Water Main Breakages

Wen Ye, Gautam Agarwal, Bryan Jin

University of Wisconsin - Madison

1 Introduction

Water is a critical resource. Many of us rely on easy access to running water, made possible through a complex system of water mains. But much of the water main infrastructure in Madison, WI dates back to World War II and the postwar era, when pipes were made of a less durable material called spun cast iron [1]. Consequently, every year Madison Water Utility repairs over 200 water main breaks [1]. As we will show, this problem with water main breaks is exacerbated during Madison's winters. Since water main infrastructure serves such an important function, we will investigate the following question: *what are the factors that influence the number of water main breakages, and how do these factors affect the number of water main breakages?*

This report has two main sections. In the Breakage Factors section, we use data about water mains and water main breaks in Madison to examine factors that may have a relationship with the number of water main breakages. After we have some idea of what factors are associated with water main breakages, we build several predictive models for water main breakages in the Forecast section. We hope these models will provide insight into future water main breaks and risk levels of active water mains in the city of Madison.

2 Breakage Factors

The following three subsections - pipe construction, Madison geography, season and temperature - discuss the breakage factors that we investigated.

2.1 Pipe Construction

This section focuses on inherent characteristics of the water mains: the diameter/size of the pipe, the depth at which the pipe is laid, the material of the pipe, and the age of the pipe.

2.1.1 Pipe Depth

Pipe depth refers to the depth at which a pipe is laid, relative to the ground surface. Figure 1 describes the distribution of water main breaks with respect to pipe depth. From 5 feet to 7 feet, the curve rises sharply from 0.1 to 0.9. This shows that in about 80% of water main breaks, the broken pipe was laid 5-7 feet deep.

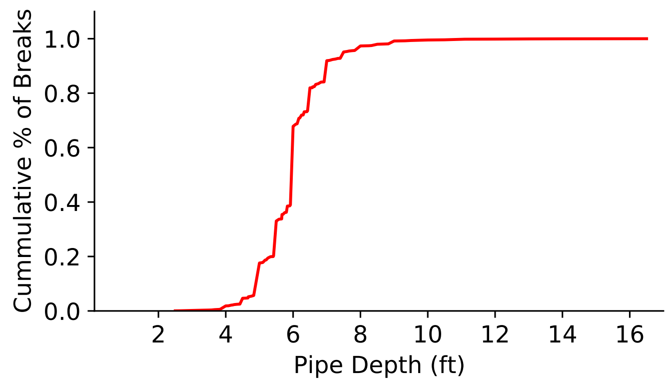

Figure 1: Cumulative Frequency of Breaks by Depth

We considered two possible explanations: most pipes are laid 5-7 feet deep, or pipes laid 5-7 feet deep are most vulnerable to breaking. We were unable to conclusively figure out to what extent each of these possibilities is true, as we did not find pipe depth data for pipes that have never broken. However, some brief research [2] suggests that most pipes at laid at 5-7 feet. Pipes laid at 5 feet are below the frost line [2][3], which is "the depth to which the groundwater in soil is expected to freeze" [3]. Laying pipes at greater depths is more expensive, so laying pipes deeper than 7 feet may be unnecessary, at least for typical Wisconsin winters [2].

While there might be a relationship between pipe depth and breakages, we did not have data that allowed us to draw any conclusions.

2.1.2 Pipe Size

Pipe size refers to the diameter of the pipe. In general, pipes that have a larger diameter need to be laid at greater depth from the ground surface. This is to make sure the distance between the ground surface and the top of the pipe is large enough to prevent the water in the pipe from freezing in the winter [4].

Figure 2 describes the distribution of water main breaks with respect to the diameter of the main's pipe. From 4 inches to 6 inches, the curve rises sharply from below

