Classifying and Mapping Wildfire Severity: A Comparison of Methods

C. Kenneth Brewer, J. Chris Winne, Roland L. Redmond, David W. Opitz, and Mark V. Mangrich

Abstract
This study evaluates six different approaches to classifying and mapping fire severity using multi-temporal Landsat Thematic Mapper data. The six approaches tested include: two based on temporal image differencing and ratioing between pre-fire and post-fire images, two based on principal component analysis of pre- and post-fire imagery, and two based on artificial neural networks, one using just post-fire imagery and the other both pre- and post-fire imagery. Our results demonstrated the potential value for any of these methods to provide quantitative fire severity maps, but one of the image differencing methods (ND4/7) provided a flexible, robust, and analytically simple approach that could be applied anywhere in the Continental U.S.

Based on the results of this test, the ND4/7 was implemented operationally to classify and map fire severity over 1.2 million hectares burned in the Northern Rocky Mountains and Northern Great Plains during the 2000 fire season, as well as the 2001 fire season (Gmelin and Brewer, 2002). Approximately the same procedure was adopted in 2001 by the USDA Forest Service, Remote Sensing Applications Center to produce Burned Area Reflectance Classifications for national-level support of Burned Area Emergency Rehabilitation activities (Orlemann, 2002).

Introduction
Forests and rangelands in the Northern Rocky Mountains and Northern Great Plains are dynamic and constantly changing living systems. Recognition of this dynamic nature and its relationship to ecosystem health and the sustainable development of natural resources has led to an increased need for accurate and current information to support natural resource planning and decision-making (Mangold, 1995). An extraordinary event occasionally occurs that requires analyses across large areas and multiple ownerships to evaluate its resource management and resource policy implications. The wildfires in the Northern Rocky Mountains and Northern Great Plains during the 2000 fire season constitute such an event in both extent and effects. Over 1.2 million hectares burned (an area greater than the State of Connecticut) with suppression costs totaling an estimated $500 million USD. In addition to the large area burned, more than 400 structures including more than 100 homes were destroyed. The extent and effects of these wildfires directly affected the lives of hundreds of thousands of people, severely disrupted important sectors of the regions’ economy, and in some locations caused lasting damage to natural resources (USDA Forest Service, 2001). Similar, albeit smaller, events also occurred in the Northern Rocky Mountains and Northern Great Plains during the 2001 and 2003 fire seasons.

One of the most fundamental information needs for evaluating the management and policy implications of such fire events is consistent and continuous mapping of fire severity. The term fire severity is defined as “the degree to which a site has been altered or disrupted by fire; a product of fire intensity, fuel consumption, and residence time” (Helms, 1998). This information is useful to the Forest Service for at least five purposes: (a) to record the effects of these fires, (b) to plan and monitor restoration and recovery activities, (c) to provide a method for updating current vegetation maps, (d) as baseline information for future monitoring, and (e) to provide an analytical basis for evaluating management and policy implications. Anticipating their information need, the Regional Foresters of the USDA Forest Service Northern and Intermountain Regions initiated an effort to evaluate methods to map fire severity for the Northern Rocky Mountains and Northern Great Plains. These methods were required to be continuous across all ownerships, use consistent classifications and methods, and be effective in identifying mortality and/or aboveground biomass consumption for common vegetation types. Additionally, the methods should be cost effective, provide timely delivery of data, and use technology that could be transferred to local administrative units and fire management teams. To this end, the Forest Service collaborated with the University of Montana, Wildlife Spatial Analysis Laboratory for this study.

Several common methods exist for obtaining data on fire severity, although these data are rarely consistently and continuously available for large areas following extensive fire events. Some of these methods have been in use for decades, while others result from recent technological developments. These methods fall into four general categories: (a) aerial and ground-based sketch mapping, (b) interpretation of multi-temporal vertical aerial photography, (c) interpretation of post-fire vertical aerial photography, and (d) digital change detection using multi-temporal satellite-based data.

C. Kenneth Brewer is with the USDA Forest Service, Remote Sensing Applications Center, 2222 W. 2300 S., Salt Lake City, UT 84119 (kbrewer@fs.fed.us).

J. Chris Winne and Roland L. Redmond are with the Department of Biological Science, The University of Montana, Missoula, MT 59812.

David W. Opitz and Mark V. Mangrich are with the Department of Computer Science, The University of Montana, Missoula, MT 59812.

© 2005 American Society for Photogrammetry and Remote Sensing

Photogrammetric Engineering & Remote Sensing Vol. 71, No. 11, November 2005, pp. 1311–1320. 0099-1112/05$11.00/0
Aerial and ground-based sketch mapping is normally conducted as part of Burned Area Emergency Rehabilitation (BAER) activities and often combines effects on soils and vegetation in one map. These maps are designed to support post-fire rehabilitation efforts and serve that purpose well. They are not, however, designed to provide fire severity mapping that is consistent across ownership boundaries (not limited to National Forest System lands) and among multiple fires across large areas. They also have inconsistencies introduced by the rotating team membership inherent in fire assignments.

