9-2001

Linking remote-sensing estimates of land cover and census statistics on land use to produce maps of land use of the conterminous United States

George C. Hurtt
University of New Hampshire - Main Campus

Lynn Rosentrater
University of New Hampshire - Main Campus

Steve Frolking
University of New Hampshire - Main Campus, steve.frolking@unh.edu

B Moore
University of New Hampshire - Main Campus

Recommended Citation

Hurtt, G. C., L. Rosentrater, S. Frolking, and B. Moore III (2001), Linking remote-sensing estimates of land cover and census statistics on land use to produce maps of land use of the conterminous United States, Global Biogeochem. Cycles, 15(3), 673–685, doi:10.1029/2000GB001299.

GLOBAL BIOGEOCHEMICAL CYCLES, VOL. 15, NO. 3, PAGES 673–685, SEPTEMBER 2001

Linking remote-sensing estimates of land cover and census statistics on land use to produce maps of land use of the conterminous United States

George C. Hurtt, Lynn Rosentrater, Steve Frolking, and Berrien Moore III

Complex Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire Durham

Abstract. Human use of the land has a large effect on the structure of terrestrial ecosystems and the dynamics of biogeochemical cycles. For this reason, terrestrial ecosystem and biogeochemistry models require moderate resolution (e.g., $\leq$0.5°) information on land use in order to make realistic predictions. Few such data sets currently exist. To create a land use data set of sufficient resolution, we developed models relating land cover data derived from optical remote sensing and a census database on land use for the conterminous United States. The land cover product used was from the International Geosphere-Biosphere Programme DISCover global product, derived from 1 km advanced very high resolution radiometer imagery, with 16 land cover classes. Land use data at state-level resolution came from the U. S. Department of Agriculture's Major Land Uses database, aggregated into four general land use categories: Cropland, Pasture/Range, Forest, and Other. We developed and applied models relating these data sets to generate maps of land use in 1992 for the conterminous United States at 0.5° spatial resolution.

1. Introduction

Human activity has a significant impact on the landscape [*Turner et al.*, 1990]. Approximately 40% of the Earth's land surface has been transformed by humans from its natural state into agricultural land, urban and industrial areas, or artificial reservoirs [*Kates et al.*, 1990; *Vitousek et al.*, 1997]. Human activity has caused major modifications in the biogeochemical cycles of carbon [e.g., *Houghton et al.*, 1999], nitrogen [e.g., *Smil*, 1999; *Vitousek et al.*, 1999], and water [e.g., *Vorosmarty et al.*, 1997; *Yang et al.*, 1998]. Clearly, it is essential that regional and global-scale terrestrial ecosystem and biogeochemistry models include the effects of human activity in their calculations. To do so, models must include information on the nature, extent, and location of land use activities. For large-scale studies, there are two primary sources of information that can be used: remote-sensing and census-based statistics. These sources have different characteristics and strengths (Table 1). Remote-sensing products generally have remarkable spatial resolution ($\leq$1 km) of land cover (e.g., vegetation type). Census statistics often include extensive information on land use (e.g., type of management). Unfortunately, neither of these sources of information is sufficient. Remote-sensing products focus more on land cover than land use, although important aspects of land use can be inferred from changes in land cover [e.g., *Skole and Tucker*, 1993]. Census statistics are defined for political domains (e.g., county, state, or national), which are usually too spatially coarse and irregular for models. Terrestrial ecosystem and biogeochemistry models would benefit from a database that combined land cover and land use information.

Recent studies have demonstrated that important benefits can be gained by considering both remote-sensing and census-based information together. For example, *Ramankutty and Foley* [1998] combined remote-sensing and census-based information to map the spatial pattern of croplands at 5-min resolution in a way that is generally consistent with census statistics on cropland area. By documenting cropland at high-spatial resolution, this study provided a useful land use product for ecosystem and biogeochemistry models. *Frolking et al.* [1999] aggregated remote-sensing estimates of cropland area in China to the county scale and compared them to Chinese census statistics on cropland area to check for consistency. They found significant discrepancies, attributed to errors in the census, misclassifications by remote sensing, and inconsistencies in spatial resolution.

These two earlier studies considered only a fraction of the land and only a single land use: cropland. There remains both a need and a potential for further connections between remote-sensing products and census statistics. In this paper, we develop models relating a common remote-sensing based land cover product (IGBP-DISCover [*Belward et al.*, 1999; *Loveland et al.*, 1999]) and a census-based land use product that classifies all land with multiple land use categories (Major Land-Uses [*U. S. Department of Agriculture (USDA)*, 1996]. We use these data sets and models to produce maps of estimated land use for the conterminous United States at 0.5° resolution, typical of current ecosystem modeling studies [e.g., *Schimel et al.*, 2000]. We also use them to estimate the land cover composition of other lands, a particularly heterogeneous category in the census data set. Our results are not intended to replace either of the input data sets. Rather, our analysis yields new products with useful information that neither data set contains alone.

2. Methods

2.1. Remote-Sensing Data

The International Geosphere-Biosphere Program's (IGBP) DISCover land cover classification [*Belward et al.*, 1999] is a commonly used remote-sensing based land-cover product [e.g., *Ramankutty and Foley*, 1998]. The product used monthly Normalized Difference Vegetation Index (NDVI) composites (April 1992 to April 1993) from the advanced very high resolution radiometer (AVHRR) and various ancillary data to develop a 1-km resolution

Paper number 2000GB001299.
0886-6236/01/2000GB001299$12.00

Table 1. Characteristics of Land-Cover and Land-Use Products

Land Cover	Land Use
Based primarily on remote sensing data	Based primarily on census data
Typically provides dominant vegetation type within a pixel, with "subclass" details (e.g., evergreen versus deciduous forest)	Dominant use, with variable detail on management
Some land use (e.g., cropland), but little detail on management	Little detail on vegetation sub-class (e.g., "Forest" class may not distinguish between evergreen and deciduous)
High spatial resolution (~1 km^2)	Low spatial resolution (~10^2-10^6 km^2)
Repeatable (i.e., monitoring for change)	Higher quality data will have longer repeat intervals (~5–10 yr)

map of the terrestrial surface with 17 land cover categories. We eliminated the category "Water Bodies" from the classification and spatially aggregated the remaining product to the state level (the resolution of the census data described below), by overlaying a political boundary map and counting pixels in each category within each state. Table 2 lists the categories and brief category definitions of this land cover product. Figure 1a shows the area of land in each of the 16 land cover classes for each state in the conterminous United States.

2.2. Census Statistics

Many sets of statistics on land use are available for the United States, from sources such as the Census of Agriculture, National Agricultural Statistics Service, the Natural Resource Conservation Service, and the Food and Agriculture Organization. Among these, the USDA Major Land-Uses data set [*USDA*, 1996] is unique in that it provides a classification of all land at the state level. While county level statistics on cropland exist, no substate level census product exists that considers all land with multiple land use categories. The Major Land-Uses data set describes five major land use classes: Cropland, Pasture, Forest, Special, and Other. We combined the particularly heterogeneous classes Special and Other into a single category, Other, and used this with the remaining three major land use classes from this data set (see Table 3 for definitions). Figure 1b illustrates the area of land in each of these categories for each state in the conterminous United States.

2.3. Models Relating the Remote-Sensing Product and the Census Statistics

Both the remote-sensing classification and the census categorization are complete descriptions of the land surface in the sense that all land is classified in each product (e.g., Figure 1). However, as Tables 2 and 3 show, the classification systems are different. The land cover product has 16 terrestrial classes, with distinctions based on characteristics of the dominant vegetation such as evergreen/deciduous, or forest/woodland/shrubland. It is based on a 1-km resolution remote-sensing product. The land use data are available in tabular form (state totals) and are aggregated into four broad land use classes. Since the data are based on survey responses, they can reflect a finer resolution view of the landscape. At the same time, because the census data

Table 2. IGBP Discover Land Cover Classification[a]

	Name	Area in United States, 10^6 km^2	Description
1	Croplands	1.144	land covered with temporary crops followed by harvest and a bare soil period
2	Cropland/natural veg. mosaic	1.047	lands with a mosaic of croplands, forests, shrublands, and grasslands in which no one component comprises more than 60% of the landscape
3	Open shrublands	1.244	woody vegetation less than 2 m tall and with shrub cover between 10–60%
4	Savannas	0.030	herbaceous and other understory systems, and with forest canopy cover between 10–30%
5	Grasslands	1.233	lands with herbaceous types of cover
6	Evergreen needleleaf forest	1.184	lands dominated by trees exceeding 2 m and with >60% canopy cover
7	Evergreen broadleaf forest	<0.001	lands dominated by trees exceeding 2 m and with >60% canopy cover
8	Deciduous needleleaf forest	0.000	trees exceeding 2 m and canopy cover >60%, with an annual leaf-on and leaf-off cycle
9	Deciduous broadleaf forest	0.953	trees exceeding 2 m and canopy cover >60%, with an annual leaf-on and leaf-off cycle
10	Mixed forests	0.380	tree communities with interspersed mixtures or mosaics of the other four forest cover types
11	Closed shrublands	0.027	woody vegetation less than 2 m tall and with >60% shrub canopy
12	Woody savannas	0.327	herbaceous and other understory systems, and with forest canopy cover between 30–60%
13	Urban and built-up	0.078	land and covered by buildings and other man-made structures
14	Permanent wetlands	0.001	permanent mixtures of water and herbaceous or woody vegetation that cover extensive areas
15	Snow and ice	<0.001	lands under snow and/or ice cover throughout the year
16	Barren	0.019	exposed soil, sand, rocks, or snow with <10% vegetated cover during any time of the year
	Total area	7.673	

[a] After *Belward* [1996]. The "Open Water" class has been masked out of data.

