Christopher Craig
STAT 5010
May 3, 2020

How Have Natural Disasters Changed Over Recent Years?

Throughout the past decade or so, climate change has gone from a highly debated phenomenon to one that has been more or less accepted to be true by society. Along with this comes a lot of speculation and studies about the impacts that climate change is having on our planet. Among this speculation is how climate change has impacted natural disasters. It seems like after every large natural disaster impacts the country we hear about how climate change has increased the frequency of natural disasters in recent years or how the magnitudes of these natural disasters are increasing. A quick google search about the impact of climate change on natural disasters reveals numerous articles and journals that have a wide range of claims about this topic being made. On the economic side of this question, there seems to be a consensus on the fact that the frequency of costly disasters has increased. An article from a journal, *Weather and Climate Extremes,* via ScienceDirect states that "the United States has seen more billion-dollar natural disaster events recently than ever before, with climate models projecting an increase in intensity and frequency of these events in the future…" (1). As the population, infrastructure, and inflation increase, it follows reason that the potential of financial losses from natural disasters has increased over time. However, this does not suggest that the raw frequency of natural disasters have increased or that these disasters have gotten worse over time. Some articles do attempt to tackle this side of problem but use some less than satisfying methods to come to their conclusions. Figure 1 shows an example of a graphic from one of these articles (2). The Intergovernmental Panel on Climate Change (IPCC) stated in their First Assessment Report from 1990 that "there (is) no evidence of an increasing incidence of extreme events over the previous few decades"(1). This report is outdated, however and in a section from a textbook published by the IPCC called *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation,* the author states that "a changing climate leads to changes in the frequency, intensity, spatial extent, duration, and timing of weather and climate extremes, and can result in unprecedented extremes" (7). Furthermore, an article from CNN quotes a chief climatologist of a reinsurance company, stating "climate change is contributing to the increase in extreme weather events" (5).

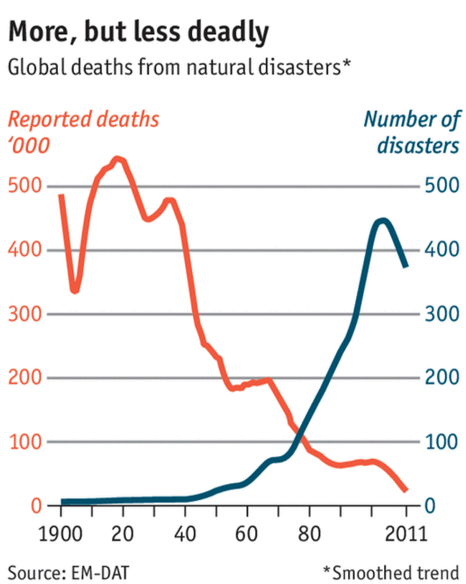


Figure 1.

Several people and organizations make claims about this topic, but none of them provide significant statistical evidence to back their claims up. So, has the frequency of natural disasters been increasing in recent years? Has the magnitude of natural disasters increased in recent years? The goal of this project is to use data from reliable sources to try and answer these questions using a combination of time series and regression techniques.

The data being used for this analysis, all obtained through Kaggle, are for US wildfires, Pacific and Atlantic hurricanes/cyclones, and global earthquakes. The US wildfire data ranges from 1991 to 2015 and is a random sample of 50,000 wildfires that originally came from the database of the Fire Program Analysis system, which was a federal program that was retired in 2015. The hurricane/cyclone data ranges from 1950 to 2015 and is from the National Hurricane Center of the NOAA, a US government agency. The earthquake data ranges from 1965 to 2017 and is from the National Earthquake Information Center. It includes all measured earthquakes that were of magnitude 5.5 or higher during that time period. Each of these datasets include a variable indicating the size of each individual event. Wildfires are categorized by classes with A being the smallest (less than a quarter acre burned) and G being the largest (greater than 5,000 acres burned). The hurricane did not include a 'category' variable for magnitude, so maximum wind speed was used as a proxy. The earthquake dataset also includes a magnitude indicating the magnitude of each earthquake. Figure 2 shows a histogram for each dataset displaying the distribution of events for each type of natural disaster over time.

