

Quantifying the impacts of bias in landcover data on global change analyses

Lyndon Estes ^{*}, Peng Chen [†], Stephanie Debats ^{*}, Tom Evans [†], Fanie Ferreira [‡], Gabrielle Ragazzo ^{*}, Justin Sheffield ^{*} and Kelly Caylor ^{*}

^{*}Princeton University, Princeton, NJ USA, [†]Indiana University, Bloomington, IN USA, and [‡]GeoTerraImage, Pretoria, RSA

Submitted to Proceedings of the National Academy of Sciences of the United States of America

Blah blah.

landcover | bias | remote sensing | agriculture | crop yield | harvested area
| carbon | agent-based model | landscape

Abbreviations: GTI, GeoTerraImage; SSA, sub-Saharan Africa

The nature and distribution of landcover is a fundamental determinant of many environmental and social processes that drive or are affected by global change (1), such as agricultural production and food security (2–4), carbon cycling (5, 6), biodiversity loss (7, 8), or demographic changes (9). Landcover maps are therefore critical for understanding the nature and impact of such changes (10), and they need to be accurate at the finest scales at which the underlying processes operate. For example, agricultural productivity and nutrient loadings can vary greatly between neighboring fields, and field sizes are often <2 hectares in regions where smallholder farming still dominates (11, 12). To understand agriculturally driven processes, it is thus necessary to accurately delineate fields at their smallest grain size, and to do so at regional to global scales to have a consistent set of maps.

Landcover data can only be developed with satellite imaging, but often the average size class of the cover type of interest is smaller than the sensor resolution, or spectrally indistinct from other neighboring covers, which propagates classification error (10, 13, 14). The result is that landcover datasets are generally inaccurate at finer scales and greatly differ between one another, particularly in those parts of the world undergoing the most rapid land use changes, where the aforementioned sources of bias tend to be most pronounced (15–17).

These errors are well-known (16, 17, 10, 18, 19), and there are a variety of efforts underway to improve landcover maps, particularly for agriculture (20, 14). What is less known is the degree to which these errors bias measurements built upon the distributional and areal information in landcover. An impediment to this understanding is that the errors are hard to quantify because spatially extensive reference data are not available for most regions of the world—particularly over Africa and other developing regions. Errors assessment therefore typically rely on a small number of ground truth points or survey data aggregated to political boundaries. For this reason, we have a better understanding of the biases between landcover datasets or in relation to country-level statistics (16, 17, 21) than we do of how error changes over spatial gradients or as a function of aggregation scale.

Being unable to fully quantify the errors in landcover maps of course makes it difficult, if not impossible, to quantify their impact on downstream analyses. There has been some work examining how such error influences climate simulations (22), agricultural land use patterns (23), and carbon flux (24) and human population estimates (9), but these either use simulated landcover errors (22) or compare relevant differences in estimates between different satellite-derived landcover maps (9, 24). The exception is (23), who use a high quality, ground-

collected reference map detailing farm land use parcels in central Belgium, but the number of sites and region were both fairly restricted, and the parcels were not spatially contiguous.

There is thus an urgent need to more precisely quantify landcover map errors and how these vary over large regions, particularly for the regions where landcover is changing most rapidly yet is most poorly known. We address this need in this study, using a unique, high accuracy agricultural landcover map for South Africa to quantify the errors in several latest generation landcover maps that are broadly used in global change studies. We use this information to examine how i) landcover properties and related classification schemes influence error, ii) how these errors change with aggregation scale, with the specific goal of determining “safe” scales for making area-based calculations, and 3) how these errors propagate through several different forms of downstream analyses that broadly represent the global change research focus areas, including biogeochemical and land use change studies, food security assessments, land surface hydrology and climatology, and human geography.

Study area and landcover data

Our study focused on South Africa, which comprises nearly 6% of sub-Saharan Africa’s (SSA) landmass, and has a large, diverse agricultural sector, ranging from large commercial operations to smallholder farms (25, 26). This diversity suggests that the country’s agricultural landcover spans the range of types that are found throughout the rest of SSA.

The South African government commissioned a whole country cropland boundary map in order to stratifying the annual aerial crop type census used to calculate harvested area estimates (27). The map was made by trained workers who visually interpreted high resolution satellite imagery and manually digitized field boundaries following a standardized mapping protocol. The resulting vectorized field maps, which were made in 2007 and updated in 2011, provide a unique, high accuracy reference dataset of both crop field distribution and size classes. We converted the vector data into a rasterized estimate of cropland percentage at 1 km resolution (henceforth the “reference map”), which was 97% accurate in distinguishing cropped from non-cropped areas.

