Summary of Data Related Issues as they Affect
The JFSP IFT-DSS Development and Deployment

H. Michael Rauscher
30 October 2008

The JFSP Software Tools and Systems (STS) study has identified the fuels treatment analysis and planning process as the most important application area to test the design and implementation of the Interagency Fuels Treatment Decision Support System (IFT-DSS), an innovative collaborative system architecture approach to DSS. During the development of the conceptual design document, it became apparent that there exist 4 more-or-less comprehensive systems that address the fuels treatment analysis and planning process. They are ArcFuels, INFORMS, NIFTT fuels treatment, and Starfire. In addition, a direct survey of field fuels treatment specialists resulted in the recognition that the majority of respondents used their own ad-hoc process rather than one of the 4 comprehensive systems.

Almost every direct contact with field fuels treatment specialists as well as discussions with developers of fuels treatment systems brought up the fact that data issues presented enormous challenges to the deployment of software support systems. It became obvious that the JFSP Fuels Working group had to find out what was going on in the data arena if the proposed IFT-DSS was to be successful.

Note: this summary of data related issues as they affect the JFSP IFT-DSS development and deployment project is intended ONLY a BEGINNING scoping of the situation and should by no means be regarded as a fully comprehensive analysis. This summary was produced with limited resources in people and time on a volunteer basis. In the opinion of its authors, the main use of this summary should be to motivate discussion and perhaps a more formal analysis of the situation leading to well reasoned and supported suggestions on how to proceed in the future.

The NIFTT fuels treatment planning process and Starfire use the LANDFIRE database layers (www.landfire.gov). These database layers are available to anyone for use and they provide wall-to-wall coverage in the lower 48 United States, forest and non-forest. The LANDFIRE project has mapped FLM and FCCS fuelbeds for the Western US and is in the process of mapping the Eastern US . Additional FLM and FCCS fuelbeds may be developed through research and may be mapped by LANDFIRE during Operations and Maintenance (LFOM). Tree-lists in FVS, FUELCALC, and FOFEM format have been developed by NIFTT in parallel with LANDFIRE National and LFOM. In 2009, LANDFIRE coverage is expected to be completed for Alaska and Hawaii. The data issues of interest concerning LANDFIRE data are (1) how to assess suitability for a particular project level analysis and match the correct questions to those that the data accuracy is able to address; data is intended for broad- and mid-scale analyses for fire and fuel related issues; with local evaluataion and editing data can be used for fine-scale analyses on fire incidents and for project fuels planning. Accuracy is expected to be less than optimum for local-scale analyses without this local evaluation and editing (Ohmann et al. below) (2) LANDFIRE National data currently available for download represent circa 2000.

LFOM was kicked off with the Rapid Refresh in the Western US, for which data is now available for download that includes enhancements for vegetation and fuel layers with updates for wildland fires through 2007. The full LANDFIRE Refresh component of LFOM has been started, which will enhance and update all vegetation and fuel layers for treatments and disturbances to circa 2008. Data for the Southeast and Northwest will become available in spring and summer of 2009 followed by staged completion across the rest of the US by September of 2010. In addition to the updated layers the state and transition models for predicting future changes across forests and non-forest will be released. The biennial and decadal components of LFOM will start in 2011 with a focus on updates of post-2008 change areas every 2 years using new techniques in change detection combined with Refresh tools and decadal updates based on new remote sensing. Release of update and editing tools and guidelines for local evaluation and editing will parallel these various components of LFOM. This may or may not be sufficient for current year analyses or for analyses of small project areas; (3) For local planning LANDFIRE data should be revised and calibrated based on available local data (NIFTT offers processes for doing such updating or editing). LANDFIRE data offers easy access and the tools developed to perform a fuels treatment analysis, along with training to learn how to use them, are available.

ArcFuels has been updated to be able to use LANDFIRE grid data to conduct a fuels treatment analysis and planning process with largely the same functionality as the NIFTT fuels treatment planning process (Ager, Pers. Comm..)

ArcFuels and INFORMS use treelist data to provide vegetation data input to a fuels treatment analysis. They offer a lot of analysis power for project level planning as well as support the landscape level analysis to place treatments into the appropriate context for their effects to be evaluated. The trouble is that wall-to-wall FVS treelist data rarely exist for a particular project area. Various data imputation methods, reviewed by Ohmann et al. below, have been developed and tested that support the generation of wall-to-wall FVS treelist data. These imputation methods work across ownership boundaries as long as representative data is available for the entire range of vegetation units encountered and as long as the vegetation units in the analysis area are forests. FVS cannot simulate dynamics for non-forested vegetation units. The only vegetation dynamics simulator that works on non-forested lands is PHYGROW, a part of the INFORMS toolkit. The use of PHYGROW as of 2008 is still in the experimentation and testing stage. Australia, New Zealand, and other locations have developed indices to track grassland conditions with regards to fire that might be explored for use in the US.

To do imputation, requires high quality, field plot based data that is available to resource specialists regardless of employing agency. Rauscher et al (see below) conducted a review and summary of the data source availability issues. As of 2008, most existing data sources restrict access to only those fuels treatment specialists working for the agency owner. The only exceptions are the LANDFIRE data and the GNN FVS treelist data in the Pacific Northwest. A national-level imputation pilot study (NaFIS) could provide the technical basis for developing nationwide data of this kind, but it may not be sufficiently reliable for local-scale analyses. Despite plans for making data sources widely available, none of the agencies have so far accomplished this. This means that USDA FS employees are able to access FSVeg data but nothing else. NPS employees can access DataStore databases in FFI format but nothing else.

BLM employees can access their data in FIREMON/FFI formats but not the NPS data in FFI format. You get the idea. The bottom line is that fuels treatment specialists in many cases cannot use the best available, ground-based data for their project analyses.