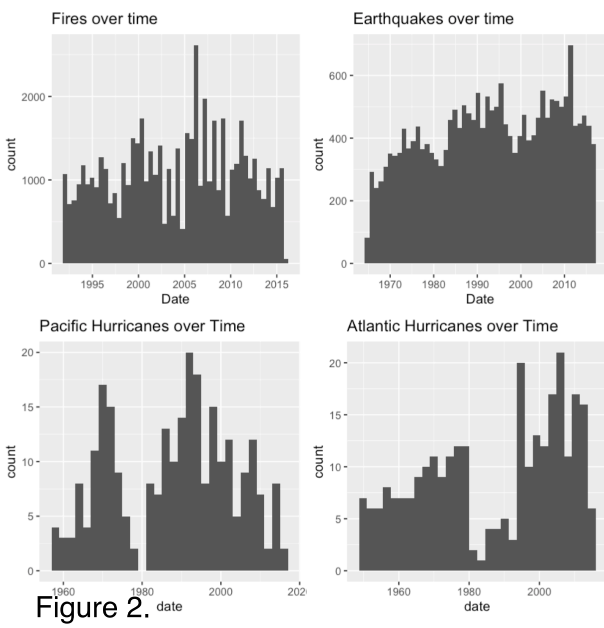


Figure 2.

There are several transformations that needed to be done to this data in order to make them suitable for time series analyses. After transforming the date variables to a 'Date' data type in R, the Pacific and Atlantic hurricane datasets were joined into one. Any unnamed hurricanes were removed from the data as it is impossible to distinguish them from each other. From each of the original source datasets, there was a new dataset created that has the raw number of respective natural disasters that occurred in each month. These datasets will be used to determine if the frequency of natural disasters has increased over time. For the US wildfire dataset, there was another set of data created with the raw counts of fires greater than 300 acres in each month (classes E through G). A second dataset was created from the original hurricane dataset that includes just the storm with the highest wind speed within each month. Finally, a second dataset was created from the original earthquake data that includes just the earthquakes with the largest magnitude within each month. These latter three datasets will be used to determine if there is statistical evidence that the magnitude of natural disasters has increased in recent years. The reason that all of the groupings are by month was to make sure the data samples were evenly spaced which makes them useable within a time series context. It is also worth noting that there is a significant limitation with the earthquake data. As fault lines are very complex, a large seismic event can be detected by many instruments at many depths and sometimes show up as several separate events within the dataset (6). For example. There is a spike in the number of earthquakes in 2011 that is largely due to the massive earthquake that hit Japan that year. Part of this problem was mitigated by grouping the earthquake data by date and depth simultaneously, so events that occurred at the same day and depth are counted as one.

A very basic and intuitive approach that could be used to see if the frequency and/or magnitude of natural disasters have increased in recent years would be to fit a simple regression model with number of events as the response and time as the predictor. Then, if the coefficient for the time parameter is positive and statistically significant, we could conclude that the specific frequency or magnitude of the natural disaster type has increased in recent years. Unfortunately this would violate multiple assumptions from the simple regression model. The most obvious of these violations is the violation of the assumption that error terms are uncorrelated or independent. Many natural phenomenon such as wildfires and hurricanes occur in very complex patterns which can depend on the time of year or at other irregular time intervals. There is also a strong possibility of autocorrelation or negative autocorrelation within the data which would also violate this assumption. Another violated assumption is that of normality of the response. For the four datasets that involve the frequency of natural disaster events, the response variable is a count per month which would come from the Poisson distribution. Due to these violations and the data's dependance on time, regression paired with time series analysis can used to answer these questions and diminish some of the impacts of these assumption violations.