Reserved for Publication Footnotes

We compared our reference percent cropland estimates to those created from four satellite-derived landcover datasets. We obtained South Africa’s 30 m resolution National Landcover map (SA-LC) for 2009 (28), the 500 m resolution MODIS Landcover for 2011 (29, 30), the 300 m resolution GlobCover 2009 (31), and the new 1 km Geo-wiki hybrid-fusion cropland map for Africa (18). We chose these particular datasets because they are nearly contemporaneous with our reference data, and represent the major types of landcover products used by researchers: SA-LC typifies the higher resolution, Landsat-derived maps that are developed individually for many countries (32), MODIS and GlobCover are widely used global-scale products (33, 34), while Geo-Wiki incorporates the first three datasets and is the current state of the art for agricultural landcover maps. We extracted the cropland classes from the first three datasets and converted these to 1 km resolution percent cropland estimates (hereafter simply the “landcover maps”), resulting in 4 maps to compare to our reference cropland map (the “reference map”).

Quantifying Error

We used these maps to first quantify error in cropland area estimates. We calculated error as the difference between the reference and landcover maps at different scales of aggregation (1 to 100 km), in order to estimate bias and how it varies with scale. Next, we assessed how bias correlates with the amount of cropland cover in agricultural landscapes, to gain insight into how landscape patterns may affect error.

We undertook five further analyses to investigate how map error can impact assessments founded on landcover maps. These include first-order analyses, in which values for a variable of interest are mapped to particular cover type(s), and second-order analyses, in which a process model draws on the cover types’ values to calculate an output value. We created four datasets to represent second order analyses. The first was a series of maps of vegetated carbon stocks created following the methodology of Ruesch and Gibbs’ (35). The second was constrained cropland percentage maps, which, following Ramanakutty et al (36) were adjusted so that their total cropland areas matched provincial-level reported cropland totals. Using these adjusted cropland percentage maps, we disaggregated district-reported maize harvested area and yields (following 37). We then compared differences between total carbon stock estimates calculated from the reference map with those from the four cropland maps, and again examined how these differences changed as a function of aggregation scale. We made the same comparisons for total maize harvested area, average yield, and total production.

For the second-order analyses, we examined how cropland cover errors influence 25 km resolution monthly evapotranspiration estimates produced using the Variable Infiltration Capacity (38) land surface hydrology model. For this example, we used the cropland maps to adjust the seasonally varying, landcover-specific leaf area index (LAI) values that VIC uses to partition water vapor fluxes into their evaporative and transpirative components. In the second example, we examined how these errors can impact the parameterization of an agent-based food security model (39). Spatially-explicit, agent-based models are frequently employed in land change science, and require an initialization step to assign landscape resources to model agents (e.g. 40–42). In this case, we used the cropland maps to allocate farmland to agents representing individual households in political districts, with the model assigning each household its initial cropland holdings using a function that considers total district cropland and how much cropland is near the agent’s location.

Results

Landcover bias. We created the 1 km reference and landcover maps, removing all croplands marked as communal or small-holder farmland in the reference vector maps (individual fields were not mapped with the same precision), as horticulture, or plantation forestry (SI), and calculated the total cropland extent in the remaining area (1,081,000 km², or 90% of South Africa). The 2011 reference map showed a cropland area of 104,310 km², which the SA-LC and GeoWiki maps overestimated by 26 and 5.8%, respectively, and GlobCover and MODIS underestimated by 21 and 26.1%.

We then aggregated each map to 5, 10, 25, 50, and 100 km resolutions, and subtracted the four landcover maps from the reference map at each scale of aggregation to assess the spatial patterns of bias (Fig. 1).

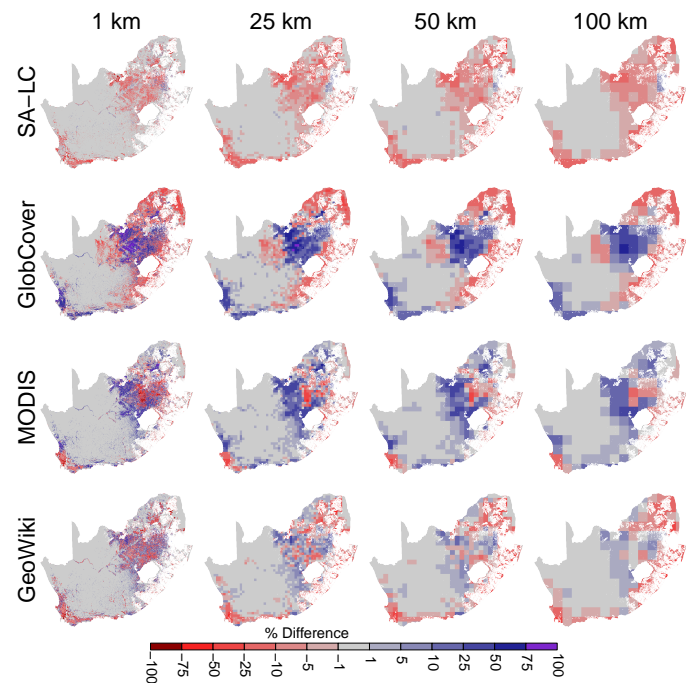


Fig. 1. Differences in percent cropland estimates between the reference map and each of the four landcover maps. Rows indicate the landcover map being assessed (by subtraction from the reference map), while columns refer to resolution of aggregation. White indicates areas with no data where communal farmlands or plantation forests were removed.

