Chapter 13

Perspectives on LANDFIRE Prototype Project Accuracy Assessment

James Vogelmann, Zhiliang Zhu, Jay Kost, Brian Tolk, and Donald Ohlen

Introduction

The purpose of this chapter is to provide a general overview of the many aspects of accuracy assessment pertinent to the Landscape Fire and Resource Management Planning Tools Prototype Project (LANDFIRE Prototype Project). The LANDFIRE Prototype formed a large and complex research and development project with many broad-scale data sets and products developed throughout its various stages. The scope of the project was defined as mapping and modeling vegetation, wildland fuel, and fire regime characteristics (Rollins and others, Ch. 2). Because of the breadth of the investigation, it is important to base our expectations for accuracy on a clear understanding of the intricacies, interdependencies, and scope of mapping and modeling LANDFIRE products. Our goals in this chapter are to: 1) provide relevant background information regarding accuracies and what was realistically achievable in the LANDFIRE Prototype, 2) provide background regarding our strategies for LANDFIRE National, 3) describe our actual LANDFIRE Prototype accuracy results in broad terms, and 4) provide recommendations for the national implementation of LANDFIRE. This chapter is not intended to provide an exhaustive list and description of all of the various accuracy-related issues and conclusions resulting from the LANDFIRE Prototype (for specific details, the reader will be referred to the appropriate chapters). Rather, this chapter is intended to be broad in scope and to place the many accuracy components within the context of the LANDFIRE Prototype and LANDFIRE National projects. Please note that Lunetta and Lyon (2004) provide an in-depth discussion of the current state of accuracy assessment within the science community.

Background

General Accuracy Tenets and Philosophy

First we will provide the reader with several broad tenets used in defining accuracy assessment for the LANDFIRE Prototype Project and thereby lay the foundation for the more in-depth discussion following.

Tenet 1: Assuming that thematic detail and spatial scale are constant, product accuracy is generally inversely correlated with the size of the region being assessed.

Within the remote sensing literature, there are many references to accuracy levels, and many of the reported values are quite high. These high levels may lead to inflated expectations regarding what types of accuracies will be achievable from LANDFIRE. Many previous studies were conducted within relatively small study

In: Rollins, M.G.; Frame, C.K., tech. eds. 2006. The LANDFIRE Prototype Project: nationally consistent and locally relevant geospatial data for wildland fire management. Gen. Tech. Rep. RMRS-GTR-175. Fort Collins: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station.

areas, often aided by high levels of "hand crafting" during the mapping process and/or in-depth knowledge of the particular study area. We do not have the luxury of spending a great amount of time and effort on any one particular region mapped through the LANDFIRE Project, and the mapping and modeling tasks need to be accomplished through largely automated processes. These limitations do not by any means reduce the value of the products being created through LANDFIRE; however, it should be stated that LANDFIRE products will likely have lower overall accuracies than do data sets derived from more localized studies characterized by large amounts of field data, increased processing effort that may include on-screen digitizing and recoding, and/or iterative refinement of modeled results.

Tenet 2: The higher the thematic detail, the lower the accuracy.

A relatively large number of vegetation classes were mapped for the LANDFIRE Prototype (Long and others, Ch. 6). While the chosen map unit classification system made sense on many levels for the LANDFIRE Prototype, it must be recognized that the proliferation of classes in this or similarly complex systems will imply a relative decrease in accuracy levels. This does not in any way diminish the value of the vegetation products, but is rather simply a result of a more complex map unit classification design. For example, a two-category classification of water and uplands is likely to result in high accuracy, with expected accuracies above 99 percent. This high accuracy does not mean that the *value* of the product is particularly high, but simply reflects that the accuracy for depicting these two classes is high. Additionally, there are difficulties that arise when categorizing continuous phenomena into rigid and discrete classes. For instance, a more detailed map unit classification system might treat juniper and pinyon – juniper ecosystems as several discrete classes even though the boundaries between them are relatively arbitrary and difficult to delineate both in the field as well as within the imagery. With complex vegetation map unit legends, such as that used in the LANDFIRE Prototype, vegetation class accuracy levels can be expected to drop. Nevertheless, LANDFIRE products reliably and consistently describe the distribution of vegetation composition, condition, and structure and associated wildland fuel and fire regimes across broad landscapes. These mapped data are useful for hazardous fuel reduction projects, for a variety of resource management projects, and for both strategic and tactical wildland fire management.

Tenet 3: Field information used for assessing accuracy is not perfect.

As mentioned under Tenet 2, the LANDFIRE Prototype vegetation map unit legends are relatively complex (Long and others, Ch. 6). The map unit classifications are developed using large quantities of field data, and all of the field plots are assigned to one of the many possible classes. Most of these plots are used to generate maps, but some are reserved for use in the accuracy assessment phase of the investigation. We recognize four major potential sources of error associated with field plot data:

- Errors occur frequently in the identification of species and measurement of vegetation structure in the field (for example, in the data for one prototype field plot, a misplaced decimal point indicated a shrub height of 60 feet).
- The vegetation on some field plots has undoubtedly changed between the time the field data were collected and when the imagery was acquired.
- Geo-location errors in plot and imagery data result in inaccurate characterization of some imagery pixels.
- The assignment of plots to specific vegetation classes will have errors associated with the wide array of opinions among professional field ecologists regarding the field classification of any given field plot.

Tenet 4: The modeled results of complex ecological systems will be characterized by ambiguity and controversy.

The products generated from the LANDFIRE Prototype represent our best approximations in depicting the current status of very complex natural phenomena. The information used in our modeling efforts is based on the best available input data and assumptions. However, although our output products represent reasonable and robust depictions of current conditions, we recognize that, due to lack of baseline research, our knowledge of certain ecological systems is imprecise. Use of such information in the modeling process may result in potential flaws in the products, and hence not all of the core LANDFIRE deliverables will be free of error and ambiguity. Nevertheless, the LANDFIRE Project represents an integration of the best available science in remote sensing, ecosystem simulation, landscape fire and succession modeling, predictive landscape mapping, and wildland fire behavior and effects prediction.

We are therefore confident that the products generated represent the best current assessments of the status of these ecosystems with regard to wildland fire and will be of great value to natural resource managers.

Accuracy Assessment Considerations for LANDFIRE

The need for conducting accuracy assessments of the spatial products created from mapping projects has been well documented (Congalton 1991; Foody 2002). Factors that influence map accuracy include (but are not limited to) the remote sensing platform, the quality of ancillary sources of information, the quality of field data, the floristic complexity of the map unit classification system used, and the sampling design. Traditional first-order map accuracy estimates involve generating an error matrix, computing overall accuracy, and estimating "producer's accuracy" and "user's accuracy" (Congalton 1991). In the past, assessment of map accuracy has involved much post-mapping fieldwork in order to develop error matrices. These formal, traditional accuracy assessments involving field campaigns can be labor-intensive, time-consuming, and cost-prohibitive, especially when dealing with projects that cover large regions of diverse and overlapping vegetation composition and conditions (Stehman and others 2000). For this reason, only a few efforts have conducted accuracy assessments across broad expanses such as the entire United States (Stehman and others 2003; Wickham and others 2004).

