Remote Sensing of Environment 156 (2015) 418–425

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A model for the propagation of uncertainty from continuous estimates of tree cover to categorical forest cover and change

Joseph O. Sexton [a,*], Praveen Noojipady [a,b], Anupam Anand [a,c], Xiao-Peng Song [a], Sean McMahon [d], Chengquan Huang [a], Min Feng [a], Saurabh Channan [a], John R. Townshend [a]

[a] Global Land Cover Facility, Department of Geographical Sciences, University of Maryland, College Park, MD 20742, USA,
[b] National Wildlife Federation, National Advocacy Center, Washington, DC 20004, USA
[c] Global Environment Facility, Washington, DC 20433, USA
[d] Smithsonian Environmental Research Center, Edgewater, MD 21037, USA

ARTICLE INFO

Article history:
Received 29 July 2014
Received in revised form 26 August 2014
Accepted 30 August 2014
Available online 11 November 2014

Keywords:
Forest
Change detection
Uncertainty
Propagation
Tree cover
Continuous fields
Landsat

ABSTRACT

Rigorous monitoring of Earth's terrestrial surface requires mapping estimates of land cover and of their errors in space and time. Estimation of error in land-cover change detection currently relies heavily on external, post hoc validation—i.e., comparison of estimated cover to independent values that are assumed to be true. However, reference data are themselves uncertain, and acquiring observations coincident with historical data is often impossible. Complementarily, modeling the transmission, or propagation, of error through the processes of classification and change detection provides an internal means to estimate classification and change-detection error at the scale of pixels. Modeling uncertainty around the estimate of fractional, "continuous-field" cover as a standard Normal distribution in each pixel at each of two times, we derive a method for propagating this uncertainty to categorical land cover-classification and change detection. We demonstrate the approach for mapping forest-cover change and its uncertainty based on bi-temporal estimates of percent-tree cover and their associated root-mean-square errors (RMSE). The method described here propagates only the imprecision component of error and not bias, so neither the resulting categorical estimates of cover nor the detection of change (e.g., forest loss) are affected by the transmission of uncertainty. However, propagating the RMSE of input estimates into probabilities of forest cover and change enables mapping and visualization of the spatial distribution of the imprecision resulting from model-based estimation of tree cover and from selection of the threshold of tree cover to define "forest". When compared to reference data with a fixed definition of forest (e.g., ≥30% tree cover) the threshold effect is an importance source of apparent error in forest-cover and -change estimates. The approach described here provides a useful description of classification and change-detection certainty and can accommodate any definition of forest based on tree cover—an especially important consideration given the variety of institutional definitions of forest cover based on remotely sensible structural characteristics.

1. Introduction

1.1. Background

Sustaining the welfare of a growing human population in a changing environment is dependent on regular and reliable ecosystem monitoring (Sexton, Urban, Donohue, & Song, 2013; Townshend & Brady, 2006). To this end, a growing number of remotely sensed datasets representing Earth's land cover now span multiple observations over time. However, error accompanies all inferences, and so rigorous land-cover monitoring must be based on maps of land cover and change accompanied by estimates of their errors (Congalton & Green, 2009; Foody, 2002; Heuvelink, Burrough, & Stein, 1989; Stehman, 2000).

Consistent with the Inter-governmental Panel on Climate Change IPCC (2006), we define *error* as the inverse of *truth*, or the degree to which a set of values differs from reality. We further partition the concept of error into systematic, i.e., *inaccuracy* or *bias*, and unsystematic, or random, error—i.e., *imprecision* (Willmott, 1982); we treat uncertainty as synonymous with imprecision. To date, error estimation in land-cover mapping and change detection has employed predominantly *validation*—i.e., post hoc comparison of estimates to external sources of reference (Congalton & Green, 2009). When based on a rigorous sampling design, remote sensing validation falls within the general statistical framework of design-based inference (Foody, 2002; Gregoire 1998; Stehman, 2000). Given a lack of error in the reference data themselves, validation can provide estimates of both accuracy and precision

* Corresponding author. Tel.: +1 301 405 8165.
E-mail address: jsexton@umd.edu (J.O. Sexton).

http://dx.doi.org/10.1016/j.rse.2014.08.038

(Willmott, 1982). However, the acquisition of accurate reference observations is an expensive—and itself often uncertain—endeavor (Berger, Gschwantner, McRoberts, & Schadauer, 2014; Breidenbach, Antón-Fernández, Petersson, McRoberts, & Astrup, 2014; Foody, 2002), and the necessarily sparse samples it yields often support only broadly aggregated, regional summaries of error (Fisher, Hurtt, Thomas, & Chambers, 2008; Foody, 2002).

Errors in remotely sensed data vary in space and time, and so description of these errors must likewise strive to reflect this complexity (Steele, Winne, & Redmond, 1998). The proper scale at which to infer error in land-cover estimates is thus equivalent to that of cover itself—i.e., at pixel resolution and extent. As an alternative to *design*-based inference, *model*-based inference (Gregoire 1998) has been used to map the estimated certainty of static land-cover categories (e.g., Liu, Gopal, & Woodcock, 2004; McRoberts, 2006; Steele et al., 1998) and of their changes over time (e.g., McRoberts & Walters, 2012). Further, the development of multi-temporal datasets representing continuous biophysical attributes (DeFries, Field, Fung, & Justice, 1995, DiMiceli et al., 2011, Hansen et al., 2011; Sexton, Song, Feng, et al., 2013; Sexton, Song, Huang, et al., 2013) and their increasing use for mapping and change detection (e.g., Hansen et al., 2013; Hansen, Stehman, & Potapov, 2010; Huang et al., 2010; Kennedy, Yang, & Cohen, 2010) prompt the development of a rigorous approach to categorical change detection based on multi-temporal continuous fields.

Errors arise both from models and from data, including: (1) the specification and parameterization of models and (2) the spatial and temporal registration, sampling, and measurement of data (Berger et al., 2014; Burnham & Anderson, 2002; Clark, 2007; Heuvelink et al., 1989). Ideally, the effects of all pertinent error sources should be communicated alongside model inferences, including estimates of model parameters and "predicted" cover values. For various components of the total error budget, this is typically accomplished in any of three ways: by sample-based methods (Stahl et al., 2014); by error propagation (Berger et al., 2014; IPCC, 2003; Stahl et al. 2014); and by Monte-Carlo—i.e., "parametric bootstrap"—methods (Breidenbach et al., 2014; Gertner & Dzialowy, 1984; Metropolis & Ulam, 1949).

Error propagation is practiced commonly in allometric estimation of tree and forest attributes and occasionally in land-cover mapping and change detection. Berger et al. (2014) incorporated imprecision from model specification and measurement of covariates into the variance of allometrically estimated tree stem volume. Breidenbach et al. (2014) used Monte-Carlo simulation to quantify model-related variability in biomass stock and change estimates. Based on a logistic regression relating binary forest cover to top-of-atmosphere reflectance, McRoberts (2006) incorporated estimation uncertainty—including spatial autocorrelation in the training data—into the uncertainty of regional forest-area estimates. McRoberts and Walters (2012) used a validation error matrix to illustrate the construction of confidence intervals for net forest-cover loss estimated from maps of forest-probability at two times.

