Mapping Settlements in the Wildland–Urban Interface: A Decision Tree Approach

Rutherford V. Platt
Gettysburg College

The wildland–urban interface (WUI) is the area where human-built structures intermingle or abut wildland vegetation. Maps of the WUI are important for resource management, particularly related to wildfire mitigation, but are often based on spatially coarse data such as housing counts from census blocks. Here, three decision tree models are used to create maps of human settlements for use in delineating the WUI. The first model uses statistics derived from image objects; the second model uses data related to topography, amenities, and accessibility; and the third model uses all available data. The accuracy of the models was evaluated in terms of the percentage of actual structures that fall within the area delineated as settlements. Overall, the three decision models performed similarly, although the third decision tree model was the best. For delineating settlements, all three decision tree models represent an improvement over a null model and the Radeloff et al. (2005) WUI mapping methodology and perform similar to the Wilmer and Aplet (2005) WUI mapping methodology. The models are also more flexible than many existing models, as they allow users to trade off accuracy and the size of the delineated settlement. The strategies described here can potentially yield improved maps of the WUI over larger areas. **Key Words: dasymetric mapping, decision trees, object-oriented, wildland–urban interface.**

荒地—城市界面（WUI）是指人类建筑结构混合或紧靠荒地植被的地区。WUI 地图对资源管理，尤其是有关野火缓解是重要的。但它往往是以空间粗略数据，如从人口普查块得来的房屋数为基础。在这里，我们用三个决策树模型来生成人类居住区地图，从而划定 WUI。第一个模型使用从图像对象派生的统计；第二个模型使用与地形，设施，和无障碍相关的数据；第三个模型使用所有可用的数据。依据落在划定的定居点区域范围内的实际结构百分比，对模型的准确性进行了评估。总的来说，三个决策模型表现类似，虽然第三个模型是最好的。对划定定居点而言，所有三个决策树模型都代表了比空模型和 Radeloff 等（2005）使用的 WUI 测绘方法的一个改进；并与 Wilmer 和 Aplet（2005）的测绘方法表现类似。这些模型也比许多现有的模型更加灵活，因为它们允许用户权衡准确性和所划定的定居点大小。这里所描述的策略可能会产生更好的较大面积的 WUI 地图。**关键词：分区密度图，决策树，面向对象，荒地—城市界面。**

La interfaz urbano-forestal (WUI) es el área en donde estructuras hechas por el hombre se entremezclan o colindan con la vegetación forestal. Los mapas de la WUI son importantes para el manejo de los recursos, particularmente en relación con la mitigación de los incendios forestales, pero a menudo se basan en datos de baja resolución espacial como el recuento de viviendas por bloques censales. Aquí, tres modelos de árboles de decisiones se utilizan para crear mapas de asentamientos humanos para su uso en la delimitación de la WUI. El primer modelo utiliza las estadísticas derivadas de objetos de imagen, el segundo modelo utiliza datos relacionados a la topografía, los servicios y la accesibilidad; y el tercer modelo utiliza todos los datos disponibles. La precisión de los modelos se evaluó en términos del porcentaje de las existentes estructuras que califican dentro del área delimitada como asentamientos. En general, los tres modelos de decisiones funcionaron similarmente, aunque el tercer modelo de árbol de decisión fue el mejor. Para delimitar los asentamientos los tres modelos de árbol de decisión representan una mejora sobre un modelo nulo y a la metodología cartográfica WUI de Radeloff et al. (2005) y funcionan de manera similar a la metodología cartográfica WUI de Wilmer y Aplet (2005). Los modelos son también más flexibles que muchos de los modelos existentes, ya que permiten a los usuarios compensar la precisión y el tamaño del asentamiento

The Professional Geographer, 64(2) 2012, pages 262–275
Initial submission, January 2010; revised submission, June 2010; final acceptance, September 2010.
Published by Taylor & Francis Group, LLC.

delimitado. Las estrategias descritas aquí pueden potencialmente producir mapas mejorados de la WUI en zonas más extensas. **Palabras claves: mapa dasimétrico, árboles de decisión, orientación a objetos, interfaz urbano-forestal.**

The *wildland–urban interface* (WUI), the area where housing abuts or intermingles with wildland vegetation, is associated with the vexing problems of structure loss due to wildfire, habitat fragmentation, spread of invasive species, and human–wildlife conflict. An important challenge for resource managers is consistently and accurately mapping the WUI. The definition of the WUI generally has three components: a "community" (henceforth called settlement), wildland vegetation, and a distance buffer representing the area where the WUI extends beyond the settlement (Stewart et al. 2007). Each component can be defined in a variety of ways depending on the purpose and assumptions of the study. The distance buffer, for example, has been defined as up to 2.4 km (1.5 miles) from a settlement (Healthy Forests Restoration Act 2003), 0.8 km (0.5 miles) from a settlement (Wilmer and Aplet 2005), or a variable distance from a settlement depending on vegetation height (Platt 2010). Wildland vegetation also has myriad definitions, most based on vegetation types in the National Land Cover Data set (NLCD; Radeloff et al. 2005; Wilmer and Aplet 2005; Hammer et al. 2007; Theobald and Romme 2007).

This article focuses entirely on improving the first aspect of the WUI definition: human settlements in sparsely settled areas. Settlements include structures, roads, lawns, and other features. The areal extent of settlements is typically derived either from distance to structure locations or from socioeconomic data. Structure locations can be estimated from well location data available from the state divisions of water resources (Aspinall 2004) or by digitizing structure location from high-resolution imagery. Unfortunately, structure location data sets are often incomplete, out of date, or prohibitively expensive to develop over large extents. For these reasons, many maps of settlements rely instead on socioeconomic data. For example, settlements can be mapped using parcel data, which are consistently collected by county assessors, up to date, although not always publicly available in geographic information system (GIS)-ready form. Housing counts from census data can also be used to map settlements, but in sparsely populated areas census blocks are very large and contain a large amount of undeveloped land. In these areas, estimates of housing location or housing density are often misleading—locally densely developed or undeveloped areas will effectively be "averaged away" within a large census block. Short of knowing the exact location of structures, one way to achieve improved areal estimates of settlements is with techniques related to dasymetric mapping.