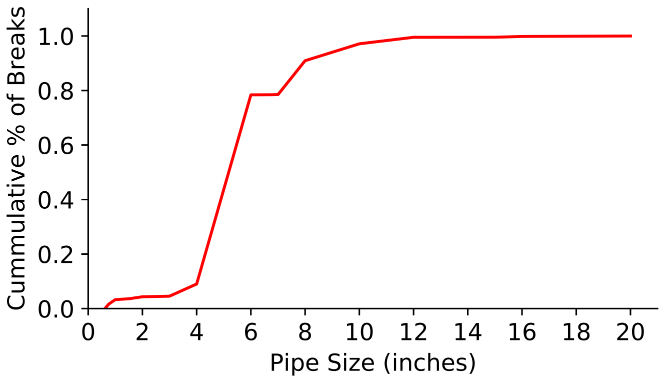

Figure 2: Cumulative Frequency of Breaks By Size

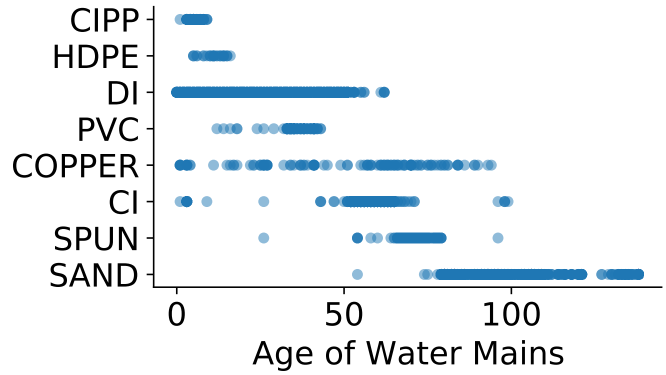

Figure 3: Material and Age of Water Mains

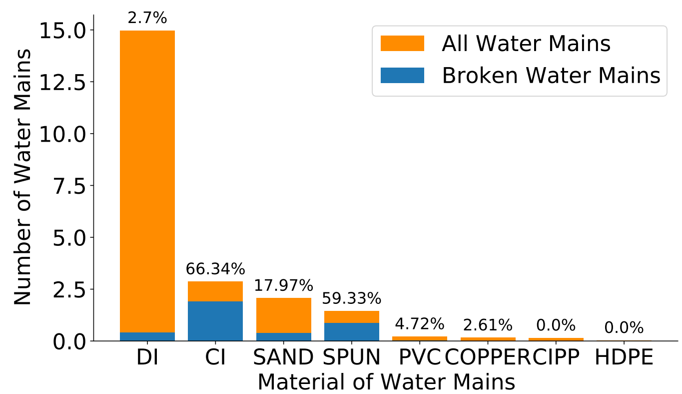

Figure 4: Material of Water Mains

0.1 to 0.8. This shows that in about 75% of water main breaks, the broken pipe has a diameter between 4-6 inches. However, only 35% of all pipes have a diameter between 4-6 inches; in particular, 35% of all pipes have a diameter of 8 inches, and 25% of pipes have a diameter between 10-16 inches. This suggests that smaller pipes are more prone to breakages.

With evidence that larger pipes break less, we made sure to incorporate pipe size in our models.

2.1.3 Material Used

In the city of Madison, 97.52% of pipes are made of some variety of iron, and these pipes account for 99.6% of breakages. Most iron pipes are laid underground, while copper pipes connect water mains with buildings (e.g. households or public facilities). As a result, copper pipes tend to be shorter than iron pipes, e.g. the average length of a copper pipe is half the average length of a ductile iron pipe.

Before going into more details, here is a list of full names and abbreviations for the different pipe materials: Cast Iron (CI), Ductile Iron (DI), Spun Cast Iron (SPUN), Sand Cast Iron (SAND), Copper (COPPER), Polyvinyl chloride (PVC), High Density Polyethylene (HDPE), and Cured In Place Pipe (CIPP).

From our preliminary research, we understood that modern pipes are made of DI, because DI is less brittle than materials such as SPUN, which were used by older pipes [1]. We wanted to see if we could find evidence for the superiority of DI.

Figure 3 is a scatter plot showing the material and age of water mains. We see that among all the iron pipes, SAND is the oldest type (spanning from 70 to 140 years old) followed by CI and SPUN pipes spanning from 50 to 80 years old. DI pipes are the youngest, most of which are under 60 years old. One thing to note is that DI pipes were invented in 1943. So, the DI pipes that are over 77 years old are likely to be a result of errors in data entry and should be ignored.

Figure 4 gives a detailed breakdown of the materials used for water pipes. The y-axis is the number of water mains in thousands. The percentage on top of each bar is the rate at which these pipes have broken. We see that even though there are many water mains made of DI, less than 3% of DI water mains have broken. In contrast, over 50% of CI and SPUN pipes have broken. One explanation for this large difference is that, as we expected, DI is a better material than CI or SPUN for preventing breakages. However, there is another explanation: Figure 3 shows that CI and SPUN pipes are generally much older than DI pipes, so they have had a longer time to break. In our models shown later, we will try to account for both possible explanations by using pipe material and pipe installation year as model inputs. We note that SAND pipes perform really well as they are old and have a low rate of breakage.

Figure 5 shows the length (in feet) of broken water mains by material. The x-axis is the number of weeks since the date of the first recorded breakage in the dataset - 1997/01/02. The x-axis ends at week 1205, which corresponds to the date of the last recorded breakage in the dataset - 2020/02/07. On the y-axis is the percentage of

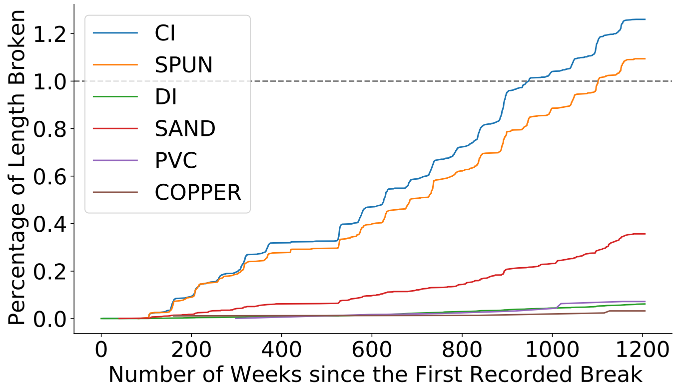

Figure 5: Broken Length per Week CDF

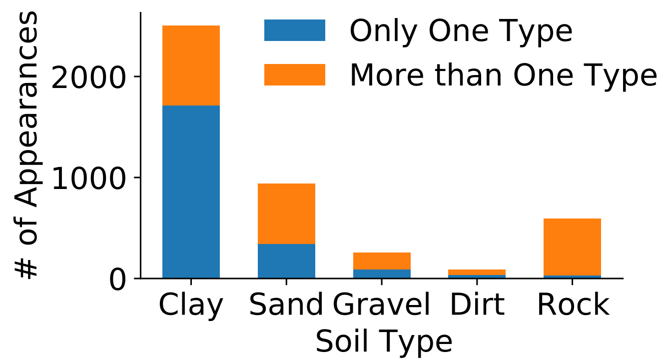

Figure 6: One Soil Type vs Multiple Soil Types

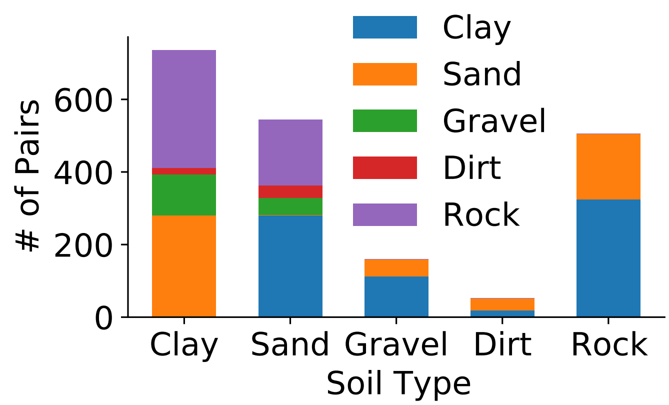

Figure 7: Common Pairs of Soil Types

total pipe length that is broken. Among the iron types, we see that CI and SPUN water mains have percentages greater than 100%. So, we can infer that many CI and SPUN pipes broke multiple times (especially since not all CI and SPUN pipes have broken, as shown by Figure 4). For comparison, DI pipes and SAND pipes perform much better.

