RDCEP WORKING PAPER SERIES

The Center for Robust Decision Making
on Climate and Energy Policy

Synthesis of a Complete Land Use/Land Cover Dataset for the Conterminous United States

Neil Best, Joshua Elliott and Ian Foster

Working Paper No.12-08

May, 2012

Synthesis of a Complete Land Use/Land Cover Dataset for the Conterminous United States

Neil Best* Joshua Elliott† and Ian Foster†

May 4, 2012

Abstract

We present a new land cover dataset for the conterminous USA called "$PEEL_0$." It is designed for climate change impact analysis based on land use change at 5 arc-minute resolution. The procedure described herein is adaptable for generating similar datasets for other regions and requirements. $PEEL_0$ is derived from existing data products – the MODIS Land Cover Type (MLCT) and National Land Cover Database (NLCD) – but its design overcomes certain limitations that hinder their use for climate change impact analysis. First, while other products that focus on agriculture neglect non-agricultural land use/land cover (LULC) categories, $PEEL_0$ contains eight distinct LULC classes in addition to a "crop" class. $PEEL_0$ features subcell area fractions for each class, increasing its depth of information over traditional single-category LULC maps. Second, $PEEL_0$ offers improved accuracy in characterizing cultivated lands, important for quantifying agriculturalactivity. $PEEL_0$ provides a more accurate spatial distribution of cultivated lands over MLCT as compared to reference datasets and improved totals for cultivated land relative to USDA Major Land Uses census data. We present here landcover data for 2001 plus $PEEL_0$s synthesis methodology, which combines information from multiple sources by establishing a common classification scheme at lower spatial resolution. $PEEL_0$ was developed as an initialization dataset for a partial-equilibrium economic land use model (PEEL) that simulates land use/land cover change in response to exogenous agricultural prices and climate change scenarios. We anticipate that similar landcover data products will be of use to other modeling efforts worldwide.

Keywords: land-use/land-cover, remote sensing, MODIS, NLCD, high-performance data processing, agriculture

JEL Classifications: Q15, C89

*Computation Institute, University of Chicago and Argonne National Laboratory; Department of Geography & Environmental Studies, Northeastern Illinois University; email: nbest@ci.uchicago.edu

†Computation Institute, University of Chicago and Argonne National Laboratory

Contents

1 Introduction

Understanding the factors that drive land use and land cover (LULC) change and developing better methods for projecting future change are important problems in the context of both climate change and agricultural production. The 3rd Assessment Report of the UN Intergovernmental Panel on Climate Change (IPCC) noted that "the emissions scenarios considered in future climate change studies need to integrate high resolution representations of land use change" and that increased coupling among the various relevant components, such as mitigation and adaptation responses to climate change and climate response to land use, should be included in a consistent framework for integrated assessment [*Jones et al.*, 2001]. For these and other reasons, models that integrate physical and socioeconomic factors at resolutions that resolve important variations in the spatial and temporal patterns of land use decisions, environmental conditions, and climate impacts are needed. Such models assist policy and decision makers in evaluating the life cycle impacts of programs intended to, for example, subsidize the production of biofuels on industrial scales; encourage the adoption of sustainable farming practices that will increase carbon uptake in the biosphere; or motivate research and development to both increase crop yields and decrease fertilizer use while anticipating shifts in the distributions of particular crops in response to climate change. Relevant projects include IMAGE [*van Vuuren et al.*, 2006, 2007], AIM [*Matsuoka et al.*, 1995], and, at the University of Chicago, the Partial Equilibrium Economic Land use (PEEL) model.

The objective of the PEEL model is to stochastically simulate global land use change driven by exogenous commodity price trajectories and climate change scenarios. These simulations need an initialization dataset that characterizes the initial state of land uses and natural covers. This paper describes the formulation of that initialization dataset which we have named $PEEL_0$ to signify

its role as the "zeroeth" time step. In order for the PEEL model to be computationally feasible we have chosen a resolution of 5 arc-minutes ($5'$) coupled with a fractional representation of per-cell cover types allowing model grid cells to have a mixture of cover types without further spatial specificity. This is also known as sub-pixel or mixed pixel analysis. Conveniently this resolution is common among the available data sets that characterize global agricultural activity, which we describe below, but those data sets are silent about non-agricultural uses and natural covers in general. At the opposite extreme, global LULC data sets have higher resolutions and use a discrete categorical representation, which is to say that any pixel is assigned one and only one LULC classification and makes no allowance for LULC mixture other than the possibility of mixed/hybrid classes as separate categories. We demonstrate here a process for aggregating high-resolution LULC datasets that incorporates principled reclassification of disparate LULC categorization systems and decomposition of hybrid LULC classes in order to create a reliable input dataset for the PEEL model. This PEEL_0 dataset must express agricultural land use in terms of area that is directly utilized for production and also fill in the non-agricultural classes to give a complete representation of the landscape for each $5'$ grid cell.

Initially we hypothesized that the PEEL model could be initialized with a simple reclassification, decomposition, and aggregation of the MODIS Land Cover Type v005 (MLCT) dataset [*Friedl et al.*, 2010, *LP DAAC*, 2008], described in greater detail below. In addition to offering global coverage another advantage of adopting MLCT for initialization would be that it provides an annual snapshot of global LULC beginning in 2001. It is derived from raw remote sensing data from a stable platform processed through a single classification algorithm thereby providing an internally consistent initialization sequence. However we soon realized that MLCT contains certain systematic biases against small-scale features that have characteristic dimensions smaller than its 15 arc-second ($15''$; 500 m, nominal) resolution, such as rural transportation networks and small, often linear, water or wetland features. Through comparison with other census-based datasets for agriculture, namely the Agricultural Lands in the Year 2000 (Agland2000) [*Ramankutty et al.*, 2008] and the Major Land Uses (MLU) datasets [*Lubowski et al.*, 2006], we found an overestimate of production area because of inclusion of other use and cover types in MLCT's "cropland" class, which represents the overall function of a landscape unit at MLCT resolution rather than a direct expression of cultivated acreage. Therefore we formulated and approach for correcting these biases using other available data. For the initial realization of PEEL_0 we have had to restrict the demonstration of this procedure to the conterminous United States (cUSA) due to the availability of an appropriate dataset from which to derive our correction offsets for the water, wetland, and urban/developed classes, specifically the 2001 National Land Cover Database (NLCD) [*Homer et al.*, 2004, 2007]. The application of these offsets not only adjusts agricultural areas, the "crop" class, but also forest, open, shrub, and barren classes to varying degrees.

The rest of the paper is as follows. Section 2 describes the datasets that we use to construct and evaluate PEEL_0. Section 3 describes the algorithm that we use to construct PEEL_0 by preparing the MLCT and NLCD datasets, selectively incorporating cover fractions for particular classes from NLCD, and decomposing the MLCT mosaic class. Section 4 describes the process by which we evaluate the PEEL_0 product and presents the results of this analysis. In Section 5 we conclude with a discussion of the merits of this endeavor and of future avenues of research.

2 Land Use and Land Cover Datasets

A number of datasets describe the distribution and intensity of global agricultural activity in various ways. Some such as the Global Irrigated Areas Map [*Thenkabail et al.*, 2008] and the Global Map

of Rainfed Crop Areas [*Biradar et al.*, 2009] are the product of applying classification techniques to large collections of remote sensing and GIS data. Others such as Agland2000, Area and Yields of 175 Crops (175Crops2000) [*Monfreda et al.*, 2008], and Spatial Production Allocation Model (SPAM) [*You et al.*, 2006] are further informed by agricultural production data published at national and subnational levels and disaggregated to grid cells within those boundaries according to an optimization method *You and Wood* [2006]. The latter datasets can complement general comprehensive LULC datasets by offering additional information on how to differentiate areas of cropland according to cultivars and farming practices such as crop rotation, multiple cropping, and irrigation but are themselves silent on non-agricultural use and cover types. We use four datasets in this work: MLCT and NLCD to construct PEEL_0 plus Agland2000 and MLU to evaluate PEEL_0's quality. We describe these datasets here.

The **2001 MODIS Land Cover Type v005** dataset is a 500 m ($\sim 15''$) resolution land cover dataset. We base our method on MLCT because of its global coverage, annual time series, and free availability. We plan to extend PEEL_0 in both time and space, so MLCT is the clear choice as a foundation dataset. MLCT provides three data values for each pixel: a primary classification, a percentage measure of classification confidence, and a secondary classification. Among the choices of classification systems MLCT offers we chose to work with the International Geosphere-Biosphere Programme (IGBP) scheme [*Friedl*, 2002] and adopted a simplified variant thereof for PEEL_0, described below in Section 3.1.1. The confidence level is intended as a measure of the likelihood of classification error. As we describe in Section 3.1.2, we reinterpret this information as an estimate of the fraction of subpixel area covered by the primary class.

The **National Land Cover Database 2001** provides a higher-resolution (30 m, $\sim 1.25''$) snapshot of LULC across the cUSA study area, plus Alaska, Hawaii and Puerto Rico, circa 2001. Because NLCD's classification was informed by ancillary datasets such as population density, buffered roads, and the National Wetland Inventory [*Homer et al.*, 2004], we expect that it will give better estimates of aggregate area for detailed features like rural transportation networks and small stream and wetland features. Although it is unclear from the work of*Homer et al.* [2004] what ancillary data was applied in what constituent mapping zones of NLCD, we accept its representation of these fine details to be the best available. As we describe in Section 3.3, we apply differences between NLCD and MLCT data as a correction, in order to compensate for MLCT bias against these finely detailed structures that results from its resolution.

The **Agricultural Lands in the Year 2000** dataset is a $5'$-resolution dataset that merges satellite-derived LULC classifications with census data of arable land, permanent crops, and pasture compiled at national or subnational levels according to availability of such data at or near the turn of the century. It uses two LULC classification datasets derived from remote sensing data as inputs, an older version of MLCT (known as BU-MODIS) and GLC2000. The "pasture" class in Agland2000 likely has much in common with the "open" class from MLCT, but we do not employ that data in this analysis. Because of its basis in census data, we use the cropland component of Agland2000 as a type of observational product for evaluating our incremental adjustments to the maps we derive from MLCT in Section 4. Figures 17 and 18 show the distribution of both cropland and pasture areas for the detail and full study areas, respectively. The classification algorithm used to construct the 2001 MLCT uses Agland2000 cropland data as a prior probability in its classification algorithm but because it is an influence rather than a constraint we consider this circularity to be weak enough to allow us to use Agland2000 as an independent reference data set.

The **USDA Major Land Uses** (MLU) dataset [*Lubowski et al.*, 2006] is a consistent census-based time-series record of US agricultural land uses produced by the USDA Economic Research Service going back to 1945. MLU contains data by state and distinguishes land use among six broad categories: cropland; grassland, pasture, and range; forest-use land; special-use land; urban

land; and other uses. Additional important subcategories for relating these census-based land-use categories to satellite based land-cover categories include cropland pasture, rural transportation areas; and farmsteads, farm roads, and lanes. For example, we include parks and other protected forest areas with the forest class; we distinguish cropland pasture from both cropland and pasture land for better comparison with NLCD statistics; we combine rural transportation networks and other developed areas, farmsteads, and other developed uses with the urban cover class; and we include miscellaneous lands such as marches, swamps, deserts, and lands designated for defense or other special purposes as other land cover. We note that MLU does not include coastal or inland water bodies and so has a lower overall count of land area compared with satellite datasets.

