SOIL Discuss.,
https://doi.org/10.5194/soil-2019-77-AC1, 2020

Interactive comment on "Soil classification based on spectral and environmental variables" *by* Andre Carnieletto Dotto et al.

Andre Carnieletto Dotto et al.

andrecd@usp.br

Received and published: 5 February 2020

Referee #1 Referee: For a good while I was also wondering if the proposed methodology, which is k-means clustering, is an adequate and sufficient method for new soil classes retrieving. Answer: K-means clustering is a widely used partitioning method not limited only in soil science but in many other areas. It is widely used as partitioning method to grouping similar sets of data. As every technique, k-means clustering also present its advantages and disadvantages. We selected k-means because its quantization advantages: relatively simple to implement, scales to large data sets, can set the positions of centroids, easily adapts to new examples, and generalizes to clusters of different shapes and sizes. Referee: the classification presented here will do a better job compared to traditional soil taxa (e.g. WRB) or not. Answer: As we have mentioned

in the manuscript, our aim is to provide a new and quantitative approach to define the soil classification based on proximal and remote sensors data. The biggest problem in soil classification is the need for experienced personal (pedologists) to perform the classification. Furthermore, the process requires tacit knowledge from the professional and is relatively subjective. Therefore, the classification process is costly, time demanding and might provide results with large uncertainty. We kindly ask the referee to observe the fact that "doing a better job" is not only related to the reduction in classification uncertainties, but also to define means to perform a fast and reliable classification, which can support soil characterization and consequently soil-related decision processes. We do acknowledge that soil spectroscopy is not capable to distinguish some of the important variables required in conventional classification, e.g. morphological attributes. Some classes share many soil properties and even environmental characteristics and are more difficult to distinguish. However, other soil classes are relatively distinctive, consequently, it is possible to categorize them. Besides that, soil type differentiation based on the Vis-NIR spectra takes into consideration some of the most important attributes, such as soil colour, iron oxides, clay minerals, carbonates and organic matter. The soil spectral classes conducted to incorporate applicable soil data for agricultural management, with less interference of personal/subjective/empirical knowledge, and more reliable on automation measurements by sensors. Referee: Please describe in detail how AIC was computed and provide an equation. Answer: We added more information in the methodology to describe with more detail the AIC analysis. "To determine which number of clusters appears to best describe the data, i.e. the optimal number of clusters, the Akaike information criterion (AIC) was performed. To calculate the AIC, we applied the function kmeansAIC from kmeansstep R package. It calculates the AIC value of a specific k-means cluster and it specified centroids. The AIC was implemented using the values from 1 to 15 clusters. The analysis was performed by data driven 30 times. The overall modal cluster with the lowest AIC value was selected and assumed to be the optimal number of clusters, which can represent the most appropriate number of SSC classes." Referee: Explain why do think the traditional classification

system WRB does not take into account climate and terrain information. Answer: We do not agree that the traditional classification systems does not take into account climate and terrain data, in fact, we have verified exactly the opposite in this manuscript. But we believe that it is difficult to take into account the climate and terrain data in the classification approach. As alternative, we applied remote sensing images as a source of soil-landscape information. By the digital elevation model is possible to extract several terrain attributes that are taken into consideration in the soil surveys, for example. Along with spectral data, we added these climatic and terrain data into the modelling. Adding climatic and terrain data into the calibration model, provided an improvement in the prediction of soil classes. The increment of climatic and terrain data aimed to incorporate the Pedologist's impersonal knowledge on environmental into the quantitative modelling.

Interactive comment on SOIL Discuss., https://doi.org/10.5194/soil-2019-77, 2019.