Dispersion Patterns of Vehicles on Highways

Ashwin Naren

May 2024

Introduction

Traffic is an issue that every driver faces, the average driver in 2022 spent over 51 hours in traffic every year [1]. Traffic is most apparent in highways, where there are often long lasting traffic jams which take hours to clear up. Highways are easy targets for statistical analysis due to two key factors: constant velocity and predictability.

On a highway, people strive to maintain a constant velocity, barring any adverse scenario, like a speed check by a police officer, which allows for easy extrapolation. This is in contrast to regular roads where there are traffic lights at regular intervals.

People on an highway are also predictable, they don't switch lanes randomly or turn, they stay in their lane and move forward.

These two factors allow us to calculate the dispersion of cars with minimal error. To calculate the dispersion, we simply divide the variance by the mean. I would suspect that more traffic leads to a random dispersion, due to cars being equidistant from each other, and less traffic leading to clumped dispersions, from personal experience. Around here, traffic heats up around 4-6pm, so I would expect the index of dispersion to be closer to 0 during those times, and the index of dispersion to be above 0 at all other times. It is likely that time and index of dispersion can be modeled with a polynomial or sinusoidal best fit line, but to be sure that time and index of dispersion are correlated, Spearman's Correlation Coefficient can be checked.

Knowing the dispersion pattern of cars can be useful in the prevention of traffic, for example triggering meters on the on-ramps to help fill the gaps in a clumped dispersion of cars.

Methods

Videos were retrieved from the California Department of Transportation [2]. This data is publically available on their website and is realtime. I choose locations that were of reasonable camera quality, with equally spaced lanes and equally spaced lane markers. The reason for the latter two will be shown later. Another important factor was the proximity to an exit, as if the spot was too close to an exit, there would be too many merges, which leads to non-constant velocities. The CCTV camera must, of course, be recording enough cars for statistical analysis. The video was scraped using a python program and saved for further processing. One crucial step of data collection was finding the appropriate camera. In the end, I settled for an isolated strip of CA-17, a seaside highway with little exits that connects the Bay Area to Santa Cruz. It is a four-lane road, which minimizes the possibility of merging while still lacking traffic lights.

Processing

Data Processing can be split into two steps. Finding the positions of the cars on the screen, so called "virtual" position calculation, and calculating their real life positions.

Virtual Position Calculation

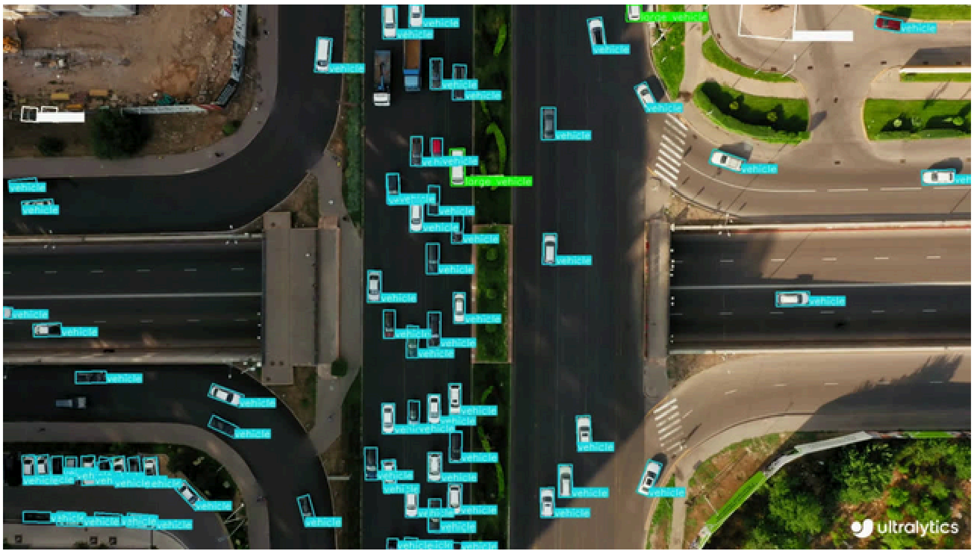

Figure 1: Image of highway on which cars are bounded with boxes using YOLO v8

To calculate the virtual position of the cars, I used the You Only Look Once (YOLO) v8 model [3]. However every model needs a dataset. To get this dataset, I took data from the

Spanish Government that labels cars from drone footage and trained the YOLO v8 model on these frames. To help with data processing I used the custom detection model on 7 second snippets. Due to the slow nature of Machine Learning, I spaced the clips 10 minutes apart, so as not to slow down data processing. I ran the model on each clip using the tracking version of the model I trained. Then, I exported all this data as JSON using a custom program. JSON is a commonly used format to share data between applications. In this case, I was exporting data from the machine learning program to the calculation program.

