Taylor & Francis
Taylor & Francis Group

Estimating area errors for fine-scale feature-based ecological mapping

E. C. ELLIS*† and H. WANG†‡

†Department of Geography and Environmental Systems, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, MD 21250 USA
‡Current address: Environmental Sciences institute, Florida A&M University Science Research Center 305 D, 1520 S. Bronough St., Tallahassee, FL 32307 USA

High spatial resolution feature-based approaches are especially useful for ecological mapping in densely populated landscapes. This paper evaluates errors in estimating ecological map class areas from fine-scale current (~2002) and historical (~1945) feature-based ecological mapping by a set of trained interpreters across densely populated rural sites in China based on field-validated interpretation of high spatial resolution ($\leqslant 1$ m) imagery. Median overall map accuracy, corrected for chance, was greater than 85% for mapping by trained interpreters, with greater accuracy for current versus historical mapping. An error model based on feature perimeter proved as reliable in predicting 90% confidence intervals for map class areas as did models derived from the conventional error matrix. A conservative error model combining these approaches was developed and tested for statistical reliability in predicting confidence intervals for ecological map class areas from fine-scale feature-based mapping by a set of trained interpreters across rural China, providing a practical basis for statistically reliable ecological change detection in densely populated landscapes.

1. Introduction

Long-term ecological changes in densely populated landscapes are responsible for a growing share of global and regional environmental change (Foley *et al.* 2003, DeFries *et al.* 2004, Ellis 2004). Although changes in resource management and ecosystem processes are a critical part of this (Vitousek *et al.* 1997, Houghton 2003), long-term ecological changes in densely populated landscapes are dominated by fine-scale changes in landscape structure (<30 m) caused by the creation, transformation, and abandonment of anthropogenic features with distinct boundaries, such as buildings, roads and small agricultural plots (Forster 1985, Lo and Shipman 1990, Ellis *et al.* 2000b). For this reason, fine-scale feature-based mapping is an especially useful approach in measuring long-term ecological changes within densely populated landscapes (Ellis *et al.* 2000b, August *et al.* 2002, Thomas *et al.* 2003, Ellis 2004, Lu *et al.* 2004, Ellis *et al.* 2006).

Feature-based environmental mapping has a long history rooted in aerial photograph interpretation (Troll 1939) and this approach is expanding in response to the wide availability of high spatial resolution satellite imagery and recent advances in geographic information systems (GIS), image segmentation and

*Corresponding author. Email: ece@umbc.edu

automated feature-extraction (Jensen and Cowen 1999, Hill *et al.* 2002, Thomas *et al.* 2003, Lu *et al.* 2004, Wang and Ellis 2005b). Even with these advances, fine-scale feature-based mapping remains a relatively resource-intensive process, as are environmental sampling and land management surveying, and all of these methods combined together are most powerful when measuring long-term ecological changes within intensively managed landscapes (Ellis *et al.* 2000a, Johnes and Butterfield 2002, Ellis 2004). Therefore, regional scale application of these methods is usually based on regionally stratified sampling designs, such as area frame sampling, that can limit fine-scale mapping and other resource-intensive methods to a smaller set of regionally-representative sample strata or sample cells (Gallego *et al.* 1994, Gallego 2000, Ellis 2004). By this approach, fine-scale measurements within a set of sample cells, such as crop area or fertilizer use, can be used to make regional estimates of these parameters by adjusting for sample cell relationships with regional data (Gallego *et al.* 1994, Gallego 2000, Ellis 2004).

Fine-scale ecological changes within sample cells can be mapped by comparing current and historical feature-based maps using GIS, as long as maps are prepared to the same standards and the spatial scale of mapped features is substantially greater than the scale of misregistration errors between maps (Verbyla and Boles 2000, Lu *et al.* 2004, Wang and Ellis 2005a). When historical imagery must be geo-corrected with limited information however, misregistration errors may reach 10 m or more, causing 'false-change' errors when smaller features are misaligned between maps, potentially obscuring ecologically significant changes (Townshend *et al.* 1992, Wang and Ellis 2005b, 2005a). Fortunately, false change errors caused by misregistration are not a significant source of error in map class area estimates for entire sample cells, if sample cells are large enough (Wang and Ellis 2005a). For this reason, fine-scale ecological changes across mapped sample cells can be detected by testing whether area changes are greater than errors in estimating map class areas within each time period (Lu *et al.* 2004).

Conventional error-matrix approaches for estimating errors in map class areas are well established (Congalton and Mead 1983, Story and Congalton 1986, Congalton and Green 1999, Foody 2002) as are methods specific to feature-based mapping (Green and Hartley 2000). In the present study, we apply both error matrix and feature-based approaches to estimate errors in map class areas from current (~2002) and historical (~1945) ecological feature-based maps by trained interpreters across China's densely populated rural landscapes as part of a regional study of long-term ecological change (Ellis 2004).

Disagreement between trained interpreters is a major source of error in thematic maps prepared by image interpretation, even when maps are field validated (Cherrill and McClean 1999, Green and Hartley 2000, Powell *et al.* 2004). This study assesses this source of error by comparing maps prepared by a set of trained interpreters across a sample cell representing the most challenging mapping conditions across the sites in our study (Ellis 2004). Although errors in feature-based mapping include position, shape and classification (Green and Hartley 2000), analysis of ecotope maps (table 4 in Wang and Ellis 2005a) demonstrated that positional error is a negligible source of error in map class areas estimated as a percentage of entire 500 m × 500 m sample cells, given that false change error was near zero for cells >463 × 463 m at the maximum coregistration error and number of map classes observed per cell (15 m, assuming maximum opposing errors at both times; 38 classes, in Jintang; Wang and Ellis 2005b). We therefore focused our analysis on

errors in map class areas caused by disagreement in feature mapping (shape-related) and feature classification between trained interpreters.

The goal of this assessment was to develop a model for predicting statistically reliable error intervals for map class area estimates from sample cell maps across the set of trained interpreters, with statistically reliable error intervals defined as those enclosing the correct value in at least 90% of estimates (Schenker and Gentleman 2001). Based on this definition, models for predicting map class area errors were developed and tested for their statistical reliability and overall error rate at the site of model development and at an independent check site, to establish a general model for predicting errors in class area estimates across interpreters.

2. Methods

2.1 *Study sites*

A set of 60 500 m × 500 m square sample cells was selected for ecological mapping and other measurements by a regional sampling design, with 12 cells allocated within each of five 100 km^2 field sites representing environmentally distinct regions across densely populated rural China (Ellis 2004). Preliminary mapping across sites confirmed that our site in densely populated eastern Jintang County, Sichuan Province in the Sichuan Basin Hilly Region of China was the most challenging for mapping based on its greater complexity of landscape management and terrain and lowest image quality across sites in China (Ellis 2004). The region has a subtropical monsoon climate and intensively managed hilly terrain composed of semi-natural bench plateaus with upland agriculture divided by steep slopes (both vegetated and barren), with rice paddies and small ponds ringed by housing and narrow roads in the elongated, gradually terraced, valleys in between. Sample cells at this site had the largest number of map classes, the most difficult terrain, dispersed built-up features obscured by tree and bamboo cover, complex patterns of land and vegetation management and relatively poor quality historical aerial imagery (Ellis 2004, Wang and Ellis 2005b).

