UNIVERSITY OF ALBERTA

PREDICTING HUMAN-CAUSED FOREST FIRE OCCURRENCE IN WHITECOURT FOREST, ALBERTA

BY

Cristina Vega Garcia

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfilment of the requirements for the degree of Master of Science.

DEPARTMENT OF FOREST SCIENCE

Edmonton, Alberta

Spring 1994

ISBN 0-612-11399-X

UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Cristina Vega Garcia

TITLE OF THESIS: Predicting human-caused forest fire occurrence in Whitecourt Forest, Alberta.

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: 1994

Cristina Vega Garcia
c/ Uruguay 11, 4°A
Madrid 28016, SPAIN

26 th Nov. 93

UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled PREDICTING HUMAN-CAUSED FOREST FIRE OCCURRENCE IN WHITECOURT FOREST, ALBERTA submitted by CRISTINA VEGA GARCIA in partial fulfilment of the requirements for the degree of MASTER OF SCIENCE.

Dr. Paul M. Woodard, Supervisor

Dr. Wiktor L. Adamowicz

Bryan S. Lee

Mr. Bryan S. Lee

William E. Phillips

Dr. William E. Phillips

S. J. Titus

Dr. Stephen J. Titus

October 5th, 1993

Dean
Faculty of Graduate Studies and Research
University of Alberta

re: Permission to use copyrighted material

This letter serves as notice that Paul M. Woodard and Bryan S. Lee approve the reproduction, copying, and use of the materials in the papers:

Vega-Garcia, C., P.M. Woodard, and B.S. Lee. 1993. Geographic and temporal factors that seem to explain human-caused fire occurrence in Whitecourt Forest, Alberta. Pages 115-119 in Proceedings of the GIS'93 International Symposium, Vancouver, British Columbia, Canada, February 15-18th, 1993, Vol 1. 591 pp.

and:

Vega-Garcia, C., P.M. Woodard, and B.S. Lee. 1993. Mapping risk of wildfires from human sources of ignition with a GIS. Pages 419-426 in Proceedings of the Thirteenth Annual ESRI User's Conference, Palm Springs, California, USA, May 24-28th, 1993, Vol.1. 608 pp.

by Cristina Vega Garcia in connection with her thesis work.

Edmonton, 25th November 1993

Paul M. Woodard
Professor
Department of Forest Science
University of Alberta

Bryan S. Lee
Project Leader Fire Management Systems
Northern Forestry Centre
Forestry Canada

Fax Memo

Eighth Annual
Symposium on
Geographic
Information
Systems

February 21 – 24, 1994

Vancouver
British Columbia
Canada

GIS

94

REPLY FAX	604-688-1573
REPLY PHONE	604-688-0188

ATTENTION	Cristina Vega
FAX	1-403-435-7359
DATE	October 26, 1993
FROM	Paddy O'Reilly

Re: Permission to Copy

This letter serves as notice that the organizers of the GIS'93 Symposium have approved reproduction, copying and distribution of the paper

"Geographic and temporal factors that seem to explain human-caused fire occurrence in Whitecourt Forest, Alberta" published in the Proceedings of GIS'93 pages 115-119, Volume 1.

by Cristina Vega Garcia in connection with her research work. This permission extends to colleagues of Ms. Vega Garcia as authorized by her.

Paddy O'Reilly
GIS Symposium Director

GIS'94
Symposium Office
Suite 207
1102 Homer Street
Vancouver
BC V6B 2X6
Canada

Phone
(604) 688–0188

Fax
(604) 688–1573

Fax toll free
1–800–661–0044

ESRI

Environmental Systems Research Institute
380 New York Street
Redlands California 92373 USA
Telefax: (909) 793-5953
Telex: 910 332 1317
Telephone: (909) 793-2853

To: Cristina Vega (Person) — Univ of Alberta, Dept of Forest Science (Company)

Edmonton, Alberta, Canada (City, State, or Country)

Fax Number: 403-435-7359 (operator # 403-435-7377)

From: Sheri Ellis

Date: 10-29-93

Time: 12:00 pm

Subject: Publication Permission

Project/Task #: ~~[illegible]~~ 6601/601

Number of pages transmitted (including this cover sheet): 1

Dear Ms. Vega,

You don't need permission from us. The 1993 Publication Permission form you signed gave ESRI "...express permission to use, copy, reproduce, publish, republish, and distribute" your paper. We no longer retain full copyright ownership rights. Feel free to do what you like ~~[illegible]~~ with your paper. Hope to see another paper from you in 1994!

Sheri

Copy To:

Citation:

Vega-Garcia, C. 1994. Predicting human-caused forest fire occurrence in Whitecourt Forest Alberta. M.Sc. Thesis, University of Alberta, Edmonton, Canada. 117 p.

To my parents

ABSTRACT

The present study models the daily human-caused wildfire occurrence in Whitecourt Forest, Alberta, using geographic and temporal variables from the forest environment. The main hypothesis in the study was that the extent and location of fire-producing activities are determined by the state of the forest environment at any given time. The following variables were identified as relevant to the human-caused fire problem: distance to closest road, town, and campsite, elevation, fuel category, land ownership, forest commerciality, and location on a certain forest district, and Fire Weather Index, Initial Spread Index, Build-Up Index, Fine Fuel Moisture Code, Duff Moisture Code, relative humidity, wind speed, and month. These variables were used for building logit models and neural network predictive models.

A binary logit model was successfully developed to predict daily human-caused fire occurrence in eight fire occurrence prediction units (all less than 5,000 km^2) in the Whitecourt Forest. This model provided a binary prediction of fire occurrence for each fire occurrence prediction unit and day within a standard fire season (April to October) using unit area (km^2), Forest District, Build-Up Index, and Initial Spread Index values to compute the probability of fire occurrence. This model correctly classified 79% of the observations used for model building, and correctly predicted 74% of the outcomes in an independent data set not used in the development of the model.

A back-propagation neural network model was developed to predict the daily probability of fire occurrence in the same eight fire occurrence prediction units. As

with the logit model, predictions were limited to days within a standard fire season (April to October), and the prediction was binary fire occurrence for each area and day. The network used the fire occurrence prediction unit area (km^2), Fire Weather Index, and Forest District as inputs. The network was able to correctly classify 81% of the outcomes in the training data set. It also predicted outcomes for a test set not previously used for model building with a 76% accuracy.

The general conclusion of the study was that both logit models and neural network models can be effectively used for human-caused wildfire occurrence prediction. These models captured very well the relationships between several geographic and temporal variables and fire occurrence. Performance of the neural network model for prediction was slightly better that the performance of the logit model, but the improvement was not sufficient to compensate for the higher computational cost involved in the network development. Reluctance by users to adopt "black box" models might also suggest the use of logit models over the neural networks at this point in time.

ACKNOWLEDGEMENTS

I wish to express my sincerest gratitude to all individuals and institutions that have made this study possible.

I am specially indebted to my supervisor, Dr. Paul M. Woodard, for his never-faltering dedication, friendly guidance, and support throughout my program, and for his patience and helpful advice during the writing of this thesis. I am also sincerely grateful to my supervisory committee members: Dr. Wiktor L. Adamowicz (Rural Economy Dept.), Mr. Bryan S. Lee (Forestry Canada), Dr. William E. Phillips (Rural Economy Dept.), and Dr. Stephen J. Titus (Forest Science Dept.), for their helpful suggestions and advice all along.

Special thanks go to Mr. Mag A. Steiestol of the Alberta Forest Service for his expert advice and continuous disposition to help. I am also particularly grateful to the staff of the Northern Forestry Centre, Forestry Canada, for their friendship and support in the last three years, specially: Mr. Kerry R. Anderson, for his helpful comments and assistance; Mr. Scott R. Henderson, for his aid when computers misbehaved; Mr. Richard M. Smith, for his advice on GIS-related matters; Mr. Marty E. Alexander, for his helpful suggestions and contribution of bibliographic material; and Mr. Max Pinedo and Mr. Kelvin Hirsch, for their help on literature citing matters.

Thanks are also due to Mr. Terry Taerum (U of A) for his statistical advice, the Department of Forest Science staff for their assistance, and fellow graduate students for their friendship and support.

I would also like to thank the following institutions: the Ministry of Education and Science of Spain for sponsoring my education at the University of Alberta; Forestry Canada, Northwest Region, for their financial and technical support and the use of their facilities; and the Alberta Forest Service, Forest Protection Branch, for graciously donating the fire data needed for the study.

Most of all, I wish to thank my husband Jorge and all my family for their support, encouragement, and patience during the course of this study.

TABLE of CONTENTS

Chapter IV

Chapter V

Chapter VI

Appendices

LIST of TABLES

LIST of FIGURES

CHAPTER I

The fire occurrence prediction problem: Introduction.

1.1. General

Human-caused forest fires are not random events. Most often, they are located in the proximity to human habitations and transportation corridors. They are believed to follow seasonal, weekly, and even daily patterns (Martell *et al.* 1987, Todd and Kourz 1991).

Experienced fire managers are usually capable of assessing future fire occurrence trends and locations (Cunnigham and Martell 1976, Todd and Kourtz 1991) but rarely are they able to provide daily predictions specific for a well-defined geographic area.

The need for adequate daily predictions of human-caused wildfire occurrence has been long recognized. In Canada, fire occurrence prediction has been considered a national research priority by universities and the Canadian Forestry Service for at least two decades (Lynham 1989). A model capable of predicting wildfires on a daily basis for areas of administrative significance would enable resource managers to more effectively and efficiently deploy fire suppression resources. This would result in a reduction in costs and losses. The development of such a model anywhere faces one great difficulty which is probably the reason why so few are in operation, that is the uncertainty associated to human behaviour with respect to fire.

Human-caused fires usually result from a combination of dangerous conditions in the forested environment and presence of human sources of ignition. Specifically, the environment at any location encompasses the topography (particularly slope, aspect, elevation), fuel (as it is affected by site quality and disturbance history), and weather (which affects plant growth, mortality, biomass, moisture content). All these factors affect the likelihood of fire ignition and determine fire behaviour (Merrill and Alexander 1987). Topographic, fuel, and weather variables can be measured, and the Canadian Forest Fire Weather Index System (Van Wagner 1987), and the Fire Behaviour Prediction System (Forestry Canada Fire Danger Group 1992) provide procedures that enable managers to quantify the effect of these physical environment variables on fire ignition and behaviour. Unlike the forest environment, human sources of ignition are more uncertain. For example, often we do not know how many people are in the forest, how many people use fire, and how safely do they use fire.

The "chance of fire starting as determined by the presence and activity of causative agents" is usually referred to as *risk* (Merrill and Alexander 1987). Data required for risk assessment, such as number of people in an area on a particular day, and the activities they are engaged in, are usually unavailable (Martell *et al.* 1987). This has led researchers to investigate techniques to obtain indirect estimates of risk, or to leave the human risk component out of their prediction models altogether.

1.2. Risk related considerations

A number of studies have been conducted to identify human-related characteristics that would account for variability in fire occurrence rates in different areas in the United States (Cole and Kaufman 1963, Johnson 1968, Christiansen and Folkman 1971, Doolittle 1972). Even though some population variables, such as rural population density, showed positive correlation to fire occurrence, they were usually poor indicators of the number of actual or future fire starts (Doolittle 1972, Altobellis 1983, Donoghue and Main 1985, Doolittle and Donoghue 1991). These approaches to risk estimation did not consider the characteristics or influences of transients in fire occurrence rates, since it was assumed that the majority of fire starters were members of the local population (Christiansen and Folkman 1971, Doolittle 1972). Doolittle (1972) suggested population variables in areas where many fires occur can sometimes indicate low risk, and therefore be misleading, if only a few individuals (arsonists) are setting the fires. Altobellis (1983) further suggested that fires in the southern United States were most likely caused by the activities of a small percentage of the total population. Doolittle (1972) concluded that further study of fire-producing activities and individuals was needed before human risk could be used as an input in any fire danger rating system.

In a study on human behaviour with respect to fire in a wildland situation, Folkman (1977) reported that all types of persons are a potential source of uncontrolled fire, but his results suggested that the activity in which a person was engaged was the major determinant of the risk they represented. This conclusion was supported by Doolittle (1972), who reported that certain activities such as camping or smoking or the operation of machinery, increased the risk of wildfire in forests. Based on the knowledge that wildland fires are commonly related to human activities, managers usually stratify fires by categories for reporting purposes (Main and Haines 1974, Donoghue 1982, Woodard and Niederleitner 1983, Higgins and Ramsey 1992) and try to infer danger levels from these data. Deeming *et al.* (1977) introduced in the U.S. Forest Service's National Fire Danger Rating System (NFDRS) a human-caused fire occurrence prediction model in which risk was estimated from previous records of fire starts and a subjective daily managerial assessment of fire-producing activities levels. This model, which is still in operation, has the common pitfalls of models that rely on subjective assessments.

The present study attempts a new approach to risk estimation based on the above considerations. The main hypothesis in the present study is that *the extent and location of fire-causing human activities are shaped by the state of the forest environment at any given time*. This hypothesis arises from the work that has been done in the recreational field. Recreation choice behaviour has been shown to be related to site attributes. Site-related geographic and temporal variables have been used to describe human use in a particular area. Recreationalists prefer certain attributes in their recreational environment, and for specific activities some attributes are absolutely required. In the same way the environment shapes recreational activities, it also shapes all other activities taking place in the forest, including the fire-causing activities. Access is the best example of a required attribute for all activities, whether it be by land, water or air. Access by manual or mechanical means is a variable that limits the number and distribution of humans as possible ignition sources in the forest. Given reasonable access, variables such as availability of fruits for picking, lakes for trophy fishing, flat spots near water for tenting, or timber of high value for cutting, in addition to the periods of time when such activities are engaged in, play an important role in determining when and where people will use fire in the forest.

This hypothesis implies that geographic and temporal variables, then, can be used to estimate indirectly the risk posed by humans in certain areas, for specific periods of time. Chou *et al.* (1990) found several geographic variables such as vegetation type, location close to road or campsite to be related to human-caused fire occurrence in the San Jacinto Ranger District, San Bernardino National Forest, California, and more importantly, suitable for fire occurrence prediction. An extension of this hypothesis, first proposed by Phillips and Nickey (1978), is that given identical conditions in two forest environments, risk levels should be equal. It is recognized that this simplification of the fire producing process only considers the influence of external stimuli, which if known are easy to measure, in human decision-making and behaviour and leaves out internal/personal motivations or predispositions, which are difficult to evaluate in comparison. Modeling these internal forces has been accomplished in the behavioral sciences, but that is beyond the scope of this work.

This approach to risk estimation, presents the advantage of being equally valid for transient and resident sources of ignition. This hypothesis also recognizes the dual impact that some factors have on fire occurrence. For instance, plant communities can be viewed as fuel types, and as attractors or detractors for certain recreational activities. Also, rain dampens dead woody material and prevents ignition, but also keeps more campers home.

1.3. Successes in predicting daily human-caused fire occurrence

There are few models that attempt to predict wildland fire occurrence on a daily basis. Subjective probability assessments have been used to predict fire occurrence (Cunnigham and Martell 1976), as they have been used to estimate risk (Deeming et al. 1977), but most models are based on statistical analyses of fire occurrence data with respect to weather data, and all are specific to well defined geographic cells or areas.

The first attempt at predicting the probability of human-caused fire occurrence was reported in Crosby (1954). He used linear regression to relate number of fires in Clark National Forest to the ratings of the Central States Danger Meter. Bruce (1963) used the negative binomial distribution to relate daily fire occurrence data from Louisiana and Missouri to fire danger rating class. Deeming *et al.* included in 1977 a human-caused fire occurrence prediction model in the U.S. National Fire Danger Rating System. In this model, fire occurrence predictions are a function of an ignition component (IC) and human risk, which is estimated from historic data and a daily subjective assessment. A multilinear regression approach was used by Haines *et al.* (1983), who related the probability of a fire day in portions of the states of Michigan, Wisconsin, and Pennsylvania to the Ignition Component (IC) of the U.S. National Fire Danger Rating System. Loftsgaarden and Andrews (1992) used a logit model to describe the probability of a fire day in Lolo National Forest, Montana, based on the

Energy Release Component (ERC) of the National Fire Danger Rating System (Deeming *et al.* 1977).

In Canada, Cunnigham and Martell (1973) used the Poisson model to relate the average number of fires per day (λ parameter of the distribution) in Ontario to the Fine Fuels Moisture Code (FFMC), which is one of the outputs of the Canadian Forest Fire Weather Index System, CFFWI, (Van Wagner 1987). This relationship was further studied using a logit model to predict the probability of a fire day (Martell *et al.* 1985, 1987) by season and cause, based on the Codes and Indices in the CFFWI. In a later version of this model, Martell *et al.* (1989) grouped causes of fires into two categories and added periodic variables (trigonometric functions of the Julian date and FFMC) which accounted for seasonal variability, to the variables FFMC and BUI (Van Wagner 1987) to predict the probability of a fire day. A similar approach developed for Ontario by the Petawawa National Forestry Institute (PNFI) resulted in the PEOPLE fire occurrence prediction model. Todd and Kourtz (1991) have tested this model in Quebec. This model uses a gamma distribution to calculate expected daily numbers of fires (λ) using FFMC, DMC (Van Wagner 1987), and wind speed as independent variables (Tithecott 1990b). The authors considered the λ parameter of the Poisson distribution to be a random variable with a gamma distribution, and revised the gamma parameters with a Bayesian process to incorporate recent trends in fire occurrence. Ontario has used the last two models for the past few years. An evaluation of: (1) Martell's model, (2) the PEOPLE model, and (3) an expert system (FUZZY) developed also by PNFI, was conducted in 1989 in Ontario. The results of these tests are reported in Tithecott (1990a, 1990b, 1990c). Davidson (1993) produced probability curves to predict a fire day for Pictou County, Nova Scotia, as a function of the FFMC (Van Wagner 1987) and the subseason. Summaries of other related works have been compiled by Martell and Otukol (1985), Lynham (1991), and Tithecott (1993).

Currently, no models attempt to account for daily differences in the presence and activity of causal agents within prediction units, although Martell *et al.* (1987) did use cause, subseason and day in the week as surrogate variables for the day to day variation in land use activities. These above consideration suggests a need for empirical prediction methods that go beyond the capabilities of current models by incorporating some parameters that account for the human risk factor.

1.4. Objectives

The objective of this study was to build a daily, human-caused, fire-occurrence, prediction model using temporal and weather variables, and some geographic variables capable of describing the arrangement and variation of risk levels and fuels (hazard) in the study area. At this point, the following questions arose:

1. Which variables have a strong association to fire occurrence?

2. What kind of model best describes the relationship between these variables and fire occurrence?

3. Can this relationship be used to predict fire occurrence by location or day with enough accuracy that it would be useful for managerial purposes?

1.5. Study design

The Whitecourt Provincial Forest in Alberta, Canada, (see Appendix A) was selected as the study area because digitized geographic information was available for this Forest. Since information pertaining to geographic characteristics change over time due to the construction of new roads, facilities or other developments, it was important to use a short time period because the accuracy of the recorded geographic information would be more reliable, but the period needed to be long enough to show trends in patterns of fire starts by humans. The 5-year period between 1986-1990 was selected for developing the model, since fire occurrence data in this period were believed to be complete, reliable, and sufficient to show trends in fire starts patterns by humans. The geographic information was also digitized during this period (1986-1987). Both data sets were used in the analysis. By using this rather short time period, errors arising from relating fires to geographic features that were non-existent at the time the fires took place were minimized.

The first objective of this study was to identify the key variables important in predicting human-caused fire starts in the Whitecourt Forest. This information was needed before proceeding to the actual model building phase of this study.

Fire occurrence models were built using these variables and both logistic regression analysis (the logit model) and neural net technology (NeuralWare Inc. 1991)[1]. The reasons for choosing these analytical procedures are presented in detail

[1] The use of trade, firm, or corporation names in this thesis is for ease in reproducing the results reported, and does not constitute endorsement.

in their corresponding chapters. The performance of these two approaches was tested using data from the 1991 and 1992 fire seasons.

This thesis is presented in chapters. Each chapter is designed to stand alone. In Chapter II, various weather, temporal, vegetation, topographic and positional variables were analyzed in an attempt to identify the variables most related to human-caused fire occurrence. The next two chapters present the models developed by using results from Chapter II. The ability of the logistic regression analysis techniques to represent human-caused occurrence data is discussed in Chapter III, while the ability of neural network technology to predict human-caused fire occurrence is presented in Chapter IV. Chapter V presents the conclusions of the study, and Chapter VI summarizes future research needs.

1.6. References

Altobellis, A.T. 1983. A survey of rural population density and forest fire occurrence in the South, 1956-1970. U.S.D.A. Forest Service, Southern Forest Experiment Station, Research Note SO-294. 4 pp.

Bruce, D. 1963. How many fires?. Fire Control Notes 24: 45-50.

