Running Head: Classification Error & LPI error

Map misclassification can cause large errors in landscape pattern indices: Examples from habitat fragmentation

William T. Langford[1], Sarah E. Gergel[2], Thomas G. Dietterich[3], Warren Cohen[4]

[1] corresponding author - National Center for Ecological Analysis and Synthesis, 735 State St., Suite 300, Santa Barbara, CA, USA 93101, Phone (805) 892-2522, Fax (805) 892-2510

[2] Centre for Applied Conservation Research, Department of Forest Sciences, 3008 – 2424 Main Mall, University of British Columbia, Vancouver, BC, Canada V6T 1Z4

[3] Computer Science Department, Oregon State University, Corvallis OR 97331, tgd@cs.orst.edu

[4] USDA Forest Service, Pacific Northwest Research Station, Corvallis OR 97331, Warren.Cohen@orst.edu

April 13, 2005

REVISED SUBMISSION TO ECOSYSTEMS

ABSTRACT

Although habitat fragmentation is one of the greatest threats to biodiversity world-wide, virtually no attention has been given to the quantification of error in fragmentation statistics. Landscape pattern indices (LPIs) such as mean patch size and number of patches are routinely used to quantify fragmentation and are often calculated using remote sensing imagery that has been classified into different land cover classes. No classified map is ever completely correct, so we asked if different maps with similar misclassification rates could result in widely different errors in pattern indices. We simulated landscapes with varying proportions of habitat and clumpiness (auto-correlation) and then simulated classification errors on these same maps. We simulated higher misclassification at patch edges (as is often observed), and then used a 'smoothing' algorithm routinely used on images to correct 'salt-and-pepper' classification error. We determined how well classification errors (and smoothing) corresponded to errors seen in four pattern indices. Maps with low misclassification rates often yielded errors in LPIs of much larger magnitude and substantial variability. While smoothing usually improved classification error, it sometimes increased LPI error and reversed the direction of error in LPIs introduced by misclassification. Our results show that classification error is not always a good predictor of errors in LPIs, and some types of image post-processing (e.g., smoothing) might result in underestimation of habitat fragmentation. Furthermore, our results suggest that there is potential for large errors in nearly every landscape pattern analysis ever published, as virtually none quantify the errors in LPIs themselves.

Key words: fragmentation; landscape metrics; landscape pattern indices; spatial error; classification error; thematic map; accuracy assessment; remote sensing; uncertainty.

INTRODUCTION

Habitat fragmentation is thought to have significant effects on community structure and composition and to be one of the greatest threats to biodiversity (Villard et al. 1999, Terborgh et al. 2001, Benitez-Malvido and Martinez-Ramos 2003, Cordeiro and Howe 2003, Ferraz et al. 2003). Effects of increasing fragmentation can come from an increase in the number of patches, the distance between patches (Saunders et al. 1991), and the amount of edge habitat within each patch (Brittingham and Temple 1983; Andrén and Angelstam 1988, Laurance et al. 2000). Landscape fragmentation is commonly characterized using measures of these and other values such as mean patch size. Because software to compute these landscape pattern indices (LPIs) is widely available, pattern analyses are routinely performed for a variety of habitats and for many different purposes ranging from habitat conservation to regional planning (Cardille and Turner 2001, Turner et al. 2001, McGarigal et al. 2002, Fahrig 2003).

LPIs are often computed over remote sensing images that have been classified into different land cover classes (Skole and Tucker 1993, Peralta and Mather 2000, Griffiths et al. 2000, Imbernon and Branthomme 2001), but there are always errors made in classifying the pixels of an image into land cover classes. This lack of accuracy raises the question of whether these classification errors could lead to substantial errors and variation in the LPIs derived from classified maps (Hess 1994). Despite its apparent importance, virtually no study measures the error in LPIs, much less accounts for the error caused by image misclassification. This oversight may have serious consequences for both science and policy because there is potential for errors of unknown magnitude in nearly every landscape pattern analysis ever derived from classified images.

While a great deal of effort has gone into techniques for assessing classification accuracy

(Congalton and Green 1999, Stehman 2001), it is generally specified in terms of the percentages of misclassified pixels, with no reference to the location of the errors. However, higher classification errors are often associated with textured areas or patch boundaries (Edwards and Lowell 1996, Plourde and Congalton 2003). While several authors have examined other causes of error in LPIs (Turner et al. 1989, Cardille and Turner 2001, Saura and Martinez-Millan 2001), the effect of classification errors on LPIs has rarely been addressed directly. When it has been addressed (Hess and Bay 1997, Wickham et al. 1997, Brown et al. 2000, Shao et al. 2001), differing conclusions have been reached. The conflicting conclusions may be partially due to the fact that each of these studies was based on a single image or landscape, and thus did not control for the impact of landscape structure.

Objective

Here, we use a series of simulated landscapes and classification errors to test the notion that LPI error is necessarily correlated with classification error. Specifically, we address the following two questions:

1) How do image classification errors and levels of smoothing affect LPIs of fragmentation?

2) How do the amount of habitat and its clumpiness affect the error in the LPIs?

Unlike previous work, these simulations allow us to examine both the magnitude and spread in LPI error among landscapes that are strictly controlled for spatial characteristics that might affect the values of the LPIs. While our simulated landscapes and classification errors have elements of real applications to them, it is important to note that we do not claim to build a predictive model of the precise amount of expected LPI error for any particular classified map.

METHODS

Overview

We created a set of 10,800 simulated landscapes designed to represent specific aspects of the variability seen in real landscapes, as well as aspects of the remote sensing image classification process. We first created 270 simulated ‘correct’ base landscapes with two land cover classes: habitat and background. These ‘correct’ base maps represented images where every pixel has been classified as either habitat or background with no error. These maps were designed to vary systematically in the proportion and clumpiness of habitat. We also created ‘incorrect’ misclassified maps from the correct base maps, representing images that have been incorrectly classified to some degree, mimicking two types of common classification errors. We ‘smoothed’ the images using a process routinely applied to correct salt-and-pepper errors on images. Afterwards, we determined the impact of classification errors and smoothing on several LPIs derived from the simulated maps. Finally, we determined whether classification errors were related to the magnitude of and spread in LPI error.

Measuring Classification Error

We refer to two commonly used measures of classification error: producer’s error and user’s error, both measured on a per class basis, as opposed to the total error for all habitat or land cover classes. We chose these measures of classification error because nearly all current research reports classification errors in this way. The producer's error (e_p) is the probability that a true land cover class will be incorrectly mapped and measures the errors of omission. It represents the classification errors that we systematically induced in our maps. In our study with two cover types, producer's error refers to the percentage of habitat pixels in the “correct” map that are incorrectly labeled in the “misclassified” map (i.e., are labeled as background):

$$e_p = fn/(tp+fn) \qquad (1)$$

where fn=false negatives, tp=true positives and we treat “habitat” as the positive class.

We also measure the resulting user's error (e_u) on our maps. It indicates the probability that a pixel from a land cover map does not match the correct (i.e., reference or ground-truthed) land cover class. User's error measures the error of commission. It is defined as the percentage of pixels that are classified as habitat but should be classified as non-habitat, and it is calculated as follows:

$$e_u = fp/(tp+fp) \tag{2}$$

where fp=false positives, tp=true positives.

While both measures are important, we examine our results primarily with respect to user's error, since it intuitively represents how much the user should “believe” that a given pixel is the class that the map claims it is. We introduce producer's errors of 20% on patch edges and 10% in patch interiors on our ‘incorrect’ maps, and then measure the resulting user's error. Thus, the highest producer's error possible on the initial (unsmoothed) simulations is always < 20%, but the user's error can vary widely. We chose the 10% and 20% values primarily so that we could investigate the impact of conservative classification errors. Also, using identical error rates to those of other studies allows for a useful basis for comparison (Wickham et al. 1997).

Map Simulation

Correct Base Maps. We created a total of 270 ‘correct’ base maps to represent stationary, isotropic landscapes with no classification error (see Figure 1 for examples). Each of these two class random map was generated using the program RULE (Gardner 1999). RULE was chosen for several reasons. The mid-point displacement fractal algorithm used by RULE is particularly useful for generating continuous variability in landscape structure because both the proportion of habitat and the clumpiness of habitat can be varied systematically and independently across a range of values. In addition, the behavior of LPIs has been systematically examined on landscapes simulated by RULE and much theoretical work in habitat connectivity is based on landscapes generated using this software (With and King 1997), thus further enhancing the utility of our results.