1.2. Objectives

In this paper we derive a model for the propagation of error from fractional, "continuous-field" estimates of cover, through classification of discrete land-cover categories, to post-classification change detection in each pixel. We demonstrate the approach by application to forest-cover change detection in a region of active clearing and regrowth, using tree cover and corresponding uncertainty estimates from a previously published global, Landsat based tree-cover dataset (Sexton, Song, Feng, et al., 2013). Although the method is applicable to any continuous, bi-temporal representation of biophysical attributes (e.g., canopy height or biomass) or land cover (e.g., impervious surface), this development is especially pertinent to mapping forest changes across the wide variety of definitions of "forest" based on remotely sensible characteristics.

Although the method is general, several specific data sources are used here to illustrate our approach, using model-based inference to propagate uncertainty from input estimates of tree cover and using design-based inference to validate the resulting maps of forest cover and change. Input rasters of estimated tree cover and error were taken from a global, percent-tree cover dataset produced at 30-m resolution for circa-2000 and -2005 (Sexton, Song, Feng, et al., 2013). These estimates were produced by an empirical regression-tree model trained on an ensemble of land-cover datasets as the response variable and Landsat-based surface reflectance as covariates. Their errors were estimated in each pixel by propagating the uncertainty from training data relative to lidar-based reference measurements of tree cover. Using design-based inference, an independent reference dataset of visually interpreted observations of binary forest/non-forest cover was used to validate the resulting forest-cover and –change maps.

2. Methods

2.1. Theory

2.1.1. Defining forest cover in terms of tree cover

Define "forest" as a class of land cover wherein tree cover, c, exceeds a predefined threshold value, c^*. The probability of belonging to "forest", $p(F)$, is therefore the probability of c exceeding the threshold c^* (Fig. 1)—i.e., the integral of the probability density function of c above c^*:

$$p(F) \overset{\text{def}}{=} p(c>c^*) = \int_{c*}^{100} p(c)dc. \tag{1}$$

Complementarily, the probability of membership in non-forest is simply $1 - p(F)$.

In any location i, tree cover c_i is commonly estimated by a model f of remotely sensed covariates $\boldsymbol{X}$ (Hansen et al., 2003; Homer, Huang, Yang, Wylie, & Coan, 2004; Sexton, Song, Feng, et al., 2013):

$$c_i = f(\boldsymbol{X};\beta) + \varepsilon_i, \tag{2}$$

where β is a set of parameters, which are estimated empirically, and ε is residual error.

Given a joint sample of locations $i = [1,2,\ldots n]$ with coincident true and estimated values of a continuous variable such as tree cover $(c_i, \hat{c}_i)$, error may be quantified as the root-mean-square error (RMSE), which for large samples approximates the standard deviation of estimates of the true value of cover:

$$\sigma_\varepsilon = \sqrt{\frac{\sum_i (c_i - \hat{c}_i)^2}{n-1}}. \tag{3}$$

Thus, given c_i, and an estimator (e.g., linear regression) producing estimate $\hat{c}_i$ and root-mean-square error $\sigma_i = \sigma$, a Normal probability distribution of possible values of c_i may be assumed (Clark, 2007; Hastie, Tibshirani, & Friedman, 2001; Snedecor & Cochran, 1989):

$$p(c_i) \overset{\text{def}}{=} N\left(\hat{c}_i, \sigma^2\right) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(c_i-\hat{c}_i)^2}{2\sigma^2}}. \tag{4}$$

Given paired estimates of cover and its RMSE, this model provides a probability density function of tree cover $p(c)$ (Eq. (1)) and therefore the probability of identifying forest for each pixel i.

2.1.2. Change detection based on bi-temporal class-probabilities

Given the probability of detecting forest in a location $i = (x,y)$ at each of two times t, four dynamic classes (D) are possible: stable forest (FF), stable non-forest (NN), forest gain (NF), and forest loss (FN).

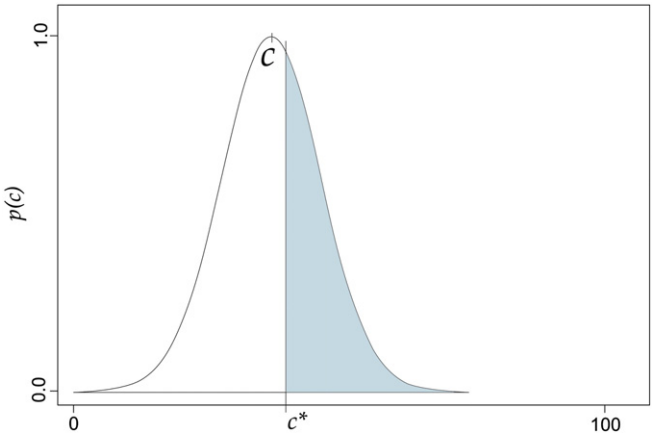

Fig. 1. Estimation uncertainty of tree and forest cover within a pixel, modeled as a standard Normal probability density function of tree cover, with probability of forest (shaded) and non-forest (unshaded) defined relative to a threshold of tree cover, c*.

Calculating the probability of each of these dynamics at that location simply requires calculating the following joint probabilities:

$$p(FF)_i = p\left(F_{i,1}, F_{i,2}\right) = p\left(F_{i,1}\right) \times p\left(F_{i,2}\right) \quad (5)$$

$$p(NN)_i = p\left(N_{i,1}, N_{i,2}\right) = \left(1-p\left(F_{i,1}\right)\right) \times \left(1-p\left(F_{i,2}\right)\right) \quad (6)$$

$$p(NF)_i = p\left(N_{i,1}, F_{i,2}\right) = \left(1-p\left(F_{i,1}\right)\right) \times p\left(F_{i,2}\right) \quad (7)$$

$$p(FN)_i = p\left(F_{i,1}, N_{i,2}\right) = p\left(F_{i,1}\right) \times \left(1-p\left(F_{i,2}\right)\right) \quad (8)$$

where subscripts denote observation times (Fig. 2). In practice, the model of error is approximate, and so carets (^) will be used to denote that the resulting values are estimates. These joint probabilities sum to unity at each location *i*, and because they are merely transformations of the original cover and error values in every pixel, they may be mapped geographically without gain or loss of information from those estimates. In order to produce a categorical map of change classes, each pixel may be assigned either the most probable class at *i*, or some other criterion of probability may be set (e.g., $p \geq 0.9$) to filter detection based on certainty of the tree cover and derived forest-cover and -change estimates.

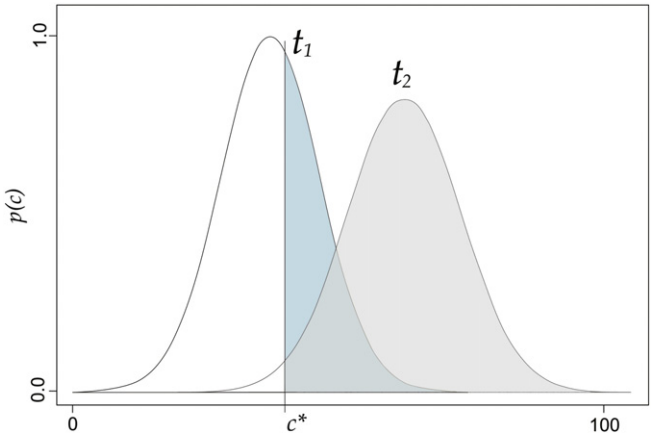

Fig. 2. Categorical (forest) change detection based on probabilistic fields of tree cover at two times, *t*1 and *t*2.

