8-9-2008

Artificial neural networks to detect forest fire prone areas in the southeast fire district of Mississippi

Mohan P. Tiruveedhula

ARTIFICIAL NEURAL NETWORKS TO DETECT FOREST FIRE PRONE AREAS IN THE SOUTHEAST FIRE DISTRICT OF MISSISSIPPI

By

Mohan P Tiruveedhula

A Thesis
Submitted to the Faculty of
Mississippi State University
in Partial Fulfillment of the Requirements
for the Degree of Master of Science
in Geosciences
in the Department of Geosciences

Mississippi State, Mississippi

August 2008

ARTIFICIAL NEURAL NETWORKS TO DETECT FOREST FIRE PRONE AREAS OF MISSISSIPPI

By

Mohan P Tiruveedhula

Approved:

William H. Cooke III Assistant Professor of Geosciences (Director of Thesis)	John C. Rodgers Associate Professor of Geosciences (Committee Member)
Jinmu Choi Assistant Professor of Geosciences (Committee Member)	Christopher P. Dewey Associate Professor and Graduate Coordinator of Geosciences
Darrel Schmitz Professor and Head of Geosciences	Gary Myers Dean of the College of Arts and Sciences

Name: Mohan P Tiruveedhula

Date of Degree: August 9, 2008

Institution: Mississippi State University

Major Field: Geosciences

Major Professor: Dr. William H. Cooke III

Title of Study: ARTIFICIAL NEURAL NETWORKS TO DETECT FOREST FIRE PRONE AREAS IN THE SOUTHEAST FIRE DISTRICT OF MISSISSIPPI.

Pages in Study: 88

Candidate for Degree of Master of Science

An analysis of the fire occurrences parameters is essential to save human lives, property, timber resources and conservation of biodiversity. Data conversion formats such as raster to ASCII facilitate the integration of various GIS software's in the context of RS and GIS modeling. This research explores fire occurrences in relation to human interaction, fuel density interaction, euclidean distance from the perennial streams and slope using artificial neural networks. The human interaction (ignition source) and density of fuels is assessed by Newton's Gravitational theory. Euclidean distance to perennial streams and slope that do posses a significant role were derived using GIS tools.

All the four non linear predictor variables were modeled using the inductive nature of neural networks. The Self organizing feature map (SOM) utilized for fire size risk classification produced an overall classification accuracy of 62% and an overall kappa coefficient of 0.52 that is moderate (fair) for annual fires.

DEDICATION

This thesis, a golden opportunity in my life is affectionately dedicated to my parents, Tiruveedhula Ramadevi and Tiruveedhula Venkateswarlu whose high expectations, constant inspiration and everlasting love form the base of my progress.

ACKNOWLEDGMENTS

I avail this opportunity to express my whole hearted sense of gratitude to Dr. Bill Cooke, my thesis major advisor for his valuable inspiration, guidance, constructive and keen criticism, prompt suggestions, constant encouragement and sympathetic attitude right from the inception of this research till the final shaping of this manuscript in the present form. I am fortunate to work under the guidance of such a dynamic personality.

I express my sincere thanks to my committee members Dr. John Rodgers and Dr. Jinmu Choi for their encouragement, cooperation and guidance during this research. I am grateful to Dr. Christopher P. Dewey, the Graduate Coordinator of Geosciences and Dr. Darrel Schmitz, Head of Geosciences department, for their continuous encouragement during my post graduation program.

My special heartfelt thanks to my brother, Raviraj Sadasivuni and sister-in-law Lalitha Sadasivuni for their friendly valuable suggestions during this research work and helping me in understanding the finer aspects of the subject, thereby imbibing in me an element of scientific thinking.

Special thanks to Dr. Fred Musser, who had given me a chance to pursue my further education in United States Of America.

It is extremely difficult to get appropriate words to pen down my feelings of gratitude towards my parents, brother, sister-in-law, sister and brother-in-law. Nothing could have been accomplished with out their blessings.

I should also thank my friends Ramesh, Rajesh, Murali and Arun for my initial stabilization in Mississippi State University. I must keep on record the cooperation and moral support extended by the Chadwick as well as the Aiken Indian friends and families.

Finally, I would like to thank Kate Grala, Rekha Pillai, John Morris and graduate students Ryan, Mitchell, Mark and Eduardo for their help and suggestions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Forests are an important asset to mankind in terms of their valuable role in preserving the environment to sustain life. Fire is useful in every day life. Natural fires and accidental fires can turn theses fires as a negative asset to the society. In United States, an average of 250,000 wildfires occurs per year that affects almost 5 million acres of forest, brush, and grass-covered lands (Wilson and Davis 1988). Annual fire protection services and loses costs more than half billion and two billion respectively in United States (Wilson and Davis 1988). Fire potential in general is lower in eastern US, still Mississippi on average has 3760 wildfires per year (Cooke et al, 2007). Natural disasters like the Hurricane Katrina have raised the fire threat of the state of MS that necessitated a Geographic information system (GIS) based fire management approach by Mississippi Forestry Commission (Gilreath, 2006). The Concept of GIS and remote sensing (RS) is becoming increasingly important to combat fires by forest management agencies. GIS can be used to monitor the fire source variables in fire risk mapping that is simple and efficient. Artificial neural networks (ANN) are recently emerged computational tools in the context of GIS and RS to model complex real world problems that are empirical in nature that provide exact solutions to precise or imprecise problems (Basheer and Hajmeer,2000). Kohonen Self organizing neural networks find application in classification problems, pattern recognition and data reduction. In the context of fire risk modeling, variables that are complex in nature can be best analyzed by self organizing feature map of artificial neural networks. Ultimately, the managers can visualize the source variables that create a particular fire pattern to combat fires.

The primary objective of this research is to implement ANN's self organizing feature map (SOM) for the associated fire input variables (human interaction, fuel density interaction, slope and Euclidean distance from perennial streams) and test the accuracy of predictions in the potential model for the Southeast Fire District of Mississippi.

There are two significant aspects associated with the present research.

1. The implementation of ANN to the associated four non linear predictor variables that are not easily described by deterministic processes.
2. The development of fuel density layer in a similar way to the human interaction layer as one of the variables in the present research.

This research is organized in to five chapters to provide readers with information pertaining to various concepts of this work, the details of which are outlined below:

Chapter I introduces the importance of forests, forest fires and the need of the forest fire research.

Chapter II discusses the background information and review of literature related to the forest fires, the variables (ignition sources from human impacts, Fuel density interaction, slope and distance from perennial streams) considered in the present research, ANN, SOM approach of the ANN, various GIS related fire models as well as ANN related fire models.

Chapter III explains the research objectives/goals of the present research.

Chapter IV deals with the materials and methods. As a part of this chapter study area, the various vector layers as well as the raster layers, data preparation used for the present research were discussed in detail.

Chapter V deals with results and discussion of the present research. The results of the four variables (city interaction, fuel density interaction, slope and Euclidean distance from perennial streams), SOM approach of ANN and accuracy obtained for the present classification are discussed in this section.

Chapter VI deals with summary and conclusions. The usefulness, limitations associated with the present research, and further research suggestions were listed in this chapter.

CHAPTER II

BACKGROUND INFORMATION AND LITERATURE REVIEW

This chapter reviews the past studies in relation to importance of forests, beneficial and harmful effects of forest fires, historic fire trends in MS, GIS fire models, four variables of the present research, ANN and SOM concepts.

2.1 Forests, Fire and wildfire

Globally, forests and other woody areas occupy 40 percent of the land surface and they serve as an important source of substantial goods and services to mankind (Wright, 2004). Forests play an important role in the carbon cycle, radiation budget and maintaining climatic balance and changes that occur in forest biomass effects these processes (Roy and Ravan 1996). Forests help to reduce carbon emission as they store large quantities of carbon and exchange it with the atmosphere by photosynthesis and respiration (Brown etal. 1999).

Fire is a disturbance process well before human interaction that altered North American land scapes approximately 12-20 million years ago (Ankica Grant, 2007). Fire as a natural process manages vegetation for fuel load reduction, regeneration and biodiversity conservation (Ankica Grant, 2007). Fire is important in everyday lives (cooking and warmth) to regenerate land and clear land (Ankica Grant, 2007). With the advent of the Europeans, people became aware of possible damage that could occur to agricultural lands, towns and cities and used fire suppression measures.

Wildland fire is one of the most ubiquitous of all terrestrial disturbance agents (Perry, 1998) and has been an important constituent in the natural environment (kemp, 1981; Cope and Chaloner, 1985) that effects ecological process such as vegetation succession and ecosystem structure and function (Kotsias and Karteris, 2003).

Forest fire is a fire burn that occurs in forested areas, brush, grass, tundra, or other vegetation (Ankica Grant, 2007). Natural fires, accidental and/or arson fires and man-controlled fires (prescribed fires) are the categories of forest fires. Of the various land covers, forests constitute the major portion of the burned area and the cost estimate to suppress these fires exceeds one billion dollars (Terry 1997). With this general information on fires, forest fires, and importance of fires a look at the fire trend in MS would provide the readers an insight in relevance to the present work.

2.2 Fires in Mississippi

Rudis and Skinner, 1991 studies the importance and distribution of fires in the South Central U.S. and stated that seventy five percent of forests burned during the last 10 years (1981-1990) were associated with wood production, livestock or wildlife production, or vegetation management and three percent is associated with natural disturbance. Fire evidence (systematic surveys along with recent inventories from private and public forest inventories) is reported to occur on 26 percent (22.4 out of 87.2 million acres) of timber land surveyed. Of this in the year 1987, 4.8 million acres fire evidence is reported in MS.

Cooke et al, 2007 stated that though a low fire potential exists in the eastern US, a 14 year (1991-2004, Figure 2.1) historic fire data analysis obtained from the Mississippi

Forestry Commission showed that MS on average experiences 3670 wildfires per year. A minimum of 1847 fires and maximum of 6616 fires were reported during the 14 year period.

Recently, with in the past six years (1999-2004) over 20,000 fire burns were reported in the state of MS. Many of these fires occur in the southeastern region of MS as such the south east part consisting of 22 counties (Figure 4.1) is regarded as south eastern fire district and leader of forest fires in MS (MFC, 2004 and Gilreath, 2006). The following figure shows the fire incidents per month in a year for the study area.

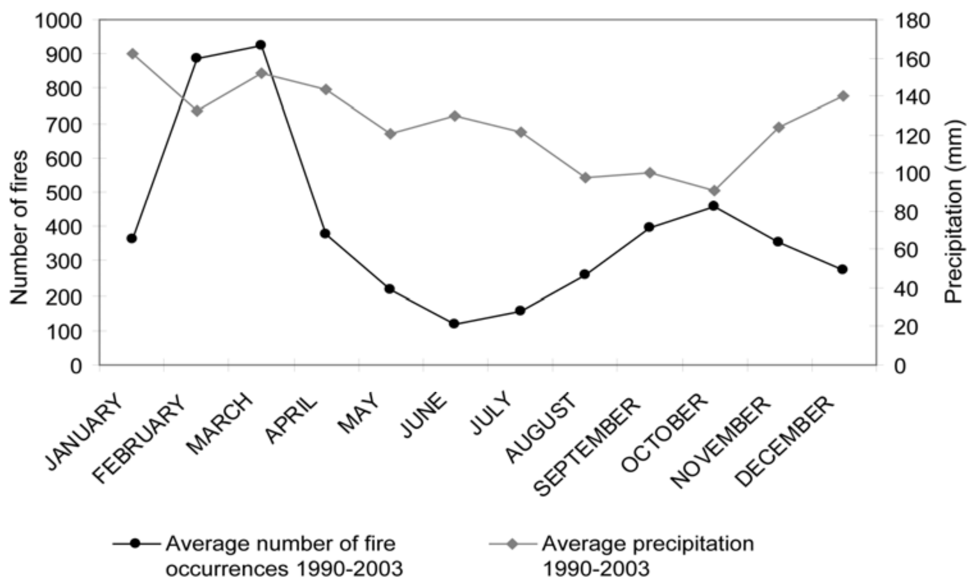

Figure 2.1 Fire incidents per month from 1990 to 2003

As such it is evident that the highest fire potential occurs in MS at two time periods (Figure 2.1). The two time periods corresponds to winter and summer seasons respectively. The fire potential in winter ranges from January to March while the summer fire potential is from July to November. In general, high fire occurrences exists in February and March in winter and in September and October in summer (Gilreath, 2006). Summer season in general experiences dry spells with fires more of anthropogenic origin. In contrast to summer, late summer and early fall witness evaporation exceeding precipitation (Gilreath, 2006). The summer dry spells together with anthropogenic factors exacerbates the fire potential to the state of MS (Gilreath, 2006). In winter one million hectares prescribed burns are done annually in the state of MS that sometimes become uncontrolled in the south east coastal plain of MS (Gilreath, 2006). In spite of high precipitation, the highest fire potential in the winter peaks (January to March) illustrates the importance of human component in fire modeling and the details of various fire models are reviewed below.

2.3 Fire models

Forest ecosystems are subjected to a wide variety of natural and anthropogenic disturbances such as fire, drought, insect and disease attacks, wind throw and breakage, air pollution, rain and surface water acidification, snow/ice damage, threatened and endangered species viability, and other small scale disturbances (Schmoldt D.L., 2001*)*. Fire is one of the serious threats to forests along with insect and disease attack (Schmoldt D.L., 2001*)*.

Risk is an attribute of loss event or disturbance as such fire risk is a fire attribute that consists of two components of potency and chance. Potency is the cost i.e. the severity and size of the fire while chance is the likelihood of the fire (Schmoldt D.L., 2001). Analysis of the fire risk prone areas may be stated as to assess the areas that have a chance of fire occurrence with the severity and size of the fire for that area or simply the areas with a chance of fire origin that can spread to other areas (Esra Erten et al., 2004). The chance component of fire risk is usually referred as the fire potential which is the likelihood or probability that a particular landscape is susceptible to fire in presence of an ignition source (Cooke et al, 2007).

Wildfire is an ecological (Ankica Grant, 2007) that is a part of ecological modeling. Ecological modeling provides methods to understand ecological systems, to assess human impacts and aid in environmental decision making (Reginald Mead, 2006). Ecological modeling can be process based, empirical based or a combination of both (Reginald Mead, 2006). A process based (expert systems) ecological process has low predictive power with high explanatory depth that tries to combine prior knowledge of the process to model it. Empirical based (Neural networks) ecological approaches are data based with high predictive power and low explanatory depth that are exactly opposite to process based models. A combined approach is a hybrid of the two approaches such as the Bayesian belief networks (Reginald Mead, 2006).

Klabokidis et al, 2004 used a fire danger rating index that consists of fire weather index, fire hazard index, fire risk index and fire risk index. The fire weather index included variables such as air temperature, wind velocity, relative humidity and precipitation. The fire hazard index included the fuel models including fuel moisture content; elevation and aspect while the fire risk index included distances related to roads, livestock, power lines, urban areas and as such related to human activities. As such the variables considered in this research are all can be grouped under fire hazard index.

Similarly Brenner, 2002 derived a final fire risk assessment system for the state of Florida. He used a weighted model of three indexes namely wildland fire susceptibility index (WFSI), fire effects index (FEI) and fire response accessibility index (FRAI) for this purpose. WFSI represented GIS variables such as historic fire locations, fuel model, canopy closure, aspect, slope, elevation, weather, and fire size. FEI included GIS variables representing fire fighting facilities, tree plantations, a quantification of urban interface, and utility corridors. FRAI included data such as roads, resource locations and water bodies. The variables under these three indexes suggest the possible fire occurrence factors for the south eastern US.

The interaction among various variables such as vegetation, slope, aspect, distance from roads and distance from settlements to determine fire risk areas in Turkey was studied by Esra Erten et al. (2004) using LANDSAT satellite imagery and GIS. They used LANDSAT TM images before (1992) and after (1998) fire to identify burned area, vegetation loss. They used a supervised classification approach to identify the forest land cover and integrated this satellite imagery with GIS parameters such as topography, vegetation type, vicinity to roads and settlements to identify fire risk areas. Their differential weights to these parameters indicated that aspect, slope and landforms are crucial in determining the fire spread. The results also indicated that that fire spread is more rapid on up-slopes and least rapid on down slope with southern slopes more prone to fire.

The roles played by abiotic, biotic and human factors in determining the spatial patterns of wildfire's origin across the upper mid western United States were studied by Jeffrey Cardille et al. (2001). Their research mainly focused on a set of factors (a biotic,

biotic and human variable) that explained fire activity, the variation that exists from these set of factors to the rest of factors and the effect of spatial scale variation on these predictive variables. The results indicated that no single factor or factor type dominates and fire pattern analysis depends on both fire size and fire activity. The factors related to these parameters could be easily interpreted and factors significant at one spatial scale are also significant at the other indicating the analysis strength. The multivariate analysis study revealed that areas with higher population density, higher road density and lower distance to non-forest areas were more likely to catch fire.

Iwan Setiawan et al, 2004 derived fire hazard map using land use, road network, slope, aspect and elevation data in Pahang, Malaysia. Based on the fire history data, the results suggested that southern aspect logged over peat swamp forests have higher fire potential (greater than 40%). The results also showed that fire hazard increases with increasing surface slope.

In general, models can be grouped in to fire behavior, fire potential and fire risk models. Fire behavior is based on fuel factors such as moisture content and fuel temperature of live and dead fuels (Dasgupta et al, 2006). Fire potential is related mainly to climate, topography, anthropogenic influences and vegetation (Cooke et al, 2007). In general, the following variables are considered under each category. Climatic variables such as Precipitation, evaporation, wind and lightening (Cooke et al, 2007), topographic variables such as slope and aspect (Esra Erten et al, 2004), ignition variables such as road density (Gilreath 2006) or interaction among cities population (Raviraj, 2007) or socio-economic factors like age distribution, education level, population density, household income, living conditions, and employment status (Baird et al., 1969) as human ignition source, fuels and landscape have been found to influence probability and distribution of forest fires. The same variables when used to estimate fire size and severity along the likelihood of fire occurrence constitute the fire risk as is the situation under the present study.

Though one cannot control nature, a forest fire risk zone map can be used to evaluate forest fire problems to find satisfactory solutions (Jaiswal et al., 2002) and minimize fire frequency. With all this available information on the GIS variables in various fire models this research had chosen the following four variables which indicates the importance of each variable with respect to the current research.

