

My final project is focused on the association between precipitation intensity and traffic speeds. If you would like to look at the code involved in this project, you can find it all at in my Github repository below.

This is the general outline we'll follow for today, for the sake of time we'll jump right in and talk about the problem at hand.

According to the Federal Highway Administration, roughly 5.8 million crashes occur every year, 21% of which are related to weather, and of that 21%, 46% are related to rainfall. That translates to roughly 560 thousand crashes each year that are associated with rainfall alone. That's a significant number of crashes and brings me to the question that I'm posing for this project.

What impact does precipitation intensity have on traffic speeds? I've broken this question up a little with the following questions, how different if at all are traffic speeds under precipitation conditions in comparison to dry conditions, and to what extent. Can we calculate a percent reduction? Is there a spatial or temporal aspect to this relationship? And finally, are there external factors that can be identified?

With the questions clearly defined, let's talk about the data. All data for this analysis is from June 2018. The traffic speeds data is from the INRIX speeds database maintained by JTRP. The data is at 1 minute resolution and has been down sampled to 2 minutes to match the precipitation data. INRIX is a company that specializes in crowd sourced traffic data and has access to roughly 300 million data sources, both commercial and private. Conveniently, there are no missing values in the data set as they take care of any necessary imputation on the back end. As a consequence, they label the data with confidence scores. High confidence indicates the data is a real time average. Medium confidence data indicates that the data is a combination of surrounding segment's real time averages. Low confidence indicates that the speeds have been estimated from a historical average.

For the purposes of this study, I've restricted my analysis to the north bound corridor of I-65 from Zionsville to Lafayette. This corridor consists of 69 segments with 98.75% of the data listed as high confidence and 1.25% listed as medium confidence. None of the data is low confidence for this domain. If you're struggling to visualize how INRIX data collections work, the image in the bottom left is a good summary of how it would work. Cars are moving along the road and as they pass geo-tagged locations, speeds are recorded and averaged every one minute.

As for the precipitation intensity data, this is derived from the National Severe Storms Laboratory's Multi Radar Multi Sensor system. This is a highly processed data set but has some great qualities for analysis. It provides a 1x1km grid and 2 minute temporal resolution. Many data sources have been combined by this system including radar, various observation types, and forecasts. One of the major problems with this data set is that due to its size, there are no archives aside from Iowa. Iowa only archives a handful of the variables, of which only

precipitation intensity was useful for my analysis. The data for a single month, cut down to just Indiana is roughly 15gb in netcdf format.

That brings me to the next problem, how do I get this massive precipitation data set down to the road segment level. This is where interpolation comes in so that I can subset my traffic data to the road segment level and dispose of the rest. I had originally tried to implement Kriging interpolation using the gstat package but it was entirely too slow, averaging about 12 minutes per time slice. With 21,600 time steps for 2 minute resolution in June, I needed something substantially faster to avoid having to write my code to run in parallel on a cluster. I discussed this problem with Dr. Shan and he had mentioned using nearest neighbor interpolation. I ended up shying away from nearest neighbor because I didn't think it was the best choice for my weather data. If you were to look at a stereotypical radar image, typically the areas of high precipitation are in red and low precipitation are in green, in between though is typically a transition from red to orange to yellow to green. This idea can be characterized with bilinear interpolation as it essentially does a weighted average in the x direction and then does the same in the y direction. Using the Akima package and interpp function, which runs on some old Fortran code, I can do bilinear interpolation in about 1.5 seconds for a single time step. The interpolation takes about six hours for the full data set.

Before we move into the analysis, it's important to define a few terms. A sense of normal was necessary to have something to judge the precipitation events by. Normal traffic flow was defined as speeds greater than or equal to 60 mph. Since I'm interested in events where traffic slows during precipitation events, I've defined a traffic event as any time when speeds are less than 60 but precipitation is greater than 0.

Also, since I suspected a temporal component to exist in the data, I've defined Day and Night using civil twilight hours for June 2018. Civil twilight is defined as the time when it is light enough out to be outside and work, even if the sun isn't above the tree line yet. For June 2018, daylight is defined as 6am to 8:30pm eastern time.

Using the criterion laid out previously, I decided to try to get a feel for the data with a scatterplot. What results is a somewhat deceptive plot. This plot originally led me to believe that there were significantly more events that occurred at night than in the day time. That actually is not true, as we'll see in a moment, but what it does confirm is that there seem to be more variation in the precipitation amounts for various speeds at night.

When counting the number of occurrences for each criterion, what we see is that the events we're interested in only account for 0.17% of the data with daytime events leading with 0.13%. This is a pretty typical problem in any analysis that deals with rare events. I've heard hurricane people discuss this problem quite regularly. The thing you want to study is rare so there are relatively few observations to study. More data can be collected but the proportions should stay roughly the same.

In addition to counts, means for each segment were calculated and are presented as box and whisker plots to show the distribution. Under normal conditions, daylight speeds are higher than those of at night. I would have expected this to be the other way around due to traffic volumes but construction at night may contribute to this. During an event though, the results are reversed, speeds during the day are lower than those at night, indicating that precipitation intensity may have a stronger impact to speeds during the day. This seems to likely be due to higher volumes in the day.

When viewed spatially, the night time events show strong clustering where as the day time events show some clustering but are more homogenous.

If we zoom in, many of these impact locations are near on and off ramps, pointing to one external factor that seems to be at play here.

There was another location that jumps out at night as having a much higher count of events. Zooming in revealed that this wasn't a ramp but was actually an underpass. Investigating this further with google street view revealed that this was a bridge under construction. The streetview timestamp was August 2018, meaning this bridge was definitely under construction in June. Most traffic construction occurs at time and seems to confirm construction as an external factor that impacts the speeds.

Finally, to quantify how much impact precipitation has on traffic speeds, a percent reduction from normal mean speeds was calculated. The code for this is shown here but for the sake of time I'm not going to go over it. A summarized formula is shown below the code.

When plotted, a band jumps out during the night from 10 to 15% reduction in speed with a clump of precipitation intensities above 15 mm/hr. During the day time a clump jumps out at the 13 to 22% reduction in speed marks with the precipitation above 8 mm/hr. In either case, it doesn't seem to be particularly sensitive to the precipitation intensity.

So to conclude, traffic speeds are most likely to be reduced by 10 to 15% from normal at night when precip is greater than 15 mm/hr. This becomes more severe during the day with 13 to 22% speed reductions with precip greater than 8 mm/hr.

There does appear to be a spatial and temporal component to these. Spatially, construction and ramps seem to impact traffic speeds more than precipitation alone. Temporally most traffic events occur during the day and result in significantly greater reductions in speed than normal.

Thank you for your time, any questions?