Eindhoven University of Technology

Modeling with Mathematics

2WH70

Traffic Jams in Houston, Texas

Author	*ID Number*
A.P.O Lefevre	1517600
E. Zambolin	1521101

September 27, 2020

Technische Universiteit Eindhoven University of Technology

Contents

1 Introduction **2**

2 Model Development **3**
2.1 Maximal Flux of Vehicles (MFV) 3
2.1.1 Safety Interval 3
2.1.2 Distribution of cars, buses and trucks 5
2.1.3 Speed difference between lanes 6
2.1.4 Flow 7

3 Data Analysis **12**
3.1 Overview 12
3.1.1 Real flux after the crossroad 12
3.1.2 Real Flux VS. Theoretical MFV 12
3.1.3 Degree of Saturation 13
3.1.4 Travel Time 14
3.2 Lane Speed combination analysis 15
3.2.1 Lanes & Speeds Statistics based on our model 15
3.2.2 Lanes & Speeds Statistics based on provided data 18
3.3 Critique and Additional Analysis 27
3.3.1 Critique of the Model 27
3.3.2 Complementary Analysis on the Real Flux 28

4 Selection & Solutions **32**
4.1 Highways/section selection 32
4.1.1 Spacial Analysis 32
4.1.2 Statistical Selection 34
4.1.3 Final Selection 36
4.2 Modification of the PNR 40
4.3 Results 43

5 Conclusion **46**

6 Annex **48**

1 Introduction

A well-known phenomenon that happens and characterizes any busy civilized cities is *traffic jam.* All around the world more and more people are using their car not only to travel long distances but on a daily routine, as it can be going to work, doing grocery, carrying kids around for school, sports, meet up with friends. Researches showed that on average there is one car for every 7 inhabitants on Earth, which means that the world park is nine-zeroed. Considering these numbers and the fact that usually the roads people choose to take are the most frequently used to head from a place A to a destination B, and that they do it in similar time slots (due to the fact that our lives are all scheduled with similar timing that society has imposed), it seems inevitable to encounter traffic at some point of the day. The main complain about high traffic is the waste of time for the users. According to the AAA Foundation for traffic Safety [4], on average an American citizen spends 50 minutes/day on driving. In this report we will focus our attention on the city of Houston, Texas. In particular we will analyze and try to improve its road network (working on its weaknesses), in order to provide the user with a better service. What we're asked to do is either reducing/modifying the traffic or improving the physical network of routes (PNR), or both. By reducing/modifying the traffic we mean asking people to adapt their behaviour on the roads by using detours, decreasing their cruising speed, use traffic lights or put parking limitation and so on. As pointed out on the project description, in general in Europe, most cities try to act upon the traffic directly, for example by asking people to use bicycles instead of cars. The Netherlands is a perfect example of this type of approach: with almost a quarter of the population cycling every day, it is the country with the biggest number of bicycles per inhabitants [2]. But this is not always the best choice as it may seem at first: even if the Dutch economy is doing well, what we've been suggested to think about is that taking such decisions can be a slowdown for the economic activity besides restricting the freedom of the user (which is not recommended). The only option left would then be the PNR modification. To improve the physical network of routes means enlarging the ways, remove obstacles, add lanes, build new routes and so on. Obviously, such an operation would be more costly due to the fact that we're dealing with the 'hardware' of the system.

As previously mentioned, the goal of this project is to reduce the traffic in the city of Houston, Texas. In order to do this, we have been provided with a set of data, characterizing the vehicles' fluxes around all Houston's highway system, on an ordinary day: January 29th 2018, between 8 a.m and 9 a.m. In particular, for each segment of highway, we've been given: its number of lanes, its speed limit, its in-going and out-going flux, it's location in the system and it's length.

We will firstly focus our attention on the identification of those segments of highway where traffic jams are mot likely to occur. In order to perform such a research we will develop a mathematical model for the maximal flux (the MVF) and we will compare the obtained results with the actual flux for each section(provided by the data set).
After identifying an locating traffic jams, we will be able to get a clear global picture that will allow us to properly analyze the data provided and how our model fits the data. Finally, after getting a full view of the nature of all the fluxes characterizing all the branches of Houston's highways system, we will propose some modifications of the PNR which could help decreasing traffic. For this last step we can proceed in different ways; the most common and user-friendly would be to choose a small and relevant number of characteristics trips, which are among the most used ones and determine benefits for these trips only. Even if this seems to be the better approach (it limits the number of simulations to perform allowing us to focus on qualitative and representative sections and it's easier to be understood by the public opinion), our analysis and the results it would lead, will stress some conclusions that will make us prefer a different method.

2 Model Development

As mentioned in the introduction, we'll first determine where traffic jams occur and how they impact the user's experience. The goal is to get a clear picture of the various fluxes on all branches of Houston's highways system.
In order to find where traffic jams occur, we will first compute the maximal flux of vehicles for each section of highway. Then we'll analyse the data-set and compare the computed MFV with the actual flux.

2.1 Maximal Flux of Vehicles (MFV)

A key concept for the project is the notion of *maximal flux of vehicles* (MFV), which measure the maximum number of vehicles which can go through a lane during one hour, without creating traffic jams and so without increasing the travel time. In our project the MFV is assumed to be constant over a 1-hour interval; which means that with a MFV equal to X vehicles per hour, we assume that in 10 minutes $\frac{X}{6}$ vehicles would pass through the section in analyis. Moreover in our case the MVF represents also the maximal flux at which the travel time is optimal. In other words, the maximal flux that still allows you to drive at maximum speed.
Anyway we have to keep in mind that if a section happens to have a Real Flux higher than our theoretical MFV, this does not mean that the section is highly congested or that traffic jam will occur automatically; but it instead represents the fact that the cruising speed will need to be adapted with the consequence of an automatic increase of the travel time.

In order to compute the MFV we need the following information:

- v : speed limit of the section
- d : safety interval between vehicles
- n : number of lanes
- $C_{\%}, B_{\%}, T_{\%}$: percentage of cars, buses and trucks on the roads
- L_c, L_b, L_t : length of each type of vehicle

From the supplied data set, we already know for each section the speed limit and the number of lanes. Using the speed limit, the distribution of vehicle types and their respective length; we can compute the safety interval needed between each vehicle in order to avoid accident.

First of all, we know that the flux of vehicles is defined for each lane. It means even for two parallel lanes of the same section, the flux can vary. This difference is especially observable on highways. Usually, the left most lane drive at an higher average speed than the right most one. In order to be as much precise as possible we should take this difference into account. Moreover, we should estimate the proportion of cars, buses and truck in the traffic to have a more representative MFV. This is mostly due to the fact that the safety distance between two vehicles should be adapted in function of the configuration (a car following a truck, a truck following a bus, etc.).

In order to compute the MFV we first need to compute ourselves some parameters (such as the safety interval and flow of vehicles); in the sub-paragraphs that follow we will get through the final result step by step.

2.1.1 Safety Interval

In order to compute the MFV, we have to determine the safety interval needed between two vehicles to avoid accidents (and so, in a way, avoid possible traffic jams related to them). To compute this interval, we need to determine the total braking distance in function of the speed. The total distance travelled in between the moment the driver realizes he has to stop and the moment the vehicles reaches a zero velocity, is composed of two parts : *the perception-reaction distance*, which is the distance travelled while the driver perceives he has to stop and acts on it and the *the braking distance*, which is instead the amount of space needed by the travelling vehicle to stop.

Perception-reaction distance

$$Dist_{p-r} = v \cdot t_{p-r} \tag{1}$$

Braking distance [7] :

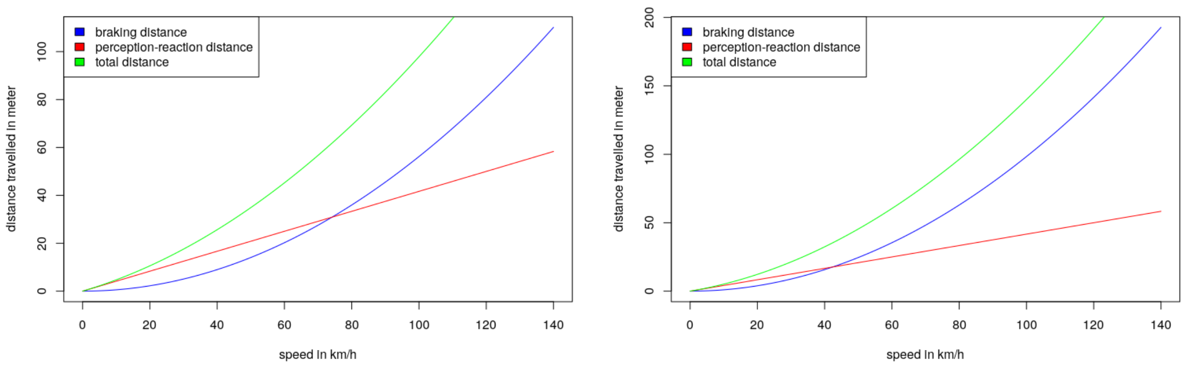
$$Dist_{braking} = \frac{v^2}{2 \cdot \mu \cdot g} \tag{2}$$

From the formulas above we get the following equation for the *total braking distance* :

$$\begin{aligned} Dist_{total} &= Dist_{p-r} + Dist_{braking} && (3) \\ &= v \cdot t_{p-r} + \frac{v^2}{2 \cdot \mu \cdot g} && (4) \end{aligned}$$

where v is the speed $[m/s]$, t_{p-r} the reaction time $[s]$, μ the coefficient of friction and g the gravity on earth $[m/s^2]$.

According to the National Highway Traffic Safety Administration, the average reaction time is 1.5s [1]. The coefficient of friction depends on the vehicle, the weather, the quality of the road and more other aspects. In general, some standard typical values for μ, that can be applied in most of the cases are $\mu_{dry} = 0.7$ for dry conditions (which will be the one we'll use in our model), and $\mu_{wet} = 0.4$ when the road is instead wet (we can notice how this value is almost half the one used for typical weather conditions). Using these values, we got the two following plots [figure 1] [*totalBrakingDistance.R*] representing the braking, perception-reaction and total distance travelled in function of speed for both cases



(a) **Dry condition**

(b) **Wet condition**

Figure 1: Braking, perception-reaction and total distance travelled in function of speed

In order to simplify our model, we will only consider the scenario of optimal weather, where highways are "dry". Using the total breaking distance, we now need to estimate a 'suitable' safety interval between vehicles. We schematized the problem (in a simple way) as below :

They are two cars c_1 and c_2 at initial distance d_i from each other and driving at the same speed v. At instant $t = 0$ the car c_1 brakes and travels a distance equal to $D_{1,braking}$. The second car c_2 will first process the action, during which it will travel a distance $D_{2,p-r}$, and then will brake, travelling a further distance $D_{2,braking}$. The total distance travelled by c_2 will then be $D_{2,total}$. Once both cars are stopped we assume no accident happened if c_2 is at a reasonable distance d_f from c_1

To solve the above problem and determine the safety interval needed between two vehicles, represented here by d_i, we should then solve this equation :

$$\begin{aligned} d_{i,minimum} &= d_f - D_{1,braking} + D_{2,total} &(5)\\ &= d_f - \frac{v^2}{2 \cdot \mu \cdot g} + v \cdot t_{p-r} + \frac{v^2}{2 \cdot \mu \cdot g} &(6)\\ &= d_f + v \cdot t_{p-r} &(7)\\ &= d_f + D_{2,p-r} &(8) \end{aligned}$$

for different values of d_f and different speeds. If we plot this equation we get the (predictable) figure (2) [*safetyInterval.R*] :

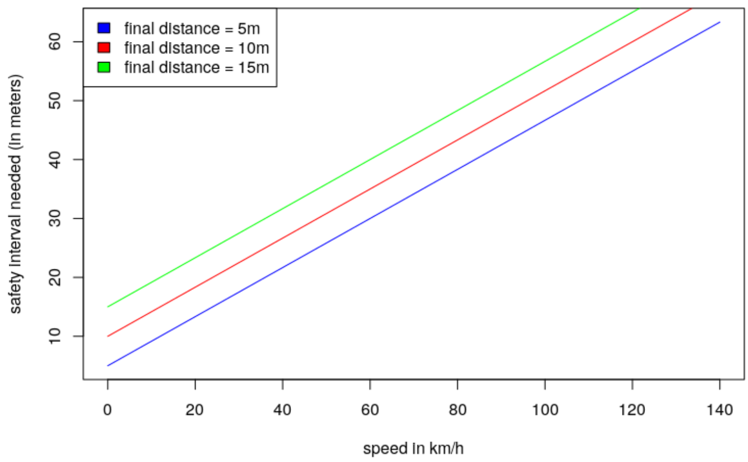


Figure 2: Minimal safety interval needed in function of the speed, for three minimal final distances

From combining all the information we have acquired so far, we can now define a function which returns the minimal safety interval needed, given the travel speed and final distance between the vehicles:

$$\Delta(v, d_f) = d_f + v \cdot t_{p-r} \tag{9}$$

2.1.2 Distribution of cars, buses and trucks

For our aim we can consider 3 main types of vehicle driving on highways : cars, buses and trucks (notice that we do not take motorcycles into account, due to the fact that they don't impact traffic as much as other types of vehicles).
To each type of vehicle we assign a characteristic length:

Vehicle Type	**Length (meters)**
Cars (L_c)	5
Buses (L_b)	10
Trucks (L_t)	15

Table 1: Length of each type of vehicles

The proportion of trucks and buses among all vehicles is left to our discretion. In order to obtain results as close as the reality we did some researches on the *Transportation in the United State*. Most of the information we found came from the governmental site web of the *Bureau of Transportation Statistics*, and especially from the heading Highway Statistics - 2018 where we found a table on State Motor-Vehicle Registrations in 2018 [6]. But after verification this table is not representative of the distribution of each type driving on highways. It only gives us the number of vehicles registered, not the actual number of vehicles on highways. We finally used figure 15.12 from the book *Transportation Decision Making:*

Principles of Project Evaluation and Programming [5] and the website *facethefactusa.org* [3] to find a suitable distribution of vehicle types on US highways and this is what we got:

Vehicle Type	**Average Daily Traffic (ADT)**	**Percentage %**
Cars (C%)	131.000	94
Buses (B%)	1.700	5
Trucks (T%)	6.950	1
Total	139.650	100

Table 2: Percentage of each type of vehicles on US's highways

We are going to use these information to estimate the MFV of each section of the highways around Houston. In fact, as each type of vehicle has a different length, knowing the distribution of each of them will allow us to estimate with more precision how many cars, buses and trucks can pass through a section of highway during 1 hour (without stressing the system); by determining the average length the vehicles in the system and the average safety distance needed.
Moreover, even if it is not absolutely realistic, for practical reasons, we will assume the driving speed to be the same for cars, buses and trucks.

2.1.3 Speed difference between lanes

As previously stated, there is a speed difference between lanes on highways, and to make our model more realistic, we will take it into consideration. In particular we associate this speed difference as probably due to two main factors. The first one takes into account the fact that you must take the left lanes (or right depending on the country) to overtake the vehicle in front of you; automatically increasing the average speed on the most left lane (but this kind of argument is not entirely relevant for US's highways). The second main reason is that when you drive on the most right lane, you will most probably take the next exit, so you have to slow down. Where vehicles driving on the most left lanes are more susceptible to travel over long distances, and by consequence are driving at a higher speed. It is very hard to find scientific report on the matter. Anyway, we found two interesting articles we focused our attention on. The first one entitled *Speed and Acceleration Characteristics of Different Types of Vehicles on Multi-Lane Highways* describes the difference of speed between the type of vehicle on different Multi-Lane highways. Even if the main goal is to observe a difference between the different types of vehicle, we can try to read between lines and compare results between six and four lanes highways. The second research entitled *Estimation of Average Vehicle Speeds Traveling on Heterogeneous Lanes Using Bluetooth Sensors* also give us some insights on the difference of speed between lanes, but it focuses on specific type of lanes (as the HOV lane, or E-ZCash) and on the software/hardware device used to compute those information. Mainly base on those researches and our own driving experience we made the following assumption on the traffic:

A section of highway has n lanes (n≥2), which are numbered from the right to the left (so lane i=1 is the most right lane while lane i=n represents the most left one). For sections with a number of lanes higher than 2 ($n \geq 3$), the first lane (i=1) will be characterized by a cruising speed equal to the 75% of the speed limit; the second lane (i=2) would have cruising speed equal to the 85% of the speed limit, while the remaining lanes ($i \geq 3$) are assumed to have a cruising speed which is exactly the speed limit of the section in analysis.

$$v_{avg,i} = \begin{cases} v \cdot 0.75 & i = 1 \\ v \cdot 0.85 & i = 2 \\ v & i \geq 3 \end{cases}, \tag{10}$$

For a section with $n = 2$ lanes instead, we will do the following assumption:

$$v_{avg,i} = \begin{cases} v \cdot 0.85 & i = 1 \\ v & i = 2 \end{cases}, \tag{11}$$

2.1.4 Flow

The main goal of this section is to model the flow (ie. the number of vehicles passing a reference point in a unit of time). Before expressing the flow we need to define the density, that can be express as the number of vehicles per unit of length. In our case we want to compute the maximal flux of vehicles, that is possible to have on each section without creating any traffic jam. It means that we have to find the maximal density so that the average driving speed stays at its highest. This can be done using two parameters: the length of a vehicle and the safety distance. We have the following equation:

$$k = \frac{1}{l + d} \tag{12}$$

where k is the density, l the length of the vehicle and d the safety interval needed between two vehicles.

We want to recall that previously we have mentioned that we have three different types of vehicle with different lengths and that the safety interval varies depending on both the speed and the type of vehicle. In order to get a model as close as possible from the reality we will compute the average length of a vehicle and the average final distance needed, depending on the configuration. We are going to use the distribution of vehicle type describe in table 2, and the length of each vehicle described in table 1.

$$\begin{aligned} L_{avg} &= C_{\%} \cdot L_c + B_{\%} \cdot L_b + T_{\%} \cdot L_t && (13) \\ &= 5.55m && (14) \end{aligned}$$

Now that we get the average length of vehicles on the traffic, we can compute the safety distance needed for the different possible configurations (which means that, if for example a truck is following a car, then the safety distance should be higher than if a car is following another car). The project description assumed that each vehicles drive at the same speed independently on the type. Then, we are not going to compute exactly the difference of the braking distance between each type of vehicle; what we will do is change the final distance needed between 2 vehicles after an emergency braking. If a vehicle is following a another vehicle of the same type (ex: car $->$ car, bus $->$ bus or truck $->$ truck) we assume that the final distance d_f should be at least 10 meters. When a car is following a bus or truck, this distance remains the same as the car brake faster than a bus/truck. When a bus is following a truck, the distance also remains the same. Conversely, if a bus is following a car, we force the final distance to 20 meters and if a truck is following a car or a bus, we then force this distance to 30 meters. In order to compute the average safety distance we have to determine the probability of each configuration. To do this, we will simply use a probability tree (see figure 3).