The most pronounced extreme biases are in the MODIS and GlobCover maps, with both underestimating cropland extent by 10–75% in the center of the country (blue areas in Fig. 1, and the dominant production region), and overestimating along the eastern to northern margins (red areas in Fig. 1). The average biases for GlobCover and MODIS are 21% and 34% at 1 km, meaning that each map underestimates cropland by that amount at this resolution. MODIS mean bias drops to 8% at 50 km of aggregation, whereas GlobCover bias is still 24% at 100 km (SI Appendix, Fig. S1).

The SA-LC map uniformly overestimates cropland throughout the country (Fig. 1), but the magnitude of mean bias is much lower at -8% at 1 km to -6% at 100 km (Fig. S1). The GeoWiki map has a very heterogeneous pattern of bias, having a slight underestimation tendency (5% mean bias) at 1 km, which changes to a smaller overestimation bias (-2 %) at 100 km (Fig. S1).

The role of landscape characteristics in shaping bias. To investigate the degree to which landscape features influence land cover map bias, we extract all pixels in agricultural areas (>0.5% cropland) of the 1 km reference map using the boundaries of 354 magisterial districts (South Africa's finest administrative unit, which average 3,445 km² in size; SI Appendix, Fig. S3), and averaged the values of these pixels within each district. These provided an estimate of the intensity of cropping within the farmed parts of the landscape (with districts defining the landscape scale), as well as the degree of mixing between cropland and other land covers. We then extracted the cropland map biases for the same locations, and calculated the mean absolute bias within each district. Absolute bias describes the magnitude of bias but not its direction, and is more informative about the likelihood of the map being biased at any given point in the landscape than actual bias, as positive and negative biases can cancel each other out within small areas.

A generalized additive model fit to district-level absolute bias (log-transformed) shows that absolute bias peaks at 50% cropland cover for all but the GlobCover map (which continued to increase with cropland cover), and is lowest when the landscape is dominated either by cropland or other cover types (Fig. 2). In other words, bias is highest when cropland cover is mixed evenly with other cover types.

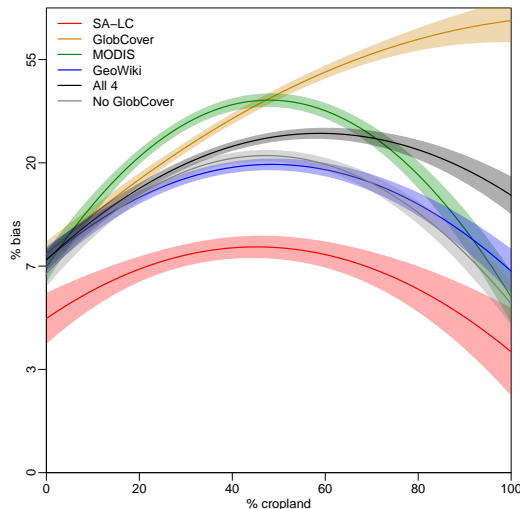


Fig. 2. The relationship between the mean absolute bias in cropland maps and cropland cover within agricultural areas (reference map pixels having >0.5% cropland), averaged within the boundaries of magisterial districts ($n = 345$), as fit with a generalized additive model. Prediction curves are color-coded to the different cropland maps, with the solid line indicating predicted absolute bias, and the lighter shading the standard error of the coefficients. Models were also fit to the mean absolute biases across all four maps (black curve), and from all maps excluding GlobCover (grey curve).

The impact of bias on calculating carbon stocks. We used the data of (35) to calculate carbon densities for African forests, shrublands, croplands, grasslands, and sparse habitats. We assigned the cropland carbon values to map cells in proportion to their cropland cover. For the non-cropland proportions, we assigned the carbon value from each of the other types, such that we created four different carbon maps for each landcover map at each aggregation scale (Fig. S4), which allowed us to test how carbon estimates vary as a function of i) bias in the cropland estimates and ii) the characteristics of adjacent cover types.

At the 1 km scale, the differences between country-wide total carbon stocks made using any of the cropland maps and

the reference were within $\pm 3\%$, regardless of which cover type was adjacent to cropland (Table S1), because differences between total stock estimates are diluted by the large area (~50-70%) of the country having no cropland. Comparing the carbon stocks from each map's agricultural area reveals much greater differences (Table S1). SA-LC overestimates carbon stocks by just 2% when the adjacent cover type is forest, and up to 15% when it is sparse cover. MODIS ranges from negligible overestimates in denser carbon classes (forest, secondary forest, and shrublands) to 8-13% underestimates for the grassland and sparse classes. GeoWiki underestimates for all types, from <1% for sparse cover to 8% for forest. GlobCover shows the largest total carbon stock bias, with over-estimates ranging from for sparse lands to 162% for forest. The magnitude of this bias is due to false positives—GlobCover identified cropland in nearly 50% of pixels, compared to 30% for the other three cropland maps.