2.4 GIS variables of ANN

2.4.1 Ignition sources from Human Impacts

Presently human based ignitions are the most common source of wild fires (Whelan, 1995) and human-caused fires have an annual average increase of nearly ten times than lightning-caused fires (Hildebrand, 2003). In the United States, two-thirds of forest fires are attributed to humans (Zhai et al., 2003) that became extremely important in the Northern US as well as the Southern United States where humans serve as fire ignition source either by carelessness or by arson (Raviraj, 2007).

Human risk can be mapped by relating location of fires to specific areas of land use or human activity such as roads, camping sites, cities, forest urban interfaces or a particular land use type are all responsible for fires. (Vliegher et al., 1993; Langhart et al., 1998; Alcazar et al., 1998; Chuvieco and Congalton 1989). The characteristics, contiguity and area size of vegetation, fire ignition sources, fire suppression, and fuel loads are all influenced directly by human factors (Pye et al., 2003). In recent years wildfire risk is an increasing concern where humans live in close proximity to forests (Greenberg and Bradley 1997, Lavin 1997).

Different parameters are used to represent the human component in forest fire modeling such as the road density (Gilreath, 2006), distance to primary roads, secondary roads, urban areas, waste disposal, railways, recreation areas, agricultural works, grazing lands, forestry works and other similar high population density areas (Kalabokidis et al, 2004). All these parameters especially road density indicates the strength of access that human's posses to areas of high fire potential.

The gravity model is one of the earliest spatial interaction models (Fischer, 2000) have been implemented in various fields related to migration and transportation (Raviraj 2007). Spatial interaction is broadly defined as movement of people, commodities, capital and information over geographic space that results from a decision process (Fischer, 2000). Raviraj, 2007, stated that gravity interaction model is generally used for market analysis to analyze the market area surrounding a shopping center laid the foundation of spatial interaction modeling. Newton's law of gravity that utilizes a mass and a friction variable to measure the attractiveness is now extended to analyze local human phenomena.

Using the gravity based spatial interaction approach, Raviraj, 2007 recently studied forest fire prone areas in the South East fire district of MS. The human risk as a spatial component along with fuel variable was used to predict fire occurrences for SE MS fire district. The results suggested that fires occur in clusters. This particular model when validated with the historic fire data proved to be a better representation and estimator of fire risk that represented fire frequency as per fire risk zone as for example very low fire risk is represented by less number of fires and very high fire risk zone contained high fire frequency. In comparison to the to road density, significant estimates were noticed for very low fire risk for all seasons and summer low fire risk while for all other fire risks (medium, high and very high) of other seasons (annual, summer and winter) the Gravity model showed better results. This indicates the importance of human component in forest fire modeling that is associated with the present study area. Thus about 41 cities are present in the study area with a population ranging from 1005 (Sumrall) to 71127 (Gulfport) with interaction among these cities posing different fire level threats in different areas, this research would like to use the recently developed gravity based city interaction as the human component in the present research.

2.4.2 ***Fuel Characteristics***

Vegetation plays an important role in the spread of fires (Countryman, 1972) and the literature review suggests the use of vegetation characteristics such as fuel type that aid fire spread in fire risk modeling (Andrews, 1986; Deeming et al., 1977). Further the forest harvest dumps, construction, mis-management practices and accidental/ intentional / prescribed fires all affect fire behavior and there by fire potential (Zhai et al., 2003).

Cooke et al, 2007 stated vegetation plays a vital role in fuel estimation as particular vegetation types are more predisposed than the other and in MS fires are common in needle leaf conifers particularly in pines, mixed coniferous and broad deciduous stands compared to broad leaf deciduous plants. More over natural disasters such as the Hurricane Katrina or other rapid environmental changes results in significant vegetation damage that increase fuel loads as well as fire potential.

Schultz, 1997 stated that MS mainly contains the even-age needle-leaf evergreen forest stands. Similarly the Southern Fire District mainly contains the needle leaf evergreen forest cover with Loblolly pine as the principal type and has the second largest Loblolly pine stands in US. This Loblolly pine (*Pinus taeda)* is more susceptible to fire compared to the well fire adopted long leaf pines and are particularly susceptible to wildfires when they reach less than 4.6 meters tall (Schultz, 1997).

Wear and Greis, 2002 studied the fuel characteristics of longleaf and Loblolly pine. They sated that the morphological natural adaptation of longleaf pine with widespread root system and rapid growth places the terminal buds of seedlings well above the height of most forest fire flames. The thickened plant parts such as the stem bark along the needles defend the buds and make these plants fire resistant. Loblolly pine though generally grows in the wet areas, presently inhabits even the abandoned agricultural and longleaf pine areas that add more fuel prone to fire.

Fuel type and age of fuel stands are also important in fire prediction in terms of the ignition level. Tanskanen et al. (2005) showed that differences in the moisture regime of surface fuels of different age classes that are dominated by pines results in significant varying ignition conditions which affirm that the age of the fuel stand plays an important role in fire prediction.

Cooke et al, 2007 studied the pre and post Katrina fuel conditions as a fire potential component for southern MS. They used fire occurrence fire data age grid to derive four similar age groups within fire frequency class. Forests of indefinite age and uneven aged mixed forest species are grouped under no origin class while relatively young (10-19), intermediate (20-25), and mature age (26-30 and older) are the other age classes included in the study. Number of fires, average fire size, number of fires normalized by area and percentage of burned area in each class were analyzed to assign the fire hazard to each group. This resulted in unique age type combinations as no origin conifers, conifers 10-19, conifers 20-25, conifers 26-30 and older, no origin mixed, mixed 10-19, mixed 20-25, mixed 26-30 and older, no origin hard wood, broad leaf 10-19, broad leaf 20-25, broad leaf 26-30 and older. Their results suggested that fire hazard has increased after the hurricane Katrina as is evident with increase in to a very high fire risk class from higher risk class in conifers 10-19. The conifers of 26-30 remained at a very high fire risk both before and after Katrina and are considered to be highly susceptible to fire.

Thus pines of 20-30 years were considered to be more fire susceptible and the fire susceptibility might vary as per areas of the forest and their spatial separation. This idea is recently explored to calculate city interaction based on population of cities and their spatial separation using Newton's Gravity Model (Raviraj, 2007). As with the population density, fuel densities with a specific fuel type and age at different spatial separation might pose a different fire threat at different locations that this research explored and used as one the variable in predicting forest fire occurrence.

2.4.3 Topography

A number of fire models exist that are based on the slope and aspect as influences on fire spread. The models include important topographic factors such as steep slopes, aspect, and elevation (Gilreath, 2006). The effect of topography on fire may be positive or negative and is not an important factor in the Coastal Plain of the US (Wade, 1988). In general, topography effects the solar radiation, wind speed, wind direction in an area, and generate wind eddies that increases fire potential (Raviraj, 2007).

Elevation influences vegetation composition, fuel moisture and humidity. More than 90% of cases of forest fires occur at 100 meters above sea level. Most of these fires occur in areas which are below sea level. Fires are less severe at higher elevations due to higher rainfall. Steep gradient increases the rate of fire spread because of more efficient convective preheating and ignition and eastern aspects dry faster since gradients facing east receive more ultraviolet light during the day (Chuvieco and Congalton, 1989).

In another finding Jo et al., 2000 stated that southern exposures have high solar and wind influences compared to the northern slopes. In Southern and South Western aspect, greater than 60% of the forest fires occur on slopes in the range of 0° and 20° and fire hazard increases with increasing slope.

Slope considered to be an important fire variable, is an insignificant factor in wildfire prediction in Mississippi (Zhai et al, 2003). In the Southeastern U.S. fires occurrence chances are high in uplands compared to bottomland. Surrogate slope factors such as spectral differences in Landsat TM images between various forest types that characterize slope are in use in Mississippi (Collins et al., 2005 and Gilreath, 2006).

Though slope is considered insignificant in fire prediction for MS, based on a number of literature reviews discussed above, slope plays an important factor in fire spread that disposes high slope areas at a greater risk. Also with Jo et al, 2000 findings as 60% of the forest fires occur in slopes of 0 to 20 degrees, and as the slope of the study area varies from 0 to 36 degrees, this research utilized slope as one the variable in fire prediction.

2.4.4 Euclidean distance to perennial streams

Water bodies are reported to function as natural firebreaks that affect fire spread (Daniel et al., 2005). Larsen, 1997 studied the spatial and temporal variations in forest fire frequency in the boreal forests of Northern Alberta. They used forest stand age, fire scar and historical data to test the hypothesis that fire frequency varies with mean waterbreak distance (MWD) around a site. The results suggested that fire cycle length varies inversely with the MWD around a site. Further, the relation is highly significant in jack pine and aspen forests than in black or white spruce forests. Thus MWD influence, respectively, variations in forest dominant and fire frequency.

Cyr et. al, 2007 studied the effects of broad scale, fine scale and intermediate factors to analyze the potential influence fire frequency using a proportional hazard model and a semi-parametric analysis. The average distance to water bodies was considered as a potential intermediate physiographic and topographic factor in determining fire frequency. The results showed that a two to six fold variation in fire frequency is related to geographic and topographic factors.

Bergeron, 1991 studied the fire regime of southern boreal forest in relation to landscapes in northwestern Quebec for islands of Lake Duparquet in comparison to the adjacent lakeshore. The results suggested that lake shores have a very few large fires where as islands due to abundant pines, witness less intense and high frequent fires. This suggests a low chance of fire occurrences near to the water bodies in comparison to far off places.

Distance to water bodies is one of the potential factors in forest fire modeling. About 12815 perennial streams are present in the study area that might have a significant influence of forest fires. In general water bodies regulate the surrounding climate and soil permeability that dispose areas with differential fire threats. As inverse relation exists with fires and distance from water bodies in many of the situations. Keeping in view of the existing majority of perennial streams in the study area this utilized Euclidean distance from perennial streams as one the variable in forest fire prediction.

2.5 Fire management

Fire suppression is important in terms of public safety, owned property and protection of resources (Ankica Grant, 2007). In the US, the government expend an annual average of greater than $800 million to mitigate wildfires (NIFC, 2006) and the annual fire protection services and loses account to more than half billion and two billion respectively (Wilson and Davis 1988).

Wright, 2004 stated that wildfire danger rating systems have been in use by many developed countries. A variety of factors influence the size, spread and suppression cost of fires. It is difficult, expensive, and requires lot of time as well as energy to analyze the fire responsible factors in order to combat wild fires. The success in combating pre-suppression and post suppression of wild fires by forest management agencies depends on their understanding and prediction of forest fire.

Burgan, 1998 stated that fire danger rating systems are being implemented by the US Forest service from 1954 and National Fire Danger Rating System (NFDRS) was the first national system. Started in 1972, the NFDRS has been recently revised in 1988. Wildland Fire Assessment System (WFAS) utilize interpolate spot measurements to map national level fire potential and the information is available to the general public and fire protection agencies in every state.

Similarly fire models are being in use by individual states and the state of Florida has developed one such descriptive model using seasonal climatic swifts and highly prone grass landscape to indicate areas of fire potential that poses threat to humans (Gilreath 2006 and Brenner, 2002). Thus fire danger rating systems and fire models are valuable to locate high fire potential areas and implement needed actions by the fire protection agencies.

2.6 GIS and forest fire modeling

Geographic Information System (GIS) is “a comport system for capturing, storing, querying, analyzing and displaying geographically referenced data” (Chang K.T, 2006). Fire modeling is based on a variety of spatially driven factors such as vegetation, topography, weather, management activity, residential pattern, location and neighborhood that needs sophisticated analytical techniques to analyze forest fires (Chou et al., 1990).

GIS is one such sophisticated tool that is capable of organizing and analyzing the complex spatial and temporal traits of forest fires (Zhai et al., 2003) that uses an information and decision support system to enhance management practices (Sunar and Ozkan 2001).

Thus using appropriate information, ArcGIS can be used in modeling to visualize the relation that exists for an event and its associated factors (Raviraj, 2007). This particular software can support various data formats to be exchanged and used for techniques available in other GIS softwares as is the use of Artificial neural works (ANN's) usage from IDRISI Andes for the present study.

2.7 Artificial neural networks

Artificial intelligence (AI) methods utilize modern methods of problem solving such as knowledge-based systems, fuzzy logic, artificial neural networks, and Bayesian belief networks by applying varying computational and algorithmic approaches that integrate human cognitive abilities and provide computers with a capability to solve problems (Schmoldt D.L., 2001). Of the AI approaches, Artificial neural networks (ANNs) deal with decision making using trained patterns with little explicit knowledge (Schmoldt D.L., 2001).

The field of AI also has many applications in the area of ecological modeling (Reginald Mead, 2006) and risk analysis (Schmoldt D.L., 2001). In forestry, ANN's of AI derive the relationship between the dependent and independent variables in a much different way similar to empirical statistical models (Schmoldt D.L., 2001). Mc Cormick et al., 2004 stated that in wild fire modeling the use of ANN is unique in contrast to modern models that integrate the relationships between fire behavior environment variables such as fuel, topography and climate, also capture and analyze the cover, landform and climate interactions that might be unique both temporally and spatially.

Over the last several years, the use of ANN's has increased considerably due to advances in computing performance (Skapura 1996). Also ANN's due to their adaptability and ability to produce classification accuracies higher than those of statistical classifiers, has become prominent in scientific community and witnessed increased research in remote sensing field (Paola and Schowengerdt 1995, Atkinson and Tatnall 1997).

ANN's have been defined as a system that includes many simple processing elements that operates in parallel and the function is determined by network structure, connection, weights and node function (Hara *et al.* 1994).

Similarly Haykin 1994 defines ANN's as a "A massively parallel distributed processor that has a natural propensity for storing experiential knowledge and making it available for use"

Sunar and Ozkan, 2001 describes the the main characteristics of ANN's as:

(i) Intrinsic ability to generalize;

(ii) Make weaker *a priori* assumptions about the statistical distribution of the classes in the dataset than a parametric Bayes classifier, and

(iii) Capable of forming highly nonlinear decision boundaries in the feature space.

Table 2.1 Advantages and disadvantages of ANN (Silva A.P., 2003)

ANN: Advantages	ANN: Disadvantages
Capable to learn non linear and very complex relations	Long training time requirement and Possible over-fitting
Easy to use, implement and integrate the results in a GIS.	Deciding most efficient network structure for a particular problem and
Ability to handle noisy data, and	Inconsistent results due to the initial weights and learning parameters
Good predictive capabilities	Difficult to understand it's internal behavior

Pijanowski et al., 2002 stated that ANN's are powerful tools that utilize machine learning approach to numerically solve relationships between inputs and outputs and are used in wide range of discipline like economics, medicine, landscape classification, mechanical engineering and remote sensing. They further stated that the ANN's posses the generalization and mathematical features to perform well on unfamiliar data that is not adversely affected by errors in the original data and can even perform well on data sets derived by imperfect satellite remote sensing land use or forest type classification that does not hamper the accuracy of the fire risk model at large.

Sunar and Ozkan, 2001 stated that several ANN"s architectures and algorithms have been derived and implemented in the areas such as classification, forecasting and modeling. Of all unsupervised or coarse training and supervised or fine learning are the two primary types presently in use. The unsupervised ANN's are similar to principal component analysis, factor analysis and cluster analysis where ANN's discover statistical regularities in its inputs that to develop different modes of behavior to represent different classes (Sunar and Ozkan, 2001). Supervised ANN"s is similar to regression and discriminant analysis that compares the actual known output with the predicted output (Sunar and Ozkan, 2001).

Hepner et al. 1990 stated that the complex relationships that exist between input variables to predict an output for a given input object can be best modeled using supervised learning approach. Presently, of the available thirty different ANN's (Peng and Wen 1999), SOM has been used in a lot of areas such as data classification, pattern recognition, image analysis, and exploratory data analysis (Jiang and Harrie, 2004). A review of the SOM as is the method utilized for the fire size risk classification is discussed below.

2.8 SOM

SOM approach of ANN is used in this research to classify fire size risk and a brief review of some of the related concepts is reviewed in the proceeding paragraphs.

Self Organizing Map or SOM or Kohonen Neural Networks hereafter referred to as SOM was developed by Kohonen (Kohonen, 2001). One of the fascinating properties of SOM is the automatic detection (self organizing) of the relationships within the set of input patterns (Brand Tso and Mather) and it aids in preserving the toplogical relations i.e. similar close input space patterns will be mapped to the associated close output space and vice-versa (Jiang and Harrie , 2004).

SOM is widely in use for classification problems and environmental studies that more than 4300 papers were present related to SOM (Tran et al, 2003 and Kohonen, 2001) as is the central concept in the present fire size risk classification.

Characteristics of SOM: Wang, S and Wang, H, 2002 stated three unique characteristics of SOM

1. Do not depend on any associated statistical test assumptions and is effective in with high-dimensional data
2. It is flexible and can be utilized in cluster analysis as it ignores the statistical assumptions of the input data and
3. It provides a way to visualize the clusters of high dimensional data that is not present in any other data analysis method.

SOM architecture: The neural network depicted in Figure 2.2 is the two-layer SOM architecture used for a three input variables.

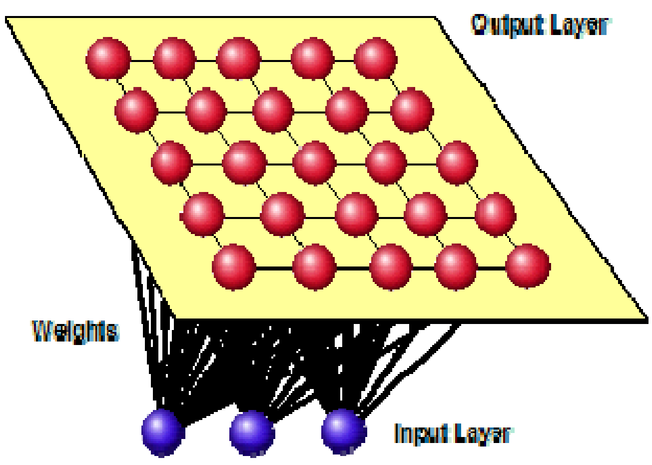

Figure 2.2 SOM architecture for three input layers (IDRISI Andes)

Wang, S and Wang, H, 2002 stated that the lower layer represent input data while the upper layer output nodes represents the organization map of the input patterns after the unsupervised or supervised approach as the case may be. Variable connection weights are used to connect every lower layer node to every upper-layer node.

