Point Pattern Analysis: Wildfires in NYS: 2008-2018

Samantha Benedict
Professor Dana Sylvan
STAT 787
City University of New York, Hunter College

Abstract

New York State has seen a fluctuation in the abundance of wildfires between the years of 2008 and 2018. In this project we study a point pattern analysis of the number of wildfires in New York State over this 11 year span and suggest a possible covariate, yearly average rainfall, to explain the specific clustering that exists in the Adirondack and Catskill regions as well as parts of eastern Long Island.

1. Introduction

1.1 Data

This data was collected by the New York State Department of Environmental Conservation Division of Forest Protection. Consisting of 28 variables and 1,643 observations, we were most interested in the latitudinal and longitudinal coordinates. The data was available in csv format and using ArcGIS, converted into shapefiles, designated by year. All of the analysis was done using R 3.6.1.

The entirety of the spatial analysis performed for this report was done using the spatial package spatstat. Many of the functions required raster files and again, these were converted from the original data in ArcGIS. Other data conversions were done in R using the maptools package.

Additionally, annual precipitation averages per latitude and longitude locations were obtained from PRISM Climate Group Repository. PRISM gathers climate observations from a range of monitoring networks and develops spatial climate datasets using modeling techniques to acquire weather averages across space at the yearly and monthly level. [1]

2. Exploratory Analysis

In this report, we will explore the question, "Are wildfires clustered in areas where there is lower than average yearly rainfall?" To do this we must first look at basic point pattern analysis techniques, where we will observe the locations of wildfire events within the bounded region of New York State and compare these analyses for each individual year within our 11 years of available data. We will now explore both 1st and 2nd order properties measuring both intensity and spatial dependence. Initially, we would like to determine if there is possible dependence between each year.

2.1 Are Wildfires Amongst Years Independent?

In order to determine if the years are independent or whether a temporal pattern exists, we must first plot the frequencies of wildfires per year to ascertain whether or not a visual pattern appears evident. We can see in **Figure 1**, there appears to be possible seasonality. Further time series analysis should be done to determine this, however 11 years might not be long enough to make a definitive conlcusion.

Regardless, it is worth looking into each individual year of point patterns to compare the results considering the vast fluctuation between years. Our goal is to determine if this yearly fluctuation could be in part influenced by average yearly rainfall.

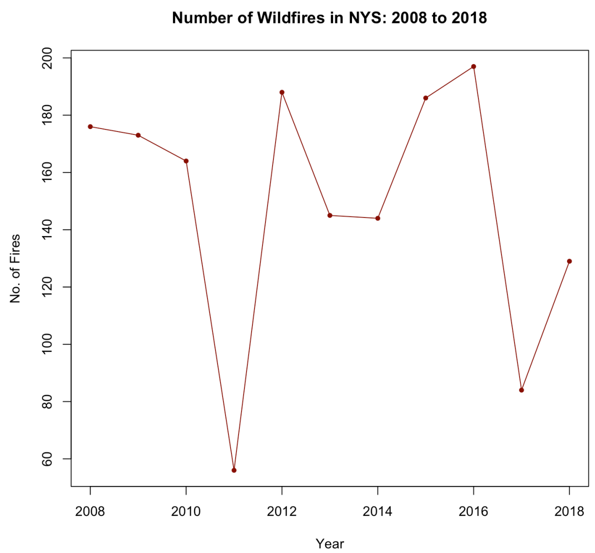

Figure 1 Frequency Plot of Wildfires

2.2 Quadrat Methods

The analysis of point patterns is focused on the spatial distribution of observed events and inferences about the underlying processes that generate them. [2] The first step is to establish whether the events exhibit a systematic pattern over the spatial bounds as opposed to complete spatial randomness (CSR). To determine this, we use a chi-squared test. Using a shapefile gathered from New York State data clearing house, the data points were constrained by state boundaries. After trial and error, an optimum grid was chosen based on point counts within each quadrat. From here, a simple chi-squared test was performed in R.