Interpretation of multi-temporal or post-fire, vertical aerial photography can provide very detailed and valuable information about fire severity, but these are dependent on the availability of current aerial photography shortly before a fire event and the acquisition of aerial photos following the fire event and before substantial additional change occurs (i.e., salvage harvest, green-up, and snow). These two conditions are rarely met because it is relatively uncommon to acquire post-fire aerial photography, and pre-fire photography is dependent on normal acquisition scheduling (approximately every ten years). Although visual qualitative change detection of aerial photography by a skilled interpreter yields, in many cases, more accurate and precise results than digital change detection (Edwards, 1990), results are not consistent due to the inherent variability and subjectivity of different interpreters. These techniques are also labor intensive and expensive, particularly when they need to be converted to planimetric map form and digitized for analysis. Additionally, they produce qualitative information on fire severity making quantification of change very difficult.

There is a substantial body of literature addressing many aspects of change detection using satellite-based remote sensing data (Jensen, 1981; Riordan, 1981; Lund, 1983; Park et al., 1983; Míline, 1988; Caselles and Garcia, 1989; Coppin and Bauer, 1994; Mattikalli, 1994; Ngai and Curlander, 1994; Sloggett et al., 1994; Franklin et al., 1995; Hugoda et al., 1995; Solheim et al., 1995; Hashem et al., 1996; Mahlke, 1996; Parra et al., 1996; Lyon et al., 1998; Brewer et al., 2000; Coppin et al., 2001). Lu et al. (2003) and Coppin et al. (2004) provide detailed reviews and discussions of change detection methods. Further detailed review and discussion of this literature is beyond the scope of this paper, but in general these vegetation change detection methods are readily abstracted to fire severity mapping and fall into four categories: (a) visual change detection, “on-screen digitizing”; (b) multiple classification comparison, “from-to post classification comparisons”; (c) image algebra “indices and ratios”; and (d) multi-temporal composite classification. Coppin et al. (2001) discuss and compare these general methods in relation to canopy change detection. This comparative discussion is easily adapted to fire severity mapping.

In contrast to aerial photo interpretation, or aerial and ground-based sketch mapping, digital change detection offers synoptic, temporal, objective and repeatable procedures. It also permits the incorporation of the infrared and microwave regions of the electromagnetic spectrum that contain relevant information on biophysical characteristics of vegetation (e.g., moisture status or photosynthetic activity). Given these considerations, we determined that digital change detection would best meet the stated objectives.

Our initial plan was to use two image dates, pre- and post-fire, in an image-differencing change detection analysis. As we thought more about the problem and as the geographic extent of wildfires in western Montana and Idaho expanded, several options became apparent and potentially desirable. Rather than limit the analytical approach to one based on temporal image differencing and/or ratioing between pre- and post-fire images, we decided to expand the comparison to include artificial neural networks (Opitz and Machlin, 1999) and principal component analysis. Thus, our study evaluates six different approaches to classifying and mapping fire severity using multi-temporal or single-date Landsat Thematic Mapper (TM) data. The six approaches tested include two based on temporal image differencing and ratioing between pre- and post-fire images, two based on multi-temporal principal component analysis, and two based on artificial neural networks. Of the six methods we compared, only one of the artificial neural network methods was based on a single post-fire image date. While it is true we could have applied the PCA methods to a single date, this would have required more extensive use of training data for classification, and we felt it was more cost-effective to rely on the change detection approach.

**Study Area**

The study area selected for this analysis was the Fort Howes wildfire complex located on the Ashland Ranger District of the Custer National Forest in southeastern Montana (Figure 1). The Ashland Ranger District is located in an “island” mountain range containing a diverse mix of forest, shrubland, and grassland vegetation types. The Fort Howes complex burned primarily during the last week of July 2000. Of the major wildfires in the Northern Rocky Mountains and Northern Great Plains, it was the only one that was completely controlled at the time the study was initiated.

Figure 1. Vicinity map of the Fort Howes wildfire complex in southeastern Montana. The area of detail indicates the area used in Figure 6 to illustrate pre-fire and post-fire imagery, as well as example classification results.
Acquisition of the Reference Dataset

The reference data used in this study were obtained by manual interpretation of vertical aerial photographs (1:15 840 scale) acquired on 17 August 2000 by the Custer National Forest. To minimize subjectivity in the collection of the reference data, one person completed all the photo-interpretation following the USDA Forest Service vegetation classification system for lifeform-level class assignment (i.e., forest lifeform ≥10 percent trees). The photo-interpretation process incorporated extensive ancillary data, including ground-based, post-fire photography of each class (Wilson, 1960). The ground-based photography was collected by Forest Service ecologists Jeff DiBenedetto and Ken Brewer within three to four weeks of the fire. Additional ancillary data included pre-fire vegetation plot data with known locations and a pre-fire Landsat TM based landcover classification (Redmond et al., 1998). Reference data locations were selected for photo interpretation after the TM imagery, aerial photography, and ground-based photography were in hand. Large homogenous locations were selected to avoid or minimize errors introduced by mixed-condition pixels. The resulting 268 reference data locations were split into a training set (n = 196) and a test set (n = 72). The former was used to train the artificial neural network classifiers, whereas the latter was used to evaluate the accuracy of all six classifiers.