Techniques that worked well in assessing mapping accuracy across large regions for the 1990s National Land Cover Database (NLCD; Vogelmann and others 2001) employed modifications of traditional accuracy assessment methodologies (Stehman and others 2003; Wickham and others 2004). As background, the 1990s NLCD database was developed using Landsat satellite imagery acquired for the Multi-Resolution Land Characteristics (MRLC) 2001 consortium using methods previously described (Vogelmann and others 1998). During development of the database, it was determined that an accuracy assessment for the large area product was required, and that such an effort would have to be modified from more traditional assessments. The modifications were necessary in part due to the scarcity of field data across the mapped regions, the large size of the area being assessed (and associated high costs of collecting data from a statistically valid number of field locations across the entire conterminous United States), difficulties in assigning unambiguous map unit labels to many field plots, and geolocational errors associated with field plot and satellite-derived mapping information.

Three important lessons learned from the accuracy assessments of the 1990s NLCD effort pertain directly to the accuracy assessment methods used during the LANDFIRE Prototype Project:

- Collecting data for and compiling custom field databases is time consuming and expensive. Similarly, combining data from disparate sources and distilling them into a training database for mapping purposes is time consuming, expensive, and can result in data inconsistencies unless special effort is made to crosswalk and/or standardize input data. On the other hand, using existing field data, rather than collecting custom field data, saves both time and money. In short, for large-area projects, it makes sense to use existing field data for conducting accuracy assessments.
- Determining accuracy values for different subregions is acceptable when mapping large regions. Accuracies are likely to vary across large mapped areas due to region-specific heterogeneity in landscape composition and structure, and it was advantageous to derive an understanding of the geographic variability of accuracies of the products developed for LANDFIRE. To this end, use of a systematic random sampling design can provide optimal results. Such a design ensures that all geographic regions are adequately sampled and thereby ensures that at least some estimates of accuracies exist throughout the entire study region.
- Some errors are more "wrong" than others. For instance, for the LANDFIRE effort, misclassification of a pinyon – juniper stand as a riparian woodland stand will likely have a greater negative impact on the predicted fire behavior than misclassification of a pinyon – juniper stand as a juniper stand. Furthermore, some vegetation types are spectrally and biogeographically very similar to other vegetation types, and even with "perfect" source material, it is difficult to adequately distinguish some of these classes. For example, Douglas-fir and white fir are spectrally very close (fig. 1), and both species inhabit similar ecological niches. In regions where both Douglas-fir and white fir occur, we can expect significant confusion between the two classes. For instance, in central Utah, cross validation accuracies for these two classes were quite low, as anticipated. Nonetheless, we suspect that the errors related to misclassifying similar vegetation types will only minimally impact predicted fire behavior, whereas

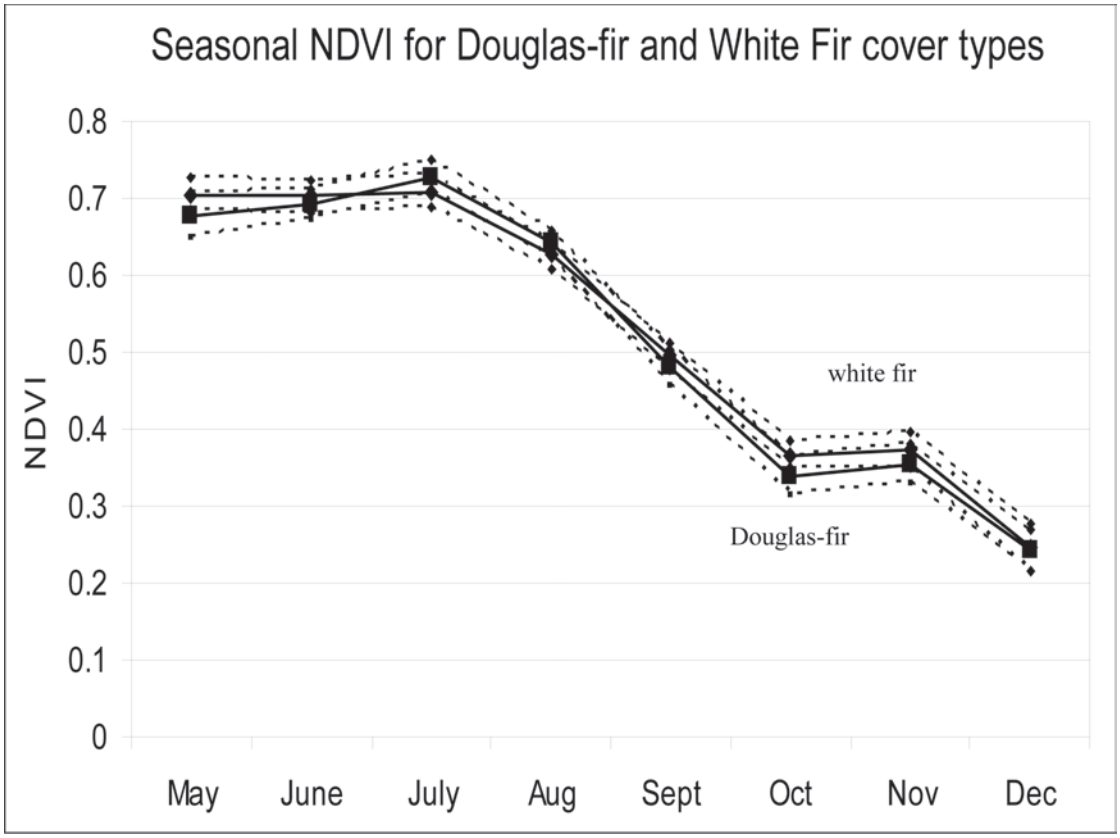

Figure 1—Seasonal normalized difference vegetation index (NDVI) spectral profiles for Douglas-fir and White Fir cover types.

errors related to misclassifications of more dissimilar vegetation types lead to greater negative impact. For this reason, both ecologists and image analysts need to critically analyze error matrices in order to fully understand and characterize the ways in which product errors may affect project objectives.

We took these lessons into consideration in the design of our LANDFIRE accuracy assessment protocol:

- Because LANDFIRE is a large-region project, we tapped into a variety of data sources and made use of existing field data to assess the accuracy of LANDFIRE Prototype products (rather than wasting time and money collecting data for and compiling a custom field database). See Caratti, Ch. 4 for details on the acquisition of data for and compilation of the LANDFIRE reference database.
- Cross-validation error matrices were generated and examined separately for both LANDFIRE Prototype regions.
- For the LANDFIRE Prototype, mappers, ecologists, and wildland fire scientists critically evaluated errors at several stages in prototype product development. These evaluations resulted in aggregation and disaggregation of classes based on the "mappability" and "model-ability" of the vegetation classes. See Keane and Rollins, Ch. 3 and Long and others, Ch. 6 for detailed descriptions of the creation of the final vegetation legends for the LANDFIRE Prototype. This expert-based process for map unit classification refinement is built into the accuracy assessment system for LANDFIRE National.

Overview of Accuracy Assessment Conducted for the LANDFIRE Prototype Project

The LANDFIRE Prototype Project involved many sequential steps, intermediate products, and interdependent processes, each involving evaluations of the accuracy

of intermediate and final products. Please see appendix 2-A in Rollins and others, Ch. 2 for a detailed outline of the procedures followed to create the entire suite of LANDFIRE Prototype products.

Role of Input Data

Field data accuracy issues—Field data played a critical role in many stages of the LANDFIRE Prototype. These data were essential inputs for developing the vegetation products, percent canopy cover and height data layers, and potential vegetation data layers. See Caratti, Ch. 4 for detailed information on data acquisition for and compilation of the LANDFIRE reference database.

Described below are a number of data quality issues that needed to be addressed in the LANDFIRE Prototype.