Let me recap the situation by highlighting an excerpt from the Ohmann et al. paper:

Sources of tree list data currently available to various users

Data for all ownerships, nationwide (mid- to national-scale): LANDFIRE data, available to all. Useful for many analytical purposes but intended for large geographic extents (regional to national) if not evaluated and edited, and for fire- and fuel-related issues. Accuracy is expected to be less than optimum for local-scale analyses unless data is locally evaluated and edited. Presumably these data will continue to be supported under the ongoing LANDFIRE program.

Data for all ownerships in the Pacific Northwest (mid-scale): GNN data, available to all. Maps contain more forest attributes than LANDFIRE but accuracy varies among attributes. GNN mapping of fuels-related variables has been explored (Pierce et al., in review), but is not part of the current implementation. Developed primarily for mid-scale analysis; accuracy at the local scale may be insufficient for local management decisions. New analytical technologies are being developed to take advantage of these data where available. The GNN project is a research effort with no long-term home or support. Ongoing updating and maintenance of GNN datasets is not within the research mission. A long-term plan to maintain and support this kind of data on existing vegetation is needed for the region and possibly beyond. The NaFIS project may provide direction for national implementation of nearest neighbors methods, but the expanded scope may result in less reliable data at the regional (mid-) scale.

Data for Forest Service lands, in FSVeg (local- to mid-scale): Tree list data available for polygons on Forest Service lands where stand exam data have been stored in the National Field Sampled Vegetation Database (FSVeg), and where data have been extracted for use with FVS. Available to Forest Service employees only. INFORMS with stand exams in FSVeg allows the user to do their own imputation, where the user also provides the necessary GIS (polygon) layer and other related datasets in addition to stand exam data of sufficient quality. The current version of INFORMS includes imputation technology that can be run locally for a project or Forest-wide with sub-projects. This application is intended for local-scale data and analyses, but there are no accuracy assessment tools in the current version. Users are trained to field-verify results and use internal statistics to evaluate the overall quality of each analysis. Accuracy assessment methods are under development. The FSVeg database has long term support, but currently is available only to users within the Forest Service.

Data for other Agency lands and other ownerships in FFI (FEAT/FIREMON Integrated): The FFI monitoring tool assists managers with collection, storage, and analysis of ecological information, and includes tree list data similar to that in the above FSVeg databases for

Forest Service lands. (In some cases Forest Service sites are included in FIREMON.) There currently are no imputation tools linked directly to these data (although GNN [and LANDFIRE???] is using some FIREMON plots). Employees of other agencies in general have no imputed landscape data available to them other than GNN and LANDFIRE as described above. Efforts are underway by INFORMS and other groups to include these data.

Please refer to the two summary papers below for a good bit more in depth analysis and understanding of the data related issues.

So what needs to happen?

The JFSP IFT-DSS project will use the collaborative system architecture design approach to make the 4 existing comprehensive fuels treatment planning systems and the most common ad-hoc approaches (along with all their supporting subsystems) available to fuels specialists in an organized, understandable, and useful way. For the IFT-DSS to be truly useful, the fuels specialists of all agencies must have easy access to all available data for a project area regardless of who "owns" the data. They must have the necessary support tools that can gather the available data based on a simple landscape identification method to define project and analysis boundaries, data mining software needs to automatically reformat data from various sources into the needed analysis standard, tools that help the users understand the different degrees of accuracy need to be available, and finally, powerful data visualization and analysis tools need to be assembled so that the user can convince themselves as well as stakeholders of the appropriateness of the input data layers that form the foundation of any effective fuels treatment analysis.

This brief summary of the state of data imputation and data sources is not adequate to answer HOW we need to proceed to achieve the data related vision stated above. It must be regarded as only a beginning evaluation and discovery of what currently exists. It appears that momentum is building from many directions, not just fire and fuels, for the development of a national tree-list dataset based on one or another of the many variants of nearest neighbor analysis. We are not talking about a huge investment in gathering new field data. The FIA plots records as well as other existing datasets are sufficient to impute mid-scale, national data treelist data sets. The richness and complexity of nearest neighbor maps comes with the added burden of user education. In fact, users of all kinds of fire and fuels related data badly need training in the appropriate use and interpretation of the various available data sets. It is noteworthy that the NIFTT fuels treatment planning and analysis process does an admirable job of training fuels specialists in the application of that type of data. Data availability and quality for non-forest lands is far below that of forested lands. It is noteworthy that other countries in Europe and Australia seem further advanced in grassland fire analysis and planning than we are here in the US. We are not convinced that it will take a huge amount of investment to improve the data situation. It is likely to be more a matter of making current data available to everyone. Investment in software improvements needs to go hand-in-hand with investments in data management.

In a late breaking development, we have recently been made aware that Dr. Karen Short of the

LANDFIRE project (kshort@landfire.org) has specialized in accessing available field inventory records for the purpose of imputing wall-to-wall vegetation, treelists, and other variables for use by Interagency Resource Specialists. We did not have time to tap into her expertise and suggest Dr. Short be a key player in any further summary efforts on this topic area.

We recommend that the JFSP and the NWCG National Interagency Fuels Treatment Coordinating Group (NIFCG) combine forces to organize and fund a special project over the next year that will examine these data issues in more depth and develop a credible and practical improvement pathway for the future. Ideally, a centralized storage process that includes FSVeg, FFI, FMA and other vegetation unit scale data sources needs to be crafted with open access to everyone, including the public, with a user friendly web interface.

Submitted for consideration by:

H. Michael Rauscher
JFSP Project Manager for the STS Study

Nearest-neighbors mapping of vegetation and 'tree lists' at landscape scales in the US

by Janet L. Ohmann
with contributions from Eric Twombly, Bob Keane, Nick Crookston, and Alan Ager
30 September 2008

The need for multi-attribute vegetation and 'tree-list' maps for large landscapes

Maps of existing vegetation and land cover are needed at a range of spatial extents: from the local stand- or project-level, to support operational decisions; to the mid-scale, defined as large landscapes, watersheds, or regions that usually span multiple land ownerships, to support regional assessments, strategic planning, and policy analysis; to national, continental, or even global scales. The specific needs for spatial vegetation information, in terms of vegetation attributes, spatial resolution, reliability, and currency (how up-to-date) vary with geographic extent and objectives. This summary focuses on meeting information needs at the *mid-scale*, typically areas of 10,000 to 25,000 acres, and broader. We approach the problem with a philosophy of developing data at the most detailed level that is practical (basic vegetation attributes, finest spatial resolution), with the notion that data can be aggregated, generalized, or summarized to meet a variety of needs and possibly across a range of scales. This approach affords the greatest analytical flexibility and cost-effectiveness.

