How Cropland Data Errors Impact Our Understanding of Socioeconomic and Environment Change

Lyndon Estes[1,2], Peng Chen[3], Stephanie Debats[1], Tom Evans[3], Stefanus Ferreira[4], Tobias Kuemmerle[5], Gabrielle Ragazzo[6], Justin Sheffield[1], Adam Wolf[6], Eric Wood[1], and Kelly Caylor[7]

[2]*Woodrow Wilson School, Princeton University, Princeton, NJ USA*
[1]*Civil and Environmental Engineering, Princeton University, Princeton, NJ USA*
[3]*Indiana University, Bloomington, IN USA*
[4]*GeoTerraImage, Pretoria, RSA*
[5]*Geography Department, Humboldt University, 10099 Berlin, Germany*
[6]*Arable Labs, Princeton, NJ USA*
[7]*Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA USA*

Abstract

Research into global environmental change is often built upon land cover data, particularly croplands maps, which can have substantial errors. How these errors impact our understanding of environmental change processes and associated policies is thus of great interest, but is difficult to fully quantify because we lack spatially comprehensive reference data, particularly in the World's most rapidly developing regions. We used a high quality, national-scale cropland map to assess bias and accuracy within the cropland classes of current generation land cover datasets and four examples of "downstream" (land cover-dependent) analyses, two related to physical processes (estimates of vegetative carbon stocks and evapotranspiration) and two to socio-economic processes (gridded crop production estimates and agent-based simulation of household food security), at resolutions ranging from 1-100 km. We found that cropland maps have substantial errors below 25 km resolution, with substantial biases and inaccuracies, particularly from maps derived from coarse resolution sensors. In some cases these error metrics remain high even when maps are aggregated to 100 km. These errors can be substantially amplified in downstream analyses (e.g. the carbon and crop production analyses), but in cases where the values of different cover types that underpin the analysis are similar (e.g. evapotranspiration estimates) the results can be relatively insensitive. To avoid these errors, and thus minimize the risk of misunderstanding global change processes or misinforming policy, substantial map aggregation of land cover maps from their base resolution is often needed. NEW TEXT HERE.

1 Introduction

The functioning of the Earth System is fundamentally connected to the characteristics of land cover (Lambin, 1997). Our increasing modification of the Earth's surface (Lambin *et al.*, 2003) means that socioeconomic and physical processes increasingly interact through land cover. To fully understand these processes, an accurate understanding of the nature and distribution of land cover is essential.

This importance is understood by a growing number of social, economic, and natural scientists, who are using land cover data to advance understanding of food security (Lark *et al.*, 2015; Licker *et al.*, 2010; Wright & Wimberly, 2013), carbon cycling (Asner *et al.*, 2010; Gaveau *et al.*, 2014), biodiversity loss (Luoto *et al.*, 2004; Newbold *et al.*, 2015), demographic shifts (Linard *et al.*, 2010), and other important facets of Earth System processes.

The value of the insights resulting from such studies depends upon the veracity of their underlying land cover data, much as a house requires a solid foundation in order to remain standing. Unfortunately, the evidence to date indicates that this house has shaky foundations. The reason for this is that land cover data can only practically be derived from satellite imaging, which has several important constraints that propagate mapping errors. First, in many regions the spatial arrangement of cover types is smaller than the sensor resolution (e.g. smallholder's farms Debats *et al.*, 2016; Jain *et al.*, 2013; Ozdogan & Woodcock, 2006), or the types of interest are spectrally indistinct from neighboring ones (Sweeney *et al.*, 2015), which are factors that increase mapping complexity (Yu *et al.*, 2014). Second, the act of defining a cover class can cause error, in that selected classes may have highly diverse

spectral properties (e.g. croplands or savannas; Debats *et al.*, 2016; Estes *et al.*, 2016a)) and thus be difficult for the classifier to distinguish. Discretizing a continuous cover type (e.g. dividing a forest into different canopy cover classes) can promote classification error, particularly near class boundaries (Foody, 2002), as well as confusion about the actual extent of the cover type (Sexton *et al.*, 2015). Furthermore, class definitions often vary between maps, complicating inter-comparison (Kuemmerle *et al.*, 2013). Third, land cover maps are often used to detect changes (e.g. Gross *et al.*, 2013), but seasonal variability and land cover changes can be easily confused. Given these multiple sources of error, land cover maps are often inaccurate at finer scales and disagree widely between products, particularly in the world's most rapidly developing regions (Estes *et al.*, 2013a; Fritz *et al.*, 2010, 2013, 2011). These errors limit our ability to obtain granular, mechanistic understanding of processes related to global change.

These problems with land cover products are known (Fritz *et al.*, 2015, 2010, 2011; See *et al.*, 2015; Verburg *et al.*, 2011), and there are a variety of map improvement efforts underway (e.g. Estes *et al.*, 2016a; Fritz *et al.*, 2012). Likewise, the importance of assessing the accuracy of land cover maps is increasingly recognized, and there are well-developed, best-practice guidelines for gathering and using ground-truth samples to robustly quantify map error (Foody, 2002; Olofsson *et al.*, 2014, 2013; Stehman *et al.*, 2012). Because comprehensive, spatially representative ground truth data are typically unavailable for rapidly changing regions (See *et al.*, 2015), what remains an open question is exactly how much the maps researchers typically use deviate from actual land cover, and how this in turn impacts our understanding of environmental change processes. Our current understanding of map accuracy over such areas is often based on scarce information or top-down "sanity checks" made in comparison to aggregated survey data (Larsen *et al.*, 2015; Yu *et al.*, 2014).

Since it is difficult to fully quantify land cover map errors, it is even more challenging to gauge their impact on downstream analyses, where there is substantial risk of error amplification (Kuemmerle *et al.*, 2013). Some work has examined how map error influences climate simulations (Ge *et al.*, 2007), agricultural land use patterns (Schmit *et al.*, 2006), carbon flux measurements (Quaife *et al.*, 2008), human population estimates (Linard *et al.*, 2010), and species distribution models (Tuanmu & Jetz, 2014), but these either use simulated land cover errors (Ge *et al.*, 2007) or compare relevant differences in estimates between different satellite-derived land cover maps (Linard *et al.*, 2010; Quaife *et al.*, 2008; Tuanmu & Jetz, 2014), or, in cases where spatially explicit ground validation maps were used for assessment, cover small and discontiguous extents (Schmit *et al.*, 2006).

Fortunately, the recent, explosive growth in public and private initiatives to develop new Earth observing capabilities, which range from small drones[1] to new high resolution satellite arrays[2] and better mapping methods (Debats *et al.*, 2016; Estes *et al.*, 2013a; Fritz *et al.*, 2012; Tuanmu & Jetz, 2014), are finally providing the means to more comprehensively interrogate the accuracy and biases in the land cover products that have become commonplace in global change research–and which are often used to make policy decisions (Searchinger *et al.*, 2015).

In this study, we take advantage of this recent growth in data to address the call to more thoroughly assess errors in land cover maps (Kuemmerle *et al.*, 2013; Olofsson *et al.*, 2014, 2012), and further examine how these errors might impact our understanding of socioeconomic and environmental conditions. Using a unique, high-quality, spatially extensive map of South African croplands, which was created by expert mappers delineating individual fields visible within high resolution imagery, we conduct spatially comprehensive, bottom-up analyses to answer the following two questions: 1) What is the extent of error in several widely used land cover products?; 2) How do these errors propagate through downstream biophysical and socioeconomic studies?

The answers to these questions provide important insights into how cropland datasets can influence our understanding socioeconomic and environmental processes in South Africa, as well as more broadly throughout the rest of sub-Saharan Africa (SSA), where our current knowledge of the extent and distribution of cropland relies heavily on land cover maps (Fritz *et al.*, 2010; See *et al.*, 2015).

Materials and Methods

Datasets

In the late 2000s, the South African government commissioned a cropland map that was made by manually interpreting and digitizing fields visible within high resolution satellite imagery (Fourie, 2009). The resulting vectorized field boundaries provide unique, highly accurate data on field sizes and distribution for 2009-2011. The accuracy of this dataset was evaluated in a previous study (Estes *et al.*, 2016a), in which visual assessment of cropland presence/absence was made within 15,225 4 ha plots (25 sub-plots within 609 1 km^2 grids) placed over high resolution satellite imagery, and compared to coverage by the vector boundaries. Measured in terms of its ability to distinguish crop fields from other cover types, the results showed these data to be 97% percent accurate, with user's accuracies of 94% and 98% for the cropland and non-cropland classes, and producer's accuracies of 84 and 99% (see SI for more description, and Estes *et al.*, 2016a).