SOM Normalization approach: Similar to GIS multi criteria decision analysis, SOM also require the variables to be normalized by a transformation technique to comparable units. A number of transformation techniques exist that are in use today of which score range method is widely used for ANN classifications (Basheer and Hajmer 2000; Anthony and Xia Li 2002). In the score range the difference that exists between the minimum score and the raw data is divided by the score range thus standardizing the scores from 0 to 1 (Malczewski, 1999). This allows all the variables in the present study to receive equal attention during the training process and match the synaptic weights range (IDRISI Andes help contents)

SOM Process: Tran et al, 2003 stated that SOM is a two-dimensional neural network whose nodes arrange multidimensional input variables. The process self organizes the input data to a lower two dimensional map of output nodes having very little or no idea of the data input data structure. The output nodes associate with the referenced variables possess the same dimension as the input vectors and represent well defined clusters with similar properties. SOM approach needs prior specification of the output nodes and the output map's configuration before the learning process.

SOM Network Parameters

i) SOM Classification/ Learning Function

As with other classification techniques, SOM also utilize unsupervised or a supervised approaches for classification purposes. Unsupervised classification has been defined as identification of natural groups, or structures, within multi-spectral data and supervised classification as processes of using samples of known identity to assign pixels to classes (Campbell 1987).

Sunar and Ozkan, 2001 studied the use of unsupervised iterative self organising data algorithm and maximum likelihood approach to identify burned fire classes in Turkey. The results suggested that this classification approach yielded a perfect unbiased assessment of burned area.

The unsupervised or coarse tuning of the SOM utilizes the clustering approach. Wang, S and Wang, H, 2002 stated that SOM is a typical artificial intelligence technique of cluster analysis. A cluster is a well defined group of close observations as per Euclidean distance and in general, hierarchical, partitioning and overlapping are the three statistical clustering methods. Each clustering uses its own approach to derive clusters such as allocating the observations to the nearest cluster centers that are renewed and observations are reallocated till a steady state is obtained. SOM uses this idea of clustering that is an alternate to statistical cluster analysis

Miller and Yool, 2002 studied different classification approaches and stated that higher accuracies were obtained using unsupervised iterative self organising data algorithm for classifying multi-spectral data. They found that stratified pre fire vegetation improve the accuracy. Supervised approaches require ground data as training sites with at least fifty points per class that might be impractical for many fires studies and could be substituted by Land sat image accuracy.

The present SOM is one such classification approach that utilizes coarse tuning as unsupervised approach and fine tuning as a supervised approach.

ii) Learning rate

Learning mainly consists of specifying the input, output determination and modifying the weights by specified training rules (Hoffmann, 2005). In general slower learning rate requires more time to produce a well trained system and faster learning rates neglect the minor discriminations that are accomplished by slower training. The learning function generally contains the learning rate parameter that is positive and lies between zero and one. If the value is greater than one, the learning algorithm easily overshoots to correct the weights, and the network will oscillate.

iii) Gain term

Hoffmann, 2005 stated that gain term and neighborhood radius is used to control the speed and accuracy of the approximation (quality of the final surface). The gain term is similar to a Gaussian function that is a measure of the movement of the grid. The gain term value is also positive that lies between zero and one. During the initial iterations (control the steepness of the convergence), a large gain term value is desired that considers the overall shape of the scattered data compared to the final iterations that requires smaller gain term values and as the value reaches to zero, convergence is achieved.

iv) Neighborhood radius

Initially training process utilizes a large neighborhood and small random numbers for connection weights to identify the input vectors of the incoming data to transmit to the output nodes using these connections (Wang, S and Wang, H, 2002).

The neighborhood function is one that decreases with the distance in the output space nearer to the winning unit that is responsible for interaction among different units. It's a time decay function during training (IDRISI andes help contents) with large initial neighborhood radius that slowly decreases in size over time to separate each unit from its neighbor effects (Wang, S and Wang, H, 2002). Many of the SOM implementations reduce the radius to 1 (Figure 2.3) to have a neighboring effect even in the final stages of training while few implementations set this parameter to decrease to zero.

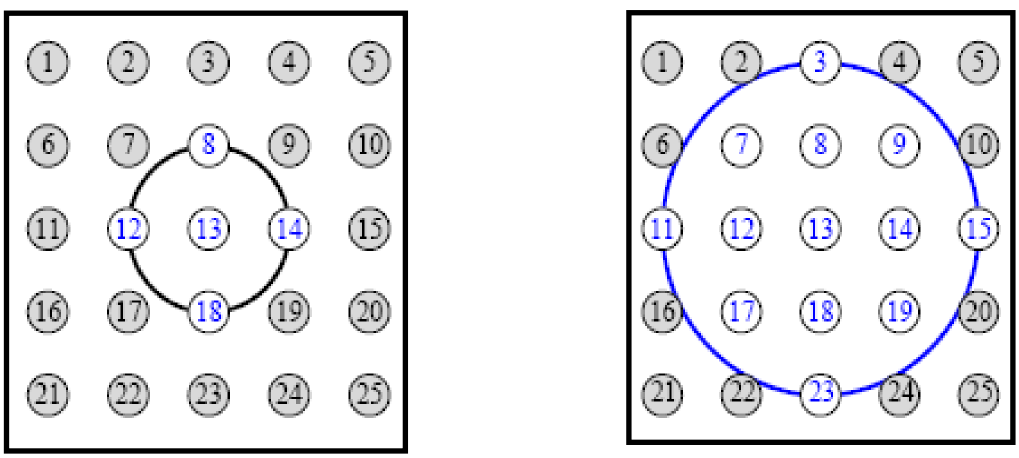

Figure 2.3: A two dimensional neighborhood of radius d=1 (left) and d=2 (right) around neuron 13 (Source: Neural network Tool box)

Later the weights are updated according to the Kohonen's learning rule (Kohonen, 1989) to find an active output node in a “winner take all” competition that is the output node with weights most similar with the local density function of the cluster centers approximately the probability density function of the input vectors (Wang, S and Wang, H, 2002). In general the weight update takes place for the active output node and its topological neighbors to represent organized output nodes that are realm clusters derived without an idea of a priori cluster centers (Wang, S and Wang, H, 2002; Fig. 4.13).

2.9 ANN based models for image classification

Filippi and Jensen, 2006 studied the utility of Fuzzy unsupervised ANN (Fuzzy Linear vector quantization ANN) in classifying the AVIRIS image. The results are at par with the traditional multilayer perceptron. A classification accuracy of 82.82 % and 84.66% were obtained for FLVQ and MLP respectively. FLVQ proved to be computationally efficient and less time consuming than all other supervised and unsupervised algorithms.

Arora et al., 2004 studied land slide hazard zontaion using artificial neural network approach in the Bhagirathi (Ganga) valley of Himalyas. The study utilized IRS – 1 B sensor data to produce a classification accuracy of around 80% with a small training data and showed same trend to existing ground truth locations.

Sanchez et al., 2003 used ANN as a tool for mineral potential mapping with GIS. A trained network is used to estimate an efficient gold potential map. The results stated that ANN is an effective tool for mineral exploration spatial data analysis.

GIS and ANN based land transformation model (LTM) was used to forecast and assess the impact of urban sprawl in coastal watersheds along eastern lake Michigan (Pijanowski et al., 2002). This LTM used a multilayer perceptron neural net to forecast urban use changes over 2020 and 2040 using non urban sprawl and urban sprawl trends. This approach was able to characterize two of the nine watersheds under study area consideration to experience the most urban change in the next 20-40 years.

ANN was used in land cover classification using remotely sensed data (Kavzoglu and Mather, 2003). They studied the optimum design of ANN for classification problem. They further compared this optimum ANN design with that of the heuristics approach and maximum likelihood classifier. The results suggested a higher classification could be achieved with the use of optimized ANN approach as it incorporates all relevant ancillary data.

In a study to characterize the relative suitability of environments for forest types in a complex tropical vegetation mosaic, Hilbert et al., (1999) used ANN to characterize the relative suitability of forest classes defined by their physiognomy and canopy structure. The study utilized seven climate, nine soil parent material classes and seven terrain variables. The model is highly successful in accurately predicting 75% of forest mosaic compared to the 28% accuracy achieved by maximum likelihood method. The authors stated that ANN approach has high potential for climate analysis and vegetation patterns where use of ANN is highly applicable.

Sunar and Ozkan, 2001 used an Artificial neural networks model for IRS-1C image data to estimate the burned area by classifying the image. The network architecture utilized two input units, five nuerons in the first hidden layer and fifteen second hidden units and six output units representing landcover classes. The study utilized 864 training pixels represnting appropriate land covers classes. The learning rate used for this purpose is 0.3 with 2000 iterations. The results of ANN in terms of the area burned (6294 ha) was little bit higher and closer to the actual area burned (7094 ha) compared to the conventional methods (6290 ha).

Pijanowski et al., (2001) used a neural network base urban change model for two metropolitan areas of the upper midwest of the united states. The study utilized three model types and four model performance metrics. The results indicated that neural net model in most cases performed well on pattern analysis and not on location using Kappa as the performance metric. Pijanowski et al., (2002) used land transformation (LTM) to study the effects of nine factors (roads, highways, residential streets, rivers, Great lakes coastlines, recreational facilities, inland lakes, agricultural density and quality of views) that influence urbanization pattern in coastal watershed. ANN was implemented to learn the patterns of development and test the predictive ability of the model. The results indicated that the model performed well with a relatively high predictive ability of 46% at a higher resolution.

Franzini et al., (2001) studied the connections using SOM in a group of variables related to social, economic and environmental to analyze Milan urban system complexity. They concluded SOM is better than classification tree and factor analysis approaches. They stated that SOM uses an unsupervised method of fuzzy and self organized approaches to analyze nonlinear correlation that don’t need any exogenous rules. Further

SOM is found to be a powerful tool to reduce and synthesize information to derive a base knowledge from a complex system.

Weijian wan and Donald Fraser (1999) developed a concept of multiple self organizing maps (MSOM's) for multisource fusion and compound classification. Their concept is based on Kohonen SOM and results indicated by kappa index of agreement stated that MSOM can be used for multi source fusion with high dimensionality, complex characteristics and disparity.

Bacao et al., (2005) in his research on self organizing maps as substitutes for K means clustering stated that Kohonen's self organizing maps is most effective method with proper training parameters.

Basheer and Hajmeer (2000) in their article on ANN stated that ANN's are recent computational tools to solve complex real world problems. These models are empirical in nature that can be used to derive accurate solutions for precise or imprecise problems.

Suwardi Annas et al., (2007) in their research compared Self organizing map and principal component analysis methods for classifying and visualizing fire risk in forest regions. The results suggested that SOM is better suited than PCA for visualizing fire risk distribution in forests. The color coding and labeling also proved to be effective in visualizing the classified fire risk.

2.10 Accuracy assessment

RMSE, error or confusion matrix and Kappa statistics were the common methods employed to assess map accuracy (Prabhu, C.L, 2006). Of these error or confusion matrix and Kappa statistics are discrete multivariate techniques to measure map accuracy

(Prabhu, C.L, 2006). The error matrix is valuable to assess the overall map and individual class accuracies enabling calculations in terms of producers and user's accuracy (Prabhu, C.L, 2006).

Filippi and Jensen (2005) stated that producer's and user's accuracy evaluates the omission and commission errors respectively. Omission and commission errors were obtained by dividing these correctly identified pixels by column row totals respectively for that class.

The kappa coefficient measure considers the proportion of agreement between data sets that is due to chance alone and offsets the chance agreement by considering off diagonal elements (Prabhu, C.L, 2006).

Simpson, 2008 stated that it's a value that is less than or equal to 1. A value of 1 indicates a perfect agreement while less than that is interpreted accordingly by different authors. The value of Kappa can be derived using equation …… (Simpson, 2008)

$$\kappa = \frac{p_o - p_e}{1 - p_e} \quad \text{……………………Equation 2.1}$$

where P_o and P_e are observed and expected level of agreements under consideration. Some of the possible interpretations of Kappa value are listed by Simpson, 2008. (Table 6.1a & b).

With all this available ready hand information, the present forest fires research is a ecological process that require ecological modeling with ANN's method of the available AI techniques. Further as SOM of ANN's is widely implemented in classification problems as is the present fire risk classification, this research would explore the use of

SOM approach with the justified variables of citi interaction, fuel density interaction, slope and distance to water bodies.

CHAPTER III

RESEARCH OBJECTIVES

Literature review justifies the use of the proposed four variables considered in this study (city interaction, Fuel density interaction, slope and Euclidean distance to perennial streams). All the four variables were treated as independent layers to derive a fire potential model. The fuel density interaction among fuels was calculated similar to cities interaction (Raviraj, 2007). Fire is an ecological process that requires an understanding of the ecological modeling. ANN's of AI has been found to be extensively useful in ecological modeling that extracts relationship between dependent and independent variables. Further SOM of ANN's finds its valuable application in the classification problems.

This study tests ANN's applicability for predicting forest fire potential for SE fire district of MS. Cities, perennial streams and historic fire locations were used as a part of the vector data. Similarly fuels, digital elevation model (DEM) and Euclidean distance to perennial streams were used as a part of raster data format. For this research city interaction layer was obtained from the previous work of Raviraj, 2007. The pine class with 20-30 years of age is considered to be highly prone to fire. The density of fuels was assessed by Newton's Gravitational theory similar to city interaction calculation. Euclidean distance to perennial streams and slope layers were derived using the spatial analyst GIS tools and relevant data sources. All the variables are treated as inputs in an artificial neural network model in order to find the dependency of historic fire occurrences with respect to these four variables to predict fire risk.

Any fire potential model should include historical fire occurrence data for model validation (Chuvieco et al., 2004). A five year period of fire occurrences from 1999 to 2005 in SE Mississippi fire district was utilized for this ANN model validation.

Therefore, the primary objective of this research is to implement ANN’s SOM model for the associated fire input variables and test the accuracy of predictions in potential model for the Southeast Fire District of Mississippi. The details pertaining to data and methods mentioned above are discussed at length in the next chapter.

CHAPTER IV

MATERIALS AND METHODS

4.1 Study area

The study area is located in south eastern MS. The Longitude and Latitude for MS are 88°7'Wto91°41'W and 30° 13' N to 35° N respectively (Net state.com).

The state is surrounded by Tennessee in the North and the Gulf of Mexico on the south. Mississippi borders Alabama on the east and Arkansas and Louisiana on the west (Net state.com). The average elevation for the MS is 300 feet above sea level (Net state.com).

Specifically the study area covers 22 counties in MS popularly called the Southeastern Fire District of MS (Figure 4.1). It includes the mid Coastal plain -gentle hill topography, developed drainage, and diverse soils (Schultz, 1997). The Lower Coastal Plain includes well drained forest soils and deep sandy alluvial soils (Schultz, 1997).

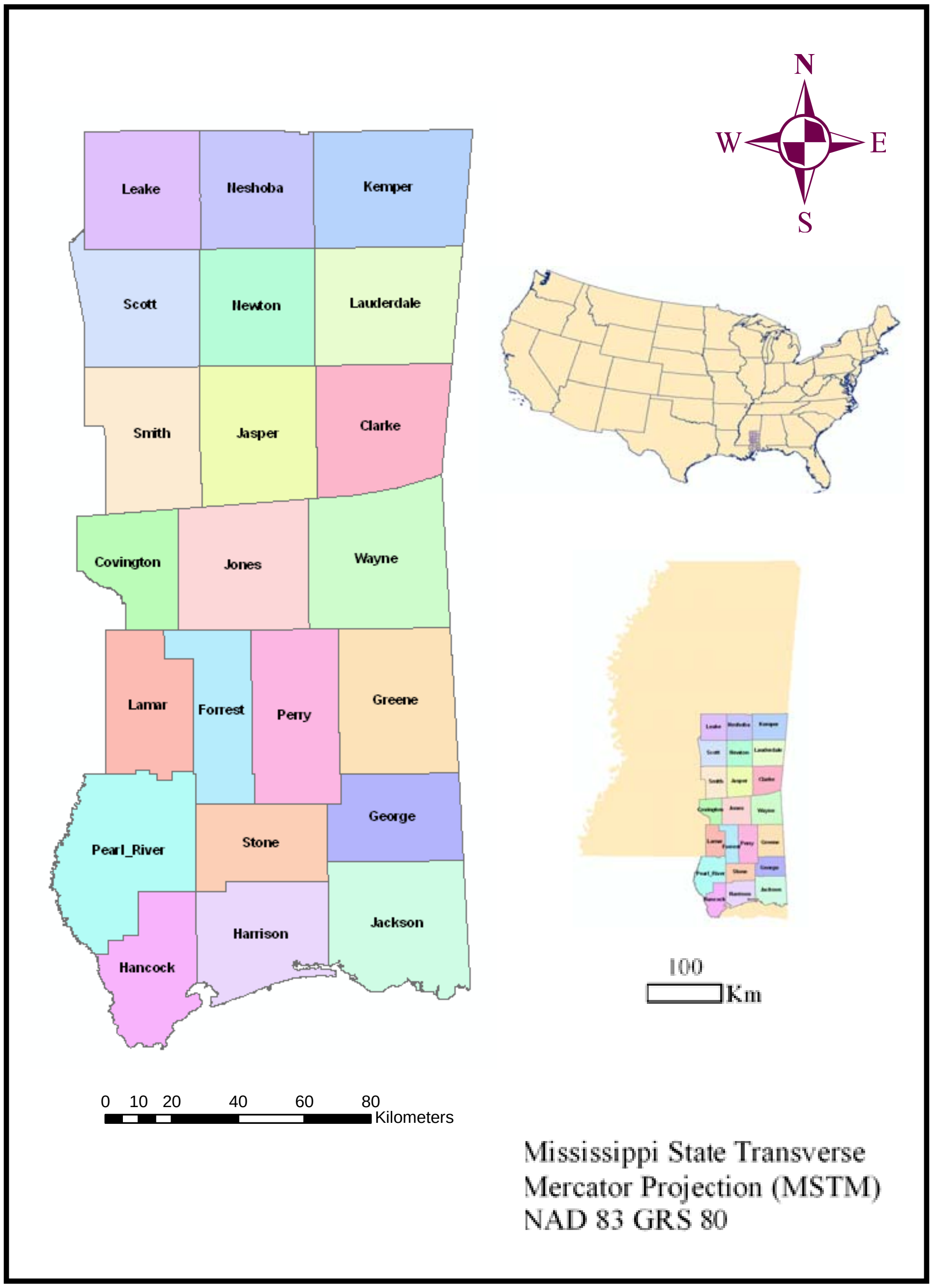

Figure 4.1: Counties in Southeast fire district of MS

4.2 Vector Data

GIS spatial data can be physically represented either in a rater or vector format. Vector data format utilizes strings of coordinates to represent features in terms of points, lines and polygons. The vector data utilized in this research is outlined in figure 4.2 and is described below.

