

## SPATIAL VARIATION IN FIRE FREQUENCY IN THE INTERIOR COLUMBIA RIVER BASIN

Donald McKenzie<sup>1</sup>, David L. Peterson<sup>2</sup>, and James K. Agee<sup>1</sup>

<sup>1</sup>College of Forest Resources and <sup>2</sup>USGS Forest and Rangeland Ecosystem Science Center  
University of Washington Field Station  
Box 352100, Seattle, WA 98195-2100  
E-mail: [dmck@u.washington.edu](mailto:dmck@u.washington.edu)

### ABSTRACT

We developed multiple regression models and tree-based (CART - classification and regression tree) models to predict fire return intervals across the Interior Columbia River basin at 1-km resolution, using georeferenced fire history, potential vegetation, cover type, and precipitation databases. We weighted the models based on data quality and performed a sensitivity analysis of the effects on the models of estimation errors due to lack of crossdating. The regression models predict fire return intervals from 1 to 375 years for forested areas, whereas the tree-based models predict a range of 8 to 150 years. Both types of models predict latitudinal and elevational gradients of increasing fire return intervals. Although the tree-based models explain more of the variation in the original data, the regression models are less likely to produce extrapolation errors. Thus, the models serve complementary purposes in elucidating the relationships among fire frequency, the predictor variables, and spatial scale. They also demonstrate the integration of qualitative and quantitative methods, and can be updated as better fire history data become available.

### INTRODUCTION

In this paper, we present statistical models for predicting coarse-scale patterns of fire frequency in the Interior Columbia River basin (ICRB), using a fire history database (hereafter FHDB) from the western United States (Heyerdahl et al. 1995). We use empirical methods to test the relationships between fire frequency and both vegetation types and environmental gradients. Our principal objectives are to evaluate the effectiveness of different modeling strategies for extrapolating model results to broad spatial scales, and to examine the sensitivity of model predictions and interpretation to uncertainties in the databases. In the process, we develop fire frequency coverages for forested areas of the ICRB. We discuss the applicability of our methods to the problem of modeling coarse-scale fire effects and potential improvements in models and databases that would make broad-scale predictions more accurate.

### METHODS

The ICRB contains those portions of the Columbia River Basin inside the United States east of the crest of the Cascade Mountains in Washington and Oregon, and portions of the Klamath River Basin in California and the Great Basin in Oregon, Utah and Nevada. The ICRB covers more than 58 million ha, 46% of which is in forested vegetation. Elevation of the forested areas ranges from 50 to 3700 m, and mean annual precipitation ranges from 130 to 3500 mm. Agriculture, grazing, and fire suppression are responsible for major changes in vegetation in both forested and non-forested areas during the last 50-60 years (Hann et al. 1997).

The ICRB Landscape Assessment (Hann et al. 1997) and simulation modeling efforts to predict future vegetation (Keane et al. 1996b) produced a wealth of GIS coverages, which provided a geographic template for our model predictions, and a source of predictor variables. Our response variable was fire frequency, expressed as the expected (mean) fire return interval (FRI). From the FHDB, we extracted the following variables for all fire history sites within the ICRB as defined by the maps included in the integrated assessment: 1) fire return interval (response variable: mean = 50.7 years, range = 6-419 years), and 2) elevation and geographic coordinates from an Albers projection (predictor variables).

The fire history data vary in quality. Most are not accurately crossdated, and more than half of the reconstructions use fewer than 10 trees. Also, beginning and ending dates vary, admitting possible confounding effects from the effects of different climatic regimes and human activities. The methods employed for calculating fire frequency also differ, and include point estimates, composite fire intervals (CFIs), and natural fire rotation (NFR), or fire cycle computation (Agee 1993, Johnson and Gutsell 1994).