---

[1] e.g. 3DRobotics, DJIA

[2] Planet Labs, Skybox

We used these vector data as a reference for evaluating four land cover products representative of the type commonly used in global change studies and related areas of research. The first was South Africa's own 30 m resolution 2009 National land cover map (SA-LC SANBI, 2009), which is typical of the higher-resolution, Landsat-based maps that are created for individual countries (e.g. Fry *et al.*, 2009). Although global-scale, Landsat-derived maps have recently become available (Chen *et al.*, 2015), their reported accuracy for cultivated areas is lower (80-85% user's accuracy) than that achieved by more intensive, national to sub-national products (e.g. 90% user's accuracy Sweeney *et al.*, 2015). The second and third were respectively the 300 m GlobCover 2009 (Arino *et al.*, 2012) and 500 m resolution MODIS land cover data (for 2011, the final year of the reference interval), which are widely used global-scale products (e.g. Gross *et al.*, 2013; Shackelford *et al.*, 2015). The fourth dataset was the 1 km GeoWiki cropland map for Africa (Fritz *et al.*, 2015), which fuses the cropland classes of multiple landcover products (including the three assessed here), with the resulting cropland percentages adjusted until their totals sum to match cropland areas reported in national agricultural statistics. This "hybrid-fusion" approach represents the state of the art in mapping agricultural land cover. Since GeoWiki provides a continuous bounded value of land cover (percent cropland cover), which cannot be feasibly converted to a categorical value, we converted all other datasets, including the reference vectors, into comparable 1 km gridded percent cropland estimates.

To develop percent coverage from the cropland reference map, we intersected the original field boundary vector maps with a 1 km grid, and calculated the percent of each cell occupied by fields. In the resulting grid, we masked out areas classified as communal farmland (18.7% of mapped croplands in the reference data), because only their outer perimeters were digitized (Fourie, 2009), which risked overestimating cropland extent because the vectors also enclose uncropped areas between adjacent smallholder fields. We also excluded from the analysis areas covered by permanent tree crops, sugarcane plantations, and commercially afforested areas, because these three classes were not common to all five cropland datasets. To remove permanent tree crops (3.1% of cropland area), we masked out reference vectors labelled as such, and to exclude the other two cover types, which are not included in the reference data, we relied on information from two other datasets. We used a 20 m resolution landcover map of KwaZulu-Natal to mask out the primarily coastal sugarcane farms (93-100% user's and 76-98% producer's accuracy for sugarcane classes; GeoTerraImage, 2013), and used the SA-LC dataset to filter out areas of commercial afforestation, which are mainly located in South Africa's montane areas and do not overlap with arable croplands. In both cases, we aggregated these classes, and then masked out any 1 km pixels that had >0% cover of each class. The resulting masked reference grid covered 90% of South Africa (1,081,000 km^2), of which 104,304 km^2 was cropland.

We extracted the cropland classes from SA-LC, MODIS, and GlobCover, and converted these into percent cropland estimates at 1 km resolution. Both MODIS and GlobCover had mixed/mosaic classes of cropland and other covers, thus we followed Fritz *et al.* (2015) by creating upper, mean, and lower cropland estimates from these classes to produce three versions of the gridded percentages. We used the mean map for the main analysis, but estimated error variability using all three versions (SI).

Assessment of land cover map accuracy

We first evaluated the quality of the land cover product-derived percent cropland maps (hereafter referred to as the "test maps"). Instead of the standard confusion matrix-based accuracy metrics (Olofsson *et al.*, 2014, 2013), which apply to categorical land cover maps, we assessed the bias and accuracy of the test maps based on the gridded residuals that resulted when each test map was subtracted from the reference map. Here bias is the mean residual value, weighted by the density of reference cropland, and accuracy is the mean of the absolute values of residuals (also weighted by cropland density), thus accuracy is inversely related to this value. We calculated these metrics for the original 1 km resolution, and for maps that were further aggregated to 5, 10, 25, 50, and 100 km resolutions, in order to evaluate how bias and accuracy changes with observational grain. For these aggregated maps, we applied a further weight in calculating error metrics, the number of pixels contributing to each aggregated pixel, to prevent pixels close to national boundaries or where non-target cover types were masked out from having outsize influence on the statistics.

We also assessed how land cover pattern impacts map performance by modeling the correlation between map accuracy and cropland density. To evaluate this relationship, we used magisterial district boundaries (n=354, mean area=3,445 km^2; SI) to provide a landscape-scaled unit for calculating characteristic cover density. We filtered out pixels with <0.05% (0.5 ha) cropland, to prevent the much larger areas of non-agricultural districts from dominating the signal, extracted the absolute values of test map errors and corresponding reference cropland percentages, and calculated their district-wide means. The relationship between mean absolute error (response) and cropland density (predictor) was then modeled using a generalized additive model (Hastie & Tibshirani, 1990), with each district weighted by its number of agricultural pixels (Wood, 2001). To account for potential spatial autocorrelation, we fit a two-dimensional smoothing spline to the coordinates of each district's centroid.

Impact of map error on downstream analyses

We then used these maps to conduct four analyses typical of global change research: 1) estimation of carbon stocks, 2) simulation of evapotranspiration, 3) disaggregation of crop yield and production, and 4) simulating household dynamics using an agent-based model. The first and third analyses were relatively simple, in that the variable(s) of interest were mapped onto land cover using empirical relationships. The second and fourth relied on more complex numerical methods, where land cover was one of several variables needed to run each model. For the simpler analyses, we examined how results were influenced by map aggregation, while for the more complex cases, our assessments were confined to each numerical model's standard output resolution.

Estimating vegetative carbon stocks

To understand the carbon cycle and climate forcing due to land use change, it is important to have accurate, high resolution maps of vegetative carbon stocks (Searchinger *et al.*, 2015). One widely used vegetative carbon dataset is that of (Ruesch & Gibbs, 2008), who mapped estimated carbon density values for different vegetation types to the classes of a global land cover product. The resulting data were intended to provide a baseline for climate policy by the Intergovernmental Panel on Climate Change (IPCC), as well as input to other land use and biogeochemical analyses (Ruesch & Gibbs, 2008).

We followed this method to create vegetative carbon maps for South Africa. Since our cropland percentage map provided no information on surrounding cover, we developed several variants representing potential surrounding cover types by assigning the average carbon densities of five biomes (forest, secondary forest, shrubland, grassland, and sparse vegetation (Ruesch & Gibbs, 2008)) to the non-cropland fraction of our maps. These hypothetical maps represented the range in potential carbon densities, and allowed us to investigate how carbon estimates can vary as a function of i) test map errors and ii) the properties of neighboring cover types. We multiplied cropland densities by cropland fractions and added these to each of the five other densities multiplied by the residual non-cropland fractions to create five different carbon density maps. We aggregated each carbon map up to the five coarser resolutions for scaling comparisons (SI).

To assess carbon estimation error, we subtracted test map-derived carbon maps from those based on the reference map, and calculated bias and accuracy scores using the method described for the cropland maps (see previous section).

Estimating evapotranspiration

Accurate estimation of hydrological fluxes is critical to understanding how land-atmosphere interactions impact the climate system and runoff (Liang *et al.*, 1994). Land surface hydrological models are used to simulate these processes, and depend on land cover maps to provide information on the characteristics of vegetation and other materials covering the surface, as these govern the rates of runoff, infiltration, and evapotranspiration. We used the Variable Infiltration Capacity (VIC; Liang *et al.*, 1994) land surface hydrology model run with the Africa Flood and Drought Monitor's meteorological data (Sheffield *et al.*, 2013) to produce monthly gridded evapotranspiration estimates for South Africa for the years 1979-2010 at 25 km resolution. VIC's land cover scheme (derived from AVHRR) provides values for leaf area index (LAI), plant rooting depth, aerodynamic roughness, and several other variables that the modeluses to partition water vapor fluxes into their evaporative and transpirative components. We adjusted VIC's base map so that its cropland fractions matched those of the 25 km reference and test maps (each reprojected and resampled to VIC's $0.25°$ resolution), then ran one instance of VIC for each of the five land cover schemes. We then compared the mean annual ET produced by the reference map variant with those from the test maps to assess the degree to which map errors impact evapotranspiration values.