The spatial patterns of carbon bias (Fig. S4) are a reflection of the cropland biases (Fig. 1). Where cropland is underestimated and the surrounding cover type is of higher carbon density than cropland, carbon density is overestimated. For lower density (grassland and sparse vegetation), there may be slight underestimates. These tendencies are reflected in the average of each map's spatial biases over agricultural areas (Fig. 3). For example, MODIS and GlobCover overestimate carbon by ~50% on average at 1 km resolution when forest is the cropland-adjacent cover (stars in semi-transparent green and gold bars, Fig. 3; Table S2). When it is sparse vegetation (open circles in Fig. 3), MODIS underestimates by nearly 3% at all scales, whereas maps that overestimate cropland (e.g. SA-LC, semi-transparent red) overestimate carbon density for this cover. Overall, GeoWiki has the lowest average spatial biases, for all cover types and all resolutions. Its maximum bias is a 12% overestimate at 1 km for forest cover, thereafter all biases are less than 10%, and converge to $\pm 2\%$ above 1 km aggregation (Fig. 3, Table S2). All maps' biases are within $\pm 10\%$ bias after aggregation to 25 km.

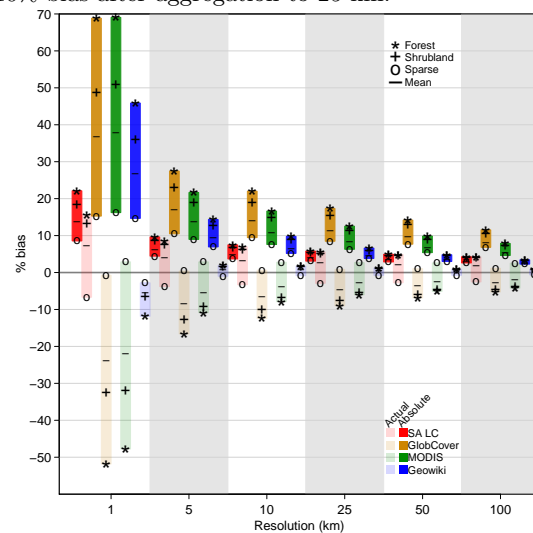


Fig. 3. The average spatial biases in carbon density estimates based on cropland maps, expressed as a percent difference relative to the reference map. The values of actual bias (incorporating bias direction and magnitude) and absolute bias (magnitude only) are presented in semi-transparent and solid colors, respectively, with colors denoting the specific cropland map and symbols indicating which cover type was used to calculate cropland adjacent carbon density. The bar represents the mean biases calculated across each of the 5 cover types. Shrubland and grassland bias values are near zero and not shown for display clarity.

The patterns of mean absolute spatial bias (solid colored bars in Fig. 3) generally follow that of actual bias, but with somewhat higher magnitudes and a few notable differences. The most notable is that Geo-Wiki has fairly large absolute bias at 1 km, averaging 27% across cover types (line in solid blue bar, Fig. 3, Table S2), which is close to the 36-37% for GlobCover and MODIS. SA-LC has the lowest absolute bias across scales, averaging (across cover types) from 14% at 1 km to 3% at 100 km. (Fig. 3, Table S2). The increase in Geo-Wiki's absolute spatial bias relative to SA-LC's can be attributed to the spatially heterogeneous nature of its cropland biases (Fig. 1) compared to the other three cropland maps.

Bias in harvested areas, yield, and production estimates.

- Used provincial boundaries to extract total cropland from reference data. Cropland totals used to adjust cropland fractions falling within those boundaries from each cropland map so cropland map totals equal the provincial-level total from the reference (SI).
- Disaggregated magisterial district level maize yield and maize harvested area onto the adjusted cropland fractions, and onto the reference map.
- Calculated differences in spatial biases of yields and total production estimates at different scales of aggregation.
- Yields aggregated to higher scale using two methods: 1) weighted mean aggregation, where harvested area fractions used to weight yields at each aggregation scale, technically more difficult; 2) mean of any harvested area contributing to aggregation, without weighting. Production aggregated by either calculating production at 1 km and then aggregating to each scale and differencing, or aggregating harvested fractions and type 2 yield, and calculating production at each level.
- Average bias normalized as a percentage of mean pixel-level yield or production calculated from the reference map.

