A Multivariate Decision Tree Analysis of Biophysical Factors in Tropical Forest Fire Occurrence

Rey S. Ofren and Edward Harvey

Abstract.—A multivariate decision tree model was used to quantify the relative importance of complex hierarchical relationships between biophysical variables and the occurrence of tropical forest fires. The study site is the Huai Kha Khaeng wildlife sanctuary, a World Heritage Site in northwestern Thailand where annual fires are common and particularly destructive. Thematic layers of several biophysical variables were combined in a GIS with field measurements of fuel loading and stand physiognomy. Canopy vegetation (NDVI), rainfall, geology, elevation, and forest type explain most of the variation in burned surface across the mountainous landscape. Pixels with normalized vegetation difference index values of 0.7284 best discriminated burned and non-burned areas. Less important decision tree model rules identified fire occurrence thresholds for annual rainfall of 1,285 mm, elevation of 700 m, and distinguished between moist evergreen and dry deciduous formations. A map of the sanctuary was prepared using GIS to illustrate spatial variation in fire hazard probabilities predicted from the decision tree model.

Forest fire is a common event in many human-altered ecosystems in the tropics (Malingreau 1990). However, land managers and policymakers in many tropical countries are not adequately prepared to cope with the increasing forest fire occurrences. Limited resources and financial constraints (Goldammer and Manan 1996) encourage continued development of effective fire management support tools. In forest fire management, there is a recognized need for longer term studies of the relationships between fire regime, landscape heterogeneity, and human land-use (Turner and Romme 1994). Effective management of wildfires requires a thorough understanding of the spatial distribution of vegetation and human land-use, as well as their complex spatial relationships (Chou 1992).

The ecology of forest fires is complex (Underwood and Christensen 1981), but fires tend to be distributed regularly in space and time within fire regimes specific to a vegetation type, topography, and climate (Pyne 1982). Spatial statistical analysis is an effective approach to quantify the environmental interactions and variability of fire at different levels of spatial and ecological organization (Bailey 1994).

The study area is the Huai Kha Khaeng Wildlife Sanctuary (HKK) in northwestern Thailand. This sanctuary is unsurpassed in mainland Southeast Asia in terms of both size and biological quality, leading to its declaration as a World Heritage Site in 1991. It supports extensive stands of rare dry tropical forests and deciduous riparian forest, and a high diversity of wildlife. It also contains a nearly complete representation of the fauna of central Indochina, enriched by some Sundaic and Burmese taxa. Forests at HKK are composed of a mosaic of four different forest types: dry dipterocarp, mixed deciduous, dry evergreen, and hill evergreen. During the wet season, the deciduous forests provide grass and bamboo for grazing wildlife. In the dry season, evergreen forests supply browse and year round shelter.

The average annual rainfall in HKK is approximately 1,500 mm. During the pronounced dry season, two of the four forest types, the dry dipterocarp and the mixed deciduous, are prone to forest fires. Both have a thick understory dominated by grasses and bamboo that can be extremely dry and flammable. However, fire is widely believed to be detrimental to the ecological integrity of the HKK forest mosaic. Seedlings of the evergreen forest are weakened or destroyed when ground fires sweep through the understory because these species have not developed a resistance to fire. Conversely, plant species common in the fire-adapted dry dipterocarp forest are provided with new opportunities for colonization along the edges of the burned evergreen forests (Bunyavejchewin and Baker 1996).

The main objective of this paper is to quantify in a multivariate decision tree model the environmental factors associated with fire occurrence. This statistical model is used to produce a fire hazard probability map of HKK.

SPATIAL MODELING OF FOREST FIRE

The need to simulate the spatial effect of wildfires at landscape scales has resulted in the development of a

r.ofren@auckland.ac.nz, Geography Department, University of Auckland, New Zealand, and eharvey@sfu.ca, Simon Fraser University, Burnaby, B.C., Canada.

variety of statistical and analytical models that differ in their complexity and data requirements. These include non-spatial thermodynamic models, and spatially explicit methods from percolation theory, cellular automata, and expert systems. The following specific issues formed the foundation for this paper:

- spatial dimension of fire peculiar to a heterogeneous environment (Campbell *et al.* 1995)
- interaction of variables to identify patterns and processes (Chou 1992),
- probability of occurrence (Vasconselos and Guertin 1992),
- appropriateness of model in tropical regions (Pickford *et al.* 1992), and
- scarcity of local research directed to tropical fire management (Goldammer and Penafiel 1990).

