A Large-Area, Spatially Continuous Assessment of Land Cover Map Error and Its Impact on Downstream Analyses

Lyndon Estes[1,2,3*], Peng Chen[4], Stephanie Debats[3], Tom Evans[4], Stefanus Ferreira[5], Tobias Kuemmerle[6,7], Gabrielle Ragazzo[3], Justin Sheffield[3,8], Adam Wolf[9], Eric Wood[2], and Kelly Caylor[3,10]

[1]*Graduate School of Geography, Clark University, Worcester, MA USA*
[2]*Woodrow Wilson School, Princeton University, Princeton, NJ USA*
[3]*Civil and Environmental Engineering, Princeton University, Princeton, NJ USA*
[4]*Indiana University, Bloomington, IN USA*
[5]*GeoTerraImage, Pretoria, RSA*
[6]*Geography Department, Humboldt University, 10099 Berlin, Germany*
[7]*Integrative Research Institute for Transformations in Human-Environment Systems, Humboldt University, 10099 Berlin, Germany*
[8]*Geography and Environment, University of Southampton, Southampton, United Kingdom*
[9]*Arable Labs, Princeton, NJ USA*
[10]*Bren School of Environmental Science and Management, University of California Santa Barbara, Santa Barbara, CA USA*

*+1-202-431-0496 (p) | +1-508-793-8881 (f) | LEstes@clarku.edu

Abstract

Land cover maps increasingly underlie research into socioeconomic and environmental patterns and processes, including global change. It is known that map errors impact our understanding of these phenomena, but quantifying these impacts is difficult because many areas lack adequate reference data. We used a highly accurate, high-resolution map of South African cropland to assess 1) the magnitude of error in several current generation land cover maps, and 2) how these errors propagate in downstream studies. We first quantified pixel-wise errors in the cropland classes of four widely used land cover maps at resolutions ranging from 1 to 100 km, then calculated errors in several representative "downstream" (map-based) analyses, including assessments of vegetative carbon stocks, evapotranspiration, crop production, and household food security. We also evaluated maps' spatial accuracy based on how precisely they could be used to locate specific landscape features. We found that cropland maps can have substantial biases and poor accuracy at all resolutions (e.g. at 1 km resolution, up to ∼45% underestimates of cropland (bias) and nearly 50% mean absolute error (MAE, describing accuracy); at 100 km, up to 15% underestimates and nearly 20% MAE). National-scale maps derived from higher resolution imagery were most accurate, followed by multi-map fusion products. Constraining mapped values to match survey statistics may be effective at minimizing bias (provided the statistics are accurate). Errors in downstream analyses could be substantially amplified or muted, depending on the values ascribed to cropland-adjacent covers (e.g. with forest as adjacent cover, carbon map error was 200-500% greater than in input cropland maps, but ∼40% less for sparse cover types). The average locational error was 6 km (600%). These findings provide deeper insight into the causes and potential consequences of land cover map error, and suggest several recommendations for land cover map users.

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, it is essential to have an accurate understanding of the nature and distribution of land cover (Verburg *et al.*, 2011). 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 (Verburg *et al.*, 2011), much as a house requires a solid foundation in order to remain standing. Unfortunately, the evidence indicates that this house has shaky foundations (e.g. Fritz *et al.*, 2011a). The reason is that land cover data can only practically be derived from satellite imaging, which has several constraints that propagate mapping errors. First, in many regions the characteristic scales of cover features are smaller than the sensor resolution (e.g. smallholders' fields Debats *et al.*, 2016; Jain *et al.*, 2013b; Ozdogan & Woodcock, 2006), or the covers of interest are spectrally indistinct from neighboring ones (Fritz & See, 2008; 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; Fritz & See, 2008; Verburg *et al.*, 2011) and can 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 (Fritz & See, 2008; Kuemmerle *et al.*, 2013). Third, land cover maps are often used to detect changes (e.g. Gross *et al.*, 2013), but seasonal variability and cover changes can be easily confused. Given these multiple sources of error, land cover maps are often inaccurate and disagree widely between products, particularly in the world's most rapidly developing regions (Fritz *et al.*, 2010, 2013, 2011b). 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, 2011b; 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, 2015). 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 (Kuemmerle *et al.*, 2013; See *et al.*, 2015), what remains an open question is exactly how much the maps researchers typically use deviate from actual land cover, how this affects analyses based on these maps, and how this in turn impacts our understanding of processes being studied. 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 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). Although previous studies have examined how map errors propagate, these are primarily assessed using either simulated errors, relative differences in existing land cover maps, or ground validation data covering relatively small areas (e.g. Ge *et al.*, 2007; Linard *et al.*, 2010; Quaife *et al.*, 2008; Schmit *et al.*, 2006; Tuanmu & Jetz, 2014).

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 (Drusch *et al.*, 2012; Hand, 2015) and better mapping methods (Debats *et al.*, 2016; Estes *et al.*, 2013; Fritz *et al.*, 2012), are providing means to more comprehensively interrogate the accuracy and biases in land cover products that have become commonplace in global change research, and which often underpin 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

[1] e.g. 3DRobotics, DJIA

environmental conditions. Using a unique, high-resolution, high-quality 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 analyses?

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 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 in high resolution satellite imagery (Fourie, 2009). The resulting vectorized field boundaries provide unique, highly accurate data on field sizes and distribution for the 2009-2011 period. A previous study evaluated the accuracy of this dataset (Estes *et al.*, 2016a), using a visual assessment of cropland presence/absence in high resolution satellite images within 15,225 individual 4 ha plots (25 sub-plots within 609 1 km^2 grids) to evaluate the ability of the vector boundaries 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).

We used these vector data as a reference for evaluating four land cover products representative of those commonly used in global change studies and related areas of research (see the SI for an illustration of all five datasets in their original form). The first was South Africa's 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 those of 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 (Friedl *et al.*, 2010) 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 Global Land Cover Share (GLC-Share; Latham *et al.*, 2014), which harmonizes and merges the classes of the best available high and medium resolution products for each country/region, providing separate maps for cropland and 10 other thematic classes. Among the datasets included in the GLC-Share's cropland map is the GeoWiki cropland percentage map (Fritz *et al.*, 2015; Waldner *et al.*, 2016), which provided an important input to GLC-Share's cropland layer (Latham *et al.*, 2014). The GeoWiki map was also constructed using fusion techniques to merge multiple cropland datasets (including SA-LC, MODIS, and GlobCover), but further calibrated to match cropland areas reported in national agricultural statistics, and validated by crowdsourced volunteers (Fritz *et al.*, 2015). GLC-Share and the multi-map fusion methodology underlying it (and the GeoWiki map it incorporates) represents a state of the art approach for mapping agricultural land cover. Since GLC-Share provides a continuous value of land cover (cropland percentage), which cannot be feasibly converted to a categorical value, we converted all other datasets, including the reference vectors, into comparable 1 km gridded cropland percentages.