- *Number of field plots*: For the LANDFIRE Prototype accuracy assessment, we used all field plot data that met the stringent quality-control criteria (Caratti, Ch. 4) and represented the large number of classes mapped during the vegetation mapping tasks (for details about the vegetation mapping procedures, see Frescino and Rollins, Ch. 7 and Zhu and others, Ch. 8) We used literally thousands of points for each of the two prototype regions. During this process, we recognized that some vegetation classes had limited numbers of field plots. Short of gathering additional plot information (see Keane and Rollins, Ch. 3 for LANDFIRE Prototype design criteria), there was no obvious solution to this problem. We attempted to map these rarely sampled vegetation types, even when we had limited numbers of field plots for those classes. We believe that most of these rare classes were under-represented in the resultant products.
- *Field plot geolocational accuracy*: Field plots must have accurate geolocational coordinates to geographically rectify with the many spatial databases involved in the LANDFIRE process. This was especially important during the vegetation cover and structure characterization phase of the LANDFIRE Prototype, wherein each field plot was matched with a single Landsat pixel and used in the mapping process. Any significant error in the field location coordinates has the potential to match the wrong spectral information with that particular field plot, thereby resulting in mapping error. For the prototype effort, we overlaid plot locations onto satellite imagery to determine whether there were plots that obviously did not match the imagery. While most plot locations appeared to be reasonable, we observed that many plots representing natural vegetation were actually located on major roads. When plot information was originally acquired for these sites, the actual Global Positioning System (GPS) measurements were apparently made at the road locations adjacent to the field plots, rather than within the field plots. Thus, the GPS locations did not exactly match the locations where the field measurements were made. For these sites in the LANDFIRE Prototype, a new set of geolocations was derived to better represent actual field plot locations.

  In another case, we noted (also based upon imagery assessment) that many putative shrub sites were located in obviously forested areas. We later discovered that those plots corresponded to a particular project in which the main focus was to describe shrub vegetation regardless of whether or not it represented the dominant vegetation type. These plots were consequently discarded from the prototype accuracy assessment. Both cases illustrate the need for assessing field plot information in conjunction with satellite imagery to ensure that the field information is accurately recorded.

  Moreover, it should be recognized that satellite imagery can have georeferencing errors as well. As a general rule, the coordinates of most pixels in the imagery used for the LANDFIRE Prototype are within 30 meters of the actual location – but exceptions occur. Even in the case where a pixel has slightly greater than a 15-meter error associated with it, this may be large enough to create a slight yet definite mismatch between the imagery and field information. While there is little that we can do about this problem, we at least need to recognize that some of the error term associated with the products generated will be attributable to this issue.
- *Assignment of field data into discrete vegetation classes*: One of the challenges in generating land cover maps is the stratification into discrete classes of a very complex natural world composed of multiple continuums. Regardless of which vegetation map unit system is used, many vegetation plots will represent elements of two or even more classes, and thus some plots will defy unambiguous categorization. As an example of one such problem, we mapped Juniper and Pinyon – Juniper (PJ) as two distinct classes. In nature, pinyon pine and

juniper often coexist, but sometimes juniper occurs as more-or-less pure stands. We used 25 percent juniper composition as the threshold separating Juniper from Pinyon – Juniper (in other words, if a stand had 75 percent or greater basal area juniper in a stand comprised of both pinyon pine and juniper, it was called "Juniper"; whereas, if it had less than 75 percent juniper, it was called "PJ"). Analysis of seasonal spectral data indicated that many juniper stands were spectrally distinct from many of the PJ stands (fig. 2); however, significant spectral overlap existed between the two classes, as well. After decision tree classification, cross-validation accuracies indicated significant error in the classification of these two cover types (fig. 3). We believe that much of this error is attributable to the artificial boundaries imposed by the classification of a continuum.

- *Temporal correlation between field data and satellite imagery*: Disturbance such as that caused by fire, insects, or logging can alter the sites enough to cause the temporal mismatches between field data and satellite imagery that result in classification problems. For the prototype, we made use of a large volume of existing field data acquired from disparate sources (Caratti, Ch. 4), and much of the field information was acquired over a long period of time. Although information from many plots was relatively old (for example, field data acquired over a 10-year time period prior to imagery acquisition), we determined that many of these plots still contained information that was useful and relevant to the LANDFIRE Prototype. For example, plots located within reasonably intact and undisturbed forests or sagebrush lands, under normal circumstances, do not change much over a 10-year span. After completing the first prototype study in Utah, we recognized the importance of using a change-detection approach and employed such an approach in the northern Rockies prototype region to discard plot information derived from areas that changed between the times when the field information was obtained and when the imagery was acquired.

Geospatial data issues—Landsat imagery data from the MRLC 2001 consortium served as the primary source of spatial data for developing the vegetation and structure products (Homer and others 2004) (refer to

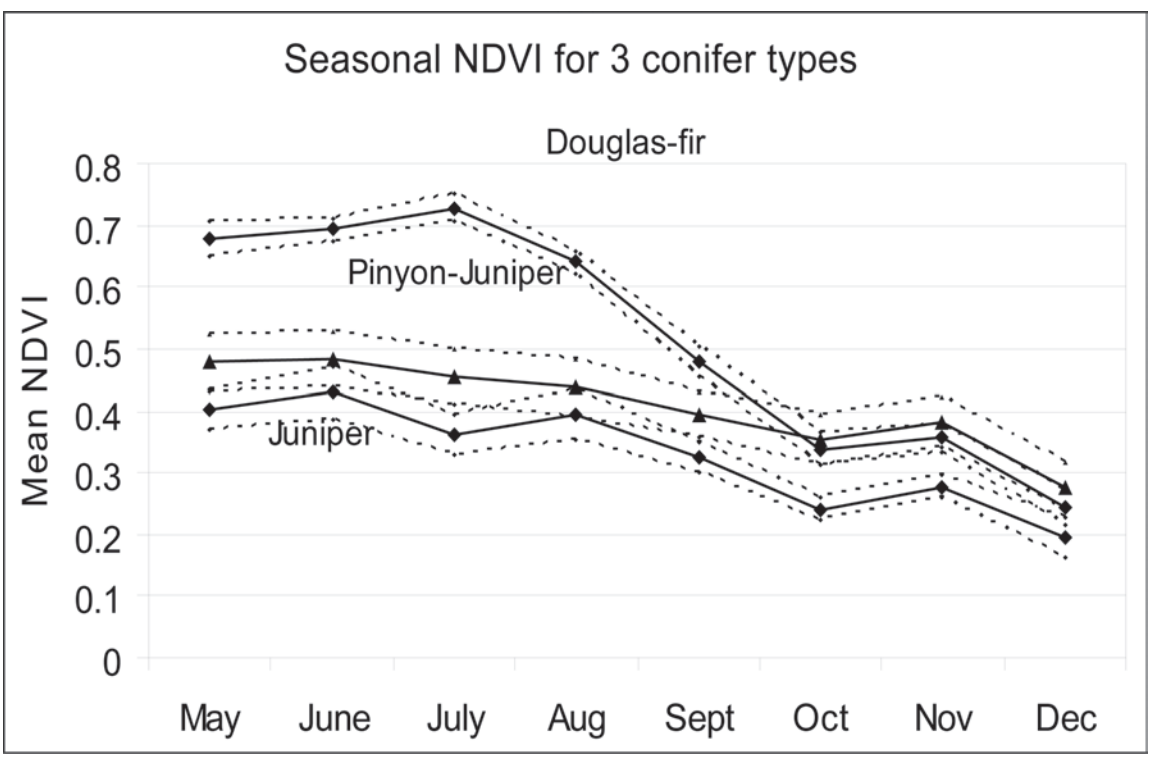

Figure 2—Seasonal normalized difference vegetation index (NDVI) spectral profiles for Douglas-fir, Pinyon – Juniper, and Juniper cover types.