4.2.1 Cities layer

The city theme layer was utilized to derive the city interaction layer using Newton's law of spatial interaction by Raviraj in 2007. The cities polygon layer was obtained from Mississippi Automated Resource Information System (MARIS). The state wide cities theme layer is clipped to the study area and projected to a Mississippi Transverse Mercator (MSTM).

4.2.2 Perennial streams

Perennial streams vector data is important in terms of calculating the Euclidean distance from sreams to historic fire occurrences. Perennial streams data is available as a vector (line) file in MARIS. This state wide theme layer is obtained from MARIS, clipped to the study area and projected to MSTM.

4.2.3 Historic fire location dataset

The historic fire data from 1999 to 2003 was obtained from the Mississippi State Forestry Commission. The points are entered as x, y coordinates with an attribute table, projected in a Mississippi Transverse Mercator (MSTM) projection, and stored as a shape file. These fire locations are important in training the SOM as well as validating the model results.

4.3 Raster data

Raster format data utilizes a two dimensional matrix of uniform grid cells that are homogeneous with square, rectangular or regular in shape. These cells are usually referred to as pixels. This research utilized the raster data that is outlined in figure 4.2 and is explained below. A 30 m resolution is utilized for these raster grids with MSTM projection.

4.3.1 Fuels

Satellite images were obtained from satellite imagery of Landsat earth observation program for approximately every five years from 1974 to 2003 (Cooke et al., 2007). The available data from Multispectral scanner (MSS), Thematic Mapper (TM) and Enhanced Thematic Mapper (ETM+) was rectified to a common base map to derive forest age of the of the available land cover in 2003. Also forest type and land cover information was used to finally obtain a unique age species layer. This work was done at the spatial information technology lab of Mississippi State University by Collins (Collins et al., 2005). These unique age and species layers are important to obtain the fuel density interaction layer.

4.3.2 *Digital elevation model (DEM)*

DEM is available on MARIS website with 30 m resolution and MSTM projection. The DEM was processed by university of Mississippi Geoformatics centre (UMGC) and is available in compressed interchange (.e00) format for individual counties. The meta data is available on line at the MARIS site for further information. The grids were sink filled and mosaicked using the mosaic tool in Arc Map platform DEM was used to calculate the slope of the study area that facilitate fire spread and their by fire risk.

4.3.3 *Euclidean distance*

Perennial stream data obtained from MARIS was used to derive the Euclidean distance from the streams in ArcMap. The resultant is a raster layer with 30 m resolution with MSTM projection that represents the distance from each stream to fire occurrences in the study area.

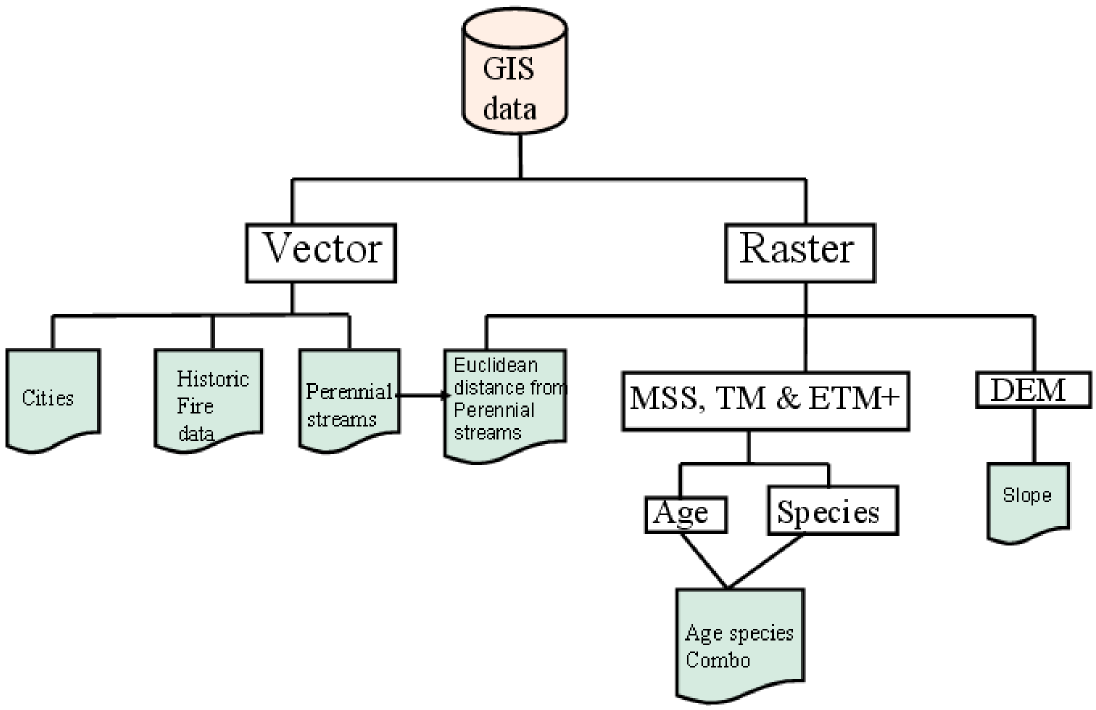

Figure 4.2: GIS data used for fire risk analysis

The common coordinate system (MSTM) ensures the accurate overlay of all the layers. GIS allow integrated analysis of spatial and attribute data. These data can be manipulated and analyzed to derive information suitable for a particular application (Malczewski, 1999). Table 4.1 and figure 4.3 outline the summary of variables and research work respectively to derive information related to fire risk analysis of the SE fire of MS.

Table 4.1 Summary of the variables utilized for fire risk analysis

	GIS Data Format	Variable	GIS Data Format	Source	Resolution
City	**Vector**	**City interaction**	**Raster**	**Raviraj, 2007**	**30x30 m**
		Fuel density interaction	**Raster**	**MSS, TM & ETM+**	**30x30 m**
DEM		**Slope**	**Raster**	**MARIS**	**30x30 m**
Perennial streams	**Vector**	**Distance to perennial streams**	**Raster**	**MARIS**	**30x30 m**

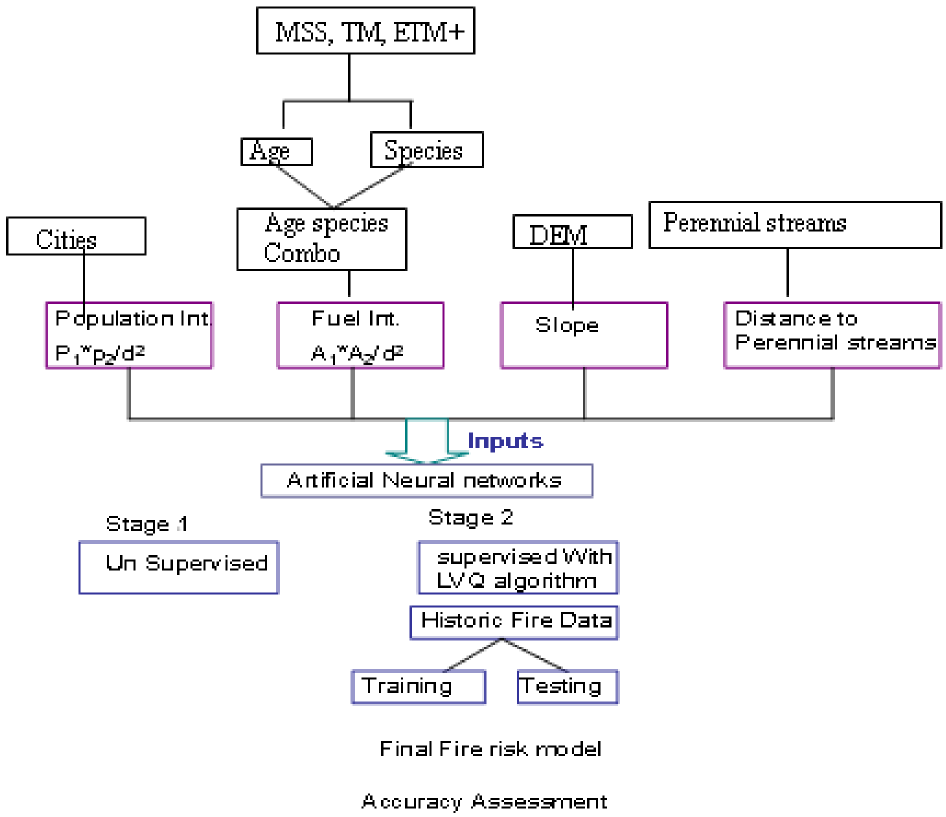

Figure 4.3: Flow chart depicting the fire risk analysis of the SE fire district

4.4 Data Preparation

4.4.1 City /population interaction

This work was carried out by previous master's student in the Dept. of Geosciences (Raviraj, 2007). The cities layer was obtained from the MARIS. The layer was clipped to the study area. Centroids were generated for each city. Distances of one city area to other city areas was calculated using a GIS distance tool. The population and distance factors of the cities layer were used to calculate the spatial interaction using the gravity equation (Equation 4.1). From this Gravity spatial interaction a population interaction layer is generated (Figure 4.4). A threshold of 1000 people is considered as the basis of population interaction. Interaction is the sum of gravitational effect that a particular city population posses as a function of distance and population of rest of the cities in the study area.

$$CityInteraction = \sum_{i}^{n} \frac{P1 * P2}{d^2} \quad \text{…………………………..Equation 4.1}$$

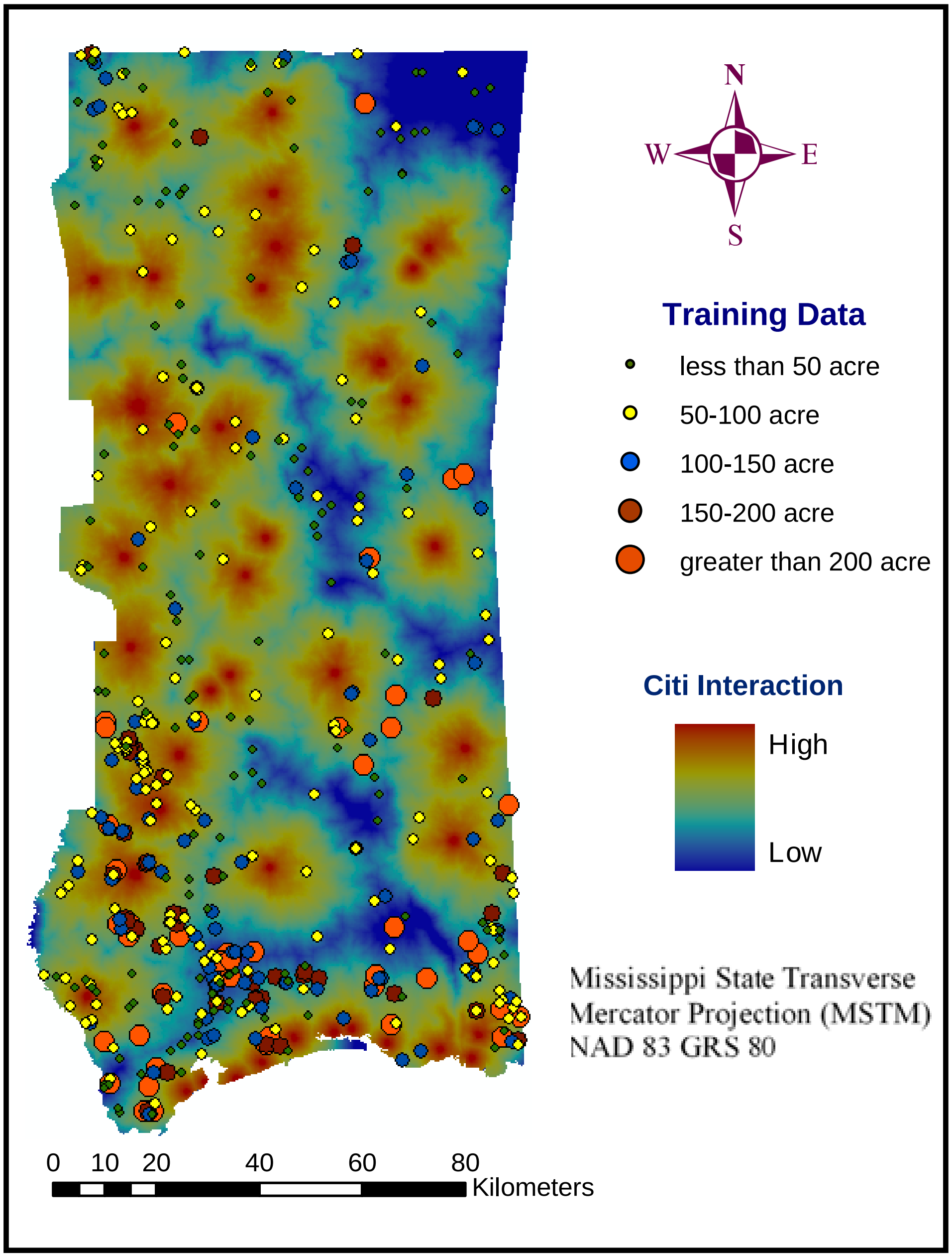

Figure 4.4: City interaction with respect to fire size of SE fire district

4.4.2 *Fuel density interaction*

Land sat MMS, TM and ETM+ images were used to derive thematic maps of land cover types (Table 4.3) and forest stand age (Table 4.2). A post classification change detection method in which two temporal images are classified individually, then compared to identify changed pixels (Cooke et al., 2007). Further unsupervised approach is used to determine forest type (Cooke et al., 2007). Once the thematic classifications were obtained, a unique reclassification scheme was used to assign unique integer values for forest age and land cover type. Additive map algebra is used in raster calculator to obtain a species age combination layer (Table 4.4 and Figure 4.5).

Table 4.2 Unique Age values used to Age Classes Derived from MSS, TM & ETM+

Age Classes	Unique Values
Open Areas	10
Regenerating Areas	20
Non-Origin (>30 years) Forest	30
Zero to Nine years	40
10 to 19 Years	50
20 to 30 Years	60
Greater than 30 Years	70

Table 4.3 Unique Values used to Land Cover Classes Derived from MSS, TM & ETM+

Land Cover Classes	Unique Values
Open Areas	1
Regenerating Areas	2
Broadleaf Deciduous Forests	3
Mixed Broadleaf Deciduous/Needle-leaf Evergreen Forests	4
Needle-leaf Evergreen Forests	5

Table 4.4 Unique species and age values used to derive pine 20-30 yr age group layer

Species age category	**Risk potential**
Non-origin HW, HW:20-30 yr, HW: >30 yr	0(Very Low)
Open, HW: 0-9 yr, HW:10-19 yr	1
Non-origin MX, MX : 0-9 yr, MX: 20-30 yr, MX: >30 yr	2
Non-origin Pine, Pine: 0-9 yr, MX: 10-19 yr, Pine:>30 yr	3
Pine(10-19 yr)	4
Pine(20-30 yr)	5(very High)

Note: HW: Hardwood; MX: Mixed vegetation; yr: age in years; > Greater than

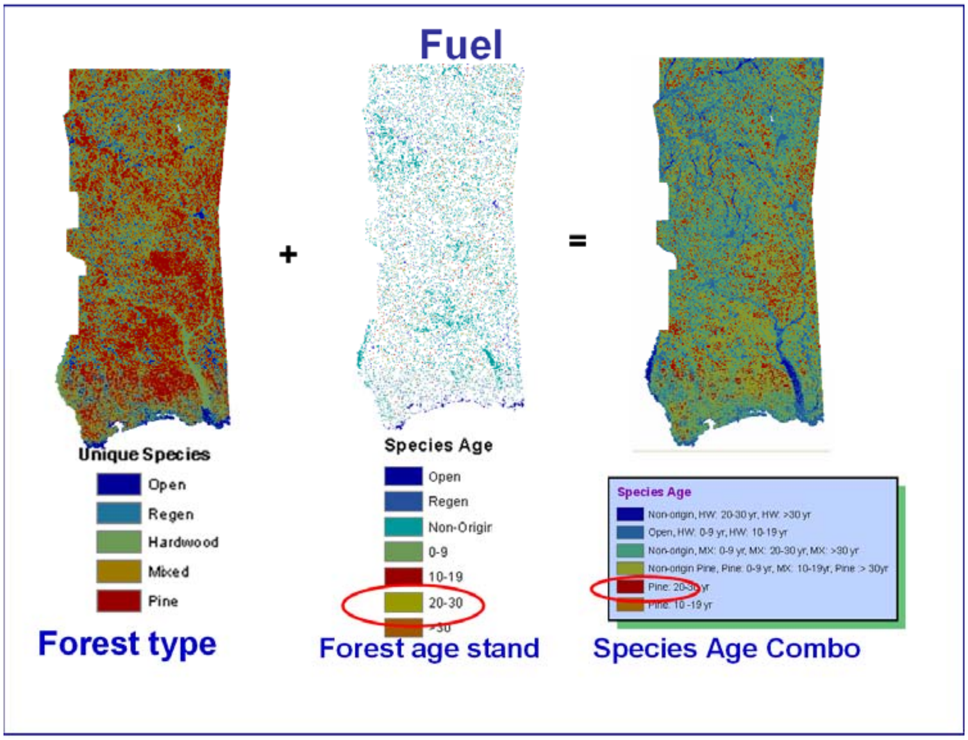

Figure 4.5: Unique Age and species used to derive pine 20-30 yr age group layer

Post Katrina studies (Cooke et al, 2007) indicated that pine class with 20-30 years age group is highly prone to fire. This particular species age combination is separated using reclassification and additive map algebra methods (Table 4.4) for further analysis to obtain fuel density interaction layer. The steps involved are Clump and sieve, filtering or sieve, centroid generation, and Point distance calculation

Clump and sieve method provides a way to generalize classified images. It is a form of contiguity analysis that can be utilized to group similar pixels called raster regions or clumps (ERDAS Field Guide). Contiguity analysis can be used either to create raster regions for a large class or eliminate too small raster regions not fit in an application (ERDAS Field Guide). Fuel size is also an important factor in determining fire frequency and generally high fire frequency is associated with large fuel size (Johnson and Van Wagner, 1985). Similar pixels in the pine of age group 20-30 yr are grouped using this technique.