2.2. Study area and data

We applied the method in the path-47, row-27 scene of the Landsat World Reference System 2 (WRS-2) (Fig. 3), using model-based tree-cover and error estimates from a global, 30-m resolution tree-cover dataset for circa-2000 and -2005 epochs (Sexton, Song, Feng, et al., 2013; available from the Global Land Cover Facility (www.landcover.org)) for initial (t_1) and final (t_2) observations, respectively. Tree-cover and error (RMSE) at t_1 and t_2 were estimated based on Landsat image dates of July 30, 2000 and July 23, 2006, respectively.

2.2.1. Input estimates of tree cover and its uncertainty

Input raster maps of tree cover and error in circa-2000 and -2005 were estimated by an empirical regression tree based on atmospherically corrected surface reflectance as covariates and an ensemble training sample of percent tree cover based on the MODIS vegetation continuous fields dataset (DiMiceli et al., 2011; Hansen et al., 2003; Sexton, Song, Feng, et al., 2013). Assuming that tree-cover change in each MODIS pixel was negligible over the period—a reasonable assumption given the rates of canopy closure in the study area—the n = 6 observations of the MODIS VCF from 2000 to 2005 were used as training against each year (2000 or 2005) of the reflectance covariates (**X**). Calculated for each terminal stratum (i.e., "node", or "leaf") of the tree, model imprecision (i.e., residual variance) was assigned to each pixel as the RMSE of tree-cover estimates relative to values of the multi-temporal response ensemble. Model errors were thus assumed to be homoscedastic within each regression stratum and incorporated uncertainty from both the model and the training data. Because they were estimated based on the data used to train the model, these errors were likely under-estimates of true error. Therefore, error (RMSE) of MODIS tree-cover training data (DiMiceli et al., 2011) relative to lidar-based reference measurements was also calculated in a global sample distributed across temperate and tropical forests. Due to scarcity of lidar data, this component of the total error budget was estimated as a global constant: RMSE = 16.83%. Estimates of the total per-pixel uncertainty were mapped by combining these global and local components as their root sum of squares (Sexton, Song, Feng, et al., 2013). Fig. 3 shows the spatial and frequency distributions of model-based tree cover and error estimates across the study scene. Following the International Geosphere-Biosphere Programme definition of forests (IGBP, 1992), the tree-cover threshold used to define forests, *c**, was defined as 30% cover.

2.2.2. Reference observations of forest-cover change

Design-based validation of forest-cover and -change maps employed an independent reference dataset from the North American Forest Disturbance (NAFD) study of the North American Carbon Program (Thomas et al., 2011). Located in the Pacific Northwest region of the United States, the study area includes widespread, intensive timber harvest and regeneration as well as Olympic National Park and the Seattle metropolitan area, where little anthropogenic clearing or regeneration occur. Collection of change-detection reference data was described by Thomas et al. (2011). Assuming the same definition of forests ($c > 30\%$ tree cover), data for the four dynamic classes (FF, NF, FN, and NN) were gathered following a design-based, stratified random sample to increase precision in rare (i.e., change) classes while adjusting for class proportions in the estimation of the error matrix. Strata were defined by persistent-forest and initial disturbance-year classes identified by the Vegetation Change Tracker algorithm (VCT) (Huang et al., 2009). In each sample pixel, trained interpreters visually evaluated time-serial Landsat imagery assisted by high-resolution imagery from TerraServer (www.terraserver.com) and/or Google Earth (www.earth.google.com). Interpreters labeled "change" and "no-change" forest conditions in each pixel using knowledge of the spectral properties, temporal changes, and spatial context of the pixel within the context of the surrounding landscape over time.

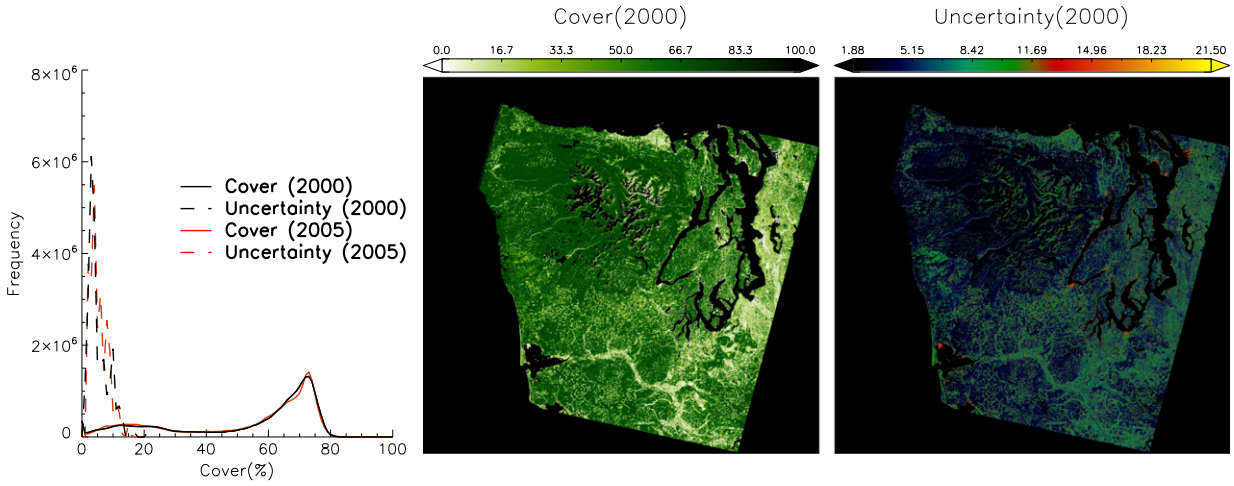

Fig. 3. Frequency and spatial distribution of tree cover and measurement uncertainty (RMSE) in the WRS-2 path-47, row-27 scene. The horizontal (i.e., longitudinal) distance across the image is approximately 180 km.

3. Results

3.1. Distribution of tree-cover estimates and of their error

For each year, the frequency distribution of the input tree-cover estimates ($\hat{c}_i$) was weakly multimodal, with a strong peak near 70% cover, a lesser peak between 10 and 25% cover, and a very small peak at 0% cover (Fig. 1). (This distribution does not represent cover values over water, snow, and ice, or values obscured by clouds or their shadows.) The scene-wide frequency distribution of estimated error (RMSE) was also bimodal, with a dominant peak around 5% and a smaller peak around 10% cover. (Because the error model assumed a mean error of zero, errors depict only varying degrees of imprecision, i.e., uncertainty.) In each year, tree-cover and error estimates were moderately, but significantly negatively correlated ($r = -0.6821$, $p < 0.001$ in 2000 and $r = -0.5439$, $p < 0.001$ in 2005).