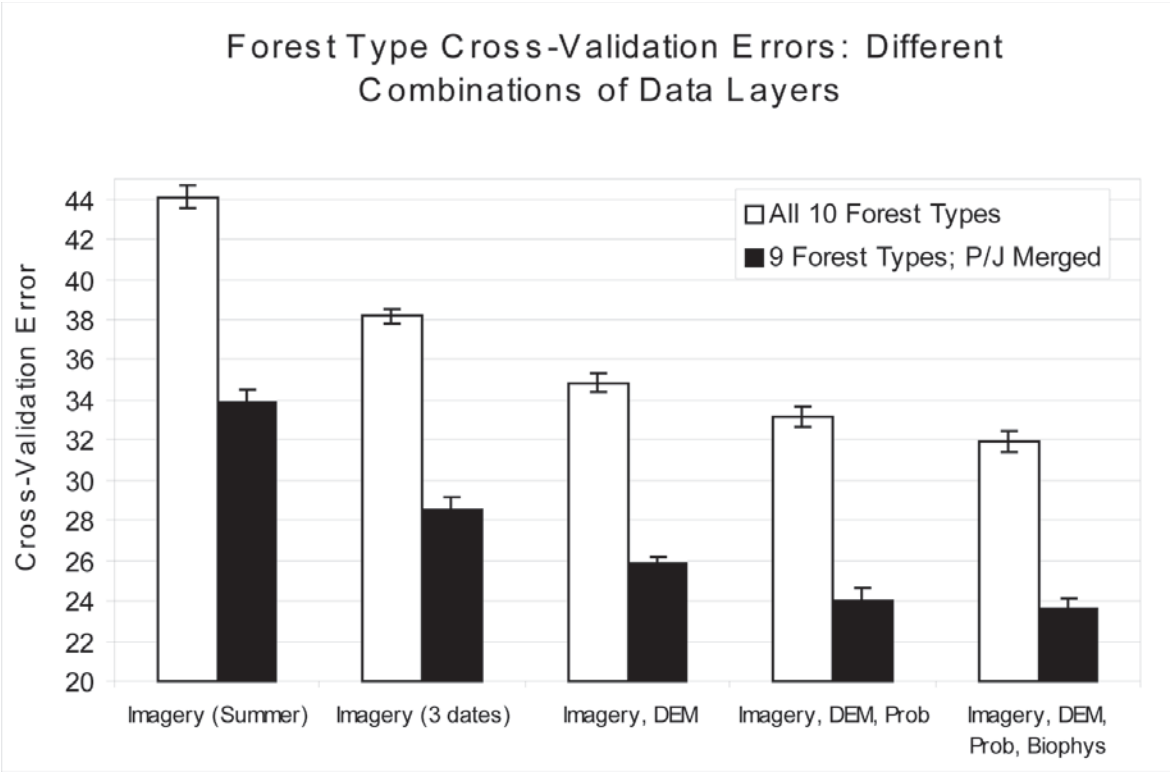

Figure 3—Cross-validation errors for forest types in the Zone 16 prototype study area as a function of different amounts of input source material. Black bars depict the effects of merging the Pinyon – Juniper and Juniper classes.

Zhu and others, Ch. 8 for further discussion regarding the imagery and ancillary data sources used for vegetation mapping in the LANDFIRE Prototype). In general, the images used for the prototype effort were the best data available during the LANDFIRE Prototype and represented three seasonal time periods (leaf-off spring, leaf-on summer, and leaf-on fall). Although the MRLC 2001 data used are of high quality, problems can arise when using any source of remotely sensed information. The foremost imagery-related problems affecting the LANDFIRE Prototype included atmospheric issues, disparate imagery acquisition dates, and geolocational problems.

- *Atmospheric issues*: Most of the acquired image scenes used in the prototype effort were of excellent quality. Even the best scenes, however, have occasional cloud and/or haze problems, which can either totally obstruct the view of portions of landscape or change the digital values enough to impact the mapping process. While not a large problem in the prototype areas, there were a few locations for which imagery quality was sub-par. These issues are inevitable and are likely to be a bigger problem in cloudier locations of the country such as the eastern United States and the upper Midwest.
- *Disparate imagery acquisition dates*: We attempted to use imagery from similar time periods as much as possible; however, due to cloud issues, optimal imagery data were not always available. Using scenes from different dates of the same year, such as using July and September data in the same "leaf-on" mosaic, resulted in problems resulting from phenological differences. Using scenes from different years, such as using one scene from 2002 and an adjacent scene from 2003, resulted in problems related to different weather patterns (for example, vegetation spectral response can be very different during wet versus dry years) and to occasional land cover changes that occurred between years. For the LANDFIRE Prototype, we attempted to minimize these problems through careful selection of scenes and use of spatial "date of acquisition" information in our decision tree and regression tree classifications.

- *Geolocational problems*: Images used in this investigation were processed using the National Landsat Archive Production System methods (USGS Landsat Website 2004). Data were corrected for terrain and projected to a standard projection (Albers Equal Area) using automated software processing. Individual pixel coordinate information was approximately 30 meters from actuality. Thus, even when field information had precise GPS coordinates, the field data were sometimes linked to the wrong pixel due to imagery registration errors. Because of technological, time, and budget constraints, we could not circumvent this problem. Registration methods needed to be consistent and automated to ensure that the process was feasible for application over the entire United States. We simply had to assume that the field data adequately characterized an area broader than the precise location of the plot and that the image pixel used was spectrally representative of its surrounding pixels. Note that in many cases, the quality-control checks performed on the field data mitigated some of these problems.

Ancillary data issues—Other sources of input information for the LANDFIRE Prototype included Digital Elevation Model (DEM) data and derivative products, 1990s NLCD land cover data (Vogelmann and others 2001), 2000s NLCD land cover data (Homer and others 2004), a suite of biophysical gradient data layers (Holsinger and others, Ch. 11; Keane and others 2001; Rollins and others 2004), and potential vegetation information (Frescino and Rollins, Ch. 7). Error terms are associated with each data type. While it is beyond the scope of this chapter to describe in detail all of the sources of errors associated with the many data layers, a few specific points should be made:

- Although not flawless, each data source used in the LANDFIRE Prototype represented the best available science and data quality.
- The source of the DEM data was the National Elevation Dataset (NED) (Gesch and others 2002). Although NED is an excellent source of digital elevation data, it came to our attention during the final stages of the prototype effort that another data source would have been more appropriate: the Elevation Derivatives for National Applications (EDNA) data set (http://edna.usgs.gov). The EDNA data represent a set of data layers derived from an earlier version of the NED. To create the EDNA data layers, the NED data were "smoothed" so that they would be better suited for hydrological modeling purposes. It should also be noted that, regardless of the source of the digital elevation model information, there are horizontal and vertical error terms associated with these data sets tracing back to the original source material. These digital elevation model data sets are regularly improved and updated.
- The 1990s and 2000s NLCD data sets were used for stratification purposes at various stages in the prototype effort, and both data sets have known error terms associated with them. See Yang and others (2001) and Homer and others (2004) for details regarding the accuracies of these products.