In situations, where very small clumps are not practical (require a minimum area for a particular application as less than one acre forest in the present study is not practical to consider as forest fire), they can be sieved out according to their sizes (ERDAS Field Guide). Areas were calculated to these fuel (pine 20-30 year age) clumps. The minimum area for classification as forest land is one acre (FIA Glossary). Consequently, a threshold size of 1 acre is used to filter these clumps.

Centroids were generated for these clumps (pine with 20-30 yr and above 1 acre) using ArcMap. Arc GIS proximity analysis tool was utilized to determine the distance of a particular centroid (point feature) to rest of the centroids (Near Features). The result is an output database file containing the fields of input feature id, near feature id and distance (Arc GIS desktop help). Default search radius is utilized for this purpose which implies, point distance is calculated from one input feature to all other available (5561 clumps) near features.

Fuel density Interaction

Interaction calculation is based on the Newton's gravity equation (Equation 4.2 and Figure 4.6). The equation is similar to the city interaction calculation that states that interaction is directly proportional to the area each clump and inversely proportional square of the spatial separation between these clumps. Thus interaction is the sum of the gravitational effect that each particular fuel type and age clump area posses as a function of distance and area of rest of the clumps. A total of 5562 clumps were found in this study area with area greater than one acre. The following equation is used to calculate the interaction of one clump to the rest of the clumps.

$$FuelInteraction = \sum_{i=1}^{n=5562} \frac{a_i * a_n}{d^2} \qquad \text{Equation 4.2}$$

a = Area of fuel clump/stand;

i = 1 to 5562;

d = distance between two fuel clumps/stands

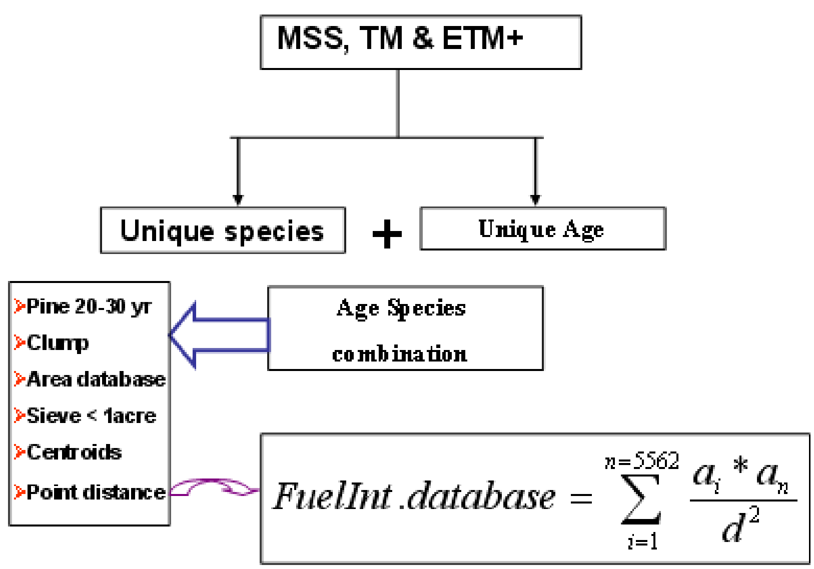

Figure 4.6: Flow diagram depicting the Fuel density Interaction variable for fire risk analysis

The summation of the interaction value (Figure 4.7) of a clump with respect to all other clumps thus obtained was populated back to the fuel attribute table thereby generating a fuel density interaction layer (Figure 4.8)

	A	B	C	D	E	F	G	H
1	Orgin	Dest	Org2Dest	A1	A2	Distance	Interaction	
2								
3	a0	a1	a0a1	443498.0	1894810.0	10035.7549098	8343.672336799604	
4	a0	a2	a0a2	443498.0	714573.0	1585.35166516	126091.85732235099	
5	a0	a3	a0a3	443498.0	368644.0	35699.3719925	128.2855329981193 8	
6	a0	a4	a0a4	443498.0	336124.0	41287.6771629	87.44808196819433	
7	a0	a5	a0a5	443498.0	1794180.0	39797.3432684	502.39986586529915	
8	a0	a6	a0a6	443498.0	337599.0	35349.1442481	119.82157060880282	
9	a0	a7	a0a7	443498.0	227797.0	54929.4517148	33.48336813877346	
10	a0	a8	a0a8	443498.0	256957.0	16983.758678	395.079502518545	
11	a0	a9	a0a9	443498.0	42349.5	58225.562306	5.540035786683749	
5550	a0	f548		443498.0	480443.0	299175.629506	2.380570835813862	
5551	a0	f549		443498.0	1343410.0	300097.466497	6.615696667390177	
5552	a0	f550		443498.0	200112.0	300234.931295	0.9845603889681006	
5553	a0	f551		443498.0	2417070.0	302266.269929	11.732796019775748	
5554	a0	f552		443498.0	66583.6	302395.779872	0.3229293260874891	
5555	a0	f553		443498.0	152412.0	302210.706003	0.7401012485972359	
5556	a0	f554		443498.0	31202.0	301546.907644	0.15218236881674718	
5557	a0	f555		443498.0	42882.7	302872.190711	0.20732658286550796	
5558	a0	f556		443498.0	21756.5	302132.431745	0.10570268140954742	
5559	a0	f557		443498.0	2741940.0	304218.702154	13.139468388452512	
5560	a0	f558		443498.0	204070.0	302900.322196	0.986441555989731	
5561	a0	f559		443498.0	133905.0	303847.226569	0.6432472028507216	
5562	a0	f560		443498.0	167350.0	305533.246805	0.7950609659080038	
5563	a0	f561		443498.0	1222090.0	254700.547216	8.354781569498424	

Total Interaction: 1023368.3365672121

Figure 4.7: Table indicating the Fuel density Interaction calculation

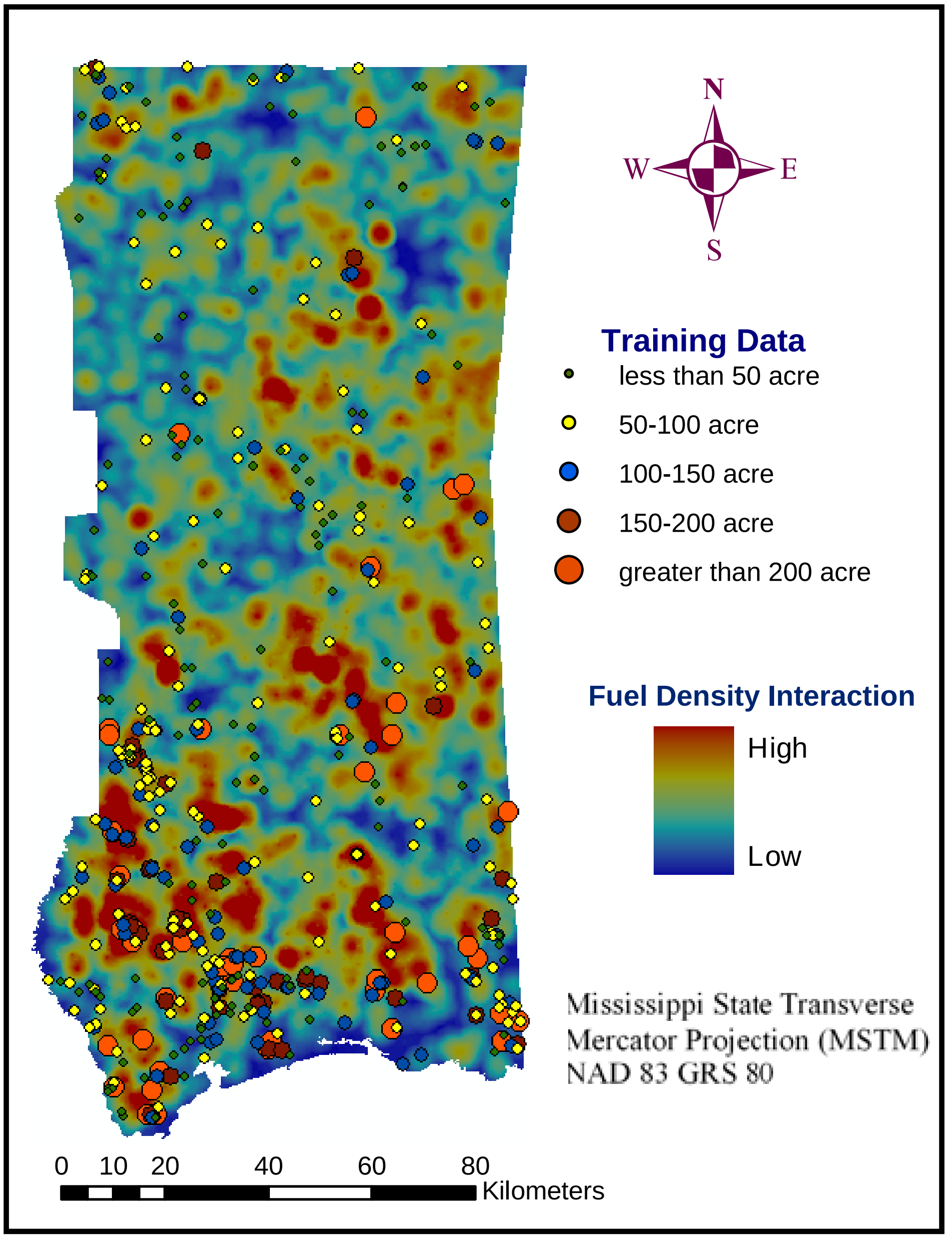

Figure 4.8: Fuel density interaction with respect to fire size of SE fire district

4.4.3 Slope

Slope is also one of the important factors considered in fire risk analysis. Slope is calculated as the rate of change in value of a cell to its neighboring cells, usually for an eight pixel window that identifies the steep down hill descent from that cell. Lower slope indicates flat terrain, and higher slope values indicate steep terrain. Literature review states that steep slopes are associated with higher fire potential compared to gentle or flat slopes. From the mosaiced DEM, slope is calculated (Figure 4.9).

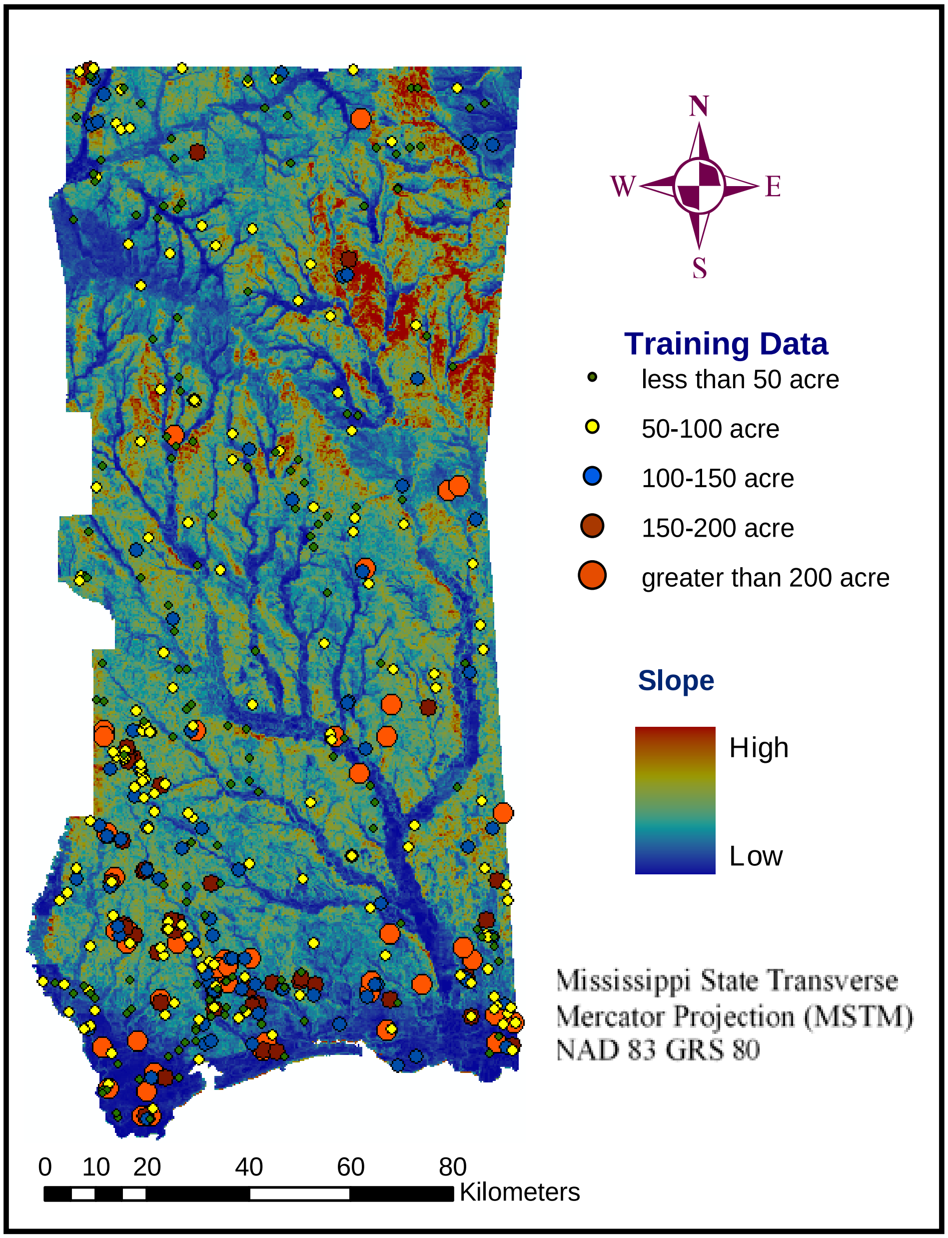

Figure 4.9: Slope with respect to fire size of SE fire district

4.4.4 Euclidean distance from perennial streams

The streams were clipped to the study area and Euclidean distance to these streams feature is calculated using Arc Map spatial analyst straight line distance tool (Figure 4.11). Euclidean distance function initially converts the streams line vector format to a raster with the specified cell size (30 m). The distance algorithm was then used to calculate the distance from the center of the stream source cells to the center of each of the surrounding cells using hypotenuse, with the x-max and y-max as the other two legs of the triangle (Figure 4.10). The shortest distance to a stream source was determined and the values were assigned to the cell location on the output raster.

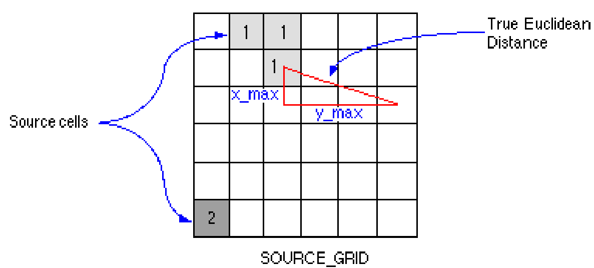

Figure 4.10: Euclidean distance calculation (Source: Arcmap desktop help)

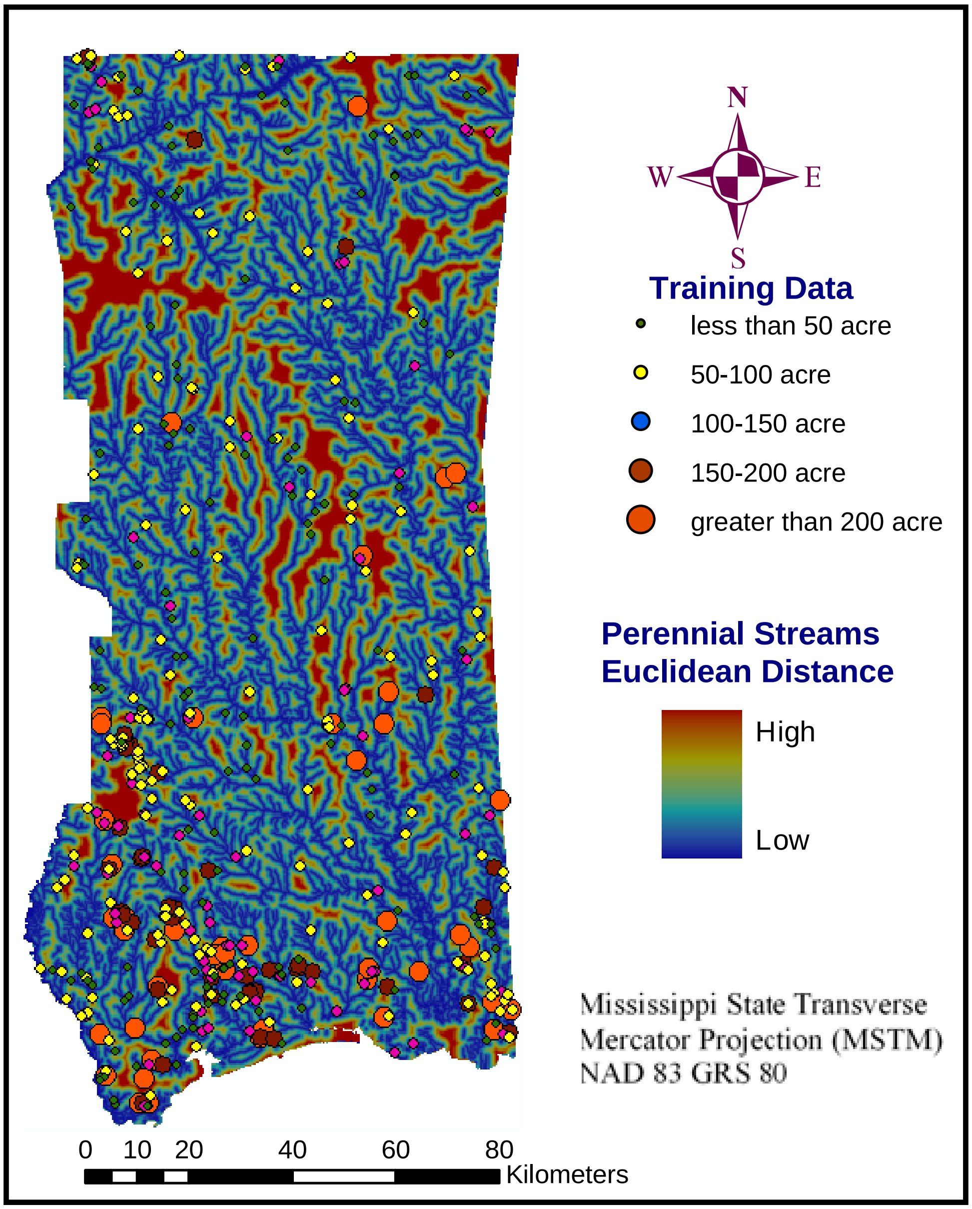

Figure 4.11: Result of Euclidean distance with respect to fire size of SE fire district

Data Normalization

The variables used in the analysis were in different ranges and different units. For example slope is measured in degrees and the values range from 0 to 36. Similarly Euclidean distance to perennial streams is measured in metere. Based on Basheer and Hajmer, 2000 the following linear transformation (Equation 4.3) is used to proportionally normalize the data between 0 and 1 in the range of the maximum and minimum values

$$x_i = \frac{z_i - z_i^{\min}}{z_i^{\max} - z_i^{\min}} \quad \text{……………………………Equation 4.3}$$

Where

x_i is the normalized value of Z_i

$Z_i^{\min}$ Minimum value of Z_i

$Z_i^{\max}$ Maximum value of Z_i

All the layers were re-projected to the Mississippi State Transverse Mercator (MSTM) projection with a 1983 North American Datum (NAD83) and the Geodetic Reference System of 1980 (GRS 80). The independent variables of fuel density interaction, human interaction, slope and distance to perennial streams were maintained as raster grids with a 30-meter resolution. These independent layers are used in ANN (SOM) to generate a fire potential prediction.