Dasymetric mapping is the division of space into zone boundaries that reflect the underlying statistical variation of a particular variable (Eicher and Brewer 2001). Typically, dasymetric mapping disaggregates coarse-resolution quantitative data to a finer resolution using ancillary data sources (Mennis and Hultgren 2006). An example illustrates the dasymetric mapping process at its simplest: Imagine a map of housing density based on census tract housing counts. The housing density estimates are poor because there are areas within the census blocks (e.g., ponds and parks) where houses cannot exist. The underlying statistical variation of housing density would be better represented if the ponds and parks had a housing density of zero and the remaining zones had a higher density. Dasymetric techniques typically use remote sensing data or other publicly available data sets (e.g., road data, land cover data, cadastral data) to derive weights or rules for distributing the population within zones (Cockings, Fisher, and Langford 1997; Eicher and Brewer 2001; Chen et al. 2004; Reibel and Bufalino 2005; Maantay, Maroko, and Herrmann 2007). The weights or rules are typically derived from "expert knowledge" or assumptions about the distribution of the variable of interest, although they can also be derived from empirical sampling (Mennis 2003; Mennis and Hultgren 2006).

Strategies for mapping the WUI either employ very simple dasymetric mapping or none at all to delineate settlements. The most spatially extensive attempt to date to delineate the WUI, the WUI assessment (Radeloff et al. 2005), maps the intermix WUI (where structures mix with wildland vegetation) and interface WUI (where structures abut wildland vegetation) across the coterminous United States using an overlay of block data from the 2000 Census and vegetation data from the NLCD. Using this method, settlements are defined as the 2000 Census blocks with a housing density of one structure per 16.2 hectares (40 acres) or more. The Radeloff et al. study did not employ dasymetric mapping, as it did not attempt to alter or refine census block boundaries. In contrast, Wilmer and Aplet (2005) and Theobald and Romme (2007) used a similar technique but employed simple dasymetric mapping—the removal of public land before calculating housing density.

Techniques related to dasymetric mapping techniques could potentially be used to delineate settlements for mapping the WUI. For example, remotely sensed imagery could be used to find areas where structures are likely to exist (e.g., bright or spectrally heterogeneous areas). Alternatively, other ancillary data sources could be used to identify areas likely to contain structures. Past research has indicated that certain factors (e.g., distance to employment opportunities, zoning) shape population patterns in both urban and rural areas. Research has also indicated that amenities such as ski resorts (Duane 1999), public lands (Riebsame, Gosnell, and Theobald 1996), and space and seclusion (Davis, Nelson, and Dueker 1994) provide a high quality of life that might override economic considerations of where to live (Rudzitis and Streatfield 1993; Rudzitis 1999). In addition, scenic natural resources such as forests, riparian areas, and lakeshores draw low-density development and are also ecologically valuable (Ball 1997; Myers et al. 2000; Hansen et al. 2002). Population growth in rural counties of the northern Rockies is associated with areas of mountainous topography, forest cover, precipitation, and conserved land (Rasker and Hansen 2000). Because population is tightly coupled with housing, it makes sense that factors related to topography, accessibility, and amenities drive and constrain the location of settlements. At the time of publication of this article, however, such variables had rarely if ever been used to refine maps of settlements within counties.

In this article, three maps of settlements were developed for the mountainous western half of Boulder County, Colorado. The maps were created using models calibrated with decision trees, a strategy for partitioning data into homogenous groups based on the explanatory variables that best distinguish the variation of the independent variable (Breiman et al., 1984). The maps use techniques similar to dasymetric mapping but for simplicity present a nominal variable (settlement vs. nonsettlement) rather than a continuous variable (e.g., housing density). The first decision tree model (DT-Objects) was calibrated with remotely sensed imagery. In this model, object statistics derived from 1-m digital ortho quarter quads (DOQQs) were used to delineate settlements. The second model (DT-Characteristics) was calibrated with variables related to topography, accessibility, and amenities at a resolution of 30 m. The third model (DT-All) was calibrated with both remotely sensed data and data related to topography, accessibility, and amenities. The project has the following goals: (1) compare the three decision tree models, a null model, and two existing WUI models in terms of their ability to delineate settlements; and (2) evaluate the relationship between location of structures and variables related to object statistics, amenities, accessibility, and topography. The decision tree approach represents a robust strategy for developing WUI maps and produces logical rules that could potentially be extended to other areas.

Methods

Study Area

The study area for this project is the private land within the mountainous areas of Boulder County, Colorado (Figure 1). This study area was chosen for several reasons. First, it is data rich; in particular, digitized locations of structures are available to calibrate and validate the models. Second, the mountainous areas of Boulder County are a quintessential WUI environment containing widespread exurban development and several small former mining towns, including Nederland, Ward, and Jamestown. The public land surrounding these areas is primarily managed by the U.S. Department of Agriculture (USDA) Forest Service, the Bureau of Land Management (BLM), the National Parks Service (NPS), and the Boulder County and City Open Space and Mountain Parks. Because virtually all structures in the study area abut or intermingle with wildland vegetation, this study is able to focus on the "human settlement" component of the WUI.

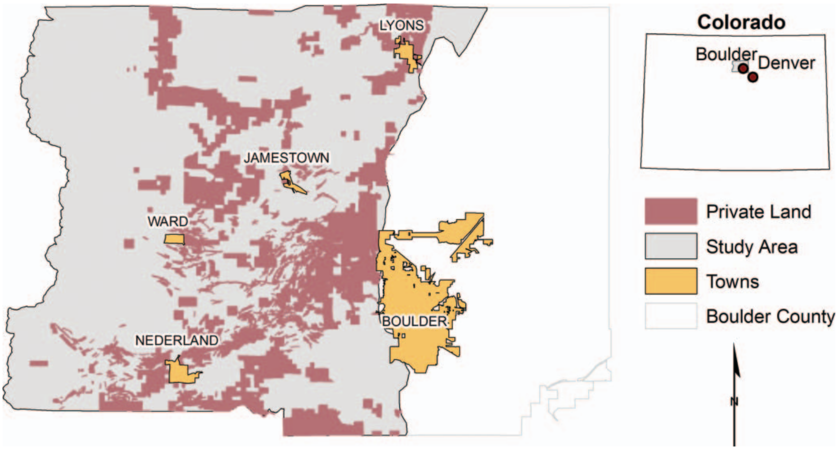

Figure 1 *Study area of the mountainous area in Boulder County, Colorado. (Color figure available online.)*

Decision Trees

Decision trees were used to delineate settlements within the study area. *Decision trees* are a set of nonparametric techniques that derive a series of rules to classify cases into discrete groups. A common statistical classifier, logistic regression, fits a logistic curve to describe the relationship between an independent variable and the probability of class membership. In contrast, decision trees do not assume a monotonic relationship between independent and dependent variables. They are able to model complex nonlinear relationships and interactions between independent variables. Studies suggest that decision trees might yield better classification accuracy than traditional statistical classifiers such as maximum likelihood (Friedl and Brodley 1997) and result in a substantial reduction in data dimensionality (Borak and Strahler 1999).