Chou, Y.H., R.A. Minnich, L.A. Salazar, J.D. Power, and R.J. Dezzani. 1990. Spatial autocorrelation of wildfire distribution in the Idyllwild Quadrangle, San Jacinto Mountains, California. Photogrammetric Engineering & Remote Sensing 56: 1507-1513.

Christiansen, J.R., and W.S. Folkman. 1971. Characteristics of people who start fires...some preliminary findings. U.S.D.A. Forest Service, Pacific Southwest Forest and Range Experiment Station, Research Note PSW-251. 5 pp.

Cole, L.W., and H.F. Kaufman. 1963. Socio-economic factors and forest fires in Mississippi counties. Mississippi State University Social Sciences Research Centre, Mississippi, USA. Preliminary Report 14. 68 pp.

Crosby, J.S. 1954. Probability of fire occurrence can be predicted. U.S.D.A. Forest Service, Central States Forest Experiment Station, Technical Paper 143. 14 pp.

Cunnigham, A.A., and D.L. Martell. 1973. A stochastic model for the occurrence of man-caused forest fires. Canadian Journal of Forest Research 3: 282-287.

Cunnigham, A.A., and D.L. Martell. 1976. The use of subjective probability assessments to predict forest fire occurrence. Canadian Journal of Forest Research 6: 348-356.

Davidson, T. 1993. A fire occurrence prediction system for Pictou County, Nova Scotia. B.Sc. Thesis, University of New Brunswick, Frederickton, Canada. 38 pp.

Deeming, J.E., R.E. Burgan, and J.D. Cohen. 1977. The National Fire Danger Rating System. U.S.D.A. Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-39. 165 pp.

Donoghue, L.R. 1982. Classifying wildfire causes in the USDA Forest Service: problems and alternatives. U.S.D.A. Forest Service, North Central Forest Experiment Station, Research Note NC-280. 5 pp.

Donoghue, L.R., and W.A. Main. 1985. Some factors influencing wildfire occurrence and measurement of fire prevention effectiveness. Journal of Environmental Management 20: 87-96.

Doolittle, M. L. 1972. The dimensions of man-caused forest fire risk: a systematic assessment. Ph.D. dissertation, University of Washington, Seattle, Washington, USA. 119 pp.

Doolittle, M.L. and L.R. Donoghue. 1991. Status of wildland fire prevention evaluation in the United States. U.S.D.A. Forest Service, North Central Forest Experiment Station, Research Paper NC-298. 26 pp.

Folkman, P.E. 1977. High-fire-risk behaviour in critical fire areas. U.S.D.A. Forest Service, Pacific Southwest Forest and Range Experiment Station, Research Note PSW-125. 12 pp.

Forestry Canada Fire Danger Group. 1992. Development and Structure of the Canadian Forest Fire Behavior Prediction System. Forestry Canada, Science and Sustainable Development Directorate, Ottawa, Ontario, Canada. Information Report ST-X-3. 63 pp.

Haines, D.A., W.A. Main, J.S. Frost, and A.J. Simard. 1983. Fire-danger rating and wildfire occurrence in the northeastern United States. Forest Science 29: 679-696.

Higgins, D.G., and G.S. Ramsey. 1992. Canadian Forest Fire Statistics: 1988-1990. Forestry Canada, Petawawa National Forestry Institute, Chalk River, Ontario, Canada. Information Report PI-X-107E/F. 76 pp.

Jonhson, F.B. 1968. Demographic and socioeconomic factors and forest fire in the southeastern United States. M.S. thesis, Mississippi State University, Mississippi, USA. 73 pp.

Loftsgaarden, D.O., and P.L. Andrews. 1992. Constructing and testing logistic regression models for binary data: Aplications to the National Fire Danger Rating System. U.S.D.A. Forest Service. Intermountain Research Station, General Technical Report INT-286. 36 pp.

Lynham, T.J. 1989. Forest fire occurrence prediction research: a discussion paper. In: Proceedings of the 34th meeting of the Canadian Committee on forest fire management, January 21-23, 1986, Ottawa, Ontario. National Research Council of Canada. Ottawa, Ontario, Canada. 24 pp.

Lynham, T.J., Editor and compiler. 1991. National Worshop on Forest Fire Occurrence Prediction. May 3-4, 1989. Government of Canada. Great Lakes Forestry Centre, Forest Fire Research Unit. [Unpublished Report]. 44 pp.

Main, W.A., and D.A. Haines. 1974. The causes of fires on northeastern national forest. U.S.D.A. Forest Service, North Central Forest Experiment Station, Research Paper NC-102. 7 pp.

Martell, D.L., E. Bevilacqua, and B.J. Stocks. 1989. Modelling seasonal variation in daily people-caused forest fire occurrence. Canadian Journal of Forest Research 19: 1555-1563.

Martell, D.L., and S. Otukol. 1985. Development of mathematical models for predicting daily people-caused forest fire occurrence in Ontario. Faculty of Forestry, University of Toronto, Ontario, Canada. 83 pp.

Martell, D.L., S. Otukol, and B.J. Stocks. 1985. A daily people-caused fire occurrence prediction. Pages 2-10 in Alexander, M.E., editor. Proceedings of the Second Central Region Fire Weather Committee Scientific and Technical Seminar, April 17, 1985, Winnipeg, Manitoba. Canadian Forestry Service, Northern Forest Research Centre, Edmonton, Alberta, Canada. 31 pp.

Martell, D.L., S. Otukol, and B.J. Stocks. 1987. A Logistic Model for Predicting Daily People-caused Fire Occurrence in Ontario. Canadian Journal of Forest Research 17: 394-401.

Merrill, D.F., and M.E. Alexander, editors. 1987. Glossary of forest fire management terms. Fourth edition. National Research Council of Canada, Canadian Committee on Forest Fire Management, Ottawa, Ontario, Canada. Publication NRCC No. 26516. 91 pp.

NeuralWare Inc. 1991. Neural Computing. NeuralWorks Professional II/plus and NeuralWorks Explorer. Pittsburgh, Pennsylvania, USA. 358 pp.

Phillips, C., and B. Nickey. 1978. The concept of "Spatial Risk" and its application to fire prevention. Fire Management Notes, 39(4): 7-8, 19 pp.

Tithecott, A.G. 1990a. Evaluation of Martell's logistic regression model for people-caused fire occurrence prediction. Ontario Ministry of Natural Resources, Aviation and Fire Management Centre. Publication No. 273. 34 pp.

Tithecott, A.G. 1990b. Evaluation of PEOPLE (a daily fire occurrence prediction system). Ontario Ministry of Natural Resources, Aviation and Fire Management Centre. Publication No. 274. 32 pp.

Tithecott, A.G. 1990c. 1989 People-caused fire occurrence prediction project. Ontario Ministry of Natural Resources, Aviation and Fire Management Centre. Publication No. 275. 18 pp.

Tithecott, A.G. 1993. Application of fire occurrence prediction models in Ontario's fire management program. Pages 57-72 in Anderson, K.R., editor. Proceedings Eight Central Region Fire Weather Committee Scientific and Technical Seminar, April 3, 1992, Winnipeg, Manitoba. Forestry Canada, Northwest region, Northern Forestry Centre, Edmonton, Alberta, Canada. 78 pp.

Todd, B., and P.H. Kourtz. 1991. Predicting the daily occurrence of people-caused forest fires. Forestry Canada, Petawawa National Forestry Institute, Chalk River, Ontario, Canada. Information Report PI-X-103. 16 pp.

Van Wagner, C.E. 1987. Development and structure of the Canadian Forest Fire Weather Index System. Government of Canada, Canadian Forestry Service, Ottawa, Ontario, Canada. Forestry Technical Report No. 35. 36 pp.

Woodard, P.M., and J. Niederleitner. 1983. Forest fire prevention data: a change in perspectives. Forestry Chronicle, 59: 229-234.

CHAPTER II

Identification of geographic and temporal variables related to human-caused wildfire occurrence[1]

2.1. Introduction

The timing and location of human-caused wildfires in a forest are dependent on temporal and spatial variables, which determine the number and distribution of human sources of ignition and the ease of ignition and fire spread. A list of time-related variables may include: long (holiday) weekends, berry picking seasons, hunting and fishing seasons, and summer breaks from school, for instance, and weather. Site-related variables are: access, towns, campsites, timber, and other attractors, and topography and fuels. All these variables vary in time and space within a forest, resulting in a variation in the risk of wildfire at every site for any specific point in time.

Chou *et al.* (1990) explored some of these geographic and spatial factors in human-caused fire occurrence using Geographic Information System (GIS) technology and found a significant relationship between them. A GIS can be described as "an

---

[1] Some material in this chapter has been published as: Vega-Garcia, C., P.M. Woodard, and B.S. Lee. 1993. Geographic and temporal factors that seem to explain human-caused fire occurrence in Whitecourt Forest, Alberta. Pages 115-119 in Proceedings of the GIS'93 International Symposium, Vancouver, British Columbia, Canada, February 15-18th, 1993, Vol 1. 591 pp.
and: Vega-Garcia, C., P.M. Woodard, and B.S. Lee. 1993. Mapping risk of wildfires from human sources of ignition with a GIS. Pages 419-426 in Proceedings of the Thirteenth Annual ESRI User's Conference, Palm Springs, California, USA, May 24-28th, 1993,Vol.1. 608 pp.

organized collection of computer hardware, software, geographic data, and personnel designed to efficiently capture, store, update, manipulate, analyze, and display all forms of geographically referenced information" (Environmental Systems Research Institute Inc. 1991b).

By using the powerful capabilities of a GIS, another subset of geographic variables (vegetation, topography, and human development factors) was chosen to study how they relate to wildfire occurrence. The goal of this study was to test the relationship of many geographic variables, together with some commonly used weather and temporal variables, to human-caused fire occurrence.

It was recognized that fire prevention efforts alter both the fire environment and risk levels in an area (Moak 1976, Doolittle and Donoghue 1991). But the effects of prevention measures on fire occurrence are very difficult to assess, and the lack of adequate data on the effectiveness of prevention efforts in the Whitecourt Forest prevented consideration of this aspect in the analysis.

Also, the available historic records of human-caused wildfire occurrence considered alluded only to fires detected or actioned by fire management agencies or "fire arrivals", as they have been named by Tithecott (1993). Thus, detection capabilities in the area should also be considered when studying the human-caused fire occurrence process. The rationale is that scarce or limited detection resources delay the report of the fires, and create a time lapse between the actual fire occurrence and its detection by the agency responsible for forest protection. Delayed "arrivals" could introduce serious errors in correlating fire occurrences to temporal and weather variables. Also, abundant and effective detection resources may increase detection efficiency and increase the number of "arrivals", thus resulting in more fires being reported. The current detection level in the study area (Whitecourt Forest, in Appendix A) is considered to be adequate and most fires are detected (recorded) on the day they occur. This is a commonly accepted assumption for human-caused fires elsewhere (Tithecott 1993), and it might be due to the important role that other humans in the same area play in detecting fires. The work described here was based on this assumption, but three detection-related variables were also analyzed as part of this study. They were: (1) atmospheric visibility (temporal), (2) distance to a lookout tower (geographic), and (3) position in area seen by 0,1,2,3 or more lookout towers (geographic), as variables that could partially affect the probability of fire arrivals because all three variables can significantly affect detection efficiency.

2.2. Methods

2.2.1. *Geographic Variables*

The 20,000 km^2 study area (the Whitecourt Forest) was divided into about 30,500 cells using a griding procedure in ARC/INFO (Environmental Systems Research Institute Inc. 1991a). Each cell measured 800 m on a side because fire locations in Alberta are recorded using the Alberta Township System, to the level of a quarter section (0.5 mile by 0.5 mile, which is approximately 800 m by 800 m). The geographic characteristics and past fire history of each cell formed the population for analysis.

Each cell was coded with information relative to the 18 attributes presented in Table 2.1. The different categories of ASPECT, ELEVATION, and SLOPE were determined by ARC/INFO software based on contour lines digitized originally by the federal Department of Energy, Mines and Resources (DEMR) from the 1:250,000 topographic map of the area. Distances to the closest CAMPSITE, LAKE, LOOKOUT tower, RIVER, ROAD, and TOWN were obtained by applying the ARC/INFO's GRID euclidean distance function (Environmental Systems Research Institute Inc. 1991a) to coverages digitized by the DEMR from the 1:250,000 base map of the forest. The location of campsites within the Forest were digitized from this same base map for this study. FUEL categories were coded using the Alberta Phase 3 Forest Inventory (Alberta Energy and Natural Resources 1985). A digital coverage for this data set was provided by the Timber Management Branch of the Alberta Forest Service. This coverage contained Alberta Phase 3 Forest Inventory stand attributes summarized for all quarter sections in Alberta. The Alberta Phase 3 Forest Inventory coverage was also used to code cells relative to PROPERTY, stand AGE, stand HEIGHT, forest COMMERCIALITY and tree DENSITY. VISIBILITY was computed with the GRID subsystem in ARC/INFO, which used position and height data for each lookout tower as provided by the Forest Protection Branch of the Alberta Forest Service, and the elevation coverage. The variable DISTRICT, included to account for general differences among forest districts other than area, was also assigned using a forest districts coverage digitized for the study.

Table 2.1. *A list of geographic variables to be tested in their relationship to wildfire occurrence.*

Abbreviation	Description	Units
ASPECT	Topographic aspect of the cell	flat, north, south, east, and west
ELEVATION	Topographic elevation	meters above sea level
SLOPE	Topographic slope in the cell	percentage
CAMPSITE	Distance to closest campsite	meters
LAKE	Distance to closest lake	meters
LOOKOUT	Distance to closest lookout tower	meters
RIVER	Distance to closest river	meters
ROAD	Distance to closest road	meters
TOWN	Distance to closest town	meters
PROPERTY	Land ownership, classified as:	forest management area, forest quota area, private land, other
FUEL	Fuel category, classified as:	FBP system categories[1]
AGE	Stand age in the cell	years
HEIGHT	Stand height in the cell	meters
COMMERCIALITY	Forest commercial value, classified as:	lumber, roundwood, low uncommercial, high uncommercial
DENSITY	Tree density in the cell	0-25%, 26-50%, 51-75%, 76-100%
VISIBILITY	Cell located on area seen from N lookout towers	N = 0, 1, 2, 3, 3+
DISTRICT	Location in Forest District M	M = 1, 2, 3, 4
OCCUR	Number of fires in the cell for the period 1986-1990	integer (0-3)

[1] Categories recognized by the Fire Behaviour Prediction System (Forestry Canada Fire Danger Group 1992).

Lastly, the number of fire occurrences for each cell (OCCUR) was coded. Fire occurrence data were obtained from the records maintained by the Forest Protection Branch of the Alberta Forest Service, in Edmonton. Only data from 1986 to 1990 were used in this study in an attempt to maintain good agreement with the dating of the geographic information available. Again, the purpose was to reduce the problems that arise from relating fire locations to geographic features that did not exist at the time of the fire occurrences. Whenever the number of fire occurrences in a cell was 2 (or 3) for the period 1986-1990, the cell was counted twice (or three times).

The null hypothesis (H_0) was stated as follows: the frequency distributions of geographic variable X in the cells with number of fire occurrences equal or greater than one (fire cells) and in the entire population of cells (the forest) are the same. Accepting this hypothesis implies that X and fire occurrence are not related. If a geographic variable X is unrelated to fire occurrence, then the values of that variable in the fire cells exhibit the characteristics of a random sample taken from the total population, because the fires are located randomly in the forest with respect to the variable. Thus, the frequency distribution of the variable X among the fire cells (the sample) is equal to the frequency distribution of X in the population encompassing all cells in the forest (except for the sampling error). If fires are related to a variable X then they tend to be associated with the occurrence of X or certain values of X, histograms in the fire cells (sample) and in the entire population of cells are different, and the null hypothesis cannot be accepted. Therefore, rejecting the null hypothesis implies a relationship between fire occurrence and the geographic variable X.

To test this hypothesis, the frequency distributions of the geographic variables for the 233 fire cells and for all cells in the forest (about 30,500 depending on the missing values of the variable) were calculated. The values obtained from ARC/INFO grids were grouped by classes using SAS (SAS Institute Inc. 1985) to remove distortion caused by partitioning in cells, prior to calculating the frequency distributions using the procedure FREQ (SAS Institute Inc. 1985). Then, the Chi-square goodness-of-fit test (Gibbons 1976) was applied to every pair of sampled and hypothesized distributions. The lack of fit is calculated in the Chi-square goodness-of-fit test through the statistic Q (Gibbons 1976):

$$Q=\sum_{i=1}^{i=r}\frac{(O_i-E_i)^2}{E_i} \tag{2.1}$$

where E_i = expected frequency in the class i

O_i = observed/sampled frequency in the class i

r = number of classes

The sampling distribution of Q is approximately the Chi-square distribution with r-1 degrees of freedom, when the sample is sufficiently large (n>30)(Gibbons 1976). In general, small Q values suggests an agreement between the two distributions, while large Q values favour rejection of the null hypothesis (H_0) of equal frequency distributions in fire cells and all cells in the forest (Gibbons 1976). The decision rule chosen was to reject H_0 when $Q > \chi^2_{.01,r-1}$ at a significant level of α =.01. This relatively low α value was chosen to reduce the probability of a type I error (reject Ho when true) and therefore weed out with great confidence (99%) non-significant variables.

2.2.2. *Temporal variables*

The units for the temporal analysis were the days in the 1986-1990 fire seasons and the weather and temporal variables associated with these days, within each of the four forest districts in the Whitecourt Forest. This weather information was obtained from historic records kept at Northern Forestry Centre, Forestry Canada, in Edmonton. The period 1986-1990 can be described as average in terms of weather, based in a comparison between the total annual precipitation at the Whitecourt weather station those years and the 30-year average from 1951-1980 (Canadian Climate Program 1982). The years 1986 and 1989 were above average, but not exceedingly so.

The 12 temporal variables that were analyzed are presented in Table 2.2. The entire collection of days in the five fire seasons was used as the population for analysis. A fire season for the Alberta Forest Service usually begins in April and ends on October. The beginning and end dates depend on the snowfall. The days with one or more fires (the fire days) were considered a sample, investigated in regards to its randomness with respect to the population. Days which suffered 2 (or more) fire occurrences were included twice (n times) in the database for analysis.

The null hypothesis was that the frequency distribution of a variable X among the fire days (the new sample) is equal to the frequency distribution of X in the population encompassing all days for the forest.

Table 2.2. *A list of temporal variables to be tested in their relationship to wildfire occurrence.*

Abbreviation	Description	Units
FFMC	Fine Fuel Moisture Code[1]	open-ended scale
DMC	Duff Moisture Code[1]	open-ended scale
DC	Drought Code[1]	open-ended scale
ISI	Initial Spread Index[1]	open-ended scale
BUI	Build-up Index[1]	open-ended scale
FWI	Fire Weather Index[1]	open-ended scale
TEMPERATURE	Daily temperature	centigrades
HUMIDITY	Relative humidity	percentage
WIND	Wind speed	km/h
VISIBILITY	Atmospheric visibility	meters
WEEKDAY	Day of the week	Saturday through Sunday
MONTH	Month	April to October[2]
OCCUR	Number of fires in the day and district for the period 1986-1990	integer (0-8)

[1] Codes and Indices from the Canadian Forest Fire Weather Index (Van Wagner 1987).
[2] Classed as the Fire Season in Alberta.

The frequency distributions of the temporal variables for the fire days (199) and for all days in the population (4,082 observations=±200 days/year x 5 years x 4 districts) were calculated with the procedure FREQ in SAS (SAS Institute Inc. 1985). The Chi-square goodness-of-fit test was applied to every pair of sampled and hypothesized distributions. The same decision rule was used in both geographic and temporal variable testing procedures.

2.2.3. *Geographic distribution of human risk*

The tests described above were expected to identify geographic factors relevant to human-caused wildfire occurrence. The geographic variables identified as being significant by the test were considered to be good descriptors of the geographic distribution of human risk levels in the Whitecourt Forest, according to the main hypothesis of the study (Chapter I). The frequency histograms of the significant variables were analyzed to establish criteria for risk definition with respect to each variable. The ARC/INFO coverages relative to these variables were combined to provide a visual representation of the geographic location of human risk in the Whitecourt Forest in the form of a map processed by ARC/INFO's GRID (Environmental Systems Research Institute Inc. 1991a). All risk-related variables were given the same importance in building the map. Then, fire data from 1991-1992 was overlaid on this map to test the validity of the geographic arrangement of risk levels found. These years were also average weather years according to the criterion used above. The standard Chi-square test was used to compare the frequency distribution of the 1991-1992 fires throughout zones in the map assigned to a certain risk level (low, moderate, high, very high) with the frequency distribution to be expected in this zones if fires were not related to human risk as defined by the significant geographic variables.

2.3. Results and Discussion

The values obtained for the test statistic Q and $\chi^2_{.01,r-1}$ are shown in Table 2.3 for all geographic variables and Table 2.4 for all temporal variables. Also present in both tables are the number of classes (r) for each variable, the result of the test, and the associated P-value. Even though this is an approximate test and P-values are asymptotic, the approximation of Q provided by the Chi-square distribution can be considered reliable in this case, because the sample size is sufficiently large and recommendations have been followed with respect to the expected class frequencies E_i (not smaller than five) (Gibbons 1976).