Artificial neural networks

The ANN analysis (SOM) was carried out according to the procedures outlined in IDRISI Andes GIS software. Figure 4.12 describes the ANN-SOM architecture and topology utilized for fire size risk prediction of SE fire district of MS. ASCII format was used to import raster data into IDRISI Andes and for output. All four variables analyzed were entered individually to the network architecture. These variables were analyzed with network parameters related to neighborhood radius, minimum and maximum learning rates, minimum and maximum gain term, fine tuning rule and the number of epochs to obtain a best classified fire risk map.

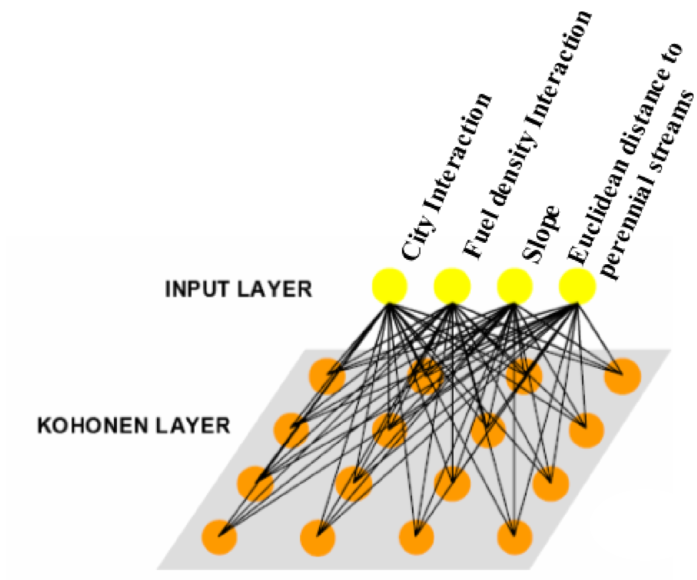

Figure 4.12: SOM architecture and topology used for forest fire risk (Source: Lidia Diappi et al 2002)

Before analyzing these variables in terms of classified fire risk, the variables are first trained using the historic fire locations. Five classes of training data sets were chosen based on fire size. These classes were appended and were brought to a usable vector training format in IDRISI Andes version. First a coarse training is used that is similar to an unsupervised approach where competitive learning and lateral interaction result in fundamental topological regions of neuron weights that represent clusters and sub clusters in the input data (IDRISI Andes help contents). The following represent the set of variables or neurons utilized for this research.

X = {City interaction, Fuel density Interaction, Slope and Euclidean distance to streams}, a four dimensional feature vector input to the neural network architecture. The Euclidean distances (Equation 4.4) between weight /reference vector and input feature vector were calculated in the SOM tool using the following equation (IDRISI Andes help contents)

$$d_j^2 = \sum_i^k (x_i^n - w_{ji}^n)^2 \quad \text{……………………… Equation 4.4}$$

Where

x_i^n is the input to neuron i at iteration n and

W_{ji}^n is the synaptic weight from input neuron i to output neuron j at iteration n.

Finally the neuron in the output layer is determined based on minimum Euclidean distance measure that is calculated as shown in equation 4.5.

$$Winner / \min . d_j = \sqrt{\sum_i^k (x_i^n - w_{ji}^n)^2} \quad \text{………………..Equation 4.5}$$

Learning rate is used to alter the weights of the winner and the specified neighborhood radius according to the equations 4.6 and 4.7 below (IDRISI Andes help contents).

$$w_{ji}^{n+1} = w_{ji}^n + \alpha^n (x_i^n - w_{ji}^n) \quad \text{If d winner i is with in } \gamma^n \text{……….Equation 4.6}$$

$$W_{ji}^{n+1} = W_{ji}^t \quad \text{If d winner i is not with in } \gamma^n \text{…….Equation 4.7}$$

α^n

Where is the learning rate at n th iteration and

d winner j is the distance of the winner neuron and other neurons in the output layer.

The above two equations implies that the winner neurons are altered by the learning rate (Figure 4.13) within the specified neighborhood radius while outside the specified radius are left unaltered (IDRISI Andes help contents). For this situation a high neighborhood radius is specified initially in coarse training in comparison to the fine tuning. The learning rate is a value that can set in the range of 0 to 1.

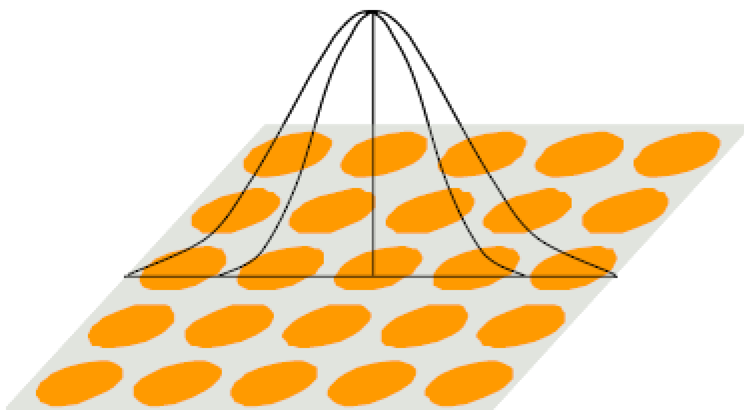

Figure 4.13: SOM weight update function (Source: Lidia Diappi et al 2002)

This is the process that occurs in the coarse or unsupervised training phase. In the fine tuning phase, the decision boundaries are refined between classes based on training data. This helps to improve the accuracy of classification. For the purpose of fine tuning Learning vector quantization (LVQ) parameter is used (IDRISI Andes help contents). The LVQ parameter utilized for this research is uses the following equations 4.8, 4.9 and 4.10.

$$w_{ji}^{n+1} = w_{ji}^{n} + \delta^{n}(x_i - w_{ji}^{n})$$ If classified x is true……….Equation 4.8

$$w_{ji}^{n+1} = w_{ji}^{n} - \delta^{n}(x_i - w_{ji}^{n})$$ If classified x is False…………..Equation 4.9

$$W_i^{n+1} = W_i^{t}$$ If i and ji are not equal…………………………………Equation 4.10

Where W_{ji} is the weight vector of the winner neuron and

δ^n is the gain term that can be adjusted in the range of 0 to 1.

In the process of this neural network architecture, learning rate and gain term are all set to decline with time between the maximum and minimum values. Similarly the neighborhood radius is also a time decay function that encloses all neurons initially and only winner in the final stage (IDRISI Andes help contents).

CHAPTER V

RESULTS AND DISCUSSIONS

To have a preliminary understanding of the dependency of fire size with respect to these variables, all the variables are plotted with respect to the fuel size. The results of which are discussed below for each variable.

5.1 City interaction

The values obtained from city interaction calculation, were normalized, scaled from 0 to 100 method and finally converted to integer grids (Gilreath, 2006) for graphical analysis. City interaction might be expressed as the number of people interacting per unit square distance. The Fire size with respect to city interaction showed a negative skewed distribution. From the graph (Figure 5.1), the distribution shows a non linear trend that could not be fit with a single equation. In general, fire size seems to be low (less than fifty acres) at low population interaction (below 0.3 percent), increases as city interaction increases (from 0.4 to 0.84 percent) and start decreasing from 0.84 percent of city interaction.

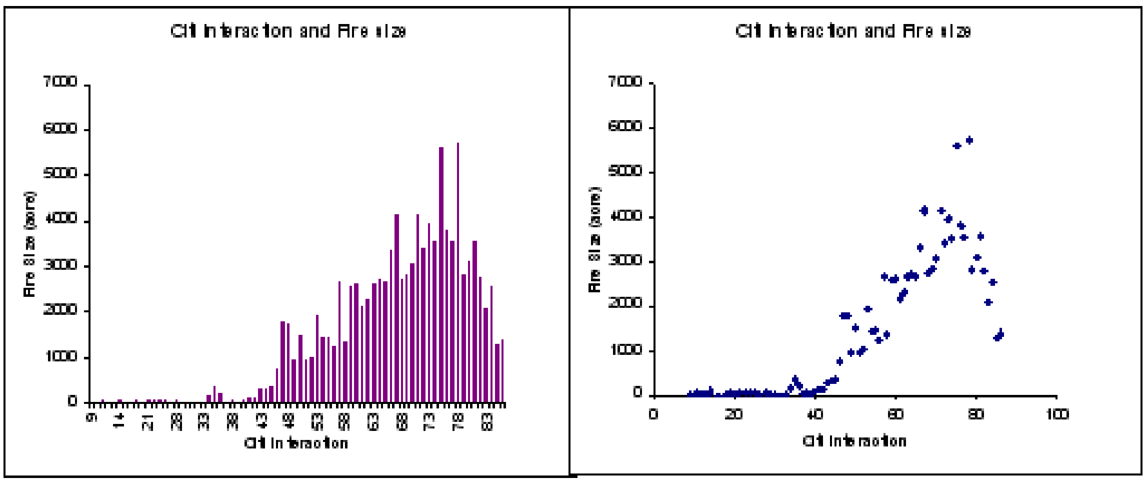

Figure 5.1: Historic fire size occurrences with respect to city interaction

5.2 Fuel density interaction

The density interaction indicates the chance of fire occurrence between fuel stands with respect to their areas. The interaction is directly proportional to the areas of the fuel stands and varies inversely to their distance. This interaction values were used to generate a risk map. The data are normalized by score range method, scaled from 0 to 100 and finally converted to integer grids (Gilreath, 2006). Higher interaction implies fuel stands with large areas and in close proximity to other fuel stands and vice versa. As a preliminary step in order to understand the dependency of historic fires on this fuel density interaction layer, a graph is plotted. The graph (Figure 5.2) showed an approximately normal distribution with peaks and shoulders at high fuel density interaction values that could not be fit with a single equation.

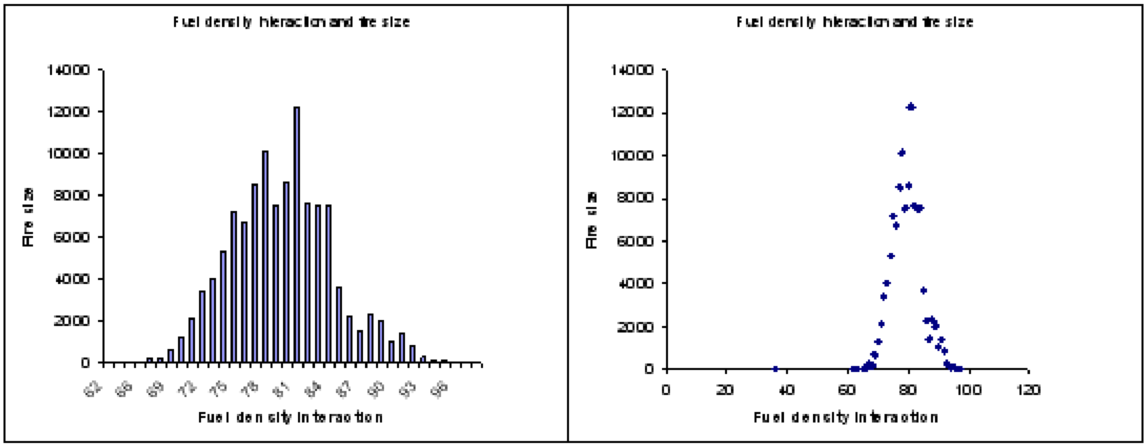

Figure 5.2: Historic fire size occurrences with respect to Fuel density interaction

5.3 Slope

Slope is calculated in degrees. The values of slope for the study area ranged from 0 to 36. The data are normalized by score range method, scaled from 0 to 100 and finally converted to integer grids (Gilreath, 2006). The values of slope with respect to historic fire data are extracted using spatial analysis tools extract values to points in Arc Map. Similar to other variables, historic fire size data is plotted with respect to the slope values (Figure 5.3). The graph too showed different trends at various regions (high fire sizes even at flat slopes indicating grass fires followed by peak fire sizes at a slope of 18 degrees and later on decreases) that could not be best fit with a simple equation.

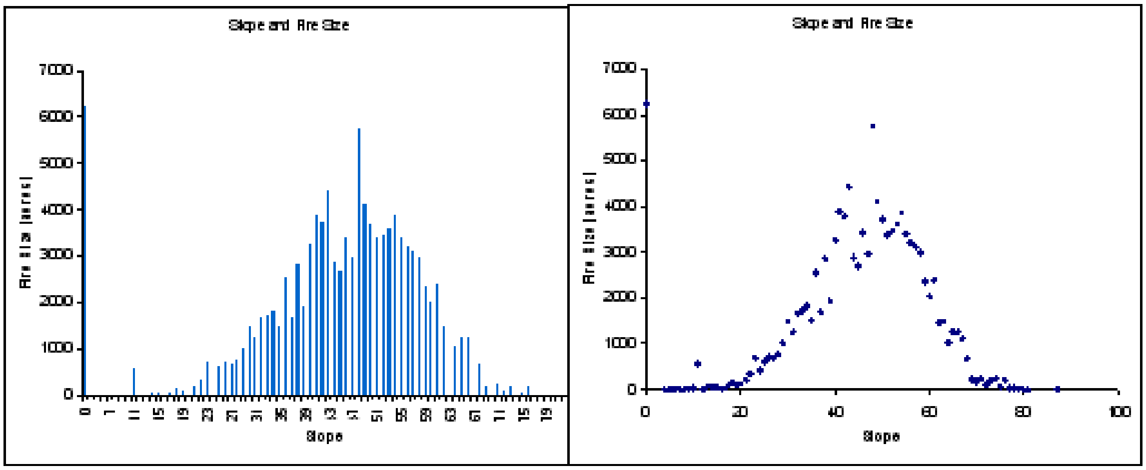

Figure 5.3: Historic fire size occurrences with respect to slope

5.4 Euclidean distance to streams

The Euclidean distance values of the historic fires to the perennial stream feature ranged from 0 to 7674 meter. Similar to the other three variables the data are normalized by score range method, scaled from 0 to 100 and finally converted to integer grids (Gilreath 2006). The euclidean distance values with respect to historic fire data are extracted using spatial analysis tools extract values to points in Arc Map. A graph (Figure 5.4) of the historic fire size occurrence data with respect to the Euclidean distance of the perennial streams showed a positively skewed distribution that could not be fit with a single equation. In general the graph showed a drastic decrease in fire sizes beyond a distance of 2302 m from perennial streams.

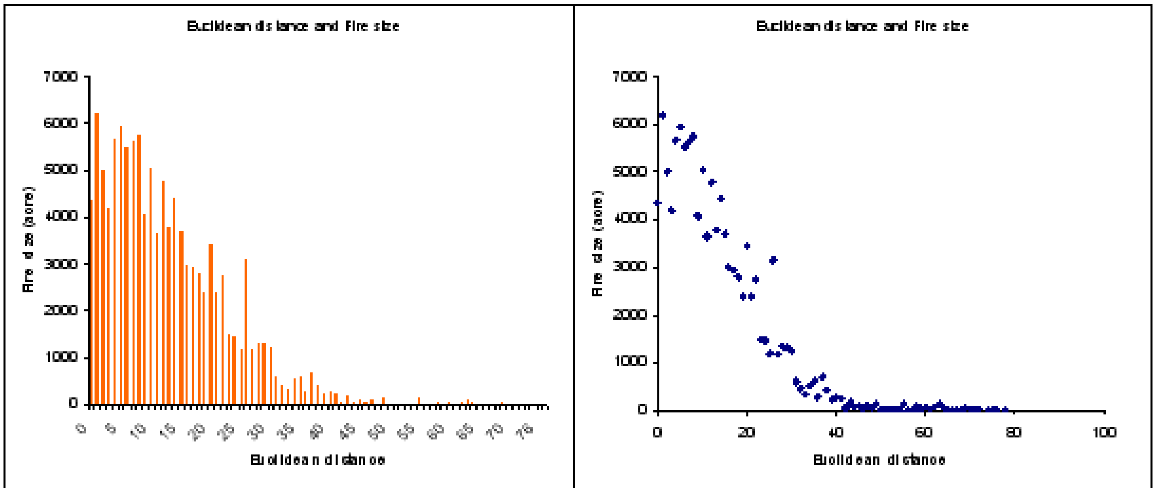

Figure 5.4: Historic fire size occurrences with respect to Euclidean distance to perennial streams

5.5 Artificial neural networks

The dependency of the historic fire occurrences to these four variables (city interaction, Fuel density interaction, slope and distance to perennial streams) could not be properly mapped with simple equations. The graphs revealed a step wise function (separate equations at different parts of the curve) to model the fire risk. At this stage SOM of neural network stream proved to be an alternative to classify fire risk. All the variables were fed to the SOM architecture. Of the several neural network parameters the following network architecture showed the highest accuracy.