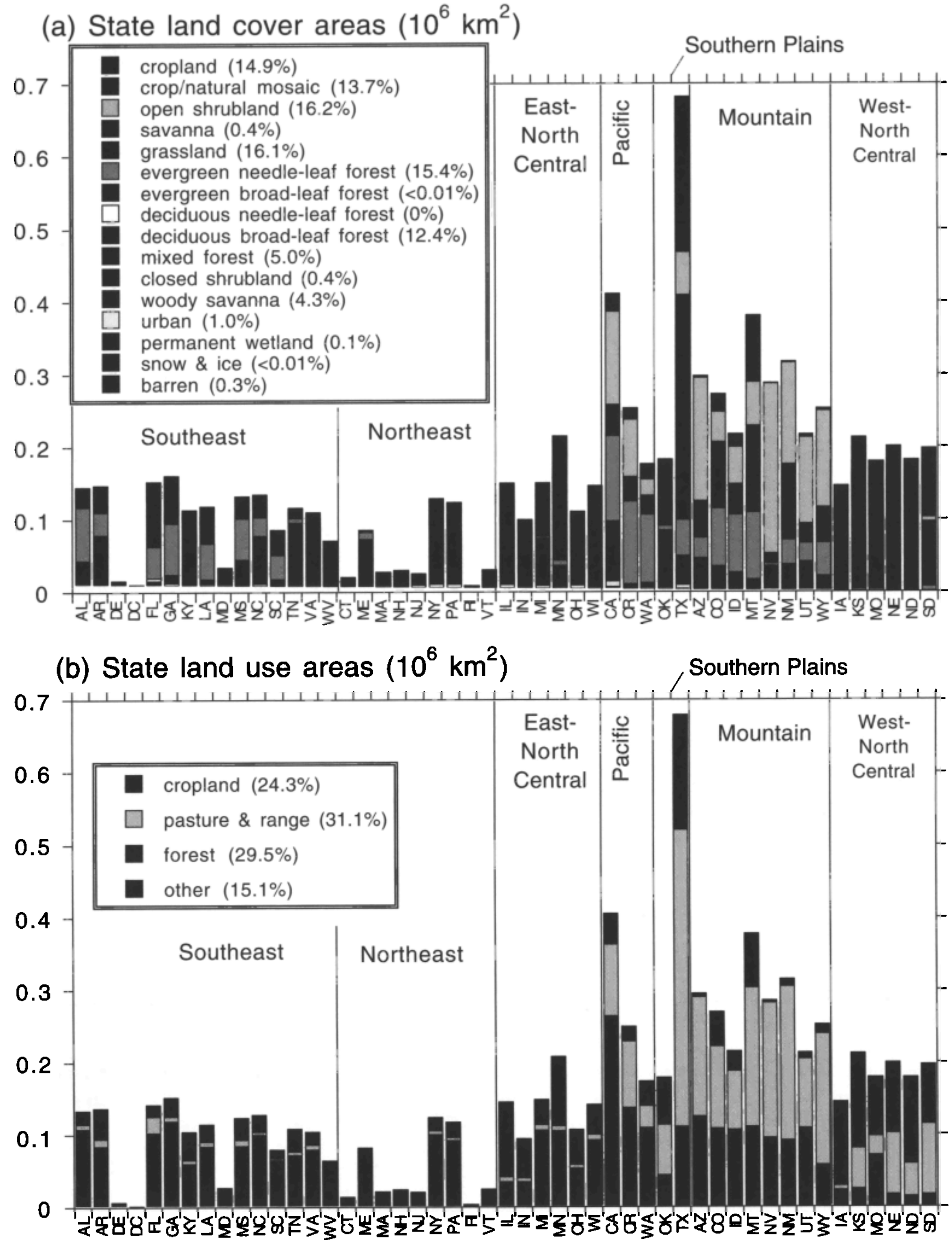

Figure 1. (a) Land cover area in each state, aggregated from the 1992 IGBP DISCover data set [*Belward et al.*, 1999]. (b) Land use area in each state according to the Major Land-Uses data set [*USDA*, 1996]. States are grouped into the seven regions used in the analysis. Values in legends are percent of national area occupied by each class (see Tables 2 and 3).

are nonspatial within states, they provide no information on the spatial pattern of land use within states.

It is possible to develop a direct translation of each land cover class into a single land use category, for example, the land cover Evergreen Needleleaf Forest mapped into the Forest land use. However, in some cases, this is clearly not the best approach. For example, the land cover category Cropland/Natural Vegetation Mosaic should have a fractional mapping into the land use Cropland, and a fraction mapping into the land use category Pasture, and/or Forest, and/or Other. In addition, the land use category Other includes national and state parks, wildlife refuges, and defense lands which might have forest, shrub, or grassland land covers, so some fraction of these land cover classes should be mapped to Other. The most general relationship between these data sets would allow each land cover class to be fractionally mapped into each land use category. We began with this model, and then simplified the model to the minimum complexity that could achieve a high level of correlation between the two landscape descriptions.

For a specified region (a state in our analysis), an estimate of the land area in each land use category can be obtained from the remote-sensing data on land cover as follows:

$$\mathbf{c}^* = A\mathbf{r} \tag{1}$$

Table 3. Main Land Use Categories From the Major Land-Uses Data Set

	Name	Area in United States, 10^6 km^2	Description
1	Cropland	1.860	total cropland: cropland used for crops; cropland idled; and cropland used for pasture
2	Pasture	2.384	grassland pasture and range: grassland and other non-forested pasture and range in farms plus estimates of open or non-forested grazing lands not in farms. Does not include Cropland used for pasture or forest land grazed.
3	Forest	2.261	total forest-use land: forest-use land grazed and forest-use land not grazed
4	Other[a]	1.160	all other land uses: urban, rural transportation, rural parks and wildlife refuges, defense and industrial, plus miscellaneous farm and other uses; unclassified uses such as marshes, swamps, bare rock, deserts, tundra, plus other special uses not estimated, classified, or inventoried
	Total area	7.665	

[a] Land use categories Other and Special from *USDA* [1996] have been combined into the category Other.

In this expression, **c*** is the (4 × 1) vector of estimated land area in each of the four census categories for a state (Table 3); **r** is the (16 × 1) vector of land area in each of the 16 remote-sensing classes for the same state (Table 2). A is a (4 × 16) matrix of coefficients that translate land cover area in **r** to land use area in **c***. All elements of A must be between zero and one ($0 \leq a_{i,j} \leq 1$, for all i and j), and, to conserve area, each column of A must sum to one ($\sum_i a_{i,j} = 1$, for all j).

We first sought a single parameterization of A that simultaneously gives good estimates of **c*** in all states. To obtain parameter values of A, we used maximum-likelihood estimation methods and a simple goodness-of-fit measure that quantifies the difference between estimates of land use for each state (**c***) and state-level data on land use (**c**) (appendix A). The 48 states and the District of Columbia, each with four land use areas, make a data set of 196 land use values.

A is a 4 × 16 matrix and thus potentially has 64 unknown parameters. However, the requirement that all land that was remotely sensed must be mapped into some land use category reduces the number of unknown parameters to 48. In addition, many of the possible relationships between particular remote-sensing classes and land use categories are not likely to be significant (e.g., "Snow and Ice" mapped to "Pasture") and thus many of the remaining parameters in A can be set to zero. To construct an efficient version of A with fewer parameters, we sequentially fit A to the census data and removed unneeded parameters as follows. The first round of parameter estimation resulted in a best fit to all state land use data with the full parameter matrix, i.e., 48 free parameters. This is referred to as the "full model." We then set to zero any parameters estimated to have very small values (<0.001) and set to one any parameters estimated to have very large values (>0.999). This generated a new matrix with fewer parameters. We then estimated a new set of best-fit parameters with this simplified matrix and again eliminated very small and very large free parameters. We repeated this procedure several times; in the final round of parameter elimination we increased the thresholds (to <0.05 and >0.95, respectively). At each stage we checked to ensure that eliminating parameters did not result in a reduction in goodness-of-fit as compared to the full model. Subsequent parameter reduction attempts caused the goodness-of-fit to decline. The resulting simplified model is referred to as the "national model."

Because of the possibility that the relationship between land cover and land use is not constant across the United States, we also estimated a separate set of parameter values with the national model A matrix for each of seven broad regions within the conterminous United States: northeast, southeast, east north-central, west north-central, southern plains, mountain, and pacific. These regions are substantially different in terms of land use and land cover (see Figure 1) and correspond to the regions used in a recent analysis of the effects of land use history on the U.S. carbon budget [*Houghton et al.*, 1999]. This third model is referred to as the "regional model."

3. Results

Seven land cover classes have total areas less than 1% of the total U.S. land area (Table 2, Figure 1a). In the model parameter elimination procedure, all parameters mapping these land cover classes to Cropland, Pasture, and Forest land use categories were very small and thus set to zero. In addition, the fitting procedure mapped urban and built-up land cover completely to the Other land use category, and mapped Grassland land cover completely to Pasture. These nine direct mappings eliminated 27 free parameters. Additional parameter reduction came from nonsubstantial mappings of open shrubland, evergreen needleleaf forest, and mixed forest into cropland; cropland, evergreen needleleaf forest, mixed forest, and woody savanna into pasture; and cropland and open shrubland into forest. These parameters were all set to zero. Thus the national model had 12 remaining parameters to be estimated.