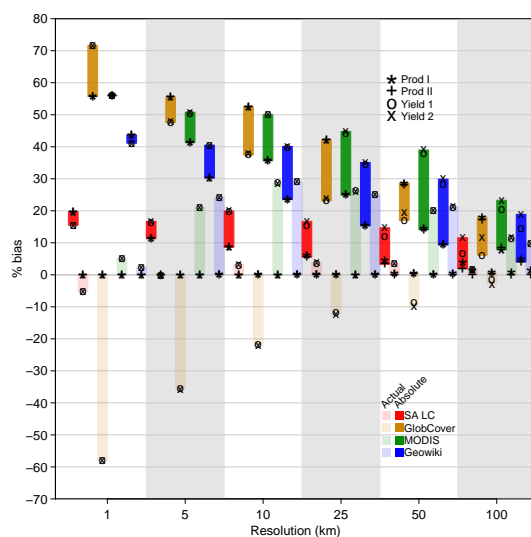


Fig. 4. The mean spatial biases in disaggregated maize yield estimates, and the bias in total crop production estimates calculated from these yields and constrained crop area estimates. The values of actual bias (incorporating bias direction and magnitude) and absolute bias (magnitude only) are presented in semi-transparent and solid colors, respectively, with colors denoting the specific cropland map and symbols indicating whether the bias is related to one of two yield aggregation methods and associated crop production estimates.

- GlobCover average bias shows large yield overestimates (nearly 60%) at 1 km, declining with aggregation scales.

Interestingly, MODIS and GeoWiki start with small yield underestimates that grow to 30% yield under-estimates at 10 km, declining thereafter. SA-LC has lowest, +/-5% (Fig. 4)

- No production biases.
- Absolute biases in both yield and production are large. All >40% for both production and yield at 1 km, except SA-LC which is between 15-20%. The biases decline with aggregation, but remain above 10% for yield for most datasets at 100 km.

Bias in evapotranspiration estimates.

- Spatial biases most pronounced for VIC run with MODIS and GlobCover. These two overestimate ET along the southern margin, but typically by less than 5-10%. (Fig. 5). In the interior, VIC-GlobCover switches between under- and over-estimation and then back again moving from West to East, but the bias is on the order of 5% magnitude. MODIS shows a similar pattern.
- GeoWiki and SA-LC shows much less discrepancy in interior. Greatest tendency is to underestimate along the southern coast.
- Difference in country-level ET estimates negligible (<2%).
- Justin, can you give some thoughts on reasons for these patterns? The tendency seems to be that where cropland is underestimated (blue areas in Fig. 1) then ET is over-estimated (red areas in Fig. 5). Here I calculate all difference as $(\text{ref} - \text{lc}) / \text{ref} * 100$, with lc being one of the four landcover maps. My expectation was that crop ET was higher than most surrounding vegetation types, so one would expect ET to be underestimated where cropland is underestimated?

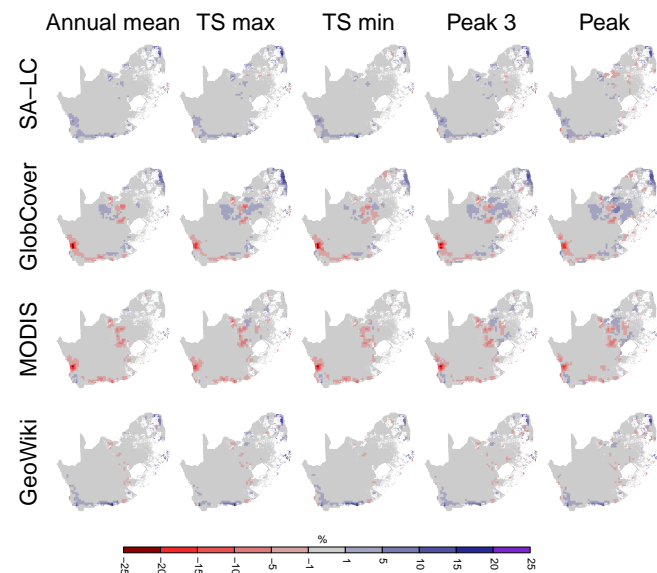


Fig. 5. Differences in evapotranspiration estimates from a 29 year run of the VIC land surface hydrology model when initialized with LAI response curves derived from the reference map, versus those from the four cropland maps. Columns indicate differences between ET estimates from different portions of the model's time series: annual mean ET; highest annual ET value in the time series (TS max); smallest annual ET value in the time series (TS min); total of mean monthly ET for the three months having the highest ET during the year (Peak 3); the mean monthly ET of the peak ET month.

Initialization errors in spatial agent-based models.

