We would like to thank the reviewers for their constructive and comprehensive comments. We have made changes to the manuscript accordingly, which are summarized below. We believe the manuscript is much stronger, but welcome additional suggestions if the reviewer feels it necessary.

Reviewer #1

1) We did not discuss the use of more detailed datasets that are certainly available for Kenya, because the intent of the manuscript was to use a rare dataset that to our knowledge is only available in Kenya to develop a LULCC model approach that can be used across SSA for land surface modeling applications. Based on the literature, we believe that the functional relationships developed here will be different than for more developed regions such as North America where arable farmland is more accessible. We have inserted the following into the discussion: "The proposed methodology when applied to other regions of the world will undoubtedly result in a different combination of socio-ecological predictors and functional relationships, because access to land varies across agrarian and non-agrarian societies, so further study is required with observed data to develop region-specific models and validate the results for countries in SSA. Kumar et al., 2013, for example, showed that in the United States pre-1900 when the country was largely agrarian and transportation networks were weak, population density and crop area were highly correlated, because crops needed to be grown close to markets. However, as the country became more industrialized and transportation networks improved, farmers moved to more biophysically suitable areas away from city centers, making biophysical determinants of crop area more important than population density in the latter half of the 20th century. Whether the analyses are performed in agrarian or non-agrarian regions, extensive preparation of observation data will be required, because the data used in this study, namely consistent sample area frames at a spatial resolution appropriate for land surface modeling and spanning multiple climatic zones through time, is quite unique."

2) This is a good observation that was not properly articulated. The bar graphs and Figure 5 essentially show that remote sensing predictors account for more variance, initially, but the incremental (and thus overall) improvement is lower than the non-remote sensing predictors. The non-remote sensing predictors have the additional advantage of being more numerous. We added to the results: "It should be noted in each case however, the highest ranked remote sensing predictors resulted in lower model error than the highest ranked non-remote sensing predictors. The non-remote sensing predictors were more numerous and generated larger incremental improvements that contributed to overall greater predictive power." We have made minor additions to reaffirm this observation.

3) The ultimate goal of this model-building exercise is to project land for SSA. This is extremely difficult to do with Earth observation data. That said, we have inserted text in the discussion to address two possible avenues for Earth observation in retrospective analyses: "An analysis of the non-remote sensing and remote sensing predictors together revealed that for agriculture, natural vegetation, savanna, and forest cover, Earth observation data provided an additional 1-2% explained deviance. If the long-term average remote sensing predictors could be downscaled using MODIS or Landsat data and then aggregated to 5x5 km2 resolution with distribution moments as predictors, for example, the explanatory power of non-remote sensing predictors could be further enhanced for retrospective analyses. Another avenue worth exploring could involve using downscaled long-term average remote sensing predictors to develop 5x5 km2 probabilities as in the Pengra et al., 2015 dataset to evaluate the non-remote sensing models proposed here." Additional figures are available for the combined analysis, but we did not include them in the manuscript, because remote sensing predictors provided little added predictive ability and they would have made the manuscript too cumbersome. We did however, include one statistic concerning the added explained deviance worded above.

4) We have changed this to reflect the fact that remote sensing predictors initially explained more variance, but additional predictors added little value when compared to non-remote sensing predictors.

5) We believe the inclusion of the paragraph in the discussion from comment #1 properly addresses the case when regions transition from agrarian to industrial.

Reviewer #2

1) This is essentially a methods manuscript. We are demonstrating a new approach that overcomes the challenges of non-remote sensing and remote sensing approaches, so the manuscript can appear technical at times. In upcoming studies, the approach will be applied to important questions in earth system science as now highlighted in the manuscript. That said, and without more specific remarks from the reviewer, we have edited the manuscript by eliminating technical jargon, e.g. multicollinearity, and more clearly describing statistical approaches. The manuscript is already long and we believe adding any more detail will make it too cumbersome.

2) Per another reviewer's suggestion, we have addressed the transferability issue. In short, the method will remain the same for other regions of the world, but the functional relationships and important predictors, as in the example of the United States, will change.

Further, it would have been difficult to conduct this analysis outside of Kenya at the time the manuscript was written, because the observed data assembled did not exist to our knowledge anywhere else in the world. The sample area frames as indicated in the methods do span multiple climatic zones and a standard split sample approach with cross-validation was used to assess the model. That said, we have added to the discussion: "Whether the analyses are performed in agrarian or non-agrarian regions, extensive preparation of observed data will be required, because the observed data used in this study, namely consistent sample frames at a spatial resolution appropriate for land surface modeling and spanning multiple climatic zones through time, is quite unique."