The parameter estimation procedure generated a set of parameter values for A (national model; Table 4) that maximized the correspondence between estimated areas of land use in each category in each state based on remote sensing data and the values provided by the census data set. In addition, using the *Metropolis et al.* [1953] algorithm, we generated a range of parameter values for each parameter with an approximately equal goodness-of-fit; these ranges are also listed in Table 4. Note that a model parameterization with all free parameters set to their minimum or all to their maximum values would not generate a good fit, nor even conserve land area. However, a large number of independent sets of parameter values had approximately equivalent goodness-of-fit (appendix A), and within this set of "good" models, the parameters varied over the ranges shown.

For the national model, correlations between land use area estimates and land use data were Cropland, 0.94; Pasture, 0.99; Forest, 0.93; and Other, 0.95 (Figure 2a). However, small land use areas in states were misestimated by more than 100%. To evaluate the model's performance at larger scales, we aggregated the census data (**c**) and the estimates based on remote sensing (**c***) to regional and national totals and performed additional comparisons. There was closer agreement between the aggregated estimates and census data at the regional scale (Figure 2b), but some discrepancies were still greater than 10%. For national totals (Figure 2c), the difference between estimated values and land use census data were less than 5% for each category. The improved fit at larger scales occurred because many of the discrepancies from the state level compensated for each other in larger-scale summaries.

Table 4. National Model Nonzero Parameter Values, Areas, and Ranges, and Regional Model Parameter Values

			National Model				Regional Model[a]						
Map To	Map From	Param.	Value	Range[b]	Area, 10^6 km^2	Area Range,[b] 10^6 km^2	PAC[c]	MTN[c]	SP	WNC	ENC	SE	NE
Cropland	Cropland	$a_{1,1}$	0.87	0.79–0.94	0.989	0.899–1.073	0.98	0.81	0.97	0.86	1.00	0.53	0.99
	Mosaic	$a_{1,2}$	0.55	0.47–0.65	0.579	0.492–0.678	0.00	0.82	0.28	0.92	0.31	0.08	0.00
	Deciduous broad	$a_{1,9}$	0.19	0.07–0.35	0.177	0.065–0.335	0.63	0.83	0.77	0.13	0.01	0.39	0.15
	Woody savannas	$a_{1,12}$	0.53	0.08–0.82	0.174	0.025–0.268	0.20	0.12	0.51	0.79	0.00	0.90	0.13
Pasture	Mosaic	$a_{2,2}$	0.17	0.05–0.21	0.173	0.054–0.215	0.00	0.14	0.67	0.00	0.08	0.30	0.00
	Open shrubland	$a_{2,3}$	0.76	0.71–0.80	0.944	0.878–1.000	0.71	0.74	0.50	0.88	0.32	0.00	0.00
	Grassland	$a_{2,5}$	1.00	1.00	1.233	1.233	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Deciduous broad	$a_{2,9}$	0.06	0.01–0.23	0.057	0.012–0.217	0.00	0.13	0.13	0.24	0.15	0.05	0.04
Forest	Mosaic	$a_{3,2}$	0.12	0.07–0.33	0.124	0.074–0.342	0.00	0.00	0.03	0.05	0.38	0.40	0.30
	Evergreen needle	$a_{3,6}$	0.88	0.76–0.93	1.039	0.900–1.101	0.89	0.88	0.92	0.00	0.00	0.98	0.99
	Deciduous broad	$a_{3,9}$	0.60	0.36–0.66	0.571	0.339–0.632	0.00	0.00	0.00	0.57	0.52	0.52	0.60
	Mixed forest	$a_{3,10}$	0.85	0.47–0.95	0.324	0.177–0.361	0.03	0.03	0.00	0.19	0.76	0.78	1.00
	Woody savanna	$a_{3,12}$	0.39	0.02–0.85	0.127	0.008–0.279	0.60	0.87	0.43	0.00	0.20	0.00	0.00
Other	Cropland	$a_{4,1}$	0.14	0.06–0.21	0.154	0.070–0.245	0.02	0.19	0.03	0.14	0.00	0.47	0.01
	Mosaic	$a_{4,2}$	0.16	0.02–0.22	0.171	0.019–0.234	1.00	0.04	0.02	0.03	0.23	0.20	0.70
	Open shrubland	$a_{4,3}$	0.24	0.20–0.29	0.299	0.244–0.365	0.29	0.26	0.50	0.12	0.68	1.00	1.00
	Savannas	$a_{4,4}$	1.00	1.00	0.030	0.030	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Evergreen needle	$a_{4,6}$	0.12	0.07–0.24	0.146	0.083–0.284	0.11	0.12	0.08	1.00	1.00	0.02	0.01
	Evergreen broad	$a_{4,7}$	1.00	1.00	<0.001	<0.001	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Deciduous needle	$a_{4,8}$	1.00	1.00	0.000	0.000	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Deciduous broad	$a_{4,9}$	0.15	0.01–0.21	0.147	0.005–0.200	0.37	0.04	0.10	0.05	0.33	0.04	0.20
	Mixed forest	$a_{4,10}$	0.15	0.05–0.53	0.056	0.020–0.203	0.97	0.97	1.00	0.81	0.25	0.22	0.00
	Closed shrubland	$a_{4,11}$	1.00	1.00	0.027	0.027	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Woody savanna	$a_{4,12}$	0.08	0.00–0.43	0.029	0.001–0.141	0.20	0.01	0.07	0.20	0.80	0.10	0.87
	Urban and built-up	$a_{4,13}$	1.00	1.00	0.078	0.078	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Perm. wetlands	$a_{4,14}$	1.00	1.00	0.006	0.006	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Snow and ice	$a_{4,15}$	1.00	1.00	<0.001	<0.001	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	Barren	$a_{4,16}$	1.00	1.00	0.019	0.019	1.00	1.00	1.00	1.00	1.00	1.00	1.00

[a] PAC, Pacific; MTN, mountain; SP, southern plains; WNC, west northcentral; ENC, east northcentral; SE, southeast; NE, northeast. (See footnote to Table 4 for states in each region).

[b] Cells with no range were not free parameters.

[c] Fits in these regions were not improved with regional parameterizations, while in other regions the fits were improved.