We developed two model databases; the first had the advantage of larger sample size, and the second the



1984). A particular advantage of tree-based models is that they can capture non-additive behavior and complex interactions between variables (Clark and Pregibon 1992). Our tree-based models were built from the same model databases as the regression models. We used an adaptive estimation method (Breiman et al. 1984) to minimize the complexity of the model (number of branches and nodes) without sacrificing goodness-of-fit, and then used a cost-complexity measure derived by Breiman et al. (1984) to prune the tree. Predictor variables on which there were no partitions in the final pruned model were thus eliminated.

For each variable in the final (tree-based or regression) models, we created a raster coverage (GRID – 1-km resolution) with data values only at forested pixels. We used the tree-based and linear regression models to predict the FRI for the new data (all forested pixels within the ICRB). We then created four raster coverages of predicted FRIs.

We used standard diagnostics, bootstrap estimates of prediction error (Efron and Tibshirani 1993) for the regression models, and 10-fold cross-validation (Venables and Ripley 1994) for the tree-based models. Because the purpose of our model was to extrapolate local relationships to the regional scale, we also produced statistical and graphical summaries of model predictions at the 1-km scale. We examined the distribution of predicted FRIs from both models for each vegetation type for obvious anomalies, using the output maps and histograms of FRIs for each vegetation type. For example, we expected FRIs to be positively correlated with latitude and elevation, and to observe differences between types in mean and range. We also expected that most predicted FRI distributions for vegetation types would not display major discontinuities or distinctly bimodal patterns. This partly qualitative procedure suggested which of the models would be more robust to extrapolation.

We expected the principal source of error in both the full and reduced datasets to be the lack of accurate crossdating for many of the reconstructions that used fire scars. From rough calculations on simulated increment cores and discussions with fire ecologists, we concluded that a typical error would be to underestimate, or less commonly overestimate, fire frequency by a factor of two. To estimate the effects of this and similar errors on the parameters and broad-scale behavior of the models, we simulated a correction factor that could be applied to non-crossdated FRI estimates in the fire history database to account for potential crossdating errors. This correction factor was a ran-

dom variable, and was calculated differently in two scenarios.

Scenario I – “lumpers.” This scenario assumes that estimates of FRI will be high because researchers would tend to adjust fire dates from different samples to be more synchronized. After some experimentation, we selected the following correction factor:

$$K_1 = 1 / [1 + 1/(U_{0,10})^{1/2}]$$

where  $K_1$  is the correction factor, and  $U_{0,10}$  is a uniform random variate on the interval (0,10). The “corrected” values of FRI (original value multiplied by  $K_1$ ) have a maximum (and mode) of  $.76 \times \text{original}$  and a mean approximately  $.5 \times \text{original}$ .

Scenario II – “lumpers” and “splitters.” This scenario assumes that errors will be equally likely on either side of the original. The correction factor is:

$$K_2 = \Gamma_{20,10} \text{ or } 1/\Gamma_{20,10} \text{ with probability } = .5$$

where  $\Gamma_{20,10}$  is a gamma random variate with shape parameter = 20, rate parameter = 10, expected value = 2, and skewness  $\approx 0.44$ . Corrected FRIs would be thus higher or lower than originals with equal probability.

Because the correction factor is a random variable in both scenarios, every correction is unique, and every realization of a set of corrections applied to the FRIs is also unique. We applied each scenario 25 times to both regression models and both tree-based models, correcting only FRIs under 30 years from non-crossdated fire scars (Heyerdahl et al. 1995), and storing the parameter estimates and their p-values, the fitted values, and  $R^2$  (regression models) and the proportional reduction in deviance (tree-based models). We compared these data to output from the original four models, and randomly selected realizations (of regression models only) to compare to model predictions at the regional scale.