We varied the proportion of the landscapes occupied by habitat from ten to ninety percent of the landscape in ten percent increments so that we could examine the impact of habitat loss on LPIs. In addition, we varied the Hurst exponent, H, which ranges from 0 to 1 to examine the impact of the clumpiness of fragmentation on our results. A Hurst exponent of 0 represents landscapes that are negatively auto-correlated and an exponent of 1 represents landscapes that are extremely positively auto-correlated. We examined H levels of 0.2, 0.5, and 0.8. Thus, we created 9 proportions of habitat and 3 levels of auto-correlation, with 10 replicates of each, for a total of 270 initial correct base maps. While this does not mimic every possible aspect of landscape structure, varying these two simple parameters created a wide range of landscape structure (Figure 1).

Incorrect and Smoothed Maps. We created 'incorrect' maps from 'correct' maps to mimic two types of classification error: (a) randomly located misclassification ("salt and pepper" error); and (b) increased misclassification near patch boundaries (as compared to the interior of patches). While this is not the only error model possible, random error and edge error are a part of any classifier's error. Moreover, this error model creates errors that may directly influence fragmentation through the generation of spurious patches. It also matches the error model used in Wickham et al. (1997). This was important for our study because we wanted to examine whether the same error model would lead to different conclusions on different landscapes. Other error models, for example, ones with more spatial autocorrelation are left for future studies.

Because our maps contained only two classes, we induced errors in the classification of individual pixels on the correct map by simply changing habitat cells to background, or vice versa, with a given probability (or error rate). This type of error is routinely termed ‘producer’s error.’ We assumed that error rates near patch edges were 10% higher than in patch interiors (e.g., due to mixed pixels, Wickham et al. 1997). For inducing errors, we defined an edge pixel as any pixel with at least one of its four neighbors belonging to the other category. Thus, we set the error rates in the patch interiors at 10% with corresponding edge error rates of 20%. We created 10 incorrect maps from each correct base map, resulting in 2700 total incorrect maps.

We then removed small, isolated regions assumed to be in error by applying a smoothing technique that removes all regions smaller than the size of a designated minimum mapping unit (MMU). Any patch smaller than the size of the minimum mapping unit is removed by changing the classification of its pixels to the classification of the patch with which it shares the longest border. In the case of only two classes, that border is always with a fully surrounding patch. We employed minimum mapping units of 2, 4 and 9 cells to remove all patches below these size thresholds, for a total of 8100 smoothed maps. We chose this method rather than a majority filter because it is simple to implement and because a majority filter tends to remove edge complexity. The sizes were chosen to roughly correspond to a graduated set of patch sizes: a pair of adjacent pixels, a 2x2 square, and a 3x3 square. Similar procedures were used in the classified WISCLAND Land Cover data set, where upland land cover categories (excluding URBAN) were 'smoothed' to patches no smaller than four contiguous pixels using a clump-sieve-fill algorithm, whereas wetlands were smoothed using a minimum mapping unit of two pixels (WISCLAND 1993). Brown et al. (2000) used sieves of size 1, 3, and 10 hectares as well as majority filters of sizes 3x3 and 5x5.

Conservative classification errors. It is unlikely that our simulated errors overestimated classification error rates found in real maps at edges or patch interiors. Reported classification error rates for some commonly studied data sets vary widely depending on land cover class, level of classification and method of accuracy assessment. For example, in the 1992 U.S. National Land-Cover Dataset where accuracy assessment was done extensively and carefully, per class user's accuracies for Anderson Level I classifications for the mid-Atlantic region range from 35% correct to 92% (Stehman et al. 2003). Values for Anderson Level II classifications for the same region range from 1% to 92%. There is also evidence for optimistic bias in the reporting of classification error in many studies (Hammond and Verbyla 1996, Stehman 2001). For the purposes of simulating edge-biased classification error, we have assumed patch edges to be one pixel wide (Wickham et al. 1997). However, true edge widths may be larger in many landscapes (Edwards and Lowell 1996), especially in heavily textured images. Thus, our edge width of one pixel is likely to lead to conservative edge errors.

Measuring Fragmentation

We applied Fragstats 3.0 (McGarigal et al. 2002) to calculate the following LPIs on simulated correct, incorrect, and smoothed landscapes: a) Mean Shape Index, b) Total Edge (m), c) Number of Patches, and d) Mean Patch Size (ha). (Details on the LPIs calculated by Fragstats can be found in the Fragstats documentation at: http://www.umass.edu/landeco/research/fragstats/fragstats.html). Increasing values of mean patch size indicate less fragmentation, whereas increases in the other LPIs are indicative of greater fragmentation. We make no claims about whether these are ecologically informative measures of fragmentation. Rather, we chose these LPIs because they are among the simplest to understand and the most widely employed measures for quantifying habitat fragmentation. They were also hypothesized to be among the most affected by edge misclassification.

Analysis of Results

For simplicity, we considered only binary maps (with only two land cover classes) and only examined the results for one class. Although real maps may have more than one class, the LPIs of fragmentation we examined are calculated on a ‘per class’ basis (e.g., for one habitat type only, assuming all other habitat types are background). As long as the LPI of interest does not distinguish among the different classes that make up the background class, the value of the LPI will be identical in the multi-class and the combined binary versions of the map.

It is also very important to examine classification error on a per class basis rather than computing total classification error for the landscape as a whole because a classifier can have low accuracy for one class of interest, but still have high overall accuracy when averaged across all classes. For example, the total classification error can be small and misleading when a map has low classification errors for abundant classes, and high classification errors for less abundant classes. Thus, examining classification error and LPI error both on a ‘per class’ basis makes the most sense for this study.

For each LPI on the incorrect and smoothed landscapes, we determined the percent error in the LPI relative to the correct landscape, which we termed % LPI Error.

$$\%\ \text{LPI Error} = ((\text{LPI}_{incorrect} - \text{LPI}_{correct}) / \text{LPI}_{correct}) * 100 \qquad (3)$$

We plotted % LPI Error versus User's Error for the initial misclassifications as well as for each level of smoothing at different minimum mapping units. While we demonstrate the spread in LPI error within each user's error value quite clearly using these graphs, we do not compute a standard deviation or variance of LPI errors. This would require arbitrarily binning data points (by base landscape, proportion, clumpiness, or user’s error, etc.) when our purpose is simply to examine whether for any given user's error, there is a large spread of values of LPI error.

RESULTS

First, we summarize the general behavior of our landscape model and error model by examining the relationship between producer's error (which we manipulated) and the resulting user's error. Second, we examine the impact of the classification errors on the computed LPIs to determine if smoothing affected LPI error. We discuss the results for each LPI in detail. Lastly, we examine the relative impact of landscape structure (clumpiness and habitat loss) on LPI error.

To fully understand our results, it is important to explain how our error model and our landscape model are related (Figure 2). Even though our simulated producer's errors in habitat classification were limited to 10 – 20%, it resulted in a larger range of user's errors, between 0 and 60% (Figure 3). User's errors were generally higher in the incorrect landscapes (before smoothing) with low proportions of habitat (Figure 3). However, ‘smoothing’ of an image reduced maximum user's error by roughly half, from around 50% in the incorrect images to less than approximately 25% when smoothed (MMU=9) (Figure 3). We have shown the full range of values for user's and producer's error in Figure 3 to show the effects of smoothing on those errors. In Figures 4 and 5 though, we have only shown values for images where both user's and producer's error are less than or equal to 15%. While we computed results for all of the simulated landscapes that we generated, we have chosen to only show results here for maps with the most conservative amounts of classification error. Even though classification rates are often higher than 15% and our LPI errors for those classification error rates were even more extreme, we have done this so that we are certain not to display results for classification error rates that any users would reject as being too high to use in a real application.

How do image classification errors and levels of smoothing affect LPIs of fragmentation?

