Title
A Statistical Analysis of Wildfire Occurrences in Los Angeles County

ISBN
9798247985600

Author
Xia, Jingfan

Publication Date
2026-06-05

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

A Statistical Analysis of Wildfire Occurrences in Los Angeles County

A thesis submitted in partial satisfaction

of the requirements for the degree

Master of Science in Statistics

by

Jingfan Xia

2026

ABSTRACT OF THE THESIS

A Statistical Analysis of Wildfire Occurrences in Los Angeles County

by

Jingfan Xia
Master of Science in Statistics
University of California, Los Angeles, 2026
Professor Frederic Schoenberg, Chair

As global warming and urban development continue, areas with warm and dry climates have become more prone to wildfires in recent years. Not only do wildfires cause instant public health issues and financial losses, but they also result in long-term damage in local biosphere. To predict and mitigate wildfires, it is important to apply statistical methods to perform a comprehensive analysis on the occurrences of wildfires to understand the trends and correlations.

Los Angeles County is an important and typical area to analyze, as it has a dry, warm Mediterranean climate, and is surrounded by hills and mountains. Those mountains are usually covered by deciduous trees and shrubs, which contributed to dry, fallen leaves that are prone to get ignited by winds in winter. Moreover, Los Angeles County is densely populated and has limited land for urban developments. As a result, many regions near foothills have been urbanized, including cities or communities of Pacific Palisades, Brentwood, West Hollywood, Los Feliz, La Canada Flintridge, Altadena, Sierra Madre, and Arcadia. As these parts of the county sit just south of either the Santa Monica Mountains or the San Gabriel Mountains, wildfires that occur in these mountain forests are a nonnegligible threat in winter. The Eaton Fire, for instance, caused severe loss in Altadena, Sierra Madre, and Arcadia, while the Palisades Fire had very negative socioeconomical influences on Pacific Palisades and Brentwood. Both wildfires occured in January 2025, when precipitation was less than normal and a major Santa Ana wind event occurred, which had synergistic effects

that dramatically elevated the risk of wildfires.

To further understand the occurrences of individual wildfires in the Los Angeles County as well as their relationship, I use statistical approaches including Stoyan Grabarnik (SG) estimation, Moran's I method, Local Indicators of Spatial Association (LISA), and Social Network Analysis to form a clear and multidimensional reflection. The Stoyan Grabarnik method is an efficient and predictive way that can help analyze the likelihood of wildfires, which is indispensable for fire prediction and prevention. Moran's I and LISA are more focused on determination of occurrence distribution, which facilitates the identification of high-risk areas in a period. Social Network representations provide us with information about relationship between each wildfire with different parameters and are thus helpful for understanding the mechanism and interaction of wildfires.

The thesis of Jingfan Xia is approved.

Jamie Goodwin-White

Mark S. Handcock

Frederic Schoenberg, Committee Chair

University of California, Los Angeles

2026

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

I would like to thank my family for their continuous support, encouragement, and understanding throughout my academic journey.

CHAPTER 1

Introduction

We have been interacting more closely with the topological, thermal, and hydrological environments as urban developments proceed in various regions all over the world, especially in the past few decades, with rapid advances in science and technology. Recently, many metropolitan areas have extended to the edge of mountains and forests, increasing the possibility of experiencing natural disasters.

Wildfires are unrestrained and destructive fires that often occur in mountainous areas with forests or shrubs. Some areas, including Los Angeles County, are densely populated, have limited precipitation and high average temperature, and sit in the middle of hills or mountains, all of which have caused frequent and risky occurrences of wildfires. As we continue to further promote public safety and thus mitigate negative effects of wildfires on urban areas, it is vital to understand the relationship between different parameters of each wildfire, as well as the relationship between wildfires. Only in this way can we not only form a more comprehensive and accurate understanding of wildfires but also prepare for risk mitigation and public health policies in advance to reduce potential negative influences of these dangerous and complicated events.

In my thesis, I aim to complete a comprehensive, multifaceted, and informative analysis of wildfire occurrences in Los Angeles County, by combining spatiotemporal covariates, spatial autocorrelation, and social networks. With data from NASA satellites about the location, time, luminosity, area, and more relevant parameters, I not only predict the likelihood of wildfire occurrence, but also investigate if these occurrences are clustered or scattered, while incorporating further analysis of the mutual relationship between each pair of occurrences.

Stoyan-Grabarnik (SG) methods, which are ideal for spatiotemporal point process model analysis, are used to estimate and predict the trends of wildfires in the area. Moran's I [1] and LISA [2] methods are then applied to the data for spatial autocorrelation, which facilitates the identification of clusters and outliers among those occurrences. Social networks are then introduced to achieve a more microscopic, interactive relationship between wildfires.

CHAPTER 2

Literature Review

2.1 Background

Wildfires are complex phenomena associated with a wide range of parameters, including temperature, vegetation, wind speed and direction, as well as precipitation. Recent research has pointed out that wildfires are multifaceted, complex situations influenced by synergistic correlation of physical, biological, social, and management parameters[3], which makes it essential to analyze, mitigate, and prevent wildfire occurrences in more comprehensive and interdisciplinary settings that fully consider human-nature interactions in diversified environments.

Moreover, wildfire risks are especially increased in Wildland Urban Interface(WUI) areas, where urban developments meet with natural vegetation, and the expansion of those areas globally has further elevated human exposure to wildfire issues[4, 5]. These regions have suffered from much higher wildfire susceptibility due to the synergistic effects of climate, vegetation, and anthropogenic influences.

Last but not least, recent large scale datasets and reports have further supported those ongoing trends, indicating the increased occurrences of WUI wildfires across the globe[6, 7]. In this way, a variety of statistical methods, including spatiotemporal modeling, spatial autocorrelation, and social networks seem pragmatic for us to better understand the essence of wildfire occurrences and dynamics. The challenges of utilizing statistical models for prediction and management of wildfire events have also been widely discussed in the literature [8]. Statistical models have been extensively developed and optimized to analyze major parameters of wildfire occurrences, including but not limited to ignition, spread, and intensity

[9].

Bivariate and multivariate statistical methods for wildfire probability mapping have been compared to exhibit the advantages of incorporating multiple environmental factors in comprehensive, predictive modeling [10]. Recent work has also combined spatial structure with the network to analyze wildfire ignition points with a linear network point process framework. In contrast, my method forms a network according to spatiotemporal proximity between wildfire occurrence events. Those two perspectives exhibit diverse ways of integrating spatial and network-based modeling in wildfire analysis[11].

I will review the four areas of current literature in the research area of wildfires that lead to this thesis: social and public health perspectives, likelihood prediction, spatiotemporal analysis, as well as correlation estimation.

2.1.1 Social and public health perspectives

For social and public health perspectives, the researchers are more focused on prevention and related policies, as well as socioeconomic outcomes of wildfire occurrences[12]. Wildfire prevention policies are an indispensable part of research about these occurrences, as they cause significant losses in a variety of areas. Prior research focuses mainly on designing proactive methods in advance to reduce risks of long term wildfire occurrences. According to the researchers, prevention seems to be more cost friendly and effective in comparison to suppression of ongoing occurrences. Fuel treatment, prescribed burning, and fire resistant design of communities provide effective ways to mitigate the possibility of wildfires before they initiate. In addition, forest management, brush clearing, and inspections before selling houses are widely utilized methods to mitigate the possibility and influence of possible wildfire occurrences in the future. These methods are supposed to be effective and beneficial, as preventing major wildfires without potential damages usually costs less than putting them out after they already occur.

The research in these perspectives, however, are not customized enough to deal with circumstances at specific locations and thus need further optimizations in clarity and accuracy. For instance, the Los Angeles County has elaborate brush clearance policies due to the risk of wildfires in winter, yet few researchers have conducted detailed research about the effects of them. Brush clearance, which is especially beneficial in areas with climates suitable for the growth of herbal and woody shrubs, can greatly reduce the amount of burnable substances including dropped leaves and branches, and thus dismantle essential material conditions necessary for wildfires to occur. Moreover, wildfire occurrences in Los Angeles are strongly related to urban developments into the wildland urban interface(WUI) areas, where residential, industrial, and business areas meet with vegetation prone to wildfires.

Recent studies have highlighted that structural limits, such as insufficient firefighting resources, vegetation controls, and infrastructure renovations, have noticeably elevated the risk of wildfire occurrences in this region[13]. These analyses can become more customized to different areas with diverse climatical and topographical parameters, as the same method can cause very different effects in different subregions. In Long Beach with limited hills as well as a more oceanic climate, and in Arcadia with a large proportion of high elevation areas and a more continental climate, the effects of the same brush clearance policies for LA County would be quite different due to local environments, so that more parameters reflecting specific conditions should be included to form a more accurate and pertinent conclusion. Moreover, statistical approaches have been included in those literature, which strengthens the validity and trustworthiness of the suggestions provided in them, yet them would provide a more comprehensive and multifaceted conclusions if more diverse parameters and methods are applied.

2.1.2 Likelihood prediction

Likelihood prediction is mostly about estimating the possibility of wildfire occurrences of a specific region in the future given wildfire data of the selected region as well as nearby regions in the past. In recent literatures, a variety of diversified statistical methods have been applied to analyze the probability of future wildfires, and they provide rich and progressive insights for a better understanding of causes and possibilities of wildfires. Learning about the odds of wildfires in each of the specific locations inside high-risk or focused areas is a key step to further optimize and elaborate on wildfire related initiatives and policies. Several statistical modeling methods to capture more subtle and region-specific details. In recent literatures, Multiscale Geographically Weighted Regression (MGWR)[14] has been widely applied to capture spatially varying relationships. If some phenomenon cannot be fully explained by the correlation between existing parameters and wildfire risks, then it is very probable that some hidden, latent variables pose influence on the relationship. Moreover, the usage of Distributed Lag Nonlinear Models (DLNM) [15] has been proven effective for quantifying lagged and nonlinear weather-wildfire relationships, so that we don't miss the nonlinear part of the correlation between heat, humidity, wind, and wildfire risks. For interpretable exposure–response estimation, generalized additive models (GAMs) allow smoother nonlinear effects of temperature, humidity, and more parameters. Few of current literatures have applied the Stoyan Grabarnik method, which splits the region into smaller grids and thus optimizes the locality and accuracy of likelihood analysis. I will use Stoyan Grabarnik estimation to characterize small-scale clustering beyond what is already explored about the occurrences of wildfires. This approach would complement the area level models and provide a fine scale understanding of local wildfire hotspots, and thus offers further insights into the identification, adjustment, and removal of high-risk areas as we update the model once new wildfires occur in the focused region.

2.1.3 Spatiotemporal analysis

As wildfires are related to complex and intrigued parameters that alter with space and time, spatiotemporal structures have been accounted for by combining spatial statistical tools (including Moran's I and LISA) with machine-learning models enriched with neighborhood-based covariates, enabling both clustering detection and predictive risk mapping. The usage of Moran's I and LISA have greatly increased the precision and efficiency of estimating trends of wildfire occurrences. Based on current literatures, the results of these machine learning methods can be used to identify districts with high risks and thus simulate the effect of adjusting to find the most optimal urban setting for minimal overall wildfire risks among different regions. A broader usage of these spatial statistic methods would provide us with even more consistent and conclusive results, which facilitate us to better understand the dynamics of wildfire occurrences through space and time. By applying several variants of Moran's I method, we would be able to illustrate spatial structure and thus detect heterogeneity in the data. Moreover, Global Moran's I could test for overall spatial autocorrelation, which helps to show the synergistic effect of multiple environmental parameters on wildfire. Furthermore, Local Moran's I would be ideal for detecting hotspots and cold spots in the map, where hotspots represent areas with clusters of elevated risks and cold spots represent areas with low-risk clusters. In addition, Bivariate Moran's I method specializes in exploring co-patterning between each of the climatic parameter and wildfire rate among subregions.