A time series is simply a series of datapoint that are indexed in an order of time. There are three types of patterns that can be found in time series data; which are trend, cycles, and seasonality. Trend correspondents to long term changes in the data, seasonality refers to the pattern that a time series shows in relation to time, and cycles refer to other fluctuations in the series that are not in a consistent pattern. These response variable in a time series is made up of these patterns added together along with a remainder component. The first step in analyzing our data is to deconstruct the data into these different components. This can be helpful in visualizing each time series and is essential for regression with the datasets that do not have a poisson response. There are several methods to deconstruct a time series but for our purposes, the classic method of deconstruction was used which keeps things simple. Figure 3 shows the time series deconstruction of the maximum hurricane wind speeds per month and figure 4 shows the deconstruction of maximum earthquake magnitudes per month. The three graphs on the bottom

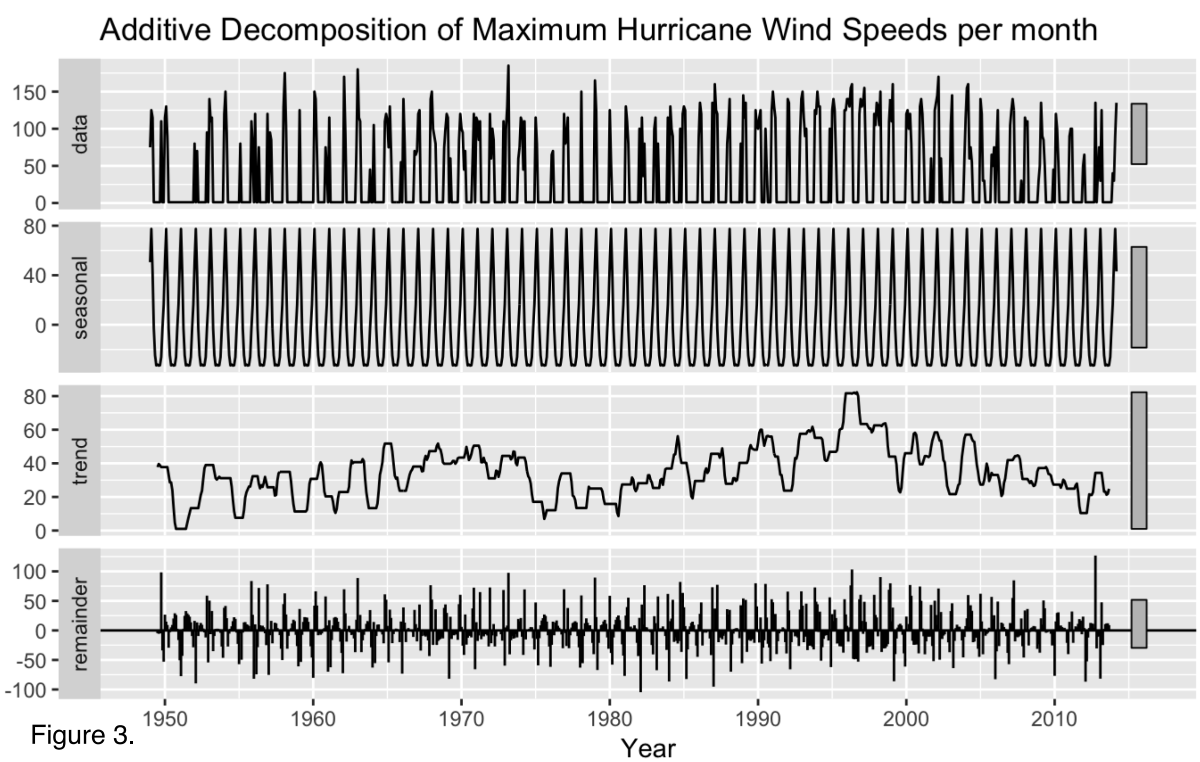


Figure 3.