Classification errors often resulted in large errors in LPIs, even at classification error rates considered low by the remote sensing community. While the relationship between the values of LPIs on the correct versus the incorrect and smoothed maps was roughly linear for Total Edge (Figure 4e-h), it was not for the other LPIs. For the Number of Patches, the relationship between the correct and incorrect values depends on the minimum mapping unit (MMU) used for smoothing. It was linear only for MMU=4 and 9. When examining percent error in LPIs relative to the correct image, the errors in LPIs were extremely high (Figure 5), in a number of cases, higher than 1000% (Figures 5e,f,i,j,p). User's error did not reliably correspond to LPI error for any LPI examined, and therefore, it was not a useful predictor of LPI error. Smoothing did not consistently reduce the magnitude of errors in LPIs relative to the original correct classification. However, smoothing using MMU=9 always reversed the direction of error caused by the initial misclassification (overestimation changes to underestimation and vice versa) for all the LPIs examined (Figure 5). Thus, while smoothing to remove salt and pepper error improved per class user's error, it sometimes increased LPI error (Figure 5). We next examine each LPI in greater detail.

LPI errors for Mean Shape Index were generally less than ±10% in the incorrect landscapes (Figure 5a), but increased to 50% with smoothing (Figure 5b-d). The errors were not consistently biased in one direction, as the LPI was underestimated for the incorrect landscapes (5a), but overestimated for the MMU=4 (5c) and 9 landscapes (5d). However, the lower user's errors in the smoothed landscapes still yielded both higher magnitude and greater spread in LPI error within a given user's error (Figure 5d).

For Total Edge, the maximum percentage LPI error approached 4000% (Figure 5e). Smoothing reduced the LPI error to the range between +150% and –60% (Figure 5g-h), but the direction of error differed depending on the minimum mapping unit. Total Edge was overestimated on the initial incorrect, MMU=2, and MMU=4 landscapes (Figure 5e-g) and underestimated by more than 50% for the most fragmented landscapes when MMU=9.

The percent error in the Number of Patches (NP) relative to the correct image ranged from approximately 0 to 10,600%. The behavior of NP was qualitatively similar to that of Total Edge, where the LPI was overestimated for the initial incorrect landscape and for the MMU = 2 landscapes (Figure 5i-j), but then underestimated by as much as 100% for MMU=9 (Figure 5*l*).

For Mean Patch Size (MPS), the initial misclassifications resulted in underestimation of the LPI by as much as 100% (Figure 5m). When smoothed using MMU=2, MPS was both over- and under-estimated (Figure 5n). Successive smoothing using MMU=4 and 9 resulted in increases in LPI error, to more than 1000% for MMU=9 (Figure 5p).

Our results suggest that the magnitude and spread of LPI errors due to misclassification can be quite large. Moreover, LPI errors do not always decrease with spatial post-processing techniques, such as smoothing, that are routinely applied to reduce user's error. In short, the spatial arrangement of classification error is at least as important as the amount. LPI errors are not always lower on maps with lower user's errors, and maps with the same user's error can frequently result in maps with LPI errors of very different magnitudes.

How do the amount of habitat and its clumpiness affect the error in the LPIs?

The magnitude of LPI errors was affected by the structure of the landscape, but it was affected more by the clumpiness (H, the auto-correlation parameter) than by the proportion of habitat. This was especially evident for Total Edge, Number of Patches, and Mean Patch Size before smoothing (Figure 5e.i.m.n). For Total Edge, LPI error was lowest for H=0.2 (dispersed), and much greater when H=0.8 (clumpy), approaching 4000% error (Figure 5e). Qualitatively similar patterns were evident for Number of Patches. The effect of landscape clumpiness on error in Mean Patch Size was dependent on MMU. Before smoothing, the smallest LPI errors were roughly -50% in the H=0.20 landscapes (Figure 5m), and LPI errors were as large as -100%. When smoothed, the largest LPI errors were in the H=0.20 landscapes (Figure 5o-p), and the lowest LPIs errors were in clumped landscapes (with H=0.80). While the proportion of habitat had less effect on our LPI error, it is correlated highly with user's error (Figure 3), and thus our results must be seen in that context. The lowest user's errors always occurred in the landscapes with the highest proportion of habitat, but this was often the region where the highest LPI errors occurred.

DISCUSSION

We are aware of only four other studies addressing the effects of classification error on LPI error and all but one of these studies were based on either a single image or landscape and did not control for the impact of landscape structure. The one study that does use more than one landscape and does control for landscape structure is Hess and Bay (1997). Like our study, they examined the effect of map error on several LPIs. Unlike our study, their indices were non-spatial ones, including percent cover and two commonly used diversity indices. They used three levels of map error and found that they did cause error in the measures. They further described a method to put confidence intervals around the measures. Of the remaining three studies, only one measured both accuracy and variability in the LPIs. Wickham et al. (1997) simulated classification errors over portions of a land cover map derived from a single TM image. As in our study, they computed the difference in the values of several LPIs between the original base map and the maps with simulated classification errors. They concluded that for the landscape and methods they used, classification error did not increase LPI error. Instead of simulating errors, Shao et al. (2001) had 23 different interpreters classify a single TM image and then measured the variation in LPIs over the 23 different maps. They found that even though there was not much variation in the accuracy of the classifications, there was a great deal of variation in the LPI error. The third related study is that of Brown et al. (2000) in which they examined estimates of error in pattern indices used for change detection. They classified images for two forested areas, subsetted the two areas into many smaller landscapes and then compared the LPI values obtained in overlapping areas of adjacent images photographed around the same time. They found that the LPIs of mean patch size and number of patches were more error prone than the edge density LPI.

Our experiments demonstrate that classification error is not always a reliable predictor of LPI error, and that one cannot assume a map with low classification error will produce accurate LPIs. In our simulated landscapes, the spread in LPI error was generally large, and the magnitude in LPI error was almost always much larger than the magnitude of classification error, even for small classification errors. Also, the common practice of reducing classification error by spatially smoothing a classified map using a minimum mapping unit sometimes increases LPI errors. This result is also consistent with the results in Brown et al. (2000) which is based on empirical data. The consistent underestimation of number of patches and consistent overestimation of total edge and mean patch size in smoothed classifications suggest that landscape fragmentation may be routinely underestimated as a result of smoothing classifications and should be investigated further. Since the landscapes and error models in this paper are artificial, the specific quantitative results obtained here do not generalize to all landscapes and error matrices. However, our results show large and unpredictable amounts of error in LPIs on images with very low classification error by the standards of the remote sensing community. In addition, our results show that different smoothing and/or classification techniques may be recommended for reducing errors in different LPIs even on the same image. In the following sections, we discuss these points in greater detail.

How do image classification errors and smoothing techniques affect LPIs of fragmentation?

Our results suggest that the accuracy of LPIs derived from classified images may not be easily predicted from the accuracy of the classification. In our simulated landscapes, the magnitude of LPI error was nearly always a great deal larger than the magnitude of the corresponding classification error, even when classification error was small. While there is no reason to expect that the value of the LPI error should *equal* the value of the classification error, we often found that the errors were not even within the same order of magnitude. Moreover, the large spread in LPI error at all levels of classification error (shown in Figure 5) means that we cannot fit any single function from classification error to LPI error in the conditions that we have simulated. We cannot even guarantee the weak criteria that the rankings of maps by classification error would yield the same results as a ranking based on LPI errors.

Predicting LPI error from classification error was further confounded by the fact that the common practice of using minimum mapping units to reduce salt and pepper classification error actually increased LPI error for mean patch shape and mean patch size. Moreover, it always reversed the direction of error (from over to underestimation or vice versa). Further, the ideal smoothing levels for reducing LPI error varied by LPI (Figure 5). For example, the smallest LPI error values and spread for Mean Patch Shape occurred in scenarios with no smoothing, while the best smoothing level for Total Edge was MMU=9; for Number of Patches, MMU=4; and for Mean Patch Size, MMU=2. This suggests that minimizing error for each LPI may require using different maps derived from different classifiers or smoothing levels for each LPI.

How do the amount of habitat and its clumpiness affect the error in the LPIs?