2.2 Madison Geographics

This section gives a picture of where these pipes are located in Madison and explores the types of soil in which the pipes are laid.

2.2.1 Soil Type

Soil type describes the soil in which a water main is laid. Many water mains are laid in soil that can be described with a single "atomic" type: clay, sand, gravel, etc. Other water mains are laid in soil that can be described with two "atomic" types: clay and sand, sand and gravel, etc. A few water mains are even laid in soil that can be described with three "atomic" types: clay, sand, rock or clay, gravel, rock.

To get a better understanding of the various soil types, Figure 6 shows how frequently the most common atomic types appeared on their own vs. in combination with other atomic types. For example, the leftmost bar shows that when clay appeared as an atomic type, the soil type was just clay over 60% of the time. In contrast, the rightmost bar shows that rock almost always appeared with other atomic types (so the soil type was "sand and rock" or "clay and rock").

Figure 7 focuses on the instances of the soil type variable where two atomic types are recorded. The leftmost bar shows that when clay appears as one of two atomic types, the other atomic type is either sand or rock in 80% of the instances. The next bar shows that when sand appears as one of two atomic types, the other atomic type is either clay or rock in 80% of the instances. The three bars on the right indicate that when gravel, dirt, or rock appear in a pair, they are always paired with clay or sand.

Overall, clay and sand seem to be the main atomic types. Gravel, dirt, and rock seem to function as secondary types that can be associated with either clay or sand. In a later part of this report, we will investigate how the distribution of water main breaks across soil type changes based on the season.

2.2.2 Location

Figure 8 shows the geographic distribution of water mains for the past 30 years. The water mains are colored according to material. One trend we see is that the water main infrastructure is extending toward the outskirts of the city. Meanwhile, the proportion of water mains that are made of DI is steadily increasing. Recall from Figure 3 that the oldest pipes are made of SAND. These pipes are mainly in the center of the city. As we move outward to the boundaries of the city, we see newer DI pipes. On the other hand, we see that non-iron type pipes such as COPPER and PVC pipes are only used in particular areas rather than throughout the city.