## RESULTS

For the full data, the best multiple regression model uses four predictor variables (Table 1), is highly significant ( $n=182$ ,  $p < 0.0001$ ), and has reasonable explanatory power ( $R^2 = 0.44$ ). The model for the reduced data uses three predictors (Table 1), is also highly significant ( $n=87$ ,  $p < 0.0001$ ), and has better explanatory power ( $R^2 = 0.57$ ). Signs of coefficients, except for interaction terms, are positive, thus an increase in

Model	Coefficient	Value (SE)	Pr(> t )
Full data	Intercept	-2.118 (0.823)	0.0109
	$\log_e(\text{COV})$	0.153 (0.065)	0.0219
	AlbersN	3.082e-6 (4.792e-7)	<0.0001
	Summer precip.	1.242e-2 (3.275e-3)	0.0002
	Elevation	2.755e-3 (5.257e-4)	<0.0001
	Precip./elev.	-9.726e-6 (2.418e-6)	<0.0001
Reduced data	Intercept	-13.915 (4.399)	0.0022
	AlbersN	8.173e-6 (2.851e-6)	0.0053
	Summer precip.	0.081 (0.017)	<0.0001
	Elevation	9.802e-3 (2.824e-3)	0.0008
	Precip./elev.	-4.679e-5 (1.298e-5)	0.0005

**Table 1. Parameter estimates for the regression models. Response variable in full data uses a  $\log_e$  transform, and reduced data uses a square root transform.**

summer precipitation, latitude, elevation, or the numerical value of COV increases predicted FRI. The range of fitted values for FRI is 8-87 years for the full data, and 3-124 years for the reduced data. The correlation (Pearson's R) between fitted values for the two regression models (on sites common to both) was .95.

The tree-based model for the full data produced 16 distinct predicted values, ranging from 8 to 131 years. The model uses the same four variables as the full data regression model. The primary partition is on AlbersN, which accounted for 49% of the total reduction in deviance. The number of sites represented by terminal nodes ranges from 5 to 26. For the reduced data, there are only 10 distinct predicted values, ranging from 11-150 years. The primary partition is also on AlbersN, accounting for 70% of the total reduction in deviance. The number of sites represented by terminal nodes ranges from 5 to 25. Proportional reduction in deviance from both tree-based models (roughly equivalent to  $R^2$ ) is 0.77; hence, they have greater explanatory power, in the statistical sense, than the regression models.

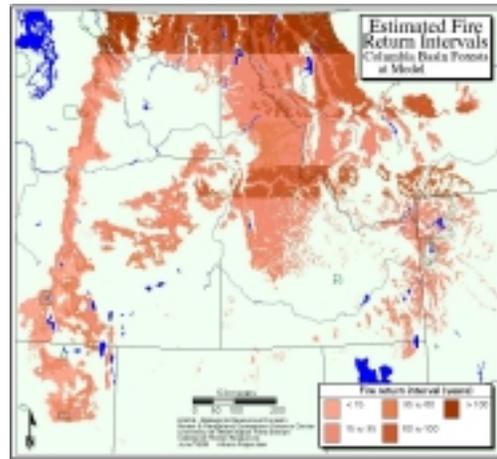
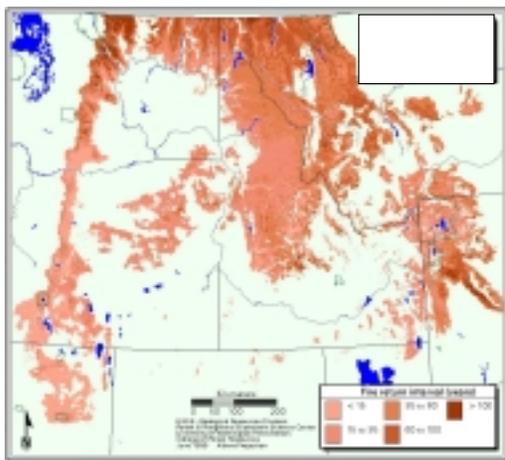
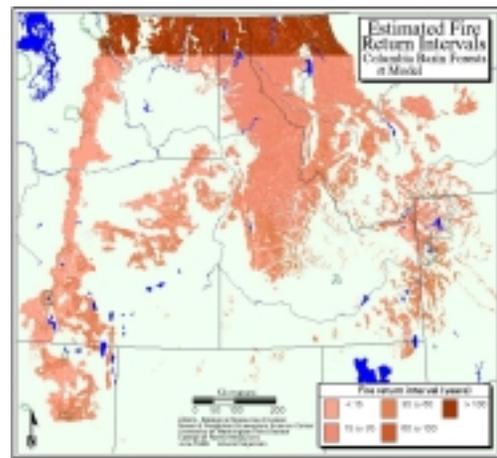
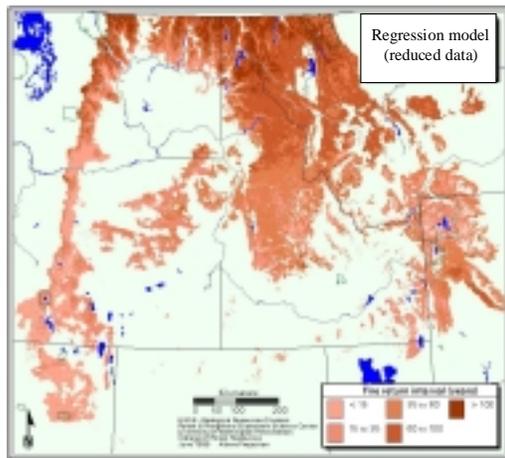
Parameter estimates in the regression models changed little (maximum change less than 1% for any parameter) in either Scenario I or Scenario II. For the full data model, the highest fitted values from Scenario I were slightly lower on average (10%) than for the "true" model. For the reduced data model, the lowest fitted values from Scenario II were 50% lower on average than for the true model. Other extrema of fitted values differed less than 1% from the true models.