Due to the limitations of object detection, I chose to only collect data during daytime, to minimize false positives and negatives.

Physical Position Calculation

Due to the cameras being positioned at an angle, and the highways lacking any standardized unit, I used the only evenly-spaced marking on the highway I could find, the lane dividers. As mentioned in the Data Collection section, the cameras were chosen partially on whether they had equally spaced lanes and equidistant lane dividers. From this I use linear interpolation to find the position of each car.

Figure 2: Lines juxtaposed over CCTV footage

As seen in this image, we can split the image into sections that can be used to calculate the index of dispersion. To do this we label the sections, the bottom left is (0,0), the one above is (0,1), etc., and we have the left-most lane be lane 0, and the right-most be lane 5.

My calculation program, which was written in rust, took this as an input

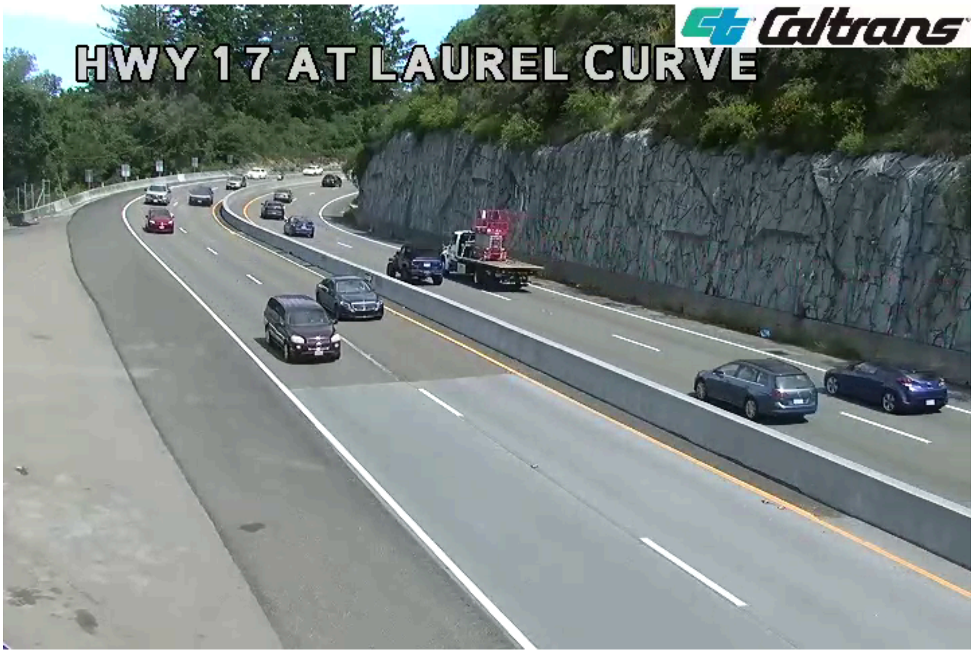

Figure 3: Image of highway on which final analysis was performed

Data

The video clips and exported car positions are available in the appendix as a link.

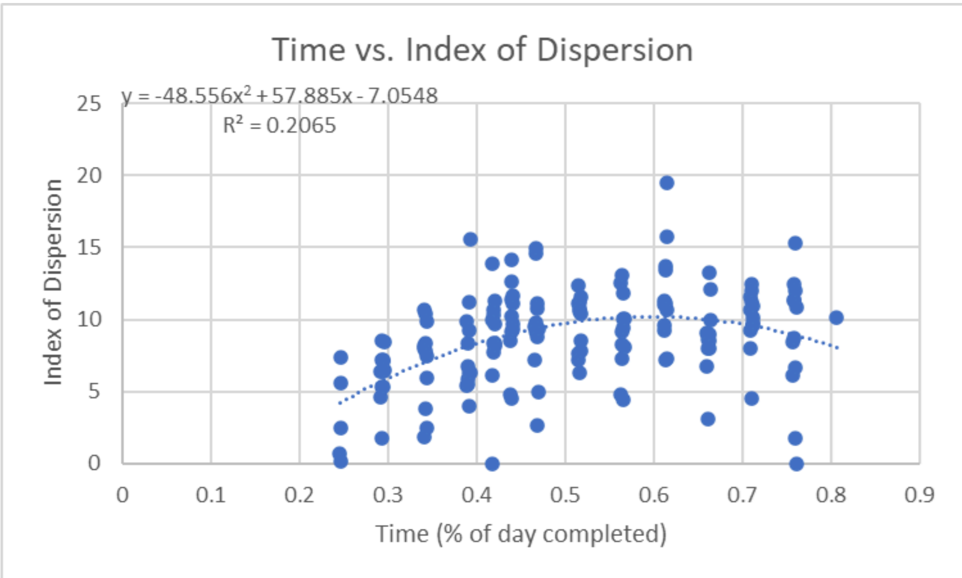

Figure 4: Time vs. Index of Dispersion

Note that time is in non-standard units, but is expressed as a percent of the day (0.5 corresponds to 12:00 PM).