At landscape to regional scales, forest managers, policymakers, and researchers increasingly desire spatially explicit, wall-to-wall, digital maps for a large array of forest attributes. Such data are needed to support a variety of applications including assessments and scenario modeling (e.g., fire, insect, pathogens, wildlife habitat) to ecosystem modeling (e.g., carbon sources/sinks, climate change, and ecosystem services). Many applications require digital maps where each map unit is attributed with a 'tree list,' defined as the tally of individual-tree-level data typically recorded on a forest field plot (species, diameter, height, live crown, and density). Tree lists can be input directly to stand projection models such as the Forest Vegetation Simulator (FVS) (Crookston and Dixon 2005, Dixon 2002, Reinhardt and Crookston 2003), and can be used to derive many attributes of forest structure and composition relevant to fuels, wildlife habitat, timber, and other forest values.

Nearest neighbors is a relatively new family of methods that is gaining in popularity, due largely to capability to provide maps of multiple forest attributes with associated measures of reliability. Some applications of nearest neighbors methods can provide tree-list maps. This paper briefly describes nearest neighbor methods, and summarizes current projects at mid-scale and broader in the US. We emphasize forest lands, only because that is where most efforts have concentrated, primarily due to the lack of regionally consistent field data for nonforested areas.

A short primer on nearest neighbors mapping

Nearest neighbors methods are used to develop estimates for unsampled areas (*target* map units) by relying on the relationship between sampled areas (*reference* dataset) and spatially comprehensive, correlated data from auxiliary sources such as remotely sensed imagery, forest type maps, physiographic data, climate models, or other relevant GIS layers (*predictor* variables). Several variants of nearest neighbors mapping are currently being used, including *k* Nearest Neighbors (*k*NN) (Tomppo 1990), Most Similar Neighbor (MSN) (Moeur and Stage 1995), and Gradient Nearest Neighbor (GNN) (Ohmann and Gregory 2002). The methods differ in how search space and nearness (distance) between target and reference points is assessed, how many neighbors (k) are selected, and how they are weighted when $k > 1$.

Actual applications of the methods also can vary in terms of the reference, predictor, and target data used. The reference data, for which complete vegetation data are available, typically are either field plots or stand exams. The target set (map units) can be pixels in a raster (grid) of any spatial resolution, or polygons (e.g., forest stands) of any size. Lastly, nearest neighbors methods can use one or many nearest-neighbor plots or stands (values of k) in the imputation process. Applications where tree lists are imputed to pixels or stands typically have used a single plot or stand ($k = 1$) (Temesgen et al. 2003). In this case, the covariance structure of trees and derived attributes within the stand or plot is maintained in the target map unit, which can be advantageous for subsequent analyses. Several techniques are available for diagnosing whether a given application of nearest neighbors has yielded results that are satisfactory for a particular purpose (McRoberts et al. 2007, Stage and Crookston 2007).

Evaluating map quality

Map accuracy can be assessed at the local (plot or stand) or regional scale (across the map area as a whole). Local-scale accuracy traditionally is evaluated using cross-validation, by comparing paired predicted (map) and observed (plot or stand) values for a subsample of the reference plots or stands that were excluded from model development. Diagnostics may include measures of precision (e.g., root mean square error, kappa statistic) and bias. Two-way error matrices, or confusion matrices, often are constructed for vegetation classes or variables of interest.

At the regional scale, distributions of map area can be constructed by summing map pixels or stands for vegetation classes or for intervals of continuous variables. For validation, the map distributions are compared against independent estimates for the region, such as from design-based inventories by Forest Inventory and Analysis. Regional-scale accuracy assessments are less commonly conducted, yet provide information on whether the full range of variability of vegetation is represented in a map, which may be more important to a landscape-level application than local accuracy. Several other map diagnostics are possible, but are beyond the scope of this paper.

Evaluation of whether the reliability of a particular map is satisfactory or 'good enough' is highly dependent on the objectives and scale of the application. Map products may be unreliable at the local stand or pixel scale while providing excellent representation at the landscape or regional level. Conversely, map quality may be quite good for a local area where supporting data are abundant and up-to-date, but map coverage may be inconsistent or biased when viewed across a broader, multi-ownership landscape. Reliability also may vary greatly among vegetation attributes. Because accuracy assessment methods differ widely among map products, caution is needed when comparing maps and accuracy assessments.

Quirks and caveats for use of nearest neighbors maps and tree lists

Much of the attractiveness of nearest neighbors methods is that they are multivariate and non-parametric, allowing simultaneous prediction of more than one variable. They also are uniquely suited to providing maps attributed with 'tree lists.' However, nearest neighbors models can be 'tuned' to emphasize one or more variables over others, which can strongly influence neighbor selection and the relative accuracies of variables in the resulting maps. As a general rule, univariate methods that focus on a single vegetation attribute tend to provide better local-scale map accuracy for that attribute compared to multivariate methods, which by their nature arrive at a 'compromise solution' across many variables. However, high local accuracy often comes at the expense of regional-scale accuracy, in the form of loss of range-of-variability (loss of the highest and lowest values) across the map as a whole. Furthermore, layering several single-attribute maps together, even if each one individually is highly reliable, may result in unrealistic combinations for specific map locations. Although nearest neighbor imputation may result in lower prediction accuracy for any single variable when compared to other methods (although this is not always the case), the maps may better represent regional distributions. If a single nearest-neighbor plot is imputed to each map unit, the covariance of vegetation attributes is maintained.