To convert the reference data to percentages, we intersected the field boundary vectors with a 1 km grid and calculated the percent of each cell covered by fields. In the resulting grid, we masked out areas classified as communal farmland (18.7% of cropland area in the reference data), because only their outer perimeters were digitized (Fourie, 2009), which risked overestimating cropland extent because those vectors enclosed uncropped inter-field areas. We also excluded areas under permanent tree crops, sugarcane plantations, and commercially afforestation, because these classes were not common to all five cropland datasets. To remove permanent tree crops (3.1% of cropland area), we masked out reference vectors labeled 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) in 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 cropland map errors

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 (to condition the analysis on cropland prevalence; see the SI for an assessment of error using other, non-weighted, variants of these measures), and accuracy is the mean of the absolute values of residuals (also weighted by cropland density), thus lower values signify higher accuracy. 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 the reference and test 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 cover 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. 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 5, 10, 25, 50, and 100 km 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).

Although this analysis ascribed hypothetical carbon densities to non-cropland areas, the selected values represent the range in potential carbon stocks in landscapes containing arable agriculture (the method for calculating error metrics only affects pixels containing cropland; pixels without cropland do not influence the results), which allowed us to counterfactually investigate how carbon estimates are affected by the interaction of i) test map errors and ii) the properties of neighboring cover types.

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 model uses 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), and correspondingly altered the fractions of non-cropland cover types to accommodate the adjusted cropland fractions. We then ran one instance of VIC for each of the five land cover schemes, and 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 crop 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. To obtain higher spatial resolutions, disaggregation of these statistics using gridded land cover data is common (Monfreda *et al.*, 2008; Ramankutty *et al.*, 2008; Schierhorn *et al.*, 2013).

We used these methods to first disaggregate the harvested area for maize (South Africa's largest crop; Estes *et al.*, 2013) onto our 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; Schierhorn *et al.*, 2013). 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 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 (Statistics South Africa, 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, to analyze socio-ecological system dynamics, and to test policy impact (Berger & Schreinemachers, 2006). To obtain robust insights, 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; Chen *et al.*, 2013). The model (described in more detail in the SI) is initialized so that each household is allocated a share of cropland, based on household number and cropland area estimated derived from survey statistics. Annual household crop production (maize) is simulated as a function of its field area, local weather, soil properties, and management actions, all of which can vary between households. The initialization process iteratively assigns households to the landscape as a function of neighbor and cropland proximity, ensuring that households are grouped into communities and that their fields are within a realistic proximity. To achieve this, the model first randomly places 100 households onto the simulated landscape, allocating each household its required cropland pixels, which must be within 1.5 km of the household. This process is iterated until all households are assigned cropland, or all available cropland is allocated. The model is considered to be well-calibrated when all households are allocated cropland, and all cropland is allocated to households.

Like many spatial ABMs, the model is computationally intensive and requires high spatial resolution to match the scale of individual fields, and therefore applied to smaller regions (e.g. districts). To meet these computational needs, we selected four contiguous magisterial districts (1,037-1,329 km^2) in eastern South Africa with similar climate and 28-45% of cropland coverage. To create cropland surfaces for each district, we disaggregated all five cropland maps into 100 m binary cropland/non-cropland cover maps, and then ran the model separately for each district and with each cropland map (20 simulations total). To examine 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, which respectively represent: 1) the share of district total cropland that was not assigned to any household (a measure of the model's effectiveness in matching households to available cropland); 2) the total area of cropland that should have been allocated to households in each district but wasn't (due to mismatches between the cropland map and the survey-based estimates of total cropland holdings); 3) 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.

The impact of map error on identifying specific locations

The bias and accuracy metrics reflect the degree to which quantitative estimates are influenced by land cover map errors. However, land cover data may also be used to identify specific locations (e.g. areas of high agricultural potential and low ecological cost; Estes *et al.*, 2016b), as opposed to general quantities. It is therefore important to also assess how map errors can impact the spatial accuracy of maps, or the ability to accurately locate specific features of interest. To evaluate this, we calculated the mean Euclidean distance (in km) between pixels representing a specific feature within the test maps relative to their nearest neighboring pixels representing the same feature in the reference map. The features in this analysis were simply those locations falling within the upper deciles of a) cropland cover, b) carbon density (based on the average carbon density of non-crop vegetation types), and c) crop yield. We confined this analysis to the 1 km resolution maps, as higher spatial resolutions are typically required when maps are used to identify locations rather than estimating quantities (e.g. Estes *et al.*, 2016b).

Results

Cropland map errors

Our 1 km reference cropland map indicated that crop fields covered 104,304 km^2 (nearly 10%) of the total study area in the 2009-2011 time period, with corresponding cropland area estimates of 131,390, 82,358, 77,090, and 110,272 km^2 resulting from the SA-LC, GlobCover, MODIS, and GLC-Share maps, respectively. Cropland area estimates from both the reference and test maps were constant for all levels of aggregation.

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 bias 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 GLC-Share 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, GLC-Share became slightly more accurate, having <5% MAE at 100 km resolution. The reason GLC-Share had relatively poor accuracy at 1 km resolution was due to the highly heterogeneous error pattern, which traded between positive and negative residuals over short distances, thereby inflating MAE at this scale.