This study used a particular tree growing method called *classification and regression trees* (CRT; Breiman et al. 1984). CRT finds thresholds of the independent variables that split data into groups that are as "pure" (homogeneous) as possible in terms of the dependent variable. The splits are based on the Gini method, which calculates impurity based on the squared probabilities of the cases belonging to a dependent variable category. The tree continues to grow until either (1) the tree grows to a maximum of five levels, a commonly used cutoff to maintain model parsimony, or (2) splitting the data results in an improvement in impurity (squared probability of an area containing a structure) of less than 0.0001. This fully grown tree has the smallest possible "risk" (the proportion of misclassified cases adjusted for prior probabilities and any defined misclassification costs). To avoid overfitting, the trees are then pruned by removing nodes (rules) to create the smallest tree that does not increase risk by more than one standard error. Generally, pruning results in a vastly simplified set of classification rules, with a minimal increase in risk.

To calibrate and validate the decision tree models, a supervised classification strategy was used. A data set of 13,908 points was compiled where points represent (1) the actual location of the 6,954 structures and (2) a random sample of 6,954 points. The random sample represents all private land beyond 200 m from a structure. The actual locations of structures were digitized by Boulder County Land Use in 2003, using DOQQs and on-the-ground Global Positioning System readings. The points represent the estimated center of the building footprint. Because areas with a density lower than one structure per 16.2 ha are commonly considered outside a community (U.S. Department of Agriculture and U.S. Department of the Interior 2001; Radeloff et al. 2005; Wilmer and Aplet 2005), and thus not part of the WUI, structures that were more than 576 m from another structure were removed. Under this definition, two or more adjacent 16.2-ha^2 parcels with structures at the center would count as a community, but a more dispersed set of structures would not. The decision trees were used to distinguish the "structure" points from the "nonstructure" points, after which the models were applied to the landscape as a whole to delineate settlements. A randomly selected 80 percent of the points were included in the calibration procedure, and the remaining 20 percent were reserved for validation.

DT-Objects

The first decision tree model was calibrated with object statistics derived from 1-m black-and-white 1999 U.S. Geological Survey (USGS) DOQQs. USGS DOQQs were used because they are the same images used to digitize individual structures, are available free of charge for large areas, and require less computational power to process than multispectral data of comparable resolution. Definiens Professional 5.0 (Definiens AG, Munich, Germany) was used to segment the image into objects and calculate object statistics, a strategy known as *object-oriented image analysis* (OBIA). OBIA is the segmentation and classification of homogeneous image polygons, or objects, rather than individual pixels. Whereas traditional classification typically relies exclusively on spectral and textural data, OBIA also utilizes spatial relationships between objects, shape characteristics of objects, and a wide variety of statistics related to spectral and textural characteristics of objects. Studies have suggested that OBIA techniques are better (or at minimum no worse) than traditional pixel classification methods (Willhauck 2000; Civco et al. 2002; Oruc, Marangoz, and Buyuksalih 2004; Whiteside and Ahmad 2005; Platt and Rapoza 2008). Decision tree models are likely to be effective in sorting through and identifying the spectral, spatial, textural, and contextual object statistics for image classification (Laliberte, Fredrickson, and Rango 2007).

Within the framework of Definiens Professional 5.0, the object segmentation process is based on a number of parameters related to scale, color, shape, smoothness, and compactness (*Definiens Professional 5.0 User Guide* 2006). The scale parameter is a unitless number that controls the size of image objects. The color and shape parameters dictate the relative influence of spectral information and shape in creating object boundaries. The shape parameter is defined by the smoothness and compactness parameters. Compactness is calculated as the ratio of the border length and the square root of the number of object pixels. Smoothness is calculated as the ratio of the border length and the shortest possible border length derived from the bounding box of an image object. Because there is no optimum set of parameters, a standard practice is to select parameters through trial and error. The following parameters were iteratively selected to derive objects: scale, 50; color, 0.7; shape, 0.3; smoothness, 0.5; compactness, 0.5. These parameters yielded objects that captured individual structures and surrounding infrastructure (Figure 2).

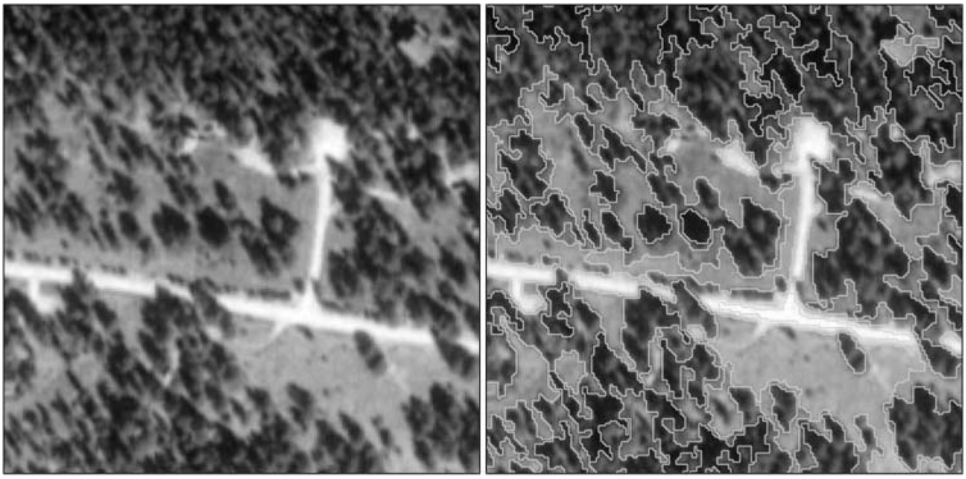

Figure 2 *Detail of image segmentation before (left) and after (right).*

Table 1 *Variables derived from image objects*

Spectral characteristics	Shape	Texture
Max pixel value	Length	GLCM contrast
Mean difference to scene	Area including inner polygons	GLCM std dev
Std dev	Area	GLCM mean
Minimum pixel value	Area excluding inner polygons	GLCM homogeneity
Mean of outer border	Perimeter	GLCM entropy
	Number of edges	GLCM correlation
	Border length	
	Width	
	Shape index	
	Compactness (polygon)	
	Border index	
	Compactness (generic)	
	Roundness	
	Length of longest edge	
	Elliptic fit	
	Rectangular fit	
	Average length of edges	
	Asymmetry	
	Density	
	Length/width	
	Std dev of length of edges	
	Number of inner objects	
	Main direction	

Note: Haralick's gray-level cooccurrence matrix (GLCM) describes how different combinations of pixel values occur within an object (Haralick, Shanmugam, and Dinstein 1973).