2.1.4 Correlation estimation

Networks are also efficient tools for determining correlations and trends in wildfire occurrences. For instance, recent studies have shown that convolutional neural networks (CNNs) and recurrent architectures can capture spatial and temporal patterns in environmental data more effectively than traditional models. CNN-based model for spatial feature extraction could help benchmark the performance of classical models. This exploratory DL component would operate as an analytical tool, allowing the analysis to benefit from modern predictive approaches without compromising interpretability. Social networks, on the other hand, can provide us with inferences about the correlation between each of the separate incidents. Basically, they are distinct yet subtly correlated with each other, so that the collective relationship is relatively nuanced. For wildfire occurrences, although each of the wildfire is independent of each other, the climatic parameters that lead to them, including wind speed and direction, humidity, as well as temperature, share many common traits. By treating each wildfire incident as a vertex and parameters of occurrences as aspects of vertices, a social network indicating the relationship between each pair of occurrences can be constructed to understand the fundamental mechanics of the hidden correlations. Social networks are widely applied to analyze the relationship between organizations or people, yet attempting to use them to understand wildfire occurrences may provide new and meaningful perspectives.

In general, undirected social networks would provide a more conclusive and comprehensive relationship among wildfires, while directed social networks would reveal more hidden and delicate correlation between each pair of them.

2.1.5 Additional Insights and Extensions

The combination of the Stoyan Grabarnik method, spatial autocorrelation, and network modeling provides a multifaceted understanding of wildfire occurrences in LA County. Global spatial autocorrelation is relatively subtle, yet local clustering patterns and network structures suggest that wildfire events have some non-random spatial and temporal dependencies.

The network-based approach also further highlights that wildfire incidents, though not directly and inherently dependent, can still be connected via shared environmental and climatic conditions, which indicates that wildfire dynamics are influenced by both local interactions and more complicated regional factors.

Apart from statistical approaches, a variety of machine learning methods have been increasingly used to analyze environmental data. For instance, Random Forest models [16] have been applied to analyze complex relationships between environmental variables and wildfire risks. Those models, with the integration of statistics, data science, and environmental science, can discern more complex nonlinear relationships, and further improve predictive performance.

CHAPTER 3

Methods

3.1 Stoyan Grabarnik Method

The Stoyan Grabarnik method is a useful statistical approach for analyzing spatiotemporal point processes. It provides a comprehensive framework for estimating the intensity function of events by aligning local empirical properties with theoretical expectations. The method had been developed using second-order characteristics of point processes, and is commonly applied in spatial statistics to examine residual analysis in point process models.

We can think wildfires as spatiotemporal point processes, in which (x_i, y_i, t_i) can be the spatial and temporal coordinates of the i-th wildfire occurrence event.

The intensity function is:

$$\lambda(x, y, t) = \exp\left(\beta_0 + \beta_1 x + \beta_2 y + \beta_3 t + \beta_4 z_1 + \cdots + \beta_p z_p\right),$$

$\beta_0, \beta_1, \beta_2$, and β_3 are parameters to estimate. $z_1, \ldots, z_p$ represent environmental covariates including vapor pressure deficit and temperature, and they had been considered in extended models.

The Stoyan Grabarnik method estimates the intensity function by partitioning the spatiotemporal domain into bins and then comparing the observed and expected event densities within each bin. The observed number of events is compared with the expected value indicated by model intensity for each bin. The estimation process minimizes the discrepancy between empirical inverse intensity measures and theoretical expectations across all bins. B shows the number of bins, and for each bin b, W_b is the set of observed points in that bin. This is based on comparing the bin volume with the sum of inverse intensities at the observed points.

$$L(\beta) = \sum_{b=1}^{B} \left(\sum_{i \in W_b} \frac{1}{\lambda(x_i, y_i, t_i)} - |B_b| \right)^2 ,$$

In the formula above, $|B_b|$ is bin b 's volume. β is calculated by minimizing the sum of squared differences.

This provides us with flexible modeling of local clustering behavior as well as a baseline framework for understanding wildfire occurrence patterns in LA County, and it can thus help us validate our model assumptions.

3.2 Social Network Analysis

Social network analysis (SNA) is a method used to estimate relationships between entities represented as nodes (actors) and edges (ties). It analyzes structural patterns beyond individual attributes, using graph theory to measure connection strength, flow, and more aspects.

Unlike traditional statistics that treat individuals as independent observations, social network analysis treats individuals as interconnected, and data are thus analyzed at a more structural level. Although wildfire occurrences in Los Angeles County are often sparsely related, social network analysis (SNA) is introduced here to further investigate their underlying relational structure in a comprehensive way.

The wildfire dataset of LA County includes 21,791 events. If all potential spatiotemporal connections in a specific window are to be considered, the resulting network will become extremely dense, with 1,790,891 edges. Such a dense structure is not only computationally expensive but also very difficult to understand, as it suggests that each wildfire event is connected to an excessively large number of subsequent events.

To deal with this issue, a sparse directed network was constructed instead, containing 65,319 edges. In this network, each wildfire event is connected to at most three temporally subsequent events based on spatial proximity. This adjustment maintains the overall local spatiotemporal structure, while offering increased interpretability as well as computational feasibility. The adjusted network has an average out-degree of approximately three, which indicates a more sparse and locally constrained interaction structure.

Due to the large scale of the full network, fitting latent space models such as the Euclidean latent position model becomes computationally intensive. To facilitate model estimation and visualization, a subset of 500 wildfire events was selected for further analysis. This subset intends to preserve the key spatiotemporal characteristics of the original dataset, while providing relatively efficient estimation of the latent structure. Similar sampling methods are often used in network analysis with large-scale graphs as they provide a reasonable approximation of the underlying interaction patterns without enormous computational cost.

The definition of edges significantly influences the structure of the generated network. Each node has at most three three outgoing edges, as each of the wildfire occurrence is connected to a subsequent occurrence in time according to spatial proximity. This should be taken into consideration when we go through the results, as it limits the degree distribution and suggests a specific structure for the network.

The latent space in this model is two-dimensional for visualization, while models with 3 or 4-dimensional latent spaces could exhibit more complicated internal structures of the network, and thus provide us with a more detailed and complex understanding of latent relationships. For further exploration, I also created both 3-dimensional and 4-dimensional graphs, and the results fit the 2-dimensional version, with no significant, observable additional separation. The 2-dimensional graph already exhibits the main features of the network with sufficient details.

3.3 Spatial Autocorrelation Analysis

Spatial autocorrelation analyzes how similar data points are according to their proximity by observing their distribution. Widely used in spatial and environmental statistics, it determines whether nearby locations have similar values of interest using methods including

Moran's I and LISA to get spatial dependency.

The global Moran's I is used to quantitatively estimate the general level of spatial autocorrelation across the area of interest, and indicate if regions with relatively similar wildfire intensities are spatially clustered. LISA strengthens this by marking localized patterns of clustering, enabling the identification of statistically significant hotspots and spatial outliers, and providing a more detailed understanding of the spatial structure of wildfire occurrences.

The global Moran's I statistic is

$$I = \frac{n}{\sum_i \sum_j w_{ij}} \cdot \frac{\sum_i \sum_j w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2},$$

where n represents the number of spatial units, and x_i is the observed value at location i, $\bar{x}$ is the mean of all observations, and w_{ij} represents the spatial weight between locations i and j. A positive Moran's I corresponds to positive spatial autocorrelation, which means that similar values are more likely to be clustered, while a negative value shows spatial dispersion. Values close to zero suggest states that are close to spatial randomness. Here, the spatial weights w_{ij} are defined according to spatial adjacency, which is basically distance. Adjacent regions have relatively higher weights, while distant regions receive lower or zero weights.

The LISA statistic further strengthens the global measures, by estimating the contribution of each location to the general spatial autocorrelation. It provides us with a measure of local spatial autocorrelation. The Local Moran's I statistic for location i is

$$I_i = \frac{(x_i - \bar{x})}{S^2} \sum_j w_{ij}(x_j - \bar{x}),$$

where x_i represents the observed value at location i, $\bar{x}$ is the mean of all observations, and w_{ij} denotes the spatial weight between locations i and j. S^2 is the sample variance.

These methods are practical in wildfire analysis in Los Angeles County, with diversified topographical, climate, and urban conditions in different subregions. Wildfire events are not uniformly distributed, and spatial autocorrelation analysis could be effective for determining geographical clustering and statistical significance of high risk areas. This is indispensable for further understanding regional wildfire dynamics as well as optimizing comprehensive risk management and resource allocation.

CHAPTER 4

Results

4.1 SG Model Comparison and Sensitivity Analysis

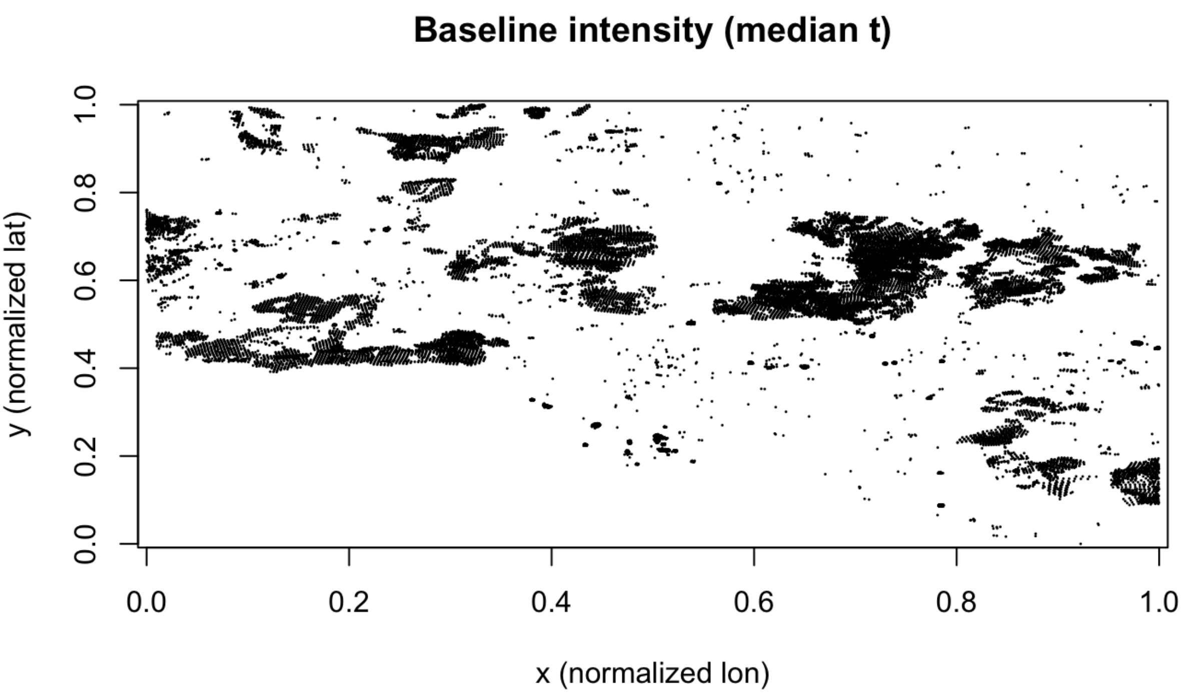

Figure 4.1: Baseline intensity without covariates (8×8×6 binning). The horizontal axis is the normalized longitude within Los Angeles County, and the vertical axis is showing the normalized latitude. Darker colors represent higher estimated intensity, and lighter colors mean lower estimated intensity. The results mostly correspond with our observed records of wildfire occurrences in the study region.