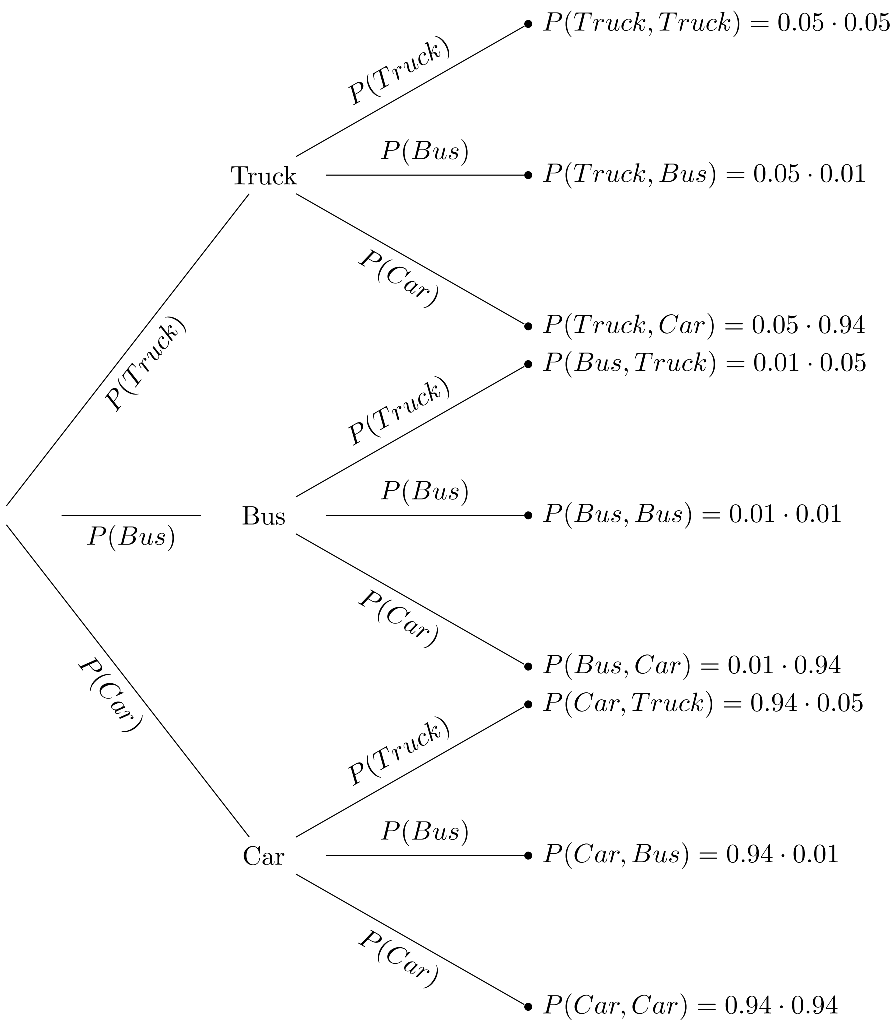


Figure 3: Different configuration and their probabilities

From figure 3 and using the required final distances described before, we now compute the average final distance needed between 2 vehicles after an emergency braking (table 3).

Identification ($config_i$)	**First Vehicle**	**Second Vehicle**	**Probability**	$d_{f,i}$ **(meters)**
1	Car	Car	0.8836	10
2	Car	Bus	0.0094	10
3	Car	Truck	0.047	10
4	Bus	Car	0.0094	20
5	Bus	Bus	0.0001	10
6	Bus	Truck	0.0005	10
7	Truck	Car	0.047	30
8	Truck	Bus	0.0005	30
9	Truck	Truck	0.0025	10

Table 3: Final distance required depending of the configuration, and the probability of the configuration

From table 3 we compute the average final distance required :

$$d_{f,avg} = \sum_{i=1}^{9} P(config_i) \cdot d_{f,i} \quad (15)$$
$$= 11.044 \quad (16)$$

On figure 4, by using (16) and the Δ function (cf. equation 9), we plotted the minimum safety distance required in function of the speed and the average finale distance required in case of emergency braking.

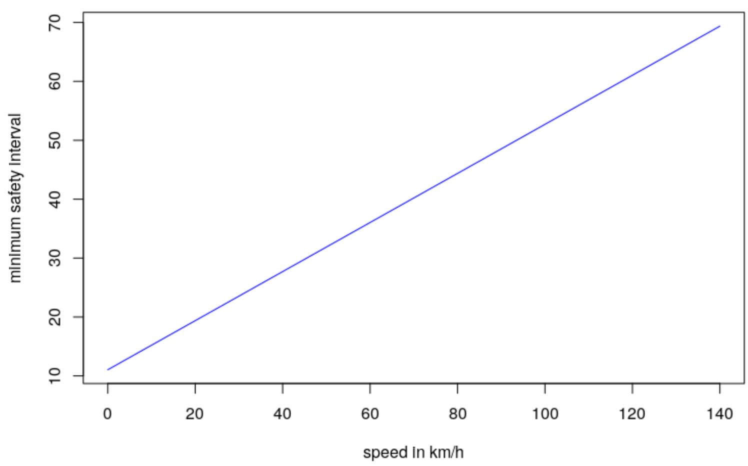


Figure 4: Minimal safety interval required in function of the speed, using the average final distance $d_{f,avg}$.

We will now update equation (12) in order to get a function that give us the average density per unit of length on the Houston's highways system for a given a speed limit. You have to keep in mind that this is the critical density, i.e.the maximum density possible such that no traffic jam occurs.
In this new equation we take into account the different lengths and the distributions of the type of vehicles among traffic.

$$\kappa(v) = \frac{1}{L_{avg} + \Delta(v, d_{f,avg})} \quad (17)$$

From equation (17) we can then establish the flow's equation as follows:

$$q = k \cdot v \quad (18)$$

where q is the flow, k the density and v the speed.

Using function (17) we can define a new function which computes the flow on one lane given the speed limit.

$$\Phi(v) = \kappa(v) \cdot v \quad (19)$$
$$= \frac{1}{L_{avg} + \Delta(v, d_{f,avg})} \cdot v \quad (20)$$

Finally, by combining equation (20), (17) and (10) we establish a formula to compute the MFV of a section in function of the speed limit and the number of lanes (n):

For n≥3:

$$MFV(v,n) = \Phi(v \cdot 0.75) + \Phi(v \cdot 0.85) + \sum_{k=3}^{n} \Phi(v) \tag{21}$$

if n=2:

$$MFV(v,n) = \Phi(v \cdot 0.85) + \Phi(v) \tag{22}$$

To visualize the progression of the maximal flux of vehicles on the highways in function of the speed limit and the number of lanes, we plotted 6 different curves representing the MFV [cf. figures 5]. We chose to plot MFV for highways with 2,3,4,5,6 and 7 lanes as they are representative of the highways' configuration in the data-set.

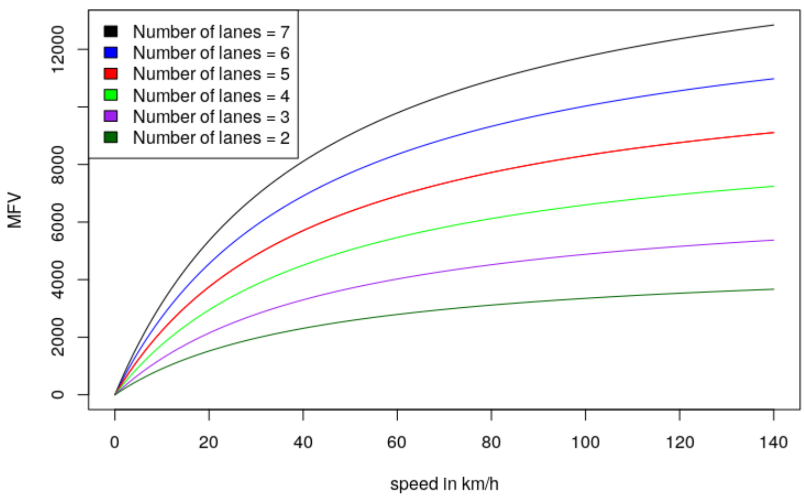


Figure 5: Maximal flux of vehicles per hour on highways in function of the number of lanes and the speed limit

On table 9 below, there's a list of the MFV for all the different configurations (available on the data set) SPEED LIMIT - NUMBER OF LANES. From now on we will use the values in this table (9) to compare the MVF with the real flux in order to find where some suggestions on where traffic jams are more likely to occur how much they will impact the user experience.

Speed Limit	**Number of Lanes**	**MFV (per hour)**
88	2	3218
88	3	4715
88	4	6367
88	5	8019
88	6	9671
88	7	11324
97	2	3319
97	3	4870
97	4	6572
97	5	8273
97	6	9975
97	7	11676
105	2	3399
105	3	4993
105	4	6733
105	5	8473
105	6	10213
105	7	11953
113	2	3471
113	3	5103
113	4	6878
113	5	8653
113	6	10427
113	7	12202

Table 4: MFV per hour for the different speed limits and number of lanes configuration in the data set

3 Data Analysis

Now that we have modeled the MVF for each possible sections (on table 9 we have computed the MVF for all the configurations), in function of the speed limit and the number of lanes, we will apply it on our data-set and compare the results with the provided real flux, in order to determine where traffic jams can occur and how.
Firstly, we will make an overview of the data-set. Secondly we will determine which sections are the most representative for a situation of congestion, and we will focus our attention on them. Before starting the data-analysis we will do a quick recap of the composition of the data set.

The data-set is composed of a list of highways *(26 in total)*, each one divided into a certain amount of sections *(647 in total)* of different lengths *(from 0,1 to 6,1 kilometers, 1,2 in average)*. For each section we also have the speed limit and the number of lanes. Finally we have the real flux of vehicles per hour thought the section and also the incoming/outgoing flux. For the data-analysis, all computations are directly made on the supplied excel file. We use .R , .m or excel scripts to plot graphs with those results and to do further analysis.

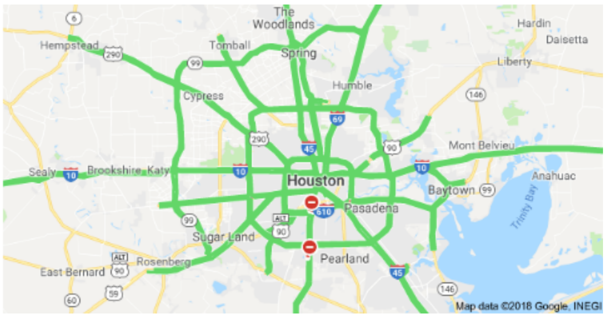


Figure 6: Simple map of the Houston highways system with highlighted highways

3.1 Overview

3.1.1 Real flux after the crossroad

For each section, the data set provides us the actual the real flux [vehicles/hour] for each sections. This can suggest us where to focus our efforts on.
In particular from the project description we know that this value is computed using the incoming and outgoing fluxes. *'For instance, on the I-10 Eastbound, at crossroad Dairy Ashford Road, the number of incoming vehicles is 1327 vehicles per hour, the number of outgoing vehicles is 2671 vehicles per hour. The traffic just before the crossroad is 14112 vehicles per hour. Then the traffic just after the crossroad is 14112 + 1327 2671 = 12768 vehicles per hour'.*
After verifying it for all sections, we found some errors; moreover taking a closer look on the data we observed that these errors mostly happen when there is an interconnection between two highways or when multiple exits are present over a small section (usually named XA, XB, XC etc. where X is the name of the exit). We have then to keep in mind that some outliers values might be due to some errors on the data.

3.1.2 Real Flux VS. Theoretical MFV

From now on, we want to get an overview of the number of sections where a traffic jam could potentially occur. We computed the MFV for each section using the formula established in section 1 (cf. equation 27). In order to know if a section is subjected to a traffic jam or not we will use the following function:

$$PotentialTJ(S) = \begin{cases} True & \bullet(S) > 0 \\ False & \bullet(S) <= 0 \end{cases}, \tag{23}$$

where S represents a general section and $\bullet$(S) the difference between the actual flux going through the section S and the MFV of that section (24)

$$\bullet(s) = RealFlux(s) - MFV(s) \tag{24}$$

Using what we've done above, we are able to compute the number of sections that respect to our model can be interested in a situation of traffic jam. Based on this, on figure 7 we can find the number of sections where traffic jams could occur. The ratio between vulnerable sections and not, is equal to 0.59. This means that almost 60% of the sections are likely to have traffic jam; but we have to be careful with this result: it only gives us sections where the real flux is greater than the MFV, but for example if the real flux is less than 5 % higher than the MFV, we can easily say that the traffic won't influence too much the user-experience. We will come back on that later.

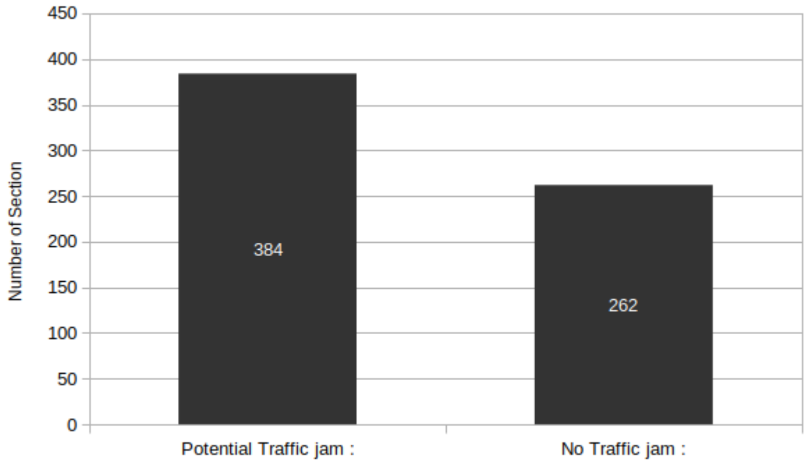


Figure 7: Potential sections subject to traffic jams vs other

3.1.3 Degree of Saturation

In the preceding section we calculated the difference between the MFV and the real flux for each highway's sections, but as mentioned before, having a real flux greater than the MFV does not automatically lead to a bad user-experience (congestion, high travel time etc.). In order to determine the impact of the such traffic on the user, we will compute the degree of saturation, also known as the volume to capacity ratio. The degree of saturation is the ratio between the demand (which in this case is the real flux RF) and the capacity (the MFV):

$$DoS = \frac{q}{C} \tag{25}$$

$$\simeq \frac{RealFlux}{MFV} \tag{26}$$

where DoS is the degree of saturation, q, the flow, C the capacity. Real flux per section is given in the supplied data set and MVF is the result of equation (27) for a given section, and can be found on table 9

This ratio gives us us information on how much congested the traffic is on a specific section of the highway. If the ratio is less than or equal to 1, it means there is no congestion and we can assume the traffic to be smooth. Conversely, the higher ratio, the more congested the section. We analyzed the distribution of this value for each section subjects to traffic jam and the result can be observed on figure 8. We decided to only consider degree of saturation (DoS) between 1 and 4 for the plot because higher values are not representatives and mostly due to errors or specific situations in the data set (outliers).
The degree of saturation is a useful indicator as it helps us determine which sections are the most congested and to what extends, we will decide which sections to focus on based on this indicator.

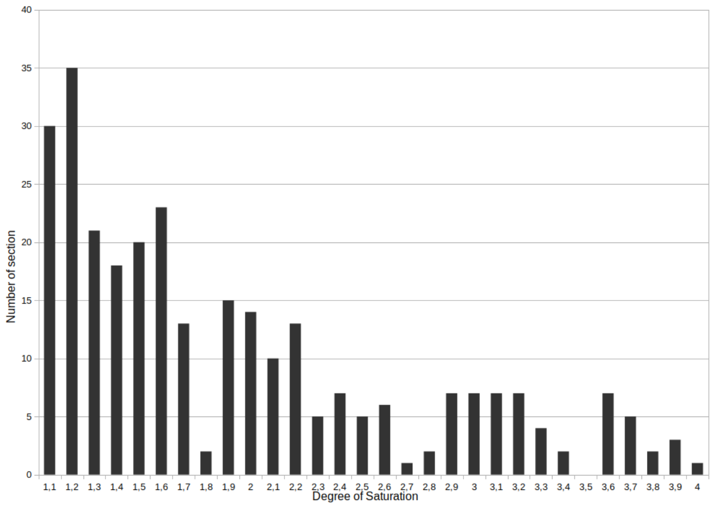


Figure 8: Number of section per Degree of Saturation

3.1.4 Travel Time

In order to get a representative overview we want to analyze one last factor: the travel time. The goal is just to get a global overview of the system. To do so we will compute the travel time for each section in the best case scenario, when the flux is less than or equal to the MFV. We can use the following formula to compute the minimum travel time over a section in the best case scenario:

$$Time = \frac{Distance}{Speed} \tag{27}$$

By using this formula we can calculate the travel time in minutes for each section. To compare it with the actual situation we have to estimate the travel time in real conditions, using the supplied data. Considering that we know the distance of each section, the speed limit for the section, the traffic flow (real flux) and the capacity (MFV) we have to model the travel time for sections with those parameters. Different solutions are proposed in the book Transportation Decision Making: Principles of Project Evaluation and Programming [5] in chapter 5 : travel time impact. The 3 mains solutions are the HCM method, COSMIS method and BPR method. HCM and COSMIS look at the average speed, where BPR method directly gives us the travel time. Finally we found one last method: the *Underwood speed estimation model* which computes the effective speed depending on the free flow speed, the actual density and the optimal density. Unfortunately since HCM and COSMIS methods are not applicable in our circumstances (cf. [8]). It leaves us two options, BPR method or the Underwood model.

