil properties on the magnitude of the visual soil quality indicators: a logistic regression approach

Fernando Teixeira*

MED–Mediterranean Institute for Agriculture, Environment and Development, Institute for Advanced Studies and Research, Universidade de Évora, Pólo da Mitra, Ap. 94, 7006-554 Évora, Portugal

* **Correspondence:** Email: fteixeir@uevora.pt.

Abstract: Understanding how different climates and soil properties affect the soil processes requires quantifying these effects. Visual soil quality indicators have been proposed to assess the robustness of the soil processes and infer their ability to function. The scores of the visual soil quality indicators covary with climate features and soil properties, and their magnitude is different in acid-to-neutral and alkaline soils. These variables show collinearities and interactions, and the assessment of the individual effect of each variable on the scores of the visual indicators and the selection of the best set of explanatory variables can only be made with a definite set of variables. Logistic regression was used to calculate the effects of six climate variables and four soil properties, and their interactions, on the scores of eight visual soil quality indicators. Simple models featuring climate and soil variables explained a substantial part of the variation of the visual indicators. Models were fitted for each visual indicator for acid-to-neutral and alkaline soils. The sample size needed was calculated, and the method and its validity were discussed. For two possible outcomes, the sample size using the events per variable (EPV) criterium ranges between 62 and 183 observations, while using one variable and a variance inflation factor, it ranges between 22 and 234. Except for the model of soil structure and consistency for acid-to-neutral soils, with a C statistic of 0.67, all others had acceptable to excellent discrimination. The models built are adequate, for example, for the large-scale spatial outline of the soil health indices, to couple with soil morphological-dependent pedotransfer functions, and so on. Future models should consider (test) other explanatory variables: other climate variables and indices, other soil properties and soil management practices.

Keywords: New Zealand visual soil assessment method; VSA; logistic regression; models; interactions; climate; soil properties; soil reaction

1. Introduction

Soil morphological features and scoring them are the basis of the visual soil assessment methods. The premise of the classification systems is the existence of correlations between the magnitude (score) of these soil morphological features (visual indicators) and the soil structure adequacy for sowing or growing a crop. Their empirical nature, nonetheless, may require an experienced observer [1], while others, with little training and the use of standards to guide the classification, can be performed by anybody [2,3].

The inexistence of an operational definition of soil structure precludes us from inferring the soil structure status based solely on measured soil properties. Conversely, the visual soil assessment methods set forth the procedures for practical, inexpensive and quick assessment of the soil structure. Although there is little doubt the visual indicators provide an authoritative classification of the ability of the soil to function, our understanding of the soil processes that govern the extent of the features observed is conjectural. How climate and soil properties affect the soil processes and the score of these visual indicators is, with few exceptions, practically an incognita. This raises some pertinent questions: “Can we measure the correlation between the score of the visual indicators and climate and soil properties? Can we measure the effects of these variables on the scores?”

To answer these questions, let us first focus on the following: the scores of the visual indicators are ordinal, i.e. there is a natural order for the scores (e.g., “poor”, “moderate” or “good”). Although the visual soil assessment methods may be robust, with different observers consistently classifying the soil structure into the same category [4], no inference can be made on the existence of a scale or the continuity of the score of the visual indicators. The inexistence of a scale between two or more scores- no physical quantity is measured- makes it impossible to calculate arithmetic means, variances or the linear correlations between the scores and continuous numerical variables (e.g. soil organic matter). However, it is possible to study the correlation between the scores of visual indicators and other variables by studying, for example, their monotonic relationship; one such approach is by employing Spearman’s rank-order correlation [5]. Regarding climate and soil properties, we were able to establish that the correlations between the scores of the visual indicators and the soil properties and climate variables, and the correlations between soil properties and climate variables, are different and depend on the soil reaction (unpublished work). This study also allowed us to establish the relative importance of the variables on the magnitude of the score of the visual indicators (the relative strength of the effect) in acid-to-neutral and alkaline soils. In general terms, regarding the soil properties measured, organic matter (soil organic matter (SOM) and labile organic carbon (LOC)) is the most important variable in acid-to-neutral soils and pH in alkaline. Regarding the climate variables, we observed, for most indicators, that the indices (aridity and continentality indices and net primary production) dominated other climate features but, as we will see in the present work, interactions between climate features (e.g. mean annual temperature and precipitation) can explain much of the variation of the scores. Second, let us focus on the following: the visual indicators should be sensitive to changes induced in the soil by different soil management practices. The indicators “soil structure and consistency” and “soil porosity”, both visual indicators of the “New Zealand visual soil assessment method” (VSA),

were found to be particularly sensitive to the soil management practices, irrespective of soil type and climate [6]. Five other visual indicators of the VSA method were not as sensitive to soil management practices- with few exceptions, the results were not statistically significant- but that is not to say that at different locations, the interactions of soil management practices with soil properties and climate will not increase the sensitivity of those indicators. Third, given these correlations and effects on the scores of the visual indicators, can we predict the scores? Score prediction is a classification problem: assigning observations to discrete categories. Thus, in the present work, we propose logistic regression as a statistical model to calculate the strength of the effect of variables and interactions on the scores of the visual soil quality indicators. Logistic regression presents numerous advantages over the alternatives; for example, the linear discriminant analysis would require the assumption that the populations are multivariate normally distributed and with equal covariance matrices- those assumptions are not needed with logistic regression [7]. The models thus created, using simple predictors which can be obtained from readily available spatial data, would allow, for example, to estimate the spatial distribution of soil quality indices based on visual soil quality indicators [2,3,8,9] or to improve the prediction performance of pedotransfer functions that depend on soil morphological features [10].

In this study, the visual soil quality indicators used are those of the New Zealand Visual Soil Assessment method [2]. The method comprises the observation of soil structure (a drop-shatter test), porosity (observable macroporosity), colour (a differential observation between managed and undisturbed soil), the presence of tillage pan, earthworm counts in a given time, susceptibility to erosion and surface ponding. The stability of the aggregates in water was also observed. The magnitude of each visual soil quality indicator falls into one of three categories, “poor”, “moderate” and “good”. We hypothesize that the visual indicators scores can be modelled with logistic regression, based on the premise (suggested by our findings) that climate features and soil properties, explain much of the variation in the scores of the visual indicators.

The main objectives were: i) to create logistic regression models of the score of the indicators of the New Zealand visual soil assessment method (VSA), and of the aggregate stability in water, with climate and soil variables as predictors; ii) to assess the contribution of each predictor by fitting single-variable models, interactions (models with two main effects and an interaction term) and their combinations (selected models); iii) to calculate the sample size needed to create models for each visual soil quality indicator, using two simple approaches: the number of events per variable (EPV) and using the effect of one variable on the scores, in the presence of other variables, and using a variance inflation factor.

2. Materials and Methods

2.1. Dataset

2.1.1. Overview

This study evaluates logistic regression models of the scores of visual soil quality indicators as a function of climate variables, soil properties and their interactions. Besides, the sample size needed to fit the models was calculated using two different approaches: the events per variable (EPV) approach and the effect of an explanatory variable and a variance inflation factor at a given significance level and power.

The dataset used to fit the models was recorded in the spring/summer of 2016 at 256 locations in 13 study sites across eight pedoclimatic zones in Europe and China. Three representative sampling points were used per location (the number of points increased if a significant variation was observed). Each record comprises eight visual soil quality indicators, four soil properties (seven variables) and six climate variables. The Pearson correlation coefficients between all continuous variables are in Table 2a and 2b in Supplemental Materials.

2.1.2. Visual Soil Assessment

The visual soil quality indicators assessed are identified in Table 1. Except for “soil stability” (slake test), which is described in [11] and was adapted to three categories, all others are described in Shepherd’s New Zealand VSA method [2]. A brief description of the visual indicators and standards used is presented in Table 1 in Supplemental Materials.