The base model creates a smooth and consistent intensity surface with moderate spatial variation, and the relatively high intensity areas are mostly clustered in specific subareas, signaling the existence of localized wildfire clustering. Figure 4.1 is the baseline intensity figure with 8×8×6 binning. We can see the estimated intensity surface in the study region. The horizontal axes are spatial coordinates, while the color scale means the magnitude of the intensity function. Regions with higher estimated intensity are mostly consistent with the observed wildfire occurrence locations, suggesting that the model has successfully reflected the major spatial distribution of events.

In contrast, the overall structure appears relatively uniform, indicating that the model with no covariates can only recognize more general spatial patterns. Fine-scale variations and sharp transitions in intensity are not clearly represented in this setting, so spatial and temporal coordinates are actually insufficient for fully understanding wildfire conditions in LA County, as other environmental covariates are expected to be included to further improve model accuracy.

Figure 4.2 exhibits the estimated intensity surface with VPD as a covariate and the same spatial coordinate system and color scale as in Figure 4.1.

Figure 4.3, 4.4, and 4.5 exhibit the estimated intensity surface with wind speed, precipitation, and temperature as covariates and the same spatial coordinate system and color scale as in Figure 4.1.

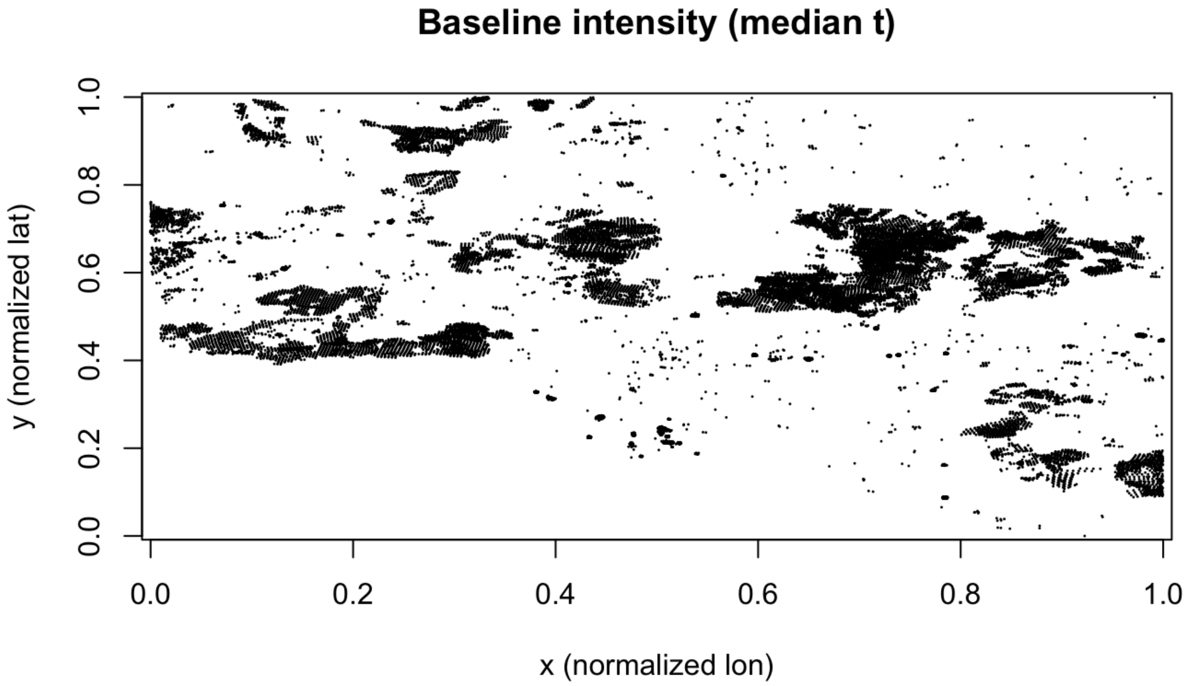

Figure 4.2: Intensity with VPD as covariate (8×8×6 binning). The results mostly correspond with our observed records of wildfire occurrences in the study region similar to Figure 1, but with increased spatial heterogeneity. The result now has more compact clusters and less scattered plots, as well as higher concentrations of high intensity areas. It now shows more significant and localized clusters in the central-east part of the map, indicating the significant influence of VPD. The adjusted model now has better accuracy and trustworthiness. For instance, the higher intensity region around x = 0.7, y = 0.6 looks more concentrated in comparison to the baseline.

The introduction of VPD results in increased spatial heterogeneity in the estimated intensity surface. Higher wildfire intensity regions are now more significant and localized, especially in the central-east area, which indicates the robust influence of VPD on wildfire occurrence variability. The model now recognizes more distinct and concentrated clustering behavior than the base model, especially in areas with environmental conditions susceptible to wildfires. These results suggest that combining climate-related variables improves the accuracy and readability of the model, and VPD seems a vital factor in estimating wildfire risks in Los Angeles County.

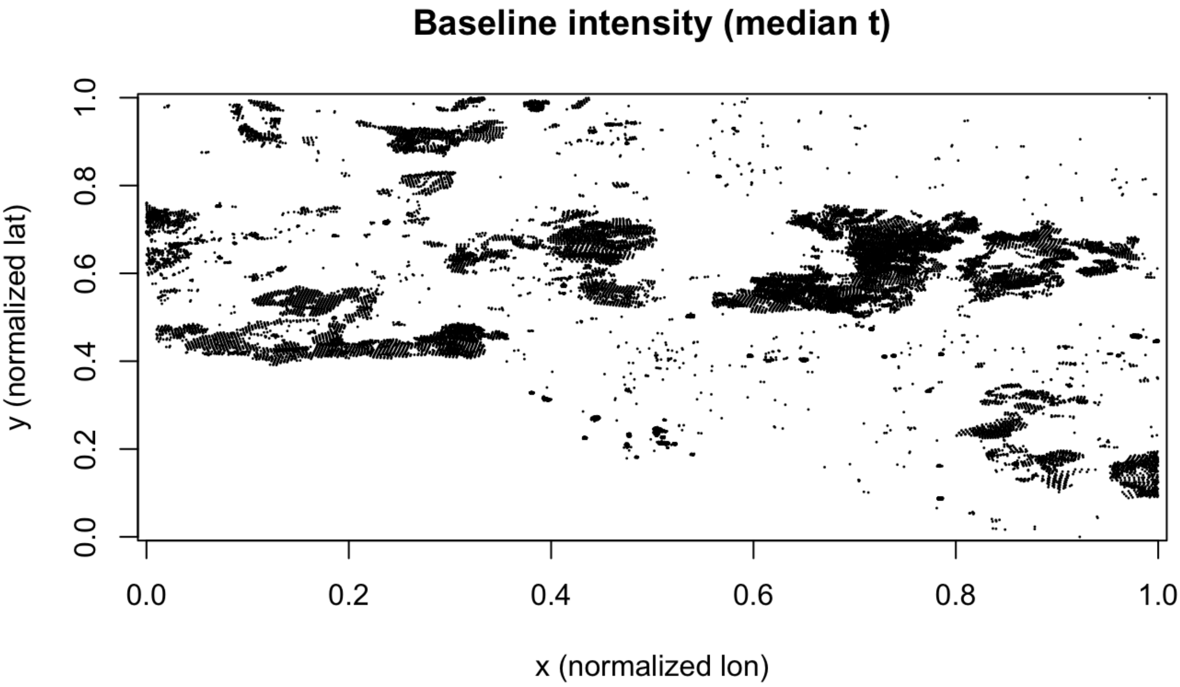

Figure 4.3: Intensity with wind speed as covariate (8×8×6 binning). The results mostly correspond with our observed records of wildfire occurrences in the study region similar to Figure 1, but with slightly increased spatial heterogeneity. The baseline result is mostly preserved in this graph.

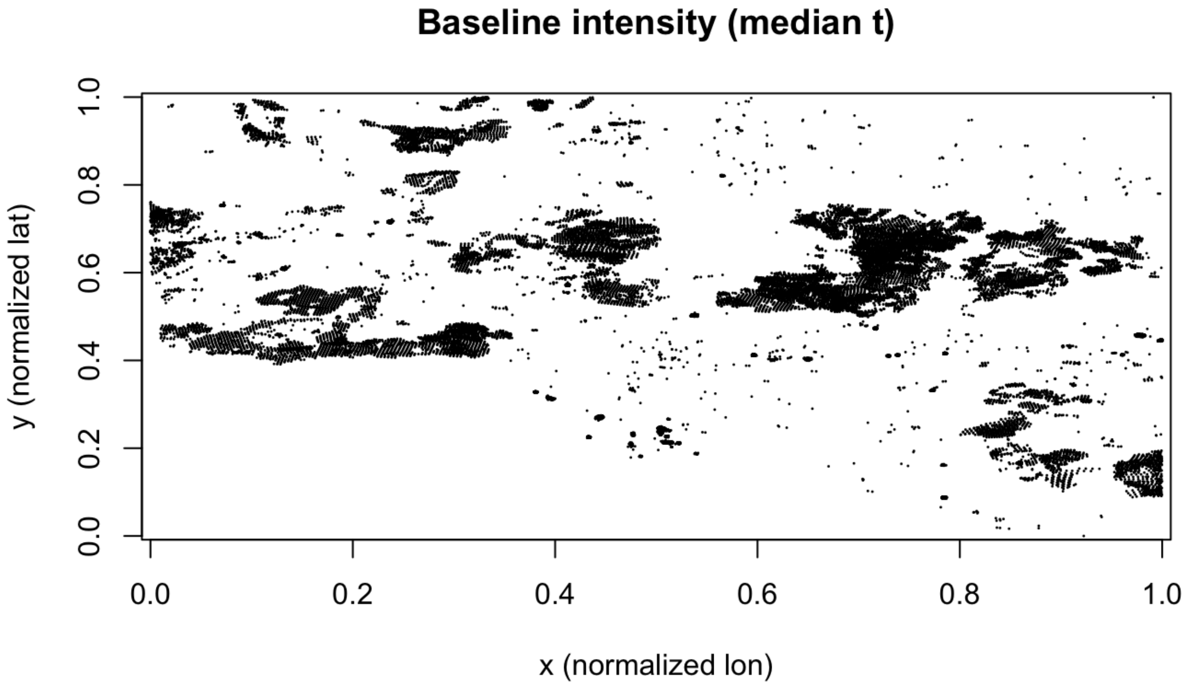

Figure 4.4: Intensity with precipitation as covariate (8×8×6 binning). The results mostly correspond with our observed records of wildfire occurrences in the study region similar to Figure 1. In general, the graph looks almost the same to the baseline result.