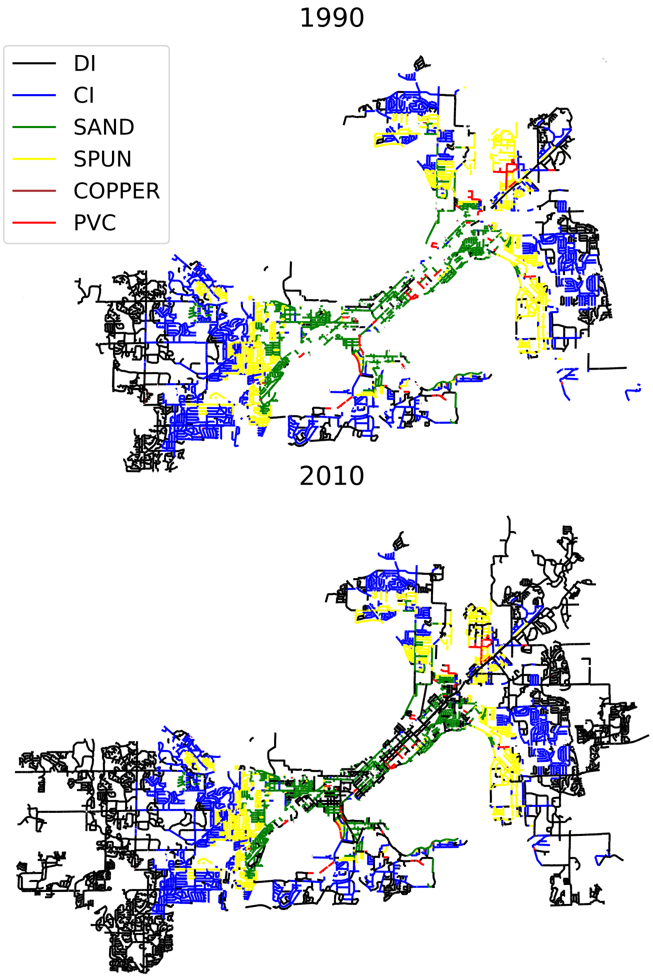

Figure 8: Water Mains by materials

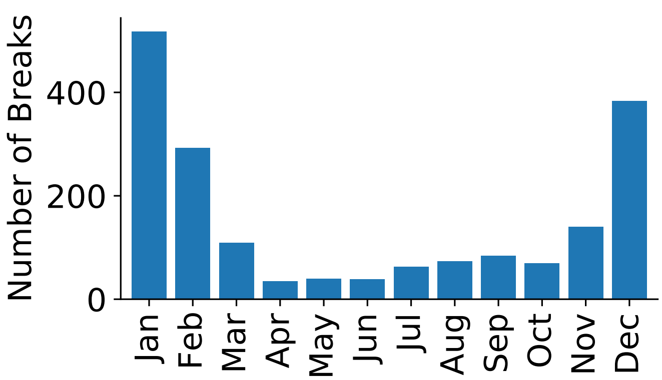

Figure 9: Number of Water Main Breaks by Month

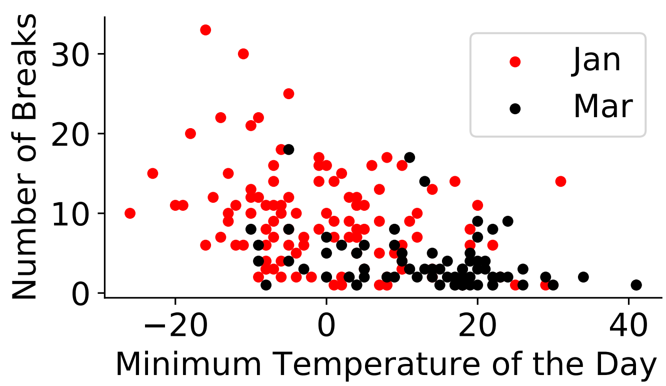

Figure 10: Jan Mar comparison

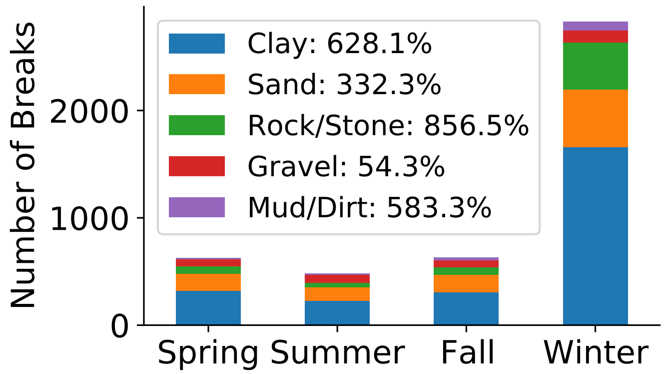

Figure 11: Percentage of Breaks by Season

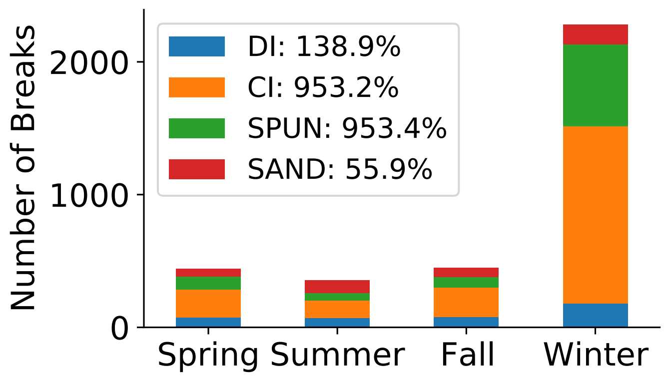

Figure 12: Iron Type Materials and Season

2.3 Season and Temperature

This section looks at how season and temperature affect the number of pipe breakages.

Figure 9 shows the total number of water main breaks by month, from 1980 to 2020. The total number of water main breaks is several times higher in winter months: December, January and February. Low temperatures and other winter conditions may contribute to this difference. After we saw Figure 9, we decided to examine more closely the relationship between temperature and number of breaks.

Figure 10 gives a more detailed comparison to tease out the influence of temperature and season. It shows the distribution of breaks on a day in January and March. Across the past 23 years, on days of the same minimum temperature, we see that the day in January almost always has more breaks than the day in March. This indicates that temperature is not the absolute determinant of the number of breaks on a given day. Season also matters as Jan is in Winter and March is in Spring.

Figure 11 shows the distribution of water main breaks across soil type for each season. Relative to summer, breaks for pipes laid in rock/stone increased by 856% during winter, whereas breaks for pipes laid in gravel only increased by 54%. During winter, rock/stone has the greatest increase in number of breaks and gravel has the lowest. We infer that pipes laid in gravel are more resistant to the effects of winter than pipes laid in other soil types: installing future water main pipes in gravel could reduce the number of breakages. One possible explanation for this phenomenon is that gravel is porous, and water drains quicker through gravel than other soil types [5]. Madison has slightly less rainfall than the national average but significantly more snowfall than the national average, suggesting a large amount of water may accumulate in Madison soil [6]. Since water flows quickly through gravel, pipes laid in gravel can be kept relatively dry compared to pipes laid in other soil types. Hence, pipes laid in gravel could be less affected by the freezing and thawing of water around the pipes.

Figure 12 shows the distribution of water main breaks across iron type for each season (here we are only focusing on iron pipes). Like with Figure 11, the percentages in the legend compare the number of breakages in the winter to the number of breakages in the summer. We see a drastic increase in the number of breakages for pipes made of CI or SPUN. On the other hand, SAND and DI pipes have a much smaller increase in breakages from summer to winter.

Going back to Figure 3, we see that starting from around 50 years ago, almost all iron pipes are made of DI. This means that DI pipes are younger than the other iron pipes. And the age interval between CI and DI pipes nearly do not overlap, suggesting that DI pipes came into use mostly to replace CI pipes. We see a similar pattern for SPUN and SAND pipes as well. However, since SAND pipes have a low rate of breaking as well as a small increase in number of breakages from summer to winter, we want to recommend the city replacing CI and SPUN pipes with DI pipes before replacing SAND pipes. Since SAND pipes are the oldest pipes and yet do not break often, we know SAND pipes perform well over a very long period of time.

3 Forecasts

3.1 Modeling Breaks from Season and Temperature

Before building a model to predict the number of breakages, we first wanted to know if the four seasons actually have distinct patterns in terms of the number of breakages. Figure 13 shows that they indeed do. The range of temperature varies as well as the maximum number of breakages reached. The number of breakages is always an integer, but to reduce overlapping and have a better sense of where all the points lie, we added a little noise with a normal distribution centered at 0 and standard deviation of 0.1.

