Earth Syst. Dynam. Discuss.,
doi:10.5194/esd-2016-33-RC5, 2016

Earth System Dynamics
Discussions

Interactive comment on "Continuous and consistent land use/cover change estimates using socio-ecological data" *by* Michael Marshall et al.

Anonymous Referee #5

Received and published: 12 November 2016

This manuscript develops a method for improving annual LULCC mapping in Kenya. A large array of socioeconomic, ecological and remote sensing variables were collected and related to LULCC categories via machine learning models. Random forest models were used to select the most important variables and generalized additive models were then used to build simplified predictive models using the reduced variable set. 70% of reference sample were used for training and 30% for error assessment. R2 was between 62% and 65% for agricultural and natural vegetation and was lower for other land cover types (e.g. urban). The authors concluded that population density is the most important predictor and that non-remote sensing predictors consistently outperformed remote sensing variables for each land cover type. All analysis was performed at 5km spatial resolution.

Overall, the data analysis is well done and the paper is well written. Given the datasets

that the authors collected and the analysis performed, the conclusions are all valid. However, the authors should be cautious when generalizing their conclusions. Major comments are below:

1. The accuracies are all quite low. The RF models with the complete variable set yielded pseudo R2 <= 0.69 for level 1 land cover legend and even lower for level 2 land cover legend. These results âĂŤtraining accuracies – were obtained over a relatively small area. It's probably safe to say that the accuracies would be worse if the approach is extrapolated over a larger region e.g. SSA.

2. Part of the low accuracy may have been resulted from the limited use of remote sensing variables. All socioeconomic and bioclimatic variables were either pre-existed or simply extra/interpolated, whereas the remote sensing-based phenological variables were derived by the authors. There is no doubt that the temporal domain of remote sensing data is important for LULCC mapping. By assuming there is one phenological transition per year, modeling vegetation phenology using harmonic functions and annual NDVI time series also make sense. However, other features, and most critically, the surface reflectance of various spectral wavelengths were not used in the study. In this regards, what was the reason of choosing the ∼8km GIMMS NDVI3g data over the ∼5km Long Term Data Record (LTDR) (http://ltdr.nascom.nasa.gov/cgi-bin/ltdr/ltdrPage.cgi)? The annual LTDR dataset contains red, NIR and thermal bands in addition to NDVI. I suppose the accuracy would be much improved when these spectral values were included in the model, although the question of whether remote sensing variables would outperform non-remote sensing variable then would still be in doubt. As such, statements on non-remote sensing variables outperformed remote sensing variables for LULCC estimation in the abstract, the discussion and the conclusion sections should be modified accordingly. These are localized conclusions constrained to the limited use of remote sensing data and the coarse spatial scale.

3. Scale plays a critical role in the analysis. At 5km spatial resolution, all we can observe from satellite data are macro-level land surface features. Fine-grain hetero-

geneities are effectively concealed. The discovered relationship – socioeconomic variables, practically population density, are positively correlated with land use intensity – at this scale likely does not hold at very fine scales. For instance, a 5km grid may include many villages and the surrounding agricultural lands, whereas a 30m grid may be part of a village. The link between population density and agriculture percentage at 5km resolution may not apply to 30m. As such, statements like Page 24 lines 1-3 indicating that fine-scale Landsat data can readily fit into the developed modeling frame should be revised or deleted.

Minor comments:

Page 6. The relationships between LULCC and socio-ecological systems are two ways. This is pointed out at the very beginning of the manuscript. It should also be mentioned conceptually when discussing the advantage of using bioclimatic and socioeconomic variables to predict LULCC in the future (e.g. 50-100 years).

Page16 lines11-12. Please explain what the variables bio7.d, bio14.d, and bio3.d mean in the text, although they are listed in Table 2.

Interactive comment on Earth Syst. Dynam. Discuss., doi:10.5194/esd-2016-33, 2016.