The tree-based models changed very little in either scenario. Primary partitions remained on AlbersN, PRDs changed only 1-5%, and no major structural changes occurred. Fitted values at terminal nodes changed less than 10%, and splits on the predictors were consistent.

The total number of regional-scale predictions from the models is three orders of magnitude greater than the number of sample sites. The regional predictions cover a larger elevational range (49-3713 m) than the model database (727-2550 m). There are eight COVH types in the regional (forested) coverage that were not represented in the model database, although these account for less than 8% of the total pixels. Predictions of FRI from the regression models range from 1 to 375 years at the regional scale for the full data model, and 2 to 290 years for the reduced data model. Predictions from the tree-based models are restricted to the 16 (full data) and 10 (reduced data) discrete values at the nodes of the respective trees.

Viewed regionally, predictions from the regression models reveal latitudinal gradients (Figure 2). The gradient is the dominant feature of the reduced data model (Figure 2a); the full data model predicts that the longest FRIs will be in the northern Cascade Mountains, Washington, the Wind River Mountains, Wyoming, and in the northwestern corner of Montana (Figure 2b). Predictions from the tree-based models display distinct horizontal bands in addition to the latitudinal gradient (Figure 3). These bands do not correspond to known biotic or abiotic gradients and are artifacts of the dominance of AlbersN in the partitioning process and of the limited number of unique predicted values (16 and 10).

At the regional scale, the predicted FRIs from realizations in the sensitivity analysis (RSAs) closely track those from the corresponding regression model. Except for a few extreme outliers (< 0.1% of pixels), differences ("true" model - RSA) are less than 10 years for all comparisons. Proportional differences are much greater in cover types predicted to have short FRIs. As



**Figure 2. Output maps of predicted fire return intervals from the regression models, collected into five categories.**

expected, RSAs from Scenario I consistently predict shorter FRIs in ponderosa pine systems (32% of total pixels) than the regression models, because in the “lumpers” scenario, FRIs for these sites in the fire history database were assumed to have been overestimated. However, this consistent bias is not apparent for “Interior Douglas-fir”, the other common vegetation type for which many FRIs were reduced in the RSAs, Scenario I.