Disaggregating crop yield and production statistics

The spatial variability of yield and production is critical for understanding food security, trade, and the potential for agricultural expansion and intensification (Licker *et al.*, 2010; Monfreda *et al.*, 2008). The most reliable source of such data are national to sub-national agricultural statistics, often available only at relatively coarse-scaled administrative boundaries. *****************HERE*************To obtain higher spatial resolutions, efforts have been made to disaggregate these statistics using gridded land cover data (Monfreda *et al.*, 2008; Ramankutty *et al.*, 2008). These disaggregated datasets, which are constrained to match the agricultural statistics within the boundaries of the areas for which they are reported, have seen increasing use because they are considered to be more accurate than single source methods, and provide consistent data on which to base studies of global change (Ramankutty *et al.*, 2008; See *et al.*, 2015)

We used these methods to disaggregate the harvested area for maize (South Africa's largest crop (Estes *et al.*, 2013a)) on top of our reference and cropland maps, followed by yields, which were assigned to cells having harvested areas greater than zero. The first step in this process entails adjusting cropland percentages so that their totals match census-derived cropland area estimates (Ramankutty *et al.*, 2008). In place of census statistics, we used the reference map to calculate total cropland areas for South Africa's nine provinces, then adjusted the pixel-wise cropland percentages in the four test maps so that their province-wise sums matched these totals (SI). We then followed Monfreda *et al.*'s (2008) procedure for disaggregating maize (South Africa's largest crop; Estes *et al.*,

2013b) planted area and yields onto the reference and adjusted test maps. The necessary statistics were obtained from magisterial district-level agricultural censuses conducted for the year 2007 (**?**).

We then used these two layers to calculate maize production, and further aggregated the yield and production grids to 5, 10, 25, 50, and 100 km resolutions before quantifying the bias and accuracy of each test map's yield and production values. In this case, we could not convert cell-wise errors into percentages of the reference map values (because many cells had zero values for one map but not the other), so we calculated bias and accuracy from the map residuals and then normalized their values to the reference map means.

Agent-based simulation of household food security: Spatially-explicit agent-based model (ABMs) are frequently employed to understand land use decision-making and socio-ecological systems, often to facilitate improved policy, particularly in the arena of human development (Berger & Schreinemachers, 2006). To obtain valid insight into the functioning of such systems, it is important to calibrate an ABM to empirical data describing the characteristics of land and land users, so that the model realistically represents the social and biophysical features of the study region (Berger & Schreinemachers, 2006).

In our example, we used an ABM of household food security that simulates food production by individual farming households (the agents) in response to weather and management capacity (Chen *et al.*, 2013). The model is initialized such that each household is allocated a designated share of cropland derived from household survey statistics, and then calculates annual household-level agricultural production based on maize yields simulated by the DSSAT cropping system model (Jones *et al.*, 2003). Yield estimates vary as a function of soil properties, management inputs, and weather. The former two factors differ between households, as do the total areas of cropland, thus total production can vary substantially between households. The initialization process iteratively assigns households to the landscape using a function that factors in neighbor and cropland proximity, to ensure that households are grouped into communities and that their fields are within a realistic proximity. To achieve this, the model first takes a weighted (by cropland area frequency) random draw of 100 households and places these within the simulated landscape, assigning each household its required number of "fields" (cropland pixels), which must be within 1.5 km of the household's location and not already assigned to another household. This process is iterated until all households are assigned cropland, or all available cropland is allocated. The model is thus considered to be well-calibrated when all households are allocated their appropriate area of cropland, and all cropland is allocated to households.

Like many spatial ABMs, the model is computationally intensive, and thus run over smaller geographic domains (e.g. districts, rather than an entire country) and at higher spatial resolutions (10s to 100s of meters) in order to represent the different land units of individual farmers. To match these computational characteristics, we selected four contiguous magisterial districts (ranging from 1,040-1,343 km^2, SI) in the eastern part of the country with similar climate and 28-45% of their areas devoted to cropland. The frequency distributions of household cropland holdings were derived from household survey data obtained from Zambia (no equivalent data were available for South Africa). We used these statistics to calculate the "true" number of households per district by dividing reference cropland areas by the mean cropland area (2.2 ha), and preserved the cropland area distributions by multiplying the total number of households by the frequencies. To create cropland surfaces for each district, we disaggregated all five cropland maps to binary cropland/non-cropland cover types with 100 m resolution, in order to match the survey-recorded average field size of 1 ha. The model was then run separately for each district, with the household agents planting and managing a maize crop throughout the growing season, with possible final maize yields drawn from a look-up table that relates yield to particular climatic outcomes and management decisions, as simulated by the DSSAT model.

We ran each ABM simulation for a single season, once per district and when initialized by each of the cropland maps (20 simulations total). To examined how map errors impacted the land allocation process and household food production estimates, we calculated three variables: the percent of unallocated cropland, land deficit, and food deficit. The percent of unallocated cropland is the share of the district's total cropland area that was not assigned to any household, which measures how effective the model was in matching household agents to available cropland resources. Land deficit is the total area of cropland that should have been allocated to households in each district but wasn't, due to the mismatch between the cropland map and the survey-based estimates of what total cropland holdings should have been. Food deficit is the percentage shortfall in the average amount of food production that should have been produced by each household but wasn't because of the land deficit.

Results

Map quality

Bias and accuracy

Our reference cropland map indicated that crop fields covered nearly 10% of the total study area in the 2009-2011 time period. Subtracting each test map from the reference maps created pixel-wise residuals, where negative and positive values respectively represent overestimates and underestimates by the test map (Fig. 1A).

The most pronounced errors were in the MODIS and GlobCover maps, which showed large positive residuals in the center of the country where cropland is most concentrated (blue areas in Fig. 1A), and negative residuals (red areas) along the eastern and northern margins. These patterns translated into substantial map bias (Fig. 1B), with GlobCover and MODIS mean error exceeding 45% and 25% respectively at 1 km resolution, meaning that each map tends to underestimate cropland by that amount at that resolution. This bias declined with each level of map aggregation, being reduced to nearly 15% for GlobCover and 5% for MODIS at 100 km. The magnitude of mean absolute error (MAE) was somewhat higher in all cases. The GeoWiki map, in contrast, was the least biased overall, showing just a 7% bias at 1 km and near 0 for all other scales of aggregation, although its accuracy (23% MAE) was only half as good as SA-LC's at 1 km (11% MAE), which despite its uniform overestimation bias (Fig. 1A) was the most accurate map at aggregation scales < 10km. Above this, GeoWiki became slightly more accurate, having <5% MAE at 100 km resolution. The reason GeoWiki had relatively poor accuracy at 1 km resolution was due to the high heterogeneous error pattern, which traded between positive and negative residuals over short distances, thereby inflating MAE at this scale.
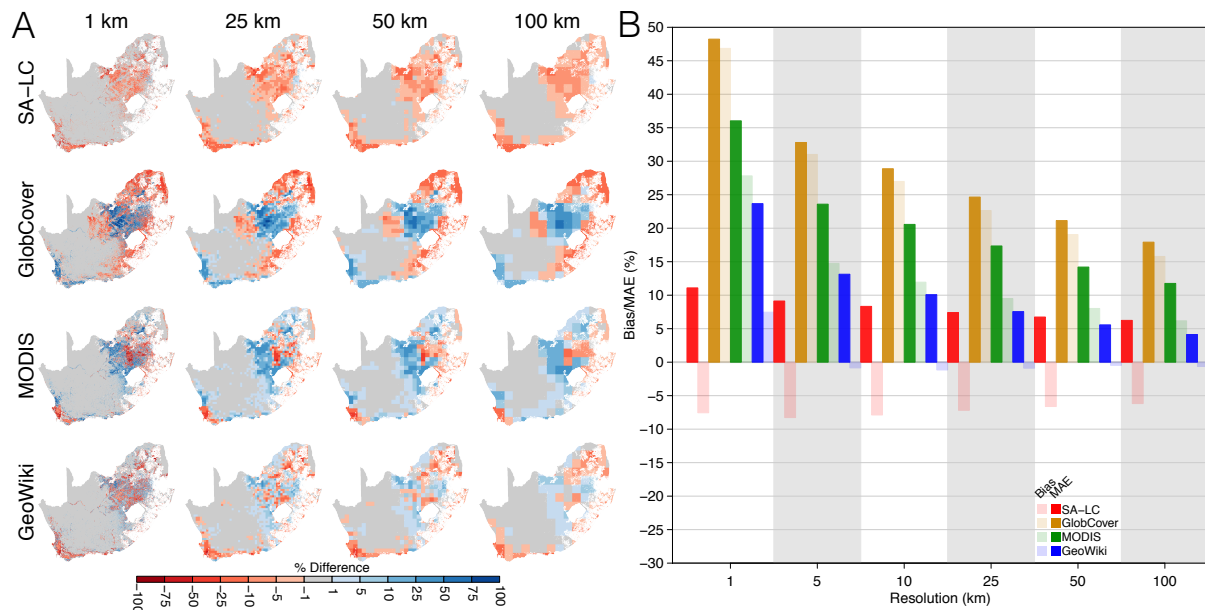

Figure 1: (A) Errors in the percent cropland estimates resulting from each of the four test maps relative to the reference map at different scale of pixel aggregation. Rows indicate the test map being assessed (by subtraction from the reference map), while columns refer to resolution of aggregation. White indicates areas where areas under communal farmlands or permanent tree crops were removed from analysis. (B) The bias (mean error) and accuracy (mean absolute error [MAE]) of each test map at each scale of aggregation, weighted by the percentage cropland in each cell of the reference map. Bias estimates are indicated by the semi-transparent bars, accuracy (lower is more accurate) by the solid bars, with bar colors coded to specific cropland maps.