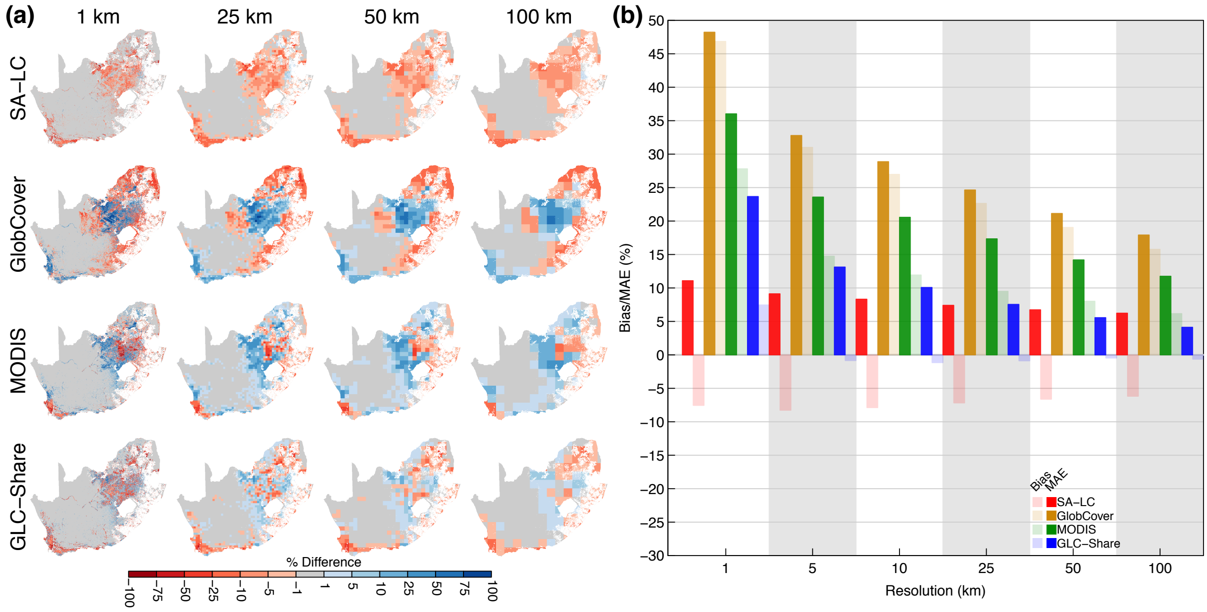

Figure 1: (a) Errors in the percent cropland estimates resulting from each of the four test maps relative to the reference map at different scale of pixel aggregation. Rows indicate the test map being assessed (by subtraction from the reference map), while columns refer to resolution of aggregation. 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.

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 was dominated either by cropland or by another type (Fig. 2). In other words, accuracy was lowest when cropland cover was 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.

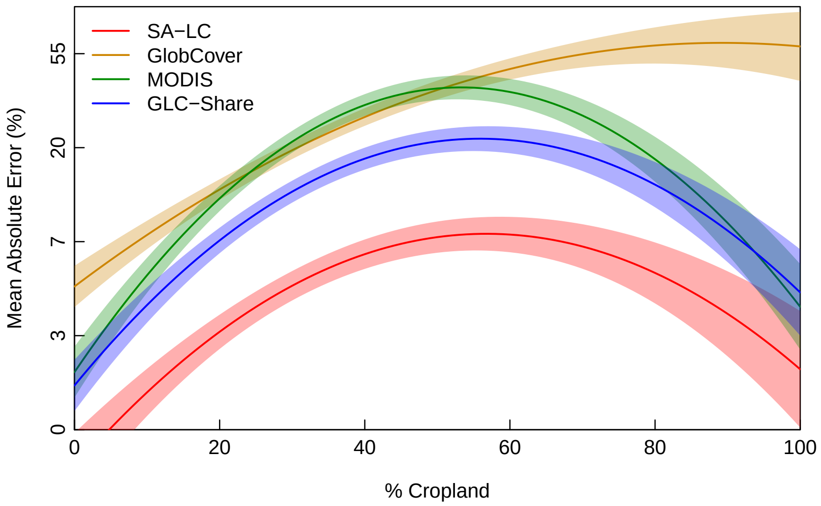

Figure 2: The relationship between map accuracy (the mean absolute error) in test maps and the actual cropland cover within agricultural landscapes (reference map pixels having >0.5% cropland), here defined by the boundaries of magisterial districts (n = 345), as fit with a generalized additive model. 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.

The impact of map error on downstream 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 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 GLC-Share had biases of 22% and -46%, 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 GLC-Share 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, GLC-Share produced the least biased carbon density estimates above 1 km resolution.

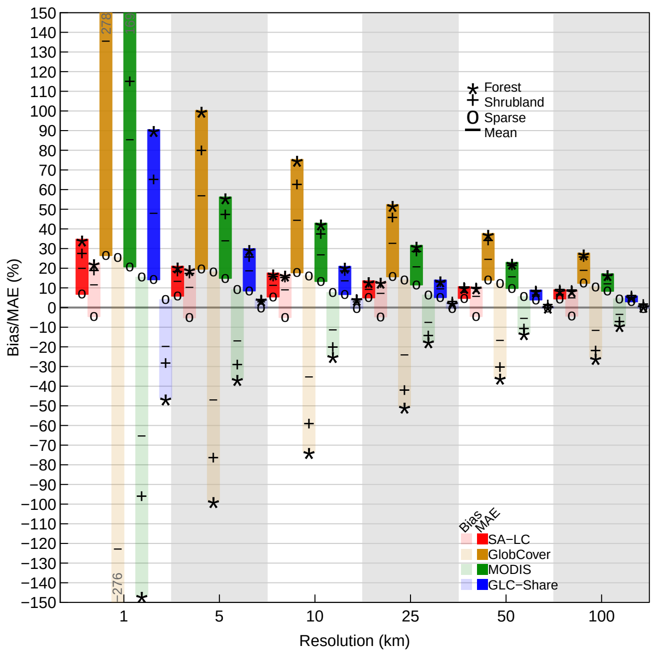

Figure 3: Biases and accuracies (mean absolute errors) of carbon densities derived from cropland maps, calculated as percents relative to the reference map. Bias estimates (represented by symbols) fall within the semi-transparent floating bars, while accuracies are contained in the solid bars. Bar colors 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 (see SI for all values). MODIS and GlobCover values at 1 km exceeding the plot's Y limits are provided near their truncated tops.

In terms of accuracy, MAE values were essentially the same as bias magnitudes, except for GLC-Share's, which were twice as large; GLC-Share's 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, where GLC-Share 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 GLC-Share 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.