PEEL$_0$ will be the result of combining information from MLCT and NLCD. The current realization of PEEL$_0$ will populate grid cells within the cUSA study area that intersect with the cUSA polygons in version 1 of the Global Administrative Areas vector dataset [*Hijmans et al.*, 2009]. Each grid cell will have a floating point value for the PEEL LULC classes: water, forest, shrub, open, wetland, crop, urban, and barren. Each class will have its own map in a stack where the stack has the constraint that each cell's values must sum to unity. This can be represented electronically as a table where each row is a grid cell and each column is a class, as a multiband raster in any appropriate geospatial format such as GeoTIFF, or as a NetCDF file that uses the LULC classes as variables. Our implementation of the procedure produces all of the above outputs for the convenience of potential users. They are available for download at `http://hdl.handle.net/10779/3ab922c0c7827ea2a0b5d6557fb8fa09`.

3 Construction of PEEL$_0$

Our multistep process employs four distinct procedures for transforming the input data sets to produce PEEL$_0$.

Reclassification: Because we deal with LULC classifications based on different sets of class definitions, we seek to find a common basis on which to compare their contents. Specifically, we define a simplified set of LULC classes and map the classes in the original datasets to them. This simplified classification is suitable for our economic LULC forecast models because subtle distinctions between ecological roles of different types of forests, for example, are not relevant.

Aggregation: We use the term aggregation to denote moving to a lower resolution raster representation while preserving as much information from the aggregated pixels as possible. Each class is assigned a fractional value for each $5'$ cell based on the relative proportion of that classes presence among the higher-resolution classified pixels of the original dataset. This preserves more information than does a naive aggregation, such as accepting the class that occurs most frequently (i.e., the mode) as the representative class.

Correction: After aggregation of MLCT and decomposition of the mosaic class we selectively incorporate aspects of the higher-resolution NLCD. NLCD has an advantage in resolving particular classes, namely water, wetland, and urban, by virtue of resolution and /or ancillary data used as prior probabilities in its classification algorithm. In this way, we adjust the landscape composition indicated by the foundation dataset of this method, MLCT. These corrections are motivated primarily by a recognition of biases against finely detailed features that get washed out by MLCT's coarser resolution.

Decomposition: The need for decomposition arises when a class definition is a hybrid of a fundamental types in the final desired classification. MLCT features a "cropland / natural vegetation mosaic" class that we unpack in order to assign a portion of its area to agricultural production and the rest to natural covers. Decomposition, in contrast to disaggregation or downscaling, holds the spatial resolution constant and makes no attempt to impart additional spatial precision to subcell distributions of LULC.

Our general algorithmic approach can be summarized as follows:

1. Prepare MLCT data.
   (a) Reproject to geographic coordinates and mask cUSA study area.
   (b) Reclassify to PEEL_0 classification.
   (c) Calculate per-pixel, per-class areas at native resolution.
   (d) Aggregate to the $5'$ grid.
   (e) Decompose the mosaic fraction into crop and natural cover components to give total MLCT cropland. This is PEEL_0^a.
2. Prepare NLCD data for use as correction layer.
   (a) Reproject to geographic coordinates and mask cUSA study area.
   (b) Reclassify to PEEL_0 classification
   (c) Calculate per-pixel, per-class areas at native resolution.
   (d) Aggregate to the $5'$ grid.
3. Combine datasets to produce PEEL_0.
   (a) Revert mosaic decomposition that produced PEEL_0^a.
   (b) Selectively incorporate into PEEL_0^a cover fractions for water, wetland, and urban classes from NLCD.
   (c) Decompose mosaic fraction into crop and natural cover components. This is PEEL_0.

3.1 Preparing the MODIS Land Cover Type

The synthesis of PEEL_0 begins with the global MLCT product. We systematically incorporate the full depth of information offered by MLCT. Rather than interpret the secondary classification as the next most likely possibility, we accept the triplet of primary class, secondary class and confidence level as an expression of the subpixel composition of that area. Aggregation of MLCT from $15''$ to $5'$ blurs the spatial precision implied by this formula by treating the local $20 \times 20 \times 3$ array as a probabilistic expression of each $5'$ cell's cover composition. We show in Section 3.1 that this approach, given a principled assumption about the relationship between confidence level and the allocation of subpixel area among the detected classes, improves the estimates of acreages in aggregate as well as their spatial distributions, particularly for cropland.

MLCT comprises a set of tiles in a global equal-area sinusoidal projection. To prepare this dataset for use in this study, we first patched those tiles together and reprojected the resulting mosaic to geographic coordinates. We then extracted the conterminous US study area using the subset of cells proscribed for PEEL_0 above in Section 2. This area includes the water bodies on the American side of the international border across the Great Lakes but not oceanic waters beyond

the coastal grid cells that intersect with any land mass. To illustrate the process of converting these datasets from their original representation, we also often include maps of an area of south-eastern Michigan to show greater detail through each step of the process. We chose this region for its diversity of land covers and uses, its relative diversity of agricultural commodities across its significant cropland area, the significant presence of the MLCT/IGBP mosaic class to illustrate our method for its decomposition, and our familiarity with the region. Where space allows, we also present limited maps of variables over the conterminous US. More detailed maps will be provided online upon publication.

3.1.1 Reclassification of MLCT to $PEEL_0$ classes

Table 1 shows the mapping of the IGBP classes used in MLCT to our $PEEL_0$ classification. Because our primary interest is in agriculture, we collapse the five forest categories into a single class. We assign woody savannas and savannas to the shrub and open classes, respectively; these assignments are supported by the IGBP class definitions because of the overlap in the forest canopy cover for those classes. These assignments make sense in the context of LULC modeling because the ecological roles, potential uses, and conversion costs of the two savanna types are dissimilar. We combine areas of "permanent snow and ice" with "barren or sparsely vegetated" areas, which includes deserts, to form the $PEEL_0$ "barren" class, based on their shared characteristics of low population density and low intensity of economic activity.

	MLCT/IGBP	$PEEL_0$
0	water	water
1	evergreen needleleaf forest	forest
2	deciduous needleleaf forest	
3	evergreen broadleaf forest	
4	deciduous broadleaf forest	
5	mixed forests	
6	closed shrublands	shrub
7	open shrublands	
8	woody savannas	
9	savannas	open
10	grasslands	
11	permanent wetlands	wetland
12	croplands	crop
13	urban	urban
14	cropland / natural vegetation mosaics	mosaic
15	permanent snow and ice	barren
16	barren or sparsely vegetated	

Table 1: Reclassification of MLCT/IGBP to $PEEL_0$ necessary for compatibility across input dataset classification systems. Adapted from *Friedl* [2002].

Figure 1 shows the result of reclassifying the MLCT data for our detailed study area. This area is dominated by the crop class in the north and the mosaic class to the south, with scattered forests and pockets of development throughout. The urban complex of Port Huron, Michigan and Sarnia, Ontario is visible in the southeast corner.

Areas in the northern and central sections of the map that were classified as crop in the primary layer have null values in the secondary class, shown as white cells. These cells coincide with values

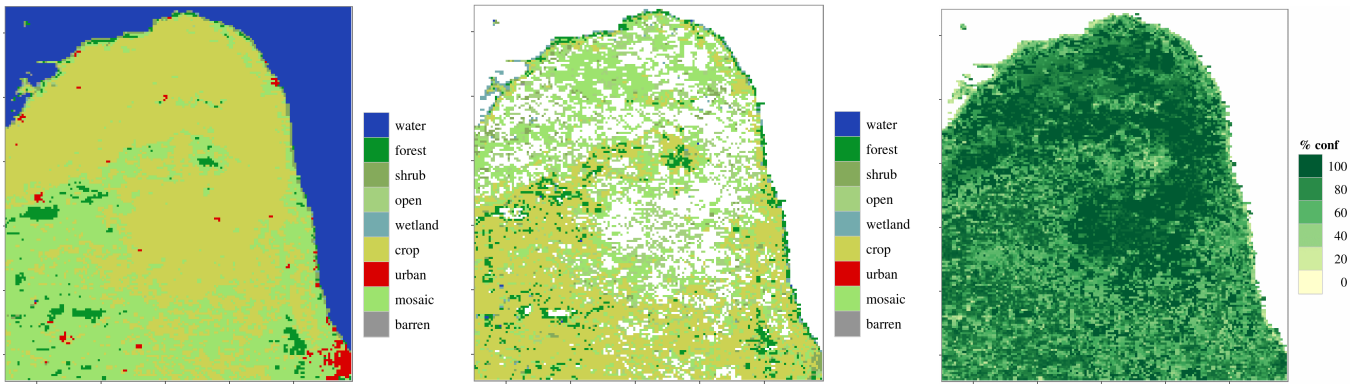

Figure 1: Detailed study area in southeastern Michigan showing MLCT primary reclassified cover (left), secondary reclassified cover (middle), and primary cover classification confidence(right) at native 500 m ($\sim 15''$) resolution.

of 100% confidence in the third layer of the data, so their areas are assigned entirely to the primary class. The high prevalence of 100% confidence cells for the crop class leads us to expect that MLCT will overestimate cropland, since any such large areas of cultivated land are certainly interspersed with homesteads, fence lines, small wood lots, roads, and other cultural features. For example, in areas such as this that were made available for settlement in the 19th century according to the Public Land Survey System (PLSS) we expect to find a more-or-less regular grid of rural roads at one-mile intervals, which is readily visible in maps and aerial photography of the region. Arguably, the IGBP cropland class is intended to indicate an overall function of the landscape rather than to measure areas directly employed in production. In that sense, we can say that the PEEL_0 cropland has a different definition because we seek to reconcile cropland areas with agricultural census data.

The reclassification step can assign primary and secondary classes to the same category. When a pixel indicates the forest class for both its primary and secondary classifications, it simply reflects a distinction between subtypes of forest in the original data, for example evergreen and deciduous. We note that crop and mosaic classes often appear in pairs (primary crop, secondary mosaic and vice versa). This coupling is not surprising given that mosaic areas comprise 40–60% cropland by definition so their exemplars must necessarily be near one another in the classification space. We explore this dynamic further in Section 3.4.

Figure 14 shows the primary classification, secondary layer, and confidence level for the entire cUSA study area. For improved visualization of the relative distributions of particular classes, we also provide, in Figure 15, facet maps for the individual classes. Familiar generalities of cUSA geography are apparent, such as the prevalence of forests in the east and northwest, cropland in the midwest, shrub lands in the southwest, and open lands across the west. The mosaic class is concentrated in the eastern portion of the study area; we attribute this phenomenon to greater population density, topography, and historical patterns of settlement resulting in characteristically smaller parcels and a greater degree of mixing among agricultural uses and natural covers.