The tests suggest fire occurrences are not randomly distributed throughout the Whitecourt Forest with respect to distance to road, land ownership, distance to town, distance to campsite, elevation, forest district, fuels and some forest characteristics. All these variables are associated with wildfires. These results agree with findings in California with respect to distance to roads, distance to campgrounds, and vegetation (Chou *et al.* 1990). Also, fire occurrences are influenced by ISI, FWI, FFMC, relative humidity, month, DMC, BUI, and wind speed. These variables are highly related to timing of fire occurrences, and in that order of importance, according to the test.

Table 2.3. *A list of geographic variables and their Chi-square test values.*

r	variable	Q	$\chi^2_{.01,r-1}$	reject H_0	p-value
10	ROAD	177.38	21.7	yes	< .001
4	PROPERTY	84.07	11.3	yes	< .001
10	TOWN	61.70	21.7	yes	< .001
11	CAMPSITE	50.76	23.2	yes	< .001
11	ELEVATION	49.06	23.2	yes	< .001
4	DISTRICT	36.68	11.3	yes	< .001
5	FUEL	33.50	13.3	yes	< .001
5	COMMERCIALITY	19.13	13.3	yes	< .001
8	AGE	19.34	18.5	yes	.005< P <.010
6	HEIGHT	17.98	15.1	yes	.001< P <.005
8	LAKE	16.99	18.5	no	.010< P <.025
5	DENSITY	9.11	13.3	no	.050< P <.100
4	VISIBILITY	8.83	11.3	no	.025< P <.050
7	LOOKOUT	6.68	16.8	no	>.100
7	SLOPE	3.92	16.8	no	>.100
7	RIVER	3.35	16.8	no	>.100
5	ASPECT	2.51	13.3	no	>.100

r=number of classes, Q=goodness-of-fit test statistic

Table 2.4. *A list of temporal variables and their Chi-square test values.*

r	variable	Q	$\chi^2_{.01,r-1}$	reject H_0	p-value
6	ISI	242.14	15.1	yes	< .001
10	FWI	237.70	21.7	yes	< .001
9	FFMC	200.43	20.1	yes	< .001
8	HUMIDITY	161.74	18.5	yes	< .001
7	MONTH	103.27	16.8	yes	< .001
9	DMC	98.13	20.1	yes	< .001
7	BUI	76.34	16.8	yes	< .001
11	WIND	29.02	23.2	yes	.001< P <.005
7	DC	16.61	16.8	no	.010< P <.025
8	VISIBILITY	14.19	18.5	no	.025< P <.050
5	TEMPERATURE	10.16	13.3	no	.025< P <.050
7	WEEKDAY	4.46	16.8	no	>.100

r=number of classes, Q=goodness-of-fit test statistic

These results concur with theoretical expectations formulated by Alexander (1986) for ISI, FWI, FFMC, and by Maxey and Lee (1973) for relative humidity. They also agree with findings by Martell *et al.* (1987) for FFMC and BUI, by Todd and Kourtz (1991) for FFMC, DMC and wind speed, and with several other studies in different locations.

Slope and weekday, which have been traditionally regarded as relevant to the occurrence of human related forest fires were not found to be significant according to this test. With regards to weekday, the result of this study agrees with the findings of Martell *et al.* (1987) for the northern region of Ontario. The lack of significance relative to slope might be due to lack of variability in data values caused by the general flatness of the study area. In all other cases, there was enough variability in data values in the populations to assume that any trend in fire occurrence would be identified by the test.

No consideration was given to the possibility that the variables were correlated. Instead, each variable was evaluated independently in the test, when in fact, strong correlations are likely among geographic variables and among temporal variables. For instance, ISI is calculated from wind speed and FFMC values, FWI is calculated from BUI and ISI indices. Roads are built low in the river valleys, and form denser networks around population centres. Grass fuels are predominant in privately owned land. This issue should be considered in future studies of fire occurrence prediction that includes the variables selected as relevant in this study.

Further analysis of the frequency histograms for the significant geographic variables yielded important information for several fire management applications in the form of a map. For numeric variables this analysis led to the use of the 90th percentile of the distributions to define risk. For instance, 90% of the human-caused fires started within 4.8 km from a road. The other 10% started within a wider range of 4.8 to 16.8 km from a road. Some of these fires were suspected of being mismatched in time with respect to the road coverage available for the study, others of being exceptions to the general trend. Therefore, areas within 4.8 km (the 90th percentile) from a road were declared at risk due to their physical location relative to the risk factor roads. A grid was then developed in ARC/INFO, reclassifying these risk areas as 1 and all others as 0. The same criteria and process were followed for: distance to closest town, distance to closest campsite, and elevation variables. The 90th percentiles for these three variables were 35 km, 30 km, and 1,000 m (rounded).

The frequency distributions for the categorical variables: fuels, land ownership, and forest commerciality were analyzed differently. The categories represented in the fire cells with percentages higher than the corresponding ones in the forest were declared categories of risk. For instance, fuels in the forest were distributed as follows: 61.8 % conifers, 34.9 % deciduous, 0.6 % open fuels, and 0.8 % slash. Fuels in the fire cells, though, were: 47.9 % conifers, 42.5 % deciduous, 2.7 % open fuels (grass), and 1.6 % slash. Therefore, areas classed as deciduous, grass, or slash were identified as risk factors, and an ARC/INFO coverage (grid) was developed reclassifying these areas as 1, and the rest (conifers, water, no fuel) as 0. The higher risk categories associated to land ownership and commerciality were: "private land"

and "highly uncommercial", respectively.

The seven partial risk grids were then summed using ARC/INFO's GRID software (Environmental Systems Research Institute Inc. 1991a) to obtain a composite grid, a "Map of Risk for the Forest". The risk values of this composite grid ranged from 0 to 7. Areas of value 0 showed no risk factor, as defined above, present. In these areas, humans posed very little risk of fire occurrence. Areas with a value of 7 had all risk factors present, and most fire occurrences were expected to start here. For displaying purposes and easier interpretation, this grid was reclassified according to the following criteria: 0-1 = low risk, 2-3 = moderate risk, 4-5 = high risk, 6-7 = very high risk, and it is displayed in Figure 2.1. Paler shades of grey indicate lower risk, whereas black depicts very high risk areas in the Forest.

Fire data from 1991-1992 was overlaid on this Map of Risk to assess the validity of the geographic arrangement of the risk levels found. The number of fires in the low-risk zone was 4 fires per million ha, and in the moderate-risk zone was 22 fires per million ha. The high-risk zone suffered 45 fires per million ha, 55 fires per million ha occurred in the very-high-risk zone. The Chi-square goodness-of-fit test (Gibbons 1976) applied established the significant relationship between the location of the fires in 1991-1992 and the distribution of human risk in the Whitecourt Forest as defined by the seven geographic variables (Q = 14.99, $\chi^2_{.01,3}$ = 11.3).

The Map of Risk seemed to properly describe the arrangement of risk levels due to people in Whitecourt Forest according to the new fire occurrence data. The Map of Risk developed also agreed with the actual geographic distribution of human

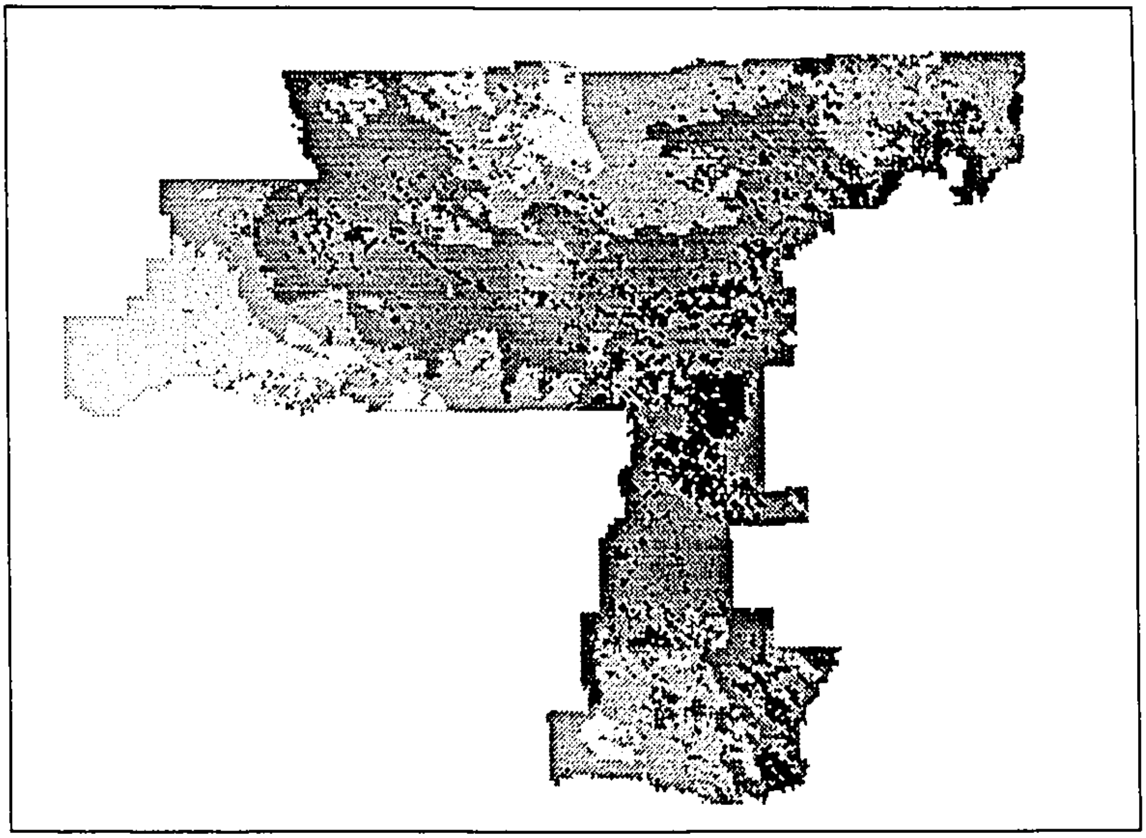


Figure 2.1. Whitecourt Forest Map of Risk, based on seven risk factors: Closeness to roads, campsites, and towns, low elevation, private land, uncommercial forest, and presence of certain fuels. Darker shades indicate higher risk.

risk estimated by the fire personnel in the Whitecourt Forest[2], and consequently, prevention efforts are currently being applied to much of the area classified as of very-high-risk.

2.4. Management Implications

Phillips and Nickey (1978) stated that similar numbers of fire occurrences should be expected in areas with similar environmental and fire-related characteristics, unless prevention efforts introduced a difference in those characteristics. Effective forest fire prevention can pay dividends (Moak 1976) and efforts are usually designed to reduce the occurrence of unwanted fires in areas where fires have historically occurred within the past five years. Yet this approach cannot account for areas of risk where fires have not yet occurred. Fire prevention programs could benefit from risk maps to identify zones where advertising and poster campaigns can be more effective, or areas where personal contact (Doolittle and Welch 1974) with the target publics should be increased, or as an aid in scheduling or routing detection patrols.

In combination with more complete information about hazard levels (fuel related) in the Forest, these maps can aid in designing efficient strategies for fuel modification where required. Several examples of spatial strategies for prescribed burning in San Bernardino National Forest, California, using ARC/INFO, were presented in Chou (1992).

---

[2] Personal communication, Mag Steiestol, Fire Prevention Co-ordinator, Alberta Forest Service, Forest Protection Branch, Edmonton.

There are other possible applications for a risk map in fire management. Incorporation of risk areas to the daily planning for aerial patrols should improve its effectiveness and increase efficiency since one of the major difficulties these resources face in operation is the uncertainty associated to the route planning process. Usually, patrol routes and patrol frequencies are established using predefined guidelines depending on the daily weather parameters and visibility (Kourtz 1987).

A typical application of GIS to fire management is the determination of visible areas from fixed lookout towers. Mapping of areas of high risk that are non-visible from lookout towers allow managers to allocate other detection resources to them and make sure they are covered. Figure 2.2 shows the locations of those non-visible high-risk and very-high-risk areas in Whitecourt Forest. They have been calculated simply by combining visibility and risk grids. Also, high-risk areas can be accounted for when positioning new detection resources, thus maximizing their productivity.

Some of the most important values-at-risk in a forest can be expected to be close to human-caused fires starts: the humans themselves. It is important that towns, campgrounds, airfields, or in general "man-made" structures are properly identified. This features, and other values-at-risk such as timber, or personal property, can be easily displayed on a risk map. Hence, measures can be taken to insure their protection.

The usefulness of a human risk map is limited by the rapidly changing conditions in the fire environment, in particular fuels and weather. This problem can be partially solved by computing seasonal or daily maps. These maps can account for

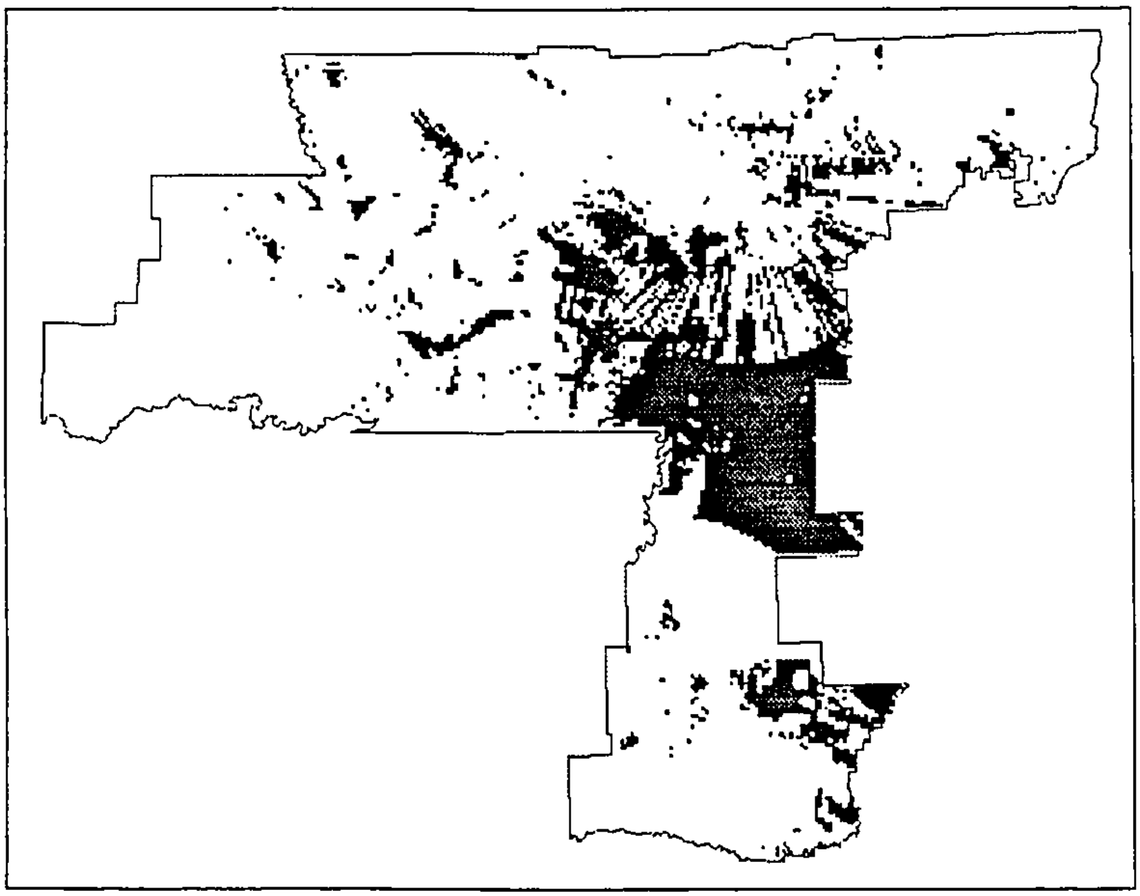


Figure 2.2. Map of non-visible high-risk areas in Whitecourt Forest

well known variations in combustibility of certain fuels throughout the fire season. The map in Figure 2.1 is applicable when all grass, deciduous and slash areas are prone to burn. These conditions frequently exist during Spring (April-June) and Fall (August-October). During the Summer, after green-up has taken place, a map including only grass and slash areas may require concentration of efforts and describe better the risk of fire occurrences.

2.5. Conclusions

Human-caused wildfire occurrences are rare events that exhibit complex relationships with geographic and temporal variables. There is a great degree of randomness associated with the fire occurrence prediction process, and we will never be able to account for singularities in human behaviour. But until now we have not taken full advantage of some geographic and temporal relationships that appear to be quite well defined as a result of this study, and may broaden our understanding of human risk and improve current wildfire occurrence predictions.

The following temporal variables are significantly related to human-caused wildfire occurrence in Whitecourt Forest: FFMC, DMC, BUI, ISI, FWI, relative humidity, wind speed, and month. The following geographic variables: distance to roads, towns, and campsites, topographic elevation, land ownership, forest commerciality, and fuels are also significantly related to fire occurrence. Furthermore, these geographic variables can be used to describe and map risk associated to human sources of ignition in the Whitecourt Forest.

Mapping of human risk can be of help to fire managers in making decisions about prevention, detection, and suppression actions. The usefulness of a map of risk is limited by the changing characteristics in the fire environment. Seasonal, even daily maps could be used for certain applications, but a preferable solution for dealing with human-caused forest fires is the development of a daily fire occurrence prediction model.

This model should incorporate risk and environmental factors and consider their geographic and temporal variation. It should also be able to deal with presumable correlations among the variables involved. Such a model was attempted in this study and it is described in the next chapters. The variables found to be significantly related to fire occurrence in this chapter were the factors considered for modeling fire occurrence in the same study area. By doing so, the human-caused fire occurrence prediction model building process was conducted more efficiently.

The local character of this selection of variables is stressed, though. Patterns of fire starts by humans vary across geographic areas and climates. There is no guarantee that these same exact results will apply elsewhere. Some of the variables selected in this study are probably not so strongly related to wildfire occurrence in other regions (month, for instance, cannot be expected to be relevant in equatorial regions where there are no seasons). Some of the variables rejected in Whitecourt might show a strong relationship to human-caused fires in other areas (distance to a lookout tower in areas with very scarce resources, slope in very steep country). A screening of variables is recommended prior to any fire occurrence prediction model building.

2.6. References

Alberta Energy and Natural Resources. 1985. Alberta Phase 3 Forest Inventory: forest cover type specifications. Alberta Forest Service, Resource Evaluation and Planning Division, Edmonton, Alberta, Canada. ENR Report No. Dept. 58. 53 pp.

Alexander, M.E. 1986. Editor's Note. Page 55 in Alexander, M.E., compiler and editor. 1986. Proceedings of the Third Central Region Fire Weather Committee Scientific and Technical Seminar, April 3, 1986, Winnipeg, Manitoba. Canadian Forestry Service, Northern Forestry Centre, Edmonton, Alberta, Canada. Study NOR-5-05 (NOR-5-191), File Report No. 16. 61 pp.

Canadian Climate Program. 1982. Canadian Climate Normals. Vol 3. Precipitation 1951-1980. Atmospheric Environment Service, Environment Canada. 602 pp.

Chou, Y.H., R.A. Minnich, L.A. Salazar, J.D. Power, and R.J. Dezzani. 1990. Spatial Autocorrelation of Wildfire Distribution in the Idyllwild Quadrangle, San Jacinto Mountain, California. Photogrammetric Engineering & Remote Sensing. 56: 1507-1513.

Chou, Y. H. Evaluating spatial strategies of wildfire prevention with a GIS. 1992. Pages 1-10 in Proceedings of the GIS'92 International Symposium, Vancouver, British Columbia, Canada, February 10-13th, 1992.

Doolittle, M.L., and G.D. Welch. 1974. Fire prevention in the Deep South: personal contact pays off. Journal of Forestry 72: 488-490.

Doolittle M.L., and L.R. Donoghue. 1991. Status of wildland fire prevention evaluation in the United States. U.S.D.A. Forest Service, North Central Forest Experiment Station, Research Paper NC-298. 26 pp.

Environmental Systems Research Institute. 1991a. GRID. Command References: Environmental Systems Research Institute Inc., Redlands, California, USA.

Environmental Systems Research Institute. 1991b. Understanding GIS: The ARC/INFO method. Environmental Systems Research Institute Inc., Redlands, California, USA.

Forestry Canada Fire Danger Group. 1992. Development and Structure of the Canadian Forest Fire Behaviour Prediction System. Forestry Canada, Science and Sustainable Development Directorate, Ottawa, Ontario, Canada. Information Report ST-X-3. 63 pp.

Gibbons, J. D. 1976. Nonparametric methods for quantitative analysis. International Series in Decision Processes No 23. Holt, Rinehart and Winston, New York, New York, USA, 463 pp.

Kourtz, P. 1987. The need for improved forest fire detection. The Forestry Chronicle 63: 272-277.