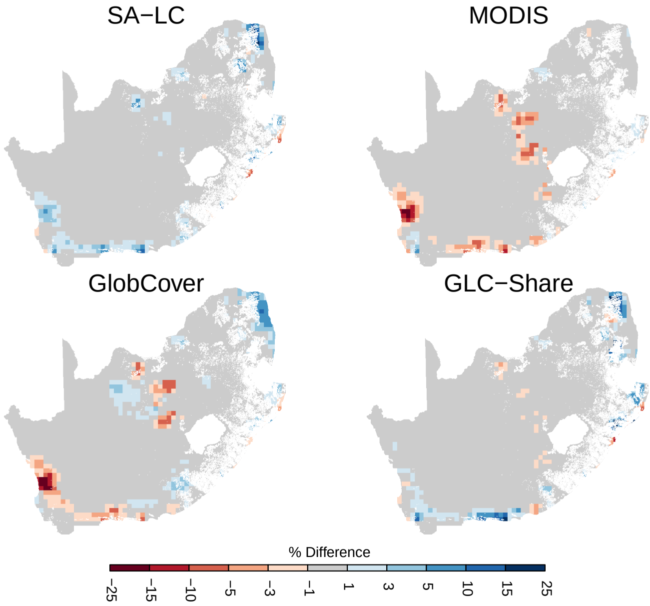

Figure 4: Differences in annual mean evapotranspiration estimates from 29-year runs of the VIC land surface hydrology model when initialized with LAI response curves derived from the reference map, versus those from the four test maps.

Downscaling 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 higher, but still low, for most datasets, with the exception of GlobCover, 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 GLC-Share. 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 spatial 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 were

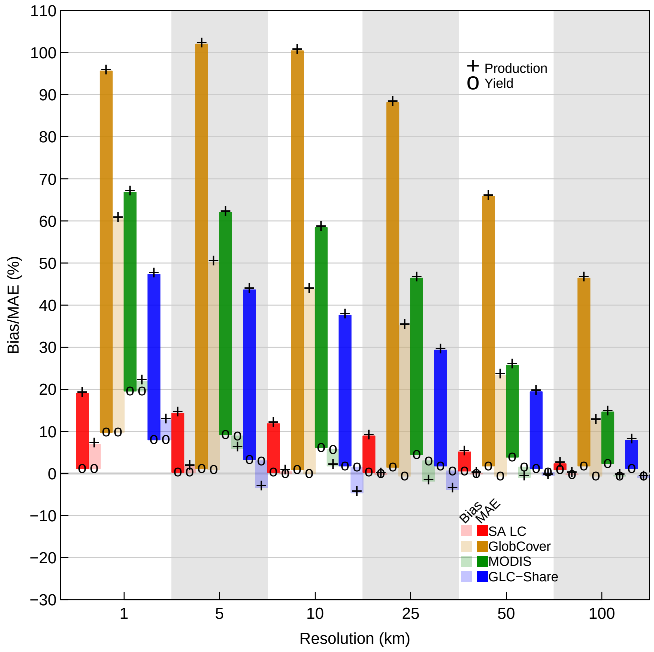

Figure 5: Bias (mean error) and accuracy (mean absolute error [MAE]) in disaggregated maize yield and production estimates. Bias estimates (represented by symbols) fall within the semi-transparent bars, mean absolute errors in the solid bars, with bar colors coded to specific cropland maps. Symbols code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

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

Location errors

The average distance between areas containing the highest cropland densities (upper decile) in the reference map and those delineated by the test maps ranged from 1.1 km for SA-LC to 18.2 km for GlobCover, with MODIS (10.1 km) and GLC-Share (2.8 km) having intermediate displacements (Fig. 7). Locational errors in maps indicating the highest yielding areas showed a

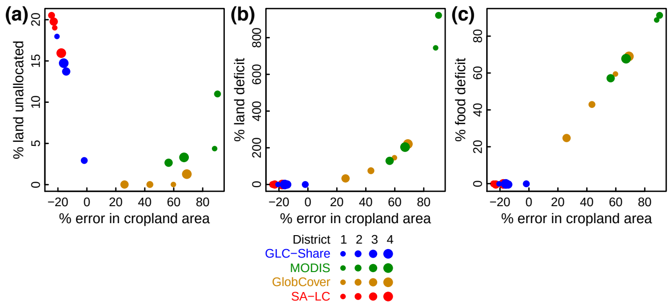

Figure 6: Biases in agent-based model results relative to the district-wise errors (as a percent) in total cropland area, in terms of a) the percent of cropland in each district that was not allocated to any household, b) the land deficit, or the total area of cropland that should have been allocated 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.

similar pattern, with a range of 0.8-14.2 km (SA-LC and GlobCover) and intermediate errors of 5.8-7.5 km (GLC-Share and MODIS). For areas of highest carbon density, locations identified by the MODIS-derived map were most distant from those shown by the reference map (11.3 km), followed by GLC-Share (7.4 km), GlobCover (6.8 km), and SA-LC (3.7 km).

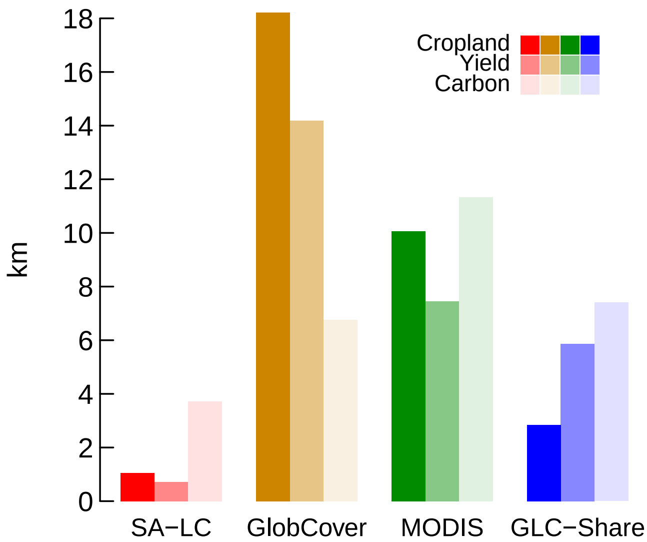

Figure 7: Average nearest neighbor distances (in km) between pixels representing features identified by the reference map versus those identified from the test maps. Bar colors indicate the different features (and thus contributing maps), which were delineated by selecting pixels with values greater than the 90th percentile: densest cropland (solid bars); highest maize yield (medium transparent bars); highest carbon density (most transparent bars).