Underwood model to estimate speed :

$$K = K_c \cdot \ln \frac{v}{v_{est}} \tag{28}$$

$$v_{est} = v \cdot \exp\left(\frac{-K}{K_c}\right) \tag{29}$$

where K is the density, K_c the density at maximum flow, and v the free flow speed

BPR Method to estimate travel time:

$$T_{travel} = T_0[1 + \alpha \cdot (\frac{trafficflux}{capacity})^n] \tag{30}$$

$$= T_0[1 + \alpha \cdot (\frac{RealFlux}{MVF})^n] \tag{31}$$

$$= T_0[1 + \alpha \cdot (DoS)^n] \tag{32}$$

where T_0 is the free flow travel time, α and n are two constants (respectively equal to 0.15 and 4 in , cf. [8] in chapter 3 : Current Speed Estimation Techniques)

Since we had more resources for the BPR method, we decided to use it to compute the actual travel time for each section and compare it with the travel time in the best case scenario. In order to visualize how much time is wasted in the traffic jam and on how many sections, we computed the ratio between the travel time compute with the BPR method and in best case scenario. We computed this ratio only for sections where the MVF was less than the real flux, otherwise we forced the value to 1. When the ratio is equal to 1 it means the actual travel time is the same than the theoretical one in perfect conditions. The higher the ratio, the more travel time is increased. For example a ratio of 1.5 means you take 50 % more time to travel a through a section, compared to the best case scenario. As for the DoS, we plotted the distribution. We choose, again, to only consider value between a certain range ([1,5] for this distribution) as some values are not representatives and due to errors. Moreover, those errors occur where errors also appeared previously when we verified the real flux after crossroads. Therefore we can assume it is also due to specific situation (exit ramps, highways interconnection, multiples exits etc.)

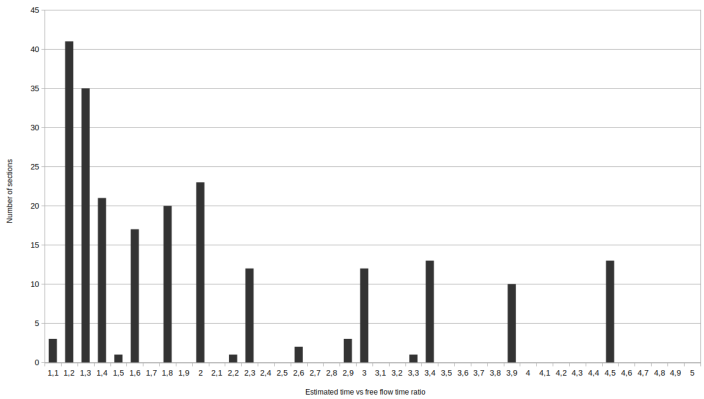


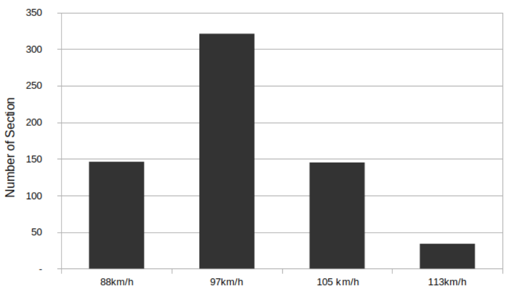
Figure 9: Number of section per Travel Time Ratio (round to 1 decimal)

3.2 Lane Speed combination analysis

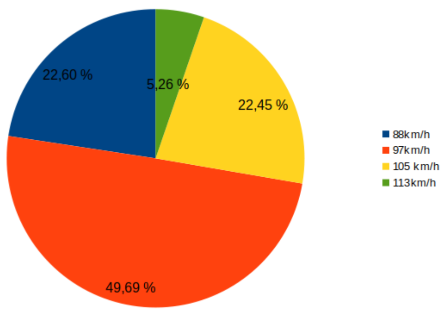
3.2.1 Lanes & Speeds Statistics based on our model

In this part we will group highway's sections by speed limits and number of lanes and perform some analysis related to the previous sections (travel time ratio, degree of saturation, etc). The main goal of this analysis is to obtain statistics for each type of sections in order to give us insights on which section we should focus on.

First of all we group each section by its speed limits. We then obtain the following charts (Figure : 10), providing the number of sections by speed limits and the distribution.



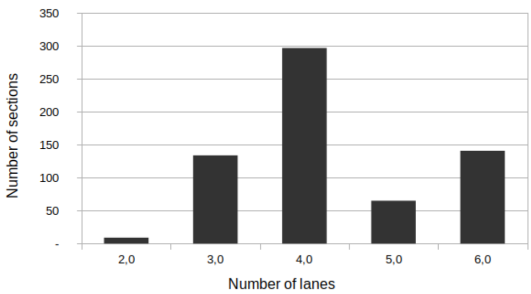
(a) **Number of section by speed limits**



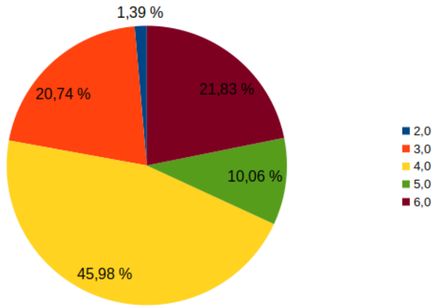
(b) **Distribution of the section in function of speed limits**

Figure 10: Statistics on the number of section for each speed limits

Secondly we can perform the same analysis, but this time we group sections by their number of lanes. We also got 2 charts (Figure : 11), giving us the number of sections and the distribution in function of the number of lanes.



(a) **Number of section by number of lanes**



(b) **Distribution of the section in function of number of lanes**

Figure 11: Statistics on the number of section for each number of lanes possible

We will now merge those criteria and group the sections by speed and limits in order to obtain the number of sections for each configuration and so a distribution. For visualization reason we decided to only plot the number of sections for each configuration (figure 12) but not the distribution. The latter is represented using a double entry table (table 5).

Lanes / Speed	2	3	4	5	6
88 km/h	0.46 %	5.88 %	11.76%	3.10%	1.39%
96 km/h	0.93 %	12.07%	28.64%	6.97%	1.08%
105 km/h	0.00%	0.31%	2.79%	0.0%	19.35%
113 km/h	0.00%	2.48%	2.79%	0.00%	0.00%

Table 5: Distribution of section in function of speed limit and number of lanes

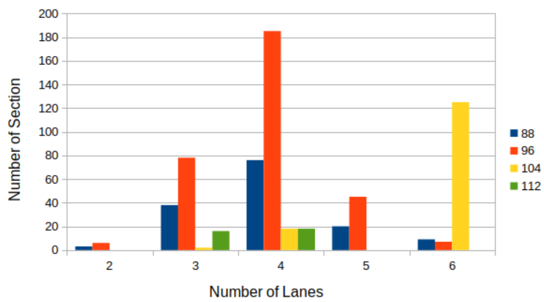


Figure 12: Number of section in function of the speed limit and and number of lanes

Using this distribution we now compute the number of sections for each configuration where the degree of saturation is greater than 1. In other words, the number of sections for each configuration where a traffic jam could occur. Once again we only plot the number of sections but not the distribution, which we decided to let under the form of a table.

Speed \ Lanes	2	3	4	5	6
88 km/h	66.67 %	50.00 %	56.58%	55.00%	33.33%
96 km/h	50.00%	73.08%	68.11%	42.22%	85.71%
105 km/h	0.00%	50.00%	27.78%	0.0%	28.80%
113 km/h	0.00%	37.50%	55.56%	0.00%	0.00%

Table 6: Distribution of section where the DoS is greater than 1 for each configuration

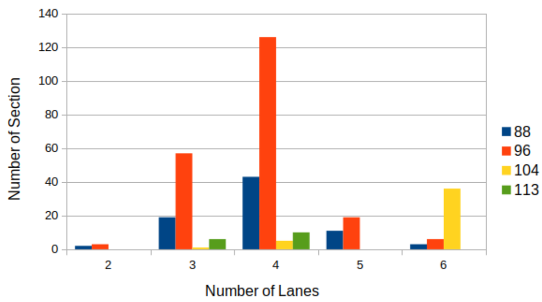


Figure 13: Number of section where the DoS is greater than 1 for each configuration

Finally, now that we have the proportion of sections subject to traffic jam for each configuration, we will compute the average *Travel Time Ratio (TTR)*, described previously in section 3.1.4, for each configuration (cf. Figure 14). When there is no potential delay on the travel time, we force the value to be 0 while we limit this value to 10 to avoid outliers which can biased our results

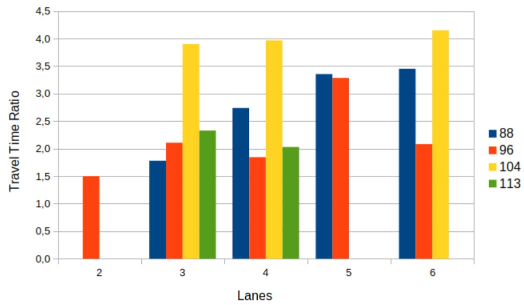


Figure 14: Average Travel Time Ratio for each configuration possible

Additionally, we decided to plot in the form of violin plot the distribution of the TTR for the 8 most represented combinations (cf. table 3). In order to have a better visualisation of the distribution of TTR we deleted all values where the ratio was higher than 10 (conversely to the figure 14 where we forced greater value than 10 to be 10). This allow us to see with more details the distribution of the TTR for each combination .This is what we obtained :

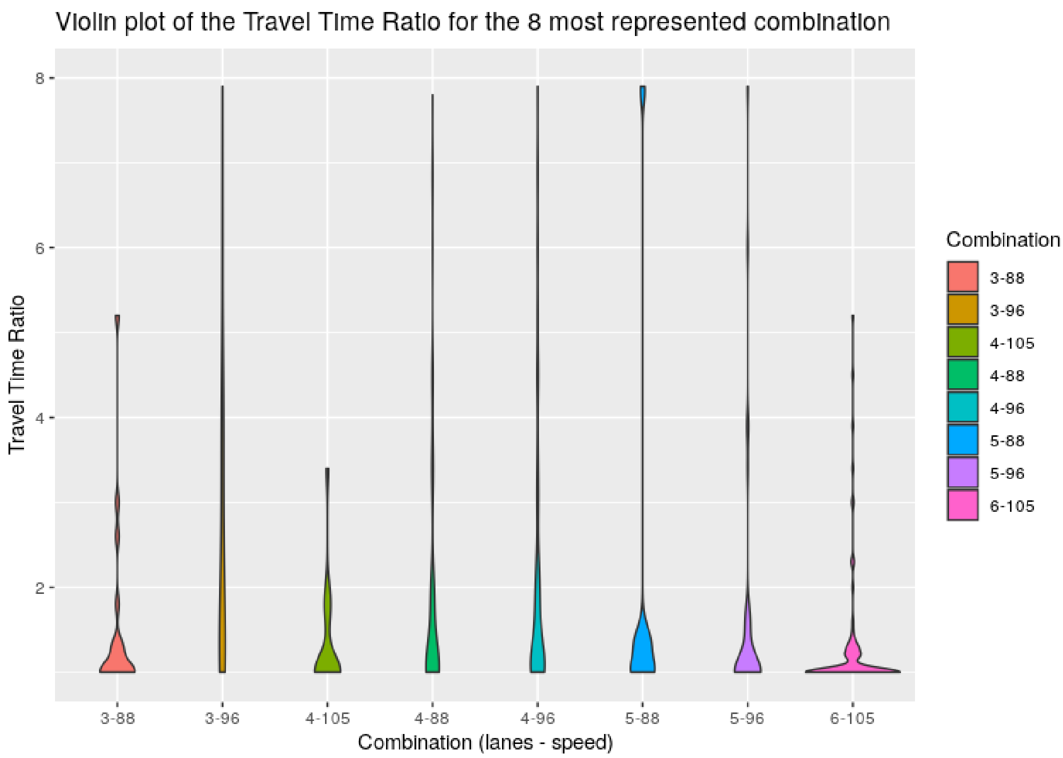


Figure 15: Violin Plot of the Travel Time Ratio for the 8 most represented combination

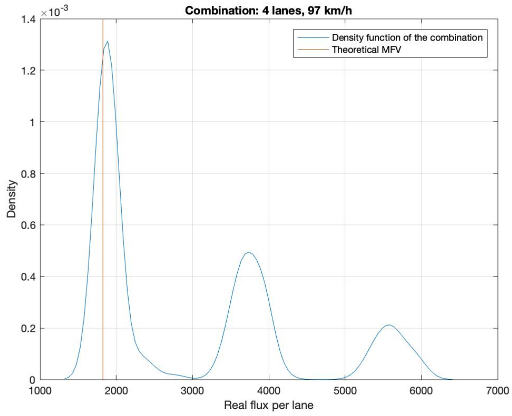
Observation: By looking closely at figure 15 we can observe some interesting things. First of all we observe that for almost all combinations, the TTR is lower than 2. Let's keep in mind that a TRR of 2 means you take twice more time to cross a section than in the best case scenario (driving at max speed without congestion). This is quite coherent as, except in highly congested section, you usually drive at a crusing speed slightly under the speed limit and even when there are traffic jams, most of them do not multiply you travel time by more than a factor 2. However we still observe that the TTR factor can still be high (4 combination have an max TTR of 8). If we consider section with TTR lower than 2 as section where congestion is still reasonable and does not impact that much the user experience on the Houston highways system; section with high TTR can be considered as section where the impact on the user experience is important. Moreover we can observe some 'bubbles' on each plot, especially the 3-88, 5-96 and 6-105 one. It let us think that the distribution of the TTR is maybe not continuous but discrete. We will observe similar phenomenon later when observing the distribution of the real flux of each combination. Several suggestion can be done on this observation but as it is a major part of the analysis we will discuss it in the section *'Complementary Analysis on the Real Flux'*

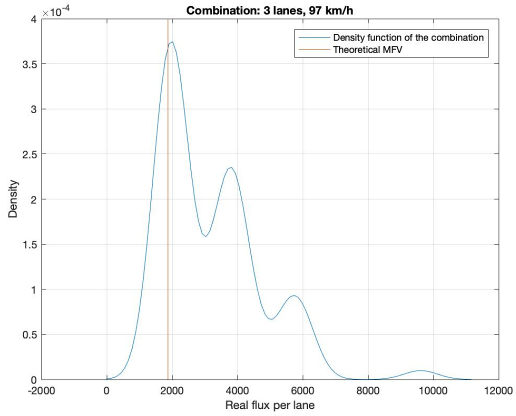
3.2.2 Lanes & Speeds Statistics based on provided data

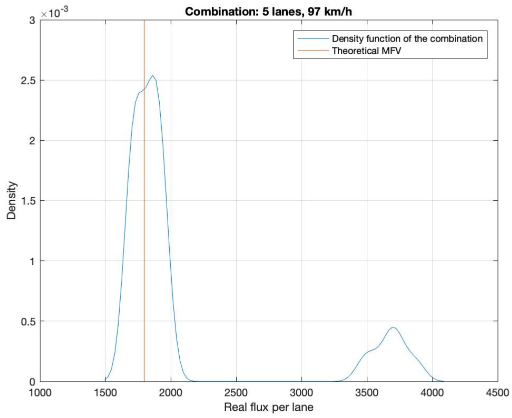
On the previous section we did a statistical analysis mainly based on our model (MVF, DoS, Travel Time Ratio, etc.). We now need to focus on the real data and perform an analysis mainly based on the provided data only. We previously demonstrated that all configurations (lane and speed) are not equally represented. Based on this, and in order to avoid biased results we decided to focus our data-analysis on the 8 most prominent combinations (cf. 3) :

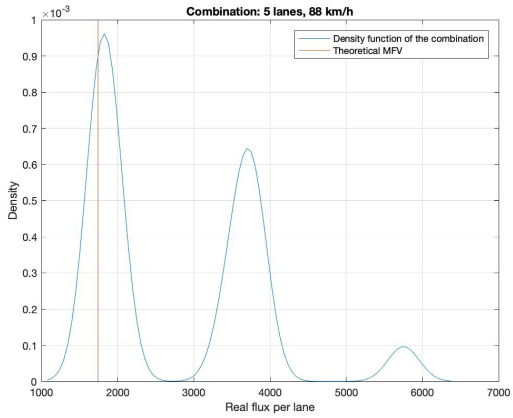
- 4 lanes, speed limit set to 96km/h
- 6 lanes, speed limit set to 105km/h
- 3 lanes, speed limit set to 96km/h
- 4 lanes, speed limit set to 88km/h
- 5 lanes, speed limit set to 96km/h
- 3 lanes, speed limit set to 88km/h
- 5 lanes, speed limit set to 88km/h
- 4 lanes, speed limit set to 105km/h

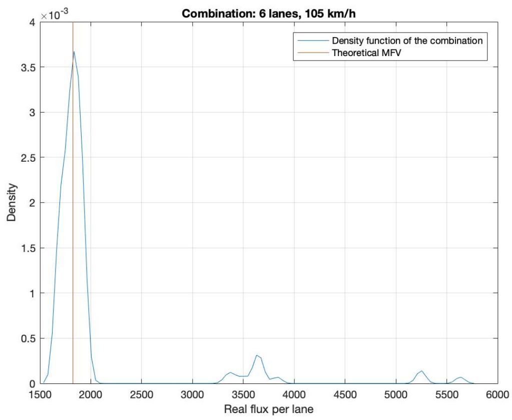
More precisely we will base our analysis on the main factor provided by the data set : the Real Flux. The idea is to observe the distribution of the real flux for each configuration. From this, we would be able to locate the theoretical MFV threshold of our model on the distribution and establish some critics on it. Moreover, this kind of analysis, will give us hints for the section's selection of phase 3 and also help us investigate some possible errors in the data. In the plots that follow, we analyze the real fluxes of each of the 8 configurations through their density functions; the flux is expressed as 'flux per lane' to better help us compare the results (since each configuration has different number of lanes), while the vertical red line represents in each case the theoretical MFV computed previously with our model (and reported in table 9).