Martell, D.L., S. Otukol, and B.J. Stocks. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. Canadian Journal of Forest Research 17: 394-401.

Maxey, W.R., and R. Lee. 1973. Some theoretical approaches to the prediction of wildfire occurrence. West Virginia Forestry Notes 1: 6-9.

Moak, J.E. 1976. Fire prevention: does it pay?. Journal of Forestry 74: 612-614

Phillips, C. and B. Nickey. 1978. The concept of spatial risk and its application to fire prevention. Fire Management Notes 39(4): 7-8,19.

SAS Institute Inc. 1985. SAS User's guide; Statistics, Version 5 Edition: SAS Institute Inc., Cary, North Carolina, USA. 956 pp.

Tithecott, A. 1993. Application of fire occurrence prediction models in Ontario's fire management program. Pages 57-72 in Anderson, K.R., compiler. Proceedings Eight Central Region Fire Weather Committee Scientific and Technical Seminar, April 3, 1992, Winnipeg, Manitoba. Forestry Canada, Northwest Region, Northern Forestry Centre, Edmonton, Alberta, Canada. 78 pp.

Todd, B., and P.H. Kourtz. 1991. Predicting the daily occurrence of people-caused forest fires. Forestry Canada, Petawawa National Forestry Institute, Chalk River, Ontario, Canada. Information Report PI-X-103. 16 pp.

Van Wagner, C.E. 1987. Development and structure of the Canadian Forest Fire Weather Index. Government of Canada, Canadian Forestry Service, Ottawa, Ontario, Canada. Forestry Technical Report No. 35. 36 pp.

CHAPTER III

A Logit Model for Daily Human-caused Forest Fire Occurrence Prediction

3.1. Introduction

Human-caused wildfires are the result of a complex interaction between human sources of ignition and the physical environment of the forests. In a certain area, at any point in time, there are only two outcomes for this interaction: either a fire occurs or it does not. The capability of predicting fires on a daily basis for a certain region, for instance a forest district, can be very useful for many fire management applications, since many fire prevention decisions are made at a district level. Tithecott (1993) has pointed out that fire control experts want reliable predictions of occurrence for their daily planning, but they do not expect exact numbers of fires, rather some indication of the severity of the fire day and where fires are likely to occur.

Previous studies have explored the relationship between human-caused fires and several weather variables (Martell *et al.* 1987, Todd and Kourtz 1991), or geographic variables (Chou *et al.* 1993). The present work builds on these studies in an attempt to combine both weather variables and geographic variables in a daily human-caused fire occurrence prediction model for Whitecourt Provincial Forest, Alberta (Appendix A).

3.1.1. *Model selection*

A dichotomous dependent variable such as Fire Yes/No can be studied through several binary data analysis techniques. Common choices for models with dichotomous dependent variables are discriminant analysis, the linear probability model, the probit model, and the logit model (Ben Akiva and Lerman 1985, Cox and Snell 1989). Models relying on discriminant analysis were ruled out because it is not logical to assume that the independent variables available for use are normally distributed. Incorporation of dummy variables was anticipated; in this case normality assumptions would be violated, and the discriminant analysis estimator would not be consistent (Maddala 1983).

The linear probability model is not well suited for this application, mainly because the predictions would not always be restricted to the interval (0,1) (Maddala 1983, Cox and Snell 1989). This limitation can be overcome by using either the logit or the probit model. Though they are equal in predictive power, the logit model offers computational advantages (Maddala 1983). Furthermore, logit models have been successfully used in many similar applications. Martell *et al.* (1987) developed a daily human-caused fire occurrence prediction model where the probability of a fire-day was given by a logit model, as a function of weather parameters, in the Northern Region of Ontario. Loftsgaarden and Andrews (1992) used a logit model to describe the probability of a fire-day in the Lolo National Forest, Montana, based on the Energy Release Component (ERC) of the National Fire Danger Rating System (Deeming *et al.* 1977). Chou *et al.* (1993) have also used the logit model to identify areas of high probability of fire occurrence in San Bernardino National Forest, California, based on environmental, human, and spatial factors.

The logit model assumes the existence of a "latent" dependent variable, in this case the fire occurrence probability for a day, that is not observable other than as a dummy variable Y of value 0 (no-fire occurrence) or 1 (fire occurrence) (Maddala 1988). In this model, the daily probability of fire occurrence (at least one fire), P_i, is assumed to be adequately described for any observation i by the logistic function,

$$P_i = P(Y=1) = \frac{\exp(Z_i)}{1+\exp(Z_i)} \tag{3.1}$$

where Z_i is a function of the independent variables, in this case;

$$Z_i = \beta_0 + \sum_{j=1}^{j=k} \beta_j x_{ij} \tag{3.2}$$

where x_{ij} are the k explanatory variables, and β_0, β_1, β_2,...,β_k the k+1 parameters to estimate. Substracting equation 3.1 from 1 and simplifying leads to the daily probability of no-fire occurrence,

$$1 - P_i = P(Y=0) = \frac{1}{1+\exp(Z_i)} \tag{3.3}$$

Because the daily probability of fire occurrence is assumed to be a logistic function of the independent variables, the logit models are often referred to as "logistic regression models" or simply "logistic models".

Dividing equation 3.1 by 3.3 and taking the logarithm on both sides of the equation gives the log-odds ratio of the two possible outcomes, a linear function of the independent variables (Maddala 1988),

$$\ln\left(\frac{P_i}{1-P_i}\right)=Z_i=\beta_0+\sum_{j=1}^{k}\beta_j x_{ij} \tag{3.4}$$

where P_i is the probability of Y = 1, for any observation *i*

$1-P_i$ is the probability of Y = 0,

x_{ij} are the k explanatory variables or covariates,

and β_0, $\beta_1,\beta_2,\ldots,\beta_k$ are the k+1 model parameters to estimate.

Estimation of the model is conducted through maximum likelihood methods when analyzing individual observations such as Fire Yes/No (Stynes and Peterson 1984). Data pertaining to fire occurrence in the study area must be collected in order to estimate the model parameters. An adequate sampling strategy must be followed. Since sampling is usually random and fires are rare events, the data on fire occurrence is not easily acquired. A random sample of daily observations in the Whitecourt Forest for any period of time includes very few fire observations and a large number of no-fire observations. This problem has been often mentioned by fire researchers in other studies. The logit model allows the use of different sampling rates for the two subpopulations of fire and no-fire observations to obtain a balanced data set for the model estimation (Maddala 1988). Prediction for the entire population is permitted by making an appropiate adjustment in the intercept (Prentice 1986, Maddala 1988, Hosmer and Lemeshow 1989). This transformation is not valid for the probit or linear probability models (Maddala 1988). This was the most important factor in selecting the logit as the model for this study.

3.2. Methods

3.2.1. *Model development*

Previous work in the Whitecourt Forest (Chapter II) identified important geographic and temporal factors as relevant to the human-caused wildfire occurrence problem. Those factors are the basis for constructing the independent variables for the present study. They include: distance to nearest road, distance to nearest town, distance to nearest campsite, topographical elevation, fuels, forest commerciality, forest district, codes and indices in the Fire Weather Index (Van Wagner 1987), except the Drought Code, relative humidity, wind speed, and month.

A Geographic Information System (ARC/INFO, Environmental Systems Research Institute Inc. 1991), was used to map eight fire occurrence prediction units in the study area. Each of the four forest districts was divided in two zones: Areas ≤ 5 km from a road, and areas > 5 km from a road (Figure 3.1). Five kilometres was found to be the threshold distance from roads within which 90% of all fires start (Chapter II). This partition was expected to increase the variability in the geographic variables for analysis, while keeping the number of prediction units low. Predictions for these units would be useful for fire managers in the Forest making decisions at the district level. The fire occurrence prediction units ranged in area from 805 to 4,660 km^2. Some units were contiguous, but most of the prediction units were formed by summation of fragmented subunits within each forest district sharing the same

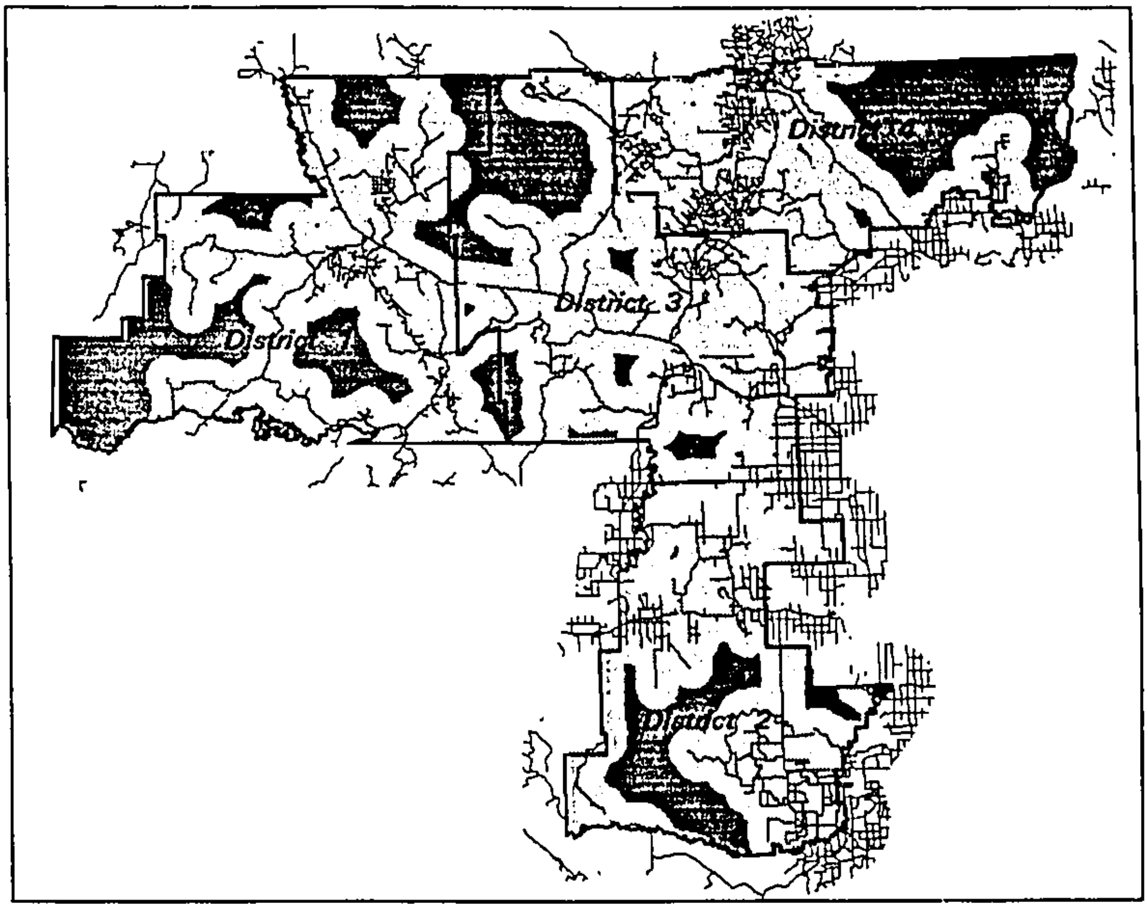


Figure 3.1. Eight fire occurrence prediction units for the logit analysis: four forest districts within the Whitecourt Forest divided in areas ≤ 5 km from a road and areas > 5 km from a road. Road network overlaid on the map.

"distance to road" characteristics.

Each fire occurrence prediction unit was coded, by using the ARC/INFO software and digital information from different sources (Chapter II), with information pertaining to the variables presented in Table 3.1:

Table 3.1. *Geographic variables assigned to each prediction unit.*

Abbreviation	Description	Units[1]
AREA	Fire occurrence prediction unit area	km^2/10,000
ROADDIS	Average distance to roads	km/100
TOWN	Average distance to town	km/100
CAMP	Average distance to campsites in the unit	km/100
ELEV	Average topographic elevation	meters above sea level / 10,000
PPA	Total area privately owned in the unit	km^2/1,000
COMH	Total area of highly uncommercial forest value	km^2/1,000
FUEL1	Total area of deciduous fuel in the unit	km^2/10,000
FUEL2	Total area of grass fuel	km^2/100
FUEL3	Total area of slash fuel	km^2/100
ROAD	Dummy variable for unit distance to road < or > than 5 km	1,0
DISTRICT	Dummy variable for location in Forest District 2 or Forest Districts 1,3,4	1,0

[1]Numeric variables were scaled down to obtain parameter estimates in the same order of magnitude.

The two dummy variables were included to account for sources of variability not included in other the factors. The dummy variable ROAD with a of value 1 for the prediction units within 5 km from roads, and 0 for units farther than 5 km from roads, was included because distance to road is the most significant geographic variable in explaining fire occurrences in the Whitecourt Forest (Chapter II). The dummy variable for DISTRICT with a value of 1 for prediction unit located in district 2, and 0 for prediction unit located in any other district was included because District 2 suffers from higher human pressure in the form of developmental activities, and is closer to Edmonton, Alberta, which has a population > 500,000 people.

The database for the analysis included observations for each day in the fire seasons 1986-1990 (April-October), in each of the prediction units described above (±200 days/fire season x 5 fire seasons x 8 units = 8,009 observations). This five-year period was chosen to keep agreement with the digitizing date of the geographic data. The daily weather variables assigned to each unit every day were averaged from the weather stations available in the district where the unit was located, and are listed in Table 3.2.

The Canadian Fire Weather Index System did not start on the same day for all stations and for any one station among years. Starting dates depend on a set of rules (Van Wagner 1987), which are influenced by local weather conditions. Days without codes and indices were not used in the analysis. The dummy variable MONTH, with a value of 1 for days in April and May, and 0 in June-October, was included to account for seasonal trends in fire occurrence in the Whitecourt Forest (50% of the

Table 3.2. *Temporal variables assigned to each prediction unit, each day.*

Abbreviation	Description	Units[1]
FFMC	Fine Fuels Moisture Code	open-ended scale[2]/100
DMC	Duff Moisture Code	open-ended scale[2]/100
BUI	Build-Up Index	open-ended scale[2]/100
ISI	Initial Spread Index	open-ended scale[2]/10
FWI	Fire Weather Index	open-ended scale[2]/100
RH	Relative humidity	percentage/100
WS	Wind speed	(km/h)/100
MONTH	Dummy variable for	1,0

[1]Numeric variables were scaled down to obtain parameter estimates in the same order of magnitude.

[2]as defined by Van Wagner (1987)

human-caused fires occurred in Spring in 1986-1990). The binary dependent variable OCCUR was assigned a value of 1 if at least one fire occurred in the unit and day of the observation, and a 0 value if there was no fire.

The result of this stratification was a data base of 8,009 observations, of which only 157 were fire observations. Since the logit analysis is not affected by unequal sampling rates (Maddala 1988, Prentice 1986), a random sample of 157 no-fire observations was obtained and used with the 157 fire observations for the logit analysis.

The SAS procedure LOGISTIC (SAS Institute Inc. 1989) was used to compute the logit models. This program computes the maximum likelihood estimates of the regression parameters using the "Iteratively Reweighted Least Squares (IRLS)" algorithm (SAS Institute Inc. 1989). Eighty-eight models, all in linear form, were built following recommendations from Cox and Snell (1989) and Hosmer and Lemeshow (1989) for model building with large numbers of explanatory variables. In the model building process:

1. Variables thought to be of special importance such as AREA, DISTRICT were forced in some of the models.

2. Totals, such as RTC = ROADDIS + CAMP + TOWN, or ratios, such as ARRODIS=ROADDIS/AREA or F=(FUEL1+ FUEL2+ FUEL3)/AREA, were calculated in an attempt to reduce the number of independent variables.

3. Stepwise regression and backward elimination were used to help assess the influence of several variables, but were not relied on when choosing the final variables. Hosmer and Lemeshow (1989) have reported that mechanical selection procedures such as these can select models containing irrelevant or noise variables as best models.

4. Pairwise correlation coefficients (Appendix B) for all variables indicated that strong multicollinearity effects should be expected. For this reason, only one variable from each subset of highly correlated variables (correlation > 50 %) was included in each model.

3.2.2. *Criteria for model evaluation*

Three criteria were used to compare the usefulness of the models developed.

1. Criteria based on the estimated parameters.

First, signs of estimated parameters were checked to make sure they agreed with theoretical expectations based on previous knowledge of the fire occurrence problem. Secondly, the Chi-square test and Wald statistic (Ben-Akiva and Lerman 1985) were used to assess the significance of estimated parameters.

2. Criteria based on goodness of fit of the models to the data.

Loftsgaarden and Andrews (1992) recommended the use of the Hosmer and Lemeshow (1989) goodness-of-fit test to assess fit in models with two or more variables. This test was computed by the LOGISTIC procedure, with the usual likelihood ratio test, and the Akaike Information Criterion and Schwartz Criterion statistics. The last two statistics are used when evaluating different models for the same data, and the lower their values, the more accurate the model is (SAS Institute Inc. 1989). These were used as secondary decision criteria in cases where similar values were obtained in two or more models for the Hosmer and Lemeshow goodness-of-fit test statistic, the main criterion.

3. Criteria based on predictive capabilities of the models.

The predictive capabilities of the models were measured using the 2x2 classification table (Table 3.3) of observed and predicted responses (SAS Institute Inc. 1989) as the most important criterion. The total percentage correctly predicted by the table is computed as A+D/A+B+C+D. Several indices of rank correlation between the predicted probabilities and observed responses given by the same procedure (% concordant, Somers' D, Gamma, Tau-a, c) were used as the secondary decision criteria in models with similar classification tables.

Table 3.3. *Classification table of observed and predicted responses.*

Frequency Table Row Pct.	Predicted no-fire	Predicted fire
Observed no-fire	A Specificity	B False Positive
Observed fire	C False Negative	D Sensitivity

The analysis of influential observations and outliers was carried out according to the logistic regression diagnostics developed by Pregibon (1981).

3.3. Results

3.3.1. *Model Selection*

SAS outputs for the best six models are given in Appendix C. They are all valid models; they are very similar in variables chosen and parameter estimates; and they are all parsimonious (only 3-5 variables). None of the 314 observations used for model building were considered to be outliers acording to the diagnostics developed by Pregibon (1981), so none was removed from the analysis. Models 4 and 5 best fitted the data. Yet Model 6 scored the highest in total percentage correctly predicted. But Model 1 was chosen as the best based on its overall good values for most selection criteria (Table 3.4, Table 3.5). Since all the models had similar total percentages correctly predicted, selection was based on the specificity (percentage of no-fires correctly predicted; Table 3.3), and false positive values (percentage of no-fires predicted as fires; Table 3.3), since these were the categories with more cases in

Table 3.4. *Criteria for Model 1 selection.*

response levels = 2
number of observations = 314

Analysis of maximum likelihood estimates

Variable	DF	Parameter estimate	Standard error	Wald Chi-square	Pr> Chi-square
INTERCEPT	1	-4.6048	0.5917	60.5551	0.0001
AREA	1	7.6590	1.2198	39.4268	0.0001
DISTRICT	1	0.7367	0.3429	4.6159	0.0317
BUI	1	2.0478	0.9936	4.2482	0.0393
ISI	1	3.9563	0.6323	39.1474	0.0001

Hosmer and Lemeshow goodness-of-fit test
Goodness-of-fit statistic = 10.94 with 8 DF (p=0.2051)

the classification table after removing the bias introduced by different sampling rates.

The estimated coefficients values and signs were as expected, and they were significant at 0.05 level. The Hosmer and Lemeshow (1989) goodness-of-fit test showed adequate fit of the model to the data (Chi-square value 10.94, p-value 0.2051). This model correctly classified 79.0 % of all the observations, 120 (76.4%) out of 157 fire days, and 128 (81.5%) of the 157 no-fire days, it failed to predict 37 of the fires, and produced 29 false alarms (Table 3.5).

Table 3.5. *Classification table for the model building data: Model 1.*

Frequency Row Pct.	Predicted no-fire	Predicted fire
Observed no-fire	128 81.5%	29 19.5%
Observed fire	37 22.4%	120 76.4%

Model 1 parameter estimates (Table 3.4) were calculated from groups with unequal sampling rates (P_1=157/157 for the fires, P_2=157/7852 for the no-fires), hence an adjustment in the intercept was needed in order to use the model for wildfire prediction (Maddala 1988). After subtracting the correction value from the intercept (3.9122 = ln(157/157)-ln(157/7852)), the probability (P_i) of at least a human-caused fire occurrence happening in any fire occurrence prediction unit in Whitecourt, on any given day, can be established using equation 3.1, where Z_i is a linear function (equation 3.2) of the independent variables for any observation *i*;

$$Z_i = -8.5171 + 7.6590 * AREA_i + 0.7367 * DISTRICT_i + 2.0478 * BUI_i + 3.9563 * ISI_i \quad (3.5)$$

3.3.2. *Strategies for testing the model*

Data collected in 1991 and 1992 were used to test the predictive capabilities of the models for independent data. This independent data included 3,294 new observations, of which only 58 were fire occurrences. In accordance with the actual proportion of fires *versus* no fires in the real-world data, most predicted values of probability of fire occurrence for each zone and day were zero or close to zero in all models.