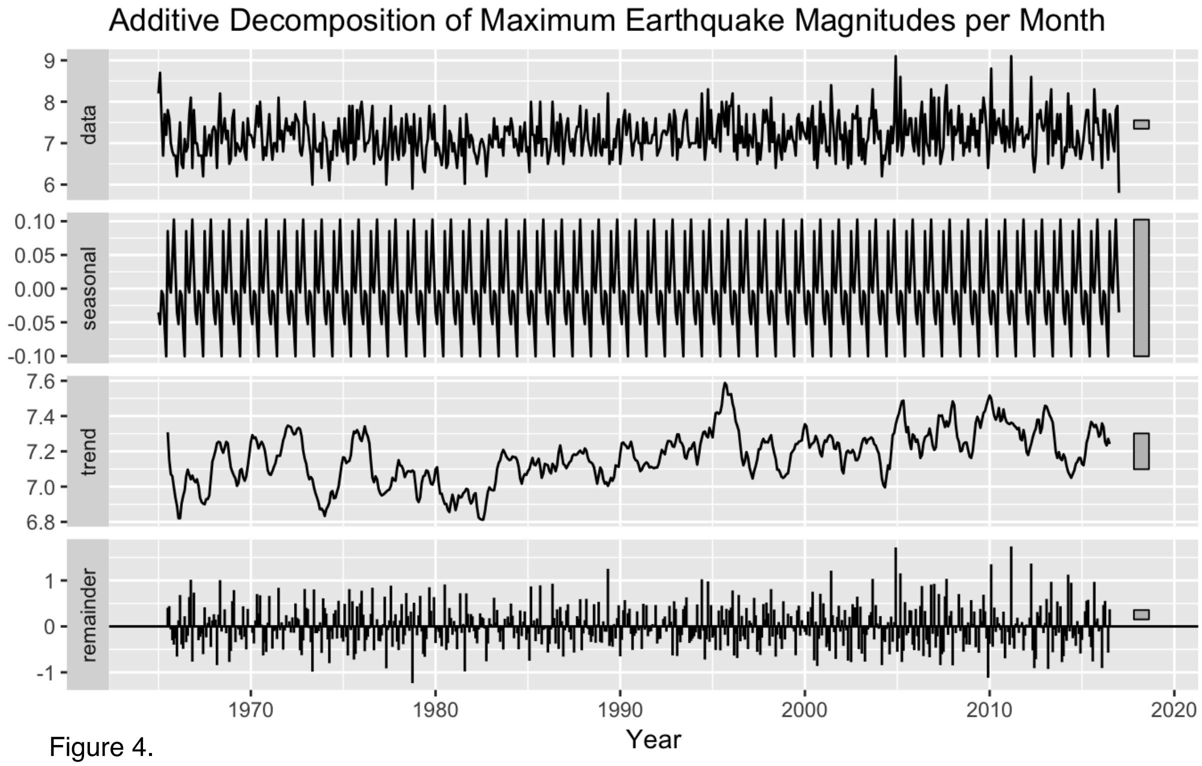


Figure 4.