We therefore selected a sample cell representing these challenging conditions at the Jintang site for detailed investigation of interpreter error (centre=104.770°E, 30.549°N). This cell included more than 90% of the feature classes observed across the 12 sample cells mapped at this site, and ranged in elevation from 396 to 433 m with a population of ~800 persons km^{-2} in 2002 and ~530 persons km^{-2} in ~1945. A check site in southern Yiyang County, Hunan Province, China was selected for error model testing (400 m × 400 m cell, centre=112.450°E, 28.385°N). This cell had a similar climate to the Jintang site, but with the lower population density (2002 ~470, 1945 ~180 persons km^{-2}) and less intensive land management typical of China's Subtropical Hilly Region, with gentle hillslopes covered by forestry and perennial crops (tea) with gently terraced rice paddies ringed by houses and roads in between (Ellis 2004).

2.2 *Imagery*

IKONOS 4 band pan-sharpened 1 m resolution GEO imagery (www.spaceimaging.com) was acquired 22 December 2001 across a 7 × 14.25 km scene (100 km^2) in eastern Jintang County, and orthorectified using a digital elevation model and 25 ground control points obtained by submeter accuracy global positioning systems (GPS) using the high resolution satellite model (i.e. rigorous model) of PCI

Geomatics orthoEngine version 8.2 (PCI Geomatics, Richmond Hill, Ontario, Canada), yielding a positional error of 4.0 m CE90 (circular error 90) as detailed in Wang and Ellis (2005b), meeting MS IIRS (MultiSpectral Imagery interpretability Rating Scale) Level 5 image quality standards (Imagery Resolution Assessments and Reporting Standards (IRARS) Committee 1995). Black and white aerial photographs for the site on 25 June 1944 were obtained from the US National Archives and Records Administration (NARA; RG-373, www.archives.gov) and orthorectified to a positional error of 6.5 m (CE90) using image tie points from orthorectified IKONOS imagery (Wang and Ellis 2005b), meeting NIIRS (National Imagery interpretability Rating Scale) Level 3 image quality standards (Imagery Resolution assessments and Reporting Standards (IRARS) Committee 1996). IKONOS imagery for the Yiyang site was acquired on 1 January 2002 and orthorectified by a similar approach (Wang and Ellis 2005b).

2.3 *Feature mapping and classification*

Ecologically distinct landscape features (ecotopes) were mapped across sample cells by field-validated interpretation of high spatial resolution imagery ($\leqslant 1$ m) by trained interpreters using a standardized ecological mapping and classification system designed to delineate stable land management and vegetation features observable at ground level at the time of image acquisition by both ecologists and local land managers (anthropogenic ecotope mapping; Ellis *et al.* 2006). Ecotope features were mapped by a scale-explicit sequential mapping strategy, beginning with linear features ($\geqslant 2$ m in width and area $\geqslant 25\,\text{m}^2$, length $\geqslant 4 \times$ width; examples are roads and ditches), followed by hard areal features ($\geqslant 5$ m in width and area $\geqslant 25\,\text{m}^2$ with clear edges and homogeneous interiors; examples are buildings and water bodies), with the remaining area divided into soft features ($\geqslant 5$ m in width and area $\geqslant 100\,\text{m}^2$ with fuzzy edges and variable cores; examples are crop plots and vegetation patches). All ecotope features were corrected by field validation to conform to stable (potentially observable for $\geqslant 2$ y) land-management boundaries at ground level in the field by the interpreter and local land managers in cases where vegetation cover, shadow, or off-nadir imagery confused land-use boundaries in imagery (Ellis *et al.* 2006). This procedure enabled the production of comparable ecotope maps of the same site from high resolution imagery acquired in different seasons (leaf on versus leaf off), and by different sensors, as long as imagery met basic quality standards as described in Ellis *et al.* (2006).

After initial mapping, features were classified using a four level *a priori* ecological classification hierarchy, FORM→USE→COVER→GROUP+TYPE, combining simple land form, use and cover classes (FORM, USE, COVER) with a set of more detailed feature management and vegetation classes (GROUPs) stratified into TYPEs. Ecotope classes are created by combining all four classification levels within each feature. For example, a forest of closed canopy regrowth deciduous trees (GROUP+TYPE=dt02) on a gentle slope (FORM=SL=Sloping) managed for harvest (USE=T=forestry) with Perennial COVER (P) is classified as the ecotope 'SLTPdt02' (FORM+USE+COVER+GROUP+TYPE).

Initial maps of sample cells were prepared by trained interpreters by direct interpretation of polygon features from imagery in a GIS (Arcinfo 8.3, Environmental Systems Research Institute, Redlands, California) followed by feature verification and correction in the field assisted by local land managers and 1 : 1200 scale image and feature maps. Historical maps were groundtruthed with the

aid of two local elders per sample cell, aged $\geqslant 16$ at the time of image acquisition, by a combination of interviews assisted by 1:1200 scale image maps and by visiting all confusing areas in the field together with elders. This interpretation and groundtruthing sequence was repeated twice by the same interpreter at each sample cell and then reviewed by another trained interpreter to ensure compliance with the standard mapping and classification rules, which included correction for height distortion, shadowing and tree cover over buildings, roads, and water bodies (Ellis *et al.* 2006). Final ecotope maps were checked against an ecotope classification code database to check for and correct invalid code combinations and feature topology was checked and corrected to ensure continuous classified polygon feature layers and an overall map area error tolerance of $+/-0.05\%$ of sample cell area.

Prior to mapping across sites, all four interpreters were trained at two different sites across China by the repeated blind comparison of their maps with standardized reference maps, to calibrate results across interpreters. The Jintang sample cell was mapped by the full set of trained interpreters on site from 3–10 July 2003 (current map) and from 10–16 July 2003 (historical map), along with two additional 'trainee' interpreters (trained at one site or less). The Yiyang sample cell was mapped from 10–17 November 2002 by the same interpreters. Interpreters were permitted to discuss the mapping process with each other, but were not permitted to view others' maps until all were completed.

2.4 *Accuracy assessment*

Map accuracy was assessed by comparing current and historical maps by three trained interpreters with a 'gold-standard' reference map of the sample cell for each time period prepared by field validation and correction of the map deemed most accurate across 4 interpreters (Powell *et al.* 2004, Ellis *et al.* 2006); the interpreter producing this base reference map was subsequently excluded from analysis of interpreter error. Use of a gold-standard reference map corrected across interpreters is an optimal strategy for thematic map accuracy assessment, producing more reliable results than those from any one interpreter, even though the reference map must still include some error (Powell *et al.* 2004).

Data for map accuracy assessment were obtained for each sample cell by placing a 3 m triangular grid of points over the maps of each interpreter and the reference map using GIS, and then determining the classification of each point by each interpreter and the reference (the triangular grid had $n=31\,837$ points in the Jintang sample cell, and had $n=20\,064$ points in the Yiyang cell). The triangular grid point sample was a better representation of map class areas than a random point sample of the same size, as the proportion of the sample cell covered by each ecotope class was found to be nearly identical between the triangular grid point estimate and the areas determined from the original classified ecotope polygons. (In Jintang, the sum of absolute ecotope area differences between the polygon data and the triangular grid sample was 0.4%, but was 1.8% for the random sample.) The classification data obtained from the grid point sample were entered into a conventional error matrix to calculate overall map accuracy (correct/total) and class user's accuracy (correct/ mapped) by pairwise comparison of each interpreter with the reference map (Congalton and Green 1999). Cohen's Kappa (κ) and Andrés and Marzo's Delta (Δ; 2004) measures of agreement corrected for random chance were calculated using the algorithm of Andrés and Marzo (2004; http://www.ugr.es/~bioest/delta.htm). κ and

Δ are related statistics, with values >0.75 indicating strong agreement above chance (Andrés and Marzo 2004, Norusis 2004b).