In the logit model, the summation of predicted probabilities is equal to the total number of observations in which Y=1 (Ben-Akiva and Lerman 1985). For the model building data the summation of probabilities of Y=1 was 157 in all models because that many fire observations were in the 1986-1990 data set. For the independent data the predicted probabilities summed 94 in Model 1, which compared to the actual number of 58 occurrences in 1991-1992 data set indicates a trend with Model 1 to overestimate the number of fires. This trend is visible also in all other models.

Model 4 best predicted the total number of fires in 1991-92 test years (71.7 *vs* 58), but it performed worst in the classification table (71.58% total percentage correctly predicted).

Classification tables (Table 3.3) were used to evaluate the performance of the models in predicting for the 1991-1992 independent data. Computation of classification tables usually involves establishing a probability level (cut-off point) to segregate observations into "likely events" (in this study prospective fire occurrences) and "unlikely events" (prospective no-fire occurrences) (Schuster 1983, Jamnick and Beckett 1987). This probability level is customarily set at 0.5, which is the midpoint of the logistic distribution. 0.5 was the cut-off point used in computing the classification tables for the model building data.

Nevertheless, this cut-off point is arbitrary, and ultimately depends on the objectives for the model or the goals of the user (Jamnick and Beckett 1987). A decision about the "best" probability level involves a trade-off between predicting correctly the fires and predicting correctly the no-fires (Schuster 1983, Jamnick and Beckett 1987). Schuster (1983) suggested this problem is similar to a Type I *versus* Type II statistical problem. The objective in this study was to obtain similar accuracy in predicting both fires and no-fires. By defining an arbitrary probability level of 0.02 in the classification table of Model 1 for the independent data set (Table 3.6), 74.10 % of the total number of new observations, 74.14 % of the fire days, and 74.10 % of the no-fire days were correctly classified.

The classification tables presented in Appendix D show the relative

Table 3.6. *Classification table for an independent data set: Model 1.*

Frequency Row Pct.	Predicted no-fire	Predicted fire
Observed no-fire	2398 74.14%	838 25.90%
Observed fire	15 25.86%	43 74.10%

performance of the best logit models on an independent data set. The cut-off point 0.02 provided the best predictions for the independent data in all 6 models.

3.4. Management Implications

Several binary logit models were successfully developed and validated for the human-caused fire occurrence data available in Whitecourt Forest, Alberta, for the 1986 through 1992 fire seasons. The best model included the covariates of AREA, DISTRICT, ISI, and BUI. This low number of variables indicates a parsimonious, practical model. Its few requirements of input data should make it easy to apply and operate in daily prediction of wildfire occurrence. An additional desirable feature of all the models was obtained from the way the fire occurrence prediction units were defined. The models predict fire occurrence for prediction units close to roads (< 5km) and for prediction units far from roads (>5km). This implies that predictions are given for areas easily accessible by road or areas located in more remote situations for the suppression forces. It also means that predictions are provided for areas more exposed to the public view for detection and areas where detection cannot rely on resources other that those provided by the fire protection agency concerned.

In its application for prediction, though, several considerations must be made:

1. The Fire Weather Index System was designed to provide numerical ratings of fire danger during a snow-free fire season that usually spans from April to October in Whitecourt. Hence, during the winter period, and other days outside of the period in which codes and indices in the FWI are computed, no predictions can be made, since the logit model developed relies on the daily BUI and ISI. Nevertheless, fires occurring out of the fire season are few in numbers and are usually easy to control. These fires are not a big concern for most fire protection agencies.

2. The prediction provided is only fire "Yes/No" for each prediction unit and day. No estimation is given for actual numbers of fires. In the period 1986-1990, there was just one occurrence per unit and day in most instances (124 of 157), but in 17 of the observations in which Y=1 (fire days) there were actually two fires, in four of the observations there were three fires, in one there were four fires and in another there were eight fires. This information about multiple fires on the same day and unit was not considered on the analysis. Nevertheless, placement of suppression resources according to this model should suffice to cope with the current fire loads, since multiple fire situations are so uncommon, and the number of simultaneous fires is low.

3. If desired, the probability that N fires will occur in one of the prediction units on a certain day can be calculated by means of a Poisson process described in Martell *et al.* (1987). The logit model in the present study gives the estimated probability that "one or more fires" will occur in a certain area and day. This

probability can be used to estimate the λ parameter of the Poisson distribution used by Martell *et al.* (1987) to model daily human-caused wildfire occurrence. These authors also determined the probability of having at least one fire occurrence per prediction unit and day by logistic regression analysis. They used the codes and indices in the Fire Weather Index (Van Wagner 1987) as explanatory variables. Their models selection criteria (average score), though, has been questioned by Loftsgaarden and Andrews (1992), who stated that it does not truly indicate "how good the models are or how well they fit the data". Loftsgaarden and Andrews (1992) recommended the use of the Hosmer and Lemeshow (1989) goodness-of-fit test instead, advice that has been followed in this study.

4. Performance of the model depends on the probability value chosen to form the classification table. The value 0.02 was arbitrarily selected in order to obtain balanced accuracy in predicting fires and no-fires. This probability level could, and should be selected based on the objectives of the Fire Agency using the model. A higher value (less false alarms but lower accuracy in predicting the fires) may be preferred if: (1) scarce suppression resources are available, (2) if they are expensive to deploy, or (3) if high protection levels are not critical for values at-risk. A lower probability value (higher accuracy in predicting the fires, but more false alarms) would be preferred for an area where values-at-risk are high and there are abundant and mobile suppression resources readily available. In this case, false alarms would be less of a concern than a fire getting away. Economic considerations should be used when determining the accuracy level to be used in each case.

5. A model that overestimates the number of fires tends to reduce its credibility when applied in prediction. False alarms will occur occasionally regardless of the probability level used to separate fires and no-fires. This problem, though, diminishes in importance when the predicted fires are plotted against the actual fires over time because false alarms very often associated with periods of high risk, which are usually associated with actual fire occurrences (Figure 3.2).

6. No geographic variable was included in the final model, and only a summation of areas of dangerous fuels divided by unit area (F) and the average distance to road divided by unit area (ARRODIS) were present in two of the other models (Appendix C), even though several have been found to be significantly related to human-caused fire occurrence (Chou *et al.* 1993, Chapter II in this thesis). The small range in their values may have masked the true importance of the variables thus resulting in their elimination. Only eight values (one for each prediction unit) were included for each geographic variable across the database for analysis. A larger study area may have encompassed more geographic variability, but the data required for such an analysis were not available. Also, it is unrealistic to assume that the area covered by fuel types in any zone would remain unchanged for a period of seven years or that new roads were not built, but again, none of these data were available on a yearly basis.

7. Todd and Kourtz (1991) suggested that patterns of fire starts by humans can change very quickly. The model proposed in this study could be re-evaluated/updated to account for new trends every year, but no provision was made to account for new

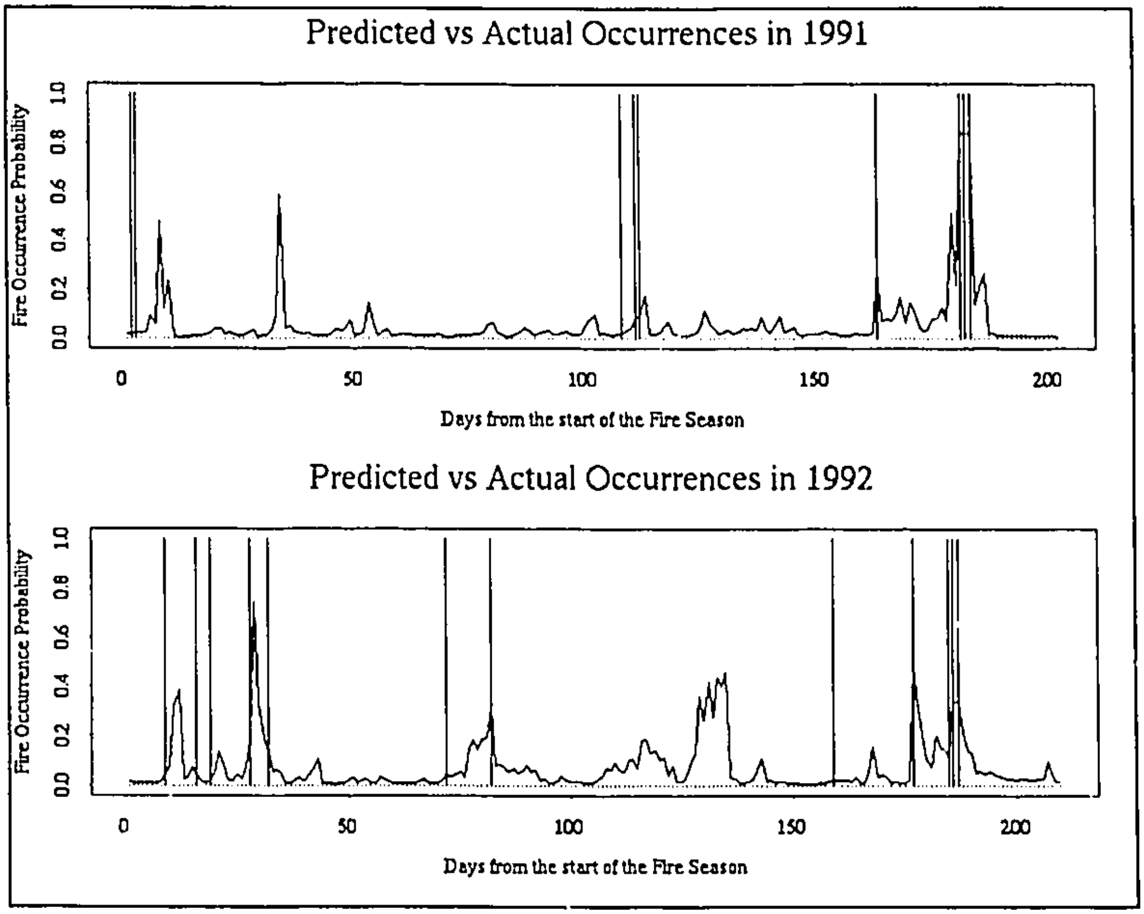


Figure 3.2. Plots of predicted vs actual fire occurrences in District 2, zone ≤ 5 km from roads. Predictions fluctuate between 0 and 1, actual fire occurrences are represented in the graph by vertical lines extending from 0 to 1.

trends throughout the current fire season.

8. Weather variables from previous days strongly influence the conditions and likelihood of fire occurrence in the current period, a fact that would suggest the need for including "lagged" variables in the logit models (for instance, "amount of precipitation in the previous three days"). The reason why no "lagged" variable was included is that this past weather influence has been built-in in the tables for computing codes and indices in the Fire Weather Index (Van Wagner 1987). In calculating its daily outputs, this system includes some outputs of the previous day (FFMC, DMC, and DC). Nevertheless, this introduces serial correlation in the data[1]. For this study serial correlation among independent variables was ignored.

9. Predictions of fire occurrence for the next day have to rely on forecasted ISI and BUI values. The use of forecasted weather indices can affect the performance of the model, because predictions become dependent on the reliability of the forecasted weather indices. Users should be aware of the effect a forecast can have on this or any other fire occurrence prediction model (Tithecott 1990).

None of these considerations prevent application of the models for wildfire occurrence prediction within their limits of applicability. Logit models can be used to predict daily human-caused fire occurrence in Whitecourt Forest. Predictions will probably be much improved once accurate and updated geographic data are obtained

---

[1] Exploratory data analysis using 1986 data in District 1 of the Whitecourt Forest suggests the current values were serially correlated to past values as far back as 8 days for FFMC, 14 days for DMC, 24 days for DC, 15 days for BUI, 9 days for ISI, and 12 days for FWI.

and made available for this and other applications.

3.5. References

Ben-Akiva, M. and S.R. Lerman. 1985. Discrete choice analysis: theory and application to travel demand. The MIT Press, Cambridge, Massachusetts, USA. 390 pp.

Chou, Y.H., R. A. Minnich, and R.A. Chase. 1993. Mapping probability of fire occurrence in San Jacinto Mountains, California, USA. Environmental Management 17: 129-140.

Cox, D.R. and E.J. Snell. 1989. Analysis of binary data. Second Edition. Chapman and Hall, New York, New York, USA. 236 pp.

Deeming, J.E., R.E. Burgan, and J.D.Cohen. 1977. The National Fire Danger Rating System-1978. U.S.D.A. Forest Service, Intermountain Forest and Range Experiment Station, General Technical Report INT-39. 63 pp.

Environmental Sytems Research Institute Inc. 1991. Understanding GIS: The ARC/INFO system. Environmental Systems Research Institute Inc., Redlands, California, USA.

Hosmer, D.W. Jr. and S. Lemeshow. 1989. Applied logistic regression. John Wiley & Sons, New York, New York, USA, 307 pp.

Jamnick, M.S., and D.R. Beckett. 1987. A logit analysis of private woodlot owner's harvesting decisions in New Brunswick. Canadian Journal of Forest Research 18: 330-336.

Loftsgaarden, D.O. and P.L. Andrews. 1992. Constructing and testing logistic regression models for binary data: Applications to the National Fire Danger Rating System. U.S.D.A. Forest Service. Intermountain Research Station, General Technical Report INT-286. 36 pp.

Maddala, G.S. 1983. Limited-dependent and qualitative variables in econometrics. Econometric Society Monographs;3. Cambridge University Press, New York, New York, USA. 401 pp.

Maddala, G.S. 1988. Introduction to econometrics. Macmillan Publishing Co., New York, New York, USA. 412 pp.

Martell, D.L., S. Otukol, and B.J. Stocks. 1987. A logistic model for predicting daily people-caused fire occurrence in Ontario. Canadian Journal of Forest Research 17: 394-401.

Pregibon, D. 1981. Logistic Regression Diagnostics. The Annals of Statistic 9: 705-724.

Prentice, R.L. 1986. A case-cohort design for epidemiologic cohort studies and disease prevention trials. Biometrika 73: 1-11.

SAS Institute Inc. 1989. SAS/STAT User's Guide, Version 6, Fourth Edition, Volume 2. SAS Institute Inc., Cary, North Carolina, USA. 846 pp.

Schuster, E.G. 1983. Evaluating nonindustrial private landowners for forestry assistance programs: a logistic regression approach. U.S.D.A. Forest Service Intermountain Forest and Range Experiment Station, Research Paper INT-320. 7 pp.

Stynes, D.J., and G.L. Peterson. 1984. A review of logit models with implications for modeling recreation choices. Journal of Leisure Research 16: 295-310.

Tithecott, A.G. 1990. 1989 People-caused fire occurrence prediction project. Ontario Ministry of Natural Resources, Aviation and Fire Management Centre. Publication 275. 18 pp.

Tithecott, A. 1993. Application of fire occurrence prediction models in Ontario's fire management program. Pages 57-72 in Proceedings Eight Central Region Fire Weather Committee Scientific and Technical Seminar, April 3, 1992, Winnipeg, Manitoba. Forestry Canada, Northwest Region, Northern Forestry Centre, Edmonton, Alberta, Canada. 78 pp.

Todd, B., and P.H. Kourtz. 1991. Predicting the daily occurrence of people-caused forest fires. Forestry Canada, Petawawa National Forestry Institute, Chalk River, Ontario, Canada. Information Report PI-X-103. 16 pp.

Van Wagner, C.E. 1987. Development and Structure of the Canadian Forest Fire Weather Index. Canadian Forestry Service, Forestry Technical Report No 35. 36 pp.

CHAPTER IV

Applying Neural Network Technology to Human-Caused Wildfire Occurrence Prediction

4.1. Introduction

Artificial Neural Networks (ANN) are a new technology for processing information. This technology is modelled after the computing system thought to be used by the human brain (Klimasauskas 1991a). Neural nets, as they are frequently referred to, are composed of many processing elements (PE), the artificial equivalents of neurons. These elements are capable of relatively simple operations, and are usually grouped in layers or slabs. There is typically an input layer where data are presented to the network, an output layer that provides the response of the network to the input data, and one or more "hidden" layers in between (NeuralWare Inc. 1991). The PEs are interconnected through connections of variable strength, called "weights", and operate in a parallel manner. A typical network, then, "consists of a sequence of layers or slabs with full or random connections between successive layers" (NeuralWare Inc. 1991).

Artificial Neural Networks have a large processing capacity because of the way they are structured. The capability to perform intelligent tasks such as: (1) learning by example, (2) generalizing learned knowledge, and (3) recognizing patterns make this technology extremely powerful and useful. These capabilities are achieved through a "learning phase" in the operation of a network. "Learning" is a process of modifying the connecting weights in response to data being presented at the input layer (and the output layer, optionally), according to a predefined mathematical algorithm or "learning rules" (NeuralWare Inc. 1991). In this way, the network shapes itself to reflect the relationships between inputs and outputs in a "training" data set (Klimasauskas 1991a).

Regression analysis is one of several statistical approximation methods which can accurately approximate functions or describe relationships. The advantages of ANN are that they do not require assumptions on the underlying distribution of the data, and allow variable interaction and non-linearity in the data (NeuralWare Inc. 1989). Neural networks may be more robust and predict better than statistical models when non-linear relationships are studied and distributions are non-normal (Neuralware Inc. 1989, Cook *et al.* 1991). In some applications, ANN have achieved reductions in error ranging from 5 to 50%; in others where the problem was well understood, it merely matched or approximated the corresponding statistical model developed for the same application (Klimasauskas 1991a).

ANN have been used successfully for signal and language processing, image compression, character recognition, combinatorial problems and servo control (NeuralWare Inc. 1991). Recent applications in the natural resources field included classifying data for land-use planning (Yin and Xu 1991), satellite imagery interpretation for cloud-type analysis (Peak and Tag 1992), modeling tree survival (Guan and Gertner 1991), and for several forecasting problems such as predicting solar flares (Klimasauskas 1991a).

4.2. Objectives

The objective in the present study was to test if neural network technology could be used to improve current human-caused fire occurrence predictions. The same data were used in this analysis as in the binary logit analysis applied to the fire occurrence prediction problem that was described in Chapter III, thus insuring comparability between the results of both model. An attempt was made to develop a simple network model for wildfire prediction from 20 geographic and temporal variables. Non-relevant variables would be eliminated, using sensitivity analysis on inputs, throughout the network development process. Guiver and Klimasauskas (1991) advise that "the less superfluous information the network is given, the better it is able to latch on to the true relationships in the data". These authors recommend the use of experts when selecting inputs. For this reason, two neural network models were also developed to match the best daily human-caused fire occurrence prediction logit models (Model 1 and Model 2) described in Chapter III, using their same variables as inputs. In this case, the expertise used for input selection was provided by the logit analysis.

4.3. Neural network model choice

"Back-Propagation feedforward networks with non-linear PEs" have become the standard choice for modeling, forecasting, and classification (Klimasauskas 1991b). In this type of network, there are no feedback connections among layers (NeuralWare Inc. 1991). The PEs reach an internal activity level by summing the weighted inputs given. Then the summation is modified by a continuous transfer function (usually a sigmoid) and passes to the output path of the PE (Neuralware Inc. 1991).

The learning rule in back-propagation networks is the generalized delta rule developed by Rumelhart *et al.* (1986). A back-propagation network uses input data to compute its own output, then compares it with the desired output. If no error is being made, no learning occurs (Rumelhart *et al.* 1986). If the output is in error, back-propagation assumes that all PEs and connections are to blame. To reduce the difference between desired and actual output, the weights are changed by propagating the output error backward through the connections to the previous layer (NeuralWare Inc. 1991). The process of presenting pairs of input/output vectors to the network for weights update is continued until a single set of weights is found that produces zero error (or an error sufficiently close to zero, such as 0.001) for all pairs presented. The network then, is said to have "converged" (NeuralWare Inc. 1991).

4.4. Model Development

The software package NeuralWorks Professional II/plus (NeuralWare Inc. 1991) was used for developing the networks. All data processing was done on a SUN Sparcstation 10 platform. This tool allowed for experimentation with different network architectures, learning rules, and transfer functions in the PEs. Klimasauskas (1991a) served as the authority for the procedures used in developing the neural nets reported in this study.

1. Data collection.

Abundant geographic, weather and fire data were available for the Whitecourt Forest, Alberta (Appendix A), so this forest was selected as the study area. These data were available from the Forest Protection Branch of the Alberta Forest Service, Edmonton; and Forestry Canada, Northwest Region, also in Edmonton.

Previous work (Chapter II) determined that human-caused fire occurrence in Whitecourt Forest is highly related to the following geographic and temporal variables: (1) distance to nearest road, (2) town, and (3) campsite, (4) topographical elevation, (5) fuels, (6) forest commerciality, (7) forest district, (8) all codes and indices in the Canadian Forest Fire Weather Index System (Van Wagner 1987) (except the Drought Code), (9) relative humidity, (10) wind speed, and (11) month.