3.1.2 Aggregation of MLCT to PEEL_0 resolution

MLCT has a nominal resolution of roughly 500 m that equates to $15''$ at the equator, a conveniently even factor-of-20 division of the $5'$ grid to which we wish to aggregate. Recall that for each pixel, MLCT provides a primary classification, a percentage confidence, and a secondary classification. Our aggregation strategy aims to extract as much information as possible from the ($20 \times 20 = 400$

MLCT pixels) × (3 values) = 1200 data values that MLCT provides for each PEEL_0 cell. MLCT's secondary cover type was originally intended to express the most likely alternative to the primary type [*Friedl et al.*, 2010], with the confidence level providing an indication of per-pixel classification error. We propose an alternative interpretation where the subpixels areas for the primary and secondary cover types in pixel x are given by:

$$A_p(x) = A_{min} + (1 - A_{min})c(x) \quad (1)$$
$$A_s(x) = 1 - A_p(x) \quad (2)$$

where $c(x)$ is the confidence level of the primary classification on $(0, 1]$ and A_{min} represents the minimum area fraction to be assigned to the primary class given $c(x)$=0. A_p and A_s are the fractional areas assigned to the primary and secondary classes, respectively. A given class must constitute at least 50% of the pixel area in order to be considered primary; therefore, settingA_{min}=0.5 affords maximum consideration to the secondary class in this scheme. Simplifying the equations by substituting this value gives:

$$A_p(x) = \frac{1 + c(x)}{2} \quad (3)$$
$$A_s(x) = 1 - A_p(x) = \frac{1 - c(x)}{2} \quad (4)$$

Instances of c<0.20 are rare, as shown by Figure 2 for a particular subset of MLCT pixels (see Section 3.4), so generally the primary class will be assigned more than 60% of the MLCT pixel area, consistent with MLCT's definition of "primary class," which is that it covers no less than 60% of a given pixel x [*Friedl*, 2002]. These definitions assume that the relationship between classification confidence, and the subpixel fraction of the primary class is a linear, monotonically increasing function. Other monotonic functions could be used, but the differences would be second-order refinements to this formulation.

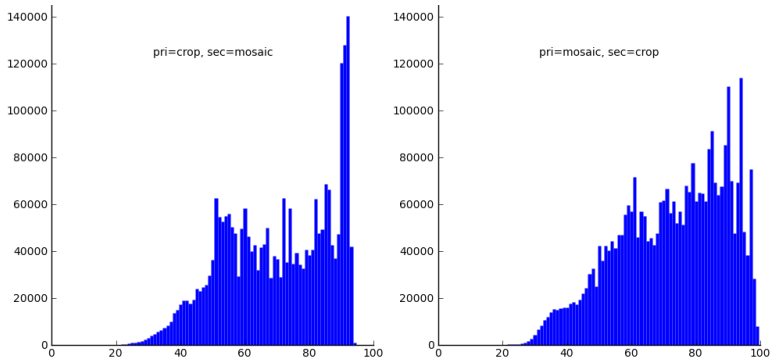

Figure 2: Histograms of the confidence measure for all cells in the cUSA classified as primary type crop, secondary type mosaic (left) and as primary type mosaic, secondary type crop (right).

In the analysis that follows, we compare the product of these assumptions with the case of A_{min}=1.0, which gives zero consideration to the secondary class and is therefore indifferent to the confidence level. In the interest of brevity, we do not consider here intermediate parameterizations in which the secondary class is used but given less than maximum consideration. Still, an advantage of our algorithm is that inclusion of the secondary class can be varied continuously by A_{min}.

For a clearer intuition of this subpixel area allocation procedure, consider the histograms of the confidence measures for MLCT pixels with primary/secondary classes equal to crop/mosaic

or mosaic/crop, shown in Figure 2. These pixels are atypical in a sense because the mosaic class is itself defined as a hybrid of crop and natural cover that contains about 40–60% cropland. By pinning this fraction at 50% our algorithm defines the fractional area of crop in crop/mosaic cells as a function of the confidence $c(x)$ to be:

$$A_{crop}(x) = \frac{1 + c(x)}{2} + \frac{1 - c(x)}{4} = \frac{3 + c(x)}{4} \tag{5}$$

since half the mosaic class is going to crop. Similarly for mosaic/crop cells:

$$A_{crop}(x) = \frac{1 + c(x)}{4} + \frac{1 - c(x)}{2} = \frac{3 - c(x)}{4} \tag{6}$$

implying that, at minimum, these cells are majority cropland.

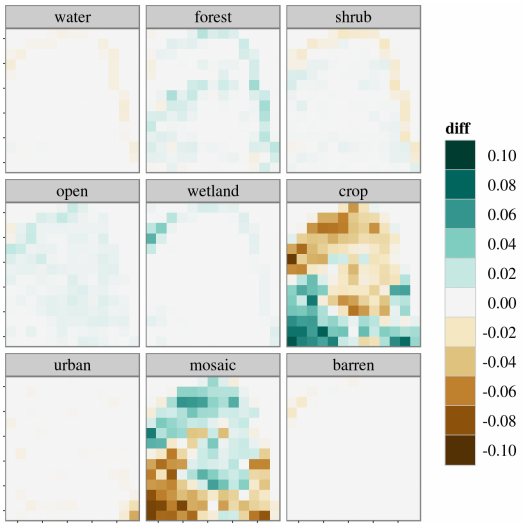

Figure 3: Difference of aggregated subpixel fractions for $A_{min} = 1.0$ vs. $A_{min} = 0.5$, positive when $f(A_{min} = 0.5)$ is greater.

Computationally the process of converting the reclassified maps to subpixel fractions at the desired $5'$ resolution is a three-step process. First, we calculate the fraction of the primary cover type as a function of the classification confidence as described above, independently of the primary and secondary classifications. Next, a subpixel fraction for each cover type is calculated at the original $15''$ resolution–this fraction is zero for all but the one or two classes indicated. Third, we aggregate to a coarser resolution by calculating the means of these fractions over the intersecting $15''$ pixels within a given $5'$ grid cell. Figure 3 uses a difference map to shows the different outputs that can result from setting $A_{min} = 0.5$ as against $A_{min} = 1.0$ when calculating the subpixel fractions and aggregating those fractions to $5'$. Positive values in the map indicate areas where $A_{min} = 0.5$ produced a greater value. Considering the secondary class results in a shift of up to 10% of total cell area from crop to mosaic in the north of the detail area and vice versa for the southern portion.

3.2 Preparing NLCD

Our reclassification for NLCD, shown in Table 2, is more complicated than that used for MLCT (Table 1). Although NLCD has fewer forest classes than MLCT, they are equally unambiguous. We

equate four NLCD developed land classes with the PEEL_0 "urban" class, to represent developed areas of all densities, but maintain them separately in anticipation of MLCT only capturing high density development. The result of this reclassification is shown in Figure 4 at native NLCD resolution. Many detailed features missing from the 500 m MLCT product are apparent in this figure, including rural transportation networks and small wooded, water, and wetland features. Figure 16 shows the result of reclassifying NLCD and aggregating it to $5'$ subpixel fractions for the full cUSA.

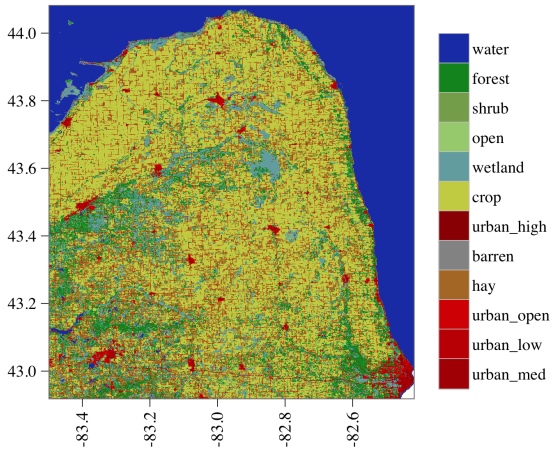

Figure 4: Detailed study area in southeastern Michigan showing NLCD reclassified at native 30 m ($\sim 1.25''$) resolution.

Though repeating the aggregation process for the entire study area is computationally expensive because of NLCD's high resolution, the algorithm is the same as for refactoring MLCT when considering only the primary cover type (i.e., setting $A_{min} = 1$), so we do not describe it here. Figure 5 shows the effect of the aggregation for the detailed study area.

	NLCD	
11[a]	water	water
41	deciduous forest	forest
42	evergreen forest	
43	mixed forest	
52	shrub/scrub	shrub
71	grassland / herbaceous	open
90[a]	woody wetlands	wetland
95	emergent herbaceous wetlands	
82	cultivated crops	crop
81	pasture / hay	hay
21	developed, open space	urban_{open}
22	developed, low intensity	urban_{low}
23	developed, medium intensity	urban_{med}
24	developed, high intensity	urban_{high}
	(no equivalent)	mosaic
12	perennial ice/snow	barren
31[a]	barren land	

[b] Additional coastal classes exist in NLCD but are not present in the lower 48 states and so are not included here.

Table 2: Reclassification of NLCD to PEEL_0. See Table 1. Adapted from *Homer et al.* [2004].

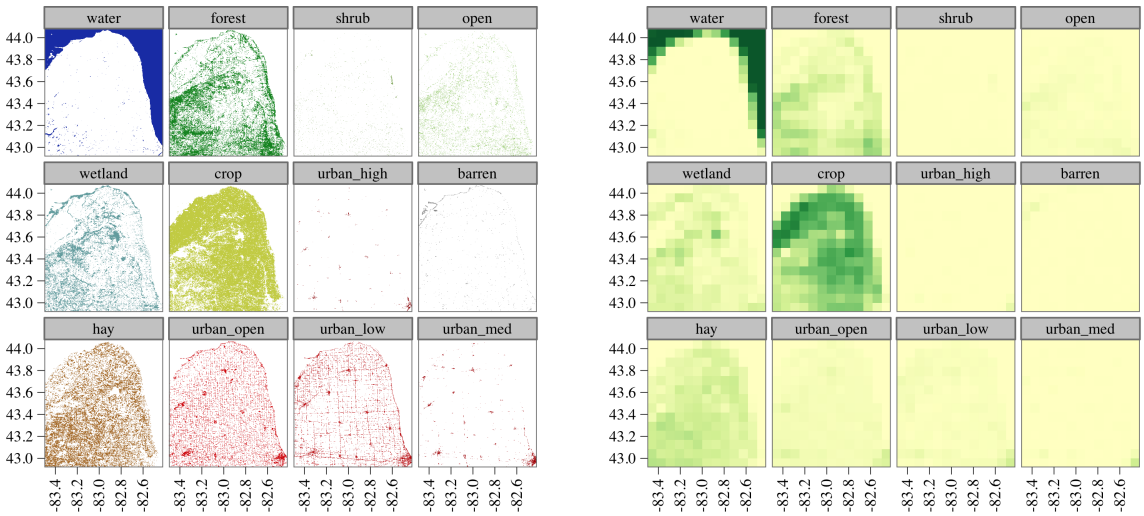

Figure 5: NLCD covers at native resolution with discrete categories (left) and aggregated cover fractions at PEEL_0 resolution with sub-pixel fractions ranging from 0 (yellow) to 1 (green) (right) shown separately for the detailed study area.