are the components that, added together, make up the data at the top. The trend and cycle of the time series are combined in this method. The grey bars on the right all show the relative scale of each component. As the scales for each plot are slightly different, the grey bars are different sizes. These grey bars can show the relative impact that each component has on the data. For example, it is clear from the size of the grey bars on the plots that seasonality plays a huge role in the hurricane data but a very small and possibly inconsequential role in the earthquake data. After decomposing the data, a linear regression model can be constructed by using the tslm() function from the forecast package in R. This is similar to the regular lm() function, however both trend and season can be used as predictors which can be useful for interpretation. Trend is treated as a continuous parameter while season is treated as a categorical parameter. For example, if the response is from April, then the parameter associated with April takes the value of 1, and the rest of the season parameters take the value of 0. Figure 5 shows the summary of the tslm() function for the maximum hurricane wind speed data which uses both seasonality and trend as predictors. In this summary all of the predictors except for the October season are significant at the .001 significance level. This indicates that the trend is significant and that there is a very strong seasonality within the hurricane data. The strong seasonality was expected due to the fact that there is an actual hurricane season that goes from May to August and very few hurricanes occur outside of those months. The trend parameter is approximately .02, meaning that each month there was a .02 mph increase in maximum observed hurricane wind speed on average. This corresponds to about a 1 mph increase of maximum hurricane wind speed on average every 4 years. This does not seem like very much, but it does indicate that the max wind speeds of the worst storms in each month have been trending upward over the last half century. The results of the summary for the maximum earthquake magnitude data were far less interesting. None of the seasonal predictors were significant in the model, so seasonality was

```
Call:
tslm(formula = windts ~ trend + season, data = windts)

Residuals:
     Min       1Q   Median       3Q      Max 
-118.541  -13.256   -2.429    7.949  146.491 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  76.325040   5.093562  14.985  < 2e-16 ***
trend         0.020520   0.005847   3.510 0.000475 ***
season2      27.661298   6.437105   4.297 1.95e-05 ***
season3      -5.101647   6.437113  -0.793 0.428292    
season4     -52.948462   6.461836  -8.194 1.05e-15 ***
season5     -74.492060   6.461823 -11.528  < 2e-16 ***
season6     -82.497195   6.461815 -12.767  < 2e-16 ***
season7     -81.825408   6.461813 -12.663  < 2e-16 ***
season8     -82.692082   6.461815 -12.797  < 2e-16 ***
season9     -82.481833   6.461823 -12.764  < 2e-16 ***
season10    -75.810046   6.461836 -11.732  < 2e-16 ***
season11    -52.984412   6.461855  -8.200 1.01e-15 ***
season12    -33.589548   6.461879  -5.198 2.58e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 36.98 on 770 degrees of freedom
Multiple R-squared:  0.5079,	Adjusted R-squared:  0.5002 
F-statistic: 66.22 on 12 and 770 DF,  p-value: < 2.2e-16
```

Figure 5.

removed from the equation. The trend parameter is significant at the .001 level and corresponds to a .0003 monthly increase of maximum observed earthquake magnitude on average. Adjusted R-squared for this summary was tiny, with a value of around .02. This indicates that trend explains an extremely small amount of the variation within the response. This combined with the fact that an earthquake magnitude increase of .0003 is almost negligible makes it difficult to draw any conclusions from this summary.

Each of the other four datasets contain a count per month, or poisson, variable as the response. A method of perform a poisson regression on tome series data is available from the tscount package in R. This package contains a tsglm() function which allows the user to perform a generalized linear regression model on a time series with a poisson or negative binomial response. This function shares some similarities between the glm() functions, but has workarounds for using time series data such as using a “quasi maximum likelihood estimation of the unknown model parameters”(8). The implementation of this function is quite complex. More information about the function can be found through the description of the ‘tscount’ package via CRAN (8). The frequency of wildfires per month, frequency of wildfires greater than 300 acres per month, frequency of earthquakes per month and frequency of hurricanes per month were all tested with tsglm() regression. Through testing each model and checking the diagnostics it was found that the wildfire and earthquake data were modeled well with the negative binomial response, which is a distribution that counts the number of failures in a set. Figure 6 is a probability integral transform (PIT) histogram for the generalized linear time series model fit for the wildfire frequency data. If the data fits the distribution being used then this histogram should be approximately uniform. Figure 7 shows the PIT histogram for the large wildfire data and is what a poor fit looks like. The hurricane and large wildfire frequency data were not modeled well