It's important to note that the reliability of nearest neighbors maps may be more a function of data quality used in map development than of the particular mapping method used. Characteristics of both reference and predictor data influence the outcome of nearest neighbors analyses. Elements of reference data quality, whether stand exams or plots, include the sampling intensity (number of observations), representativeness of conditions within the mapping area, timeliness, temporal match to imagery or other predictors, within-plot or -stand sampling error, and completeness of the vegetation measurements (which vegetation components are tallied).

Most of the nearest neighbors maps currently available fall into two categories: polygon maps constructed from a stand map and stand exams, and raster (grid) maps constructed using plots and satellite imagery. Debates over the merits of these two approaches are best focused on the underlying data quality and on suitability of the map relative to the scale of the application (e.g., operational treatment decisions vs. regional strategic planning). These two map types also differ greatly in terms of their spatial patterning or 'look-and-feel,' determined by interactions among the spatial resolution and pattern of the target map units, and characteristics of the reference and predictor data. Although spatial configuration can be quantified using various metrics, there are no standard accepted measures of ‘accuracy,’ and map choice is more a matter of subjective preference and practical considerations regarding the map application. For example, land managers typically work with stand maps, whereas ecosystem modelers operate with grids.

Several caveats for use of nearest neighbors maps and tree lists apply equally to any of the methods or products discussed in this paper. As a general rule, these tree list maps are expected to be reliable for analysis at the landscape scale; local applications should be avoided or undertaken with extreme caution. As discussed above, relative accuracies of individual map attributes will vary, depending on particulars of the nearest neighbors analysis. For example, tree list maps constructed to emphasis fuels (or any other single use) may provide better accuracy for this purpose than alternative maps, but also may be less suited to other analyses. The tree-list data imputed to map units are dependent on the sample of trees included in the source plot data. For example, FIA plots do not provide an adequate sample of seedlings, so they are not included in imputed tree lists, which may impact derived canopy fuels variables.

Most inventory programs come from a traditional focus on timber resources and therefore forest lands and live trees. Over recent decades, inventories have expanded to more completely sample all vegetation, including snags, large down wood, understory vegetation, and in some cases surface fuels. However, the population characteristics and sampling properties for these vegetation components are not well understood, particularly in the context of nearest neighbors imputation of tree lists, and within-plot sampling error can be quite high.. Dead wood and understory vegetation are notoriously variable in time in space, with patterns strongly influenced by disturbance history and poorly correlated with overstory conditions. Although some nearest neighbors maps include attributes of these other vegetation components, accuracy assessment is problematic. Furthermore, many nearest neighbor models rely on affordable satellite imagery such as Landsat, which is not particularly sensitive to understory conditions.

Fuels mapping at regional scales is particularly challenging – see Keane et al. (2001) and Pierce et al. (in review) for detailed discussions. Characteristics of the tree canopy, including derived canopy fuels variables, may be mapped with acceptable accuracy for many applications. However, mapping of surface fuels based on the reference and predictor data that currently are widely available and affordable is much more difficult. Mapping of fuel models is particularly challenging. Tools available in FVS-FFE for generating fuel models for use in fire simulations does not always work well in regional applications. Furthermore, most current applications of nearest neighbors methods are confined to forest land. Areas of nonforest are mapped using ancillary data sources such as the National Land Cover Data which contain their own errors and biases. For fuels and fire applications, reliable depiction of burnable nonforest (e.g., grasslands and shrublands) and non-burnable nonforest (non- or sparsely-vegetated) is a critical need.

Current broad-scale nearest neighbors mapping projects in the US

A number of groups in the US and internationally have developed and applied various nearest neighbor methods, and research in this area is active and ongoing. This paper briefly describes several projects underway in the US that are relevant to the goal of providing tree list data for large landscapes and broader. There are many local projects by researchers or land managers, but they are beyond the scope of this paper.

<u>yaImpute.</u> yaImpute is a tool that can be used for developing tree list maps for an area of

interest, rather than a map product *per se*. yaImpute is a statistical package (Crookston and Finley 2008), written in R, that performs several popular nearest neighbor routines including kNN, MSN, GNN, and a novel nearest neighbor distance metric based on the random forest proximity matrix (Breiman 2001). The yaImpute user can define the search space, subsequent distance calculation, and imputation rules for a given number of nearest neighbors. The package offers a suite of diagnostics for comparing results and a set of functions for mapping results.

Contact: Nick Crookston (*ncrookston@fs.fed.us*)

Websites: *http://www.jstatsoft.org/v23/i10/*, *http://cran.r-project.org/web/packages/yaImpute/index.html*

INFORMS. INFORMS (Integrated Forest Resource Management System) is another tool that can be used for developing tree list maps. INFORMS is decision support software developed for the USDA Forest Service. The current version includes an MSN tool set and links to vegetation data (stand exam polygons) in the USDA FS corporate database (FSVeg) and to FVS. INFORMS is being updated to accommodate several alternative imputation methods by integrating the yaImpute software, to utilize raster (grid) vegetation data in addition to polygons, and to accommodate nonforest within a landscape. A key component of INFORMS will be a way to sample or otherwise scale-up pixel-level imputation data (e.g., from GNN grids, see below) to map polygons, to create tree lists for input to FVS. Other practical issues to be addressed are a mechanism for updating out-of-date plots or stand exams used in imputation, and issues related to access to corporate databases and proprietary plot locations by those outside the USDA FS.

