Spatio-Temporal Modeling of Wildfire Risks

in the U.S. Forest Sector

Xuan Chen

Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695, e-mail: xchen8@ncsu.edu

Barry K. Goodwin
William Neal Reynolds Distinguished Professor,
Department of Agricultural and Resource Economics, North Carolina State University, Raleigh, NC 27695

Selected Paper prepared for presentation at the Agricultural & Applied Economics Association's 2011 AAEA & NAREA Joint Annual Meeting, Pittsburgh, Pennsylvania, July 24-26, 2011

Abstract

In the US forestry industry, wildfire has always been one of the leading causes of damage. This topic is of growing interest as wildfire has caused huge losses for landowners, residents and governments in recent years. While individual wildfire behavior is well studied (e.g. Butry 2009; Holmes 2010), a lot of new literature on broadscale wildfire risks (e.g. by county) is emerging (e.g. Butry et al. 2001; Prestemon et al. 2002). The papers of the latter category have provided useful suggestions for government wildfire management and policies. Although wildfire insurance for real estate owners is popular, the possibility to develop a forestry production insurance scheme accounting for wildfire risks has not yet been investigated. The purpose of our paper is to comprehensively evaluate broadscale wildfire risks in a spatio-temporal autoregressive scenario and to design an actuarially fair wildfire insurance scheme in the U.S. forest sector. Our research builds upon an extensive literature that has investigated crop insurance modeling. Wildfire risks are closely linked to environmental conditions. Weather, forestland size, aspects of human activity have been proved to be crucial causal factors for wildfire (Prestemon et al. 2002; Prestemon and Butry 2005; Mercer et al. 2007). In light of these factors, we carefully study wildfires ignited by different sources, such as by arson and lightning, and identify their underlying causes. We find that the decomposition of forestland ecosystem and socio-economic conditions have significant impacts on wildfire, as well as weather. Our models provide a good fit to data on frequency and propensity for fires to exist (e.g. R-square ranges from 0.4 to 0.8) and therefore provide important fundamental information on risks for the development of insurance contracts. A number of databases relevant to this topic are used. With the Florida wildfire frequency and loss size database, a complete survey of four measurements of annual wildfire risks is implemented. These four measurements are annual wildfire frequency, burned area, fire per acre and burned ratio at county level. In addition, the national forestry inventory and analysis (FIA) database, Regional Economic Information Systems (REIS) database and the national weather database have supplied forestland ecosystem, socioeconomic, and weather condition information respectively.

With our spatio-temporal lattice models, impacts of environmental factors on wildfire and implications of wildfire management policies are assessed. Forestland size, private owners' share of forestland, population and drought would positively contribute to wildfire risks significantly. Cold weather and high employment are found to be helpful in lessening wildfire risks. Among the forestland ecosystem, oak / pine & oak / hickory forestland would reduce wildfire risks while longleaf / slash & loblolly / shortleaf pine forestland would have a mixed impact. An interesting finding is that oak / gum / cypress forestland would reduce wildfire frequency, but would enhance wildfire propensity at the same time. Hurricanes could intensify wildfire risks in the same year, but would significantly decrease the next year's wildfire risks.Meanwhile, cross sample validation verifies that our method can forecast wildfire risks adequately well. Since our approach does not incorporate any fixed-effect indicator or trend as in the panel data analysis (Prestemon et al. 2002), it offers a universal tool to evaluate and predict wildfire risks. Hence, given environmental information of a location, a corresponding actuarially fair insurance rate can be calculated.

Key word: wildfires, forestry, weather, socio-economic, Spatio-Temporal autocorrelation

Introduction

Forests cover a large land area in the United States. Since the early twentieth century, the forestland area has been stable around 302 million hectares, compromising about 1/3 of the total land of the United States. In a world context, the U.S. makes up about 10\% of the world's total forestland, and its timber production for industrial products accounts for about 1/4 of the world's production (Brad 2004). On average, 3 pounds of forestry industry products are consumed by each U.S. resident every day. This means that every year an amount equivalent to 100-foot tree will be consumed by each American (Bonson).

Although forest and timber industries play an important role in the U.S., they are constantly threatened by wildfire outbreaks. A wildfire is any uncontrolled fire in combustible vegetation that occurs in the countryside or in a wilderness area. The temperature of a wildfire could rise to 2600 degrees Fahrenheit. Since this temperature can melt down iron, properties and trees in its way are destroyed immediately. Also wildfires usually spread rapidly over large areas. The forward blasts could be as wide as 60 feet and flames could rise up to 325 feet and move as fast as 100 miles per hour. This is especially true for violent crown fires - called “firestorms “or “blowups”, that engulf the top of huge trees as they sweep across the landscape (Bronson). These characteristics make it difficult to contain a large wildfire within a small space scale and extinguish it in a short period. The vector of wildfire transmission risks, involving significant weather events and idiosyncratic fuel buildups on the ground, makes containment a major challenge.

In December 2003, Healthy Forest Restoration Act (Act 2003) was signed by President George W. Bush with the aim of protecting land from wildfire disasters. Preceded by the two worst wildfire seasons (2000 and 2002) after World War II, this act intends "*to improve the capacity of the Secretary of Agriculture and the Secretary of the Interior to conduct hazardous fuels reduction projects on National Forest System land and Bureau of Land Management lands aimed at protecting communities, watersheds, and certain other at-risk lands from catastrophic wildfire, to enhance efforts to protect watersheds and address threats to forest and rangeland health, including catastrophic wildfire, across the landscape, and for other purposes*"

---------Healthy Forest Restoration Act, page 108

After the passage of HFRA (Healthy Forest Restoration Act), many more hazardous fuel reduction projects on federal lands have been expedited to protect forest-adjacent-communities from wildfire. This act has proved to be a significant effort in wildfire prevention.

Disaster relief, a form of ad hoc assistance, is usually used to compensate property owners after disastrous wildfires. Some national organizations such as the American Red Cross offer immediate aid to victims after large wildfires. Other local non-profit funds, such as the Georgia Wildfire Relief Fund, provide assistance to affected residents and engage in local ecosystem restoration in a long term. However, affected private timber business owners are always highly dependent on the government disaster relief programs. For example, southern California got attacked by large wildfires during two consecutive years from 2007 to 2008. In 2007, the Internal Revenue Service (IRS) granted tax relief for southern California wildfire victims. After the 2008 wildfire season, both the IRS and the California state government granted tax relief for affected business owners in southern California.

Widely spread disasters such as wildfire pose a significant hazard to timber production and thus warrant consideration of a relevant single-peril forest insurance product. First, such an approach can provide an actuarially fair rate, which may attract insurance companies and forestland owners to engage in a private insurance market. The possibility of removing the government externality in the disaster payment market will likely result in a more efficient market scheme. Second, the potential economic benefits from mitigation and reduction of the further spread of wildfire may be enhanced under such a specific-peril plan.

The first benefit stems from the notion that comprehension of a particular hazard and its spatiotemporal transmission mechanism warrants the development of a class of single-peril insurance products that measure wildfire risks accurately. Given the fact that wildfire risks are usually catastrophic, if actuarially fair rates can be implemented in a single-peril insurance plan, risk-averse forestland owners will purchase such insurance products once insurance companies offer them to the market. Such a private insurance market can ease the destructive losses of forestland owners even in the absence of government interferences. Furthermore, as forest disaster relief is becoming a fast growing burden for governments worldwide (Holecy 2006), developments of private wildfire insurance products can lessen the government financial stress if unexpected ad hoc aids eventually become unnecessary.

The second benefit stems from the notion that understanding spatiotemporal aspects of wildfire risks and recognizing the potential spatial externality can provide benefits to forestland owners, insurance companies, local and state governments and society generally. To fully capture those benefits requires a comprehensive study of spatiotemporal relationships of wildfire risks and observable forest characteristics and environmental factors. In addition, a practical effective insurance policy needs to minimize adverse selection and moral hazard issues and induce incentive-compatible actions by forestland owners to prevent wildfire risks. A fair premium insurance plan also needs to evaluate compliance policies that decrease outbreak chances by reducing hazards in advance. Prescribed burning permits could be an example of efforts made by forestland owners and governments to reduce wildfire risks.

The State of Florida, with a significant forestland portion of its total land area and a history of frequent wildfire outbreaks, presents an ideal case study for modeling forest losses associated with wildfire risks. As many as 16 million acres of forestland cover almost half of Florida's total land area. Ranked among top four tree-planting states, Florida plants over 82 million trees every year, with 5 trees planted for each tree harvested. The forest and forest products industries have an economic impact of $16.5 billion, including 133,000 jobs . At the same time, Florida suffers over 4000 wildfire occurrences on average every year with approximately 200,000 acres of forestland burned. Moreover, the fact that more than 300,000 private (non-industrial) landowners own half of Florida's forestland suggests a potential demand for forest wildfire insurance protection.

This paper studies the spatiotemporal correlated risks of Florida wildfire outbreaks between 1981 and 2005. To model the spatial and temporal aspects of wildfire, it is critical to understand the underlying causes and propensity for wildfire. An extensive literature, including research by Prestemon et al. 2002 and Prestemon et al. 2005, investigated the temporal and spatial autocorrelation of wildfires and the relation of risk to underlying factors. Wildfire risks can be transmitted temporally and spatially and are affected by significant weather conditions as well as socio-economic factors.

We use statistical models to quantify wildfire risks and estimate associated insurance indemnity and premium rates. The wildfire risks and associated premium rates are estimated for a county-level annual contract which would pay pre-specified indemnities to insured forestland owners of a specific county in the event that the wildfire frequency or propensity exceeds the pre-specified levels.

The remainder of this chapter is organized a follows. Section 1 will derive the wildfire risks functions and introduce several statistical spatio-temporal models. Section 2 will discuss the data and present some preliminary analysis results. Section 3 will analyze the empirical results. Section 4 will make conclusion remarks and discuss future extension of our studies.

1. Conceptual Framework

   i. Functional forms of wildfire risks

In an actuarially fair insurance plan, the insurance premium should be set equal to the total expected loss. In a general term, if we define z as a loss event, the expected loss should be expressed as

$$E(Loss) = \int f(z) * E(loss|z)\, dz, \quad (1.1)$$

where $f(z)$ is the probability density for the event z, and $E(loss|z)$ is the conditional expected loss when the loss event z occurs. If z is a discrete variable with N possible outcomes, (1.1) can be rewritten as

$$E(Loss) = \sum_{i=1}^{N} P(z = i) * E(loss|z = i). \quad (1.2)$$

If we use $z = 1$ to denote a loss event, and $z = 0$ otherwise, the expected loss can be expressed as

$$E(Loss) = P(z = 1) * E(loss|z = 1). \quad (1.3)$$

Some insurance plan pays a fixed amount of money. For example, a life insurance policy will only pay the beneficiary in the event of death of the insured, without any possible partial payment, e.g. $E(loss|z = 1)$ is a predetermined amount. Because in an actuarially fair insurance scheme, the premium should be set equal to expected loss $E(Loss)$, our main task is to model the loss probability $P(z = 1)$. The loss probability is assumed to be contingent upon a set of observable covariates X

$$P(z = 1) = F(X\beta), \tag{1.4}$$

where β is the associated parameter vector.

Although we have already gathered data of relevant observable covariates X for wildfires such as weather and forestry conditions, how to define $P(z = 1)$ is still a question. One difficulty stems from the fact that most of the wildfire counts in the Florida counties are more than once each year. The other difficulty is that a wildfire would probably never burn down a whole county, and burned size of each incidence varies enormously. However, if $P(z = 1)$ can be viewed as the risk probability of one site being burned, it is reasonable to assume that this probability is uniform across a county in a specific year

$$z = \begin{cases} 0 & if\ the\ site\ does\ not\ get\ burned \\ 1 & if\ the\ site\ gets\ burned \end{cases}. \tag{1.5}$$

If each county is divided into *n* equally sized small sites, from the law of large numbers, it is known that

$$\lim_{n\to\infty} P(\bar{z} - E(z)) = \lim_{n\to\infty} P\left(\frac{\#\{z=1\}}{n} - P(z = 1)\right) = 0,$$

or

$$\frac{\#\{z=1\}}{n} \xrightarrow{P} P(z = 1). \tag{1.6}$$

Since $\#\{z = 1\}$ is the total number of burned sites among the total *n* sites, $\frac{\#\{z=1\}}{n}$ becomes the burned ratio of the total *n* sites in each county. If *Y* is denoted as the burned ratio of a county, the ratio $\frac{\#\{z=1\}}{n}$ will converge to the burned ratio when the total number of sites *n* increases to the infinity

$$\frac{\#\{z=1\}}{n} \xrightarrow{n} Y. \tag{1.7}$$

During a specific year in a county, both Y and $P(z = 1)$ are fixed values. Then combining (1.6) and (1.7), the burned ratio *Y* becomes an indicator of the burning risk probability

$$Y = P(z = 1). \tag{1.8}$$

After plugging (1.8) into (1.4), loss probability function can be formalized as

$$Y = F(X\beta), \tag{1.9}$$

where Y is the burned ratio. However, not only do we want to develop a reasonable insurance scheme, but also we hope to help with the wildfire management policies. Therefore, in addition to regressing on the burned ratio, the wildfire frequency, intensity and density will also be used as the dependent variable *Y* respectively in this essay.

To estimate the general loss probability function (1.9), we start by investigating the wildfire causes. The occurrences of wildfire are due to many sources, among which arson and lightning are two major leading causes. In average, arsonists set 1.5 million fires each year in the United States, resulting in over 3 billion in damages, about 500 fatalities, and thousands of injuries (TriData Corporation). The analysis of Florida wildfire causes (Figure 1.A) confirms this statement and shows that over 25% of wildfires are caused by arson and over 15% are caused by lightning. Since these two causes take a large proportion of all the wildfires, it is necessary to look into the crucial factors affecting the functions of each one.

Following is the classical crime function (Becker 1968) as applied in Butry & Prestemon(2005)

$$EU_i(O_i) = \pi_i U_i\big(g_i - c_i - f_i(W_i, w_i)\big) + (1 - \pi_i)U_i(g_i - c_i), \tag{1.10}$$

where EU_i expresses the expected utility of committing a crime, g_i and c_i are the benefit and the cost of the incendiary respectively. $f_i(W_i, w_i)$ is the loss when being caught, where w_i is the wage and W_i is the employment status. π_i is the probability of being caught, which should be a function of law enforcement. Some analysis indicates that law enforcement effort may be simultaneously determined along with crime (Becker; Fisher and Nagin), so it is natural to consider π_i as a function of g_i, c_i, W_i, w_i

$$\pi_i = \pi(g_i, c_i, W_i, w_i). \tag{1.11}$$

If we assume that the benefits of arson crimes are homogeneous in an area, which means that g_i is a constant g within the same area, the arson crime function should be a function of W, w, c_i. The production cost c_i is a function of time available (Jacob and Lefgren), fuels, and weather (Gill and et al., Vega Carcia and et al., Prestemon and et al. 2002)

$$c_i = c(L_i, fuel_{s,t}, weather_{s,t}), \tag{1.12}$$

where s is the location and t is the time point. The leisure time available for each individual is also associated with the employment status and the wage

$$L_i = L(W_i, w_i). \tag{1.13}$$

Fuels are determined by land ground coverings. Since wildfires always happen on the forestland, it is natural to assume

$$fuel_{s,t} = F(forestry_{s,t}), \tag{1.14}$$

where $forestry_{s,t}$ is the forestland condition at time *t* and location *s*. After plugging (1.13) and (1.14) into (1.12), the individual wildfire production cost function can be expressed as

$$c_i = c(L(W_i, w_i)\ , F(forestry_{s,t}), weather_{s,t}). \tag{1.15}$$

Then the individual arson crime utility function is a function of the wage, the employment status, the forestland type and the weather condition

$$\begin{aligned}EU_i(O_i) = \pi\big(g, c\big(L(W_i, w_i), F\big(forestry_{s,t}\big), weather_{s,t}\big), W_i, w_i\big)\\ * U_i\left(g - c\big(L(W_i, w_i), F\big(forestry_{s,t}\big), weather_{s,t}\big) - f_i(W_i, w_i)\right) +\\ \big[1 - \pi\big(g, c\big(L(W_i, w_i), F\big(forestry_{s,t}\big), weather_{s,t}\big), W_i, w_i\big)\big]\\ * U_i\big(g - c(L(W_i, w_i)\ , F(forestry_{s,t}), weather_{s,t})\big).\end{aligned} \tag{1.16}$$

The arson decision made by each person is therefore made by maximizing (1.16) given $g, W_i, forestry_{s,t}$ and $weathter_{s,t}$. Hence the individual arson crime decision must be a function of all these covariates

$$Arson_i = A_i(g, g_i, W_i, forestry_{s,t}, weather_{s,t}). \tag{1.17}$$

Consequently, the aggregate arson incidences of an area within a period should be determined by population, employments, forestry types and weather conditions,

$$\#Wildfire_{arson} = \sum_{i=1}^{n} A_i\left(g, W_i, forestry_{s,t}, weather_{s,t}\right), \quad (1.18)$$

where *n* is the population.

Another important wildfire cause is lightning, especially "dry lightning". A dry lightning is a lightning that happens outside the raining area. After the dry lightning strikes the ground, whether a wildfire could happen does not only depend on the weather, but also on the forestland situation. Naturally the weather condition and the forestland type are the two main factors for a lightning to cause a wildfire. Hence the lightning caused wildfire ignitions can be expressed as

$$\#Wildfire_{lightening} = Lightening\left(forestry_{s,t}, weather_{s,t}\right). \quad (1.19)$$

Debris burnings also cause many wildfires in the south (Figure 1). As it is a relatively inexpensive option to get rid of debris, debris burning is usually associated with socio-economic conditions. Besides, a wildfire caused by unsafe debris burning, always a consequence of setting a fire at the wrong place (land ground) and the wrong time (weather), must be related to the forestry and weather conditions

$$\#Wildfire_{Debris} = Debris\left(forestry_{s,t}, weather_{s,t}\right). \quad (1.20)$$

Therefore, combining (1.18), (1.19) and (1.20), the aggregate wildfire incidences from all sources can be expressed as

$$\#Wildfires = \sum_{Source} \#Wildfire_{Source} = F\left(g, W, n, forestry_{s,t}, weather_{s,t}\right), \quad (1.21)$$

where W is the vector of all the people's wages in this community. As the arson benefit is non-monetary in most times, and assumed to be constant, we would like to drop this variable. Hence, the analysis of the main causes of wildfires leads us to investigate how forestland types, weather conditions and socio-economic conditions affect the wildfires incidences. Since wildfire propensity, density and burned ratio can be similarly derived from this theory modeling, we can write

$$Y = F\left(W, forestry_{s,t}, weather_{s,t}\right), \tag{1.22}$$

where Y is a measurement of wildfire risks.

ii. Statistical models

Pooled Regressions

As usual, a pooled linear regression is used at the beginning. The OLS regression takes a form of

$$Y = X\beta + \varepsilon, \tag{1.23}$$

For each observation *i*,

$$y_i = x_i\beta + \varepsilon_i \ , \tag{1.24}$$

where ε_i is i.i.d. white noise.