- We selected four magisterial districts with similar climates, and differing abundances of cropland ranging from X-Y%, and differing patterns of map bias (Fig. S8).
- Using household survey statistics from a project in Zambia that provides information on the average field area of farming households (X ha, SD = Y ha), we assigned each district a number of households determined by the total area of reference cropland in each district divided by mean farm size. **Tom, Peng, please correct as needed—maybe there is some variability to account for agent heterogeneity drawn by resampling from the distribution?** This number of households represents the “true” number of households in the district.
- Disaggregated cropland to 100 m resolution, which is required by the model.
- An agent-based model was then initialized wherein the household locations were seeded throughout the district, and assigned the required amount of cropland with a radius of X from the household location.
- These steps were repeated for each cropland map using the same number of household agents.
- We then examined the differences that occurred between the agent initialization from the reference map and each of the four cropland maps, in terms of the percent of households that did not have cropland allocated, and the amount of cropland left unallocated in each district (Fig. 6).
- The percent households without farms is 0 when cropland is overestimated (because there is plenty of land to choose from), but increases in a 1:1 relationship with the amount of cropland underestimation (positive values on X-axis in Fig. 6).

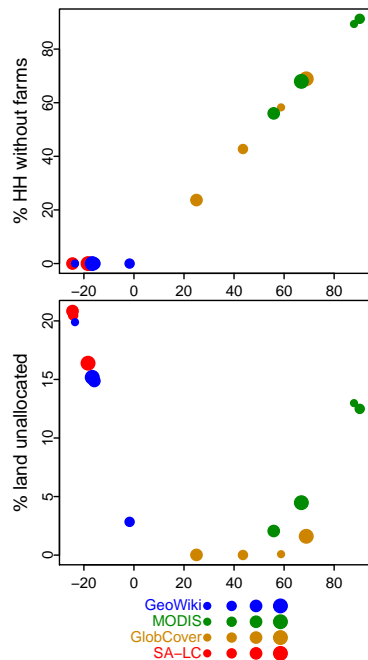


Fig. 6. Biases in agent-based model initialization relative to cropland map biases, measured in terms of the percent of households having no cropland allocated (top) and the percent of cropland left unallocated (bottom). Dot sizes represent a district number (smallest = district 1; largest = district 4), colors represent the landcover map. **to fix:** add x-axis legend, add district number to legend.

- The percent of land unallocated, however, shows a U-shaped relationship with cropland biases, declining to near

zero when cropland bias is between zero and up to ~50% under-estimated, thereafter increasing. Thus in cases where cropland bias insufficient farmland for households, a significant share of available cropland can go unallocated (nearly 15%). **Tom, Peng, I assume this has to do something with seeding and the spatial pattern of cropland, but not sure exactly what. Model not sure which HH to give the land too because there are too many for the amount of cropland available, so it gives up?**

Discussion

Some points to make thrown out now because out of time. Others will be added. Please add any important implications/considerations you see from the results.

Some points from my notes last year:

• Bias as a function of scale

1. At 1 km resolution all landcover products are still fairly biased.
2. Bias drops to acceptable levels quickly for geowiki—at 5X5 km, mean bias is just 1% (overestimated). The absolute bias for this dataset is 10% or lower from 10X10 km resolution and coarser.
3. The SA dataset’s bias is fairly consistent but low across all levels of aggregation, amounting to no more than an 8% overestimate of cropland with absolute bias of similar magnitude.
4. MODIS and GlobCover biases (mostly of underestimation) do not dissipate until the higher levels of aggregation. MODIS’s actual bias (under-estimation) falls below 10% at 20 km resolution, but the absolute bias remains above 10% until more than a 100-fold aggregation is done (>100 km resolution). For GlobCover, it is still too high.

• Bias as a function of cropland cover

1. Classification algorithms are thus more error-prone where landcover is mixed/heterogenous.
2. The exception to this lies in the GlobCover dataset, where bias primarily increases as a function of cropland cover. The reason for this is that GlobCover’s cropland classes do not provide for 100% cropland cover, so aggregation tends to exacerbate underestimates.
3. Thus caution is needed when aggregating a mixed pixel class.
 - An example illustrates this: take 4 1 km pixels, 2 of which are 100% cropland, 2 of which are other cover types. Imagine a landcover product classifies 3 of these as cropland (2 correct, 1 an error of commission), using a cropland class that is defined as 50% cropland. Aggregating the actual fraction by a factor of 4 will result in a new 4 km pixel having 50% cropland, whereas aggregating the landcover product’s pixels will give just 38% cropland, even when factoring in the incorrect classification.)

• Bias as a function of method

1. Higher resolution and ensemble-based approaches have less bias.
2. geowiki represents a fusion of multiple coarse resolution data sources that has undergone extensive validation using a crowdsourcing approach
3. the SALC dataset is based on 30 m landsat data, but incorporates a range of ancillary data and expert judgement

4. MODIS and GlobCover data are effectively single source/single algorithm.

5. **Newer points begin here**

6. Statistically constrained constrained landcover estimation approaches provide accurate area-based inferences when aggregated. But spatial errors are still high, as seen with GeoWiki and production/yield estimates. Using these to identify yield gaps at specific map locations is inappropriate, or even for a larger location if it does not coincide with the geographic boundaries of the statistical unit.