Contact: Eric Twombly (*etwombly@fs.fed.us*)

Website: *http://www.fs.fed.us/informs/index.php*

The Nationwide Forest Imputation Study (NaFIS). This study is evaluating alternative nearest neighbor techniques with the goal of recommending an approach for nationwide implementation. The vision for a national nearest neighbor application is to rely on FIA and plots as the primary reference data for forest land, and Landsat as the primary remotely sensed data. Other regional plot datasets may be considered as regional options (e.g., Current Vegetation Survey (CVS) in the Pacific Northwest, FIREMON plots). The study is investigating nearest neighbor methods through a pilot study focused on seven large mapping zones across the US that vary in terms of ecological conditions and availability of FIA data. The analyses are evaluating efficient nearest neighbor algorithms, variance estimators and other diagnostics, and data processing techniques for broad-scale mapping. Spatial data products will depict a national core set of forest variables at moderate spatial resolution (30-m pixels). Only a subset of the various nearest neighbors methods will result in tree-list maps. Lessons learned from the pilot study will provide operational guidance for efficient implementation nationwide. Implications of findings for various applications may be explored in a follow-on study. NaFIS partners are from the USDA Forest Service (Western and Eastern Wildlands Environmental Threats Assessment Centers, Forest Health Technology Enterprise Team, FIA, PNW and Northern Research Stations, Remote Sensing Applications Center), Michigan State University, and Oregon State University.

Contacts: Janet Ohmann (western US) (*johmann@fs.fed.us*), Andrew Finley (eastern US) (*finleya@msu.edu*)

Website: *http://blue.for.msu.edu/NAFIS/*

Gradient Nearest Neighbor (GNN) mapping in the Pacific Coast States. The GNN variation of imputation mapping was developed to support analysis of forest policy effects on large, multi-ownership landscapes. Tree-list data are needed to support stand and landscape projection systems and response models for wildlife, timber, and other values. As currently implemented, GNN uses gradient modeling to impute a single plot (with tree list) to each pixel in a raster map. Reference data are from regional plot datasets (FIA, CVS, FIREMON, etc.). Predictors are derived from Landsat imagery, climate models, digital elevation models, soils, disturbance maps, and other spatial data. The GNN maps are rasters at 30-m resolution with multiple joined attributes describing live trees, snags and down wood, and understory vegetation. Several map diagnostics are provided for both local and regional scales, and for all individual vegetation variables. Prediction accuracy varies widely among vegetation attributes, and users are expected to evaluate the sufficiency of the map data for their applications.

GNN data are being developed for all of Washington and Oregon and much of California for use in many applications, including the Interagency Mapping and Assessment Project (IMAP), National Forest Planning in Region 6, BLM cumulative effects analysis, Effectiveness Monitoring for the Northwest Forest Plan, strategic planning by state agencies and non-governmental organizations, and many research studies. The GNN method has been evaluated specifically for mapping fuels (Pierce et al., in review; Wimberly et al. 2003). The GNN grids are being linked to ArcFuel's method of using a few ideotypic tree lists to run various fuels treatment scenarios and to develop "correction factors" for LANDFIRE data values. An interactive landscape visualization system based on computer gaming technology is available for GNN and other tree list maps.

Contacts: Janet Ohmann (GNN) (*johmann@fs.fed.us*), Alan Ager (ArcFuels) (*aager@fs.fed.us*)

Websites: *http://www.fsl.orst.edu/lemma* (GNN), *http://www.fs.fed.us/wwetac/arcfuels/index.html* (ArcFuels)

A spatially explicit tree-list for the US linked to LANDFIRE data products. The LANDFIRE group is developing methods to summarize FIA plot data to create a tree list for every combination of LANDFIRE’s existing vegetation type, biophysical setting, successional class, and canopy bulk density (Herynk and Drury, in prep.). Although the tree lists can be input to FVS, they were developed primarily to input to FOFEM, FUELCALC, and other fire-related programs to aid in spatial analysis of fuel treatments and fire effects. Because of the emphasis on predicting fire-related tree mortality, selection of reference plots is based on bark thickness, and the tree lists were not found to predict basal area and tree density very well. This implies that other approaches to building tree lists might be needed for each management or analysis objective. This approach is viewed by LANDFIRE as a stop-gap measure to create a LANDFIRE tree list for the Rapid Refresh and Fire Severity mapping project. Other approaches for developing tree-list maps may be better in the long term, but may require more work to prepare wall-to-wall US layers needed by LANDFIRE. The same caveats and limitations described above for all nearest neighbors products apply to the LANDFIRE data.

Contact: Bob Keane (*rkeane@fs.fed.us*)

Website: *http://www.landfire.gov/*

Sources of tree list data currently available to various users

Data for all ownerships, nationwide (mid- to national-scale): LANDFIRE data, available to all. Useful for many analytical purposes but intended for large geographic extents (regional to national), and for fire-related issues. Accuracy is expected to be less than optimum for local-scale analyses. Presumably these data will continue to be supported under the ongoing LANDFIRE program.

Data for all ownerships in the Pacific Northwest (mid-scale): GNN data, available to all. Maps contain more forest attributes than LANDFIRE but accuracy varies among attributes. GNN mapping of fuels-related variables has been explored (Pierce et al., in review), but is not part of the current implementation. Developed primarily for mid-scale analysis; accuracy at the local scale may be insufficient for local management decisions. New analytical technologies are being developed to take advantage of these data where available. The GNN project is a research effort with no long-term home or support. Ongoing updating and maintenance of GNN datasets is not within the research mission. A long-term plan to maintain and support this kind of data on existing vegetation is needed for the region and possibly beyond. The NaFIS project may provide direction for national implementation of nearest neighbors methods, but the expanded scope may result in less reliable data at the regional (mid-) scale.

Data for Forest Service lands, in FSVeg (local- to mid-scale): Tree list data available for polygons on Forest Service lands where stand exam data have been stored in the National Field Sampled Vegetation Database (FSVeg), and where data have been extracted for use with FVS. Available to Forest Service employees only. INFORMS with stand exams in FSVeg allows the user to do their own imputation, where the user also provides the necessary GIS (polygon) layer and other related datasets in addition to stand exam data of sufficient quality. The current version of INFORMS includes imputation technology that can be run locally for a project or Forest-wide with sub-projects. This application is intended for local-scale data and analyses, but there are no accuracy assessment tools in the current version. Users are trained to field-verify results and use internal statistics to evaluate the overall quality of each analysis. Accuracy assessment methods are under development. The FSVeg database has long term support, but currently is available only to users within the Forest Service.

