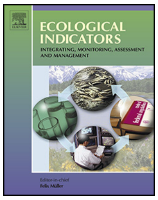
Ecological Indicators 63 (2016) 366–373

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



Explore the influence of soil quality on crop yield using statistically-derived pedological indicators

Katalin Juhos[a,*], Szilárd Szabó[b], Márta Ladányi[c]

[a] Department of Soil Science and Water Management, Faculty of Horticultural Science, Corvinus University of Budapest, 29-43 Villányi St., H-Budapest, Hungary
[b] Department of Physical Geography and Geoinformatics, Faculty of Science and Technology, University of Debrecen, 1 Egyetem tér, H-4032 Debrecen, Hungary
[c] Department of Biometrics and Agricultural Informatics, Faculty of Horticultural Science, Corvinus University of Budapest, 29-43 Villányi St., H-Budapest, Hungary

ARTICLE INFO

Article history:
Received 27 May 2015
Received in revised form 4 December 2015
Accepted 16 December 2015
Available online 6 January 2016

Keywords:
Principal component regression analysis
Soil productivity function
Soil quality index
Winter wheat
Maize
Solonetz
Chernozems

ABSTRACT

The aim of this paper is (1) to find and discuss the best multivariate statistical method in exploring the soil productivity function in an East-Hungarian region; (2) to evaluate and interpret the edaphic indicators and Hungarian soil quality index (HSQI); and (3) to identify the main determinant factors and indicators in this region. Soil pH, carbonate content, soluble and exchangeable Na^+, clay, humus, available phosphorus and potassium content were analyzed. Topographical position and HSQI were evaluated as well. Yield data (maize, winter wheat, sunflower) of 10 years were standardized using calculated relative yield of each crop. Having simple indicators, stepwise linear regressions for mean relative yield were inadequate for choice uncorrelated indicators which have significant influence on yields. The variables were analyzed using principal component analysis (PCA) with Varimax rotation. According to the eigenvalues greater than 1, the PCA yielded three principal components (PCs) explaining a total of 89.471% of the variance for the entire data set. These factors could be well interpreted as derived complex indicators. Having the three PCs, a stepwise linear regression process (PCR) was conducted with dependent variables mean relative yield. The explained variance for mean relative yield was as high as adjusted $R^2 = 0.771$ ($p < 0.001$). The three PC factors together explained the mean relative yield better than the simple indicators and the HSQI. So, the variables can effectively explain the yield and the variability together with other variables as linear combinations. Consequently, PCR is a successful method to reveal the site specific relationship between soil properties and yields and to revision the HSQI at local level.

1. Introduction

Understanding the variability of landscape and soil properties and their effect on crop yield is a critical component of site-specific and sustainable management systems and land use planning. Soil quality has been defined in several ways. The main point of the most cited definitions is that soil quality is the capacity of soils to function within ecological and land-use boundaries, to sustain productivity (Doran and Parkin, 1994; Karlen et al., 1997). This concept is often considered non-objective. For these reasons Letey et al. (2003) and Rousseau et al. (2012) claimed that soil quality could not be measured itself and that only management practices should be compared to determine which one enhanced some particular soil indicator at stake. According to our definition, soil quality for the agricultural region and soil use we studied means (1) limitations and the capabilities of the soil properties for land uses, (2) the variety of growable cultures and the level of production (suitability). The so-defined soil quality (productivity function) can be expressed by yields of crops and pedological indicators which define the most limited and important factors. Determination of soil quality is very difficult because the relationship between crop yield and soil quality is very complex and depends on complex interactions among physical and chemical properties of soil and other external natural factors. Several methods of soil quality evaluation have been developed that use several statistical methods. The key to an appropriate method is the choosing and interpretation of the indicators.

Various statistical models were employed to get prediction of yield. In particular, the functional linear model with scalar response is surely the model that has attracted more attention, e.g., several authors found that pedological indicators have a significant

---

* Corresponding author. Tel.: +36 15826468.
E-mail address: katalin.juhos@uni-corvinus.hu (K. Juhos).

http://dx.doi.org/10.1016/j.ecolind.2015.12.029
1470-160X/

relationship with crop yield by using stepwise multivariate linear regression analysis (SMLR) (Andrews and Carrol, 2001; Drummond et al., 2003; Rezaei et al., 2006; De Araujo et al., 2009). Two of the most important methodologies used to estimate the parameters of the functional linear model with scalar response are principal component regression (PCR) and partial least-squares regression (PLS) (Febrero-Bande et al., 2015). Loadings from linear combinations of variables in PLS allowed identifying the soil properties that have the greatest influence on yields (Corwin et al., 2003; Ping et al., 2004). Another option is combining variables based on their linear correlation using principal component analysis (PCA). PCA is an effective dimension reduction technique, but the derived PC factors do not explain the total variance of the entire set of variables. Several author analyzed the relationship between soil properties and yields by conducting principal component regression process (Mallarino et al., 1999; Cox et al., 2003; Shukla et al., 2004b; Ayoubi et al., 2009; Cheng et al., 2016). Classification and regression trees (CARTs) are more robust techniques due to its low prediction error (Tittonell et al., 2007; Zheng et al., 2009; Ahman and Bhatti, 2015). Drummond et al. (2003) investigated several types of supervised feed-forward neural networks in identifying methods capable of relating soil properties and grain yields on a point-by-point basis within 10 individual site years. Suitability of these methods depends on the structure, size of the database.