Figure 14 gives the visualization of the prediction model

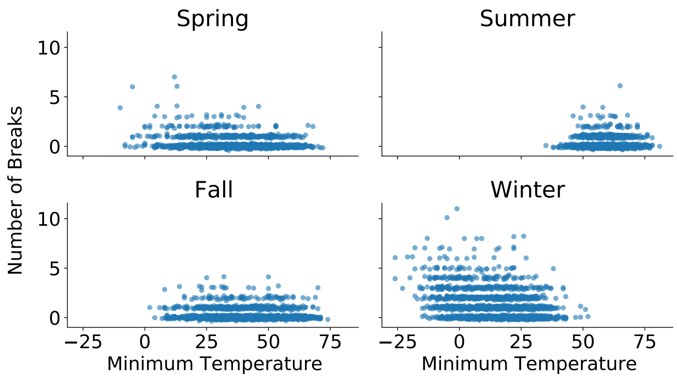

Figure 13: Number of Breakages by Soil Type categorized by Season

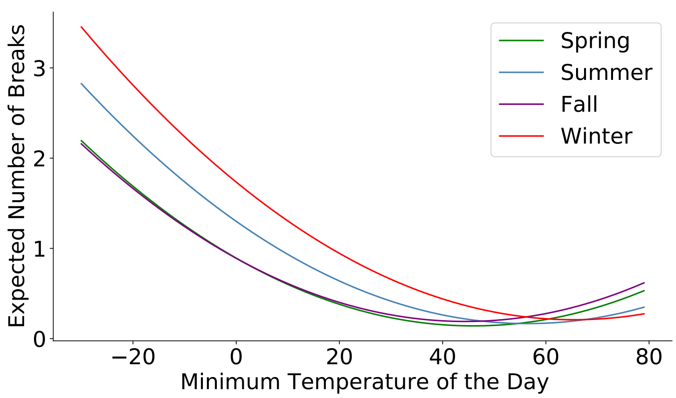

Figure 14: Season and Temp Model

given the minimum temperature of a day and what season we are in. This is a second degree polynomial regression. Season is fed in using One Hot Encoding.

The minimum temperature of the day is better at predicting the number of breaks than the maximum temperature of the day - a 3% difference in terms of explained variance score. This makes sense because most of the breaks happen in the winter, and min temperature is probably more representative of the weather on a winter day than max temperature. We tried feeding in month of the year (using One Hot Encoding) instead of season, but this changed the explained variance score by very little. So, we think splitting each year into 12 months is probably too specific, and season is a good split. We also fed in the change in temperature from the previous day to the current day, and surprisingly, this also did not really improve the model. In the end, the combination of minimum temperature and season is the best input for predicting the number of breaks on a given day, explaining 26% of the variance.

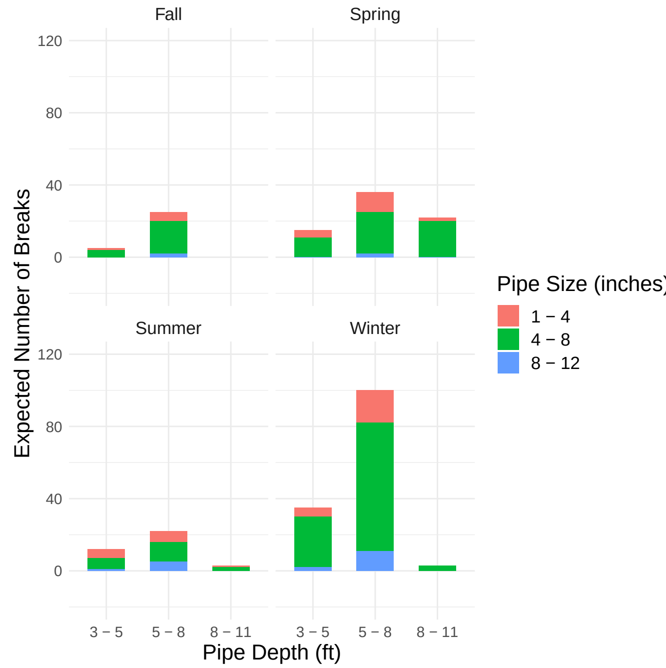

Figure 15: Predicting Breaks for 2021

3.2 Modeling characteristics of breaks

Figure 15 predicts of the number of breakages for different seasons of 2021 with the pipe diameter in inches and pipe depth in feet. Using the data on the pipe mains broken in the month of January during the past 20 years, this model forecasts the number of pipe breaks in future seasons based on diameter and depth. From Figure 15, we infer that during Winter we expect around 125 pipes to break. This will be the highest among the seasons and a majority of these pipes will be between 5 feet and 8 feet deep and between 4 metres and 8 metres in diameter.

The model in Figure 15 incorporates 12 least squares regression for separate months of the year and generalizes the season, resulting in increased accuracy. The features used in regression form cubic polynomials for pipe size and depth and square factors for subsequent years.

The accuracy of this model is estimated in Figure 16 where the same least squares problem is trained on data before 2018 and tested against the actual pipe breakages for 2018. A positive difference indicates the actual breakages are more than predicted while a negative difference indicates otherwise. The model is fairly accurate. Accuracy can further be increased by incorporating the complete data set for all the pipes that are present. Most of the error revolves around the common categories.