Discussion

The preceding analyses contributes to existing work investigating land cover map error and its consequences (e.g. Fritz *et al.*, 2011a; Olofsson *et al.*, 2013; Verburg *et al.*, 2011). Previous studies have assessed map errors either by using point-based accuracy assessments (e.g. Foody, 2002; Frey & Smith, 2007; Olofsson *et al.*, 2013), by evaluating between-map discrepancies (e.g. Fritz & See, 2008; Fritz *et al.*, 2011a, 2010), or by comparing map-derived estimates to aggregated statistics (e.g. Fritz *et al.*, 2010; Larsen *et al.*, 2015; Yu *et al.*, 2014). A smaller number based their assessments on contiguous ground truth maps, but these covered relatively small regions (<3000 km^2, or <0.03% of the area covered here; Dendoncker *et al.*, 2008; Schmit *et al.*, 2006).

Other studies have also examined how map errors impact downstream analyses, including simulated rainfall (Ge *et al.*, 2007), carbon stocks and emissions (Goetz *et al.*, 2009; Jain *et al.*, 2013a; Olofsson *et al.*, 2013; Quaife *et al.*, 2008), nitrogen fluxes (Jain *et al.*, 2013a; Nol *et al.*, 2008), human population density (Linard *et al.*, 2010), species distributions (Tuanmu & Jetz, 2014), and landscape patterns (Langford *et al.*, 2006). The majority of these used either point validation, map inter-comparison, or a combination of both to assess errors (Goetz *et al.*, 2009; Jain *et al.*, 2013a; Linard *et al.*, 2010; Olofsson *et al.*, 2013; Quaife *et al.*, 2008; Tuanmu & Jetz, 2014). Others have used simulated map errors (Ge *et al.*, 2007; Langford *et al.*, 2006) or differences relative to small area ground truth maps (<1000 m; Nol *et al.*, 2008) to examine error propagation.

Our study builds on and goes beyond these previous efforts by providing a large-area, spatially continuous quantification of cropland classification errors, and by examining how these actual errors influence several common downstream applications. The spatially comprehensive nature of these analyses provides deeper insight into the causes and consequences of error than would otherwise be obtained from either a point-based accuracy assessment or through the imposition of simulated error. By assessing errors within a continuous estimate of cropland, we were also able to examine how pixel-wise errors change with scale, while minimizing the confounding effects of aggregating a categorical variable (Marceau & Hay, 1999; Moody & Woodcock, 1995, and see discussion in subsequent Recommendations section). These analyses were enabled by a high accuracy reference map that likely provides the truest measure of cropland area and distribution for this region. Although this reference map is not perfect, being affected by the map-makers' occasional interpretation errors (mostly of omission, SI), while temporal mismatches between the reference and test maps may account for some of the error we identified, our assessment (SI) suggests that such discrepancies do not appreciably impact our findings.

Sources of error in cropland maps

Our findings showed that the most accurate cropland estimates, across all spatial scales, were produced by SA-LC followed by GLC-Share. In the reverse order, these two datasets were also the least biased, with GLC-Share having effectively zero bias at aggregation levels of 5 km or coarser. GlobCover was highly inaccurate and biased at all scales, with MODIS estimates being nearly as inaccurate but substantially less biased. These error patterns are attributable to two factors: sensor resolution and methodology. With respect to the former, SA-LC's higher accuracy is largely because the 30 m (0.09 ha) resolution of Landsat imagery is smaller than the average area of South African crop fields. Previous work in agricultural remote sensing has shown that the sensor resolution should be finer than average field size to accurately estimate both the area and location of croplands(Ozdogan & Woodcock, 2006; Pax-Lenney & Woodcock, 1997). When pixel sizes are small relative to the objects being mapped, the number of "mixed pixels" (those where the spectral signature is defined by more than one cover) is relatively small, and their number naturally increases as sensor resolution decreases (Ozdogan & Woodcock, 2006).

Mixed pixels introduce another potential source of error related to methodology, which stems from the need to define thresholds for allocating pixels to different cover types (Ozdogan & Woodcock, 2006). This error is evident in the MODIS and GlobCover results, both of which cope with the mixing problem by assigning sub-pixel proportions of cropland to mosaic classes (Arino *et al.*, 2012; Friedl *et al.*, 2010). These classes place upper thresholds on cropland, causing underestimation error where actual cropland proportions are higher (Fig. 2B). Over South African croplands, the GlobCover map was dominated by its mosaic pixel classes, leading to substantial underestimation bias that persisted even with aggregation. MODIS, on the other hand, classified more areas as pure cropland, and thus had lower underestimation bias.

The modeled relationships between map accuracy and cropland density (Fig. 2) further demonstrate how pixel mixing and class definition influence error. For MODIS, SA-LC, and

GLC-Share, error was highest where pixels were evenly divided between cropland and other cover types, reflecting earlier work showing that classification accuracy is lowest when cover types are most mixed (Gross *et al.*, 2013; Verburg *et al.*, 2011). On the other hand, GlobCover was least accurate over pixels with 100% cropland, an error pattern imposed by the proportions defining cropland within GlobCover's dominant mosaic classes; these set an upper bound that necessarily led to underestimation error over dense croplands. Previous work by Ozdogan & Woodcock (2006) shows that the threshold used for separating agricultural from non-agricultural classes is a significant source of error.

GLC-Share's cropland errors demonstrate three other possible ways in which methodology affects error pattern. The first is that constraining remotely sensed cropland proportions to match census-statistics may be effective in reducing bias (this point is attributable to the GeoWiki product (Fritz *et al.*, 2011a, 2015), a statistically constrained map that is a major component of GLC-Share, although it is unclear how much GeoWiki's results dominate those of other maps in the GLC-Share fusion process). The second is that such a constraint cannot correct the errors of commission and omission within the individual landcover datasets that were merged to create the GLC-Share map (Fritz *et al.*, 2015), which is why its accuracy is relatively low at 1 km resolution (Fig. 1c). However, GLC-Share's accuracy was substantially higher than MODIS and GlobCover's (Fig. 1b, 3, & 5), which reveals the third point, namely that the landcover fusion process used to create GLC-Share does help minimize such error. Fusion may be particularly effective for minimizing underestimation errors caused by mixed classes (as with GlobCover) in areas of substantial sub-pixel heterogeneity (Fritz *et al.*, 2015; Tuanmu & Jetz, 2014). This fusion approach mirrors the ensemble methods used by various modeling sub-disciplines (e.g. crop (Asseng *et al.*, 2013), climate (Giorgi & Mearns, 2002), and ecological modeling (Araújo & New, 2007)) to increase prediction confidence.