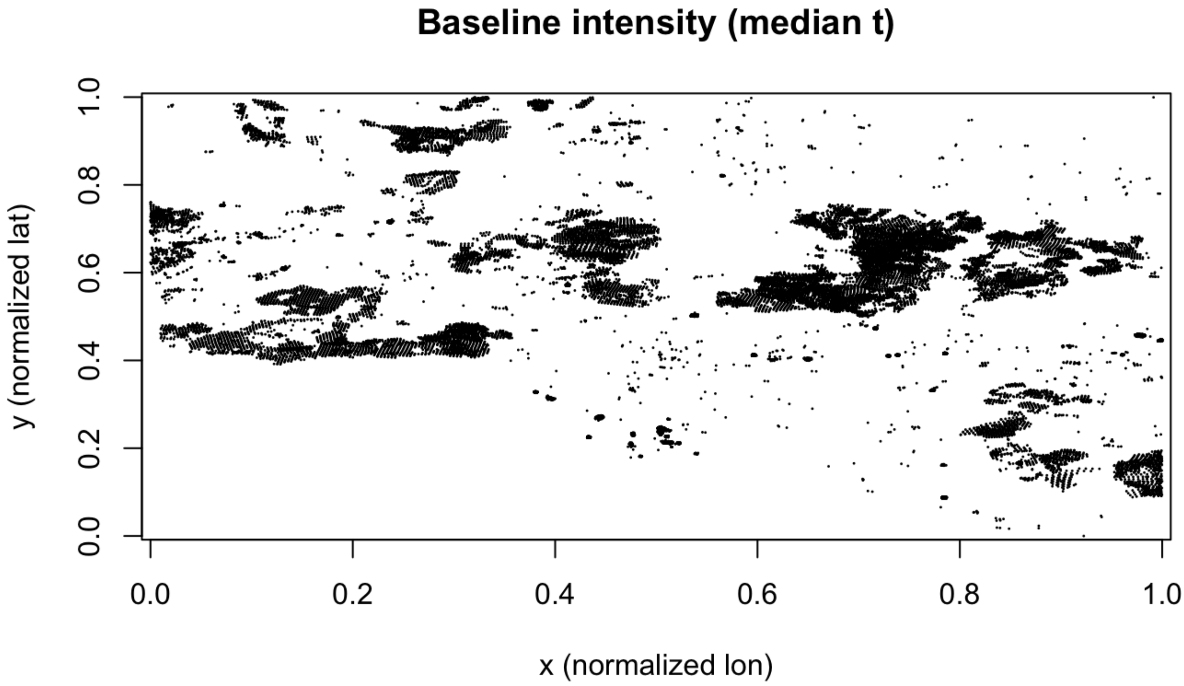

Figure 4.5: Intensity with temperature as covariate (8×8×6 binning). The results mostly correspond with our observed records of wildfire occurrences in the study region similar to Figure 1. The clusters remain mostly unchanged in location and concentration.

The introduction of wind speed, precipitation, or temperature has no significant influence on spatial heterogeneity in the estimated intensity surface. There are only minor changes in very limited local areas. Higher wildfire intensity regions are still mainly in the central-east area, and are only slightly more significant and localized, which indicates the weak influence of those parameters on wildfire occurrence variability. These parameters do not significantly change the spatial structure of wildfire occurrences in Los Angeles County.

Figure 4.6 represents the estimated intensity surface with a finer binning resolution, while keeping the same axes and color scale.

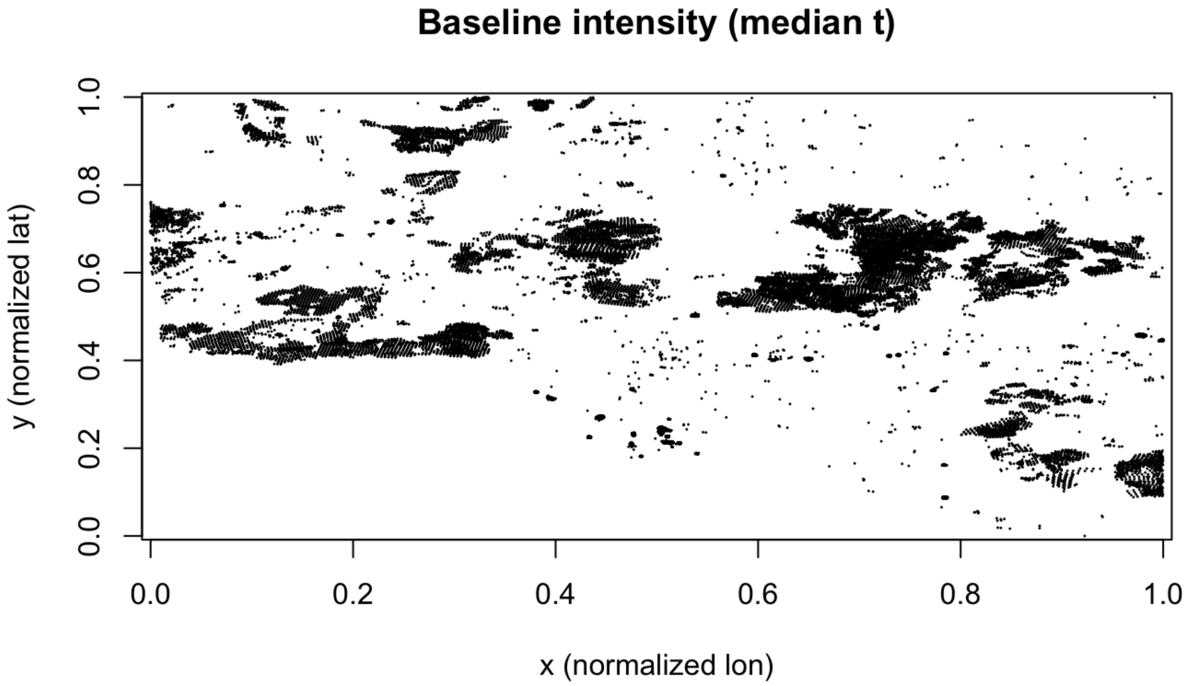

Figure 4.6: Intensity with finer binning resolution (10×10×8 binning). The results mostly correspond with our observed records of wildfire occurrences in the study region similar to Figure 1, but with increased details. The clusters are mostly in the original locations, with slightly increased spatial granularity. It also exhibits the stability and robustness of the model in response to changes to the binning resolution parameter.

A finer binning structure results in more localized variation in the estimated intensity surface. Compared to the setting with coarser binning, the model can now capture more detailed spatial features and smaller-scale clustering patterns. In spite of these local variation disparities, the overall spatial structure remains largely consistent, which indicates that the model is mostly robust to changes in binning resolution. The stability of such results suggests that our conclusions are not sensitive to moderate adjustments in parameter. Such robustness is ideal for precise and effective interpretations of wildfire patterns across different modeling settings.

As a whole, the estimated intensity surfaces exhibit similar large-scale spatial patterns across different model specifications. While local variations may still exist, the overall consistency of the results indicates that the model is sufficiently robust to parameter choices, and thus provides a stable, trustworthy representation of wildfire occurrence patterns in LA County.

4.2 Social Network Analysis

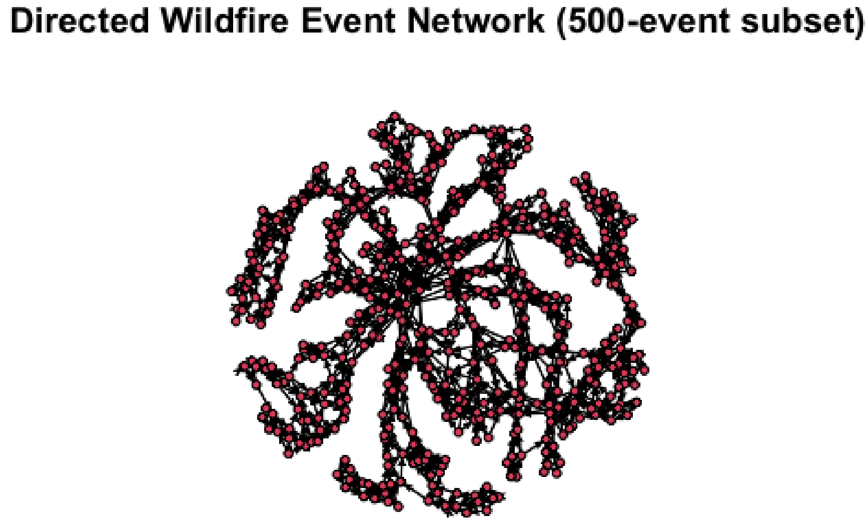

Figure 4.7: Directed wildfire event network (500-event subset). Each node means a wildfire occurrence event, and each directed edge is a connection from a wildfire occurrence to a temporally subsequent event, based on spatial proximity. Each event is connected to up to three most spatially proximal subsequent events. There are 1,488 directed edges, and the graph reflects that each wildfire occurrence event is connected with a limited number of later events. There are several relatively dense local regions in the network, which means that nodes are connected to a large number of subsequent events. In some other regions, the network is sparse. It indicates that wildfire occurrences are often followed by events in limited spatiotemporal ranges. We can also notice that connections usually have limited branching as time passes by.

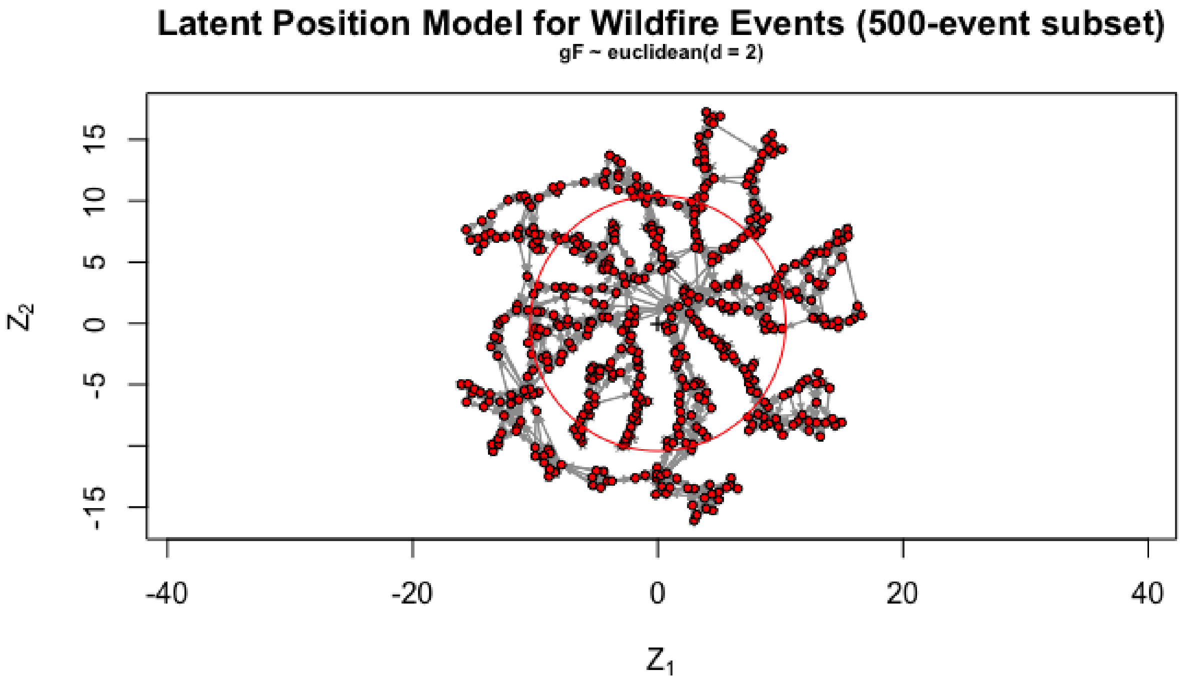

Figure 4.8: Euclidean latent position model with 2D latent space for wildfire events (500-event subset). Each point is a wildfire occurrence event, and the distance between different points now represents the probability of connection in the network. Loosely clustered regions can result from similarities in spatial proximity, temporal occurrence, as well as environmental conditions, and indicate localized interaction patterns among wildfire events. Several distinct clusters can be observed, and they indicate that those events have a higher likelihood to be connected with each other, which could represent wildfire occurrences that are close in both space and time. Nodes that are further apart are less likely to be connected, suggesting limited relationship between distant wildfire occurrences. The network shows non-random connectivity. There are localized structures between those events, as the separated clusters do not align with uniform connections.