Accuracy of Thematic Maps

Cross-validation and points for independent validation—Accuracy assessment is an integral component of land cover mapping work. When a large number of field points are available, a reasonable alternative to generating traditional first-order accuracy estimates (see the above section *Accuracy Assessment Considerations for LANDFIRE*) is cross-validation. To create the LANDFIRE vegetation products, we employed decision tree analysis implemented within the See5 program (Quinlan 1993) using Landsat, DEM, slope, aspect, biophysical gradient, and potential vegetation data layers. The program enables cross-validation, which consists of repeated experiments in which a subset of the sample is used to train a classification model and an unseen subset is used to evaluate the model. In model runs for the prototype effort, we found that a five-fold cross-validation was appropriate. In each model run, the original field point data sets were divided into five subsets of equal size, and each subset was used to evaluate the algorithm trained using the remaining four subsets. Theoretically, this approach is not as thorough as a rigorous, statistically designed post-mapping field accuracy assessment campaign. It has been shown, however, that cross-validation can provide accuracy estimates comparable to these time-consuming and expensive methods (Huang and others 2003). See Frescino and Rollins, Ch. 7 and Zhu and others, Ch. 8 for actual accuracy results and cross-validation error matrices for the vegetation products derived for the LANDFIRE Prototype. For LANDFIRE National, we recommend reserving a set percentage of plots from the decision and regression tree analyses for independent accuracy assessment. See the *Recommendations for National Implementation* section below for details.

Field verification—Although it is not always feasible to conduct a detailed field verification and validation campaign, when possible, field visits at various stages of product development can be highly useful. Field visits, both during and after the product generation phase, provide the technical teams conducting the mapping work with a good basic understanding of the natural vegetation and ecology of the regions in which they are working. Further, field checks of particular sites to determine if they match the modeled results can be very instructive and useful for improving mapping accuracies. For the LANDFIRE Prototype, we made three separate field visits of approximately five days each. We traveled to the central Utah highlands region twice (once before mapping and once after the products were created), and we traveled once to the western Montana region (post-mapping). In all cases, images and/or maps were evaluated in the field, and actual plot measurements were made. Although not statistically rigorous, such efforts provided a better understanding of potential problem areas for future methods improvement. For example, an area of western hemlock was overestimated in the map products, and we were able to trace the overestimation back to problems in the original field sampling methods used to help generate the training data in the mapping process. Although no obvious solution to the problem was apparent, the case illustrates the importance of field visits in methods improvement. In another field activity, spectral measurements of shrub and herbaceous vegetation density were made by one team in the western Montana region to help refine shrub and herbaceous canopy cover methodology. This activity was undertaken in an attempt to improve canopy cover mapping and is being considered for the National Implementation of LANDFIRE.

Consistency checks with data from other sources—Related data sets, generated by other projects and for other applications, are often available and can be used for comparison purposes. The USGS Gap Analysis Program (GAP), for example, generates detailed vegetation maps for conservation management and planning (http://www.gap.uidaho.edu). We compared the GAP products created for the central Utah highlands prototype area with the cover type maps created for the LANDFIRE Prototype. The two sources of data compared reasonably well in some cases and less so in others (see figs. 4 and 5). It should be noted that the GAP products were created using different field databases than those used for the LANDFIRE Prototype. In addition, the vegetation map unit classification systems used were different, which limited the utility of direct, parallel comparison between the GAP products and LANDFIRE products. Although such comparisons may lack statistical rigor, they indicate where major qualitative similarities and differences exist between products and in turn may indicate which classes and regions are the most suspect. In addition, vegetation and structure products should be reviewed by regional experts whenever possible to determine whether noteworthy mapping problems exist and whether additional work is warranted. Such a review is recommended for national implementation of LANDFIRE.

Accuracy of Potential Vegetation Type and Canopy Fuel Maps

We generated potential vegetation type (PVT) data sets using decision tree software and cross-validation routines very similar to those used for generating vegetation maps. We also produced coinciding maps of confidence, which depict the relative prediction errors representing a spatial and visual representation of PVT map accuracy. See Frescino and Rollins, Ch. 7 for detailed descriptions and results of these activities. We estimated the accuracy of canopy fuel layers using regression tree procedures in which correlation coefficients were generated to measure the agreement between the predicted values and actual values. Additionally, we compared with predicted values a set of points randomly selected from the LANDFIRE reference database from each prototype zone. As in the case of PVT, we also produced coinciding maps of confidence. See Keane and others, Ch. 12 for a detailed description of canopy fuel accuracy.

Accuracy of Maps Based on Landscape Simulation Models

Accuracy evaluation of vegetation maps created from satellite imagery and ancillary data is straightforward and is based on a foundation of scientific literature (Foody 2002; Lunetta and Lyon 2004). In contrast, it is often conceptually very difficult to ascertain the quantitative accuracy of many of the products that are generated through complex modeling efforts, such as those employed to create the historical reference conditions for quantifying ecological departure in LANDFIRE. Moreover, it is difficult — if not impossible — to assign an absolute measure of accuracy to an ecological departure product because such a product represents deviation from conditions modeled under a variety of limitations in terms of baseline ecological data. Modeling assumptions, while based on the best available disturbance ecology science, may or may not be completely valid. Without the luxury of time-travel, it is very difficult to validate what the "normal" or historical vegetation condition actually was.