The frequency distribution of estimated tree cover ($\hat{c}_i$) was nearly identical among the two years, but the bi-temporal correlation coefficient ($r = 0.7495$, $p < 0.001$) was notably less than unity. Although the deficit ($1 - r = 0.2505$) also included noise in the data, this phenomenon of fine-scale gross dynamics summing to broad-scale net stasis is a characteristic of "dynamic equilibrium", or "shifting mosaic steady state" conditions (Heinselman, 1973; Remmert, 1991). Error was moderately correlated among the two years [($r = 0.5489$, ($p < 0.001$))], with errors slightly larger in 2005. The 30% cover threshold, c^*, was located between the middle and upper modes of the frequency distribution of cover in each year; this association with relatively few pixels led to little error in discriminating forest from non-forest in either year.

Olympic National Park is visible as a large, intact region of certain forest cover in the northwestern portion of the scene. The greater Seattle metropolitan area is visible as a large region of sparse, uncertain tree cover in the northeastern portion of the scene. Uncertainty in these pixels was likely due to spectral mixing of trees with lawns, impervious cover, and other surfaces with widely variable reflectance properties. Across the rest of the scene, where industrial forestry is prevalent, dense tree cover was mixed with patches of sparse to intermediate tree cover. Uncertainty estimates larger than 20% cover were extremely rare, and values between 10 and 20% were clumped in isolated patches near water bodies; further inspection revealed these to be associated with estuaries, which have reflectance characteristics similar to tree cover. Such information is useful to data developers, who would likely focus future algorithm refinements here.

3.2. Distribution of forest-change classes and probability

Estimates of the certainty (i.e., probability) of post-classification change estimates varied in space and among the classes. Among the four classes, stable forest (FF) was associated with highest estimated certainty, with a large proportion of pixels having $\hat{p}(\mathrm{FF}) > 0.9$, especially within the national park (Fig. 4). Stable non-forest (NN) was also associated with high model and measurement certainty, with the largest number of pixels having $\hat{p}(\mathrm{NN}) > 0.9$ located outside the park, especially in valley bottoms and near water where urban and agricultural land uses are prevalent. The large region of highly certain, stable non-forest pixels in the northeastern portion of the scene is the greater Seattle metropolitan area. Mapped with varying certainty, forest-loss (FN) pixels were associated predominantly within areas of industrial forestry located outside the national park and the Seattle metropolitan area. Detections of forest gain (NF) were rarer and less certain, but were patterned as distinct patches interspersed with forest-cover losses in areas of heavy clear-cutting and regrowth, also outside of the park and metropolitan area. Due to the relatively slow and highly variable rates of regrowth in this region (Schroeder, Cohen, & Yang, 2007), the certainty of forest-gain pixels was low compared to the other classes.

Fig. 5 shows the most probable class and its estimated probability in each pixel for a small sub-set of the study scene. Most pixels showed very high certainty ($\hat{p} > 0.9$), and those identified as stable forest (FF) were especially certain. Smaller probabilities were associated with change classes, especially of regrowth and in regions of great spatial heterogeneity. Such spatially distributed assessments of accuracy could be applied to identify and prioritize methodological improvements (Breidenbach et al., 2014), filter or weight data used in subsequent analyses (Kim et al., 2014), or adjust regional summaries of areal coverage (McRoberts, 2006).

3.3. Independent validation of cover estimates

Comparison of the most probable classes to the change-detection (NAFD) reference data revealed a slight deviation toward forest in the underlying tree-cover estimates ($\hat{c}_i$) at t_1 and t_2 (Table 1). Overall accuracy was 95%, with Kappa = 0.88. Stable forest (FF) was mapped with nearly perfect accuracy, with User's accuracy [$\mathrm{UA} =^{\mathrm{def}} \mathrm{p}\left(\mathrm{D}|\hat{D}\right)$] equaling 95% and Producer's accuracy [$\mathrm{PA} =^{\mathrm{def}} \mathrm{p}\left(\mathrm{D}|\hat{D}\right)$] equaling 99%. The majority of commission errors toward FF occurred due to errors at only one time—either as erroneous non-forest in t_1 or t_2, although misclassification of non-forest as forest in both t_1 and t_2 rivaled the rate of this particular type of error in t_1 alone. (Commission error is the

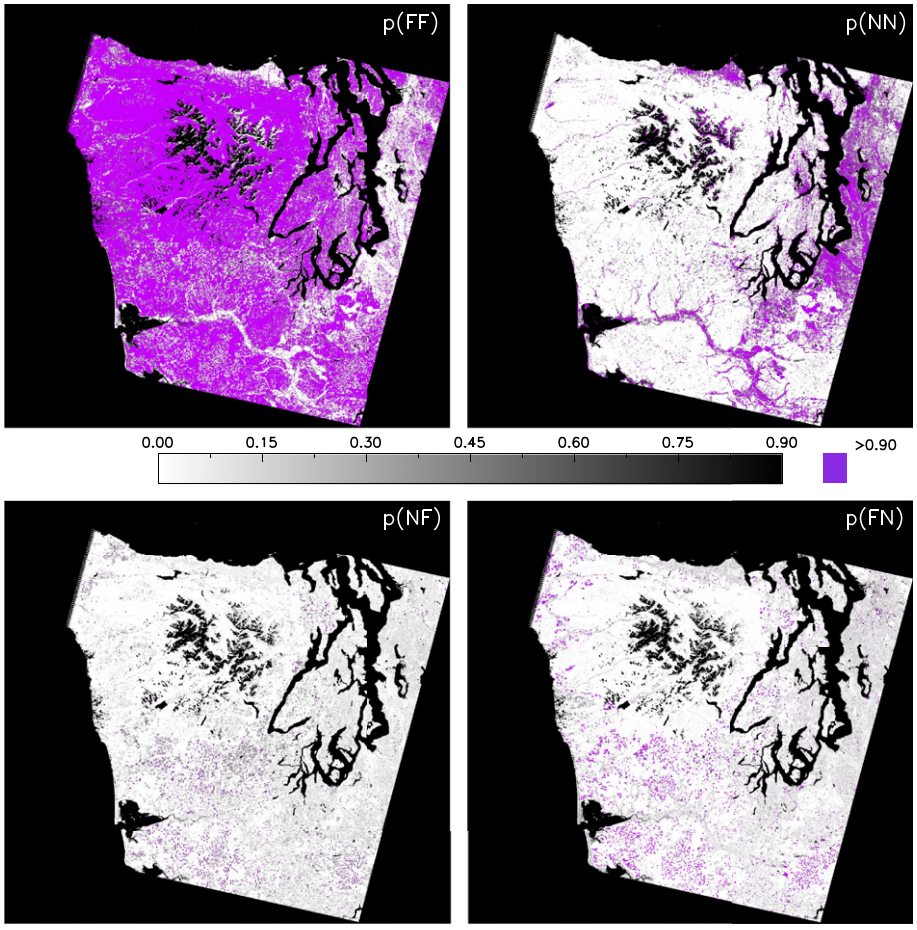

Fig. 4. Spatial distribution of classification probabilities for stable forest (FF), stable non-forest (NN), forest loss (FN), and forest gain (NF). Probabilities greater than 0.9 are highlighted in purple. The horizontal (i.e., longitudinal) distance across the image is approximately 180 km.

inverse of User's accuracy (100 − UA); omission error is the inverse of Producer's accuracy (100 − PA)). Stable non-forest (NN) was also mapped with high accuracy, with PA = 93% and UA = 100%. Errors in NN were overwhelmingly accounted for by omissions, the slight majority of which were misclassified as FF.