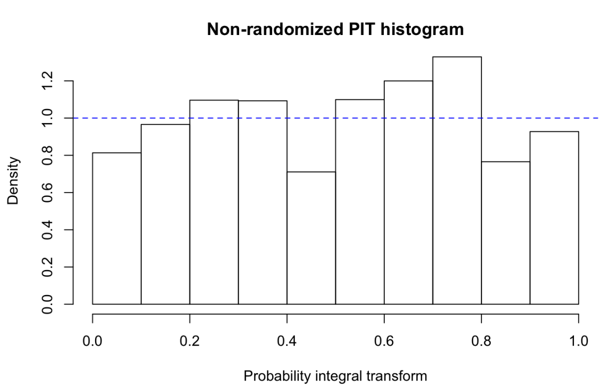


Figure 6.

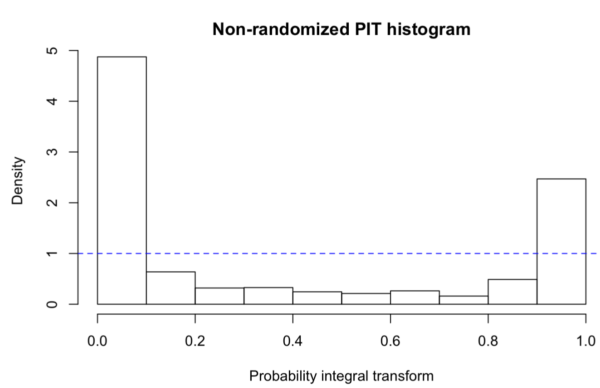


Figure 7.

by either the negative binomial or the poisson response options, so they were modeled by the tslm() function from earlier with a non-normal response being a source of error for the regression. Figure 8 shows the results of the summary of the two fitted tsglm() models. Frequency of wildfires is on the top while frequency of earthquakes is on the bottom. The beta_1 parameter corresponds to regression on the previous measured value, which can show autocorrellation in the data. The beta_12 parameter is the 12th order autocorrellation and can be

	Estimate <S3: AsIs>	Std.Error <S3: AsIs>	CI(lower) <S3: AsIs>	CI(upper) <S3: AsIs>
(Intercept)	23.6197	11.3647	1.345	45.894
beta_1	0.4007	0.0615	0.280	0.521
beta_12	0.3651	0.0705	0.227	0.503
alpha_12	0.0995	0.1076	-0.111	0.310
sigmasq	0.2139	NA	NA	NA

	Estimate <S3: AsIs>	Std.Error <S3: AsIs>	CI(lower) <S3: AsIs>	CI(upper) <S3: AsIs>
(Intercept)	4.4956	1.6051	1.3496	7.642
beta_1	0.2583	0.0409	0.1780	0.338
beta_12	0.1171	0.0395	0.0398	0.194
alpha_12	0.4875	0.0785	0.3336	0.641
sigmasq	0.0656	NA	NA	NA

5 rows

Figure 8

used to show if there is seasonality within the data. The alpha_12 data corresponds to regression on values of the conditional mean 12 seasons (or one year) previous. For the regression on wildfire frequency (top), we can see that there is a correlation between the number of wildfires in a month and the number in the previous month. The values in beta_12 show that there is some seasonality in the data as the standard error is quite small relative to the estimate. The alpha_12 value is not significant, though, as the standard error is relatively large and the confidence interval includes both positive and negative values. These means that there is not statistical evidence to conclude that the frequency of wildfires has been increasing in recent years. For the earthquake data we can see that the beta_1 and beta_12 values are significant but quite small indicating that there is some slight autocorellation between subsequent values and some slight seasonality. The alpha_12 value also appears to be statistically significant, meaning that if we take the data at face value there is some evidence that the frequency of earthquakes have been increasing in recent years.