## DISCUSSION

The models reveal highly significant relationships between fire frequency and the predictor variables. Because estimates of fire frequency are necessary for modeling fire effects and succession, the data represented by the output maps represent a wealth of new information, which will complement the ICRB assess-

**Figure 3. Output maps of predicted fire return intervals from the tree-based models, collected into five categories.**

ment and assist coarse-scale modeling efforts in the region. In contrast to the coverages from the ICRB assessment, which delineate five broad ranges of fire frequency (Morgan et al. 1996), our models produce estimates of fire frequency at the resolution of one year. Also in contrast, the ICRB models assigned fire regime classes to cover types (Hann et al. 1997), whereas our models predict fire frequency principally from environmental and geographic variables.

The negative coefficient for the elevation and precipitation interaction in both regression models indicates that at higher elevations, FRI is less strongly correlated with precipitation than at low elevations. For example, the models predict that the differences in FRI between low-elevation ponderosa pine forests (drier) and low-elevation cedar-hemlock forests (wetter) would be proportionally greater than between high-elevation

whitebark pine forests (drier) and high-elevation mountain hemlock forests (wetter).

The tree-based and regression models serve complementary purposes in understanding the relationships among FRI and the predictor variables. The tree-based models explain more variation in the response, but when extrapolated to the regional scale, they produce anomalous results. The inability of tree-based models to predict new values is also a significant drawback in extrapolations of this magnitude. Conversely, the regression models, although they have weaker (statistical) explanatory power at the scale of the model database, provide a simple and robust method of prediction. Thus, although the tree-based models show a better statistical fit, only the regression models are suitable for broad-scale predictions.

Of the two regression models, the reduced data model has a more homogeneous response variable and produces more homogeneous predictions at the regional scale (Figure 2a). The full data model incorporates vegetation, albeit weakly, and isolates geographic areas of long FRIs independent of the latitudinal gradient (Figure 2b). Sensitivity analyses suggest that errors in computing the response variable would be slightly greater for the full data, and proportionally greater in systems with short FRIs. Predictions from the reduced data model are probably more accurate in systems with short FRIs, because errors from lack of crossdating are less severe, but estimates of FRI for low fire frequency systems are probably better from the full data model. The ICRB assessment model does not appear to produce any latitudinal gradient, probably because it is focused on vegetation types and broad classes of fire frequency. Our full data model displays both the broad gradient and isolated areas of high FRIs, while our reduced data model displays only environmental and geographic gradients.

The models could be improved by additional fire history information for the ICRB, particularly if data collection and interval estimation were standardized, providing better confidence to FRI estimates in model databases. Crossdating all tree-ring records would significantly improve the accuracy of the response variable. Some fire history studies should be initiated specifically to improve regional scale models – that is, with non-local objectives. A sparse grid of fire history sites, while not providing detailed local information, could include more vegetation types and be amenable to rigorous quantitative methods of spatial aggregation (Dutilleul 1993, Legendre 1993, Rossi et al. 1993).

The models have implications for local fire management, simulation modeling, and ecological scale concepts (Peterson and Parker 1998). The output maps provide a coarse-scale component to local databases and can be helpful in estimating characteristics of local fire regimes, particularly in the absence of local fire history information. The extent to which the contagious nature of fire can be incorporated into a coarse-scale fire frequency model is unknown. We were unable to discern spatial autocorrelation among the sites in the existing fire history database, but new sampling designs for fire history reconstructions might address this problem. For example, if grids were established to measure point FRIs in different systems, autocorrelation structure could be more easily determined, and interpolated values could be compared to predictions from a model that assumes independence.

Ecosystem management is being applied under hierarchical frameworks at multiple spatial scales. Informed decisions are needed at increasingly broad spatial scales, but in most cases, detailed quantitative data are not, and may never be, available (McKenzie et al. 1996, Lertzman and Fall 1998, McKenzie 1998). Integration of existing databases, complementary use of qualitative and quantitative methods, resolution of scale incompatibilities in spatial data, and more efficient approaches to data collection will improve our understanding of broad-scale interactions among fire, vegetation, and the physical environment.