Over-representation of the forest class at either time resulted in under-representation of the change classes, forest loss (FN) and forest gain (NF). Class-probability maxima within each pixel and independent validation both revealed this avoidance of change classes. Compared to the reference data, change classes were correspondingly identified with less accuracy than the stasis classes, and omission errors were due to false commission to forest in one or both times. Forest loss (FN) had moderate PA (61%), but low UA (31%). Forest gains (NF) were detected with UA of 47% but PA of only 23%. Errors in FN were dominated by misclassification of non-forest as forest in t_2, and errors in NF were dominated by misclassifications of non-forest as forest in t_1. However, these

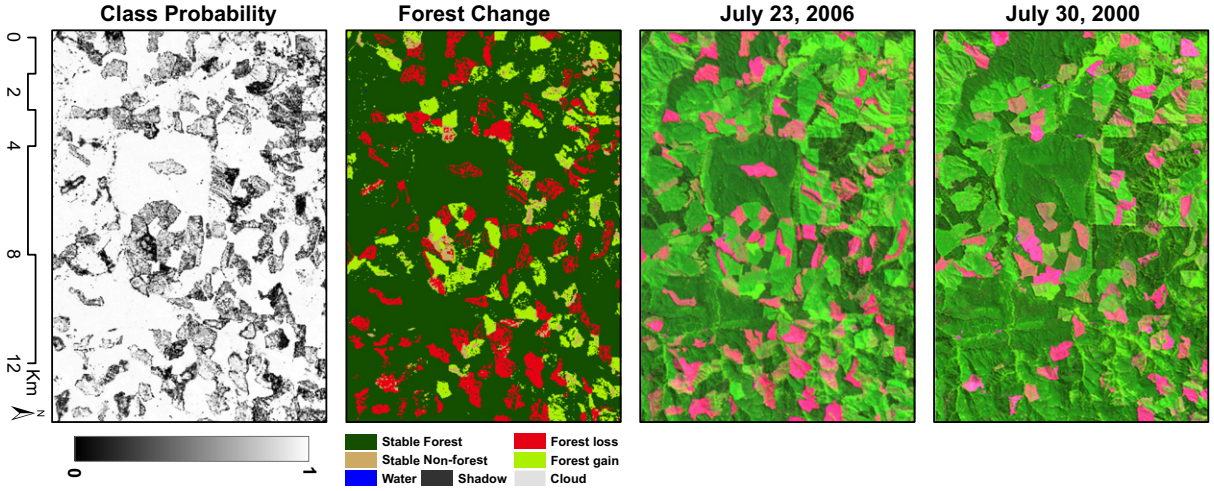

Fig. 5. Reflectance composites (bands 7, 4, 2), maximum-likelihood change map, and probabilities of maximum-likelihood classes for a subset of the study area.

Table 1
Confusion matrix between most probable (expected) dynamic classes and classes observed by visual interpreters (n = 456). Expected values are derived from a model with tree-cover threshold defined as $c^* = 30\%$. Contingency rates are adjusted for sampling biases.

		Expected					Producer's accuracy
		FF	FN	NF	NN	n (%)	
Observed	FF	65.66	0.59	0.31	0.00	66.62	98.55
	FN	2.50	0.48	0.00	0.00	0.79	60.88
	NF	2.09	0.00	0.64	0.00	2.74	23.53
	NN	1.07	0.47	0.41	27.9	29.85	93.46
	n (%)	69.07	1.54	1.36	28.02	100	
User's accuracy		95.05	31.29	47.26	99.56	Overall	94.68

errors—especially those of the change classes—were sensitive to the definition of forest (c^*).

3.4. Sensitivity of change-detection error to the definition of forest

Accuracy varied widely over the range of tree-cover thresholds (c^*) distinguishing "forest" from "non-forest" (Fig. 6). This sensitivity varied among types of error, with User's and Producer's accuracies of each of the four classes more sensitive than Overall accuracy. User's and Producer's accuracies of the change classes—i.e., forest gain and loss—were more sensitive to the definition of forest than were those of stable forest and stable non-forest.

Errors in the input tree-cover estimates ($\hat{c}_i$) were transmitted to change-detection estimates. Although its effects were not obvious at $c^* = 30\%$ tree cover, the documented saturation of the input tree-cover data at $c = 80\%$ cover (Sexton, Song, Feng, et al., 2013) affected all aspects of accuracy relative to independent data as c^* approached and exceeded 80% cover. Only Producer's Accuracy of NN remained high above 80% tree cover, indicating a loss of omission error as the class definition was broadened to include all possible values.

Decreases in commission errors toward FF at $c^* = \sim 15\%$ cover and omission errors toward NN at $c^* = 25\%$ cover suggest an implicit tree-cover threshold of $c^* = 20 \pm 5\%$ cover in the human-interpreted (NAFD) data used as reference. However, errors were less consistent in the change classes—agreement between post-classification and human-interpreted changes was maximized anywhere between $c^* = 10$ and $c^* = 60\%$ tree cover, depending on the type of error and change class. Because the tree-cover values at t_1 and t_2 were independently estimated by the same model, this discrepancy suggests a loss of precision in human observers when identifying changing relative to static forest cover.

Bias may be estimated as the difference between commission error and omission error; positive bias is associated with commission > omission and negative bias with commission > omission errors. As with error in general, bias was more prevalent in the change classes than in the stable classes. However, even in the stable classes, bias increased rapidly above 55% cover—from forest to non-forest. In the change classes, bias was generally toward forest loss and away from gains across the range of thresholds, although bias shifted slightly toward forest gain above 55% cover.

3.5. Comparison to non-probabilistic post-classification change detection

Categorical maps of forest cover in t_1 and t_2 were generated by assigning the class "forest" to each pixel for which $\hat{c}_i > c^*$ at each time. These static maps were then combined into a bi-temporal change map of classes {FF, FN, NF, NN}. The resulting post-classification change map and its accuracy relative to the NAFD reference data (Table 1) were identical to those produced by the probabilistic method developed here. That is, this incorporation of probabilistic information enabled mapping of the uncertainty of categorical land-cover classification and change detection from continuous fields but did not affect the estimates themselves.