Pure STAR (Spatio-Temporal Auto-Regressive) Model

Annual wildfire counts have been found positively auto-correlated both temporally and spatially in Table 2. Therefore, it is needed to accommodate this data with an auto-correlated structure. Whittle (1954) proposed a pure spatial auto-regression model in the form of

$$Y = \rho WY + \epsilon, \tag{1.28}$$

where ρ is the spatial dependence parameter and W is a spatial weight matrix. However, the temporal dependence isn't incorporated in (1.28). Thus we decided to combine (1.28) with an AR(1) process, and obainted a pure spatio-temporal auto-regression model

$$Y = \rho_s W_s Y + \rho_t W_t Y + \epsilon, \tag{1.29}$$

where the residuals of the vector ϵ are i.i.d. white noise such that $\epsilon \sim W(0, \delta^2 I_{ST})$, given *S* is the number of counties, and *T* is the number of years. In (1.29), ρ_s measures the spatial dependence of wildfires, and ρ_t measures the temporal dependence. The dependent vector is $Y = (y_1\ y_2\ y_3 \ldots y_t \ldots y_T)'$. Each element y_t is a vector of all counties' dependent variable observations in the year *t*

$$y_t = (y_{1,t}\ y_{2,t}\ y_{3,t}\ \cdots y_{s,t} \cdots y_{S,t})',$$

where $y_{s,t}$ is the observation of the dependent variable in the county *s* during the year *t*. As a result, the dependent vector can be decomposed as

$$Y = (y_{1,1}\ y_{2,1}\ y_{3,1}\ \cdots y_{s,1} \cdots y_{S,1}\ y_{2,1}\ y_{2,2} \cdots y_{S,2} \cdots y_{S-1,1} \cdots\ y_{S-1,T}\ y_{S,T})'. \tag{1.30}$$

To avoid the singularity problem in the estimation stage, the spatial weight matrix W_s in (1.29) should be carefully constructed as

$$W_s = I_T \otimes Border, \tag{1.31}$$

where ⊗ is the Kronecker product and $Border$ is defined as

$$Border(i,j) = \begin{cases} \frac{1}{\#\{k: k\ is\ a\ neighbor\ of\ i\}} & if\ j\ is\ a\ neighbor\ of\ i \\ 0 & if\ j\ is\ not\ a\ neighbor\ of\ i \end{cases}. \tag{1.32}$$

In this weight matrix, the sum of each row is 1, and in other words, the weighted sum of each individual county's neighbors equal to 1. Hence, each element of $W_s Y$ in (1.31) will be the average of *Y* values of a county's all neighbors. To model the temporal dependence, a temporal weight matrix W_t is also required. The most straight forward and simplest way is to introduce an AR(1) process with

$$W_t = \begin{bmatrix} L & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & L \end{bmatrix}_{ST \times ST}, \tag{1.33}$$

where L is a lag operator. Now for each element of Y, the equation (1.30) can be simplified as

$$y_{s,t} = \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + \varepsilon_{s,t}, \tag{1.34}$$

where $y_{\bar{s},t}$ represents the average of neighboring y's of the county *s* in the year *t*.

Mixed STAR model

Extraordinarily simple is the Pure-STAR model, but exclusion of the covariates *X* ignores the influences of the environmental factors on wildfires. Withstanding this problem, another form of auto-regressive model can be written as

$$Y = X\beta + \rho_s W_s Y + \rho_t W_t Y + \epsilon. \tag{1.35}$$

In contrast to the Pure-STAR model, independent variables are addressed into this model, and thus it is viewed as a Mixed-STAR model. This model admits that not only the wildfires of spatial neighbors and temporal neighbors have direct impacts on the wildfires of a specific county in a specific year, but also the independent variables have direct influences on the wildfires. The equation (1.35) is in a matrix form, and each element of Y can be written as

$$y_{s,t} = x_{s,t}\beta + \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + \varepsilon_{s,t}\ . \tag{1.36}$$

Residuals-STAR model

The above two STAR models already took spatio-temporal autocorrelation into consideration, and the equations (1.34) and (1.36) indicate that the average of neighboring dependents and the lagged dependent can directly affect the dependent. However, if the residuals of OLS regression are auto-correlated, which are usually found in our regressions, a model which can incorporate spatially and temporally correlated residuals is needed. A Residuals-STAR (residuals-saptio-temporal-auto-regressive) model is constructed as

$$Y - X\beta = u, \tag{1.37}$$

where

$$u = (\rho_s W_s + \rho_t W_t)u + \epsilon. \tag{1.38}$$

Basically this Residuals-STAR model assumes that, other than the part explained by the regressors, the remaining unexplained part of the dependent's deviations are still auto-correlated. Combining (1.37) and (1.38), we got

$$Y - X\beta = (\rho_s W_s + \rho_t W_t)(Y - X\beta) + \epsilon, \tag{1.39}$$

which is equivalent to

$$Y = \rho_s W_s Y + \rho_t W_t Y + X\beta - \rho_s W_s X\beta - \rho_t W_t X\beta + \epsilon\ . \tag{1.40}$$

Each element of Y can be written as

$$y_{s,t} = \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + x_{s,t}\beta - \rho_s x_{\bar{s},t}\beta - \rho_t x_{s,t-1}\beta + \varepsilon_{s,t}, \tag{1.41}$$

where $x_{s,t} = (x_{s,t}^1, x_{s,t}^2 x_{s,t}^3\ , \dots, x_{s,t}^K)$ is the observation of independent variables in the county *s* at the year *t*, $x_{\bar{s},t} = (x_{\bar{s},t}^1, x_{\bar{s},t}^2 x_{\bar{s},t}^3\ , \dots, x_{\bar{s},t}^K)$ is the neighboring average of the independents, and $x_{s,t-1}$ is the independent variables' observation in year *t-1*. Due to the nonlinear structure in (5.20), a FIML (full information maximum likelihood) estimation method is to be used. It assumes i.i.d. error term ε with

$$E(\varepsilon_{s,t}) = 0, \tag{1.42}$$

$$Var(\varepsilon_{s,t}) = \Sigma, \tag{1.43}$$

and it tries to minimize the objective function

$$constant + \left[\frac{ST}{2}\right] ln(det(S^*)) - \textstyle\sum_{i=1}^{ST} ln\,|J|, \tag{1.44}$$

where S^* is the estimate for Σ, and

$$J = \frac{\partial\left(y_{s,t} - (\rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + x_{s,t}\beta - \rho_s x_{\bar{s},t}\beta - \rho_t x_{s,t-1}\beta)\right)}{\partial y_{s,t}}.$$

The covariance of the parameter vector $(\rho_s, \rho_t, \beta')'$ is $\left[\hat{Z}\left({S^*}^{-1} \otimes I\right)\hat{Z}\right]^{-1}$. Suppose there are *p* (here *p=K+2*) parameters in total, then

$$\hat{Z} = (\hat{Z}_1, \hat{Z}_2, \cdots \hat{Z}_p).$$

If we denote $q_{s,t} = y_{s,t} - \left(\rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + x_{s,t}\beta - \rho_s x_{\bar{s},t}\beta - \rho_t x_{s,t-1}\beta\right)$, $Q = (q_{1,1}, q_{2,1}, \dots q_{s,t}, \dots q_{S,T})'$ and $Q_i = \frac{\partial Q}{\partial \theta_{\mathrm{i}}}$, each element of $\hat{Z}$ can be written as

$$\hat{Z}_i = \varepsilon \frac{1}{ST} \sum_{s=1}^{S} \sum_{\mathrm{t}=1}^{\mathrm{T}} \left(\frac{\partial \left(y_{s,t} - \left(\rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + x_{s,t}\beta - \rho_s x_{\bar{s},t}\beta - \rho_t x_{s,t-1}\beta\right)\right)}{\partial y_{s,t}} \right)^{-1} *$$

$$\frac{\partial^2 (y_{s,t} - (\rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + x_{s,t}\beta - \rho_s x_{\bar{s},t}\beta - \rho_t x_{s,t-1}\beta))}{\partial y_{s,t} \partial \theta_{\mathrm{i}}} - Q_i, \tag{1.45}$$

where θ_{i} is the *ith* element of the parameter vector $(\rho_s, \rho_t, \beta^{'})^{'}$ and ε is the error vector.

Pooled Regression with Prediction

Similar to the analysis in the first sector, spatio-temporal autocorrelation is put aside at the beginning and pooled regressions are considered first. To forecast next year's wildfire risks, the independent variable should be known values. Therefore, the function (1.22) should be changed to

$$y_t = F(y_{t-1}\beta), \tag{1.46}$$

Among all the pooled regression models, the OLS regression has the simplest linear form,

$$y_t = x_{t-1}\beta + \varepsilon_t \ ,$$

given

$$t = 1,2 \cdots T,$$

where $y_t = (y_{1,t}, y_{2,t}, \cdots y_{S,t})$, *S* is the total number of counties. The error vector ε_t, is composed of i.i.d. white noise $\varepsilon_{s,t}$, with $E(\varepsilon_{s,t}) = 0$ and $Var(\varepsilon_{s,t}) = \sigma^2$. Each element of y_t is

$$y_{s,t} = x_{s,t-1}\beta + \varepsilon_{s,t}. \tag{1.47}$$

The parameter estimate vector $\hat{\beta}$ can be obtained from the OLS regression. When the data of T periods is available, the forecast of next year's dependent variable's observation $y_{s,T+1}$ should be

$$\widehat{y_{s,T+1}} = x_{s,T}\hat{\beta}. \tag{1.48}$$

When the dependent variable is a count data, the Poisson regression and the Negative-Binomial regression can be used. Similar with (1.26), for these two models, the regression on $y_{s,t}$ is

$$log(y_{s,t}) = x_{s,t-1}\beta + \varepsilon_{s,t}, \tag{1.49}$$

and the forecast of $y_{s,T+1}$ would be

$$\widehat{y_{s,T+1}} = e^{x_{s,T}\hat{\beta}}. \tag{1.50}$$

Forecast with Mixed-ST model

Following the spatio-temporal structure utilized by Anton et. al. (2008), Goodwin and Piggott (2009), a Mixed–ST (mixed spatio-temporal) model is assumed as

$$y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t w_t y_t + \varepsilon_t. \tag{1.51}$$

where $w_s = Border$, which is defined in (1.32), and

$$w_t = \begin{bmatrix} L & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & L \end{bmatrix}_{S\times S}.$$

This model assumes that the lagged neighboring wildfires as well as the lagged wildfires have direct effects on the current wildfire incidences. For each element of y_t,

$$y_{s,t} = \rho_s y_{\bar{s},t-1} + \rho_t y_{s,t-1} + x_{s,t-1}\beta + \varepsilon_{s,t}, \tag{1.52}$$

The forecast of $y_{s,T+1}$ would be

$$\widehat{y_{s,T+1}} = \widehat{\rho_s} y_{\bar{s},T} + \widehat{\rho_t} y_{s,T} + x_{s,T}\hat{\beta}\ . \tag{1.53}$$

Forecast with Pure-ST Model

In contrast with the Mix-ST model, which excluded all the covariates, this Pure-ST model is to determine whether most wildfires can be solely explained by the direct effects by lagged neighboring wildfires and lagged wildfires in the form of

$$y_t = \rho_s w_s y_{t-1} + \rho_t w_t y_t + \varepsilon_t, \tag{1.54}$$

In this equation, the explanatory variables are excluded from the regression, and we only use the first order auto-regressive structure. In the former segment, the 1st order spatial dependence is modeled as the dependence between each county's wildfires and its neighbors' in the same period. However, in this Pure-ST model and the above Mixed-ST mode the spatial dependence is the dependence between each county's wildfires and its neighbors' lagged wildfires. Each element of y_t has the form of

$$y_{s,t} = \rho_s y_{\bar{s},t-1} + \rho_t y_{s,t-1} + \varepsilon_{s,t}. \tag{1.55}$$

Consequently, wildfires in the next period is predicted by

$$\widehat{y_{s,T+1}} = \widehat{\rho_s} y_{\bar{s},T} + \widehat{\rho_t} y_{s,T}. \tag{1.56}$$

Forecast with Pure-STAR Model

To do the forecast, the Pure-STAR model is used again. Same as before, in this model the spatial auto-correlation is still modeled between the dependents in the same year

$$y_t = \rho_s w_s y_t + \rho_t w_t y_t + \varepsilon_t, \tag{1.57}$$

which is equivalent to

$$y_t = \rho_s w_s y_t + \rho_t y_{t-1} + \varepsilon_t, \tag{1.58}$$

and each element of y_t

$$y_{s,t} = \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + \varepsilon_t.$$

However, unlike (1.56), which directly predicts the observation in the next period $y_{s,T+1}$ using the information at the period *T*, some matrix manipulation is needed to forecast the dependent in (1.58). From (1.58), we can get

$$(I_S - \rho_s w_s)y_t = \rho_t y_{t-1} + \varepsilon_t,$$

Hence

$$y_t = (I_S - \rho_s w_s)^{-1}\rho_t y_{t-1} + (I_S - \rho_s w_s)^{-1}\varepsilon_t. \tag{1.59}$$

To get the wildfire prediction in the time period *T+1*, most people would suggest using the estimator as

$$\widehat{y_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}\widehat{\rho_t} y_T. \tag{1.60}$$

Although this estimation is unbiased, it omits the autocorrelation structure between residuals, which is in the second part at R.H.S. of (1.59). Therefore, another estimate of y_{T+1} is considered, which includes an estimate of ε_{T+1} and is in the form of

$$\widehat{y_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}\widehat{\rho_t} y_T + (I_S - \widehat{\rho_s} w_s)^{-1}\widehat{\varepsilon_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}(\widehat{\rho_t} y_T + \widehat{\varepsilon_{T+1}}). \tag{1.61}$$

As the residuals part ε_t is assumed to follow the same distribution, a natural way to estimate $\widehat{\varepsilon_{T+1}}$ is to get the average of residuals from the existing *T* periods,

$$\widehat{\varepsilon_{T+1}} = \frac{\sum_{t=1}^{T} \widehat{\varepsilon_t}}{T}. \tag{1.62}$$

For each $\widehat{\varepsilon_t}$,

$$\widehat{\varepsilon_t} = (I_S - \widehat{\rho_s} w_s)y_t - \widehat{\rho_t} y_{t-1}. \tag{1.63}$$

Combing (1.62) and (1.63), we have

$$\widehat{\varepsilon_{T+1}} = \frac{\sum_{t=1}^{T}(I_S - \widehat{\rho_s} w_s)y_t - \widehat{\rho_t} y_{t-1}}{T}. \tag{1.64}$$

After we combined (1.59) and (1.64), we got another forecast for the dependent variable,

$$\widehat{y_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}\widehat{\rho_t} y_T + \frac{1}{T}\sum_{t=1}^{T}[y_t - (I_S - \widehat{\rho_s} w_s)^{-1}\widehat{\rho_t} y_{t-1}]. \tag{1.65}$$

For comparison, both (1.59) and (1.65) are used to forecast future wildfires. With the expressions of these two estimates, one important thing should be noticed. That is, each county's $y_{s,T+1}$ could not be directly predicted as it was before, but instead the vector of all the future values $\widehat{y_{T+1}}$ can only be simultaneously forecasted. The same out-of-sample-check methods are also used with these two estimators (1.61) and (1.65).

Forecast with Mixed-STAR Model

Similar to the above segment, a Mixed-STAR model is also used to forecast wildfires. Unlike in the *Mixed-STAR Model* segement, where the contemporaneous independent variables were used, lagged covariates are used in this part. The reason is that when wildfires in the time period *T+1* are being forecasted at the end of the year *T*, only the information of first T period is available. The regression takes a form of

$$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t w_t y_t + \varepsilon_t, \tag{1.66}$$

which is equivalent to

$$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1} + \varepsilon_t. \tag{1.67}$$

The regression on each element $y_{s,t}$ of the vector y_t can be expressed as

$$y_{s,t} = x_{s,t-1}\beta + \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} + \varepsilon_{s,t}. \tag{1.68}$$

Similar to (1.60) and (1.64), future wildfires in the period *T+1* can be estimated in two ways, either

$$\widehat{y_t} = (I_S - \widehat{\rho_s} w_s)^{-1}(x_{t-1}\hat{\beta} + \widehat{\rho_t} y_{t-1}), \tag{1.69}$$

or

$$\widehat{y_t} = (I_S - \widehat{\rho_s} w_s)^{-1}\left(x_{t-1}\hat{\beta} + \widehat{\rho_t} y_{t-1}\right)$$

$$+\frac{1}{T}\sum_{t=1}^{T}[y_t - (I_S - \widehat{\rho_s} w_s)^{-1}\left(x_{t-1}\hat{\beta} + \widehat{\rho_t} y_{t-1}\right)] \tag{1.70}$$

Forecast with Residuals-STAR Model

Other than imposing the auto-correlation of the dependent into the model, we have considered another possibility that the residuals are auto-correlated. The same model is applied as the one used in the *Residuals-STAR Model* segment, but lags of the independent variables are used instead of contemporaneous covariates.

$$y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta) + \varepsilon_t \Rightarrow$$

$$y_t = x_{t-1}\beta + (\rho_s w_s + \rho_t w_t)y_t - (\rho_s w_s + \rho_t w_t)x_{t-1}\beta + \varepsilon_t \Rightarrow$$

$$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t w_t y_t - \rho_s w_s x_{t-1}\beta - \rho_t w_t x_{t-1}\beta + \varepsilon_t \Rightarrow$$

$$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1} - \rho_s w_s x_{t-1}\beta - \rho_t x_{t-2}\beta + \varepsilon_t$$

Therefore, each element of y_t is

$$y_{s,t} = x_{s,t-1}\beta + \rho_s y_{\bar{s},t} + \rho_t y_{s,t-1} - \rho_s x_{\bar{s},t-1}\beta - \rho_t x_{s,t-2}\beta + \varepsilon_{s,t}. \quad (1.71)$$

Similarly with the above segment, two forecasts of y_{T+1} are derived

$$\widehat{y_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}(x_T\hat{\beta} + \widehat{\rho_t} y_T - \widehat{\rho_s} w_s x_T\hat{\beta} - \widehat{\rho_t} x_{T-1}\hat{\beta}), \quad (1.72)$$

or

$$\widehat{y_{T+1}} = (I_S - \widehat{\rho_s} w_s)^{-1}\left(x_T\hat{\beta} + \widehat{\rho_t} y_T - \widehat{\rho_s} w_s x_T\hat{\beta} - \widehat{\rho_t} x_{T-1}\hat{\beta}\right)$$

$$+\frac{1}{T}\sum_{t=1}^{T}[y_t - (I_S - \widehat{\rho_s} w_s)^{-1}\left(x_{t-1}\hat{\beta} + \widehat{\rho_t} y_{t-1} - \widehat{\rho_s} w_s x_{t-1}\hat{\beta} - \widehat{\rho_t} x_{t-2}\hat{\beta}\right)]. \quad (1.73)$$

2. Data and Preliminary Analysis

    i. Databases

The wildfire data used in this chapter is the Florida wildfire database obtained from the Florida State Forestry Division. Each observation has the initial time, location (the county), duration, fuel type, fire cause, and burned acreage of the wildfire. The time span is from

1981 to 2005. We have got every county's yearly wildfire count and annual burned acreage by aggregating the ignitions in each county every year.

Different wildfire causes were analyzed and it is found that one wildfire is a function of some environment factors and socio-economic factors. Other than that analysis, in this part, the mechanism of wildfire ignitions will be checked comprehensively again. The fire environment triangle (Countryman 1972) of fuels, topography and weather will be evaluated. Also in the functional form of wildfire risks

$$Y = F\left(W, forestry_{s,t}, weather_{s,t}\right)$$

a number of independent variables should be selected.

For almost all the wildfires, fuels are the vegetation on the ground, so it is important to understand the land ground conditions in Florida. The FIADB (Forest Inventory and Analysis Database) obtained from the U.S.D.A., which consists of 4 national forest inventory surveys from 1980 until 2007, contains all the forest characteristics of interest, such as owner types and timber categories. In the total forestland acreage for each county each year, the acreages of privately owned forestland, publicly owned forestland, and each forestland type are available. Then several important variables are derived: total forestland acreage, private owners' share, public owned share and the share of each forest type group (e.g. Oak / pine group, Longleaf / slash pine group). However, these observations are not consecutively available since the FIADB has only 4 inventories (1980, 1987, 1995 and 2007). To make up for the missing values, we interpolated them with a "join" method (join points with straight lines) to obtain the whole time series. As different types of trees grow at different altitudes, the forest types information from this dataset not only directly provides the vegetation distributions, but also implicitly discloses the related topography conditions, which is the second leg of fire environmental triangle.