The segmentation procedure produced objects of variable size (M = 747 m^2, SD = 683 m^2). Within the objects, a total of thirty-five object statistics were calculated, each related to a spectral characteristic, texture, or shape of the objects (Table 1; see *Definiens Professional 5.0 User Guide* 2006 for further details). The object statistics were then assigned to the points that fell inside (6,954 representing structures and 6,954 representing nonstructures). The calibration points were then used to calibrate the decision tree model to develop rules for distinguishing the structure points from the nonstructure points.

DT-Characteristics

The second decision tree model was calibrated on site characteristics related to topography (slope, topographic position, solar radiation), amenities (percentage canopy cover, distance from stream, distance to trailhead, distance to public lands), and accessibility (distance to road, distance to city; Table 2). Rasters representing these variables were derived from commonly available data sets from the USGS, USDA Forest Service, and Boulder County Land Use. All rasters have a spatial resolution of 30 m. It was hypothesized that houses would most likely be built in areas of low slope, close to the city, in areas that receive lots of sunlight, in valleys, in open canopy areas, on south-facing slopes, on areas close to roads and trails, and near streams (Table 2). As with the object statistics, the site characteristics were assigned to the points (6,954 representing structures and 6,954 representing nonstructures). As with DT-Objects, the calibration points were then used to calibrate the decision tree.

DT-All

The third decision tree model was calibrated on the same validation points, using the data from both DT-Objects (object statistics) and DT-Characteristics (topography, accessibility, and amenities).

Delineation of Settlements from Decision Tree Rules

Decision trees are made up of a series of rules, each associated with the proportion of observations that belong to a particular class.

Table 2 *Topography, accessibility, and amenities variables*

Name	Description	Source data
Slope	Slope (degrees)	USGS 30 m DEM
Dist2city	Distance to Boulder by roads	USDA Forest Service road data
Radiation	Total annual solar radiation received by site	USGS 30 m DEM
TPI	Topographic position index* (Jenness 2006)	USGS 30 m DEM
Cancover	Canopy cover (%)	LANDFIRE
Publanddist	Euclidean distance to public land (m)	Boulder County public lands data
Roaddist	Euclidean distance to closest road (m)	USDA Forest Service road data
Disttrail	Distance to trailheads along roads	USDA Forest Service and Boulder County Open Space Trail data
Streamdist	Euclidean distance to stream (m)	USGS streams

Note: Topographic position index (TPI) is calculated as the difference between a cell and the neighborhood of the cell. If TPI is positive, it is higher than the surrounding neighborhood (for large neighborhoods, interpreted as a ridge or hill); if it is negative it is lower than the surrounding neighborhood (for large neighborhoods, interpreted as a valley). TPI is strongly scale dependent; in this case a 1,000-m circular neighborhood is used. USGS = U.S. Geological Survey; USDA = U.S. Department of Agriculture; DEM = digital elevation model.

For example, a single decision tree rule might state that of points that are more than 90 m from roads and more than 15 m from public land, 84 percent represent nonstructures and 16 percent represent structures. In this example, all cells more than 90 m from roads and more than 15 m from public land would be classified as nonstructures because the 84 percent proportion is above the 50 percent cut point (the proportions are used as a proxy for probability). The 50 percent cut point is arbitrary and in fact does not represent the actual proportion of cells containing a structure; whereas 50 percent of the calibration points represent structures, only a small percentage of cells in the landscape actually contain structures. By design, the delineated settlement was overpredicted to minimize errors of omission.

All maps were produced using 30-m grid cells. The input data for DT-Characteristics is already represented in 30-m grid cells, but the objects used as input to DT-Objects are of variable size and based on 1-m orthophotos. To make the maps comparable, the image objects were converted to a 30-m grid, using the center point rule in cases where multiple objects intersect a grid cell.

Model Comparison and Evaluation

After delineating the settlements, the models were compared using three techniques: (1) a visual comparison, (2) classification matrices, and (3) a graph of the percentage of structures that fall within settlements (high is better) versus the percentage of study area classified as settlements (low is better). This graph allows the models to be compared at every possible cut point, not just 50 percent. For reference, the models are compared to a null model, where the percentage of land in settlements is equal to the correctly classified structures. The models are also compared to the settlements defined by two existing WUI mapping strategies: Radeloff et al. (2005) and Wilmer and Aplet (2005).

Results

DT-Objects

Variables with the most explanatory power appear earlier in the tree and more frequently (Lagacherie and Holmes 1997). The most important variable for the DT-Objects tree is maxpixel (Figure 3): A point has a higher probability of representing a structure when the maxpixel value is high. Farther down the tree, variables related to object texture appear frequently. When gray-level cooccurence matrix (GLCM) contrast, GLCM stddev, and GLCM entropy are high, this indicates that the pixel values within the object are heterogeneous. It makes intuitive sense that structure points tend to be located in objects that are heterogeneous with some bright pixels, as structures have a variety of building materials, some of which (e.g., cement, bare ground) are highly reflective and some of which (e.g., asphalt, most shingles) are dark. Variables related to shape (e.g., length)