Error propagation in downstream products

Cropland map errors were either amplified or muted within the various downstream applications we assessed. Both tendencies were evident in the Tier-1 carbon maps, the simplest of the downstream methods. Using the ratio between each carbon map's accuracy score (MAE) and that of its foundational cropland map to calculate error propagation (values >1 means error was exacerbated, <1 means it was muted), we see that errors in the 1 km carbon map errors were

200 (SA-LC) to 500% (GlobCover) larger than cropland map errors when forest was the adjacent cover type, but ∼40% lower for the sparse cover type. Aggregation helped to reduce error magnitude, but carbon maps nevertheless had 30-50% more error than cropland maps at 100 km resolution when forest or shrubland were sharing the pixel (the error ratio associated with the sparse vegetation carbon class remained relatively constant with aggregation).

The error propagation patterns in the carbon map were therefore determined by the differences between the carbon density of cropland and that of the adjacent cover types; forest and shrubland have higher carbon densities than cropland, whereas the sparse cover type is lower. A similarity in the values assigned to the cover types adjacent to cropland may also explain the low error rates in the evapotranspiration estimates produced by the VIC model, which were all 90-95% lower than those in the input cropland maps. This dampening of error contrasts with results from elsewhere showing that map errors can substantially alter rainfall simulations (Ge *et al.*, 2007). In the case of the ET simulations, VIC's map-related variables (e.g. LAI curves, effective rooting depth) were relatively similar between cropland and the adjacent landcover types, which we did not alter beyond adjusting their percentages to accommodate altered cropland proportions.

The disaggregated yield and crop production maps we created also showed both error amplification and muting, which in this case depended on the particular analysis. Errors within the yields maps were uniformly lower (50-90% less at 1 km) than those in the input cropland maps, whereas errors in the production maps were 70-100% higher at 1 km, and actually were exacerbated at intermediate levels of aggregation (10-25 km) for GlobCover, MODIS, and GLC-Share (170-290%). This latter tendency was mostly caused by accuracy in the production maps improving more slowly with aggregation (Fig. 5) than in the original cropland maps (Fig. 1B). These contrasting results reflect differences within disaggregation methods (Monfreda *et al.*, 2008). First, the yield methodology is a simple form of disaggregation that paints district-level yields onto pixels with >0% cropland within each district without attempting to map within-district yield variability. This simplicity means that it is only sensitive to errors in classifying cropland presence/absence, but not to errors in cropland proportions. Production, on the other hand, is calculated from disaggregated harvested area maps, which are created by proportionally allocating district-level crop harvested area onto cropland fractions, making them highly sensitive to errors in cropland proportion. Of particular note is that harvested area maps are subject not only to this statistical constraint (that pixel-wise harvested area fractions sum to district totals), but cropland fractions are also adjusted to match district-level cropland area estimates (see Methods; Ramankutty *et al.*, 2008). This suggests that statistical constraints are therefore not necessarily helpful in preventing pixel-level error propagation.

Sensitivity to cropland area also was evident in the food security model. The ABM's most important metric of food security–household-level crop production–was only impacted by underestimates of cropland area, which lowered the models' estimates of average household production (in turn overstating the degree of food *in*security) because individual households were allocated insufficient cropland. Cropland overestimates did not cause the opposite effect, because total households were constant and the allocation routine prevented cropland holdings from exceeding their assigned, census-derived hectarage. The less predictable result was that the model sometimes left cropland unallocated when maps substantially underestimated cropland area (e.g. MODIS in Fig. 6A). This initialization error was caused by the spatial arrangement of croplands in the district, which in the MODIS maps was clumped in relatively small islands within the four selected districts. This error interacted with the ABM's household placement routine, in that some cropland islands fell beyond the 1.5 km search radius of their nearest randomly placed households, resulting in those croplands being left unallocated. This result demonstrates how map errors can propagate through more complex models as a function of both model assumptions (here the choice of search radius) and model structures (the randomized household placement routine).

This latter finding also highlights the types of errors that can be caused by spatial inaccuracies in land cover maps, which was explicitly evaluated by the analysis of spatial distance between pixels containing upper decile of values in reference and test maps (Fig. 7). The relative size of offsets tends to follow the patterns of accuracy seen in other assessments, with SA-LC producing the most spatially accurate results, and GlobCover the least, with the exception that GlobCover's 90th percentile carbon density locations are closer to those in the reference map than those of MODIS or GLC-Share. Numerically, none of these nearest neighbor differences seem large, but they are akin to the root mean square error term used when measuring geometric distortion in a satellite image. Under this conception, the error in all but SA-LC's cropland and yield examples

(Fig. 7) is >3 pixels (or 300%), with an overall average of nearly six pixels.

Broader implications

Although our study focused on a single country and a subset of possible land cover-derived analyses, its findings highlight issues with broader geographic and practical relevance. In geographical terms, the key question is whether the error patterns revealed here will be similar outside of South Africa? Land systems in many regions differ substantially from those in southern Africa, and further research on the bias and accuracy of cropland maps is thus desirable. Yet, if one considers the rest of Sub-Saharan Africa (SSA), a region notorious for its lack of high-quality cropland maps, the answer is almost certainly yes. Previous work showing substantial disagreement between different cropland maps in other SSA countries (Fritz *et al.*, 2010) supports this contention. Another reason is that farming elsewhere in SSA is dominated by smallholders whose fields are substantially smaller than those in South Africa (Samberg *et al.*, 2016), which increases mixed pixel classification error. Additionally, smallholders' fields often contain residual trees, which create a park-like appearance that classifiers struggle to distinguish from the savannas that dominate the region (Debats *et al.*, 2016; Estes *et al.*, 2016b; Sweeney *et al.*, 2015).

These factors suggest that the cropland map errors are likely to be even larger than we found, as well as errors in downstream products. For example, carbon map errors should be on the higher end of those found here (Fig. 3), given the higher potential for error in the base cropland maps, combined with the fact that SSA's croplands lie mostly within savanna or forest biomes where the differences in carbon density between croplands and native vegetation would be higher (Searchinger *et al.*, 2015). A presumably greater difference between the LAI and rooting depths of crops and these dominant vegetation types may also increase ET estimation errors.

