

AH2174 Traffic Simulation Modeling and Applications

Lab 2: Microscopic Traffic Simulation Modeling using SUMO

Alaa Eltayeb

Alaae@kth.se

For Q1, could you explain how you introduce randomness to simulation in addition to by vehicle generation? How many vehicles are generated? Any issues?

random seeds are used for the random number generator, for each simulation run, the current clock time is used as the random seed to be generated as poison process as explained in the vehicle generating part. The total number of vehicles generated is 3000 vehicles the attribute departure lane in route file was set to = best, as best lane among the 3 lanes, Different random seeds used for the random number generator the vehicle dept with different seeds 5 and 10 simulation runs are done to simulate the dynamic of traffic scenarios

one of the issues were the reduce of the numbers of generate vehicles and this is solved by changing the route attribute to best lane as mentioned above, below are the process used to generate the random vehicle

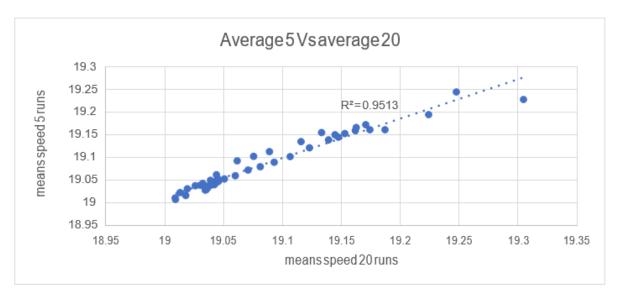
```
def generateRandomVehicles(rate, maxtime, routeid, veh type):
    #Ex: <vehicle depart="30" id="veh0" route="r1" type="Car"/>
    tdep = 0.0
    vehseq = []
    vehid = 0
    while tdep <= maxtime:
        tint = -math.log(random.random()) / rate
        tdep += tint
        vehseq.append('<vehicle depart="{}" id="veh{}" route="{}" type="{}"/>'
    return vehseq
def writeToFile(strList, filename):
  with open(filename, 'a') as filehandle:
       for str in strList:
            filehandle.write(str + '\n')
if __name__ == "__main__":
    # Example of usage
   seq = generateRandomVehicles(0.8333, 3600, "r1", "Car") # Note: remember
    writeToFile(seq, "testfile.xml")
```

I am bit confused with Fig 6-8, not sure what they mean, also Fig 9. It is good that you probe with statistical thoughts but you might need to explain what they mean.

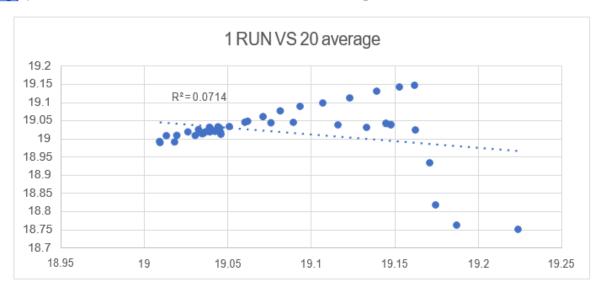
In fig6-8, this figures were meant to described and explained the correlation and similarity between the different runs the purpose of this is to indicate which of the average values of 5 and 20 runs are more presentable to the mean speed of the actual dynamic of traffic the main question was is it worth it to conduct more runs, is there is a significant differences between them when It comes to the average speed, to check this, the figures below were represented to see the similarity, correlation and closeness of values between the different runs (1, 5 and 20)

The conclusion made that, there is a simiularity between the values of 5 and 20 runs up to 95% thus, an average of 5 runs can be taken as the preventive of the dynamic of the traffic .

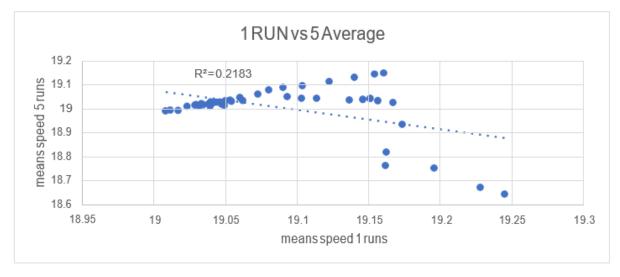
Fig(9): this fig were made to check the outliers values of the speeds in each runs, since I went deeper in the statistical part I wanted to check the extreme values using the box plot methods for further calculations, however all the data of the different runs were taken into consideration in this report,