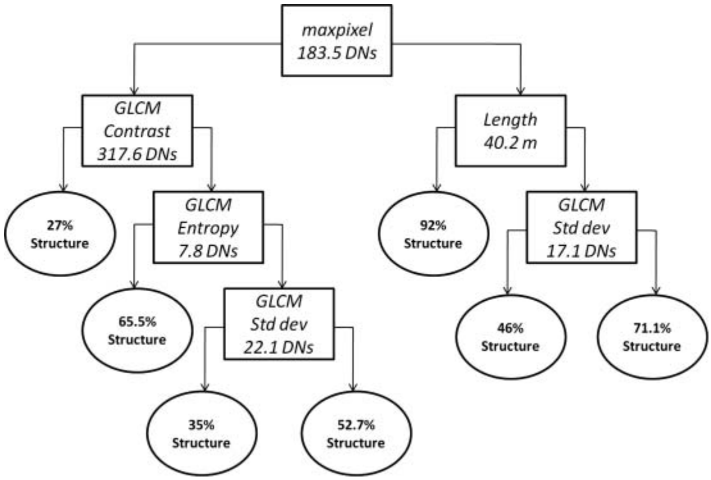

Figure 3 *Decision tree for DT-Objects. Each node (shown as a rectangle) includes a variable and threshold value. Values greater than the threshold are split to the right, whereas values less than or equal to the threshold are split to the left. Terminal nodes (shown as ovals) include the probability that a point represents a structure.*

are also important but appear less frequently than texture. For example, points in objects with high maxpixel value that also have low length (e.g., are compact) are likely to represent structures.

DT-Characteristics

The second decision tree model, DT-Characteristics, shows that only a few variables are important for distinguishing the classes. At the top of the tree, roaddist is the most important variable (Figure 4). Farther down the tree, slope and publanddist are also important.

The rules can be interpreted as follows: Structures tend to be within 92 m of the road network. Structures also tend to be located in areas directly adjacent to public land. Indeed, many of the few structures located far from roads are within 15 m of public land. Structures tend to be located on land with less than 16 degrees slope in places farther than 36 m from a road.

DT-All

The third decision tree model is calibrated with both of the aforementioned data sets. As in DT-Objects, maxpixel appears at the top of the tree (Figure 5). Roaddist appears on the second level. On the third level, object statistics such as GLCM contrast, length, and perimeter are important. On the lowest levels, topographic position index (TPI), publanddist, and GLCM stddev help distinguish structures from nonstructures.

Model Comparison and Validation

To compare the models, maps were developed using the 50 percent cut point and visually compared (only a subset shown here; Figure 6). The map confirms that the models are successful in identifying the areas that contain structures but also illustrate some error (e.g., actual structures located outside of the delineated settlement).

The models were then compared and validated using classification matrices constructed with the 20 percent of the observations removed from the calibration procedure. Again, the 50 percent cut point was used. For DT-Objects, it was found that 71.9 percent of the points representing structures were correctly identified and that 71.1 percent of the points

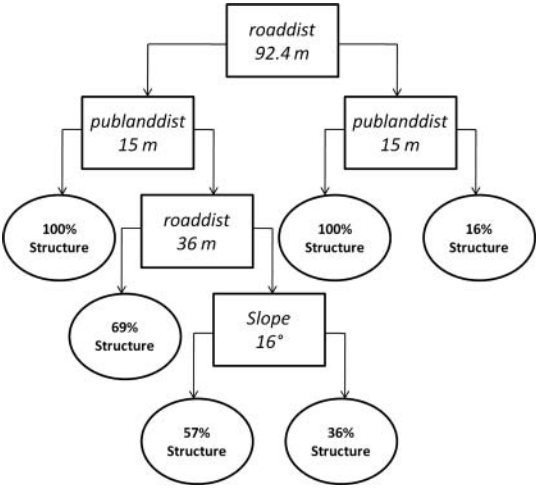

Figure 4 *Decision tree for DT-Characteristics. Each node (shown as a rectangle) includes a variable and threshold value. Values greater than the threshold are split to the right, whereas values less than or equal to the threshold are split to the left. Terminal nodes (shown as ovals) include the probability that a point represents a structure.*

representing not containing structures were correctly identified (Table 3).

For the DT-Characteristics model, it was found that 80.6 percent of the points representing structures were correctly identified and that 60.4 percent of the points not representing structures were correctly identified (Table 4). Compared to DT-Object, DT-Characteristics did a better job identifying points that represent structures but also predicted that many points are likely to represent structures when in fact they do not.

The DT-All model was found to be the best of the three but only marginally (Table 5). It was found that 80.8 percent of the points representing structures were correctly

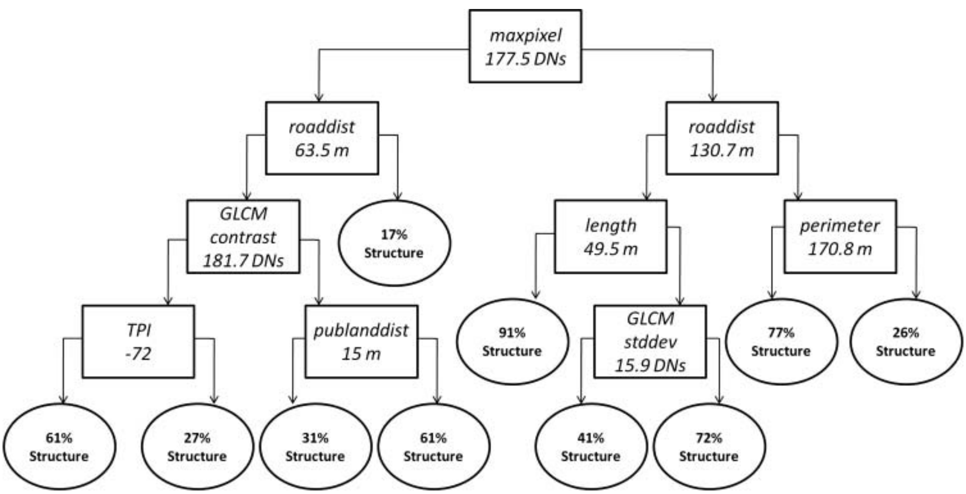

Figure 5 *Decision tree for DT-All. Each node (shown as a rectangle) includes a variable and threshold value. Values greater than the threshold are split to the right, whereas values less than or equal to the threshold are split to the left. Terminal nodes (shown as ovals) include the probability that a point represents a structure.*