It is important to note that the same grid size was used for each wildfire year. This allowed for easy comparison between yearly maps. Let us first look at the wildfires from 2018.

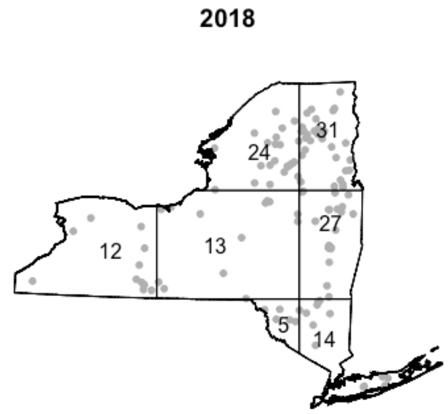

Figure 2(a) Quadrat Map of Wildfires from 2018

Figure 2(a) shows a plotted point pattern of all the wildfires that occurred in New York State in 2018. The spatial distribution of events appears not to follow a homogenous poisson process. Upon further investigation, quadrat analysis confirmed this statement.

We consider the null hypothesis: The point pattern is CSR. After examining the chi-squared test at an alpha level of .05, we obtain a P-value less than 2.2e-16. Therefore we can reject the null hypothesis and determine with 95% confidence that our point pattern is not CSR.

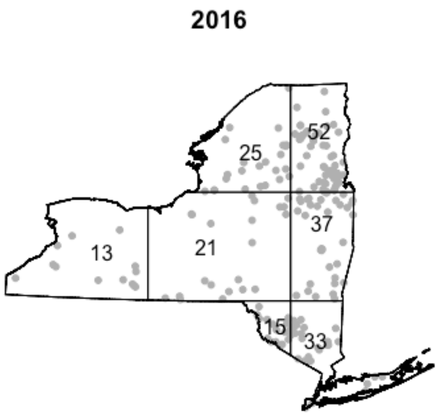

Figure 2(b) Quadrat Map of Wildfires: 2016 (year with most wildfires)

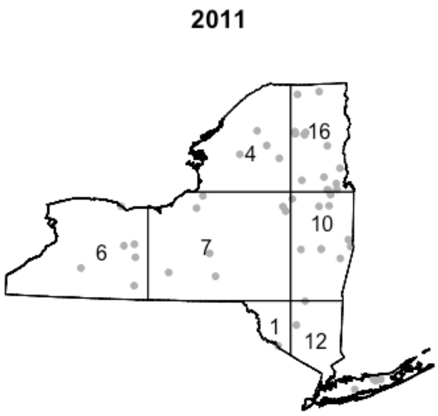

Figure 2(c) Quadrat Map of Wildfires: 2011 (year with least wildfires)

This same quadrat analysis was done for the remaining ten years, each yielding us with the same conclusion: none of the 11 years of wildfire point patterns exhibted complete spatial randomness. **Figure 2(b)** shows the point pattern from the year with the most wildfires and **Figure 2(c)** shows the point pattern from the year with the least wildfires. For the remainder of this report, we will continue to look at the years 2018, 2016, and 2011, the most current year with 129 wildfires, the highest wildfire year with 197 wildfires, and the lowest wildfire year with 56 wildfires, respectively all for purposes of comparison.

Some weaknesses pertain to the quadrat method for determining spatial intensity. In order to use a chi-squared test, each quadrat should contain at least five points. [3] Due to the nature of dispersion, this was an impossible task which could lead to inaccurate results.

2.3 Kernel Density

After determining that CSR is not present in any of the 11 years of spatial data, we must decide whether uniformity exists within the spatial presence of wildfires or if there is systematic clustering. We can visualize this using kernel smoothing which measures intensity of events based on a specified search radius around each event. [2] We call this a bandwidth which we must postulate in R using trial and error. **Figures 3(a)-(c)** depict kernel density heat maps all with specified bandwidths.