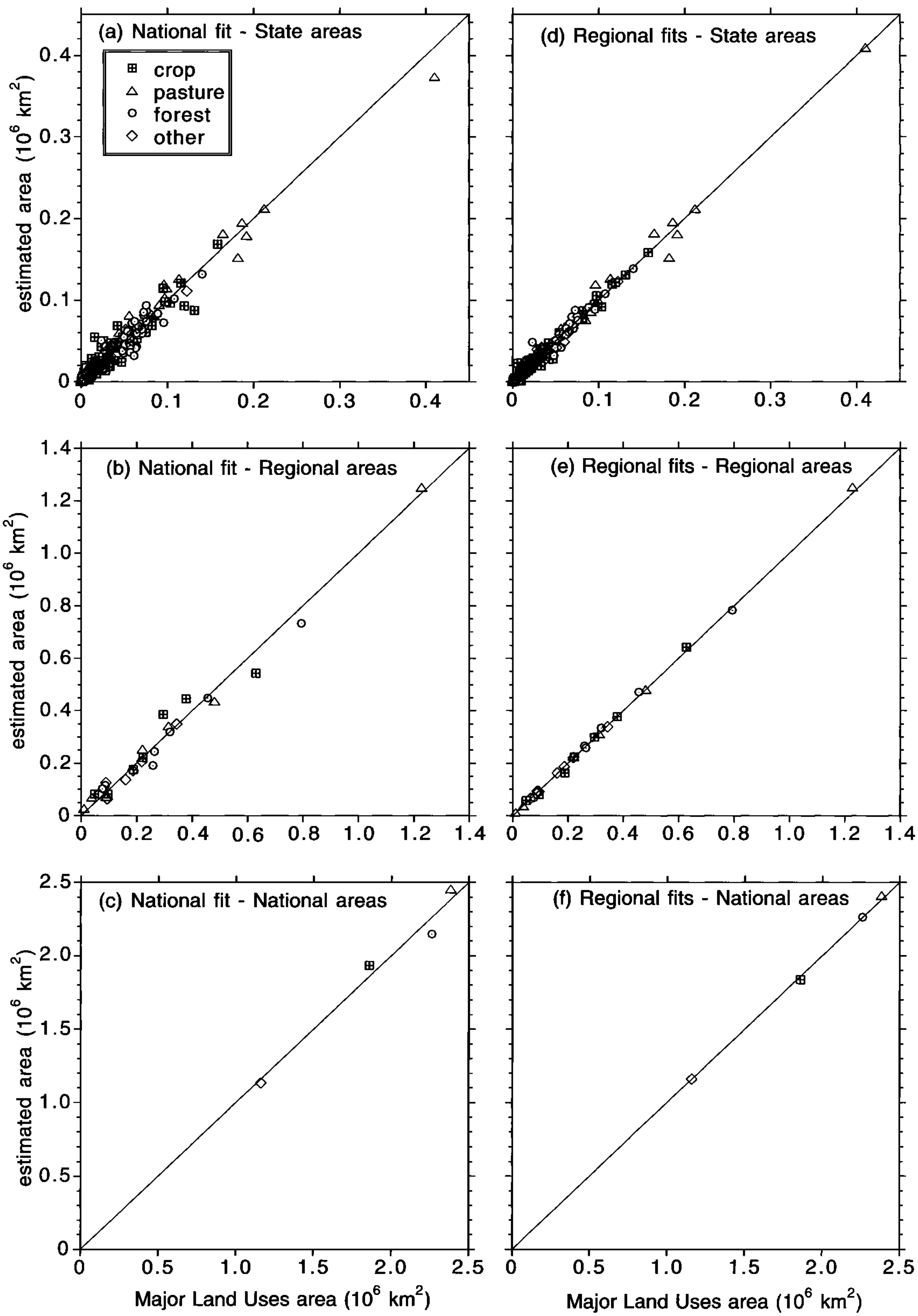

Figure 2. (a) Estimates of land use in each state using the national model of A (y axis) plotted versus the corresponding data from the Major Land-Uses data set (x axis). (b) Estimates in Figure 2a aggregated to the regional level compared to census data aggregated to the regional level. (c) Estimates in Figure 2a aggregated to the national level compared to land use data aggregated to the national level. (d)–(f) Same as Figures 2a–2c, but using the regional model of A. Each panel has a 1:1 line to aid interpretation.

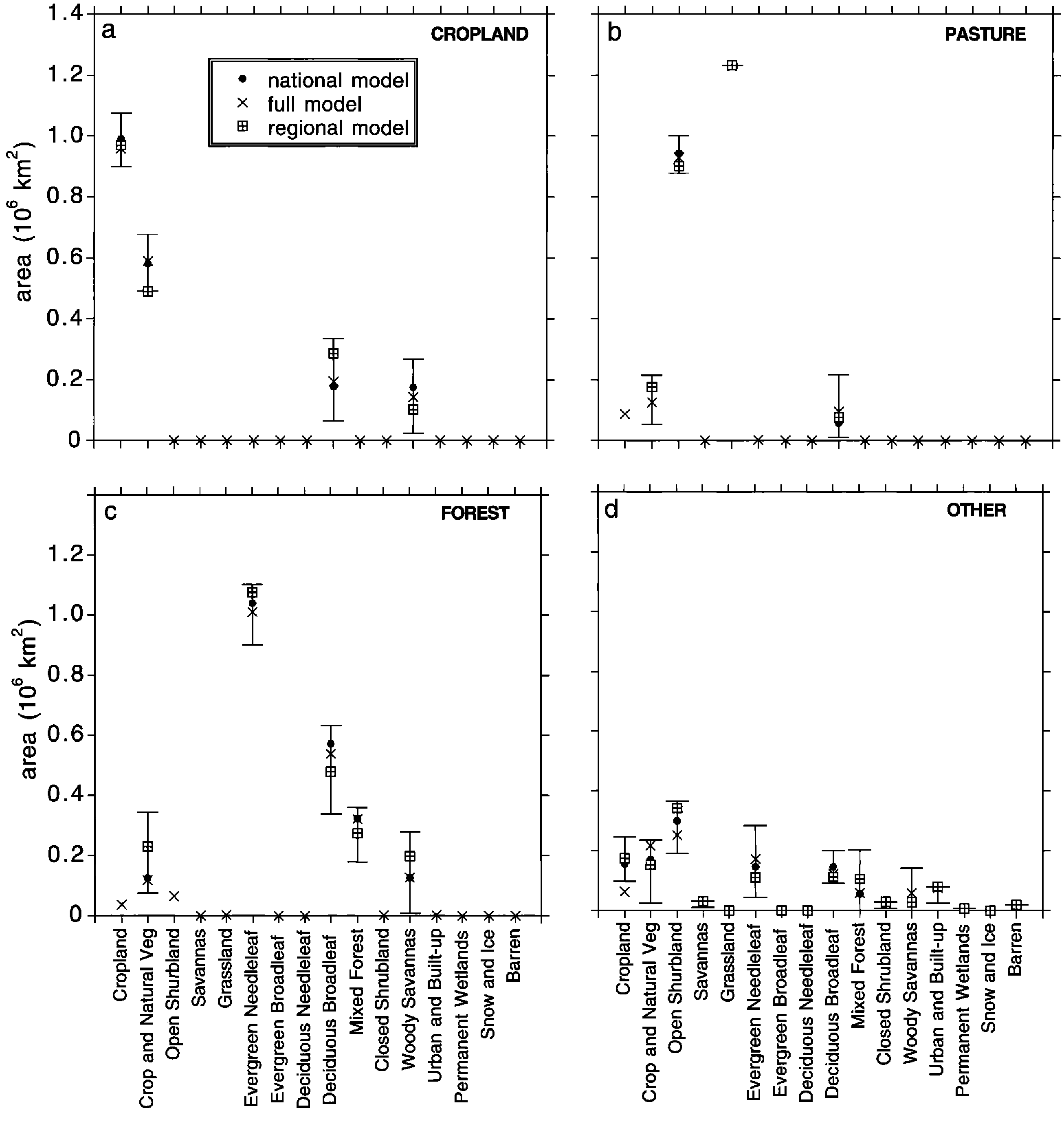

Figure 3. The estimated composition of the major land use classes in terms of remotely sensed land cover. Solid circles are area totals based on best fit values for the national model; uncertainty bounds are area totals based on ranges parameters can have with equal likelihood (see appendix A for discussion and Table 4 for parameter values). Crosses represent area totals based on best-fit parameter values for the full model. Hatched squares represent national area totals based on best-fit parameter values for the regional model (see Table 4).

We also estimated separate parameter values of A for each of the seven regions (Table 4). This version resulted in closer agreement to census data (Figure 2d), with correlations of Cropland, 0.99; Pasture, 1.00; Forest, 0.98; and Other, 0.98. This reflects the fact that five of the seven regions had substantial improvement in the estimates when given a regional parameterization; only the Mountain and Pacific regions were unimproved. The improved estimates at the state level translated into closer agreement between the estimates and census data at regional and national scales (Figures 2e–2f). The improvements were gained by increasing the number of estimated parameters from 12 to 84 (Table 4).

3.1. Cropland

Despite the fact that the area of Cropland in the census is ~60% greater than the area in the Cropland land cover class estimated by remote sensing nationally (Tables 2 and 3), the national model only

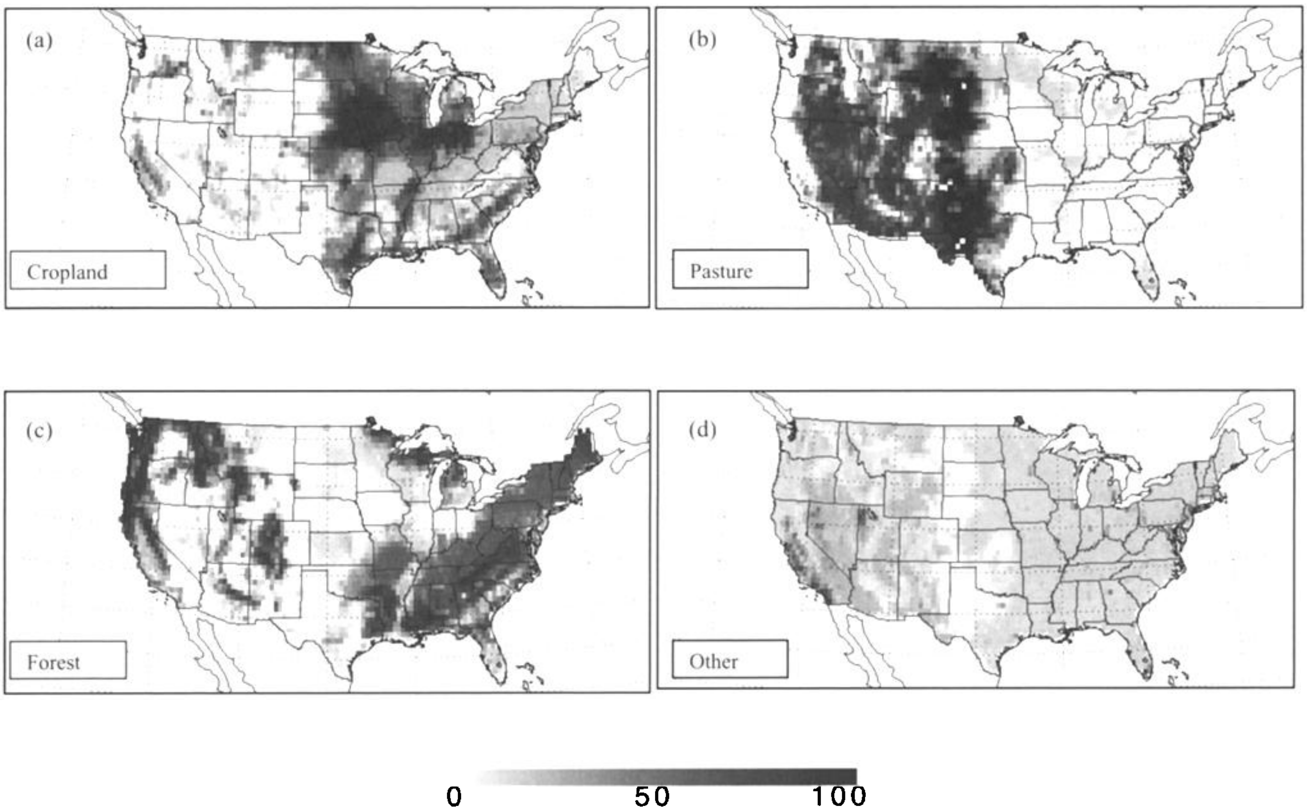

Figure 4. Maps of estimated land use at 0.5° resolution using the national model of A. Each 0.5° × 0.5° grid cell is shaded according to the fraction (0–100%) of that grid cell that is estimated to be in the particular land- use category: (a) Cropland, (b) Pasture, (c) Forest, and (d) Other.

