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

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

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

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

Blah blah.

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

Abbreviations: GTI, GeoTerraImage

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

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

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

Not being able to full quantify the errors in landcover maps of course makes it difficult, if not impossible, to quantify their impact on downstream analyses. There has been some work examining how such error influences climate simulations [21], agricultural land use patterns [23], and carbon flux [22] and human population estimates [9], but these use either simulated landcover errors [21] or compare relevant differences in estimates between different satellite-derived landcover maps [9,22]. The exception is [23], who use a high quality, ground-collected reference map detailing farm land use parcels in central Belgium, but the number of sites and region were both fairly restricted, and the parcels were not spatially contiguous.

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

Data Sources.

Discussion

Blather.

More blather. Materials and Methods

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

Digital RCD Analysis.

Appendix: App 1

Appendix

This is an example of an appendix without a title.

ACKNOWLEDGMENTS. I thank everyone tearfully.

Reserved for Publication Footnotes

1. Lambin EF (1997) Modelling and monitoring land-cover change processes in tropical regions. Progress in Physical Geography 21(3):375–393.
2. Lark TJ, Salmon JM, Gibbs HK (2015) Cropland expansion outpaces agricultural and biofuel policies in the United States. Environmental Research Letters 10(4):044003.
3. Wright CK, Wimberly MC (2013) Recent land use change in the Western Corn Belt threatens grasslands and wetlands. Proceedings of the National Academy of Sciences 110(10):4134–4139.
4. Licker R et al. (2010) Mind the gap: how do climate and agricultural management explain the yield gap of croplands around the world? Global Ecology and Biogeography 19(6):769–782.
5. Asner GP et al. (2010) High-resolution forest carbon stocks and emissions in the Amazon. Proceedings of the National Academy of Sciences 107(38):16738–16742.
6. Gaveau DLA et al. (2014) Major atmospheric emissions from peat fires in Southeast Asia during non-drought years: evidence from the 2013 Sumatran fires. Scientific Reports 4.
7. Newbold T et al. (2015) Global effects of land use on local terrestrial biodiversity. Nature 520(7545):45–50.
8. Luoto M, Virkkala R, Heikkinen RK, Rainio K (2004) Predicting bird species richness using remote sensing in boreal agricultural-forest mosaics. Ecological Applications 14(6):1946–1962.
9. Linard C, Gilbert M, Tatem AJ (2010) Assessing the use of global land cover data for guiding large area population distribution modelling. GeoJournal 76(5):525–538.
10. See L et al. (2015) Improved global cropland data as an essential ingredient for food security. Global Food Security 4:37–45.
11. 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.
12. Debats S, Luo D, Estes L, Fuchs T, Caylor K (year?) A generalized computer vision approach to mapping agricultural fields in Sub-Saharan Africa. Remote Sensing of Environment.
13. Lobell DB (2013) The use of satellite data for crop yield gap analysis. Field Crops Research 143:56–64.
14. Estes L et al. (2015) DIYlandcover: Crowdsourcing the creation of systematic, accurate landcover maps. PeerJ PrePrints 3:e1266.
15. Estes LD et al. (2013) Projected climate impacts to South African maize and wheat production in 2055: a comparison of empirical and mechanistic modeling approaches. Global Change Biology 19(12):3762–3774.
16. Fritz S, See L, Rembold F (2010) Comparison of global and regional land cover maps with statistical information for the agricultural domain in Africa. International Journal of Remote Sensing 31(9):2237–2256.
17. Fritz S et al. (2011) Cropland for sub-Saharan Africa: A synergistic approach using five land cover data sets. Geophysical Research Letters 38:L04404.
18. Fritz S et al. (2015) Mapping global cropland and field size. Global Change Biology 21(5):1980–1992.
19. Verburg PH, Neumann K, Nol L (2011) Challenges in using land use and land cover data for global change studies. Global Change Biology 17(2):974–989.
20. Fritz S et al. (2012) Geo-Wiki: An online platform for improving global land cover. Environmental Modelling & Software 31:110–123.
21. Ge J et al. (2007) Impacts of land use/cover classification accuracy on regional climate simulations. Journal of Geophysical Research: Atmospheres 112(D5):D05107.
22. Quaife T et al. (2008) Impact of land cover uncertainties on estimates of biospheric carbon fluxes. Global Biogeochemical Cycles 22(4):GB4016.
23. Schmit C, Rounsevell MDA, La Jeunesse I (2006) The limitations of spatial land use data in environmental analysis. Environmental Science & Policy 9(2):174–188.