The proportion of habitat had less effect on LPI error than habitat clumpiness did. As the smoothing MMU increased, clumpiness had less effect on LPI errors for Mean Patch Shape and Mean Patch Size, but greatly affected LPI errors for Total Edge and Number of Patches. While habitat proportion was not related closely to LPI error in our simulations, it had a large effect on the difference between user's error and producer's error generated by our model. User's errors were generally higher in low proportion landscapes (Figure 3). This error pattern would not be unreasonable to expect using any classifier for several reasons. First, there are likely to be relatively more edge and mixed pixels in lower proportion landscapes, exactly the type of pixels where more classification error is expected. Second, the small sample size of pixels in low proportion landscapes results in fewer training data available for training a classifier. Lastly, higher user's error in low proportion landscapes is likely simply given the equation for User's Error: fp/(tp+fp). In a low proportion landscape, this equation would be dominated by the false positives, given the small number of true positives possible even if 100% of habitat was recognized correctly. In future experiments, we could totally eliminate this effect by writing a more complicated error model that draws pixels in a different way to control for user's error as well. We did not rerun the experiments here with the more complicated model because our conclusions remain the same even if we remove all points with user's error greater than 15%.

Bounds of Generalization Our primary goals were to characterize a certain type of classification error over a range of landscape structures and to test the hypothesis that a map with small classification error necessarily yields spatially reliable LPIs. Our results should not be generalized quantitatively to all landscapes, all LPIs, or all types of classification error. Our results are limited to the landscape structures, LPIs, smoothing and classification error models tested. In spite of these limitations, we argue that the scenarios we have presented have enough realism to reasonably reflect some of the complexity of a real application, and demonstrate that this complexity can generate subtle and counterintuitive outcomes.

Just as we recommend not over-generalizing from our results, our study strongly illustrates the problem in drawing general conclusions based on classifications derived from one original base landscape or image. As an example, even though our experiments employed classification error assumptions similar to those of Wickham et al. (1997), our study yielded much larger errors in LPIs not only because we used different LPIs, but more importantly we used a range of different landscape structures.

Simulations vs. Real Landscapes The danger of over-generalizing from a single map is made particularly clear by the use of multiple simulated landscape maps that we have presented here. Simulated landscapes are often employed to directly control, manipulate, and replicate features of landscape structure (Gardner et al. 1987, Gardner et al. 1991) such as the proportion of habitat types or clumpiness (With and King 1997, Turner et al. 2001, Gergel 2002). Controlling for the variability in real landscapes is extremely problematic because they vary widely in many of these features that influence the behavior of LPIs. Often, they are also highly anisotropic and non-stationary (containing different gradients of clumpiness or auto-correlation, and in different directions). This makes their patterns uneven within one landscape. Thus, any subset of real landscape imagery used for a study such as this is unlikely to systematically span the range of proportions or variability in auto-correlation. Furthermore, the proportions of different habitat types on a remote-sensed image are ultimately dependent on the spatial resolution and on the level of thematic resolution of land cover classes chosen to create the map (e.g., 3 vs. 5 cover types).

Equally important is that when using a classified image as the ‘correct’ base landscape for simulating classification errors, the ‘correct’ patch boundaries and LPI values cannot be known due to LPI errors already introduced by the classifier. The range of LPI values on the initial correct map (if obtained this way) would incorporate the classifier's bias and need not represent the true range of LPI values. For example, smoothing tends to reduce the complexity of patch edges and the number of small patches.

We argue that until we understand the behavior of LPIs in highly controlled and replicated scenarios, we cannot expect to understand or predict their behavior on real maps which vary in many uncontrolled ways.The numerous sources of error and variability make it difficult to isolate the factors influencing the behavior of LPIs on real landscapes or those from any classified image. With simulated landscapes, we can systematically vary proportion and spatial autocorrelation and we can generate multiple realizations of these landscapes using the same parameters.

Error models. One limitation in our error model is that spatial structuring of error is not directly accounted for beyond edge effects. Directly modeling spatial autocorrelation in the errors might have different results and should be investigated since many errors in our model come from the splitting off of small patches. However, there are several factors that suggest that results derived from this model are still of interest. First, our experiments showing smoothing with a minimum mapping unit of 9 cells do address a certain amount of autocorrelation because there is spatial grouping of errors when no patch can be smaller than 9 cells. Second, as the spatial resolution of imagery used increases, there is less averaging of the signal; consequently, the local characteristics of the image are less uniform. Per pixel classifiers that do not take this texture into account are more likely to generate varying classifications of neighboring pixels. These classifications are likely to require post-classification smoothing similar to that employed in our simulations, possibly with similar results. Third, the amount of salt and pepper error that occurs in a classified image is partly a function of the skill of the operator. The experiments in Shao et al. (2001) with multiple operators showed many different outcomes even given exactly the same input. Anyone with access to an image processing program can push a button to classify an image and then compute LPIs on that output. Users whose primary expertise is not in image processing, may have little idea how to deal with issues such as texture in an image and produce speckled classifications which they may or may not smooth afterwards. In either case, their errors may be qualitatively similar to those expressed in our error model. Finally, our results have shown that even if there is less than 1 or 2% error of this kind, it can still generate huge amounts of error in some LPIs (e.g., see Figure 5p).

Normalization of LPI errors for comparison Another important issue is that the ability to compare and rank different LPIs by error requires a fair method for normalizing LPI errors. Throughout this paper we have examined LPI error as a percentage of the correct LPI value, but the total *possible* percent error can differ greatly among LPIs. When the range of an LPI is infinite or allows values to become infinitessimally small (i.e., arbitrarily close to zero), there will be no upper bound on the percentage error possible. In contrast, any LPI whose magnitude is bounded above and below and does not include any neighborhood of zero will have bounds on the percentage error it can possibly attain. It may thus appear better (lower error) than an LPI with errors that can be unbounded. For example, the fractal dimension of a planar region can only vary between 1 and 2. Consequently, it can never have an error greater than 100% because the denominator can never be less than one and the numerator can never be greater than one. Similarly, an LPI may be *effectively* bounded if it is very difficult or uncommon to observe values outside a narrow range (e.g., the Shape Index in our studies). In either case (actual or effective bounding of error), it is inappropriate to directly compare the magnitude and variation of the error of bounded LPIs with unbounded ones. The bounded LPI may appear to have very little percentage error, but this may be simply because its range is so small that the ecologically relevant distinctions in the LPI's values require high precision that is irrelevant in the unbounded LPI.

One way to normalize the errors would be to normalize them against what is considered to be the smallest ecologically relevant distinction in the application of interest. For general comparison of LPIs however, it would be useful to have a normalization method that was more independent of specific applications. One way to do this would be to compare the errors to the largest error "reasonably possible" for that LPI. For example, we could assume that a classifier should not do any worse than randomly guessing classifications. More research is necessary to determine what are the appropriate characteristics and candidates for a normalization method to allow comparison of LPI errors. However, it is clear that any future studies that attempt to *rank* LPIs in terms of LPI error will need to normalize their error measures in some way so that the results are meaningful.

Accuracy Assessment Our work suggests that both the creators and the users of classified images need to do more to measure spatial aspects of classification errors in maps. In particular, better techniques for accuracy assessment of LPIs need to be developed for several reasons. First, as we have shown, non-spatial measures of accuracy (user's error) do not necessarily correspond to errors in spatially-explicit LPIs. To date, studies that do measure spatial aspects of classification error (Foody 2002, Hagen 2003, Pontius et al. 2004), do not examine how those error measures relate to the errors in the LPIs. Second, many users of classified remote sensing imagery are often unlikely or unable to verify the spatial accuracy of classifications themselves due to budget or expertise or because not all of the original reference data is available to the user (e.g., because of agreements with land owners covered by the map). In such cases, the producers of classified imagery would help their users by evaluating spatial characteristics of classification error instead of just counting misclassified pixels.

However, it is not clear whether it is possible to develop predictive models of LPI error that can be generalized to real-world applications. The reason for this is that the pattern of errors in any classified map is the result of complex interactions among many factors. These factors include the classification algorithm, its training data, the user's skill, the landscape, the class structure (e.g., relative abundance of classes, relative importance of different types of errors, patch shapes, patch interlacing, anisotropy) as well as the spatial distribution of classification errors (e.g., location of easily confused classes relative to each other). It may be more useful to develop project-specific predictive models for LPI error based on image characteristics, as Brown et al. (2000) did for change detection in LPIs in forest cover maps. At a minimum however, analogous to classification error analysis, subsamples of the images of interest should have patch boundaries identified through some more reliable method of assessment. These subsamples could then be used to provide a standard for assessing the correctness of spatial measures derived from their classification.