mapped 87% (range, 79–94%) of remotely sensed cropland into the Cropland land use. This was primarily because six states (Alabama, Florida, Geogia, Iowa, Louisiana, and South Carolina) have the opposite pattern, namely more area in the Cropland land cover class than the census reports, and three other states (Indiana, Nebraska, and Ohio) have nearly equal remote sensing and census Cropland areas. Most of these states also have large areas in the Mosaic land cover class, a significant part of which must be mapped to cropland to achieve cropland agreement in other states ($a_{1,2}$ = 0.55; range, 0 47–0.65). Thus the best trade off was to map most, but not all, remotely sensed Cropland to Cropland land use. The estimate of 55% of Cropland/Natural Vegetation Mosaic as Cropland is consistent with the IGBP definition of this land cover category, that no more than 60% of the land cover is cropland (Table 2). Together, the Cropland and Mosaic land cover classes account for 82% of the Cropland land use area. Deciduous broadleaf forest and woody savanna land cover classes each contribute ~9% to Cropland land use (Figure 3a). Six states (Connecticut, Kentucky, New York, Pennsylvania, Vermont, and West Virginia) have ≥80% or their land area classified as deciduous broadleaf forest and require some of this forest to be mapped as cropland ($a_{1,9} > 0$, Table 4) to achieve adequate crop area. Of the six states with ≥10% area in woody savanna (Arizona, California, Kansas, New Mexico, Oklahoma, and Utah, five (all but Kansas) have significantly more Cropland land use than Cropland plus Mosaic land cover, hence some of the woody savanna land cover must be mapped as cropland as well ($a_{1,12} > 0$, Table 4). These parameters had little impact in other states where there is not much Woody Savanna or Deciduous Forest land cover.

The full model generated very similar parameter values for the mapping of Cropland (Figure 3a). However, the regional model had some large differences in parameters values (Table 4) In four regions the Cropland to Cropland mapping parameter ($a_{1,1}$) was ≥98%, while in the Southeast, this parameter dropped to 53% (Table 4). Four states in the Southeast (Florida, Georgia, Lousiana, and South Carolina) have remote sensing Cropland land cover more than 50% greater than the census. This could be due to misclassification of remote sensing imagery, underreporting to the census for land use data, or a mismatch in class definitions (Tables 2 and 3) None of these model parameters caused large differences in the total amount of Cropland mapped nationally (Figure 3a).

3.2. Pasture

The Grassland land cover class was fully mapped into the Pasture land use ($a_{2,4}$ = 1.0; Table 4). In all but two of the 27 states with >5% of their area in Pasture, the area in Grassland is less than the area in Pasture. Also, in the subset of those states with significant open shrubland (all in the mountain and pacific regions), the Grassland land cover area is much less than the Pasture land use area. Grassland and Open Shrubland ($a_{2,3}$ = 0 76, range:0.71–0.80) together contributed ~91% of the Pasture area, with the Mosaic land cover contributing ~7% and deciduous broadleaf forest contributing ~2% (Table 4, Figure 3b). In the full model, there was also a small contribution to Pasture from Cropland (Figure 3b), but the parameter ($a_{2,1}$) was less than 0.05 and was eliminated in the reduced models without significant reduction in goodness-of-fit nationally.

3.3. Forest

Two forest land cover classes (evergreen broadleaf and deciduous needleleaf) were mapped not to the Forest land use but to the Other land use (Table 4). However, because of their insignificant areas in the conterminous United States (deciduous needleleaf has

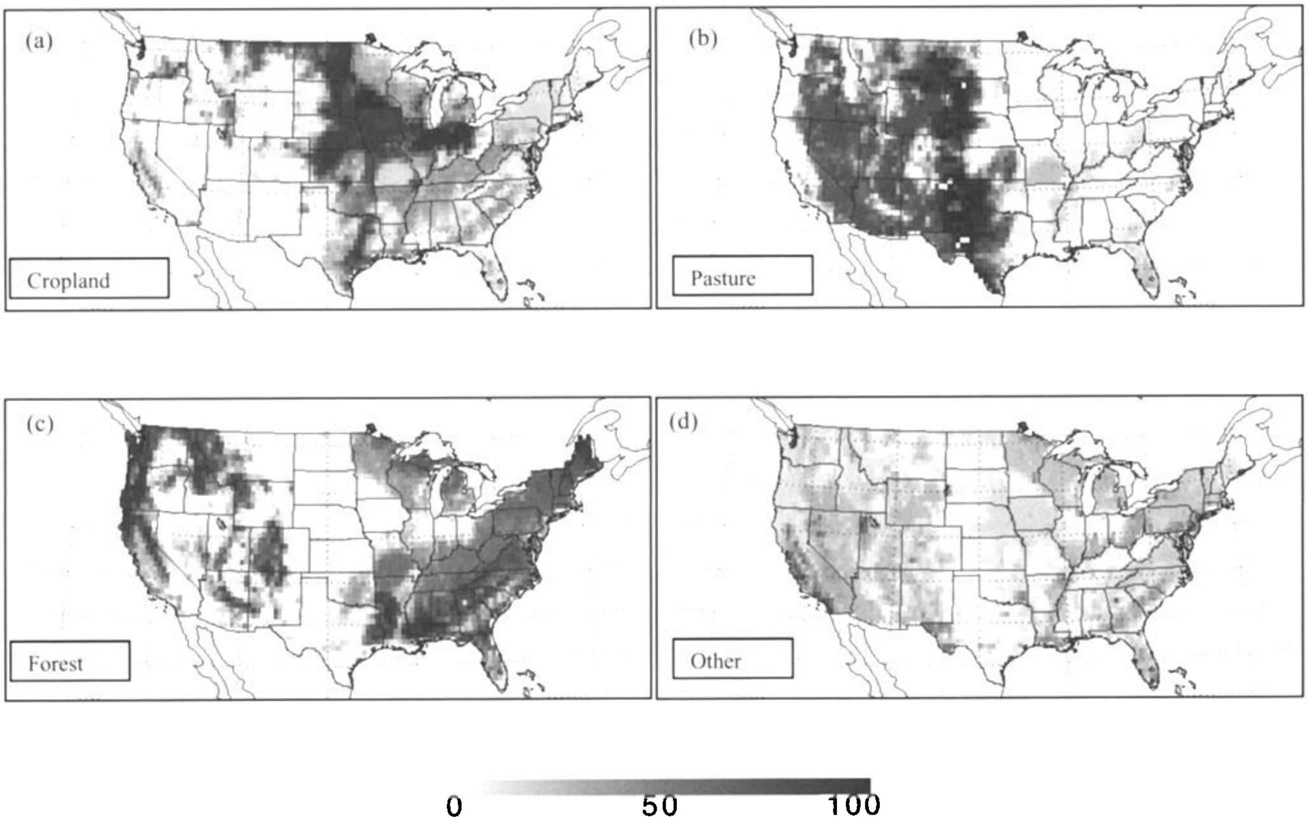

Figure 5. Maps of estimated land use at 0.5° resolution using the regional model of A. Each 0.5° × 0.5° grid cell is shaded according to the fraction (0–100%) of that grid cell that is estimated to be in the particular land- use category: (a) Cropland, (b) Pasture, (c) Forest, and (d) Other.

zero area), this has no real effect on the overall goodness-of-fit of the model. About half the area mapped to Forest land use came from evergreen needleleaf forest ($a_{3,6}$ = 0.88; range, 0.76–0.93), one-quarter from deciduous broadleaf forest ($a_{3,9}$ = 0.60; range, 0.36–0.66), one-sixth from Mixed Forest ($a_{3,10}$ = 0.85; range, 0.47–0.95), and about 5% each from Mosaic and woody savanna (Table 4, Figure 3c). Again, the full model had small contributions to Forest from Cropland and also Open Shrubland (Figure 3c), but the parameters were less than 0.05 and were eliminated without significant reduction in goodness-of-fit nationally. The seven states with significant Open Shrubland (Arizona, California, Indiana, Nevada, Oregon, Utah, and Wyoming) either have small areas of Forest land use or sufficient Forest and woody savanna land covers to match the Forest land use total.

3.4. Other

All land cover classes except Grassland contributed to the Other land use category (Table 4, Figure 3d), though not all land cover classes occur in all states (Figure 1a). Because of the very broad definition of Other land use in the Major Land-Uses database (Table 3), only for Washington, D.C., does the area of the combined urban, wetland, snow and ice, and barren land cover classes exceed 30% of the area of the Other land use. In 25 states these four land cover classes supply <10% of the necessary Other land use area. Most national-model parameters mapped 10–20% of the major land cover classes to the Other land use category (Table 4) Full-model parameters were similar (Figure 3d), except that Cropland to Other ($a_{4,1}$) was smaller because in the full model Cropland was also mapped to Pasture and Forest In the regional model case, 10 of the parameters mapping to Other took on high values (>0.80), but they were always associated with relatively small areas (<1% of a regional area).