For the wildfire network, I used a subset with 500 events and 1,488 directed edges for latent space modeling, and the network visualization (Figure 4.7) exhibits a relatively sparse structure with localized connections, with each event connected with a small number of subsequent events. Each node means a wildfire occurrence, and each directed edge is a connection to a temporally subsequent event based on spatial proximity.

To further investigate the relational structure, a Euclidean latent position model with two-dimensional latent space is utilized. The latent representation created (Figure 4.8) puts wildfire events into a low-dimensional space, and the distance between nodes shows their connection likelihood. Each point means a wildfire event, and the distance between different points represents the probability of connection in the network.

From the latent space plot, we see several loosely clustered regions that suggest the existence of localized interaction patterns among wildfire events. These clusters could result from similarities in spatial proximity, temporal occurrence, or/and shared environmental conditions. The estimated intercept term is also statistically significant and the constructed network exhibits non-random connectivity.

After taking consideration of the latent structure, the baseline connection tendency is now relatively balanced, indicating that the observed network patterns are majorly caused by latent proximity instead of global density.

The latent space model, with its clear representation of wildfire interactions, reveals hidden relational structures over the original geographic coordinates.

4.3 Spatial Autocorrelation Analysis

The Moran scatter plot has an apparent positive slope, and thus the spatial autocorrelation in wildfire occurrences of LA County is positive. Regions with higher wildfire counts are likely to be surrounded by areas with similarly elevated activity levels, suggesting global spatial clustering. In Figure 4.9, the horizontal axis is the standardized wildfire count at each location, and the vertical axis represents the spatially lagged values according to neighboring regions.

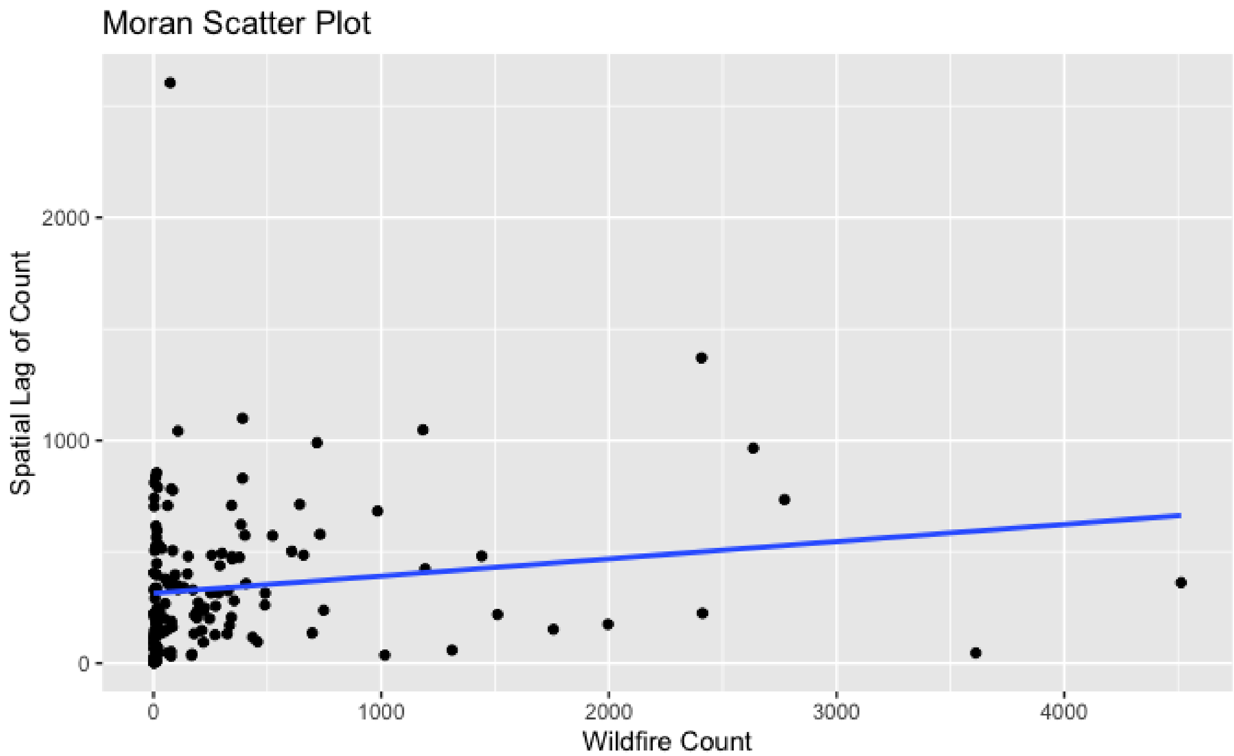

Figure 4.9: Moran scatter plot for wildfire counts using a 14×14 spatial grid. The horizontal axis shows wildfire count at each location, and the vertical axis shows the spatially lagged count according to nearby locations. The slope is positive, and indicates the positive spatial autocorrelation of wildfire events in Los Angeles County.

The LISA cluster map further reveals localized spatial patterns. I identified several High-High regions and indicated concentrated wildfire occurrence hotspots which are surrounded by Low-High regions, exhibiting transitional zones where local wildfire intensity differs from that of nearby areas. In Figure 4.10, each grid cell means a spatial unit, and different colors indicate different types of local spatial association. Most regions aren't statistically significant, showing the dispersed and heterogeneous pattern of wildfire occurrences in LA County. High-High clusters in Figure 4.10 are mostly foothill regions, including parts of Glendale, La Canada Flintridge, Angeles National Forest, and the Palos Verdes Peninsula. The results mostly correspond to WUI areas with increased possibility of wildfire events.

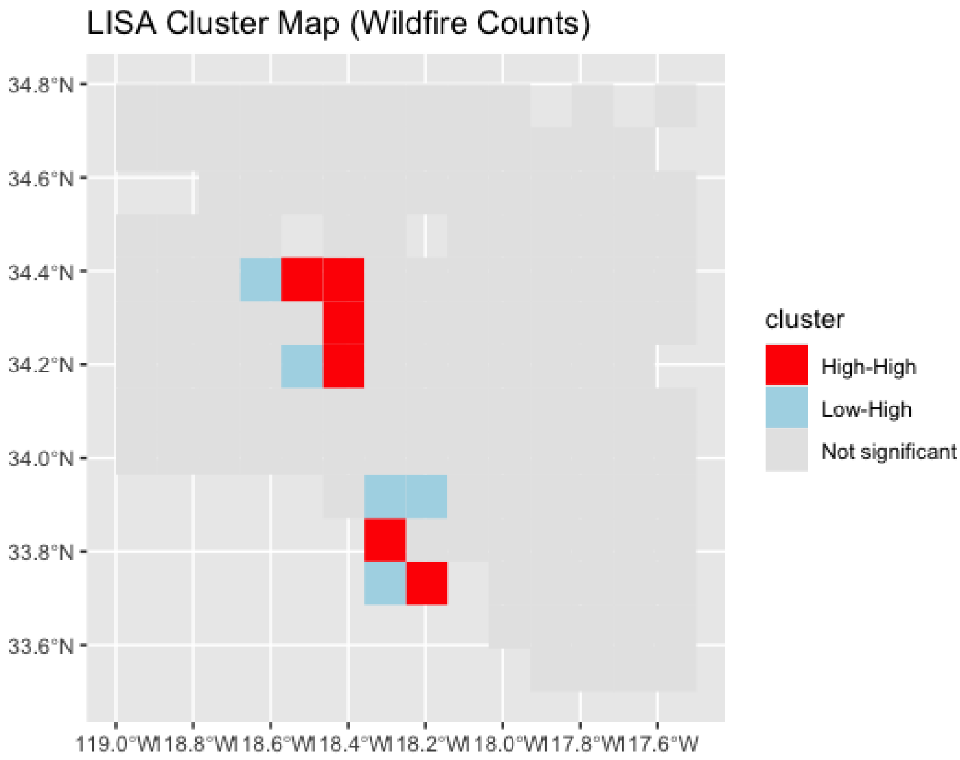

Figure 4.10: LISA cluster map for wildfire counts using a 14×14 spatial grid. The High-high regions marked in red represent regions with high local wildfire intensity that are surrounded by nearby regions with high local wildfire intensity. The Low-high regions marked in blue represent regions with low local wildfire intensity that are surrounded by nearby regions with high local wildfire intensity. Regions marked in darker gray have no significant clusters, and regions marked in lighter gray are not applicable as they are not in Los Angeles County, which is the area of study.

I tested different grid resolutions during the analysis and this 14×14 spatial grid provides a balanced result between spatial detail and statistical stability for viable detection of global and local spatial structures.

To examine the sensitivity of the results to grid resolution, I considered finer 16×16 and 18×18 grids. In comparison to the original grid, 16 by 16 and 18 by 18 grids resulted in more fragmented and less coherent clustering trends, so that overly fine spatial partitioning can potentially increase sparsity and reduce clarity. In Figure 4.11 and 4.13, the axes have the same meaning as in Figure 4.9. The horizontal axis is standardized wildfire counts, and the vertical axis represents spatially lagged values. In Figure 4.12 and 4.14, each grid cell represents a spatial unit as in Figure 4.10. Colors still mean different types of local spatial association. The difference here is that Figure 4.11 and 4.12 results are now based on a 16×16 spatial grid, while Figure 4.13 and 4.14 results are now based on a finer 18×18 spatial grid.