There are several methods to define the soil quality quantitatively as an index. These soil quality indices are the most commonly used today to evaluate impacts of agricultural practices on the soil quality. The development of a soil quality index should follow three steps: (1) selection of indicators, (2) scoring the selected indicators and (3) integration of indicators in an index. Among the pedological indicators selection methods, total data set (TDS) and minimum data set (MDS) have been widely used. The MDS is a selected collection of indicators chosen according to expert opinion (Andrews et al., 2002; Lima et al., 2013) or several multivariate statistical methods (Andrews et al., 2002; Sharma et al., 2005; Masto et al., 2008; Qi et al., 2009; D'Hose et al., 2014; Rahmanipour et al., 2014). The second step is normalizing the MDS indicators by different numerical scales (usually between 0 and 1) using linear (Liebig et al., 2001; Sharma et al., 2005) and non-linear scoring functions (Andrews et al., 2002; Zhang et al., 2004; Qi et al., 2009; Lima et al., 2013; D'Hose et al., 2014; Rahmanipour et al., 2014). The integration of normalized indicators into quality indices is possible through additive (Karlen et al., 1997; Andrews et al., 2002; Lima et al., 2013; D'Hose et al., 2014), multiplicative (Zhang et al., 2004; Sharma et al., 2005; Masto et al., 2008) and other (Qi et al., 2009; Rahmanipour et al., 2014) techniques. Since the indicators have different importance in land productivity, almost all of authors weighted the normalized indicators. We consider these indices partly subjective in view of their indicator interpretation and weighting methods. To develop an index which expresses the productivity function of soils, first it is necessary to reveal the complex relationship between soil properties and yields, then the main limiting factors are needed to be identified.

The traditional Hungarian soil quality evaluation has been based on the units of genetic soil taxonomy (Stefanovits et al., 1972). Every genetic soil type is assigned a minimum and maximum number and the Hungarian soil quality index (HSQI) of any specific soil unit is calculated through some deductions from maximum values according to the specific properties of the soil. However, the reasons for the yield variability are still poorly understood (Máté, 1999; Dömsödi, 2007). According to our hypothesis, neither simple pedological indicators nor HSQI are suitable to reveal the relationship between soil properties and crop yield in our region. Other kind of derived or complex indicators are necessary to be calculated using multivariate statistical tools. The specific objectives of this study were: (1) to find a suitable multivariate statistical method to explore the relationship between soil properties and crop yield in an East Hungarian region; hereby (2) to evaluate and interpret the edaphic indicators and HSQI in regard to soil productivity function; (3) to identify the main determinant factors and the most usable indicators in this region.

2. Materials and methods

2.1. Site description

The research site is an alluvial plain located in East Hungary (21°13′ E, 47°17′ N). The warm temperate climate is characterized by cool winters and hot, dry, droughty summers, mean annual precipitation of 580 mm and mean annual temperature of 10.5 °C (Fábián and Matyasovszky, 2010). Depth of groundwater table is approximately 50–300 cm. Soils were developed on alluvial deposits with loam, loamy clay and clay texture. Soils can be classified as *Chernozems*, *Solonetz* and *Gleysols* (FAO, 2014).

The agricultural management practice and crop rotation were uniform on every plot and characterized by conventional tillage in a nonirrigated system and the fertilization was done by nitrogen only. Plot areas ranged from 1.04 to 31.90 ha and the total area covers 225 ha (Fig. 1).

2.2. Processing of yield data (productivity indices)

Winter wheat (*Triticum aestivum* L.), maize (*Zea mays* L.) and sunflower (*Helianthus annuus* L.) yield data were collected at each plot for 10 years. Productivity of plots were expressed by standardization of yield data as follows:

Relative yield was first calculated:

$$\mathrm{RY}_p = \frac{Y_p}{Y_{\max}},$$

where RY_p is relative yield of plot p (a value between 0 and 1), Y_p – yield of parcel p (t ha^{-1}), $Y_{\max}$ – maximum yield on the total research site over all parcels (t ha^{-1}). Then mean relative yields ($\overline{RY_p}$) for 10 years were calculated for each plots p.

2.3. Soil survey and analysis (simple indicators and HSQI)

Soil survey was conducted for analysis of physical and chemical properties that were previously considered important in land-use management in our region; microbiological parameters are considered too dynamic attributes to these objectives (Tóth, 2011). Soil samples were collected at 0–100 cm at 20 cm depth increments (Fig. 1). Soil pH was measured in a 1:2.5 soil/KCl mixture potentiometrically. Determination of the total carbonate content ($CaCO_3$) was conducted according to the volumetric method (MSZ, 1978a). Determination of soluble and exchangeable sodium (AL-Na) based on extraction with acid ammonium lactate (Egnér et al., 1960). Particles smaller than 0.002 mm (clay) were determined by the pipette method using preparation with sodium pyrophosphate (MSZ, 1978b). Humus was measured by the Tyurin method (Kotroczó et al., 2014) and it was given as t ha^{-1} based on the content values and soil bulk density. Homogeneous soil units were delimited based on soil survey. Mean soil properties (at 0–100 cm depth) and surface elevation were calculated for the plots (Fig. 1). Soil samples composed of 20 subsamples were collected for analysis of potentially available potassium and phosphorus nutrients at 0–30 cm depth from the 28 plots every year (2004–2013). Available phosphorous (AL-P_2O_5) and potassium (AL-K_2O) content were determined with acidic ammonium lactate extraction (Egnér et al., 1960). These annual nutrient contents were averaged because they did not change significantly over the 10 years. The nutrient contents were given as mg kg^{-1} since the soil bulk density in the