Spatial Statistical Modeling of Wildfire Incidence in Arizona

STAT574E | Group 3

Raymond Owino, Alex Salce, Matthew Wallace

December 2024

Executive Summary

The *Wildland Fire Incident Locations* dataset we selected for study (provided by the National Interagency Fire Center) contains spatial coordinates of a wildfire's origin and its resulting size in acreage (among several other useful attributes) for all wildfire incidence in the US as archived in the IRWIN system since 2014. Our research goals included the study of factors that may influence the resulting size of a wildfire, and to build spatial models that may aid as a tool for wildfire risk assessment and resource allocation, forecasting, and other possible real-world applications. We were also interested whether wildfire incidence, when treated as a point pattern, occurs as a completely spatially random process, or if there really is inhomogeneity in spatial intensity for incidence and thus truly higher risk in some spatial regions than others.

Attributes contained in the dataset allowed us to take three different approaches to spatial statistical modeling. Using the incident size as a continuous response, we built a spatial linear model that predicts wildfire size based on spatial and selected covariate inputs. Filtering the data using incident size as well as other incident attributes also allowed us to treat wildfire incidence data as a point process. We fit a Log-Gaussian Cox Process (LGCP) spatial model to study spatial intensity, as well as a binary response spatial logistic regression model to map spatially the regions of highest probability of large wildfire (≥ 1000 acres) incidence in the region of study.

Our modeling efforts studied Coconino County (CC) in Northern Arizona, which has the highest overall wildfire incidence as compared to the rest of the state. We gathered data for incidence proximity to roads, environmental, and census data corresponding to spatial coordinates to use as covariates for our models to support our research questions. These covariates improved performance of our models by AIC metrics. **Our LGCP model indicated that wildfires are not CSR, and there is thus greater risk of wildfire incidence in certain regions of CC**. Our models require some further refinements, but serve as a solid foundation and proof of concept for our real-world application goals.

Introduction

Wildfire incidence data describes the spatial origin of a wildfire, and is one of many ways that wildfires can be studied; it is also a type of data that lends itself well to be studied by methods in spatial statistical analyses. Spatial statistical models can be valuable in real applications for wildfire prevention, giving inferential insights into wildfire risk as well as tools to aid in resource allocation for mitigation efforts and planning. Our project seeks to build models that could be useful for these kinds of efforts, and our research questions aim to provide insight to what kinds of environmental and human factors can influence the size of a resulting wildfire based upon where it originates spatially.

Wildfire Incidence Data

Accompanying this paper is `wfigs_az_sf_EPSG32612.RData` which imports `wfigs_az_sf` to the R environment, an `sf` object with 18089 observations of wildfire incidence in the state of Arizona as well as corresponding covariate data for each entry (to be discussed later). This data was originally acquired via Wildland Fire Incident Locations from the National Interagency Fire Center. It contains spatial point data indicating the origin of each wildfire recorded in the IRWIN database, and includes many useful data attributes for each entry. The attributes from this dataset that we used were

- `x` and `y` | Spatial coordinates in lat/lon
- `IncidentSize` | Size of the resulting wildfire in acres
- `FireCause` | Human, Natural, Unknown, Undetermined
- `FireDiscoveryDateTime` | Date & time of incident reporting
- `IncidentTypeCategory` | WF (wildfire) or RX (prescribed burn)

Other covariate data that was captured for each incidence point will be discussed later in the report.

Research Questions

Our research questions were driven by our curiosities about the ways in which we could approach modeling the data. We were also motivated in part by our findings during our explorations, and we wanted to evaluate whether real data we felt could influence wildfire size could serve as useful covariate information for our models. Additionally, we wanted to evaluate spatial randomness for different types of incidence to be sure what we are investigating is more than a completely spatially random process.

In what ways can we approach spatial modeling of this data to produce useful insights?

During our data explorations, we rationalized different ways we could model the dataset. First, we could treat `IncidentSize` as continuous response data, and spatial coordinates as fixed observations, in order to build a spatial linear model. The model could be used to predict wildfire size if an incident were to be observed at a new location with corresponding covariate conditions.

We could also use `IncidentSize`, as well as other data attributes, as filters to subset incidence of types that we would like to study, for example large, naturally caused wildfires

Proximity to roads (major or backroads) During data exploration, one

Can we find useful covariate data that can improve our models?

.

Are the patterns of human or non human caused fires spatially CSR, or do they exhibit an inhomogeneous spatial intensity?

If we filter our data to a desired "type" to study (for example large, naturally occurring wildfires) as a realization of some Poisson process with intensity $\lambda(\mathbf{s})$, can we identify higher risk areas of the fire type in question? Or are they not more than completely spatially random (CSR). As an extension, can we model the probability of these types of fires we choose to assess risks in terms of proababilities given conditions (covariates) that we choose for our model fits?

Exploratory Data Analysis

Statistical Analyses

Fixed, Continuous Response

Spatial Linear Model

Point Process

Log-Gaussian Cox Process

Binary Response GLM Logistic Regression

Conclusions