3) The primary purpose of the manuscript was to introduce new geospatial data and a new methodology using a unique observed dataset. The retrospective analysis performed, as stipulated, was for illustrative purposes only. To add a prospective analysis to the manuscript would make the manuscript, which is already long, too cumbersome. We are however using newly acquired observed data to project the models into the future across SSA. We added to the conclusion: "In an upcoming study, the modeling approach proposed here will be used with a newly acquired sample area frame dataset to estimate baseline LULCC and project land suitability across SSA mid-21st century with AFRICLIM and other geospatial data."

4) We have rephrased page 4, lines 18-21 to merely reflect the difficulty in applying integrated models, as opposed to their frequent use. We did not give an exhaustive review of models, but simply gave two examples which illustrate the primary

categories. The introduction is targeting regional (more than global) land surface modeling, because of the deficiency of current LULCC models in capturing land-atmosphere feedbacks. That said, we have added appropriate sentences in the introduction that current integrated models run at a very coarse resolution. The IMAGE model, which you elude to runs at 50 km spatial resolution, which is much too coarse to capture land-atmosphere interaction and feedbacks. The model we propose runs at 5 km resolution, which is more appropriate. Remote sensing-based models run at even higher spatial resolution (30 m – 500 m), but have their own deficiencies. In this manuscript, we are essentially striking a balance between integrated and remote sensing-based models.

5) We have explained the BIOCLIM term and why it is more useful than other climatic variables for LULCC estimation.

6) The ".d." extension was used to indicate dynamic, as opposed to static or slowly-changing predictors. We have changed the caption for Table 2 and throughout the main body of the manuscript to reflect this.

Reviewer #3

1) We agree with the reviewers that the methods section needed a strong introduction. We have inserted the following at the beginning of Section 2.2: "The development of the functional relationships from the sample area frames involved four major steps illustrated in Figure 2. Non-remote sensing and remote sensing predictors were selected after an exhaustive online search that are freely and seamlessly available across SSA, so that the relationships can be used in future studies across the continent for retrospective or prospective analyses. Given the large number of predictors collected, machine learning was used to identify a subset of the most powerful predictors before constructing the functional relationships. The functional relationships were then evaluated against remote sensing predictors with hold-out samples and finally, used to demonstrate how the relationships can be used to reconstruct LULCC estimates continuously through time." We have also inserted a workflow outlining major milestones as Figure 2.

2) We have inserted the following after defining the two levels of classification in the methods: "These two levels of specificity allowed us through model-building to understand the level of detail that can be captured by coarse resolution geospatial data." It is expected that coarse resolution data will not be able to capture the same level of classification detail as higher resolution data. Since we had the data to explore this hypothesis, we performed the analysis. For the most part, coarse resolution data is inappropriate for detailed classification. This is important, because regional to global scale analyses often run with detailed classification descriptors and it is clear from our analysis that coarse resolution data is good for understanding, for example, the impact of the transition of natural vegetation to agriculture, but perhaps not for forest to crops. We have added the following as a major finding to the discussion: "3) coarse resolution data was able to capture general classification descriptors, but was unable to capture more detailed descriptors." and expounded on this finding in the final paragraph of the discussion.

3) Per a previous reviewer's comments, we have addressed the transferability issue.

4) We have removed reference to pre-1981 throughout the manuscript, but have simply stated that the models can be used for retrospective analysis. Per a previous reviewer's comments, we reiterated that the time series analysis was purely for demonstrative purposes and that population and climate (main drivers) are projectable and could be used with AFRICLIM, for example, to project LULCC into the future. To address the uncertainty with projections, we added to the discussion on the population data used: "In addition, the extrapolation method used is efficient

and can be projected indefinitely, but does not capture complex demographics that other methods do and can lead to 'runaway' growth/decline and unrealistic mid- to late- 21st century projections for scenario-building (Baker et al., 2008)." In reference to using other biophysical and socio-economic data in the discussion, we added "Many biophysical predictors are available mid- and late-21st century and are therefore widely used for prospective analyses, so methods should be explored to project soils and socio-economic data into the future to improve LULCC estimates."

Reviewer #4

Major Comments

1) Per previous reviewer comments, we have addressed the transferability issue.

