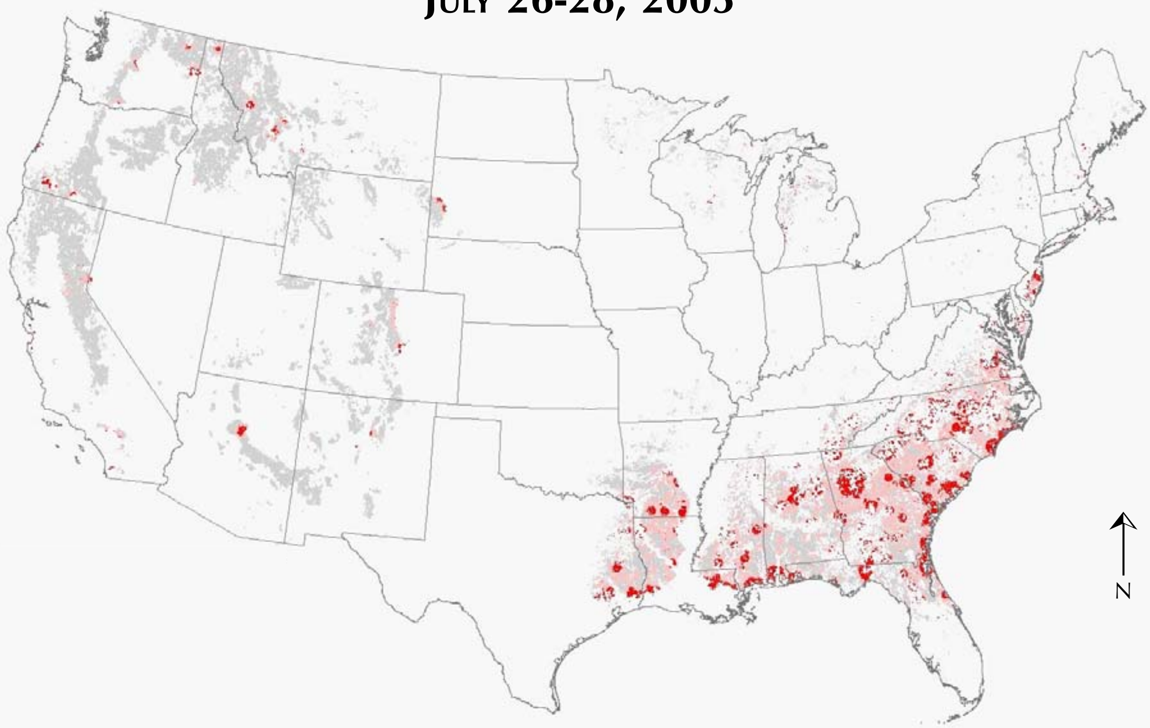
Forest Health
Technology Enterprise Team

RISK MAPPING

Deriving Data

WORKSHOP PROCEEDINGS: QUANTITATIVE TECHNIQUES FOR DERIVING NATIONAL-SCALE DATA

WESTMINSTER, COLORADO
JULY 26-28, 2005



Compiled by Michael Marsden, Marla Downing, and Mark Riffe

Forest Health Technology Enterprise Team—Fort Collins, Colorado

United States Department of Agriculture

Forest Service

FHTET-2005-12
September 2005

Workshop Proceedings: Quantitative Techniques for Deriving National Scale Data

Westminster, Colorado
July 26-28, 2005

Compiled by Michael Marsden, Marla Downing,
and Mark Riffe

USDA Forest Service
Publication FHTET-05-12

ACKNOWLEDGEMENTS

Thanks are extended to all the participants who gave of their time and efforts to make this event a success. We especially wish to thank those who made formal presentation at the workshop and those who contributed to the final session to define needs and direction attributes of a national risk system.

FOREWORD

A national risk assessment system is needed to identify and display the present and potential major impacts from forest pests in map format. This is an important tool for information transfer in communication within the Forest Service, between government agencies, and to clients outside these agencies. Pests' impacts from native pests such as southern pine beetles and dogwood anthracnose in the East or spruce beetle and dwarf mistletoes in the West have been reported by Regions in a large-scale map format. Although these provided a summary of condition, they did not allow for close management assessment and were difficult to verify. Now we have the added assignment of mapping the potential areas of impact for introduced pests and producing maps at a high scale of resolution consistent by location and with a readily defendable risk status.