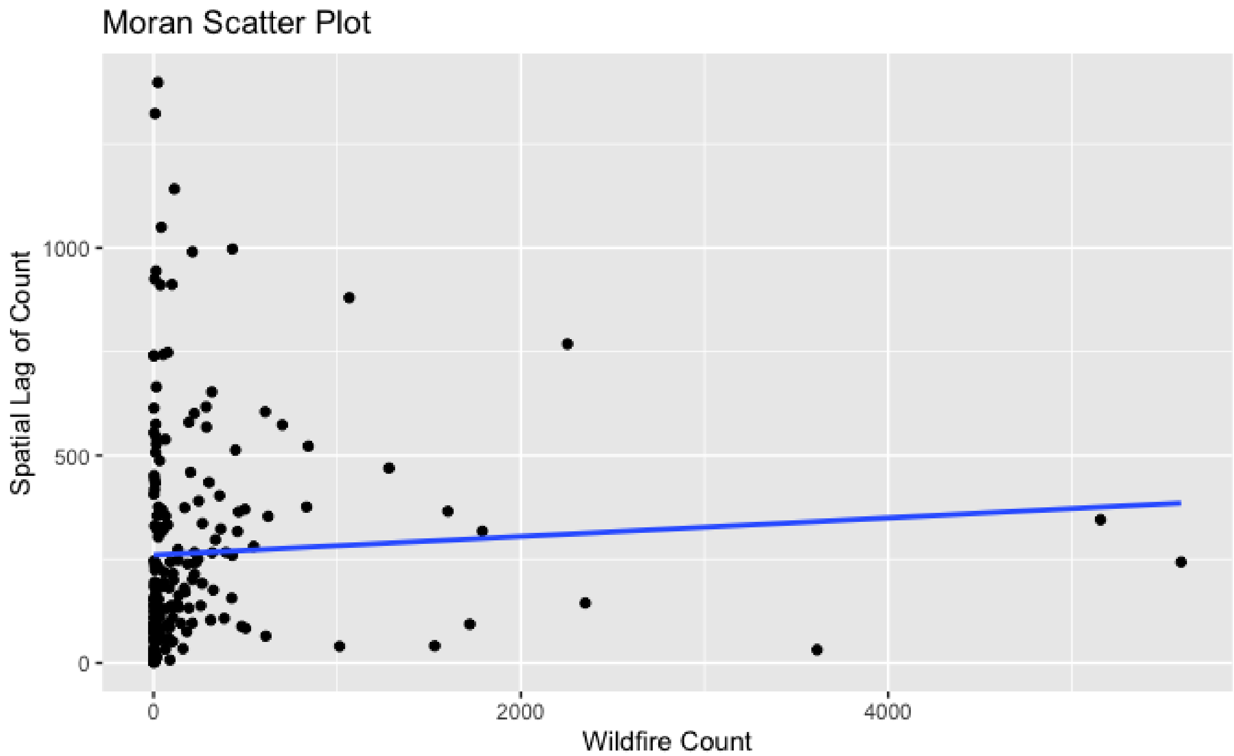

Figure 4.11: Moran scatter plot using a 16×16 spatial grid. The slope is positive, and indicates the positive spatial autocorrelation of wildfire events in Los Angeles County, which aligns with the result shown in Figure 4.9.

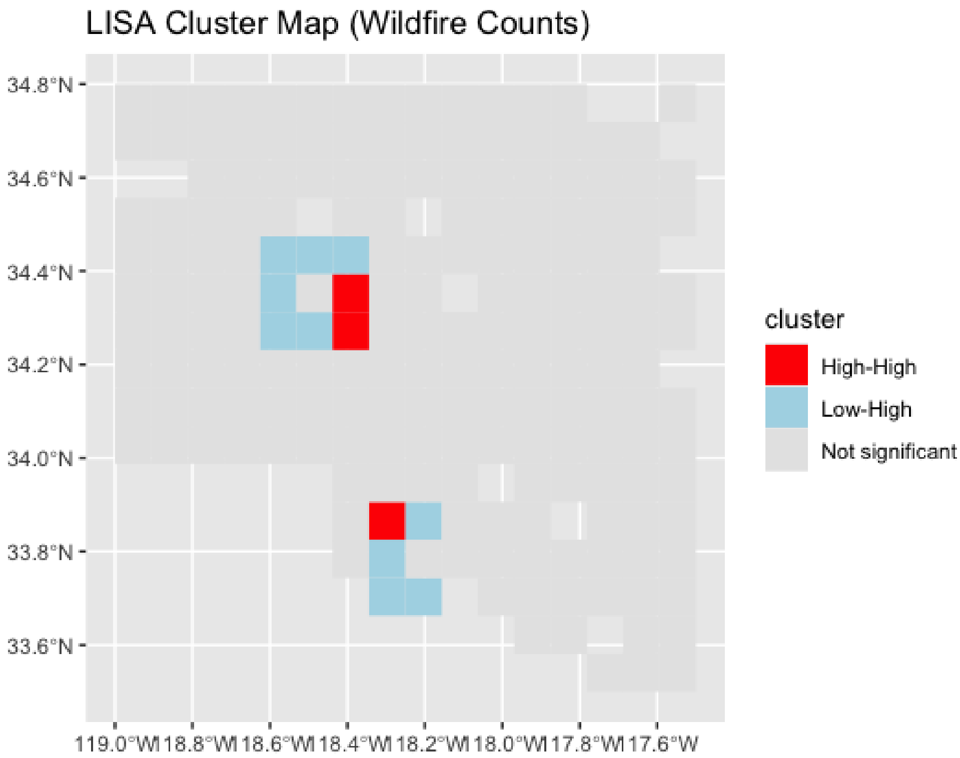

Figure 4.12: LISA cluster map using a 16×16 spatial grid. This graph has increased sparsity and decreased clarity in comparison to the LISA Cluster Map generated using a 14×14 spatial grid, but still shows consistent results. Moreover, the model exhibits some sensitivity to the selection of grid resolution parameter.

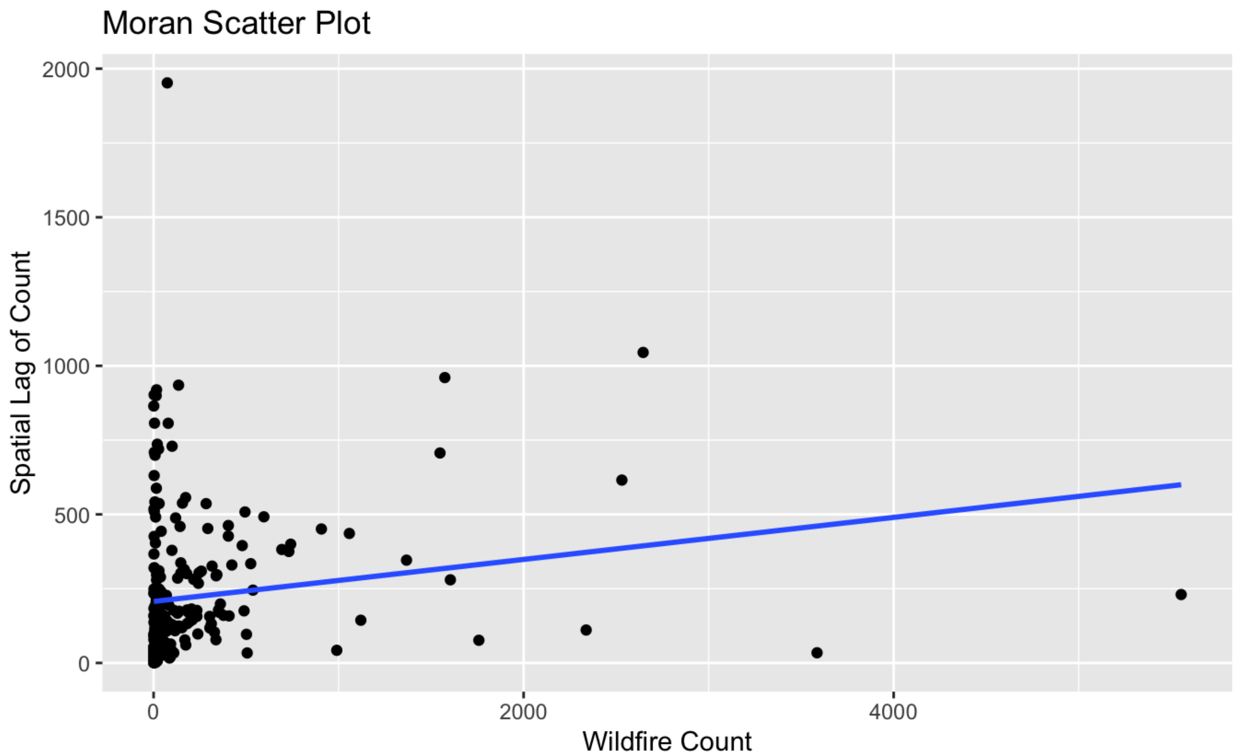

Figure 4.13: Moran scatter plot using a 18×18 spatial grid. The slope is positive, and indicates the positive spatial autocorrelation of wildfire events in Los Angeles County, which aligns with the result shown in Figure 4.9.

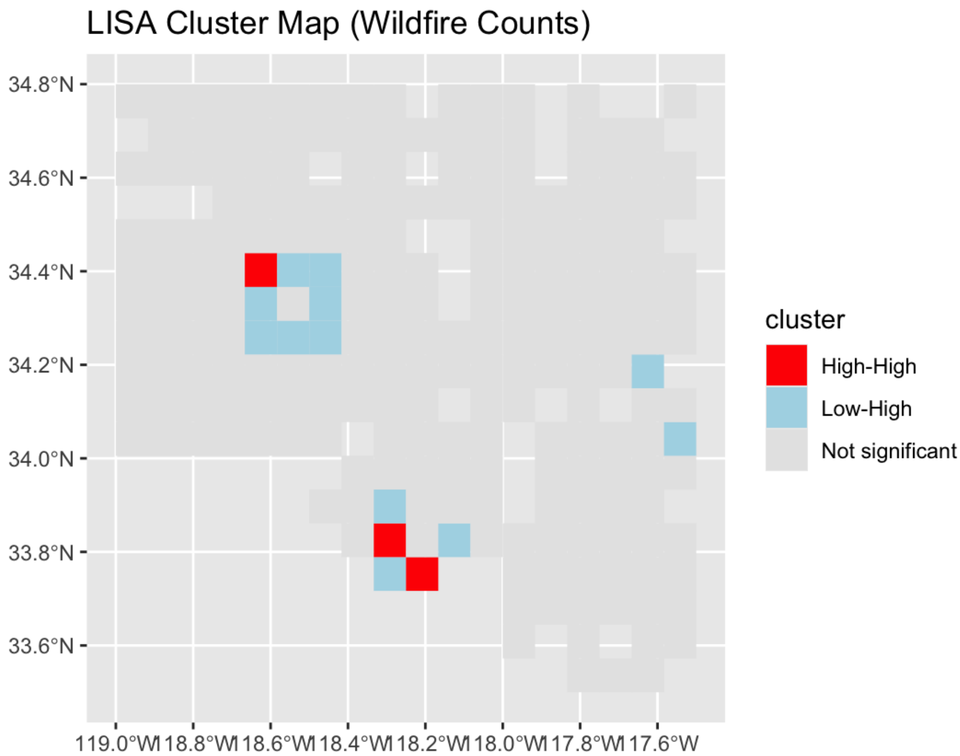

Figure 4.14: LISA cluster map using a 18×18 spatial grid. This graph has significantly increased sparsity, decreased clarity, as well as more noisy and fragmented spatial patterns in comparison to the LISA Cluster Map generated using a 14×14 spatial grid, suggesting the drawbacks of overly fine spatial partitioning.

I also considered coarser 12×12, 10×10 and 8×8 grids. In comparison to the original grid, 12 by 12 grid resulted in weaker positive slope in the Moran Scatter Plot and less significant clustering pattern in the LISA Cluster Map, so that coarser spatial partitioning can potentially smooth out local spatial patterns and weaken spatial autocorrelation.

10×10 and especially 8×8 grids resulted in negative slopes in the Moran Scatter Plot and insignificant clustering pattern in the LISA Cluster Map, suggesting a weak, negative spatial autocorrelation. This indicates that overly coarse spatial partitioning could even distort the underlying spatial structure.