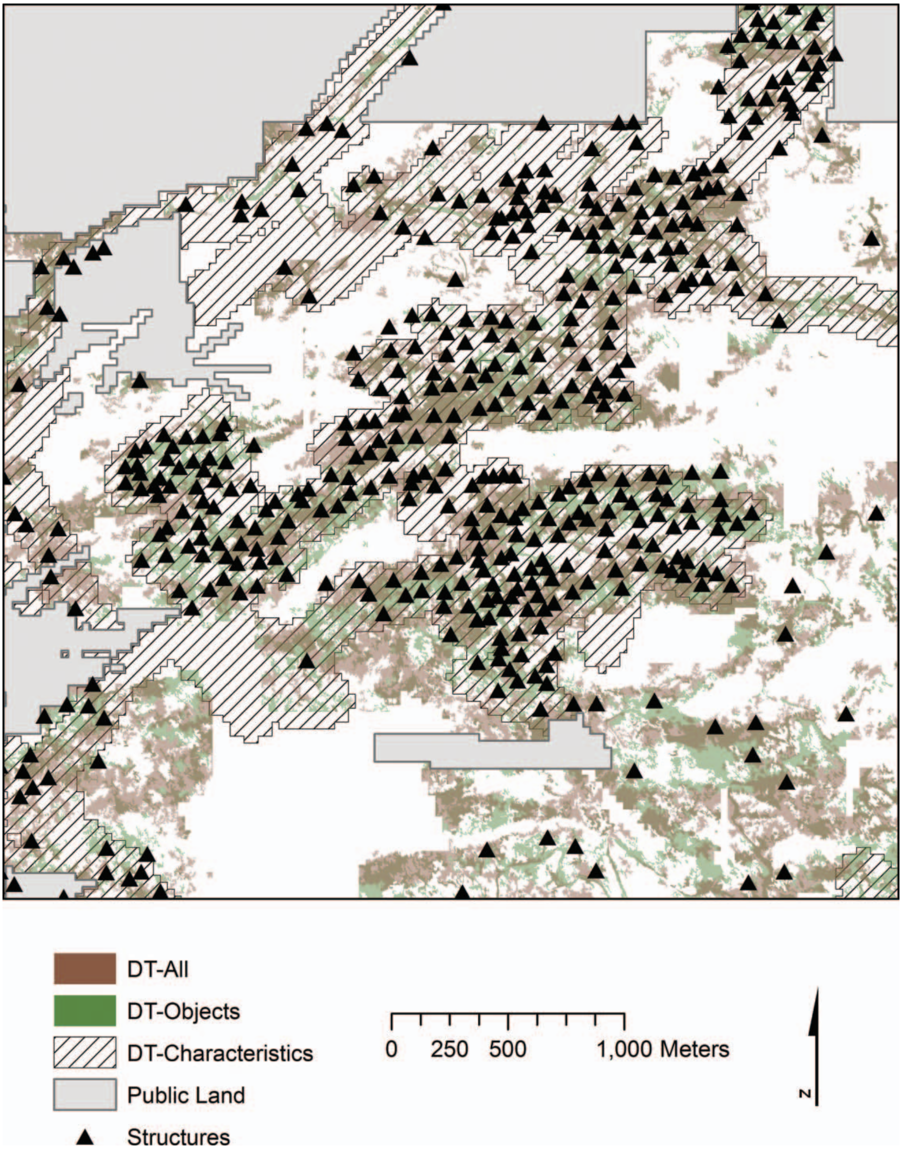

Figure 6 *Detail of settlement map. DT = decision tree. (Color figure available online.)*

identified and that 70 percent of the points not representing structures were correctly identified.

The three models were then compared and validated using a variation on a receiver operating characteristic (ROC) curve (Figure 7). Like a traditional ROC curve, the y axis shows the accuracy—also called *true positive* or *1-error of omission*. In this study, it can be interpreted as the percentage of nonremote structure points (i.e., within 576 m of another point) that fall within cells classified as settlements. The x axis shows the percentage of all cells classified as settlements. This can be interpreted similarly to the *false positive* rate or *1-error of commission*, as only a small (but unknown) percentage of the land area includes structures and surrounding infrastructure.

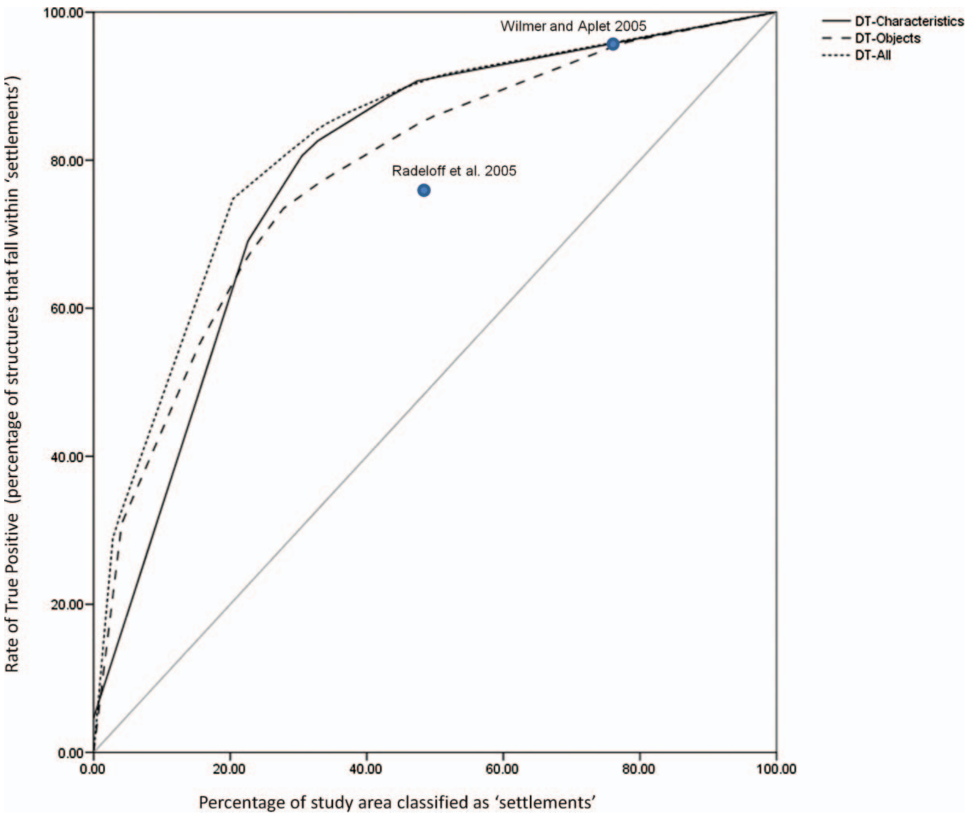

Figure 7 *Receiver operating characteristic (ROC) curve showing the trade-off between "true positive" rate (i.e., the percentage of nonremote structure points within settlements) and the percentage of the study area classified as settlements. DT = decision tree. (Color figure available online.)*

The greater the area under the curve in Figure 7, the better the model. For reference, the curve of a hypothetical model with no predictive power (null model) would fall along the diagonal line (Figure 7); the percentage of cells classified as settlements is equal to the percentage of structures that fall within the settlements area. Also, for reference, two existing WUI mapping strategies (Radeloff et al. 2005; Wilmer and Aplet 2005) have been placed on the graph as single points.