According to the common knowledge to prevent wildfire, some categories of trees and bushes on the ground can increase wildfire risks while some other ones can decrease wildfire risks. We will evaluate the effects of forestland types on wildfires, and expect some type of trees like short leaf pines would be a crucial factor.

In addition to the shares of different forestry groups, the forestry ownership structure is also important. From the arson crime theory, it is known that law enforcement plays a role in stopping arsons (Butry & Prestemon 2005). Compared with privately owned forestlands, government owned forestlands usually have more forest police and forest rangers, which should reduce arson risks and careless fire burnings.

Weather, the third leg of fire environmental triangle, though easy to observe day by day, is difficult to measure at an annual basis. To get weather observations, we used the National Climate Database from the N.O.A.A.. This data provides monthly weather observations, such as temperature, precipitation and drought index, for the 8 climate zones in Florida. After matching all the 67 counties of Florida with the 8 climate zones, we got monthly observations for all the counties. In general, there are two main categories of weather condition variables. One is temperature and the other is drought. After a careful selection, we used the HDD index as the temperature indicator and the SP12 index as the drought measurement in this chapter.

Although monthly temperatures are already reported in the database, they turned out not to be a good measurement of coldness or hotness. The reason is that the reported monthly temperature is the mean of all the days' temperatures in a month, which neglects the variations. Instead, the HDD index is preferred. HDD, or "Heating degree days", is a measure of how much (in degrees), and for how long (in days), outside air temperature is lower than a specific "base temperature" (or "balance point"). This index is used to calculate related energy consumptions required to heat buildings. HDD's are calculated by subtracting the average temperature for a given day from 65°F. For example: A 45°F day results in 20 HDD's. Conversely, a 70°F day results in -5 HDD's, which is less than zero, in such case the HDD's for that day will automatically default to zero, as no significant energy for heating is necessary. Obviously, a higher HDD is equivalent to a colder weather. The monthly HDD index in the database is a sum of every day's HDD's in a month. In other words, HDD measures the coldness of that month. Since our empirical analysis is focused on the yearly wildfire frequency, we aggregated all the twelve months' HDD's of the same calendar year together. Because the sum of HDD's is the total heating energy consumption in a specified year, it is a reasonable measurement of how cold that year is. However, in some cases, it is found that the relevant coefficient is too small to report, so instead of the yearly HDD index, we decided to use the daily averaged HDD index as an explanatory variable by dividing the yearly HDD index by 365. In contrast, another index CDD, or "cooling degree days", measures hotness. Unfortunately it is not appropriate to add the CDD index into the model due to an obvious reason: HDD and CDD tend to be linear correlated, and using both of them at the same time would cause a multi-colinearity problem.

SP12 (Standardized Precipitation Index), measurement of drought, is a transformed form of the probability of observing a given amount of precipitation in 12 months. A zero index value reflects the median of the distribution of precipitation, a -3 indicates a very extreme dry spell, and a +3 indicates a very extreme wet spell. The more the index value departs from zero, the drier or wetter the recent 12 months is, when compared to the long-term climatology of the location. Therefore, the value of the SP12 index in December tells the drought condition of the past whole year.

Another important weather phenomenon affecting wildfires is hurricane. As mentioned in some wildfire reports (eg. President Report 2000), strong storm winds could exaggerate the spread and the intensity of large fires. Therefore hurricanes are also considered part of weather variables in this paper. In forestry literatures, there is a long debatable hypothesis that wildfire risks grow when trees are knocked down by hurricanes. Also, there is another argument that since government agencies will immediately remove the fallen trees and impose strict legal restrictions on wildfire burnings after the hurricane season, the wildfire hazards in the following year will be reduced. In this paper, these two hypothesizes will be tested.

A hurricane is categorized as a powerful storm that begins over a warm sea with a wind speed of at least 65 knots. The hurricane dataset is also obtained from the N.O.A.A., and it traces all the hurricanes since 1851. It records each hurricane's location and speed every six hours. We concentrated our attention on the hurricanes after 1980, and counted one incidence only if a hurricane trespasses a Florida county. Then we accumulated the incidences for each county every year and got the yearly hurricane counts in the Florida counties.

To account for the opportunity costs to ignite a wildfire, the economic status of people in society will also be evaluated. Since poor people are more likely to set fires to burn fuels or for other purposes, it is likely that the poorer the area, the higher wildfire risk it faces (Butry and Prestmon 2005). Hence, socio-economic conditions can be an influential factor to cause wildfires, and especially incendiaries. For the socio-economic conditions, the REIS dataset, which is obtained from US Census of Labor and Statistics, is used. It tells the socio-economic situations in the Florida counties every year. Though there are many socio-economic variables available, population, employment and income are the main factors of concern which are dictated by the theory. Moreover, a measurement of low income population, the government transfer payment is considered. However, it turns out that the average income, the average transfer payment and the employment rate contain very similar information: how wealthy or poor this area is. Not surprisingly they are so correlated that only one of them can be used. After evaluating those variables carefully, we chose the employment rate, along with the population, to measure socio-economic conditions.

Table 1 presents the definitions and summary statistics of the wildfire dependent variables and other relevant explanatory variables. Each dependent variable has records of all the 67 counties in Florida between years 1981-2005. This results in a total of 1675 county-year combined observations. Meanwhile every explanatory variable covers the years from 1980 to 2005, consisting of 1742 observations. The reason that we need one more year's observations of the covariates than those of the dependent is that the lags of regressors will be utilized in the prediction models.

ii. Preliminary analysis

The dataset used in this paper consists of all the Florida counties' annual wildfire records and the associated environmental conditions from 1980 to 2005. If the fact that those counties are geographically adjacent is not important, then this data can be viewed as only a panel data. In such a circumstance, the ordinary panel data analysis methods, like the fixed-effect model and the random-effect model, are suitable for analysis. However, a couple of facts should not be ignored: the observations in this panel data could be both spatially and temporally auto-correlated. Figure 2 is the plot of four neighboring counties' annual wildfire counts. Obviously, the movements of these four curves tend to have a similar pattern, so spatial autocorrelation may be the case. For each county, its time series has a downward trend, and does not follow a white noise pattern. This implies that a non-stationary stochastic process may exist. Consequently, both spatial correlation and temporal correlation should be considered.

3. Empirical analysis

i. Contemporaneous Wildfire Risks

Annual Wildfire Counts

Table 3 presents the parameter estimates of different models when the dependent variable is annual wildfire count. The results tell some useful information. First of all, across all those models, most independent variables have consistent impacts on the wildfire ignitions respectively. Secondly, not only most coefficients are significant, but also the directions of the impacts are as expected.

Scale factors, total forestland size and population size, are always significantly positively linked to wildfires. This is due to the fact that a larger forestland or population implies a higher probability of wildfire incidence. For forestland conditions, the higher the private share of forestland, the more frequently wildfires happen. It is because that there are more police forces and forest rangers on the federal and state owned forestlands than on private owners' forestlands. As a consequence, incendiaries and careless wildfire uses are more likely to take place on privately owned forestlands. Different forestland types affect wildfires in different ways. The longleaf / slash pine forestland & loblolly / shortleaf pine forestland group can decrease wildfire incidences. This effect is significant negative except in the Mixed-STAR model (column 3.c). Conversely, the impact of the oak / pine forestland & oak / hickory forestland group is consistently positive, and is significant in all the models expect in the Residuals-STAR model. Meanwhile, the oak / gum / cypress forestland group significantly reduces wildfires, except in the Mixed-STAR model. Nearly all the weather conditions have significant effects on wildfires. A higher HDD index, equivalent to a cooler year, significantly decreases wildfire risks. A drier weather, which is presented by a smaller value of the December's SP12 index, significantly increases wildfires. Though only significantly in Poisson model, hurricanes are positively linked to wildfires in all the models that are without auto-regressive configurations. Employment, proxy of the opportunity cost to set arson ignitions (Prestemon and Butry 2005), significantly decreases wildfires when no autocorrelation is imposed. Unfortunately, the coefficient of employment ratio becomes insignificant in the Mixed-STAR model and it becomes positive in the Residuals-STAR model. However, these two unpleasant estimates are largely attributable to the multicollinearity problem, so generally the impact of employment on wildfires is still significantly negative.

In order to better evaluate the impacts of those independent variables, the elasticities of all the independents in the Poisson model and in the Negative-Binomial model are also calculated. The results are similar between these two models. The private share of forestland has the highest elasticity while the forestland size and the HDD index also have relatively large impacts. For example, if the private share increase by 1%*(*this 1% increase is not representing an increase from 1% to 2%, but representing an increase by 1%, i.e. 1% to 1.01%), wildfire incidences will increase by 0.65%. Although none of the factors are elastic, most of them have elasticities bigger than 0.1. Among all the covariates, the hurricane has the smallest elasticity due to the fact that hurricanes rarely happen.

When there are many independent variables in a single regression it is essential to make sure that no multi-collinearity exists. All the three rules of thumb-- VIF, tolerance and condition index-- are satisfied in the OLS, Pure-STAR and Mixed-STAR models. Therefore, no multicollinearity problems exist in those regressions. In the Residuals-STAR model, as the regression is in a non-linear form, no multi-collinearity analysis is available.

The model fitness is fine in the OLS regression, as the R-square is 26%. The Pure-STAR model enormously improves the R-square to around 56%. It means that over a half of the wildfires variations can be explained by the neighboring average and lagged wildfires. The Mixed-STAR model and the Residuals-STAR model have even higher R-squares because more regressors are used. Between these two models, the Residuals-STAR model has a better fitness and it implies that the Residuals-STAR model is preferred in wildfire count analysis.

According to the previous analysis, if the pooled regressions, i.e. OLS, Poisson and Negative-Binomial models, produced uncorrelated residuals, a further consideration of auto-regressions is not needed. Therefore, spatio-temporal correlation of the pooled regression residuals is examined.

Moran's *I* index and Geary's *c* index are still valid to check residuals' spatial autocorrelation, by contrast, the Durbin-Watson test for temporal autocorrelation is illegitimate as long as the lagged dependent variables are involved in the regression (Nerlove and Wallis (1966), Durbin (1970), Dezhbaksh (1990)). Instead, the Breusch(1978)–Godfrey(1978) serial correlation LM test, which is a robust statistic for lagged regressions, will be adopted. This Breusch-Godfrey statistic is used for a test of H_0: no autocorrelation versus H_1: residuals follows AR(P) or MA(P). After every regression, this test is carried out against the AR(1) model of the residuals and the autocorrelation sign is recorded at the same time.

In the bottom of Table 3, the spatial correlation tests (Moran's I and Geary's c) results and the temporal correlation tests (DW and BG) results of the residuals in each model are reported. It is found that the residuals are still strongly positively spatio-temporal correlated, which confirms the need to use auto-regressive models.

After the auto-regressive models are imposed, positive spatial correlation is mostly eliminated. Of all the 24 years, only in at most 5 years, significant positive spatial autocorrelation exists in Table 3.B-D. However, the residuals have significant negative spatial correlation in some years. For example, the Moran's *I* tests show that the negative spatial correlation of residuals is significant in 10 out of the 24 years at 0.1 level in the

Residuals-STAR model (column 3.c), while in 10 of 24 in the Mixed-STAR model and in 15 of 24 in the Pure-STAR model. These negatively correlated residuals may be caused by overestimating the positive spatial dependence, or may be due to the heterogeneity of the spatial autocorrelation between different counties.

As DW tests are invalid for lagged regressions, BG tests are preferred in the temporal autocorrelation tests (column 3.b-d). The positive temporal autocorrelation problem is almost solved by the three STAR models, and negative autocorrelation of the residuals only exists in very few years in the Pure-STAR model and the Mixed-STAR model. The Pure-STAR model and the Mixed-STAR model have incorporated the spatio-temporal dependence of the dependent variable directly into the models, on the contrary, the Residuals-STAR model is constructed on the dependence of the residuals. As a matter of fact, the temporal autocorrelation of the Residuals-STAR regression residuals is significantly negative in 17 out of 67 counties. This means that in some years, the positive temporal dependence of residuals is overestimated by the Residuals-STAR model. Also it implies that the spatial autocorrelation of residuals is not as strong as that of the dependent.

In general, the STAR models fit the data better than the pooled regressions and most independent variables have significant impacts on wildfires.

Trans-log Regressions of Annual Wildfire Count and Burned Acreage

One concern with the OLS result in Table 3 is the low R-square values. This is understandable because the distribution of wildfire counts, as in Figure 4.A, is unlikely to be normal. Therefore, some data transformation techniques are considered. Figure 4.B represents the distribution of wildfire counts after they are taken logarithm of. Since it looks much more alike a normal distribution than before, it is better to use the logarithm of wildfire counts as the dependent variable. Meanwhile, we are going to adopt the same set of explanatory variables in the regression analysis. However, for consistency, the positive numerical independent variables, including the two scale factors (the forestland size and the population size), will also be in trans-log forms. The weather factors, either already transformed from the weather observations, or with a lot of zeros in the case of hurricanes, are not suitable to be taken logarithms of. All the other variables are originally measured in percentages hence they don't need any transformations.

In addition to the frequencies of wildfires, the damages of wildfires are another important subject to study. The annual burned size measures the wildfire damages well. Figure 5.A depicts the distribution of annual wildfires burned sizes. Among wildfires, a lot of them only result in burning less than 5 acres, which is not a big loss. However, a disastrous wildfire could burn down thousands of acres of forestlands after it spreads widely, causing a loss much bigger than the total damages caused by hundreds of small wildfires. In other words, the variation of damages is so big that the largest burned size is thousands larger than the smallest ones. As a result, the distribution is far away from a normal distribution. Since there are a proportion of gigantic values, the popular count data models, such as the Poisson model and the Negative-binomial model, are not appropriate to use, even if the burned acreage can be thought as the total count of burned one-acre-sites. In the same way, the natural logarithm of burned sizes are preferred, and the transformed data distribution, as in Figure 5.B, complies with a normal distribution well. Thus the trans-log form of annual burned sizes is used as the dependent variable. For the convenience to do comparisons, the same set of regressors are used as in the trans-log annual wildfire count regressions. The results are presented in Table 4.

The overall results in Table 4 are analogous to those in Table 3. For example, the coefficients of the forestland size and the population size are always significantly positive in all the models. Although most variables have similar patterns of impacts on the wildfires in trans-log regressions as they did in Table 3, several things should be noticed. First of all, the R-square of the OLS regression (column4.a) on the logarithm of wildfire counts is improved a lot to 0.465 compared with 0.364 in Table 3. Similarly, in Table 4, the R-squares in the Pure-STAR model, the Mixed-STAR model and the Residuals-STAR model are all enhanced by around 10% compared with those in Table 3. Therefore, the logarithm transformation has improved the model fitness. Also the model fitness for burned acreage data is moderately good, with the R-square values ranging from 0.35 in the OLS model to around 0.5 in all the STAR models.

Secondly, the two scale factors, forestland size and population, are still always significantly positive in all the regressions. But they affect the wildfire frequency in a different way from the way they affect the intensity. For the forestland size variable, the elasticity of its impact on wildfire counts is bigger than that on wildfire burned sizes. In the meantime, the impact of the population variable is also much more elastic on the burned acreages than on the wildfire count.

Forestland ecosystem affects wildfires through various ways. The private owners' share always significantly increases wildfires both in frequency and intensity, which coincides with our expectation. The impact of the longleaf / slash pine forestland & loblolly / shortleaf pine forestland group is insignificant in all the models except the one in the OLS regression on wildfire counts, and the sign of its coefficient flips irregularly. The share of oak / pine forestland & oak / hickory forestland group significantly decreases both wildfires and burned sizes in most models, which exhibits a different pattern compared with Table 3. The elasticity of this forestland type's share is moderate in the regressions on wildfire count while its elasticities in the regressions on burned sizes are one of the highest. The oak / gum / cypress forestland share significantly increases burned sizes while it significantly decreases wildfire ignitions. This interesting finding implicates that this forestland type can alleviate the chance of wildfire ignition, and at the same time it can help with wildfire contaminations.

The weather factors, temperature and drought, affect wildfires in the direction as we expected, except that the HDD index is positive in the Mixed-STAR model and the Residuals-STAR model. This issue is due to the multi-collinearity problem in these models. With an unstable coefficient and a big variance at most times, hurricanes have no definite impact detected.

As the socio-economic indicator, the employment ratio has the highest elasticity in most models, and is significantly negatively linked to wildfires. However, its coefficient becomes significantly positive in the Residual-STAR model, which is a side-effect caused by the multi-collinearity problem.

In these trans-log regressions, each coefficient represents the related elasticity since all the variables are either percentages or logarithms. It is found that for each independent variable, the elasticity of its impact on wildfire counts is smaller than that on the wildfire burned sizes. This stems from the fact that burned sizes always have bigger values than wildfire counts.

With the values of tolerance and VIF, no multi-collinearity is detected. However, some big condition index values imply that multi-colinearity problems exist sometimes. This issue may have lead to some abnormal values of the coefficients of the HDD index and the employment ratio in the Residuals-STAR model.

The spatial autocorrelation tests suggest that the OLS residuals are still mostly spatial-correlated, and the STAR models almost eliminate the spatial dependence between residuals. The temporal autocorrelation test results for residuals, however, suggest that the temporal dependence may be overestimated. The DW and BG tests results for the OLS regression residuals have shown that in most counties wildfire count residuals are positive correlated, while only in a few counties burned acreage residuals are positively correlated. However, after imposing spatio-temporal autoregressive structures, no counties have significant positive correlated residuals, but around 1/4 – 1/3 of the 67 counties have significant negative correlated residuals. It implies that the STAR models have assumed such a strong homogenous positive temporally dependence for all the counties that the regression residuals in some counties revert to be negatively correlated.

In conclusion, the logarithm transformation of data enhances the model fitness. Most variables showed desired signs, but occasionally the emerged multi-collinearity problems caused some coefficients to be irregular. The forest factors affected wildfires in different ways. The STAR models solved the spatio-dependence problem, but might have over-estimated the temporal dependence.

Wildfire Density and Burned Ratio

In the former two segments, both the wildfire frequency and the wildfire intensity are modeled. Besides those two wildfire attributes, another important topic to study is wildfire densities. Two kinds of fire density measurements (Prestemon et al. 2002) can be used. One is the wildfire count per acre of forestland, which is usually called the "wildfire density", and the other is the wildfire burned ratio, which is the proportion of burned acreage out of the total forestland acreage in a county. These two measurements represent the densities of wildfire frequency and wildfire intensity respectively.

Figure 6.A depicts the percentage distribution of wildfire count per acre and obviously the distribution is not normal. Since this is not a count data, the Poisson model and the Negative-Binomial model are inappropriate. However, after a logarithm transformation, the distribution is muck likely to be normal (Figure 6.B).

The same thing happens to the burned ratio. The burned ratio distribution (Figure 7.A) is very long tailed and is concentrated mostly close to 0. The reason why it has some values beyond 1 is that some wildfires had developed so large that they crossed the county borders. Since in this data the records only have wildfire origins and spread sizes and no specific paths, we accredited all the burned acreages to the originated counties. In this way, some burned ratios could be bigger than 1. Simply dropping them would cause sample shrinking, and setting them to 1 would also cost some useful information. Therefore, we decided to keep these "outliers" since they contain the information about the intensity of those wildfires. A transformation is needed for burned ratio data, and like the wildfire count per acre variable, after the natural logarithm is taken the data is much more like a normal distribution (Figure 7.B). Though the burned ratio, which is already a percentage variable, may be a better choice to use as the dependent, its abnormal distribution forces us to use its logarithm instead.