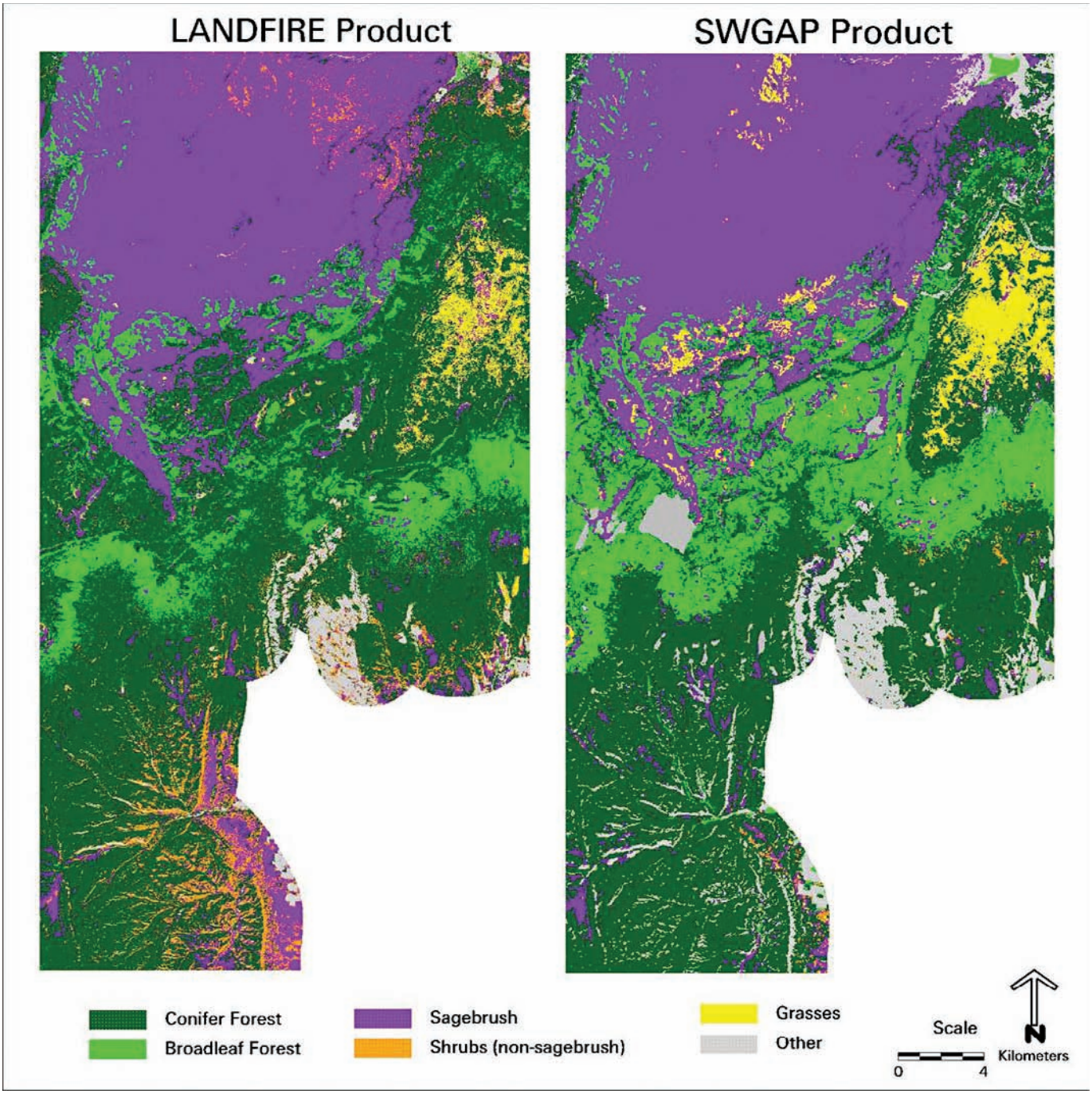

Figure 4—Comparison between a LANDFIRE vegetation type product and a product developed by the Southwest GAP Project in southern Utah. Multiple thematic classes have been combined to facilitate visual comparisons.

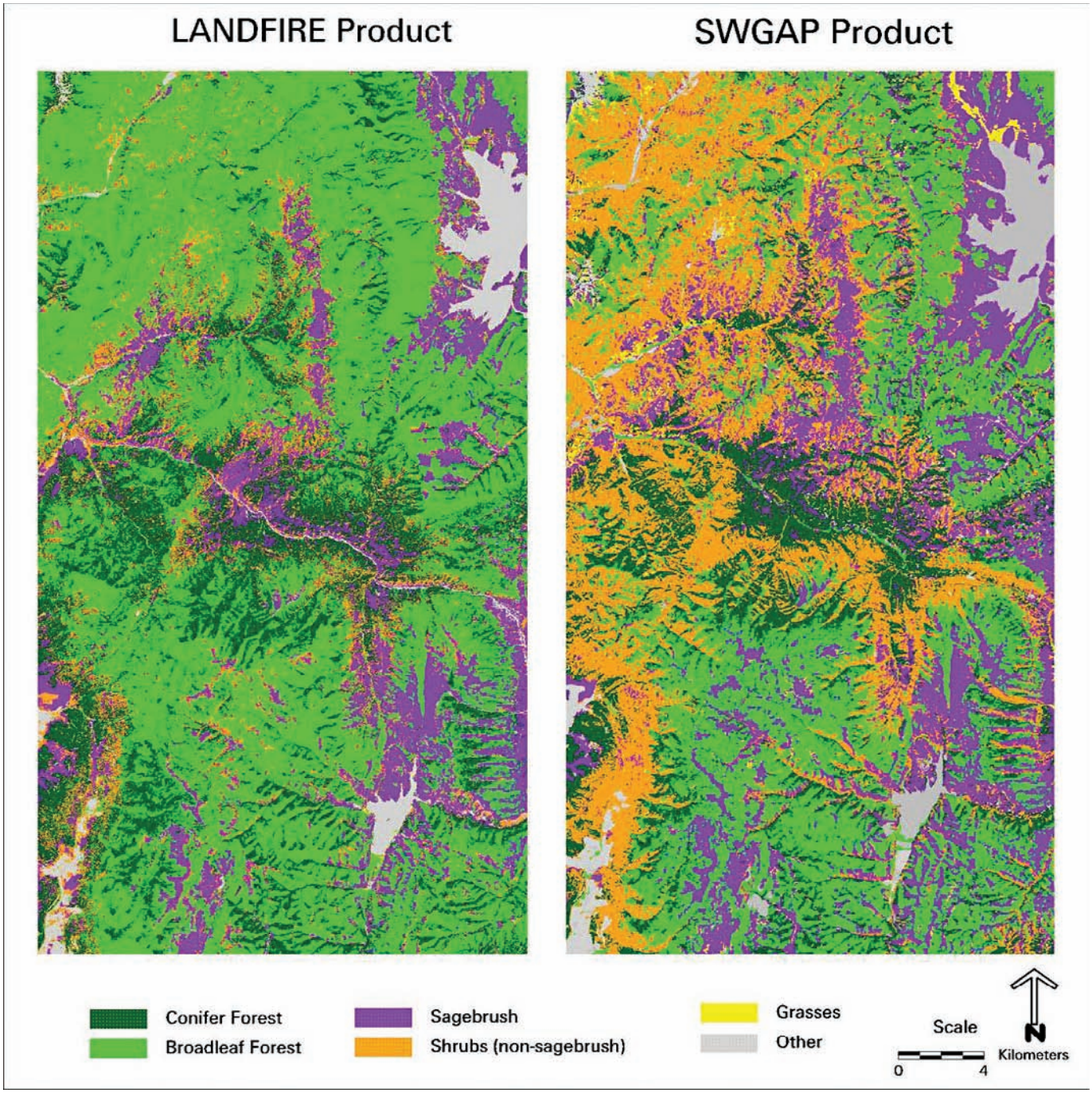

Figure 5—Additional comparison between a LANDFIRE vegetation type product and a product developed by the Southwest GAP Project in southern Utah. Multiple thematic classes have been combined to facilitate visual comparisons. Major differences between shrub and broadleaf forest classes can be traced back to differences in classification systems (Gambel oak and bigtooth maple were categorized as trees in the LANDFIRE map unit classification and as shrubs by GAP).

For accuracy assessment approaches used to evaluate LANDFIRE products based on landscape simulation models, see Pratt and others, Ch. 10 and Holsinger and others, Ch. 11. In addition, see the *Recommendations for National Implementation* section below for suggestions on improving the accuracy assessment of LANDFIRE products based on landscape simulation models.

Recommendations for National Implementation

Source Data

All source data need to be inspected carefully. This is especially true for field data and imagery, which form important foundations for much of the ensuing LANDFIRE tasks. As a matter of course, if field data used for training are inaccurate, then the resulting products will likely have lower levels of accuracy. Imagery quality can also greatly affect accuracy levels of derived products. Although optimal imagery data sets are not always available for a given location, there are usually several excellent options. It is important to ensure that the best possible imagery data sets are used. Below are some specific recommendations regarding the selection of source data.

Number of field plots—As general rule, the more field reference plots, the better. For each LANDFIRE National mapping zone, we anticipate using literally thousands of field plots in order to develop adequate characterizations. These must represent the entire range of conditions that occur throughout the mapping zones. For vegetation map unit classification development, for example, we have a target number of at least 100 plots per class. Fewer plots per class would diminish our confidence in our ability to map that class accurately and would likely result in the inadequate mapping of that particular feature. Rare classes (land cover features limited in occurrence across the landscape) are notoriously difficult to map accurately, largely because there are relatively few field plots representing these classes that can be used for training data. For national implementation of LANDFIRE, we recommend 1) generating vegetation products using all plots, 2) evaluating results, 3) determining which vegetation classes were represented by too few plots, and 4) re-running the map unit classification without these rare classes.

