Suggestions for improvement

While I appreciate the effort made by the authors in responding to my comments, I still believe they have not addressed the main problems, i.e., (1) mismatch between their map (i.e., EthioSoil Grid map) and the Soil Atlas of Africa, and (2) relatively low accuracies of classification. The EthioSoil Grid map is also inconsistent with the SOTER Map of Ethiop (Figure 3.3) produced by Elias (2016)[1]. In my opinion it is better to direct efforts towards identifying the reasons for these discrepancies and fine-tuning the analysis accordingly to produce a map that achieves greater accuracy. The overall accuracy of 56.2% reported in the paper does not inspire confidence. The Kappa coefficient of 48% is even less comforting; it tells us that the classification is not better than can be found by chance alone (random). If Kappa was higher (e.g., >60% or more, we could confidently say that the classification was significantly better than a random association. I extracted the producer and user accuracy from Table 1 of Ahenafi et al and calculated the omission and commission errors for each reference soil groups (RSGs) (see table below). Evidently, the errors of classification are large (orange) or unacceptable (red) for most RSGs. The reason is partly a sampling problem. Note in the table below and the following graph that as the sample size for RSGs decreased, the omission and commission errors increased. Even with large sample sizes some RSGs (e.g., Luvisols and Cambisols have been classified with >50% error.

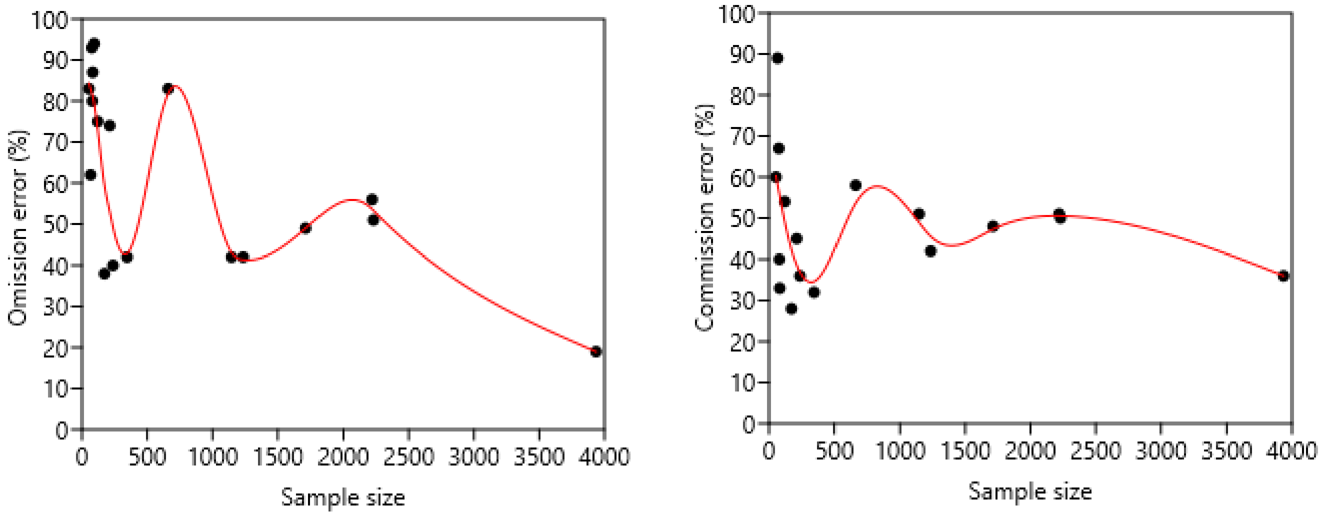
Reference soil group	Sample size (N)	Accuracy		Error		
		Producer	User	Omission	Commission	Comment
Vertisols	3935	0.81	0.64	19	36	Acceptable error (AE)
Cambisols	2219	0.44	0.49	56	51	Unacceptable error (UE)
Luvisols	2229	0.49	0.50	51	50	UE
Leptosols	1713	0.51	0.52	49	48	Large (LE)
Fluvisols	1235	0.58	0.58	42	42	LE
Nitisols	1146	0.58	0.49	42	51	LE
Regosols	660	0.17	0.42	83	58	UE
Alisols	345	0.58	0.68	42	32	LE
Andosols	237	0.60	0.64	40	36	LE
Arenosols	212	0.26	0.55	74	45	UE
Calcisols	172	0.62	0.72	38	28	AE
Solentz	120	0.25	0.46	75	54	UE
Lixisols	93	0.06	1.00	94	0	UE due to very small producer N = 1
Phaeozeme	82	0.13	0.67	87	33	UE due to very small user N
Soleschaks	79	0.20	0.60	80	40	UE due to very small user N
Acrisols	75	0.07	0.33	93	67	UE due to very small user N
Planosols	66	0.38	0.11	62	89	UE due to very small user N
Gleysols	53	0.17	0.40	83	60	UE due to very small user N
Umbrisols	28					
Ferralsols	14					Conditions in Ethiopia do not favour
Gypsisols	11					
Plinthosols	6					Conditions in Ethiopia do not favour
Stagnosols	2					