Error and land cover density

The generalized additive model revealed primarily non-linear relationships between district MAE and cropland density that were best approximated by a first order polynomial function of cropland density (for all four cropland maps: p<0.001 on both terms of quadratic and on smoothing function applied to district centroids; > 85% deviance explained). Map accuracy was typically lowest at intermediate levels of cropland density (50-60% cover) for all but the GlobCover map (where accuracy continues to decline with cropland cover), and was highest where the landscape is dominated either by cropland or by another type (Fig. 2). In other words, accuracy is generally lowest when cropland cover is mixed evenly with other cover types. GlobCover's accuracy continued to decrease with cropland density because the dominant agricultural cover class contributing to the test map was defined as 50-70% crops mingled with other vegetation, thus the maximum percentage was constrained by this mixture range.

The impact of map error on physical analyses

Carbon estimates

The spatial patterns of test map errors transmitted into substantial carbon estimation errors, with the sign varying as a function of the density of carbon adjacent to croplands (SI). Where cropland was underestimated and the surrounding cover type was more carbon dense than cropland, carbon density was overestimated, but when the
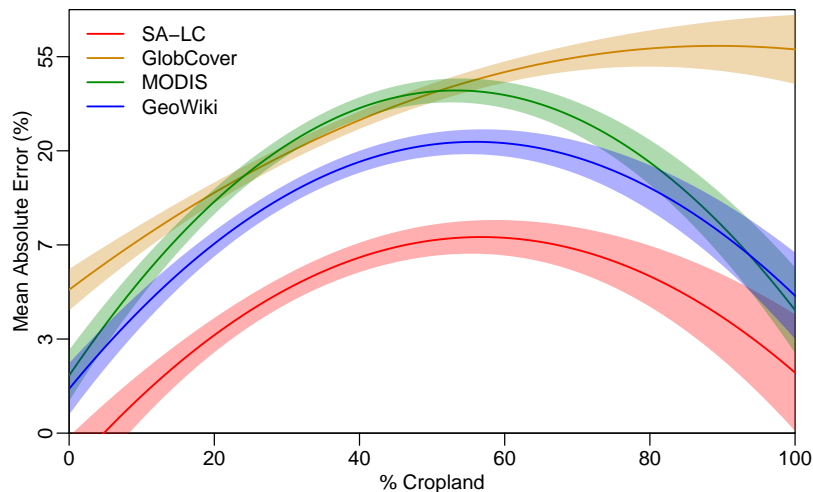
Figure 2: The relationship between map accuracy (the mean absolute error) in test maps and the actual cropland cover within agricultural landscapes (reference map pixels having >0.5% cropland), here defined by the boundaries of magisterial districts (n = 345), as fit with a generalized additive model. Prediction curves are color-coded to the different test maps, with the solid line indicating predicted absolute bias, and the lighter shading the standard error of the coefficients.

cover type was less dense than croplands (e.g. sparse vegetation), then carbon density was underestimated. The inverse was true where cropland was overestimated.

The magnitude of carbon errors varied as a function of the carbon density of surrounding cover, as demonstrated by the bias statistics (Fig. 3). Bias was near zero when grassland was the adjacent cover type (SI), as its carbon density is nearly the same as cropland. However, when forest was adjacent then bias was a three- to five-fold multiple of cropland map bias (Fig. 1B). At the most extreme, GlobCover's bias was -276% at 1 km, but even SA-LC and GeoWiki had biases of 22% and -50%, respectively. Bias could be substantial even for the least carbon dense vegetation type (sparse), as evidenced by the 15-25% mean error at 1 km for MODIS and GlobCover under this class. The mean bias across the different potential adjacent vegetation classes ranged between -20 for GeoWiki and -123% for GlobCover at 1 km (with MODIS in between these), while SA-LC's average bias was 11%. Biases declined fairly rapidly with aggregation, with all datasets having an average (across cover types) bias magnitude of <10% at ≥25 km of aggregation, except for GlobCover, which was -12% at 100 km (SI). As with cropland percentages, GeoWiki produced the least biased carbon density estimates above 1 km resolution.

In terms of accuracy, MAE values were essentially the same as bias magnitudes, except for GeoWiki's, which were twice as large. The average MAE across vegetation classes was 47% at 1 km, dropping to <10 only with 25 km of aggregation. In contrast, SA-LC's carbon estimates were twice as accurate at 1 km, and were slightly more accurate up to 25 km of aggregation were GeoWiki achieved parity.

Evapotranspiration estimates

Compared to the carbon analysis, the bias and accuracy in evapotranspiration (ET) calculated using the VIC model was negligible, averaging less than than +/-2%. However, there were several error hotspots in the resulting ET residual maps (Fig. 4). The most pronounced of these were the 5-15% overestimates in the center of the country caused when VIC was initialized with MODIS and GlobCover, while overestimates along the southern and western coasts reached 25%. These locations correspond primarily to the margins of major crop production regions–in the center is the westernmost boundary of the summer rainfall growing region, marked approximately by the 400 mm isohyet, where maize is the primary crop. The west coast hotspot falls at the western edge of the wheat-dominated winter rainfall region (Hardy *et al.*, 2011), where growing season rainfall is approximately 200 mm.

SA-LC and GeoWiki also resulted in ET errors estimates along the southern and western coasts, but here the tendency was to underestimate ET, while biases in the center of the country were either negligible to absent. All but MODIS underestimated ET by 5-15% in the northern tip of the country.
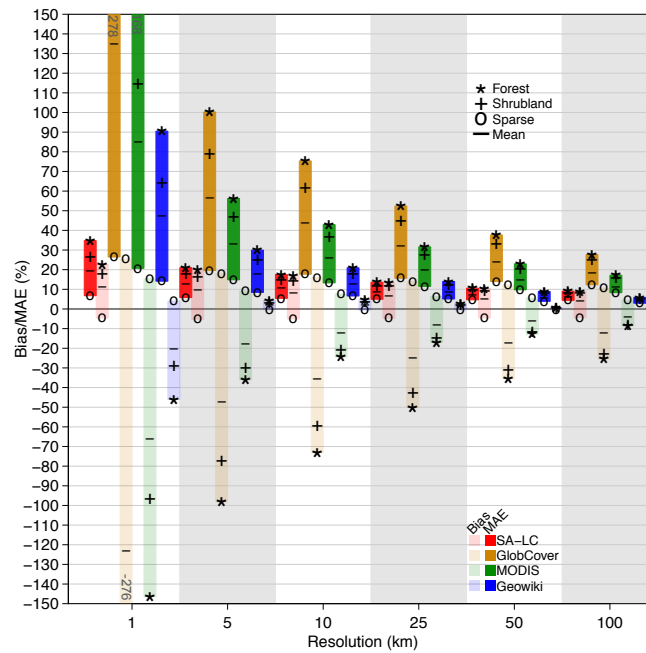
Figure 3: Biases and accuracies (mean absolute errors) of carbon densities derived from cropland maps, calculated as percents relative to the reference map. Bias estimates (represented by symbols) fall within the semi-transparent floating bars, while accuracies are contained in the solid bars. Bar colors are coded to specific cropland map, symbols indicate 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 were near zero, while secondary forest values were close to forest values, and thus these are not shown for display clarity (but see Table S2). MODIS and GlobCover values at 1 km exceeding the plot's Y limits are provided near their truncated tops.

Socio-economic analyses

Gridded crop yield and production data

Maize yields disaggregated onto the test maps showed some marked differences relative to the reference map, but only at the margins of the major crop production areas where cropland is sparser (SI). These differences resulted when a yield value was mapped onto a grid cell where the reference map had no harvested area, and thus zero yield. In more densely cropped areas, such discrepancies were less frequent because both the reference and test maps were both likely to have some maize harvested area, and therefore a yield value. Yield biases were thus fairly low (and accuracy high), with the largest being 20% for MODIS at 1 km, following by GlobCover with 10% (Fig. 5). These dropped to <10% with aggregation.