2.5 Area measurement errors

Three independent methods were used to estimate errors in map class areas. Two of these, class user's error (CUE) and class area error (CAE), were calculated from conventional error matrix data obtained as above and map class areas calculated as the sum of polygon areas, respectively. CUE, or 'error of commission', includes errors in position and was calculated as 1 – class user's accuracy from the row marginals of the error matrix for each interpreter (Story and Congalton 1986). CAE, the 'non-site-specific error' of Congalton and Green (1999), excludes positional error, and was calculated as the difference between interpreter (CA_{obs}) and reference (CA_{ref}) map class areas, divided by interpreter area (CA_{obs}):

$$CAE = \frac{\left|CA_{ref} - CA_{obs}\right|}{CA_{obs}} \tag{1}$$

A third, feature-based method for estimating area errors, feature area error *(FAE)*, was derived from analysis of a limited set of individual ecotope features (*j*) mapped by each interpreter (i), as the absolute difference in area between corresponding interpreter (*FA(obs)*) and reference (*FA(ref)*) features

$$FAE_{ij} = \left|FA(ref) - FA(obs)_i\right| \tag{2}$$

in theory, FAE for each map class (FAE_k) could be calculated as the sum of FAE_{ij} across all of the features in the class divided by the class area, thereby obtaining error estimates comparable with CUE or CAE. This is not possible however, because interpreters do not always recognize exactly the same features within each sample cell. We therefore measured FAE_{ij} for a subset of current and historical features reproduced by all interpreters and analysed these for relationships with feature classification, area, perimeter and other factors to determine whether models based on these factors might facilitate prediction of FAE across map classes. Features selected for FAE analysis varied across the range of feature sizes and included only those with perimeters mostly within the sample cell ($>75\%$ mapped versus clipped edges). For current maps, FAE_{ij} was calculated from five independent mappings of 35 features including 16 hard and 19 soft features (3 trained + 2 trainee interpreters, total = 175 features). FAE_{ij} for historical maps was calculated from three independent mappings of 14 features, including 7 hard and 7 soft features (three trained interpreters, total = 42 features).

2.6 Error estimators for map class areas

Errors in map class area estimates from the three interpreters of the Jintang sample cell were tested for relationships with map class (ecotope, USE), feature type (hard versus soft), feature area and feature perimeter to determine whether all or some of these factors were useful predictors of error in map class area estimates. Ecotope classes smaller than 0.25% of sample cell area were excluded from this and other analyses because these smaller ecotope classes were both statistically unreliable (more than half had $<70\%$ user's accuracy; Foody 2002) and unimportant: in total these never accounted for $>2.4\%$ of any sample cell area. We tested for relationships

between FAE_{ij} and feature area and perimeter using linear regression, and then combined these variables together with USE class and hard vs. soft feature type in a Univariate general linear model (GLM) to test for ecotope FAE_{ij} differences and for the factors linked to these differences using data for individual features averaged across three interpreters (Norusis 2004b). This approach produced a complete model of FAE_{ij} for each time period, including all variables as fixed factors, along with their interactions. To test for differences among ecotope CUE and CAE estimates, linear mixed models (LMM; Norusis 2004b) were used, with ecotope as fixed factor, because this test is more reliable than GLM when observations are not fully independent with equal variance, as was observed for ecotope CUE and CAE estimates three interpreters per time period, 20 current ecotopes, 10 historical ecotopes; Levene's test $P < 0.05$; (Snedecor and Cochran 1980).

Two models for predicting error intervals for class area estimates were evaluated in terms of the total amount of error they produced and their statistical reliability. Statistical reliability was quantified as the probability that the rate of successful error predictions produced by an estimator (intervals covering the correct value) was at least as great as expected for an estimator that is 90% successful in error prediction (90% reliable). The first and simplest model was to predict class area error as a direct percentage of ecotope class area (class area error=class area × error). CUE and CAE already describe class area errors as a percentage of class area, so general estimates of these errors across ecotope classes and interpreters ($CUE_{overall}$, $CAE_{overall}$) were made by bootstrapping their 95% confidence upper limit (pooled CUE and CAE were not normally distributed; current $n=60=20$ ecotopes × 3 interpreters, CUE=17.7%, CAE=29.2%; historical $n=30=10$ ecotopes × 3 interpreters, CUE=31.0%, CAE=33.8%; 3000 parametric bootstrap runs; Efron and Tibshirani 1986).

A second error model based on feature perimeter was derived from the relationship between feature perimeter and FAE_{ij}, based on the concept of 'epsilon bands' of error around mapped features (reviewed by Green and Hartley 2000). In this model, the mapped perimeter of each ecotope feature (FP_j; the portion of feature perimeter not cut by sample cell boundaries) was multiplied by an error factor ($FAE_{overall}$, in m) to estimate FAE_j for each mapped feature. FAE_j was then divided by feature area (FA_j) and added across all of the features in each map class (k) to estimate map class FAE as a percentage of map class area (FAE_k):

$$FAE_k = \sum_J \frac{FP_j \times FAE_{overall}}{FA_j} \qquad (3)$$

$FAE_{overall}$ was estimated as the 95% confidence upper limit of the slope of the FAE_{ij} to FP_j relationship calculated by linear regression (current=0.93 m, historical=3.57 m, regression $R^2 > 0.75$, $P < 0.001$ for both times).

90% confidence intervals for map class areas were calculated for the three error estimators ($CUE_{overall}$, $CAE_{overall}$, $FAE_{overall}$), by multiplying each error estimator by $Z_{0.05}$ (1.645; 90% confidence for the normal distribution), given that ecotope-level error across interpreters (CUE and CAE) approximated a normal distribution (Shapiro-Wilk test P>0.05 for 56 out of 60 current+historical estimates; Norusis 2004a). The statistical reliability of these error intervals was then tested by calculating them across the ecotope area estimates of the three interpreters of each sample cell (Jintang current $n=60$, historical $n=30$), tabulating the number of successful error predictions (error intervals containing the reference value), and

testing their resulting error prediction success rate against the hypothesis of 90% prediction success using the binomial test (Taylor 1997). This test compared the error prediction success rate with the chance of observing less than or equal to this number of successes given the number of trials and the hypothesis that the error estimator is 90% reliable (error prediction is successful in ⩾90% of predictions). When this binomial test gives a less than 10% chance of the observed success rate, the hypothesis of 90% reliable error prediction is rejected with 90% confidence. When the test yields a ⩾10% chance of the observed success rate, a 90% reliable error model is not rejected, and when the binomial test yields a ⩾90% chance of the observed success rate, then it is ⩾90% probable that the error model is >90% reliable.

To investigate relationships between error prediction success and total error introduced by area-based (e.g. CUE, CAE) versus perimeter-based error models (e.g. FAE), we calculated error intervals for ecotope area estimates by the three interpreters of the Jintang sample cell across a range of values for area and perimeter error. At each increment of error predicted by each model (area, perimeter), ecotope error intervals for the three interpreters were tested for their error prediction success rate and sum of errors across ecotope classes. Based on results of this analysis, area- and perimeter-based models were then tested for their prediction success rate and overall error across three interpreters at the Yiyang sample cell (16 current ecotope classes).

3. Results and discussion

3.1 *Mapping accuracy*

Overall accuracy of current (2002) and historical (1944) ecological maps of the Jintang sample cell (figure 1) was >85% across interpreters at all classification levels, with median accuracy >90%, meeting widely accepted map accuracy standards (tables 1 and 2; Foody 2002). Map accuracy corrected for chance using Cohen's κ was greater for current maps (>85%) than for historical maps (>75%), and was even higher using the related and more robust Δ statistic (>86% for current versus >83% for historical maps; Andrés and Marzo 2004). Overall map accuracy for individual classification levels (USE and GROUP) was greater than that of combined classes, such as ecotope and USE+COVER, confirming that the accuracy of area estimates can be increased by lowering their thematic resolution (Petit and Lambin 2001, Smith *et al.* 2003; median overall accuracy for current GROUP maps =92.6%, USE+COVER =91.6%). As expected, with less ecotope classes (16) and less intensive land management, ecotope map accuracy at the Yiyang sample cell was even higher than in Jintang, with overall accuracy ranging from 87% to 93% across three interpreters, and with κ estimates ranging from 84% to 91% (2002 map; overall Δ ranged from 86–92%).