Therefore, the logarithm of wildfire count per acre and the logarithm of burned ratio will be used as the dependents, and the same set of explanatory variables will be used. However, for consistency, as the dependents are densities, the scale factors will also be changed to densities, i.e. using forestland ratio instead of forestland size, and using the average number of residents per acre of county land instead of the population size. Other factors remain the same forms as in Table 3 and Table 4.

The regression models are as same as those in Table 4, and the results (Table 5) are similar to Table 4. Most predictors have desired signs, but several things should be paid attention to. One scale factor, population density, still has significant a positive coefficient. Somehow, the other scale factor, the forestland ratio, has a significant negative impact on wildfire density and burned ratio. At first, it is thought to be abnormal. However, it is of perfect sense when it is carefully evaluated. On one hand, the negative sign stems from the way how the variables are constructed. For example, the dependent variable wildfire count per acre is the wildfire count divided by forestland size, and the forest ratio variable is the proportion of the forestland size in a county. As the forestland size actually appears on the both sides of the regression equation, the coefficient of the forestland ratio may be distorted. On the other hand, the negative coefficient means that given a fixed land size of each county, if the forestland size is higher in this county, fewer wildfire incidences would happen on each acre of forestland. This explains why a higher forest ratio reduces wildfire densities.

The forestry conditions have similar impacts on both densities as they did on wildfire frequencies and intensities. The coefficient of the private share is always positive, but not significant in the Residuals-STAR model (column 5.D).The impact by the share of longleaf / slash pine forestland & loblolly / shortleaf pine forestland group is insignificant in most models and the sign of its coefficient changes irregularly. The share of oak / pine forestland & oak / hickory forestland group is significantly linked to both wildfire densities in most regressions. Moreover, the oak / gum / cypress forestland share increases the burned ratio while it significantly decreases the wildfire count per acre. This phenomenon confirms our findings that an oak / gum / cypress forestland is not a good place to ignite wildfires, but is helpful to worsen damages.

Like in Table 4, in almost all the regressions of Table 5 the weather factors and the socio-economic conditions have significant coefficients with expected signs, except the HDD index in some scenarios and the employment ratio in the Residuals-STAR model (column 5.D). Again these two cases are both victims of multi-collinearity problems. Multi-collinearity, as implied by the high condition index value, could be a problem in the Mixed-STAR model and Residuals-STAR model. However, another weather factor, hurricane, has no significant impacts on wildfires.

The overall fitness in Table 5 is best among all the three tables in this chapter. Even in the OLS regression, the R-squre is 68% when the dependent is wildfire count per acre and 55% when the dependent is burned ratio, implying that the selected predictors explain the variations of wildfire densities forcefully well.

The spatial autocorrelation tests suggest that the OLS residuals are significantly positively auto-correlated in nearly a half of the total surveyed years. According to Moran's I index and Geary's c index, the STAR models sufficiently alleviate the spatial dependences of residuals. Although the OLS residuals exhibit strong positive temporal autocorrelation in most counties, the imposing of STAR models pushes the regression residuals more inclined to be negatively temporally auto-correlated.

Conclusion

In this part, six regression models and five dependent variables are utilized to study the wildfire behavior. We started by investigating the annual wildfire count. In addition to the OLS regression, the Poisson regression and the Negative-Binomial regression were applied to this count data. In addition, STAR models were adopted to take spatio-temporal dependence into consideration. Moreover, logarithms of wildfire counts and burned acreages were used as dependent variables, but only the OLS model and STAR models are applied. In the final part, densities of wildfire frequency and intensity are also studied.

Generally the model fitness is good enough, especially after the logarithm function of wildfires was utilized. Most independent variables have significant coefficients and some useful wildfire management policy implications have been suggested. Forestland size and population always increase wildfire risks except that higher forestland ratio is linked to a lower wildfire density. A county with a bigger share of privately owned forestland faces a higher wildfire risk. This implies that, in the opposition of the suggestions to better protect national forests (President Report 2000), it is more efficient to decrease wildfires by increasing surveillance efforts on private forestlands.

Different types of forestland groups affect wildfire risks differently. It is indefinite whether the longleaf / slash pine forestland & loblolly / shortleaf pine forestland group can decrease wildfires. Meanwhile, the oak / pine forestland & oak / hickory forestland group is significantly negatively linked to wildfires*(* Although its coefficient is positive sometimes in Table 3, since we prefer the trans-log regression, we follow the conclusions derived from Table 4 and 5). In addition, the oak / gum / cypress forestland can decrease the wildfire frequency and increase the wildfire intensity at the same time. Therefore, an appropriately arranged forestland system can reduce wildfires, i.e. with more coverage of oak / pine forestland & oak / hickory forestland.

Weather conditions have direct influences on the wildfires. A higher drought index, which is equivalent to a wetter ground condition, significantly decreases wildfire risks. A cooler temperate, which is represented by a higher HDD index, decreases wildfires in most cases. The hypothesis about the casual relationship between hurricane and wildfires has been tested. A hurricane incidence has a relatively small impact on wildfires ignitions, and its coefficient doesn't have a consistent sign. However, as its coefficient in the Poisson Regression is significantly positive, it is possible that hurricanes may have enhanced wildfire risks in the same calendar year.

Socio-economic conditions are also closely related to wildfire risks, i.e., it is found that a county with a good record of employment percentage is likely to have less wildfires.

Unfortunately, multi-collinearity exists in some scenarios when the lags of dependents are added into regressions. Sometimes this problem distorted the coefficients of HDD index and employment ratio in the Mixed-STAR models and the Residual-STAR models.

Because wildfires are significantly positive spatio-temporal auto-correlated, STAR models are incorporated into the analysis. As found in the regressions, nearby wildfires and recent wildfires have significantly positive effects. Therefore, wildfire management agencies should be more alert when experiencing a recent “bad” year or being surrounded by “hot spots”. Although the STAR models may have over-estimated the residuals’ temporal autocorrelation, they successfully removed spatial dependence between residuals.

Overall, our analysis makes it possible that wildfire mechanisms are well understood.

ii. Forecast Wildfire Risks

Substantial analysis has been focused on how wildfires originated and how physical and social factors affect wildfires within a same period. The majority of wildfire risks can be explained by those models, and appropriate forest management policy implications have been proposed. However, another even more important question arises, whether wildfire risks can be predicted. If so, precautious advice on wildfires preventions can be offered, and more importantly, a fair insurance scheme can be designed.

For most agricultural insurance contracts such as the all perils output insurance contract, farmers sign the contracts at the beginning of each calendar year, and the insurance covers the subsequent whole year. It is reasonable to assume a same scenario for the wildfire insurance contract. At the beginning of each year, wildfire occurrences, environmental conditions and socio-economic statistics of the past year available, the tasks is to estimate the next year’s wildfires based on those known information. To predict the future wildfires, the following models are used.

Forecast Annual Wildfire Count

In Table 6 all the mentioned forecast models from the section 1.2 are applied when the dependent is annual wildfire count. Although all the independent variables used are lags, the parameter estimations for most of them are still significantly desirable.

First of all, the scale factors, forestland size and population, are always significantly positive. Secondly, the forestry structure matters in predicting wildfires. Private share is significant in worsening the next year's wildfire count, also with the biggest elasticity. The oak / gum / cypress forestland, with significant negative coefficient, can reduce the future wildfires. However, neither the oak / pine forestland & oak / hickory forestland group nor the longleaf / slash pine forestland & loblolly / shortleaf pine forestland group has a consistent effect.

In these prediction models weather conditions behave differently from the current models. Despite the fact that the HDD index is always significantly negative as expected, the coefficient of drought index SP12 flips its signs across different models and its magnitude varies enormously. One interesting thing to notice is that hurricane now plays a negative impact on next year's wildfire count. This is not counter intuitive, and instead it implicitly suggests that the hypothesis of hurricane causing more wildfires right now is true. The reason stems from the truth that the hurricane rarely happens: a surge of hurricane incidences in the past year usually means no chance of hurricane strikes in the coming year. In other words, the hurricanes are temporally negatively correlated. Therefore, an increase of hurricanes in the current year implies a decrease of hurricanes next year, and consequently results in a decrease of wildfires. Moreover, many wildfire agencies have claimed that trees knocked down by the hurricanes may cause wildfires later on. Hence trees removal actions are always taken immediately after hurricanes and the local government usually adds stricter constraints on the wildfire uses such as prescribed fires permissions. Those posterior disastrous steps are likely to put off the wildfire risk in the following year.

The employment ratio, just like it did in the chapter 5, always has a significantly negative coefficient except in the Residuals-STAR model (column 6.f).

According to the elasticities values of the explanatory variables in the Poisson model and the Negative-Binomial model, none of the independents is elastic and the forestland size and the private share are most elastic.

The temporal dependence parameter ρ_t is significantly positive in all the scenarios, which confirms the preliminary analysis. However, the spatial dependence ρ_s is negative in the Mixed-ST model (column 6.b) and Pure-ST model (column 6.c). Since the spatial parameter in these two scenarios is modeled as the impact from the lagged neighboring wildfires in these two models, the negative signs and the relatively small values suggests that this impact is more likely to be indirect. In contrast, the direct spatial impact from neighboring wildfires in the same year (column 6.d, 6.e & 6.f) is always significantly positive.

The model fitness is good enough, with the R-square value of 0.5-0.6, when the spatial-temporal structure is incorprated. However, the R-square values in Table 6 are smaller than those explanatory models in chapter 5 in every corresponding model. This is not surprising because all the independent variables in this part are observations from the lagged year. Those lagged explanatory variables can't explain wildfires as comprehensively as the contemporaneous explanatory variables.

According to the three rules of thumb, multi-collinearity does not exist in Table 6. Moreover, the spatial and temporal autocorrelation of the regression residuals is checked using Moran's *I* index and Geary's *c* index. As long as the STAR models are not used, the positive spatial autocorrelation is strong, even in the Pure-ST and Mixed-ST models (column 6.b & 6.c). This suggests that only modeling the dependence between lagged neighbors' wildfires and current wildfires can't solve the spatial autocorrelation problem. However, the STAR models (column 6.d, 6.e, 6.f) have residuals tend to be more negatively spatially auto-correlated. In the view of the DW test, the positive temporal autocorrelation is strong in the Pooled regressions (column 6.a). After the spatial-temporal dependence structure is applied, the results of BG test (column 6.b-f) indicate that the temporal auto-correlation is almost solved.

Finally, As long as the forecasts for the dependent variables can be derived, it is necessary to do an out-of sample-check. This database has the records of wildfires from 1981 to 2005. The parameter vector β is estimated using the data of 1981 to 2004, and with this estimate the observations of dependent variable in 2005 can be forecasted. Therefore, the sum of the squared differences between forecasts and observed values measures the forecast power to predict the wildfires in 2005, which is reported in the row labeled with "check up: 2005". In a similar way, after the parameter vector is estimated by using the 1981-2003 data, another out-of sample check can be performed with the year 2004's observations. The results are reported in the row labeled with "check up: 2004". These two out-of-sample-checks will be performed in all the prediction models. The out of sample checks suggest the Pure-STAR model (column 6.d) with the forecast (1.61) bring in the best prediction of the year 2005, and the Negative-Binomial model (column 6.a) has the best prediction of the year 2004. This is a little surprising, but may be due to the idiosyncratic data structure in the forecasted year.

Overall, the independents behave closely like they did in chapter 5, because all those factors fluctuate merely from year to year. Hurricanes variable is now always significant with a negative sign. Because the ST models didn't model spatial dependence directly, in this case the results of STAR models are more favorable.

Forecast Annual Wildfire Count with Natural Logarithm Transformation

Due to the fact that the distribution of the logarithm of wildfire counts looks much likely to be normal, the trans-log form of wildfire counts is chosen to be the dependent variable. Moreover, since this is not a count data, the Poisson Model and Negative-Binomial model are dropped in this part.

From the results, we found similar conclusions with the former segment. The scale factors are significantly positive, and the employment ratio significantly decreases next year's wildfire count except in the Residuals-STAR model (column 7.f). Among the forestry conditions, the share of privately owned forestland has a significantly positive impact and the share of oak / gum / cypress forestland group has a significantly negative impact on the wildfires. Somehow the impacts by the group of longleaf / slash pine forestland & loblolly / shortleaf pine forestland and the group of oak / pine forestland & oak / hickory forestland are ambiguous. As weather conditions, analogous to the results in Table 6, in Table 7 the coefficient of HDD index and hurricane are significantly negative while the coefficient of Drought Index SP12 is unstable. Employment ratio can significantly decrease wildfires, except in the Residual-STAR model (column 7.f). The multi-collinearity problem exists if the spatial-temporal dependence model is applied. It is because that the lagged independents must have a strong linear relation with the lagged dependent, as we found in chapter 5.

After the dependent is taken a logarithm of, the model fitness is improved compared with Table 6, though not as well as the current analysis models (Table 4). Again, spatial autocorrelation can only be eliminated after STAR models (column 7.d, 7.e & 7.f) are imposed. In the ST models (column 7.b & 7.c), the spatial dependence parameter ρ_S is not stable and the residuals are still positively spatially auto-correlated. The positively temporal autocorrelation is diminished after the spatio-temporal structure is applied (column7.b-f), but may have been over-estimated by the use of ρ_t because the residuals in those models tend to be negatively temporal correlated (BG test).

The out-of-sample-checks suggest that the Poisson model (column7.a) and the estimator (1.61) in Pure-STAR model (column 7.d) have the best forecasts for the year 2004 and 2005 respectively, but this may be solely due to the data idiosyncratic attributes in these two years.

Forecast Annual Burned Acreage with Natural Logarithm Transformation

All the conclusions drawn from Table 8 are exactly same as in Table 7 except two forestry conditions. Figure 5 shows that the logarithm of burned acreage is more appropriate to use as the dependent variable, and the same models as in Table 7 are applied. The share of the oak / pine forestland & oak / hickory forestland group now consistently decreases the next year's burned acreage. Similar to what we found in chapter 5, the oak / gum / cypress forestland helps to enhance the future annual burned acreage significantly.

The out of sample check indicate that Poisson model (column 8.a) and the estimator (1.61) in Pure-STAR model (column 8.e) provides the best forecasts of 2004 and 2005 respectively in this case.

Forecast the Logarithm of Wildfire Count Per Acre

To predict the density of wildfire frequency, the logarithm of wildfire count per acre is adopted as the dependent in view of the fact that the tans-log form is close to the normal distribution.

Most results in Table 9 are similar to those in Table 6, which is used to predict the logarithm of wildfire count. However, several things need to pay attention. The oak / pine forestland & oak / hickory forestland group now consistently decreases the next year's wildfire frequency density, while it didn't in Table 7. Beside the SP12 index, now the HDD index is inconsistent on how to affect the wildfire frequency density. The forest ratio has a negative impact on the wildfire density and the reason for the coefficient to be negative is as same as that in chapter 5. That is, the more forestland one county has, the lower chance each acre gets burned. The model fitness is fairly high, with the R-square of 0.6-0.8.

The out of sample check shows that the Poisson model (column 9.a) and the estimator (1.61) in Pure-STAR model (column 9.d) have the best forecasts for 2004 and 2005 respectively.

Forecast the Logarithm of Annual Burned Ratio

At last, the burned ratio is studied, which is the main concern of our interest. If the future burned ratio could be predicted appropriately, a method to design an actuarially fair wildfire risk insurance scheme is established.

Except that the multi-collinearity problem is not severe at all in this part, the results in Table 10 are similar with Table 8. The out-of-sample-checks suggest the best forecasts for 2004 and 2005 are provided by the Poisson model (column 10.a) and the estimator (1.61) in the Pure-STAR model (column 10.d) respectively. Since the R-square is as good as at least 0.6 across the models, it provides a reliable forecast method after all.

Conclusion on the Prediction Analysis

In this section several statistical models are adopted to predict future wildfires. The underlying idea is to regress current wildfires on the lagged explanatory variables. If the results are good enough, current information can be used to forecast next year's wildfire risks.

In the view of model fitness, the logarithm transformation is indeed helpful. In the regressions, most explanatory variables have significant impacts on future wildfires. Moreover, most of the coefficients of the independents have same signs as in chapter 5, because forestry conditions, socio-economic status and weather factors don't change a lot from year to year.

The forestland size and population always significantly increase wildfires, but the forestland ratio significantly decreases wildfire densities. Among forestry conditions, the private share of forestland has a positive impact on wildfires. The oak / gum / cypress forestland can reduce wildfire frequency as well as increase wildfire intensity.

As a weather condition, a high HDD index, which is equivalent to a cool year, can significantly decrease the next year's wildfires, and an outbreak of hurricane implies a low wildfire risk next year. The employment ratio can significantly reduce wildfires except in the Residuals-STAR model.

The fitness of this model is substantially well although sometimes the multi-collinearity problem exists. The spatio-temporal models alleviated the positive temporal autocorrelation of the residuals. But they may also over have emphasized the temporal dependence so that the residuals are inclined to be negatively temporally auto-correlated. The ST models and STAR models have incorporated the spatial dependence in different ways. The ST models modeled spatial dependence ρ_S as the dependence between lagged neighbors' wildfires and the current wildfires. Meanwhile, the STAR models modeled took the spatial dependence ρ_S as the dependence between counties in the same period. The results suggest that ρ_S in the ST model is not stable and the residuals are still positively spatial auto-correlated. Therefore, the spatial dependence between lagged neighbors' wildfire and current wildfires is very likely to be indirect and the STAR models are preferred.

The out-of-sample-checks for the year 2004's wildfires and the year 2005's wildfires didn't provide a universally best forecast estimator, most likely due to the idiosyncrasy of the observations in each year.

4. Concluding Remarks

In this chapter, wildfires' behavior is studied and several econometric models are applied to explain wildfire mechanism and to predict wildfire risks. Wildfire causes have been analyzed when forestry conditions, socio-economic factors and weather are taken into consideration. More importantly, it is found that wildfires are both spatial and temporal auto-correlated.

With the wildfire data in the Florida state, several models, including pooled regressions such as OLS and Poisson and Negative Binomial models and spatio-temporal autoregressive models are used. In section 1.3 how those explanatory variables influence the wildfires in the same year are analyzed. In order to do forecasts, the lagged explanatory variables are adopted to explain the current wildfires.

In most cases, the independent variables are influential and the spatio-temporal auto-regressive structure is needed. Generally, the influences of most independent variables are identified and found to be significant, i.e. the oak / gum / cypress forestland decreases the wildfire frequency as well as increases the intensity. It is also verified that hurricane strikes increase wildfires in the same calendar year, but significantly decreases wildfire risks in the subsequent year. Therefore, a number of forestry management policy suggestions and precautious actions are proposed.

Though sometimes the multi-collinearty problem exists, the overall fitness of the models suggests that those models are sufficiently well to explain and forecast the wildfires.

Meanwhile, cross sample validation verifies that our method can forecast wildfire risks adequately well. Since our approach does not incorporate any fixed-effect indicator or trend as in the panel data analysis (Prestemon et al. 2002), it offers a universal tool to evaluate and predict wildfire risks. Hence, given environmental information of a location, a corresponding actuarially fair insurance rate can be calculated.

However, as we found in the spatial-temporal tests for the residuals and out of sample checks, the idiosyncratic characteristic of the data raises a challenge to find a universally perfect model. Therefore, the robustness of the theory needs to be further looked into.

In this chapter, wildfire risks are calibrated and an insurance plan is designed by measuring the burned ratio. However, one potential problem arises from the theoretical assumption that each wildfire will cause a total loss. In reality, the loss ratio is least likely to be homogeneous. One solution is to figure out the loss density distribution associated with the wildfire intensity. As long as both of these two factors can be measured accurately, a perfectly actuarial fair wildfire insurance scheme should be executed.