A Geographic Information System (ARC/INFO, Environmental Systems Research Institute Inc. 1991), was used to map eight fire occurrence prediction units in the study area. Each of the four administrative forest districts within the Whitecourt Forest was divided in two zones: areas $\leq$ 5 km from a road, and areas > 5 km from a road (Figure 4.1). Five kilometres was found to be the threshold distance from roads within which 90% of all fires start (Chapter II). These fire occurrence prediction units ranged in area from 805 to 4,660 km^2. These areas were in general fragmented within each Forest District, but some were contiguous.

Each prediction unit was coded with information pertaining to the following variables using ARC/INFO software: (1) unit area (AREA) in km^2, (2) average distance to roads (ROADDIS) in km, (3) average distance to towns (TOWN) in km, (4) average distance to campsites (CAMP) in km, (5) average topographic elevation

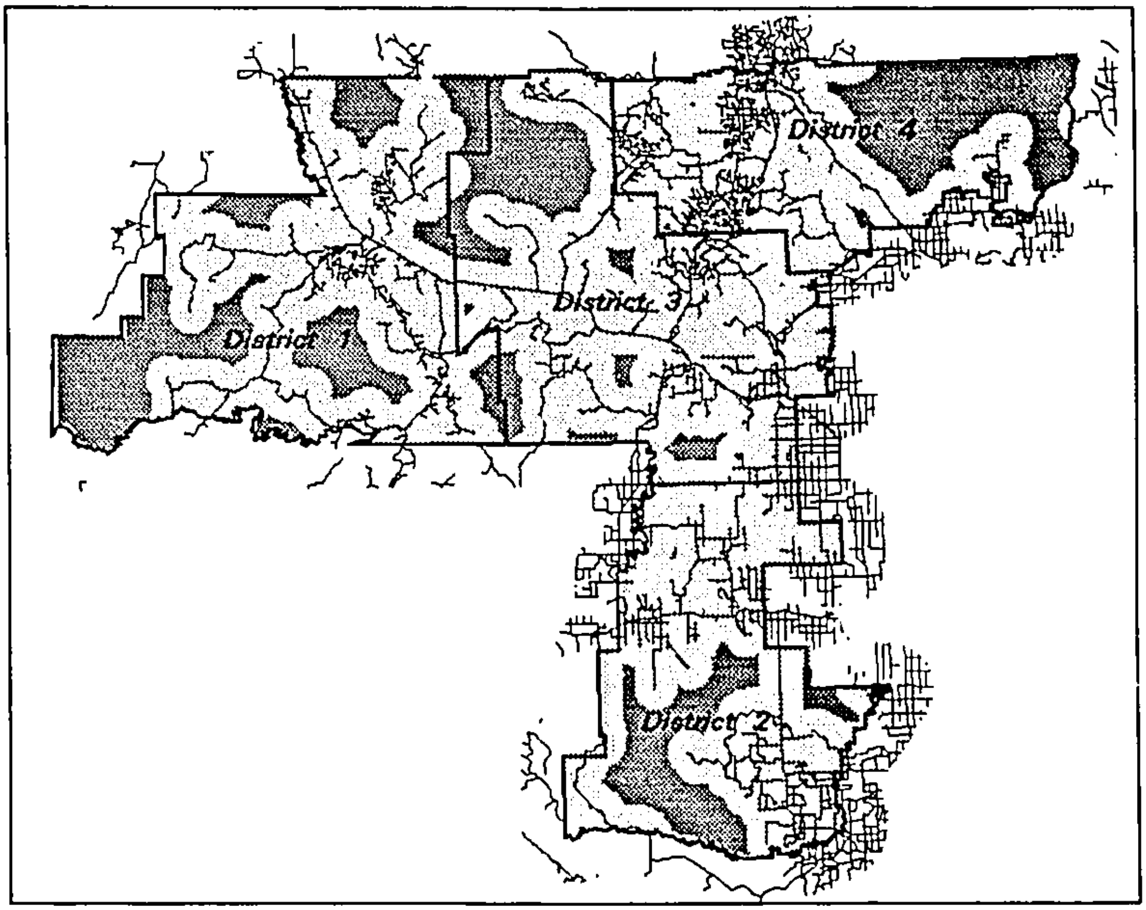


Figure 4.1. The four forest districts were divided in areas ≤ 5 km from a road and areas > 5 km from a road, for a total of eight fire occurrence prediction units. Also shown is the road network.

(ELEV) in m, (6) total area privately owned (PPA), (7) total area of highly uncommercial forest value (COMH), (8) total area of deciduous fuel (FUEL1), (9) total area of grass fuel (FUEL2), (10) total area of slash fuel (FUEL3), all in km^2. Two dummy variables were also included. The dummy variable (11) ROAD which had a value of 1 for zone within 5 km from roads, and 0 for zone farther than 5 km from roads, was included because distance to road is the most significant geographic variable in explaining fire occurrences in the Whitecourt Forest (Chapter II, this thesis). The second dummy variable (12) for DISTRICT had a value of 1 for area in District 2, and 0 for area in any other District. This was done because District 2 suffers from higher human pressure in the form of developmental activities, and is closer to Edmonton.

The database for the analysis included observations for each day in the fire seasons 1986-1990 (April-October), in each of the prediction units described above. This five-year period was chosen because it coincided with the digitizing date of the geographic data. The weather variables assigned to each prediction unit every day were averaged from the weather stations available in the district where the unit was located. These variables were: Fine Fuel Moisture Code (FFMC), Duff Moisture Code (DMC), Build-up Index (BUI), Initial Spread Index (ISI), Fire Weather Index (FWI), as defined by Van Wagner (1987), and relative humidity (RH) and wind speed (WS). A dummy variable MONTH, with a value of 1 for days in April and May, and 0 for days between June and October, was used to account for the traditional high fire occurrence rate in Spring. The binary dependent variable OCCUR was assigned value

1 if at least one fire occurred in the management unit and day of the observation, and 0 value if there was no fire. The result of this stratification was a data base of 8,009 observations (±200 days/fire season x 5 fire seasons x 8 units), of which only 157 were fires.

2. Data separation into training and test sets.

A random sample of 157 no-fire observations was obtained from the data base of 8,009 observations, and analyzed with the 157 fire observations. All training data sets included these same 314 daily observations in the eight prediction units in the Whitecourt Forest for the period 1986-1990. This same data set had been used in the logit analysis, Chapter III, but this was not the reason to chose this proportion of fire/no-fire observations in the training data sets. Rather, it was important to distribute the number of training observations *equally* in each output category (Klimasauskas 1991b) to avoid having the networks learn that by classifying all outputs as no-fires, they would be right most of the time. The data were presented to the networks in ASCII format. For comparison with Logit Model 1, an ASCII file was prepared including only the variables: AREA, DISTRICT, Initial Spread Index (ISI), and Build-Up Index (BUI) (Van Wagner 1987), and the dependent variable OCCUR of value 1 or 0 depending on if at least a fire occurred in that unit and day or not. For comparison with Logit Model 2, the training data set included the variables AREA, DISTRICT, and Fire Weather Index (FWI) (Van Wagner 1987), and OCCUR.

Test sets were also required to avoid overtraining. Overtraining may result in the network "memorizing" the data, or learning very specific and non-desirable features in the training data set (Everly 1993). Overtrained networks are not able to generalize the learned knowledge, and predict properly for new data. This problem can be avoided by checking the learning process periodically and evaluating the performance of the network in predicting fire occurrence for an independent test data set at every step (Everly 1993). The ideal test data should be representative of the real-world conditions (Klimasauskas 1991b), and should not include data previously used in as part of the training exercise. Hence, fire occurrence data from 1991-1992 were used in testing the networks. This data set included 3,294 new observations, of which 58 were fire observations.

3. Data transformation into network appropriate inputs.

Numeric inputs such as AREA, ROADDIS, PPA, ISI, BUI, or WS, were re-scaled for input to the network (typically in the ranges from 0 to 1, or -1 to 1) by specifying the MinMax Table option in the Professional II/plus. DISTRICT, ROAD, MONTH and OCCUR were categorical. Categorical inputs were encoded using a "one of N" encoding, which means that each category was assigned to a separate input node (Klimasauskas 1991b) where 1 would be coded as (1,0) and 0 would be coded as (0,1).

4. Select, train, and test the network.

Exploratory analysis was used to select the more suitable transfer functions (sigmoid or hyperbolic tangent) and learning rules (delta rule, cumulative delta rule, or normalized cumulative delta rule) (NeuralWare Inc. 1991). Trials indicated that the best results for this problem could be achieved by using the sigmoid transfer function, and the generalized delta rule with very small learning rates, where the α in output layer=0.01-0.005, and the α in hidden layer=0.1-0.05, and the momentum= 0.4-0.7. A variation of the back-propagation algorithm (Fast Back-propagation (NeuralWare Inc. 1991)) was used to speed training.

Different network architectures were tried; yet, all trials employed just one hidden layer. Klimasauskas (1991c) suggests most problems can be solved with one hidden layer. The number of PEs in the hidden layer (H) were estimated from the following guiding formulae (Neuralware Inc. 1989):

$$H = (1/2)*(I+O)$$

$$H = (I+O)^{1/2}$$

$$H = I*O$$

$$H = (I+O)^2$$

where I=number of inputs (20 for the general model; 5 for model 1; or 4 for model 2) and O=number of outputs (2 nodes to map the output 2-dimensional vector).

Network architectures for the general model with all twenty variables ranged from 5 to 48 PEs in the hidden layer, according to the formulae above. Network architectures for Model 1 had 3,4,5,7, and 10 PEs in the hidden layer, also according to the formulae. Network architectures for Model 2 were built with 2,3,4,6, and 8 PEs in the hidden layer. Layers were fully connected to the previous one, but connections were not allowed to jump layers, since previous trials showed that this did not improved performance of the networks. Five copies of each of the several networks thus developed were created and their weights were randomized to obtain different starting points for training.

The training process was conducted differently for the networks with all variables and for the networks created for comparison with Logit Models 1 and 2. The observations were presented to the 50 comparison networks randomly by selecting the "File Rand." option in the IO Menu in NeuralWorks Professional II/plus. Every 1,000 iterations, the performance of the network being trained was tested with respect to the independent test set, and the network was saved only if improvement had taken place. An improvement was defined as a reduction in the Root Mean Square (RMS) Error in the output layer for the test set. The RMS error adds up the squared errors of each output PE (two in this case), the total is divided by the number of output PEs to average, and then takes the square root of the average (Neuralware Inc. 1991). Classification tables of observed and predicted responses were then calculated to select the best network within each group of networks with the same architecture.

For the network model with all 20 variables, several architectures were built, each with five randomizations. Training examples were also presented to the networks in random order. No pre-set limit of iterations was imposed. The networks were trained to their best performance by testing every 1,000 interactions on the test data set. Then, the "Explain" function within NeuralWorks Professional II/plus was used to perform sensitivity analysis on the inputs. This function allows one to identify inputs that consistently exhibit very little influence on the outputs (NeuralWare Inc. 1991). These inputs were dropped, and the model building process was re-started with a smaller set of inputs. New architectures were built, with different randomizations, and trained. Different combinations of the inputs that exhibited the most effect on the output were also explored. Classification tables of observed and predicted responses (Table 3.3, Chapter III) were also calculated to select the best network within each group of networks with the same architecture. Selection of the best network of all developed was based on its best performance on the training and test sets. A network-specific C code was generated by making use of the "FlashCode" option in the Professional II/plus for deployment of the best neural network model.

4.5. Results and Discussion

Sensitivity analysis of the inputs in the general model led to the selection of RH, FWI, ISI, FFMC, and ROAD as the more consistent contributors to the good performing networks. DMC, BUI, and MONTH were also important but contributed less to the performance of the networks derived. Only once out of 15 networks, was DISTRICT the main contributor. Hence, several network architectures were built including the five selected variables, plus AREA. AREA was included because observations are tied to prediction units of variable size and it was considered a key factor in the prediction. Since it is recommended to have at least 5 training examples for each weight in the network (Klimasauskas 199[illegible]), network sizes had to be limited to a maximum of 63 weights (314 training examples/5 = approx. 63). This limitation applied also to the networks for comparison with Logit Models 1 and 2. Subsets of the selected variables were chosen for network development in an attempt to obtain small-sized networks to be trained more efficiently. These subsets were formed by excluding highly correlated variables one at a time. Performance of the networks is not affected by high correlations among the input variables (NeuralWare Inc. 1989). This elimination was motivated only by a desire to reduce redundant information in the inputs and achieve a reduction in network size and complexity.

The classification tables for the various networks which best modelled the variables or the architectures used as part of this analysis are presented in Appendix E. All the networks presented smooth RMS error graphs, which descended slowly during learning, but none reached convergence. In fact, the RMS error dropped only from 0.31/0.30 to 0.24/0.23 over the training phase, but this was not considered to be a problem since a zero error in this case would likely indicate memorization of the training data. Training was completed after 19,000 iterations for most of the networks. The entire training set of 314 observations was presented to the networks some 60 times in those 19,000 iterations (19,000/314=60), each time in a different random order. Performance did not improve, in general, beyond this point. The weights for each layer grew at the same rate through the operation of the network adopting a bell shape when plotted in a weights histogram. There was no PE saturation during the initial stages of training to indicate values outside the range of the transfer function. These two facts, plus the fact that the RMS graphs descended slowly suggested that the networks behaved properly throughout the training process. Professional II/plus provided several graphs or instruments to monitor training progress. These instruments are displayed in Figure 4.2 which presents the best network developed for the training data set.

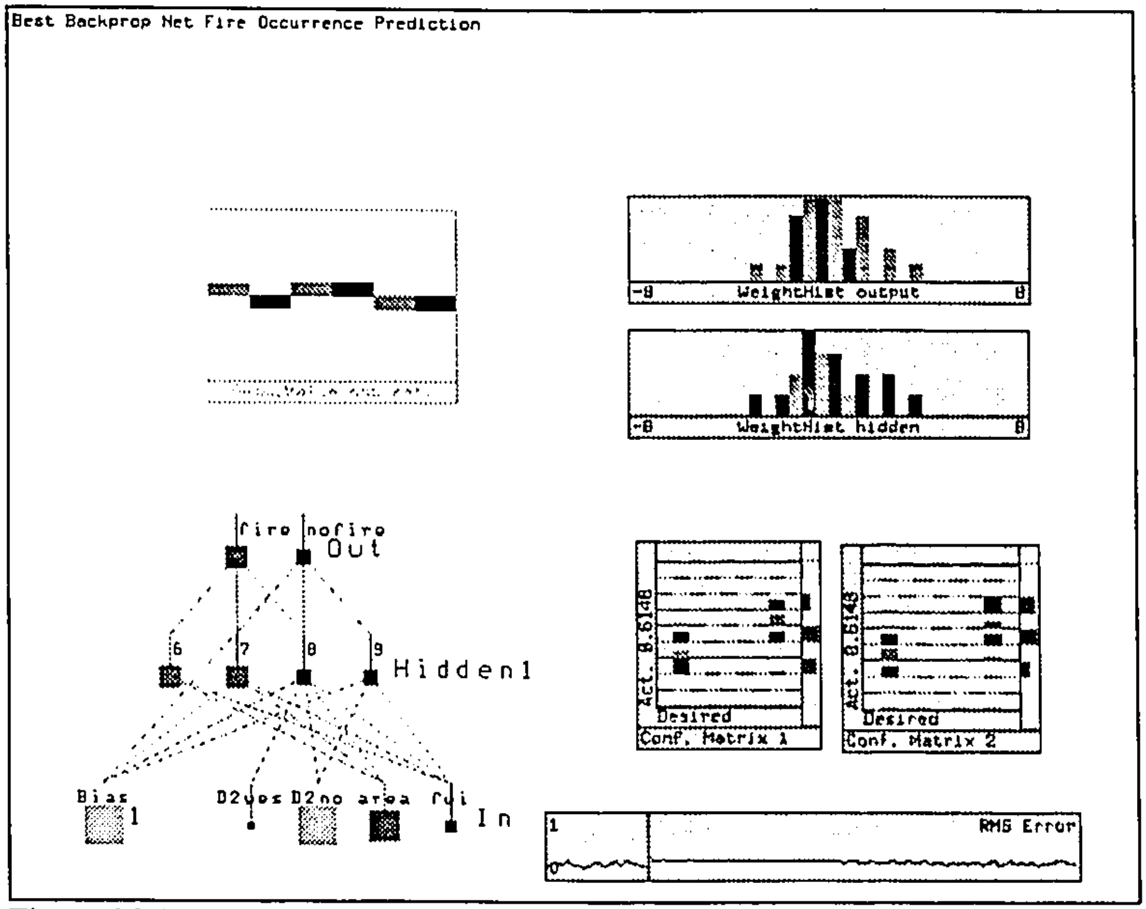


Figure 4.2. Best neural network model, and instruments provided to monitor and evaluate training.

The best network used the input variables of AREA, DISTRICT, and FWI, which were the same variables selected for Logit Model 2. It had four input nodes and two output nodes (as a result of the one of N encoding). This network was able to predict correctly 85% (133) of the no-fire observations, and 78% of the fire observations (122) in the training data set. The total percentage correctly predicted was 81%. The classification table for the training data set is presented in Table 4.1.

Table 4.1. *Classification table to evaluate the performance of the best neural network model on the training data set.*

Frequency Row Pct.	Predicted no-fire	Predicted fire
Observed no-fire	133 84.71	24 15.29
Observed fire	35 22.29	122 77.70

For the test set the total percentage correctly predicted was 76%. The network was able to classify correctly 76% of the no-fire observations and 75.8% of the fire observations in the period 1991-1992. The classification table for the test set of this model is presented in Table 4.2.

The classification tables were calculated assuming the Huang and Lippmann (1987) criterion for classifying network outputs on encodings similar to the ones in this study. According to this criterion, the output node with the largest value over 0.5 was the correct one (Huang and Lippmann 1987, Cook *et al.* 1991). So, if the network produced an output of (0.7, 0.1), it was considered equivalent to (1,0), and classified as a predicted fire. Likewise, a network output of (0.45, 0.65), for instance,

Table 4.2. *Classification table to evaluate the performance of the best neural network model on an independent test set.*

Frequency Row Pct.	Predicted no-fire	Predicted fire
Observed no-fire	2462 76.08	774 23.92
Observed fire	14 24.14	44 75.86

would be classified as a no-fire prediction.

The network-specific C code generated for the Professional II/plus for deployment of the best network is presented in Appendix F.

The improvement in predictions provided by the ANN model with respect to more traditional systems was not as dramatic as it has been in other similar applications. The total percentage correctly predicted by the best network for an independent data set improved only by 2% with respect to the total percentage correctly predicted given by the best logit model available for the Whitecourt Forest (Table 3.6, Chapter III) for the same independent data. Data available for model building were limited in time and space, and data are the most important factor in determining a model's performance.

Any model is only as good as the data it relies on. The fact that two very different techniques, the logit models in Chapter III and the ANN model in this Chapter, reached almost identical accuracy in predicting human-caused wildfire occurrence in the Whitecourt Forest seem to indicate limitations in the data, not in the techniques used to develop the models.

4.6. Management Implications

The limitations of this neural network model must be weighted against its usefulness before it could be used for daily human-caused wildfire occurrence prediction. Some of these limitations are shared by other prediction models of wildfire occurrence developed before (such as the logit models in Chapter III).

1. This model can not provide predictions for the days outside of the fire season period because codes and indices produced by the Canadian Forest Fire Weather Index System (CFFWIS) are not computed for these days. This limitation is unavoidable in any model that relies on the CFFWIS. Nevertheless, fires occurring out of the fire season are infrequent and usually weather conditions make it easy to control such fires. This objection does not reduce the benefits of including among the variables for prediction, codes and indices developed to provide a general estimation of forest fire danger in Canada (Van Wagner 1987).

2. This model do not estimate the actual number of fires to occur for each prediction unit and day, only that wildfire (one or more) would or would not occur. Nevertheless, this is not a serious limitation, since ANN can be built to predict numbers of fires, when enough cases or data are available for training in each desired category. Currently, there is a shortage of data, which prevents the development of such a sophisticated model. In the five-year period considered, only 23 out of 157 fire observations included more than one fire. Among those 23, only 6 included more than two fires. It can be argued that a larger area or a longer period of time should have been considered for the analysis, but the geographic data needed for such an analysis

were not available.

3. The number of "false alarms" (no-fire observations predicted as fires) given by the model was high (774 over a two-year period with 3294 observations). This was considered to be the worst pitfall in using the model for daily fire occurrence prediction. False alarms reduce credibility in the prediction system among fire management personnel. Neural network parameters could be tuned to reduce "false positives" by using "non-standard error tables" during training (Klimasauskas 1991d). This is an advantage these models possess over logit models. Likewise, this technique can be used to reduce false negatives (fires the system failed to predict). Economic considerations and priorities set by the corresponding fire protection agency indicate which way the analysis should proceed. Inclusion of these economic goals in ANN development was considered to be beyond the scope of this study.

