Four-century history of land transformation by humans in the United States: 1630-2020

ESSD-2022-135

The paper takes on the very substantial challenge of recreating historical land use for the United States. The authors are correct that outside of coarse-level reconstructions such as HYDE, that there is nothing of higher resolution that goes back to pre-settlement by European colonists. The authors pitch this methodology and dataset as the solution to that data gap.

I don't think the authors are completely successful in making that argument. After stating the need for higher-resolution data, and presenting a methodology that should allow for some higher-level of spatial detail than something like HYDE, the authors inexplicably leave out any kind of spatial analysis of how well their model result performs. It's difficult when spatially explicit data on land use is hard to come by (and hence the need for this work!), but some analysis of the full-resolution data compared to a data set such as NLCD (available 2001 to 2019) could have helped establish confidence in the model to capture spatial patterns well. Even a comparison to historical county-level data would help. However, the only "spatial" analysis of the data are some very coarse regional assessments that don't provide a reader much of a feel for the model's capability for generating realistic, high-resolution spatial patterns.

Validation overall is a weak point of the paper. On the one hand, I understand the difficulties in trying to "validate" results such as this, when consistent reference data is absent or scattered. However, it's not acceptable from a modeling perspective to use HYDE, NLCD, and other data to parameterize the model, and also use those data in what's labeled as "validation" of results. Given the lack of spatially explicit reference data, I have no issue with the authors doing "consistency checks" with other data sets, including missed opportunities such as county-level ag census data that could have been used to provide a better feel for the spatial patterns produced by the model. But don't try to sell it as a real "validation" of model results.

Overall, the authors continually note the "Uncertainties" associated with coarser data such as HYDE, but fail to conclusively demonstrate their results are superior in terms of those uncertainties.

A major need for the paper is recognition of the differences between source datasets, and the uncertainties it introduces into the modeling. For example, many parts of the model are parameterized by using multiple datasets that have inherent differences. Pasture, for example, is quantified at a state level by NRI, and then HYDE for older dates. Given how variable definitions are for "pasture" in the first place, it's asking for trouble to mix and match datasets such as that. The authors do seem to make some attempts to harmonize differences in datasets, but the methodology isn't well-defined enough to let me know if that's really being done. In short, historical land use reconstruction is difficult because 5 different datasets may give you 5 different answers for how much "forest", "urban", "cropland", or "pasture" is actually there for a given date!

The spatial allocation of "change" on the landscape is also quite simplistic, based solely on probability surfaces with no stochasticity. The authors do attempt to mitigate the 'static probability surface' problem present for past applications such as CLUE and FORE-SCE by some simple weighting with population. But the actual allocation is fully dependent on the probability surfaces at the end of the day,

and as a result, as you go back in time, you tend to see classes such as cropland and pasture become concentrated in the high-probability locations, with less fragmentation that's there in later dates.

Finally, note I did download and look at some of the output results. For brevity I'll keep my comments here to the "Boolean" land cover. Note that while results look reasonable at broad scales, the approach of parameterizing state-by-state does seem to cause some issues as you go back in time, as does the spatial allocation methodology. Going back in time reveals a number of obvious state and even what appear to be county boundaries, hard obvious lines where land use clearly differs on either side of a political boundary. Given the complete reliance on probability surfaces alone for the spatial allocation of change, land use looks more concentrated on the landscape for some classes going back in time. For example, on the 1850 map, cropland is concentrated in very large contiguous chunks in many areas, and very sharp and obvious political boundaries are present.

Individual comments follow.