As such, this study incorporates the strength of the integration capabilities of GIS and multivariate statistical modeling to handle spatial data of various forms in analyzing behavior pertinent to the occurrence of fire. Once the susceptibility of the study area to forest fire has been mapped together with a knowledge of its factors' interrelationship, it can be used in forest management as an early warning tool in helping to refine firefighting strategies, direct human land-use, and improve fire prevention and control methods.

MULTIVARIATE DECISION TREE MODEL OF FIRE OCCURRENCE

The statistical methods in this study follow a data-driven (exploratory data analysis) modeling approach in an attempt to identify the variables most strongly associated with fire occurrence. Known locations of previous fire events were used as the baseline information to analyze the biophysical variables associated with fire occurrence and to develop the decision tree model. Multivariate decision tree modeling was adopted because some environmental variables are categorical, others are continuous, and many variables are hierarchically interrelated. These properties are difficult to accommodate in standard linear regression methods.

Decision tree modeling has its origins in artificial intelligence research where the aim was to produce a system that could identify existing patterns and recognize similar patterns in the future (Quinlan 1986 as cited by Moore *et al.* 1991). The hierarchical structure provides an efficient means for sorting observations into classes because at each step, alternative path and/or class assignments are eliminated.

Decision tree models use predictor variables to sequentially split the sample into smaller groups with more pure class membership. The rules are determined by a procedure known as recursive partitioning. This involves splitting the data into subsets based on the first predictor and then identifying entirely different relationships with other predictors in the two resulting subsets. In this manner, the tree model attempts to construct a binary decision tree by selecting the most useful variables from a set of candidate predictor variables (Baker 1993). It provides information about the relative importance and hierarchical interrelationships of variables related to fire occurrence. In this connection, it was adopted to identify environmental variables or groups of variables associated with the occurrence of forest fire.

Data Preparation

Biophysical factors related to the occurrence of fire were drawn from maps of topography, vegetation type, climate, and geology, and field measurements. Maps of the distribution of previous fires were the main source of spatial information of fire occurrence. Field measurements yielded data about fuel type, green weight load, fuel bed depth, and variables related to vegetation physiognomy. Laboratory measurements generated data about fuel moisture content, fuel type proportion, and the estimated fuel bed volume and density. Additional tree canopy data expressed in normalized vegetation difference index (NDVI) were extracted from Landsat Thematic Mapper (TM) data. A Digital Elevation Model (DEM) was used to calculate slope, elevation, and aspect variables. Field sampled fuel and tree physiognomy variables were interpolated over the entire study area. In this respect, topographic gradients and forest cover were used as a basis in interpolating various levels of fuel and tree structures.

The Decision Tree Model

Stratified random sampling was used to select a training sample (Lillesand and Keifer 1987). Spatial variation in forest cover is considered to be the most appropriate variable to stratify the landscape because the type and amount of fuel vary among vegetation formations (Clarke and Olsen 1996). A total of 1,000 randomly selected training pixels were sampled, representing 0.03 percent of the total number of pixels in the study area. Eighty-two percent of the sample was in unburned areas, and 180 training samples were derived from burned sites.

Fourteen biophysical variables were used as predictors to model the binary burned and unburned variable. Slope exposure, forest, and geological variables are categorical variables, while elevation, slope, fuel bed density and load, fuel moisture content, tree density and height, basal area, rainfall distribution, and vegetation index (NDVI) are continuous variables.