Reference

Andies S. Brandsma and Ronald H. Ketellapper, 'Further evidence on alternative procedures for testing of spatial autocorrelation among regression disturbances', Chapter 5 of *Exploratory and explanatory statistical analysis of spatial data*, Bartels and Ketellappers, 1979

Bartels and Ketellappers, *Exploratory and explanatory statistical analysis of spatial data*, 1979

Conelis P.A. Bartels, 'Operational statistical methods for analyzing spatial data', chapter 2 of *Exploratory and explanatory statistical analysis of spatial data*, Bartels and Ketellappers, 1979

Cressie, *Statistics For Spatial Data*, 1993

Dubbelmaan C., Abrahamse J., Koert J.C., "A new class of disturbance estimators in the general linear model", *Statistica Neerlandica*, Volume 26, Issue 4, Date: December 1972

Kent, B., K Gebert, S. McCaffrey, W. Martin, D. Calkin, E. Schuster, I. Martin, H.W. Bender, G. Alward, Y. Kumagai, P.J. Cohn, M. Carroll, D. Williams, and C. Ekarius. "Social and Economic Issues of the Hayman Fire." Hayman Fire Case Study, *USDA Forest Service General Technical Report* RMRS-GTR-114 (Revision), R.T. Graham, ed., p. 315-95. Fort Collins, CO: USDA Forest Service, 2003. 17

Becker, G.S. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76(1968):169–217.

Fisher, F., and D. Nagin. “On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rates and Sanction Levels.” In A. Blumsten, J. Cohen, and D. Nagin, eds. *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, DC: National Academy of Sciences, 1978, pp. 361-99.

Cameron, C.A., and P.K. Trivedi. *Regression Analysis of Count Data*. Cambridge: Cambridge University Press, 1998.

Marvell, T., and C. Moody. “Specification Problems, Police Levels, and Crime Rates.” *Criminology* 34(1996):609-46.

Eck, J., and E. Maguire. “Have Changes in Policing Reduced Violent Crime? An Assessment of the Evidence.” *The Crime Drop in America*. A. Blumstein and J. Wallman, eds., pp. 207-65. New York: Cambridge University Press, 2000.

Butry D. & Prestemon J. “Spatio-Temporal Wild land Arson Crime Functions”, selected paper prepared for presentation at the *American Agricultural Economics Association Annual Meeting*, Providence, Rhode Island, July 24-27, 2005

Prestemon, J.P., and D.T. Butry. *“Time to Burn: Modeling Wildland Arson as an Autoregressive Crime Function.”* American Journal of Agricultural Economics, 2005, 756-770 “Managing the Impact of Wildfires on Communities and the Environment: A Report to the President In Response to the Wildfires of 2000”, September 8, 2000

Prestemon JP, Pye JM, Butry DT, Holmes TP, Mercer DE, “Understanding broad scale wildfire risks in a human-dominated landscape”, *Forest Science* 48, 2002, 685–693.

O'Brien, Robert M. "A Caution Regarding Rules of Thumb for Variance Inflation Factors," *Quality and Quantity* (2007): 41(5)673-690.

Gordon, R. A.. “Issues in multiple regression”, *American Journal of Sociology* (1968) 73:592–616.

Gordon, R. A., “Citation-classic – issues in multiple regression”. Current Contents/*Social and Behavioral Sciences* (1987) 36: 18–18.

Gujarati, D.N., *Basic econometrics*, 4th ed., (2003) Boston, McGraw-Hill

Durbin, J., and Watson, G. S. "Testing for Serial Correlation in Least Squares Regression, I." *Biometrika* (1950) 37: 409–428.

Durbin, J., and Watson, G. S. "Testing for Serial Correlation in Least Squares Regression, II." *Biometrika* (1951) 38: 159–179.

Van den Poel Dirk, Larivière Bart, “Attrition Analysis for Financial Services Using Proportional Hazard Models”, *European Journal of Operational Research* (2004),157 (1): 196-217.

Upton, G. and B. Fingleton, *Spatial Data Analysis By Example*, ,1990

Belsley, Kuh & Welch, *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity* , 1980

O'Brien, Robert M. *"A Caution Regarding Rules of Thumb for Variance Inflation Factors,"* Quality and Quantity (2007) 41(5)673-690.

Countryman, C. “Fuel moisture and fire danger - some elementary concepts.” *USDA Forest Service Region 5, Aviation and Fire Management* (1972) 4p.

Cameron, A.C. and P.K. Trivedi, *Regression analysis of count data, Cambridge University Press*. (1998). ISBN 0-521-63201-3

Hilbe, J.M., *Negative Binomial Regression* (2007), Cambridge University Press. ISBN 978-0-521-85772-7

P. Whittle, “On Stationary Processes in the Plane, Biometrika”, Vol. 41, No. 3/4 (Dec., 1954), pp. 434-449

Breusch, T.S., “Testing for autocorrelation in dynamic linear models”, *Australian Economic Papers*(1978), 17: 334–355. doi: 10.1111/j.1467-8454.1978.tb00635.x

Godfrey, L. G., “Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables”, *Econometrica*, Vol. 46, No. 6 (Nov., 1978), pp. 1293-1301

Durbin, J.: "Testing for Serial Correlation in Least Squares Regression When Some of the Regressors are Lagged Dependent Variables," *Econometrica*, 38 (1970), 410-421

Marc Nerlove and Kenneth F. Wallis, “Use of the Durbin-Watson Statistic in Inappropriate Situations”, *Econometrica*, Vol. 34, No. 1 (Jan., 1966), pp. 235-238

Hashem Dezhbakhsh, “The Inappropriate use of Serial Correlation Tests in Dynamic Linear Models”, *The Review of Economics and Statistics*, Vol. 72, No. 1 (Feb., 1990), pp. 126-132

Greene, William H., *Econometric Analysis*, 5th edition

Cliff, A. D. and Ord, J. K., “Model Building and the Analysis of Spatial Pattern in Human Geography”, *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 37, No. 3 (1975), pp. 297-348

Cliff, A.D. and Ord, J.K., The choice of a test for spatial autocorrelation, pp. 54-77 in *Display and analysis of spatial data* (Davis, J.C. and McCullagh, M.J., Eds) Wiley: Chichester, 1975

Cliff, A.D. and Ord, J.K., *Spatial Processes: Models and Applications*. Pion: London.1981

Williams, G.W. *Aboriginal use of fire: are there any "natural" communities?* Washington, DC: U.S. Department of Agriculture, Forest Service, Washington Office p. 2002.

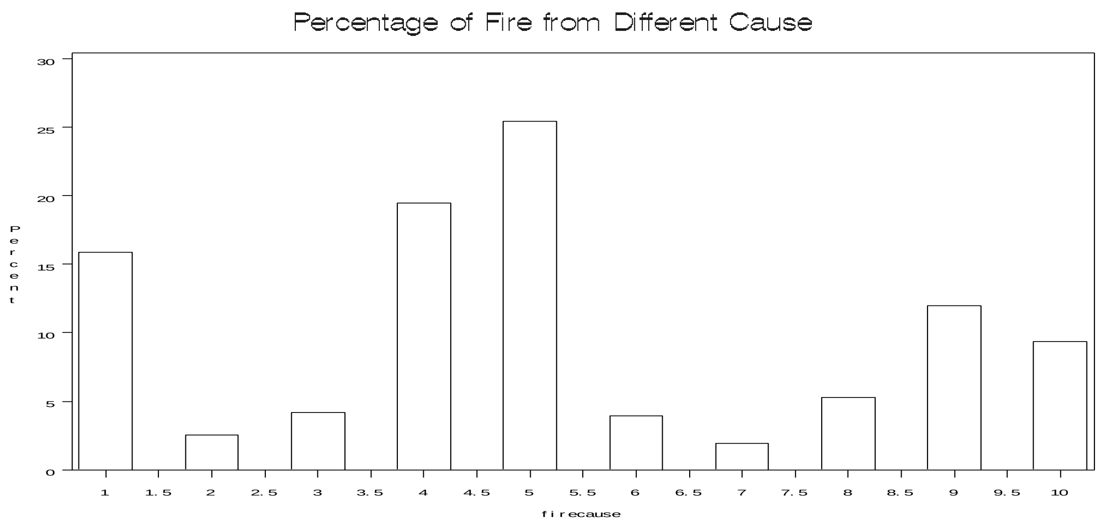


Figure 1.A. Causes: 1=lightning; 2=campfire; 3=cigarette; 4=debris burning; 5=incendiary/arson; 6=equipment; 7=railroad; 8=children; 9=unknown; 10=misc

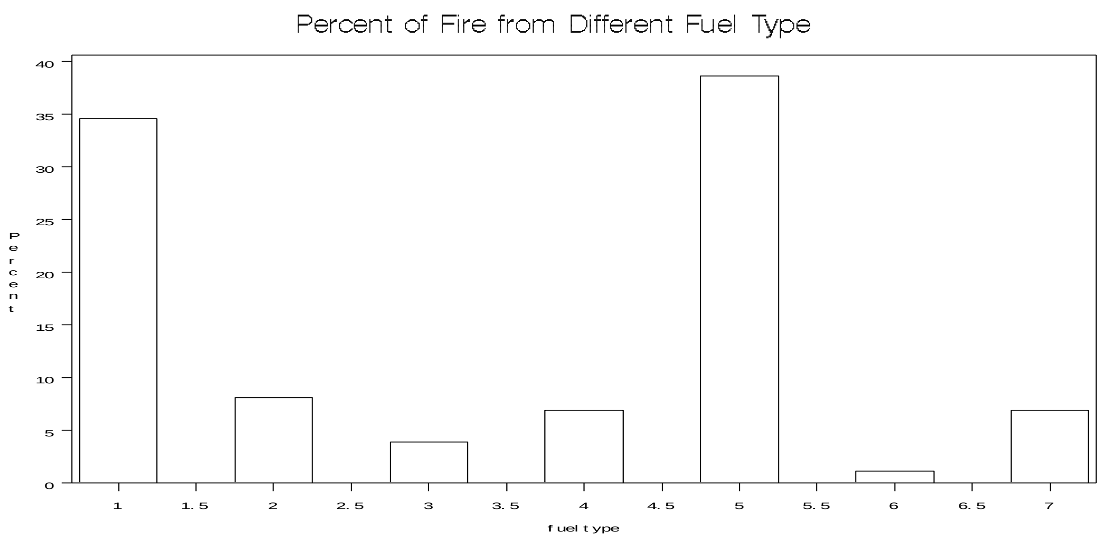


Figure 1.B. Fueltype: 1=palmetto-gallberry; 2=dense pine (fire in crown); 3=swamp; 4=blowy leaf (eg. turkey oak); 5=grassy fuels; 6=muck (organic soils); 7=other

Figure 1. Wildfires Decomposition by different causes and fuel types

Table 1. Definition and statistics of variables

Variable	Definition	N	Mean	Std. Dev.
Dependent Variables				
Wildfire count	Yearly wildfire count in each county	1675	77.7665672	67.52843
Log(wildfire count)	Log of Yearly wildfire count in each county	1675	4.027236	0.84778
Log(wildfire count per acre)	Log(wildfire incidence per acre of forestland in each county)	1675	-8.11495	1.03162
Burned acreage	Annual burned acreage in each county	1675	2977.11869	11471
Log(burned acreage)	Log(Annual burned acreage in each county)	1675	6.517001	1.61070
Log (burned ratio)	Log (Annual burned acreage in each county/ total acreage of forestland in each county)	1675	-5.62518	1.87544
Independent variables				
Forestland (in 10,000 acres) [1]	Size of forestland in each county (in 10,000 acres)	1742	2.458959	1.54222
Log(forestland) [1]	Log(size of forestland in each county, acres)	1742	12.14218	0.83690
Forestland ratio	Size of forestland/ size of the county	1742	0.516821	0.28138
Private owned share of forestland	Privately owned forestland share of total forestland in the county	1742	0.743458	0.25859
longleaf / slash pine & loblolly / shortleaf pine share	Aggregate shares of longleaf / slash pine group and loblolly / shortleaf pine group of total forestland in the county	1742	0.414437	0.18408
Oak / pine & oak / hickory share	Aggregate shares of oak / pine group and oak / hickory group of the total forestland in the county	1742	0.207606	0.12438
Oak / gum / cypress share	Share of Oak / gum / cypress group of the total forestland in the county	1742	0.220949	0.12413
Daily average of HDD Index [1]	Total of HDD index for a year divided by 365	1742	2.792059	1.42474
December SP12 index	Probability of observing a given amount of precipitation in this year	1742	0.253104	0.97988
Hurricane incidences	Annual hurricane incidences in each county	1742	0.181975	0.47771
Population (in 10,000)	Size of population (in 10,000 heads)	1742	2.074341	3.53070
Log(population) [1]	Log(population in heads)	1742	11.19606	1.47601
Log(population density)	Log(population in heads/total forestland size in acres)	1742	-1.86473	1.35473
Employment ratio [2]	Employed workers/total population size	1742	0.426678	0.11656

[1] We divided population and forestland size by 10,000 and sum of HDD by 365 so as to avoid the case that coefficients are too small (less than 0.000001) to repoart, but we use original measurement when we take natural logarithm of them.

[2] Instead of the traditional definition which is percentage of total workforce who are unemployed and are looking for a paid job, employment rate here is defined as employed workers divided by total population.

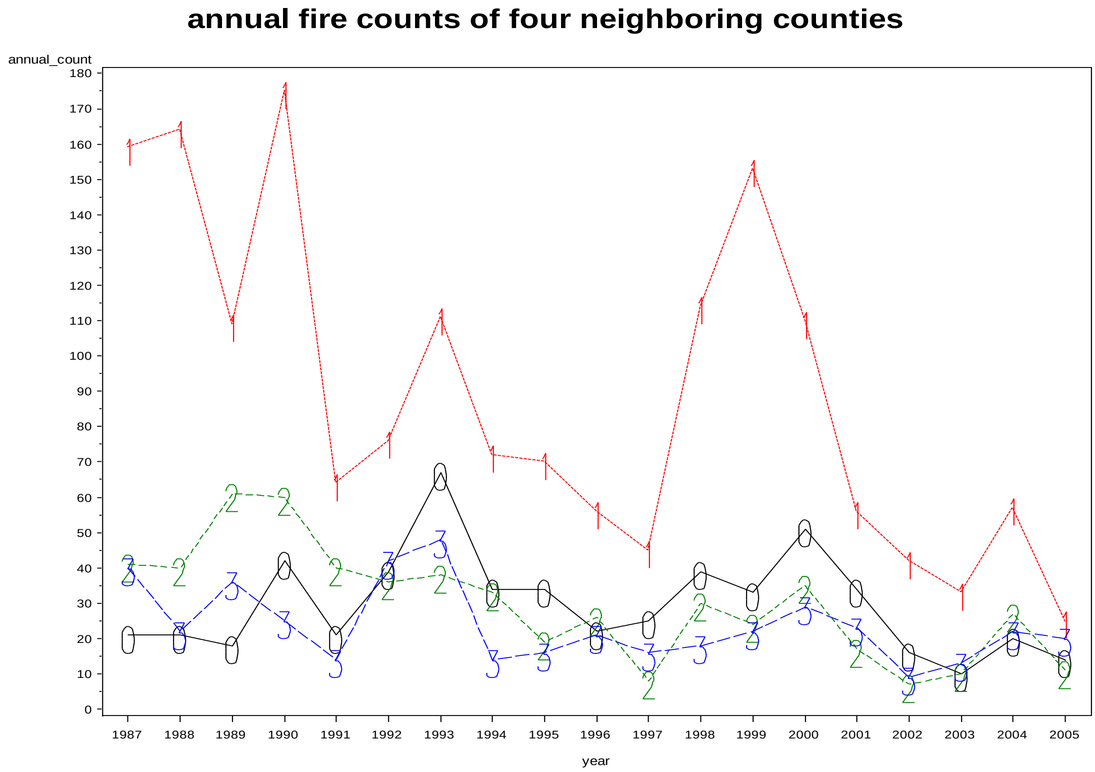


Figure 2. Wildfire counts in 4 neighboring counties

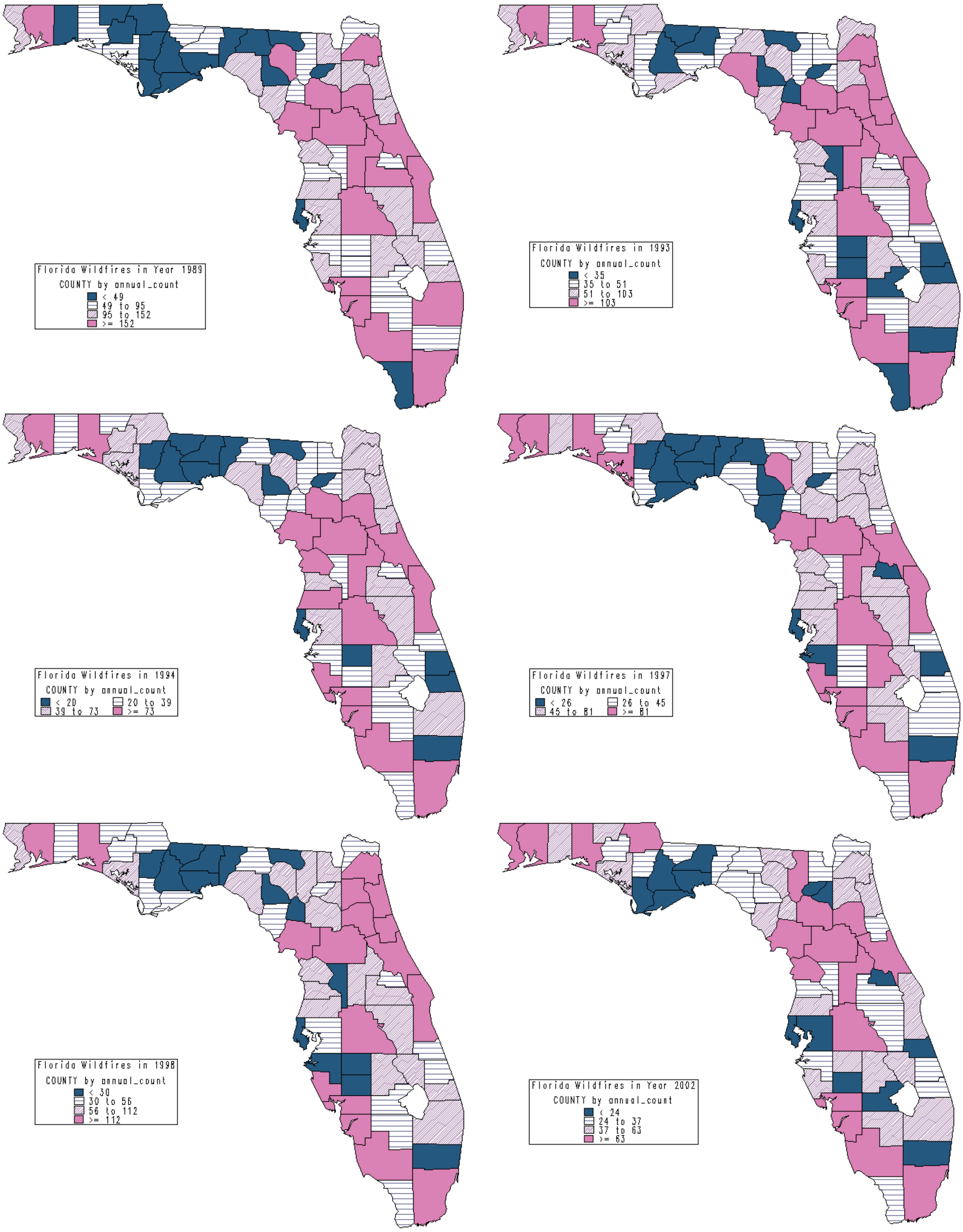


Figure 3. Florida wildfire counts in 6 selected years

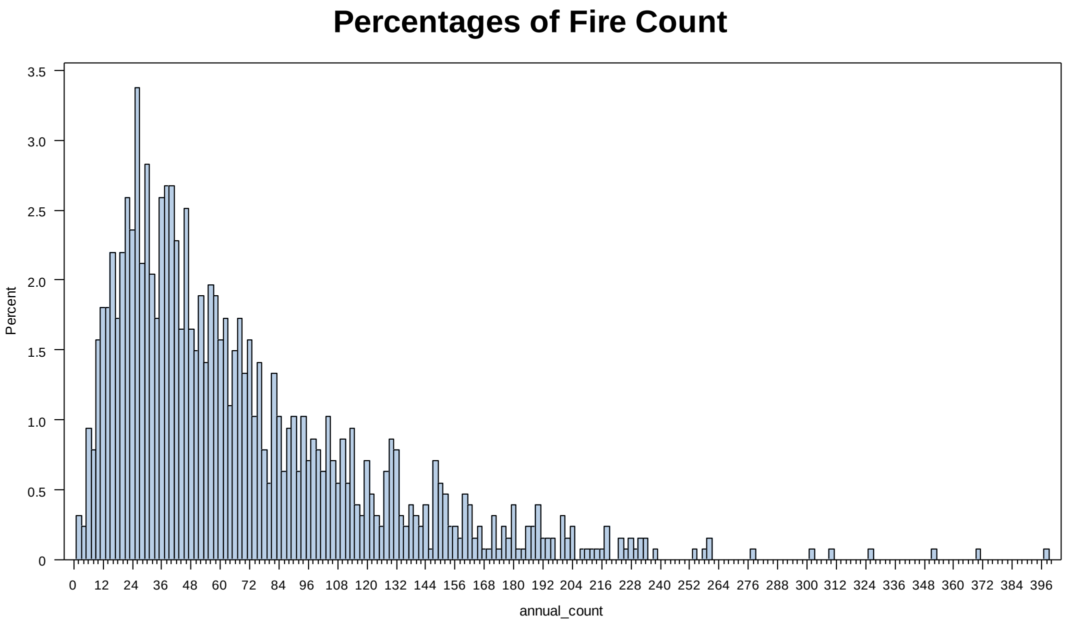
Table 2: Spatial autocorrelation tests with P values for annual wildfire count