Production biases were generally slightly higher, but still low, for most datasets, with the exception of Glob-Cover, which had a large underestimation bias of >60% (relative to mean production) at 1 km, which remained above 10% even at 100 km of aggregation. MODIS production bias was above 20% at 1 km, but declined to below 10% at higher levels of aggregation.

In contrast, the accuracy of production estimates was poor. Here all datasets but SA-LC had MAE values of ≥30% below 25 km of aggregation (Fig. 5), reaching as high as 100% for GlobCover at 1 km, followed by 65% for MODIS and 45% for GeoWiki. SA-LC estimated production was most accurate, having between 10-20% MAE between 1 and 10 km, and <10% at 25 km and higher. This low accuracy relative to the gridded yield measures relates to the disaggregation process for harvested area, which allocates a fractional value to each pixel, which is itself a fraction. The process of adjusting the gridded values so that their totals match reported statistics does relatively little to correct the map's underlying commission or omission errors, and this constraint in fact appears to shorten the distance between negative and positive residuals (SI), thereby increasing absolute errors.

Agent-based model of household food security

In terms of impact to agent-based model simulation, where cropland map errors were negative (indicating a cropland overestimate by the test maps), the percent of land left unallocated had a straight one-to-one relationship with the percentage of overestimation (Fig. 6A). When cropland was underestimated, all croplands were allocated up until the underestimation exceeded 50%. The MODIS-based simulation for districts 1 and 2 was most pronounced for this tendency, with 5-10% of cropland remaining unallocated despite the fact that the majority of households
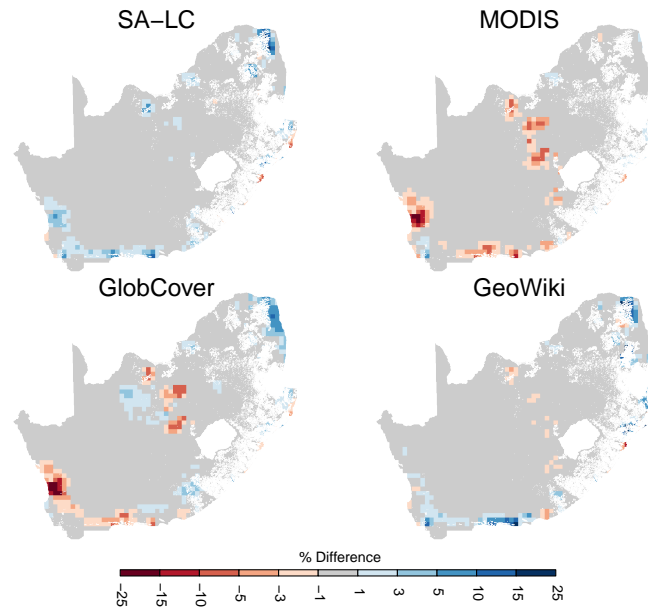
8

Figure 4: Differences in annual mean evapotranspiration estimates from 29-year runs of the VIC land surface hydrology model when initialized with LAI response curves derived from the reference map, versus those from the four test maps.

were not assigned cropland (because cropland was underestimated by 85%). This non-linear relationship occurred because croplands tend to cluster, and when underestimated clusters tend to be small and isolated, they are more likely to fall outside of the search radius used by the model for allocating fields to households when they are initially seeded onto the landscape.

Land deficit (the total area of cropland that should have been allocated to households in each district, but wasn't) increased exponentially in relation to cropland underestimation–reaching around 800% for MODIS in districts 1 and 2 (Fig. 6B)–and would become infinite in the case of a 100% underestimate. This contrasted with food deficit (the percentage shortfall in the average amount of food production that should have been produced by
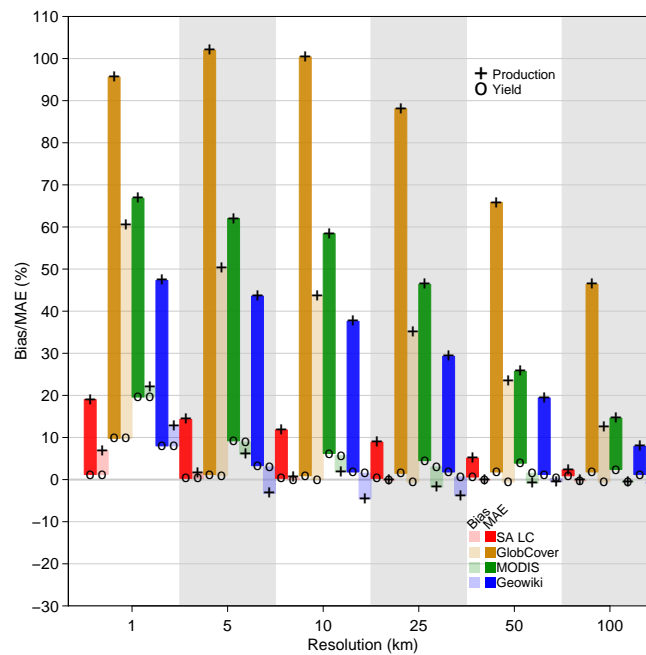

Figure 5: Bias (mean error) and accuracy (mean absolute error [MAE]) in disaggregated maize yield and production estimates. Bias estimates (represented by symbols) fall within the semi-transparent bars, mean absolute errors in the solid bars, with bar colors coded to specific cropland maps. Symbols code the different variables (production and yield), normalized to their respective means.

9

337 each household but wasn't), which increased linearly with the percentage of cropland underestimate (Fig. 6C).
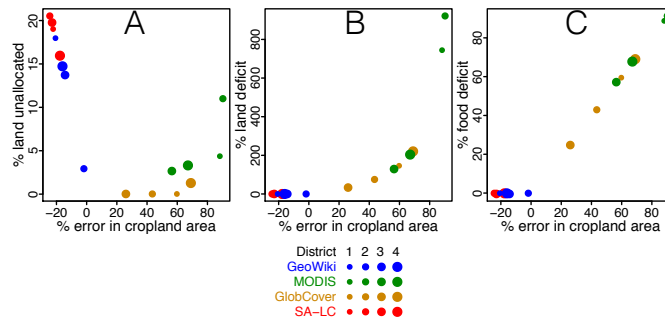

Figure 6: Biases in agent-based model results relative to the district-wise errors (as a percent) in total cropland area, in terms of A) the percent of cropland in each district that was not allocated to any household, B) the land deficit, or the total area of cropland that should have been allocated to households in each district but wasn't (expressed as a percent of total district cropland, as determined by test maps), and C) the food deficit, or the percentage shortfall (relative to the reference simulation) in mean household food production resulting from inadequate cropland allocation. Dot sizes correspond to district numbers, colors represent the land cover map.

Discussion

339 This spatially comprehensive, bottom-up land cover map assessment provides new insight into the extent and
340 causes of map error, and its consequences for understanding global change. The analysis was made possible
341 by a spatially comprehensive, high accuracy dataset that likely provides the truest measure of cropland area and
342 distribution currently available for this region. Although this reference map is not perfect, being affected by the
343 map-makers' occasional interpretation errors (mostly of omission, SI), while temporal mismatches between the
344 reference and test maps may account for some of the error we identified, our assessment (SI) suggests that such
345 discrepancies do not appreciably impact our findings. Our results are also supported by previous, continental-scale
346 assessments that revealed substantial errors and inconsistencies between land cover maps (Fritz *et al.*, 2010; Gross
347 *et al.*, 2013). These similarities, combined with the larger extents of these studies, also suggest that our findings
348 are applicable beyond South Africa's borders.