The formal presentations at this workshop are arranged into three parts: **An Overview of the Risk Mapping Question**, **A Review of Methods of Analysis Used in Risk Mapping**, and **Current Examples of Risk Mapping**, though there is some degree of overlap of presentations in the three groups. Some in the presentations made in the Overview section define both the methods by which the maps were constructed and make comments on the analyses. In the Methods presentations, examples are used with actual forest data sets. In the Current Examples section, some comments on the attributes of the methods used are made. It is hoped that grouping the presentations will help the reader grasp the key issues of this meeting.

Following the presentation is a summary of the discussion session: **Where do we go from here with National Risk Mapping?**

CONTENTS

AN OVERVIEW OF THE RISK MAPPING QUESTION

Frank J. Sapio made a presentation, titled "**Why Invasives Species Risk Mapping is Important.**" In it, he sited the creation of the National Invasive Species Council and the charge made to it. He then proceeded with a case study – ash decline in Michigan. This included the history of the damage, the identification of the pest *Agrilus planipennis*, the development of the Ash "Risk" Model, and the mapping for the risk ratings as an aid to control of the outbreak.

Borys Tkacz presented an update on the Data Quality Act in "**Information Quality – Implications of Recent Legislation.**" The Data Quality Act passed in 2001 is having a trickle down effect as different levels of government provide more details on the expected changes in the management of data and information in the Forest Service.

Following the Office of Management and Budget (OMB) definitions and guidelines, the USDA has issued their directions. Within the Forest Service there are some special cases, most notably the Forest Inventory and Assessment (FIA) data sets. To do the risk-mapping assignment, we will be dealing with large data sets, mathematical models, statistical analyses, and mapping. Not only are we to keep records of the maps, we need also to keep records of information that went into the construction of these maps and make the information available if requested by congress or the general public. This charge makes the record keeping of risk-mapping a mandate.

Bill Smith's presentation was "**Data Mining and Mind Mining: Principles and Limitations of National Risk Mapping for Exotic Insects and Diseases**." In it he applied some very basic tents in developing a risk/hazard map for Sudden Oak Death (SOD). The problem appeared nationwide. The model was kept simple, but includes key pieces such as host identification, climate where the pest can survive, and pathways for the spread. This was a conceptual model not a data-driven one. It was followed up with both surveys and evaluation of FIA data plots. Bill also referenced the Data Quality Act as an additional incentive to document well how we construct any risk map.

Ken Brewer presented "**Development and Production of a Moderate Resolution Forest Type Map of the United States**." In 1993, forest group types were mapped for the entire country. Since that time the data available for mapping forest group types has changed. Today, we have a higher resolution data set (250-m pixel data instead of 1-km pixel data), continuous nationalwide geo-spatial data, new modeling techniques, and an improved computing environment. The key analysis tool used in the update of the forest groups was Cubist/See5, which calculates a set of classification trees. Examples of the old and new maps were presented side by side. Accuracy assessments were made using the FIA plots that were held out from the model calibration steps.

Andy Lister put a different spin on the risk-mapping task in using the mapping technique to enhance the data set in his presentation, "**Creation of an n-Dimensional Spatial Database Instead of a Static Map.**" He used a K-NN approach in which k nearest neighbors are considered for imputation of a value to each element in the spatial database. Then summaries of the k levels of imputation are computed. The average value imputed can be mapped, as can the variance of that value and the mean distance it is from a known value. This allows for mapping the risk and two measures associated with its precision.

Frank J. Krist Jr. made two presentations at the workshop. One, "**Standardizing the National Risk Map Utilizing A GIS Based Multi-Criteria Modeling Framework**," is included in the overview presentations because it presents the latest risk mapping procedure used by FHTET. Reports such as "The Forest Insect and Disease Conditions in the United States 2003" are national reports and they include pest risk maps. It is not desirable to see a change in the host, pest, or risk values along a state or other political boundary unless one really exists: to eliminate political boundaries, it is necessary that maps on both sides of the boundary be built using the same tools. Frank's presented the five steps to the multi-criteria modeling process and some examples.