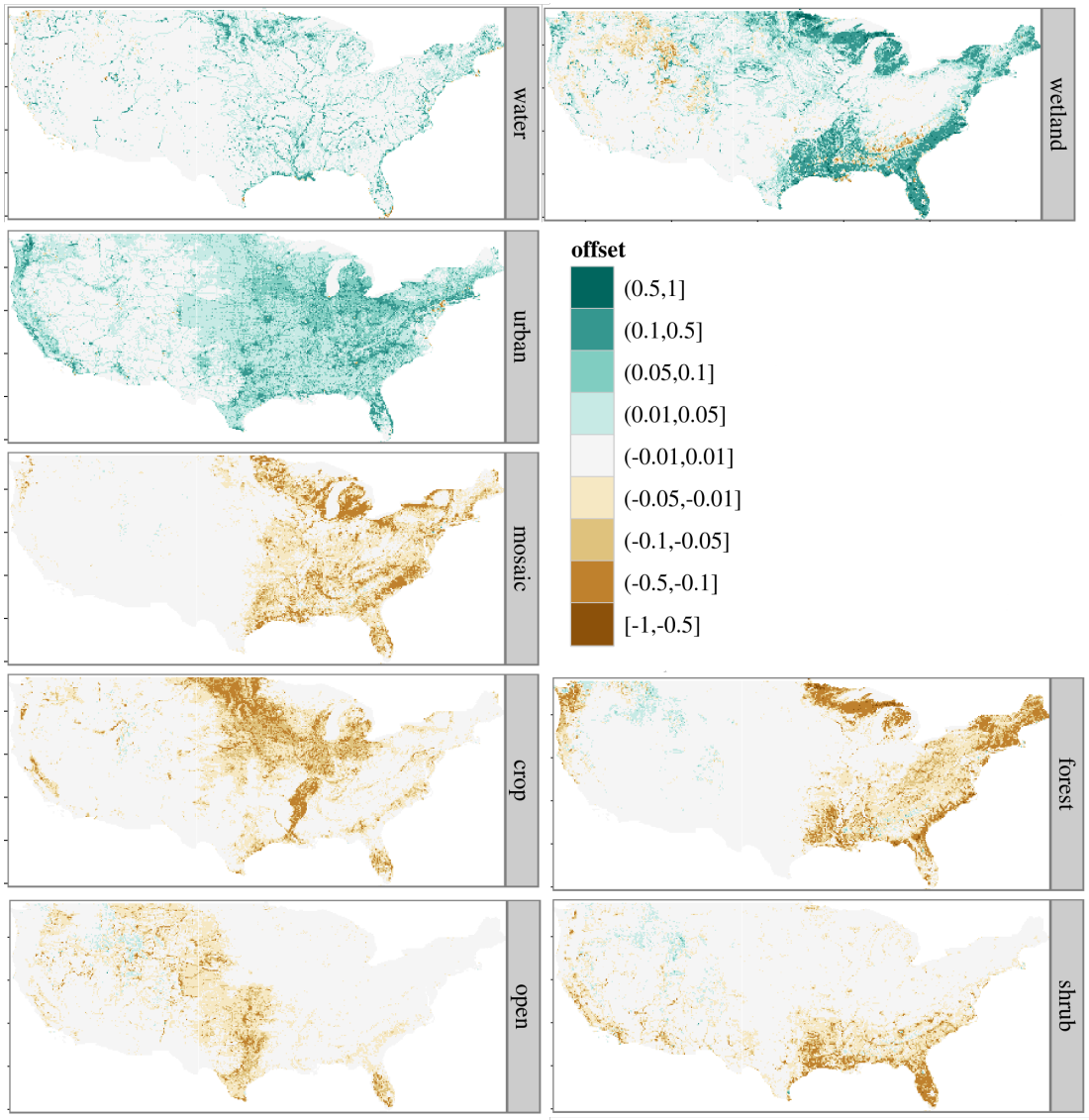

Figure 6: NLCD offsets (log scale). Water, wetland, and urban cover fractions are tuned to match those in NLCD, and the difference is accommodated as described in Section 3.3.

3.3 Applying NLCD Offsets to $PEEL_0$

Table 3 shows that MLCT is negatively biased in the total areas assigned to water, wetland, and urban features relative to NLCD. Visual inspection shows that features of these classes tend to have smaller characteristic dimensions, causing them to be overlooked in MLCT because of its resolution. One example is the rural transportation networks in areas surveyed under PLSS, where roads have been laid out on a generally regular square-mile grid. $PEEL_0$ includes this infrastructure in the urban class as another form of developed land.

To merge this information from NLCD, we first accept the areas for water, wetland, and urban classes in the reclassified, 5′-aggregated version of NLCD that we have computed as truth. We then calculate offsets for those classes versus our 5′ MLCT data by direct subtraction. Where NLCD is greater, the difference will be positive, and so a positive offset will be added to the fraction already present for any one of the "truth" classes from NLCD. The other classes are then adjusted so that they are present in proportion to each other as indicated by MLCT but in the area remaining after accepting the water, wetland, and urban areas from NLCD. Figure 6 shows the spatial distributions of the offsets calculated based on our assumptions about the water, wetland, and urban classes in NLCD. The logarithmic scale used in these maps makes apparent both areas of significant adjustment, greater than 10%, and the extent to which small adjustments on the order of 1–5% occur. We see the detailed structure of drainage networks in the water class and population centers in the urban class that could easily be confused with the vegetative classes in the MLCT classification, perhaps in heavily wooded suburbs where transportation infrastructure is obscured and difficult to resolve, for example. The offsets for the NLCD truth classes (urban, water, and wetland) are generally positive, although not strictly so because the algorithm does not preclude the possibility that MLCT may locally overestimate these classes in particular regions and still suffer an aggregate deficit relative to NLCD.

3.4 Decomposing the MLCT mosaic class

The MLCT "cropland/natural vegetation mosaic" class is problematic for the economic models for which $PEEL_0$ is intended, because it combines developed land use and natural land cover. This class is defined as a hybrid of cropland and some mixture of natural covers (forest, shrub, or open) with no single component exceeding 60% [*Friedl*, 2002]. We wish to differentiate the cropland from the natural vegetation in order to calculate a more meaningful total for cropland area and thereby eliminate the mosaic class from the final tabulation. To this end, we make three simple assumptions about the composition of area identified as mosaic lands:

1. 50% of mosaic area is assigned to the crop class.
2. The other 50% is a blend of forest, open, and shrub in relative proportion to the expression of those classes in the same 5′ cell.
3. In the absence of any natural classes in the 5′ cell the natural component of the mosaic is an equal blend of all three.

We make these simplifying assumptions so that we can proceed with the evaluation of this analytical framework. It might be interesting to vary the proportion of mosaic land allocated to cropland. However, we have no principled basis for doing so, despite the definition's implication that this proportion is variable. Our chosen 50% level reflects the assertion that the mosaic is a cultural class grouped with cropland and urban in the IGBP classification scheme without overstating the degree of development. MLCT provides adequate variability in this dimension by commonly pairing

cropland and mosaic in the primary/secondary class data. The second assumption imposes that 15″ mosaic cells' noncrop portion have the same relative composition of forest, open, and shrub as does the nonmosaic portion of the 5′ grid cell in which it falls. Therefore, mosaic pixels in a 5′ cell where only forest is found in the three non-crop mosaic components, are assigned 50% crop and 50% forest.

4 Evaluation of $PEEL_0$

We submit two standards for assessing the credibility of the $PEEL_0$ dataset that we have produced: aggregate areas and average cell-by-cell error. For an independent reference measurement of aggregate areas we use totals for the cUSA study area from the MLU data set. This comparison is appropriate because we have converted MLCT's cropland class, a bulk measurement of a functional landscape area, to the $PEEL_0$ crop class, the portion of that landscape that is actively involved in agricultural production at that point in time and is therefore commensurate with the MLU statistics. However, because MLU only offers state-level statistics it cannot tell us very much about the quality of the distribution of $PEEL_0$'s crop class. Therefore we also compare $PEEL_0$ to NLCD's agricultural components, crop plus "hay", as well as Agland2000. Both of these data sets exhibit close correlation with published agricultural statistics in the aggregate but each exhibit their own particular distributional qualities with respect to $PEEL_0$. In this section we show that our process of incrementally adjusting $PEEL_0$ produces consistent improvement in the distribution of crop area in terms of average per-cell error with respect to both reference data sets. Along the way we also compare the affect of varying the parameter A_{min} over values of 1.0 and 0.5 which correspond to considering only MLCT's primary classification and maximum incorporation of the secondary class, respectively, as described above to show that there is useful information content in the subsidiary variables offered by MLCT on the same basis.

4.1 Comparison of Aggregate Areas

We start by tabulating the aggregate areas by class for MLCT, NLCD, and Agland2000. After decomposing the mosaic class, MLCT indicates 495.4 Ma (200.5 Mha) of cropland for A_{min}=0.5 and 488.1 Ma (197.5 Mha) for A_{min}=1.0 in the cUSA in 2001, the NLCD indicates roughly 448.9 Ma (181.7 Mha) combined of "cultivated crops" and intensively managed "pasture/hay", Aglands2000 indicates roughly 446.5 Ma (180.7 Mha) of cropland, and MLU indicates roughly 441.3 Ma (178.6 Mha) of combined cultivated crops and "cropland pasture. " The areas for all the major comparison and intermediate datasets, and most parameter choices considered here, are shown in Figure 7, and key values are collected in Table 3. Aggregate cultivated land is within 0.8% of the MLU value, compared with 16.2% for the Modis Land Cover Type (MLCT) primary cover product, 1.8% for the National Land Cover Database (NLCD), and 1.2% for the Agricultural Lands in the Year 2000 (Agland2000) dataset. Aggregate water, natural, and urban cover classes are within 2.2, 1.0, and 6.1%, respectively, compared with deviations of 24.0, 0.01, and 69.2% for MLCT primary and 2.2, 1.35, and 6.1% for NLCD. Third, the spatial distribution of cultivated land is improved; $PEEL_0$'s per cell subpixel fraction root mean square error for cropland relative to NLCD is 0.149 versus 0.175 for the 2001 MLCT primary classification, a 16% improvement.

We are concerned here primarily with the ∼10% overestimation of cropland in MLCT, especially in light of the relative agreement among the other products considered here. The inability of MLCT to resolve rural transportation networks, minor settlements, and small water or wetland features is a major contribution to its surplus of cropland acreage indicated. NLCD's greater resolution, ∼30 m vs. ∼500 m, makes it better suited to discerning developed areas in rural landscapes, ranging

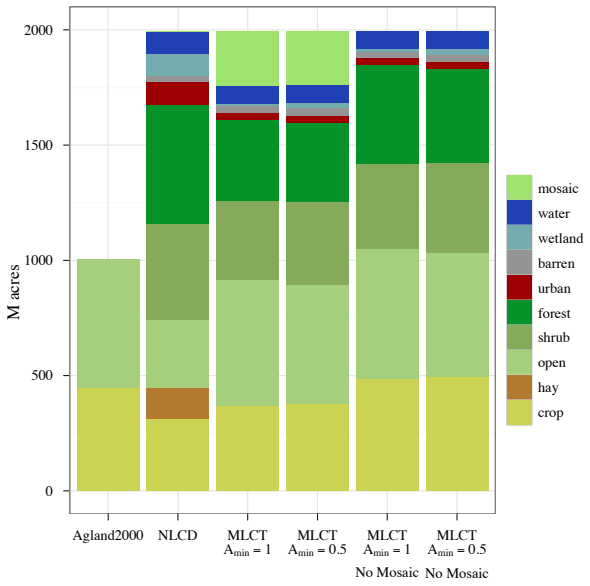

Figure 7: Total acreages by map and cover showing significance of MLCT mosaic class and underestimation of water, wetland and urban classes relative to NLCD.