General guidelines for selecting and using land cover data

350 Our findings regarding bias and accuracy suggest several guidelines for selecting and using land cover data in global
351 change research (Table 1), and suggested aggregation scales. These guidelines vary according to the intended
352 application, what error property the users wants to minimize or maximize, and what tradeoff the user is willing
353 to make between error and resolution. For the purpose of clarity in these guidelines, we assume that a maximum
354 error magnitude of 10% is acceptable, but the results of our analyses (Fig. 1, 3, 5) allow different thresholds to be
355 evaluated.
356 The first general guideline is that land cover products derived from coarse resolution sensors, such as MODIS
357 and GlobCover, need substantial aggregation to reduce the large biases and inaccuracies that manifest at finer
358 resolutions. Researchers might need to use these datasets for a variety of reasons, for instance, because they
359 provide consistent, multiple land cover classes at a global scale, or allow changes to be tracked across large scales
360 over time (Luoto *et al.*, 2004). The appropriate scale of aggregation then depends on which error property the
361 user wants to minimize. If the objective is to minimize systematic errors in order to calculate an unbiased mean
362 (e.g. average carbon density), then at least 25-50 km of aggregation is needed to reduce map bias to within +/-10%
363 (Table 1, Fig. 1B & 3). If the user wants to maximize accuracy, which is important for comparing differences
364 between locations on a single map, or between dates at fixed locations, then even coarser aggregations are needed
365 to reduce mean absolute error below 10% (Fig. 1, 3, 5).
366 The characteristics of cover at the landscape scale may also require greater levels of aggregation to achieve
367 sufficient accuracy. We found that accuracy is lowest where mixing of cover types is greatest (Fig. 2), a result
368 echoed in an Africa-wide change detection study made with MODIS (Gross *et al.*, 2013). Outside of South Africa,
369 where smallholder-driven agriculture predominates (Lambin *et al.*, 2013), such mixed landscapes are more com-
370 mon, thus greater levels of aggregation may be needed over these areas. Coarser aggregations are also necessary
371 if the cover type of interest belongs primarily to a mixed class (as with GlobCover). This leads to underestimation
372 bias (Fig. 2) and high inaccuracy, which will persist until the aggregated pixel exceeds the characteristic size of

the geographical area within which the cover type dominates. In South Africa's farming regions, this can exceed 1000 km^2.

Maps derived from higher resolution sensors, such as the SA-LC dataset, do not have this mixture problem because the pixels are fine enough to be assigned to a single cover type. This characteristic leads to higher accuracy, which is what we observed here: the SA-LC dataset needed the least aggregation–just 5-10 km–to achieve <10% MAE in most of the example applications we tested (note, however, another guideline in this result: even with high resolution maps some aggregation is needed to achieve higher accuracy).

Unfortunately, Landsat-scaled maps are typically developed for specific countries, using varying methods, and can be hard to obtain (an exception to this is the newly released GLC30, which has a reported classification accuracy of 80% (Chen *et al.*, 2015)). Our results also suggest that, despite their higher accuracy, they may not necessarily be the least biased maps. Except for a few results (notably the 1 km carbon and maize production estimates), GeoWiki-based maps had the lowest bias across most scales, which is the largely the result of calibrating the cropland percentages to match reported agricultural statistics (Fritz *et al.*, 2015, 2011). This procedure is similar to the one we used in disaggregating maize harvested areas (Monfreda *et al.*, 2008; Ramankutty *et al.*, 2008), which led to fairly unbiased estimates of maize production and yield (Fig. 5) above 1 km (except for GlobCover's production estimates). This result, together with GeoWiki's unbiased cropland estimates (Fig 1.B), indicates the value of fusing inventory data with remote sensing to reduce map bias.

Of course, for the constraint to be effective, the inventory data themselves must be faithful to reality, but in many places, particularly Africa, agricultural statistics are suspect (Carletto *et al.*, 2013; FAO & Bank, 2013), which is a vulnerability of the fusion approach (See *et al.*, 2015). Beyond this limitation, this method does not improve map accuracy, which is evident in GeoWiki's relatively high MAE for cropland percentage at 1 km (23%, Fig. 1B), and by all maps' (except SA-LC's) highly inaccurate maize production estimates (Fig. 5). The reason that accuracy is not improved is because the statistical constraint cannot correct the omission and commission errors within the land cover maps. To minimize these types of errors, different techniques are needed, such as creating a single consensus map from multiple land cover products (e.g. Fritz *et al.*, 2010; Tuanmu & Jetz, 2014). This second kind of data fusion was used in the first stage of creating the GeoWiki map (before statistical adjustment), and it accounts for GeoWiki's higher accuracy scores relative to MODIS and GlobCover (Fig. 1B, 3, & 5). It may be particularly effective in minimizing underestimation errors that result when a mixed class type becomes dominant (as with GlobCover), or when dealing with substantial sub-pixel heterogeneity (Fritz *et al.*, 2015; Tuanmu & Jetz, 2014)

Table 1: Guidelines for selecting and using different classes of land cover product (based on the characteristics of the satellite sensor(s) and map-making methods), according to the general purpose of the application, the error minimization objective, and the range of aggregation needed to achieve the error objective (here we use a maximum acceptable error magnitude of 10%). The preferred land cover product class for each application is indicated in boldface.

Application	Objective	Source of land cover Map	Minimum aggregation*
High frequency change detection	Maximize accuracy	Coarse resolution sensor (e.g. MODIS)	>100 km
		High resolution sensor (e.g. Landsat)	1-25 km
Mean values (e.g. carbon density)	Minimize bias	Coarse resolution sensor	25-50 km
		if mixed class dominates	>100 km
		Fusion product: statistical constraint	1-5 km
		High resolution sensor	1-10 km
Spatial comparisons (e.g. crop production)	Maximize accuracy	Coarse resolution sensor	25-100 km
		if mixed class dominates	>100 km
		Fusion product: map consensus**	10-50 km
		High resolution sensor	1-25 km

*Aggregation ranges account for variation in the results from the different map analyses.
**This assumes that consensus-based fusion products (e.g. Fritz *et al.*, 2015, 2011; Tuanmu & Jetz, 2014) incorporate maps from coarse resolution sensors, making them less accurate at finer scales than products created using high resolution sensors.

These last few points suggest two final guidelines. For high accuracy, users should select well-validated datasets from higher resolution sensors–this option should be increasingly available as computational power grows and high resolution image archives expand (Chen *et al.*, 2015; Hansen *et al.*, 2013); 2) where low bias is needed over large areas, the best option is to use newer generation, statistically constrained maps such as GeoWiki (and possibly the GLC-Share [3] datasets for other cover types). But in either case some aggregation–1-25 km, depending on application and whether low bias or high accuracy is the objective–will still likely be necessary to achieve sufficiently low error. Additionally, these higher accuracy, lower bias products often represent just a single time

---

[3]GLC-Share. www.glcn.org

period (e.g. Fritz *et al.*, 2015; Tuanmu & Jetz, 2014), thus applications requiring higher frequency (e.g. annual), large area land cover maps will depend on MODIS or similar sensors, and therefore much coarser aggregations.

Implications for global change analyses

Besides the preceding land cover data selection and usage guidelines, our findings also have broader implications for global change analyses and associated policy decisions. The simpler, most direct uses of land cover data have substantial potential to skew our understanding of global change processes. For example, the global carbon carbon density map of (Ruesch & Gibbs, 2008), of which we created facsimiles here, has been widely used, but our results suggest that any analyses of the spatial variability of carbon stocks, or of mean carbon densities, can be highly misleading if the map is not first sufficiently aggregated. For example, a study of deforestation risks in the Democratic Republic of the Congo estimated forest rents as a function of carbon densities based on this map. Although the map was aggregated to ∼50 km, we found that mean absolute errors can still reach 40% at this scale (Fig. 3). The same carbon map provided country-specific mean forest carbon densities for an evaluation of climate mitigation policies (Cattaneo *et al.*, 2010), which our results suggest could be substantially underestimated (Fig. 3). In contrast to these, an assessment of the congruence between carbon and biodiversity (Strassburg *et al.*, 2010) aggregated the carbon densities to ∼110 km, where accuracy is much higher.

A related example can be seen in high resolution maps (10 km) showing the potential food production benefit of closing yield gaps (e.g. Figure 3 in Foley *et al.*, 2011). Such maps may give an inflated sense of the precision with which such locations can be identified. Although the study in this example is careful to report aggregated statistics and maps the per hectare values, and the map-makers stress that their products should not be used at the grid cell scale[4], it is not inconceivable that users might make decisions based on their visual interpretations of maps. For this reason, it is advisable to aggregate presented maps until their accuracy reaches an acceptable standard (Table 1).

What then, of the impacts of land cover map error on more complex analysis? Here the results appear to be mixed. Map error appears to have had little impact on the overall bias and accuracy of VIC's evapotranspiration (ET) estimates, which contrasts with a related study that examined how land cover maps impact rainfall simulations (Ge *et al.*, 2007). There are two reasons for this. First, map accuracy increased nearly two-fold by aggregating to 25 km, the resolution of VIC. Second, VIC's land cover-derived variables (e.g. LAI curves, effective rooting depth) could dampen the effect of map errors on ET calculations when the values of these variables in the adjacent land covers are similar to that of cropland, which is the case in much of VIC's land cover scheme for South Africa. In other regions, where the matrix land cover's LAI values depart substantially from cropland's (e.g. in forest), the ET errors would likely be much larger. The carbon example illustrates how cover properties can mute or amplify map error.