Topology: 7507 columns and 10290 rows.

Input layer neurons i.e. the four input variables of city interaction, fuel density Interaction, slope and Euclidean distance to streams.

Output layer neurons: five classes of fire risk i.e. fire size risk of less than 50 acre, 50 to 100 acre, 100-150 acre, 150 to 200 acre and greater than 200 acre.

Various network parameters such as learning rate, gain term, neighborhood radius and iterations were involved in using the SOM approach as such they are explained here under.

A 1 by 3 interval was chosen that will sample all pixels in the input variables. (A 1 by 1 interval will sample all pixels in the input images and increase the time for calculation. Conversely, increasing the interval will decrease the time for calculation- IDRISI help contents).

Learning rate: It is a time decay function that is set in the range of 0.5 to1

These parameters reduced the quantization error to 0.0005 and fine tuning was utilized to obtain higher accuracy. Linear vector quantization is used for fine tuning with 2000 iterations and a gain term of 0.0001 to 0.0005. In coarse tuning the neighborhood radius is at large initially while it was decreased to 1 in fine tuning. A minimum mean distance algorithm is used to classify the unknown pixels in this research. All the above network parameters resulted in a final fire risk map that is described below.

Final map: The SOM process yielded the final map as shown in figure 5.5. The final map along the training data used for classifying fire size risk is shown in figure 5.6. Blue colour represents a fire size risk of less than 50 acres while red colour represents a fire size risk of greater than 200 acre. In terms of the associated four variables from the map it is evident that fire size risk of less than 50 acre are likely the places where few people interact with fuel sizes less than 50 acre that are very nearer to the water bodies on almost gentle slope. Similarly fire size risks of greater than 200 acre are likely the places that find the best in relation to four associated variables and are generally the places with more number of people interacting with fuel sizes greater than 200 acre on medium slopes (13-18 degrees) that lie within 100 to 4000 meter from the water bodies.

Training data: A total of 250 training points are used for fire size risk analysis. Five classes (Table 5.1) of fire sizes were used for training with SOM of ANN. Historic fire occurrence data was separated in to these fire size classes and training points were selected randomly from these sizes with in the study area (Figure 5.5 & 5.6).

Accuracy assessment: The SOM approach uses in general 50% of the training data for training and the other 50% for testing. The result of the IDRISI SOM classification also yielded an error matrix and Kappa coefficient to analyze the fire prediction accuracy for the variables considered in the classification. The error matrix thus obtained is a symmetrical matrix that represents fire size risk prediction by SOM along the rows and fire occurrence or training data along the columns (Table 5.1).

Table 5.1 Error matrix analysis obtained for SOM fire size risk analysis

Fire size risk	Less than 50 acre	50-100 acre	100-150 acre	150-200 acre	Greater than 200 acre	Total	Errors of Comission
Less than 50 acre	36	3	8	3	6	56	0.3571
50-100 acre	5	33	0	3	1	42	0.2143
100-150 acre	6	9	29	2	2	48	0.3958
150-200 acre	3	4	9	22	1	39	0.4359
Greater than 200 acre	5	6	9	11	34	65	0.4769
Total	55	55	55	41	44	250	
Errors of Omission	0.3455	0.4000	0.4727	0.4634	0.2273		0.3840

This errors matrix was used to obtain the accuracies (Table 5.2) using the procedure stated by Filippi and Jensen (2005). In the present situation, producer's accuracy refers to the chance that the used fire occurrence training point correctly predicts fire size risk while the user's accuracy determines the likelihood that a particular predicted fire size risk class represents the same category on the ground (Prabhu, 2006).

Table 5.2 Accuracy obtained for classified fire size risk of SE MS

Fire size risk	Less than 50 acre	50-100 acre	100-150 acre	150-200 acre	Greater than 200 acre
Accuracy (%) Producer's	65.45	60.00	52.73	53.66	77.27
User's	64.29	78.57	60.42	56.41	52.31

The kappa coefficient is a second measure that is used to determine the predicted fire size risk map accuracy. The overall kappa coefficient of 0.5201 is obtained for the present classification of fire size risk. The kappa index of agreement for individual fire size classes (Table 5.3) varied from 0.4 to 0.6 that considered fair and moderate (Simpson, 2008).

Table 5.3 Kappa index of agreement derived from SOM for fire size risk classification

Fire size class	Kappa Index of Agreement
Less than 50 acre	0.5548
50-100 acre	0.5192
100-150 acre	0.4149
150-200 acre	0.4509
Greater than 200 acre	0.6929

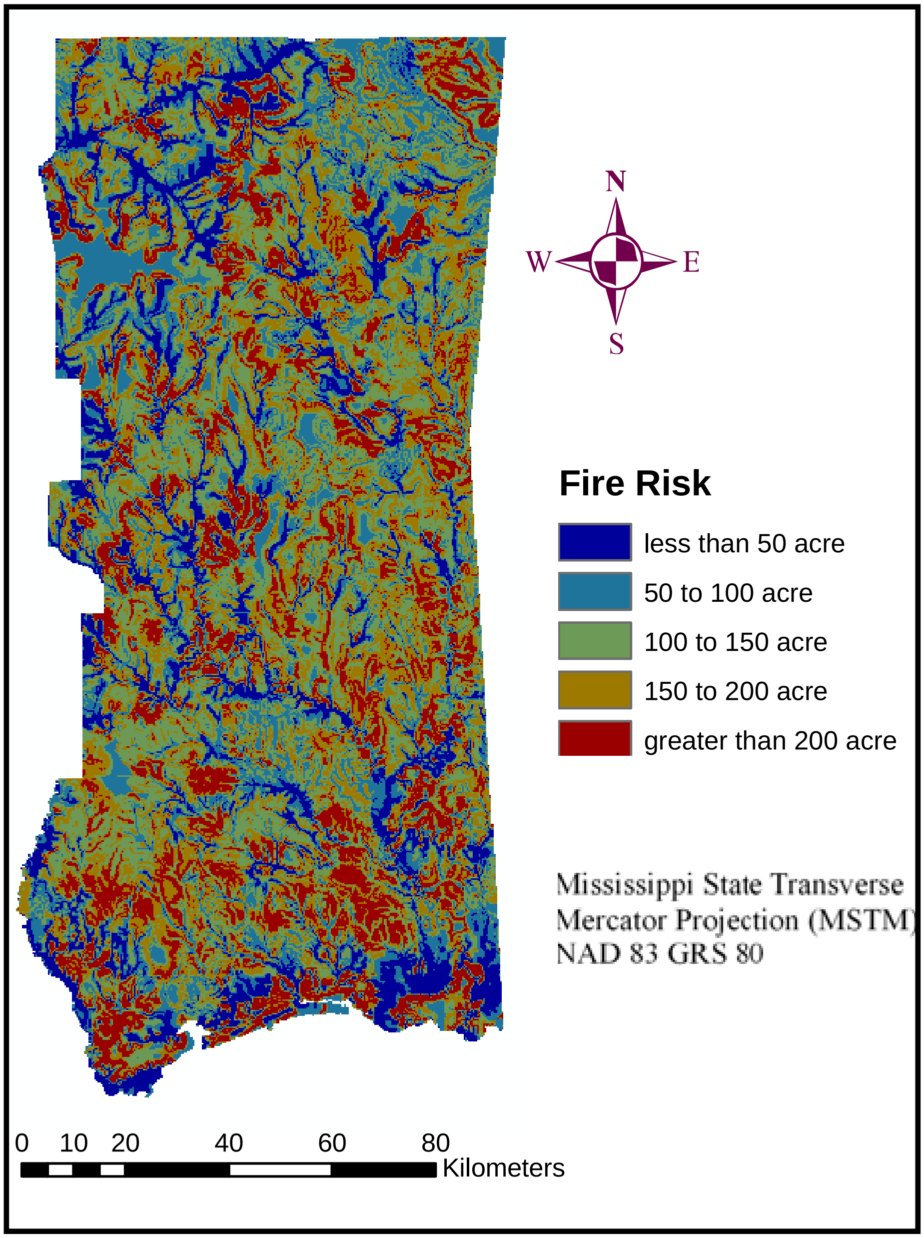

Figure 5.5: ANN SOM final map indicating classified fire risk for SE MS

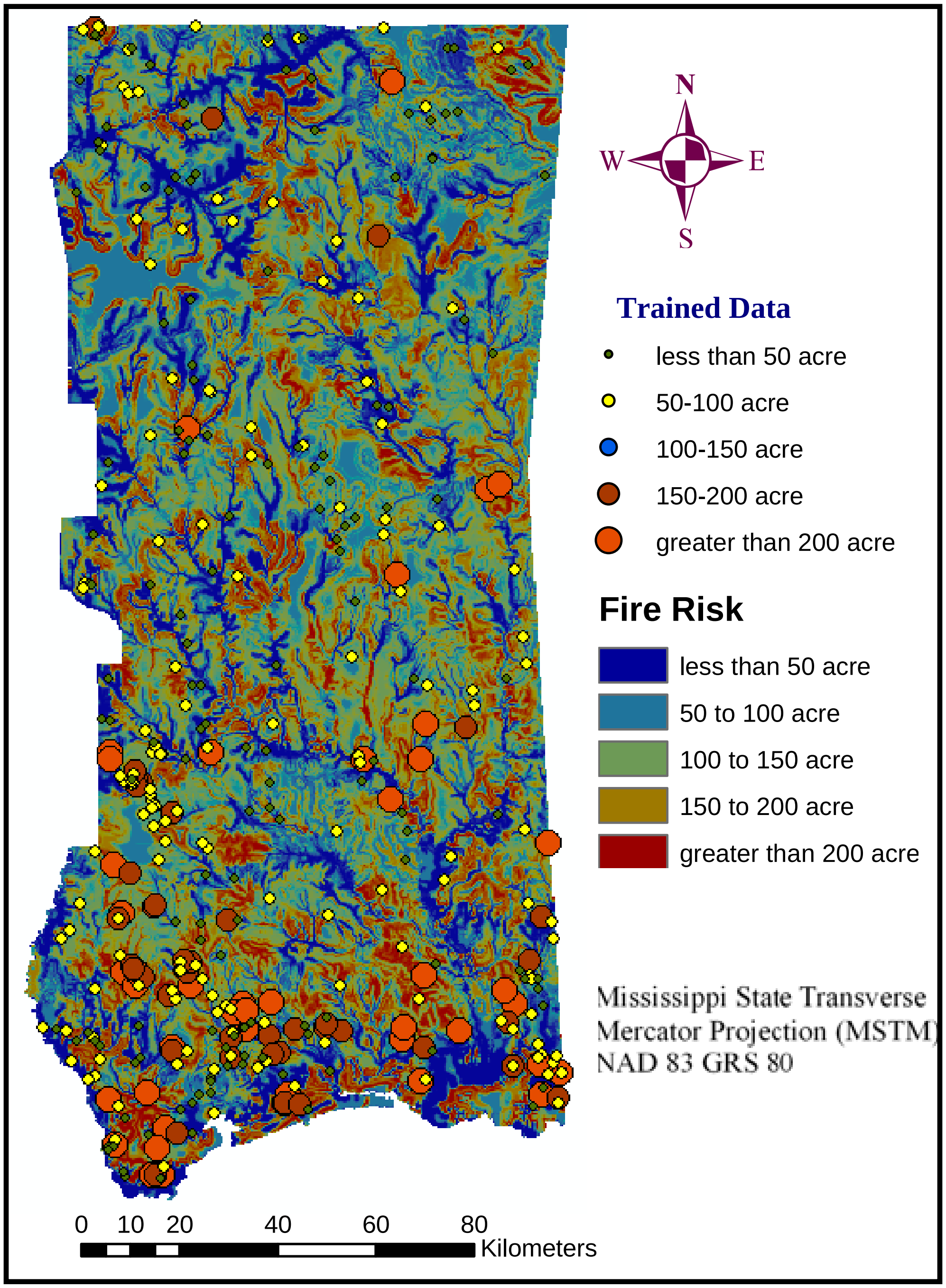

Figure 5.6: ANN SOM Classified fire risk for SE MS along trained data

Discussion

ANN's form a part of Artificial intelligence group. Artificial intelligence techniques are being employed in risk analysis especially in the areas of ecological modeling due to their ability to deal with uncertainty, vagueness, incomplete and inexact specifications, intuition, and qualitative information (Schmoldt D.L., 2001). As a part of fire risk modeling this research figured a non linear relation ship that existed among the four variables in relation to fire size. To deal with such complex ecological processes the inductive nature of ANN is found to be ideal to model fire patterns.

The analysis of the fire size occurrences and the four variables was accomplished using GIS and IDRISI softwares. Arc GIS was used to derive the four variables and IDRISI Andes version was used to implement SOM part of ANN. Initially it has taken time to optimize network parameters and the usable data interchange formats. ASCII format found to be the best data interchange format from GIS to IDRISI Andes and vice versa.

Once all the data is in usable format, one can try with various network parameters to increase the accuracy of classification.

This research is peculiar in comparison to the earlier fire researches by Gilreath (2006) and Raviraj (2007). While Gilreath focused mainly on road density, Raviraj utilized additive model on gravity based human interaction and fuels. This research considered a group of variables in relation to fire occurrences and also classified fire sizes. This can be extended to any number of variables provided all have the same topology and for further research recommends the use of more training data over longer time periods (eg., 15 years…fires & climate) to develop seasonal models that are of greater interest to fire managers.

CHAPTER VI

SUMMARY AND CONCLUSIONS

Artificial neural networks proved to be a best alternative in situations where the data distribution is irregular that do not follow a particular regular pattern and required stepwise functions at various stages of the distribution. SOM proved to be fascinating in the area of neural networks for classification purposes. Learning is also efficient, effective, and suitable for classification purposes. IDRISI SOM module aided the integration of Arcmap data to perform the analysis. The module result also yielded an error matrix and kappa coefficient as the measures of classification accuracy.

Error matrix produced an overall classification accuracy of 63 % and overall kappa coefficient value of 0.5201. These values lie in the range of fair or moderate as per the kappa values interpretation outlined by Simpson, (2008) (Table 6.1a & b).

Table 6.1 (a & b) Kappa coefficient interpretations (Source: Simpson 2008)

(a)

Category	Kappa value
Poor agreement	Less than 0.20
Fair agreement	0.20 to 0.40
Moderate agreement	0.40 to 0.60
Good agreement	0.60 to 0.80
Very good agreement	0.80 to 1.00

(b)

Category	Kappa value
Poor	Less than 0 .40
Fair	0.40- 0.59
Good	0.60- 0.74
Excellent	Greater than 0.74

One of the limiting factors found in this research is the historic fire data locations used for training. Only few Training data points are available for classes of 150 to 200 acre and greater than 200 acre fire risk classes. Training data is also limiting in terms of fire type and age in obtaining higher accuracies of classified fire risk. There is no clear cut demarcation of the forest type and in what age class the fires occurred while the research considered only pine forest type with 20-30 years age group to calculate fuel density interaction. That might be one aspect that has hindered the improvement of classification accuracy.

In terms of input variables of ANN, linear trend existed in all the variables only to a certain part of the data set as is evident from the graphs (Figure 5.1, 5.2, 5.3 & 5.4). A common observation from the graphs was that the fire size risk is low at low city interaction as well as at low fuel density interaction. Fire size risk is also low with respect to low Euclidean distance and majority of the fire size is accumulated with in the first 40 percent of the perennial stream distance. With respect to slope, other than a fire size peak on flat terrain that could probably be bush land fires located in the study area it showed a similar trend of low fire size in lesser slopes.

A general observation associated with the four variables was that, fire size risk was low at low variable values, increased to a certain extent and started declining. This particular irregular trend that existed in the associated variables for forest fire size risk was very well predicted by the present SOM approach. All the results can be summarized and concluded as follows:

- High fire sizes were associated with 13 to 21 degrees of slope for the study area with peak fire sizes occurring around 18 degrees.
- A significant drop in fire sizes occurs beyond 2302 meter distance from perennial streams.
- Neural networks are good for modeling complex ecological processes using variables that are not linear predictors
- The inductive nature of ANN is ideal for modeling fire patterns that are not easily described by deterministic processes
- The accuracy is moderate (fair) 62% for annual fires

- Accuracies are likely to improve for seasonal fire prediction
- Implementation is easy in the GIS analytical environment
- Implementation for fire management agencies will require analyst training and development of a user-friendly interface
- SOM is good for classification purposes

Similarly, the limitations associated with the present research can be summarized as follows:

- Topology is a constraint
- Historic fire data – no clear cut for fire type and age
- Lack of adequate amounts of training data made seasonal modeling difficult
- Needs time for training and optimizing the network parameters
- Low explanatory depth
- Neural networks are sensitive to local conditions and training data and extrapolation to different geographic areas and landscape conditions is often a drawback to wide use.

REFERENCES

Andrews, P.L. (1986). BEHAVE: Fire behavior prediction and fuel modeling system-BURN subsystem Part 1. USDA For. Serv. Gen. Tech. Rep. INT-194. 130 pp.

Ankica Grant. (2007). Fighting Fire with Fire: Investigating Prescribed Burns for Fuel and Fire Management in Quetico Provincial Park, Ontario. Thesis. University of Waterloo, Ontario, Canada.

Anthony Gar-On Yeh and Xia Li. (2002). Urban simulation using nural networks and cellular automata for land use planning. Symposium on Geospatial theory, processing and applications, Ottawa, 2002.

Arc GIS 9.2 desk top help Environmental Sciences Research Institute (ESRI).

Arora,M.K., Gupta,A.S.D., Gupta,R.P. (2004). An artificial neural network approach for landslide hazard zonation in the Bhagirathi (Ganga) Valley, Himalayas. International Journal of Remote Sensing, 2004 (Vol. 25) (No. 3) 559-572.