Table 1. Visual soil quality indicators and frequency of the scores.

| VSA indicator | All records | | | | Acid to neutral soils (pH ≤ 7) | | | | Alkaline soils (pH > 7) | | | |
|---|-------------|------|-----|------|--------------------------------|------|-----|------|-------------------------|------|-----|------|
| | n | Poor | Mod | Good | n | Poor | Mod | Good | n | Poor | Mod | Good |
| <i>Soil structure and consistency</i> | 256 | 33 | 127 | 104 | 133 | 16 | 61 | 56 | 123 | 16 | 61 | 46 |
| <i>Soil porosity</i> | 256 | 30 | 120 | 114 | 133 | 10 | 54 | 69 | 123 | 20 | 59 | 44 |
| <i>Soil stability (slake test)</i> | 255 | 34 | 142 | 87 | 132 | 8 | 78 | 46 | 123 | 26 | 57 | 40 |
| <i>Soil colour</i> | 256 | 32 | 129 | 103 | 133 | 17 | 59 | 57 | 123 | 14 | 64 | 45 |
| <i>Presence of cultivation pan</i> | 256 | 27 | 140 | 97 | 133 | 8 | 63 | 62 | 123 | 19 | 69 | 35 |
| <i>Earthworm count</i> | 256 | 133 | 88 | 43 | 133 | 60 | 49 | 24 | 123 | 65 | 39 | 19 |
| <i>Surface ponding (under cropping)</i> | 232 | 4 | 50 | 186 | 114 | 0 | 21 | 93 | 118 | 4 | 29 | 85 |
| <i>Susceptibility to wind and water erosion</i> | 256 | 11 | 56 | 189 | 133 | 4 | 25 | 104 | 123 | 7 | 31 | 85 |

2.1.3. Soil properties

The dataset comprises the following measured soil properties: soil organic matter (SOM) (different methods according to the location: Walkley-Black method, Tjurin method and loss on ignition), soil labile organic carbon (LOC) (following the protocol proposed by Weil et al. [12], with a diluted solution of 0.02 M of KMnO₄), soil pH (measured in water, 1:1 ratio, [13]), soil penetration resistance (Eijkelkamp penetrometers to a depth of 0.40 m) and soil particle-size fractions (different methods according to the location: sieving and sedimentation, and interaction with radiation). The range of values of all soil properties for all study sites is in Table 2. Regarding the results of SOM and the models containing this variable, the uncertainty brought about by using different methods is

unknown, and thus they should be used with caution. For the remainder of the text, acid is used to denominate acid-to-neutral soils.

Table 2. Soil properties range (number of measurements, minimum, maximum and median values).

| Soil property | All | | | | Acid to neutral (pH≤7) | | | | Alkaline (pH>7) | | | |
|---------------|-----|-----|-----|-----|------------------------|-----|-----|-----|-----------------|-----|-----|-----|
| | n* | Min | Max | Med | n* | Min | Max | Med | n* | Min | Max | Med |
| Sand (%) | 227 | 4 | 93 | 47 | 117 | 10 | 93 | 50 | 110 | 4 | 85 | 43 |
| Silt (%) | 227 | 6 | 80 | 34 | 117 | 7 | 80 | 36 | 110 | 6 | 78 | 33 |
| Clay (%) | 227 | 0 | 56 | 13 | 117 | 0 | 44 | 11 | 110 | 0 | 56 | 15 |
| PR (MPa) | 178 | 0.5 | 5.9 | 2.2 | 97 | 0.6 | 5.0 | 2.2 | 81 | 0.5 | 5.9 | 2.4 |
| pH | 255 | 4.2 | 8.7 | 7.0 | 133 | 4.2 | 7.0 | 6.3 | 122 | 7.1 | 8.7 | 7.7 |
| LOC (mg/g) | 249 | 0.1 | 8.9 | 0.8 | 127 | 0.1 | 8.9 | 0.7 | 122 | 0.1 | 6.9 | 0.8 |
| SOM (%) | 175 | 0.3 | 9.2 | 2.2 | 80 | 0.3 | 9.2 | 2.2 | 95 | 0.4 | 8.6 | 2.4 |

*Number of observations (not all properties were recorded at all sites). LOC: labile organic carbon (mg/g); SOM: soil organic matter (%); PR: penetration resistance

2.1.4. Climate variables and indices

Estimates of the climate variables at the 256 georeferenced locations were obtained from the “Local Climate Estimator”, which uses the FAOCLIM 2 dataset (worldwide agroclimatic data) [14]. The pedoclimatic zones, defined in [15], are in Table 3. Climate variables are in Table 4.

Table 3. Pedoclimatic zones of the sampling locations.

| Europe | China |
|--------------------------|----------------------------|
| Mediterranean temperate | Central Asia tropical zone |
| Mediterranean semi-arid | Middle temperate zone |
| Atlantic | |
| Southern sub-continental | |
| Northern sub-continental | |
| sub-Boreal to Boreal | |

Table 4. Climate variables range (the minimum, maximum and median values).

| Climate variable/index* | Minimum | Maximum | Median |
|-------------------------|---------|---------|--------|
| T | 4.7 | 18.7 | 10.5 |
| P | 250 | 1393 | 665 |
| PET | 504 | 1507 | 728 |
| AI | 0.20 | 2.24 | 0.96 |
| NPP | 459 | 1794 | 1035 |
| GCI | 1.9 | 80.4 | 27.5 |

*T = mean annual temperature (°C); P = mean annual precipitation (mm); PET= mean annual potential evapotranspiration (mm); AI = Aridity index =P annual mean/PET annual mean; NPP = net primary production potential (g (DM) m⁻² yr⁻¹), NPP lim = limiting value (NPP temperature or NPP precipitation); GCI=Gorczyński Continentality Index.

2.2. Logistic regression analysis

2.2.1. Models

Logistic regression models allow the study of the relation between a categorical dependent variable (in our study, the scores of a visual soil quality indicator) and a set of explanatory variables (in our case, climate and soil variables) and their interactions. The first step is to calculate the logit transformation. For example, consider a binary dependent variable, and “0” and “1” are the two possible outcomes, where “0” represents a negative response and “1” a positive then if p is the proportion of positives, $1-p$ will be the proportion of negatives and the logit transformation is given by Equation 1:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{P_1}{P_0}\right) + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

where P_1 and P_0 are the prior probabilities of a random observation belonging to one or the other outcome, and the β_s are the regression coefficients to be estimated from the data (by the maximum likelihood method, corresponding to the values that maximize the log-likelihood equation) and the X_s are the explanatory variables. In this study, the prior probabilities were set as equal. The logistic transformation equation (the inverse of the logit transformation) is used to calculate the probability of outcome 1 (Equation 2):

$$p = \frac{e^{\text{logit}(p)}}{1 + e^{\text{logit}(p)}} \quad (2)$$

The coefficients (β_n) are the natural logarithm of the odds ratio ($\ln(\text{OR})$). The odds ratio is the ratio between the odds after increasing one unit and the initial odds (odds ($X_1 + 1$)/ odds (X_1)). A positive coefficient means that increasing the units of that variable increases the probability of outcome 1; a negative outcome is the opposite, i.e. increasing the number of units of that variable decreases the probability of outcome 1 (increasing the probability of outcome 0) and a coefficient equal to zero means that the variable has no explanatory value (no change in the odds ratio by increasing one unit, and thus the odds ratio is equal to 1 ($\ln(1) = 0$)).

In our case, with three possible outcomes, one would be chosen as a reference outcome, and the other two would be compared to that reference using two regression equations ($3 - 1 = 2$). However, the frequencies of the category “poor” are too low for all indicators but “earthworm count”, for which “good” was the lowest (Table 1), and, a priori, that would prevent us from creating meaningful models for those categories (see section 2.2.3.). Thus, categories “poor” and “moderate” were collapsed into a single category ($Y = 0$ (poor + moderate) and $Y = 1$ (good)), except for “earthworm count”, for which “moderate” and “good” were collapsed.

Interaction models. The interaction models, with two main effects and an interaction term, take the form:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \ln\left(\frac{P_1}{P_0}\right) + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 \quad (3)$$

When categorical independent variables are present, interpreting the coefficients of the main effects and interaction terms is straightforward. However, when the variables are continuous, the

results require an analysis of the evolution of the logit with different combinations of values of the main effects. The reason lies in the way the coefficients of the variables are estimated and thus reflect the effect structure (the main effects are forcibly highly correlated with the interaction term). The interpretation of the coefficients of interactions is different from that of models without interaction terms and probably best apprehended graphically (see Figure A1 in Annexe 1 in Supplemental Materials for three examples). The existence of a potential interaction was considered for an increase of the pseudo-R-squared (R_L^2) higher than 0.10 compared to the R_L^2 of the main effects (models with two variables).

Models were created using NCSS Statistical Software (version 11).

2.2.2. Statistical tests

The data used to fit each model has “repeats”, i.e. depending on the variables chosen for the model, some records can be identical. For example, models containing only climate variables will have several repeats, while a model containing, for example, a climate variable and labile organic carbon will probably have few or no repeats. Because we are primarily interested in comparing models to assess which variables and interactions account the most for the variations observed in the scores, we opted to use the pseudo-R-squared statistic.