3.2 *Error in map class areas*

Map class accuracy and map class area were related, but not in a simple way. At all levels of classification, classes with larger areas tended toward greater accuracy (Δ and class user's accuracy =1−CUE), even though some small classes had surprisingly high accuracy (mostly hard features such as houses), and some large classes had surprisingly low accuracy (mostly soft features with Disturbed, Forestry, or Fallow USE; tables 1 and 2). As a result, user's accuracy was only weakly related
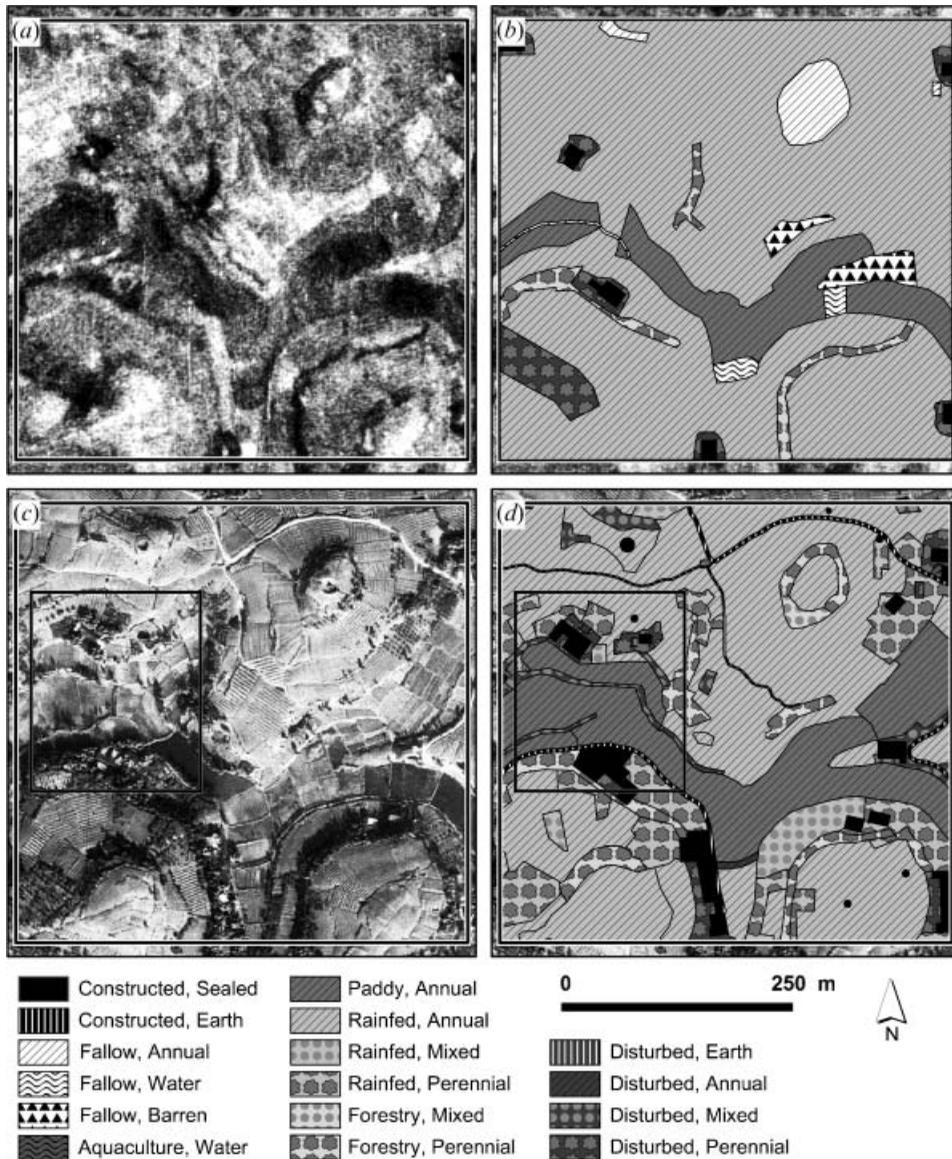
Figure 1. Imagery and ecotope reference maps of the Jintang 500×500 sample cell (UTM projection). (*a*) 1944 aerial photograph and (*b*) historical ecotope map. (*c*) 2001 IKONOS imagery and (*d*) current ecotope map. Ecotope features are symbolized with USE class symbols overlaid by COVER class symbols, as described in the legend. A box highlights area expanded in figure 3.

to map class area either directly or after log-transformation of area (regression $R^2 < 0.31$, $P < 0.05$). On the other hand, more than 90% of current ecotope classes with satisfactory user's accuracy ($>70\%$; Foody 2002) had larger areas ($>0.25\%$ of sample cell area), and more than half of ecotopes with poor user's accuracy ($<70\%$) were smaller classes ($\leqslant 0.25\%$ of sample cell area), similar to previous results (Cherrill and McClean 1995). As these smaller ecotopes never covered $>2.4\%$ of any sample cell across sites, we consider it both prudent and practical to flag all map

Table 1. Current (2002) USE and ecotope class areas (reference ± max difference from reference), accuracy corrected for chance using Andrés and Marzo's Δ (median ± max difference from median), mean class user's error (CUE) and class area error (CAE) across three interpreters, along with predicted feature area error (FAE) as % class area for the Jintang sample cell (figure 1). USE classes are sorted by area and partitioned into ecotopes, with FORM class indicated in parentheses where needed.

USE	Ecotope	Area (% cell)	Δ (%)	CUE (%)	CAE (%)	FAE (%)
Rainfed agriculture		65.1±0.6	93.2±2.8	3.4	0.7	6.5
	Small-scale staple crops (Bench Plateau)	36.4±0.8	93.3±2.5	4.4	0.9	5.7
	" (Sloping)	10.7±0.1	95.8±1.0	3	1.2	3.3
	" (Summit)	2.3±0.1	89.5±0.5	7.7	1.9	8.5
	" (Foot Slope)	1.4±0.2	84.6±4.0	3.6	11.6	8
	Small-scale immature pear orchard (Bench Plateau)	2.8±0.1	93.3±10.3	6.9	2	8.7
	Small-scale mature mandarin orange orchard (Bench Plateau)	11.0±0.3	90.4±4.6	10.7	1.9	10.5
	" (Sloping)	0.4±0.3	60.3±44.0	21.4[*]	103.0[*]	13.3
Paddy		18.2±0.7	95.2±2.0	4.0	1.8	5
	Reservoir paddy (Foot Slope)	12.7±0.6	95.0±2.1	2.9	2.2	4.9
	Rice paddy (Foot Slope)	3.8±0.2	92.7±2.2	6.7	2.7	4.3
	" (Bench Plateau)	1.8±0.1	89.6±0.7	8.3	3.3	9.9
Forestry		5.7±1.2	75.6±9.0	27.9	12.6	16.8
	Regrowth open woody vegetation (Steep Slope)	0.4±0.1	77.4±13.6	19.2[*]	10.3	16.5
	Planted conifer forest (Steep Slope)	4.8±1.2	86.0±5.3	24.9[*]	14.6	16.7
	Regrowth open wooded brush (Steep Slope)	0.5±0.0	71.0±0.9	16.5[*]	2.8	18
Constructed		5.5±0.2	81.4±5.0	18.1	3.0	29.1
	Unpaved local roads	1.8±0.0	68.3±7.9	29.3[*]	1.5	56.7
	Single story attached housing	1.9±0.1	86.5±7.0	11.9	1.8	14
	Multistory attached housing	0.9±0.1	84.2±11.2	12.6	9.7	10.8
	Single story detached housing	0.4±0.1	73.8±15.3	15.8[*]	13.7	19.6
Disturbed		5.4±0.5	68.2±2.9	31.8	5.8	24.6
	Disturbed woody vegetation with debris (Steep Slope)	1.7±1.2	37.7±14.9	23.9[*]	127.9[*]	13.9
	Disturbed trees with debris (Bench Plateau)	1.4±0.4	54.6±17.2	27.9[*]	23	22.2
	" (Steep Slope)	1.2±0.9	81.2±8.7	47.3[*]	38.4	36.4
USE classification, overall κ=88.5±3.3		99.9±3.3	90.9±2.1	16.9	20.4	16.2
Ecotope classification, overall κ=86.2±2.9		96.0±5.2	87.8±2.6	15.2	18.7	15.1
Sum of ecotope errors as % sample cell area				8	5.7	8.5