In this paper, we do not mean to suggest that deriving a more accurate value of the LPI as defined for the given single landscape will solve all problems with LPIs. Even if we could identify the correct value of an LPI from a classified map, there are other issues about whether that value has ecological meaning. For example, if a landscape represents a stochastic process, the many possible realizations of that process may each yield different values for an LPI and in that sense, there is no single correct value for the LPI. Questions such as this have led some researchers to advocate approaches based on stochastic models (Remmel et al. 2002, Fortin et al. 2003). For example, to determine whether the LPI values of two landscapes are significantly different, they suggest measuring the overlap of confidence intervals on the LPI values obtained over a large number of realizations from an ecologically relevant stochastic model. These questions are beyond the scope of this paper, however, they raise the following important point.

The stochastic modeling approach requires the estimation of model parameters such as proportion and autocorrelation from the same erroneously classified maps of land cover used to calculate LPIs. Therefore, the stochastic modeling approach also requires an analysis of model errors and LPI errors induced by estimating model parameters from maps containing classification errors. In short, *any* process that computes spatial results based on a classified image must assess the errors in the results themselves rather than assuming that the classification error is a reliable estimate of the amount of error in the results.

CONCLUSION

Several important considerations result from this work that merit further attention. In particular, fragmentation statistics may be suspect regardless of the accuracy of the classifications on which they were based, as quite small errors in classification can lead to quite high errors in LPIs. This has implications not only for forest fragmentation statistics, but for wetland and riparian zones as well. Many states in the U.S. are currently spending millions of dollars developing GIS coverages and other spatial databases to quantify wetland abundance and wetland losses. Even simple measures such as “the number of wetlands” and “average wetland size” may require further accuracy assessment in light of the results presented here. Errors in LPIs for all habitat types deserve further consideration, particularly as remote sensing data becomes available at ever increasing levels of resolution. Such imagery provides an unprecedented opportunity to quantify and monitor all habitats, but especially, smaller habitats such as ephemeral wetlands and narrow, linear riparian zones which are often missed in routine mapping using coarser scale data sources. The linear nature of riparian habitats might render them particularly sensitive to errors in classification – as a misclassified pixel or two could shatter one long contiguous patch into several smaller patches. While presenting valuable new sources of data at the resolutions needed for many ecological applications, spatial measures derived from classifications of such data sources must be used with care. A recent review of papers on habitat fragmentation discussed the challenges in discriminating the impact of habitat loss from that of habitat fragmentation on biodiversity, and found that the impact of fragmentation was weaker and less consistent than that of habitat loss (Fahrig 2003). This conclusion is interesting in light of the fact that errors in measuring habitat *loss* are routinely quantified (via classification errors), whereas errors in measuring landscape *pattern* are generally not quantified, and may be substantial. The kinds of errors found in LPIs in our study suggest there may be some danger that fragmentation may be routinely underestimated as a result of smoothing and we recommend this possibility be investigated further.

In summary, our work shows that the spatial arrangement of classification errors affects the amount of error in LPIs. Classification errors often resulted in large errors in LPIs, even at classification error rates considered low by the remote sensing community. The amount of map classification error is not necessarily a reliable predictor of LPI error. One cannot assume that a map with low classification error will produce relatively accurate LPIs. This suggests that the results of any fragmentation study where LPI errors have not been measured may be incorrect, and potentially to a large degree. No one would classify an image and then claim that the classifications were accurate without testing that claim, yet virtually no study using LPIs derived from a classified image actually measures the accuracy of the LPIs derived from those maps. More emphasis must be placed on evaluating the sources and spatial nature of error in land cover data used for conservation purposes, because over- or under-estimation of the degree of fragmentation can have significant impacts on both public policy and scientific conclusions related to habitat fragmentation. Simply saying that it is difficult or expensive to measure LPI accuracy will not correct erroneous conclusions derived from faulty LPI values.

ACKNOWLEDGEMENTS

The authors thank D. Brown, J. Cardille, Y. Carmel, M-J. Fortin, G. Hess, M. Kinnaird, C. McCain, J. Parrish, K. Phrodite, V. Radeloff, and D. Vazquez for valuable comments on this paper. We also thank an anonymous reviewer of an earlier version of this paper for comments on normalization methods. W.T.L. was supported by NSF grants IRI-9204129, IRI-9626584, ITR-0085836, ONR grant number N00014-95-1-0557, AFOSR grant number F49620-98-1-0375, the NASA Terrestrial Ecology Program, EPA STAR fellowship number U 915196-01-1. Part of this work was conducted while W.T.L. and S.E.G. were Postdoctoral Associates at the National Center for Ecological Analysis and Synthesis, a Center funded by NSF (Grant #DEB-0072909), the University of California, and the Santa Barbara campus.

LITERATURE CITED

Andrén, H. and P. Angelstam. 1988. Elevated predation rates as an edge effects in habitat islands: experimental evidence. Ecology 69:544-547.

Benitez-Malvido, J. and M. Martinez-Ramos. 2003. Impact of forest fragmentation on understory plant species richness in Amazonia. Conservation Biology **17**: 389-400.

Brown, D. G., J. Duh, and S. A. Drzyzga. 2000. Estimating error in an analysis of forest fragmentation change using North American landscape characterization (NALC) Data. Remote Sensing of Environment **71**:106-117.

Brittingham, M.C. and S. A. Temple 1983. Have cowbirds caused forest songbirds to decline? Bioscience 33:31-35.

Cardille, J.A. and M.G. Turner. 2001. Understanding landscape metrics I. Pages 85-100 in *Learning Landscape Ecology: A Practical Guide to Concepts and Techniques*. S.E. Gergel and M.G. Turner, editors. Springer-Verlag, New York, New York, USA.

Congalton, R.G., and K. Green. 1999. Assessing the Accuracy of Remotely Sensed Data: Principles and Practices. Lewis Publishers, Boca Raton, Florida, USA.

Cordeiro, N. J. and H. F. Howe. 2003. Forest fragmentation severs mutualism between seed dispersers and an endemic African tree. PNAS **100**: 14052-14056.

Edwards, G., and K. E. Lowell. 1996. Modeling Uncertainty in Photointerpreted Boundaries. Photogrammetric Engineering and Remote Sensing **62**:377-391.

Fahrig, L. 2003. Effects of habitat fragmentation on biodiversity. Annual Review of Ecology Evolution and Systematics **34**:487-515.

Ferraz, G. et al. 2003. Rates of species loss from Amazonian forest fragments. PNAS **100**: 14069-14073.

Foody, G. 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment **80**: 185-201.

Fortin, M. J., B. Boots, F. Csillag, and T.K. Remmel. 2003. On the role of spatial stochastic models in understanding landscape indices in ecology. Oikos **102**: 203-212.

Gardner, R.H. 1999. RULE: map generation and spatial analysis program. Pages 280-303 in J.M. Klopatek and R.H. Gardner, editors. *Landscape Ecological Analysis: Issues and Applications.* Springer-Verlag, New York, New York, USA.

Gardner, R. H., B. T. Milne, R. V. O'Neill, and M. G. Turner. 1987. Neutral models for the analysis of broad-scale landscape patterns. Ecosystems **1**:19-28.

Gardner, R. H., and R. V. O'Neill. 1991. The use of neutral models for landscape analysis. Pages 289-307 *in* M. G. Turner and R. H. Gardner, editors. Quantitative Methods in Landscape Ecology: The Analysis and Interpretation of Landscape Heterogeneity. Springer-Verlag, New York, New York, USA.

Gergel, S.E. 2002. Cumulative impact of levees and dams on the duration of temporary floodplain ponds: a terrain model approach for assessing multiple disturbances at broad scales. Ecological Applications **12**:1740-175.

Griffiths, G. H., J. Lee, and B. C. Eversham. 2000. Landscape pattern and species richness; regional scale analysis from remote sensing. International Journal of Remote Sensing **21**:2685-2704.

Hagen, A. 2003. Fuzzy set approach to assessing similarity of categorical maps. International Journal of Geographical Information Science. **17**: 235-249.

Hammond, T. O., and D. L. Verbyla. 1996. Optimistic bias in classification accuracy assessment. International Journal of Remote Sensing **17**:1261-1266.

Hess, G. 1994. Pattern and error in landscape ecology: A commentary. Landscape Ecology **9**:35.

