Crowdsourcing and Data Analytics for Near-Miss Statistics

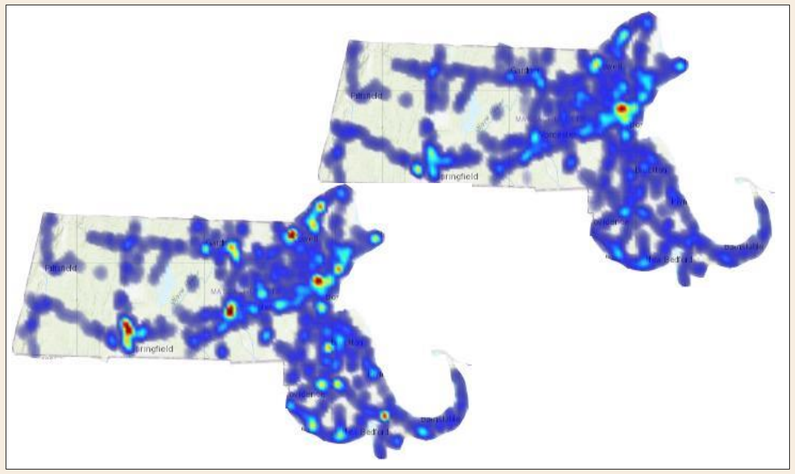
Mazdak Tootkaboni (CEN, UMassD), Arghavan Louhghalam (CEE UML), Franz-Josef Ulm (CEE, MIT)
Mohammad PourghasemiSaghand (CEE, UMassD), Meshkat Botshekan (previously at MIT)

Motivation

Near-miss statistics for different traffic situations offer significant insight for stakeholders but they are hard to quantify. This is because they are represented by extreme statistics rather than simple averaging.

Methodology

Recent research [1] integrates the physic-based model of Nagel-Schrekenberg (NaSch) with the speed profiles to compute the probability of harsh breaking, i.e., extreme deceleration from maximum to zero speed (P_5). The results show agreement with the collision risk.



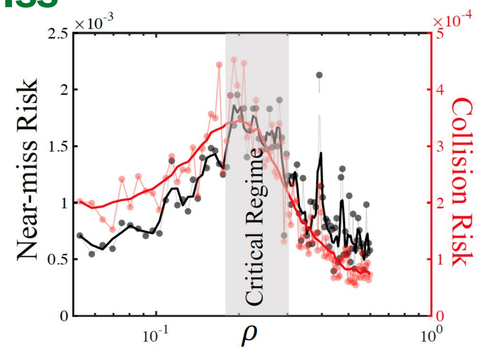
Collision risk (top) from crash data & extreme deceleration (P_5) (bottom) for the state of Massachusetts shows agreement

Maximum deceleration

Is P_5 enough to describe the likelihood of near-miss events for different traffic conditions? Do we need a more comprehensive measure to represent the near-miss event and its corresponding probability?

Proposed near-miss event

We observe deviation between P_5 and collision risk at low and high traffic densities.

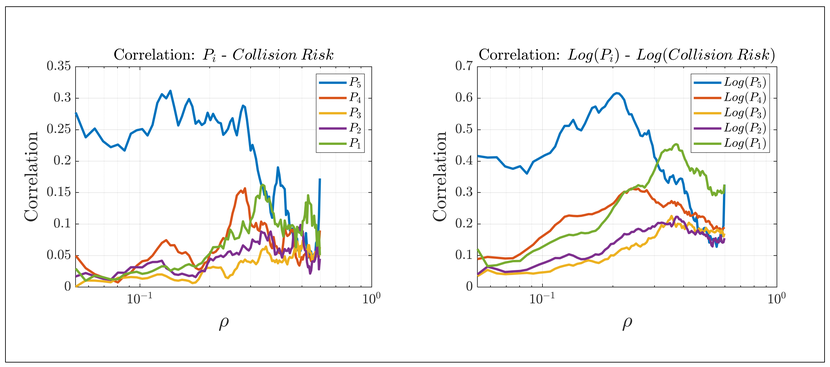


Collision risk and predicted number of max. deceleration

Our physics-based model calibrated by speed data provides the means to consider the impact of probability of other levels of deceleration (i.e., decelerating from lower speeds to zero). To this end, we discretize the speeds into five different categories V_i with i ∈ $\{1, 2, 3, 4, 5\}$, where V_5 is the maximum speed, and $V_i = i \times V_5/5$.

Results

The correlation between collision risk and probability of maximum deceleration (P_5) decreases for high density traffic road. Higher correlation with probability of lower deceleration, e.g., P_1 is observed in high density areas.



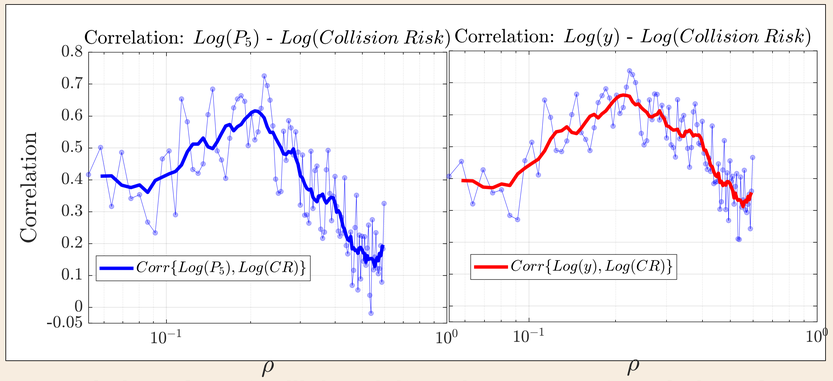
Correlation between collision risk and P_i for different traffic densities in linear and logarithmic space

We use a comprehensive definition for the near-miss event that is representative of different levels of deceleration. Noting the higher correlations between the logarithm of collision risk and $\log(P_i)$ model we utilize a power function for near-miss even:

$$y = 10^{b_0} \times P_1^{b_1} \times P_2^{b_2} \times P_3^{b_3} \times P_4^{b_4} \times P_5^{b_5}$$

Conclusions

Higher correlation between the proposed near-miss definition ($\log y$) and logarithm of collision risk is observed, which indicates the enhancement of the prediction accuracy especially in high density traffic roads.



Correlation of Log(*collision risk*) and $\log(P_5)$ & Correlation of $\log$(*collision risk*) and $\log(y)$

Acknowledgments

Financial support from Michelin North America and Michelin DDI is acknowledged.

References

[1] M. Botshekan and F.-J. Ulm, “Spatial and temporal memory effects in the Nagel-Schreckenberg model for crowdsourced traffic property determination,” *Physical Review E*, vol. 104, no. 4, p. 044102, 2021.

[2] Massachusetts Crash Query and Visualization https://apps.impact.dot.state.ma.us/cdv/