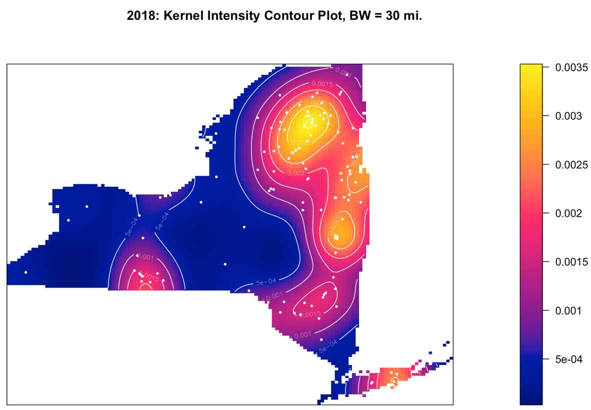

Figure 3(a) Kernel Estimation with bandwidth=30 miles (2018)

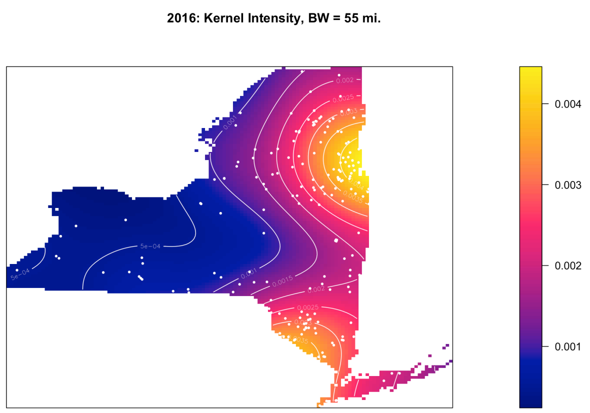

Figure 3(b) Kernel Estimation with bandwidth=55 miles (2016)

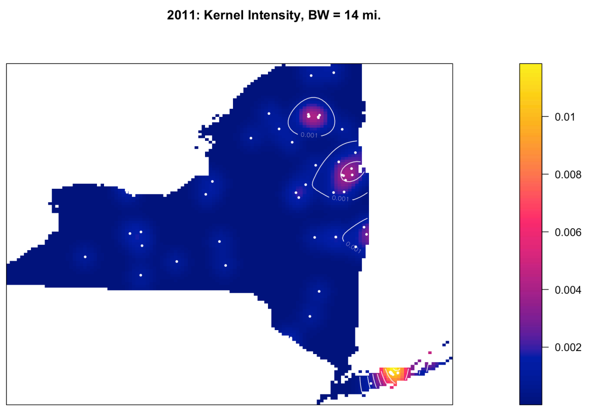

Figure 3(a) Kernel Estimation with bandwidth=14 miles (2011)

After much trial and error, the optimal bandwidths for the heat maps in **Figures 3(a), (b) and (c)** were 30 miles, 55 miles, and 14 miles, respectively. In **Figures 3(a)-(c),** the concentrated zones are displayed by the yellow, orange and slightly pink shaded areas and as the color darkens, the events disperse. Contour lines have been added over top the heat maps to strengthen this visualization. Additonally, the default isotropic Gaussian kernel was used for each of the kernel density maps in this report due to the minimal change that each of the other kernel choices produced.

In **Figure 3(a)** we can see evident clusters in the northern Adirondack region of New York, trickling down the east coast (upper Catskills) as well as clustering in eastern Long Island (Long Island State Pine Barrens). 2018 also shows a cluster in the bottom western part of the state. Both **Figures 3 (b) and (c),** portray similar locations of clusters, again in the Adirondack region, upper Catskills, and Long Island Central Pine Barrens area. These concentrated areas localized in the same general regions for all of the remaining years of wildfire point patterns.