3.5. Maps of Land Use

The results to this point have focused on scales ranging from states to national totals. Close agreement between remotely sensed land cover and census statistics on land use can be achieved at these scales with simple models. In this section, we use the models that we developed (As) to produce maps of estimated land use with relatively high-spatial resolution (0.5°). To produce these maps, we aggregated the remote-sensing data to 0.5° and applied the national model (Figure 4) and the regional models (Figure 5) to generate maps of estimated land use at 0.5° resolution. Each map uses a gray scale to depict the fraction of each 0.5° grid cell that is estimated to be in each of the four land use categories.

The maps illustrate many recognizable features and are similar to one another. In both, Forest dominates the east and areas of the pacific and mountain west. Pasture is dominant in much of the mountain region and southern plains states. Cropland predominates in the west north central and east north central regions. What is also clear from both maps is that most grid cells are heterogeneous and composed of more than one land use. Subgrid-scale mixtures of Forest, Crop, and Other occur in the eastern states, mixtures of Crop and Other in the north central states and mixtures of all four land uses in the west. There are differences between these maps, however. Perhaps the most visible differences are the facts that the regional model estimated less Cropland and more Forest and Other in the coastal Southeast, the southern Mississippi Valley, and the

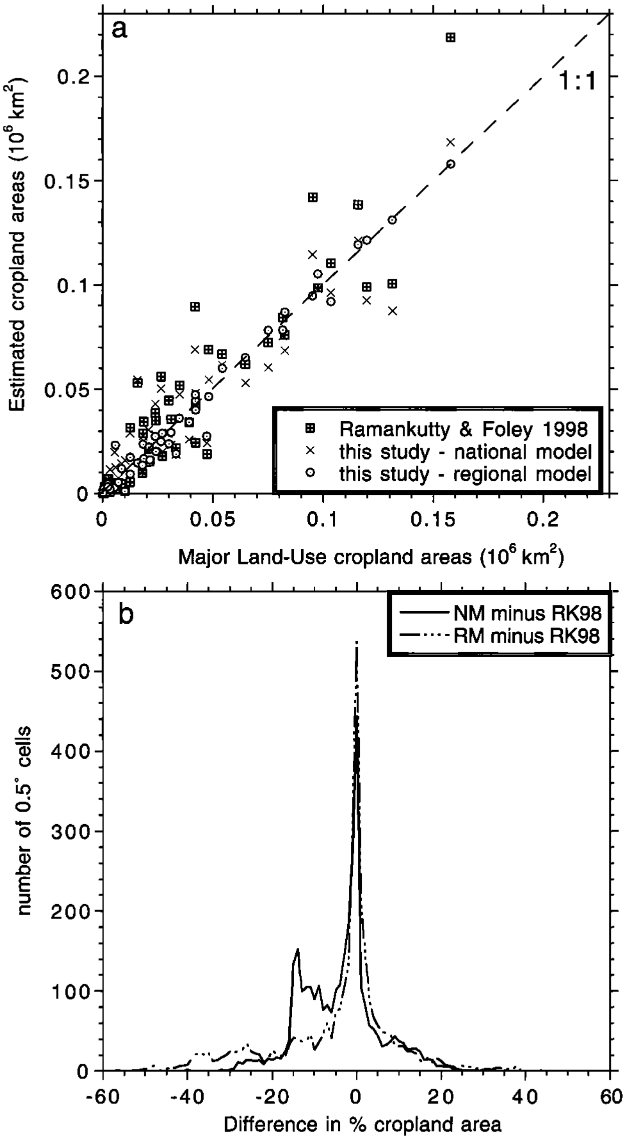

Figure 6. (a) Scatter plot comparing Major Land Uses Database [*USDA*, 1996] cropland area for each state with estimates using the national model, the regional model, and the estimate of *Ramankutty and Foley* [1998]. (b) Histogram of differences in percent cropland area for each 0.5° cell in the conterminous United States. NM, national model; RM, regional model; RK98, *Ramankutty and Foley* [1998]; see text for details. The high peak in the histogram at low discrepancy is dominated by cells with little or no cropland area in both estimates.

southern parts of Michigan, Wisconsin, and Minnesota (Figure 5). The regional model map also shows a greater density of Other land use in western Texas.

3.6. Comparison to Other Products

It is highly desirable to test our estimates of the spatial patterns of land use. However, the lack of sufficiently spatially resolved data on multiple land uses over the United States makes this problematic at present. Even county level census data on Cropland, which do exist, are not straightforward to use as a check on our substate estimates of Cropland because of corrections and adjustments made at the state level when producing the Major Land Uses data set (M. Vesterby, personal communication, 1999). It is possible, however, to compare our estimates of Cropland to corresponding estimates from the study of *Ramankutty and Foley* [1998]. We emphasize that this comparison provides interpretation for just one of our four land use estimates.

Ramankutty and Foley [1998] generated a global map of cropland at 5-min resolution using a procedures comparable to the procedures used here. For North America they aggregated the 205 land cover classes of the DISCover 1-km SLCR land cover product into six categories of cropland density. Using the Major Land-Uses data for United States, provincial crop data from Statistics Canada, and FAO data for Mexico and Central America, they found the best fit for the fractional cropland cover values (equivalent to elements in the first row of our A matrix but including data for Canada, Mexico, and Central America). The parameters for the fraction of cropland were Crops (1.0; not a free parameter), Crops with Other Vegetation (0.75), Crop/Other Vegetation mosaic (0.65), Other Vegetation/Crop mosaic (0.65), Other Vegetation with Crops (0.25), and Other Vegetation (0.0, not a free parameter). They used these parameters to generate maps of Cropland land use from remote sensing data, at 5-min and 0.5° resolution. We acquired their 0.5° resolution map for the conterminous United States for comparison with our results (Figure 6; Table 5). The two methods estimate fractional cropland area in 0.5° cells to within 10% of each other for 72% (national model) or 67% (regional model) of the cells in the conterminous United States (Figure 6b). Our national-model estimate consistently differs from their estimate in cells that are remotely sensed as pure (or nearly pure) cropland because our national model estimates that only 87% of pure cropland pixels are actually cropland. In the regional model estimate, this difference is less pronounced because in many regions the relevant parameter ($a_{1,1}$) is estimated to be near 100% (Table 4). However, other large discrepancies exist. Regionally, the differences between our estimates and that of *Ramankutty and Foley* [1998] are perhaps most pronounced in the southeast and east north-central, where Ramankutty and Foley estimated a much larger area in cropland than either our model or the Major Land-Uses data set (Table 5).

3.7. Interpretation of Lands in the Other Land Use Category

Other land is a very heterogeneous land use category that occupies a large fraction of the country (>15%). Clearly, knowledge of specific land use activities and land cover types present on Other lands is essential to make any use of this category in ecosystem modeling studies. Some additional land use information can be gained from the Major Land Uses data set by disaggregating Other into its few component subcategories. However, the land use data set says nothing about the land cover on those lands.

Our models can be used in a secondary analysis to estimate the land cover associated with each land use, perhaps most importantly Other land. Just as our models estimated that the amount of Other land varies spatially because the land cover information used to estimate the presence of Other lands varies at these scales, our models also estimate that the composition of Other varies spatially. To illustrate this point, we produced maps and corresponding pie charts of the estimated composition of Other lands across the United States (Figures 7 and 8). Forested regions in the northeast and northern southeast are estimated to have Other lands which appear forest-like to remote sensing. Generally, nonforested regions in the central and western parts of the country have Other lands which appear nonforested. Concentrations of Other land that are estimated to be nonvegetated can be seen in parts of Utah, Texas, and the urban centers along the northeast coast.

4. Discussion and Conclusions

Because of the large effects of land use on ecosystem structure and dynamics, ecosystem models must incorporate spatially resolved

Table 5. Cropland Area Estimates (10^6 km^2) for Regions in the United States

Region[a]	Major Land-Uses [*USDA*, 1996]	*Ramankutty and Foley* [1998]	This Study: National Model	This Study: Regional Model
Pacific	0.10	0.06	0.08	0.08
Mountain	0.19	0.13	0.18	0.16
Southern plains	0.22	0.28	0.22	0.22
West north-central	0.63	0.60	0.54	0.64
East north-central	0.38	0.53	0.44	0.38
Southeast	0.30	0.41	0.39	0.30
Northeast	0.05	0.07	0.08	0.06
National total	1.86	2.08	1.93	1.84

[a] Pacific: California, Oregon, and Washington; mountain: Arizona, Colorado, Idaho, Montana, Nevada, New Mexico, Utah, and Wyoming; southern plains Oklahoma and Texas; west north-central: Iowa, Kansas, Missouri, Nebraska, North Dakota, and South Dakota; east north-central: Illinois, Indiana, Minnesota, Ohio, and Wisconsin; southeast: Alabama, Arkansas, Delaware, Washington, D.C., Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, South Carolina, Tennessee, Virginia, and West Virginia; northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont.

land use information in order to make realistic predictions. Previous studies that have sought connections between remotely sensed land cover products and census-based land use statistics considered cropland only [*Ramankutty and Foley*, 1998; *Frolking et al.*, 1999]. There are several reasons for this. First, cropland is generally the most intensive level of land management aside from urban/industrial development, i.e., the vegetated land with the highest degree of land use Second, cropland is commonly a class in both land cover products and census statistics. Finally, cropland may have some distinctive characteristics that make it readily detectable by remote sensing. However, considering cropland only has limitations; for many regions the majority of land use is not as cropland. For example, pasture occupies 31% and Other lands, which could have any of a multitude of land uses, occupy more than 15% of the conterminous United States according to the Major Land Uses data set.