Acquisition of the Image Dataset

Landsat TM imagery was chosen for this work because the mid-infrared reflectance of vegetation is strongly related to important vegetation canopy characteristics relative to fire effects. Landsat TM data are also acquired continuously and archived data could, therefore, be purchased to meet the time and area needs associated with the 2000 wildfires. Because the areas burned in the Fort Howes complex actually spanned portions of two Landsat TM scenes, Path 35/Row 28 and Path 35/Row 29, we had to find and purchase two images for each date. For the pre-fire images, we acquired a 05 July 2000, overpass by the Landsat 5 satellite; whereas the two post-fire images were collected by the Landsat 7 satellite on 27 August 2000. No major weather events or phenological change that would alter ground reflectance occurred between the time the fire was extinguished and the post-fire imagery was acquired. Copies of all four images were obtained from the EROS Data Center according to the following parameters:

- 30 meter pixel size
- Grid north map orientation
- NAD-27 datum (Clarke 1866 ellipsoid), and
- Albers equal area projection.

Pre-processing of Image and Ancillary Data

Both images, for each date, were ortho-rectified to previously terrain-corrected images for the respective scenes using the Geometric Correction Module and the Landsat orbit model in ERDAS IMAGINE® (ERDAS, 1997), as well as 7.5-minute digital elevation models. A minimum of at least 100 ground control points (GCP) were placed and located throughout each of the un-rectified images with additional points concentrated in and around each of the burned areas, as determined visually, and from the burned area perimeters supplied by the Custer National Forest. Actual rectification involved the cubic convolution algorithm and a 30 m pixel size. The cubic convolution re-sampling algorithm was selected to minimize the loss of spectral and spatial integrity of the data and to maximize their consistency between the two image dates (Lillesand and Kiefer, 1987). The resulting Root Mean Square Error RMSE was less than one-half pixel or 15 m.

All images were corrected for atmospheric scattering by adjusting the histograms of the TM bands such that their minimum values were close to zero (Song et al., 2001). The data from each image date were clipped to a rectangular area that encompassed the entire Fort Howes wildfire complex, and each band was checked to make sure it had the same number of pixels as its corresponding band in the other image date.

Finally, ancillary data layers, such as lifeform and topography, were assembled, co-registered, and clipped to the same study area boundary. Lifeform data came from the National Land Cover Dataset (NLCD) (Vogelmann et al., 2001) and topography was derived from 7.5-minute digital elevation models downloaded from the U.S. Geological Survey.

Data Processing and Analysis for Image Differencing Methods

Image differencing and image ratioing are the most frequently applied procedures for detecting change through comparison of datasets from different acquisition dates (Mattikalli, 1994). The image differencing method subtracts values of one image from the other, pixel-by-pixel. These values are often standardized to avoid the ambiguity resulting from identical difference values created from different original pixel values. This approach is fast, simple, and requires minimal user intervention (Sloggett et al., 1994). Image ratioing divides one image’s pixel value by the other. Single-band radiometric responses are often transformed to strengthen the relationship between spectral data and biophysical characteristics of vegetation (e.g., moisture status or photosynthetic activity). Vegetation indices are often used to accomplish these objectives (Coppin and Bauer, 1994; Yin and Williams, 1997). Many vegetation indices are described in the literature; therefore, careful selection is required because an index derived for a given application may not be appropriate for another (Wallace and Campbell, 1989; Lyon et al., 1998). A review of the literature identified only one index specifically derived to enhance the biophysical relationships with burned vegetation (Lopez-Garcia and Caselles, 1991). This Normalized Difference (ND) index, subsequently modified slightly by Key et al. (2002) and named the Normalized Burn Ratio, was identified as appropriate to meet the objectives of this study. The Normalized Burn Ratio (hereafter referred to as ND4/7) is calculated as: (TM4 - TM7)/(TM4 + TM7). In addition, we modified the original ratio by substituting TM5 for TM7 to produce a second ratio (hereafter referred to as ND4/5) calculated as: (TM4 - TM5)/(TM4 + TM5). Both these normalized difference values provided a quantitative continuous change image linked directly to actual quantitative change in vegetation condition (e.g., tree canopy mortality from fire). This approach offered comprehensive information that can be classified in a variety of ways for different analysis objectives (e.g., three equal-interval tree mortality classes versus five mortality classes).

We calculated these normalized ratios for the two image dates (pre-fire versus post-fire) and developed a difference image for each ratio stratified by different lifeforms. Lifeform (tree, shrub, and grass) was assigned to each pixel through an overlay procedure using the NLCD (see Table 1). The NLCD was developed to provide “a reasonably consistent and seamless 30 meter product for the conterminous United States” (Vogelmann et al., 2001). Given its geographic extent and development objectives, the NLCD was well suited for the development of a fire severity mapping procedure that could be applied nation-wide. The NLCD was developed from Landsat TM imagery following standardized methods and defined 21 land-cover classes in a hierarchical classification system (Vogelmann et al., 1998; Vogelmann and Wickham, 2000). The classification logic used in the NLCD did not
match the USDA Forest Service classification taxonomic logic used for the reference data (e.g., the NLCD defines both forests and shrublands as greater than or equal to 25 percent canopy cover of trees or shrubs, respectively; whereas the USDA Forest Service classification and mapping standards define forests and shrublands as greater than or equal to 10 percent). In addition to differences in the classification logic, classification errors from the stratification data propagate into the fire severity classification and mapping.