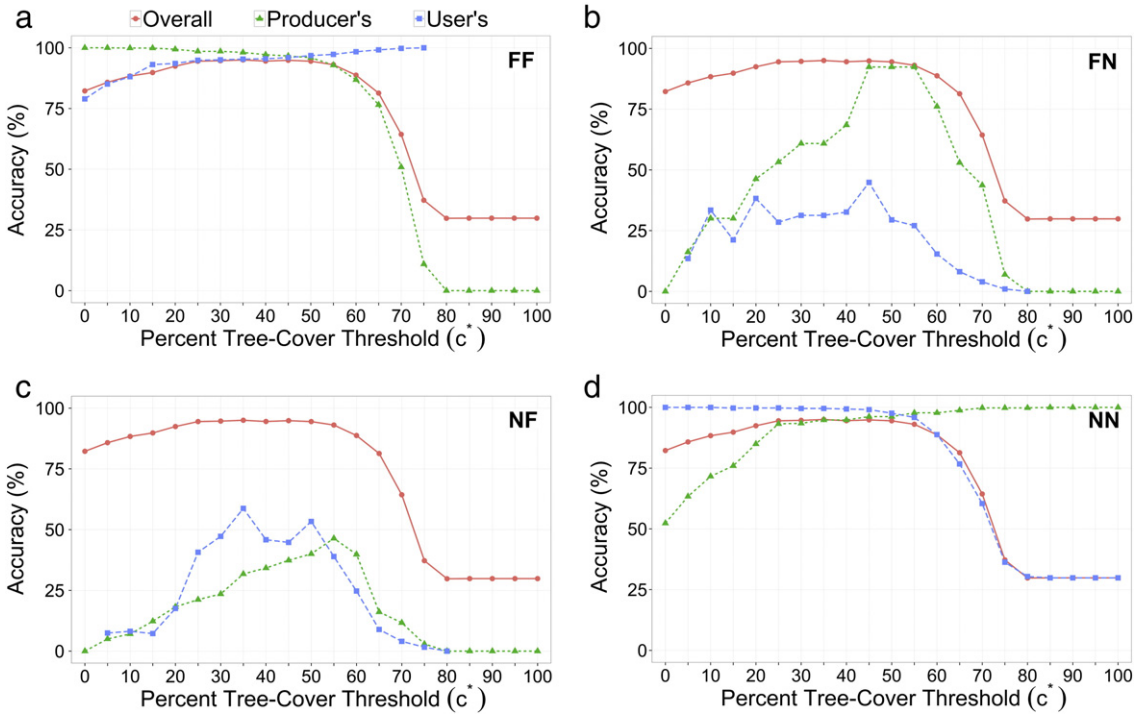

Fig. 6. Sensitivity of change-detection accuracy to the threshold of percent tree cover used to define the static class "forest". Producer's and user's accuracy are plotted for each of the four dynamic classes (FF: stable forest; NN: stable non-forest; FN: forest loss; NF: forest gain). Overall accuracy (which is calculated across all classes) is plotted on each of the class' graphs.

4. Discussion and conclusions

Land cover is most commonly represented as discrete classes, i.e., categorically, through which changes are inferred either by direct identification of change in the spectral domain or by classification and subsequent comparison of classes over time (Lunetta & Elvidge, 1999; Singh, 1989). Alternatively, cover or other biophysical attributes may also be represented as "continuous fields" (DeFries et al., 1995)—in the case of land cover, in terms of fractions or proportions of pixel area occupied by pure or ideal classes. The quantitative thematic scale of continuous fields allows greater flexibility in mathematical operations (Stevens, 1946), including change detection by simple subtraction (i.e., "differencing") as well as post-classification change detection. Whereas direct (spectral) change classification produces estimates of error relative to the training data, post-classification change detection—as well as image subtraction or differencing—ignore classification errors unless the errors are modeled explicitly (e.g., McRoberts and Walters (2012)).

Much effort has been focused on error estimation through validation, especially toward the compilation of public reference databases, sampling designs, and inference of population parameters (Fritz et al., 2009; Olofsson et al., 2012; Stehman, Olofsson, Woodcock, Herold, & Friedl, 2012). Rosenfeld, Fitzpatrick-Lins, and Ling (1982) used a binomial error model to derive minimum sample-size requirements for estimating accuracy of land-cover classifications. Card (1982) estimated corrections for deviations due to sampling variation in accuracy assessments, based on prior information on areal coverage of land-cover classes; Sexton, Urban, Donohue, et al. (2013) later applied this logic to correct the "prevalence problem" (Byrt, Bishop, & Carlin, 1993) of the Kappa statistic (Cohen, 1968). Stehman (2005) compared errors in area estimates arising from mapping versus sampling approaches. Stehman and Wickham (2006) discussed sampling designs, error metrics, and areal estimation of land-cover change, and Stehman et al. (2009) derived design-based population estimators of several accuracy metrics for two-stage cluster sampling. McRoberts (2010) compared probability- (i.e., design-) and model-based approaches for inferring the true area of land-cover classes from image-based maps and independent reference data. McRoberts and Walters (2012) used a validation error matrix to adjust estimates of net forest-cover loss from remote sensing. McRoberts, Cohen, Næsset, Stehman, and Tomppo (2010) provide a recent review of these developments.

Although these advances greatly improve the rigor of inferences derived from validation, the necessary acquisition of independent reference data remains among the most expensive and time-consuming tasks related to production and analysis of land cover and change maps. Given an estimate of the error of the land-cover estimate in each pixel, modeling the propagation of that error into subsequent inferences provides a useful complement to validation that enables data producers to identify and prioritize algorithmic refinements. For users, the resulting per-pixel error estimates could as well refine inferences such as areal coverage, transition rates, and carbon budgets, all of which are currently carried out at regional levels of aggregation (Card, 1982; Hall, Botkin, Strebel, Woods, & Goetz, 1991; Olofsson, Foody, Stehman, & Woodcock, 2013; Pontius & Li, 2010).

Modeling estimates of land cover and its error as a probability distribution in each pixel represents estimation error locally and enables modeling its transmission into inferences such as change detection. Work remains to apply methods being developed to scale local (i.e., pixel-level) cover and error estimates to unbiased regional inferences of areal coverage. Work also remains in the development and application of alternative models of error. The standard Normal error model we employed assumed symmetric errors with a mean of zero, thus assimilating any bias or skew into a single variance (i.e., imprecision) term. Heteroscedasticity was minimized by our use of regression trees to estimate tree cover. However, mis-specification of the error model, which was necessitated by the use of root-mean-square error (RMSE) by the input data, could still lead to inflated precision of classification and change detection, especially at the extremes of tree cover, where errors are most likely to be skewed. Although the assumption of symmetrical errors cannot be true at the extremes of the range of tree cover (implying true cover <0 or $>100\%$), certainty tends to be higher there than near the forest-cover criterion (here, $c = 30\%$). Their distance from the ($c^* = 30\%$) threshold and their relative certainty minimize artifacts on detecting forest-cover and change. Application of more representative error models will require research into alternative probability density functions, as well as methods for estimating their parameters at pixel-resolution. Further efforts will thus assess alternative models for the frequency distribution of error and develop methods to scale local cover and error estimates to unbiased regional inferences of areal coverage.

At its current level of development, the method we present here transmits classification errors while leaving the estimates themselves unaffected. Neither the translation from continuous to categorical cover nor the detection of categorical changes are affected by the incorporation of uncertainty into the modeling process; neither forest cover nor change estimates differed from those of non-probabilistic image differencing. The method is thus superior to simple, non-probabilistic change detection (e.g., image differencing, post-classification change detection) in its ability to communicate the uncertainty in each pixel. Practically, the method thus provides a flexible means of qualifying and filtering change-detection estimates based on measurement uncertainty.