Year	Moran's I index - P value	Geary's c index-P value
1981	0.007007	0.366366
1982	0.2672673	0.8558559
1983	0.6256256	0.965966
1984	0.1621622	0.9309309
1985	0.004004	0.7717718
1986	0.1201201	0.6936937
1987	0.041041	0.3453453
1988	0.043043	0.6866867
1989	0	0.034034
1990	0.029029	0.6756757
1991	0.004004	0.2282282
1992	0.025025	0.4824825
1993	0.001001	0.027027
1994	0	0.1261261
1995	0.004004	0.1871872
1996	0.049049	0.3793794
1997	0.005005	0.04004
1998	0	0.016016
1999	0.029029	0.2672673
2000	0.001001	0.0540541
2001	0.032032	0.5855856
2002	0	0.0800801
2003	0.016016	0.1511512
2004	0.003003	0.0610611
2005	0.049049	0.3493493

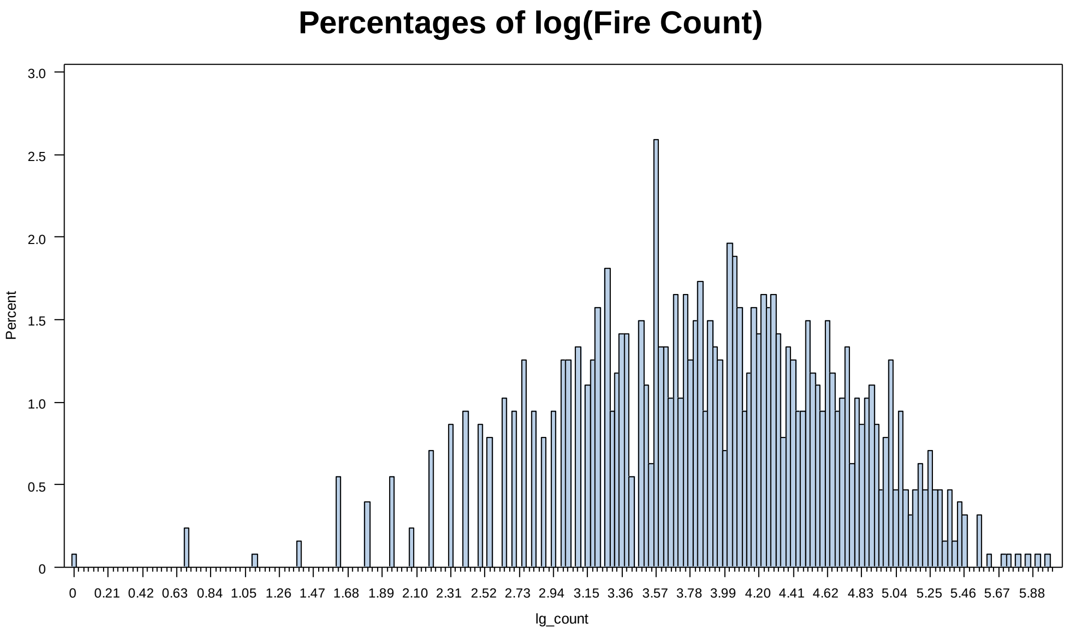
Table 3. Spatial and temporal autocorrelation test

Dependent variable		Fire count	Burned acreage	Log(annual count)	Log(acreage)	Log(Fires per acre)	Log(Burned ratio)
Moran	Positive	21* 21** 12***	23* 20** 14***	21* 21** 12***	23* 20** 14***	25* 25** 24***	22* 20** 16***
	Negative	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Years			25			
Geary	Positive	7* 4** 0***	13* 5** 1***	7* 4** 0***	13* 5** 1***	25* 24** 22***	12* 9** 5***
	Negative	2* 1** 0***	0* 0** 0***	2* 1** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Years			25			
DW	Positive	63* 56** 37***	32* 23** 9***	60* 53** 41***	37* 28** 11***	59* 53** 44***	36* 31** 15***
	Negative	0* 0** 0***	0* 0** 0***	0* 0** 0***	1* 0** 0***	0* 0** 0***	1* 0** 0***
	Counties			67			

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.



4.A



4.B

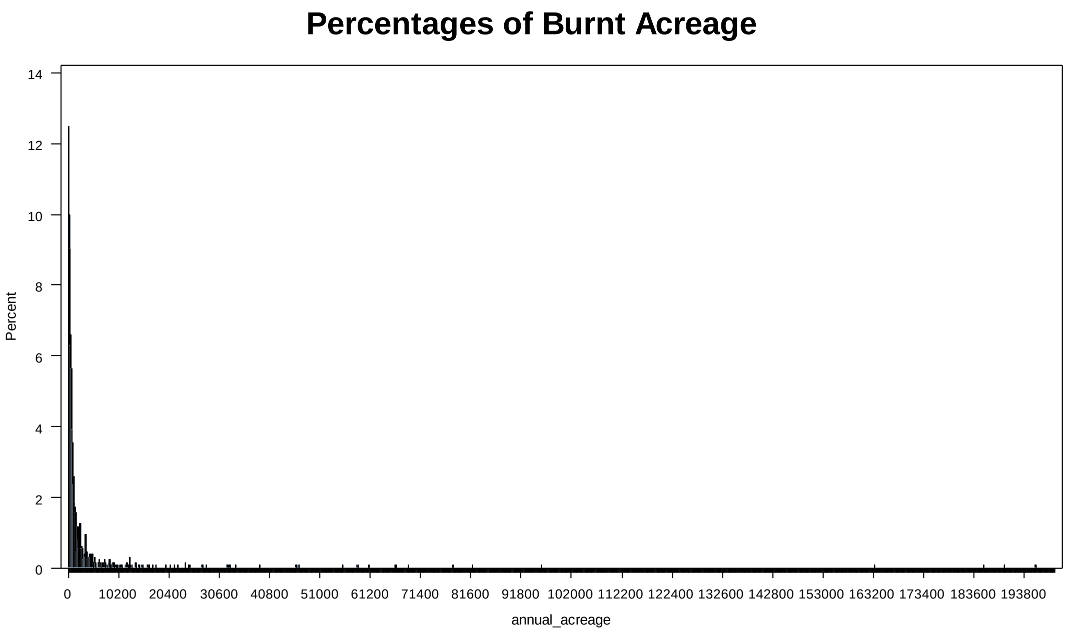
Figure 4. Distributions of wildfire count and logarithm(wildfire count)

Table 3. Different regression models for wildfire count

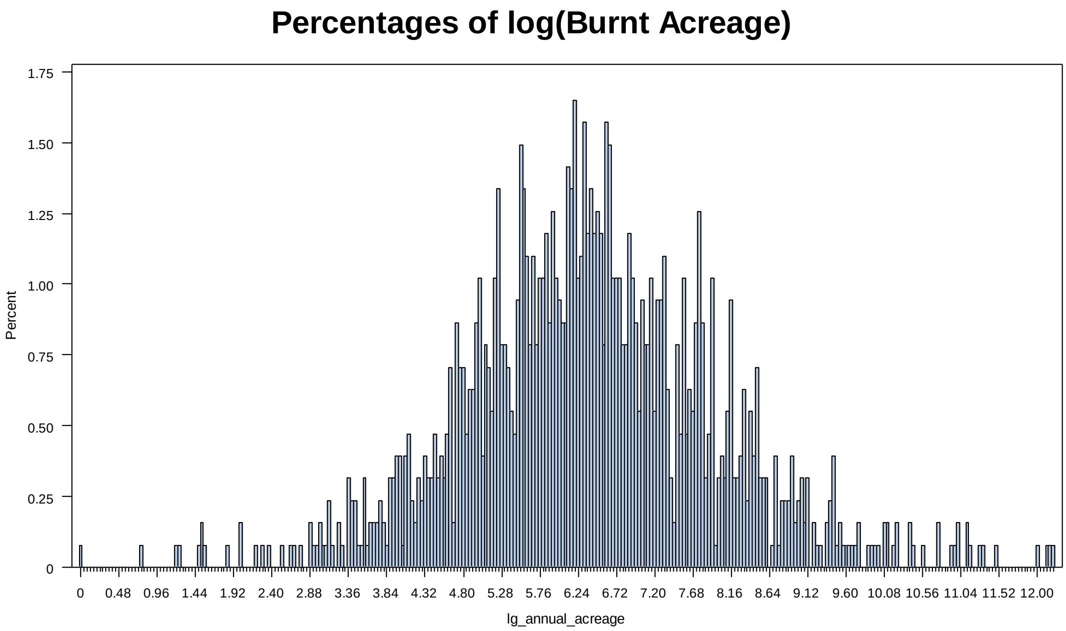
Table	(3.a)					(3.b)	(3.c)	(3.d)
Model	**OLS Model**	**Poisson Model**		**Negative Binomial Model**		$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t = x_t\beta + \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t - x_t\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_t\beta)$
Parameter	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Elasticity[1]**	**Estimate (Std. Error)**	**Elasticity[1]**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t						0.53002*** (0.01432)	0.48721*** (0.01548)	0.533683*** (0.00807)
ρ_s						0.40934*** (0.02448)	0.32610*** (0.02652)	0.400541*** (0.0197)
Intercept	25.84599*** (9.85694)	3.6282*** (0.0195)		3.7893*** (0.1097)		-0.80831 (2.14588)	-13.79226** (6.70002)	-25.9446 (31.8432)
Forestland (in 10,000 acres)	19.20304*** (1.16463)	0.2098*** (0.0020)	0.5155401	0.2314*** (0.0138)	0.5687887		6.79788*** (0.82161)	18.30473*** (1.5536)
Private owned share of forestland	69.96086*** (8.92214)	0.8866*** (0.0185)	0.6566387	0.8628*** (0.1076)	0.6400208		9.62310* (5.74662)	19.13293* (13.8456
longleaf / slash pine & loblolly / shortleaf pine share	-29.10498** (14.06510)	-0.3945*** (0.0275)	-0.161534	-0.4032** (0.1713)	-0.1659		-5.41713 (9.12389)	-48.3589* (25.5486)
Oak / pine & oak / hickory share	2.86475 (13.57592)	0.0772*** (0.0265)	0.0170621	0.0109 (0.1546)	0.0026999		19.84621** (8.67917)	-26.1469 (26.8554)
Oak / gum / cypress share	-6.24797 (15.04022)	-0.0861*** (0.0298)	-0.018092	-0.2895* (0.1750)	-0.063647		-10.95672 (9.54568)	-69.6704**** (26.9592)
Daily average of HDD Index	-11.00835*** (1.61592)	-0.1106*** (0.0032)	-0.3091	-0.1502*** (0.0182)	-0.419518		-0.53304 (1.15255)	12.14503*** (1.9405)
December SP12 index	-23.4624*** (1.45450)	-0.2797*** (0.0027)	-0.070521	-0.2914*** (0.0165)	-0.073739		-9.91210*** (1.05098)	-14.3669*** (1.0198)
Hurricane incidences	2.06111 (6.80376)	0.0327*** (0.0135)	-0.001386	0.0248 (0.0792)	-0.000087		-2.47853 (4.22363)	-4.29628 (4.8533)
Population (in 10,000)	5.19468*** (0.59446)	0.0625*** (0.0010)	0.129779	0.0574*** (0.0076)	0.119019		1.99316*** (0.38390)	4.351747*** (0.6889)
Employment ratio	-20.23241 (14.33557)	-0.2806*** (0.0280)	-0.119223	-0.3148* (0.1719)	-0.134162		-0.46723 (8.98291)	5.987855*** (1.2623)
Scale/Dispersion		1.0000		0.4195 (0.0143)				
Min. Tolerance	0.30090					0.94309	0.28191	
Max. VIF	3.32338					1.06035	3.54722	
Max. Con. Index	21.06790					4.74331	24.48811	
R-Square	0.2640					0.5622	0.6105	0.6556
Moran Pos.	21* 15** 11***	19* 16** 12***		20* 18** 15***		2* 2** 2***	5* 4** 3***	2* 2** 2***
Moran Neg.	0* 0** 0***	0* 0** 0***		0* 0** 0***		15* 15** 8***	6* 5** 2***	10* 5** 4***
Geary Pos.	8* 4** 2***	10* 6** 2***		10* 4** 2***		3* 2** 1***	4* 2** 0***	1* 1** 0***
Geary Neg.	2* 2** 0***	2* 2** 0***		2* 1** 0***		15* 14** 11***	11* 8** 3***	10* 8** 4***
years	25	25		25		24	24	24
DW Pos.	41* 33** 19***	42* 36** 22***		41* 34** 21***				
DW Neg.	0* 0** 0***	0* 0** 0***		1* 0** 0***				
BG Pos.	10* 7** 3***	12* 9** 4***		12* 9** 4***		2* 1** 0***	3* 2** 0***	2* 0** 0***
BG Neg.	0* 0** 0***	1* 1** 0***		1* 1** 0***		4* 2** 0***	3* 1** 0***	9* 4** 1***
counties	67	67		67		67	67	67

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Elasticities are evaluated at the mean values of explanatory variables. For both Poisson and Negative binomial models, the elasticity is given by $\bar{X}_k \beta_k$.



5.A



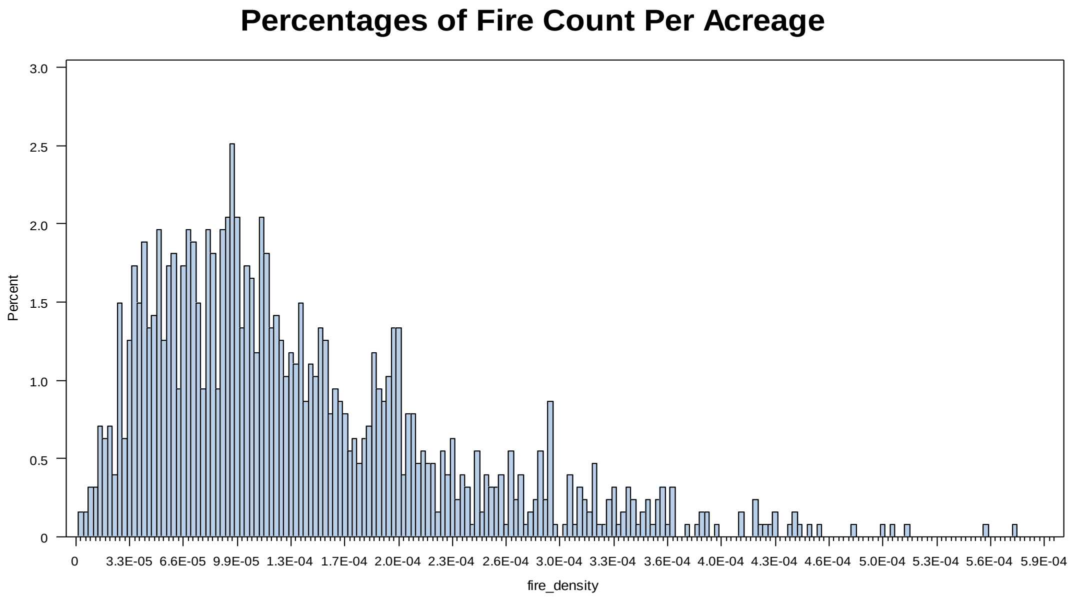
5.B.