Data for other Agency lands and other ownerships in FFI (FEAT/FIREMON Integrated): The FFI monitoring tool assists managers with collection, storage, and analysis of ecological information, and includes tree list data similar to that in the above FSVeg databases for Forest Service lands. (In some cases Forest Service sites are included in FIREMON.) There currently are no imputation tools linked directly to these data (although GNN [and LANDFIRE???] is using some FIREMON plots). Employees of other agencies in general have no imputed landscape data available to them other than GNN and LANDFIRE as described above. Efforts are underway by INFORMS and other groups to include these data.

Where to from here?

Many existing applications of nearest neighbors methods at broad spatial extents are *ad hoc* efforts that have coalesced around particular information needs and funding. These projects have been led by the research community, often with partners in land management. The widespread recognition of the value and flexibility of imputation-based maps for meeting a variety of research and management needs has led to discussions at many levels (e.g., IMAP in the Pacific Northwest, NaFIS at the national level) on how to ensure continued availability of this kind of data. Specifically, institutional structures (people and funding) are needed that will ensure the continued availability of up-to-date maps of existing vegetation and land cover based on the best available technology and data.

Such an endeavor will be most effective if it involves partners in both research and management, to keep current with evolving technology while assuring the relevance of products. There also are compelling advantages to an interagency approach, for cost efficiency and to minimize proliferation of contradicting vegetation datasets. Furthermore, it would be important to avoid allowing a single resource or issue to dominate the development of broad-scale imputation datasets, in order to maximize product utility for a variety of uses, to facilitate integrated analyses of multiple values, and to foster 'ownership' and investment in the process by many groups. Many jurisdictional and institutional challenges will need to be overcome to make this happen, but the payoff would be large.

Data needed to support local-scale analyses also are lacking in many -- if not most -- locations. Minimum data requirements need to be defined, particularly if data other than FIA plots are to be used, and in regards to developing information for nonforested landscapes. Improving technology in remote sensing, in particular LiDAR, may allow map accuracy based on extensive datasets like FIA to be improved to a level that is acceptable for local analyses.

References

Breiman, L. 2001. Random Forests. Machine Learning 45:5-32.

Crookston, NL, and AO Finley. 2008. yaImpute: an R package for *k*NN imputation. Journal of Statistical Software 23:1-16.

Crookston, NL, and GE Dixon. 2005. The forest vegetation simulator: A review of structure, content, and applications. Computers and Electronics in Agriculture 49:60-80.

Dixon, G.E. (comp.). 2002. Essential FVS: A user's guide to the Forest Vegetation Simulator. Internal Report. USDA Forest Service, Forest Management Service Center, Fort Collins, CO; 189 pp.

Herynk, J, and S Drury. In prep. A spatially explicit tree list for the United States. Gen. Tech. Rep. Rocky Mountain Research Station, USDA Forest Service.

Keane, RE, R Burgan, and J Van Wagtendonk. 2001. Mapping wildland fuels for fire management across multiple scales: integrating remote sensing, GIS, and biophysical modeling. Int. J. of Wild. Fire 10:301-319.

McRoberts, RE, EO Tomppo, AO Finley, and J Heikkinen. 2007. Estimating areal means and variances of forest attributes using the k-Nearest Neighbors technique and satellite imagery. Remote Sensing of Environment 111:466-480.

Moeur, M, and AR Stage. 1995. Most similar neighbor: an improved sampling inference procedure for natural resource planning. For. Sci. 41: 337–359.

Ohmann, JL, and MJ Gregory. 2002. Predictive mapping of forest composition and structure with direct gradient analysis and nearest neighbor imputation in coastal Oregon, USA. Canadian Journal of Forest Research 32:725-741.

Pierce, KB, JL Ohmann, MC Wimberly, MJ Gregory, and JS Fried. In review. Imputation mapping of wildland fuels and forest structure for management and simulation modeling.

Canadian Journal of Forest Research.
Reinhardt, ED, and NL Crookston. 2003. The fire and fuels extension to the forest vegetation simulator. General Technical Report RMRS-GTR-116. USDA Forest Service Rocky Mountain Research Station, Fort Collins, CO.
Stage, AR, and NL Crookston. 2007. Partitioning error components for accuracy-assessment of near-neighbor methods of imputation. Forest Science 53:62-72.
Temesgen, H, VM LeMay, KL Froese, and PL Marshall. 2003. Imputing tree-lists from aerial attributes for complex stands of south-eastern British Columbia. Forest Ecology and Management 177:277-285.
Tomppo, E. 1990. Designing a satellite image-aided national forest survey in Finland. In SNS/IUFRO Workshop on the Usability of Remote Sensing for Forest Inventory and Planning, p. 43-47. Umea, Sweden.
Wimberly, MC, JL Ohmann, KB Pierce, MJ Gregory, and JS Fried. 2003. A multivariate approach to mapping forest vegetation and fuels using GIS databases, satellite imagery, and forest inventory plots. Proceedings of the Second International Wildland Fire Ecology and Fire Management Congress. Orlando, FL; 16-20 November 2003. American Meteorological Society. On the web (*http://www.ams.confex.com/ams/FIRE2003/techprogram/paper_65758.htm*) and CD-ROM.

Accessing Available Field Inventory Records for the Purpose of Imputing Wall-to-Wall Vegetation, Treelists and Other Variables for use by Interagency Resource Specialists

October 7, 2008

Mike Rauscher, Eric Twombly, Duncan Lutes, Tessa Nicolet, Nate Benson

Introduction

“At landscape to regional scales, forest managers, policymakers, and researchers increasingly desire spatially explicit, wall-to-wall, digital maps for a large array of forest attributes. Such data are needed to support a variety of applications including assessments and scenario modeling (e.g., fire, insect, pathogens, wildlife habitat) to ecosystem modeling (e.g., carbon sources/sinks, climate change, and ecosystem services). Many applications require digital maps where each map unit is attributed with the variables of interest. For example, a ‘tree list,’ defined as the tally of individual-tree-level data typically recorded on a forest field plot (species, diameter, height, live crown, and density) can be input directly to stand projection models such as the Forest Vegetation Simulator (FVS) (Dixon 2002, Reinhardt and Crookston 2003), and can be used to derive many attributes of forest structure and composition relevant to fuels, wildlife habitat, timber, and other forest values (Ohmann et al. 2008, in review).”