[*] Significant ecotope fixed effect in LMM (α=0.05).

E. C. Ellis and H. Wang

Table 2. Historical (1944) USE and ecotope class areas (reference±maximum difference from reference), map class accuracy corrected for chance using Andrés and Marzo's Δ (median±maximum difference from median), mean class users error (CUE) and class area error (CAE) across 3 interpreters, and predicted feature area error (FAE) as percent of class area for the Jintang sample cell (figure 1).

USE	Ecotope	Area (% cell)	Δ (%)	CUE (%)	CAE (%)	FAE (%)
Rainfed agriculture	Small-scale staple crops (Bench Plateau)	77.5±6.7	90.6±8.2	2.7	3.9	8.2
Paddy	Reservoir paddy (Foot Slope)	10.0±2.1	94.9±9.5	16.8	11.2	23.0
Fallow		5.2±1.3	84.9±8.5	21.7	9.7	37.6
	Broadleaf herbaceous regrowth vegetation (Summit)	2.3±0.1	96.4±1.5	3.6	2.0	17.6
	Exposed rock (Bench Plateau)	1.2±0.2	77.7±11.4	22.8	72.7[*]	37.1
	" (Steep Slope)	0.5±1.0	74.2±11.5	12.2	11.9	45.3
	Small pond	0.8±0.3	82.3±38.9	28.1	12.8	50.3
Disturbed		3.8±1.2	68.6±12.9	29.8	19.4	50.5
	Tree-covered regrowth grave	2.2±0.3	82.9±1.6	17.2[*]	9.5	22.4
	Disturbed trees with debris (bench plateau)	1.6±0.5	50.1±14.2	48.4[*]	25.0	78.5
fromestry	Regrowth conifer forest (Steep Slope)	2.6±1.8	77.6±16.6	41.7[*]	38.0	52.4
Constructed	Single-story attached housing	0.9±0.5	86.6±23.6	45.6[*]	41.5	56.5
USE classification, overall κ=79.0±5.7		100.0±13.6	87.7±4.4	26.4	20.6	38.0
Ecotope classification, overall κ=78.5±5.6		99.5±14.3	87.6±4.4	23.9	22.8	39.1
Sum of ecotope errors as % sample cell area				7.0	7.2	14.5

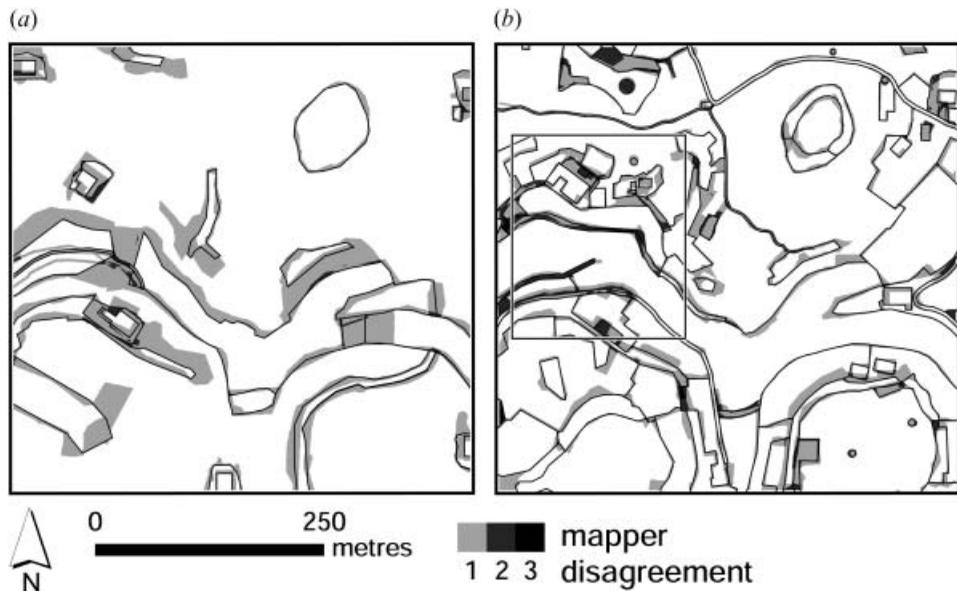[*] Significant ecotope fixed effect in LMM ($\alpha=0.05$).

Figure 2. Classification disagreement between 3 trained interpreters across the Jintang sample cell (figure 1) in terms of (*a*) historical and (*b*) current ecotope map class (blank=all agree). A box highlights area expanded in figure 3.

classes covering $\leqslant 0.25\%$ of a sample cell as unreliable and to eliminate them from further analysis, although up to half of these smaller estimates might in fact be accurate. For example, the gold standard current reference map of the Jintang sample cell had 32 ecotope classes, and interpreters used from 30 to 33 classes, but all had the same 20 larger ecotope classes after eliminating ecotopes smaller than 0.25% of the sample cell.

Comparison of maps across trained interpreters confirmed that errors in feature alignment and edge mapping were ubiquitous across current and historical maps of the Jintang sample cell (figure 2, with detail in figure 3; Green and Hartley 2000). It was also apparent that misclassification was far more common for smaller features (higher perimeter/area), likely because these features attracted less attention from interpreters and were often more difficult to classify than larger features (figure 2). Together, these observations indicate that feature perimeter should be a strong predictor of error in class area estimates (Green and Hartley 2000).