2.4 Nearest Neighbor Analysis: G and F-Functions to Detect Clustering

After exploring 1^{st} order properties to measure intensity, we must now consider 2^{nd} order properties to measure spatial dependence among events within each year of the 11 years of point patterns. The first method of nearest neighbor analysis that will be discussed is the use of the G-Function. This is the simplest method of measuring spatial dependence and examines the cumulative frequency distribution of the nearest neighbor distances between all the events within the spatial bounds of the data. [4]

In **Figures 4(a) and 4(b)** the G-Functions have been graphed for the years 2018 and 2011. Both figures show the observed events indicated by the solid black line. If this line is above the grey confidence interval, this indicates the presence of clusters. **Figure 4(a)** shows the observed line above the envelope and displays that at lower distances, the points are closer together than at higher distances thus satisfying the requirements for clusters for the G function**. Figure 4(b)** also indicates the presence of clusters, however the results are not as dramatic. These results make sense since the clustering appears to be less intense for the year 2011.

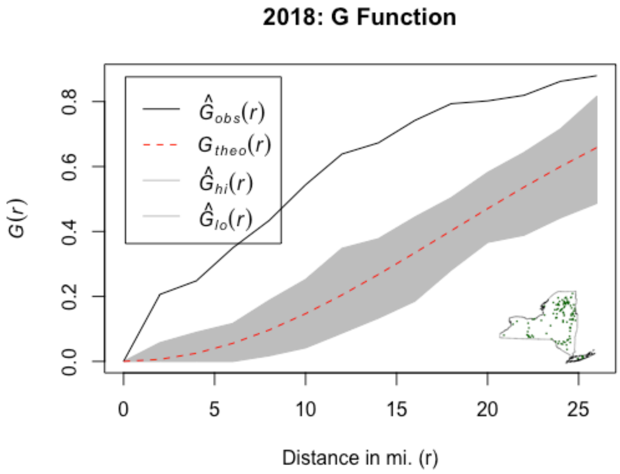

Figure 4(a) G-Function to detect clusters for the year of 2018

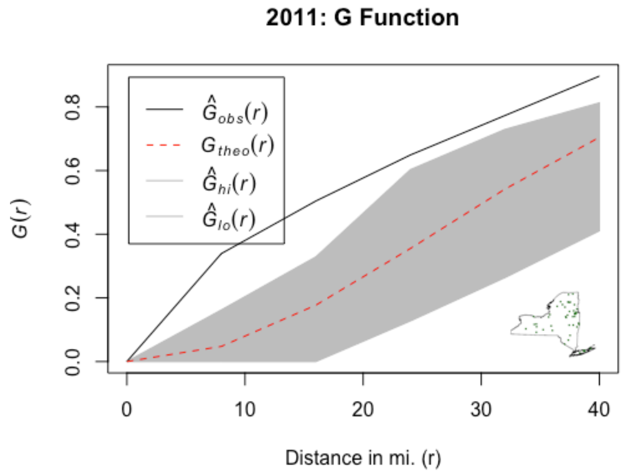

Figure 4(b) G-Function to detect clusters for the year of 2011

Like the G-Function, the F-Function measures spatial dependence among events in a point pattern. The F-Function selects a sample of event locations anywhere in the boundaries of a point pattern at random and calculates the minimum distance to every other event in the spatial bounds.[4]

The F-Function for 2018 and 2011 can be seen in **Figures 4(c) and (d).** Similarly to the G-Function, clustering is present if the observed line falls below the confidence envelope for the F-Function. Again both of these plots indicate the presence of clusters. If the observed values had fallen within the envelopes, CSR would exist but if they had fallen above, we could have considered the point pattern regularly spaced.

Again, **Figure 4(d)** confirms our conclusion that 2011 had less extreme clustering and after looking at all 11 G and F-Functions, 2011 was deemed to be the year that exhibited events closest to CSR but still slightly clustered. All 11 years showed the presence of wildfire clustering.