Ohmann et al. (2008, in review) have identified and summarized all of the ongoing FVS treelist imputation projects in the United States and provided a brief introduction to the various tools that have been developed to support imputation. As Ohmann noted above, FVS treelists are useful for many resource analyses and planning problems including fuels treatment analysis and planning. Within the fuels treatment analysis domain, two comprehensive fuels treatment planning software packages, INFORMS and ArcFuels, need FVS treelist data as a primary input.

Through the efforts of the LANDFIRE project, there currently exists wall-to-wall vegetation data in the lower 48 states. LANDFIRE data is projected to become available for Alaska and Hawaii in 2009. FCCS is also a source of fuels information. There is an existing 1-km grid map for the entire United States. NIFTT is also developing an online training package to teach people about FCCS. FCCS fuelbeds are being developed in parallel with LANDFIRE releases, with current coverage of the western U.S. at 30-m grids. At the end of 2009, FCCS fuels information will also be available for the eastern U.S., Alaska, and Hawaii. The scale and accuracy of LANDFIRE data is not always the best for project level fuels treatment analysis. For many fuels treatment project analyses, FVS-ready treelist data is preferred and few, if any, new fuels treatment planning analyses have a wall-to-wall treelist vegetation data layer available for the subject landscape and project area. More commonly, some subsample of the vegetation units within a landscape will have field plot records available, others will have none. In addition, the available field plot records will have been sampled over some time period before a project is initiated, so the attributes may not represent the current conditions at the beginning of the project analysis year.

The following summary of data source availability is focused on FVS-ready treelist vegetation data with some consideration given to other vegetation data for fuels treatment planning. To help make this situation more tractable for resource manager, there is a need to be able to “impute” missing treelist data from known sample records and then to use the FVS Vegetation Dynamics simulator to “grow” all plots to a common year. This then provides the wall-to-wall treelist vegetation data layer that can be the foundation of the ID Team analysis and planning project whether for fuels treatment or other resource objectives. **Data to support the imputation process exists and it can be aggregated if desired to make it more broadly accessible and useful.**

It is the **objective** of this summary paper to describe the database sources available to Interagency specialists that contain field records eligible for use in the data imputation process for predicting FVS-ready treelist data. Furthermore, this paper describes the desirability of improving access and availability to these various data sources so that an ID Team in any agency can be sure to assemble the best available data for the landscape under study. The ultimate goal is to use the data and imputation procedures to create data layers for the entire U.S. These data layers would include all the fuels information needed by managers for fuel treatment planning.

Tree List and Other Vegetation Data Sources

USDA Forest Service - FSVeg Database

The Forest Service - FSVeg (Field Sampled Vegetation) database is currently only available within the Forest Service firewall. This data is currently stored in an Oracle database at the USDA - data center at Kansas City. There is currently work underway to build a duplicate data warehouse. This data would then be available for use in imputation by any application. This is anticipated to be complete in the next calendar year. This was confirmed with the NRIS (National Resource Information System) staff.

USDI FFI (FEAT-FIREMON)

Recently FEAT and FIREMON (fire effects monitoring tools developed to assist managers with collection, storage, and analysis of ecological information) were integrated into FFI. However some FIREMON users have elected not to upgrade at this time. Thus, data is available in both the FIREMON and FFI databases. These databases reside on PC's or on servers depending upon the agency involved. The databases on servers are behind a firewall (either BLM, FWS, or NPS) and thus generally not available except to users who operate behind each particular agency firewall. The databases that reside on PC's are accessible only by the person operating that particular PC, no one else. The BLM has taken pains to seek out these dispersed databases and consolidated them into FIREMON and FFI. All of NPS fire effects monitoring data, with the exception of a few parks, are in FFI and could be easily consolidated into one database if desired. Indeed, effort is underway within the NPS to consolidate the meta-data for 91 locally maintained FFI databases on the NPS DataStore site (science.nature.nps.gov/nrdata/). Access will be initially limited to NPS employees and it will be updated annually. The NPS DataStore application provides data producers and users with standardized, integrated metadata and data management and dissemination capabilities. Part of the DataStore design allows for distribution of metadata and data to other federal and non-federal clearinghouse sites. FWS uses FFI and FIREMON for fire effects monitoring as well. The BIA has fire effects monitoring data mostly in FIREMON with some in FFI. This tribal data may or may not be available for imputation purposes depending upon the level of restrictions a particular tribe places on the data access.

GNN Database and grid Layers

Another source of already imputed tree list data is the GNN (gradient nearest neighbor) product for all of Oregon and Washington States. This data is currently stored in export files and on the Oregon State University server, but this is not considered a long term solution. Storing this data at FRAMES (Fire Research and Management Exchange System) could be a solution for data storage and would make this data available to interagency users. (ArcFuels currently uses this data and INFORMS is developing the ability to use this data)

Forest Inventory and Analysis (FIA) Data

Many imputation technologies including LANDFIRE utilize FIA data as part of their source data. These tree lists are used above in GNN as well but the initial plot locations are not available. The tree lists are now available for use after imputation (GNN example) because it is impossible to tell where the actual plot is located. By law FIA can not divulge the plot locations. An on-going discussion with FIA has been the development of technology that would allow for an extensive data set to be available for needed imputation relationships without having to disclose plot locations. This would allow the imputation technologies to take advantage of FIA data to update data layers without going through agreements and storing these sensitive locations.