Indeed, feature perimeter was a stronger linear predictor of interpreter error in feature area estimates (FAE_{ij}) than feature area itself (current linear regressions, figure 4: perimeter $R^2=0.76$, area $R^2=0.62$; historical linear regressions, not shown: perimeter $R^2=0.86$, area $R^2=0.84$; $P<0.001$ for all regressions). Although feature area was nearly as strong a predictor of error as feature perimeter using power regression, this relationship can be explained by the strong relationship between feature area and feature perimeter (Figure 4(c), $P<0.001$ for both regressions). Given that perimeter was the stronger predictor, and that feature area was a nonsignificant predictor of FAE when area and perimeter were regressed together, feature perimeter was the more powerful predictor of FAE (multiple linear regression, current perimeter $P<0.001$, area $P=0.16$, $R^2=0.77$ overall $P<0.001$; historical perimeter $P=0.07$, area $P=0.25$, $R^2=0.88$ overall $P<0.001$). Surprisingly,
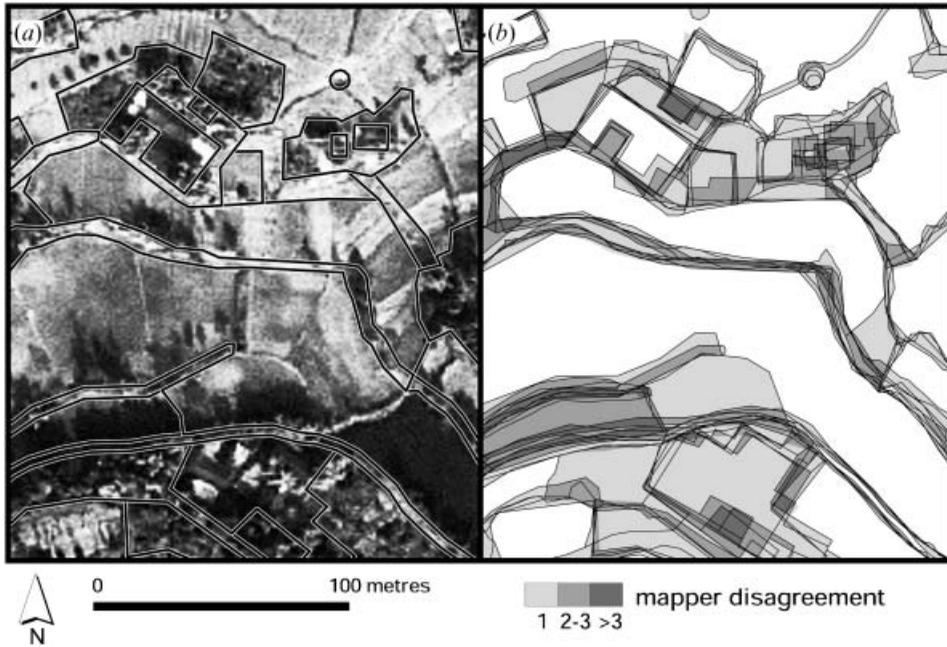
Figure 3. Detail of current feature mapping errors at the Jintang sample cell (figure 1). (*a*) Reference feature boundaries over 2001 IKONOS image. (*b*) Feature boundaries for 5 interpreters (3 trained, 2 trainee) overlaid on interpreter classification disagreement (*b*).

FAE did not differ significantly between land USE class or feature types (hard vs. soft), nor did these factors affect the relationship between FAE and feature perimeter or area ($P > 0.2$, figure 4). Therefore, the simple linear relationship between FAE_{ij} and feature perimeter provides a straightforward model for predicting errors in map class areas from feature perimeters by multiplying these by the slope of the perimeter to FAE relationship (the line intercept was statistically non-significant).

Error estimators based on the conventional error matrix produced similar, but not identical results. CUE includes errors in feature position and estimates error as the misclassified percentage of each class, while CAE is independent of feature position and agreement with other classes (Congalton and Green 1999). As a result, even though their means across ecotope classes were not significantly different, CUE was not correlated with CAE and also varied less between classes than CAE (tables 1 and 2). As expected, CUE also showed a tendency toward higher error estimates for ecotopes with smaller features (roads and detached houses) and features with less discrete edges (Disturbed and Forestry land USE), while CAE differences between ecotopes were without clear trend (tables 1 and 2).

3.3 Error estimators for ecological map class area estimates

Given the goal of producing statistically reliable error estimates for ecological map class areas from sample cell maps by a set of trained interpreters across five rural sites in China, we developed a practical strategy to predict these errors based on four principles. First, to avoid false detection of changes and differences in class areas (Type I error), larger, more conservative error intervals were estimated, allowing
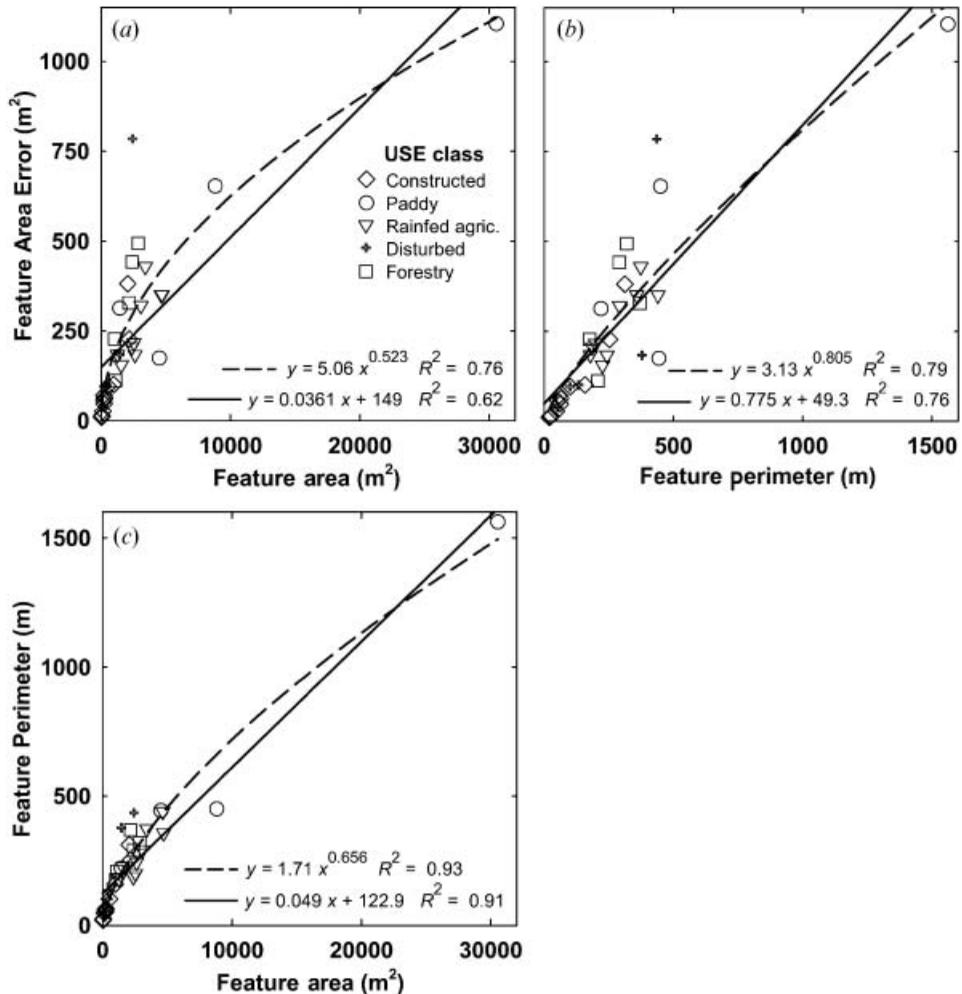
Figure 4. Relationships of feature area error (FAE) with (*a*) feature area and (*b*) perimeter, and (*c*) the relationship between feature area and perimeter for current maps of the Jintang sample cell. USE class of features is indicated by symbols, solid lines are linear regression, dashed lines are power regression.