	MLU 2002[a]	Agland 2000	NLCD	MLCT A_{min}=1	MLCT A_{min}=.5	MLCT No Mosaic	WWU Offset	MLCT Adjusted	$PEEL_0$
water			96.5	75.0	74.3	74.3	22.3	96.5	96.5
forest	657.1[b]		513.2	353.6	344.7	410.8	-44.7	300.1	355.7
shrub			420.1	341.8	358.7	387.2	-23.8	334.9	358.0
open	584.2	557.1	291.2	545.8	516.9	538.7	-21.0	495.9	514.9
wetland			95.0	11.0	26.0	26.0	69.0	95.0	95.0
crop	379.5[c]	446.5	310.8	369.6	378.9	495.4	-39.0	339.9	437.6
pasture	61.8 [c]		138.4						
urban	96.9[d]		102.8[e]	29.8	27.3	27.3	75.4	102.8	102.8
mosaic				237.0	232.9		-37.4	195.5	
barren			24.5	28.9	32.8	32.8	-0.9	31.9	31.9
other	112.0[f]								
(all)	1893.8	1003.7	1992.5	1992.5	1992.5	1992.5	0.0	1992.5	1992.5

[a] Data from MLU does not include coastal or large inland water bodies.
[b] We include parks and other protected forest areas with the forest class.
[c] For MLU we distinguish "cropland pasture" from harvested and idle cropland for better comparison.
[d] We combine rural transportation networks and other development with the urban class.
[e] This includes all NLCD developed classes: high, medium, and low density and "developed open."
[f] This includes miscellaneous lands, marshes, swamp, desert, etc., plus 14.8 Ma designated for defense or other special purpose.

Table 3: Effect of NLCD offsets on total acreages, $A_{min} = 0.5$ everywhere, except where explicitly noted.

from rural roads to farmsteads to small communities that do not show up in MLCT. A total area of roughly 75.4 Ma (30.5 Mha) of land developed to one extent or another remains after subtracting

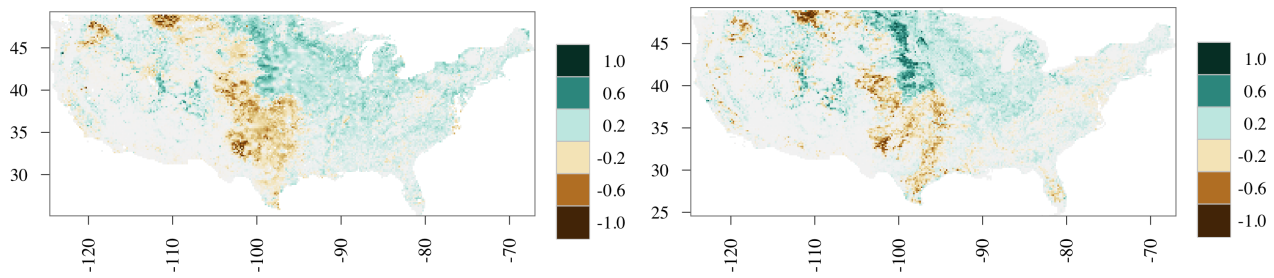

Figure 8: Difference between MLCT (no mosaic, A_{min}=0.5) and Agland2000 (left) and NLCD (right) crop.

the MLCT urban class from all developed classes in NLCD. Applying this difference as an offset brings us significantly closer to the expected acreage under cultivation in 2001; see Section 3.3. This offset also brings the national area of urban and developed cover much more in line with MLU census data.

4.2 Comparison of Root Mean Square Errors

The purpose for processing the MLCT for two values of A_{min} as described in Section 2 is to evaluate whether information from the secondary cover type contributes positively to the accuracy of the dataset we seek to synthesize. The primary objective of this synthesis is to achieve accuracy in cropland distribution. Although MLCT overstates cropland acreage for both A_{min}=0.5 and A_{min}=1.0, we discriminate among the two on the basis of error distribution rather than aggregate error. Figure 8 shows the cell-by-cell differences in crop area fractions between an intermediate MLCT-derived data set and both Agland2000 and NLCD. In these figures A_{min}=0.5 and the crop area includes half of any mosaic area by virtue of mosaic decomposition. The maps for A_{min}=1.0 are not visually distinguishable from this result but calculating a per-cell average magnitude of error, defined as the root mean square error (RMSE), does reveal the advantage of accepting A_{min}=0.5. RMSE is defined mathematically as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}{n}} \tag{7}$$

where $\hat{\theta}_i$ are the "predictions" derived from the respective MLCT derivations and θ_i are the "observations" taken from the Agland2000 or NLCD datasets in turn. The RMSE values, collected in Table 4, show clearly that choosing A_{min}=0.5 improves the spatial correlation of the MLCT cropland layer compared with both Agland2000 and NLCD.

4.3 Comparison of Hexbin Plots

To further examine the relationships between the distributions of cropland that we derive from MLCT, on the one hand, and from Agland2000 and NLCD, on the other, we create "hexbin" plots. These are essentially two-dimensional histograms that show the number of grid cells that occur within discrete regions of the space defined by coordinates that are cropland fractions for the two datasets. A hexbin plot operates much like a common scatter plot; but for datasets with as many observations as we wish to include, it gives a cleaner representation of that structure. We employ

WWU[a] Offset	A_{min}	RMSE (Agland2000)	RMSE (NLCD)
no	1.0	0.180	0.175
no	0.5	0.165	0.161
yes	0.5	0.151	0.149

[a] "WWU offset" refers to the water, wetland, and urban offsets calculated from NLCD and applied to the data product in Section 3.3.

Table 4: RMSE vs. Agland2000 and NLCD for a variety of intermediate and final data products.

a logarithmic scale for the bin counts to obtain a more complete picture of the overall dispersion and local concentration of the observations.

Figure 9 plots the crop fractions of MLCT with A_{min}=1 and 0.5 versus the Agland2000 crop and NLCD "crop+hay" layers. As one would expect, an overall correlation exists among these variables, especially given that Agland2000 provides prior probabilities to the MLCT classification (and that an early version of MLCT, BU-MODIS, provides a key input to the construction of Agland2000). The MLCT primary class clearly exhibits a positive bias overall, although a subset that is negatively biased is also apparent for low values of the Agland2000 crop fraction in the interval $[0.1, 0.5]$. Note also the drastic decrease in correlation when Aglands2000 reaches 1.0 relative to the stronger relationship over the interval $[0.8, 1.0)$. Something peculiar about the Agland2000 allocation procedure appears to drive the crop fraction to its maximum in areas where the remote sensing data clearly resists such a characterization. One possible explanation is systematic errors in the agricultural census data that drive the Agland2000 algorithm, forcing unrealistically high concentrations in order to satisfy the algorithm's constraints.

Figure 9 also shows the correlation of MLCT vs. the NLCD crop and hay/pasture layers. We see clearly the positive bias of MLCT relative to NLCD in general, and especially over the interval $[0.8, 1.0]$. These are the cells in which MLCT sees nearly complete cropland, whereas NLCD sees sometimes substantial mixing with other classes, which we attribute to small features missed by MLCT. A curious feature arises here as well, similar to the decorrelation where Agland2000 showed 100% crop, namely a decorrelation where NLCD shows close to 0% crop. The origin of this feature is not clear to us, but we suspect a classification error in MLCT.

Setting A_{min}=1 produces a greater overall error by assigning entire MLCT pixels to the cropland class and not allowing for the possibility of mixed covers, not too mention the finer details that exceed the instrument's resolution. The results on Table 4 indicate that A_{min}=0.5 is more representative of the distribution of cropland as compared with both NLCD and Agland2000 because even before adjustments. Although the total area indicated is higher according to Table 3, there is less error on a cell-by-cell basis, indicating that it does a better job of representing the spatial distribution than A_{min}=1.0. It reduces the RMSE against Agland2000 from 0.180 to 0.165, and against NLCD from 0.175 to 0.161. These results are reflected in the structure revealed by Figure 9, where fewer cells in the MLCT data are assigned 100% crop because of the secondary class. Additionally, where crop was included in a secondary class, it also caused cells of near-zero value for MLCT to lift away from the x-axis. The uncorrelated observations for Agland2000 equal to 1.0 and NLCD equal to zero are still present, however.

This result persuades us that considering the secondary class with $A_{min} = 0.5$ is indeed the correct approach. Because of the improved fit with Agland2000 and NLCD cropland and its full

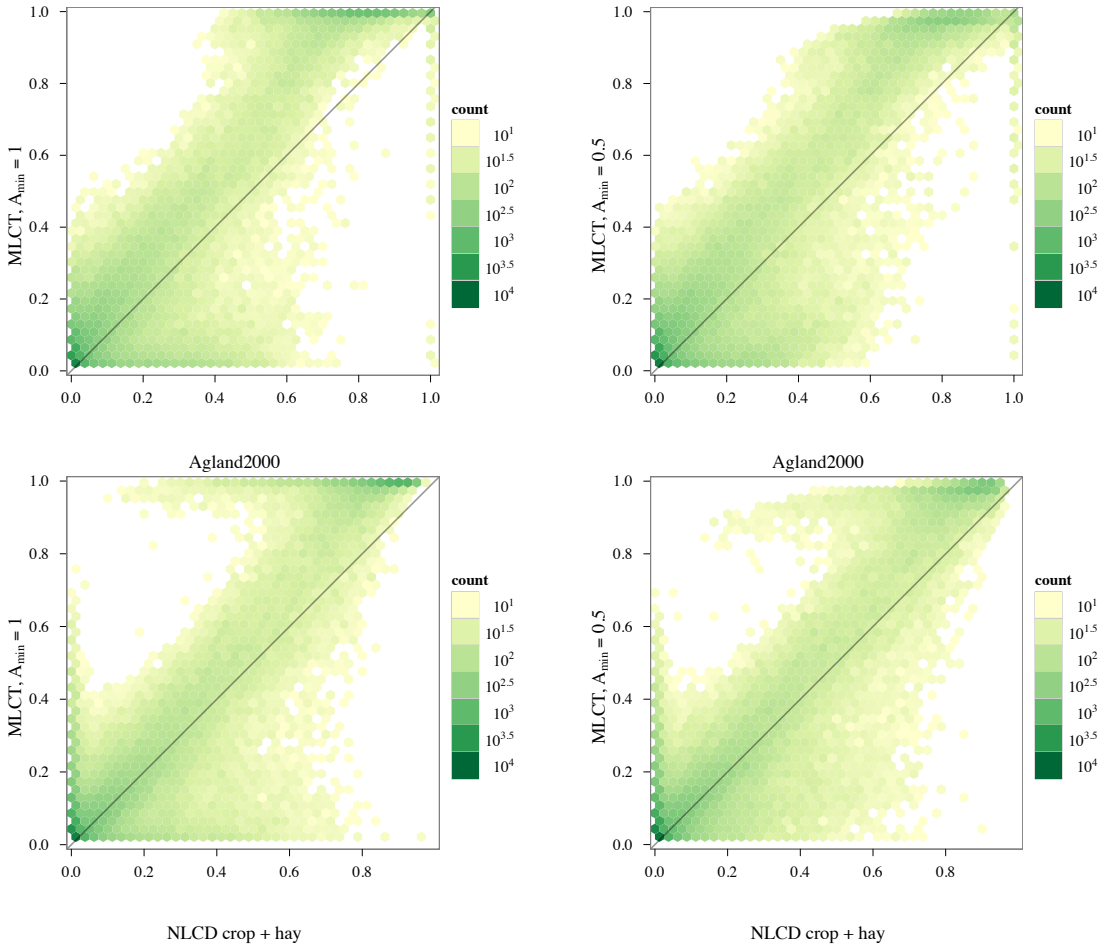

Figure 9: Hexbin plot of MLCT crop $A_{min} = 1.0$ (left) and $A_{min} = 0.5$ (right), both with mosaic removed, versus Agland2000 crop (top) and NLCD cropland plus hay/pasture (bottom).

consideration of all information imparted by the MLCT data this is the result that is adopted as PEEL_0^a indicated in the algorithm outline in Section 3 above and utilized in subsequent analysis.