Despite the low values of the overall error metrics, the hotspots of ET errors (Fig. 4) may have outsize significance because they occurred at the climatic margins of the major crop growing regions, where a substantial share of farms are irrigated. Since VIC does not simulate irrigation, the magnitude of these errors is likely to have been underestimated, because irrigation substantially increases latent heat flux and can substantially modify regional climate (Mueller *et al.*, 2015; Sacks *et al.*, 2008). This also suggests the potential for map error to magnify in a coupled land-atmosphere model (which could explain the difference between this result and that of Ge *et al.*, 2007), and thereby give misleading insight into the climatic effects of land use.

Relative to the ET example, map errors had a much larger impact on the agent-based model's results. For the model's most important metric of food security–household-level crop production–the result was fairly straightforward for a class of models where the interactions of multiple agents often produce analytically intractable outcomes (Janssen & Ostrom, 2006). In this case, the model was only sensitive to cropland underestimates, which had the perfectly predictable effect of lowering average household food production, which would in turn overstate the degree of food *in*security. Overestimates did not matter in this case, because the number of households remained fixed, and their cropland holdings were constrained to match the survey statistics (but this would matter for models that allow these parameters to vary). Less predictable was the finding that the spatial distribution of cover can impact model parameterization, with land being left unallocated even when needed, thus such models also require spatial, not just statistical, accuracy in their base land cover maps.

The errors in the cropland data we tested were large, and it is unlikely that current ABM-based studies, mostly of small regions, are based on such inaccurate land cover data. However, as computing power increases, so, too, does the feasible application extent of ABMs, which will then become more reliant on larger-scaled, more erroneous, land cover data, with the attendant risk that insights into socio-ecological systems arising from these models could be highly misleading.

Beyond these examples, there are other applications where land cover map errors could alter understanding and policy. One of these lies in assessments of land availability for agricultural expansion, which could be particularly erroneous if they use a "residual approach" (Lambin *et al.*, 2013), where existing cropland and other occupied cover

---

[4]www.earthstat.orgwp-contentuploadsMETADATA_HarvestedAreaYield175Crops.pdf

types are used to filter a map of potential suitability. Our results suggest that the pronounced underestimation biases (Fig. 1) we found could inflate land availability estimates, and therefore increase the risk of unjust land allocation policies (Rulli *et al.*, 2013). Another example is the growing efforts to minimize the environmental and climate impacts on new agricultural expansion frontiers, particularly Africa's savannas, which requires identifying areas that provide the highest agricultural benefit for the lowest environmental cost (Searchinger *et al.*, 2015). This sort of analysis entails finding "win-win" land use solutions, which are essentially areas that deviate from an expected development cost-benefit relationship; for example, regions of unusually low carbon density but high crop fertility (Searchinger *et al.*, 2015). The validity of such analysis are entirely dependent In order to be effective, such analyses must be undertaken at the fine scales–hectares, rather than square kilometers–at which land use decisions are made, thus their value for informing land use policy is contingent on the accuracy of their underlying land cover/land use data (Searchinger *et al.*, 2015). The validity of such analyses, which must be undertaken at the fine-scales where land use decisions are made (hectares rather than square kilometers), depends on having high quality map data, in order to have confidence that the identified tradeoffs are based on real biophysical properties, not simply map error.

The way forward

Our analysis demonstrates that the land cover data that increasingly inform much global change research and policy can be substantially misleading, depending on the dataset selected, its application case, the scale to which it is aggregated, and whether the insight being drawn from the map depends on how accuracy or low bias. We provide some basic guidelines (Table 1) for selecting and using land cover data that can help increase confidence in the conclusions drawn from land cover-based studies.

Beyond these guidelines, we recommend updating several important, global-scale datasets using the latest generation land cover maps. For example, rebuilding the widely used cropland distribution and yield maps (Monfreda *et al.*, 2008; Ramankutty *et al.*, 2008) using the GeoWiki map could substantially improve our knowledge about current global agricultural distribution and productivity. Similarly, reconstructing the IPCC's Tier 1 vegetative carbon map (Ruesch & Gibbs, 2008) using a combination of the GeoWiki and new global land cover consensus maps (e.g. GLC-Share or Tuanmu & Jetz, 2014) the GLC-Share datasets would improve our confidence in regional to global assessments of vegetative carbon stocks. Alternatively, the new high-resolution, Landsat-based GLC30 (Chen *et al.*, 2015), aggregated and converted to fractional cover classes at 1 km–and fused with high quality national land cover maps, where available (similar to Fritz *et al.*, 2015, 2011)–may provide a preferable base map for rebuilding these datasets.

Fusing human judgement with expanding high resolution satellite data archives, improved classification algorithms, and expanding computing power appears to offer the greatest promise for improving the quality of land cover maps. The GeoWiki project and new high resolution global forest change maps (Hansen *et al.*, 2013; Sexton *et al.*, 2015) are prominent examples of these newer methods. Another approach is that of the Mapping Africa project[5], which, inspired by quality of the reference map used in this study, enlists internet-based workers to manually digitize cropland boundaries (Estes *et al.*, 2016a). Merging this crowdsourced approach for creating vector maps of discrete cover types with newer computer vision-based mapping algorithms (Debats *et al.*, 2016) is a near-term goal of this project, as the former can interactively provide high quality training and test data (by using humans' superior capacity to delineate discrete cover types in noisy images (Estes *et al.*, 2016a)) for the latter, which has the advantage of speed and the ability to handle high-dimensional datasets. Approaches such as these will be needed to create the next generation of land cover maps, so that we can gain a clearer, unbiased understanding of global change.

Acknowledgments

References

Arino O, Perez JJR, Kalogirou V, Bontemps S, Defourny P, Bogaert EV (2012) *Global Land Cover Map for 2009 (GlobCover 2009)*. PAN-GAEA.

Asner GP, Powell GVN, Mascaro J, *et al.* (2010) High-resolution forest carbon stocks and emissions in the Amazon. *Proceedings of the National Academy of Sciences*, **107**, 16738–16742.

Berger T, Schreinemachers P (2006) Creating agents and landscapes for multiagent systems from random samples. *Ecology and Society*, **11**, 19.

---

[5]http://mappingafrica.princeton.edu

Carletto C, Jolliffe D, Banerjee R (2013) The emperor has no data! Agricultural statistics in Sub-Saharan Africa. *World Bank*.

Cattaneo A, Lubowski R, Busch J, Creed A, Strassburg B, Boltz F, Ashton R (2010) On international equity in reducing emissions from deforestation. *Environmental Science & Policy*, **13**, 742–753.

Chen J, Chen J, Liao A, *et al.* (2015) Global land cover mapping at 30 m resolution: A POK-based operational approach. *ISPRS Journal of Photogrammetry and Remote Sensing*, **103**, 7–27.

Chen P, Plale B, Evans T (2013) Dependency Provenance in Agent Based Modeling. In: *2013 IEEE 9th International Conference on eScience (eScience)*, pp. 180–187.

Debats SR, Luo D, Estes LD, Fuchs TJ, Caylor KK (2016) A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment*, **179**, 210–221.

Estes LD, Beukes H, Bradley BA, *et al.* (2013a) Projected climate impacts to South African maize and wheat production in 2055: A comparison of empirical and mechanistic modeling approaches. *Global Change Biology*, **19**, 3762–3774.

Estes LD, Bradley BA, Beukes H, *et al.* (2013b) Comparing mechanistic and empirical model projections of crop suitability and productivity: Implications for ecological forecasting. *Global Ecology and Biogeography*, **22**, 1007–1018.

Estes LD, McRitchie D, Choi J, *et al.* (2016a) A platform for crowdsourcing the creation of representative, accurate landcover maps. *Environmental Modelling & Software*, **80**, 41–53.

Estes LD, Paroz LL, Bradley BA, *et al.* (2014) Using changes in agricultural utility to quantify future climate-induced risk to conservation. *Conservation Biology*, **28**, 427–437.

Estes LD, Searchinger T, Spiegel M, *et al.* (2016b) Reconciling agriculture, carbon and biodiversity in a savannah transformation frontier. *Phil. Trans. R. Soc. B*, **371**, 20150316.

FAO, Bank W (2013) *Action Plan of the Global Strategy to Improve Agricultural and Rural Statistics: For Food Security, Sustainable Agriculture and Rural Development*. FAO, Rome.

Foley JA, Ramankutty N, Brauman KA, *et al.* (2011) Solutions for a cultivated planet. *Nature*, **478**, 337–342.

Foody GM (2002) Status of land cover classification accuracy assessment. *Remote sensing of environment*, **80**, 185–201.

Fourie A (2009) Better Crop Estimates in South Africa. *ArcUser Online*.

Fritz S, McCallum I, Schill C, *et al.* (2012) Geo-Wiki: An online platform for improving global land cover. *Environmental Modelling & Software*, **31**, 110–123.