that smaller differences might go undetected (Type II error; Schenker and Gentleman 2001). Second, interpreter error was considered the dominant source of error in ecological map class areas from fine-scale feature-based mapping (Cherrill and McClean 1999). Third, a set of trained and calibrated interpreters following a standardized scale-explicit ecological mapping and classification procedure, such as the system used in this study, should produce consistent types of interpreter error across sites (Cherrill and McClean 1999, Powell *et al.* 2004). Finally, the assessment of interpreter error under the most challenging mapping conditions should produce conservative estimates of interpreter error across sample cells in general. Based on these principles, and the observation that misregistration error was nonsignificant for ecotope area estimates across 500×500 m sample cells (Wang and Ellis 2005a), we developed a conservative error prediction model based on analysis of interpreter error at a sample cell in the Jintang site, representing the
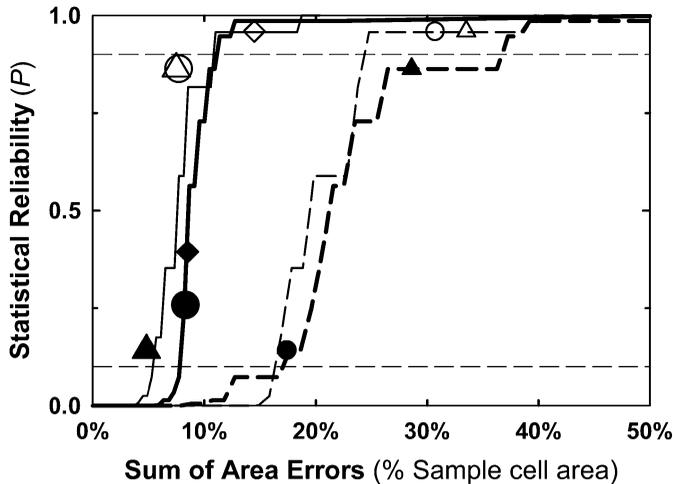
Figure 5. Statistical reliability of error intervals estimated for ecotope areas versus the error produced by the estimator across three interpreters of the Jintang sample cell. Horizontal control lines at $P=0.1$ and 0.9 highlight minimum and greater than expected P for a 90% reliable estimator. Large filled (current) and hollow (historical) symbols represent ecotope-level estimates of CUE ($\bullet$, $\circ$) and CAE ($\blacktriangle$, $\triangle$; tables 1 and 2), small symbols of the same type are CUE and CAE estimators pooled across ecotopes, along with FAE predicted from feature perimeters ($\blacklozenge$, $\diamond$). Solid lines (heavy=current, light=historical) describe error estimated as a factor of feature perimeter (i.e. FAE), dashed lines are error estimated as a factor of feature area (e.g. CUE and CAE).

most challenging current and historical mapping conditions across the sites in this study (Ellis 2004).

Figure 5 illustrates the statistical reliability and sum of errors introduced by different models for estimating error intervals for ecological map class areas across the three interpreters of the Jintang sample cell. This figure demonstrates that all error estimators were acceptable as 90% reliable under both current and historical mapping conditions, but that their reliability ranged from the minimum acceptable to significantly greater than required of a 90% reliable error estimator. As expected, CUE and CAE error intervals estimated for each ecotope class (tables 1 and 2) were more powerful error predictors than CUE or CAE pooled across ecotope classes, and produced less overall error than all other models (figure 5). However, FAE was nearly as powerful an error predictor as ecotope-level CUE and CAE even though FAE was predicted by a pooled error model across ecotope classes (equation (3); figure 5; tables 1 and 2). Surprisingly, historical error estimates were usually more reliable than current estimates, most likely owing to their wider error intervals and smaller numbers of map classes; for every sample cell in this study, historical maps had fewer ecotope classes than current maps.

Error prediction by a simple perimeter-based model (e.g. FAE, equation (3) in section 2.6) was more powerful than prediction by a simple area-based model (error as constant factor of class area, e.g. pooled CUE and CAE), producing far less overall error at all levels of prediction reliability (figure 5). In contrast with area-based errors, perimeter-based errors varied substantially between ecotope classes, with current FAE predictions ranging from <40% to >6 times the overall map error (table 1). Moreover, ecotope-level FAE predictions followed similar trends as those observed in ecotope CUE (linear regression $R^2=0.61$, $P<0.001$), with larger

errors for ecotopes with smaller features and for features with more complex edges (tables 1 and 2; CAE and FAE were unrelated, $P=0.77$). Taken together, these results demonstrate that a simple perimeter-based error model can yield statistically reliable error predictions that approximate observed CUE, a robust error estimator based on the conventional error matrix (Congalton and Green 1999), while introducing much less overall error than simple area-based models.

Analysis of perimeter-based error demonstrated that highly reliable error intervals with a ⩾95% chance of >90% prediction success were produced by error factors of ⩾1.3 m for current maps and ⩾2.7 m for historical maps, introducing overall errors of about 11% for each time period (figure 5, refer to equation (3) in section 2.6). To achieve the same reliability, area-based errors of 39% and 25% were required for current and historical maps, respectively (figure 5). Given that for the same level of prediction reliability, perimeter-based error estimates were sometimes much higher than area-based estimates for the same map classes, a mixed error model utilizing only the lower of the two error estimates for each map class should decrease overall error without sacrificing statistical reliability. When this mixed error model was applied to the Jintang sample cell using the highly reliable error factors noted above, error was reduced in 6 of 60 current ecotope estimates, decreasing overall error from 11.8 to 11.1%, and also decreased error in 17 of 30 historical ecotope estimates while lowering overall error from 11.0 to 9.5%, without any effect on prediction success rates for either time period.

Conservative error models based on interpreter error under the most challenging conditions across sites should be even more reliable when predicting errors at sites with less challenging conditions. We tested this hypothesis by applying the highly reliable error models from the Jintang sample cell to current ecotope maps of a sample cell in Yiyang, Hunan by the same set of interpreters. From 16 ecotope classes mapped by 3 interpreters ($n=48$), the area-based error model yielded 46 successful predictions with a 39% sum of errors across the sample cell (96% chance of >90% reliability), while the perimeter-based error model produced 45 successful predictions and a 13.8% sum of errors (87% chance of >90% reliability). The mixed error model was as reliable as the perimeter-based model, but lowered error in 7 of 48 of estimates, yielding a 12.6% sum of errors across the sample cell. These results confirm that conservative error models based on interpreter error at the Jintang sample cell performed equally well under substantially different mapping conditions at a sample cell more than 800 km distant across rural China.

Statistically reliable ecological map class area estimates are a fundamental part of long-term ecological change measurements in densely populated landscapes (Johnes and Butterfield 2002, Ellis 2004). Though no error model can predict the full range of errors across all mapping conditions, our analysis of interpreter error under the most challenging mapping conditions across sites in rural China facilitated the development of statistically reliable error interval prediction models for map class area estimates across sample cells mapped by a set of trained and calibrated interpreters using our standardized feature-based ecological mapping system. Though our error models are specific to ecological mapping by trained interpreters in rural China, it should be possible to apply a similar error modelling approach, potentially incorporating feature type and image quality, to other standardized feature-based mapping systems, including automated feature detection, providing a practical basis for statistically reliable estimates of long-term ecological changes in landscape structure across densely populated landscapes.

Acknowledgements

References

ANDRÉS, A.M. and MARZO, P.F., 2004, Delta: a new measure of agreement between two raters. *British Journal of Mathematical and Statistical Psychology*, **57**, pp. 1–19.

AUGUST, P., IVERSON, L. and NUGRANAD, J., 2002, Human conversion of terrestrial habitats. In *Applying Landscape Ecology in Biological Conservation*, K.J. Gutzwiller (Ed.) (New York: Springer), pp. 198–234.

CHERRILL, A. and MCCLEAN, C., 1995, An investigation of uncertainty in field habitat mapping and the implications from detecting land-cover change. *Landscape Ecology*, **10**, pp. 5–21.

CHERRILL, A. and MCCLEAN, C., 1999, Between-observer variation in the application of a standard method of habitat mapping by environmental consultants in the UK. *Journal of Applied Ecology*, **36**, pp. 989–1008.

CONGALTON, R.G. and GREEN, K., 1999, *Assessing the Accuracy of Remotely Sensed Data* (Boca Raton, Fl: CRC Press).

CONGALTON, R.G. and MEAD, R.A., 1983, A quantitative method to test for consistency and correctness in photointerpretation. *Photogrammetric Engineering and Remote Sensing*, **49**, pp. 69–74.