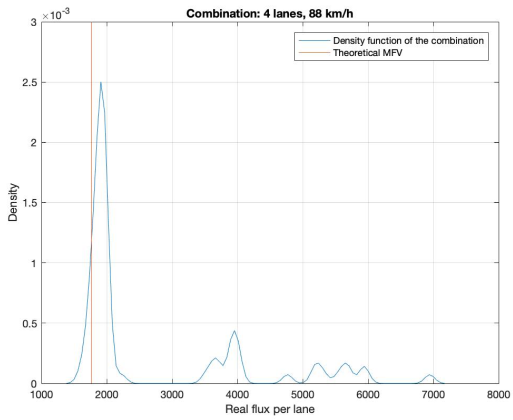












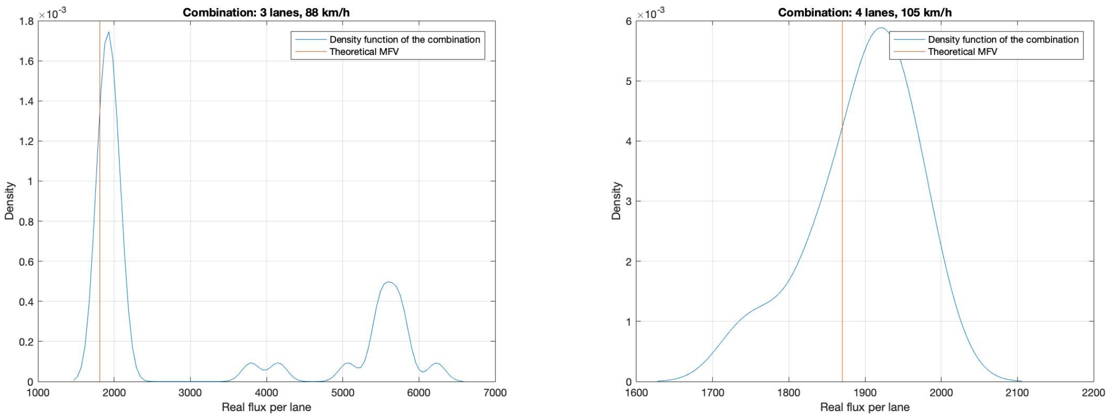


Figure 16: Density Function with MVF threshold for the 8 combinations

Looking at figure (16) we will first observe that the density functions of the Real Flux recall the density function of a Normal distribution. What's mainly different is the presence of additional "bells", meaning that there is not just one flux interval that characterizes the combination but more something like a 'major' flux interval in which the majority of cases fits in, and some others windows with higher fluxes that still represent a noticeable portion of cases. In particular, we notice how for most of the cases the main bell is in a window of values that are around 2000 veh/h per lane, while the secondary bells appear in the interval [3000   6000] (more precisely for values of flux per lane around 3000-4000 veh/h and 5000-6000veh/h).
Observation: this already gives some interesting information, and might be a sign of the fact that the presence of higher fluxes and congestion are not due to the particular configuration lane-speed limit, but are instead related to the interested part of highway (we will return on this later).

What could these additional bells represent? From our perspective these secondary bells could represent either sections where traffic jam appears, or errors in the data - for most of the cases, like for example in combination 5 lanes- 96Km/h or combination 6 lanes- 88Km/h, in which the secondary bells are more pronounced and representative in terms of density. The first guess would be the leading idea; in other cases,like for configuration 6 lanes-105km/h or 4 lanes-88km/h, where the other bells are quite 'low' in term of density, compared to the main one, we may consider the idea of the presence of some errors in the data - after a proper investigation on the interested sections-.

From these observations, we can then hypothesize that the first bell represent the Real Flux when there is low congestion or no congestion at all, while the additional bells stand for those fluxes where congestion is present and high. We still have to be careful about the main bell: what we assume does not mean that all real fluxes under those bells are fluxes without traffic jam. In section 'Critique of the Model' we will analyze this aspect in a more accurate way.

Observation: Looked from another point of view, the presence of more than a 'main bell' seems to be reasonable since apart from the physical structure and characteristics of an highway system (such as the number of lanes and the speed limit), the flux of vehicles in a section also depends on the geographical position of it with respect to the highway network (which also takes into account the presence of crossroads and their length (*)). We will encounter this again, later in this section (see observation 2 about figure 19)

(*)Based on the observation above we decided to take into consideration also the length of the highway sections and divide the sections in each combination in two groups (sections with length $\leq 1km$ and $> 1km$) and look at their distribution (see figure 17). This subdivision is mainly due to the fact that exits/entries/crossroads..., due to the lower travel speed that vehicles should respect/may encounter, can be more congested than long stretches of straight road; influencing in the wrong way our analysis.

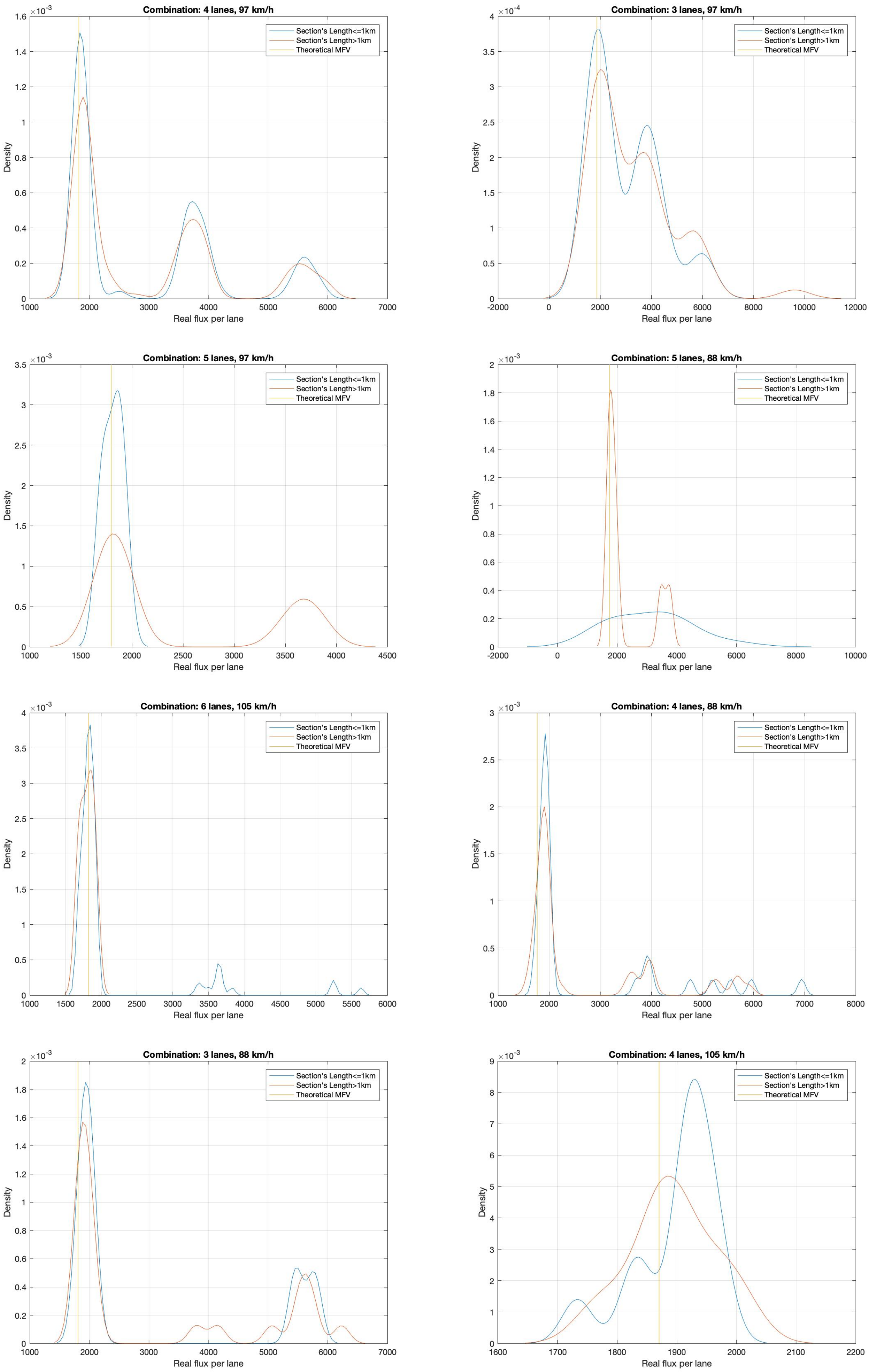


Figure 17: Density Function with MVF threshold for the 8 combinations

From figure 17 we can notice how in the majority of the cases the two configurations have similar distributions, there are anyway some exceptions, from which we can make some observations:

- Configuration 5 lanes - 96km/h : the second bell is due only by sections of length $> 1km$
- Configuration 5 lanes - 88km/h : we observe totally different behaviors between the sections of length $\leq 1km$ and those of length $> 1km$. While in the first case we observe approximately a constant distribution, in the second case we visibly notice two separate bells, a first one, with higher density, and a smaller one. Moreover both the two orange bells (sections' length $> 1km$) are inside the range of values covered by the blue bell (sections' length $\leq 1km$).
- Configuration 6 lanes - 105km/h : secondary bells are due to sections of length $\leq 1km$ (but are of very small density).
- Configuration 4 lanes - 105km/h : the two cases cover the same range of values but differently.

Observation: the number of sections we are dealing with in each configuration is totally different, we have configurations with almost 200 sections, and others with less than 20; this implies that for some of them some values and results weight way more than they should or than they would in a different scenario. We will keep this in mind for a later section and look at other representations that would better help us weight in the proper way this results (see figure 20).
For now we can say that the fact that in the majority of the configurations the two cases have similar distributions, and the fact that those configurations in which the distributions differs are those for which we have less amount of data, might suggest the fact that the section length isn't that significant for the point of view of the flow of vehicles through it.

About the MFV threshold: Looking at the MFV we can notice how the average MFV related to each possible combination is located not on the "most-right-bell", which would imply an high capacity of the road to 'handle' high fluxes (those due to the presence of these secondary bells, are present and do happen); but around the "peak-density" flux. This might be one of the reasons for which our percentage of traffic jam, detected with our mathematical model, is quite high.

Another data, that would help us better understanding the capacity of the highway system related to the demand, is the maximum flux registered for each of the eight combinations we have focused our attention on. Since the eight configurations have different number of lanes, in figure 18 we plotted both the maximal total flux registered for each configuration and the maximal flux per lane, to better visualize and 'compare' the fluxes of the cases in analysis.

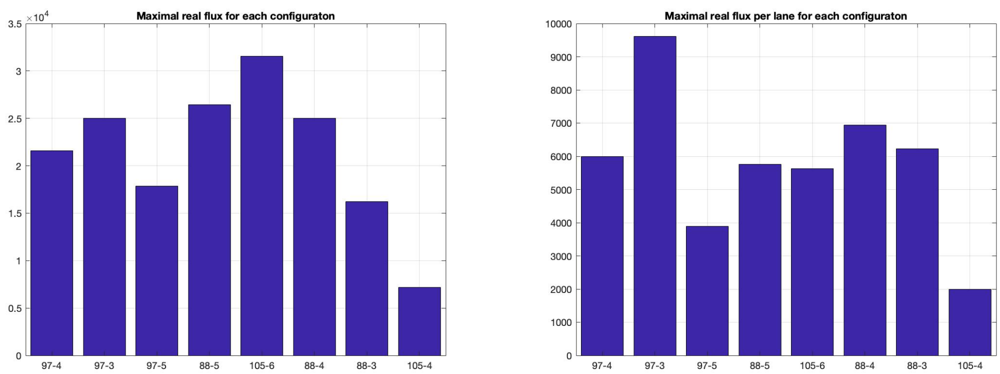


Figure 18: Maximal flux

Looking at the histogram on the left side we can notice how the configuration *105Km/h-6lanes* is the one that had registered the maximum flux, followed by the configurations *88Km/h-5lanes, 97Km/h-3lanes, 88Km/h-4lanes.*

The fact that the configuration *105Km/h-6lanes* is the one with higher real flux registered seems reasonable taking into consideration the speed and the number of lanes that characterize it. Moreover we

should recall that this configuration is the second most frequently used in this highway system (19.35%- see table 5) which makes this result more even consistent. On the other hand the three following results are instead quite interesting and if we look at table 7 we can notice how those sections (in which a peak flux for the specific combination was registered) are not regarding particular crossroad (small section length) for which a peak would have been explained but actual straight roads that would then might be taken into consideration in our selection.
Looking now at the second plot, we then realize that actually most of the configurations have quite similar peak values -in terms of actual flux per lane-, while the case *97Km/h-3lanes* stands out. This again may be taken into consideration for our selection of sections.

Moreover if we look at the sections to which correspond the peak values of each configuration we can notice that all sections have different length; so only reasoning in terms of peak values, there is not a "trend" which correlates traffic with the length of the section in analysis (and this supports our previous guess on the significance of the section's length in our analysis). It also highlights again the importance of the geographical location of a section respect to the road network, and checking the correspondence of those values on the map, it confirms the geographical analysis done before.

Section with max flux for each configuration				
Config.	Highway	Crossroad	Length (Km)	Real Flux
97-4	10 Westbound	Lockwood Drive	0.3	21600
	69 Westbound	10 Eastbound	0.2	21600
97-3	69 Westbound	45 Northbound	1	25000
97-5	69 Eastbound	Edloe Street,Buffalo Speedway	0.8	17880
88-5	10 Westbound	45 Southbound	2.2	26460
105-6	8 CC Northwest	Gessner Road, Fairbanks North Houston Road	3.2	31536
88-4	10 Westbound	Washington Avenue, Westcott Street	2.3	25000
88-3	45 Northbound	Spur 5 south-University of Houston	0.5	16200
105-4	8 CC Southeast	Beamer Road, Sabo Road, Hughes Road	0.6	7200

Table 7: Maximum flux for each combination

Observation : while for some configurations the real flux is homogeneously focused on similar values for all the sections that belongs to it, other configurations present some sections with registered fluxes that are far from the general trend of the combination it self (as we can notice in figure 19). This confirms that speed-limit and number of lanes do play a role in the traffic flow but aren't the only components that should be considered (we will come back on that later).

In figure 19 we can better understand the observations made above. In particular the graphs represent the real flux of each section belonging to that particular lane/speed configuration. The sections are ordered with respect to the excel file order of the data. Keeping in mind that some configurations have different number of lanes than others, in order to better visualize the situation and to make comparisons more easy and accurate, all the graphs deal with the registered reals fluxes/lane. Moreover, based on the considerations made before about the possible influence of the length's section on the data, we decided to have a closer look on the sections and plotted each flux of each section (in the proper configuration) for the two cases -see figure 20.

General consideration: Looking at the trends of each configuration we can notice how some sections have fluxes 2 to 4 times higher than the main trend characterizing the majority of the sections in that configuration. This is an interesting data and possible indicator for the selection of the most congested sections. Moreover, in most cases, these behaviours do not concern a single section but groups of them adjacent to each other.

Based on these observations we decided to detect these "outstanding" sections and find a possible explanation/correlation of these behaviours. More precisely, after a deep "configurations analysis", we will see how the analysis of those "outstanding" fluxes would lead to a selection of highly congested sections, that grouped, will highlight some particular highway's portions (see section 2.2.2.8 and figure 32 on the Annex).

About the section's length: form figure 20 and the considerations above we can deduce that the length of the section isn't significant for our analysis, so we can keep considering them all together (without any distinction), without biasing our analysis.