The pseudo-R-squared value for logistic regression (R_L^2) is given by Equation 4:

$$R_L^2 = \frac{L_p - L_0}{L_0 - L_s} \quad (4)$$

where L_p is the log-likelihood of the model containing the independent variables, L_0 is the log-likelihood of the model containing only the intercept, and L_s is the log-likelihood of the saturated model.

As a measure of the goodness-of-fit of the selected models, i.e. of the ability of the model to discriminate, we used the C statistic (C_{sta}). This choice was because of the ease of calculation, as it is quantitatively equal to the area under the ROC (receiver operating characteristic) curve [16]. However, it cannot be used when modelling dependent categorical variables with more than two outcomes. To learn more about the calculation of the C_{sta} , see Arboretti Giancristofaro and Salmaso [17]. C_{sta} ranges from zero to one. One interpretation of the C_{sta} is that, given a random pair of observations, one positive and one negative, it represents the probability that the probabilities calculated by the model for the positive are higher than for the negative. Thus, a C_{sta} of 0.5 means that the model has no discrimination and a value of 1, complete discrimination. In this study, following the values proposed in [17], a model has acceptable discrimination for $0.7 \leq C_{sta} < 0.8$ and excellent discrimination for $0.8 \leq C_{sta} < 0.9$.

The Wald test was used to test the significance of the regression coefficients (Equation 5):

$$Z_m = \frac{b_j}{SE(b_j)} \quad (5)$$

where b_j is the coefficient estimate, and $SE(b_j)$ is an estimate of the standard error of b_j . The null hypothesis is that the variable has no explanatory value ($b_j = 0$).

The models were selected for further study if presenting all coefficients with a p-value < 0.20 .

2.2.3. Determining the sample size needed to fit the models

There's much controversy about the adequacy of different methods to calculate the sample size in logistic regression. Using the events per variable (EPV) criterion, a rule of thumb that the number of recorded events of the category with the lowest frequency should be equal to or higher than ten events per variable has been shown inappropriate when modelling different subject matter [18]. On the other hand, the work of other authors supports this criterion. For example, in a simulation study [19], the authors found that an EPV of ten or higher would avoid many problems and that an EPV equal to or lower than five could result, among others, in a high increase in the variance of the coefficient estimates.

Initially, the EPV criterion was a guiding rule to decide the number of variables of each model. The results were then compared with a simple method proposed by Hsieh et al. [20]. It consists of starting with a logistic regression model with one independent variable (Equation 6), where X_1 is a continuous variable that we assume has a normal distribution and is a good predictor of the dependent variable:

$$\text{logit}(p) = \ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1 \quad (6)$$

To test the null hypothesis $H_0: \beta_1 = 0$ against the alternative $H_1: \beta_1 = \beta^*$, with $\beta^* \neq 0$, for a given significance ($\alpha = 0.05$) and power ($\beta = 0.20$), the minimum sample size needed is given by Equation 7:

$$n = \frac{\left(Z_{1-\frac{\alpha}{2}} + Z_{1-\beta}\right)^2}{p_0(1-p_0)\beta^{*2}} \quad (7)$$

where n is the size of the sample, Z_{γ} s are the critical values of Z , $p_0 = P(y = 1 | X_1 = \mu)$ and $\beta^* = \ln(\text{OR})$ (the natural logarithm of the odds ratio). If $p_1 = P(y = 1 | X_1 = \mu + \sigma)$, then we can use p_1 to calculate OR (Equation 8) since:

$$\text{OR} = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} \quad (8)$$

The estimated sample size for the models containing all variables and interactions was calculated by multiplying n with a variance inflation factor ($1/(1-R^2)$) (Equation 9):

$$n_{\text{total}} = \frac{n}{1-R^2} \quad (9)$$

where R^2 (determination coefficient) was calculated by regressing X_1 on all other explanatory variables using multiple regression.

Logistic regression and multiple regression models were calculated using NCSS Statistical Software (version 11). All other calculations were performed in Excel (Office 2016).

3. Results

The effects of the climate and soil properties on the scores of the visual soil quality indicators depend on the soil reaction (Figures 1 and 2). A total of 85 potential interactions were detected, with

an R_L^2 increase ≥ 0.1 compared to the models with only the main effects, 53 in acid and 32 in alkaline soils (Figure 3).

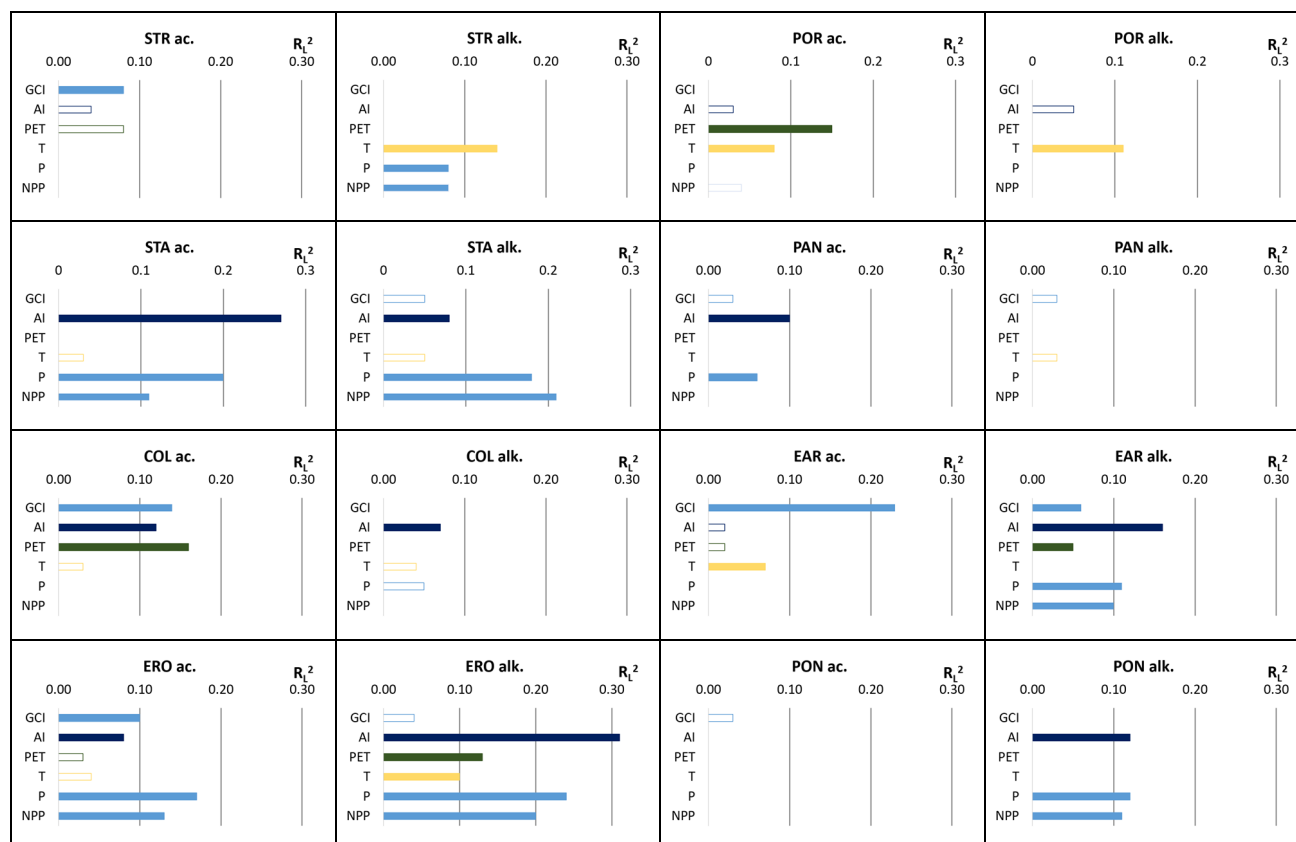


Figure 1. R_L^2 s of the models of visual indicators featuring a single climate variable (left panels for acid soils and right panels for alkaline ones). Models' coefficients with Wald p-value < 0.05 are featured with coloured bars (filled); models' coefficients with Wald p-value < 0.20 (but ≥ 0.05) are featured with white bars. T: mean annual temperature ($^{\circ}\text{C}$); P: mean annual precipitation (mm); PET: mean annual potential evapotranspiration (mm); AI: Aridity index: P annual mean/PET annual mean (dimensionless); NPP: net primary production potential ($\text{g (DM) m}^{-2} \text{ yr}^{-1}$). GCI: Gorczyński continentality index (dimensionless). Str: Soil structure; Por: Soil porosity; Sta: Soil stability (Slake Test); Pan: Presence of a tillage pan; Col: Soil colour; Ear: Earthworm count; Ero: Susceptibility to wind and water erosion; Pon: Surface ponding.

