Solid Earth Discuss., 6, C1526–C1532, 2015
www.solid-earth-discuss.net/6/C1526/2015/

Solid Earth
Discussions
Open Access

Interactive comment on "Evaluation of soil fertility in the succession of karst rocky desertification using principal component analysis" *by* L. Xie et al.

Anonymous Referee #1

Received and published: 29 January 2015

Manuscript/Discussion paper: Solid Earth Discuss., 6, 3333-3359, 2014 «Evaluation of soil fertility in the succession of karst rocky desertification using principal component analysis, by Lianwu Xie et al.»

General comments

In this paper the authors investigated the changes in 19 different soil fertility-related variables along a gradient of karst rocky desertification (RD), in five different counties belonging to the central Hunan Province (SW China). By applying PCA analysis to the soil data matrix they obtained a standardized integrate fertility indicator whose averaged scores matched (some of) the four predefined RD grades. Taking into account the correlations between the calculated integrated fertility and the measured variables they concluded that the most useful variables to be used as indicators to evaluate soil integrated fertility were: total soil organic C and P; C, N and P content in the soil microbial biomass; cation exchange capacity and bulk density. In my opinion, the research is relevant and matches the scope of SE. Besides, the authors gathered a valuable set of soil chemical, physical and biological data on different sites/plots affected by different levels of RD. The sampling design, sample processing and lab parameter determination seem also to be solid. My main criticism to this work is related to the data analysis and interpretation, and the extent to which the statistical analyses carried out in this paper are suitable and actually support the conclusions of the paper.

Specific comments.

In the following paragraphs I will comment in detail the different sections of the paper:

1) The introduction is clear and focused on the importance of efficiently assessing soil fertility in order to study the changes occurring along the RD process while minimizing the cost of determining unsuitable or redundant soil indicators. The objectives of the work are indicated, although no working hypotheses are stated.

2) Methods -Sections from 2.1 to 2.4 are, in general, clear and informative. Nevertheless, regarding the table 1, it is no clear how the grades used to select the plots were defined. Did you mean that all the plots belonging to a given RD grade must simultaneously satisfy the pre-established ranges for each of the three variables (soil depth, vegetation coverage and bedrock exposure) used here for the classification? What happens when the value of one (or two) of these variables is outside the range defined for the same row (Grade) in the remaining variable(s)?

-Statistical analysis In section 2.5. Authors said that they used ANOVA “to test homogeneity of variance”, while ANOVA is used to test the overall homogeneity of MEANS. However, homogeneity of variances is a pre-requisite to apply parametric ANOVA, and is tested by using different ad-hoc tests (Bartlett, Brown-Forsythe, Levene etc). If the homocedasticity hypothesis is rejected you must either transform variables or use specific approaches (e.g. Welch) or use a non-parametric test (as Kruskal-Wallis, for example). Soil variables are frequently highly skewed and transformations are usually necessary for meeting the assumptions required for the parametric test used by the authors, but no indication related to this problem was found in this section. On the other hand, they used repeated paired t-tests to make post-hoc comparisons between the four RD grades for each of the measured variables. In my opinion, it was not a good idea, unless you apply an additional procedure to avoid the associated inflation of type I error (such as a Bonferroni or Bonferroni-like correction). Alternatively ad hoc test to perform post hoc comparison while controlling Type I error (Tukey, Duncan, etc) could be used.

In the section 2.6, authors proposed an integrated soil fertility index whose value for each sample is calculated by summing the scores of the sample in each of the selected ("meaningful") components, weighted by the fraction of the total variance explained that is associated to each component. However, they do not argue at all why this particular numerical combination of latent variables is an optimal integrated indicator of the overall soil fertility (i.e. the ability of soil to store and supply nutrient for plants / to serve as a suitable substrate for plant growing). A priori, I do not see any advantage of using this particular mix of standarized components, which are also difficult to interpret, instead a standardized set of well-known measured variables whose meaning is broadly known for everybody. I will expand on this issue later.