The post-fire ratios were subtracted from the pre-fire ratios for each pixel within each lifeform, and then the differences were plotted as frequency histograms (for each lifeform). Change thresholds between burned and unburned pixels were established based on the visual inspection of burned and unburned areas of each lifeform on the TM imagery and aerial photography. For the tree lifeform, burn severity was further established by: (a) visually interpreting the approximate 100 percent crown removal threshold, and (b) based on this value in relation to the lower, burn/non-burn threshold, splitting the difference between these two values in three proportional groups representing non-lethal burn (<20 percent of the change difference), mixed-lethal burned forest (20 to 80 percent of the change difference), and lethal burn forest (>80 percent change difference). For example, if the first threshold between burned and unburned forest was set at 400 and the second threshold for 100 percent crown removal was set at 600, then the difference between the two thresholds was 200; 20 percent of this difference at the low end (i.e., non-lethal burn) would include change values from 400 to 440; the intermediate 60 percent of the difference (assigned to the mixed-lethal burn class) would correspond to change values from 440 to 560; this would leave all change values greater than or equal to 560 as assigned to the lethal burn class. Once the various thresholds were established, the process of assigning each pixel to one of seven fire severity classes was run as a model in ERDAS IMAGINE®. Note that ground-reference data were not used to define the thresholds between change and no-change for any lifeform, nor were they used to further subdivide change into Lethal Tree and Mixed Tree burn classes for either normalized difference ratio.

**Table 1. Lifeform Assignment from National Land Cover Dataset**

<table>
<thead>
<tr>
<th>NLCD Cover Type</th>
<th>Lifeform</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evergreen forest</td>
<td>Forest</td>
</tr>
<tr>
<td>Mixed forest</td>
<td>Forest</td>
</tr>
<tr>
<td>Transitional</td>
<td>Forest</td>
</tr>
<tr>
<td>Woody wetlands</td>
<td>Forest</td>
</tr>
<tr>
<td>Shrubland</td>
<td>Shrub</td>
</tr>
<tr>
<td>Grasslands/herbaceous</td>
<td>Grass</td>
</tr>
<tr>
<td>Open water</td>
<td>Other</td>
</tr>
<tr>
<td>Perennial ice/snow</td>
<td>Other</td>
</tr>
<tr>
<td>Low intensity residential</td>
<td>Other</td>
</tr>
<tr>
<td>High intensity residential</td>
<td>Other</td>
</tr>
<tr>
<td>Commercial/industrial/transportation</td>
<td>Other</td>
</tr>
<tr>
<td>Bare rock/sand/clay</td>
<td>Other</td>
</tr>
<tr>
<td>Quarries/strip mines/gravel pits</td>
<td>Other</td>
</tr>
<tr>
<td>Orchards/vineyards/other</td>
<td>Other</td>
</tr>
<tr>
<td>Pasture/hay</td>
<td>Other</td>
</tr>
<tr>
<td>Row crops</td>
<td>Other</td>
</tr>
<tr>
<td>Small grains</td>
<td>Other</td>
</tr>
<tr>
<td>Fallow</td>
<td>Other</td>
</tr>
<tr>
<td>Urban/recreational grasses</td>
<td>Other</td>
</tr>
<tr>
<td>Emergent herbaceous wetlands</td>
<td>Other</td>
</tr>
<tr>
<td>Deciduous forest</td>
<td>Other*</td>
</tr>
</tbody>
</table>

*Note. for Fort Howes complex, all Deciduous Forest appeared to fall outside burned area perimeters and consequently intent was to limit fire severity to Evergreen or Mixed Forest types.

---

**Data Processing and Analysis for Principal Components Analysis Methods**

Principal components analysis (PCA) is frequently used to generate new transformed and uncorrelated data from multispectral satellite imagery (Lillesand and Kiefer, 1987). If based on multiple dates of imagery, such analysis can simplify change detection by isolating important spectral indicators of landscape change within fewer bands (Fung and LeDrew, 1987; 1988). For this study, all 14 bands from the pre- and post-fire images were combined and subjected to a PCA. The first seven components were converted to image files and visually inspected to identify the two most closely associated with burned areas, in this case PC2 and PC5. With a spectral scattergram of these two components linked to the two-date imagery in ERDAS IMAGINE®, specific areas of burned and unburned vegetation were examined in relation to their position in the scattergram. A new set of rotated and translated axes was drawn on the scattergram (Figure 2) to more closely identify the burn/no burn threshold. A first-order burn index was calculated for each pixel based on its distance in spectral space to the transformed

---

**Figure 2. Scattergram showing numbers of pixels**

(represented by a forested vegetation class in the National Land Cover Data) in relation to their calculated values for principal components PC-2 and PC-5. Darker areas correspond to combinations of PC-2 and PC-5 values that were shared by more pixels, whereas light or white areas indicate combinations that were less common or absent. Pixels with high values for PC-5 and low values for PC-2 (shown in upper left quadrant) tended to represent forest that was burned. The line separating burned from unburned pixels was drawn manually on the basis of PC-5 and PC-2 values as shown.
algorithms are commonly used to classify unknown
and post-fire methods produced a more homoge-
TM was least sensitive to ratio) (Key bands plus a Normalized Difference Vegetation
ANN software. was nega-
results were

were assigned lifeform type to each pixel based on
representing shrub and grass vegetation. To accomplish this,
vegetation differently (i.e., two severity classes) from pixels
hence, we needed to classify pixels representing forest
necessary to classify fire severity differently by lifeform;
split into classes based on histogram modality. For shrubs
Index values then were displayed as separate frequency
als from a forested vegetation class in the National Land Cover Data.