3.1. Climate effect on the scores of the visual soil quality indicators

Climate affects the scores of the visual soil quality indicators differently, and concurrently the explanatory value of each variable also differs according to soil reaction (Figure 1). The coefficients of each climate variable (models containing only the intercept and one variable), the Wald p-value, the odds ratio change, the R_L^2 value and other statistical measures are in Tables 4 and 5 in Supplemental Materials.

Except for surface water ponding in acid soils and the presence of tillage pan in alkaline soils, all other indicators had at least one climate variable (single-variable model) with statistically significant

coefficients (the null hypothesis of $\beta = 0$ is rejected for $\alpha < 0.05$). The highest R_L^2 observed was for the soil erosion model in alkaline soils, featuring the aridity index as the independent variable ($R_L^2 = 0.31$). Single-variable models were not sufficient to explain the variation in the scores observed. However, interaction models of the climate variables, containing two variables and an interaction term, show that the effect of some climate variables mainly depends on the value of other climate variables (all interaction models between climate variables are in Tables 6 and 7 in Supplemental Material). Figure 3, in the body of the article, summarizes the interaction models deemed of interest- those with higher R_L^2 differences between the interaction model and the R_L^2 of the sum of the main effects (models featuring only two terms). The width of the lines denotes the difference in R_L^2 . The highest R_L^2 of the interaction models between climate variables was for the earthworm count in acid soils, featuring mean annual temperature and precipitation and the interaction term as independent variables ($R_L^2 = 0.72$, a 0.53 increase compared to the R_L^2 of the model featuring only the main effects).

3.2. Soil properties' effects on the scores of the visual soil quality indicators

Soil particle-size fractions. Models of visual indicators containing only a single particle-size fraction have low explanatory value and, depending on the soil reaction, explain the variability of the score of these indicators very differently (Figure 2). Only three models had an R_L^2 higher than 0.10, and the highest- soil colour in acid soils featuring clay as the independent variable- had 0.24. Complete models are found in Tables 8 and 9 in Supplemental Materials.

No interaction models featuring the three soil particle-size fractions (models with intercept plus four terms) had all coefficients statistically significant or at least $p < 0.20$ (Tables 10 and 11 in Supplemental Materials). Interaction models of interest, with two particle-size fractions at a time (plus the interaction term), are in Figure 3. These interactions were observed for soil erosion and surface water ponding. The interaction model with the highest R_L^2 was for surface water ponding in acid soils (sand x clay), with an increase of R_L^2 of 0.20. In alkaline soils, the interaction model of surface ponding featuring sand x silt has an R_L^2 of 0.20, an R_L^2 increase of only 0.04 compared to the main effects model (an R_L^2 of 0.16); the main effects' R_L^2 is exactly the R_L^2 of the surface ponding's model with clay as the independent variable, reinforcing the suggestion that clay content, per se, may explain the effect of soil texture on the score of the indicator (although in the future, more detailed models with other variables, such as clay mineralogy, may explain a higher proportion of the variation observed).

Soil manageable variables. Models featuring a single manageable variable (SOM, LOC, pH or PR) differ in the acid and alkaline groups (Figure 2). The highest R_L^2 (0.29) was observed for the model of "the presence of a tillage pan" in alkaline soils featuring pH as the independent variable.

Interaction models of interest containing two manageable variables at a time (plus the interaction term) were only observed in the acid soils group. They were observed for three visual indicators and between SOM and PR (Figure 3). Models are in Tables 14 and 15 in Supplemental Materials.

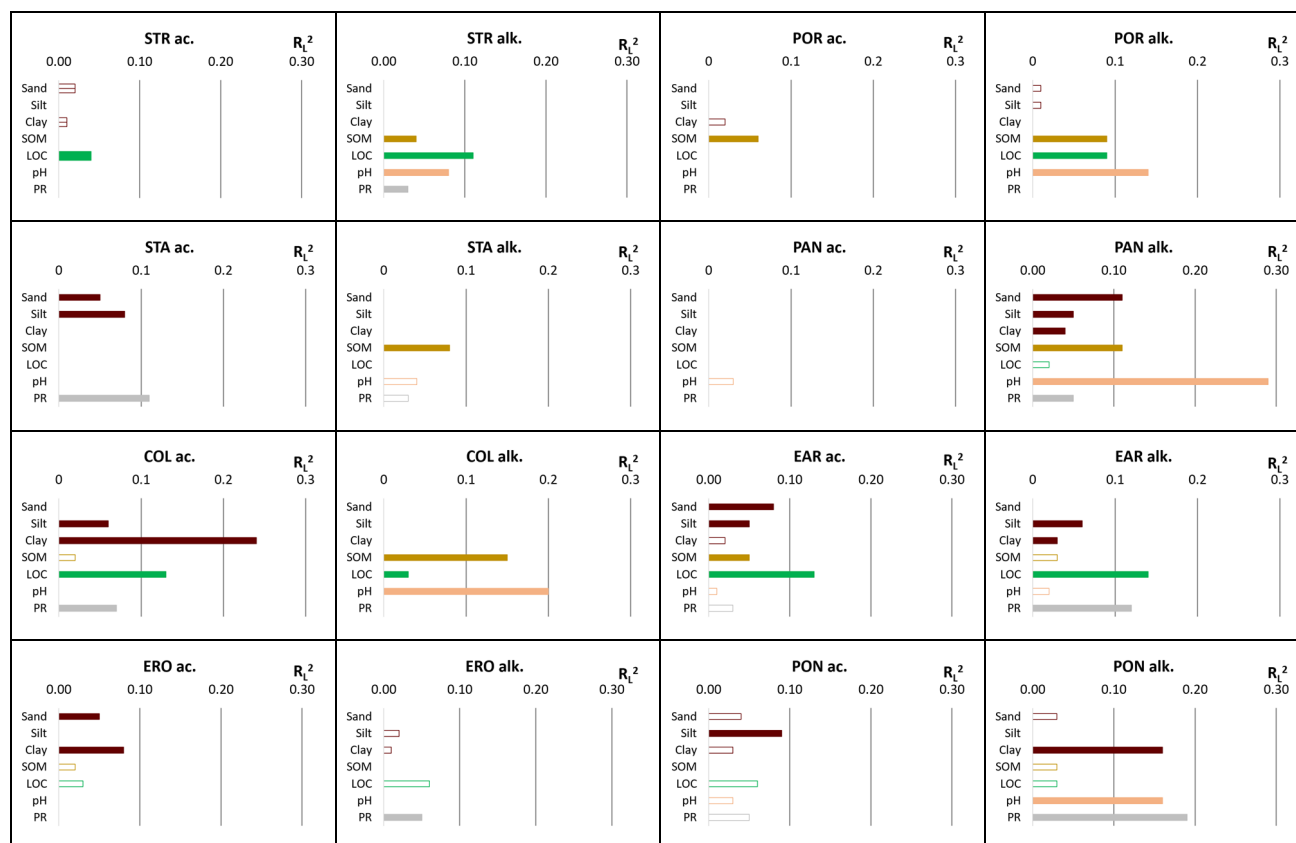


Figure 2. R_L^2 s of the models of visual indicators featuring a single soil variable (left panels for acid soils and right panels for alkaline ones). Models' coefficients with Wald p -value < 0.05 are featured with coloured bars (filled); models' coefficients with Wald p -value < 0.20 (but ≥ 0.05) are featured with white bars. LOC: labile organic carbon (mg/g); SOM: soil organic matter (%); pH: ranked; PR: penetration resistance (MPa); Sand, Silt and Clay (%). Str: Soil structure; Por: Soil porosity; Sta: Soil stability (Slake Test); Pan: Presence of a tillage pan; Col: Soil colour; Ear: Earthworm count; Ero: Susceptibility to wind and water erosion; Pon: Surface ponding.