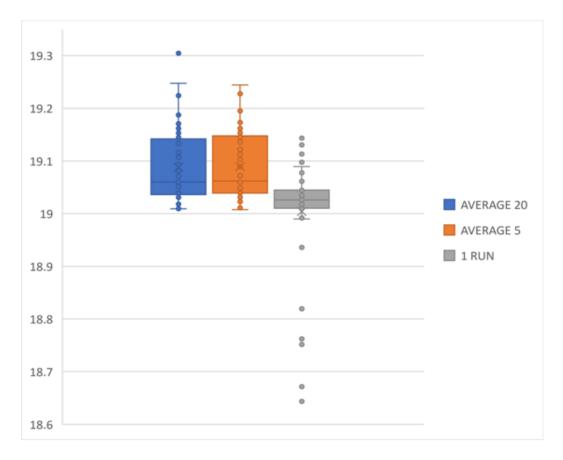
Fig(6): Coefficient of determination R2 between the average 5 and 20 runs



Fig(7): Coefficient of determination R2 between the average 20 and 1 run



Fig(8): Coefficient of determination R2 between the average 5 and 1 run



Fig(9) outliers checking

For Q2, I am also confused with what you are testing under table 2. Please explain clearly.

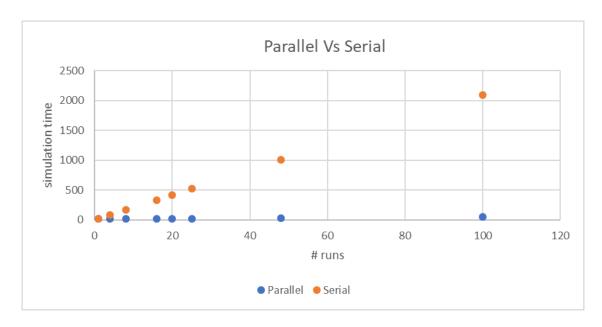
AVERAGE 2	20	AVERA	GE 5		1 RUN
Mean	19,08810976	Moan	19,08913198	Moan	19,00399662
Standard Error	•	Standard Error	•	Standard Error	0,016102704
Median	19,06017768		19,06169045		19,02602961
		Standard		Standard	
Standard Deviation	0,069775916	Deviation	0,062220394	Deviation	0,108020224

Table (2): descriptive analysis

In this table mainly, some of the variables were important not all, as the mean for example in the hypothesis testing part it is required, another way to check the similarity and the differences between the runs in

For Q4, I hope that you can provide a graph on the relation between computational time and simulation runs for both serial and parallel simulation.

Condition of the parallel of t								
Simulation	1	4	8	16	20	25	48	100
Runs								
Parallel	16.111	16.467	17.586	18.586	18.116	19.309	27.704	55.359
Serial	21.477	83.6	169.41	335.643	419.867	524.638	1004.14	2093.22



Looking at the graph above, as the numbers of runs increased the simulation time increased respectively however in the serial simulation the increment is significantly larger than the parallel , looking at the table above, a single run compare to a 100 runs, the increment of time in serial is more than 138 times the first run compare to 3,5 in the parallel. the serial simulation is costing more time than the parallel simulation, the parallel simulation can be easily conducting in the case of availability of the multi-processors CPUs as in the PDC labs .

Introduction:

SUMO is an open source microscopic traffic simulation used for modelling microscopic vehicle behavior. A simple model for a traffic network constructed manually to determine the vehicle behavior during every simulation step with inputs such as traffic demands. Also, a model created to understand the car following model and their impacts through the simulation experiments.in last the simulation output were analyzed and discussed.

Simulation of traffic flows:

Create simulation networks

A simple road network of a highway section composed of five edges (edge 1, edge 2, edge 3, edge 4 and edge 5) of one kilometer each (e.g. from (0,0) to (0, 1000), ..., to (0, 5000)) with a three lanes with speed limit 70 km/h, was constructed manually, the network modeling includes the following steps

(i) Nodes were created under the file name"sumo test.nod.xml".

(ii) Then edges were Defined between nodes in the file"sumo test.edge.xml";

(iii) The edge types were defined in the file "sumo_test.type.xml"; the 3 lanes inputs as 3 lane with speed70 (km/hr), and as shown the speed inputs is in m/second.