Specific Comments

- Lines 28 – Would add one word…"In particular, *managing* agriculture and forest-related activities…"
- Line 34 – Would add words…"…arrival of Europeans, indigenous communities practiced agriculture and crop planting in the…"
- Line 36 – Would change "mainly occurred" to "initially occurred" to indicate these activities first started here, but expanded elsewhere later (as noted by the next sentences)
- Line 37 – "Driven" not "driving".
- Page 2 – Overall the paragraph at the top of page 2 could use some work. It's a rather disjointed history of US land change. For one it doesn't really talk about land change west of the Mississippi River, it's focused solely on Eastern US change. The organization is also a bit odd and disjointed. The sentence on line 44, for example, seems like a very abrupt and odd ending to the final statement as to why a long-term land use dataset is needed. Perhaps a better organization by period (colonial, 19th century, 20th century), with a description of what occurred in each century? And perhaps a modification of the last sentence, adding "While general trends in historical US landscape change are known, we still lack a long-term dataset…"
- Line 55-56 – Be careful about highlighting "uncertainties" in datasets such as HYDE, as your historical landscape construction will also have substantial uncertainties. Your workflow itself uses HYDE data. There's limited spatially explicit data available from which to base a model-based landscape reconstruction, and many of the datasets you're using were also used by HYDE.
- Lines 87-88 – I wouldn't call the use of these other datasets "validation". It's a consistency check, not a validation, as these data sets too have uncertainties, and some are modeled just as you're modeling.
- Line 90-91 – How was resampling done to get to 1-km grid cells? Is it fractional LULC within a given 1-km cell for datasets with native resolution <1 km?
- Section 2.2.1 – This is an extremely simplistic methodology for calculating urban land area. To start, it's all based on one current dataset, NLCD. How was NLCD used? First of all, NLCD tends to underestimate low-density residential lands, which can bias your results. Secondly, NLCD "urban" classes also include extensive representation of road networks, which if counted as "urban", greatly overestimates urban land. For a rural state, for example, NLCD classes not only

major roads, but every small section road has a 1-pixel-wide "urban" class representing it. Unless measures were taken to account for NLCD's underrepresentation of low-density residential lands, and to account for all the "urban" pixels that are really roads, it biases the results.

The other problem is the very simplistic method for calculating land area. You're assuming the relationship between urban land per capita and total population is constant through time. Clearly it's not. Without accounting for that changing relationship, urban estimates can easily be biased.

- Section 2.2.2 – You're using (at least) three different data sources to help establish cropland area. For historical land use, estimates vary widely, dependent upon methodology, data source, thematic definitions of a land use, etc. As a result, when switching from USDA-based data, for example, after 1889, and using HYDE before 1889, you'd expect an obvious break in estimated "cropland" amounts. How were those inconsistencies among historical land use datasets harmonized?
- Paragraph starting on line 120 – You're assuming the relationship between harvested area and planted area from 1978 to 2017 is consistent decades and centuries before those data…a very dangerous assumption.
- Line 123 – Yet another dataset, Borchers et al. 2014, was used to establish double-cropping at a regional level. Again, consistency among most of these datasets isn't great.
- Line 124-125 – Another basic assumption that likely isn't true through time.
- Line 125-126 – Because the 1879 number was different you assumed it was incorrect? But data >1889 were "correct"? Was the reconstructed cropland area in 1879 substantially lower or higher than 1889?
- Line 130-131 – Another basic assumption that likely doesn't hold region to region.
- Line 132-133 – Again…how did you account for differences in the HYDE data, and the (mostly) USDA-based data after 1889? Is there an obvious break in cropland amount pre- and post-1889?
- Line 135-136 – One of the greatest difficulties in historical landscape reconstruction is the definition of "pasture", vs. "grassland", vs. "hay" vs. "rangeland", etc. There is no one definition that's universally accepted. Your definition here states Pasture includes areas "for the production of seed or hay crops". Many definitions of "cropland" include alfalfa, hay, and other crops in "planted area" or "cultivated crop" area. For "Pasture", you're introducing yet another completely new dataset to establish pasture area, NRI. Are the definitions of "pasture" for NRI the same for NLCD, HYDE, and the US Census of Ag?
- Line 139 – Note Wasianen and Bliss took great pains to harmonize those definitional differences across their harmonized dataset.
- Section 2.2.3 – Again…it's extremely simplistic to assume things such as "pasture per capita" and that that ratio is consistent over time, and space.
- Section 2.2.4 – Definitions of what is "Forest" vary greatly among data sets. You're introducing yet another data set in FIA that may have a definition of "forest" that differs from HYDE or from NLCD. How closely does the FATD data match with HYDE estimates, for example?
- Lines 156-157 – Yeah you lost me here with what you're trying to do, needs a better explanation.