A REVIEW OF METHODS OF ANALYSIS USED IN RISK MAPPING

Raymond L. Czaplewski presented some work done by **Gretchen Moisen** and other FIA scientists and collaborators entitled "**Mapping Forest Attributes in the Interior West: Comparing Predictive Modeling Tools**." The presentation first described the five-step process for building risk-related maps. It then summarized findings of several studies comparing the performance of a variety of statistical tools for modeling forest attributes collected on forest inventory ground plots as functions of satellite-based information. These tools include generalized linear and generalized additive models, tree-based methods, multivariate adaptive regression splines, neural networks, kriging, stochastic gradient boosting, hybrid methods, and simple linear models. Recommendations were made for production mapping efforts.

Valerie LeMay made a presentation based on work by herself and **H. Temesgen**: "**Comparison of Nearest Neighbor Methods for Estimating Basal Area and Stems per Hectare Using Aerial Auxiliary Variables**." The presentation provided background information on imputation methods and demonstrated how these methods were employed to generate tree-lists from aerial attributes for non-sampled polygons, and improve forest inventories, analyses, and management. Examples were given using data from multi-species and multi-aged stands from southeastern British Columbia.

Kenneth B. Pierce Jr. presented some work done with **Janet Ohmann**, "**Gradient Nearest Neighbor Imputation for Mapping Forest Attributes and Variability**." This presentation concerned work done on the construction of maps of forest stand conditions including forest fuel attributes in three very large and very different areas, one each in Washington, Oregon, and California.

Albert R. Stage's presentation, "**Some New Twists in the Art and Science of Imputation**," first defined the basic nearest neighbor imputation method and then answered some question about the errors associated with the method. This presentation explored the components of the error terms for the predictions and then gave examples of the predictions and their associate errors. Al's presentation also addressed several statistical techniques that should be considered when assessing the accuracy of nearest neighbor imputations.

B. Tyler Wilson made a presentation of an alternative method of testing the mean difference between two populations, "**Validation of Geospatial Models using Equivalence Tests.**" Instead of testing the usual null hypothesis "means difference equals 0," Tyler tests an alternative "mean difference does not equal 0." He uses a regression based validation posture to test two aspects of the hypothesis: intercept and slope. The presentation includes application to three competing models for predicting forested areas in Minnesota.

CURRENT EXAMPLES OF RISK MAPPING

Raymond D. Watts made a presentation on the North American's Road Network, "**USGS National Overview Road Metrics**." Roads are presented, not just as another overlay that gives some measure of access to a given area, but as a feature of the landscape with their own ecological implications. Raymond showed that statistics based on the road network like road density and distance to nearest road can be used as environmental indicators for some species.

Albert R. Stage made a presentation of **Gerald Rehfeldt's "Geographic Clines in Genetic Variation**." It included the mapping of 48 climate variables derived from temperature and precipitation monthlies. These monthlies were then used to predict genetic differences among populations within species, and the resulting geographic clines for these western conifer species then could be mapped. The key contribution to risk-mapping is that point predictions of climate can be used to produce maps that will show where certain trees species should be competitive on the landscape (at least as far as the climate is concerned) and where the species would be stressed.

Raymond L. Czaplewski presented an approach titled, "**Re-Sampling Remotely Sensed Data to Improve National and Regional Mapping of Forest Conditions with Confidential Field Data**." In it, he advanced the idea of overlaying the 150-m pixel data now available over the MODIS 250-m pixels grid, matching the FIA field data plots on the 150-m grid, and then repositioning the FIA plots to three other spots using specified misregistration errors associated with the FIA data points. Using a nearest neighbor scheme to encode the 150-m pixel database, this re-sampling method would have the added advantage of obscuring the location of the FIA plots.

Robin M. Reich presented a summary of his project in Mexico, "**Analysis of Statistical Strategies for Advancing Monitoring and Assessment of Terrestrial Ecosystems in the Mexican State of Jalisco: Pilot Study Synthesis**." This was a complete project description covering the planning stage, definition of conditions to be mapped, obtaining the available data sets (Landsat TM Bands, elevation, slope, aspect, and vegetation type), design of ground sampling, methods of statistical analysis, and mapping for conditions. The modeling task was split into two parts: large-scale trend surface model and small-scale binary regression tree model. The final product was a 10-m resolution map and relevant summary statistics.