4.4 Evaluation of NLCD Offsets

In Section 3.3 we discussed how to correct for particular classes of features that cannot be resolved in the MLCT data by accepting water, wetland, and urban fractions from NLCD and adjusting aggregated MLCT fractions accordingly. Figure 10 shows the totals by class of the offsets that result from this calculation. The item labeled "total" appears blank because a value of zero is plotted there indicating that area was conserved in this operation (i.e., that area subtracted from one class was reallocated to another). We perform the same error calculations and aggregate comparisons as above to assess whether adding NLCD offsets improves overall cropland accuracy. Table 4 gives the RMSEs for the various intermediate products. We see that each step in the algorithm improves the overall error relative to both Agland2000 and NLCD.

The most important outcome with respect to our stated objective of bringing total cropland areas in line with the total from NLCD and Aglands2000 is the reduction of crop areas by 39 Ma (15.8 Mha) and mosaic areas by 37.4 Ma (15.2 Mha). This total reduction of 57.8 Ma (23.4 Mha) of the final crop class after mosaic decomposition brings the national cropland area for $PEEL_0$ into close agreement with NLCD, MLU, and Agland2000, 437.6 Ma vs. 449.2, 441.3, and 446.5 Ma, respectively. See Table 3 and Figure 11 for a detailed comparison of the effects of the offset. Figure 12 shows two additional hexbin plots using the offset-adjusted MLCT fractions, which is the $PEEL_0$ dataset.The six RMSE values given in Table 4 correspond to the six hexbin plots in Figures 9 and 12. The significance of this result is that it is not conditioned by the desired cropland area estimate; rather, it shows a convergence in these estimates by selectively incorporating information about other classes from NLCD. Figure 13 shows the $PEEL_0$ dataset incorporating these offsets.

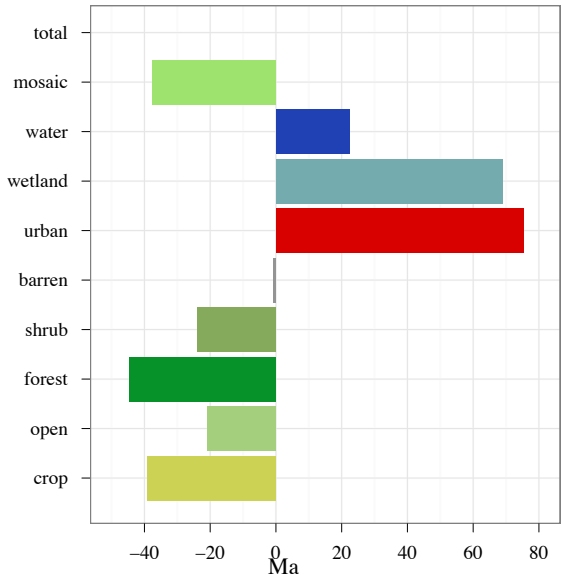

Figure 10: Total offsets calculated from NLCD.

4.5 Potential Further Offsets and Adjustments

Little consensus exists among the datasets regarding the relative coverage of forest, shrub, and open land, or indeed the definitions of these classes. Thus, we find it impractical to compare among them. MLCT shows 1220.3 Ma of land in one of these three classes, just slightly less than the final amount (1228.6 Ma) in $PEEL_0$, the additional area from mosaic decomposition being roughly cancelled by reductions due to WWU offsets. This number compares favorably to the 1224.5 Ma distributed among these three classes in NLCD and the 1241.3 Ma distributed among forest and open in MLU, which lacks a class analogous to "shrub."

We attribute these differences to inconsistent class definitions caused by the fairly continuous transition from open to shrub and from shrub to forest. For instance, NLCD defines forest land to be any land with greater than 25% coverage by a tree canopy that is generally greater than six meters tall. Shrubland in NLCD is similarly defined as any area with a shrub canopy (defined as generally less than 5 or 6 meters high) typically covering greater than 20% of the pixel. Open herbaceous land is thus roughly everything else that is not subject to intensive management (besides wetlands),

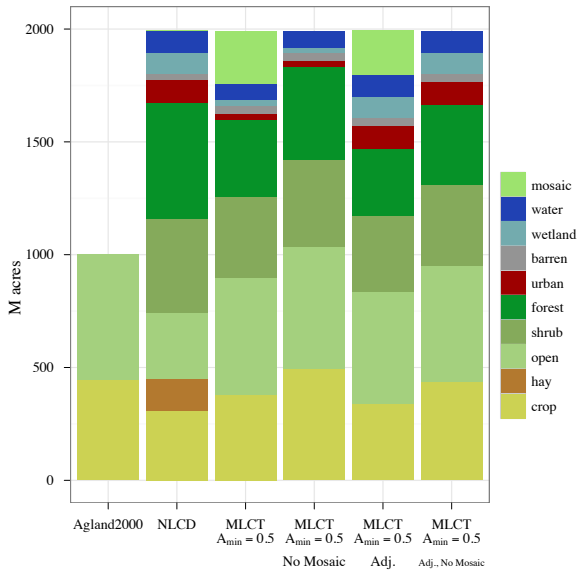

Figure 11: Total acreages after NLCD adjustment.

meaning the pixel is at least 80% open grassland. This definition is rather strict compared with the definitions of open grasslands used in MLU and MLCT, which leads to substantial disagreement among these datasets.

5 Conclusions

We have described and validated a new algorithm for combining LULC data from multiple sources, scales, and extents to construct hybrid data products with improved accuracy and fidelity. Using this algorithm, we have constructed a customized data product, PEEL_0, designed for use in studies of the economic drivers of LULC change. We stress, however, that the output of this work is more than just a static land cover dataset; it is an algorithmic framework and open source code base for producing global or regional land cover data products with customized cover representations, which integrates high-value information at multiple scales to improve and validate land cover characterizations. Variants of this method could construct other custom datasets that may offer significantly improved accuracy and fidelity for other purposes. We hope that by making our methods available in a form that permits sharing, modification, and extension, we will increase both the transparency and quality of LULC research. *Best* [2011] provides further details on the software environment and reproducible research framework used in this work. Because many interesting steps in our algorithm depend on high-quality, high-resolution LULC data that is not yet available on a global basis, we have focused our description on the US. Nevertheless, the improved accuracy that we obtain in the US lead us to expect that our procedures for aggregating MLCT to a coarser resolution using a statistical interpretation of subpixel cover and distinguishing likely components of the mosaic class can also help to more accurately characterize the distribution of global land cover, especially cropland.

As stated previously, an ultimate objective of this method is to produce a global version of this dataset that exhibits the same qualities of delineating cultivated areas and providing a complete

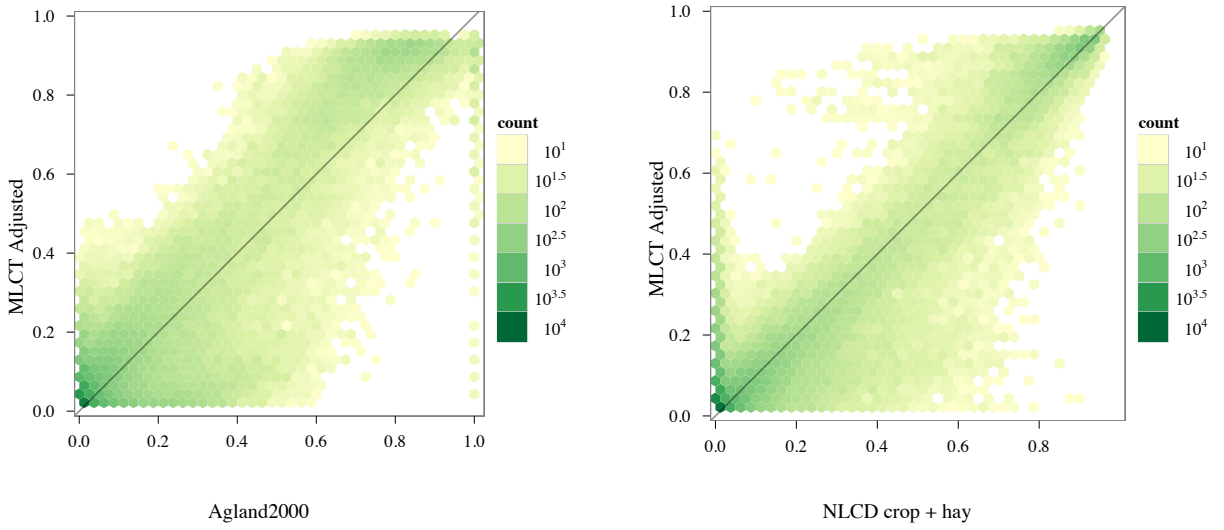

Figure 12: Hexbin plot of MLCT adjusted crop versus Agland2000 cropland (left) and versus NLCD crop plus hay/pasture (right).

LULC characterization. In the absence of high-resolution data on rural development, the lower-density portions of the "urban" class that falls below MLCT's detection threshold, it may be possible to model the overestimation factor of the MLCT cropland class. This factor would be defined as the ratio of total area encompassed by the MLCT cropland classification to area actually under cultivation and could potentially be modeled as a function of classification confidence and secondary class using the data described and produced here as a training set. The null hypothesis in the formulation of such a model is that enough diversity exists among agricultural landscapes in our cUSA study area to adequately characterize agricultural landscapes world in this regard. Similarly it may be possible to directly model the "urban" percentage below the MLCT detection threshold as a function of population density and agricultural productivity, identifying said threshold in the process. There is a clear dependency between these offsets in agriculturally productive regions so modeling them in conjunction may be constructive. We expect that global offsets for the water and wetland classes will be harder to model without corresponding proxy statistics. However, we may expect greater availability of spatially explicit catalogs of ecological services and sensitive/protected areas in the future that will close these gaps in available information. We will pursue these directions in future work.