3) Results In the section 3.1, authors present the results of the pairwise comparisons between the different RD grades, for the 19 measured indicators (table 2). As I commented above, the significance of the results in table 2 may be flawed and should be revised/clarified taking in account my previous warnings on normality, homocedasticity and methods for correctly performing post hoc multiple comparisons, avoiding type I error inflation and spurious significances. Additionally, note that the high number of tests performed in table 2 produce a large tablewise type I error inflation. To avoid this prob-

SED
6, C1526–C1532, 2015

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

lem, I suggest to use some Bonferroni or FDR corrections (see Rice 1989 and García 2004). The same can be applied for the multiple significances marked in the correlation matrix in table 3. Some of them cannot be probably considered as significant in such a context of multiplicity. The same applies for correlations in table 5.

Section 3.2. Even in the case that the significance of the results shown in Table 2 were not flawed, I could not understand why have been included in the set of variables to be analyzed by PCA up to seven variables that, according to table 2, were not significantly different for any pair RD grades. If one of the main objectives of the paper is to find an integrated indicator to evaluate fertility in lands with different RD grades, eliminating noise and reducing the number of variables to be measured, it could be reasonable to retain into the ulterior PCA analysis only those measured variables that exhibited significant differences between, at least, two RD grades, and exclude those variables with no significant differences between any pair of them. Aditionally, perfectly correlated variables (as BD and TOP in table 3) should not be included simultaneously in the PCA analysis, since it could generate artifacts. On the other hand, I disagree with the interpretation (and the names) given by the authors to the different extracted components in the page 3422 (lines 10-26). I think that they are rather arbitrary and may induce to confusion. For example, PCA1 is interpreted as a synthetic indicator of soil water holding capacity / permeability, because of its high correlation with CMC, FMC and TOP. However, according to loadings in table 4, this PC also exhibited high correlations with total nutrient contents (TN, TP) and nutrient holding capacity (CEC). Additionally, organic carbon (TOC) had its highest loading in this component. Therefore, one could think that this component summarizes a set of soil physico-chemical properties (bulk density, water and nutrient holding capacity and nutrient content) which may by directly related with soil organic content. PC2 is considered as an “organic matter component”, despite that its correlation to TOC is lower than that with PC1. Some indicators (available P, microbial P, microbial C and Microbial C) showed significant positive correlations with this component and with TOC. But a similar number of indicators -having similar loadings in PC2 (0,5-0,7, in absolute value): pH, TK, CAP- were nor correlated with

SED
6, C1526–C1532, 2015

Full Screen / Esc
Printer-friendly Version
Interactive Discussion
Discussion Paper

TOC. Looking at these values one could think that PC2 may be interpreted as a combined mixed gradient revealing relationships between soil TOC and MBC, and between soil P availability and MBP, modulated by soil pH. That could be a manifestation of the coupling between soil nutrient content/ availability and the activity of the microbial community. In summary, the main (first few) components are complex mixed gradients which are closely related to variables of different nature and have not an unambiguous meaning. On the other hand, the average integrated soil fertility scores largely fluctuated among sites for all RD grades (fig. 1), and even the mean across-site values for the four defined RD grades (fig. 2) did not show significant differences for most of the paired RD grade comparisons, except for the extreme ones (comparisons of PRD or LRD with IRD). In summary, you cannot adequately estimate neither the mean value of the integrated fertility of an area by knowing its RD grade nor viceversa, but only in the extremes of the RD gradient.

Discusion In the section 4.1 one expected to find some clear ideas about what happens with the measured soil fertility-related parameters along the succession of RD grades and the implications of these changes. Instead, a confusing discussion including references to some previously named PCA components is found. However the authors produced a nice table of data (table 2) on many interesting parameters averaged for each RD grade. This table, easily understandable, clearly shows that the average values of TOC, TN, AP, MBC and MBN perfectly matched the pre-defined succession of RD grades, and that TP and MBP showed a similar tendency. I think that PCA analysis did no add nothing to these results, but only confusion.