3.3. Interactions between all variables

The interaction models of climate variables with particle-size fractions and manageable variables, and those of particle-size fractions and manageable variables, with all coefficients with a Wald p -value < 0.20 and an increase of R_L^2 of at least 0.10, are presented in Figure 3. A similar figure is in Supplemental Materials (Figure 1), featuring interaction models with lower R_L^2 (higher than 0.03). Selected statistical properties of all models are in Tables 16 to 21 in Supplemental Materials.

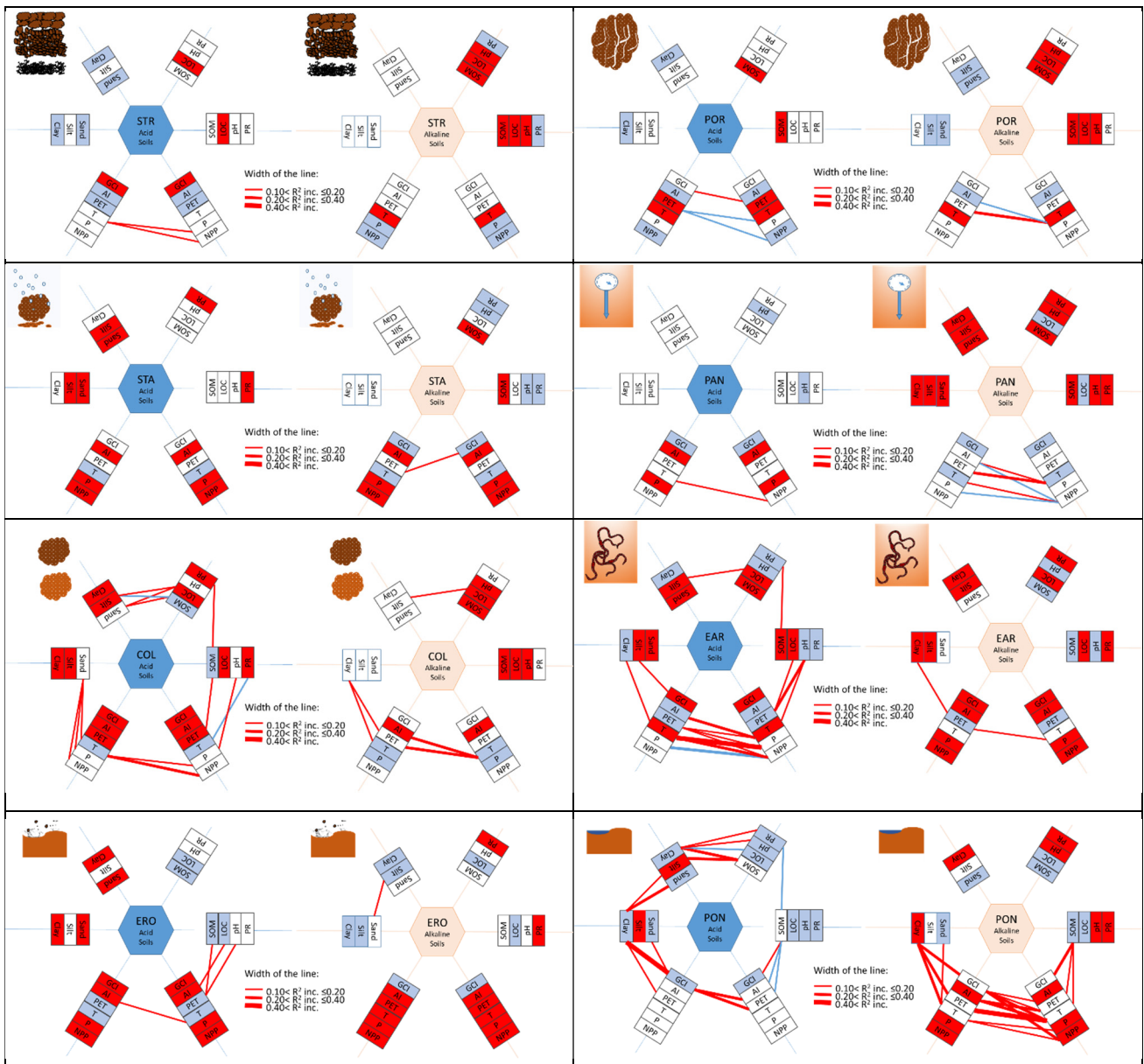


Figure 3. Potential interactions between the variables. Blue lines mean all coefficients of the interaction models had a p -value < 0.20 , and red lines that all coefficients had a p -value < 0.05 . The lines' width indicates the magnitude of the increase of the RL^2 in comparison to the RL^2 of the models containing only the main effects (two terms). Blue, red and white rectangles indicate the p -value of the coefficients of the single-variable models, and the colour holds the same interpretation (blue for p -value < 0.2 , red for p -value < 0.05 and white for p -values > 0.2). The hexagons have the name of the visual indicator within, and the colour corresponds to blue = acid soils and pink = alkaline. Models of the interactions can be found in Tables 6, 7, 10, 11 and 14 to 21 in Supplemental Materials.

Climate features and soil particle-size fractions. Interaction models of the visual indicators, containing one climate variable and one particle-size fraction (plus the interaction term), were observed in three visual indicators, with one exception the increase of the RL^2 was relatively low. The

exception, that of “surface water ponding”, had an R_L^2 increase of 0.41 in the acid group (GCI x clay), while in the alkaline group, the model featuring the same interaction had an increase of 0.26.

Climate features and soil’s manageable variables. Interactions of interest (R_L^2 increase of at least 0.1) were detected for four visual indicators: biologically mediated processes- soil colour and earthworm count- and soil erosion in acid soils and surface ponding in both soil reaction groups. The interaction model GCI and SOM for surface ponding in the alkaline group had the highest R_L^2 increase, 0.21; in the acid group, the R_L^2 increase for the same model was lower, 0.17.

Soil’s manageable variables and particle-size fractions. Only three visual indicators had potentially meaningful interactions (an R_L^2 increase > 0.10) between manageable variables and particle-size fractions: soil colour (the highest was LOC and silt in both soil reaction groups), earthworm count (LOC and silt in the acid group) and surface ponding (the highest was SOM and clay in the acid group).

3.4. Variable selection and the sample size for modelling the large-scale spatial variation of the score of visual soil quality indicators

Variable selection. The preselection of the explanatory variables for each visual indicator, different for each soil reaction group, was made taking into account the Wald p-value of the coefficients of the single-variable and interaction models ($p < 0.20$), the R_L^2 of the models and the Pearson correlations between the explanatory variables as a first approach to multicollinearity (Tables 2 to 21 in Supplemental Materials).

A definite set of variables for each visual indicator was made by fitting models and respecting an EPV (Events Per Variable, see section 2.2.3.) higher than eight, except for “surface ponding” in acid soils and “soil stability” and “soil colour” in alkaline, because of the small size of the sample of the smallest category of these indicators (Table 5). Except for the “soil structure and consistency” model for acid soils, with a low discrimination performance, all other models had acceptable ($C_{sta} \geq 0.7$) to excellent ($C_{sta} \geq 0.8$) discrimination (Table 5). Alternative and complementary models that help justify or illustrate different possibilities and effects are in Tables 22 and 23 in Supplemental Materials. Other combinations of the variables can produce similar results. The choice is rather empirical; for example, the surface ponding interaction model, in alkaline soils, with AI and T as the main effects and AI x T as the interaction term, increases the R_L^2 by 0.10 (0.50) and could be used instead of the model selected and present in Table 5. However, this interaction model has a lower C_{sta} (0.86) and adding a variable, e.g. pH, would increase the discrimination slightly (C_{sta} of 0.87 and an R_L^2 of 0.37, see Table 23 in Supplemental Materials) below the selected model.