The curves of the three models are visually similar to each other and lie above the diagonal line (null model). Starting in the upper right-hand corner of Figure 7, it is clear that

Table 3 *Classification matrix for DT-Objects*

	Predicted		
Observed	**Not a structure**	**Structure**	**% correct**
Not a structure	948	385	71.1
Structure	385	985	71.9
Overall percentage	49.3%	50.7%	71.%

Table 4 *Classification matrix for DT-Characteristics*

	Predicted		
Observed	**Not a structure**	**Structure**	**% correct**
Not a structure	839	551	60.4
Structure	196	1,206	86.0
Overall percentage	37.1%	62.9%	73.2

Table 5 *Classification matrix for DT-All*

	Predicted		
Observed	**Not a structure**	**Structure**	**% correct**
Not a structure	989	423	70.0
Structure	273	1,146	80.8
Overall percentage	44.6%	55.4%	75.4

if 100 percent of the study area is a community, then 100 percent of the structures will be within a the community. By applying any of the three decision tree models, settlements can be shrunk to 75 percent of the study area while retaining ~95 percent of the structures. The Wilmer and Aplet (2005) model performs similarly. A null model with no predictive power, by contrast, would retain only 75 percent of the structures. Continuing down the curve, DT-All and DT-Characteristics allow us to shrink the settlements to 50 percent of the study area while retaining ~90 percent of the structures, whereas DT-Objects retains only ~82 percent of the structures. The Radeloff et al. (2005) model retains 70 percent of the structures at 50 percent of the study area and so is less accurate by this standard than the three decision tree models. At the bottom of the curve, DT-All allows us to shrink the settlements to 20 percent of the study area while retaining ~75 percent of the structures, whereas DT-Objects and DT-Characteristics retain only ~62 percent of the structures. The results show that overall DT-All is the most accurate model, although the advantage depends on the particular cut point.

Discussion and Conclusions

In western Boulder County, Colorado, the location of structures is related to characteristics such as distance to road, distance to public land, and slope, as well as to spectral characteristics (e.g., maxpixel), texture (e.g., GLCM contrast), and shape (e.g., length). These independent variables were successfully used to construct maps of settlements in the WUI, similar in nature to dasymetric maps. The "best" of the three decision tree models depends on the specific application. For small refinements of the estimate of structure location, the performance of the three decision tree models is similar; all three models can reduce the settlements to 75 percent of the study area while retaining ~95 percent of the structures. In this case, the DT-Characteristics model might be "best" because acquiring and processing the data for this model is straightforward. For larger refinements, however, the DT-All model performs better than the others (higher accuracy at a given size of delineated settlements). Unfortunately, calculating object-level statistics requires time, computational power, and expensive software.

The modeling strategies described in this study represent a potential step forward in consistently mapping the WUI. In the study area, the strategies more accurately delineate settlements than the Radeloff et al. (2005) method, although they are not as conceptually simple or as easily extended to large areas. The strategies are more flexible (if not better) than the Wilmer and Aplet (2005) method because they allow users to trade off accuracy and size of the delineated settlement.

To apply the decision tree models to larger areas, it would be important to first evaluate the relationships and rules to see if they extend to other places. Many of the relationships and rules might indeed apply broadly to other places, especially in the Rocky Mountain Region. It is expected that across the country, most settlements would be spectrally heterogeneous with some highly reflective elements. Furthermore, it is expected that settlements would be close to roads and on relatively low slopes. If the relationships hold, the decision tree rules could be used in conjunction with census block housing counts to create improved maps of housing density, which could be overlaid on layers of wildland vegetation to create a full WUI map. Some regional differences will doubtless emerge, however, that could complicate general applicability. For example, the study area is amenity-rich and has extensive public land that constrains development. The relationship between public land and location of settlements could be very different in a nonmountainous environment, in an area with less extensive public lands, or in an area with different zoning practices from the study area. Exploring the regional differences would be a rich area of study. Should major regional differences in these relationships exist, it would be important to extend the method presented here to allow "rules" to vary spatially. Even if large-scale applicability should ultimately prove difficult, the process of creating the decision tree models is itself valuable, as it helps reveal the characteristics (topographic, accessibility related, amenity related, and spectral) that are associated with settlements. ■

Literature Cited

Aspinall, R. 2004. Modelling land use change with generalized linear models—A multi-model analysis of change between 1860 and 2000 in Gallatin Valley, Montana. *Journal of Environmental Management* 72 (1–2): 91–103.

Ball, J. 1997. On the urban edge: A new and enhanced role for foresters. *Journal of Forestry* 95 (10): 6–10.

Borak, J. S., and A. H. Strahler. 1999. Feature selection and land cover classification of a MODIS-like data set for a semiarid environment. *International Journal of Remote Sensing* 20 (5): 919–38.

Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and regression trees*. Belmont, CA: Wadsworth.

Chen, K., J. McAneney, R. Blong, R. Leigh, L. Hunder, and C. Magill. 2004. Defining area at risk and its effect in catastrophe loss estimation: A dasymetric mapping approach. *Applied Geography* 24: 97–117.

Civco, D. L., J. D. Hurd, E. H. Wilson, M. Song, and Z. Zhang. 2002. A comparison of land use and land cover change detection methods. In *Joint Annual Conference of the American Society for Photogrammetry & Remote Sensing (ASPRS), the American Congress on Surveying and Mapping (ACSM) and the International Federation of Surveyors (FIG)*, 22–27. Washington, DC.

Cockings, S., P. Fisher, and M. Langford. 1997. Parameterization and visualization of the errors in area interpolation. *Geographical Analysis* 29 (4): 314–28.