x-axis. For each pixel the index was further modified by the
following four equations to increase its power to discrimi-
nate between burned and unburned vegetation:

\[
\text{Index} = \text{Index} + \text{(pre - fire TM4 -post-fire TM4)}
\]

\[
\text{Index} = \text{Index} + \text{PC5}
\]

\[
\text{Index} = \text{Index} + (-\text{PC2})
\]

\[
\text{Index} = \text{Index} + \text{conditional (NBR-global mean NBR > 0, then } \sqrt{\text{NBR, else 0})}
\]

Because brightness values from TM channel 4 tended to
be lower in burned areas, the first modifier would increase
the index value for burned pixels. Similarly, because PC5 was
positively correlated with burned areas, and PC2 was nega-
tively correlated with them, the second and third modifiers
also increased the index value of burned pixels. Finally, if
the modified index for a given pixel was greater than the
global mean value of the index, then the square root of that
pixel’s index value was added to create the final index value.

To facilitate comparison with the other methods, it was
necessary to classify fire severity differently by lifeform;
hence, we needed to classify pixels representing forest
vegetation differently (i.e., two severity classes) from pixels
representing shrub and grass vegetation. To accomplish this,
we assigned lifeform type to each pixel based on NLCD.
Index values then were displayed as separate frequency
histograms representing tree, shrub and grass lifeforms, and
split into classes based on histogram modality. For shrubs
and grasses, a bimodal split was required to discriminate
between burned and unburned vegetation. But, for pixels
representing forest, unburned needed to be distinguished
from two classes of burn severity; hence the histogram from
forested pixels was split in three groups (Figure 3).

Data Processing and Analysis for Machine Learning Methods
Image features were extracted through a series of hierar-
chical and contextual classifications based on two inductive
learning algorithms: an artificial neural network (ANN),
specifically a back propagation one (Rumelhart et al., 1986),
and k-nearest neighbor (K-NN). Artificial neural networks are
widely used for image classification and more recently for
change detection (Abuelgasim et al., 1999; Gopal and
Woodcock, 1999; Dai and Khorram, 1998) in part because
they are fast and robust to noisy training data (Mitchell,
1997). Learning is achieved through the adjustments to
weights assigned to each node in the neural net; and back
propagation allows for the creation of disjunctive functions,
which makes this type ideal for learning tasks involving
feature extraction and pattern recognition (Mangrich, 2001).
The K-NN algorithms are commonly used to classify unknown
units based on the labels of the “k” nearest known (e.g.,
training) examples in Euclidean space (Mitchell, 1997).
Although K-NN algorithms may produce more accurate results
than other classifiers, their relative performance can be slow
in terms of processing time (Bain, 2000). For this reason,
they were used sparingly in the machine learning sequence
and primarily to refine the results from faster artificial neural
net classifications (Mangrich, 2001). Finally, in contrast to
traditional image processing techniques that rely on spectral
information for classification of spatial units, the machine
learning approach takes into account the context in which
units occur. In other words, learning decisions are based not
only on a unit’s spectral information, but also on its shape
and on the shape and color of all units within a surrounding
area (in this case a \(5 \times 5\) Manhattan pattern). Direct benefits
to classification accuracy from incorporating spatial context
in this manner were shown by Opitz (2002).

For the single-date, post-fire analysis, all seven TM bands
from the 27 August 2000 image were used for classification.
For the two-date, pre-fire versus post-fire analysis, 12 bands
were used. Eight came from the pre-fire image: the seven
original TM bands plus a Normalized Difference Vegetation
Index (NDVI) based on the ratio of \(\text{TM4 - TM3}/(\text{TM4 + TM3})\)
(Nemani et al., 1993); and four came from the post-fire image
(bands 4, 6, and 7, plus a ND4/7 ratio) (Key et al., 2002).
Finally, two of us (Opitz and Mangrich) assumed primary
responsibility for this portion of the study. Further details
about the methods may be found in Mangrich (2001). Since
the project was completed, a more sophisticated ML algorithm
was developed and released as Feature Analyst (www.feature-
analyst.com), a commercial extension to ArcGIS® software.

Results
Area and Patterns
The total area burned and unburned varied relatively little
among the six methods (Table 2). ND4/5 was least sensitive to
burns (18,829 hectares), whereas the PCA4/7 classified 9,503
additional hectares as having burned. Within the burned and
unburned categories, however, the results varied among the
methods. The ANN methods classify more lethal tree and
shrub classes, whereas the PCA methods classified more
mixed tree and grass types. In addition to differences in the
totals and relative class proportions, there were pattern
differences. The ANN methods produced a more homoge-
neous and blocky pattern compared to the PCA and ND
methods (Plate 1). This difference can be attributed to two
processing steps in the ANN methods. First, they were based
on a \(3 \times 3\) moving window, or pixel neighborhood, that
was selected to best match the scale of the local landscape
pattern; second, after classification, the ANN results were
merged to a 0.9 ha minimum mapping unit (see the Methods
Section). In contrast, the PCA and ND methods were pixel
classifications that were neither filtered nor merged.