Frank Krist made the last presentation of the workshop, "**A "Keep It Simple Stupid" approach to estimating BA**." Frank made the point that most risk models make use of data on stand density by tree species. In his example, he used individual species basal area (BA), quadratic mean density (QMD), and stand density index (SDI) data taken from 437 plots on the Barlett Experimental Forest and Landsat data to map the BA condition classes. Extending this process, he used FIA data plots as a base to produce a national map. The key point was that, by keeping the model relationships simple, the maps could be built using Arc/Info.

DISCUSSION SESSION: WHERE DO WE GO FROM HERE WITH NATIONAL RISK MAPPING?

THE PRESENT RISK-MAPPING SYSTEM

The present risk-mapping system is designed to produce national risk maps at the 1-km scale. It produces and maps functions of risk for a variety of pest impacts. The risk mapped includes conditions that may produce catastrophic mortality by tree species. To define this event the normal tree stocking level by area by species needs to be known, as does the normal expected mortality for such stands. Past criterion was 25% BA mortality over a 15-year period. Today this criterion must be considered in terms of economic, ecological and social impacts. This leads to the need to risk-map more that one criterion.

The system must include an update capability. Checking the map against new data will show where the map captured high-risk conditions correctly and where it missed them. It should also allow the user to identify areas where the information on which the risk is based is strong (complete data) and where it is weak (due to lack of data).

Validating a risk map is a part of the job. There were two approaches discussed to determine the accuracy of a risk map. The first was the bottom up or error propagation approach, but that has problems of dependence among error terms. The second approach was to move to the output and determine in advance what can be measured to compare to the risk map. To some degree this has been done in the Michigan study. The aerial survey data (from sketch maps) were laid over the risk map. Because the risk map was based on data taken in 1984, it should be relevant to pest impacts that occurred since that date. A key point to remember is that risk should be viewed as an ordinal output—a greater value indicates a risk greater than the risk for a lesser value. To validate such data, you need to rank the points selected to test the model. You also need to design a method to sample conditions to test the model, and not simply take opportunistic data from some highly impacted area.

Exotic insects and diseases. For some risk maps of exotics, the results cannot be measured because the pest is not here yet. When the risk concerns a rare event or an endangered species, you have to broaden the sampling to include areas of both high and low risk, getting enough data to test for this occurrence. For invasive species there are several components including the susceptibility and exposure or path of entry. This suggests that the problem be seen as having two parts: determining the presence and conditions of the host trees and tracing the shortest path for the potential exotic pest to possible introduction points. Attributes that go into the computation of the risk, such as BA of the host species, can be measured and compared to the value used for the point in the risk map. Inspections of sites within the map will confirm the conditions required by the exotic pest, but relies on an expert opinion at this point.

NEEDS AND OPPORTUNITIES TO IMPROVE THE PRESENT RISK-MAPPING SYSTEM

There is a need to further test the possibilities for using the risk-mapping system with data taken at different scales. Since the risk-mapping system is scale independent we need to test to see what happens when it is run at different scales (example with 1-km pixels or with 250-m pixels) with the same application. If there is a difference we need to explain which is the better scale for the application. Land planners are working with stand level data and they will want to use the risk-mapping system at this scale. There should be instructions or at least cautions about doing this.

For some applications the risk-mapping system needs to be linked with stand-level models. The relationship of the risk map to results of stand dynamics models is the more reasonable linkage at present. High values from the risk-mapping system computed at the 1-km pixel level sometimes can be interpreted as the increased probability of an insect outbreak. This can indicate as a need to use pest extensions for modeling stands within that pixel to estimate impact in greater detail. The direction from stand dynamics to risk mapping has dimensionality problems. A tree list useful in running a FVS model has little utility for a 1-km pixel. However, the base mortality level for a species over a 15-year period can be predicted by FVS and used as a base level for the risk-mapping system. The FVS pest models can also be used to predict future conditions, and this information can be transferred to the risk-mapping system to generate a map of future risk conditions. With the present models, professional judgment must be used to transfer information between the two systems.