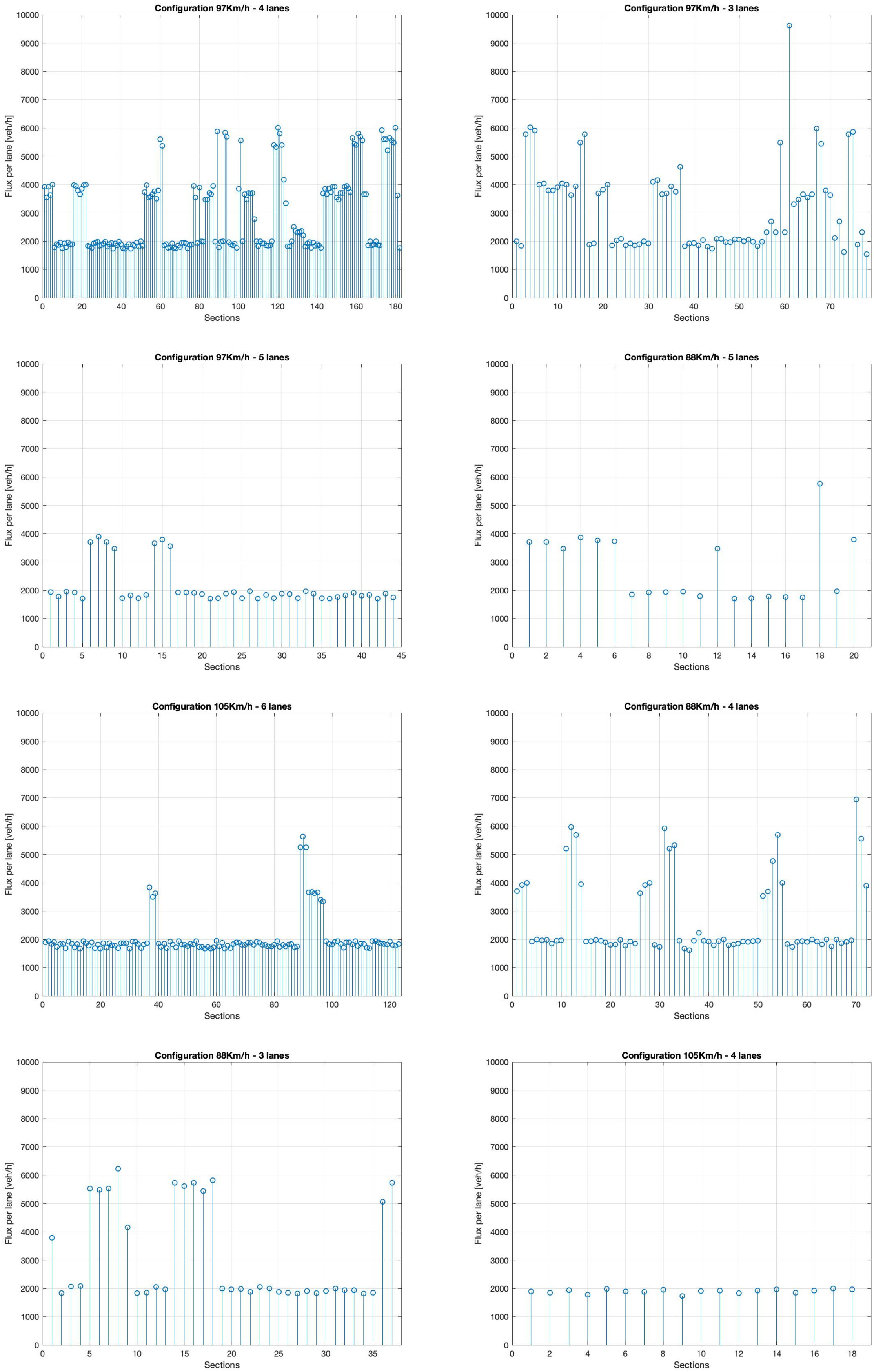


Figure 19: Fluxes of each section for each combination

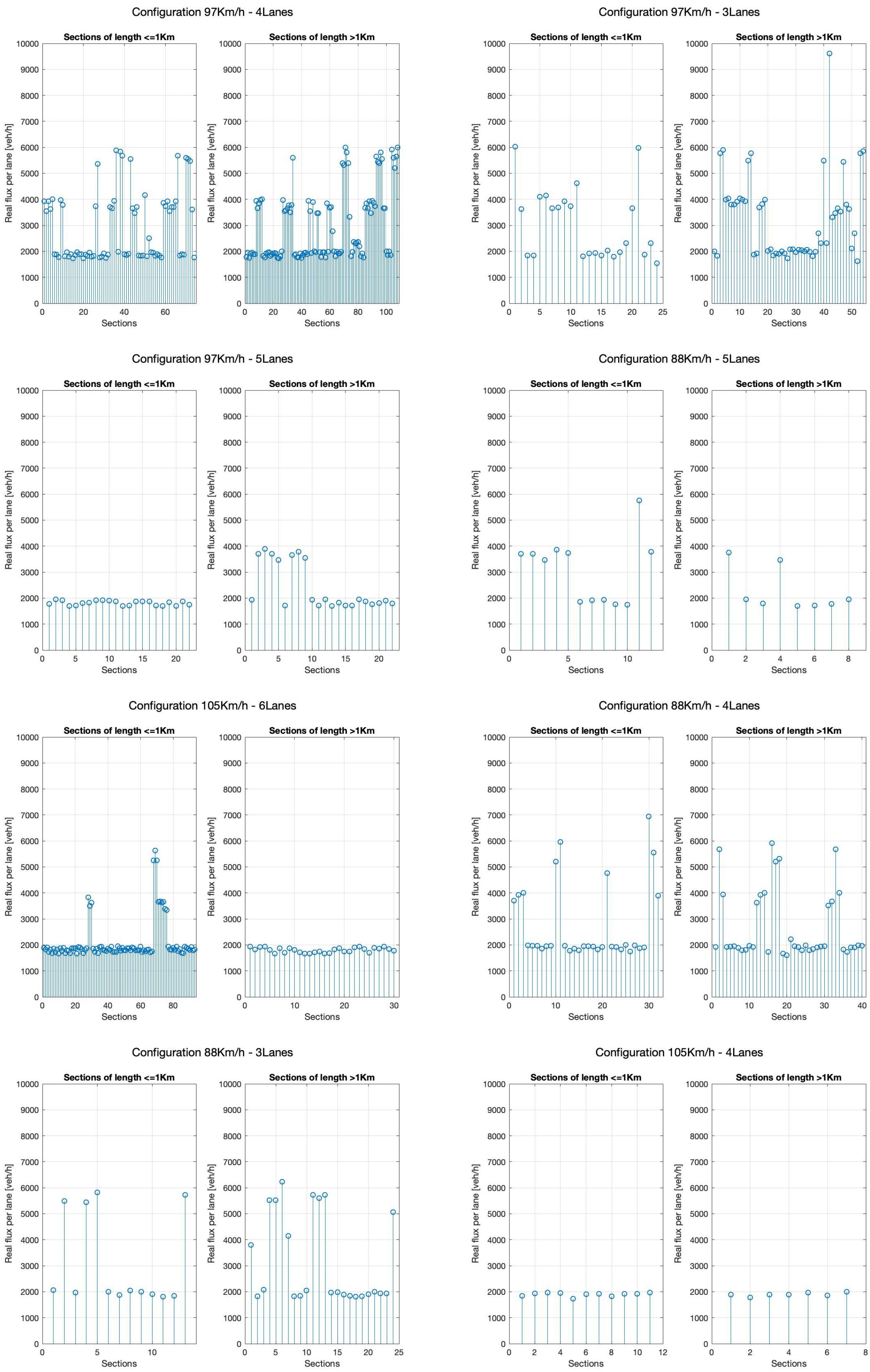


Figure 20: Fluxes in each combination with sections divided by their length into two cases

3.3 Critique and Additional Analysis

3.3.1 Critique of the Model

In this section we are going to emit some critic on our model based on the analysis done previously. Even if the values found for the MFV are coherent, the model still have some weakness. From the data overview we can see that the major problem is that our model is quite 'severe'. Using our MFV model we determine that around 59% of the section were subject to traffic jam. But this is mainly due to the definition of MFV we use, which indicate us the maximal flux such that you can travel at the maximal speed. Even if it does not tell us at which point the section is congested, this is still a high result. Moreover from figure 16 we observe that the MFV is each time too much on the left of the main bell. If we compare our distribution with a normal distribution, the MFV should be located between 0 and 1σ (cf. figure 21) on the x-axis in order to have 45% congested section. Even if we know Houston is a busy city, we can imagine that 45% of section subjects to traffic jam is more realistic than 59%.

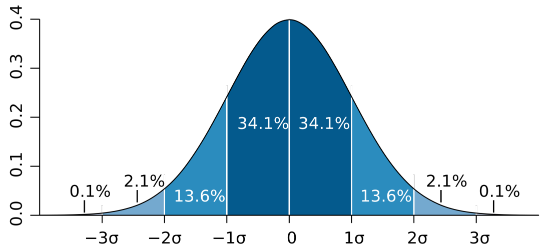


Figure 21: Standard deviation representation for a Normal Distribution

Different reasons could have lead to this result. For example we know that in practice the safety distance is usually not (strictly) respected. Then the critical density becomes higher, and the maximal flux could be augmented. In our model we used strict safety distance avoiding risk of accident in case of emergency breaking. For example, changing this value by adding a random variable for each vehicle tells us the probability that the user respect (strictly) or not the safety distance could have lead to a higher MFV. We can also contest the distribution of cars, buses and trucks on the highways system around Houston. We based our distribution using daily data not specially designed for Houston where the data set gives us the real flux for 1hour only (8a.m to 9a.m) on a specific day and for Houston. By reducing the number of trucks on the system we could reduce the safety distance between the vehicle and the density. Then it would also have increased the MFV. Those are only 2 examples, but you can find in [5] a lot of different models uses to compute the MVF using different inputs - and all have their advantages and flaws depending on the situation you want to use them. Another argument that can be used to critic our model is the fact that we use the length of section. We use it to compute the DoS, and also the Travel Time Ratio. In order to determine if our model is unbalanced between small and large section when deciding if yes or not the section is subject to traffic jam we can plot the 2 distribution. The first one is the density function of the section length for all sections. The second one is the density function of the section length for all section subject to traffic jam according to our model. Those functions can be found on figure 22.

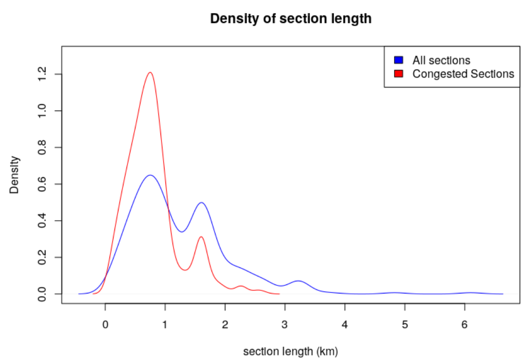


Figure 22: Density function for section length (all vs congested only)

Observation: From figure 22 we can observe several things which can lead to different conclusions. First of all we see that the results obtained by our model (red line) are coherent with the results obtain from the provided data set. In fact we can observe that for each density function, we have 2 bells situated at the same x-coord (800m and 1,8 km section length). But the more interesting observation is on the difference of heights between those bells. For the first bell (800m), the density is 2 times higher for congested sections than for all the other sections. Conversely, for the second bell (1.8km) the density is 2 times lower for congested section than for all section. It shows a big difference between small and big sections in term of probability to be congested. But why ?

Suggestion: There are 2 main explanations for this observation. One of them shakes our model when the other does not. In fact we know that our model take into account the length of the section when computing the MFV, DoS and TTR. As we defined a section subjects to congestion as section with a DoS greater than 1, we can think that the formula is biased and 'penalized' small section more than big one. But we need to find another logical reason to this phenomenon. This lead us to the second explanation. We know that small sections are more likely to be intersections, or in-between sections (for example between 2 sections of high length we can have a small section with a different amount of lanes and smaller speed limit). Those type of sections are by definition more subject to traffic jam. It can be due to a deceleration effect, a change of direction, lanes reduction and a lot of other reasons. Then we should not worry too much about this observation as it is more probably due to natural phenomenon than a lack of precision on our model. Moreover we have seen previously that the length of the section do not have a deterministic effect.

Finally, we have to remember what the the MFV represents in our situation. It is the maximal flux of vehicle at which every drivers can drive at a cruising speed equal to the speed limit of the section. This is the 'best case scenario'. But if a section has a real flux slightly higher than our theoretical MFV for a given combination of speed and lanes, it does not necessarily mean that the section is highly congested and that a traffic jam occurs. This is why we performed a deep analysis on the data. We will continue this analysis in order to find the most qualitative and representative sections.

In order to select a qualitative and representative set of sections to focus on for the final analysis and find an adequate solution, we will take into account the weakness of our model and use the data analysis previously done combined with a geographical analysis and a statistical analysis. But we still have to clarify some observations seen previously, especially on density functions of the real flux for the most represented combination.

3.3.2 Complementary Analysis on the Real Flux

From the section *'Lanes & Speeds Statistics based on provided data'*, we revealed some interesting phenomenon on the distribution of the Real Flux. Our main observation was the 'repetitive' bell-schema present on almost each distribution of the 8 combinations. On figure 16 we clearly observe the 'multiple bells' phenomenon, described on precedent sections. In this sub-paragraph we will analyze in a deeper way this phenomenon, our goal is to emit some hypotheses on the presence and possible meaning, of these secondary.

In order to perform a deeper analysis and explain the results we have previously found, we decided to analyze the Real Flux for each combination using violin plots. A violin plot is a mix between a bar chart and a distribution function plot, and it's particularly useful when the data distribution is multimodal as it is in our case (more than one peak/bump). It will show us the presence of different peaks, their position and relative amplitude; allowing us to emit further hypotheses that may confirm or not our previous guesses.

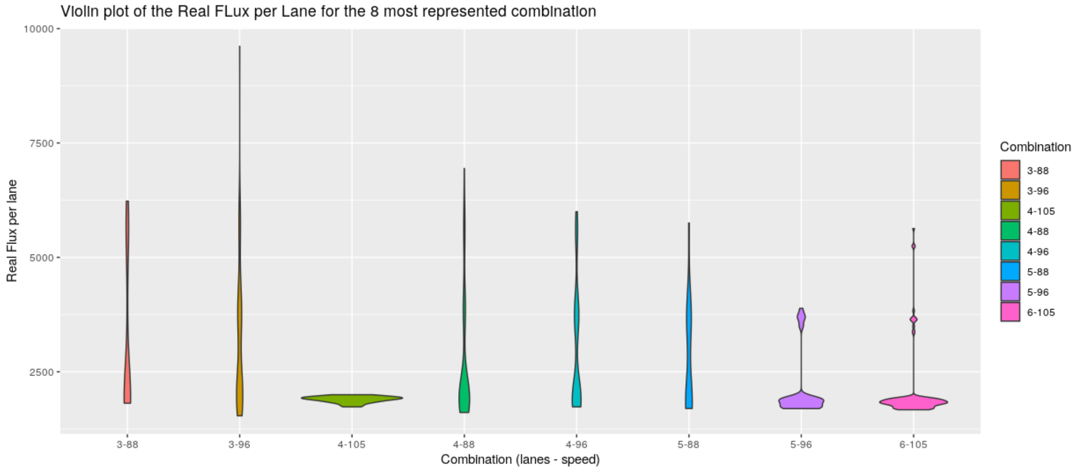


Figure 23: Violin representation of the real flux by lane for the main 8 combination

In figure 23 we have a first general view of the distribution of the 8 most frequent combinations. We can notice how, apart from 2 configurations, which are 3-97 and 4-88, the real flux per lane covers values up to 6000 veh/h. Moreover, if we take a look not only at the distributions of these two configurations (from which we can notice that the density for those outliers is really low and negligible), but also at figure 19, we can notice how actually the cases of real flux per lane $> 6000 veh/h$ are represented by only one section on both configurations (suggesting a possible error in the data). Based on this observation, on figure 24 we decided to neglect those outliers and focus on the main values of interest for the real flux.

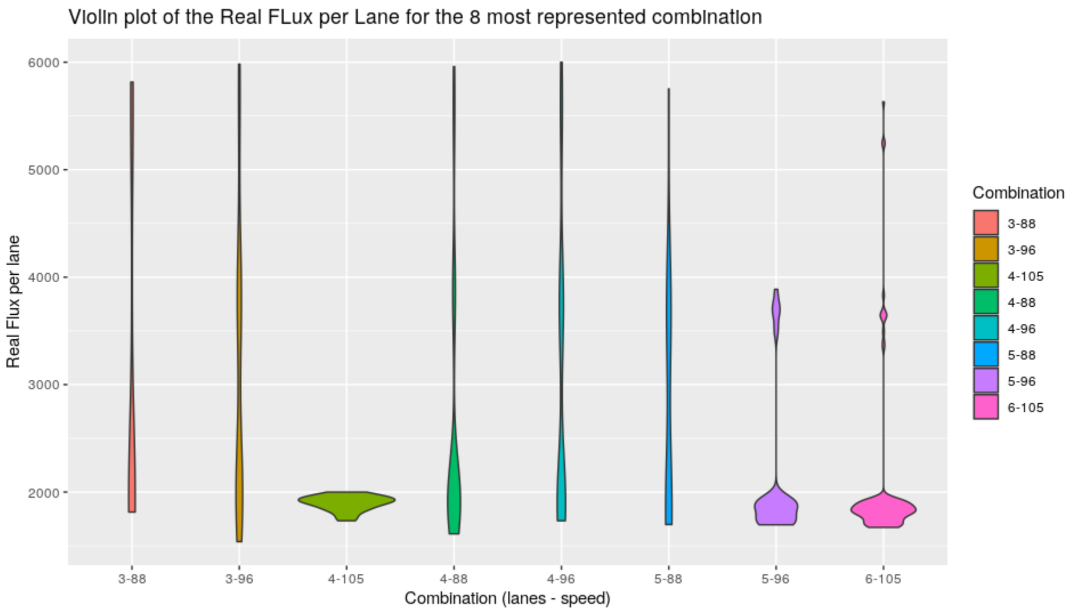


Figure 24: Violin representation of the real flux by lane for the main 8 combination, with a 6000 flux/lane limit

The first thing we can notice looking at figure 24 is that we can distinguish two main "trends": one that group in with the configurations 3-88, 3-97, 4-88, 4-97, 5-88; and one that group in with the configurations 4-105, 5-97, 6-105. While the first group covers "all" the range of values in between 1500 veh/h and 6000 veh/h which reflect the presence of those secondary bells with noticeable amplitude; the second group ha density which is mostly all concentrated in a smaller and lower interval, which correspond to the interval of values for which the other cases of the first group present a the main bell. From a first analysis we can then say that the main bell of each configuration is mostly in the same range of values for all the 8 cases.

In order to better analyse the multi-modal behaviour investigated in the previous section, on figure 25 we plotted the two cases in separate graphs.

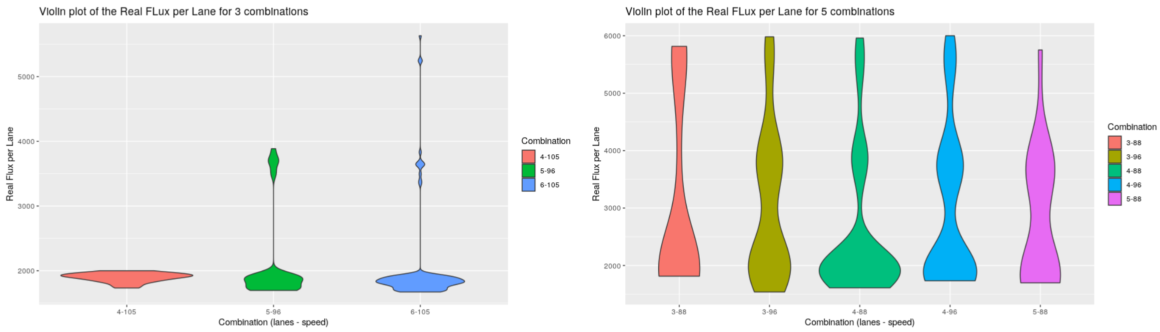


Figure 25: Violin representation of the real flux by lane for the main 8 combination

On the right side of figure 25 we can observe the density distribution of the configurations: 4-105, 5-97, 6-105. These three configurations have in common that they are basically characterized by a unique interval of values for the real flux per lane which is around 2000veh/h. While the configuration 4-105, has all its sections within that interval, configurations 5-97 and 6-105 also present come sections that registered higher values (around 4000veh/h in both the cases and also around 6000veh/h for configuration 6-105); but these outliers (which are 2 and 4 times higher in values respect to the general trend), as we can notice from figures 19 and 16, represent only a small percentage of the sections in analysis (4 over 44 sections, which is a 15% for the configuration 5-97; 9 over 122 and 3 over 122 sections for the configuration 6-105, which is the 7% and 2% of it). We can then conclude that, configurations 4-105, 5-97, 6-105 are not interesting situations of big congestion.

Looking now at the left side of figure 25 we observe the density distribution of the configurations: 3-88, 3-96, 4-88, 4-96, 5-88. We can at first notice how these distributions differ form the previous three: the presence of more "bumps" that stand for a multi-modal behaviour. In particular we can make the following observations:

- there are 3 main modes: the first one, around 1500-2000 veh/h, the second one around 3000-4000 veh/h and the third one around 5000-6000 veh/h
- the first mode interests all the configurations and moreover is the one which registers the highest density (so the biggest amplitude) in all the 8 combinations: it represents the main bell and it has the same interval of values in all the cases
- the configuration 3-88 is the only one that doesn't have any bump for values around 3000-4000 veh/h (only 2 sections over a total of 37 are around those values,while the 32% of the sections are in the upper extremes).
- configuration 5-88 presents a high amplitude (density pretty close to the density of the main bell) around the values 3000-4000, which is consistent with figure. Moreover we can notice how that second bump is represented by 40% of the sections (8 over a total of 20) and how there's only one section that reaches the value of 6000 veh/h.