The *tree* function in *S-PLUS v3.4* software (Statistical Science 1993*)* was used to create a decision tree model for classifying fire occurrence as a binary categorical event with an associated probability value:

$$tree(response \sim predictor_1 + predictor_2 + \ldots + predictor_n)$$

Decision tree models capture interactions between variables without explicit specification. This *tree* function also automatically distinguishes between regression and classification trees according to whether the response variable is continuous or categorical, respectively.

Burned and unburned pixels were classified according to the level of the predicted probability. The full tree model yielded 52 terminal nodes representing 52 classification rules. The tree revealed a residual mean deviance (RMD) of 0.2278, equivalent to the discrepancy between the observed and model fitted values. The misclassification error of the model was calculated as 0.053 (94 out of 1,000 observed values did not fall into the terminal leaves). This is an estimate of the predictive skill of the model. The least important predictors (tree canopy closure, average tree height, and basal area) were excluded in the final tree model.

The first, most important rule in the hierarchical model of fire occurrence is the vegetation index (NDVI), then rainfall (second level left split), and then elevation (second level right split) (fig. 1). Pixels in the rainfall subset were further subdivided according to forest (fourth level left split) and geological types (fourth level right split). The latter branch produced the largest number of decision rules.

The position of a predictor variable at the major branches of the tree suggests its dominance in explaining fire occurrence, compared with another variables. The predictor variable with the highest deviance determines the splitting of the data set. Likewise, comparing the deviance of the dominant predictor at a major split with the root node deviance is a simple estimate of the proportion of information explained at each branch of the model. Two parameters—*residual mean deviance (RMD)* and *misclassification error (MCE)*—were also used to quantify the level of improvement when each variable was incorporated into the model generation (fig. 2). In a stepwise manner, a tree model with a particular predictor variable omitted was compared with a tree model based on all variables.

Assessment of Predictors and Attributes

Model Improvement

The overall effect of a predictor variable on the tree model performance is particularly evident in the terminal nodes. A significant difference in *residual mean deviance* and *misclassification error* parameters signifies a variable important for the structure of the full tree model of fire occurrence.

The inclusion of the annual precipitation variable in the full tree model has the most influence on the model structure. Including this rainfall variable reduces the average deviance of the residual by almost 40 percent and increases the predictive ability of the model by 17 percent. This corroborates observations by Malingreau (1990), who asserts that annual rainfall is the most important factor in fire occurrence and frequency in the tropics, due to variations in fuel moisture and loading. Annual precipitation determines the succession of wetness and dryness periods, green biomass accumulation and fuel loading, and soil microorganism activity and litter decomposition.

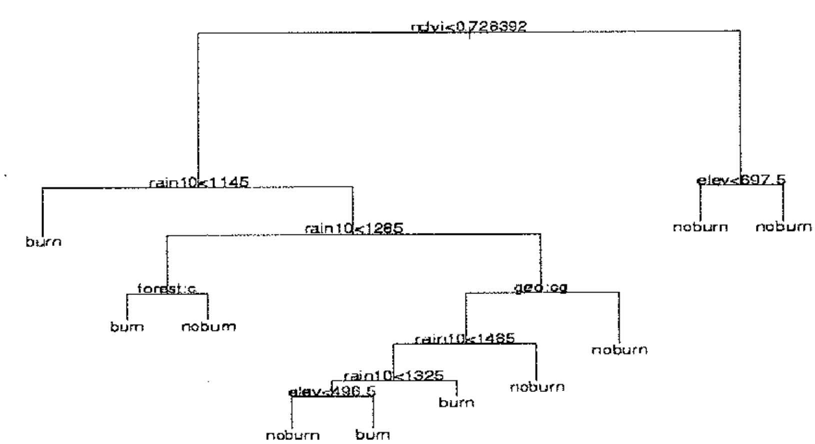

Figure 1.—*Classification tree for fire occurrence (pruned at 10 terminal nodes).*