4. The time series character of the fire occurrence data can be incorporated in the neural network development process. ANN have been successfully applied to prediction problems involving time series data (financial applications, for instance). In this study, however, build-up of dangerous conditions in the fire environment due to past weather circumstances was considered to be accounted for by the codes and indices of the CFFWIS. Yesterday's FFMC, DMC, and DC are inputs for calculation of today's FFMC, DMC, and DC in this system. Yet, this introduces problems for such regression analyses as the logit, due to serial correlation in the data. These problems add to the difficulties in the regression analyses due to high correlations among the geographic variables, and among the temporal variables (multicollinearity).

ANN have an advantage in this study over statistical techniques because the performance of the ANN is not affected by high correlations among the variables used for input to the network (NeuralWare 1989).

5. Several back-propagation neural networks were successfully developed for the human-caused fire occurrence problem. Most of them achieved performance levels comparable or superior to that of logit models in Chapter III for predicting fire occurrences. This study indicated that ANN could be used to predict fire occurrence at least as well as logit models. The best of these networks outperformed the best logit models in the classification table of observed versus predicted responses for the model building data and the test data. However, this best network benefitted from the expertise provided by the logit analysis in selecting its inputs. This apparently confirms the importance placed on expert guidance to select ANN's inputs by Guiver and Klimasauskas (1991).

6. An advantage this model shares with the logit models in Chapter III over models developed elsewhere is that the fire management units considered for prediction were relatively small (all less that 5,000 km^2). This is a desirable feature in a prediction model from the point of view of the fire protection agency concerned. Furthermore, the division of areas within a District in zones < 5km from roads, and zones > 5km from roads, provides an indication of accessibility or remoteness to the fires should they occur, and of their probability of being easily detected by unplanned detection resources.

7. Some considerations must be made with respect to the deployment and

maintenance of the developed model. Professional II/plus features a "FlashCode" command to produce a network-specific generated C code for deployment of the best network (in Appendix F). Managers must consider that the acceptance of "black box" systems such as this one among all possible users would be expected to face more difficulties than the acceptance of a more "familiar" system, such as a regression equation. Also, maintenance levels should be similar for both the ANN model or any other model, since updating of any model is driven by fire occurrence trends (changes in land use patterns, for instance).

Logit models have been successfully built for human-caused wildfire occurrence prediction in the Whitecourt Forest (Chapter III, this thesis) and other geographic locations (Martell *et al.* 1987, Chou *et al.* 1990). Therefore, perhaps one would ask the question: "Should this neural network model be recommended for field testing with preference to others available to managers?". In deploying a neural network instead of a more traditional statistical system two factors must be considered (Klimasauskas 1991d): (1) the improved performance of the ANN, if any, and (2) the additional computational cost that implies executing the ANN. The best ANN developed in this study outperformed the best logit model available for the Whitecourt Forest only by 2% in total percentage correctly predicted, and predictions involved a considerable computational cost. Predictions with the logit model required a few simple operations that even a pocket calculator could handle. It would appear that the cost-effective system for deployment in this case would be the logit model.

However, the full potential of neural computing was not explored in this study.

There are many types of ANN with specialized feedforward architectures (self-organizing maps, for instance) or automatic learning rules (delta bar delta, directed random search, and so on) that have been developed for prediction. The problem is that there is a general lack of bibliographic references on the use of those newly developed networks. The process of network building remains an art, even without considering the difficulty of selecting inputs. But as this technology evolves, many applications will benefit from what Klimasauskas (1991a) calls its "ability to approximate complex mathematical mappings". This study suggests that fire occurrence prediction will be one of those applications.

4.7. References

Chou, Y.H., R.A. Minnich, L.A. Salazar, J.D. Power, and R.J. Dezzani. 1990. Spatial Autocorrelation of Wildfire Distribution in the Idyllwild Quadrangle, San Jacinto Mountain, California. Photogrammetric Engineering & Remote Sensing. 56: 1507-1513.

Cook, D.F, J.G. Massey, and R.E. Shannon. 1991. A neural network to predict particleboard manufacturing process parameters. Forest Science 37: 1463-1478.

Environmental Systems Research Institute Inc. 1991. Understanding GIS: The ARC/INFO system. Environmental Systems Research Institute Inc., Redlands, California, USA.

Everly, R. 1993. NeuralWorks tips. NeuralWare Connection 1(1): 9-11.

Guan, B.T. and G. Gertner. 1991. Modeling red pine tree survival with an artificial neural network. Forest Science 37: 1429-1440.

Guiver, J.P. and C.C. Klimasauskas. 1991. Applying neural networks. Part IV: Improving performance. PC Artificial Intelligence July/August 1991: 34-41.

Huang, W.Y. and R.P. Lippmann. 1987. Comparisons between neural net and conventional classifiers. Pages 485-493 in: IEEE International Conference on Neural Networks, San Diego, California.

Klimasauskas, C.C. 1991a. Applying neural networks. Part I: An overview of the Series. PC Artificial Intelligence January/February 1991: 30-33.

Klimasauskas, C.C. 1991b. Applying neural networks. Part II: A walk through the application process. PC Artificial Intelligence March/April 1991: 2734.

Klimasauskas, C.C. 1991c. Applying neural networks. Part III: Training a neural network. PC Artificial Intelligence May/June 1991: 20-24.

Klimasauskas, C.C. 1991d. Applying neural networks. Part VI: Special topics. PC Artificial Intelligence November/December 1991: 46-49.

Martell, D.L., S. Otukol, and B.J. Stocks. 1987. A logistic model for predicting daily people-caused forest fire occurrence in Ontario. Canadian Journal of Forest Research 17: 394-401.

NeuralWare Inc. 1989. Applying neural networks in business, industry and government. NeuralWare Training Group. Pennsylvania, USA.

NeuralWare Inc. 1991. Neural Computing. NeuralWorks Professional II/plus and NeuralWorks Explorer. Pittsburgh, Pennsylvania, USA. 358 pp.

Peak, J.E. and P.M. Tag. 1992. Toward automated interpretation of satellite imagery for Navy shipboard applications. Bulletin of the American Meteorological Society 73: 995-1008.

Rumelhart, D.E., G.E. Hinton and R.J. Williams. 1986. Learning internal representations by error propagation. Parallel distributed processing: Explorations in the microstructures of cognition, Vol. 1: Foundations, Rumelhart, D.E., and J.L. McClelland (eds). MIT Press, Cambridge, Massachussets, USA.

Van Wagner, C.E. 1987. Development and structure of the Canadian Forest Fire Weather Index. Canadian Forestry Service. Forestry Technical Report No. 35. 36 pp.

Yin, Y. and Xu, X. 1991. Applying neural net technology for multi-objective land use planning. Journal of Environmental Management 32: 349-356.

CHAPTER V

Conclusions

The present study aimed to model daily human-caused wildfire occurrence in Whitecourt Forest, Alberta (Appendix A) utilizing geographic and temporal variables from the forest environment. A preliminary study based on the Chi-square goodness-of-fit test identified the following geographic and temporal variables as relevant to the human-caused fire problem in this Forest: (1) distance to closest road, (2) land ownership, (3) distance to closest town, (4) distance to closest campsite, (5) elevation, (6) location in a certain forest district, (7) fuels, (8) forest commercial value, age, and height, and (9) Initial Spread Index, (10) Fire Weather Index, (11) Fine Fuel Moisture Code, (12) relative humidity, (13) month, (14) Duff Moisture Code, (15) Build-Up Index, and (16) wind speed. The tests also suggested the following geographic and temporal variables to be irrelevant in human-caused wildfire occurrence: (1) distance to closest lake, (2) forest density, (3) location in area seen by lookout towers, (4) distance to closest lookout tower, (5) slope, (6) distance to closest river, and (7) aspect, and (8) DC, (9) visibility, (10) temperature, and (11) weekday.

A Map of Risk for the Whitecourt Forest was delineated by using a Geographic Information System and the geographic variables related to wildfire occurrence. Such a map could have important applications in fire prevention, detection, and presuppression planning. Nevertheless, rapidly changing conditions in the fire environment, in particular fuels and weather, are not accounted for in a Map of Risk, unless they are computed daily. This led me to consider the development of a daily prediction model, which is a much more effective planning tool.

A binary logit model was successfully developed to predict daily human-caused fire occurrence in eight fire occurrence prediction units in the Whitecourt Forest. This model provided a prediction of fire "yes/no" for each unit and day within a standard fire season (April to October) using the corresponding unit area (km^2), location in forest district 2 (dummy), BUI, and ISI values to compute the probability of fire occurrence. This model was able to predict correctly 74% of the outcomes in a validation data set not used in the development of the model.

A back-propagation neural network model was successfully developed to predict the daily probability of fire occurrence in eight fire occurrence prediction units within the Whitecourt Forest. As with the logit model, predictions were limited to days within a standard fire season (April to October), and the prediction was fire "yes/no" for each area and day. The network was presented randomly the same data set of 314 examples of fire and no-fire observations until it "learned" the relationship between unit area (km^2), FWI, location in forest district 2 (dummy), and fire occurrence. The network was able to correctly predict 76% of the outcomes in a test data set not previously used for training.

Both techniques achieved similar accuracy in predicting daily human-caused forest fire occurrence in eight fire occurrence prediction units in the Whitecourt Forest. The models developed for the 1986-1990 data set correctly predicted 79-81% of the time. When these models were used on the 1991-1992 data set, fire occurrence prediction was 74-76% accurate. The fact that two very different techniques, the binary logit and the ANN, reached almost identical accuracy in predicting fire occurrence suggests limitations in the data, not in the models. Geographic data available for this study was certainly limited, and the variation of geographic characteristics of the study area during 1986-1990 could not be accounted for (for instance, building of new roads). Hence, there is much room for improvement in this respect but several techniques currently available do adequately model this complex process of human-caused fire occurrence. Logit models and neural network models can both be used, but economic and user-related considerations might advise the use of logit models over the deployment of ANN at this point in time.

CHAPTER VI

Future Research

The principal hypothesis in this study was that human risk could be estimated from the state of the forest environment at any given time. This assumption ignored differences in perception, reasoning, and decision making capabilities in the human population. As a result, there were situations in the original data in which all physical variables indicated high risk of fire, but no fire happened, and there were situations in which everything pointed to the impossibility of a fire occurring, and a fire did occur. Data about human presence, their distribution and activities in the forest areas are necessary in future studies. These data must be collected in the field, and a methodology defined to obtain a continuous flow of information to the fire protection agencies concerned. This continuous flow of information will enable managers to update the data sets used in developing models and if ANN systems are used learning will be facilitated.

Artificial neural networks have shown great potential in their application in this field. Networks specifically developed for prediction should be tried in future models. Both logit models and neural nets allow for much improvement in current predictions, through further model refinement, the addition of higher-order terms, and the addition of new variables, current models can be enhanced.

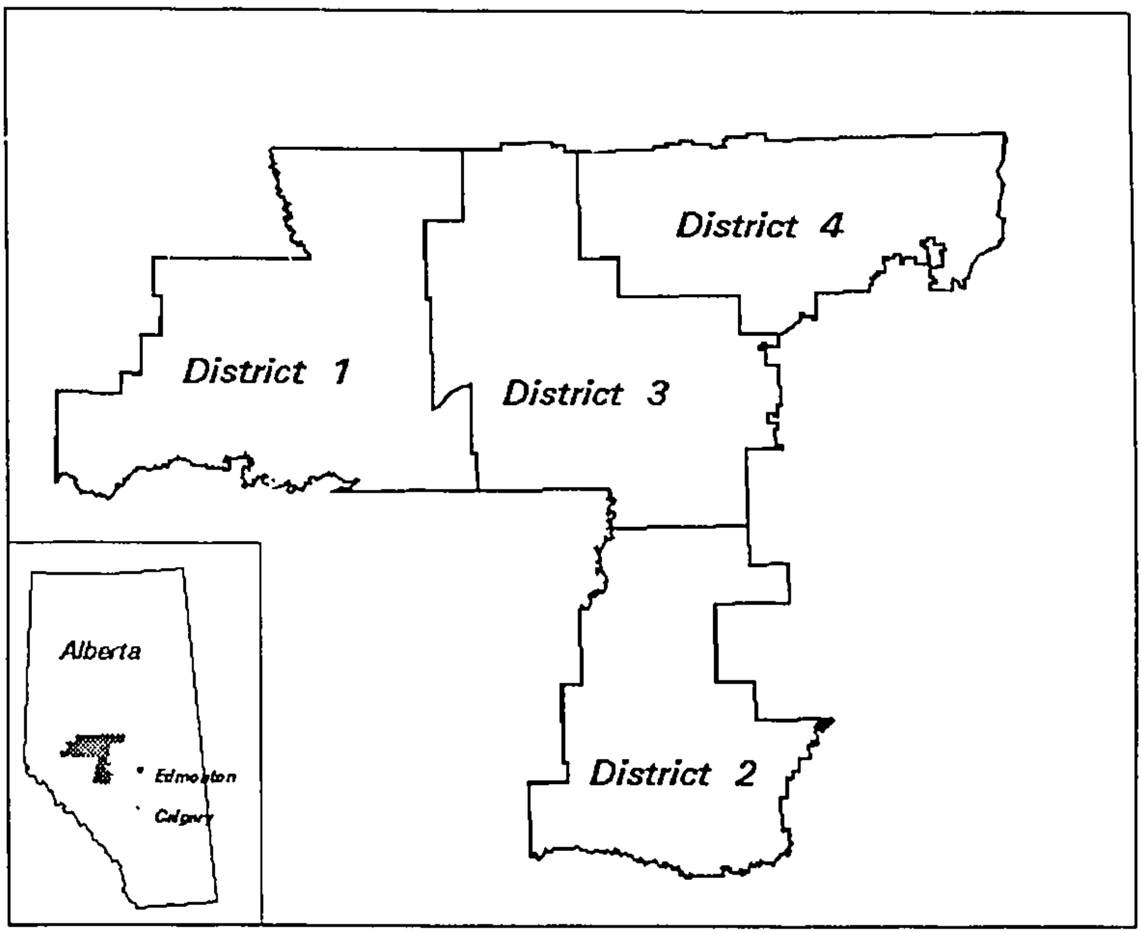
There were some limitations in the data available for this study. There was no information on how the geographic data available changed over the 5-year period used to model fire occurrence. The variable "distance to roads", which it has been found to be relevant to fire occurrence prediction, is expected to have changed in value continuously due to construction of new roads. In Whitecourt Forest, there are about 1,000 road construction proposals submitted every year. Other geographic factors may change less rapidly over time, such as property, or more rapidly, such as fuels. In future studies it will be important to incorporate accurate geographic data and keep track of the variability of these geographic data in time. A yearly update of these data is probably sufficient to improve current predictions.

The adequacy of the study area chosen must also be considered. This Forest was selected because sufficient geographic and weather information was available to attempt modeling fire occurrence in the Forest. A change in the physical location of the test site is more than likely to affect the results obtained in this study. Re-evaluation of the models using the same techniques is encouraged, but considering again all geographic and temporal factors available to the future researcher that could have an effect on human-caused wildfire occurrence. The absence of certain variables in the models developed might have been more related to the peculiarities of the data used in this study (peculiarities of the Whitecourt Forest), than to the reality of human-caused wildfires elsewhere. Future studies will provide insight on this issue.

APPENDIX A

The location of the Whitecourt Forest in Alberta

The four forest districts shown are management or administrative areas within the Forest.



APPENDIX B

Pearson correlation analysis for all

geographic and temporal variables

The SAS System : Correlation Analysis

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 314

	AREA	ROADDIS	TOWN	CAMP	ELEV	PPA	COMH	FUEL1	FUEL2
AREA	1.00000 0.0	-0.85085 0.0001	-0.68374 0.0001	-0.39916 0.0001	-0.21423 0.0001	0.54297 0.0001	0.71001 0.0001	0.88180 0.0001	0.63995 0.0001
ROADDIS	-0.85085 0.0001	1.00000 0.0	0.72409 0.0001	0.69218 0.0001	0.00098 0.9862	-0.56348 0.0001	-0.54370 0.0001	-0.74994 0.0001	-0.68174 0.0001
TOWN	-0.68374 0.0001	0.72409 0.0001	1.00000 0.0	0.71799 0.0001	0.57235 0.0001	-0.63319 0.0001	-0.15185 0.0070	-0.81968 0.0001	-0.45354 0.0001
CAMP	-0.39916 0.0001	0.69218 0.0001	0.71799 0.0001	1.00000 0.0	0.13921 0.0135	-0.49861 0.0001	0.03735 0.5097	-0.55152 0.0001	-0.32208 0.0001
ELEV	-0.21423 0.0001	0.00098 0.9862	0.57235 0.0001	0.13921 0.0135	1.00000 0.0	-0.32874 0.0001	0.18758 0.0008	-0.44515 0.0001	0.09545 0.0913
PPA	0.54297 0.0001	-0.56348 0.0001	-0.63319 0.0001	-0.49861 0.0001	-0.32874 0.0001	1.00000 0.0	-0.16908 0.0026	0.82072 0.0001	-0.15364 0.0064
COMH	0.71001 0.0001	-0.54370 0.0001	-0.15185 0.0070	0.03735 0.5097	0.18758 0.0008	-0.16908 0.0026	1.00000 0.0	0.30427 0.0001	0.83066 0.0001
FUEL1	0.88180 0.0001	-0.74994 0.0001	-0.81968 0.0001	-0.55152 0.0001	-0.44515 0.0001	0.82072 0.0001	0.30427 0.0001	1.00000 0.0	0.30015 0.0001
FUEL2	0.63995 0.0001	-0.68174 0.0001	-0.45354 0.0001	-0.32208 0.0001	0.09545 0.0913	-0.15364 0.0064	0.83066 0.0001	0.30015 0.0001	1.00000 0.0
FUEL3	0.55124 0.0001	-0.74213 0.0001	-0.66778 0.0001	-0.57338 0.0001	-0.20681 0.0002	0.88993 0.0001	-0.04194 0.4590	0.70602 0.0001	0.10013 0.0764
WS	0.07928 0.1611	-0.08179 0.1482	-0.11337 0.0447	-0.01232 0.8278	-0.00875 0.8773	-0.09957 0.0781	0.14934 0.0080	0.00601 0.9155	0.23559 0.0001
RH	-0.22852 0.0001	0.19905 0.0004	0.18939 0.0007	0.11464 0.0424	0.10803 0.0558	-0.24175 0.0001	-0.07198 0.2034	-0.25813 0.0001	-0.04571 0.4196
FFMC	0.15579 0.0057	-0.14013 0.0129	-0.13711 0.0150	-0.08826 0.1186	-0.06880 0.2241	0.18750 0.0008	0.03250 0.5661	0.18695 0.0009	0.01605 0.7769
DMC	0.07003 0.2159	-0.11203 0.0473	-0.09718 0.0856	-0.04116 0.4674	-0.08210 0.1467	0.15409 0.0062	-0.02527 0.6556	0.08917 0.1148	-0.00753 0.8943
BUI	0.02622 0.6435	-0.09060 0.1091	-0.08805 0.1195	-0.03368 0.5521	-0.08234 0.1455	0.13652 0.0155	-0.06303 0.2654	0.04857 0.3910	-0.01621 0.7748
ISI	0.12409 0.0279	-0.17505 0.0018	-0.17143 0.0023	-0.13583 0.0160	-0.04887 0.3881	0.15155 0.0071	0.01991 0.7253	0.13859 0.0140	0.08589 0.1288
FWI	0.07754 0.1705	-0.14351 0.0109	-0.15093 0.0074	-0.11264 0.0461	-0.06903 0.2226	0.16494 0.0034	-0.03964 0.4840	0.10846 0.0549	0.03161 0.5769