Apart form these more focused consideration, looking at these last 5 cases we can highlight the presence of fluxes higher than those characterizing the main trend (which was basically the only trend of the first 3 cases on the left hand side of figure 25) and that happens to be all in the same intervals of values. This can suggest the fact that the configurations 3-88, 3-96, 4-88, 4-96 and 5-88, are those more likely to be interested -with respect to the data provided- in congestion.

What can we say then, about our data? We can interpret these "multi-modal" behaviours thinking of car flux and traffic as discrete events that change rapidly from "no jam" to "traffic jam" due to some "stressful" situations. If we think of flux as a discrete instead of a continuous concept, then the main "bells" of our distributions can be interpreted as "pre-jam"/"regular" situations, where the traffic is busy but still flowing. In this prospective, then the secondary bells would instead highlight those sections in which intense congestion does occur (we say intense since, if we look at the previous plots we can notice how the "congested sections" are actually often twice as high, in terms of Real Flux values, as those representing the main bell- and half the size of the distribution). The idea would be then to focus on these secondary bells, as the "most congested sections".

Consideration: Looking at traffic as a discrete variable would also help us better understand/interpret some behaviours we have observed so far. If traffic begins to be way busy than the regular -a situation of "stress" occurs- we would encounter some parameters changes, like the safety distance (drivers would drive closer) or the travel speed (that would decrease), some entries and/or exits would be way more congested than expected, and some roads would be interest in an increase of vehicle flux due to traffic convey/delay.

We analyzed the sections that are in the "second bells", and noticed how they were all related to some specific and adjacent parts of the highway network; their geographical location can be seen in figure 32 on the annex. In particular we can notice how these sections happens to be located in some of the areas we previously identified - through our analyses - as most likely to be congested (figure 31), showing a consistency in our three different analysis (and in particular with the statistical one).

4 Selection & Solutions

4.1 Highways/section selection

Now that we have established an overview of the data-set on different factors, we will use what we have done so far in order to select relevant sections to perform a deeper analysis and based our solution on those sections. We want to select qualitative sections which are representative of the traffic in Houston. We will first perform a spatial analysis, based on geographic properties, to choose representative highways sections. Then we will used the different factors (travel time ratio, DoS etc.) to select qualitative sections. After selecting those sections we will perform an analysis and determine adequate solutions in order to reduce traffic jams in Houston. For each analysis we will also provide some statistics on the selected sections. Finally we will take the intersection of both analysis in order to have a qualitative and representative set of section.

4.1.1 Spacial Analysis

Let's analyze at first Houston's highway-network:

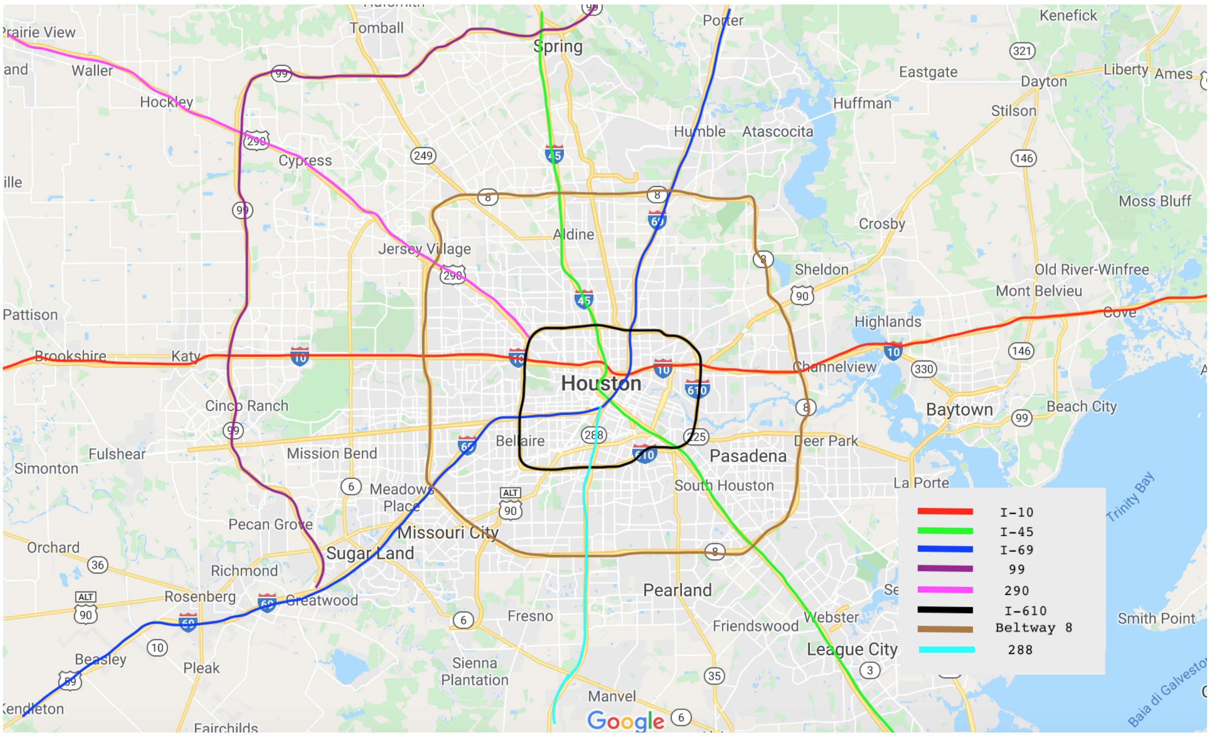


Figure 26: Highlight of Houston's highways we are going to consider

I-10 Interstate Highway, runs from east to west; also called Katy Freeway or Baytown Fast Freeway. Is the major east–west Interstate Highway in the Southern United States. In the U.S. state of Texas, it runs east from Anthony, at the border with New Mexico, through El Paso, San Antonio and Houston to the border with Louisiana in Orange, Texas.

I-45 Interstate Highway, located only within Texas; also called The Gulf Freeway (I-45 South) or the North Freeway (I-45 North). It connects the cities of Dallas and Houston, continuing southeast from Houston to Galveston over the Galveston Causeway to the Gulf of Mexico.

U.S. 59 (*Now called* **I-69**) Highway, runs from the U.S. Mexico-Border in Laredo to Texarkana, Arkansas; also called The Southwest Freeway and the Eastex Freeway (the last one will take you to east Texas).

U.S. 290 Highway, runs east to west, entirely within Texas; also known as The Northwest Freeway or simply 290.

I-610 Intestate highway, the 610 loop circles downtown; commonly identified with North/South/East/West loop or inner loop

Beltway 8 State highway, a second-most outer loop that's entirely within Harris County; also known as Sam Houston Parkway (/Tollway) or more commonly, the Beltway.

Texas 99 State highway, Houston's third outer loop (once completed), also called The Grand Parkway.

SH 288 State highway, Known also as South Freeway.

As the fourth largest city in the United States, Houston is no stranger to traffic issues. The city is very spread out and has limited public transportation system, forcing the majority of Houston's residents to commute by car.
The Texas A&M Transportation Institute has compiled a list of Texas' most congested roadways and six of the state's worst roads can be found in Houston:

- W Loop Freeway from Katy Freeway to Southwest Freeway (I-610: from I-10 to I-69)
- Southwest Freeway from W Loop Freeway to South Freeway (I-69: from I-610 to 288)
- Eastex Freeway from SH 288 to IH 10 (I-69: from 288 to I-10)
- Katy Freeway from N Eldridge Parkway to Sam Houston Tollway W (I-10: from 99 to loop 8)
- Gulf Freeway from IH10 to S Loop E Freeway (I-45 south: from I-10 to 610 south)
- North Freeway from Sam Houston Tollway N to N Loop Freeway (I-45 north: from 8 north to 610 north)

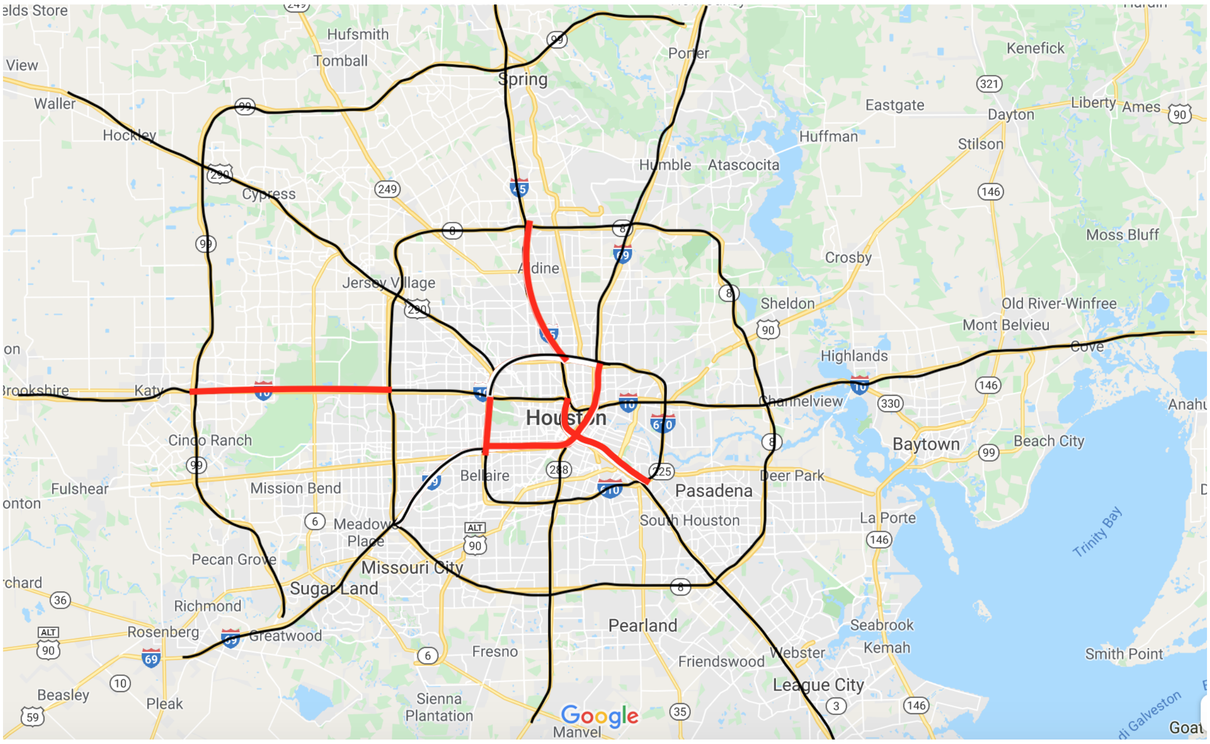


Figure 27: Highlight of the most congested sections

In particular the 610 Loop from I-10 to I-69 in Uptown is the most congested highway in Texas, as stated by *Houston Transtar*. Drivers spend 1.6 million wasted hours in traffic on this roadway per mile.
The Southwest Freeway (I-69) from the 610 Loop to Highway 288 comes in second.

These data reflect also the geographical importance of some highways' sections as links to different part of the city. The highway system from the center of the city spreads in all possible directions and viceversa, and as a network, the links in the middle are expected to be the most congested since they allow to change directions (a part from important sections that represent the only fast way to go from

one side of the city to another, we don't expect to find as most congested sections those not related to more than a highway).
Other geographical considerations might refer to tourist attractions, malls, airports (there are two in Houston, one in the north and one in the south) etc... their geographical sites would lead to similar conclusions.

From the data analysis we related the probability of a traffic jam to occur, to some characteristics of the roads such as the number of lanes and the speed limit. In particular we noticed how the cases we identified as most likely to be interested in a traffic jam are for around the 88% related to roads that have 3, 4 or 6 lanes and almost all of them ($\sim 94\%$) with speed limit between 88 Km/h and 104Km/h. Based on these observations, in the next subsection we'll investigate a method that will allow us to focus our attention only to some cases, narrowing the amount of data into the most critical ones.

4.1.2 Statistical Selection

BASED ON OUR MODEL

Using the overview analysis done on the data, especially the section *'Lanes & Speeds Statistics based on our model'*, we select qualitative lanes. In order to find which configurations is worth to analyze we will use 3 parameters :

- % of section for each configuration
- % of section with DoS greater than 1 for each configuration
- Travel Time Ratio for each configuration

By multiplying all those parameters together we obtain a value, that we named *Cross result*, for each configuration. The higher it is, the higher the configuration is worth to focus on. The result of this computation can be found in table 8.

Speed \ Lanes	2	3	4	5	6
88 km/h	0.00	5.24	18.24	5.71	1.60
96 km/h	0.70	18.60	35.99	9.66	1.93
105 km/h	0.00	0.60	3.07	0.0	23.12
113 km/h	0.00	2.16	3.14	0.00	0.00

Table 8: Cross result value for each configurations

From this table we now select the more relevant configurations. We see 4 configuration standing out from the others. Those are the following:

1. Sections with 4 lanes and speed limit set to 96km/h
2. Sections with 6 lanes and speed limit set to 105km/h
3. Sections with 3 lanes and speed limit set to 96km/h
4. Sections with 4 lanes and speed limit set to 88km/h

In order to get some information on those configurations we compute 3 things. Firstly we compute the number of sections corresponding to one of those configurations. Secondly, we compute the number of section where traffic jam could occur (eq. when DoS is greater than 1) and compute the ratio. Finally we compute the average Travel Time Ratio for all sections corresponding to one of the selected configuration. This is what we obtained :

- Number of Sections : 464
- Sections vulnerable to traffic jam : 272
- Ratio : 58,62 %

- Average Travel Time Ratio : 3,2

Those results are quite satisfying. From 382 vulnerable sections, we can now limit our analysis to 271 sections more relevant. The Ratio of vulnerable sections over non vulnerable sections is almost the same as before (59%). This means that we did not over-fitted data by taking only sections subject to traffic jam. Finally the average Travel Time ratio is equal to 3,2 which is also similar to the one computed for the entire data set (3.04). So once again we did not 'over-fitted' the data and we can consider our selected sections using statistics as quite relevant sections. But as mentioned, our model have some weakness and we need to use the analysis made on provided data only in order to select a really qualitative and representative set of section.

Based on data provided

Using the overview analysis done on the data, especially the one in *'Lanes & Speeds Statistics based on provided data'*, we select qualitative lanes. We have previously seen that our model is a bit too strict. Moreover we know that even if a section is subjected to congestion, there are different degrees of congestion. In this section we will try to separate sections subject to high congestion from those who are only a bit congested. This method should be based on provided data only and not link to our model.

The method we will use is based on the fact that the density function of the Real Flux for the 8 pre-selected combinations is similar to the density function of a normal distribution. Then as mentioned in the section *'Critique of the Model'*, we would like to have a threshold limit situated between 0 and 1σ. This threshold limit would represent the maximum Real Flux for each combination at which there is no congestion. To find this limit we apply the following method :

1. We select real Flux under the main 'bell' of the density function of the real flux
2. From the new set of real flux we compute the mean (μ) and the standard deviation (σ) of the set
3. then we compute the threshold doing $\mu + \frac{1}{2}\cdot\sigma$

In order to better visualize this method, you can find bellow an example of the result we should obtain (graphically) :

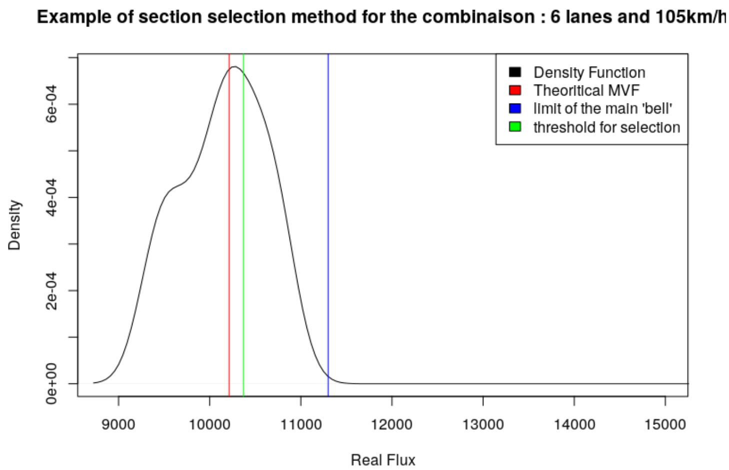


Figure 28: Real flux threshold example

We apply this method for the 8 pre-selected combinations and we obtain the following table . On the 3rd column you can observe the theoretical MVF, on the 4th column it is the threshold computed with the method mentioned before.

Speed Limit	Number of Lanes	Theoritical MFV	Maximal Flux from Real Flux
96	4	6572	7162
105	6	10213	10369
96	3	4976	5489
88	4	6367	7071
96	5	8273	8566
88	3	4824	5164
88	5	8019	8622
105	4	6733	6956

Table 9: MFV per hour for the different speed limits and number of lanes configuration in the data set

By comparing the two MFV we observe that as expected the Maximal Flux computed from the real flux is slightly higher and should allow us to select sections with higher probability to be congested and especially ones with higher congestion. By merging this analysis with the one done with our model we can now select a set of sections from the 4 most representative combination and compute some additional information. We obtain :

- Number of Sections : 464
- Sections vulnerable to traffic jam : 256
- Ratio : 55,17 %
- Average Travel Time Ratio : 2,9

Nota bene: if the ratio between the number of sections selected and the number of vulnerable sections is still high, it is mostly due to the fact that we voluntarily selected sections with high percentage of vulnerable section (cf. cross result) in order to find solutions for combinations who mostly needed it.

Now, based on the 'bell's analysis' we know that for each configuration we have similar trend. Remember that the main bell is always situated in the interval [1300-2200] veh/h, we then have secondary bells on intervals [3000-4500] and [5000-*]. As we said previously, we will consider sections under the main bell as sections not interested on traffic jam. Or more specifically : sections with small congestion only, not worth finding solution for. Thus we'll focus in the remaining 2 groups of sections: those with real flux per lane between 3000veh/h and 4500veh/h and those with real flux per lane greater or equal than 5000veh/h. They cover respectively 112 and 62 sections. Moreover, all of them are, according to our model, sections subjected to traffic jam with high DoS and high TTR. In the next sub-paragraph we will see explain how we are going to use all these information in order to select the most interesting groups of sections for the PNR modification. Finally, note that we don't make distinction between high length sections (1km$\leq$) and small one ($\leq$1km). We already proved, during the data analysis, that the length of the section was not affecting the probability of having traffic jam in a consistently way.