Omission error = 100-producer accuracy (in %); Commission error = 100-user accuracy (in %);

Note that the discrepancy between omission and commission errors is higher when sample sizes are small as in Gleysols, Planosols, Acrisols, Solenshaks, Phaeozemes and Lixisols. Note also in the figure below that both omission and commission errors vary widely when sample sizes are small, but they tend to be small with increasing sample size. That means that for the map to be improved there is a need for soil sampling in

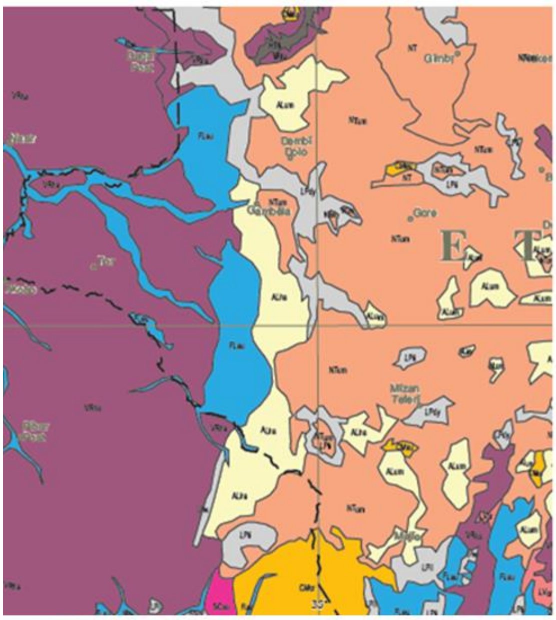
[1] Eyasu Elias, 2016. Soils of the Ethiopian Highlands: Geomorphology and Properties. CASCAPE Project, ALTERA, Wageningen University and Research Centre (Wageningen UR). The Netherlands. 385pp

the areas not covered in this analysis. No amount of modelling sophistication can compensate for small sample size.