Hess, G. R., and J. M. Bay. 1997. Generating confidence intervals for composition-based landscape indexes. Landscape Ecology **12**:309-320

Imbernon, J., and A. Branthomme. 2001. Characterization of landscape patterns of deforestation in tropical rain forests. International Journal of Remote Sensing **22**:1753-1765.

Laurance, W. et al. 2000. Rainforest fragmentation kills big trees. Nature **404**: 836.

McGarigal, K., S. A. Cushman, M. C. Neel, and E. Ene. 2002. FRAGSTATS: Spatial Pattern Analysis Program for Categorical Maps. Computer software program produced by the authors at the University of Massachusetts, Amherst. Available at the following web site: www.umass.edu/landeco/research/fragstats/fragstats.html

Peralta, P., and P. Mather. 2000. An analysis of deforestation patterns in the extractive reserves of Acre, Amazonia from satellite imagery: a landscape ecological approach. International Journal of Remote Sensing **21**:2555-2570.

Plourde, L., and R. G. Congalton. 2003. Sampling method and placement: how do they affect the accuracy of remotely sensed maps? Photogrammetric Engineering and Remote Sensing **69**:289-298.

Pontius, G., Huffaker, D., and K. Denman. 2004. Useful techniques of validation for spatially explicit land-change models. Ecological Modelling. **179**: 445-461.

Remmel, T.K., F. Csillag, S.W. Mitchell, and B. Boots. 2002. Empirical distributions of landscape pattern indices as functions of classified image composition and spatial structure. Proceedings of Symposium on Geospatial Theory, Processing and Applications, Ottawa 2002.

Saunders, D.A, Hobbs, R.J., and C.R. Margules. 1991. Biological consequences of ecosystem fragmentation - a review. Conservation Biology 5(1): 18-32.

Saura, S., and J. Martinez-Millan. 2001. Sensitivity of Landscape Pattern metrics to Map Spatial Extent. Photogrammetric Engineering and Remote Sensing **67**:1027-1036.

Shao, G., D. Liu, and G. Zhao. 2001. Relationships of Image Classification Accuracy and

Variation of Landscape Statistics. Canadian Journal of Remote Sensing **27**:33-43.

Skole, D., and Tucker, C. 1993. Tropical Deforestation and Habitat Fragmentation in the Amazon: Satellite Data from 1978 to 1988. Science **260**:1905-1910.

Stehman, S. V. 2001. Statistical Rigor and Practical Utility in Thematic Map Accuracy Assessment. Photogrammetric Engineering and Remote Sensing **67**:727-734.

Stehman, S. V., J. D. Wickham, J. H. Smith, and L. Yang. 2003. Thematic accuracy of the 1992 National Land-Cover Data for the eastern United States: Statistical methodology and regional results. Remote Sensing of Environment **86**:500-516.

Terborgh, J. et al. 2001. Ecological Meltdown in Predator-Free Forest Fragments. Science **294**: 1923-1926.

Turner, M.G., R.H. Gardner, and R.V. O'Neill. 2001. Landscape ecology in theory and practice: pattern and process. Springer-Verlag, New York, New York, USA.

Turner, M. G., R. V. O'Neill, R. H. Gardner, and B. T. Milne. 1989. Effects of changing spatial scale on the analysis of landscape pattern. Landscape Ecology **3**:153-162.

Villard, M.-A., M. K. Trzcinski, and G. Merriam. 1999. Fragmentation Effects on Forest Birds: Relative Influence of Woodland Cover and Configuration on Landscape Occupancy. Conservation Biology **13**: 774-783.

Wickham, J. D., R. V. O'Neill, K. H. Riitters, T. G. Wade, and K. B. Jones. 1997. Sensitivity of Selected Landscape Pattern metrics to Land-Cover Misclassification in Land-Cover Composition. Photogrammetric Engineering and Remote Sensing **63**:397-402.

WISCLAND (The Wisconsin Initiative for Statewide Cooperation on Landscape Analysis and Data) Land Cover Grid. 1993. User’s Guide to WISCLAND Land Cover Data. Wisconsin Department of Natural Resources.

With, K. A., and A. W. King. 1997. The use and misuse of neutral landscape models in ecology. Oikos **79**:219-229.

FIGURES

Figure 1. Examples of simulated ‘correct’ landscapes. Landscapes were designed to represent a range of landscape structural features: 3 levels of proportion (20%, 40%, 60% remaining habitat) are shown with varying levels of spatial auto-correlation (H = 0.2, 0.5, 0.8). Habitat is shown in white, non-habitat background in black. These nine proportions were used in this study, with 3 levels of autocorrelation, and 10 replicates of each scenario for a total of 270 correct maps.

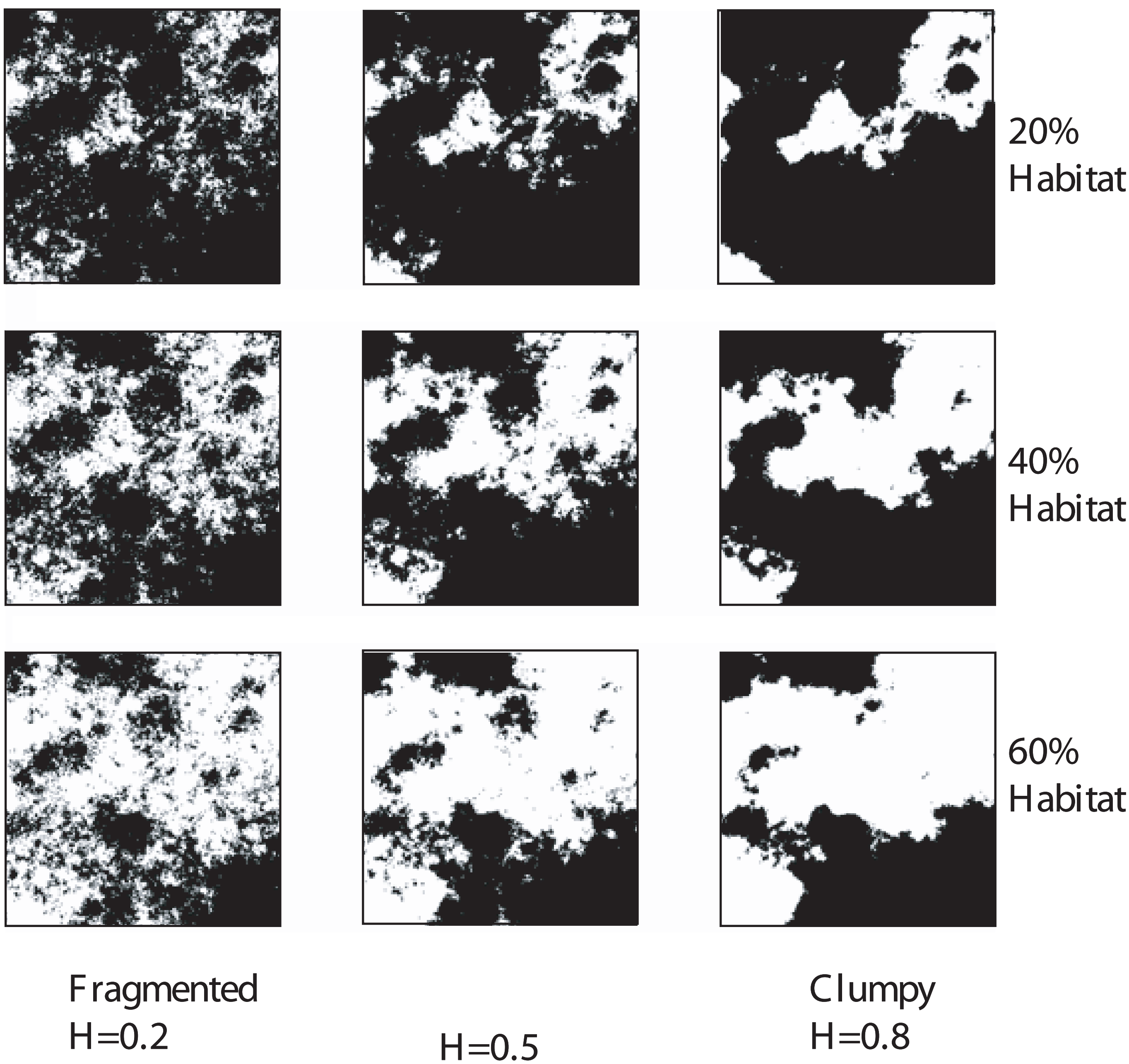
Figure 2. Examples of simulated classification errors and smoothing. In this example, a correct landscape is shown (40% habitat, H=0.5) at the top. The 'incorrect' map is derived from the correct map but with simulated classification errors of 20% at patch edges and 10% in patch interiors. Smoothing to remove salt-and-pepper error is shown for 3 minimum mapping units (MMU = 2, 4, and 9) whereby all patches smaller than the MMU are reverted to the class of the matrix surrounding the patch.