## REFERENCES

- Agee, J. K. 1993. *Fire Ecology of Pacific Northwest Forests*. Island Press, Washington DC.
- Bates, D. M. and M. J. Lindstrom. 1986. Nonlinear least squares with conditionally linear parameters. *American Statistical Association - Proceedings of the Statistical Computing Section*, pp. 152-157.
- Breiman, L., J. H. Friedman, R. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Wadsworth International Group, Belmont, CA.
- Clark, L. A., and D. Pregibon. 1992. Tree-based models. Chapter 9 in Chambers, J. M. and T. J. Hastie. *Statistical Models in S*. Wadsworth and Brooks, Pacific Grove, CA.
- Daly, C., R. P. Neilson, and D. L. Phillips. 1994. A statistical-topographical model for estimating climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* 33:140-158.

- Dutilleul, P. 1993. Spatial heterogeneity and the design of ecological field experiments. *Ecology* 74:1646-1658.
- Efron, B. and R. J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman and Hall, New York, NY.
- Hann, W. J., J. L. Jones, M. G. Karl, P. F. Hessburg, R. E. Keane, D. G. Long, J. P. Menakis, C. H. McNicoll, S. G. Leonard, R. A. Gravenmeir, and B. G. Smith. 1997. Landscape dynamics of the Basin. In Quigley, T. M. and S. J. Arbelbide, eds. *An assessment of ecosystem components in the Interior Columbia Basin including portions of the Klamath and Great Basins: vol. 2*. USDA Forest Service General Technical Report PNW-GTR-405. Pacific Northwest Research Station, Portland, OR.
- Heyerdahl, E. K., D. Berry, and J. K. Agee. 1995. Fire history database of the western United States. EPA Office of Research and Development: Report EPA/600/R-96/081. Washington, DC.
- Johnson, E. A., and S. L. Gutsell. 1994. Fire frequency models, methods and interpretations. *Advances in Ecological Research* 25:239-287.
- Keane, R. E., D. G. Long, J. P. Menakis, W. J. Hann and C. D. Bevins. 1996b. Simulating coarse-scale vegetation dynamics using the Columbia River Basin Succession Model CRBSUM. USDA Forest Service General Technical Report INT-GTR-340. Rocky Mountain Research Station, Fort Collins, CO.
- Legendre, P. 1993. Spatial autocorrelation: trouble or new paradigm. *Ecology* 74:1659-1673.
- Lertzman, K., and J. Fall. 1998. From forest stands to landscapes: spatial scales and the role of disturbances. Pages 339-368 in Peterson, D. L. and V. T. Parker. *Ecological Scale: Theory and Applications*. Columbia University Press, New York, NY.
- McKenzie, D. 1998. Fire, vegetation, and scale: toward optimal models for the Pacific Northwest. *Northwest Science* (special issue) 72:49-65.
- McKenzie, D., D. L. Peterson, and E. Alvarado. 1996. Extrapolation problems in modeling fire effects at large spatial scales: a review. *International Journal of Wildland Fire* 6:165-176.
- Morgan, P., S. C. Bunting, A. E. Black, T. Merrill, and S. Barrett. 1996. Fire regimes in the interior Columbia River basin: past and present. Report on file at USDA Forest Service Intermountain Fire Sciences Laboratory, Rocky Mountain Research Station, Missoula, MT.
- Rossi, R. E., P. W. Borth, and J. J. Tollefson. 1993. Stochastic simulation for characterizing ecological spatial patterns and appraising risk. *Ecological Applications* 3:719-735.
- Venables, W. N. and B. D. Ripley. 1994. *Modern Applied Statistics with Splus*. Springer-Verlag, New York, NY.