Acknowledgments

This work was supported in part by NSF grant SES-0951576 and DOE contract DE-AC02-06CH11357, and by the Computation Institute at the University of Chicago and Argonne National Laboratory. We appreciate the significant editorial contributions made by Elizabeth Moyer in the process of preparing this work for distribution.

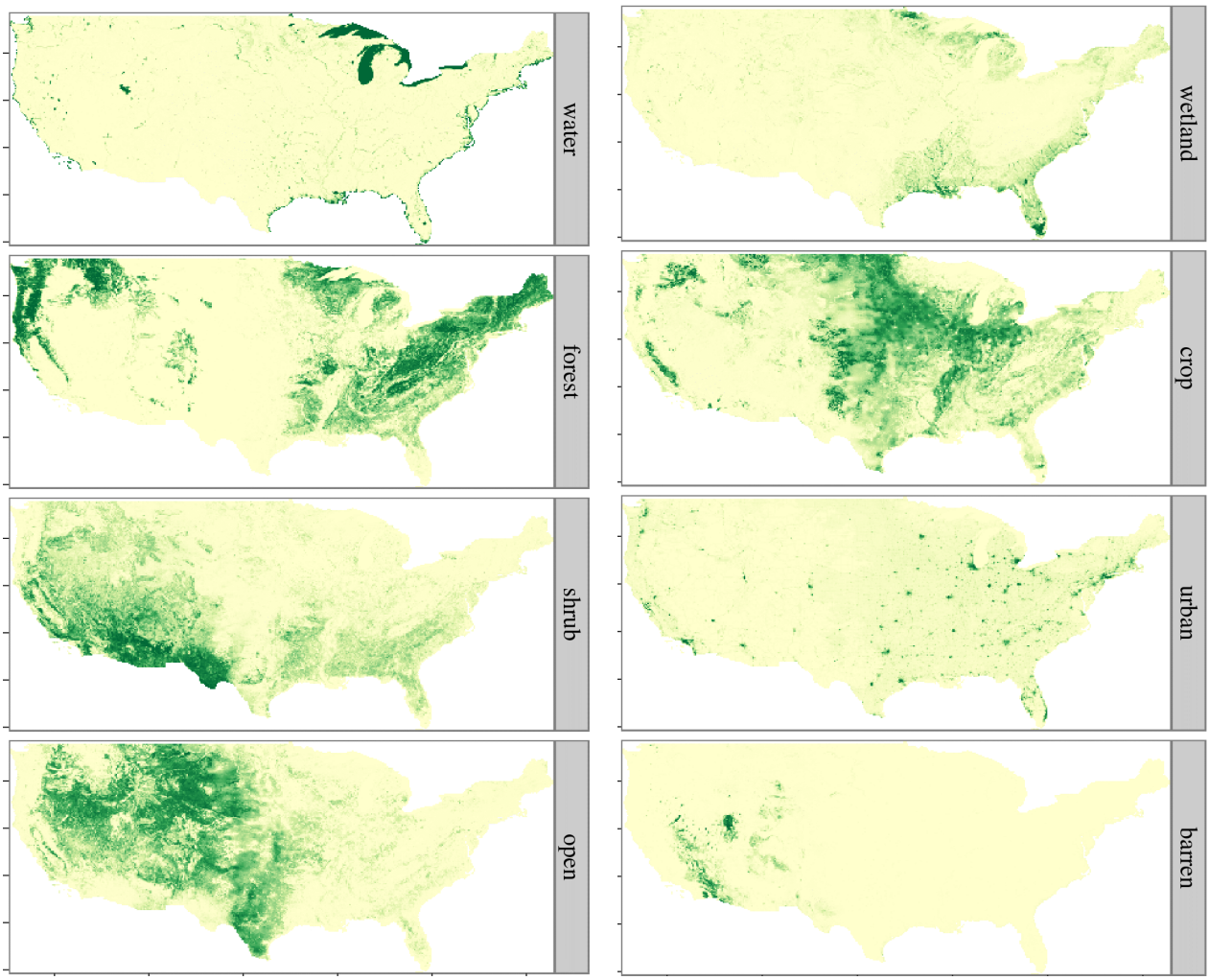

Figure 13: Final PEEL_0 maps.

References

Best, N. (2011), Synthesis of a complete land use/land cover data set for the conterminous united states emphasizing accuracy in area and distribution of agricultural activity, Master's thesis, Northeastern Illinois University.

Biradar, C. M., et al. (2009), A global map of rainfed cropland areas (GMRCA) at the end of last millennium using remote sensing, *International Journal of Applied Earth Observation and Geoinformation*, *11*(2), 114 – 129, doi:DOI:10.1016/j.jag.2008.11.002.

Friedl, M. A. (2002), Global land cover mapping from MODIS: algorithms and early results, *Remote Sensing of Environment*, *83*(1-2), 287–302.

Friedl, M. A., D. Sulla-Menashe, B. Tan, A. Schneider, N. Ramankutty, A. Sibley, and X. Huang (2010), MODIS Collection 5 global land cover: Algorithm refinements and characterization of new datasets, *Remote Sensing of Environment*, *114*(1), 168–182, doi:10.1016/j.rse.2009.08.016.

Hijmans, R., N. Garcia, and J. Wieczorek (2009), Global administrative areas (GADM) database, http://gadm.org/.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan (2004), Development of a 2001 National Land-Cover Database for the United States, *Photogrammetric Engineering Remote Sensing*, *70*(7), 829–840.

Homer, C., et al. (2007), Completion of the 2001 National Land Cover Database for the Counterminous United States, *Photogrammetric Engineering and Remote Sensing*, *73*(4), 337–341.

Jones, R., R. Leemans, L. Mearns, N. Nakicenovic, A. Pittock, S. emenov, and J. kea (2001), *IPCC AR-3 Working Group 2*, chap. 3. Developing and Applying Scenarios, pp. 147–190, Cambridge University Press.

LP DAAC (2008), Modis land cover type (MLCT, MCD12Q1 v005), https://lpdaac.usgs.gov/lpdaac/products/modis_products_table/land_cover/yearly_l3_global_500_m/mcd12q1, these data are distributed by the Land Processes Distributed Active Archive Center (LP DAAC), located at the U.S. Geological Survey (USGS) Earth Resources Observation and Science (EROS) Center (lpdaac.usgs.gov).

Lubowski, R. N., M. Vesterby, S. Bucholtz, A. Baez, and M. J. Roberts (2006), Major Uses of Land in the United States, 2002. (Economic information bulletin; no. 14), *Tech. rep.*, United States Department of Agriculture, Economic Research Service.

Matsuoka, Y., M. Kainuma, and T. Morita (1995), Scenario analysis of global warming using the Asian Pacific Integrated Model (AIM), *Energy Policy*, *23*(4-5), 357–371, doi:10.1016/0301-4215(95)90160-9.

Monfreda, C., N. Ramankutty, and J. A. Foley (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*(1), 1–19, doi:10.1029/2007GB002947.

Ramankutty, N., A. T. Evan, C. Monfreda, and J. A. Foley (2008), Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, *Global Biogeochemical Cycles*, *22*(1), 101,029/, doi:10.1029/2007GB002952.

Thenkabail, P., et al. (2008), A Global Irrigated Area Map (GIAM) Using Remote Sensing at the End of the Last Millennium, international Water Management Institute. Colombo, Sri Lanka. http://www.iwmigiam.org/info/gmi-doc/GIAM-world-book.pdf.

van Vuuren, D. P., B. Eickhout, P. L. Lucas, and M. G. J. den Elzen (2006), Long-Term Multi-Gas Scenarios to Stabilise Radiative Forcing Exploring Costs and Benefits Within an Integrated Assessment Framework, *Energy Journal*, *3*(Special Issue #3), 201–234.

van Vuuren, D. P., M. G. J. den Elzen, P. L. Lucas, B. Eickhout, B. J. Strengers, B. van Ruijven, S. Wonink, and R. van Houdt (2007), Stabilizing greenhouse gas concentrations at low levels: an assessment of reduction strategies and costs, *Climatic Change*, *81*(2), 119–159, doi:10.1007/s10584-006-9172-9.

You, L., and S. Wood (2006), An entropy approach to spatial disaggregation of agricultural production, *Agricultural Systems*, *90*(1-3), 329–347, doi:10.1016/j.agsy.2006.01.008.

You, L., S. Wood, and U. Wood-Sichra (2006), Generating global crop distribution maps: from census to grid, in *Selected paper at IAEA 2006 Conference at Brisbane, Australia*, 202, pp. 1–16.

A cUSA maps of datasets

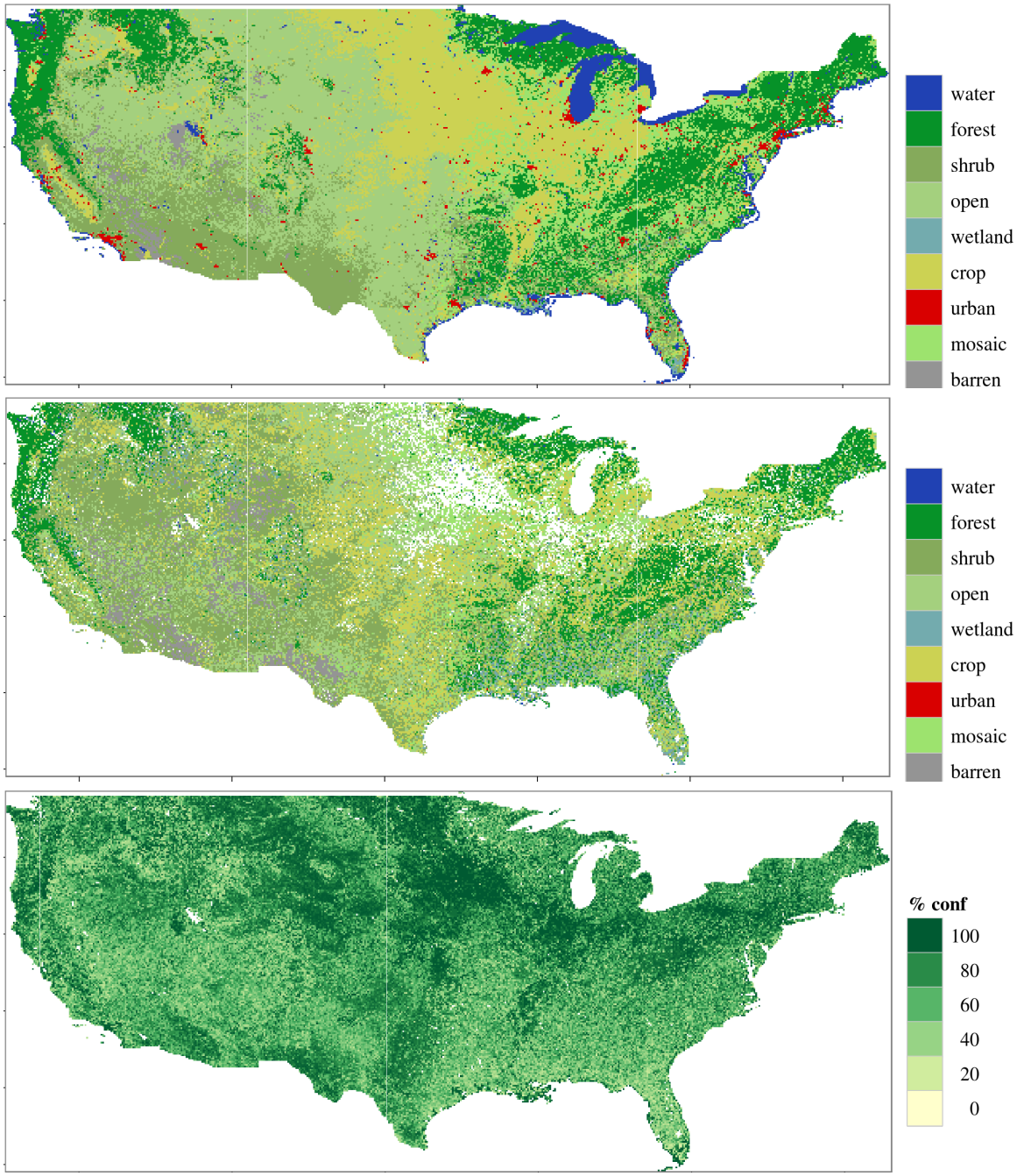

Figure 14: MLCT primary reclassified cover (top), secondary reclassified cover (middle), and primary cover classification confidence(bottom) for the cUSA.