Classification Accuracy
Considering just the ability of each method to classify
burned versus unburned vegetation without regard to
lifeform, we note that all performed exceptionally well. The users' accuracies for all six methods were 100 percent for the burned class, and ranged from 68.9 percent for ND4/5 to 100 percent for ANN2 for the unburned class (Table 3). The producers' accuracies were similarly high but reversed, with the unburned class being consistently perfect (100 percent), and the burned accuracies ranging from 65.9 percent for ND4/5 to 100 percent for ANN2.

The PCA methods and the ND methods were stratified using lifeform information from the NLCD. To account for the differences in classification logic and thematic errors in the NLCD (see the Methods Section), we adjusted the accuracy assessment data to allow for differences in lifeform, as long as they were correct at the burn/unburn level. This resulted in the test data fitting the PCA3/4, PCA4/7, and the ND4/7 substantially better than either of the two ANN methods (95.8
Table 2. AERIAL EXTENT, IN HECTARES, BY GROUP AND CLASS FOR THE SIX METHODS TESTED

<table>
<thead>
<tr>
<th>Group</th>
<th>Class</th>
<th>ANN1</th>
<th>ANN2</th>
<th>PC3/4</th>
<th>PC4/7</th>
<th>ND4/5</th>
<th>ND4/7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burned</td>
<td>Lethal Tree</td>
<td>8,576</td>
<td>10,654</td>
<td>5,282</td>
<td>7,591</td>
<td>5,158</td>
<td>8,827</td>
</tr>
<tr>
<td></td>
<td>Mixed Tree</td>
<td>6,810</td>
<td>5,442</td>
<td>8,764</td>
<td>2,168</td>
<td>2,228</td>
<td>2,096</td>
</tr>
<tr>
<td></td>
<td>Shrub</td>
<td>6,774</td>
<td>6,164</td>
<td>9,867</td>
<td>8,863</td>
<td>4,760</td>
<td>6,617</td>
</tr>
<tr>
<td></td>
<td>Grass</td>
<td>2,399</td>
<td>2,396</td>
<td>104,251</td>
<td>103,456</td>
<td>107,559</td>
<td>105,701</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>24,559</td>
<td>24,646</td>
<td>24,281</td>
<td>25,422</td>
<td>18,829</td>
<td>22,607</td>
</tr>
<tr>
<td>Unburned</td>
<td>Tree</td>
<td>92,223</td>
<td>89,218</td>
<td>56,976</td>
<td>56,690</td>
<td>58,621</td>
<td>57,129</td>
</tr>
<tr>
<td></td>
<td>Shrub</td>
<td>36,021</td>
<td>36,982</td>
<td>18,618</td>
<td>18,557</td>
<td>19,118</td>
<td>18,690</td>
</tr>
<tr>
<td></td>
<td>Grass</td>
<td>39,228</td>
<td>47,289</td>
<td>104,251</td>
<td>103,456</td>
<td>107,559</td>
<td>105,701</td>
</tr>
<tr>
<td></td>
<td>Other</td>
<td>18,848</td>
<td>12,731</td>
<td>6,748</td>
<td>6,748</td>
<td>6,748</td>
<td>6,748</td>
</tr>
<tr>
<td></td>
<td>Subtotal</td>
<td>186,320</td>
<td>186,220</td>
<td>186,593</td>
<td>185,451</td>
<td>192,046</td>
<td>188,268</td>
</tr>
<tr>
<td>Totals</td>
<td></td>
<td>210,879</td>
<td>210,866</td>
<td>210,874</td>
<td>210,873</td>
<td>210,875</td>
<td>210,875</td>
</tr>
</tbody>
</table>

Table 3. CLASS LEVEL FRACTION-CORRECT AND PERCENT-CORRECT USER’S AND PRODUCER’S ACCURACIES FOR BURNED AND UNBURNED

<table>
<thead>
<tr>
<th></th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burned (User)</td>
<td>38/38</td>
<td>100.0%</td>
<td>10/10</td>
<td>100.0%</td>
<td>37/37</td>
<td>100.0%</td>
<td>27/27</td>
<td>100.0%</td>
<td>37/37</td>
<td>100.0%</td>
</tr>
<tr>
<td>Unburned (User)</td>
<td>31/34</td>
<td>91.2%</td>
<td>10/10</td>
<td>100.0%</td>
<td>31/35</td>
<td>88.6%</td>
<td>31/45</td>
<td>68.9%</td>
<td>31/35</td>
<td>88.6%</td>
</tr>
<tr>
<td>Mean User’s Acc.</td>
<td>95.6%</td>
<td>100.0%</td>
<td>94.3%</td>
<td>100.0%</td>
<td>85.3%</td>
<td>94.3%</td>
<td>85.5%</td>
<td>94.3%</td>
<td>85.5%</td>
<td>94.3%</td>
</tr>
<tr>
<td>Burned (Producer)</td>
<td>38/41</td>
<td>92.7%</td>
<td>41/41</td>
<td>100.0%</td>
<td>37/37</td>
<td>86.8%</td>
<td>31/35</td>
<td>74.4%</td>
<td>37/41</td>
<td>89.2%</td>
</tr>
<tr>
<td>Unburned (Producer)</td>
<td>31/31</td>
<td>100.0%</td>
<td>31/31</td>
<td>100.0%</td>
<td>31/31</td>
<td>100.0%</td>
<td>31/31</td>
<td>100.0%</td>
<td>31/31</td>
<td>100.0%</td>
</tr>
<tr>
<td>Mean Producer’s Acc.</td>
<td>96.3%</td>
<td>100.0%</td>
<td>95.1%</td>
<td>100.0%</td>
<td>95.1%</td>
<td>100.0%</td>
<td>95.1%</td>
<td>100.0%</td>
<td>95.1%</td>
<td>100.0%</td>
</tr>
<tr>
<td>Overall</td>
<td>69/72</td>
<td>95.8%</td>
<td>72/72</td>
<td>100.0%</td>
<td>68/72</td>
<td>94.1%</td>
<td>68/72</td>
<td>80.5%</td>
<td>68/72</td>
<td>94.4%</td>
</tr>
</tbody>
</table>