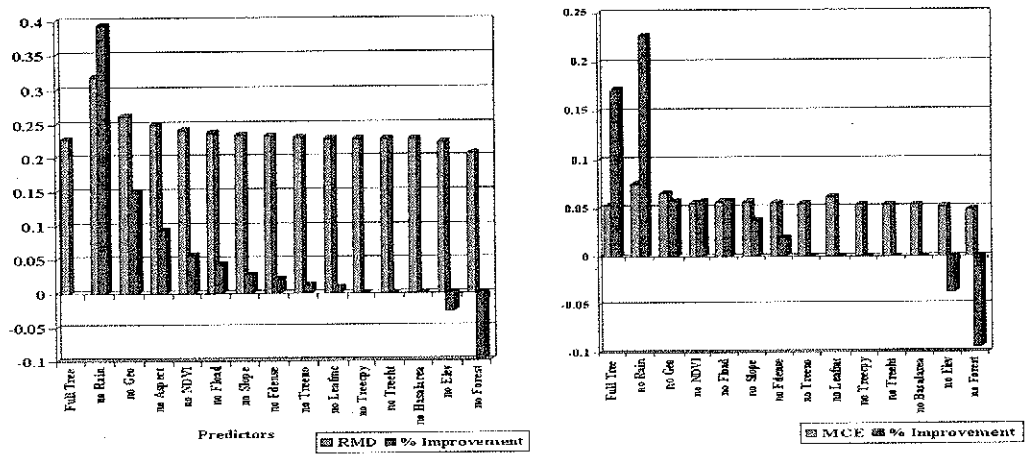

Figure 2.—*Predictors' improvement based on residual mean deviance and misclassification error.*

Geological substrate is the second most important variable in the fire hazard tree model. Parent material influences vegetation indirectly by determining soil nutrient composition (Moore *et al.* 1991). The occurrence of fire among deciduous forests is associated with decreasing soil nutrients and decreasing soil moisture. In the Huia Kha Khaeng wildlife sanctuary, there is a strong association (Pr(2) < 0.0001) between the distribution of forest type and geology. Including the geology variable decreases the average deviance of the residuals by 15 percent and increases prediction by 23 percent.

Aspect has more fire discriminating ability in the decision tree model than the slope variable. Cheney (1981) stressed that changes in fire behavior, even with increasing elevation, are largely associated with changes in exposure. In particular, slope facing the predominant wind direction and solar heating have a considerable effect on fuel moisture content.

The amount of dead fuel load available for burning is the most important fuel-related variable in the decision tree model, other than fuel bed density and dry leaf moisture content. Previous ANOVA statistical analyses revealed that fuel loading is different (Pr(F) = 0.0001) between burned and unburned sites. However, dry leaf moisture content is no different in burned and unburned sites (Pr(F) = 0.619). Stand physiognomy, vegetation index (NDVI), elevation, and slope have a weak correlation with moisture content. Stand physiognomy variables (tree canopy closure, average tree height, and basal area) did not improve the full tree model. They are strongly interrelated and highly correlated with fuel variables, NDVI, and forest type.

Excluding forest type and elevation variables from the model improved the model by 9.7 percent and 2.6 percent, respectively. An examination of the decision tree structure in the next section supports the insignificant contribution of forest type and elevation. In contrast to stand physiognomy variables, both forest type and elevation were retained in the full tree model.

Model Explanation

NDVI did not significantly improve the predictive ability of the decision tree model, but it was an important variable to explain the final model (fig. 3). The appearance of NDVI in the main split of the tree model indicates that burnt and unburnt cases can be best distinguished by a particular vegetation index value (i.e., 0.7284). The relevance of vegetation index in classifying fire occurrence lies in distinguishing evergreen forest against the deciduous type. The distinction between the two major forest types eventually segregates soil parent materials of low (e.g., *gr* – red yellow podzolic soil) or high in nutrient potential and elevation gradient below and above 600 m. Many variables related to NDVI also appear in the deepest branches of the model. Rainfall and elevation variables are the next most important factors in explaining the occurrence of burnt areas. Rainfall accounts for 67 percent of the amount of information in the model compared with 16 percent for elevation.