4.1.3 Final Selection

In this section we choose a set of sections to focus on in order to find solutions decreasing the number of traffic jam in Houston. The idea is to combine both geographical and statistical analysis in order to get the most qualitative and representative set of sections. We made the choice to select sections who need it the more, eq. sections with high congestion and huge traffic jam. In fact, it is more realistic to change the worst part of the system than the entire system, remembering that changing the system's architecture has a cost.

From the geographical part we mainly focus on Figure 32 to select sections that are highly congested. It give use precious information on where vulnerable sections are and we can use it to find them on our data set who gives us the highway's name and the beginning/end of each section in form of crossroad. More over we will try to select 'groups' of adjacent sections in order to obtain a realistic solution for the Houston's highways system. In practice it is more likely and convenient to add a lane on a large section of 5km (composed of multiple section) than add a lane for 800m on 10 different section distributed over the all system.

Now, from the statistical analysis we have to choose between different possibilities. Sections with Real Flux greater than our theoretical MFV, sections with real flux not situated in the main bell of their configuration, all combinations or only the most represented ones etc. In order to select representative sections we first decided to focus on those which are part of the 8 most represented combinations, as we did almost all our data analysis on these sections. Moreover, we also want to select qualitative sections. As we decided to focus on the worst sections only, we selected sections with a flux per lane higher than 5000veh/h. By doing this we reduced considerably the number of sections we have to find solution for, but we improved the quality of our set of sections as it now contains only sections who really need improvement.

Combining these two criteria we are able to identify 7 groups of adjacent highly congested sections. The exhaustive list of those section can be found on table 10. In the next page you can find a quick summary of the 7 groups and a plot (figure 29) that shows the density function of all section's fluxes per lane and the average flux per lane for each of the 7 selected groups. We clearly observe that there are 3 bells (centered in 1800,3500 and 5500 veh/h per lane), as described in the data analysis section. Especially we observe that the groups of adjacent sections we selected are, as wanted, situated in the third and last bell. It confirms the fact the we chose groups of sections with high congestion that really need PNR modifications.

- Group 1 :
  - Number of section : 12
  - Total length : 9,81
  - Avg flux per lane : 5618 veh/h
- Group 2 :
  - Number of section : 9
  - Total length : 9,11
  - Avg flux per lane : 5711 veh/h
- Group 3 :
  - Number of section : 6
  - Total length : 6,24
  - Avg flux per lane : 5657 veh/h
- Group 4 :
  - Number of section : 4
  - Total length : 5,63
  - Avg flux per lane : 5391 veh/h
- Group 5 :
  - Number of section : 7
  - Total length : 4.02
  - Avg flux per lane : 6145 veh/h
- Group 6 :
  - Number of section : 5
  - Total length : 5,63
  - Avg flux per lane : 5665 veh/h
- Group 7 :
  - Number of section : 3
  - Total length : 2,09
  - Avg flux per lane : 5480 veh/h

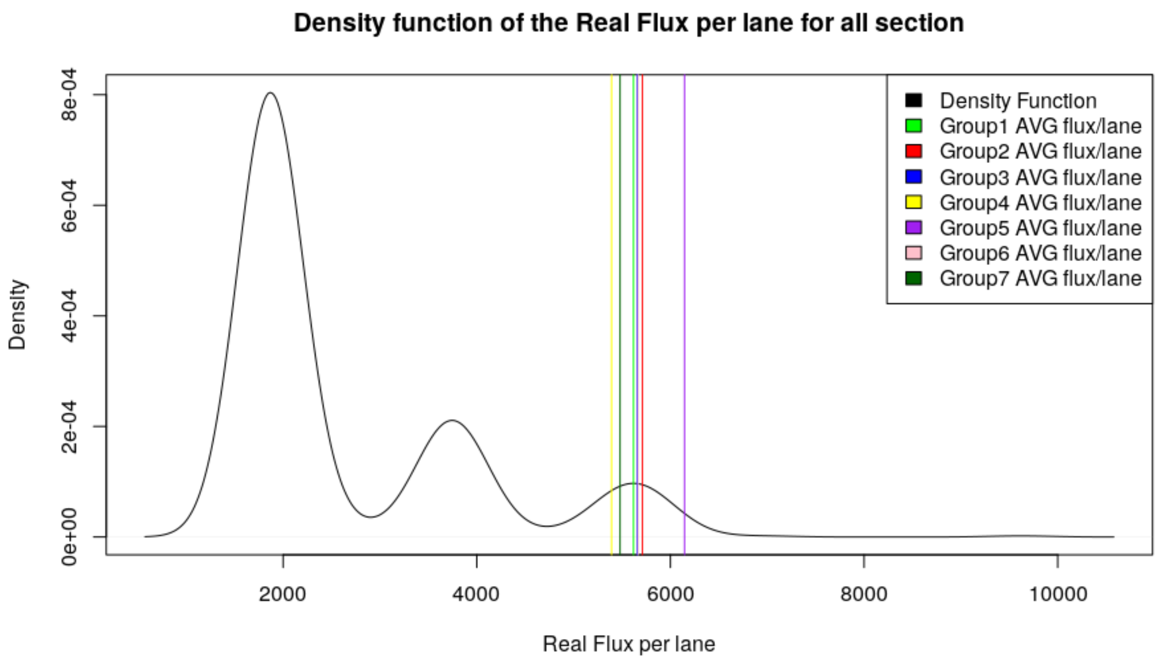


Figure 29: Real flux per lane for all section, with avg flux per lane for each group

Selected sections					
Group	Combina - tion	Highway	Crossroad	Length (Km)	Real Flux per lane
Group 1	97-4	10 Westbound	90	0,24	5920
			610 Counter-Clockwise Northeast	1.85	5600
			610 Clockwise Northeast	0.32	5600
			Gellhorn Drive	0.32	5200
			McCarty Drive	0.23	5640
			90	2.03	5560
			Kress Street, Lathrop Street	2.25	5480
			Lockwood Drive	0.32	6000
	97-3	10 Westbound	Waco Street	0.42	5774
			69 Westbound	0.23	5857
	88-3	10 Westbound	69 Eastbound	0.32	5064
			McKee Street, Hardy Street	1.29	5732
Group 2	88-4	45 Northbound	Woodridge Drive	1,74	5200
			Griggs Road, Broad Street	1.45	5960
			90	0.19	5680
	88-3	45 Northbound	Telephone Road	0.68	5525
			Tellepsen Street	1.06	5483
			Elgin Street, Lockwood Drive, Cullen Boulevard – University of Houston	0.77	5525
			Spur 5 south – University of Houston	0.48	6231
	96-3	45 Northbound	Cullen Boulevard, Elgin Street, Lockwood Drive – University of Houston	0.68	5774
			Scott Street, Pease Street, Saint Joseph Parkway – Downtown Houston	1.16	6023
Group 3	96-3	45 Southbound	Frontage Road	1.42	5525
	96-4	45 Southbound	610 Clockwise Northeast	0.87	5525
			Cavalcade Street, Link Road	0.97	5680
			Patton Street	0.77	5680
			North Main Street, Houston Avenue	0.7	5680
			Quitman Street	1.45	5680
Group 4	96-4	69 Eastbound	Fountainview Drive	0.8	5600
			Chimney Rock Road	1.61	5360
	96-6	69 Eastbound	610 Clockwise Southwest	2.28	5284
			610 Counter-Clockwise Southwest	0.93	5323
Group 5	96-4	69 Westbound	Lyons Avenue / Quitman Street	0.42	5400
			10 Westbound	0.39	5320
			10 Eastbound	0.16	6000
			Jackson Street – Downtown Destinations	0.97	5800
	96-3	69 Westbound	Polk Street – Downtown Destinations	0.48	5483
			45 Northbound	0.97	9615
	96-4	69 Westbound	Gray Avenue / Pierce Avenue – Downtown Destinations	0,64	5400
Group 6	88-3	288 Northbound	Almeda-Genoa Road	0.80	5525
			Orem Drive	0.80	5680
			Airport Boulevard	0.80	5680
			Reed Road	0.80	5680
			Bellfort Avenue	1.61	5680
Group 7	88-4	288 Northbound	South MacGregor Way / North MacGregor Way -Texas Medical Center	0,64	5920
			Binz Street / Calumet Street	0,97	5200
			Southmore Boulevard / Blodgett Street	0,48	5320

Table 10: Selected section to modify (PNR)

4.2 Modification of the PNR

We now focus on each group separately and propose some possible modification of the PNR that from our point of view could help better manage the traffic on those sections.

GROUP 1

Looking at table 10 and focusing on sections of group 1, we can notice that the section that covers crossroad "Lockwood Drive" is the one with higher flux per lane. Moreover, if we look at table 7 we realize it represent a peak section for the configuration 97-4. Looking at its length we can instead notice that the nature of this section of highway 10 is mostly a crossroad and that the sections that immediately follow are with one lane less and higher fluxes respect to those preceding it. Our suggestion would then be to add a lane for sections from "Lockwood Drive" to "Hardy Street". Doing this we would ideally obtain the following (since we are not simulating the flux of vehicles, and our only data are the real fluxes, in the table below we just ideally updated the real flux per lane computing it with the new configuration).

Group 1					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
97-4	10 Westbound	90	0,24	5920	4633
		610 Counter-Clockwise Northeast	1.85	5600	4382
		610 Clockwise Northeast	0.32	5600	4382
		Gellhorn Drive	0.32	5200	4070
		McCarty Drive	0.23	5640	4414
		90	2.03	5560	4351
		Kress Street, Lathrop Street	2.25	5480	4289
97-5		Lockwood Drive	0.32	6000	4696
97-4	10 Westbound	Waco Street	0.42	5774	4170
		69 Westbound	0.23	5857	4230
88-4	10 Westbound	69 Eastbound	0.32	5064	3656
		McKee Street, Hardy Street	1.29	5732	4140

If we then considerate all the sections in the group and relate them to the adjacent portions of the related highway we can notice that the sections we are analyzing come up sections at higher travel speed (113 veh/h); so instead of adding lanes we could suggest to higher the limit speed at 105 veh/h changing combinations 97-4 to 105-4.

GROUP 2

From 10 we observed that sections on group 2 are divided into 3 configurations, 88-4, 88-3 and 96-3. Also in section 'Spur5 south–University of Houston' we observe the maximal flux per lane for the combination 88-4. Firstly, as there is a subgroup of sections with speed limit at 96km/h we can imagine that we can change the speed limit of all section in the group to 96 km/h. But, looking at the map we observe that theses sections are situated at the intersection of 3 highways. So instead of increasing the speed limit, which can be dangerous, we will increase the number of lanes in order to decongest the traffic at this intersection. By doing this we expect to avoid the occurrence of huge traffic jam that can paralyzed several highways. If we know look at the result presented on table for group 2 we observe that almost all section have now a flux less than 4500 veh/h so it means they are now situated under the second bell and not anymore under the third one, which is what we are looking for.

Group 2					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
88-4	45 Northbound	Woodridge Drive	1,74	5200	4070
		Griggs Road, Broad Street	1.45	5960	4664
		90	0.19	5680	4445
88-3	45 Northbound	Telephone Road	0.68	5525	3990
		Tellepsen Street	1.06	5483	3960
		Elgin Street, Lockwood Drive, Cullen Boulevard – University of Houston	0.77	5525	3990
		Spur 5 south – University of Houston	0.48	6231	4500
96-3	45 Northbound	Cullen Boulevard, Elgin Street, Lockwood Drive – University of Houston	0.68	5774	4170
		Scott Street, Pease Street, Saint Joseph Parkway – Downtown Houston	1.16	6023	4350

GROUP 3

Looking at the highway portion in which sections of group 3 are located we can notice that mainly all the 45 Southbound highway has speed limit of 97Km/h. It seems then more reasonable to change the number of lanes instead of the limit speed, cause change the velocity only in this portion of highway would just probably cause a congestion at the end of it due to the need of slowing down the travel speed.

Group 3					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
97-4	45 Southbound	Frontage Road	1.42	5982	4320
97-5	45 Southbound	610 Clockwise Northeast	0.87	5441	4414
		Cavalcade Street, Link Road	0.97	5640	4257
		Patton Street	0.77	5440	4226
		North Main Street, Houston Avenue	0.7	5400	4539
		Quitman Street	1.45	5800	4445

GROUP 4

Looking at the higways portion of group 4 we observe that the speed limit is mostly set at 96km and that only the lane configuration change over the highway. Moreover looking at the maps we notice that once again we are in at the intersection of two highways : 69 Eastbound and 610. This could explain the high traffic. Intersections play a major role into the creation of traffic jam, so it is important to treat them. As said previously, speed limit is set at 96km/h in almost all the highway so it is not relevant to modify the speed limit for only 4 kilometers, moreover it could be risky to increase the speed at an intersection. We can still modify the number of lanes for each section by constructing a new lane. Moreover we observe that there are already 6 lanes for the intersection where there are only 4 and 5 lanes before and after the intersection. The idea is to add one lane at each section to fluidify incoming and outgoing flux and add a lane for user that are not exiting the highways. We can see that, after modification, all section are now part of the 2nd bell and not anymore of the 3rd one.

Group 4					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
96-5	69 Eastbound	Fountainview Drive	0.8	5600	4383
		Chimney Rock Road	1.61	5360	4195
96-7	69 Eastbound	610 Clockwise Southwest	2.28	5284	4484
		610 Counter-Clockwise Southwest	0.93	5323	4516

GROUP 5

Looking at the fluxes of sections in group 5 we can notice the presence of high values for crossroads "10 Eastbound" and "45 Northbound"; in particular if we look at table 7 we can verify that those two sections of highway registered peak-flux values for their respective configurations (97-4 and 97-3). Looking at the configurations of the sections preceding and succeeding this portion of highway we can notice that the speed limit is mostly set at 97 km/h for all the 69 Westbound highway, while what changes along different portions of it is the number of lanes (which is, near these sections of interest, frequently higher than the one characterizing the sections of group 5). Therefore, reasoning in a similar way as for group 3, we would suggest to increase the number of lanes.

Group 5					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
96-5	69 Westbound	Lyons Avenue / Quitman Street	0.42	5400	4226
		10 Westbound	0.39	5320	4163
		10 Eastbound	0.16	6000	4696
		Jackson Street – Downtown Destinations	0.97	5800	4539
96-4	69 Westbound	Polk Street – Downtown Destinations	0.48	5483	3960
		45 Northbound	0.97	9615	6944
96-5	69 Westbound	Gray Avenue / Pierce Avenue – Downtown Destinations	0,64	5400	4226

GROUP 6

Section from group 6 are situated on route 288, which as a speed limit set to 88km/h. This time we are not in presence of intersection with another highways, so we could eventually increase the speed. But looking closely at the map we observe that we are near a residential district and increasing the speed will probably disturbing the neighbourhood, moreover we are on a route, not an highways. Instead, as group 6 is only composed of one composition (88-3), we add 1 lane to all section.

Group 6					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
88-4	288 Northbound	Almeda-Genoa Road	0.80	5525	4140
		Orem Drive	0.80	5680	4050
		Airport Boulevard	0.80	5680	4140
		Reed Road	0.80	5680	3930
		Bellfort Avenue	1.61	5680	4200

GROUP 7

Since all route 288 has speed limit set at 88km/h and these three sections of group 7 are sections of small length, it wouldn't make sense to increase the speed limit. Again, we would suggest to increase the number of lanes to 5

Group 7					
New Combination	Highway	Crossroad	Length (Km)	Old RF per lane	New RF per lane
88-5	288 Northbound	South MacGregor Way / North MacGregor Way -Texas Medical Center	0,64	5920	4633
		Binz Street / Calumet Street	0,97	5200	4069
		Southmore Boulevard / Blodgett Street	0,48	5320	4163

Consideration if we observe the values of the real flux before and after the suggested adjustments, we can notice how such a change would ideally move all the fluxes in the "third bell/interval of values" to the second one.

We would like to stress the fact that we are not simulating the fluxes but we are instead using real data regarding the fluxes of vehicles along the highway network on a particular day and at a particular time; so the tables above are just to better show how, ideally, such modifications of the PRN would affect and change the results.

4.3 Results

So far we have been analyzing the characteristics of the traffic in Houston's highway network, based on data of the real flux trough each section in which the system is divided in (in a certain hour of the day - more precisely we recall that the data set provided is regarding the flux of vehicles on the day of January 29th 2018 between 8 a.m. and 9 a.m. assuming the flux to be constant in that portion of time). In particular we have noticed how the flux presents a recurrent multi-modal behaviour (more precisely 3-modal, with modes identified by what we've been calling "bells") on which we have focused our attention on. Talking with percentage, $69,1\%$ of the sections in analysis have registered fluxes with values in the main bell (i.e. up to 2000 veh/h per lane), which we have defined as representative of a non stressed traffic; while the remaining $30,9\%$ are within the second and the third bell and represents instead stressed sections where the congestion is high and traffic jams are more likely to occur. Being more precise, this $30,9\%$ of highly stressed sections are made up by a $20,6\%$ of sections with fluxes around the interval [3000 4000] veh/(h lane), i.e. sections in the "second" bell; and a $10,3\%$ of sections that are in the most left bell with fluxes in between 5000veh/(h lane) and 6000veh/(h lane).

Consideration : around 90% of the sections are contained in the first two main bells. (For these proportions we have neglected values of fluxes per lane higher than 6000 veh/h because they represent few isolated cases that we assumed to be associated with errors on the data.) The sections that fit in the third "mode" are half those with values within the second trend interval but represent still the 10% of the cases; which is an high portion of the data and this justifies the choice to modify the PRN based on these sections instead of considering them outliers and focusing on the second main interval of values (i.e 3000-4000 veh/h per lane).

What is interesting is the fact that all fluxes appear to fit perfectly in one of the three windows of values that characterize the multi-modal behaviour we have analyzed trough the "bells" of figure 16.

In figure 32 we have highlighted in the map those sections that are not in in the main bell; i.e that 30% of sections that are most likely to be interested in a traffic jam and we can notice, as previously mentioned, that those sections are in fact adjacent sections that identify some portions of the highways; suggesting that that behaviour could be due to a geographical reason (that can be due to the location of that portion of highway respect either to the map of Houston and the location of it services and attractions, either to the highway network and more in general transport network).