In Figure 4.15, 4.17 and 4.19, the axes have the same meaning as in Figure 4.9. The horizontal axis is standardized wildfire counts, and the vertical axis represents spatially lagged values. In Figure 4.16, 4.18 and 4.20, each grid cell represents a spatial unit as in Figure 4.10. Colors still mean different types of local spatial association. The difference here is that Figure 4.15 and 4.16 results are now based on a 12×12 spatial grid, while Figure 4.17 and 4.18 results are now based on a 10×10 spatial grid. Figure 4.19 and 4.20 results are now based on a 8×8 spatial grid.

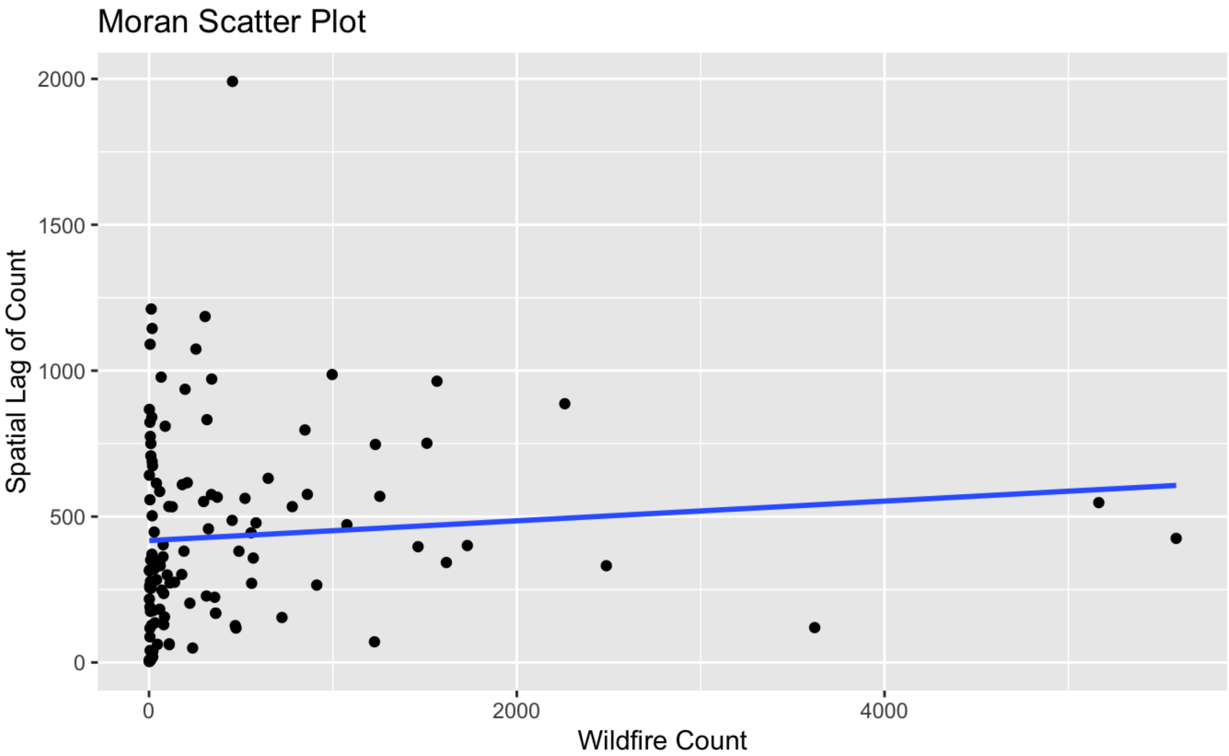

Figure 4.15: Moran scatter plot using a 12×12 spatial grid. The slope is positive, and indicates the positive spatial autocorrelation of wildfire events in Los Angeles County, which aligns with the result shown in Figure 4.9. 12 by 12 grid resulted in a weaker positive slope, though.

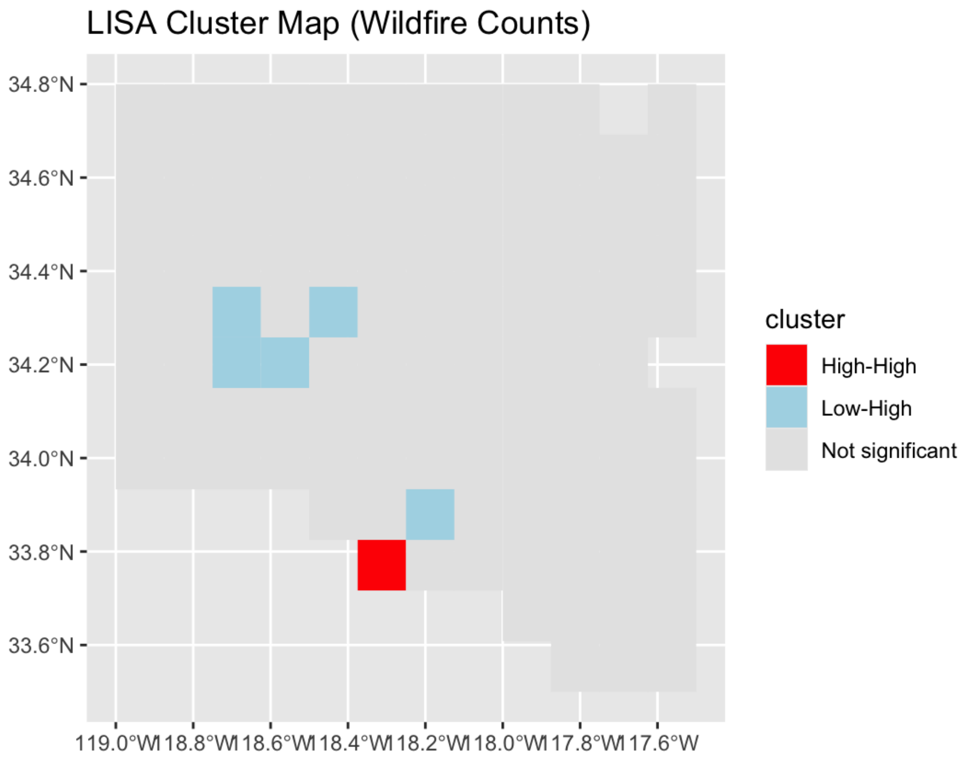

Figure 4.16: LISA cluster map using a 12×12 spatial grid.This graph has a less significant clustering pattern in comparison to the LISA Cluster Map generated using a 14×14 spatial grid, suggesting that coarser spatial partitioning can potentially smooth out local spatial patterns and thus weaken spatial autocorrelation.

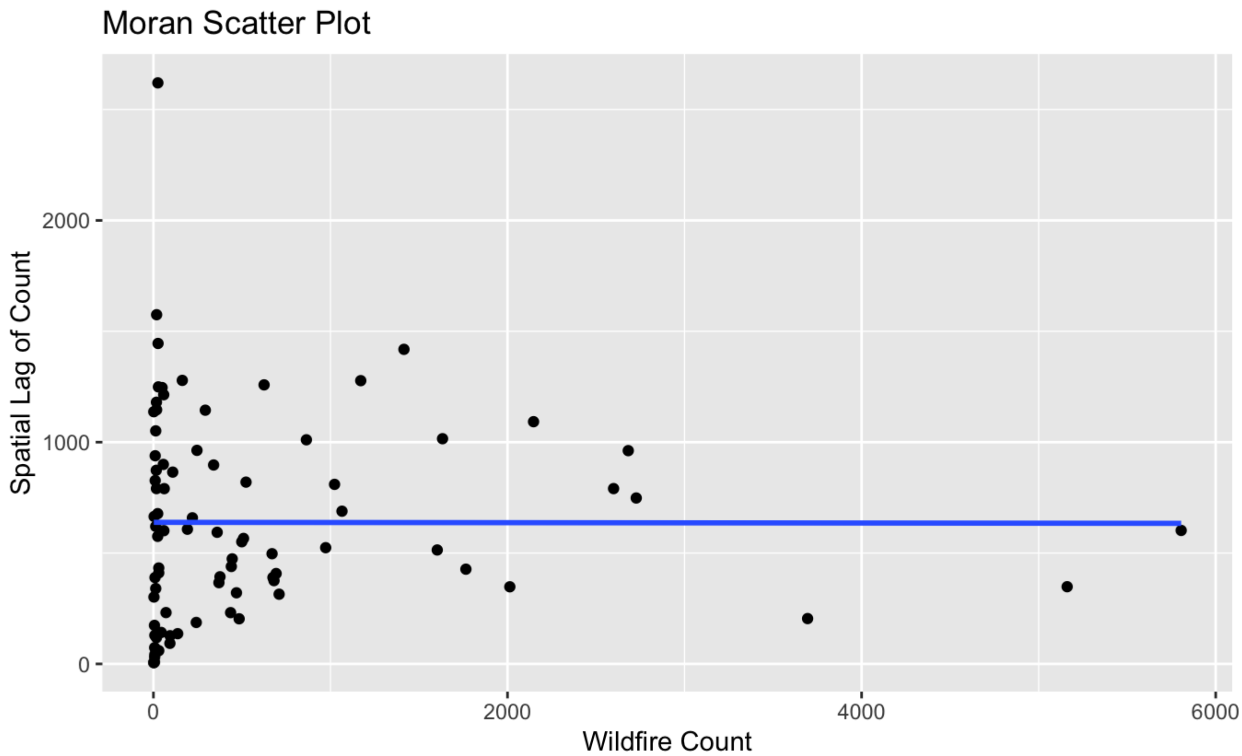

Figure 4.17: Moran scatter plot using a 10×10 spatial grid. The slope is now slightly negative(but very near 0), and indicates the very weak negative spatial autocorrelation of wildfire events in Los Angeles County, which does not align with the result shown in Figure 4.9. This indicates that overly coarse spatial partitioning could even distort the underlying spatial structure.

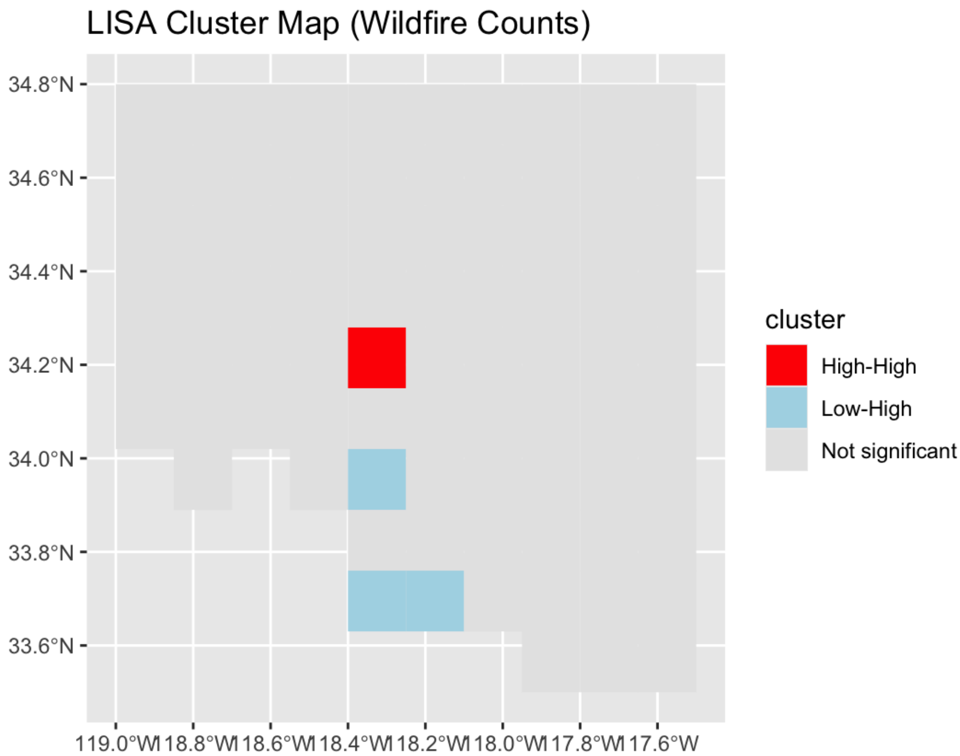

Figure 4.18: LISA cluster map using a 10×10 spatial grid. This graph has a less significant clustering pattern in comparison to the LISA Cluster Map generated using a 14×14 spatial grid, suggesting that overly coarse spatial partitioning can smooth out local spatial patterns and thus distort spatial autocorrelation.