Atkinson, P. M. and Tatnall, A. R. L. (1997). Neural networks in remote sensing, *International Journal of Remote Sensing*, **18,** 699-709.

Bacao, F., Lobo, V., Painho, M. (2005), "Self-organizing Maps as Substitutes for K-Means Clustering". In: V.S. Sunderam, G. van Albada, P. Sloot, J. J. Dongarra (Eds.), International Conference on Computational Science 2005.

Baird, W., Robinson Jr. J.W., and Jones Jr. A.R.,(1969). Beliefs and practices of selected rural residents toward forest consideration and management. Social Science Research Center, Mississippi State University, Mississippi. 40 P.

Basheer, I.A. and Hajmeer, M.N. (2000). Artificial neural networks: Fundamentals, computing, design, and application. J. Microbiol. Meth**.** 43(1)**,** 3–31.

Bergeron, Y. (1991).The influence of island and mainland lakeshore landscapes on boreal forest fire regimes. Ecology, Vol. 72, No. 6., pp. 1980-1992.

Brandt Tso & Mather P.M., Classification methods for remotely sensed data. Taylor and Francis, New York.

Brenner, J.D. (2002). Fire Risk Assessment System User Guide. Division of Forestry, State of Florida, Tallahassee, FL.

Brown, S.L., Schroeder, P., and Kern J.S. (1999). Spatial distribution of biomass in forests of eastern USA, Forest ecology and Management, vol.123, No.1, pp. 81-90.

Burgan, R.E., Klaver, R.W., Klaver, J.M. (1998). Fuel Models and Fire Potential from Satellite and Surface Observations. *International Journal of Wildland Fire*, 8(3), pp.159-170

Campbell, J.B. (1987). Introduction to Remote Sensing, 2nd ed. (New York: The Guildford Press).

Cooke, W.H., Grala, K., Evans, D.L., Collins, C. (2007). Assessment of pre and post Katrina fuel conditions as a component of fire potential modeling for southern Mississippi. Journal of forestry, vol. 105, no 8 pp. 389-397.

Chang, K. (2006). Introduction to Geographic information systems. Third edition. McGraw-Hill companies, Inc. Newyork.

Chou, Y.H., Minnich, R.A., Salazar, L.A., Power, J.D., and Dezzani, R.J. (1990). Spatial autocorrelation of wildfire distribution in the Idyllwild Quadrangle, San Jacinto Mountain, California. Photogrammetric Engineering and Remote Sensing. 56(1): 1507-1513.

Chuvieco, E., Cocero, D., Riano, D., Martin, P., Martinez-Vega, J., De la Riva, J.,Perez, F. (2004). Combining NDVI and surface temperature for the estimation of live fuel moisture content in forest fire danger rating. *Remote Sensing of the Environment*, 92, pp. 322-331.

Chuvieco, E., Congalton, R.G. (1989). "Application of remote sensing and geographic information system to forest fire hazard mapping", *Journal of Remote Sensing Environment*, Vol. 29 pp.147-59.

Collins, C.A., Wilkerson, D.W., Evans, D.L. (2005). Multi-temporal Analysis of Landsat Data to Determine Forest Age Classes for the Mississippi Statewide Forest Inventory. Analysis of Multitemporal Remote Sensing Images, nternational Workshop, Institute of Electrical and Electronics Engineers. May 16-18, Biloxi, MS. CD-ROM.

Cope, M.J. and Chaloner, W.G. (1985). Wildfire: an interaction of biological and physical processes. In Tiffney, B.H., editor, Geological factors and the evolution of plants, New Haven, CT: Yale University Press, 257-77.

Countryman, C. (1972). The Fire Environment Concept. Pacific Southwest Forest and Ranger Experiment Station, U.S. Forest Service. Berkeley, CA.

Cyr, D., Gauthier, S., and Bergeron, Y. (2007). Scale dependent determinants of heterogeneity in fire frequency in a coniferous boreal forest of eastern Canada. Landscape Ecol 22:1325–1339.

Daniel J. Grenier, D. J., Bergeron, Y, Kneeshaw, D and Gauthier, S. (2005). Fire frequency for the transitional mixedwood forest of Timiskaming, Quebec, Canada Can. J. For. Res. 35: 656–666.

Dasgupta, S. Qu, J.J. and Hao, X. (2006). Design of a susceptibility index for fire risk monitoring. IEEE Geoscience and remote sensing letters, VOL. 3, NO. 1, JANUARY 2006.

Deeming, J.E.; Burgan, R.E.; Cohen, J.D. (1977). The National Fire-Danger Rating System-1978. United States Department of Agriculture, Forest Service, General Technical Report INT-39, Intermountain Forest and Range Experiment Station, Ogden, Utah. 66 pages.

ERDAS Field Guide. Leica Geosystems GIS and Mapping, LLC. (2003).
(p. 672). (7th ed.) Atlanta, GA: Leica Geosystems GIS and Mapping, LLC.
Erten, Esra., Kurgun, V., Musaoglu N., (2004). Forest Fire Risk Zone Mapping From Satellite Imagery and GIS a Case Study. XX^th ISPRS Congress, Istanbul, Turkey.

FIA Glossory (http://www.pwrc.usgs.gov/brd/DefFiaGlossary.htm).

Filippi, A.M., Jensen, J.R. (2006). Fuzzy learning vector quantization for hyperspectral coastal vegetation classification. Remot. Sens. Environ. 100, 512–530.

Fischer, M.M. (2000). Spatial interaction models and the role of geographic information systems. In Wegener, M. and Fotheringham, S. (eds.): Spatial Models and GIS: New Potential and New Models. Taylor & Francis, London, pp. 33-43 [ISBN 0-748-40846-0].

Franzini, L, Paola Bolchi, and Lidia Diappi. (2001). Self Organizing Maps: A Clustering Neural Method for Urban Analysis, V^e Rencontres de Théo Quant.

Gilreath. (2006). Validation of variables for the creation of a descriptive fire potential model for the southeastern fire district of Mississippi. Thesis. Mississippi state university.

Greenberg, Josha D. and Gordaon A. Bradley. (1997). Analyzing the urban-wildland interface with GIS: two case studies. *J. of Forestry* 95(10):18-22.

Hara, Y., Atkins, R., Yueh, S., Shin, R., and King, J. (1994). Application of neural networks to radar image classification. IEEE transactions on Geoscience and Remote Sensing, 32, 100–109.

Haykin, S. (1994). Neural Networks: A Comprehensive Foundation. New York: Macmillan College Publishing.

Hepner, G.F., Logan, T., Ritter, N., and Bryant, N. (1990). Artificial neural network classification using a minimal training set: Comparison to conventional supervised classification. Photogramm. Eng. Rem. Sens.56:469-473.

Hilbert, D.W., and van den Muyzenberg, J. (1999). Using an artificial neural network to characterise the relative suitability of environments for forest types in a complex tropical vegetation mosaic; *Diversity Distrib.* **5** 263–274.

Hildebrand, D. (2003). Lightning tracking an important tool in land management and wildland fire suppression. Vaisala News 162:32-33. Available online at: http://www.vaisala.com/DynaGen_Attachments/Att27112/VN162_p32_33.pdf.pdf.

Hoffman, M. (2005). Numerical control of kohonen neural network for scattered data approximation. Numerical Algorithms (2005) **39:** 175–186.

IDRISI Andes Tutorial. (2006). By Eastman R.J Clark Labs, Clark University (15 th version).

Iwan Setiawan, Mahmud, A.R., Mansor, S., Mohamed Shariff, A.R., and Naruddin, A.A. (2004). GIS-grid-based and multi-criteria analysis for identifying and mapping peat swamp forest fire hazard in Pahang, Malaysia. Disaster prevention and Management. Vol 13: No. 5: 379-386.

Jaiswal, R.K., Mukherjee, S., Raju, D.K., Saxena, R. (2002). Forest fire risk zone mapping from satellite imagery and GIS.*International Journal of Applied Earth Observation and Geoinformation*, 4 (2002), pp.1-10.

Jeffrey, A. Cardille., Stephen J. Ventura, and Monica G. Turner. (2001). Environmental and socialfactors influencing wildfires in the upper Midwest, United States. Ecological applications 11(1). 2001pp.111-127.

Jiang ,B., and Harrie,L. (2004). Selection of Streets from a Network Using Self-Organizing Maps. *Transactions in GIS* 8(3): 335-350.

Johnson, E.A. and Van Wagner, C.E. (1985). The theory and use of two fire history models. Can. J. Forest Res. 15: 214-220.

Jo, M.H., Lee, M.B., Lee, S.Y., Jo, Y.W., Baek, S.R. (2000). "The development of forest fire forecasting system using internet GIS and satellite remote sensing", available at: www.gisdevelopment.net/aars/acrs/2000/ps3/ps310.shtml.

Kalabokidis, K., Karavitis, C., Vasilakos, C. (2004). Automated fire and flood danger assessment system. In Xanthopoulos, G. (ed.), Proceedings of the International Workshop on Forest Fires in the Wildland- Urban Interface and Rural Areas in Europe: an integral planning and management challenge, Athens, Greece. Published by MAICh, Crete, Greece: 143-153.

Kavzoglu, T., and Mather, P. M. (2003). The use of backpropagating artificial neural networks in land cover classification. International Journal of Remote Sensing, 24, 4907– 4938.

Kemp, E.M. (1981). Pre-Quaternary fire in Australia. In Gill, A.M., Groves, R.H. and Noble, I.R., editors, Fire and the Australian biota, Canberra: Australian Academy of Science, 3-21.

Kohonen, T. (1989). Self-Organization and Associative Memory (New York: Springer-Verlag).

Kohonen, T. (2001). Self-Organizing Maps. Berlin-Heidelberg, Springer.

Koutsias, N., and Karteris, M. (2003). Classification Analyses of Vegetation for Delineating Forest Fire Fuel Complexes in a Mediterranean test site Using Satellite Remote Sensing and GIS. *International Journal of Remote Sensing*. 24(15). pp. 3093-3104.

Langhart, R., Bachmann, A., Allgower, B. (1998). Spatial and temporal patterns of wildfire occurrence (Canton of Gnscm, Switzerland). Proc. I11 Int. Conf. on *Forest Fire Research,* Luso Coimbra, Portugal, 16. -- 20. November **2:** 2279 - 2292.

Lavin, Mary Jo. (1997). Managing fire risk to people, structures, and the environment. Fire Management Notes 57(4): 4-6.

Malczewski, J. (1999). GIS and multicriteria decision analysis. John Wiley and Sons. Inc. USA.

McCormick, R.J., Thomas A. Brandner, and Timothy F. H. Allen. (2004). Towards a theory of meso-scale wildfire modeling -- a complex systems approach using artificial neural networks. Technical report, University of Wisconsin-Madison.

Miller, J.D., Yool, S.R. (2002). Mapping forest post-fire canopy consumption in several overstorey types using multi-temporal LANDSAT TM and ETM data. *Remote Sensing & Environment.* 82. pp. 481-496.

Mississippi Forestry Commission. (2004). *Information about theMississippi Forestry Commission, Forestry Facts: Number of Wildfires.* Jackson, Mississippi.

National Interagency Fire Center.(2005). Wildland Fire Statistics. http://www.nifc.gov/stats/wildlandfire.html. Accessed April, 20, 2006. National Interagency Fire Center, Boise, Idaho.

Net state.com (http://www.netstate.com/states/geography/ms_geography.htm). Accessed on Jan 16, 2008.

Paola, J.D., and Schowengerdt, R.A. (1995). A review and analysis of back propagation neural networks for classification of remotely sensed multi-spectral imagery. Int J Rem Sens 16(16):3033–3058.

Peng, C., and Wen, X. (1999). Recent applications of artificial neural networks in forest resource management: an overview. P. 15-22 *in* Environmental Decision Support Systems and Artificial Intelligence, Corte, U., and M. Sanchez-Marre (eds.). Tech. Rep. WS- 99-07, AAAI Press, Menlo Park, California.

Perry, G.L.W. (1998). Current Approaches to Modelling the Spread of Wildland Fire: a Review.*Progress in Physical Geography*. 22:222-245.

Pijanowski, B.C., Shellito, B., Pithadia, S., Alexandridis, K. (2002). Forecasting and assessing the impact of urban sprawl in coastal watersheds along eastern Lake Michigan. Lakes and Reservoirs: Research and Management 7, 271–285.

Pijanowsky, B. C., Shellito, B. A., Bauer, M. E., and Sawaya, K. E. (2001). Using GIS, Artificial Neural Networks and Remote Sensing to Model Urban Change in the Minneapolis-St.Paul and Detroit Metropolitan Areas. Proceedings, ASPRS Annual Conference, St. Louis, MO.

Prabhu, C.L. (2006). Comparison of the utility of regression analysis and K- nearest neighbor technique to estimate above ground biomass in pine forests using Land sat ETM+ imagery. Thesis. Mississippi state university.

Pye, J.M., Prestemon, J.P., Butry, D.T. (2003). Prescribed Burning and Wildfire Risk in the 1998 Fire Season in Florida. Proceedings, RMRS-P-29, Unites States Department of Agriculture, United States Forest Service, Southern Research Station, Research Triangle Park, NC. pp. 15-29.

Raviraj. (2007). Gravity model to detect forest fire prone areas for the southeastern fire district of Mississippi. Thesis. Mississippi state university.

Reginald Mead. (2006). A Review of AI Techniques for Ecological Modeling. January 16, 2006. Available online at (http://labs.rightnow.com/colloquium/papers/mead-modelling_lit_review.pdf). Accessed on May 6, 2008.

Roy, P.S. and Ravan, S.A. (1996). Biomass estimation using satellite remote sensing data—an investigation on possible approaches for natural forest. Journal of Bioscience, 21, pp. 535–561.

Rudis, V.A. and Skinner, T.V. (1991). Fire's importance in South Central U.S. forests: Distribution of the fire evidence. In: Nodvin, S.C and T.A Waldrop (eds). Fire and the environment: Ecological and cultural perspectives. Proceedings of An International Symposium. March 20-24, 1990. Knoxville, TN. General Technical Report. SE-69. Southern Forest Experiment Station, USDA Forest Service, Asheville, NC.p.240-251.

Rigol Sanchez, J.P., Chica-olmo, M., and Abarca-Hernandez, F. (2003). Artificial neural networks as a tool for mineral potential mapping with GIS. Int. J. Remote Sensing 24: No. 5 pp. 1151-1156.

Schmoldt, D.L. (2001). Application of artificial intelligence to risk analysis for forested ecosystems. Risk Analysis in Forest Management, 49–74.

Schultz, R.P. (1997). Loblolly Pine: The Ecology and Culture of Loblolly Needleleaf evergreen: The Ecology and Culture of Loblolly Needle-leaf evergreen (*Pinus taeda* L.).

Silva, A.P. (2003). Neural networks application to spatial interpolation of climate variables. STSM on the Framework of COST 719 from 6-10 October 2003 at Vienna – Austria.

Simpson, S. (2008). Stats: What is a Kappa coefficient? (Cohen's Kappa). Online at (http://www.childrens-mercy.org/stats/definitions/kappa.htm). Accessed on 16 January 2008.

Skapura, D. (1996). *Building Neural Networks.* ACM Press, New York.

Sunar, F., Ozkan, C. (2001). "Forest fire analysis with remote sensing data" International Journal of Remote Sensing 22: No. 12 pp. 2265-2277.

Suwardi Annas, Takenori Kanai and Shuhei Koyama. (2007). Principal Component Analysis and Self-Organizing Map for Visualizing and Classifying Fire Risks in Forest Regions. Agricultural Information Research, Vol. 16 (2007), No. 2 pp.44-51.

Tanskanen, H., Venalainen, A., Puttonen, P., and Granstrom, A. (2005). Impact of stand structure on surface fire ignition potential in Picea abies and Pinus sylvestris forests in southern Finland. Canadian Journal Of Forest Research. 35 (2): 410-420.

Terry, J. (1997). Prevention: An emerging service, Fire Management Notes.57 (3):4-6.

Tran, L.T., Knight, C.G., O'Neill, R.V., Smith, E.R., O'Connell, M. (2003). "Self-organizing maps for integrated environmental assessment of the U.S. Mid-Atlantic region" Environ. Manage. 31: 822-835.

Veblen, T.T., Kitzberger, T., Villalba, R., and Donnegan, J. (1999). Fire history in northern Patagonia: The roles of humans and climatic variation. Ecological Monographs, 69(1):47-67.

Vliegher, B.Md. (1993). Risk assessment for environmental degradion caused by fires using remote sensing and GIS in a Mediterranean Region (South-Euboia, Central Greece). IGARSS'92 Int. Geosciences and Remote Sensing Symposium, Houston, pp 44-47.

Wade,D.D.(1988).United States. Southern Region, US Forest Service, USDepartment of Agriculture, US Department of the Interior. A Guide for Prescribed Burning in the South. Technical Publication R8-TP 11, US Forest Service.

Wang, S and Wang, H. (2002). Knowledge discovery through self organizing maps: data visualization and query processing. Knowledge and information Systems 4: 31-45.

Wear, D. N., and Greis, J. G. (2002). The southern forest resources assessment: summary report. Gen. Tech. Rep. SRS-54. Asheville, NC: USDA, Forest Service, Southern Research Station.

Weijian Wan., and Donald Fraser. (1999). A self-organising neural network framework for multisource data and contextual analysis. In *Proc. 7th Australasian Remote Sensing Conference, Melborne, Australia*, pages 145-150. Remote Sensing and Photogrammetry Association Australia, Ltd, 1994.

Wilson, C. C. and Davis, J. B. (1988). Forest fire laboratory at Riverside and fire research in California: past, present, and future. Source: Forest Service, U.S. Department of Agriculture 22 p. online accessed on Jan 15, 2008. (http://www.fs.fed.us/psw/publications/documents/psw_gtr105/psw_gtr105.pdf).

Whelan, R.J. (1995). The ecology of fire. Cambridge: Cambridge University Press.

Wright, D. B. (2004). A feasibility study for real time detection and georeferencing of forest fire hot spots. Thesis. University of Calgary, Calgary, Alberta.

Zhai, Y., Munn, I.A., Evans, D.L. (2003). Modeling Forest Fire Probabilities in the South Central United States using FIA data. South. J.*Appl. For*. 27: 11-17.