Other Potential Sources

Investigation should be made with other agencies regarding the availability of their “corporate” data. For example BLM has the FORVIS database, BIA has Continuous Forest Inventory data, and NPS has data from their Inventory and Monitoring program. Each of these sources likely has more data than is currently stored in FFI/FIREMON that has potential utility to the support the treelist data imputation process.

Issues Affecting the Availability and Use of Treelist Data Sources

Data Access: There are a lot of potentially useful data floating around out there that are not accessible or that nobody knows exists. Some of the issues that limit access are: (1) the data are behind a firewall that only a particular segment of the Interagency resource specialists have access to; (2) legal concerns, such as sensitive species, limit data access; (3) personal concerns on the part of the data “owner” about sharing the data; (4) propriety data (BIA); (5) lack of time and energy to place data into open use systems.

It appears from the information that we have collected that the USDA Forest Service has centralized their field data and is about one year away from placing a copy of the entire national database into an open access warehouse system at their Kansas City server farm. The USDI agencies appear also to be moving in this direction with the ongoing FFI project. One answer might be to consolidate USDI data at a single source such as FRAMES (Fire Research and Management Exchange System) which is an interagency cooperative project with the University of Idaho. Greg Gollberg, the FRAMES project leader, is very open to hosting a copy of the various source data from which treelists can be imputed. USGS/NBII hosts FRAMES at the present time. If only two database repositories could be created, FFI databases on FRAMES and FSVeg on the KC warehouse site, it would be much easier for data mining software tools to locate, capture, and process relevant data for particular projects.

Missing data mining software: The key issue is that the data can be extracted into FVS ready form and that proper expansion factors are stored to determine the proper use of the data in the model. Methods must be developed to determine if each tree list to be used contains adequate sampled attributes to support imputation needs. A proof-of-concept approach might first provide a tool to gather data for the imputation process and then a basic delivery system. The proof-of-concept may provide some useful insight into the feasibility of a full-blown approach, which would serve data for the multitude of models and reporting tools used by managers. It is important to understand that most projects could benefit from a larger base of field records to support the imputation effort, even if the project itself is wholly within one agency ownership.

Considerations of the type of data made generally available: Development of a data source that can make available more than just basic tree data is an important goal. What has been discussed to date are tree lists. A database of just trees limits our ability to use the data for fire behavior modeling or, just as importantly, fire effects. Information for tree lists are just a subset of the information available in FFI, FIREMON, FSVeg, and other data sources. Most monitoring databases store much more vegetation information such as Coarse Woody Debris, Duff, Litter and other attributes that are very important when analyzing vegetation. While imputation of non-treelist vegetation data is important and desirable, it is currently vulnerable with significant prediction accuracy problems (see the discussion on this topic by Ohmann et al.). This is especially a problem for fuels treatment analysis and planning because the existing fuels load could easily be a residual from some previous vegetation community rather than the current vegetation community and thus completely confounds the imputation process.

Vegetation data (tree lists) are not the only data needed to generate imputations. Global data and Vegetation Polygon data are generally needed to run imputation models and will also need long term availability and storage solutions. It should be clearly understood that project level fuels treatment analysis often is based not on Grid-type GIS layers but rather on polygon-type GIS layers. Where LANDFIRE data can be used for project level fuels treatment analysis, the following points do not apply.

> **Global Data**: Global data includes Landsat Data and DEM data transposed into various forms such as slope, aspect, solar insolation, solar duration, slope catchment area etc. Will need to be stored and made available just as the tree list data. This may not be required initially. Further discussions will be needed to decide how to handle all these data which are mostly grid data but there are issues around joining tiles and how it will work. There are some solutions that could be developed but we need to negotiate standard solutions.
>
> **Vegetation Polygon Layers**: Vegetation polygons are common in many management areas. Many users want to interact with polygons and really cannot or will not deal with grid or pixel data. There will also need to be a minimum data attribute sets associated with these layers. This attributes can be very minimal. If these can be stripped down to the minimum a layer joining technology could be

built (some already exist). This would allow vegetation layers convering multiple ownerships to be used. Many land management agency units don't currently have up to date vegetation polygon layers. There are new technologies on the horizon, which need more testing, that may allow inexpensive and efficient delineation of vegetation.

What's Needed Next

It is generally agreed that data exists which can be aggregated for imputation purposes. This data is often inaccessible because it is scattered, behind firewalls, difficult to locate, often on PC's, and so on. Particularly from an Interagency operations point of view, it is critical that these expansive and valuable data be found, organized, aggregated, and a copy placed in some public warehouse system that is accessible to everyone to use.

The initial steps for bringing the existing data together would be: a) determine the data variables important for imputation, fire behavior modeling and fire effects modeling (what the managers need), b) determine the amount of field data available (including permission to use the data), c) create a home for tree list data to be imputed and spatial references to these treelists to be made.

Steps for serving the data layers back to managers would be: d) determine the data layers that managers need, e) include agency representatives to assist technical approval to access the data, f) provide funded positions to assist with data management (FRAMES?), g) provide tools to access the data (be able to view and clip the desired scenes/data), h) provide documentation and training for the tool. The LANDFIRE project has had to deal with much of steps d – h and may be able to provide assistance based on their experience.

References

Dixon, Gary E. comp. 2002. *Essential FVS: A User's Guide to the Forest Vegetation Simulator*. Internal Rep. Fort Collins, CO: U. S. Department of Agriculture, Forest Service, Forest Management Service Center. 193p.

Ohmann et al. 2008, in review. Nearest-neighbors mapping of vegetation and 'tree lists' at landscape scales in the US, a state of the science status report, unpublished manuscript, pp. 8 (available upon request from Joint Fire Science Program, Boise, ID).

Reinhardt, Elizabeth; Crookston, Nicholas L. (Technical Editors). 2003. The Fire and Fuels Extension to the Forest Vegetation Simulator. Gen. Tech. Rep. RMRS-GTR-116. Ogden, UT: U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station. 209 p.