The rainfall-dominated branch separates into forest and geology subdivisions at 1,285 mm rain per annum. In an area lying in rainshadow where annual precipitation is less than 1,285 mm, the occurrence of fire is primarily dictated by forest type, particularly of deciduous formation. However, the occurrence of fire in the wetter region of the sanctuary (i.e., south to southwest) is no longer governed by forest type but by the complex interactions between geological material, topographic gradients, and fuel variables. The data subset determined by geology

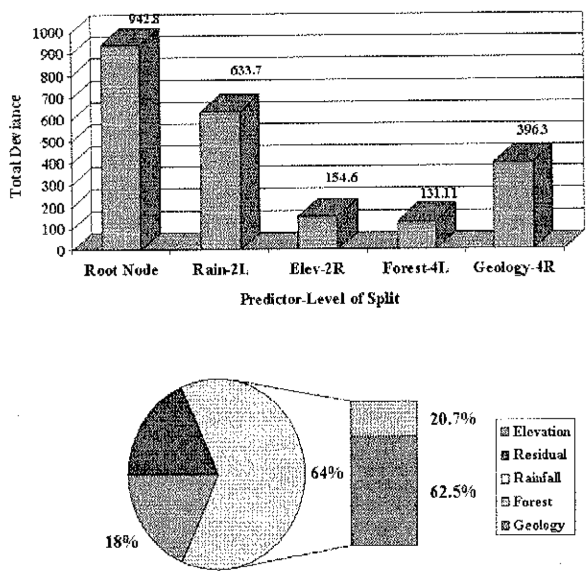

Figure 3.—*Amount and proportion of deviance explained by major predictors in the decision tree model.*

explained 42 percent (62.5 percent of the rainfall subset contribution) of the deviance for the whole model. On the other third branching level, forest type explained 14 percent (20.7 percent of the rainfall subset contribution) of the total deviance. The geology-dominated split exhibits deeper subbranches that accommodate a large combination of rules. Therefore, the main predictor variables can be ordered according to the proportion of total model deviance they explain: vegetation index (NDVI), rainfall, geology, elevation, and forest type.

Several attributes of forest type and elevation appear in distant branches of the model. In contrast to geology, the relatively small contribution of elevation and forest type is attributed to fire distribution, mostly in the moist evergreen forest formation. At higher elevations where moist evergreen forests are abundant, training samples did not provide a distinctive representation of burned and unburnt sites. They dominate unburned pixels at higher altitudes.

SPATIAL IMPLEMENTION OF FIRE HAZARD MODEL

An important use of a decision tree-based model is to predict the value of the response variable for a known set of predictor variables. In this case, an interface between the statistical decision tree model and GIS extends the decision tree into a fire hazard probability map. Decision tree rules were converted into Arc Macro Language (AML) code to map the predicted fire probability for each pixel. The function *docell* in the GRID module of *ARC/INFO* works on a cell-by-cell basis, interpreting the hierarchical rules as a sequence of spatial operations (ESRI 1997). The specific number assigned by the decision tree model to a given tree node represented the navigation process in the tree. Thematic layers of each predictor variable were combined, and a map of predicted probability of fire occurrence was created (see figure 4).

Table 1.—*Discriminatory power of predictors at first major split in decision tree model*

Predictor	Maximum deviance	Attribute
NDVI	154.40	0.7284
Elevation	142.56	593 meters
Forest	105.64	evergreen/dry dipterocarp, mixed deciduous with and without bamboo
Rain	80.26	1,145 millimeter
Canopy closure (Treecpy)	76.60	70.0%
Geology	76.41	O,Q1,gr/C,E,P,SD,pE
Fuel load (Fload)	72.32	8.32 ton/hectare
Tree height (Treeht)	63.33	33.5 meter
Basal area	59.19	48.19 m^2
Fuel bed density (Fdense)	58.22	0.069 g/m^3
Dry leaf moisture content (Leafmc)	57.16	12.76 %
Tree density (Treeno)	52.49	27 trees/hectare
Slope	7.56	10.5 %
Aspect	6.98	east,flat,north,northeast,southeast,southwest/northwest, south,west

Using the known fire location map as reference, the modeled fire occurrence produced an overall mapping accuracy of 90 percent. A fire probability threshold of 0.60 was used to distinguish between burned and unburned locations. This probability level produced a higher percentage of mapping improvement against the possible random occurrence of burned and unburned locations.