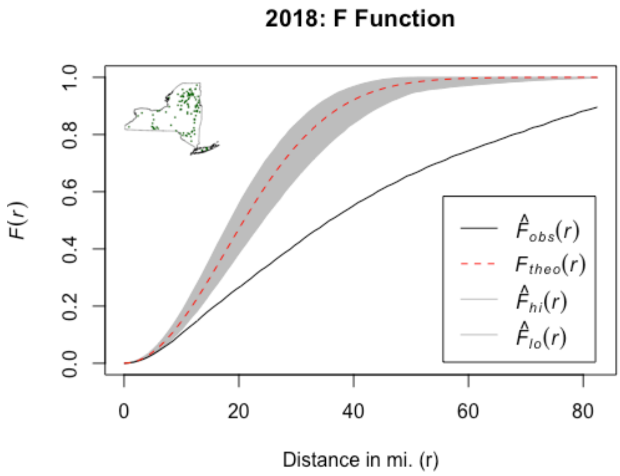

Figure 4(c) F-Function to detect clusters for the year of 2018

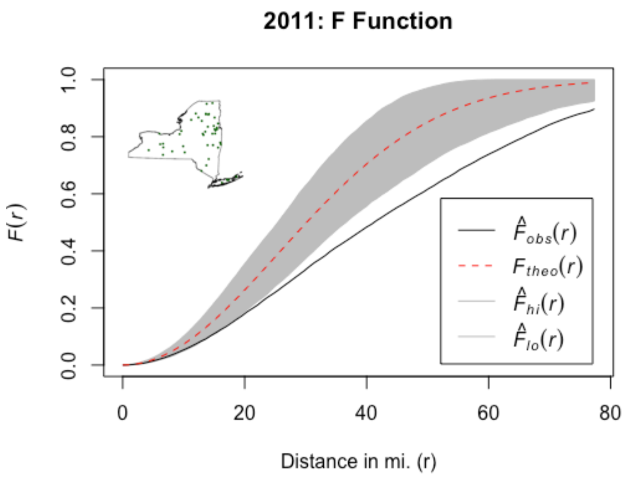

Figure 4(d) F-Function to detect clusters for the year of 2011

3. Modeling

We know clustering is evident within the 11 years of spatial wildfire data, specifically in the Adirondack and Catskill Regions as well as the Long Island Central Pine Barrens area. We now propose the question, “Do wildfires cluster in areas where there is lower than average rainfall?” To determine this we will first explore whether a linear correlation might exist between average yearly rainfall and frequency of wildfires.

3.1 Rainfall Vs. Wildfires

To decide whether or not a relationship might exist between our covariate and frequency of wildfires, frequency of wildfires was aggregated per quadrant for each year. The same was done for average yearly rainfall per quadrat per year. The quadrat size was the same as the quadrat size used previously for measuring spatial intensity.

A regression between frequency of wildfires per quadrant per year and average yearly rainfall per quadrat per year was performed in order to determine if any sort of correlation existed. This produced an adjusted R^2 value of 0.18, which is quite low but a correlation coefficient of -0.434, indicating a moderately negative correlation between amount of rainfall per year and the frequency of wildfires in that specific year. To visualize this, a scatterplot with the fitted regression line can be seen in **Figure 5**.

The fit isn’t by any means a perfect fit and while it may be unlikely, there was enough of a correlation to prompt further exploration into whether or not average yearly rainfall might be a contributor into explaining frequency of wildfires.