We have then decided to analyse in a more accurate way the fluxes of theses "risky" sections, looking at the in-going flux and out-going flux characterizing each of them. We observed the presence, in both directions, of extremely high values (respect to the average values in the provided data set) that might be responsible of such increases in the flux interval. Some of these values, once matched with the respect crossroad and identified in the map, can be explained by their location (for example: presence of airport, medical center, museum district, zoo, intersection and/or union on entries/exits with other important highways, and so on) or the way traffic normally evolves (these values were registered for sections we highlighted as highly congested, so we can interpret an higher throughput as due to the fact that while in a situation of traffic, cars slow down their speed travel but at the same time drivers feel more confident to reduce the safety distance, leading to a bigger number of cars passing by that crossroad/intersection). Other values are more logically suggesting errors in the data while for others more we weren't able to find a reason based on own knowledge of Houston's territory ; but still, we haven't been able to find an appropriate reason of why the values of the total real flux are exactly multiples by 2 or 4 with each other.

After modeling traffic jam, performing different analysis on the provided data set and selecting qualitative and representative sections to analyse, it is now time to discuss and interpret our results. In the previous subsections we have selected seven groups of adjacent sections using both geographical and statistical factors. We decided to select only sections with the highest flux per lane (i.e. sections with flux per lane in the 3rd bell's interval [5000,+] veh/h) which as motivated before was a reasonable choice. Another reason of why we decided to focus only on the most left interval of values instead of focusing on the whole 30% of sections that do not fit those values characterizing the main bell interval is based on the fact that 10% is a more reasonable amount of sections to focus on, instead of working on changes that affects 30% of the whole highway system (so it's a more realistic and convenient choice also from in the prospective of the costs involving such changes). Moreover the selected groups are of adjacent sections since it is more efficient and effective to optimize large portion of highways together than small spread spread sections, not related to each other.

On each selected group, the suggestions about possible changes in the physical network of route (PNR) were for the majority of the cases regarding changes in the number of lanes. The main result of such modifications was the shifting of the values of flux per lane that went from the third interval of values (old configuration) to the second (with the new configuration); as we can observe in the tables of the previous section and on figure 30.

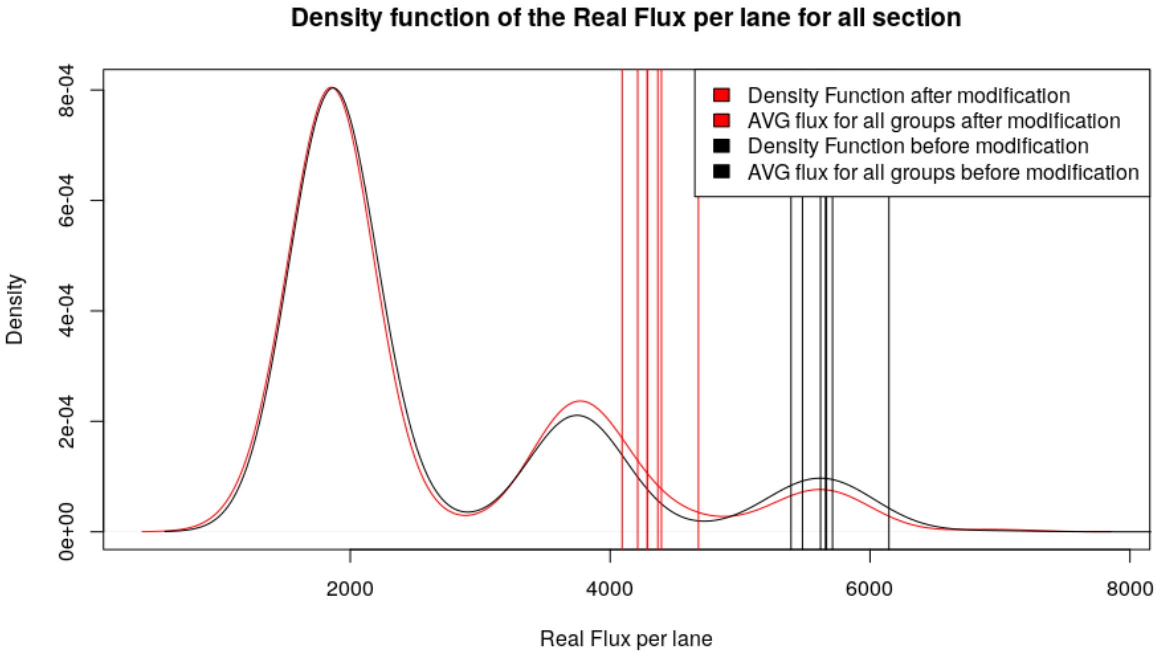


Figure 30: Real flux per lane for all section, with avg flux per lane for each group

The main goal of this last phase was to provide some suggestion on the modification of the PNR in order to reduce the impact of traffic on the most vulnerable sections of Houston's highway system. Since to modify 30% of the network is a huge work, what we did was focusing on 1/3rd of that percentage, i.e. we worked on the the most "risky" 10% of sections in the whole network, and provided solutions in order to eliminate the third trend and obtain a bi-modal system, made at the 70% by sections that fits in the main bell and are not likely to be interested in traffic jams and at the 30% by sections that are still at risk of congestion but that are all represented and defined by the same interval of value (i.e. the one of the second bell) - which is the second lower one. Considering that flux is discrete and not continuous and then that each bell represents a level of congestion, we can see that by adding a lane to each section we 'downgrade' those section to a lower level of congestion and therefore limit the intensity of traffic jams.

We have to keep in mind that modification of the PNR are costly and cannot be applied everywhere: even adding a single lane can be, in some cases, pretty herd or not possible (it requires space, money and usually creates conflicts with the neighborhood, especially when the section is located in the city). Other solutions, such as changing the speed limit or adding/modifying exits/entrances can be valid options, but with our model and the provided data set, we don't have the tools to motivate such choices neither analyse precisely their impact on the traffic.

Further consideration: We know that the provided data set regards data acquired on January 29th 2018 at the peak hour between 8am and 9am. By looking closely at the geographic situation of the selected groups, we can observe that most of them regard intersections needed/used to enter Houston's center. Based on this we could then associate our results and observation to the fact that at 8am most of the users of the Houston's highways system are probably heading to work. It follows that a similar scenario could also appear around 5pm when people are instead heading to home, leading to other congestion in the opposite way of travel. It would be then appropriate to optimize the fluxes in both directions.

5 Conclusion

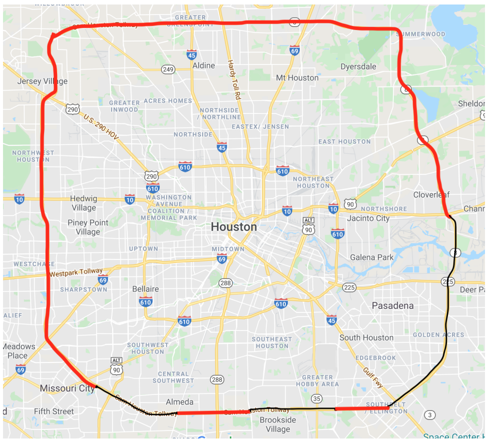
All major cities, anywhere on Earth, experience traffic jams, and this doesn't surprise us. However, in a society where everything runs fast and is accessible in timescales that only a decade ago would have been unimaginable, the value of time has become more and more important to the point of becoming a cost. That's why planification is not only important and more in demand, but a real need.

In the whole report we analyzed the present situation in Houston in term of fluxes and traffic jams, with two main goals: give a description of Houston's highways system and try to find solutions that best suits it, in order to improve the network system and with it the user's experience.

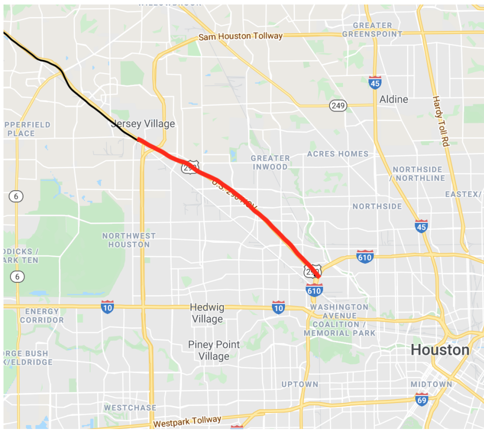
Let's a have a quick recap of what we have done: in order to get a global view of the system, we first had to model a key concept: *'the maximal flux of vehicles'* (MFV). We considered the MFV as the maximal flux of vehicle that allows users to drive at maximum speed and therefore experience no delay on a travel from A to B compared to the minimum theoretical travel time. Using the MFV we were able to compute different factors as the degree of saturation of a section, the estimated travel time for a specific section etc. Then, to have a complete analysis we deeply analyzed the provided data set which were regarding an ordinary week day in Houston's highway network, at a peak traffic hour. Form these data we were able to obverse some really odd behaviours, that we tried to investigate and that lead us to some interesting consideration (like the interpretation of traffic as a discrete variable/event) but also to some particular results which left us with some curiosity. Lastly, we used the knowledge of the system's weaknesses, we achieved with our analysis, to selected a set of representative and qualitative sections to optimize by some PNR modifications.

Beyond all things we have highlighted and also questions we have raised, this project left us with some deeper thoughts. Traffic jams are one of the consequences of years of unrestrained growth economically as well as demographically. We always want more, and keep constructing aiming at what's more comfortable for our situation without wondering if there's a different systematic of acting and thinking. Instead of enlarging routes and road networks, another valid option can always be to find a compromise. Netherlands can be an example in this case: it's the European country with the highest number of bikes per inhabitants and the PNR has been adapted in order to promote such a way of travelling. In a crucial moment of our history, with the development of new technologies but also the growth of inequality and the ecological awareness of the new generation, one may wonder if the real solution is to adapt the world to our needs or our needs to the world.

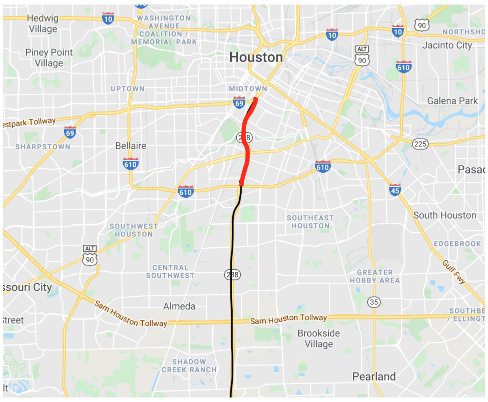
6 Annex



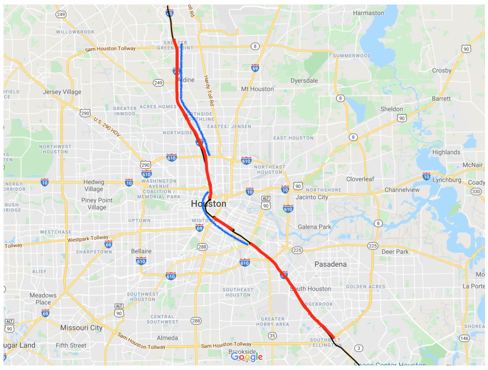
(a) **Beltway 8**



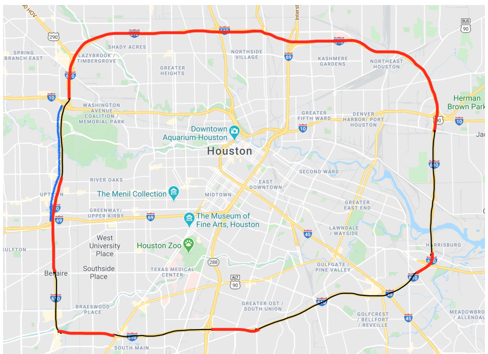
(b) **U.S. 290**



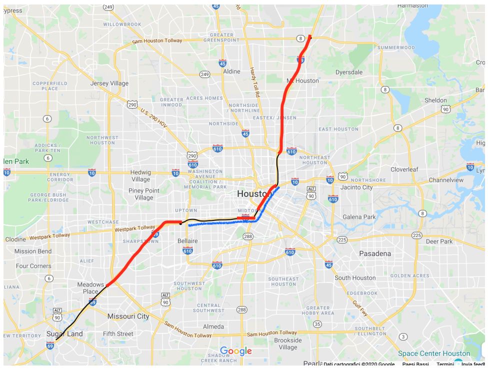
(c) **SH 288**



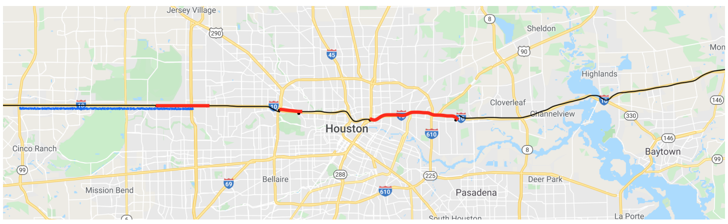
(d) **I-45**



(e) **I-610**

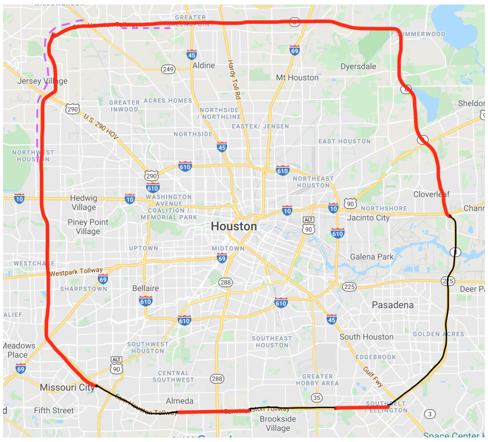


(f) **I-69**

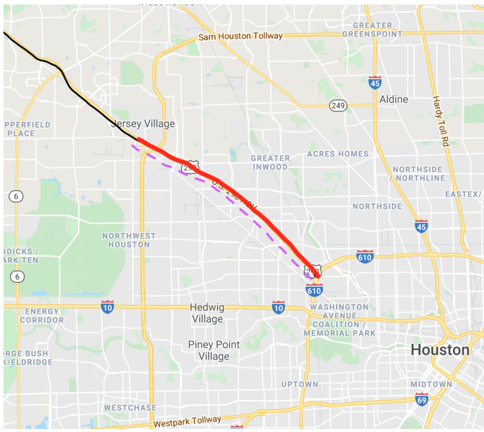


(g) **I-10**

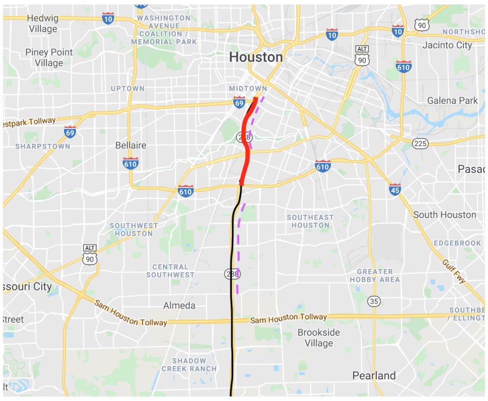
Figure 31: Traffic jams representation. In red: section **most likely to be congested** according to the statistical analysis. In blue: section **highly congested** according to the geographical analysis



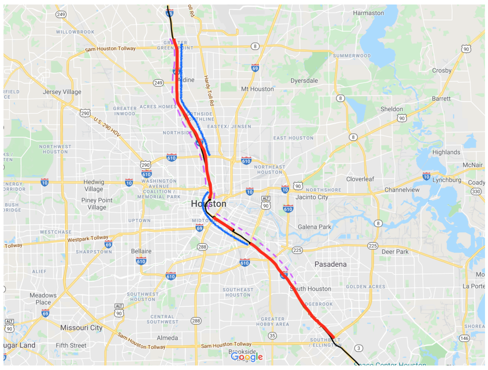
(a) **Beltway 8**



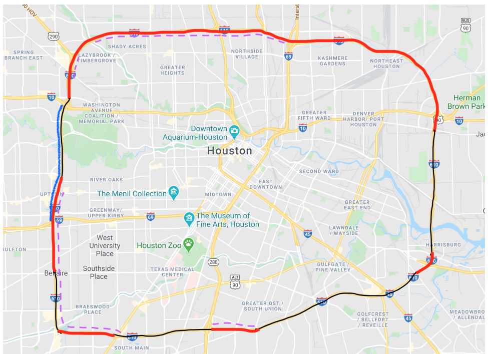
(b) **U.S. 290**



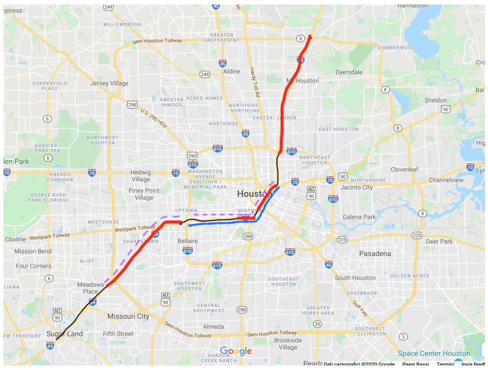
(c) **SH 288**



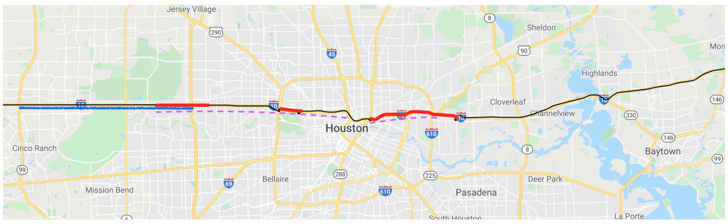
(d) **I-45**



(e) **I-610**



(f) **I-69**



(g) **I-10**

Figure 32: Traffic jams representation. In violet: most congested sections according to the **data analysis**; i.e. sections identified in the **secondary bells**

References

[1] National Highway Traffic Safety Administration. Why your reaction time matter at speed, 2015.

[2] Bicycle Dutch. Dutch cycling figures, 2018.

[3] Project of the George Washington University Face the fact. Get the numbers of that truck, 2013.

[4] AAA Foundation and DC Urban Institute, Wahsinghton. American driving survey, 2015 – 2016, 2018.

[5] Samuel Labi Kumares C. Sinha. *Transportation Decision Making: Principles of Project Evaluation and Programming*. John Wiley Sons, 2007.

[6] Bureau of Transportation Statistics. State motor-vehicle registrations - 2018, 2018.

[7] Wikipedia open source community. Braking distance, 2020.

[8] National Cooperative Richard Dowling. *Planning Techniques to Estimate Speeds and Service Volumes for Planning*. National Research Council, 1997.