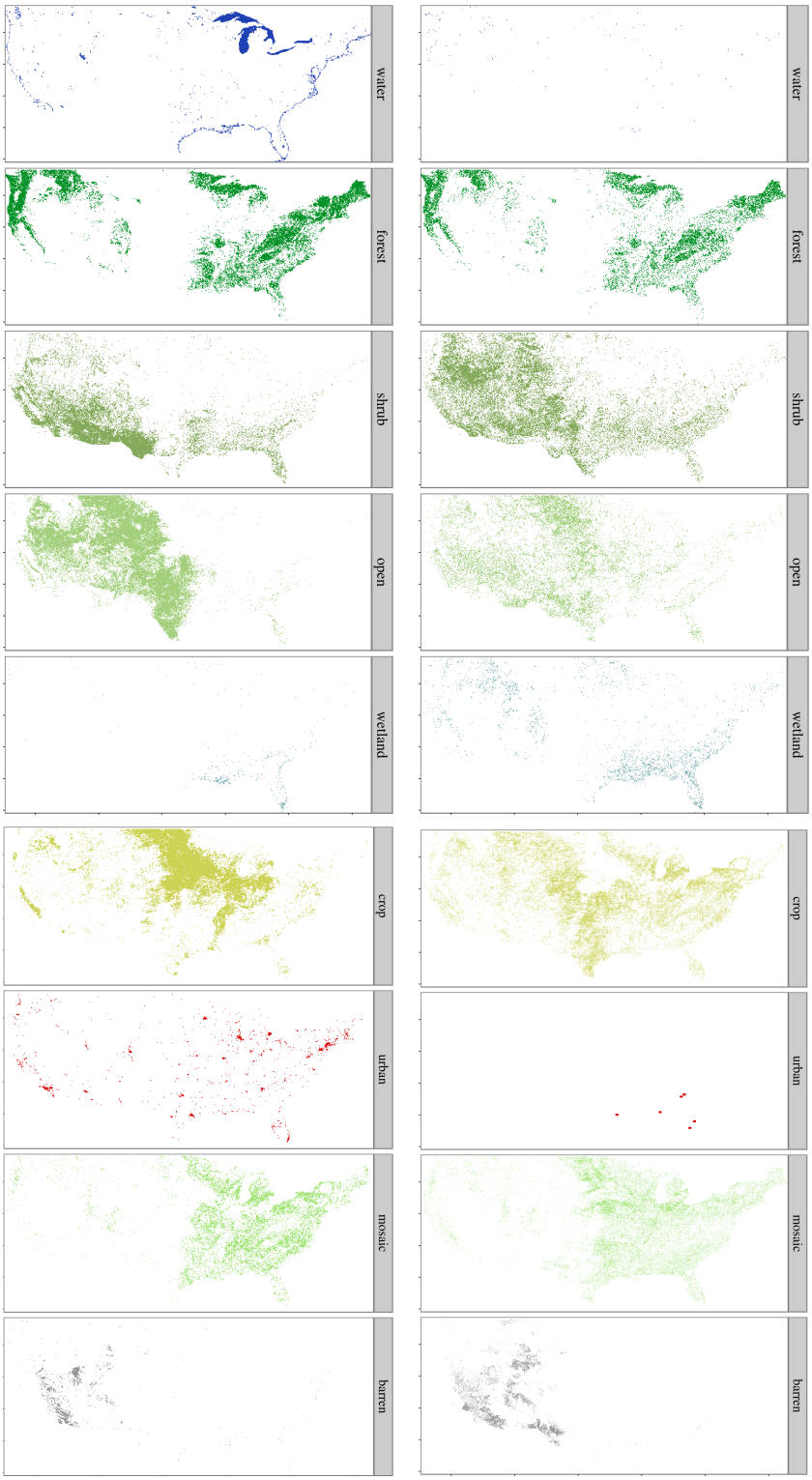

Figure 15: MLCT primary (left) and secondary covers (right).

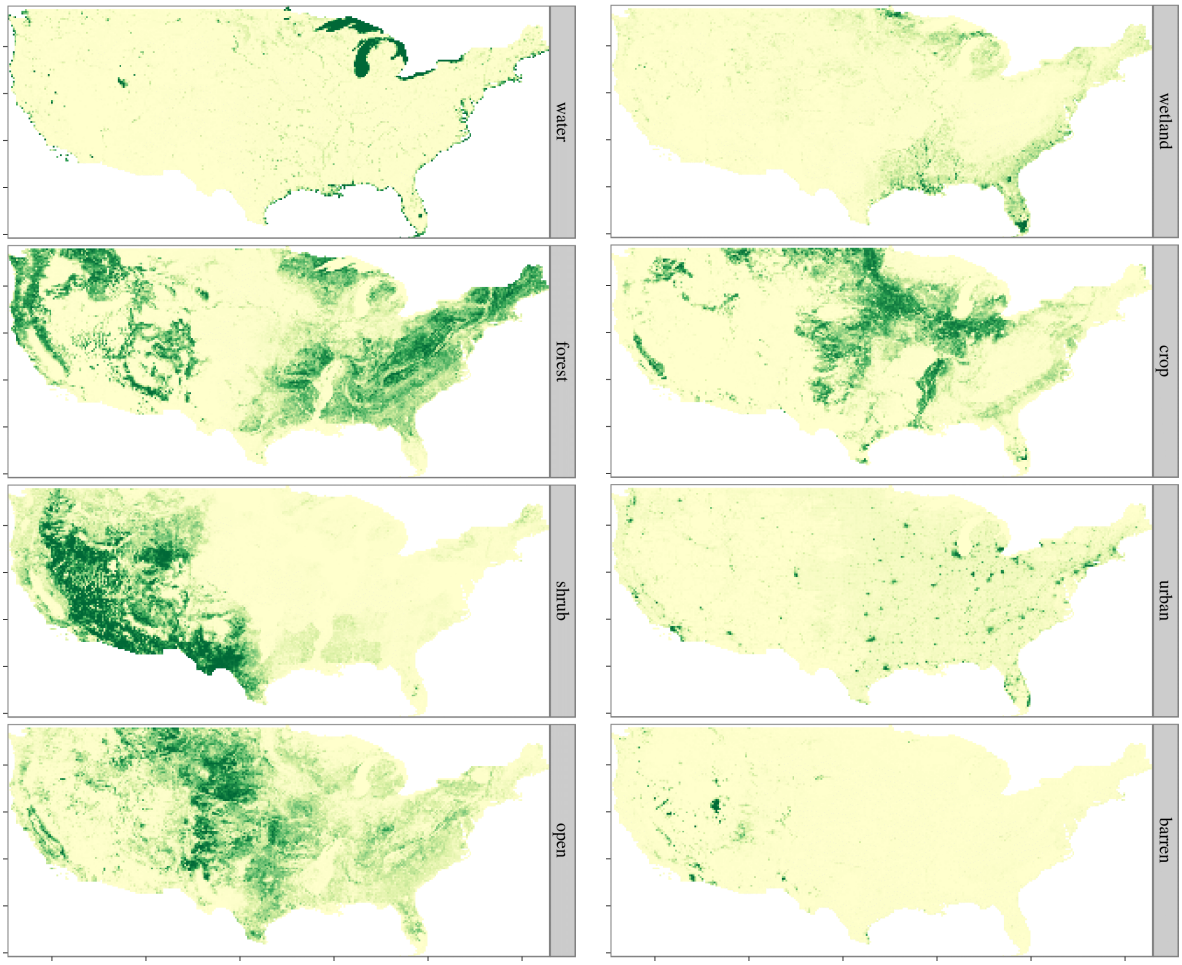

Figure 16: NLCD aggregated cover fractions.

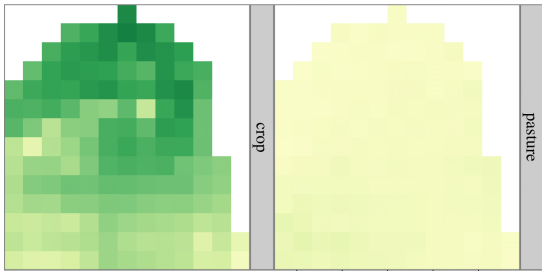

Figure 17: Agland2000 distribution in detail area.

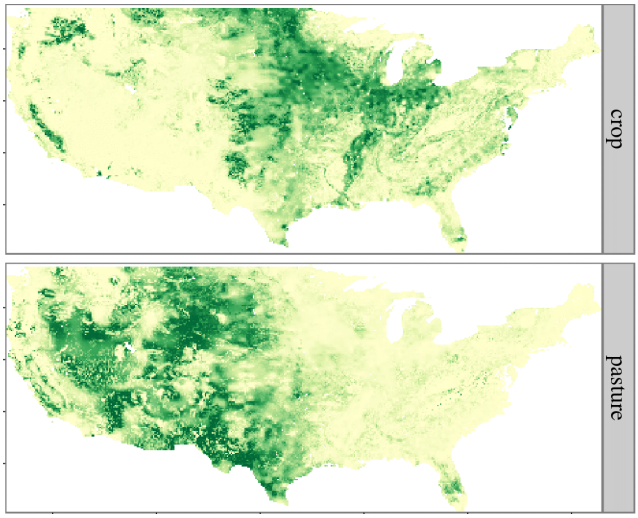

Figure 18: Agland2000 distribution in cUSA study area.

About RDCEP

The Center brings together experts in economics, physical sciences, energy technologies, law, computational mathematics, statistics, and computer science to undertake a series of tightly connected research programs aimed at improving the computational models needed to evaluate climate and energy policies, and to make robust decisions based on outcomes.

RDCEP is funded by a grant from the National Science Foundation (NSF) through the Decision Making Under Uncertainty (DMUU) program.

For more information please contact us at
info-RDCEP@ci.uchicago.edu
or visit our website:
www.rdcep.org

RDCEP
Computation Institute
University of Chicago
5735 S. Ellis Ave.
Chicago, IL, 60637 USA
+1 (773) 834 1726