The final two datasets to be fit with a model are the large wildfire and hurricane frequency data. Figure 9 shows the distribution of residuals of these two models after they were fit with the tslm() model. The residuals for the hurricane frequency data are on the left and the residuals for the frequency of fires greater than 300 acres is on the right. Both histograms show

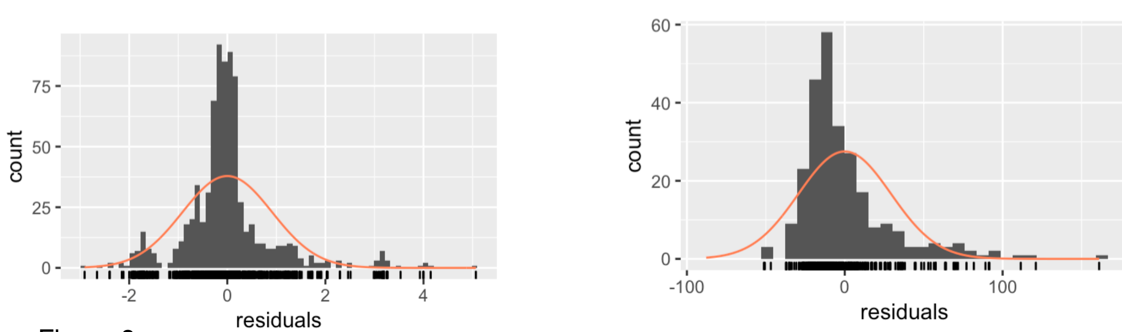
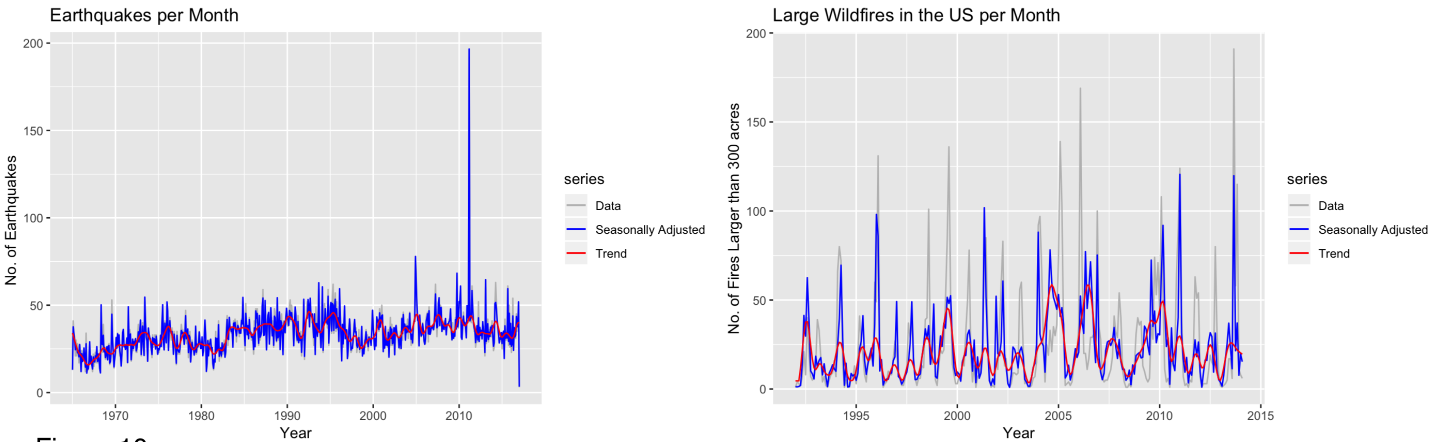


Figure 9.

that the residuals are semi normally distributed, with many values grouped near 0, and a slight skewness for the large wildfire data. Noting that these are not quite normal, we move on the the summary of the models. Both models were fit with season and trend parameters. The model fit for the large wildfire frequency data had a few statistically significant seasons at the .01 significance level and a few at the .05 significance level indicating that there is some slight seasonality within the data. The trend parameter is significant at the .01 significance level with a value of .05. This means that there is an increase of .05 large wildfires per month on average. This would be about an increase of 1 fire per month on average every 2 years which is evidence that there has been an increase in the frequency of large fires in recent years. The hurricane frequency data has a very similar model as the maximum hurricane wind speed data. All season parameters outside of October and the trend parameter were significant at the .001 significance level. The trend parameter has a value of .0007, which is the monthly increase in the frequency of hurricanes on average. While significant, this is a very small relative to the time period of observed values and does not point to an increase in the frequency in hurricanes in recent history.