Table 5. Selected models. Values were rounded*.

| | ACID SOILS | | | | | | | |
|-----------|-------------|------------|-------------|------------|------------|--------------|-------------|-------------|
| | STR | POR | STA | PAN | COL | EAR | ERO | PON |
| Variables | T | PET | AI | AI | GCI | T | P | AI |
| | GCI | SOM | PR | SILT | LOC | GCI | CLAY | GCI |
| | LOC | PR | SILT | LOC | SILT | SAND | | CLAY |
| | | | | | CLAY | | | |
| Terms | Int (-3.53) | Int (4.61) | Int (-3.69) | Int (0.24) | Int (2.52) | Int (-14.59) | Int (-3.99) | Int (17.36) |

Continued on next page

| ACID SOILS | | | | | | | | |
|-------------------|-----------------|----------------|-----------------|----------------|----------------|-----------------|---------------------|-----------------|
| | STR | POR | STA | PAN | COL | EAR | ERO | PON |
| and (β s) | T (0.25) | PET (-0.01) | AI (0.63) | AI (1.60) | GCI (-0.08) | T (1.55) | P (0.00) | AI (-19.36) |
| | GCI (0.07) | OM (1.43) | PR (0.63) | SILT (-0.05) | LOC (-0.34) | GCI (0.10) | CLAY (0.10) | GCI (-0.53) |
| | LOC (0.26) | PR (-0.68) | SILT (0.03) | LOC (-0.16) | SILT (0.04) | SAND (0.08) | | CLAY (-0.12) |
| | GCI x T (-0.01) | | | | CLAY (-0.07) | GCI x T (-0.02) | | GCI x AI (0.68) |
| Model R_L^2 | 0.08 | 0.31 | 0.20 | 0.13 | 0.39 | 0.44 | 0.22 | 0.42 |
| C_{sta} | 0.67 | 0.87 | 0.80 | 0.75 | 0.89 | 0.90 | 0.80 | 0.87 |
| SE(C_{sta}) | 0.05 | 0.05 | 0.04 | 0.05 | 0.04 | 0.03 | 0.05 | 0.07 |
| Intercept p-value | 0.004 | 0.031 | 0.000 | 0.71 | 0.039 | 0.000 | 0.000 | 0.000 |
| Bs' p-value range | 0.005 to 0.016 | 0.006 to 0.100 | 0.012 to 0.183 | 0.000 to 0.064 | 0.001 to 0.029 | 0.000 to 0.085 | 0.003 to 0.008 | 0.000 to 0.003 |
| Single rows | 117 | 54 | 92 | 105 | 106 | 111 | 104 | 87 |
| Rows used | 125 | 58 | 96 | 111 | 111 | 117 | 117 | 98 |
| EPV | 14 | 9 | 12 | 17 | 13 | 12 | 13 | 4 |
| ALKALINE SOILS | | | | | | | | |
| | STR | POR | STA | PAN | COL | EAR | ERO | PON |
| Variables | P | LOC | NPP | pH | PET | AI | AI | P |
| | LOC | pH | PET | SOM | T | LOC | SILT | CLAY |
| | pH | SOM | SOM | SAND | SOM | SILT | SAND | pH |
| | | T | pH | | | PR | | |
| Terms | Int (13.09) | Int (6.50) | Int (-5.23) | Int (8.65) | Int (13.83) | Int (2.38) | Int (-6.78) | Int (18.05) |
| and (β s) | P (-0.00) | LOC (0.45) | NPP (0.02) | pH (-1.50) | PET (-0.02) | AI (-1.42) | AI (2.70) | P (0.00) |
| | LOC (0.80) | pH (-0.95) | PET (0.01) | SOM (0.50) | T (-1.08) | LOC (-0.83) | SILT (0.17) | CLAY (-0.09) |
| | pH (-1.57) | SOM (0.50) | SOM (0.65) | SAND (0.03) | SOM (0.79) | SILT (-0.04) | SAND (0.10) | pH (-2.31) |
| | | T (-0.08) | pH (-1.74) | | PET x T (0.00) | PR (0.41) | SAND x SILT (-0.00) | |
| | | | NPPxPET (-0.00) | | | | | |
| Model R_L^2 | 0.17 | 0.24 | 0.28 | 0.25 | 0.40 | 0.29 | 0.22 | 0.40 |
| C_{sta} | 0.74 | 0.80 | 0.83 | 0.81 | 0.89 | 0.83 | 0.82 | 0.89 |
| SE(C_{sta}) | 0.04 | 0.04 | 0.04 | 0.04 | 0.03 | 0.04 | 0.05 | 0.04 |
| Intercept p-value | 0.002 | 0.178 | 0.33 | 0.091 | 0.001 | 0.066 | 0.003 | 0.017 |
| Bs' p-value range | 0.002 to 0.009 | 0.002 to 0.179 | 0.000 to 0.012 | 0.001 to 0.028 | 0.000 to 0.003 | 0.008 to 0.092 | 0.005 to 0.016 | 0.000 to 0.040 |
| Single rows | 114 | 89 | 87 | 89 | 85 | 77 | 101 | 96 |
| Rows used | 122 | 95 | 95 | 95 | 95 | 81 | 110 | 105 |
| EPV | 15 | 9 | 6 | 10 | 7 | 9 | 9 | 10 |

*Do not use rounded coefficients to calculate probabilities because they can hide important effects (use the coefficients in Tables 22 and 23 in Supplemental Materials); Int= Intercept; Csta= C statistic; SE(Csta)= standard error of the C statistic. T: mean annual temperature ($^{\circ}$ C); P: mean annual precipitation (mm); PET: mean annual potential evapotranspiration (mm); AI: Aridity index: P annual mean/PET annual mean (dimensionless); NPP: net primary production potential (g (DM) m⁻² yr⁻¹). GCI: Gorcezynski continentality index (dimensionless). LOC: labile organic carbon (mg/g); SOM: soil organic matter (%); pH: ranked; PR: penetration resistance (MPa); Sand, Silt and Clay (%). Str: Soil structure; Por: Soil porosity; Sta: Soil stability (Slake Test); Pan: Presence of a tillage pan; Col: Soil colour; Ear: Earthworm count; Ero: Susceptibility to wind and water erosion; Pon: Surface ponding.

Climate plays a role in the spatial distribution of acid and alkaline soils. The climate effect on the scores was conveniently modelled with one climate variable for ten indicators. For five visual

indicators, the score variations were better accounted for by interactions between climate variables. Additionally, for the “presence of a tillage pan” in alkaline soils, the selected model has no climate variable. The effect of pH on the scores of the latter visual indicator is very high (the model featuring only pH has an R_L^2 of 0.29 and a C_{sta} of 0.77) and may mask the climate effect (see Discussion). In the alkaline soils group, the variation of the scores of four visual indicators can be accounted for to a great extent by interaction models between climate variables and especially by interactions between PET and one other variable (Figure 3); however, in a frame of a limited number of terms per model, most models that were fitted with these interactions either performed below the selected models or the coefficient estimator function failed to converge. In the acid soils group, three visual indicators had the variation of the scores better accounted for by models including climate variables interactions. These models have the continentality index in common and, with one exception, the surface ponding model featuring the interaction GCI x AI, the interaction effect is relatively small. Nevertheless, in the acid group, other interaction models are promising- e.g. the earthworm count interaction model with T and P (C_{sta} 0.91 and R_L^2 of 0.72)- and candidates to integrate future models; however, when other variables were added to, most of these models failed to converge. The lack of convergence was probably because of the small sample for that particular combination of effects; these models are identified in Tables 16 to 21 in Supplemental Materials with the acronyms QS (Quasi-separation) or NED (Not Enough Data).

Sand and silt are strongly correlated, with coefficients of -0.82 and -0.75 , in acid and alkaline soils, respectively (Table 2a and 2b in Supplemental Materials). In the acid group, the models of six visual indicators had their performance improved by featuring one particle-size fraction. On the other hand, in the alkaline group, only four visual indicators had models with higher discrimination by featuring particle-size fractions, and one- soil erosion- was the only selected model featuring one interaction between soil variables (sand x silt). Some interaction models between particle-size fractions and climate variables should be considered in future models fitted with a larger sample. For example, the surface ponding interaction model in acid soils featuring GCI and clay (R_L^2 of 0.48 and C_{sta} of 0.87) had a similar performance to the selected model but performed poorly when other variables were added.

Except for the soil erosion models of both soil reaction groups, featuring only climate and soil texture variables, all other models feature at least one soil manageable variable. The variables increasing the most the R_L^2 s in acid soils were LOC (three models), PR (two models) and SOM (one model), while in alkaline soils, pH was present in five selected models, two models featured LOC, two featured SOM and one PR.

Sample size. Because of the small number of observations, the categories with fewer observations were collapsed with a contiguous category into one category (see section 2.2.1.), i.e. two possible outcomes.

The sample size necessary to fit each model identified in Table 5 was calculated following the rule of thumb of ten events per variable (EPV) of the category with the lowest frequency (Table 6). The sample size necessary to fit models for three possible outcomes was also calculated (the number of models to be fitted per indicator and soil reaction group would be: number of models = number of outcomes -1). Comparing the size of the sample used (number of rows used in Table 5) and the sample size calculated (two outcomes in Table 6), the sample size used was lower than calculated for two indicators in the acid group and five in the alkaline.