SUMMARY AND CONCLUSIONS

Rainfall, geology, aspect, and NDVI variables significantly improved the predictive ability of the decision tree model. NDVI, rainfall, geology, forest type, and elevation explain most of the processes underlying the hierarchical rules that distinguish between varying probability levels of fire occurrence. NDVI and soil parent materials are highly correlated with forest type, so their importance in the model was apparent.

The overall importance of each environmental variable for improving the predictive ability of the decision tree model was assessed from *residual mean deviance* and *misclassification error* parameters. Vegetation index (NDVI) was ranked fourth after aspect, geology, and rainfall in terms of model improvement, but it appeared to have the highest discriminatory power to distinguish burnt and unburnt pixels.

The insignificant contribution of forest type and elevation in the overall improvement of the predictive ability is attributed to few pixels of burned area in dense vegetation at higher altitudes. At higher altitudes, fire locations are usually confined to a narrow range of slope exposure. These are the manifestations that the decision tree is robust to outliers that tend to be segregated into distinct separate branches. Fire occurring either in dense type of forest or at higher elevations is considered to be an extreme case, but these predictors are not regarded as redundant. Nevertheless, decision tree modeling eliminates the least important redundant predictors (i.e., tree physiognomy) before growing the tree model. Assessing the overall importance of a particular predictor variable for predicting fire occurrence ignores the complex interactions among variables.

The dominance of climate over fuel variables suggests that fire in HKK may be driven by an extreme climatic condition (Bessie and Johnson 1995). This explains the low correlation ($r = 0.056$) between dry leaf moisture content and stand physiognomy. The dryness of the tree canopy, as represented in NDVI values, integrates the many environmental variables leading to fire situations.

Geographic information systems (GIS) contribute primarily by integrating spatial environmental data with a spatial statistical model and mapping the model predictions. This knowledge-based spatial approach developed predictor-response relationships outside the system and applied them to data in the geographic database. Additional research by the senior author goes beyond the evaluation of fire initiation in point basis but incorporates

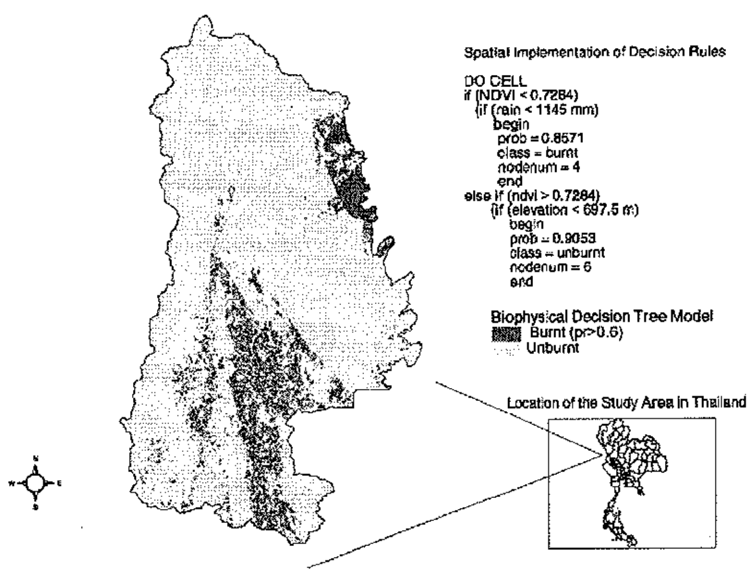

Figure 4.—*Spatial implementation of decision rule model as fire occurrence map.*

the role of neighboring fire. The proximity from the main source of ignition (i.e., human activities) was further integrated towards a more comprehensive model of fire occurrence.