In terms of broader practical implications, these findings also suggest how maps errors could impact understanding of social and environmental processes and related policy. For example, assessments of land availability for new agricultural development could be misleading if they use a "residual approach" (Lambin *et al.*, 2013), in which potential lands are identified by masking out existing croplands and un-cultivatable lands (e.g. Estes *et al.*, 2016b). In such cases, cropland underestimates (such as those of MODIS and GlobCover; Fig. 1), could inflate estimates of available land, and thereby encourage erroneous land policy (Rulli *et al.*, 2013). Similarly, spatial errors in the cropland maps (Fig. 6) could cause the wrong land to be developed, by mislocating areas with preferred development characteristics (e.g. high agricultural potential and low environmental cost; Estes *et al.*, 2016b; Gasparri *et al.*, 2015). In addition to these possibilities, maps errors may misinform land use-focused emissions policies informed by analyses that rely on Tier-1 carbon maps (e.g. Cattaneo *et al.*, 2010; Phelps *et al.*, 2013). Disaggregated yield and harvested area maps could also mislead efforts to close crop production gaps, if specific interventions are targeted using finely resolved maps (e.g. the 10 km map shown in Figure 3 in Foley *et al.*, 2011). Improper understanding of complex, coupled human-natural systems could also result from models that have calibration errors caused by land cover maps.

Recommendations

Our findings suggest a number of recommendations for using land cover maps, complementing those suggested in earlier work (Verburg *et al.*, 2011). First, to minimize error, users should typically prefer maps derived from imagery with resolutions substantially finer than the scale of individual objects of interest (e.g. agricultural fields), assuming that available maps were created with rigorous classification methods accompanied by appropriate error metrics (Olofsson *et al.*, 2014), and are thematically appropriate for the intended use (Verburg *et al.*, 2011). Finer resolution not only helps to improve classification accuracy (see *Sources of error in cropland maps)*, but can minimize the *aggregation problem*, one of two fundamental components of the modifiable area unit problem (MAUP; Openshaw & Taylor, 1979), in which the shape and placement of the non-overlapping units used to extract map values influence analyses of those values (Dark & Bram, 2007; Marceau, 1999). In remote sensing, the image's pixels define the fundamental mapping unit, and mismatches between the pixels' dimensions and the characteristic shapes and scales of natural features impact subsequent analysis (Dark & Bram, 2007). However, if the sensor resolution is fine enough, such mismatches can be minimized–a natural feature's shape can be approximated by aggregating several square pixels–giving the analyst greater ability to minimize errors associated with this aspect of MAUP (Dark & Bram, 2007; Hay *et al.*, 2003).

Of course, high quality, fine-scaled maps such as SA-LC, which was carefully developed for South Africa, do not exist for many countries. Development of a new generation of Landsat/Sentinel-based (i.e., 30m resolution or finer) land-cover maps is underway, as exemplified by the new 30 m GLOBELAND30 map (Chen *et al.*, 2015). These maps will likely prove very useful for countries and regions lacking their own focused maps, as well as for cross-border analyses. To assess whether such maps are fit for the specific purpose, users should first conduct their own, thematically-focussed accuracy assessments to better understand error rates within their region of interest.

If high quality, high-resolution maps are not available, users interested in agricultural cover could select a new fusion map, such as GLC-Share, GeoWiki, or its derivatives (Fritz *et al.*, 2015; Waldner *et al.*, 2016). Although these maps are relatively coarse-scaled, the fusion methodology helps to greatly minimize error, as does the use of fractional cover values. Calibrating against inventory data may also reduce pixel level biases (e.g. Fig. 1b). However, users should be aware that this adjustment technique introduces confounding errors where census-reported statistics are unreliable, such as in some SSA countries (Carletto *et al.*, 2015, 2013). Alternatively, users could directly apply the fusion methodology (Fritz *et al.*, 2011b) to create their own improved maps.

If a particularly erroneous map is all that is available, and the user is chiefly interested in minimizing cell-wise error and less concerned about the spatial configuration of cover, then this may be achieved by aggregating maps expressing continuous values (e.g. fractional cover) to coarser scales. This approach must be undertaken with care, as its poses a number of complications related to the *scale problem*, the other half of MAUP (Openshaw & Taylor, 1979), which include progressive declines in variance with increasing scale (even if means remain constant; Dark & Bram, 2007), and reduced efficiency in estimating regression parameters from coarser map values (Avelino *et al.*, 2016). Nevertheless, aggregating continuous variables can reduce pixel errors without biasing some statistical properties (e.g. total cropland area remains constant using the aggregation approach demonstrated here; see Results: Cropland map errors), and has been shown to reduce other MAUP-related analytical problems (Avelino *et al.*, 2016). However, the user must ensure that the scale of aggregation is appropriate for the particular analysis.

Finally, users (or makers) of downstream products should rigorously ascertain how error propagates from the base land cover map into their derived maps (Verburg *et al.*, 2011). In some cases, the providers of "off-the-shelf" products report pixel-level uncertainty values (e.g. Ramankutty *et al.*, 2008), which may be sufficient. In other cases, downstream maps may lack quantified confidence intervals (e.g. Monfreda *et al.*, 2008). Although such maps may provide guidelines for appropriate usage[2], which our analyses help to further illustrate, users should undertake their own error propagation assessments. The most straightforward way may be to use a Monte Carlo approach that generates artificial datasets (e.g. Avelino *et al.*, 2016), using introduced errors drawn from reported or user-determined accuracy statistics, for both the base land cover and the downstream maps. For downstream products based on more complex models, users should examine how land cover map errors interact with particular model assumptions or structures, and alter these where necessary and possible to minimize confounding effects. Quantifying error propagation is important to understanding how map error may influence subsequent understanding of the phenomenon of interest.

Acknowledgments

This work was supported by the Princeton Environmental Institute Grand Challenges program, the NASA New Investigator Program (NNX15AC64G), and the National Science Foundation (EAR-1534544, SES-1360463, and BCS-1026776). We thank six anonymous reviewers for their constructive comments on earlier versions of the manuscript. Disclosure: S.F. is a director of GeoTerraImage, the owner of the reference map used in this study.