There is a need for the risk-mapping system to use data and functions that are consistent with those used in planning. An important problem is the consistency of data between systems used by the Forest Service. Some of the terms basic to both are “stocking levels,” “growth rates,” and “mortality rates.” The levels and rates used in the risk-mapping system and those used in planning need to be in agreement. They also need to be well documented so that when someone asks, the values and assumptions that went into generating a risk-map can be recalled. Different standards can be used across the nation but within an ecoregion the models and criteria should be consistent – land management plans and risk-maps must use the same set of standards.

There is expected to be a link between pest impacts today and fuel conditions in the future. It is thought that the maps of pest impact today will closely relate to the forest fuels maps in the next decade. Many of the same people who are working on the risk-mapping system are also working on the mapping for forest fuels. There will be a project to add maps of past fires (25-27 years of records) to the mix, giving us ability to produce maps of what has burned, what is prime to burn and what will have prime fuel conditions in the next ten years. To do so we must keep track of the process that generated the change is forest conditions. There are large differences in insect caused mortality and stands that were thinned to the same live basal area conditions. In one case the fuel are dead trees with associated limbs and needles, in the other the slash has been burned or treated so that it is no longer a high fire danger.

DATA CONSIDERATIONS

The key data sets used today. Complete coverage of the area can be expected from the 1-km pixel data set. The FIA data plots must be matched with a pixel or combination of near pixels. Data of the type measured on the FIA plots must also be imputed to the rest of the pixels for the map. This involves two processes; a training set of data and an imputation process. The presentations at this workshop have explored what we know and where we need to do some further research on the imputation process. The training set of data used to develop a classification function is complicated by what is called "messy" plot data: plots that contain more then one cover type classification. There are cases where half the training data set would be put aside because it did not have a single forest cover type.

Linking other data sets to the risk-mapping system. Another item of interest to add to the list of overlays in the risk-mapping system is the sketch map data of pest impacts. Here, the cause is as much a variable of interest as the damage. There is also a great deal of data on large-scale wind events. This is a national risk-mapping system, and we need to acknowledge that different variables are of interest at the national level, the regional level, and the local level.

Stratification and macro-data elements. Ecoregion classes, landforms, and even soil conditions can be used to sub-divide an area into more homogeneous units for modeling or to add directly a variable supporting the expectation of either the host species or the pest agent of interest. An ecoregion map is a hierarchical classification that deals with geophysical site conditions that exist everywhere; landform types at the phase level are consistent. Potential vegetation classification systems provide good accessory information: when you have data on potential vegetation, it should be used. This approach may work better in the arid and mountainous West than in areas where water and elevation play less of a role in determining forest composition. Region 5 and 6 are doing this with ecoregion at the section level. At the 1-km scale, units should be classified correctly using ecoregion data, but this may not be sufficient for smaller units.

Measured, modeled, and imputed. In the model building process, there is a need to know which variables in specific data sets are measured, which are modeled from the measured variables, and which are imputed from another data set. In the present system, each pixel is assigned a risk class. Within each pixel there may be information on the percentage of the area in each risk class condition.

Spatial data base attributes. One aspect of the data set we use most concerns understanding the relationship of the measurements taken at different scales. At the national, regional, and local levels, there is a need to better understand the relationship between the FIA plots measurements and the measurements taken on the 1-km pixel. The opportunity exists to enhance the FIA plots data with photographs of the area around the plots to estimate a neighborhood condition. These photographs may help bridge the gap between what the FIA plots are measuring and what the 1-k pixels are seeing. This will be tested at three locations in the country in the next few years.

The data sets available for risk mapping. There are several data sets available for risk mapping: most prominent is the FIA Predictors Data Set, which contains 250-m roster data for 230 variables and has the capacity to include others. Continuous Vegetation Survey (CVS) on the West Coast is a double-intensity inventory set, but it is not double-intensity in Colorado. The Resources Planning Act (RPA) requires an assessment every ten years and has a database with BLM, FS, FIA, and Natural Resource Inventory (NRI, the Natural Resource Conservation Service's system of strategic inventory program—the equivalent of FIA data for agricultural lands) data layers. Most all of this data was taken on some systematic sampling scheme except stand inventory data, which always has some random component. Stand survey data can be added to the master set if it has a GIS component to it. This National Predictor Data Base is the place to add the climate and species functions developed by Gerald Rehfeldt.