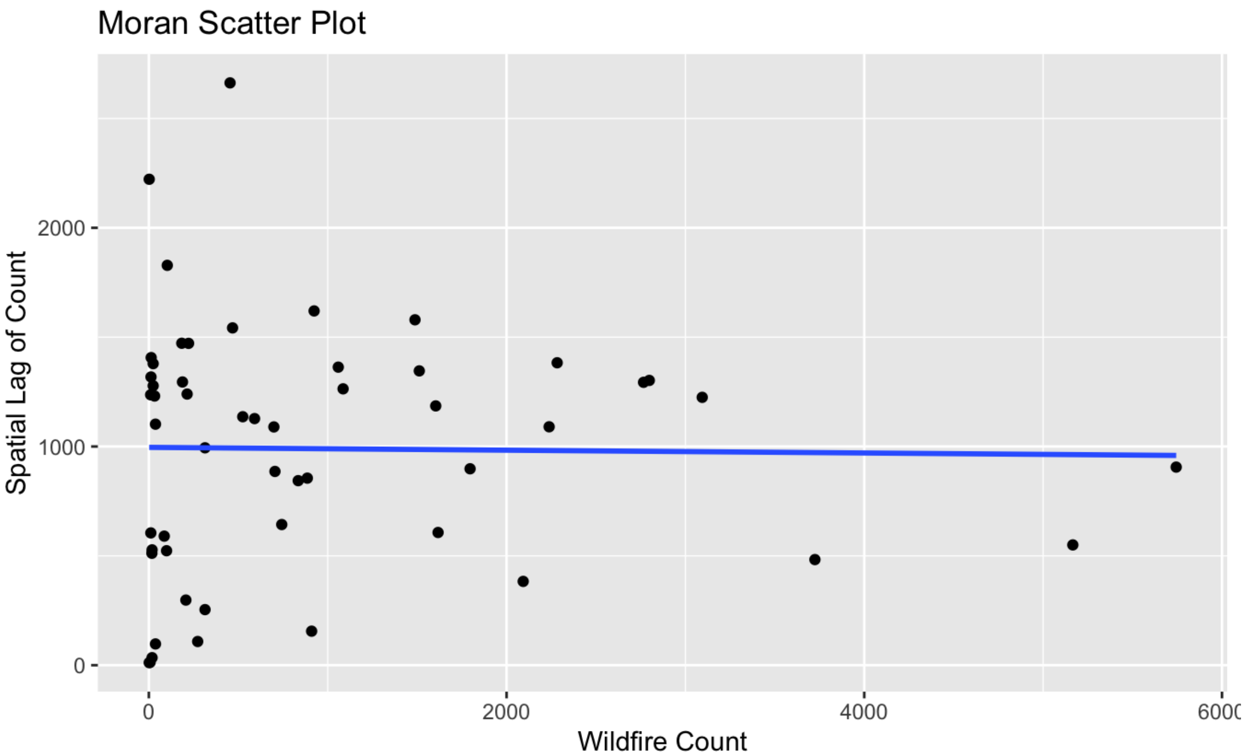

Figure 4.19: Moran scatter plot using a 8×8 spatial grid. The slope is now apparently negative, and indicates the negative spatial autocorrelation of wildfire events in Los Angeles County, which does not align with the result shown in Figure 4.9. This indicates that overly coarse spatial partitioning could even distort the underlying spatial structure.

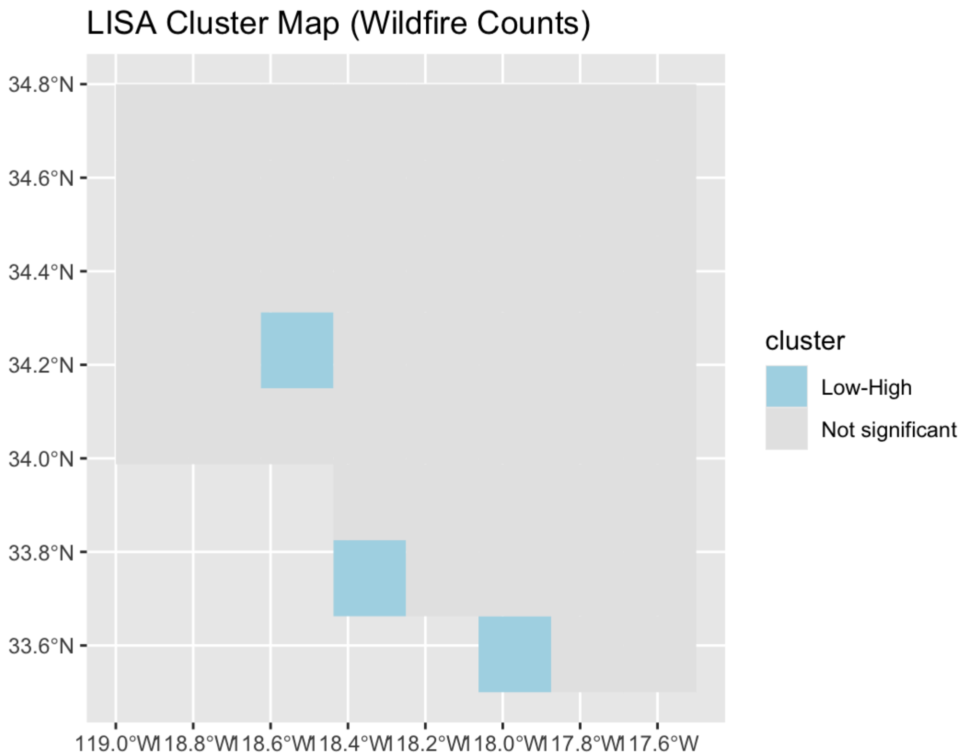

Figure 4.20: LISA cluster map using a 8×8 spatial grid. This graph has a less significant clustering pattern in comparison to the LISA Cluster Map generated using a 14×14 spatial grid, suggesting that overly coarse spatial partitioning can smooth out local spatial patterns and thus distort spatial autocorrelation.

CHAPTER 5

Conclusion

This thesis conducts a comprehensive analysis of wildfire occurrences in Los Angeles County by utilizing spatial, spatiotemporal, as well as network-based approaches.

The Stoyan Grabarnik method provides a baseline estimation of wildfire intensity, and shows spatial and temporal trends of wildfire occurrences. The results indicate that wildfire intensity varies across space and time, yet certain subregions exhibit higher baseline risks.

Spatial autocorrelation analysis based on Moran's I and LISA methods concludes that localized clustering patterns exist, although global clustering is relatively weak. Several hotspots have been identified, suggesting that wildfire occurrences are not entirely random and have spatial dependence at a local scale.

The social network approach further supports these prior findings by modeling relationships between individual wildfire events. Although wildfire occurrences are not directly dependent, the constructed network suggests that events may be indirectly connected via shared environmental and climatic conditions.

The combination of these methods provides a comprehensive, multifaceted understanding of wildfire dynamics, exhibiting both global trends and local structures. There are several limitations in this study, though. The choice of spatial resolution and threshold parameters may influence the results, and the constructed network is based on simplified proximity assumptions. Additionally, the analysis focuses on interpretability rather than predictive performance.

Future wildfire research can integrate more advanced models, including various machine learning methods to understand more complex nonlinear relationships, and thus further improve efficiency and accuracy. Further interactions of environmental covariates and higher-resolution data may also facilitate our understandings of wildfire behavior in regions with different topological, climatic, and urban development patterns.

In conclusion, it indicates that the combination of statistical and network-based approaches provides indispensable insights into wildfire occurrence patterns, and thus provides a useful angle for future research and risk control plans.

REFERENCES

[1] P. A. P. Moran. Notes on continuous stochastic phenomena. *Biometrika*, 1950.

[2] Luc Anselin. Local indicators of spatial association—lisa. *Geographical Analysis*, 1995.

[3] Yavar Pourmohamad, John T. Abatzoglou, Erica Fleishman, Karen C. Short, Jacquelyn Shuman, Amir AghaKouchak, Matthew Williamson, Seyd Teymoor Seydi, and Mojtaba Sadegh. Inference of wildfire causes from their physical, biological, social and management attributes. *Earth's Future*, 13(1), 2025.

[4] Bin Chen, Shengbiao Wu, Yufang Jin, Yimeng Song, Chao Wu, Sergey Venevsky, Bing Xu, Chris Webster, and Peng Gong. Wildfire risk for global wildland–urban interface areas. *Nature Sustainability*, 7:474–484, 2024.

[5] Yongxuan Guo, Jianghao Wang, Yong Ge, and Chenghu Zhou. Global expansion of wildland-urban interface intensifies human exposure to wildfire risk in the 21st century. *Science Advances*, 10(45), 2024.

[6] Wenfu Tang, Cenlin He, Louisa Emmons, and Junzhe Zhang. Global expansion of wildland-urban interface (wui) and wui fires: Insights from a multiyear worldwide unified database (wuwui). *Environmental Research Letters*, 19(4), 2024.

[7] Katie Hoover and Laura A. Hanson. Wildfire statistics. Technical report, Congressional Research Service, 2021.

[8] Stephen W. Taylor, Douglas G. Woolford, C. B. Dean, and David L. Martell. Wildfire prediction to inform fire management: Statistical science challenges. *Statistical Science*, 28(4):586–615, 2013.

[9] Dexen D. Z. Xi, Stephen W. Taylor, Douglas G. Woolford, and C. B. Dean. Statistical models of key components of wildfire risk. *Annual Review of Statistics and Its Application*, 6(1):197–222, 2019.

[10] Abolfazl Jaafari, Davood Mafi-Gholami, Binh Thai Pham, and Dieu Tien Bui. Wildfire probability mapping: Bivariate vs. multivariate statistics. *Remote Sensing*, 11(6):618, 2019.

[11] Medha Uppala and Mark S. Handcock. Modeling wildfire ignition origins in southern california using linear network point processes. *The Annals of Applied Statistics*, 14(1):339–356, 2020.

[12] Heloise Hesseln. Wildland fire prevention: A review. *Current Forestry Reports*, 2018.

[13] Vytenis Babrauskas. The palisades fire of los angeles: Lessons to be learned. *Fire*, 8(8):303, 2025.

[14] A. Stewart Fotheringham, Wenbai Yang, and Wei Kang. Multiscale geographically weighted regression (mgwr). *Annals of the American Association of Geographers*, 2017.

[15] Antonio Gasparrini, Ben Armstrong, and Michael G. Kenward. Distributed lag non-linear models. *Statistics in Medicine*, 2010.

[16] Leo Breiman. Random forests. *Machine Learning*, 2001.