Moreover, because the method is based on the continuous attribute (e.g., tree cover) upon which the land-cover category (forest) and its changes are defined, the method inherits the flexibility of continuous fields to be adapted to other class-definitions. Thus, for forests, the method is equally applicable to any definition of forest based on tree cover: the International Geosphere-Biosphere Programme (30% tree cover), the United Nations Food and Agriculture Organization (FAO) (10% cover), or to any others developed regionally, nationally, or internationally (Bennett, 2001). We have demonstrated the method specifically for changes in forest cover between two times, but the approach can be applied to the cover and changes of any categorical class that can be unambiguously derived from a continuous field.

Acknowledgments

Funding support for this study was provided by the following NASA programs: Making Earth System Data Records for Use in Research Environments (NNX08AP33A-MEASURES), Land Cover and Land Use Change (NNX08AN72G-LCLUC), and Earth System Science Research Using Data and Products from Terra, Aqua, and Acrimsat Satellites (NNH06ZDA001N-EOS). Xiao-Peng Song's contribution was supported by NASA's Earth and Space Science Fellowship (NESSF) Program (NNX12AN92H). Work was performed at the Global Land Cover Facility (www.landcover.org) in the Department of Geographical Sciences at the University of Maryland, in service of the Global Forest Cover Change Project (www.forestcover.org), a partnership of the University of Maryland Global Land and NASA Goddard Space Flight Center. Statistical analyses were conducted in R using the "raster" package. Image processing was performed in ENVI/IDL version 4.8. Jeff Masek and Eric Vermote at the NASA Goddard Space Flight Center provided feedback and insights on algorithms and results. R.E. McRoberts and three anonymous reviewers greatly improved the quality of this manuscript.

References

Bennett, B. (2001). What is a forest? On the vagueness of certain geographic concepts. *Topoi*, *20*, 189–201.