Figure 5. Percentage distributions of annual burned acreage and its logarithm

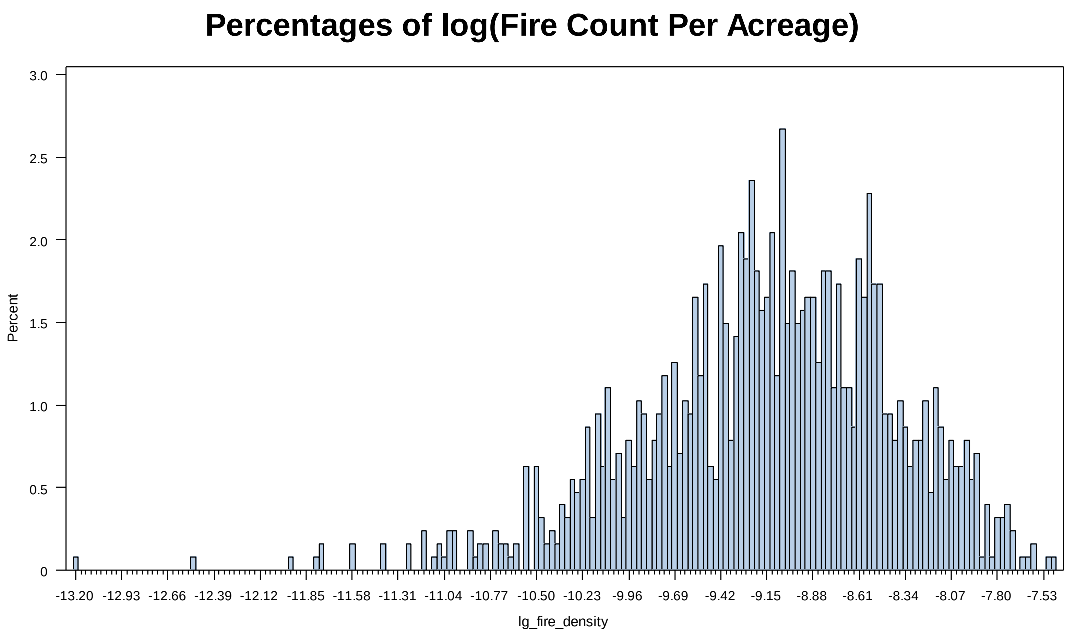
Table 4. Different Models to regress logarithms of wildfire count and burned acreage

Table		(4.a)		(4.b)		(4.c)		(4.d)	
Model		$y_t = x_t\beta$		$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$		$y_t = x_t\beta + \rho_s w_s y_t + \rho_t y_{t-1}$		$y_t - x_t\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_t\beta)$	
Dependent Variable		**Log (Annual Wildfire Count)**	**Log(Annual burned acreage)**	**Log (Annual Wildfire Count)**	**Log(Annual burned acreage)**	**Log (Annual Wildfire Count)**	**Log(Annual burned acreage)**	**Log (Annual Wildfire Count)**	**Log(Annual burned acreage)**
Parameter		**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t				0.67268*** (0.01532)	0.47445*** (0.01914)	0.57495*** (0.01817)	0.40473*** (0.02144)	0.623657*** (0.0153)	0.431506*** (0.0143)
ρ_s				0.38869*** (0.02271)	0.43309*** (0.02426)	0.25946*** (0.02308)	0.36484*** (0.03174)	0.384161*** (0.0202)	0.526584*** (0.0264)
Intercept		-6.40046*** (0.34207)	-3.66092*** (0.71693)	-0.37525*** (0.09476)	0.30288* (0.17697)	-2.97406*** (0.29107)	-4.06166*** (0.67154)	-27.9071 (47.4404)	4.852337 (11.5053)
Log(Forestland)		0.57919*** (0.02360)	0.39003*** (0.04946)			0.19174*** (0.02101)	0.17900*** (0.04501)	0.58103*** (0.0532)	0.468562*** (0.0730)
Private owned share of forestland		1.37143*** (0.09349)	1.12269*** (0.19594)			0.31326*** (0.07392)	0.38185** (0.17508)	0.389999** (0.1732)	0.445987 (0.3336)
Longleaf / slash pine & loblolly / shortleaf pine share		-0.42725*** (0.14471)	-0.31985 (0.30330)			-0.16994 (0.10936)	0.19668 (0.27478)	0.08652 (0.3351)	0.367765 (0.5651)
Oak / pine & oak / hickory share		-0.91321*** (0.14653)	-2.71642*** (0.30710			-0.14007 (0.10992)	-0.83585*** (0.27952)	-0.46911 (0.3933)	-1.71897** (0.7247)
Oak / gum / cypress share		-0.60064*** (0.16042)	0.88070*** (0.33623)			-0.20714* (0.11908)	0.58914** (0.29881)	-0.30745 (0.3426	1.132981* (0.5969)
Daily average of HDD Index		-0.14725*** (0.01783)	-0.21694*** (0.03736)			-0.0008739 (0.01481)	-0.0008099 (0.03680)	0.095016*** (0.0223)	0.038008 (0.0720)
December SP12 index		-0.29373*** (0.01557)	-0.45961*** (0.03263)			-0.14543*** (0.01312)	-0.20425*** (0.03209)	-0.17867*** (0.0135)	-0.21917*** (0.0409)
Hurricane incidences		0.07203 (0.07284)	-0.04577 (0.15267)			0.00225 (0.05295)	-0.21100 (0.13276)	-0.02379 (0.0463)	-0.22526* (0.1362)
Log (Population)		0.37148*** (0.01609)	0.64565*** (0.03373)			0.12768*** (0.01356)	0.33380*** (0.03282)	0.258266*** (0.0255)	0.501272*** (0.0409)
Employment ratio		-1.87775*** (0.17059)	-3.27142*** (0.35754)			-0.51562*** (0.12988)	-1.63149*** (0.32104)	0.308647*** (0.0342)	0.589009*** (0.0582)
Min. Tolerance		0.32577	0.32577	0.88298	0.89211	0.28705	0.29166		
Max. VIF		3.06963	3.06963	1.13253	1.12093	3.48366	3.42868		
Max. Con. I.		82.08846	82.08846	18.24734	13.61331	109.40284	95.41638		
R-Square		0.4648	0.3487	0.6588	0.4610	0.7044	0.4968	0.7384	0.5091
Moran	**Pos.**	13* 9** 4***	14* 10** 5***	1* 1** 0***	1* 1** 0***	5* 2** 1***	2* 1** 0***	1* 1** 0***	0* 0** 0***
	Neg.	1* 1** 0***	2* 2** 0***	7* 5** 0	12* 6** 1***	2* 2** 1***	9* 5** 2***	8* 5** 2***	15* 13** 6***
Geary	**Pos.**	9* 8** 6***	18* 16** 12***	1* 1** 1***	1* 1** 1***	8* 5** 2***	4* 2** 1***	4* 3** 0***	1* 1** 0***
	Neg.	1* 1** 0***	0* 0** 0***	3* 2** 1***	3* 1** 0***	2* 1** 1***	2* 0** 0***	2* 2** 2***	5* 3** 0***
year		25	25	24	24	24	24	24	24
DW	**Pos.**	44* 38** 24***	13* 10** 3***						
	Neg.	1* 1** 1***	4* 3** 2***						
BG	**Pos.**	26* 19** 9***	3* 2** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Neg.	1* 1** 1***	5* 3** 1***	16* 5** 1***	24* 14** 7***	10* 2** 0***	22* 14** 6***	26* 15** 1***	25* 16** 7***
counties		67	67	67	67	67	67	67	67

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

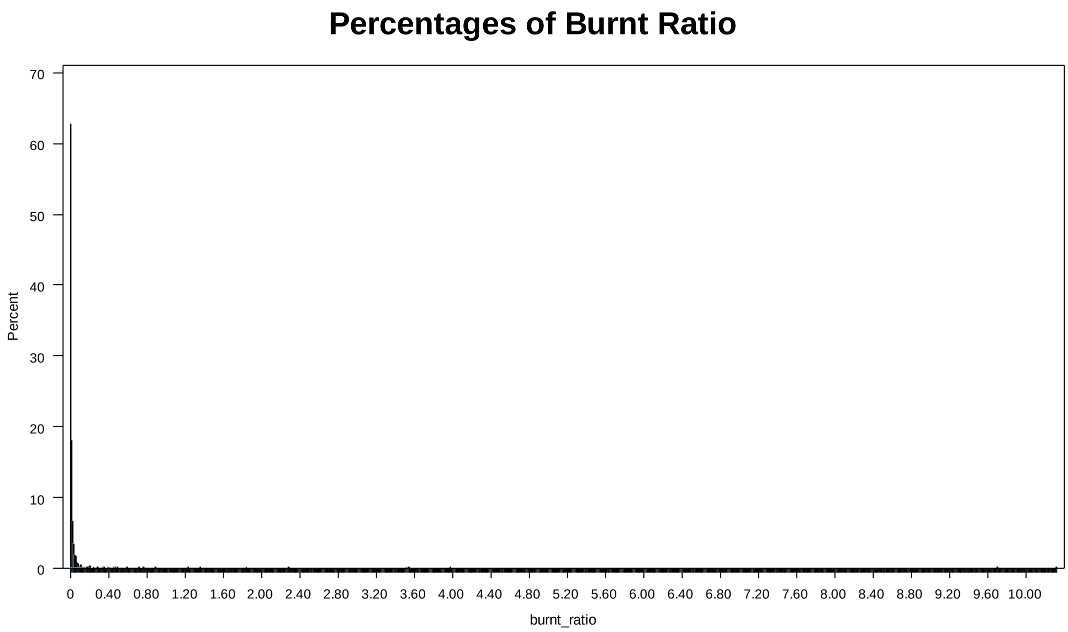


6.A

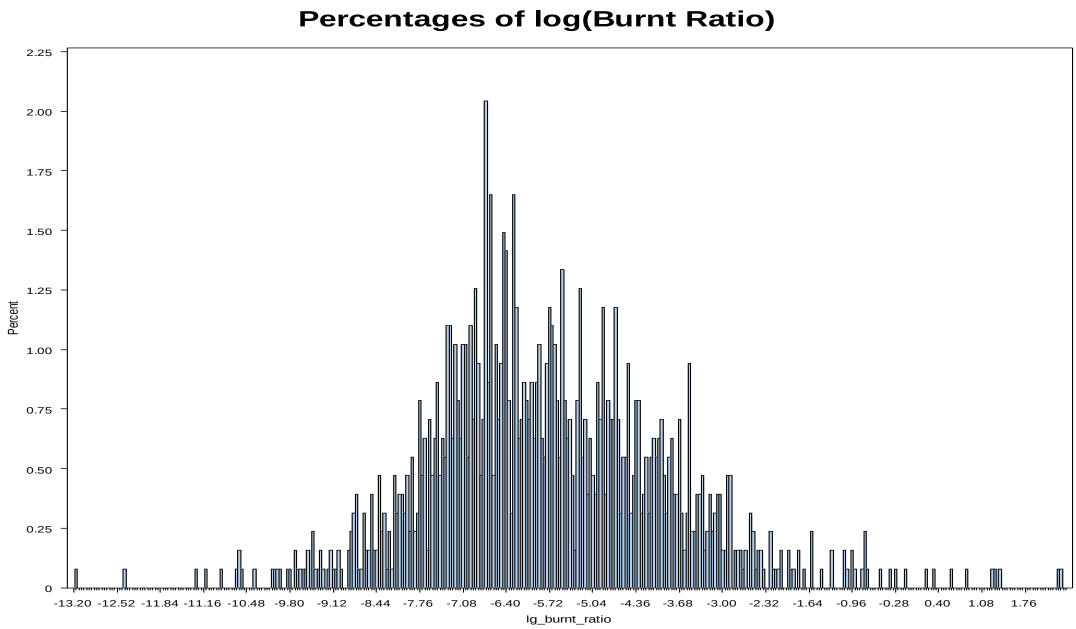


6.B

Figure 6. Percentage distributions of annual wildfire density and its logarithm



7.A



7.B

Figure 7. Percentage distributions of annual wildfire burned ratio and its logarithm

Table 5. Different Models to regress logarithms of fire count per acre and burned ratio

Table		(5.a)		(5.b)		(5.c)		(5.d)	
Model		$y_t = x_t\beta$		$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$		$y_t = x_t\beta + \rho_s w_s y_t + \rho_t y_{t-1}$		$y_t - x_t\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_t\beta)$	
Dependent Variable		Log (Wildfire count per acre of forestland)	Log (Burned ratio)	Log (Wildfire count per acre of forestland)	Log (Burned ratio)	Log (Wildfire count per acre of forestland)	Log (Burned ratio)	Log (Wildfire count per acre of forestland)	Log (Burned ratio)
Parameter		Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)	Estimate (Std. Error)
ρ_t				0.69895*** (0.01714)	0.55717*** (0.02125)	0.53596*** (0.01761)	0.39164*** (0.02102)	0.588656*** (0.0167)	0.411483*** (0.0145)
ρ_s				0.34664*** (0.02095)	0.41311*** (0.02538)	0.25996*** (0.02080)	0.34272*** (0.03002)	0.410631*** (0.0238)	0.5203*** (0.0277)
Intercept		-5.82880*** (0.10242)	-1.54386*** (0.22329)	0.33298** (0.13769)	-0.36759*** (0.13111)	-1.01448*** (0.17762)	-0.01525 (0.22107)	-4.62167 (82.5361)	-6.20862*** (1.5124)
Forestland Ratio		-2.62855*** (0.10010)	-4.23759*** (0.21822)			-1.11324*** (0.08579)	-2.21738*** (0.21285)	-2.68915*** (0.2134)	-4.11477*** (0.3792)
Private owned share of forestland		0.95877*** (0.08484)	0.30890* (0.18496)			0.17490*** (0.06721)	-0.04651 (0.16536)	0.317795** (0.1473)	0.101368 (0.3120)
Longleaf / slash pine & loblolly / shortleaf pine share		0.04948 (0.13904)	0.47948 (0.30311)			0.00943 (0.10703)	0.53092* (0.27603)	0.238136 (0.3027)	0.649365 (0.5203)
Oak / pine & oak / hickory share		-1.13957*** (0.13577)	-2.90665*** (0.29599)			-0.35124*** (0.10452)	-1.13294*** (0.27289)	-0.494 (0.3193)	-1.68376*** (0.6504)
Oak / gum / cypress share		-0.86334*** (0.13577)	0.48957 (0.32331)			-0.53171*** (0.11318)	0.13895 (0.29081)	-0.59182** (0.2995)	0.989263* (0.5470)
Daily average of HDD Index		-0.86334 (0.01927)	0.07042* (0.04201)			0.12288*** (0.01621)	0.21533*** (0.04123)	0.12955*** (0.0233)	0.171322** (0.0762)
December SP12 index		-0.29429*** (0.01456)	-0.46167*** (0.03174)			-0.15626*** (0.01241)	-0.22945*** (0.03144)	-0.17793*** (0.0141)	-0.21763*** (0.0413)
Hurricane incidences		0.07657 (0.06810)	-0.04448 (0.14846)			0.00822 (0.05065)	-0.20214 (0.13058)	-0.00649 (0.0466)	-0.19415 (0.1336)
Population density		0.23003*** (0.01571)	0.35223*** (0.03424)			0.06834*** (0.01271)	0.17157*** (0.03169)	0.109052*** (0.0252)	0.251253*** (0.0477)
Employment ratio		-1.89963*** (0.15339)	-2.92526*** (0.33439)			-0.52504*** (0.12111)	-1.39226*** (0.30357)	0.118312*** (0.0359)	0.346029*** (0.0723)
Min. Tolerance		0.25481	0.25481	0.59977	0.72632	0.21942	0.21238		
Max. VIF		3.92445	3.92445	1.66732	1.37680	4.55752	4.70848		
Max. Con. Index		24.17543	23.99894	27.73875	8.77367	67.56304	29.46754		
R-Square		0.6839	0.5455	0.7946	0.6155	0.8195	0.6437	0.8290	0.6470
Moran	Pos.	11* 8** 5***	10* 9** 4***	1* 0** 0***	1* 0** 0***	5* 1** 0***	1* 1** 0***	1* 1** 0***	0* 0** 0***
	Neg.	1* 0** 0***	2* 2** 0***	5* 4** 2***	11* 5** 1***	6* 3** 1***	6* 4** 3***	8* 5** 3***	17* 15** 7***
Geary	Pos.	11* 9** 4***	17* 13** 10***	2* 1** 0***	1* 1** 0***	6* 3** 0***	2* 2** 1***	4* 2** 0***	1* 1** 0***
	Neg.	1* 1** 0***	1* 0** 0***	2* 2** 2***	3* 2** 0***	2* 2** 2***	3* 1** 0***	5* 2** 2***	5* 2** 1***
years		25	25	24	24	24	24	24	24
DW	Pos.	39* 37** 29***	14* 8** 4***	1* 1* 0*	0* 0** 0***				
	Neg.	1* 1** 0***	7* 4** 1***	13* 5** 1***	26* 18** 8***				
BG	Pos.	30* 26** 11***	4* 2** 0***	0* 0** 0***	0* 0** 0***	1* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Neg.	1* 1** 0***	6* 4** 0***	13* 3** 1***	24* 14** 7***	6* 1** 0***	22* 12** 4***	23* 10** 0***	24* 15** 7***
counties		67	67	67	67	67	67	67	67

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

Table 6. Different models to predict annual count

Table	(6.a)					(6.b)	(6.c)	(6.d)	(6.e)	(6.f)
	$y_t = x_{t-1}\beta$	**Poisson Model**		**Negative Binomial Model**		$y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta)$
Parameter	**Estimate (Std. Error)**	**Estimate (Std. Error)**	Elasticity[1]	**Estimate (Std. Error)**	Elasticity[1]	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t						0.60453*** (0.01842)	0.66624*** (0.01719)	0.53002*** (0.01432)	0.51138*** (0.01533)	0.437214*** (0.00718)
ρ_s						-0.13923*** (0.02473)	-0.20527*** (0.02305)	0.40934*** (0.02448)	0.43408*** (0.02365)	0.500419*** (0.0203)
Intercept	20.92755** (10.35073)	3.6007*** (0.0196)		3.7812*** (0.1150)		19.22080*** (7.15957)	36.29542*** (1.95879)	-0.80831 (2.14588)	-33.0332*** (6.48249)	-7.94793 (34.2725)
Lag of Forestland (in 10,000 acres)	20.02549*** (1.23120)	0.2196*** (0.0020)	0.5396171	0.2348*** (0.0147)	0.5770473	7.72605*** (0.86282)			8.06296*** (0.78446)	20.35515*** (1.3265)
Lag of Private owned share of forestland	80.15939*** (9.44461)	1.0208*** (0.0186)	0.7652139	1.0205*** (0.1126)	0.7649991	15.29453** (6.23272)			12.85277** (5.72110)	28.29427** (12.1110
Lag of Longleaf / slash pine & loblolly / shortleaf pine share	-31.95911** (14.70106)	-0.4165*** (0.0274)	-0.172516	-0.6004*** (0.1798)	-0.248659	17.74991* (9.68105)			28.21510*** (8.82986)	-37.5238 (22.8597)
Lag of Oak / pine & oak / hickory share	-4.71984 (14.32984)	0.0072 (0.0266)	0.0014607	-0.2078 (0.1632)	-0.042297	27.36288*** (9.34320)			37.20301*** (8.51917	2.251537 (25.2867)
Lag of Oak / gum / cypress share	-9.41198 (15.80864)	-0.1386*** (0.0298)	-0.030829	-0.3526* (0.1829)	-0.078425	5.80155 (10.26105)			-1.70059 (9.42315)	-56.4083** (22.9999)
Lag of Daily average of HDD Index	- 12.55794*** (1.67588)	-0.1357*** (0.0031)	-0.382	-0.1455*** (0.0193)	-0.409562	-9.27841*** (1.10784)			-6.88770*** (1.02594)	-3.69993* (2.0729)
Lag of December SP12 index	-9.90406*** (1.53949)	-0.1213*** (0.0028)	-0.023542	-0.1313*** (0.0180)	-0.025482	4.21859*** (1.12564)			8.20500*** (0.96839)	-0.63636 (1.4856)
Lag of Hurricane incidences	-3.44590 (7.18044)	-0.0451*** (0.0142)	-0.00148	-0.0696 (0.0831)	-0.002286	-2.41062 (4.56311)			-3.13451 (4.18743)	-2.46899 (4.9050)
Lag of Population (in 10,000)	5.24713*** (0.64073	0.0628*** (0.0011)	0.1273566	0.0563*** (0.0083)	0.1141138	1.77842*** (0.42217)			2.21017*** (0.38535)	4.64926*** (0.5877)
Lag of Employment ratio	-22.23841 (15.33700)	-0.3010*** (0.0286)	-0.127493	-0.3234* (0.1854)	-0.136978	-15.83370 (9.82083)			-8.27262 (9.01396)	6.713863*** (1.1002)
Scale/Dispersion		1.0000		0.4738 (0.0159)						
Min. Tolerance	0.30875					0.29432	0.73309	0.94309	0.29792	
Max. VIF	3.23884					3.39764	1.36408	1.06035	3.35658	
Max. Con. Index	20.88127					24.20283	4.45534	4.74331	23.79281	
R-Square	0.1799					0.5486	0.5102	0.5622	0.6199	0.6163
Moran **Pos.**	16* 10** 4***	19* 13** 6***		20* 16** 6***		15* 13** 7***	18* 16** 12***	2* 2** 2***	5* 5** 4***	2* 2** 0***
Moran **Neg.**	0* 0** 0***	0* 0** 0***		0* 0** 0***		0* 0** 0***	0* 0** 0***	15* 15** 8***	10* 9** 8***	20* 19** 12***
Geary **Pos.**	6* 4** 1***	6* 5** 1***		6* 5** 1***		9* 6** 5***	12* 9** 6***	3* 2** 1***	5* 5** 1***	1* 1** 0***
Geary **Neg.**	2* 0** 0***	2* 0** 0***		2* 0** 0***		2* 1** 0***	0* 0** 0***	15* 14** 11***	12* 11** 8***	19* 16** 9***
years	25	25		25		24	24	24	24	24
DW **Pos.**	46* 37** 22***	48* 38** 20***		47* 36** 20***						
DW **Neg.**	0* 0** 0***	0* 0** 0***		0* 0** 0***						
BG **Pos.**	9* 6** 4***	12* 5** 3***		10* 6** 3***		3* 1** 0***	1* 0** 0***	2* 1** 0***	5* 4** 2***	11* 7** 1***
BG **Neg.**	0* 0** 0***	0* 0** 0***		0* 0** 0***		2* 0** 0***	3* 1** 0***	4* 2** 0***	4* 2** 0***	2* 1** 0***
counties	67	67		67		67	67		67	67
check up (2005) [2]	159432.76	164416.41		163712.13		74843.497	58757.096	21521.843 32841.464 [4]	62525.812 425607.37 [4]	44074.975 483100.7 [4]
check up (2004) [3]	85523.385	86139.126		84245.262		19533.545	30775.609	72236.266 362352.85 [4]	32984.376 338786.34 [4]	50579.731 380313.65 [4]

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Elasticities are evaluated at the mean values of explanatory variables. For both Poisson and Negative binomial models, the elasticity is given by $\bar{X}_k \beta_k$.