Using linear models and GLM's in the context of time series, we see some evidence both for and against the idea that the frequency and magnitude of natural disasters have increased over recent years. There was no statistically significant evidence that the frequency of wildfires or hurricanes have increased over time and there was no significant evidence that the magnitude of earthquakes have increased over time. However, there is some statistically significant evidence in our analysis that the frequency of earthquakes has increased in recent years, the frequency of large wildfires in the US has increased in recent years, and that the max wind speeds in the most severe hurricanes each month have increased in recent years. There are, unfortunately, several shortcomings of this analysis. First is the trustworthiness of the earthquake data. While the data comes from a trustworthy source, it is difficult to distinguish which samples in the dataset come from the same earthquake events and it would likely include some geologic theory to determine what should be considered an individual earthquake. While grouping the data my date and earthquake depth helped with this issue I am confident that there still remains some replications of the same earthquake events within the data (particularly given the glaring outlier from 2011 as seen on figure 10). All conclusions made about the frequency of earthquakes in this paper should be read with that in mind. Another shortcoming of this analysis is that the large fire frequency data and the hurricane frequency data do not have a normal response but were modeled as if they do. If this analysis were to continue there is much more that could be done with the tscount package buts was not delved into for the sake of time. It is possible that there are some workarounds with model parameters for the tsglm() functions that would fit better models to the data. Also, data from droughts, floods, and other large precipitation events could be added to this analysis to encompass a larger variety of natural disasters. In summary, regression on this time series data reveals that there is something to the sentiment that the frequency and magnitude of natural disasters have increased over time. But, any observed increases are very small and would not likely be perceivable outside of an analysis such as the one in this paper. It is far more likely that this common sentiment comes from the fact that the destruction that these events cause has increased over time, due to increases in population and infrastructure. The increase in technology and the most recent evolutions of media allow us to hear about all major natural disasters that happen each year around the world, heightening our fear of natural disaster events and increasing our feelings of mortality.



Earthquakes per Month
No. of Earthquakes
Year
series
Data
Seasonally Adjusted
Trend
Large Wildfires in the US per Month
No. of Fires Larger than 300 acres
Year
series
Data
Seasonally Adjusted
Trend

Figure 10.

Sources:

(1) - https://www.sciencedirect.com/science/article/pii/S2212094715300165
-quoted scientific journal

(2) - https://www.economist.com/graphic-detail/2017/08/29/weather-related-disasters-are-increasing
-image source

(3) - https://www.nhc.noaa.gov/climo/
-hurricane information

(4) - https://wildfiretoday.com/2019/02/02/average-size-of-wildfires-continued-to-increase-in-2018/
-article about wildfires

(5) - https://www.cnn.com/2020/01/08/business/munich-re-climate-change-natural-disasters/index.html
-quoted CNN article

(6) - https://www.usgs.gov/natural-hazards/earthquake-hazards/earthquake-processes-and-effects?qt-science_support_page_related_con=4#qt-science_support_page_related_con
-information on earthquakes

(7) - https://otexts.com/fpp2/intro.html
-online textbook on time series

(8) - https://cran.r-project.org/web/packages/tscount/vignettes/tsglm.pdf
-information about the recount package

Data via Kaggle:

Wildfire - https://www.kaggle.com/rtatman/188-million-us-wildfires

Earthquake - https://www.kaggle.com/usgs/earthquake-database

Hurricane - https://www.kaggle.com/noaa/hurricane-database