7. Constrained estimates are also dependent on the accuracy of the statistics.

- Fix above to have section on bias for global change studies

- Scales at which it is safe to estimate values of say carbon stocks.
- Above point about bias in disaggregated yield estimates - no point mapping these out. A new approach might be to take these statistically reported yields and then combine them with satellite data to estimate yield variability within the district. That way would have meaningful reason for disaggregating yields, and would be pegged to real yield values, which would help minimize errors in remote sensing of yields.
- Something on ET - doesn't seem to matter much, but land-atmosphere interactions can make these discrepancies meaningful, particularly since biases occur in arid areas where a lot of irrigation happens—can cause significant impacts on regional climate. etc. etc. Also we didn't change out land cover types, and the vegetation in SA around the cropland will have reasonably similar LAI and ET responses (I think), thus impact more muted than it might be elsewhere (e.g. in forested landscapes).
- Agent-based models. **Tom, Peng, something of significance/implications of this, please**

- Will need a section on way forward for data, etc. Key role of accurate landcover, particularly agricultural. New methods, vectorized field boundaries seem to be highly valuable, Mapping Africa, Stephanie's paper, Geo-Wiki, etc are the way ahead.

Mixed landscapes increase the chances of omission and commission errors by increasing the number of cover classes, or because such landscapes are less spectrally distinct (14)

More blather. Materials and Methods

Methods. Perhaps it is right **SI Materials and Methods**.

Describe weighted mean bias reasons.

Digital RCD Analysis.

Appendix: App 1

Appendix

This is an example of an appendix without a title.

ACKNOWLEDGMENTS. I thank everyone tearfully.

References

- Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. *Progress in Physical Geography* 21(3):375–393.
- Lark TJ, Salmon JM, Gibbs HK (2015) Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environmental Research Letters* 10(4):044003.
- Wright CK, Wimberly MC (2013) Recent land use change in the Western Corn Belt threatens grasslands and wetlands. *Proceedings of the National Academy of Sciences* 110(10):4134–4139.
- Licker R et al. (2010) Mind the gap: how do climate and agricultural management explain the yield gap of croplands around the world? *Global Ecology and Biogeography* 19(6):769–782.
- Asner GP et al. (2010) High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences* 107(38):16738–16742.
- Gaveau DLA et al. (2014) Major atmospheric emissions from peat fires in Southeast Asia during non-drought years: evidence from the 2013 Sumatran fires. *Scientific Reports* 4.
- Newbold T et al. (2015) Global effects of land use on local terrestrial biodiversity. *Nature* 520(7545):45–50.
- Luoto M, Virkkala R, Heikkinen RK, Rainio K (2004) Predicting bird species richness using remote sensing in boreal agricultural-forest mosaics. *Ecological Applications* 14(6):1946–1962.
- Linard C, Gilbert M, Tatem AJ (2010) Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal* 76(5):525–538.
- See L et al. (2015) Improved global cropland data as an essential ingredient for food security. *Global Food Security* 4:37–45.
- Jain M, Mondal P, DeFries RS, Small C, Galford GL (2013) Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sensing of Environment* 134:210–223.
- Debats S, Luo D, Estes L, Fuchs T, Caylor K (year?) A generalized computer vision approach to mapping agricultural fields in Sub-Saharan Africa. *Remote Sensing of Environment*.
- Lobell DB (2013) The use of satellite data for crop yield gap analysis. *Field Crops Research* 143:56–64.
- Estes L et al. (2015) DIYlandcover: Crowdsourcing the creation of systematic, accurate landcover maps. *PeerJ PrePrints* 3:e1266.
- Estes LD et al. (2013) Projected climate impacts to South African maize and wheat production in 2055: a comparison of empirical and mechanistic modeling approaches. *Global Change Biology* 19(12):3762–3774.
- Fritz S, See L, Rembold F (2010) Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. *International Journal of Remote Sensing* 31(9):2237–2256.
- Fritz S et al. (2011) Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets. *Geophysical Research Letters* 38:L04404.
- Fritz S et al. (2015) Mapping global cropland and field size. *Global Change Biology* 21(5):1980–1992.
- Verburg PH, Neumann K, Nol L (2011) Challenges in using land use and land cover data for global change studies. *Global Change Biology* 17(2):974–989.
- Fritz S et al. (2012) Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling &*