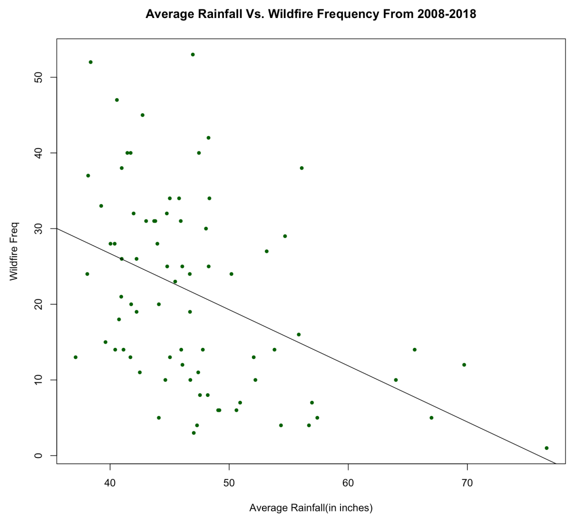

Figure 5 Regression Model of Average Rainfall Vs. Wildfire Frequency From 2008-2018

In order to answer the previously proposed question, “Do wildfires cluster in areas where there is lower than average rainfall?” we must first define what lower than average rainfall is. **Figure 6** represents a heat map of the 30-year-normal precipitation across the United States. Using ArcGIS, New York specific data was extracted to find the average 30-year-rainfall per location in New York State. Essentially, the goal was to ascertain a longstanding average for yearly rainfall.

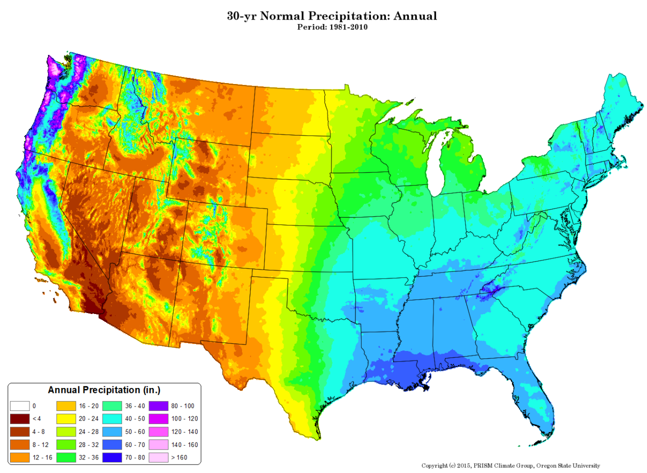

Figure 6 Prism Climate Group Repository heat map using the 30 year normal precipitation by year

Figure 7 represents summary statistics from the 30-year average rainfall data for New York State. The data was skewed right however the median and mean showed little difference, with values of 43.181 and 43.303, respectively. From this, average yearly rainfall was determined as approximately 43 inches.

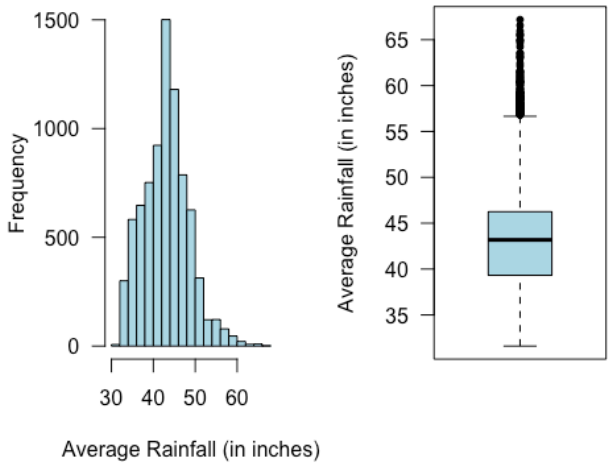

Figure 7 Summary statistics from 30-year rainfall data for NYS

We consider 43 in. as a fence that can now be applied to yearly precipitation data for each year of the corresponding wildfire years (2008-2018). Raster files were created for each year with spatial data of average rainfall less than 43 in. Using R, heat maps of the lower than average rainfall were super-imposed on each year of point patterns.

Figures 8(a)-(c) display the dispersion of wildfires atop of lower than average areas of rainfall for the years 2018, 2016 and 2011. The green to pink areas represent parts of the state that received, on average, less than 43 in. of rainfall in the corresponding year. The grey parts of the map indicate areas with greater than 43 in. of rainfall.