Fritz S, See L, McCallum I, *et al.* (2015) Mapping global cropland and field size. *Global Change Biology*, **21**, 1980–1992.

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**, 2237–2256.

Fritz S, See L, You L, *et al.* (2013) The Need for Improved Maps of Global Cropland. *Eos, Transactions American Geophysical Union*, **94**, 31–32.

Fritz S, You L, Bun A, *et al.* (2011) Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets. *Geophysical Research Letters*, **38**, L04404.

Fry J, Coan M, Homer C, Meyer D, Wickham J (2009) Completion of the National Land Cover Database (NLCD) 1992-2001 Land Cover Change Retrofit Product. USGS Numbered Series 2008-1379, U.S. Geological Survey.

Gaveau DLA, Salim MA, Hergoualc'h K, *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**.

Ge J, Qi J, Lofgren BM, Moore N, Torbick N, Olson JM (2007) Impacts of land use/cover classification accuracy on regional climate simulations. *Journal of Geophysical Research: Atmospheres*, **112**, D05107.

GeoTerraImage (2013) 2011 KZN Province land-cover mapping. Data users report and meta data.

Gross D, Dubois G, Pekel JF, *et al.* (2013) Monitoring land cover changes in African protected areas in the 21st century. *Ecological Informatics*, **14**, 31–37.

Hansen MC, Potapov PV, Moore R, *et al.* (2013) High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science*, **342**, 850–853.

Hardy M, Dziba L, Kilian W, Tolmay J (2011) Rainfed Farming Systems in South Africa. In: *Rainfed Farming Systems* (eds. Tow P, Cooper I, Partridge I, Birch C), pp. 395–432. Springer Netherlands.

Hastie T, Tibshirani R (1990) *Generalized Additive Models*. Chapman and Hall.

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.

Janssen MA, Ostrom E (2006) Empirically based, agent-based models. *Ecology and Society*, **11**, 37.

Jones J, Hoogenboom G, Porter CH, *et al.* (2003) Modelling Cropping Systems: Science, Software and Applications. *European Journal of Agronomy*, **18**, 235–265. 3-4.

14

Kuemmerle T, Erb K, Meyfroidt P, *et al.* (2013) Challenges and opportunities in mapping land use intensity globally. *Current Opinion in Environmental Sustainability*, **5**, 484–493.

Lambin E, Gibbs H, Ferreira L, *et al.* (2013) Estimating the world's potentially available cropland using a bottom-up approach. *Global Environmental Change*.

Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. *Progress in Physical Geography*, **21**, 375–393.

Lambin EF, Geist HJ, Lepers E (2003) Dynamics of land-use and land-cover change in tropical regions. *Annual Review of Environment and Resources*, **28**, 205–241.

Lark TJ, Salmon JM, Gibbs HK (2015) Cropland expansion outpaces agricultural and biofuel policies in the United States. *Environmental Research Letters*, **10**, 044003.

Larsen AE, Hendrickson BT, Dedeic N, MacDonald AJ (2015) Taken as a given: Evaluating the accuracy of remotely sensed crop data in the USA. *Agricultural Systems*, **141**, 121–125.

Liang X, Lettenmaier DP, Wood EF, Burges SJ (1994) A simple hydrologically based model of land surface water and energy fluxes for general circulation models. *Journal of Geophysical Research*, **99**, 14415.

Licker R, Johnston M, Foley JA, Barford C, Kucharik CJ, Monfreda C, Ramankutty N (2010) Mind the gap: How do climate and agricultural management explain the yield gap of croplands around the world? *Global Ecology and Biogeography*, **19**, 769–782.

Linard C, Gilbert M, Tatem AJ (2010) Assessing the use of global land cover data for guiding large area population distribution modelling. *GeoJournal*, **76**, 525–538.

Luoto M, Virkkala R, Heikkinen RK, Rainio K (2004) Predicting bird species richness using remote sensing in boreal agricultural-forest mosaics. *Ecological Applications*, **14**, 1946–1962.

Monfreda C, Ramankutty N, Foley JA (2008) Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000. *Global Biogeochemical Cycles*, **22**, GB1022.

Mueller ND, Butler EE, McKinnon KA, Rhines A, Tingley M, Holbrook NM, Huybers P (2015) Cooling of US Midwest summer temperature extremes from cropland intensification. *Nature Climate Change*.

Newbold T, Hudson LN, Hill SLL, *et al.* (2015) Global effects of land use on local terrestrial biodiversity. *Nature*, **520**, 45–50.

Olofsson P, Foody GM, Herold M, Stehman SV, Woodcock CE, Wulder MA (2014) Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, **148**, 42–57.

Olofsson P, Foody GM, Stehman SV, Woodcock CE (2013) Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, **129**, 122–131.

Olofsson P, Stehman SV, Woodcock CE, *et al.* (2012) A global land-cover validation data set, part I: Fundamental design principles. *International Journal of Remote Sensing*, **33**, 5768–5788.

Ozdogan M, Woodcock CE (2006) Resolution dependent errors in remote sensing of cultivated areas. *Remote Sensing of Environment*, **103**, 203–217.

Quaife T, Quegan S, Disney M, Lewis P, Lomas M, Woodward FI (2008) Impact of land cover uncertainties on estimates of biospheric carbon fluxes. *Global Biogeochemical Cycles*, **22**, GB4016.

Ramankutty N, Evan AT, Monfreda C, Foley JA (2008) Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000. *Global Biogeochemical Cycles*, **22**, 19 PP.

Ruesch A, Gibbs HK (2008) New IPCC Tier-1 global biomass carbon map for the year 2000. *Carbon Dioxide Information Analysis Center (CDIAC), Oak Ridge National Laboratory, Oak Ridge, Tennessee. Available online at: http://cdiac. ornl. gov/epubs/ndp/global_carbon/carbon_documentation. html*.

Rulli MC, Saviori A, D'Odorico P (2013) Global land and water grabbing. *Proceedings of the National Academy of Sciences*, **110**, 892–897.

Sacks WJ, Cook BI, Buenning N, Levis S, Helkowski JH (2008) Effects of global irrigation on the near-surface climate. *Climate Dynamics*, **33**, 159–175.

SANBI (2009) National Landcover 2009. Tech. rep., South African National Biodiversity Institute; National Department of Environmental Affairs and Tourism, Pretoria, South Africa.

Schmit C, Rounsevell MDA, La Jeunesse I (2006) The limitations of spatial land use data in environmental analysis. *Environmental Science & Policy*, **9**, 174–188.

Searchinger TD, Estes L, Thornton PK, *et al.* (2015) High carbon and biodiversity costs from converting Africa's wet savannahs to cropland. *Nature Climate Change*, **5**, 481–486.

See L, Fritz S, You L, *et al.* (2015) Improved global cropland data as an essential ingredient for food security. *Global Food Security*, **4**, 37–45.

Sexton JO, Noojipady P, Song XP, *et al.* (2015) Conservation policy and the measurement of forests. *Nature Climate Change*, **advance online publication**.

Shackelford GE, Steward PR, German RN, Sait SM, Benton TG (2015) Conservation planning in agricultural landscapes: Hotspots of conflict between agriculture and nature. *Diversity and Distributions*, **21**, 357–367.

Sheffield J, Wood EF, Chaney N, *et al.* (2013) A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security. *Bulletin of the American Meteorological Society*, **95**, 861–882.

Stehman SV, Olofsson P, Woodcock CE, Herold M, Friedl MA (2012) A global land-cover validation data set, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing*, **33**, 6975–6993.

Strassburg BB, Kelly A, Balmford A, *et al.* (2010) Global congruence of carbon storage and biodiversity in terrestrial ecosystems. *Conservation Letters*, **3**, 98–105.

Sweeney S, Ruseva T, Estes L, Evans T (2015) Mapping Cropland in Smallholder-Dominated Savannas: Integrating Remote Sensing Techniques and Probabilistic Modeling. *Remote Sensing*, **7**, 15295–15317.

Tuanmu MN, Jetz W (2014) A global 1-km consensus land-cover product for biodiversity and ecosystem modelling. *Global Ecology and Biogeography*, **23**, 1031–1045.

Verburg PH, Neumann K, Nol L (2011) Challenges in using land use and land cover data for global change studies. *Global Change Biology*, **17**, 974–989.

Wood SN (2001) Mgcv: GAMs and Generalized Ridge Regression for R. *R News*, **1**, 20–25.

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**, 4134–4139.

Yu L, Liang L, Wang J, *et al.* (2014) Meta-discoveries from a synthesis of satellite-based land-cover mapping research. *International Journal of Remote Sensing*, **35**, 4573–4588.