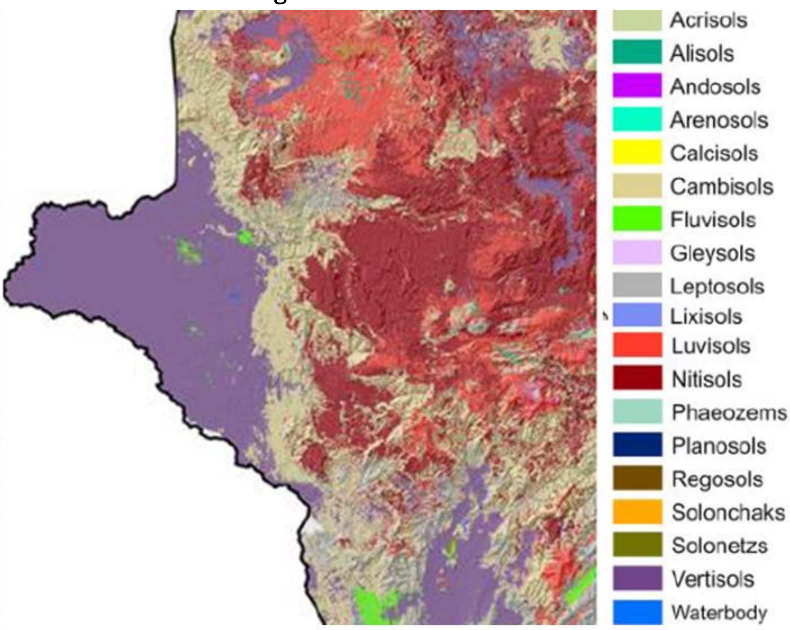


Trends in omission and commission errors with sample size. Red lines represent smoothing splines

Even for Vertisols, which had the largest sample size, the commission error is large (36%), and the probability that the points on the map will actually be present on the ground are lower than we are made to believe. Note that the producer's accuracy for the Vertisols is 81% but the user accuracy is 64%. This means that although 81% of the reference Vertisols have been correctly identified as "Vertisols" on the map, only 64% of the areas identified as "Vertisols" in the classification were actually Vertisols. A good example of this problem is revealed by the following maps covering the Western part of Ethiopia. The EthioSoil Grid map shows an entire region covered purely by Vertisols. According to the Soil Atlas of Africa, large areas are covered by Fluvisols (along the Akobo, Gilo, Baro Rivers and their tributaries). Another discrepancy between the Soil Atlas of Africa and the EthioSoil Grid map is absence of Alisols in this region. I believe these discrepancies were caused by inadequacy of legacy data available to the authors (see Figure 5 in Ashenafi et al.). Note that Alisols are one of the most problematic soils (Al toxicity, low pH, susceptibility to drought) and their misclassification in this map can create problems for their sustainable management. A third problem with the EthioSoil Grid map is that it shows other soils (Acrisols? Cambisols?) in areas typically dominated by Nitisols. Nitisols are prone to acidification with fertilizer use, and their misidentification as other soils may create problems for fertilizer management.