ACKNOWLEDGMENTS

The Royal Thai Forestry Department, the National Research Council of Thailand, and staff of Huai Kha Khaeng Wildlife Sanctuary provided various forms of assistance. Funds from the University of Auckland Graduate Research Fund, the New Zealand Overseas Development Authority, and BIOTROP-GCTE: Southeast Asian Impact Center supported this research. The following people reviewed the manuscript: Jochen Albrecht (Geography Department, University of Auckland) and Kam Suan Pheng (GIS Center, International Rice Research Institute).

LITERATURE CITED

Bailey, T.C. 1994. GIS and spatial analysis for ecological modeling. In: Fotheringham, S.; Rogerson, P., eds. Technical issues in GIS. Taylor and Francis Ltd.

Baker. 1993. Classification and regression tree analysis for assessing hazard of pine caused by *Heterobasidion annosum*. Plant Disease. 77(2): 136-139.

Bessie, W.C.; Johnson, E.A. 1995. The relative importance of fuels and weather on fire behavior in subalpine forests. Ecology. 76(3): 747-762.

Bunyavejchewin, S.; Baker, P.J. 1996. Forest fire and the Huai Kha Khaneg Wildlife Sanctuary (http://www.si.edu/organiza/centers/stri/forest/inside.html).

Campbell, J.; Weinstein, D.; Finney, M. 1995. Forest fire behavior modeling intergating GIS and BEHAVE. In: Proceedings of the 9th Annual symposium on geographic information systems. Canada: GIS World, Inc.

Cheney, N.P. 1981. Fire behavior. In: gill, A.M.; groves, R.H.; Noble, I.R., eds. Fire and the Australian Biota. Canberra: Australian Academy of Science.

Chou, Y.H. 1992. Spatial autocorrelation and weighting functions in the distribution of wildland fires. International Journal of Wildland Fires. 2(3): 162-176.

Clarke, K.C.; Olsen, G. 1996. Refining a cellular automation model of wildlfire propagation and extinction. In: Goodchild, M.F., *et al.*, eds. GIS and environmental modeling: progress and research issues. Colorado: GIS World Books.

ESRI. 1997. ARC/INFO version 7.1.2. California: Environmental Systems Research Institute, Inc.

Goldammer, J.; Manan, S. 1996. Fire in tropical forests. ITTO Tropical Forest Update. (6)1: March.

Goldammer, J.G.; Penafiel, S.R. 1990. Fire in the pine-grassland biomes of tropical and subtropical Asia. In: Goldhammer, J.G., ed. Fire in the tropical Biota: ecosystem processes and global challenges. Third symposium on fire ecology, Volkswagen Foundation. IV, Series: Ecological Studies; v.84.

Lillesand, T.M.; Kiefer, R.W. 1987. Remote sensing and image interpretation. 2d ed. John Wiley and Sons.

Malingreau, J.P. 1990. The contribution of remote sensing to the global monitoring of fires in tropical and subtropical ecosystems. In: Goldhammer, J.G., ed. Fire in the tropical Biota: ecosystem processes and global challenges. Third symposium on fire ecology, Volkswagen Foundation. IV, Series: Ecological Studies; v.84.

Moore, D.M.; Lees, B.G.; Davey, S.M. 1991. A new method for predicting vegetation distributions using decision tree analysis in a geographic information system. Environmental Management. 15(1): 59-71.

Pickford, S.; Suharti, M.; Wibowo, A. 1992. A note on fuelbeds and fire behavior in Alang-alang (*Imperata cylindrica*). International Journal of Wildland Fire. 2(1): 41-46.

Pyne, S.J. 1982. Fire in America: a cultural history of wildland and rural fire. Princeton, NJ: Princeton University Press.

Statistical Science. 1993. S-Plus guide to statistical and mathematical analysis, Version 3.2. Washington: Mathsoft Inc.

Turner and Romme. 1994. Landscape dynamics in crown fire ecosystems. Landscape Ecology. 9(1): 59-77.

Underwood, R.J.; Christensen, P.E.S. 1981. Forest fire management in western Australia. Forest Department of Western Australia.

Vasconselos, M.J.; Guertin, D.P. 1992. FIREMAP: simulation of fire growth with a geographic information system. International Journal of Wildland Fire. 2(2): 87-96.