In this paper, we present a complete mapping of all land from a common land cover classification to a set of accepted land use categories. Clearly, categorizing all land into multiple land use

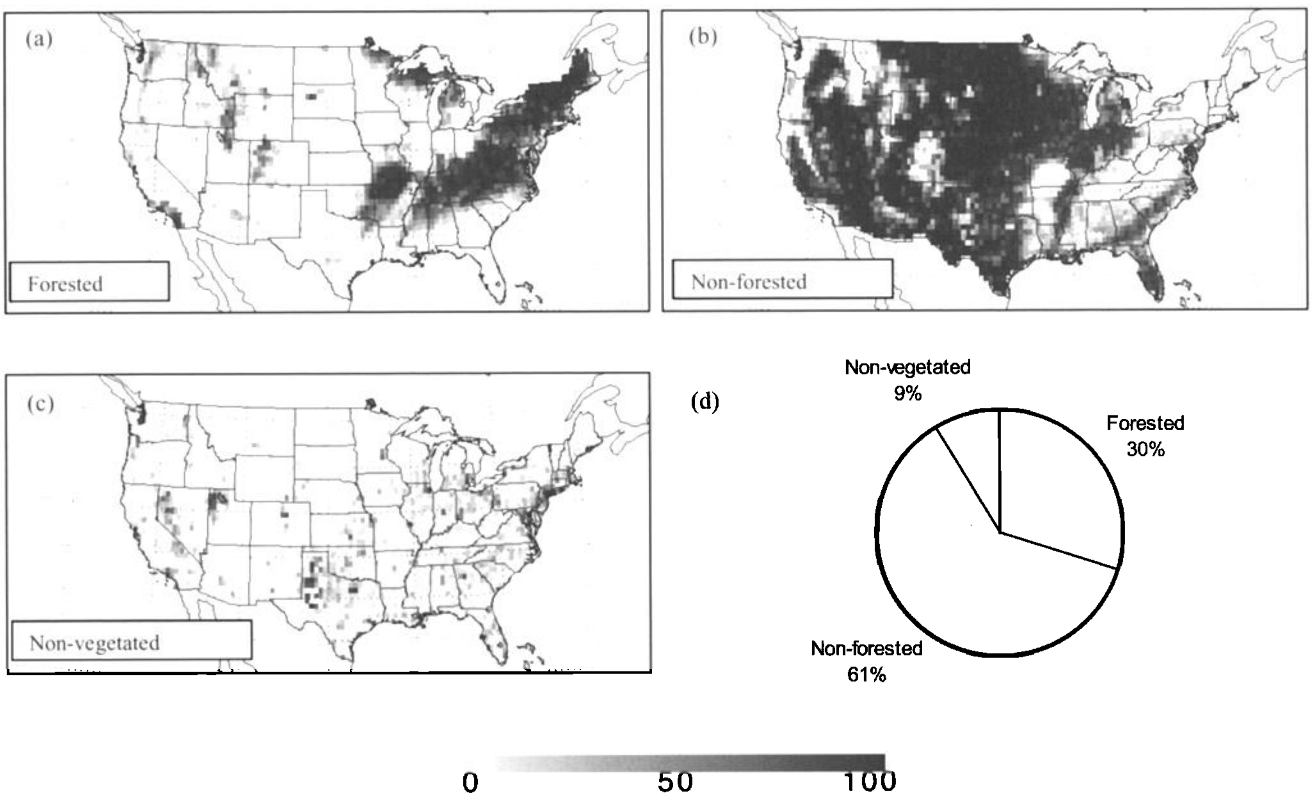

Figure 7. Map of the estimated composition of Other lands in the national model aggregated into three broad categories: (a) forested, (b) nonforested, and (c) nonvegetated. Forested is the sum of the five forested land cover classes in the IGBP classification and Closed Shrubland. Nonvegetated is the sum of urban, barren, and snow/ice classes. Nonforested is the sum of the remaining classes. Each grid cell on this map is shaded according to the fraction (0–100%) of the Other land in the cell that is of the particular type. Note that the land area of each grid cell occupied by Other land use is highly variable, and always less than 100% (see Figure 4d). (d) A corresponding pie chart of the estimated composition of Other land in terms of land cover classes for the coterminous United States

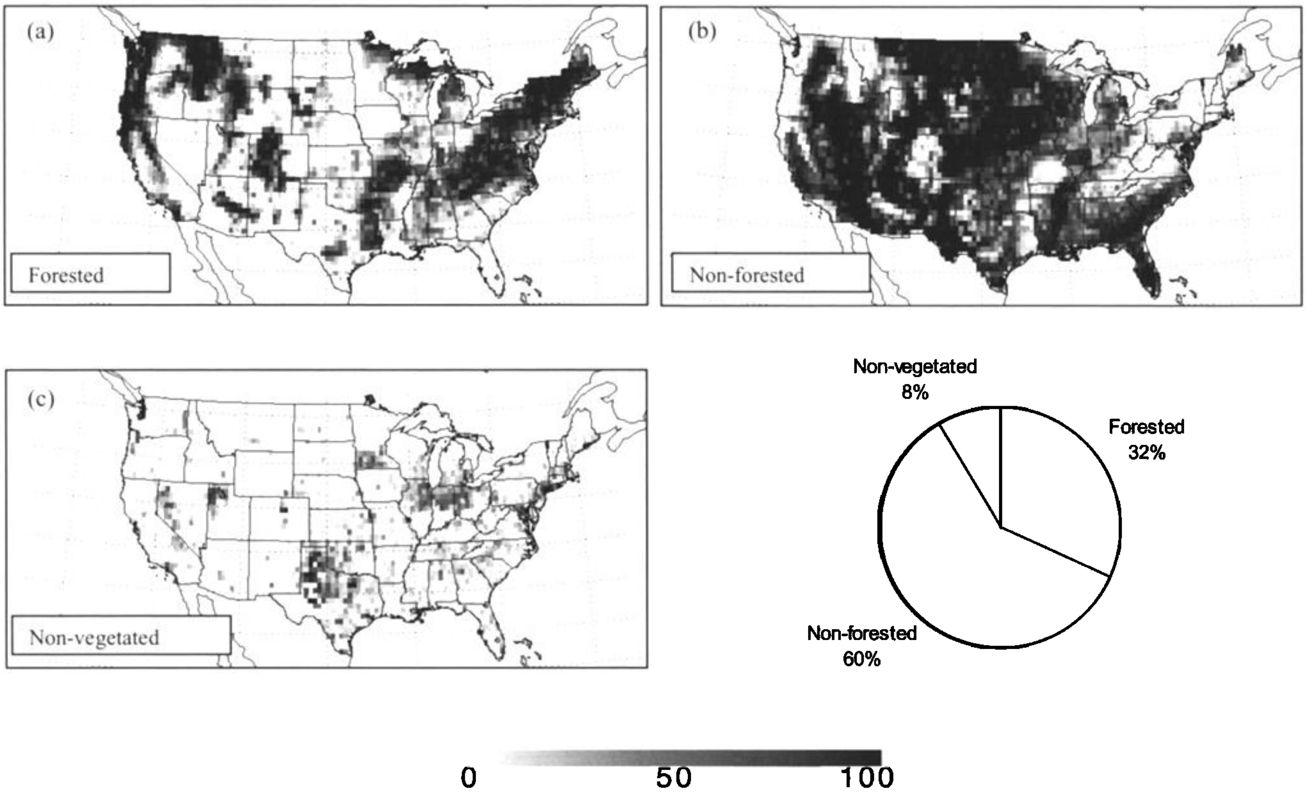

Figure 8. Map of the estimated composition of Other lands in the regional model aggregated into three broad categories: (a) forested, (b) nonforested, and (c) nonvegetated. Forested is the sum of the five forested land cover classes in the IGBP classification and Closed Shrubland. Nonvegetated is the sum of urban, barren, and snow/ice classes. Nonforested is the sum of the remaining classes. Each grid cell on this map is shaded according to the fraction (0–100%) of the Other land in the cell that is of the particular type. Note that the land area of each grid cell occupied by Other land use is highly variable, and always less than 100% (see Figure 5d). (d) A corresponding pie chart of the estimated composition of Other land in terms of land cover classes for the coterminous United States.

categories results in a more complete and useful land use product than does cropland identification alone. Even these heterogeneous categories of land use contain valuable information. They formed a key component of a recent study of the effects of land use history on the U.S. carbon budget [*Houghton et al.*, 1999], a study that relied heavily on the Major Land-Uses data set. They distinguish agricultural lands and separate Cropland and Pasture which have very different management practices. They also identify Forest and Other lands. Forest land is likely to have forest harvesting and other forest management practices. Protected parks and reserves, some of which are forested, have special management issues and are categorized as Other in the census. Although, the Other category is so heterogeneous that it is difficult to interpret on its own for ecosystem modeling, this study provides useful information by estimating the land cover composition of Other land in every grid cell. Clearly more land use information than these broad land use categories provide is needed for ecosystem models; our study has made an enhanced version of this set of land use information available at resolution that is useful for ecosystem and biogeochemistry models.