DEFRIES, R.S., FOLEY, J.A. and ASNER, G.P., 2004, Land-use choices: balancing human needs and ecosystem function. *Frontiers in Ecology and the Environment*, **2**, pp. 249–257.

EFRON, B. and TIBSHIRANI, R., 1986, Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, **1**, pp. 54–77.

ELLIS, E.C., 2004, Long-term ecological changes in the densely populated rural landscapes of China. In *Ecosystems and Land Use Change*, R.S. Defries, G.P. Asner and R.A. Houghton (Eds) (Washington, DC: American Geophysical Union), pp. 303–320.

ELLIS, E.C., LI, R.G., YANG, L.Z. and CHENG, X., 2000a, Changes in village-scale nitrogen storage in China's Tai Lake region. *Ecological Applications*, **10**, pp. 1074–1089.

ELLIS, E.C., LI, R.G., YANG, L.Z. and CHENG, X., 2000b, Long-term change in Village-scale ecosystems in China using landscape and statistical methods. *Ecological Applications*, **10**, pp. 1057–1073.

ELLIS, E.C., WANG, H., XIAO, H.S., PENG, K., LIU, X.P., LI, S.C., OUYANG, H., CHENG, X. and YANG, L.Z., 2006, Measuring long-term ecological changes in densely populated landscapes using current and historical high resolution imagery. *Remote Sensing of Environment*, **100**, pp. 457–473.

FOLEY, J.A., COSTA, M.H., DELIRE, C., RAMANKUTTY, N. and SNYDER, P., 2003, Green surprise? How terrestrial ecosystems could affect Earth's climate. *Frontiers in Ecology and the Environment*, **1**, pp. 38–44.

FOODY, G.M., 2002, Status of land cover classification accuracy assessment. *Remote Sensing of Environment*, **80**, pp. 185–201.

FORSTER, B.C., 1985, Examination of some problems and solutions in monitoring urban areas from satellite platforms. *International Journal of Remote Sensing*, **6**, pp. 139–151.

GALLEGO, F.J., 2000, Double sampling for area estimation and map accuracy assessment. In *Quantifying Spatial Uncertainty in Natural Resources: Theory and Applications from GIS and Remote Sensing*, H.T. Mowrer and R.G. Congalton (Eds) (Chelsea, Mich.: Ann Arbor Press), pp. 65–77.

GALLEGO, F.J., DELINCE, G. and CARFUGNA, E., 1994, Two stage area frame on squared segments for farm surveys. *Survey Methodology*, **20**, pp. 107–115.

GREEN, D.R. and HARTLEY, S., 2000, Integrating photointerpretation and GIS for vegetation mapping: some issues of error. In *Vegetation Mapping: From Patch to Planet*, R.W. Alexander and A.C. Millington (Eds) (Chichester; New York: Wiley), pp. 103–134.

HILL, R.A., SMITH, G.M., FULLER, R.M. and VEITCH, N., 2002, Landscape modelling using integrated airborne multi-spectral and laser scanning data. *International Journal of Remote Sensing*, **23**, pp. 2327–2334.

HOUGHTON, R.A., 2003, Why are estimates of the terrestrial carbon balance so different? *Global Change Biology*, **9**, pp. 500–509.

IMAGERY RESOLUTION ASSESSMENTS and REPORTING STANDARDS (IRARS) COMMITTEE 1995, *Multispectral Niirs Reference Guide, Imagery Resolution assessments and Reporting Standards*, (IRARS) Committee.

IMAGERY RESOLUTION ASSESSMENTS and REPORTING STANDARDS (IRARS) COMMITTEE 1996, *Civil Niirs Reference Guide, Imagery Resolution Assessments and Reporting Standards*, (IRARS) Committee.

JENSEN, J.R. and COWEN, D.C., 1999, Remote sensing of urban/suburban infrastructure and socio-economic attributes. *Photogrammetric Engineering and Remote Sensing*, **65**, pp. 611–622.

JOHNES, P.J. and BUTTERFIELD, D., 2002, Landscape, regional and global estimates of nitrogen flux from land to sea: errors and uncertainties. *Biogeochemistry*, **57–58**, pp. 429–476.

LO, C.P. and SHIPMAN, R.L., 1990, A GIS approach to land-use change dynamics detection. *Photogrammetric Engineering and Remote Sensing*, **56**, pp. 1483–1491.

LU, D., MAUSEL, P., BRONDÍZIO, and MORAN, E., 2004, Change detection techniques. *International Journal of Remote Sensing*, **25**, pp. 2365–2401.

NORUSIS, M., 2004a, *SPSS 12.0 Guide to Data Analysis* (Upper Saddle River, New Jersey: Prentice Hall).

NORUSIS, M., 2004b, *SPSS 12.0 Statistical Procedures Companion* (Upper Saddle River, New Jersey: Prentice Hall).

PETIT, C.C. and LAMBIN, E.F., 2001, Integration of multi-source remote sensing data for land cover change detection. *International Journal of Geographical Information Science*, **15**, pp. 785–803.

POWELL, R.L., MATZKE, N., DE SOUZA J., C., CLARK, M., NUMATA, I., HESS, L.L. and ROBERTS, D.A., 2004, Sources of error in accuracy assessment of thematic land-cover maps in the Brazilian Amazon. *Remote Sensing of Environment*, **90**, pp. 221–234.

SCHENKER, N. and GENTLEMAN, J.F., 2001, On judging the significance of differences by examining the overlap between confidence intervals. *American Statistician*, **55**, pp. 182–186.

SMITH, J.H., STEHMAN, S.V., WICKHAM, J.D. and YANG, L., 2003, Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, **84**, pp. 342–349.

SNEDECOR, G.W. and COCHRAN, W.G., 1980, *Statistical Methods*, 7th edition (Ames, Iowa, (USA: The Iowa State University Press).

STORY, M. and CONGALTON, R.G., 1986, Accuracy assessment: a user's perspective. *Photogrammetric Engineering and Remote Sensing*, **52**, pp. 397–399.

TAYLOR, J.R., 1997, *An Introduction to Error Analysis: the Study of Uncertainties in Physical Measurements*, 2nd edition (Mill Valley, California: University Science Books).

THOMAS, N., HENDRIX, C. and CONGALTON, R.G., 2003, A comparison of urban mapping methods using high-resolution digital imagery. *Photogrammetric Engineering and Remote Sensing*, **69**, pp. 963–972.

TOWNSHEND, J.R.G., JUSTICE, C.O., GURNEY, C. and MCMANUS, J., 1992, The impact of misregistration on change detection. *IEEE Transactions on Geoscience and Remote Sensing*, **30**, pp. 1054–1060.

TROLL, C., 1939, Luftbildplan Und Ökologische Bodenforschung (Aerial Photography and Ecological Studies of the Earth). *Zeitschrift Der Gesellschaft Für Erdkunde*, pp. 241–298.

VERBYLA, D.L. and BOLES, S.H., 2000, Bias in land cover change estimates due to misregistration. *International Journal of Remote Sensing*, **21**, pp. 3553–3560.

VITOUSEK, P.M., MOONEY, H.A., LUBCHENCO, J. and MELILLO, J.M., 1997, Human domination of Earth's ecosystems. *Science*, **277**, pp. 494–499.

WANG, H. and ELLIS, E.C., 2005a, Image misregistration error in change measurements. *Photogrammetric Engineering and Remote Sensing*, **71**, pp. 1037–1044.

WANG, H. and ELLIS, E.C., 2005b, Spatial accuracy of orthorectified Ikonos imagery and historical aerial photographs across five sites in China. *International Journal of Remote Sensing*, **26**, pp. 1893–1911.