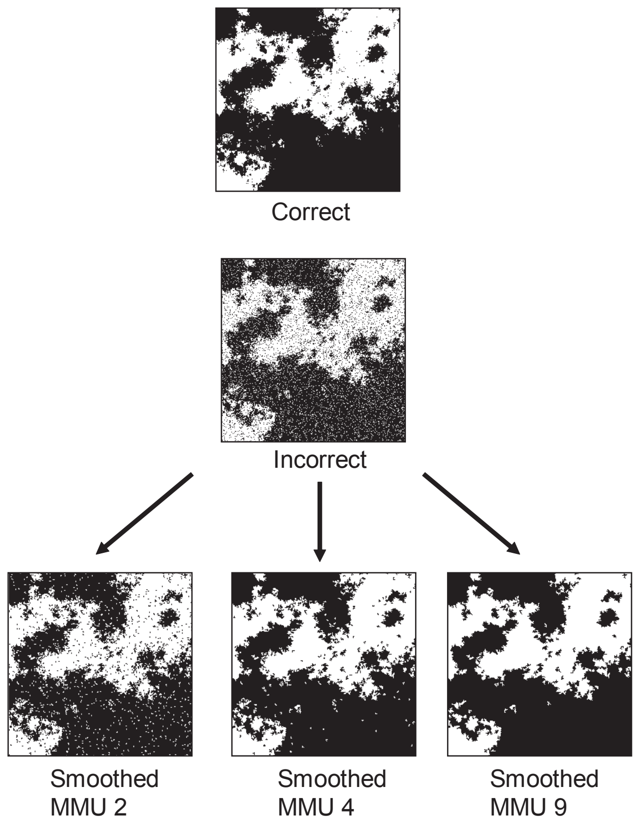
Figure 3. User's error vs. producer's error for the Incorrect and Smoothed landscapes at different MMUs. Top panel shows relationship with habitat proportions labeled and grouped into low (10-30%) medium (40%-60%) and high (70-90%) levels. Bottom panel shows the same data points labeled with the 3 levels of spatial auto-correlation (see Figure 1b for details). Points on smoothed landscapes are labeled according to the H level of the corresponding original correct landscape. Horizontal and vertical lines on the plots mark 15% user's error and producer's error respectively. Figures 4 and 5 only show results from inside this 15% region.

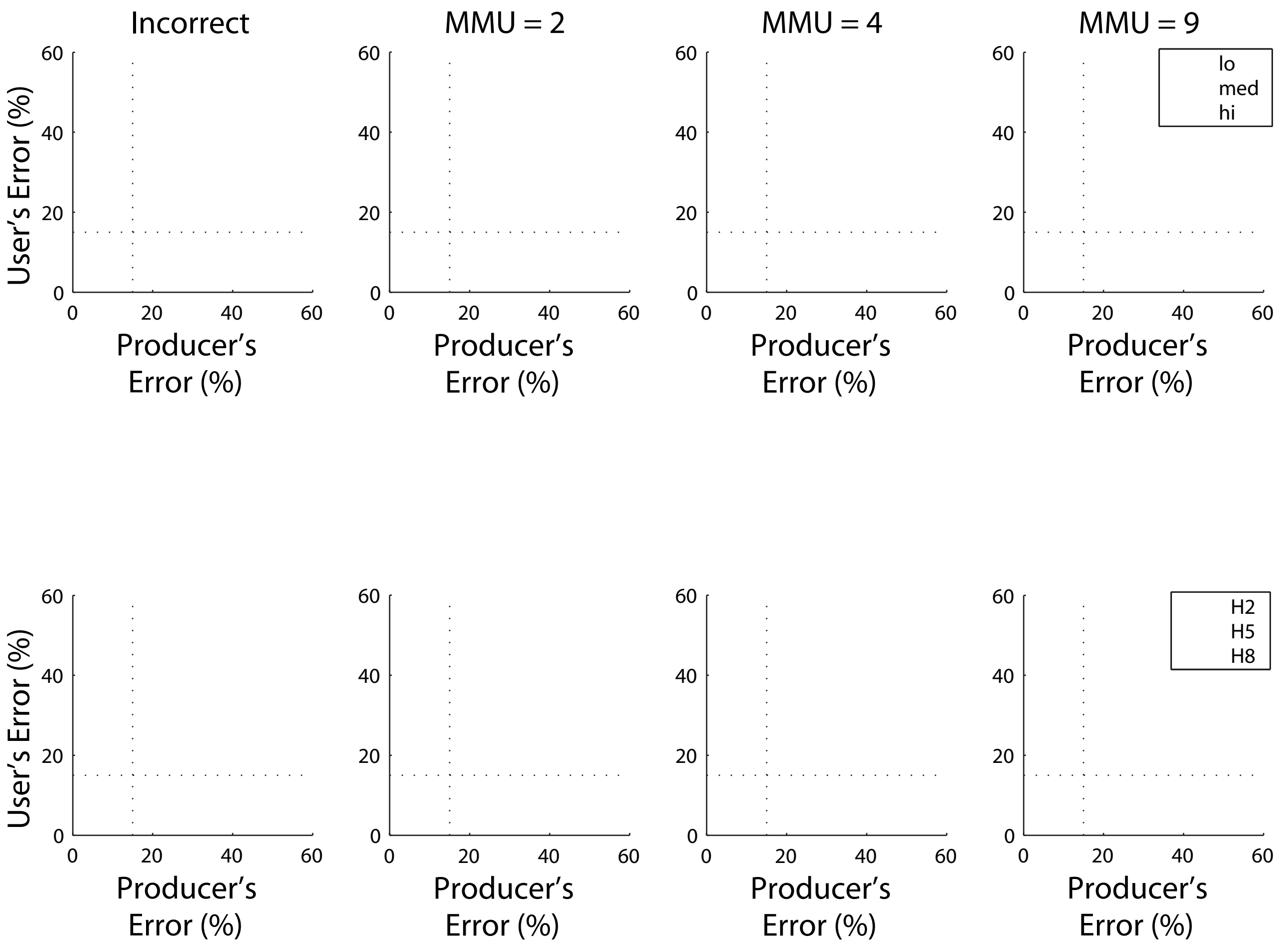
Figure 4. Values of raw LPIs on Correct base maps vs. Incorrect and Smoothed landscapes resulting from pattern analysis using Fragstats. Each graph shows one point for each map that

