Forecasting Roadway Reliability and Travel Times

Congestion is caused by many factors, including travel demand, events, and roadway features. FHWA's literature on travel time reliability characterizes congestion in terms of volume / capacity ratios, signal timing (on signalized roads), incidents, weather, work zones, and special events.

To model and predict travel time reliability, this research draws together hourly data from January 1, 2016 through June 30, 2019 describing roadway traffic volumes, traffic incidents, weather, work zones, and special events. The results below focus specifically on westbound I-70 to the west of the I-695 interchange near Baltimore.

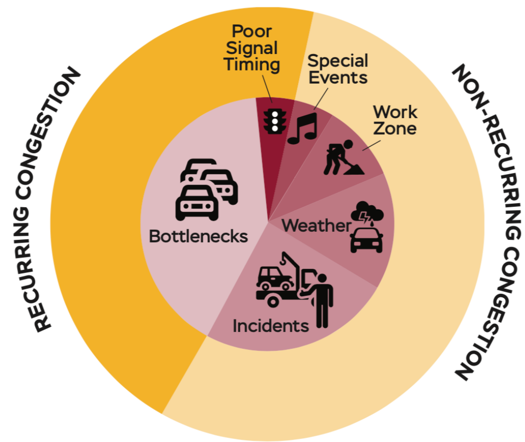

Figure 1 Causes of Congestion (FHWA)

Previous MDOT SHA work has established a statistical approach for forecasting future travel time reliability based on roadway volumes, capacity enhancements, and ITS investments. Under this task, MDOT SHA and High Street expand on this work with additional (and more granular) data. This additional data yields models with approximately twice as much explanatory power as previous work.

Modelling Approach

Multiple Linear Regression: Prior modelling used multiple linear regression to relate roadway attributes (including capacity) and volume to travel time reliability. This prior modelling explained 20 – 30% of observed variation in travel time reliability. With the benefit of additional data and greater granularity, the best performing multiple linear regression model under this new work achieved an R^2 of 0.38 (that is, explained approximately 38% of total variation in travel times) in a complex model with 31 coefficients.

Decision Trees & Random Forest Regression: Decision Trees are effective tools for estimating non-linear response variable relationships. For example, traffic volumes and travel speeds exhibit a strong but non-linear relationship. When employed in a decision tree, traffic volumes explain some 18% of travel time variation compared with only 2% of variation when used in a linear model.

The decision tree shown in Figure 2 is able to explain some 36% of travel time variation.

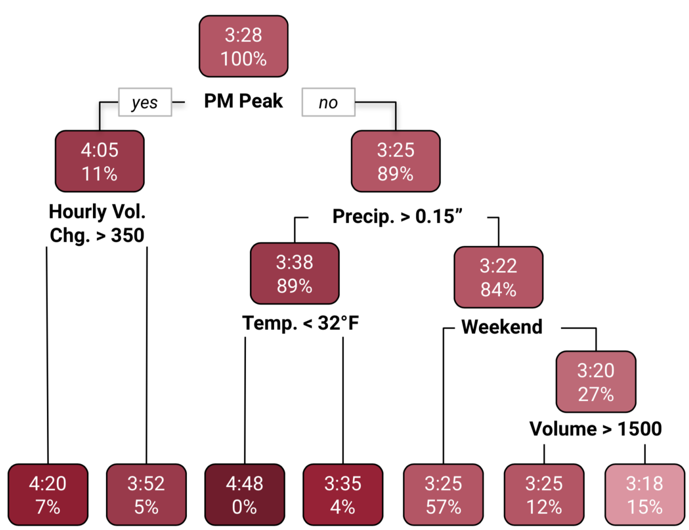

Figure 2 Travel Time Decision Tree

Impressively, this decision tree exhibits similar explanatory power with six variables to the best performing multiple linear regression model with 31 variables.

Random Forest regression is a powerful extension of decision tree regression. Training a Random Forest regression model (based on 500 individual decision trees) resulted in a model explaining 48% of total travel time variation.

Dataset	Source
Events	MDOT SHA CHART
Weather	NOAA Local Climatological Data
Volume	MDOT SHA Automatic Traffic Counters
Travel Time	NPMRDS (FHWA / Inrix)

Which Attributes Are Most Predictive of Travel Times?

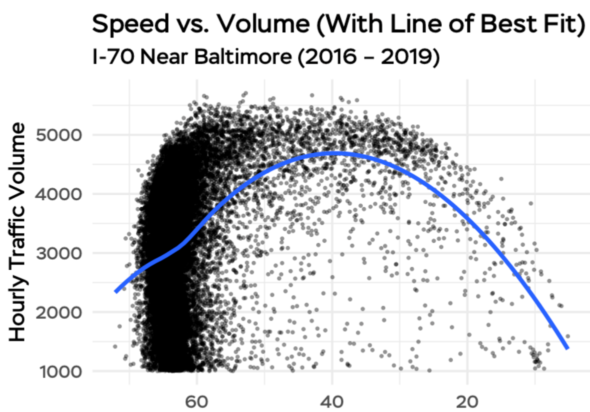

Figure 3 Volume vs. Speed

Traffic Volumes

Hourly traffic volumes are collected by MDOT SHA's Automatic Traffic Recorders. Traffic volumes correlate with both lower and higher travel times. For example, as illustrated in the scatter plot in Figure 3, lower speeds are associated with lower traffic volumes below 40 mph. Above 40 mph, lower traffic volumes are associated with higher speeds.