Table 4. CLASS LEVEL FRACTION-CORRECT AND PERCENT-CORRECT USER’S ACCURACY WITH ADJUSTMENTS FOR NATIONAL LAND COVER DATASET LIFEFORM ASSIGNMENT

<table>
<thead>
<tr>
<th></th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
<th>Fraction</th>
<th>% Correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-lethal Tree</td>
<td>10/10</td>
<td>100.0%</td>
<td>10/12</td>
<td>83.3%</td>
<td>7/10</td>
<td>70.0%</td>
<td>7/10</td>
<td>70.0%</td>
<td>5/8</td>
<td>70.0%</td>
</tr>
<tr>
<td>Mixed Tree</td>
<td>10/11</td>
<td>90.9%</td>
<td>10/10</td>
<td>100.0%</td>
<td>7/8</td>
<td>70.0%</td>
<td>7/10</td>
<td>100.0%</td>
<td>7/10</td>
<td>100.0%</td>
</tr>
<tr>
<td>Lethal Tree</td>
<td>10/12</td>
<td>83.3%</td>
<td>10/10</td>
<td>100.0%</td>
<td>12/12</td>
<td>100.0%</td>
<td>10/10</td>
<td>100.0%</td>
<td>10/11</td>
<td>100.0%</td>
</tr>
<tr>
<td>Burn Grass</td>
<td>1/3</td>
<td>33.3%</td>
<td>5/6</td>
<td>83.3%</td>
<td>16/16</td>
<td>100.0%</td>
<td>16/16</td>
<td>100.0%</td>
<td>10/9</td>
<td>100.0%</td>
</tr>
<tr>
<td>Unburn Grass</td>
<td>5/10</td>
<td>50.0%</td>
<td>4/4</td>
<td>100.0%</td>
<td>20/21</td>
<td>95.2%</td>
<td>20/21</td>
<td>95.2%</td>
<td>20/28</td>
<td>95.2%</td>
</tr>
<tr>
<td>Burn Shrub</td>
<td>7/12</td>
<td>58.3%</td>
<td>10/15</td>
<td>66.7%</td>
<td>1/1</td>
<td>100.0%</td>
<td>1/1</td>
<td>100.0%</td>
<td>0/0</td>
<td>100.0%</td>
</tr>
<tr>
<td>Unburn Shrub</td>
<td>9/14</td>
<td>64.3%</td>
<td>9/15</td>
<td>60.0%</td>
<td>4/4</td>
<td>100.0%</td>
<td>4/4</td>
<td>100.0%</td>
<td>4/5</td>
<td>100.0%</td>
</tr>
<tr>
<td>Mean User’s Acc.</td>
<td>68.6%</td>
<td>84.8%</td>
<td>93.2%</td>
<td>95.0%</td>
<td>93.2%</td>
<td>95.0%</td>
<td>83.4%</td>
<td>95.0%</td>
<td>83.4%</td>
<td>95.0%</td>
</tr>
<tr>
<td>Mean Producer’s Acc.</td>
<td>72.2%</td>
<td>80.4%</td>
<td>92.9%</td>
<td>94.3%</td>
<td>79.2%</td>
<td>94.3%</td>
<td>95.0%</td>
<td>95.0%</td>
<td>95.0%</td>
<td>95.0%</td>
</tr>
</tbody>
</table>

Discussion
Burn severity or fire severity have been reported in a variety of research applications, including surface runoff and sediment yields (Robichaud and Waldtrop, 1994), burned area relationships to natural reforestation (Lopez-Garcia and Caselles, 1991), forest stand conversion and regeneration establishment (Blackwell et al., 1992; Blackwell et al., 1995), restoration of natural fire regimes with prescribed fire programs (Brown et al., 1995; Keifer and Stanzler, 1995), and wildlife habitat components (Hutto, 1995). This variety

to 93.1 percent versus 80.6 to 72.2 percent: Table 4). The unadjusted results are also reported and reflect a reversed relationship (50.0 to 55.6 percent versus 80.6 to 72.2 percent: Table 5). We fully acknowledge that adjusting the accuracy assessment data over-estimates, and not adjusting under-estimates, the actual user accuracies of these methods. The actual range may be closer to the ANN2 method, thus we cannot conclude that the PCA3/4, PCA4/7, and the ND4/7 methods outperform the ANN2 method, just that they likely perform about as well.
of analysis objectives responds differently to alternative characterizations of landscape pattern, as well as to error structure, suggesting that efforts should be made to match the data model and error structure to the intended analyses. Evaluation and comparison of these results should be interpreted in the context of both methodological differences and analytical objectives.