Minor Comments

1) We have inserted the Ward et al., 2014 reference
2) We agree that more recent attempts have been made to couple LULCC to land surface models. We have changed the introduction accordingly: "Traditionally, spatially-explicit LULCC was not an input to land surface models, but was instead represented by structural (e.g. leaf area index) or physiological (e.g. stomatal resistance) changes in vegetation. LULCC was then mapped in parallel to characterize these changes. These early attempts have been replaced by fully coupled LULCC and land surface models (e.g. Shevliakova et al., 2009; Lawrence et al., 2012)." The main point here is that land-air coupling research exists and models do quantify its impact, but LULCC estimates need to be improved in order to better quantify coupling.
3) There are other simulated population datasets. We did not use these, because they use different methods based on varying availability of district-level population statistics. That said, we are considering for the SSA paper to use Afripop to estimate the growth/decay constant from 2000-2015. In the discussion we include the following: "Finally, there is no consensus on which population product to use however, in the future, other products (e.g. Afripop) should be compared against the product used here, used to adjust the growth/decay coefficient for population density estimates beyond 2000, or combined to make a model ensemble."
4) Based on a previous reviewer's comments, BIOCLIM has been defined and properly cited accordingly.
5) Isothermality decreases significantly from the equator- even in Kenya. We have defined isothermality instead as "more pronounced seasons."
6) We changed to "The only decrease in agricultural lands was in Kitale town (-1.40% per year)"
7) We put the discussion on country and regional-specific differences in predictors and functional relationships just after the major finding paragraphs of the discussion to highlight its importance, since it is a major caveat of the study and was mentioned by each reviewer.
8) We added detailed URL's in Table 2.
9) We included percent agriculture for 1983 and 2012 in Figure 9 (now Figure 10) to illustrate differences between a traditional LULCC study (two time slices) versus a

trend analysis of continuous LULCC. The manuscript is already quite large and we feel that adding more figures would be too cumbersome. Especially considering that the data is over multiple (and not only two) years.

Reviewer #5

Major Comments

1) We agree with the reviewer that the accuracies are "quite low" compared to high resolution LULC classification approaches. However, when dealing with coarse resolution data, R^2's of 0.65 are acceptable from the literature. Per other reviewers' comments, we added a paragraph and major caveat to the discussion, which addresses transferability issues. It will be interested to see in our upcoming paper for SSA to what degree this approach will capture agriculture and natural vegetation change. We did not have the data necessary at the time to do a SSA manuscript, but now we do.
2) Based on a previous reviewer's comments, we have added statistics to show that blending remote sensing predictors with non-remote sensing predictors did not change the results much. The biggest gains after population density actually came from introducing the ISRIC soils data, but we omitted it from the final analysis, because it is a one-time snapshot and soil properties are quite variable. In addition, it adds a layer of complexity that we felt could not be addressed properly in the discussion. As the reviewer suggests, the remote sensing approach taken has been used in previous studies. The temporal signature had to be exploited, because of the lack of spectral information (i.e. we were only using NDVI). As cited (Tian et al., 2015), the reason we used GIMMS 3g over LTDR and other long term coarse resolution remote sensing records is because it is the most appropriate for trend analysis. LTDR is a blended AVHRR and MODIS product. As such, the blending of two different data records leads to artificial jumps in 2000, which can have a significant impact on trends. Perhaps this issue has been addressed in the new LTDR Version 4. At the time the analysis was performed however, version 4 was not available. Based on previous reviewer comments, we have deemphasized non-remote sensing predictors over remote sensing predictors, addressed the real benefit of non-remote sensing predictors more clearly, and highlighted important opportunities for both. We believe the real opportunity is with downscaling.
3) Indeed, the relatively poor performance of the remote sensing predictors lies with the problem of heterogeneity and we discuss this. However, we did add "If the long-term average remote sensing predictors could be downscaled using MODIS or Landsat data and then aggregated to 5x5 km^2 resolution with distribution moments as predictors, for example, the explanatory power of non-remote sensing predictors could be further enhanced for retrospective analyses. Another avenue worth exploring could involve using downscaled long-term average remote sensing predictors to develop 5x5 km^2 probabilities as in the Pengra et al., 2015 dataset to evaluate the non-remote sensing models proposed here."

Pengra, B., Long, J., Dahal, D., Stehman, S.V., and Loveland, T.R., 2015, A global reference database from very high resolution commercial satellite data and methodology for application to Landsat derived 30m continuous field tree cover data

Minor Comments

1) We have added at various points in the discussion concerns about the complexity of interactions and feedbacks between the predictors and LULCC. Most notably we observe, that 50-100 year projections may not be realistic as SSA transitions from more agrarian (population driven) to industrialized (ecologically) driven crop area.
2) We have added to the discussion that "As seen in the functional plots, low populated areas with more temperature seasonality, or inter-annual variation, and lower bio3.d (isothermality) tended to have higher proportions of natural vegetation (savanna and shrubs). Isothermality is the ratio of mean diurnal temperature range (bio2.d) to the temperature annual range (bio7.d), which is the difference between the annual maximum and minimum temperatures." Bio14.d is already defined in Table 2 and is self-explanatory: Precipitation of Driest Month.