Berger, A., Gschwantner, T., McRoberts, R.E., & Schadauer, K. (2014). Effects of measurement errors on individual tree stem volume estimates for the Austrian National Forest Inventory. *Forest Science*, *60*, 14–24.
Breidenbach, J., Antón-Fernández, C., Petersson, Hans, McRoberts, R.E., & Astrup, R. (2014). Quantifying the model-related variability of biomass stock and change estimates in the Norwegian National Forest Inventory. *Forest Science*, *60*, 25–33.
Burnham, K.P., & Anderson, D.R. (2002). *Model selection and multi-model inference: A practical information–theoretic approach* (2nd ed.). New York: Springer-Verlag.
Byrt, T., Bishop, J., & Carlin, J.B. (1993). Bias, prevalence, and Kappa. *Journal of Clinical Epidemiology*, *46*, 423–429.
Card, D.H. (1982). Using known map category frequencies to improve estimates of thematic map accuracy. *Photogrammetric Engineering & Remote Sensing*, *48*, 431–439.
Clark, J.S. (2007). *Models for ecological data: An introduction*. Princeton University Press (632 pp.).
Cohen, J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, *70*, 213–220.
Congalton, R.G., & Green, K. (2009). *Assessing the accuracy of remotely sensed data: Principles and practices*. Boca Raton, FL, USA: Taylor and Francis.
DeFries, R., Field, C., Fung, I., & Justice, C. (1995). Mapping the land surface for global atmosphere–biosphere models: Toward continuous distributions of vegetation's functional properties. *Journal of Geophysical Research*, *100*(D10), 20867–20882.
DiMiceli, C.M., Carroll, M.L., Sohlberg, R.A., Huang, C., Hansen, M.C., & Townshend, J.R.G. (2011). *Annual global automated MODIS vegetation continuous fields (MOD44B) at 250 m spatial resolution for data years beginning day 65, 2000–2010, collection 5 percent tree cover*. College Park, MD, USA: University of Maryland (http://www.landcover.org/data/vcf/).
Fisher, J.I., Hurtt, G.C., Thomas, R.Q., & Chambers, J.Q. (2008). Clustered disturbances lead to bias in large-scale estimates based on forest sample plots. *Ecology Letters*, *11*, 554–563.
Foody, G.M. (2002). Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, *80*, 185–201.
Fritz, S., McCallum, I., Schill, C., Perger, C., Grillmayer, R., Achard, F., et al. (2009). Geo-Wiki.org: The user of crowdsourcing to improve global land cover. *Remote Sensing*, *1*, 345–354.
Gertner, G.Z., & Dzialowy, P.J. (1984). Effects of measurement errors on an individual tree-based growth projection system. *Canadian Journal of Forest Research*, *14*, 311–316.
Gregoire, T. G. (1998). Deign-based and model-based inference in survey sampling: appreciating the difference. *Canadian Journal of Forest Research*, *28*, 1429–1447.
Hall, F.G., Botkin, D.B., Strebel, D.E., Woods, K.D., & Goetz, S.J. (1991). Large-scale patterns of forest succession as determined by remote sensing. *Ecology*, *72*, 628–640.
Hansen, M.C., DeFries, R.S., Townshend, J.R.G., Carroll, M., Dimiceli, C., & Sohlberg, R.A. (2003). Global percent tree cover at a spatial resolution of 500 meters: First results of the MODIS vegetation continuous fields algorithm. *Earth Interactions*, *7*, 1–15.
Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., et al. (2011). Continuous fields of land cover for the conterminous United States using Landsat data: First results from the web-enabled Landsat data (WELD) project. *Remote Sensing Letters*, *2*(4), 279–288.
Hansen, M.C., Stehman, S.V., & Potapov, P.V. (2010). Quantification of global gross forest cover loss. *Proceedings of the National Academy of Sciences*, *107*, 8650–8655.
Hansen, M. C., Potapov, P. V., Moore, R., Hancher, M., Turubanova, S. N., Tyukavina, A., Thau, D., Stehman, S. V., Goetz, S. J., Loveland, T. R., Kommareddy, A., Egorov, A., Chini, L., Justice, C. O., & Townshend, J. R. G. (2013). High-resolution global maps of 21st-century forest cover change. *Science*, *342*, 850–852.
Hastie, T., Tibshirani, R., & Friedman, R. (2001). *Elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag.
Heinselman, M.L. (1973). Fire in the virgin forests of the Boundary Waters Canoe Area, Minnesota. *Quaternary Research*, *3*, 329–382.
Heuvelink, G.B.M., Burrough, P.A., & Stein, A. (1989). Propagation of errors in spatial modeling with GIS. *International Journal of Geographical Information Systems*, *3*, 303–322.
Homer, C., Huang, C., Yang, L., Wylie, B., & Coan, M. (2004). Development of a 2001 national land-cover database for the United States. *Photogrammetric Engineering & Remote Sensing*, *70*(7), 829–840.
Huang, C., Goward, S.N., Masek, J.G., Thomas, N., Zhu, Z., & Vogelmann, J.E. (2010). An automated approach for reconstructing recent forest disturbance history using dense Landsat time series stacks. *Remote Sensing of Environment*, *114*, 183–198.
Huang, C., Goward, S.N., Schleeweis, K., Thomas, N., Masek, J.G., & Zhu, Z. (2009). Dynamics of national forests assessed using the Landsat record: Case studies in the eastern United States. *Remote Sensing of Environment*, *113*, 1430–1442.
IGBP (1992). Improved global data for land applications. In J.R.G. Townshend (Ed.), *IGBP global change report no. 20*. Stockholm, Sweden: International Geosphere-Biosphere Programme.
IPCC (2003). Good practice guidance for land use, land-use change and forestry. In J. Penman, M. Gyartsky, T. Hiraishi, T. Krug, D. Kruger, R. Pipatti, L. Buendia, K. Miwa, T. Ngara, K. Tanabe, & F. Wagner (Eds.), Japan: IGES.
IPCC (2006). IPCC guidelines for National GHG Inventories IPCC National GHG Inventories Programme. In H.S. Eggleston, L. Buendia, K. Miwa, T. Ngara, & K. Tanabe (Eds.), Japan: IGES.
Kennedy, R.E., Yang, Z., & Cohen, W.B. (2010). Detecting trends in forest disturbance and recovery using yearly Landsat time series: 1. Landtrendr — Temporal segmentation algorithms. *Remote Sensing of Environment*, *114*, 2897–2910.
Kim, D. -H., Sexton, J.O., Noojipady, P., Huang, C., Anand, A., Channan, S., et al. (2014w). Global, Landsat-based forest-cover change from 1990 to 2000. *Remote Sensing of Environment*, *156*, 178–193.
Liu, W., Gopal, S., & Woodcock, C.E. (2004). Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogrammetric Engineering and Remote Sensing*, *70*, 963–971.
Lunetta, R.S., & Elvidge, C.D. (Eds.). (1999). *Remote sensing change detection: Environmental monitoring methods and applications*. London: Taylor and Francis.
McRoberts, R.E. (2006). A model-based approach to estimating forest area. *Remote Sensing of Environment*, *103*, 56–66.
McRoberts, R.E. (2010). Probability- and model-based approaches to inference for proportion forest using satellite imagery as ancillary data. *Remote Sensing of Environment*, *114*, 1017–1025.
McRoberts, R.E., Cohen, W.B., Næsset, E., Stehman, S.V., & Tomppo, E. (2010). Using remotely sensed data to construct and assess forest attribute maps and related spatial products. *Scandinavian Journal of Forest Research*, *25*, 340–367.
McRoberts, R.E., & Walters, B.F. (2012). Statistical inference for remote sensing-based estimates of net deforestation. *Remote Sensing of Environment*, *124*, 394–401.
Metropolis, N., & Ulam, S. (1949). The Monte Carlo method. *Journal of the American Statistical Association*, *44*, 335–341.
Olofsson, P., Foody, G.M., Stehman, S.V., & Woodcock, C.E. (2013). Making better use of accuracy data in land change studies: Estimating accuracy and area and quantifying uncertainty using stratified estimation. *Remote Sensing of Environment*, *129*, 122–131.
Olofsson, P., Stehman, S.V., Woodcock, C.E., Sulla-Menashe, D., Sibley, A.M., Newell, J.D., et al. (2012). A global land-cover validation data set, part I: Fundamental design principles. *International Journal of Remote Sensing*, *33*, 5768–5788.
Pontius, R.G., Jr., & Li, X. (2010). Land transition estimates from erroneous maps. *Journal of Land Use Science*, *5*, 31–44.
Remmert, H. (1991). *The mosaic-cycle concept of ecosystems—An overview. Ecological Studies, vol. 85*. (pp. 1–21). Springer-Verlag, 1–21.
Rosenfeld, G.H., Fitzpatrick-Lins, K., & Ling, H.S. (1982). Sampling for thematic map accuracy testing. *Photogrammetric Engineering and Remote Sensing*, *48*, 131–137.
Schroeder, T.A., Cohen, W.B., & Yang, Z. (2007). Patterns of forest regrowth following clearcutting in western Oregon as determined from a Landsat time-series. *Forest Ecology and Management*, *243*, 259–273.
Sexton, J.O., Song, X. -P., Feng, M., Noojipady, P., Anand, A., Huang, C., et al. (2013b). Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS continuous fields and lidar-based estimates of error. *International Journal of Digital Earth*, *6*, 427–448.
Sexton, J.O., Song, X. -P., Huang, C., Channan, S., Baker, M.E., & Townshend, J.R. (2013c). Urban growth of the Washington, D.C.–Baltimore, MD metropolitan region from 1984 to 2010 by annual estimates of impervious cover. *Remote Sensing of Environment*, *129*, 42–53.
Sexton, J.O., Urban, D.L., Donohue, M.J., & Song, C. (2013a). Long-term landcover dynamics by multi-temporal classification across the Landsat-5 record. *Remote Sensing of Environment*, *128*, 246–258.
Singh, A. (1989). Digital change detection techniques using remotely-sensed data. *International Journal of Remote Sensing*, *10*, 989–1003.
Snedecor, G.W., & Cochran, W.G. (1989). *Statistical methods* (8th ed.). Ames, Iowa: Iowa State University Press.
Stahl, G., Heikkinen, J., Petersson, H., Repola, J., & Holm, S. (2014). Sample-based estimation of greenhouse gas emissions from forests-a new approach to account for both sampling and model errors. *Forest Science*, *60*, 3–13.
Steele, B.M., Winne, J.C., & Redmond, R.L. (1998). Estimating and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment*, *66*, 192–202.
Stehman, S.V. (2000). Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment*, *72*, 35–45.
Stehman, S.V. (2005). Comparing estimators of gross change derived from complete coverage mapping versus statistical sampling of remotely sensed data. *Remote Sensing of Environment*, *96*, 466–474.
Stehman, S.V., Olofsson, P., Woodcock, C.E., Herold, M., & Friedl, M.A. (2012). A global land-cover validation data set, II: Augmenting a stratified sampling design to estimate accuracy by region and land-cover class. *International Journal of Remote Sensing*, *33*, 6975–6993.
Stehman, S. V., & Wickham, J. D. (2006). Assessing accuracy of net change derived from land cover maps. *Photogrammetric Engineering and Remote Sensing*, *72*, 175–185.
Stehman, S.V., Wickham, J.D., Fattorini, L., Wade, T.D., Baffetta, F., & Smith, J.H. (2009). Estimating accuracy of land-cover composition from two-stage cluster sampling. *Remote Sensing of Environment*, *113*, 1236–1249.
Stevens, S.S. (1946). On the theory of scales of measurement. *Science*, *103*(2684), 677–680.
Thomas, N.E., Huang, C., Goward, S.N., Powell, S., Rishmawi, K., Schleeweis, K., et al. (2011). Validation of North American Forest Disturbance dynamics derived from Landsat time series stacks. *Remote Sensing of Environment*, *115*, 19–32.
Townshend, J.R., & Brady, M.A. (2006). *A revised strategy for GOFC-GOLD: GOFC-GOLD*.
Willmott, C.J. (1982). Some comments on the evaluation of model performance. *Bulletin of the American Meteorological Society*, *63*(11), 1309–1313.