To the National Predictor Data Base we can add layers that are actual risk maps and overlay sketch maps. Conditions under which these impacts occur must be included because some impacts happen on non-forested lands—gypsy moth would be a good example.

For the invasive species, the mapped risk is expressed in terms of potential based on information about the pest in other countries. A data layer with information on global distribution of the important host species and invasion pests can be included in the larger database. These data would be used to develop the risk model for the pest should it arrive in this country.

Database management. The data sets for the risk map should be easy to update when new plot data is available, and should carry the history of the plots used when the map was created. This requires either taking snapshots of the data and archiving them or taking snapshots of the plot list used in the map. If stand simulation models were used as part of the analysis that produced the risk map, the model must be clearly identified and archived. That brings up the question as to who keep all these maps and related information—there may be 40 to 60 maps.

Unique problems in using the FIA data. There is a limitation in getting access to FIA data needed to build risk maps. There was a case in which the FIA map produced a few bulls-eyes: these could be assumed to be FIA plot locations, so the keepers of the data withdrew the 1-pixel data map and presented a 10-km map with no bulls-eyes. For risk mapping, we need to streamline the process for getting data and ensure that we have access to the right data when dealing with FIA plots.

Misclassification of pixels. Another FIA data problem is pixel misclassification. For example, an FIA plot at inventory might represent a timber stand, but five years later, the satellite might indicate a clear-cut there. One source of error might stem simply from incorrect plot location: misregistration of field plots can lead to pairing existing plot data with the wrong pixel. If the surrounding pixels are the same (or almost so) as the data's intended pixel, then there is no serious error due to misregistration: otherwise, the data is paired with incorrect pixel values. A recourse lies in taking a weighted average of likely pixel matches for use with the FIA plot data. Moving from a cover criteria to structure criteria may allow for building a data surface with mixed conditions.

RISK MODEL CONSIDERATIONS

The data elements. When developing a risk model with data taken at different scales, care must be taken: some variables are sensitive to the scale at which they were sampled. For example, BA at the point level is more variable than at the stand level, and may create extreme predictions in the FVS stand level models. The sampling scheme and the scale at which you wish to use the model must be taken into account when the model is calibrated and when data is sampled for use with the model. If you plan to build a risk-model that is scale-independent, then it will likely be couched as a point model. The sampling frame for the data should be a part of the record as should any error estimates for computed variables.

Risk model development strategies. Generalizations can be made at the back end of the model and avoided up-front, minimizing imputation error. An example of the former would be in running FVS/pest models with stand level data, doing a risk analyses at this level, and imputing the risk to the next higher level of data (pixels). A recent analysis in Region 6 ran at the plot-level and ended up with a multiple-criteria set that was imputed to the larger-scale data set. A risk-surface was computed at the pixel level. These risk computations could have been made at that front end of the system with plot data that would be modeled directly to produce rating of risk to a given pest or pest impact. These risk ratings would be imputed to the map along with errors associated with the predictions, and any surface trending or smoothing would be done directly on the risk-surface and would not be affected by the factors that go into computing risk.

A two-part approach to constructing risk maps. (This approach is not to be confused with sub-dividing the area into two or more homogenous units.) Methods such as CART, Regression Tree Analyses or the fitting of splines can be used to fit a surface to the physical and temporal data variable for some variable measuring a key host or pest requirement (temperature, precipitation, soil type, or land-use). A series of analyses using values from this fitted surface can then be done to predict or impute values for variables measured at the smallest level of resolution (usually plot and point samples) across the map. The final risk-mapping functions were then derived from analyses of this last data set using some criteria of host present, pest present, or host damage.

An alternative to the analyses approach above is a hybrid system in which stand- or plot-level variables are used in the same analyses with the pixel-level variables—even larger scale variables (ecosystem and land form) can be included. This is not a recommendation to throw all the data into one analysis, but to judicious choose which variables of each scale are appreciate to the series of analyses needed to define the risk model. A key decision is still "when do you impute data to a higher layer of the system and when do you predict a smooth surface of values to that final set of points?" For example when risk modeling for western spruce budworm, not only are the host conditions in the pixel (or stand) important, so is the general surrounding conditions. A variable could be computed from a sample of the conditions in adjacent pixels to put the center pixel in a susceptible landscape or not in a susceptible landscape. In modeling this risk to the pixel, this and other area-based variables could be used with the within stand variables.