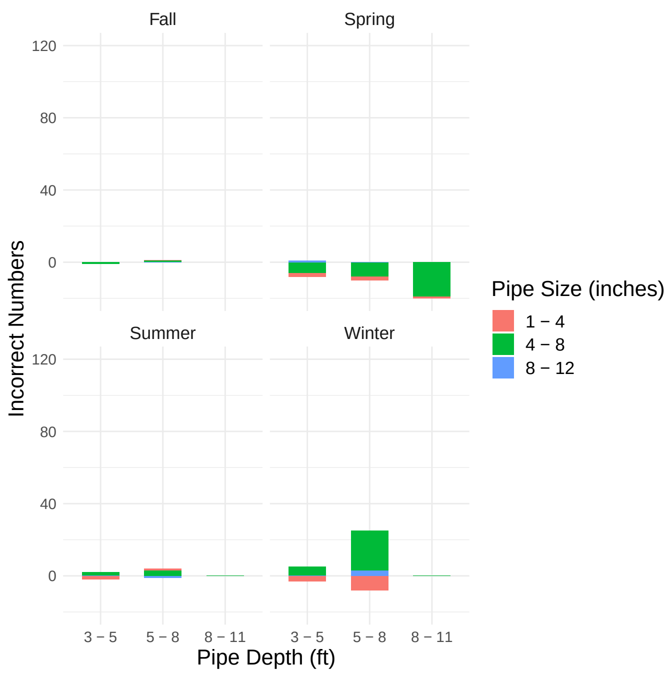

Figure 16: Testing the Predictive Model

3.3 Time interval between breaks

We also want to investigate the duration between breaks. For example, suppose a pipe main breaks and the city fixes the pipe. On average, how long will it take before the pipe breaks again? What factors influence how long the pipe can last before breaking again?

First, we want to examine our pipe main breaks data set and analyze how often each unique pipe appears. Many of the pipe breaks in the data set were recorded to have occurred on January 1, 1970, a placeholder date. Since we are interested in duration between breaks, we filtered out the data where the date of the break was unknown. Without this placeholder date, the dates in the data set ranged from the beginning of 1997 to the beginning of 2020.

Figure 17 is the CDF for the distribution of the number of breaks recorded for each individual pipe in the filtered data set. The majority of pipes only have one recorded break. Of the pipes that have more than one recorded break, most pipes have 2 or 3 recorded breaks. There were a few pipes that broke over a dozen times over a 20 year span.

For each individual pipe, we are interested in the duration between the first recorded break and the second recorded break. Figure 18 is the CDF for the distribution of these duration lengths. Note that the end of the CDF does not reach 50 percent since the majority of pipes in the data set have only one recorded break.

Looking at the y-intercept, we notice that some pipes actually have 0 duration between first break and second break, meaning two breaks for the same pipe on the same day were recorded in the data set. There are also pipes with two breaks differing by only one or two days.

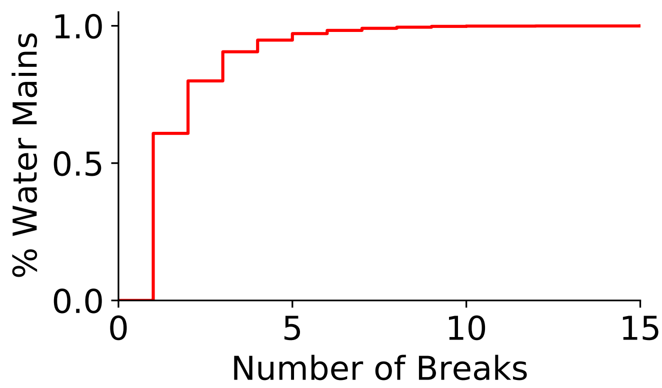

Figure 17: Number of Recorded Breaks

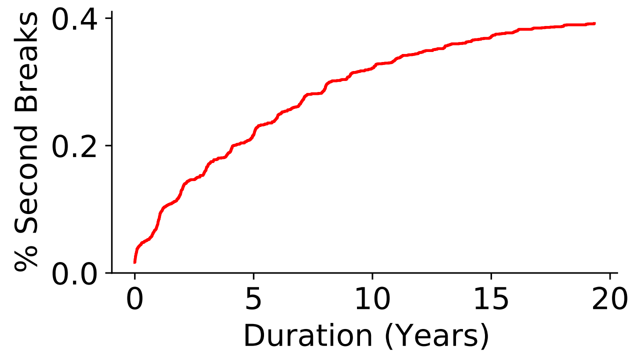

Figure 18: Time between First and Second Break

We tried to model the length of time between the first time a pipe breaks and the second time a pipe breaks. Using year of the first break, pipe depth, season during which the first break occurred, and whether rock was present in the soil type, we tried a linear regression model. Unfortunately, the model suggested that none of these input variables had a statistically significant effect (using a typical alpha level of 0.05) on the length of time between first and second break.

Looking more closely at Figure 18, we realize that within the first 5 years after a pipe breaks for the first time, around 20% of pipes break a second time. However, from 5 years to 10 years after a pipe breaks for the first time, only around 10% of pipes break a second time. This suggests that pipes that have broken more recently have a higher risk. We investigate this possibility in the next part.