[2]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2005 using 1981-2004 data.

[3]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2004 using 1981-2003 data.

[4]: Those are got when we include the estimate of residuals in the forecast equation.

Table 7. Models to predict logarithm of annual wildfire count in each county

Table	(7.a)	(7.b)	(7.c)	(7.d)	(7.e)	(7.f)
Models	$y_t = x_{t-1}\beta$	$y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta)$
Parameter	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t		0.70523*** (0.02157)	0.83936*** (0.02068)	0.67268*** (0.01532)	0.62849*** (0.01874)	0.536464*** (0.0153)
ρ_s		-0.06451* (0.02601)	0.06893*** (0.02576)	0.38869*** (0.02271)	0.37560*** (0.02094)	0.485871*** (0.0213)
Intercept	-6.46542*** (0.36981)	-1.93481*** (0.32115)	-0.80342** (0.15146)	-0.37525*** (0.09476)	-3.24757*** (0.28487)	-15.0037* (8.2618)
Lag of log(Forestland)	0.58267*** (0.02554)	0.19261*** (0.02231)			0.18354*** (0.02036)	0.585121*** (0.0450)
Lag of Private owned share of forestland	1.48377*** (0.10140)	0.38765*** (0.08144)			0.21628*** (0.07469)	0.456972** (0.1545)
Lag of Longleaf / slash pine & loblolly / shortleaf pine share	-0.54242*** (0.15523)	0.05707 (0.11790)			0.15622 (0.10755)	0.073747 (0.2877)
Lag of Oak / pine & oak / hickory share	-1.04412*** (0.15898)	-0.09141 (0.12133)			0.08994 (0.11071)	-0.43744 (0.3462)
Lag of Oak / gum / cypress share	-0.68685*** (0.17327)	-0.17194 (0.13060)			-0.07180 (0.11937)	-0.22936 (0.3031)
Lag of Daily average of HDD Index	-0.14491*** (0.01901)	-0.08712*** (0.01462)			-0.05543*** (0.01346)	0.021564 (0.0265)
Lag of December SP12 index	-0.12124*** (0.01693)	0.08981*** (0.01475)			0.11704*** (0.01257)	-0.01148 (0.0159)
Lag of Hurricane incidences	-0.08231 (0.07895)	-0.11012* (0.05810)			-0.11619** (0.05310)	-0.07266* (0.0435)
Lag of log (Population)	0.37157*** (0.01747)	0.10782*** (0.01510)			0.08869*** (0.01384)	0.258499*** (0.0223)
Lag of Employment ratio	-1.93565*** (0.18760)	-0.60001*** (0.14443)			-0.35062** (0.13228)	0.303846*** (0.0287)
Min. Tolerance	0.33475	0.30434	0.72119	0.88298	0.30268	
Max. VIF	2.98731	3.28585	1.13253	1.13253	3.30380	
Max. Con. Index	81.77689	108.50138	18.65215	18.24734	106.43400	
R-Square	0.3710	0.6439	0.6028	0.6588	0.7025	0.7061
Moran Pos.	8* 6** 2***	15* 11** 7	17* 16** 11***	1* 1** 0***	1* 1** 1***	1* 0** 0***
Moran Neg.	1* 1** 0***	0* 0** 0***	0* 0** 0***	7* 5** 0	4* 3** 2***	14* 12** 7***
Geary Pos.	9* 7** 2***	17* 12** 8***	17* 14** 11***	1* 1** 1***	4* 1** 1***	1* 0** 0***
Geary Neg.	1* 1** 0***	1* 0** 0***	0* 0** 0***	3* 2** 1***	3* 1** 1***	6* 3** 2***
years	25	24	24	24	24	24
DW Pos.	48* 46** 26***					
DW Neg.	0* 0** 0***					
BG Pos.	26* 19** 6***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
BG Neg.	0* 0** 0***	8* 4** 0***	10* 5** 0***	16* 5** 1***	17* 10** 2***	12* 3** 0***
Counties	67	67	67	67	67	67
check up (2005) [1]	63.263602	35.57923	32.406817	26.424517 21.214065 [3]	34.732176 56.944448 [3]	44.402395 95.296688 [3]
check up (2004) [2]	20.571851	11.164984	15.507054	42.692813 82.499196 [3]	26.889703 65.591233 [3]	21.503793 107.36948 [3]

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2005 using 1981-2004 data.

[2]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2004 using 1981-2003 data.

[3]: Those are got when we include the estimate of residuals in the forecast equation.

Table 8. Models to predict the logarithm of annual burned acreage in each county:

Table		(8.a)	(8.b)	(8.c)	(8.d)	(8.e)	(8.f)
Models		$y_t = x_{t-1}\beta$	$y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta)$
Parameter		**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t			0.45774*** (0.02430)	0.57937*** (0.02322)	0.47445*** (0.01914)	0.41602*** (0.02193)	0.41302*** (0.0144)
ρ_s			-0.15002*** (0.03444)	0.01732*** (0.02887)	0.43309*** (0.02426)	0.44233*** (0.02990)	0.548024*** (0.0271)
Intercept		-3.50051*** (0.74315)	-0.91785 (0.72268)	2.51242*** (0.18328)	0.30288* (0.17697)	-4.81392*** (0.66986)	5.323736 (12.3817)
Lag of log(Forestland)		0.37510*** (0.05132)	0.20583*** (0.04734)			0.20072*** (0.04464)	0.470579*** (0.0717)
Lag of Private owned share of forestland		1.22829*** (0.20376)	0.48311** (0.18826)			0.39448** (0.17762)	0.503412 (0.3394)
Lag of Longleaf / slash pine & loblolly / shortleaf pine share		-0.53194* (0.31194)	-0.06752 (0.28942)			0.51968* (0.27200)	0.398547 (0.5628)
Lag of Oak / pine & oak / hickory share		-2.97625*** (0.31948)	-1.41312*** (0.3008)			-0.61555** (0.28450)	-1.66832** (0.7301)
Lag of Oak / gum / cypress share		0.70855** (0.34820)	0.48719 (0.31912)			0.63875** (0.30067)	1.182445** (0.5980)
Lag of Daily average of HDD Index		-0.18791*** (0.03821)	-0.17340*** (0.03586)			-0.04622 (0.03405)	-0.04618 (0.0712)
Lag of December SP12 index		-0.27441*** (0.03401)	-0.10333*** (0.03444)			0.05481* (0.03115)	-0.13663*** (0.0439)
Lag of Hurricane incidences		-0.30302* (0.15866)	-0.23380 (0.14215)			-0.26482** (0.13395)	-0.11805 (0.1313)
Lag of log (Population)		0.64615*** (0.03511)	0.35729*** (0.03537)			0.31171*** (0.03351)	0.499127*** (0.0398)
Lag of Employment ratio		-3.33819*** (0.37698)	-1.83036*** (0.34852)			-1.61800*** (0.32907)	0.58155*** (0.0553)
Min. Tolerance		0.33475	0.31885	0.72572	0.89211	0.32121	
Max. VIF		2.98731	3.13629	1.37795	1.12093	3.11319	
Max. Con. Index		81.77689	95.09572	14.06757	13.61331	93.83948	
R-Square		0.2964	0.4225	0.3540	0.4610	0.4861	0.5026
Moran	**Pos.**	12* 9** 3***	7* 6** 4***	13* 11** 6***	1* 1** 0***	1* 0** 0***	0* 0** 0***
	Neg.	1* 0** 0***	1* 1** 0***	0* 0** 0***	12* 6** 1***	9* 7** 2***	17* 15** 9***
Geary	**Pos.**	19* 18** 9***	11* 9** 4***	12* 10** 6***	1* 1** 1***	2* 1** 0***	1* 0** 0***
	Neg.	0* 0** 0***	1* 0** 0***	0* 0** 0***	3* 1** 0***	1* 1** 0***	4* 2** 0***
years		25	24	24	24	24	24
DW	**Pos.**	25* 13** 7***					
	Neg.	2* 1** 0***					
BG	**Pos.**	7* 5** 1***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Neg.	1* 1** 0***	12* 6** 2***	17* 14** 4***	24* 14** 7***	19* 12** 7***	19* 13** 6***
Counties		67	67	67	67	67	67
check up (2005) [1]		171.19602	136.12877	123.56706	102.74964 102.28712 [3]	128.26596 404.93751 [3]	261.7547 549.96744 [3]
check up (2004) [2]		114.84316	84.122016	112.17891	156.45708 255.07911 [3]	119.31252 306.06023 [3]	93.80009 264.77826 [3]

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2005 using 1981-2004 data.

[2]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2004 using 1981-2003 data.

[3]: Those are got when we include the estimate of residuals in the forecast equation, like (1.65), (6.26) & (6.31).

Table 9. Models to predict logarithm of wildfire count per acre in each county each year:

Table	(9.a)	(9.b)	(9.c)	(9.d)	(9.e)	(9.f)
	$\boldsymbol{y_t = x_{t-1}\beta}$	$\boldsymbol{y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t y_{t-1}}$	$\boldsymbol{y_t = \rho_s w_s y_{t-1} + \rho_t y_{t-1}}$	$\boldsymbol{y_t = \rho_s w_s y_t + \rho_t y_{t-1}}$	$\boldsymbol{y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1}}$	$\boldsymbol{y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta)}$
Parameter	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
$\boldsymbol{\rho_t}$		0.65819*** (0.02284)	0.83936*** (0.02068)	0.69895*** (0.01714)	0.57874*** (0.01991)	0.505073*** (0.0159)
$\boldsymbol{\rho_s}$		-0.04078* (0.02417)	0.06893*** (0.02576)	0.34664*** (0.02095)	0.34732*** (0.01977)	0.523495*** (0.0237)
Intercept	-5.86372*** (0.11187)	-2.40850*** (0.18621)	-0.80342** (0.15146)	0.33298** (0.13769)	-0.32957* (0.17662)	-4.19785* (2.4170)
Lag of forestland ratio	-2.56629*** (0.10888)	-0.83153*** (0.10204)			-0.58663*** (0.09400)	-2.6966*** (0.1916)
Lag of Private owned share of forestland	1.06870*** (0.09311)	0.31121*** (0.07669)			0.14925** (0.07044)	0.405818*** (0.1324)
Lag of Longleaf / slash pine & loblolly / shortleaf pine share	-0.08481 (0.15091)	0.17018 (0.11974)			0.22684** (0.10954)	0.097471 (0.2649)
Lag of Oak / pine & oak / hickory share	-1.27076*** (0.14909)	-0.21053* (0.12020)			-0.09240 (0.11009)	-0.61274** (0.2898)
Lag of Oak / gum / cypress share	-0.98075** (0.16174)	-0.27112** (0.12921)			-0.42279*** (0.11820)	-0.62218** (0.2669)
Lag of Daily average of HDD Index	0.02308 (0.02069)	-0.04352** (0.01691)			0.03784** (0.01559)	0.07203*** (0.0275)
Lag of December SP12 index	-0.12217*** (0.01598)	0.07299*** (0.01452)			0.09953*** (0.01264)	-0.01184 (0.0166)
Lag of Hurricane incidences	-0.07438 (0.07455)	-0.10482* (0.05762)			-0.10775** (0.05278)	-0.05528* (0.0410)
Lag of Population density	0.23195*** (0.01721)	0.07813*** (0.01446)			0.05586*** (0.01328)	0.110338*** (0.0234)
Lag of Employment ratio	-1.92807*** (0.16986)	-0.65819*** (0.13936)			-0.30280*** (0.12803)	0.144654*** (0.0303)
Min. Tolerance	0.25742	0.25330	0.50269	0.59977	0.17788	
Max. VIF	3.88475	3.94791	1.98929	1.66732	5.62191	
Max. Con. Index	23.90997	62.04974	30.85670	27.73875	63.22879	
R-Square	0.6212	0.7664	0.7493	0.7946	0.8039	0.8129
Moran Pos.	9* 6** 4***	16* 11** 6***	17* 17** 13***	1* 0** 0***	1* 1** 0***	0* 0** 0***
Moran Neg.	1* 0** 0***	0* 0** 0***	0* 0** 0***	5* 4** 2***	7* 3** 1***	16* 12** 8***
Geary Pos.	8* 7** 2***	15* 12** 8***	18* 15** 13***	2* 1** 0***	3* 3** 1***	1* 0** 0***
Geary Neg.	3* 1** 0***	1* 0** 0***	1* 0** 0***	2* 2** 2***	2* 2** 1***	9* 6** 2***
years	25	24	24	24	24	24
DW Pos.	48* 38** 30***					
DW Neg.	0* 0** 0***					
BG Pos.	29* 23** 6***	1* 0** 0***	0* 0** 0***	0* 0** 0***	1* 0** 0***	0* 0** 0***
BG Neg.	0* 0** 0***	6* 3** 0***	14* 6** 1***	13* 3** 1***	13* 5** 2***	9* 3** 0***
Counties	67	67	67	67	67	67
check up (2005) [1]	55.435587	34.470847	26.607207	24.106342 21.126487 [3]	32.58941 62.393279 [3]	25.07344 49.413273 [3]
check up (2004) [2]	19.339284	10.315427	24.543754	36.756875 78.479552 [3]	18.243575 56.561765 [3]	49.969648 84.666755 [3]

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2005 using 1981-2004 data.

[2]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2004 using 1981-2003 data.

[3]: Those are got when we include the estimate of residuals in the forecast equation, like (1.65), (6.26) & (6.31).

Table 10. Models to predict the logarithm of annual burned ratio in each county:

Table		(10.a)	(10.b)	(10.c)	(10.d)	(10.e)	(10.f)
		$y_t = x_{t-1}\beta$	$y_t = x_{t-1}\beta + \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_{t-1} + \rho_t y_{t-1}$	$y_t = \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t = x_{t-1}\beta + \rho_s w_s y_t + \rho_t y_{t-1}$	$y_t - x_{t-1}\beta = (\rho_s w_s + \rho_t w_t)(y_t - x_{t-1}\beta)$
Parameter		**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**	**Estimate (Std. Error)**
ρ_t			0.42077*** (0.02455)	0.66737*** (0.02532)	0.55717*** (0.02125)	0.38480*** (0.02231)	0.390448*** (0.0147)
ρ_s			-0.12876*** (0.03255)	0.13843*** (0.03079)	0.41311*** (0.02538)	0.40470*** (0.02854)	0.544464*** (0.0275)
Intercept		-1.53705*** (0.23075)	-1.40730*** (0.22886)	-1.22664*** (0.14033)	-0.36759*** (0.13111)	0.01957 (0.22040)	-6.22138*** (1.6033)
Lag of forestland ratio		-4.29716*** (0.22458)	-2.66444*** (0.23215)			-1.84148*** (0.22336)	-4.1018*** (0.3649)
Lag of Private owned share of forestland		0.39762* (0.19206)	0.06160 (0.17913)			-0.00346 (0.16959)	0.160963 (0.3180)
Lag of Longleaf / slash pine & loblolly / shortleaf pine share		0.28850 (0.31128)	0.41987 (0.29313)			0.80195*** (0.27662)	0.615741 (0.5182)
Lag of Oak / pine & oak / hickory share		-3.16144*** (0.30752)	-1.61871*** (0.29600)			-0.90658** (0.28132)	-1.67927** (0.6608)
Lag of Oak / gum / cypress share		0.24948 (0.33361)	0.29661 (0.31104)			0.18815 (0.29463)	0.933587* (0.5416)
Lag of Daily average of HDD Index		0.09964** (0.04268)	-0.02231 (0.04100)			0.13311*** (0.03872)	0.100721 (0.0747)
Lag of December SP12 index		-0.27623*** (0.03297)	-0.11429*** (0.03397)			0.03003 (0.03099)	-0.13497*** (0.0446)
Lag of Hurricane incidences		-0.29529* (0.15376)	-0.23385* (0.14048)			-0.26058* (0.13290)	-0.09709 (0.1318)
Lag of Population density		0.34905*** (0.03549)	0.20520*** (0.03405)			0.17438*** (0.03231)	0.247498*** (0.0469)
Lag of Employment ratio		-2.96901*** (0.35037)	-1.78497*** (0.33022)			-1.33106*** (0.31353)	0.333219*** (0.0689)
Min. Tolerance		0.25742	0.20630	0.61406	0.72632	0.19983	
Max. VIF		3.88475	4.84737	1.62851	1.37680	5.00420	
Max. Con. Index		23.90997	28.23983	9.66561	8.77367	28.40827	
R-Square		0.5123	0.5872	0.5386	0.6155	0.6299	0.6437
Moran	**Pos.**	10* 7** 4***	7* 5** 5***	14* 13** 11***	1* 0** 0***	1* 0** 0***	0* 0** 0***
	Neg.	1* 0** 0***	1* 0** 0***	0* 0** 0***	11* 5** 1***	5* 4** 1***	16* 15** 8***
Geary	**Pos.**	18* 15** 7***	9* 6** 5***	14* 12** 9***	1* 1** 0***	2* 1** 0***	1* 1** 0***
	Neg.	0* 0** 0***	0* 0** 0***	0* 0** 0***	3* 2** 0***	2* 2** 0***	5* 2** 1***
years			24	24	24	24	24
DW	**Pos.**	22* 13** 7***					
	Neg.	2* 0** 0***					
BG	**Pos.**	8* 6** 0***	1* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***	0* 0** 0***
	Neg.	2* 0** 0***	11* 7** 2***	19* 14** 5***	24* 14** 7***	13* 11** 5***	19* 12** 6***
Counties		67	67	67	67	67	67
check up (2005) [1]		154.82813	126.88859	112.15762	99.252767 98.331199 [3]	118.61563 407.15129 [3]	104.3755 370.90999 [3]
check up (2004) [2]		113.54965	89.515101	131.81784	166.81928 275.96375 [3]	105.23999 277.73453 [3]	167.04101 266.77623 [3]

Notes: Single (*), double (**), and triple (***) denote significance at 0.10, 0.05, and 0.01 levels, respectively.

[1]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2005 using 1981-2004 data.

[2]: Each reported value is the sum of squared forecast errors when we forecast the wildfires of year 2004 using 1981-2003 data.

[3]: Those are got when we include the estimate of residuals in the forecast equation.