Field plot geolocational accuracy—Field plots with inaccurate coordinates have the potential to cause significant error in mapping results. We recommend that field plot locations be overlaid onto the imagery and that the plot locations be visually inspected to determine if attribute data for each plot are consistent with the imagery. Points located on roads or other locations clearly not characterized by the reference plot should be either omitted or shifted to the appropriate location.

Field data temporal issues—Much of the field information available for the national implementation of LANDFIRE is likely to have been acquired by various organizations over a relatively long period of time. As discussed above, inclusion of plots located in areas where the vegetation has changed between the time the field information was collected and when the imagery was acquired can cause significant mapping problems. The ideal situation is for field data and imagery to be acquired at approximately the same time, but this is impractical due to the large volume of field data necessary for product generation. One option is to discard plots with relatively old information (by imposing an arbitrary cutoff of five or more years); however, including as many plots as possible, even if some include older information, is preferable because even old plots can contain useful information. For this reason, for national implementation, we recommend using the change-detection approach developed for the western Montana prototype area. We recommend using normalized difference vegetation index (NDVI) change between 1990s and 2000s NLCD imagery to locate and isolate plots that have changed markedly over the last 10 years. If a plot is located within a region of high spectral change (based upon imagery analysis) and if the change appears to be related to a land cover change event (such as fire, logging, or insect disease) as opposed to a cloud or cloud shadow, the plot should be flagged and omitted from further analyses.

Imagery data—Imagery acquired by Landsat will likely continue to be the primary source of spatial data for developing vegetation and structure products for LANDFIRE National. The MRLC 2001 consortium, of which the LANDFIRE Project is a partner, is the best source for imagery in part because it is readily obtained and has been consistently pre-processed. Although this imagery represents the best data available, we do anticipate some issues that will need to be addressed. As with the prototype effort, we anticipate the primary imagery-related problems impacting LANDFIRE National to include atmospheric issues, disparate imagery acquisition dates, and geolocational issues (see above section *Geospatial data issues*). It is anticipated that haze and cloud problems will be especially prevalent in the eastern U.S., upper Midwest, and in the Pacific Northwest. Imagery differences related to phenological

variables are also likely to impact mapping on a grander scale than was experienced in the prototype effort. When current MRLC data are deemed insufficient for LANDFIRE purposes (based upon visual inspection), additional scenes should be purchased and processed and incorporated into the mosaicking process.

Ancillary data—LANDFIRE will continue to use the best available source data for national implementation. One change that we recommend is using the EDNA data set (USGS EDNA website 2004) as the primary source of digital elevation data. These data are more refined than the data used in the prototype effort. The 1990s and 2000s NLCD data sets will continue to be used for stratification purposes at various stages of LANDFIRE National.

Accuracy of Output Products

Output product inspection—All LANDFIRE products must initially undergo an inspection phase during which the following question is asked: "Do these products make sense?" Although admittedly subjective, many errors will be caught early in the process through such inspections. If performed properly, such an initial evaluation provides a valuable safeguard that can save time and prevent the need to recreate the products.

Cross-validation and error matrices—As in the LANDFIRE Prototype, we recommend the use of cross-validation for approximating accuracies, especially for existing vegetation type and potential vegetation type. Correlation coefficients derived from regression tree analyses should be used when generating continuous variable data sets. Error matrices should be evaluated to facilitate better understanding of the strengths and weaknesses of the vegetation products. Regarding creation of the mapping models, we recommend using 5- or 10-fold cross-validation for each of the individual LANDFIRE mapping zones.

Points for independent validation—For national implementation of LANDFIRE, we recommend reserving a set percentage of plots from the decision tree and regression tree analyses solely for assessing accuracy. Note, however, that the field-referenced data used as input are collected from various projects and agencies, and thus the original source of field data cannot be considered a "random" sample of plots. Any sample of plots selected from a non-random set of points cannot be considered statistically random. Nonetheless, we have determined that withholding a limited number of points for validation purposes provides worthwhile accuracy information.

Nevertheless, we determined that it's better to produce a more accurate set of products with imperfect accuracy information than a less accurate set of products with better known accuracy estimates. We do not want to withhold plots that would best be used for model and product development. As a compromise, we recommend that two percent of the plots be withheld from the modeling activities. These plots will then be used to estimate accuracies for aggregations of LANDFIRE mapping zones or "superzones". We plan to merge data sets from three to four adjacent mapping zones and conduct validation activities for these regions. A target of at least 50 plots for each vegetation class per superzone provides useful information for estimating accuracies.

Stratification of accuracy assessment—In addition to providing general accuracy information at the superzone and individual mapping zone levels, we recommend providing more local estimates of accuracy nested within these other levels. This will be accomplished through spatial stratification of broad areas using biophysical gradient modeling information and other sources of spatial data and through thematic aggregation of similar vegetation types for localized regions. The process of stratifying mapping zones into zones based on the biophysical gradient layers developed for LANDFIRE (see Holsinger and others, Ch. 5) will be used as a basis to further our understanding of product errors, which in turn will enable refinement of future mapping procedures. This stratification process may facilitate the discrimination of different vegetation types with similar spectral signatures that occupy sites having very different environmental characteristics.

Field verification—As discussed above, we recommend conducting a modest level of field verification throughout LANDFIRE National. Field visits provide the technical teams with a basic understanding of the natural vegetation and ecology of the regions in which they are working, and field visits to particular sites serve to verify (or invalidate) the modeled results. Ideally, a field visit should take place at the beginning of each zone's mapping activities for familiarization purposes, and an additional field visit should occur near the end of the mapping process to verify and refine the mapping process.

Consistency checks with data from other sources—Whenever possible, products should be compared with existing independently produced data sets. In some cases, products unrelated to LANDFIRE have been generated for certain local areas, and these can be used to help assess accuracies of LANDFIRE products. Spatial and tabular data potentially provide good

general information. In addition, we recommend that LANDFIRE support the generation of local validation data sets, where appropriate.

Accuracy of Maps Based on Landscape Simulation Models

As discussed above, it is generally very difficult to ascertain the quantitative accuracy of products generated through complex landscape modeling efforts. Even so, there are some approaches suitable for assessing the validity of certain LANDFIRE modeled products, such as modeled historical fire regimes.

Although as of yet there are no examples of complete data sets representing historical vegetation conditions for the entire United States at the spatial grain of the LANDFIRE products, there are local historical data sets that can be used to "spot check" the validity of the products generated. For instance, historical aerial photographs and field-based data sets may provide useful information for assessing modeled historical fire regime products. Although not a true quantitative analysis, comparisons with historical data will likely provide information regarding the validity of the products.

As described above, it is important that the outputs from complex modeling activities be scrutinized carefully and checked for obvious flaws or deviations from expected results. As obvious as this seems, we are aware of numerous investigations in which this avenue has been neglected and in which spatial products were produced but not carefully examined. Although this type of evaluation does not yield quantitative error estimates, it can provide valuable insight regarding probable accuracies.

Finally, users of the LANDFIRE data sets should recognize that the inputs to the modeling process, while not always perfect, reflect the most accurate and current information available and are based upon ecologically sound assumptions. For these reasons, LANDFIRE products represent state-of-the-art modeling and technology and thus a significant improvement over other current options.