How do we decide which data layers to use and what modeling approaches to use in developing a risk map? In some cases, a risk model is simply an overlay of the conditions necessary for the pest to survive. Using expert knowledge, you can list the variables that are likely to be driving the process, such as host plant distribution, and these are the primary building blocks for your model. For native pests, you may also have the history of past outbreak episodes. If you combine the right overlays (host species and climate conditions), you should see areas with conditions similar to those in previous outbreaks. This is base data-mining, and includes exploring both the spatial data and the tabular data. Many of the routines used here are automated and can be done quickly: create a response matrix—spatial data identifies locations of interest—and then overlay the tabular components for them. Forest type maps were based on USGS maps and expert opinion; now, they form a basis for susceptible forest type models for risk mapping.

Such modeling is a response to objectives, but new books on methods appear each year. One tool not mentioned is the envelope construction from data samples or imputed populations. Gerald Rehfeldt used this method to construct maps of potential to support a tree species as opposed to mapping where the species now occurs.

Univariate vs. multivariate models. Consider the task of creating a risk map for more than one pest on a host or more than one host for a pest: should a series of univariate techniques be used in the analyses or should a multivariate approach be used? If you have a multivariate objective, you should start with multivariate analyses methods. If you have a univariate objective, you start with univariate methods until someone comes along and poses a second question, which might require you to look at multivariate objectives.

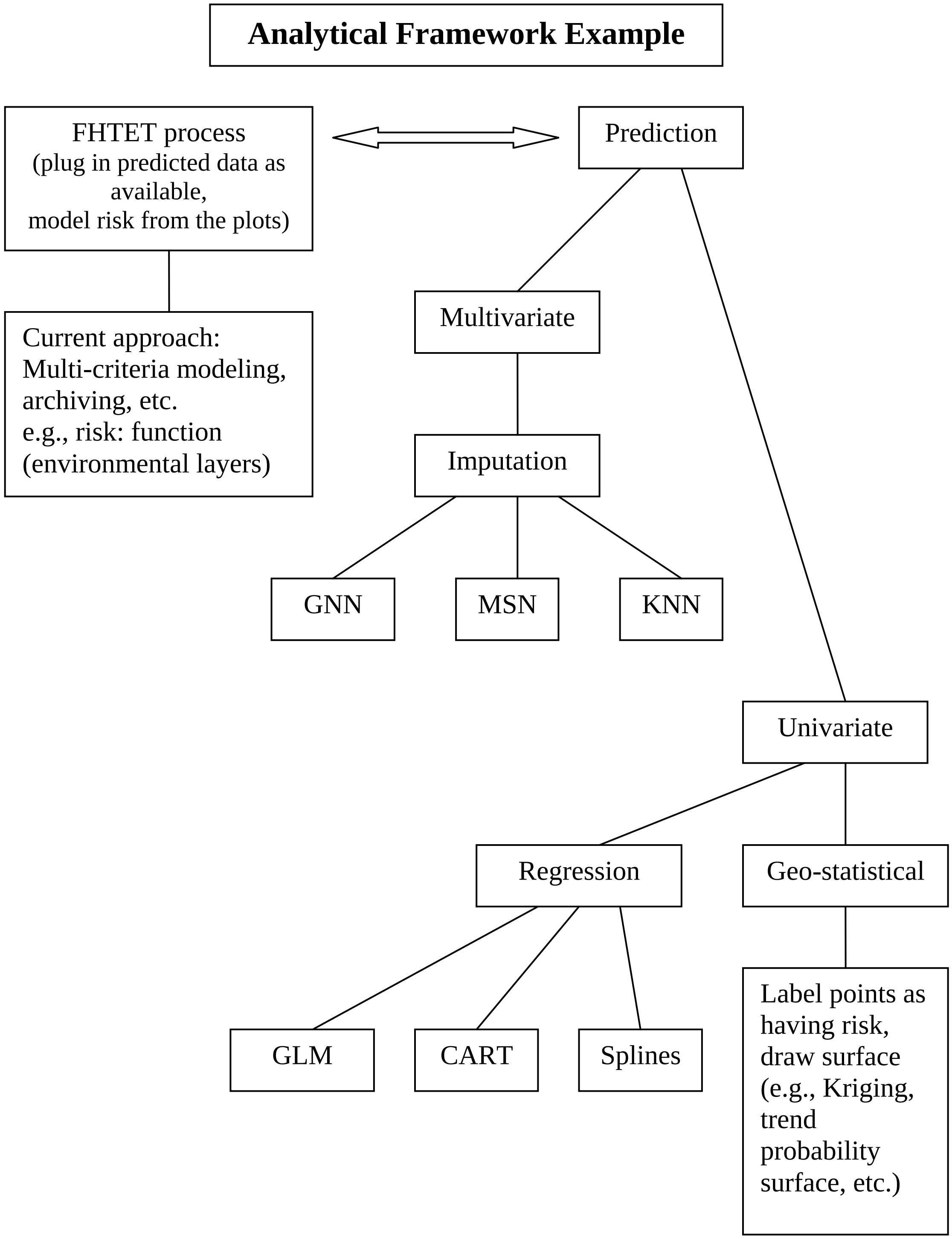
Choosing a modeling approach in risk modeling. There is no single best method for all risk mapping. When you know the objective, you have a good idea where to start. There are several considerations to be made: does the objective deal with multivariate or univariate criteria? Is a static model or a predictive model needed? The multivariate criteria problem is becoming more common. At present, the predictive function lies in stand-level models; most likely, you will also need to impute data to fill the map. If you just want a static forest type map, there are simpler options (CART, Geo-statistical, etc.).