Supporting information

Additional information on supplemental methods and results can be found in the Supplementary Methods and Results document. The paper manuscript, code, and all non-proprietary data are available as part of an R package at https://github.com/agroimpacts/croplandbias.

References

Araújo M, New M (2007) Ensemble forecasting of species distributions. *Trends in Ecology & Evolution*, **22**, 42–47.

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

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.

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

Asseng S, Ewert F, Rosenzweig C, *et al.* (2013) Uncertainty in simulating wheat yields under climate change. *Nature Climate Change*, **advance online publication**.

Avelino AFT, Baylis K, Honey-Rosés J (2016) Goldilocks and the Raster Grid: Selecting Scale when Evaluating Conservation Programs. *PloS one*, **11**, e0167945.

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

Carletto C, Gourlay S, Winters P (2015) From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis. *Journal of African Economies*, **24**, 593–628.

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.

Dark SJ, Bram D (2007) The modifiable areal unit problem (MAUP) in physical geography. *Progress in Physical Geography*, **31**, 471–479.

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.

Dendoncker N, Schmit C, Rounsevell M (2008) Exploring spatial data uncertainties in landuse change scenarios. *International Journal of Geographical Information Science*, **22**, 1013–1030.

Drusch M, Del Bello U, Carlier S, *et al.* (2012) Sentinel-2: ESA's Optical High-Resolution Mission for GMES Operational Services. *Remote Sensing of Environment*, **120**, 25–36.

Estes LD, Beukes H, Bradley BA, *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**, 3762–3774.

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

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*.

Frey KE, Smith LC (2007) How well do we know northern land cover? Comparison of four global vegetation and wetland products with a new ground-truth database for West Siberia. *Global Biogeochemical Cycles*, **21**, GB1016.

Friedl MA, Sulla-Menashe D, Tan B, Schneider A, Ramankutty N, Sibley A, Huang X (2010) MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, **114**, 168–182.

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 (2008) Identifying and quantifying uncertainty and spatial disagreement in the comparison of Global Land Cover for different applications. *Global Change Biology*, **14**, 1057–1075.

Fritz S, See L, McCallum I, *et al.* (2011a) Highlighting continued uncertainty in global land cover maps for the user community. *Environmental Research Letters*, **6**, 044005.

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.* (2011b) 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.

Gasparri NI, Kuemmerle T, Meyfroidt P, le Polain de Waroux Y, Kreft H (2015) The emerging soybean production frontier in Southern Africa: Conservation challenges and the role of South-South telecouplings. *Conservation Letters*, pp. n/a–n/a.

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.

Giorgi F, Mearns LO (2002) Calculation of average, uncertainty range, and reliability of regional climate changes from AOGCM simulations via the "Reliability Ensemble Averaging" (REA) method. *Journal of Climate*, **15**, 1141–1158.

Goetz SJ, Baccini A, Laporte NT, *et al.* (2009) Mapping and monitoring carbon stocks with satellite observations: a comparison of methods. *Carbon Balance and Management*, **4**, 2.

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.

Hand E (2015) Startup liftoff. *Science*, **348**, 172–177.

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.

Hay GJ, Blaschke T, Marceau DJ, Bouchard A (2003) A comparison of three image-object methods for the multi-scale analysis of landscape structure. *ISPRS Journal of Photogrammetry and Remote Sensing*, **57**, 327–345.

Jain AK, Meiyappan P, Song Y, House JI (2013a) CO2 emissions from land-use change affected more by nitrogen cycle, than by the choice of land-cover data. *Global Change Biology*, **19**, 2893–2906.

Jain M, Mondal P, DeFries RS, Small C, Galford GL (2013b) Mapping cropping intensity of smallholder farms: A comparison of methods using multiple sensors. *Remote Sensing of Environment*, **134**, 210–223.

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.

Langford WT, Gergel SE, Dietterich TG, Cohen W (2006) Map Misclassification Can Cause Large Errors in Landscape Pattern Indices: Examples from Habitat Fragmentation. *Ecosystems*, **9**, 474–488.

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.

Latham J, Cumani R, Rosati I, Bloise M (2014) Global land cover share (GLC-SHARE) database beta-release version 1.0-2014. Tech. rep., FAO, Rome, Italy.

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.

Marceau DJ (1999) The scale issue in the social and natural sciences. *Canadian Journal of Remote Sensing*, **25**, 347–356.

Marceau DJ, Hay GJ (1999) Remote sensing contributions to the scale issue. *Canadian Journal of Remote Sensing*, **25**, 357–366.

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.

Moody A, Woodcock CE (1995) The influence of scale and the spatial characteristics of landscapes on land-cover mapping using remote sensing. *Landscape Ecology*, **10**, 363–379.

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

Nol L, Verburg PH, Heuvelink GBM, Molenaar K (2008) Effect of Land Cover Data on Nitrous Oxide Inventory in Fen Meadows. *Journal of Environmental Quality*, **37**, 1209–1219.

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.

Openshaw S, Taylor PJ (1979) A million or so correlation coefficients: three experiments on the modifiable areal unit problem. *Statistical applications in the spatial sciences*, **21**, 127–144.

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

Pax-Lenney M, Woodcock CE (1997) The effect of spatial resolution on the ability to monitor the status of agricultural lands. *Remote Sensing of Environment*, **61**, 210–220.

Phelps J, Carrasco LR, Webb EL, Koh LP, Pascual U (2013) Agricultural intensification escalates future conservation costs. *Proceedings of the National Academy of Sciences*, **110**, 7601–7606.

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.

Samberg LH, Gerber JS, Ramankutty N, Herrero M, West PC (2016) Subnational distribution of average farm size and smallholder contributions to global food production. *Environmental Research Letters*, **11**, 124010.

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

Schierhorn F, Müller D, Beringer T, Prishchepov AV, Kuemmerle T, Balmann A (2013) Post-Soviet cropland abandonment and carbon sequestration in European Russia, Ukraine, and Belarus. *Global Biogeochemical Cycles*, pp. n/a–n/a.

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*, **6**, 192–196.

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.

Statistics South Africa (2007) Building statistics, 2007. Tech. Rep. 50-11-01, Statistics South Africa, Pretoria.

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.

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.

Waldner F, Fritz S, Di Gregorio A, *et al.* (2016) A unified cropland layer at 250 m for global agriculture monitoring. *Data*, **1**, 3.

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.