When considering the hypothesis that more wildfires exist in areas with less than average yearly rainfall, or essentially "drier" parts of the state, we would expect concentrations of events to occur in the green, yellow and mostly pink areas. However, this does not seem to be the case. Both 2018, **Figure 8(a)**, and 2011, **Figure 8(c)**, showed clusters of wildfires in areas with above average rainfall. In 2011, almost no wildfires occurred where the average yearly rainfall was the lowest. It can be seen in **Figure 8(b)** that 2016 was a fairly dry year with most of the state exhibiting a yearly average rainfall of 43 inches or less and also happened to be the year with the most wildfires.

In general, no strong pattern presented itself in the remaining years. When addressing the initial question: "Do wildfires cluster in areas where there is lower than average rainfall?" the answer seems inconclusive.

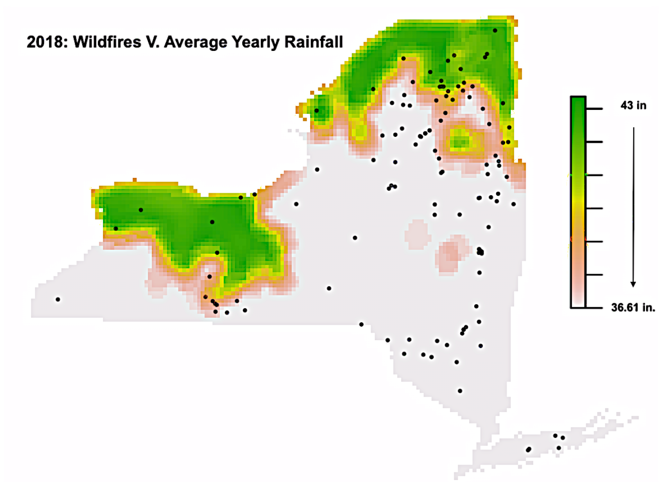

Figure 8(a) 2018 wildfires super-imposed with lower than average rainfall for that year

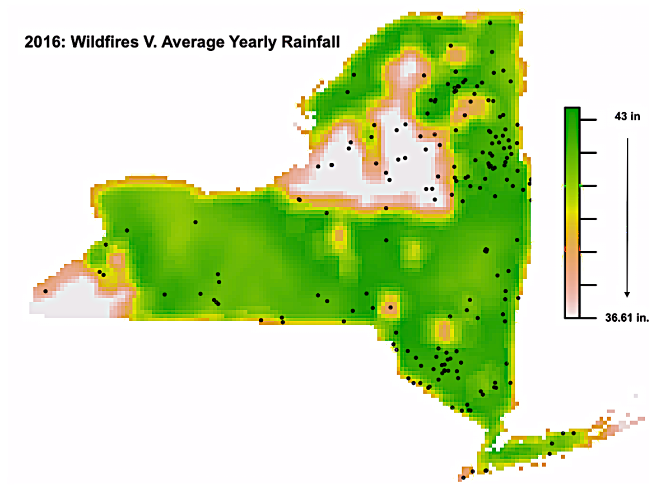

Figure 8(b) 2016 wildfires super-imposed with lower than average rainfall for that year

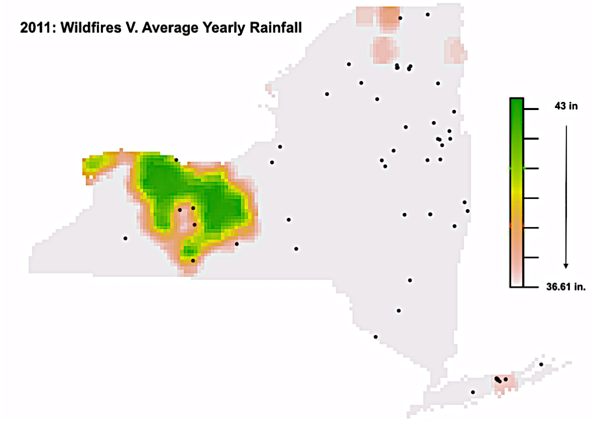

Figure 8(c) 2011 wildfires superimposed with lower than average rainfall for that year