- 542 *Software* 31:110–123.
- 543 21. Kaptu Tchuent AT, Roujean JL, De Jong SM (2011) Com-
544 parison and relative quality assessment of the GLC2000,
545 GLOBCOVER, MODIS and ECOCLIMAP land cover
546 data sets at the African continental scale. *International*
547 *Journal of Applied Earth Observation and Geoinformation* 589
548 13(2):207–219.
- 549 22. Ge J et al. (2007) Impacts of land use/cover classifica-
550 tion accuracy on regional climate simulations. *Journal of*
551 *Geophysical Research: Atmospheres* 112(D5):D05107.
- 552 23. Schmit C, Rounsevell MDA, La Jeunesse I (2006) The lim-
553 itations of spatial land use data in environmental analysis.
554 *Environmental Science & Policy* 9(2):174–188.
- 555 24. Quaife T et al. (2008) Impact of land cover uncertainties
556 on estimates of biospheric carbon fluxes. *Global Biogeo-*
557 *chemical Cycles* 22(4):GB4016.
- 558 25. Hardy M, Dziba L, Kilian W, Tolmay J (2011) Rainfed
559 Farming Systems in South Africa in *Rainfed Farming Sys-*
560 *tems*, eds. Tow P, Cooper I, Partridge I, Birch C. (Springer
561 Netherlands), pp. 395–432.
- 562 26. Estes LD et al. (2014) Using changes in agricultural util-
563 ity to quantify future climate-induced risk to conservation.
564 *Conservation Biology* 28(2):427–437.
- 565 27. Fourie A (2009) Better Crop Estimates in South Africa.
566 *ArcUser Online* (1).
- 567 28. SANBI (2009) National Landcover 2009, (South African
568 National Biodiversity Institute; National Department
569 of Environmental Affairs and Tourism, Pretoria, South
570 Africa), Technical report.
- 571 29. DAAC) LPDAACL (2011) MODIS MCD12q1 Land Cover
572 Type Yearly L3 Global 500 m SIN Grid. Version 5.01,
573 (NASA EOSDIS Land Processes DAAC, USGS Earth Re-
574 sources Observation and Science (EROS) Center, Sioux
575 Falls, South Dakota), Technical report.
- 576 30. Friedl MA et al. (2010) MODIS Collection 5 global land
577 cover: Algorithm refinements and characterization of new
578 datasets. *Remote Sensing of Environment* 114(1):168–182.
- 579 31. Arino O et al. (2012) *Global land cover map for 2009*
580 *(GlobCover 2009)*. (European Space Agency & Univer-
581 sit Catholique de Louvain).
- 582 32. Fry J, Coan M, Homer C, Meyer D, Wickham J (2009)
583 Completion of the National Land Cover Database (NLCD)
1992-2001 Land Cover Change Retrofit Product, (U.S. Ge-
ological Survey), USGS Numbered Series 2008-1379.
- 585 33. Gross D et al. (2013) Monitoring land cover changes in
586 African protected areas in the 21st century. *Ecological*
587 *Informatics* 14:31–37.
- 588 34. Shackelford GE, Steward PR, German RN, Sait SM, Ben-
589 ton TG (2015) Conservation planning in agricultural land-
590 scapes: hotspots of conflict between agriculture and na-
591 ture. *Diversity and Distributions* 21(3):357–367.
- 592 35. Ruesch A, Gibbs HK (2008) New IPCC Tier-1 global
593 biomass carbon map for the year 2000. *Carbon*
594 *Dioxide Information Analysis Center (CDIAC),*
595 *Oak Ridge National Laboratory, Oak Ridge, Ten-*
596 *nessee. Available online at: [http://cdiac.ornl.](http://cdiac.ornl.gov/epubs/ndp/global_carbon/carbon_documentation.html)*
597 *gov/epubs/ndp/global_carbon/carbon_documentation.*
598 *html.*
- 599 36. Ramankutty N, Evan AT, Monfreda C, Foley JA (2008)
600 Farming the planet: 1. Geographic distribution of global
601 agricultural lands in the year 2000. *Global Biogeochemical*
602 *Cycles* 22:19 PP.
- 603 37. Monfreda C, Ramankutty N, Foley JA (2008) Farming the
604 planet: 2. Geographic distribution of crop areas, yields,
605 physiological types, and net primary production in the
606 year 2000. *Global Biogeochemical Cycles* 22:GB1022.
- 607 38. Liang X, Lettenmaier DP, Wood EF, Burges SJ (1994) A
608 simple hydrologically based model of land surface water
609 and energy fluxes for general circulation models. *Journal*
610 *of Geophysical Research* 99(D7):14415.
- 611 39. Chen P, Plale B, Evans T (2013) *Dependency Provenance*
612 *in Agent Based Modeling*. pp. 180–187.
- 613 40. Manson SM, Evans T (2007) Agent-based modeling of de-
614 forestation in southern Yucatn, Mexico, and reforestation
615 in the Midwest United States. *Proceedings of the National*
616 *Academy of Sciences* 104(52):20678–20683.
- 617 41. Evans TP, Kelley H (2004) Multi-scale analysis of a house-
618 hold level agent-based model of landcover change. *Journal*
619 *of Environmental Management* 72(1-2):57–72.
- 620 42. Kelley H, Evans T (2011) The relative influences of land-
621 owner and landscape heterogeneity in an agent-based
622 model of land-use. *Ecological Economics* 70(6):1075–1087.