The methods in this paper included a goodness-of-fit criterion and a parameter estimation procedure to define optimal relationships between a land cover and land use data set at the state scale. Developing models that map land cover areas to land use areas for each state required trade-offs between fitting areas for each of the land use categories in different areas. While the results from this study are encouraging, a perfect fit between the data sets was not possible. There are several possible causes for discrepancies, but it is difficult to discriminate between these causes. There is error in the remote sensing classifications of land cover [*Scepan*, 1999], and there is likely to be error in the census statistics on land use (though we cannot evaluate the magnitude of this). Each of these could lead to difficulty in matching areas. A major cause of the discrepancies may be the heterogeneity inherent in the land cover and land use categories themselves. *Loveland et al.* [1999] noted that significant discrepancies in land area estimates exist between different land cover products and attributed this to different map legends (i.e., different available classes of land cover). This same issue arises in our analysis. Perhaps the largest factor is the fact that the same land cover types can have different uses in different areas For example, land classified as woodland or shrubland land cover might be in the Forest land use, or grazed and classified in the census as Pasture/Range, or set aside as a state or national park and classified as Other.

We developed two sets of maps of land use corresponding to two models relating land use and land cover data: the national model and the regional model. The first uses a single parameter set for the entire country. The second uses the same model form but separate sets of parameter values in each of seven regions. A visual comparison of these maps reveals how similar the large-scale spatial patterns are between the national and regional parameterizations of our model. However, the regional parameterization generally produced a better match to the land use statistics at state, regional, and national scales (Figure 2, Table 4) but at the expense of having many additional parameters. The overall improvement of the regional parameterizations is likely due to both the variability in dominant land cover types across the United States (see Figure 1) and also variability in how particular land cover types are used. The fact that regional parameterizations of the Mountain and Pacific regions did not lead to substantial improvement in the fit suggests that there may be different uses of important land cover classes within those ecologically diverse regions. Alternatively, there may be errors in either the census or remote sensing products at a subregional scale, which preclude a common good-fit mapping between the land cover and land use products for those two regions.

The major result of this study is a set of maps of land use in multiple categories at 0.5° resolution (Figures 4 and 5). It is anticipated that these maps will be useful for new ecosystem modeling studies working to incorporate the effects of land use [e.g., *Hurtt et al.*, 1998; *Schimel et al.*, 2000]. However, these maps should be used with caution until tested with independent fine-scale information. We compared our estimates of cropland at the 0.5° scale to those of *Ramankutty and Foley* [1998] and found both broad regions of agreement and substantial differences between our estimates in some locations. We know that our estimates of Cropland are generally as close or closer to the Major Land Uses data set (Figure 6, Table 5), but we do not know which if either data set is accurate at finer spatial scales. We also do not have independent information on the accuracy of the Major Land Uses data set. What is clear from other studies is that land use is having a major impact on terrestrial ecosystems, and ecosystem models attempting to address this must have spatially resolved information on that land use. The study is a nascent attempt to provide such information.

Appendix A

To estimate parameter values in A (see (1) in text), we implemented a version of the Metropolis simulated annealing algorithm [*Metropolis et al.*, 1953; *Press et al.*, 1992) to find parameter values that maximize the simple log-likelihood (l) equation

$$l = \sum_{s=1}^{N} \sum_{i=1}^{4} -\ln\sigma - \frac{(c^*_{i,s} - c_{i,s})^2}{2\sigma^2}. \qquad (A1)$$

In this equation, s is an index for each state, N is the number of states in the nation or a particular region, i is an index for each census land use category, $c^*_{i,s}$ is the estimate of land area in a census category (equation (1)), $c_{i,s}$ is the actual area of land in census category as reported in the Major Land Uses data set, and σ^2 is the estimated variance. For the national model, N equals 49. For the regional model, there are separate values of N and σ^2 for each region. We emphasize that this log-likelihood metric is used only as a "goodness-of-fit" metric and not as a formal statistic. To produce ranges of "approximately equal" fit for the national model, we recorded sets of parameters within 21 log-likelihood points (α = 0.05; 12 degrees of freedom) of the best set of parameter values [*Brower and Zar*, 1984].

Acknowledgments. We gratefully acknowledge financial support from NASA and NOAA. In addition, ~20% of this research was funded by the National Institute for Global Environmental Change through the U.S. Deparment of Energy (Cooperative Agreement DE-FC03-90ER61010). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the DOE. We thank Xiangming Xiao and Michael Routhier for assistance with GIS. Marlow Vesterby (USDA-Economic Research Service) provided thoughtful discussions on land use data. We thank Steve Pacala and three anonymous referees for providing suggestions that improved this manuscript. We thank the UNH ESIP project for posting the model results at www.esip.unh.edu.

References

Belward, A. S., The IGBP-DIS Global 1km Land-cover Data set (DIS-Cover): Proposal and Implementation Plans, *IGBP-DIS Working Pap.13*, 61 pp., Int. -Geosphere Biosphere Prog., Data and Info. Syst., Toulouse, France, 1996.

Belward, A. S., J. E. Estes, and K. D. Kline, The IGBP-DIS global 1-km land cover data set DISCover: A project overview, *Photogram. Engineer. Rem. Sens.*, *65*, 1013–1020, 1999.

Brower, J. E., and J. H. Zar, *Field and Laboratory Methods for General Ecology*, 2nd ed., 226 pp., William. C. Brown, Dubuque, Iowa, 1984.

Frolking, S., X. Xiao, Y. Zhuang, W. Salas, and C. Li, Agricultural land use in China: A comparison of area estimates from ground-based census and satellite-borne remote sensing, *Global Ecol. Biogeogr. Lett.*, *8*, 407–416, 1999.

Houghton, R. A., J. L. Hackler, and K. T. Lawrence, The U. S. carbon budget: Contributions from land use change, *Science*, *285*, 574–577, 1999.

Hurtt, G. C., P. R. Moorcroft, S. W. Pacala, and S. A. Levin, Terrestrial models and global change: Challenges for the future, *Global Change Biol.*, *4*(5), 581-590, 1998.

Kates, R. W., B. L. Turner, and W. C. Clark, The great transformation, in *The Earth as Transformed by Human Action*, edited by W. C. Clark et al., pp. 1–17, Cambridge Univ. Press, New York, 1990.

Loveland, T. R., Z. L. Zhu, D. O. Ohlen, J. F. Brownn, B. C. Reed, and L. M. Yang, An analysis of the IGBP global land cover characterization process, *Photogramm. Eng. Remote Sens.*, *65*, 1021–1032, 1999.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.*, *21*, 1087–1092, 1953.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in C: The Art of Scientific Computing*, Cambridge Univ. Press, New York, 1992.

Ramankutty, N., and J. A. Foley, Characterizing patterns of global land use and analysis of global croplands data, *Global Biogeochem. Cycles*, *12*, 667–685, 1998.

Scepan, J., Thematic validation of high resolution global land cover data sets, *Photogramm. Eng. Remote Sens*, *65*, 1051–1060, 1999.

Schimel, D., et al., Contribution of increasing CO_2 and climate to carbon storage by ecosystems in the United States, *Science*, *287*, 2004–2006, 2000.

Skole, D., and C. J. Tucker, Tropical deforestation, fragmented habitat, and adversely affected habitat in the Brazilian Amazon: 1978–1988, *Science*, *260*, 1905–1910, 1993.

Smil, V., Nitrogen in crop production: an account of global flows, *Global Biogeochem. Cycles*, *13*, 647–662, 1999.

Turner, B. L., W. C. Clark, R. W. Kates, J. F. Richards, J. T. Mathews, and W. B. Meyer (Eds.), *The Earth as Transformed by Human Action*, Cambridge Univ. Press, New York, 1990.

U. S. Department of Agriculture (USDA), Major Land-Uses (1945–1992), Tech. Rep., Washington, D.C., 1996. (Available at http://usda.mannlib.cornell.edu/data-sets/land/89003)

Vitousek, P. M., H. A. Mooney, J. Lubchenco, and J. M. Melillo, Human domination of the Earth's ecosystems, *Science*, *277*, 494–499, 1997.

Vitousek, P. M., J. D. Aber, S. E. Baley, R. W. Howarth, G. E. Likens, P. A. Matson, D. W. Schindler, W. H. Schlesinger, and G. D. Tilman, Human alteration of the global nitrogen cycle: Causes and consequences, *Ecol. Iss.*, *1*, 1–15, 1999.

Vorosmarty, C. J., K. P. Sharma, B. M. Fekete, A. H. Copeland, J. Holden, J. Marble, and L. A. Lough, The storage and aging of continental runoff in large reservoir systems of the world, *Ambio*, *26*, 210–219, 1997.

Yang, Z. S., J. D. Milliman, J. Galler, J. P. Liu, and X. G. Sun, Yellow River's water and sediment discharge steadily decreasing, *Eos Trans. AGU*, *79*, 589–592, 1998.

S. Frolking, G. C. Hurtt, B. Moore III, and L. Rosentrater, Complex Systems Research Center, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, NH 03824. (george.hurtt@unh.edu)

(Received May 11, 2000; revised September 11, 2000; accepted September 21, 2000.)