In the section 4.2, the fact that the average values of the integrated fertility-PCA based index do not match the RD grades is discussed. The explanation for this lack of concordance is that the reference RD classification system was not probably satisfactory. But this lead to a circular argument: while testing the ability of the calculated integrated fertility to predict the predefined RD levels, the resulting lack of matching led them to question the reference RD system itself, instead to questioning/ rejecting the proposed

SED

6, C1526–C1532, 2015

integrated index. However, looking at table 2 one can conclude that many relevant variables related to soil fertility (see the precedent paragraph) change, on average, as expected in a gradient of soil degradation. Mixing these key relevant variables with other non-relevant or non-affected ones, by applying PCA and making a confusing integrated index, lead to obscure the relevant relationships that one can directly observe in table 2.

In the section 4.3 authors selected as the best (sensitive) indicators to evaluate RD those that were best correlated to the calculated integrated fertility index. Most of them (TOC, TN, TP, MBC, MBN, MBP) changed individually across the RD succession as expected (table 2), while the two remaining (CEC, BD) did not changed at all across RD grades (table 2), despite the integrated index was significantly different when extreme grades were compared (p. 3343, line 6-10). In summary, there is no reason to think that a significant correlation with the calculated integrated fertility is a good criterion to select a key soil indicator to evaluate RD.

Finally, I would like to remark some points related to possible improvements for future versions: -the authors gathered a valuable set of multivariate soil data from different sites affected by Rocky Desestification.. -by (correctly) comparing statistically the average values of the measured parameters in each of the pre-defined RD-level they can detect the most relevant. -by analyzing redundancies among the relevant set (analizing thee correlation matrix and/or using PCA) they could refine the selection of indicators without applying complex, not useful and unnecessary statistical procedures. -An elegant multivariate alternative which would overcome most of the above-cited problems could be a canonical analysis using the three indicators used to define RD grades (i.e. soil depth, vegetation coverage and bedrock exposure) as dependent set. This will avoid the constraints associated to the RD grades defined in table 1. As an independent set, the non-redundant measured variables could be used. In this way you get the best combination of soil indicators that maximizes the optimal combination of RD indicators. Both manually (by analizing the correlations between canonical axis and the measured variables) or by objective procedures (stepwise or AIC criterion) an optimal soil variable set (i.e. taking into account both explained variance and model parsimony) could be obtained.

In summary, - In my opinion the work is not publishable in the present form and should be completely rewritten on the basis of the valuable data base generated in the work, and resubmitted for reevaluation. -In think that both PCA and the integrated fertility index introduce unnecessary complexity and are useless to reach the objectives of the papers.

Minor comments. - p. 3335, lines 10 -15: information is redundant with that in p.3337, lines 18-20, and partially redundant with Table 1. Please simplify. -An abbreviation (as ISF) is needed to avoid repeating "integrated soil fertility" many times throughout the paper. - The PCA info (p. 3336, lines 17-21) should be moved to the Methods section. - Table 1, PRD row: conversation or conservation ? -Figure 2: Explain in the legend what the bars and whiskers represent (mean $\pm$ std.error?). -Figure 2, legend: The pairwise comparisons made here seem not to match those performed in the text (p. 3343, lines 7-10). For example, the comparison PRD-IRD in the legend of fig. 2 seem to have an associated p=0.120 (i.e. a not significant difference), while the same comparison in the text seem to have an associated p=0.008 (i.e. a highly significant difference). Please clarify this.

Cited References. Rice, W.R. (1989). Analyzing tables of statistical tests. Evolution 43: 223-225. García, L.V. (2004) Escaping the Bonferroni iron claw in ecological studies. Oikos

Interactive comment on Solid Earth Discuss., 6, 3333, 2014.