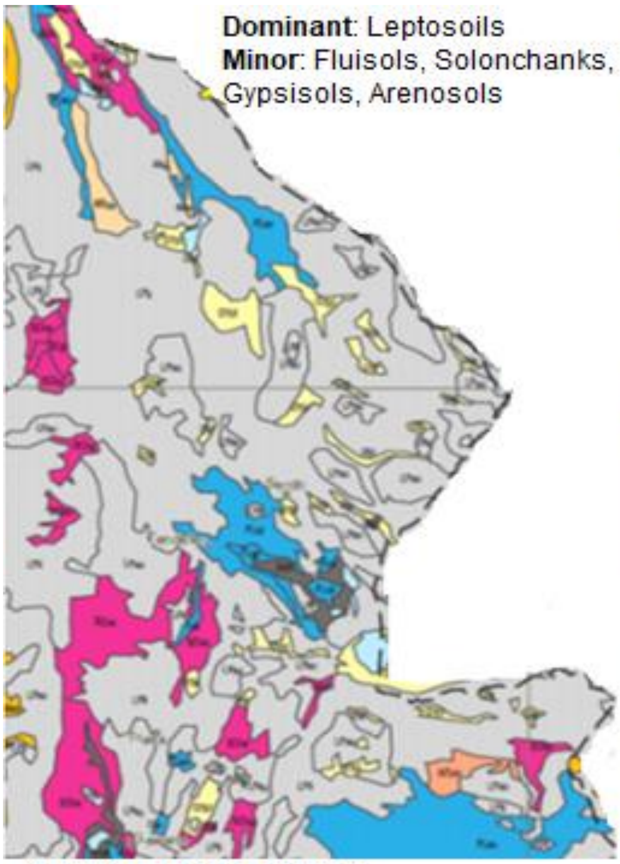


Soil Atlas of Africa

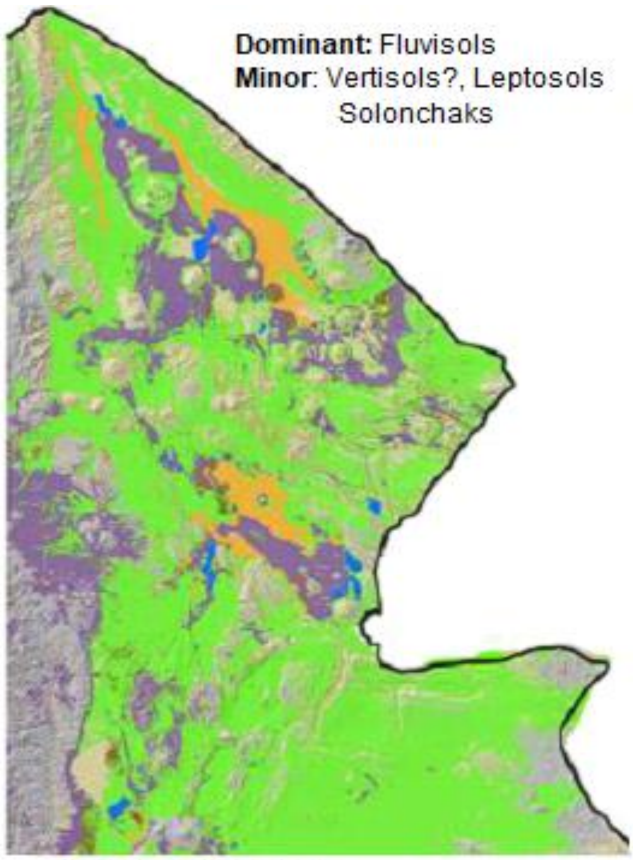


EthioSoilGrids 1.0

The classification of Cambisols and Luvisols is also problematic. Note that the producer and user accuracy for Cambisols are 44% and 49%, respectively. This means that even though 44% of the reference Camisols have been correctly identified as “Cambisols”, only 49% of the areas identified as “Camisols” in the classification were actually Cambisols. The omission error (56%) and commission errors (51%) are too big. Examination of the Confusion Matrix reveals that Cambisols were misclassified in large number of cases as Luvisols (47), Leptosols (35), Fluvisols (28), Vertisols (28), Nitisols (16), Regosols (16), etc. Similarly, Leptosols were misclassified as Cambisols (47), Regosols (32), Luvisols (27), Vertisols (24), Fluvisols (11), etc. That is probably why Cambisols and Fluvisols were found on the map in areas where normally other soil types were expected to occur. For example, the following maps clearly shows the mismatch between the Soil Atlas of Africa and the EthioSoil Grid map (Ashenafi et al.). I suspect the mismatch was partly caused by inadequacy of legacy data from the region mapped here (see Figure 5 in Ashenafi et al.). Evidently, EthioSoil Grid map shows Fluvisols and Vertisols? even in mountainous areas, where conditions for the genesis of Fluvisols and Vertisols do not exist in this region. The Soil Atlas of Africa, on the other hand, shows Fluvisols only around water bodies, where Fluvisols are likely to develop. It also clearly shows other RSGs including the qualifiers (divisions within RSGs), for example as Calcic Fluvisols, Eutric Fluvisols, Salic Fluvisols). That kind of detail is more useful to decision-makers than just saying Fluvisols.