Table 6. Sample size necessary to fit the selected models following the events per variable approach (for two and three outcomes), and following the “one variable and a variance inflation factor” approach.

| | | | ACID SOILS | | | | | | | |
|----------|-------------------|-------------|----------------|-----|-----|-----|------|------|------|------|
| Number | | | STR | POR | STA | PAN | COL | EAR | ERO | PON |
| outcomes | | | | | | | | | | |
| EPV | | Terms | 4 | 3 | 3 | 3 | 4 | 4 | 2 | 4 |
| 2 | | Sample size | 95 | 62 | 86 | 64 | 93 | 89 | 109 | 183 |
| 3 | | Sample size | 333 | 399 | 495 | 499 | 313 | 222 | NED | 1330 |
| | One Exp. Var. and | Term | LOC | OM | PR | AI | CLAY | SAND | CLAY | CLAY |
| | VIF | Sample size | 234 | 76 | 56 | 166 | 22 | 86 | 43 | 231 |
| | | | ALKALINE SOILS | | | | | | | |
| Number | | | STR | POR | STA | PAN | COL | EAR | ERO | PON |
| outcomes | | | | | | | | | | |
| EPV | | Terms | 3 | 4 | 5 | 3 | 4 | 4 | 4 | 3 |
| 2 | | Sample size | 80 | 112 | 154 | 82 | 141 | 85 | 143 | 97 |
| 3 | | Sample size | 231 | 246 | 237 | 194 | 351 | 259 | 1180 | 527 |
| | One Exp. Var. and | Term | LOC | OM | OM | pH | OM | PR | AI | pH |
| | VIF | Sample size | 54 | 85 | 121 | 32 | 54 | 41 | 30 | 39 |

NED: not enough data to calculate the sample size.

The sample size for each indicator calculated with the second approach, based on a single independent variable with a “good” explanatory value and a variance inflation factor (VIF), is in Table 6. Compared to the values obtained with the EPV approach and two outcomes, the single explanatory variable approach leads to a larger sample size for four indicators in the acid group and one in the alkaline. If the assumption is that the independent variable has good explanatory value to discriminate other categories (three categories in the present work), the sample size needed would be lower than the one calculated with the EPV approach (Table 6, three outcomes vs Ind. Var. and VIF). However, this assumption could not be checked because the sample for the categories with the lowest frequencies was too small to calculate a reliable logistic single-variable model (see Table 1).

Models of the visual indicators with six and nine variables and two or three possible outcomes for an EPV of ten need sample sizes ranging between 326 and 2993 and 215 and 2655 in acid and alkaline soils, respectively (Table 25 in Supplemental Materials).

4. Discussion

The importance of each variable on the score of the visual indicators can only be assessed in the frame of a definite set of variables. The reason for this is the existence of collinearities, i.e. the independent variables are not truly independent (see the Pearson correlation coefficients between the variables in Table 2 in Supplemental Materials). The effects portrayed in Figures 1, 2 and 3 cannot be straightforwardly added. When increasing the number of variables of a model, the overlapping effect of two or more variables and interactions on the scores of a visual indicator may increase or reduce the R_L^2 of that model, change the value and the variance of the coefficients, and increase or reduce the discrimination ability of the models [21].

The dichotomy observed between soil reaction groups concerning the effect of climate variables- and interactions- can be explained by the relationship between mean annual precipitation and evapotranspiration and soil pH [22], with acid soils occurring predominantly in humid climates and alkaline soils in arid ones. The Pearson correlation between the climate variables can be very high, which could suggest collinearity issues when modelling (see Table 2a and 2b in Supplemental Materials). However, those variables are present in the same models only in the frame of interactions. Furthermore, the p-value of the coefficients of those interaction models can be very low, and an inflation of the variance of the coefficients doesn't seem to be present. The problem is better analyzed by comparing the coefficient values and p-values of single-variable models (Tables 4 and 5 in Supplemental Materials), two-variable models and the interaction models (selected models in Tables 22 and 23 in Supplemental Materials). The hypothesis that the variance of the coefficients is inflated is put aside. Concurrently, the R_L^2 s of the models and their discrimination ability increase steeply with the inclusion of the interaction term compared to the models featuring only the main effects (selected models in Tables 22 and 23 in Supplemental Materials). An example is the "earthworm count" interaction model featuring T and P and a T x P term in the acid soils group: it has an R_L^2 of 0.72 and C_{sta} of 0.91. A model featuring only T and P has an R_L^2 of 0.19 and a C_{sta} of 0.76.

Interactions were observed in both acid and alkaline soils (Figure 3). The most common were T x P in acid soils and PET x T in alkaline. However, adding an explanatory variable to the interaction models between climate variables allowed, in some cases, to improve the discrimination ability of the models (increasing the specificity/sensibility of the models) and, in other cases, the coefficient estimator either did not converge or the model's performance dropped significantly. Other interactions were detected between soil particle-size fractions, manageable variables and climate variables. However, only the soil erosion model in alkaline soils had its discrimination improved by including one of these interactions (sand x silt). A larger sample should allow the creation of models including other variables and interactions identified in this work and possibly increase the prediction accuracy of the models.

The models with lower R_L^2 and C_{sta} are "soil structure" in both soil reaction groups and the "presence of tillage pan" in the acid group. Not coincidentally (hypothetically), soil structure and porosity are the visual indicators more sensitive to agricultural soil management practices (AMPs) [6]. The effect of different AMPs on the scores of these two indicators, the odds ratio of the score "good" of the visual indicators of AMPs to the odds of the score "good" of control groups, are in Table 24 in Supplemental Materials. The selected management practices/control odds ratio of the score "good" for soil structure and soil porosity, with statistically significant p-values (the null hypothesis that the different frequencies for the score "good" between the AMP and control were due to chance were rejected for $\alpha < 0.05$), varied between 4.7 and 21. Thus, soil management practices represent a fraction of the variation observed in the scores of these two indicators and should be considered in future models. Other variables not used in this study are also excellent candidates. For example, the soil parent material (presence of carbonate rocks) plays a definite role in soil reaction; it may be used to improve the discrimination ability of the models, as it allows the discrimination of soils that developed under different climates which do not conform with modern ones, i.e. acid soils in arid climates and vice-versa. The candidate variables to be assessed in the future for the different soil visual indicators should include, amongst others, physical (e.g. bulk density), chemical (e.g. total nitrogen) and biological properties (e.g. microbial carbon). Nevertheless, a better understanding of the processes that

govern the magnitude of the visual soil indicators that are, to an extent, conditioned by climate is the sine qua non to identify variables with a higher explanatory value.

The performance of the selected models for new observations is an open question, as it was not a goal to validate the models for standalone use. The models were created primarily to assess the climate effect- and of selected soil properties- on the scores of the visual indicators. One application might be to couple these models with pedotransfer functions that depend on the input of surface soil morphological features. Other applications might include, for example, the spatial distribution of the magnitude of these visual indicators, helping to identify and plan interventions on a large scale. For local usage, at a spatial scale for which climate variables may be considered constant, models should feature other variables (soil properties, soil management practices, irrigation and so forth).

To calculate the sample size, two assumptions had to be made: 1st, the explanatory variables hold a similar explanatory value to discriminate the different categories (when calculating for three categories); 2nd, the frequencies in the sample should be close to the frequencies in the population to be sampled.

The main problem of a low EPV is the increase in the variance of the coefficients [19]; however, when we cross-check the p-value of the coefficients of the models with the EPV used to fit the models (Table 5), it becomes apparent that the EPV approach is not an assurance of a convenient sample size. For example, the surface ponding model for acid soils has an EPV of four, a low variance of the coefficients (easily gauged by the p-value of the coefficients) and good performance (R_L^2 of 0.42 and C_{sta} of 0.87), while others with much higher EPV (up to 17) have coefficients with higher variance and poorer performance. Therefore, these findings suggest that the value of the explanatory variables to discriminate the categories is a much more rigorous way to calculate the sample size. In principle, the “one variable and a variance inflation factor” approach could be used to calculate the size of the sample regardless of the number of categories and their frequency proportions in the population to be sampled, as long as the explanatory variable has a “good” discrimination ability to differentiate all categories. However, if planning to include other explanatory variables in the models- for which we might know nothing- unless they are truly independent, we can expect the R^2 calculated by regressing the chosen variable on all other explanatory variables to increase- and a larger sample will be needed. Another problem is representativity. To account for the spatial distribution and the effect of the main interactions that are hypothetically at play and to validate the assumption that the chosen variable to calculate the sample size has similar discriminating power to all categories, a larger sample may be in order.