3.2 Kernel Density Adjusted for Covariate

Exploring the covariate even further, we can create another kernel density map, this time adjusted for our covariate. **Figure 9** depicts the predicted point density when controlled for average yearly rainfall. If the covariate, areas of low rainfall, did a good job in explaining the wildfire density, the predicted density map should be nearly uniform across the map, depicting a solid color throughout. However, this is not the case, since we see darker pink areas in the northern and north western part of the state. Again, this suggests that low rainfall does not explain the clusters that we've seen.

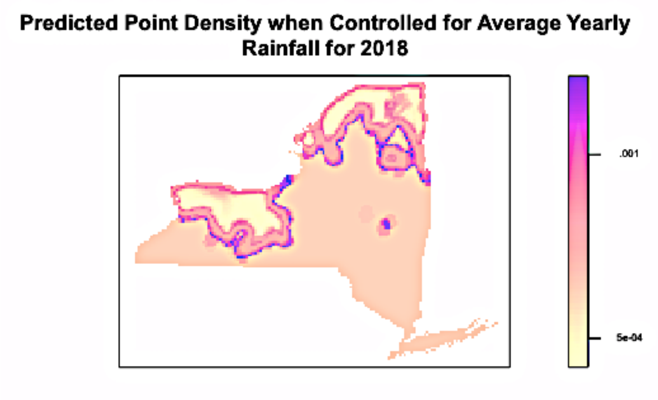

Figure 9 Kernel Density map adjusted for covariate for 2018

4. Results and Conclusion

Overall, a point pattern analysis revealed that the wildfires that occurred in NYS throughout the years of 2008 to 2018 did not exhibit complete spatial randomness. The G and F functions indicated clear spatial clusters with inconstant intensities.

Throughout the 11 years, fluctuation in amount of wildfires was clear, suggesting possible dependence among years. For all years, it seemed as though the clusters resided in the Adirondack and upper Catskill regions as well as in the area of the Long Island State Pine Barrens. Some likely reasons for this could be seasonal activity such as hiking and camping. One of the major ignition sources of wildfires in New York State is due to campfires. [5]

Considering the original question, "Do wildfires cluster in areas where there is lower than average yearly rainfall?" this seems unlikely. It is not to say that rainfall is not a contributor to wildfires but rather, frequency of rainfall or average amount per month may be more explanatory than average yearly rainfall. [6] Future research should be done to explore other metrics associated with rainfall including ground cover and soil types to determine rainwater absorption rates in these clustered areas. [7]

References

[1] PRISM Climate Group, Oregon State U [WWW Document], 2010. . Oregonstate.edu. URL http://www.prism.oregonstate.edu (accessed 12.6.19).

[2] Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V., 2008. Applied Spatial Data Analysis with R, Chap. 7: Point Patterns. ed. Springer, New York.

[3] Baxton, J., n.d. Vegetation Sampling Using the Quadrat Method [WWW Document]. Sacramento State University. URL https://www.csus.edu/indiv/b/baxterj/bio 221b/vegetation sampling quadrat.pdf.

[4] Diggle PJ. On parameter estimation and goodness-of-fit testing for spatial point patterns, Biometrics , 1979, vol. 35

[5] Wildfire Causes and Evaluations (U.S. National Park Service) [WWW Document], 2017. . Nps.gov. URL https://www.nps.gov/articles/wildfire-causes-and-evaluation.html.

[6] Mclemore, S., 2017. Spatio-Temporal Analysis of Wildfire Incidence in the State of Florida.

[7] Stratton, R., 2006. Guidance on Spatial Wildland Fire Analysis: Models, Tools, and Techniques.