Davis, J. S., A. C. Nelson, and K. J. Dueker. 1994. The new ' burbs: The exurbs and their implications for planning policy. *Journal of the American Planning Association* 60:35–59.

Definiens Professional, 5.0. Munich, Germany: Definiens AG.

Definiens Professional 5.0 User Guide. 2006. Munich, Germany: Definiens AG.

Duane, T. P. 1999. *Shaping the Sierra: Nature, culture and conflict in the changing West*. Berkeley: University of California Press.

Eicher, C., and C. Brewer. 2001. Dasymetric mapping and areal interpolation: Implementation and evaluation. *Cartography and Geographic Information Science* 28 (2): 125–38.

Friedl, M. A., and C. E. Brodley. 1997. Decision tree classification of land cover from remotely sensed data. *Remote Sensing of Environment* 61 (3): 399–409.

Hammer, R. B., V. C. Radeloff, J. S. Fried, and S. I. Stewart. 2007. Wildland-urban interface housing growth during the 1990s in California, Oregon, and Washington. *International Journal of Wildland Fire* 16:255–65.

Hansen, A. J., R. Rasker, B. Maxwell, J. J. Rotella, J. Johnson, A. Wright Parmenter, U. Langner, W. Cohen, R. Lawrence, and M. V. Kraska. 2002. Ecological causes and consequences of demographic change in the new West. *Bioscience* 52 (2):151–62.

Haralick, R. M., K. Shanmugam, and I. Dinstein. 1973. Textural features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics* 3:610–21.

Healthy Forests Restoration Act. 2003. H. R. 1904. http://www.gpo.gov/fdsys/pkg/BILLS-108hr1904enr.pdf (last accessed 21 July 2011).

Jenness, J. 2006. Topographic Position Index (tpi_jen.avx) extension for ArcView 3.x. Jenness Enterprises. http://www.jennessent.com/arcview/tpi.htm (last accessed 8 January 2010).

Lagacherie, P., and S. Holmes. 1997. Addressing geographical data errors in a classification tree for soil unit prediction. *International Journal of Geographical Information Science* 11 (2): 183–98.

Laliberte, A., E. L. Fredrickson, and A. Rango. 2007. Combining decision trees with hierarchical object-oriented image analysis for mapping arid rangelands. *Photogrammetric Engineering & Remote Sensing* 73 (2): 197–207.

Maantay, J. A., A. R. Maroko, and C. Herrmann. 2007. Mapping population distribution in the urban environment: The cadastral-based expert dasymetric system (CEDS). *Cartography and Geographic Information Science* 24 (2): 77–102.

Mennis, J. 2003. Generating surface models of population using dasymetric mapping. *The Professional Geographer* 55 (1): 31–42.

Mennis, J., and Hultgren, T. 2006. Intelligent dasymetric mapping and its application to areal interpolation. *Cartography and Geographic Information Science* 33 (3): 179–94.

Myers, N., R. A. Mittermeier, C. G. Mittermeier, G. A. B. daFonseca, and J. Kent. 2000. Biodiversity hotspots for conservation planning. *Nature* 403:853–58.

Oruc, M., A. M. Marangoz, and G. Buyuksalih. 2004. Comparison of pixel-based and object-oriented classification approaches using Landsat-7 ETM spectral bands. Paper presented at the ISPRS Conference, Istanbul, Turkey.

Platt, R. V. 2010. The wildland–urban interface: Evaluating the definition effect. *Journal of Forestry* 108 (1): 9–15.

Platt, R. V., and L. M. Rapoza. 2008. An evaluation of an object-oriented paradigm for land use/land cover classification. *The Professional Geographer* 60 (1): 87–100.

Radeloff, V. C., R. B. Hammer, S. I. Stewart, J. S. Fried, S. S. Holcomb, and J. F. McKeefry. 2005. The wildland–urban interface in the United States. *Ecological Applications* 15 (3): 799–805.

Rasker, R., and A. J. Hansen. 2000. Natural amenities and population growth in the greater Yellowstone region. *Human Ecology Review* 7 (2): 30–40.

Reibel, M., and M. E. Bufalino. 2005. Street-weighted interpolation techniques for demographic count estimation in incompatible zone systems. *Environment and Planning A* 37:127–39.

Riebsame, W. E., H. Gosnell, and D. M. Theobald. 1996. Land use and landscape change in the Colorado Mountains 1: Theory, scale, and pattern. *Mountain Research and Development* 16 (4): 395–405.

Rudzitis, G. 1999. Amenities increasingly draw people to the rural West. *Rural Development Perspectives* 14 (2): 9–13.

Rudzitis, G., and R. A. Streatfield. 1993. The importance of amenities and attitudes—A Washington example. *Journal of Environmental Systems* 22 (3): 269–77.

Stewart, S. I., V. C. Radeloff, R. B. Hammer, and T. J. Hawbaker. 2007. Defining the wildland urban interface. *Journal of Forestry* 105:201–07.

Theobald, D. M., and W. Romme. 2007. Expansion of the US wildland–urban interface. *Landscape and Urban Planning* 83 (4): 340–54.

U.S. Department of Agriculture and U.S. Department of the Interior. 2001. Urban-wildland interface communities within vicinity of federal lands that are high risk from wildland fire. *Federal Register* 66:751–77.

Whiteside, T., and W. Ahmad. 2005. A comparison of object-oriented and pixel-based classification methods for mapping land cover in northern Australia. In *Proceedings of SSC2005 Spatial intelligence, innovation and praxis: The national biennial conference of the Spatial Sciences Institute*, 1225–31. Melbourne, Australia: Spatial Sciences Institute.

Willhauck, G. 2000. Comparison of object-oriented classification techniques and standard image analysis for the use of change detection between SPOT multispectral satellite images and aerial photos. *ISPRS* 33:214–21.

Wilmer, B., and G. Aplet. 2005. Targeting the community fire planning zone: Mapping matters. Washington, DC: The Wilderness Society. http://www.wilderness.org/Library/Documents/upload/TargetingCFPZ.pdf (last accessed 8 January 2010).

RUTHERFORD V. PLATT is an Associate Professor in the Environmental Studies department at Gettysburg College, Gettysburg, PA 17325. E-mail: rplatt@gettysburg.edu. His current research focuses on spatial models of human and ecological systems in the wildland–urban interface, with a particular emphasis on wildfire.