We frame the discussion in the context of three key methodological considerations: 1) lifeform stratification, 2) image data temporal requirements, and 3) quantitative image interpretation of fire severity.

1. As previously noted, the PCA and ND methods were stratified using lifeform information from the NLCD. NLCD lifeform stratification provides a consistent and continuous basis for standardizing a national application and eliminates new lifeform-level errors in the fire severity product, however there are also some disadvantages. The relative abundance of lifeforms in a project area, the differences between the NLCD classification logic and other vegetation classification logic, as well as the need for national or regional consistency should all be considered in selecting a local vegetation dataset or the NLCD for stratification. An alternative approach could be to use the pre-fire image data to generate a lifeform-level classification specific to the site. This approach would provide a local stratification that reflects current land-cover conditions and have known error. Conversely, the machine learning methods (ANN1 and ANN2) were not stratified to lifeform. The absence of lifeform stratification provides the opportunity for reference data to reflect local vegetation conditions and information needs, but it increases the need for consistency among different project areas.

2. The second methodological consideration useful for evaluating and comparing these methods is the temporal data requirements. ANN1 is the only method tested that was based on a single post-fire image date. If a project area had no pre-fire image data available, the ANN1 method would be the only viable option of the methods tested.

3. The third methodological consideration is the amount of quantitative image interpretation required of the image analyst. ANN1 and ANN2 used reference data from photograph interpretation and required no quantitative image interpretation. Within the lifeform stratifications, the PCA methods required no quantitative image interpretation as the results come solely from the analysis of the PCA index. Conversely, the ND methods identified threshold values through image interpretation that established the class memberships within lifeforms.

Within the context of the methodological differences, all of the methods tested provide an adequate basis for most analysis applications. The accuracy estimated for the ANN1 method suggests higher error than the other methods; but as previously noted, this method used only post-fire image data. Similarly, the ND4/5 method has a higher error estimate than the other image differencing methods (PCA3/4, PCA4/7, and ND4/7). Although this level of error may be acceptable, the other image differencing methods can be implemented with comparable data requirements and analysis time to produce more accurate results.

The accuracy assessment results for PCA3/4 and PCA4/7 suggest both these methods produce highly accurate results. However, the degree of analyst input required to apply the same analytical logic from selection of best principal components to the consistent calculation of burn index values for image pixels representing different TM scenes and geographic areas could become problematic. This leaves the ND4/7 method, which produced results equal to the PCA4/5, but in fewer, less complicated steps and with more consistently applied analytical logic.

At the time of this test, the software used for the ANN1 and ANN2 was not readily available for production work. Since the project was completed, a more powerful version was developed and released as Feature Analyst®, a commercial extension to ArcGIS® software. The accuracy assessment suggests that the ANN methods can provide flexible and robust analytical approaches to meet similar objectives, but they do require more extensive reference data than are typically available, and extensive coordination would be required to maintain consistency between different project areas.

Conclusions

The Normalized Burn Ratio (ND4/7) provided a flexible, robust, analytically simple approach that could be applied anywhere within the continental United States. The single-date, machine-learning method (ANN1) provided the only viable method tested for use when no pre-fire image data are available. The two-date, machine-learning method (ANN2) provided the greatest opportunity for consistency with local or regional vegetation maps, given the availability of pre-fire and post-fire reference data.

Classified remote sensing datasets developed through one or more of these processes can be an important information source for exploring the relationships between fire severity, pre-fire conditions, and biophysical settings. Additional data and analyses may be required, however, before inference can be made about fire effects.
Epilogue

Based on the results of this test, the ND4/7 was implemented operationally to classify and map fire severity over 1.2 million hectares burned in the Northern Rocky Mountains and Northern Great Plains during the 2000 fire season as well as the 2001 fire season (Glennin and Brewer, 2002). Approximately the same procedure was adopted in 2001, by the USDA Forest Service, Remote Sensing Applications Center to produce Burned Area Reflectance Classifications for national-level support of Burned Area Emergency Rehabilitation (BAER) activities (Ore�mann, 2002). Maps of wildland fires with associated burn severity are an effective means of analyzing fire effects and have been proven successful on an incident-by-incident basis for Burned Area Emergency Response (BAER) mapping (Bobbe et al., 2001).

Acknowledgments

This study would not have been possible without considerable assistance from people at both the USDA Forest Service and the University of Montana. Jeff DiBenedetto at the Custer National Forest provided post-fire aerial photography and ground-based photography of the study area. Judy Tripp at the Northern Regional Office provided editorial assistance. At the University of Montana, Gary Gooch interpreted aerial photographs, Jim Schumacher prepared all the graphic illustrations, and Shane Mason helped parameterize implementations of the machine learning algorithm. We graciously thank these individuals, as well as three anonymous reviewers whose constructive comments greatly improved the final version of this paper.

References


Edwards, C., 1990. Image segmentation, cartographic information and knowledge-based reasoning: getting the mixture right, Proceedings of the IGARSS '90 Symposium, University of Maryland, College Park, Maryland.


(Received 05 January 2004; accepted 23 March 2004; revised 20 October 2004)