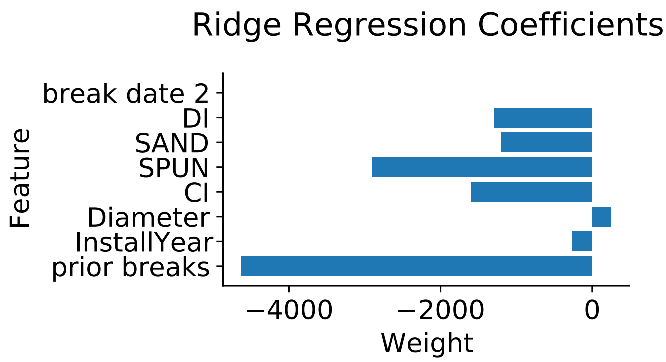

Figure 19: coefficients of Model Parameters

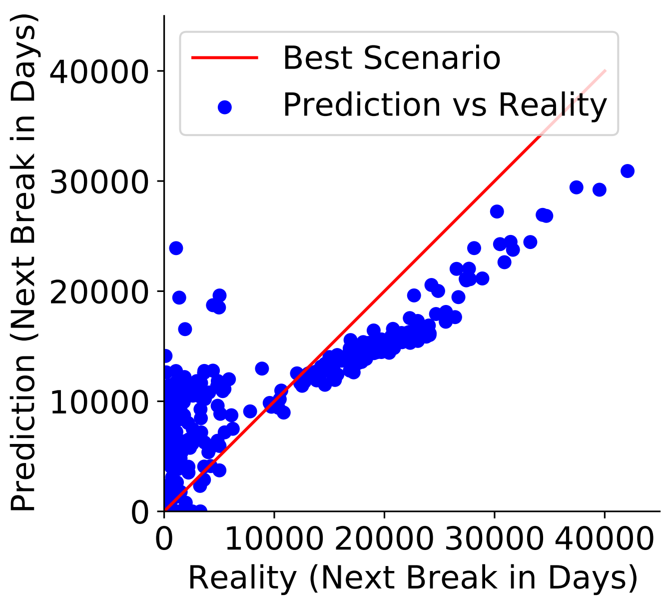

Figure 20: coefficients of Model Parameters

3.4 Risk level of active water mains

Finally, we built a machine learning model to assess the risk levels of active water mains in the city of Madison. We used a ridge regression model that incorporates the useful factors discussed in the previous section of the report. These factors include information on the material and size of the pipe, the pipe installation year, the number of times the pipe has already broken, and the last time it broke. Figure 19 shows the resulting weights for every factor in our model. And we can see that the number of prior breaks weighs very heavily.

In Figure 21, the water mains are color coded on a continuous spectrum with respect to the risk level. A pipe with 100% risk is a pipe that the model expects to break within one year from the last recorded date in our dataset, 2020/02/06. We expect pipes with 75% risk to break within 5 years, 50% risk to break within 15 years, 30% risk to break within 25 years, and 15% risk to break over 25 years from 2020/02/06. We see that there are a couple dark red pipes scattered throughout the city and a decent amount of bright red pipes that we expect to break within five years on the outskirt of the city. We consider a pipe to be safe if we do not expect the pipe to break within 25 years. Looking at Figure 18, the longest time interval for all the pipes that have broken more than once is 20 years, which is less than 25 years.

We tested the model by training the model on 90% of the data and testing the model on the remaining 10%. Figure 22 compares the predictions and reality for past breaks. The predictions for the test data closely imitate reality.

4 Conclusion

From examining pipe characteristics, we found that pipes with smaller diameter are more prone to breaks. Pipes with diameter 4-6 inches account for 75% of the breaks, but only 35% of all pipes have diameter 4-6 inches.

We also found that pipes made of CI or SPUN have a very high rate of breakage and become much more prone to breaking in the winter. Therefore, we recommend eliminating the installation of new CI and SPUN pipes. When CI and SPUN pipes break, we recommend replacing them with DI pipes if possible. On the other hand, SAND pipes perform decently despite their age. If the city considers replacing SAND pipes, we recommend not replacing SAND pipes with DI pipes until the CI and SPUN pipes are replaced.

Outside of the pipe itself, if the city has choices on the soil type when installing new pipes, we recommend installing water mains in soil that includes gravel because the nature of gravel allows water mains to be influenced by water in the soil to a smaller extend.

By looking at season and temperature, we see that not only does temperature itself impact the number of water main breakages on a given day, but season and the change in temperature also have detrimental effects on the number of water main breaks.

To conclude, these are the recommendations we want to make based on the characteristics of the pipes that make them vulnerable. And by assigning risk levels to water mains, we hope we can bring insights to the city when it comes to planning ahead. Some future directions for this project may include looking at road pavement together with water mains or examining the relationship between water utilization and water main breakages.