It was out of the scope of this article to identify and validate the simplest way to calculate the sample size needed to fit logistic models of the scores of the visual indicators of the VSA, built on climate and soil explanatory variables. Whichever method is chosen to calculate the sample size needed, not constrained to the two used in this study, it is apparent that the problem of concern is the variance of the coefficients of the logistic models; unfortunately, there is no obvious solution.

5. Conclusions

The results of this study show that simple logistic regression models of the VSA indicators can be used to assess the effects of climate and soil properties on the scores. The variables - and interactions- effects on the scores of the indicators also suggest that models with a higher discrimination ability are possible with the set of variables used. On the other hand, the close

association between climate and soil reaction hints that other variables, such as soil parent materials, might improve the accuracy of these models.

This study was thought to analyze the effect of climate and soil properties on the scores of visual indicators at a large spatial scale. These models are not adequate locally. At a farm or regional scale, the climate variables used are constants; other variables, such as soil management practices, are likely to affect the scores of the visual indicators, at least of the more sensitive ones (soil structure and porosity). The small sample meant a cap on the number of terms and outcomes the models could feature. Furthermore, missing values of soil organic matter and penetration resistance in some records might have caused some models not to converge, thus not allowing a complete assessment of the explanatory value of these variables.

By calculating the probability of a particular outcome and the weight (the odds ratio) of the different variables on that probability, the models allow the study of the soil processes that depend on soil morphological features. Possible applications of these models are, as mere examples, to calculate the rate of change of the probability of an outcome along a gradient of a climate feature, the determination of thresholds, and the improvement of the prediction accuracy of soil morphological-dependent pedotransfer functions.

As an on-farm tool, the benefits of these models are limited. Nevertheless, the models can calculate target values for manageable variables (LOC, SOM, pH and PR) for a particular location and help choose the appropriate soil management practices. On a larger scale, these and similar models can help local and national authorities in soil management planning and policy-making.

Future work should include studying the effect of soil management practices and interactions with climate variables and soil properties as they are, at least for the most sensitive visual indicators, an important source of variation in the scores. The effects of other climate variables and indices and different soil properties on the scores of the visual indicators should be assessed.

Use of AI tools declaration

The author declares he has not used Artificial Intelligence (AI) tools in creating this article.

Acknowledgements

The database used in this study resulted from the iSQAPER research project (Interactive Soil Quality Assessment in Europe and China for Agricultural Productivity and Environmental Resilience-which ended in April 2020). The project was funded by the European Union's Horizon 2020 Program for Research & Innovation, under grant agreement n° 635750, the Chinese Ministry of Science and Technology (grant n° 2016YFE011270), the Chinese Academy of Sciences (grant n° 16146KYSB20150001) and the Swiss State Secretariat for Education, Research and Innovation (contract: 15.0170-1). Special thanks go to the case study sites' teams who performed the field assessments and laboratory measurements.

Conflict of Interest

The author declares no conflict of interest.

References

1. Peerlkamp PK (1959) A visual method of soil structure evaluation. *Meded Landbouwhogesch Opzoekingsstn Staat Gent* 24: 216–221.
2. Shepherd T (2000) Visual Soil Assessment. Volume 1. Field guide for cropping and pastoral grazing on flat to rolling country. horizons.mw & Landcare Research, Palmerston North.
3. Ball B, Batey T, Munkholm L (2007) Field assessment of soil structural quality – a development of the Peerlkamp test. *Soil Use Manage* 23: 329–337. <https://doi.org/10.1111/j.1475-2743.2007.00102.x>
4. Van Leeuwen M, Heuvelink G, Wallinga J, et al. (2018) Visual soil evaluation: reproducibility and correlation with standard measurements. *Soil Tillage Res* 178: 167–178. <https://doi.org/10.1016/j.still.2017.11.012>
5. Mueller L, Kay BD, Hu C, et al. (2009) Visual assessment of soil structure: Evaluation of methodologies on sites in Canada, China and Germany. *Soil Tillage Res* 103: 178–187. <https://doi.org/10.1016/j.still.2008.12.015>
6. Teixeira F, Lemann T, Ferreira C, et al. (2023) Evidence of non-site-specific agricultural management effects on the score of visual soil quality indicators. *Soil Use Manage* 39: 474–484. <https://doi.org/10.1111/sum.12827>
7. Press S, Wilson S (1978) Choosing Between Logistic Regression and Discriminant Analysis. *J Am Stat Assoc* 73: 699–705. <https://doi.org/10.2307/2286261>
8. Mueller L, Shepherd G, Schindler U, et al. (2013) Evaluation of soil structure in the framework of an overall soil quality rating. *Soil Tillage Res* 127: 74–84. <https://doi.org/10.1016/j.still.2012.03.002>
9. Jafari A, Finke A, Vande Wauw J, et al. (2012) Spatial prediction of USDA-great soil groups in the arid Zarand region, Iran: comparing logistic regression approaches to predict diagnostic horizons and soil types. *Eur J Soil Sci* 63: 284–298. <https://doi.org/10.1111/j.1365-2389.2012.01425.x>
10. Lilly A, Lin H (2004) Using soil morphological attributes and soil structure in pedotransfer functions, In Pachepsky Y, Rawls WJ, *Developments in Soil Science* 30, Amsterdam: Elsevier Ltd., 115–141. [https://doi.org/10.1016/s0166-2481\(04\)30007-3](https://doi.org/10.1016/s0166-2481(04)30007-3)
11. Tongway D, Hindley N (1995) Manual for Soil Condition Assessment of Tropical Grasslands. Csiro. Australia. Available from: <https://publications.csiro.au/rpr/download?pid=procite:fa79052c-07eb-4345-a001-43e68c86ec4a&dsid=DS1>.
12. Weil R, Islam K, Stine M, et al. (2003) Estimating active carbon for soil quality assessment: A simplified method for laboratory and field use. *Am J Altern Agric* 18: 3–17. <https://doi.org/10.1079/AJAA200228>
13. McLean E (1982) Soil pH and lime requirement, In Page AL (ed.), *Methods of soil analysis Part 2*. 2nd ed. Agron. Monogr. 9. ASA, Madison, WI, 199–224. <https://doi.org/10.2134/agronmonogr9.2.2ed.c12>
14. FAO: Locate Climate Estimator (New_LocClim), 2005. Available from: <http://www.fao.org/land-water/land/land-governance/land-resources-planning-toolbox/category/details/en/c/1032167/>.
15. Tóth G, Hermann T, Tóth B (2016) Report on hierarchical and multi-scale pedoclimatic zonation. Report iSQAPER-W2-D2.1-001.

16. Hanley J, McNeil B (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143: 29–36. <https://doi.org/10.1148/radiology.143.1.7063747>
17. Arboretti Giancristofaro R, Salmaso L (2003) Model performance analysis and model validation in logistic regression. *Statistica* 63: 375–396. <https://doi.org/10.6092/issn.1973-2201/358>
18. Van Smeden M, Moons KG, de Groot JA, et al. (2019) Sample size for binary logistic prediction models: Beyond events per variable criteria. *Stat Methods Med Res* 28: 2455–2474. <https://doi.org/10.1177/0962280218784726>
19. Peduzzi P, Concato J, Kemper E, et al. (1996) A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* 49: 1373–1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)
20. Hsieh F, Bloch D, Larsen M (1998) A Simple Method of Sample Size Calculation for Linear and Logistic Regression. *Statist Med* 17: 1623–1634. [https://doi.org/10.1002/\(SICI\)1097-0258\(19980730\)17:14<1623::AID-SIM871>3.0.CO;2-S](https://doi.org/10.1002/(SICI)1097-0258(19980730)17:14<1623::AID-SIM871>3.0.CO;2-S)
21. Garosi Y, Ayoubi S, Nussbaum M, et al. (2022) Effects of different sources and spatial resolutions of environmental covariates on predicting soil organic carbon using machine learning in a semi-arid region of Iran. *Geoderma Reg* 29: e00513. <https://doi.org/10.1016/j.geodrs.2022.e00513>
22. Slessarev EW, Lin Y, Bingham NL, et al. (2016) Water balance creates a threshold in soil pH at the global scale. *Nature* 540: 567–569. <https://doi.org/10.1038/nature20139>



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)