Traffic Incidents and Events

MDOT's Coordinated Highways Action Response Team (CHART) office produces a dataset of traffic incidents, weather events, special events, and lane closures.

Three types of traffic incidents were considered: collisions, disabled vehicles, and lane closures. Events were tabulated for upstream and downstream segments, road segments in the opposite direction of travel, as well as parallel arterial routes. In many cases, reported collisions in the dataset occur after the spike in travel times (indicating a delay between when the incident occurred and when the incident was recorded, see Figure 4). In other cases, collisions go unreported altogether (an unpublished internal study found that Waymo reports 5 – 10 times more incidents than what the CHART data captures).

Lane closure data added negligible explanatory power, possibly due in part to limitations in the data. Special events were not independently investigated given their

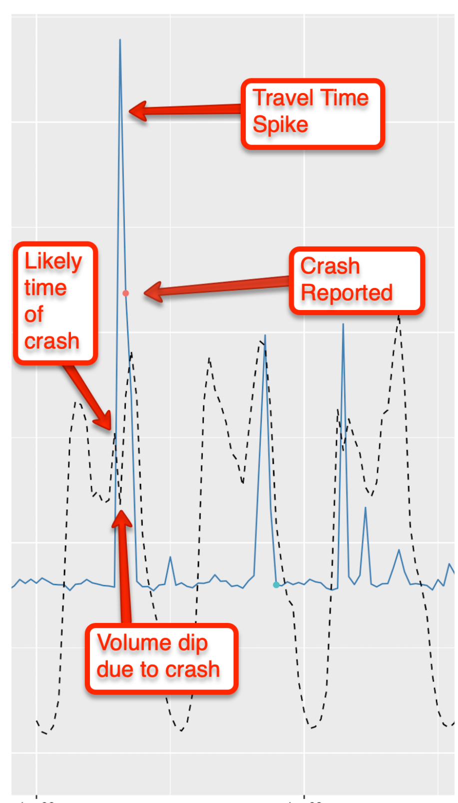

Figure 4 Delayed Crash Reporting

relative infrequency and the assumption that special events are reflected in the volume data.

Traffic incidents explain approximately 5% of travel time variation. Given that only ~1% of hourly observations in the dataset are affected by nearby incidents, this measure of explanatory power is impressive.

The Random Forest model ranked events as follows in terms of explanatory power, from most explanatory to least explanatory: 1) a collision on the same segment; 2) a collision on a downstream segment; 3) a disabled vehicle on the same segment; and, 4) a collision on the segment on the opposite side of the median.

Other events (including incidents further away) had limited explanatory power. Notably, this research fails to find evidence that events on parallel arterials impacts roadway reliability at the selected count site.

Weather

Hourly weather data was retrieved from NOAA's nearby weather station at the BWI Marshall Airport. Precipitation (and, especially, a combination of freezing temperatures and precipitation), wind gusts, and visibility collectively explain some 4 – 5% of total travel time variation.

On this westbound route, the timing of sunset explains 0.5% of total travel time variation and is strongly associated with increased collisions. Fitting a zero-inflated negative binomial model finds that, controlling for volumes, sunset increases the rate of traffic incidents by 2.2 times.

Conclusion

At the selected study site, traffic volume, incident, and weather data are together (when modelled using Random Forest regression) sufficient to explain approximately half of observed variation in segment travel times. The approximate proportions of explanatory power are illustrated in Figure 5. This represents a significant advance over prior work, but nevertheless leaves a large portion of variation unexplained. Future improvements in data and modelling approaches may yield additional improvements in explanatory power. For now, current results suggest incomplete data, limitations in modelling, or as yet not fully understood causes of travel time variation.

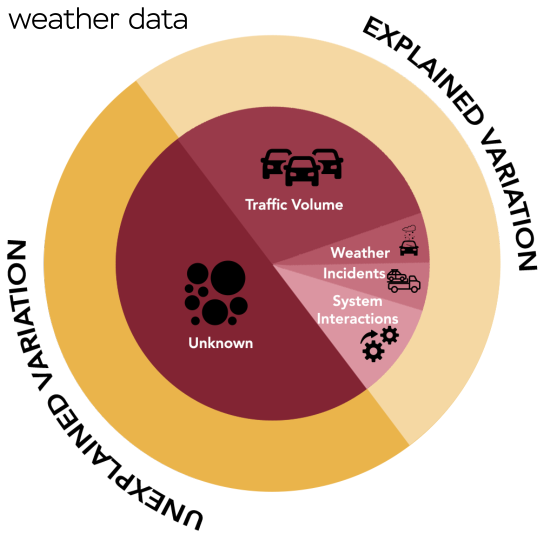

Figure 5 Modelling Results

For more information, contact:

Subrat Mahapatra, MDOT SHA (smahapatra@mdot.maryland.com), or

Mark Egge, High Street (egge@highstreetconsulting.com)