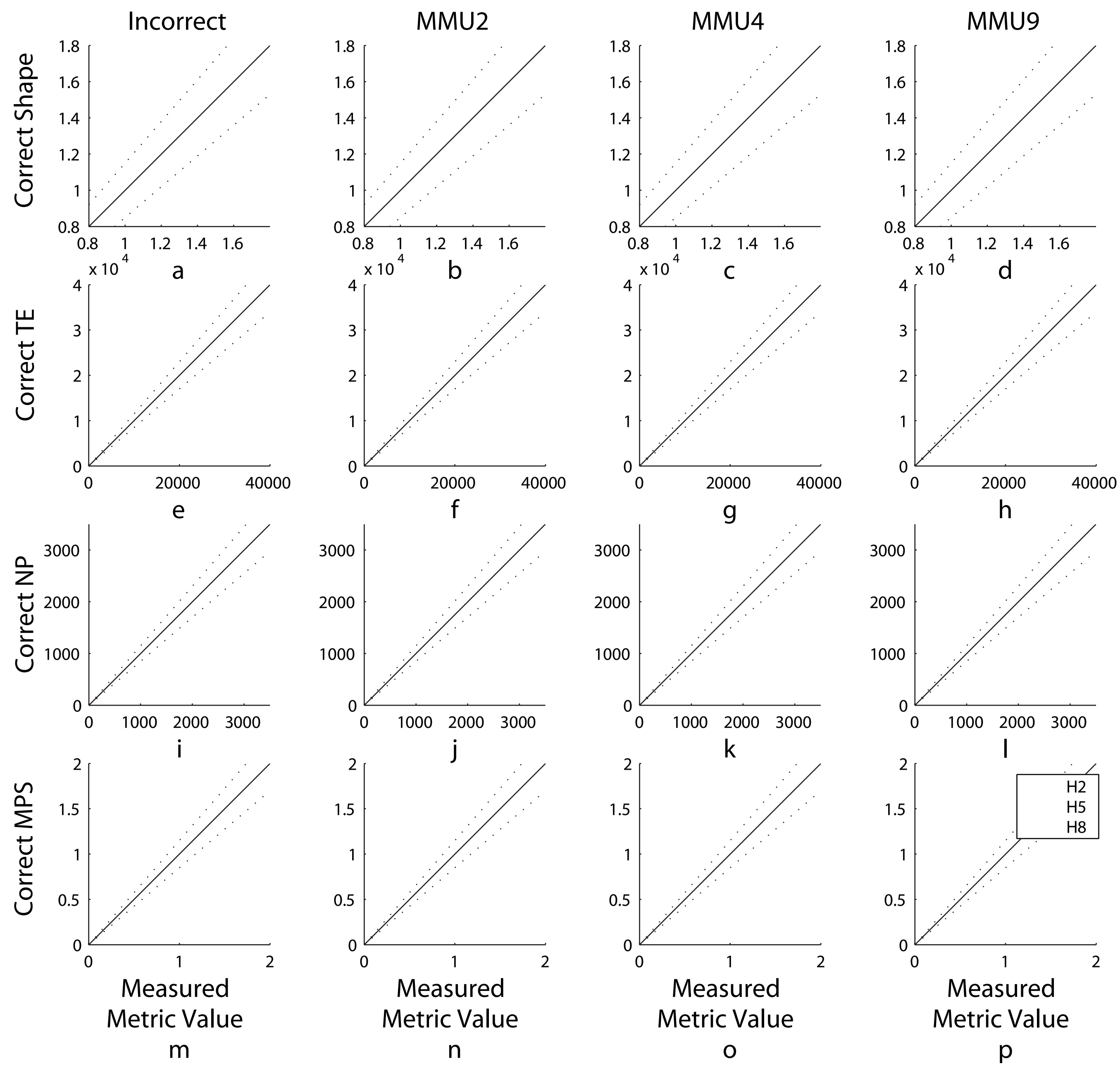
had no more than 15% user's error and producer's error. The x axis shows the measured raw LPI value and the y axis shows the correct raw LPI value that the measured value should predict. There are 3 rays on each plot. The central one shows the line that would result if there was no LPI error, that is, correct=measured. The other two rays show errors of +15% and -15% of the correct value for reference. (a-d) Mean Patch Shape, (e-h) Total Edge, (i-l) Number of Patches, (m-p) Mean Patch Size

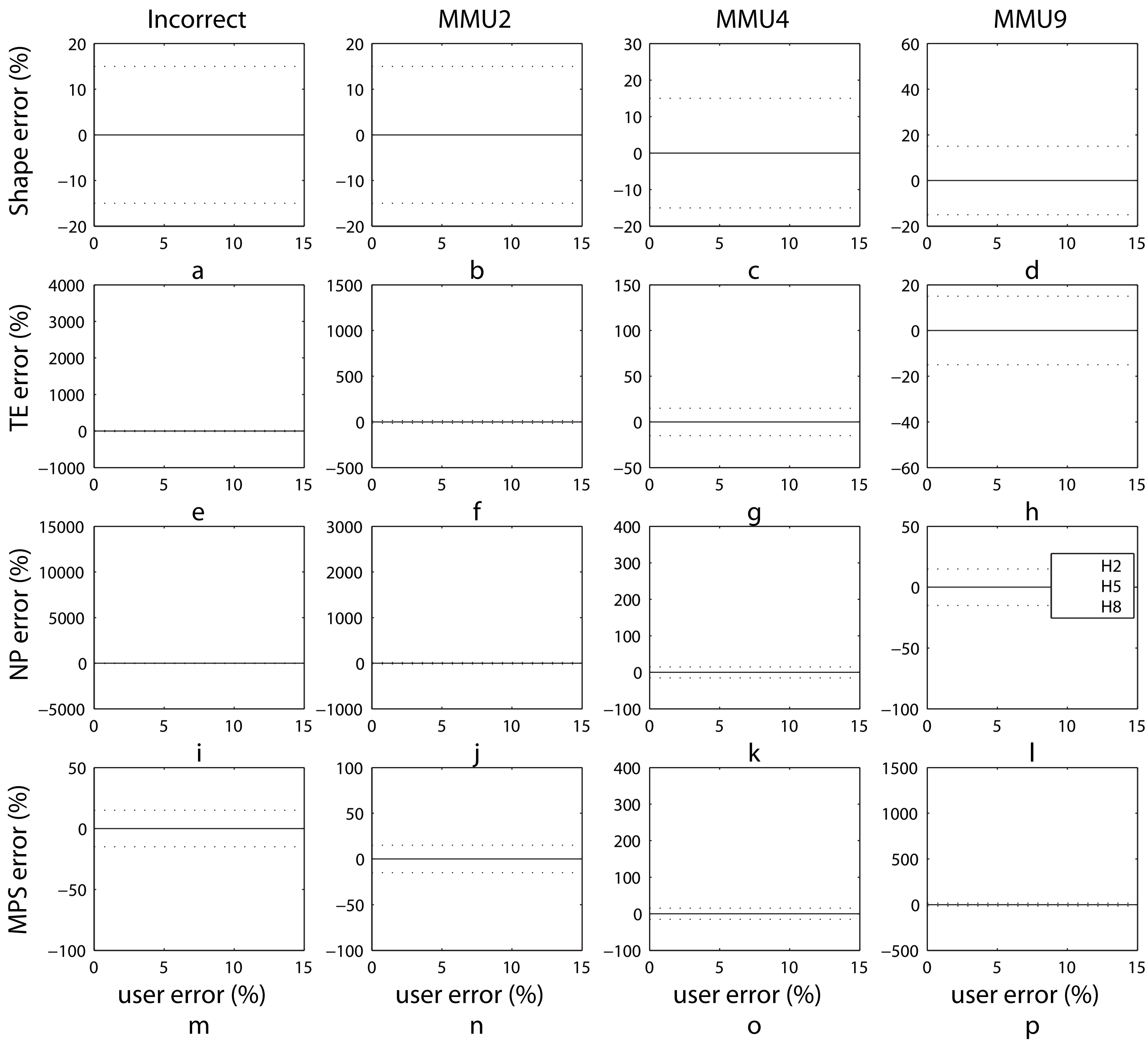
Figure 5. Percent LPI Error Relative to the LPI Value for the Correct landscape for four LPIs. Each graph shows one point for each map that had no more than 15% user's error and producer's error. The x axis shows the user's error and the y axis shows the percent error in the LPI. There are 3 lines on each plot. The central line marks the line of no LPI error. As in Figure 4, the other two lines show the boundaries of the interval of LPI errors between +15% and -15% for reference. Points are shaded according to the H value of the corresponding original correct map. (a-d) Mean Patch Shape, (e-h) Total Edge, (i-l) Number of Patches, (m-p) Mean Patch Size


20% Habitat
40% Habitat
60% Habitat
Fragmented
H=0.2
H=0.5
Clumpy
H=0.8


Correct
Incorrect
Smoothed MMU 2
Smoothed MMU 4
Smoothed MMU 9


Incorrect
MMU = 2
MMU = 4
MMU = 9
lo
med
hi
H2
H5
H8
User's Error (%)
Producer's Error (%)
0
20
40
60


Incorrect
MMU2
MMU4
MMU9
Correct Shape
Correct TE
Correct NP
Correct MPS
x 10^4
a
b
c
d
e
f
g
h
i
j
k
l
H2
H5
H8
Measured Metric Value
m
n
o
p


Incorrect
MMU2
MMU4
MMU9
Shape error (%)
TE error (%)
NP error (%)
MPS error (%)
user error (%)
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
H2
H5
H8