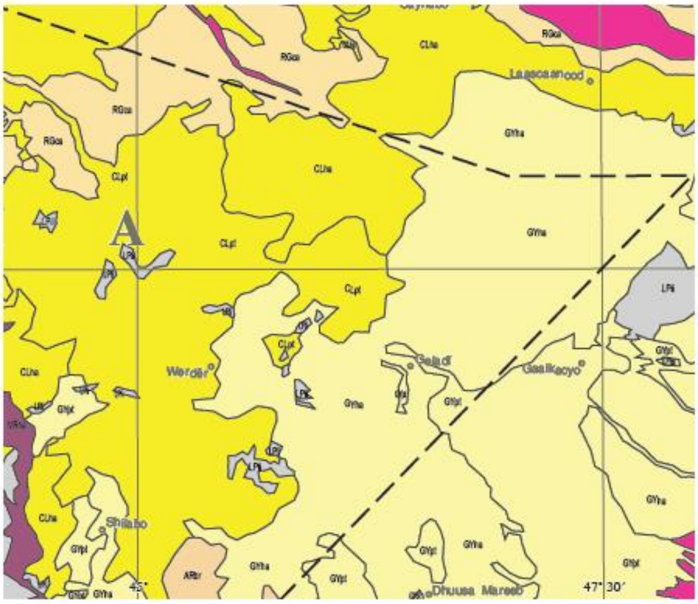
Soil Atlas of Africa (2014)



Ethiosis SoilGrid 1.0

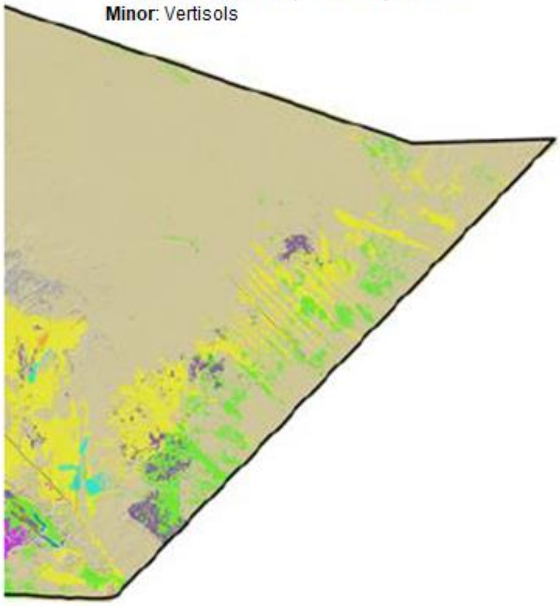
The following maps shows another huge discrepancy between the Soil Atlas of Africa and EthioSoilGrid. Here, EthioSoilGrid shows Cambisols and Fluvisols covering a large proportion of the land although Cambisols and Fluvisols are least expected. I suspect the mismatch was partly caused by inadequacy of legacy data from this region (see data gaps in Figure 5 of Ashenafi et al.). Indeed, the sample size for Gypsisols was very small. So, I also suspect that Gypsisols were not included in the training dataset. Therefore, the algorithm misclassified the RSGs as Cambisols based on the information it has. In machine learning this is called **association bias**. As far as your machine learning model is concerned, Gypsisols do not exist.

Dominant: Calcisols, Gypsisols, Regosols
Minor: Vertisols, Arenosols



Soil Atlas of Africa

Dominant: Cambisols, Calcisols, Fluvisols
Minor: Vertisols



EthioSoilGrids 1.0

In summary, I believe that inadequacy of sample sizes for some RSGs, paucity of samples from some areas and the over-representation of the dataset by a few RSGs (e.g., Vertisols, Cambisols and Luvisols), has probably led to large classification errors. In addition, the covariates included/not included in the machine learning algorithm and the use of validation methods which do not sufficiently control overfitting could bias results with small sample size. Simulation studies[2] show that K-fold Cross-Validation produces strongly biases performance with small sample sizes, even with sample size of 1000. There is a strong need for closely looking at the previous analysis and redoing the map so that it can be used as a reliable decision support tool. I encourage the authors to explore (1) opportunities to include other variables not included in the present analysis, (2) dimension reduction, (3) use of other cross-validation methods, and (4) use of an ensemble approach to see whether overall accuracy could be improved and classification errors reduce for individual RSGs.

[2] Vabalas A, Gowen E, Poliakoff E, Casson AJ (2019) Machine learning algorithm validation with a limited sample size. PLoS ONE 14(11): e0224365. https://doi.org/10.1371/journal.pone.0224365