Conclusion

There is no single recommended procedure for deriving accuracy estimates for LANDFIRE products. Because time- and cost-related constraints, it will not be possible to conduct traditional accuracy assessments for the LANDFIRE mapping region (the entire U.S.). Yet at the same time, we recognize that evaluations of quality and accuracy increase the credibility of the final LANDFIRE products. Additionally, we can learn much by assessing error terms in the products, and this knowledge can be invaluable for future mapping and modeling endeavors. We suggest conducting a suite of accuracy assessment methods for LANDFIRE National, ranging from mostly qualitative assessments (such as the critical inspection of products, consultation with regional experts, and comparisons with existing data sets) to more quantitative analyses (such as cross-validation assessments, traditional accuracy assessments at the superzone level, and select evaluations at local levels). These combined approaches will provide LANDFIRE data users with the accuracy information necessary to facilitate the appropriate use of the data.

For further project information, please visit the LANDFIRE website at www.landfire.gov.

The Authors

James Vogelmann is a Principal Scientist with the Science Applications International Corporation (SAIC), contracting with the USDOI Geological Survey Center for Earth Resource Observation and Science (EROS). Vogelmann's research focuses on large-region land cover characterization and change assessment using remote sensing and ancillary sources of spatial data. His role in the LANDFIRE Prototype Project has been to assess different methods for mapping vegetation types, serve on the LANDFIRE Vegetation Working Group, and help direct project research activities. He received his B.A. degree in Botany from the University of Vermont in 1978 and his Ph.D. degree in Plant Sciences from Indiana University in 1983.

Zhiliang Zhu is a Research Physical Scientist with the USDOI Geological Survey Center for Earth Resource Observation and Science (EROS). Zhu's research focuses on mapping and characterizing large-area land and vegetation cover, studying land cover and land use change, and developing remote sensing methods for the characterization of fire fuel and burn severity. His role in the LANDFIRE Prototype Project has been to design and test a methodology for the mapping of existing vegetation cover types and vegetation structure and to direct research and problem-solving for all aspects of the methodology. He received his B.S. degree in Forestry in 1982 from the Nanjing Forestry University in China, his M.S. degree in Remote Sensing in 1985, and his Ph.D. degree in Natural Resources Management in 1989, both from the University of Michigan.

Jay Kost is a Research Physical Scientist with the Science Applications International Corporation (SAIC),

contracting with the USDOI Geological Survey Center for Earth Resource Observation and Science (EROS). Kost's work focuses primarily on mapping existing vegetation and vegetation structure (percent canopy and height) for the LANDFIRE Project using decision and regression tree models. Optimization of these models and high map accuracy results are paramount in his work and improvements in methodology and results are continually pursued. He received his B.S. in Electronic Engineering Technology in 1987 from Minnesota State-Mankato and his M.S. degree in Space Studies from the University of North Dakota, Grand Forks. In addition, Kost has completed four years of post-graduate study in the Atmospheric, Environmental, and Water Resources Ph.D. program at South Dakota State University, Brookings.

Brian Tolk is a Research Scientist with the Science Applications International Corporation (SAIC), contracting with the USDOI Geological Survey Center for Earth Resources Observation and Science (EROS). Tolk's research focuses on the mapping and characterization of large-area land and vegetation cover, as well as on the use of close-range remote sensing methods to aid and improve LANDFIRE mapping techniques. His role in LANDFIRE has been to map land cover and structure variables for the prototype zones, implement a data management scheme, and produce promotional products for the project. He received his B.A. degree in Geography from Augustana College, Sioux Falls in 1990 and his M.A. degree in Geography from the University of Nebraska, Lincoln in 1996.

Donald Ohlen is an Environmental Scientist for the Science Applications International Corporation (SAIC) at the USDOI Geological Survey Center for Earth Resource Observation and Science (EROS). Ohlen's research and interest focus on land cover mapping for fire science applications, including the characterization of satellite data for fuel mapping and post-fire burn mapping. He earned his B.S. (1976) and M.S. (2000) degrees in Geography from South Dakota State University.

Acknowledgments

We wish to acknowledge the funding provided by the USDOI Office of Wildland Fire Coordination and the USFS Fire and Aviation Management. The cooperation of the USFS Forest Inventory Analysis office in Ogden, Utah, and the help and support provided by the staff at the USFS Fire Sciences Laboratory in Missoula, Montana, are gratefully acknowledged. We also thank Terry Sohl, Sherry Chen, Matt Rollins, and Christine Frame for helpful technical and editorial review through the creation of this manuscript. The research was performed, in part, by the Science Application International Corporation under U.S. Geological Survey Contract 1434-CR-97-40274 and 03CRCN0001.

References

Congalton, R.G. 1991. A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment. 37: 35-46.

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

Gesch, D.; Oimoen, M.; Greenlee, S.; Nelson, C.; Steuck, M.; Tyler, D. 2002. The National Elevation Dataset. Photogrammetric Engineering & Remote Sensing. 68: 5-12.

Homer, C.; Huang, C.; Yang, L.; Wylie, B.; Coan, M. 2004. Development of a 2001 National land-cover database for the United States. Photogrammetric Engineering & Remote Sensing. 70: 829-840.

Huang, C.; Homer, C.G.; Yang, L. 2003. Regional forest land cover characterization using medium spatial resolution satellite data. In: Wulder, M.A.; Franklin, S.E., eds. Remote sensing of forest environments: concepts and case studies. Amsterdam, Netherlands: Kluwer Academic Publishers. Pp. 389-410.

Keane, R.E.; Burgan, R.; van Wagtendonk, J.V. 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. International Journal of Wildland Fire. 10: 301-319.

Lunetta, R.S.; Lyon, J.G. 2004. Remote Sensing and GIS Accuracy Assessment. Boca Raton, FL: CRC Press. 320 p.

Quinlan, J.R. 1993. C4.5: programs for machine learning. San Mateo, CA: Morgan Kaufman Publishers. 302 p.

Rollins, M.G.; Keane, R.E.; Parsons, R.A. 2004. Mapping Fuels and Fire Regimes Using Remote Sensing, Ecosystem Simulation, and Gradient Modeling. Ecological Applications. 14: 75-95.

Stehman, S.V.; Zhu, Z.; Czaplewski, R.L.; Nusser, S.M.; Yang, L. 2000. Combining accuracy assessment of land-cover maps with environmental monitoring programs, Environmental Monitoring and Assessment. 64: 115-126.

Stehman, S.V.; Wickham, J.D.; Smith, J.H.; Yang, L. 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.

USGS EDNA Website, [Online]. (2004). Available: http://edna.usgs.gov

USGS Landsat Website, [Online]. (2004). Available: http://landsat.usgs.gov

Vogelmann, J.E.; Sohl, T.; Howard, S.M. 1998. Regional characterization of land cover using multiple sources of data. Photogrammetric Engineering & Remote Sensing. 64: 45-57.

Vogelmann, J.E.; Howard, S.M.; Yang, L.; Larson, C.R.; Wylie, B.K.; Van Driel, N. 2001. Completion of the 1990's National Land Cover Data set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogrammetric Engineering & Remote Sensing. 67: 650-662.

Wickham, J.D.; Yang, L.; Stehman, S.V.; Smith, J.H. 2004. Thematic accuracy of the 1992 National Land-Cover Data for the western United States. Remote Sensing of Environment. 91: 452-468.

Yang, L.; Stehman, S.V.; Smith, J.H.; Wickham, J.D. 2001. Thematic accuracy of MRLC land cover for the eastern United States. Remote Sensing of Environment 76: 418-422.