The role of outliers in a multivariate data set. If outliers are included in the analyses, they may become pivotal data points; if dropped from the analyses, some information about the predicted variable's distribution may be missed. In some cases, you may wish to tone down the effect of a very large value in a data set by setting a maximum value: thus, that one outlier point is still considered, but it will not have an inordinate effect on the model. Weighting the data points is another option. Knowledge of the sampling design is needed to assign weights to individual observations. There are rigorous techniques for dealing with observations of different weight or measured with different errors. First, incorporate the expansion factors for the observations into the model building and, when needed, use these factors as weights in the analysis. This is one way of dealing with errors in the variables. There are also analytical methods for dealing with these errors, but you need to be aware that the weights can change through time: small trees measured on 1/300-acre plot have a large expansion factor; the same tree after a growth simulation will be of much larger diameter, but still carry the same large (now, incorrect) expansion factor. You can estimate the error of the predictive function from the errors of the variables used: when you apply the function, the variable is altered. The same error structure resides in the data you are observing as in the data used to build the model, so they cancel out and you can compare the prediction directly with the observed.

CONSIDERATIONS AND DECISION TREE

The following table presents some considerations in the choice of data imputation techniques. The following decision tree then suggests the decision points that affect choice of those techniques.

Risk Map Objective	Univariate criterion	Multivariate criteria
Static	Observed data or known function of data that indicates a risk level in a single factor - the value is mapped directly or put into classes that are mapped - example historic high levels of pest damage	A set of measurements or data points that taken in combination indicate a risk level - the risk levels are mapped directly or put into classes that are mapped - the risk may be an overlay of two or more layers to create different levels of risk - example high BA levels of host and nearness of pest
Predictions into the future	The observed data is taken at one time period and conditions are to be simulated for a future condition that can be used as a single risk factor - example future areas with high BA of host	The observed data sets are to be predicted into the future and then an analyses of the future values will yield a single risk factor or as set of risk factors - example future areas of high BA of host, drought conditions and nearness to past outbreaks


Analytical Framework Example
FHTET process
(plug in predicted data as available,
model risk from the plots)
Current approach:
Multi-criteria modeling,
archiving, etc.
e.g., risk: function
(environmental layers)
Prediction
Multivariate
Imputation
GNN
MSN
KNN
Univariate
Regression
Geo-statistical
GLM
CART
Splines
Label points as having risk, draw surface (e.g., Kriging, trend probability surface, etc.)