Here I plot time with the index of dispersion. The index of dispersion is far higher than the threshold for statistical significance for each sample size. The best regression line was quadratic, with the equation $y = -48.556x^2 + 57.885x - 7.0548$.

It is hard to prove that this is really the best way to fit this data, but instead we can be certain that the index of dispersion is correlated with time by using Spearman's Correlation Coefficient. This is one of the few ways that I learned that can test for non-linear association. PRCC and Type 1/2 linear regression are not appropriate for this reason, and having a coefficient is helpful, so SRCC was chosen.

When I run the test (calculations linked to from appendix), I get a value of **0.4**, which suggests that these is an upwards trend in the index of dispersion until the day is 70% over, when it becomes too dark to collect further data, but if we see the quadratic best fit line, we do see the downwards trend at the end.

Discussion

These results are stronger than expected, although I am surprised by the amount of random variation. The results, however, disprove my hypothesis, the cars are more clumped during rush hour instead of more random. I suspect this is because of a lack of real traffic due to the post-COVID enviorment. We can conclude that this part of the data follows a curve that peaks at around 60-70% of the day before starting to decline. During the night, I have observed there to be no cars in the frame at times, which would lead to a dispersion of 0, I suspect that the theoretical best fit line would likely be sinusoidal however because it hits a low after it sunset and then progressively rises back for the next sunrise.

We also can tell that the cars are very clumped at all times, the data does not offer an explanation, and I cannot offer one either. The more cars, like there would be during 5:00, the more clumped-like they become. I suspect that this can lead us to the conclusion that cars on a highway prefer to stay clumped.

If I were to repeat this project, I would try even harder to get better, higher quality data. I would attempt to get a roadway that I can fly a drone or UAV over, to measure the Index of Dispersion in a less convoluted fashion, I would get higher resolution data that machine learning models can better understand, and I would use a more powerful, but computationally taxing model, like SAM (Segment Anything Model).

Further research could involve modifying on-ramp traffic lights to see its effect on dispersion patterns, or simply looking at the amount of cars over time.

Acknowledgements

I would like to thank Alistair Keiller for his invaluable guidance pertaining to model training and usage.

Bibliography

[1] "Global Traffic Scorecard | INRIX Global Traffic Rankings." [Online]. Available: https://inrix.com/scorecard/

[2] "Caltrans CCTV Map." [Online]. Available: https://cwwp2.dot.ca.gov/vm/iframemap.htm

[3] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8." [Online]. Available: https://github.com/ultralytics/ultralytics

Appendix

All the data, analysis, and scripts in this experiment (including the source fort his paper) can be found at https://github.com/arihant2math/msb_final_project.

seaside.xlsx contains the analysis of the Index of Dispersion, as well as their actual values, which I will append in comma format:

(format = YYYYMMDD_HHMMSS+".json", index of dispersion)