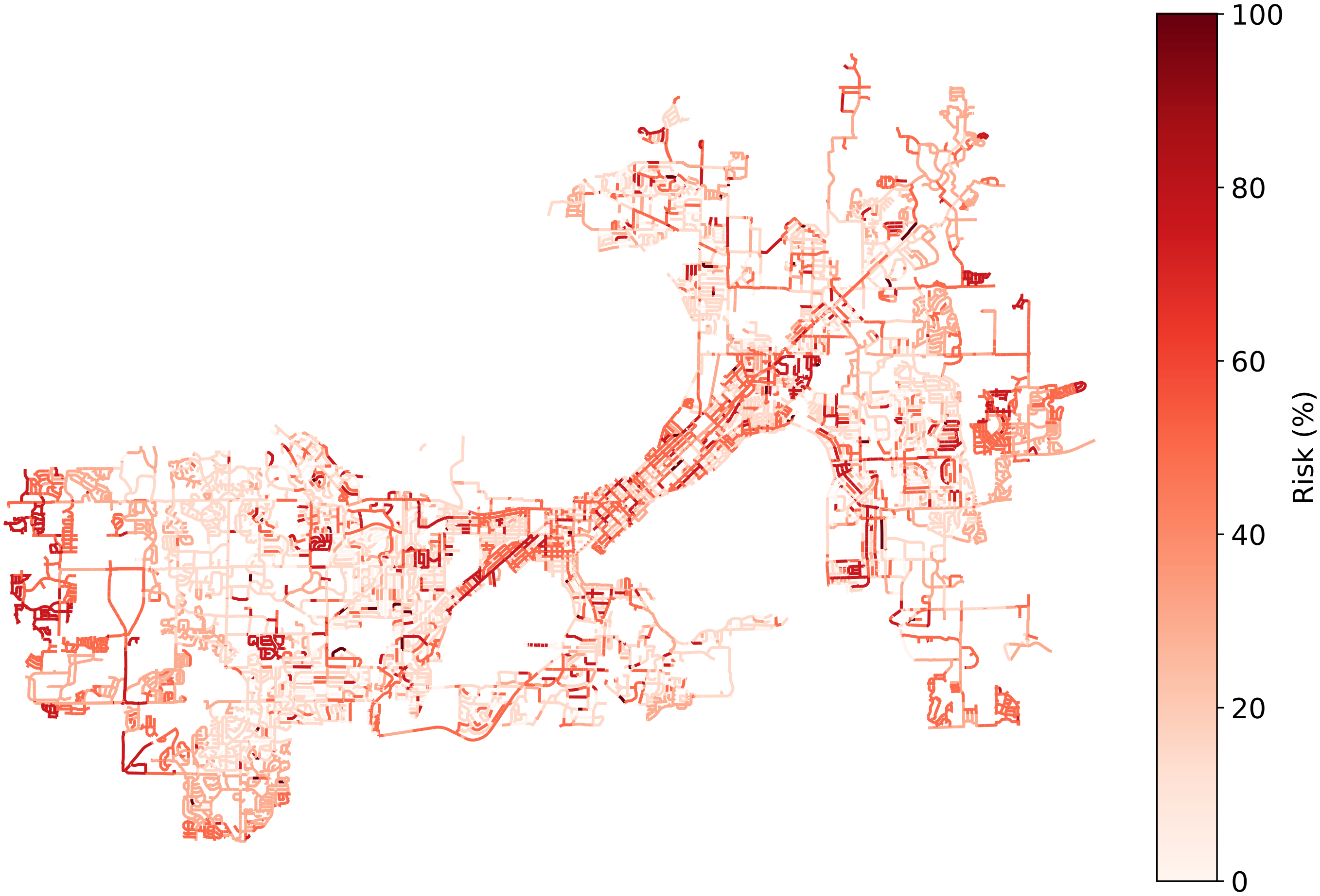

Figure 21: Risk Level For Different Pipe Breaks

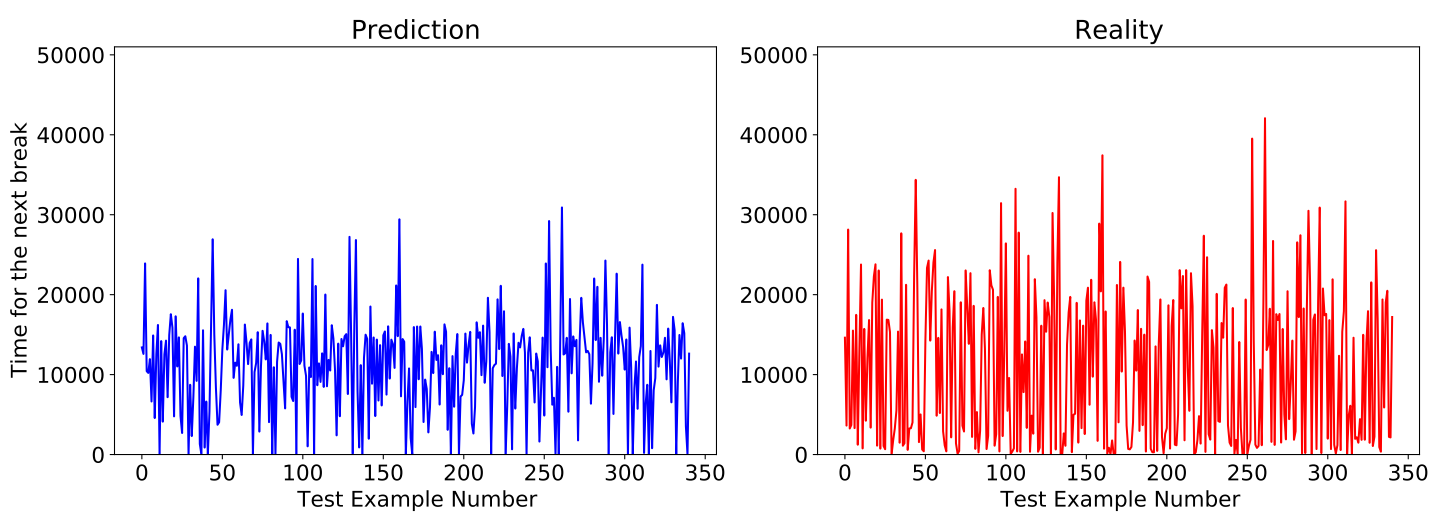

Figure 22: Testing The Accuracy Of Risk Assessment Model

References

[1] "City of Madison - Water Utility - Sustainability - Infrastructure Overhaul." *City of Madison*, www.cityofmadison.com/water/sustainability/infrastructure-overhaul. Accessed 30 Nov. 2020.

[2] Wodalski, Ed. "Frost Gone Wild: Wisconsin Utilities Hit Hard by Abnormal Winter." *Municipal Sewer and Water*, 4 Mar. 2014, www.mswmag.com/online_exclusives/2014/03/frost_gone_wild_wisconsin_utilities_hit_hard_by_abnormal_winter.

[3] "Frost line." *Wikipedia*, Wikimedia Foundation, https://en.wikipedia.org/wiki/Frost_line. Accessed 6 Dec. 2020.

[4] "Design Criteria for Water Distribution System." *Western Municipal Water District*, Jan. 2011, https://www.wmwd.com/DocumentCenter/View/239/Developer-Handbook-Section-2?bidId=#:~:text=The%20depth%20shall%20be%203.0,depth%20shall%20be%204.0%20feet.

[5] Gillespie, Evan. "Gravel to Control Water Accumulation." *SFGATE*, Hearst, https://homeguides.sfgate.com/gravel-control-water-accumulation-45821.html.

[6] "Climate in Madison, Wisconsin." *Sperling's Best Places*, https://www.bestplaces.net/climate/city/wisconsin/madison#:~:text=Madison%2C%20Wisconsin%20gets%2036%20inches,inches%20of%20snow%20per%20year. Accessed 7 Dec. 2020.