Review ESSD-2018-18, East Africa Land and Vegetation

Despite its provocative title, this paper does not qualify for nor belong in ESSD. Following from its self-professed keywords " Land Use, Mathematical Programming, Dynamic Programming, Multiple Testing, Spatial Data and Analysis, False Discovery Rate", the manuscript addresses mathematicians and statisticians but not earth system scientists. Essentially the authors want to demonstrate a better method to assign 2-D spatial patterns to terrestrial environments to improve detection and monitoring of land use change. Their basic premise - that improved mapping should enable improved detection - seems entirely plausible. One hopes they would have combined knowledge of land vegetation patterns and remote sensing with statistical techniques for delineating land use patterns. Instead they fail completely on vegetation, demonstrate an appalling ignorance of remote sensing, and may or may not have made appropriate application of mathematical tools. They proclaim - in line 172 on page 6 - that they "provide theoretical proof that the mdFDR is still controlled under sub-region independence"; they should submit 'proofs' elsewhere. If they want to contribute substantially, via publication in ESSD, to understanding of land use change and its social implications, they need to offer a much different, much improved approach. They wish to convince readers (page 18, lines 434, 435) that their "findings are consistent with historical evidence and other climate change investigations done in this region" but they fail to provide such evidence from their own work and can not, evidently, cite any third-party confirming evidence.

In generating this review I assembled a long list of mistakes and mis-statements, in many sections several per paragraph. The authors have not met journal expectations of clear precise language or of well-documented assertions. Because I recommend that they discard the present manuscript - through formal rejection by the journal or withdrawal by the authors - I will not inflict those many pages of comments and objections. If the authors want to start fresh, I make the following suggestions.

1) Start by analysing a region of known land use distribution and patterns. Make a choice that provides desired combinations of topographic variation, record duration and confirmed accuracy; many examples exist. A quick search in ESSD reveals: the UK (Wood et al., 40 years of country-side surveys, https://doi.org/10.5194/essd-10-745-2018); the continental USA (e.g. Chen et al., 80 years of plantation forest data, https://doi.org/10.5194/essd-9-545-2017);) Siberia (Ottle et al. https://doi.org/10.5194/essd-5-331-2013); agricultural regions in Korea (Seo et al. https://doi.org/10.5194/essd-6-339-2014); even a global satellite-based view covering nearly 25 years from which these authors might select interesting regions (Li et al. https://doi.org/10.5194/essd-10-219-2018). I suspect they could find good examples for mid-Atlantic Appalachia-to-coast regions.

2) Use - as most of the above studies will have done - standard land use and plant functional type terminology. Start with the IGBP land use categories, a quick google search will turn up hundreds of examples. For example: http://glcf.umd.edu/data/lc/. By adopting standard terminologies these authors will access a wide variety of pre-existing work while making their own work accessible and comparable to many subsequent users. Standard easily-accessible PFT (plant functional type) terminologies also exist.

3) Demonstrate your block assignment techniques on one of these known land use maps. Document improvements (?) relative to standard gridded approaches based on, again, existing well-described land use categories (from #2 above).

4) Compare outcomes based on known land use categories with remotely sensed parameters, specifically NDVI. Essentially, use the IGBP land use categories in test mode as a transfer standard to NDVI.

5) Get a thorough education in remote sensing of land surfaces. The current manuscript reveals a serious deficiency in this regard: uninformed and out-of-date. For this particular data set the authors want Pinzon and Tucker 2014 (Remote Sensing 2014, 6(8), 6929-6960; doi:10.3390/rs6086929) rather than Tucker et al. 2005); they should give particular attention to Panel E of Figure 4. Demonstrate, by narrative and perhaps as well by including remote sensing colleagues, that you appreciate the long, detailed and well-documented caution about quantitative application of NDVI. Demonstrate a clear understanding of the many factors that confuse or even obscure NDVI. Understand why readers and users will wonder greatly at a supposed average value of NDVI expressed as 575.8716664 (row 10 column D of the .csv file): ten (!!) significant figures? Many knowledgeable readers would rather expect to see 0.6 ± 0.1!

This reviewer had to generate my own frequency distribution of average NDVI (column D, rescaled by dividing by 1000 and rounding back to 2 significant figures) to convince me that the data represented NDVI; the resulting distribution differs substantially from Figure 4 Panel E of Pinzon and Tucker already mentioned. A subsequent improved version of this manuscript will need to demonstrate serious understanding and restraint in manipulation of satellite remote sensing data. Abundant literature exists! The authors could start by looking closely at Du et al., A global satellite environmental data record derived from AMSR-E and AMSR2 microwave Earth observations in ESSD (https://doi.org/10.5194/essd-9-791-2017).

6) Finally, if warranted by all previous steps, demonstrate the utility of your improved, tested and validated mapping technique to East Africa. Include an uncertainty assessment, which, not least, addresses the use of NDVI in that case due to absence, weakness or rapid change of land use categories.

In a substantially revised and improved manuscript, a reader would expect clear and precise language, more documentation of local conditions and changes, and much less jargon about "gift exchange problems" (page 7 line 150).

If the authors want to understand current issues in land use change they should read the recent summaries of land use modelling issues available in the CMIP special issue of GMD (e.g. https://doi.org/10.5194/gmd-9-2973-2016 and https://doi.org/10.5194/gmd-9-2809-2016). The authors clearly need assistance with land use terminology! They use, variously, "vegetation life cycles", "vegetation trends", "vegetation structure" and "land cover trends". These terms and phrases convey independent and distinct concepts to terrestrial ecologists but the authors use them casually without recognising the confusion they sow.

Abstract, line 25: data link does not work. ESSD generally does not use figshare, finding it unreliable. Many alternatives exist: 4TU, Pangaea, Zenodo, etc. Suggest you deposit the data at a reliable doi-based repository with a good browse and search capability. A figshare url does not qualify as a permanent identifier.

The authors clearly imply (page 9, line 241) that the data set provided (as .csv) derives in fact from previous work, e.g. Clements et al. 2014. Should an ESSD reader conclude that these authors have submitted a manuscript to a data publishing journal based on someone else's data? Unacceptable (and scarcely believable) if true!