The SAS System : Correlation Analysis

Pearson Correlation Coefficients / Prob > |R| under Ho: Rho=0 / N = 314

	FUEL3	WS	RH	FFMC	DMC	BUI	ISI	FWI
AREA	0.55124 0.0001	0.07928 0.1611	-0.22852 0.0001	0.15579 0.0057	0.07003 0.2159	0.02622 0.6435	0.12409 0.0279	0.07754 0.1705
ROADDIS	-0.74213 0.0001	-0.08179 0.1482	0.19905 0.0004	-0.14013 0.0129	-0.11203 0.0473	-0.09060 0.1091	-0.17505 0.0018	-0.14351 0.0109
TOWN	-0.66778 0.0001	-0.11337 0.0447	0.18939 0.0007	-0.13711 0.0150	-0.09718 0.0856	-0.08805 0.1195	-0.17143 0.0023	-0.15093 0.0074
CAMP	-0.57338 0.0001	-0.01232 0.8278	0.11464 0.0424	-0.08826 0.1186	-0.04116 0.4674	-0.03368 0.5521	-0.13583 0.0160	-0.11264 0.0461
ELEV	-0.20681 0.0002	-0.00875 0.8773	0.10803 0.0558	-0.06880 0.2241	-0.08210 0.1467	-0.08234 0.1455	-0.04887 0.3881	-0.06903 0.2226
PPA	0.88993 0.0001	-0.09957 0.0781	-0.24175 0.0001	0.18750 0.0008	0.15409 0.0062	0.13652 0.0155	0.15155 0.0071	0.16494 0.0034
COMH	-0.04194 0.4590	0.14934 0.0080	-0.07198 0.2034	0.03250 0.5661	-0.02527 0.6556	-0.06303 0.2654	0.01991 0.7253	-0.03964 0.4840
FUEL1	0.70602 0.0001	0.00601 0.9155	-0.25813 0.0001	0.18695 0.0009	0.08917 0.1148	0.04857 0.3910	0.13859 0.0140	0.10846 0.0549
FUEL2	0.10013 0.0764	0.23559 0.0001	-0.04571 0.4196	0.01605 0.7769	-0.00753 0.8943	-0.01621 0.7748	0.08589 0.1288	0.03161 0.5768
FUEL3	1.00000 0.0	-0.03497 0.5370	-0.22145 0.0001	0.16752 0.0029	0.21049 0.0002	0.21423 0.0001	0.19899 0.0004	0.22403 0.0001
WS	-0.03497 0.5370	1.00000 0.0	-0.16147 0.0041	0.09104 0.1074	0.02521 0.6563	0.01705 0.7635	0.50995 0.0001	0.39056 0.0001
RH	-0.22145 0.0001	-0.16147 0.0041	1.00000 0.0	-0.80298 0.0001	-0.39268 0.0001	-0.37587 0.0001	-0.66541 0.0001	-0.64889 0.0001
FFMC	0.16752 0.0029	0.09104 0.1074	-0.80298 0.0001	1.00000 0.0	0.47954 0.0001	0.47885 0.0001	0.63971 0.0001	0.65562 0.0001
DMC	0.21049 0.0002	0.02521 0.6563	-0.39268 0.0001	0.47954 0.0001	1.00000 0.0	0.96114 0.0001	0.42726 0.0001	0.69047 0.0001
BUI	0.21423 0.0001	0.01705 0.7635	-0.37587 0.0001	0.47885 0.0001	0.96114 0.0001	1.00000 0.0	0.40950 0.0001	0.68246 0.0001
ISI	0.19899 0.0004	0.50995 0.0001	-0.66541 0.0001	0.63971 0.0001	0.42726 0.0001	0.40950 0.0001	1.00000 0.0	0.92352 0.0001
FWI	0.22403 0.0001	0.39056 0.0001	-0.64889 0.0001	0.65562 0.0001	0.69047 0.0001	0.68246 0.0001	0.92352 0.0001	1.00000 0.0

APPENDIX C

SAS outputs for the six best logit models

Logit Model 1

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	296.670	.
SC	441.046	315.417	.
-2 LOG L	435.296	286.670	148.626 with 4 DF (p=0.0001)
Score	.	.	119.597 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-4.6048	0.5917	60.5551	0.0001	.	0.010
AREA	1	7.6590	1.2198	39.4268	0.0001	0.579763	999.000
DISTRICT	1	0.7367	0.3429	4.6159	0.0317	0.185163	2.089
BUI	1	2.0478	0.9936	4.2482	0.0393	0.192431	7.751
ISI	1	39.5628	6.3232	39.1474	0.0001	0.759288	999.000

Association of Predicted Probabilities and Observed Responses

Concordant	= 86.9%	Somers' D	= 0.740
Discordant	= 12.9%	Gamma	= 0.741
Tied	= 0.1%	Tau-a	= 0.371
(24649 pairs)		c	= 0.870

Logit Model 1

Hosmer and Lemeshow Goodness-of-Fit Test

Group	Total	OCCUR = 1 Observed	OCCUR = 1 Expected	OCCUR = 0 Observed	OCCUR = 0 Expected
1	31	2	1.37	29	29.63
2	31	6	2.99	25	28.01
3	31	5	6.20	26	24.80
4	31	7	9.76	24	21.24
5	31	8	12.96	23	18.04
6	31	20	16.74	11	14.26
7	31	21	20.98	10	10.02
8	31	27	24.93	4	6.07
9	31	28	27.50	3	3.50
10	35	33	33.58	2	1.42

Goodness-of-fit Statistic = 10.94 with 8 DF (p=0.2051)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	120	128	29	37	79.0	76.4	81.5	19.5	22.4

Logit Model 2

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	304.197	.
SC	441.046	319.194	.
-2 LOG L	435.296	296.197	139.100 with 3 DF (p=0.0001)
Score	.	.	115.204 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-4.0956	0.5273	60.3169	0.0001	.	0.017
AREA	1	7.8722	1.2103	42.3099	0.0001	0.595905	999.000
FWI	1	20.3576	2.6952	57.0521	0.0001	0.759206	999.000
DISTRICT	1	0.7479	0.3380	4.8969	0.0269	0.187961	2.112

Association of Predicted Probabilities and Observed Responses

Concordant = 86.0%	Somers' D = 0.724
Discordant = 13.6%	Gamma = 0.727
Tied = 0.5%	Tau-a = 0.363
(24649 pairs)	c = 0.862

Logit Model 2

Hosmer and Lemeshow Goodness-of-Fit Test

		OCCUR = 1		OCCUR = 0	
Group	Total	Observed	Expected	Observed	Expected
1	31	2	1.58	29	29.42
2	31	7	3.15	24	27.85
3	31	3	6.48	28	24.52
4	31	9	10.57	22	20.43
5	31	9	13.06	22	17.94
6	31	19	16.73	12	14.27
7	31	22	20.69	9	10.31
8	31	24	24.38	7	6.62
9	31	28	27.10	3	3.90
10	35	34	33.26	1	1.74

Goodness-of-fit Statistic = 11.752 with 8 DF (p=0.1626)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	120	126	31	37	78.3	76.4	80.3	20.5	22.7

Logit Model 3

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	290.880	.
SC	441.046	309.627	.
-2 LOG L	435.296	280.880	154.417 with 4 DF (p=0.0001)
Score	.	.	122.817 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-7.3712	1.2390	35.3971	0.0001	.	0.001
AREA	1	5.1693	1.3259	15.1987	0.0001	0.391299	175.784
FFMC	1	5.2138	1.6164	10.4046	0.0013	0.573788	183.797
FWI	1	10.7595	3.5790	9.0380	0.0026	0.401260	999.000
F	1	0.5636	0.1965	8.2246	0.0041	0.289622	1.757

Association of Predicted Probabilities and Observed Responses

Concordant	= 87.8%	Somers' D	= 0.758
Discordant	= 12.1%	Gamma	= 0.758
Tied	= 0.1%	Tau-a	= 0.380
(24649 pairs)		c	= 0.879

Logit Model 3

Hosmer and Lemeshow Goodness-of-Fit Test

		OCCUR = 1		OCCUR = 0	
Group	Total	Observed	Expected	Observed	Expected
1	31	2	0.61	29	30.39
2	31	5	2.57	26	28.43
3	31	5	5.03	26	25.97
4	31	6	9.47	25	21.53
5	31	10	14.27	21	16.73
6	31	20	17.97	11	13.03
7	31	20	22.51	11	8.49
8	31	27	25.23	4	5.77
9	31	28	26.83	3	4.17
10	35	34	32.51	1	2.49

Goodness-of-fit Statistic = 13.511 with 8 DF (p=0.0954)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	129	121	36	28	79.6	82.2	77.1	21.8	18.8

Logit Model 4

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for As.essing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	304.359	.
SC	441.046	323.106	.
-2 LOG L	435.296	294.359	140.938 with 4 DF (p=0.0001)
Score	.	.	107.901 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-9.1193	1.2395	54.1261	0.0001	.	0.000
AREA	1	7.3877	1.1437	41.7252	0.0001	0.559228	999.000
DISTRICT	1	0.6658	0.3348	3.9548	0.0467	0.167328	1.946
FFMC	1	7.7069	1.4558	28.0244	0.0001	0.848149	999.000
BUI	1	2.1224	1.0002	4.5025	0.0338	0.199437	8.351

Association of Predicted Probabilities and Observed Responses

Concordant	= 86.1%	Somers' D	= 0.724
Discordant	= 13.7%	Gamma	= 0.725
Tied	= 0.1%	Tau-a	= 0.363
(24649 pairs)		c	= 0.862

Logit Model 4

Hosmer and Lemeshow Goodness-of-Fit Test

Group	Total	OCCUR = 1 Observed	OCCUR = 1 Expected	OCCUR = 0 Observed	OCCUR = 0 Expected
1	31	1	0.33	30	30.67
2	31	4	2.55	27	28.45
3	31	6	6.28	25	24.72
4	31	6	10.12	25	20.88
5	31	12	14.80	19	16.20
6	31	20	18.76	11	12.24
7	31	25	22.52	6	8.48
8	31	26	24.60	5	6.40
9	31	26	25.99	5	5.01
10	35	31	31.04	4	3.96

Goodness-of-fit Statistic = 7.3617 with 8 DF (p=0.4982)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	131	119	38	26	79.6	83.4	75.8	22.5	17.9

Logit Model 5

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	299.080	.
SC	441.046	314.077	.
-2 LOG L	435.296	291.080	144.217 with 3 DF (p=0.0001)
Score	.	.	114.419 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-4.1342	0.5244	62.1490	0.0001	.	0.016
AREA	1	7.4631	1.1944	39.0458	0.0001	0.564936	999.000
DISTRICT	1	0.8007	0.3369	5.6483	0.0175	0.201230	2.227
ISI	1	44.9417	6.0236	55.6652	0.0001	0.862520	999.000

Association of Predicted Probabilities and Observed Responses

Concordant	= 86.2%	Somers' D	= 0.727
Discordant	= 13.5%	Gamma	= 0.729
Tied	= 0.3%	Tau-a	= 0.365
(24649 pairs)		c	= 0.863

Logit Model 5

Hosmer and Lemeshow Goodness-of-Fit Test

		OCCUR = 1		OCCUR = 0	
Group	Total	Observed	Expected	Observed	Expected
1	31	2	1.69	29	29.31
2	31	5	3.31	26	27.69
3	31	7	6.41	24	24.59
4	31	6	10.12	25	20.88
5	31	13	13.14	18	17.86
6	31	14	15.86	17	15.14
7	31	22	20.30	9	10.70
8	31	26	24.79	5	6.21
9	31	29	27.66	2	3.34
10	35	33	33.73	2	1.27

Goodness-of-fit Statistic = 5.7816 with 8 DF (p=0.6717)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	115	127	30	42	77.1	73.2	80.9	20.7	24.9

Logit Model 6

The LOGISTIC Procedure

Data Set: WORK.NWTDATA
Response Variable: OCCUR
Response Levels: 2
Number of Observations: 314
Link Function: Logit

Response Profile

Ordered Value	OCCUR	Count
1	1	157
2	0	157

Criteria for Assessing Model Fit

Criterion	Intercept Only	Intercept and Covariates	Chi-Square for Covariates
AIC	437.296	286.182	.
SC	441.046	301.180	.
-2 LOG L	435.296	278.182	157.114 with 3 DF (p=0.0001)
Score	.	.	122.378 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Standardized Estimate	Odds Ratio
INTERCPT	1	-3.9084	1.1142	12.3056	0.0005	.	0.020
FFMC	1	4.7607	1.6123	8.7184	0.0032	0.523916	116.824
ISI	1	26.2120	7.8885	11.0411	0.0009	0.503061	999.000
ARRODIS	1	-3.0691	0.4861	39.8587	0.0001	-0.599055	0.046

Association of Predicted Probabilities and Observed Responses

Concordant	= 87.9%	Somers' D	= 0.760
Discordant	= 12.0%	Gamma	= 0.761
Tied	= 0.1%	Tau-a	= 0.381
(24649 pairs)		c	= 0.880

Logit Model 6

Hosmer and Lemeshow Goodness-of-Fit Test

Group	Total	OCCUR = 1 Observed	OCCUR = 1 Expected	OCCUR = 0 Observed	OCCUR = 0 Expected
1	31	2	0.62	29	30.38
2	31	5	2.54	26	28.46
3	31	3	5.17	28	25.83
4	31	7	9.28	24	21.72
5	31	11	14.27	20	16.73
6	31	18	17.84	13	13.16
7	31	21	22.05	10	8.95
8	31	28	25.13	3	5.87
9	31	28	27.20	3	3.80
10	35	34	32.90	1	2.10

Goodness-of-fit Statistic = 11.726 with 8 DF (p=0.1638)

Classification Table

	Correct		Incorrect		Percentages				
Prob Level	Event	Non-Event	Event	Non-Event	Correct	Sensi-tivity	Speci-ficity	False POS	False NEG
0.500	131	121	36	26	80.3	83.4	77.1	21.6	17.7

APPENDIX D

Classification tables for independent datc

Logit Models 1-6

Classification Table for Model 1
Variables: Intercept,Area,District,BUI,ISI

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2398 72.80 74.10	838 25.44 25.90	3236 98.24
Actual Fire	15 0.46 25.86	43 1.31 74.14	58 1.76
Total	2413 73.25	881 26.75	3294 100.00

total percentage correctly predicted=74.1%, Σ pred.prob.= 94.387

Classification Table for Model 2
Variables: Intercept,Area,District,FWI

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2410 73.16 74.47	826 25.08 25.53	3236 98.24
Actual Fire	15 0.46 25.86	43 1.31 74.14	58 1.76
Total	2425 73.62	869 26.38	3294 100.00

total percentage correctly predicted=74.5, Σ pred.prob.= 93.848

Classification Table for Model 3
Variables: Intercept,Area,FFMC,FWI,F

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2303 69.91 71.17	933 28.32 28.83	3236 98.24
Actual Fire	15 0.46 25.86	43 1.31 74.14	58 1.76
Total	2318 70.37	976 29.63	3294 100.00

total percentage correctly predicted=71.2%, Σ pred.prob.= 84.806

Classification Table for Model 4
Variables: Intercept,Area,District,FFMC,BUI

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2345 71.19 72.47	891 27.05 27.53	3236 98.24
Actual Fire	17 0.52 29.31	41 1.24 70.69	58 1.76
Total	2362 71.71	932 28.29	3294 100.00

total percentage correctly predicted=72.4, Σ pred.prob.= 71.679

Classification Table for Model 5
Variables: Intercept,Area,District,ISI

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2435 73.92 75.25	801 24.32 24.75	3236 98.24
Actual Fire	17 0.52 29.31	41 1.24 70.69	58 1.76
Total	2452 74.44	842 25.56	3294 100.00

total percentage correctly predicted=75.2%, Σ pred.prob.= 94.011

Classification Table for Model 6
Variables: Intercept,Arrodis,ISI,FFMC

Frequency Percent Row Pct	Predicted No Fire	Predicted Fire	Total
Actual No Fire	2339 71.01 72.20	897 27.23 27.72	3236 98.24
Actual Fire	15 0.46 25.86	43 1.31 74.14	58 1.76
Total	2354 71.00	940 29.00	3294 100.00

total percentage correctly predicted=72.3%, Σ pred.prob.= 85.821, p=0.022

APPENDIX E

Classification tables for independent data

Best Neural Network Models

ANN Models for comparison with logistic regression models

Classification Table for Model 1 with 10 hidden PEs
Variables: Area,District,BUI,ISI

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2356 71.52 72.80	880 26.71 27.19	3236 98.24
Actual *Fire*	17 0.52 29.31	41 1.24 70.69	58 1.76
Total	2373 72.00	921 28.00	3294 100.00

After 19,000 iterations, RMS=.248, total pct correctly predicted=72.8%

Classification Table for Model 2 with 4 hidden PEs
Variables: Area,District,FWI

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2462 74.74 76.08	774 23.50 23.92	3236 98.24
Actual *Fire*	14 0.43 24.14	44 1.34 75.86	58 1.76
Total	2476 75.17	818 24.83	3294 100.00

After 19,000 iterations, RMS=.246, total pct correctly predicted=76.1%

ANN General Models

Classification Table for Model with 14 hidden PEs
Variables: Area,FWI,RH,FFMC,ISI,ROAD

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2422 73.52 74.84	814 24.71 25.15	3236 98.24
Actual *Fire*	17 0.51 29.31	41 1.24 70.68	58 1.76
Total	2439 74.04	855 25.95	3294 100.00

After 19,000 iterations, RMS=.238, total pct correctly predicted=74.8%

Classification Table for Model with 4 hidden PEs
Variables: Area,FWI,RH,ISI

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2428 73.70 75.03	808 24.53 24.96	3236 98.24
Actual *Fire*	16 0.48 27.58	42 1.27 72.41	58 1.76
Total	2444 74.19	850 25.80	3294 100.00

After 29,000 iterations, RMS=.247, total pct correctly predicted=75.0%

Classification Table for Model with 4 hidden PEs
Variables: RH,ISI,ROAD

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2356 71.52 72.80	880 26.71 27.19	3236 98.24
Actual *Fire*	14 0.42 24.13	44 1.33 75.86	58 1.76
Total	2370 71.94	924 28.05	3294 100.00

After 19,000 iterations, RMS=.243, total pct correctly predicted=72.9%

Classification Table for Model with 4 hidden PEs
Variables: Area,FWI,ROAD

Frequency *Percent* *Row Pct*	*Predicted* *No Fire*	*Predicted* *Fire*	*Total*
Actual *No Fire*	2393 72.64 73.94	843 25.59 26.05	3236 98.24
Actual *Fire*	16 0.48 27.58	42 1.27 72.41	58 1.76
Total	2409 73.13	885 26.86	3294 100.00

After 19,000 iterations, RMS=.246, total pct correctly predicted=73.9%

APPENDIX F

A network-specific generated C code

```
/* Fri Sep  3 15:58:27 1993 (m2.c) Recall-Only Run-time for <bestever> */
/*    Control Strategy is: <bkpfast> */

#if __STDC__
#define  ARGS(x) x
#else
#define  ARGS(x) ()
#endif /* __STDC__ */

/* --- External Routines --- */
extern double exp   ARGS((double));
/* *** MAKE SURE TO LINK IN YOUR COMPILER's MATH LIBRARIES *** */

#if __STDC__
int m2_start( void *NetPtr, float Yin[4], float Yout[2] )
#else
int m2_start( NetPtr, Yin, Yout )
void  *NetPtr; /* Network Pointer (not used) */
float Yin[4], Yout[2];  /* Data */
#endif /* __STDC__ */
{
   float  Xout[12];  /* work arrays */
   long   ICmpT;  /* temp for comparisons */

   /* *** WARNING: Code generated assuming Recall = 0 *** */

   /* Read and scale input into network */
   Xout[2] = Yin[0];
   Xout[3] = Yin[1];
   Xout[4] = Yin[2] * (0.00025943704) + (-.20905436);
   Xout[5] = Yin[3] * (0.027777778);
LAB110:

   /* Generating code for PE 0 in layer 3 */
   Xout[6]  = (float)(-.82123846) + (float)(-.23006813) * Xout[2] +
      (float)(-.74295259) * Xout[3] + (float)(1.5722935) * Xout[4] +
      (float)(2.3422184) * Xout[5];
   Xout[6] = 1.0 / (1.0 + exp( -Xout[6] ));
```

```
   /* Generating code for PE 1 in layer 3 */
   Xout[7]  = (float)(-1.4312067) + (float)(-.37953216) * Xout[2] +
      (float)(-.97151035) * Xout[3] + (float)(2.1379476) * Xout[4] +
      (float)(3.5678072) * Xout[5];
   Xout[7] = 1.0 / (1.0 + exp( -Xout[7] ));

   /* Generating code for PE 2 in layer 3 */
   Xout[8]  = (float)(1.0924038) + (float)(.28197211) * Xout[2] +
      (float)(.75670433) * Xout[3] + (float)(-1.7059669) * Xout[4] +
      (float)(-2.9353254) * Xout[5];
   Xout[8] = 1.0 / (1.0 + exp( -Xout[8] ));

   /* Generating code for PE 3 in layer 3 */
   Xout[9]  = (float)(.24200568) + (float)(-0.063673533) * Xout[2] +
      (float)(.15872669) * Xout[3] + (float)(-.60328162) * Xout[4] +
      (float)(-1.174696) * Xout[5];
   Xout[9] = 1.0 / (1.0 + exp( -Xout[9] ));

   /* Generating code for PE 0 in layer 4 */
   Xout[10]  = (float)(-.2924175) + (float)(.82507396) * Xout[6] +
      (float)(1.3176024) * Xout[7] + (float)(-1.1102298) * Xout[8] +
      (float)(-.41101122) * Xout[9];
   Xout[10] = 1.0 / (1.0 + exp( -Xout[10] ));

   /* Generating code for PE 1 in layer 4 */
   Xout[11]  = (float)(.25909081) + (float)(-.84139597) * Xout[6] +
      (float)(-1.2789248) * Xout[7] + (float)(1.1355991) * Xout[8] +
      (float)(.43569073) * Xout[9];
   Xout[11] = 1.0 / (1.0 + exp( -Xout[11] ));

   /* De-scale and write output from network */
   Yout[0] = Xout[10] * (1.6666666) + (-.33333333);
   Yout[1] = Xout[11] * (1.6666666) + (-.33333333);
   return( 0 );
}
```