- Section 2.2.5 – See main comments above related to how you balanced the four LULC classes.
- Line 195 – What was used to establish the "land use change boundary"? That is, what was the source of "settled area" data"?

  I don't mind the use of something like this to constrain the allocation of change, but do wonder about full-resolution results. Are there any hard border issues obvious in the data when change occurs at the edge of those defined boundary layers? Overall with the boundary and effect of population density, I appreciate you trying something other than assuming a static probability surface through time.
- Section 2.3.2 – There needs to be more explanation here. You've basically summarized the entire actual allocation to the pixel level in one sentence. I certainly get that higher probability areas will likely have a higher proportion of a given LULC class, but it's all deterministic and it's all based solely on the probability surface? There's no stochasticity? With such a sparse description of methodology, it's also hard to see how this simple description of the methodology ends up with the aggregate totals from the allocation stage matching the quantitative estimates you established for each of the LULC classes.
- As noted in the main comments, I have other concerns about the allocation strategy.
- Section 2.4 – Comparison to other LULC datasets isn't a validation, it's a consistency check. That's particularly true when every dataset has it's own production methodologies, data sources, and thematic definitions, all of which makes even direct comparison problematic.

  Beyond that, you're comparing your results to some of the same datasets from which you parameterized your modeling, as noted in the overarching comments. Also note there aren't any details as to what methodologies you're actually using for "validation" in this very short, one-paragraph section.
- Section 3.1 – As noted previously, this isn't very useful for inferring confidence in your results, when you're using the same datasets to parameterize the model as you are to "validate" model results.
- Line 273 (and throughout the results section) – If you're going to refer to a specific driving force of change, and, for example, point to a specific policy, you should name the policy and reference it (Immigration and Naturalization Act of 1965). While it certainly did change the nature of immigration to the country, you do give it too much focus as "the" causes of urban land increases after 1965. There's a lot more at play there than immigration policy.
- Lines 276-277 – You state "cropland area did not change significantly" from 1930 to present day. First, it's always problematic to use the term "significantly" in a journal paper, given the scientific meaning of the word. Secondly, I would argue there were "substantial" trends in agriculture after 1930, including some of those you mention (e.g., biofuel impacts).
- Figure 7 – On a national-scale map figure, it's difficult to see patterns of the individual land use transitions. Perhaps it would be augmented by a complementary confusion matrix of changes or some other tabular data approach that allows you to see (and easily quantify) transition types.
- Section 3.4 – This isn't the most effective section to me. As noted in the main comments, a major premise of the paper was to provide a "high resolution" historical landscape reconstruction for the US. Much of the "regional" information here is also discussed in the

overall results above. I'd have much rather seen some real examples (and preferably validation) of landscape pattern at finer scales, given the focus on higher resolution with this paper.

- Lines 338-339 – Agreed about the "per capita" approach.
- Lines 339-341 – This doesn't serve as any kind of adequate validation or even consistency check between datasets. Showing a national-scale map and stating the patterns are "consistent" isn't valuable, and is very subjective at that scale.
- Line 357-358 – Exactly why it's not very valuable to compare your model results to HYDE…those data were used to help establish the model parameters themselves.
- Line 362 – Your product has higher spatial resolution than something like HYDE, but there's no quantitative analysis of that spatial pattern that proves the superior value of that higher native resolution.
- Figure 9 – It is difficult to compare all of these datasets given the definitional differences between them, particularly for pasture and cropland.
- Lines 382-383 – I'm not sure it's more "reliable", as sample-based, inventory approaches have flaws, just as satellite-based approaches have flaws. The bigger concern to me are the definitional differences, not the methodological differences.
- Section 4.2 doesn't add a lot to the paper for me, particularly since you've already tried to explain some driving forces in the previous paragraphs of the paper. I'd much rather have the drivers woven into the story of what's happening in your results, than as a separate section.
- Lines 435-436 – I think reconstruction of historical land use is limited more by reliable, consistent historical data than methodology. Machine learning methods aren't going to be that valuable for historical reconstruction given the paucity and inconsistency of historical data for training.
- Section 4.3 – Somewhere in here you absolutely need to highlight the difficulties with trying to harmonize data sets with different definitions, data sources, and methodologies.