20240524_053500,3.267498594829846

20240524_054003,12.437032185799124

20240524_054500,0.7134075762952933

20240524_055003,27.781480130232044

20240524_055503,36.653484257126365

20240524_060000,22.81399452871773

20240524_060503,32.09928286042136

20240524_061001,35.893167642154175

20240524_061503,26.392751154817546

20240524_062000,42.622460045089056

20240524_062504,8.834373200635055

20240524_063002,36.063539505229414

20240524_063500,35.710835211720735

20240524_064000,26.533945994178247

20240524_064500,31.93738411835125

20240524_065005,32.47044659763315

20240524_065502,42.00956256969747

20240524_070003,9.10991849265757

20240524_070501,40.790908143438486

20240524_071004,53.30119122171527

20240524_071503,40.139267125031594

20240524_072003,41.73623516968639

20240524_072514,39.00689138489241

20240524_073002,19.036124933565713

20240524_073510,52.17365512155714

20240524_074004,37.2876909963228

20240524_074500,49.14446956015713

20240524_075002,29.711365688225285

20240524_075502,12.504957264919666

20240524_080000,49.34749318911952

20240524_080502,26.982498566838775

20240524_081004,33.72371129419498

20240524_081500,29.48346373477898

20240524_082000,41.56983366146502

20240524_082501,27.56183547137761

20240524_083003,19.882912886547807

20240524_083501,31.39865016163384

20240524_084001,46.075660391699465

20240524_084503,55.9672494325468

20240524_085008,31.362891895479223

20240524_085502,77.8133644493189

20240524_090004,23.759931325628123

20240524_090503,42.454373746077046

20240524_091005,22.783058631279612

20240524_091503,51.317813112088544

20240524_092004,56.92088331530573

20240524_092500,63.29339163912089

20240524_093002,70.6097586293669

20240524_093500,45.59109860720423

20240524_094000,48.55629959957149

20240524_094502,46.837440648128556

20240524_095001,55.39118054505781

20240524_095501,58.084996884136324

20240524_100003,69.51494514857872

20240524_100500,0

20240524_101004,49.67843077569439

20240524_101511,30.44941146355561

20240524_102002,51.66213343037239

20240524_102500,53.47771700049475

20240524_103002,38.48013832285836

20240524_103502,41.822509052084705

20240524_104004,48.346123692000695

20240524_104500,41.892065266581845

20240524_105002,40.8875748327038

20240524_105504,56.70050669353515

20240524_110002,47.342557411484535

20240524_110502,35.98644181797252

20240524_111003,45.711595226016115

20240524_111501,73.20401370539557

20240524_112000,49.076383714099904

20240524_112504,74.87683582468338

20240524_113004,13.417795571484673

20240524_113501,55.55659295300571

20240524_114002,44.021067967231424

20240524_114504,53.657051462856685

20240524_115002,46.855321344655685

20240524_115504,25.029353238638897

20240524_120003,61.93821019744323

20240524_120504,36.035981287173215

20240524_121002,55.78523949891508

20240524_121501,38.311167968952155

20240524_122003,31.303869717697356

20240524_122500,52.99102807260895

20240524_123001,53.7391449082386

20240524_123500,55.86122530562069

20240524_124003,42.70488437493717

20240524_124505,58.03549286366056

20240524_125004,52.009461901908566

20240524_125506,39.04184735634228

20240524_130003,24.071372311316196

20240524_130502,62.927378425121155

20240524_131002,36.42568564350972

20240524_131502,45.91899759682412

20240524_132006,41.25374445151864

20240524_132506,65.60010795622966

20240524_133005,47.18045180666048

20240524_133505,22.050741797367756

20240524_134005,59.12291484797663

20240524_134505,50.1335220707122

20240524_135003,40.56704253090642

20240524_135506,50.09567872769869

20240524_140004,46.31142495719675

20240524_140505,55.26470355175694

20240524_141003,48.04734131265402

20240524_141503,56.478902737380004

20240524_142004,68.61542010228504

20240524_142506,55.39084045424274

20240524_143002,67.2096346903906

20240524_143502,36.11382525759746

20240524_144006,53.51313948058165

20240524_144502,97.67053026746467

20240524_145005,36.44603826042448

20240524_145511,78.90675450690212

20240524_150006,33.684366226863766

20240524_150506,45.190227904609586

20240524_151005,15.309141215606154

20240524_151504,39.79328867621166

20240524_152006,42.920634422410004

20240524_152503,44.884634591940056

20240524_153003,42.49595243552165

20240524_153504,44.93303619850728

20240524_154005,66.34249321615907

20240524_154502,40.08247006690017

20240524_155002,60.586359369031875

20240524_155504,49.763546389198154

20240524_160002,57.819552046464814

20240524_160503,46.06372060940436

20240524_161002,40.14015953534413

20240524_161505,53.458035701889735

20240524_162003,62.37371595988831

20240524_162502,56.19610295208847

20240524_163003,59.830938828184884

20240524_163504,22.406410001553255

20240524_164007,48.17359608450836

20240524_164506,49.05645529490801

20240524_165004,50.489419491015205

20240524_165504,54.57602600725727

20240524_170003,42.22175753561779

20240524_170503,30.709980424680005

20240524_171006,62.3021774228678

20240524_171504,43.706600220627436

20240524_172006,56.97665637058366

20240524_172503,56.42947725810916

20240524_173003,8.850707546957967

20240524_173503,60.15063215327072

20240524_174004,33.072434780169935

20240524_174505,76.43197861719408

20240524_175006,0

20240524_175505,54.401253046047735

20240524_180006,50.68830214502809