(iv) To create the network file, the netconvert command were applied using the following syntax: netconvert --node-files sumo_test.nod.xml --edge-files sumo_test.edge.xml -t sumo test.type.xml -o sumo test.net.xml

```
netconvert --node-files sumo_test.nod.xml --edge-files my_edge.edge.xml -t sumo_test.type.xml -o sumo_test.net.xml
```

Prepare vehicle inputs for simulation:

In this report the vehicle demand in SUMO were created by simply adding the vehicle type, route and simulated vehicles by editing the file sumo_test.rou.xml:

```
∃<routes>
      <vType accel="2.5" decel="5.0" id="Car" length="5.0" maxSpeed="40" sigma="0.0" />
      <vType accel="1.5" decel="4.0" id="Bus" length="18.0" maxSpeed="25" sigma="0.0"/>
      <route id="r1" edges="link1 link2 link3 link4 link5"/>
 8
      <vehicle depart="30" id="veh0" route="r1" type="Car"/>
      <vehicle depart="35" id="bus0" route="r1" type="Bus"/>
<vehicle depart="120" id="veh1" route="r1" type="Car"/>
 Q
      <vehicle depart="140" id="bus1" route="r1" type="Bus"/>
      <vehicle depart="180" id="veh2" route="r1" type="Car"/>
13
      <vehicle depart="235" id="bus2" route="r1" type="Bus"/>
      <vehicle depart="330" id="veh3" route="r1" type="Car"/>
14
      <vehicle depart="335" id="bus3" route="r1" type="Bus"/>
15
     L</routes>
17
```

In order to run the simulation using sumo-gui and load the simulation a SUMO configuration file were created "file sumo_test.sumo.cfg", the net-file were defined also the route files in the input section, and simulation time (3600) the length of the step (0.1) in addition to some output of the simulation as FCD, summary and tripinfo files to customize the simulation model. After this step SUMO was opened and simulate the network and configuration

```
<configuration>
   <net-file value="sumo test.net.xml"/>
   <route-files value="sim_work.rou1.xml"/>
   </input>
   <time>
   <begin value="0"/>
   <end value="3600"/>
   <step-length value= "0.1"/>
10 </time>
   <traci server>
   <remote-port value="8813"/>
   </traci_server>
13
   <report>
14
   <no-duration-log value="true"/>
15
16
   <no-step-log value="true"/>
17
   </report>
   <output>
   <netstate-dump value="fcd.xml"/>
19
20
   </output>
   <output>
22
   <tripinfo-output value="c.xml"/>
23
    </output>
    <output>
2.4
25 <summary value="summary.xml"/>
26
   </output>
27 </configuration>
```

Vehicle Generation:

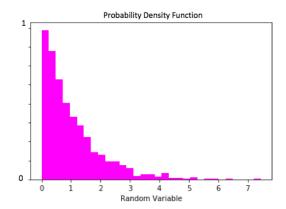
The methodology used to generate a random numbers of vehicle departures in a given duration is build on the following assumptions:

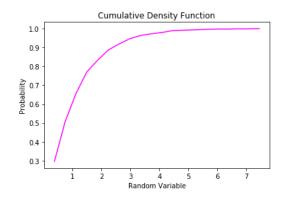
- ☐ The vehicles departure/arrival is occurred independently; knowing when one vehicle departure gives no information about when another vehicle will departure.
- Another assumption is that at any hour of the day or at any length of the time at this point/segment of the street is no different than any other hour, in other words, the probability that an event occur in a given length of time doesn't change through the time; the theoretical rate at which the event are occurring doesn't change.

This led us to conclude that the vehicles departure occurs randomly and independently, if these conditions met then, X which is the number of events in a fixed unit of time has a passion distribution.

The steps used to generate the vehicles departure:

- 1. Generate a random uniform values that has a uniform distribute from U ~Unif(0,1), where U is a random uniform variable then use this values to generate a random values from an exponential distribution $F_X(x) = 1 e^{-\lambda x}$ for x>0, $u = 1 e^{-x\lambda}$ fig(1)
- 2. Then the inverse CDF of the exponential distribution is applied, $F_X^{-1}(u)$, fig(2)
- 3. The $X = F_X^{-1}(U)$ The computed random variable X has distribution $F_X(x)$ is computed
- 4. $X = F_X^{-1}(u) = -\frac{\log(1-u)}{\lambda}$





Fig(1) fig(2):

Fig (1): Probability Density function values of vehicle departure

Fig(2): Cumulative probability distribution values of vehicle arrivals

Simulation output:

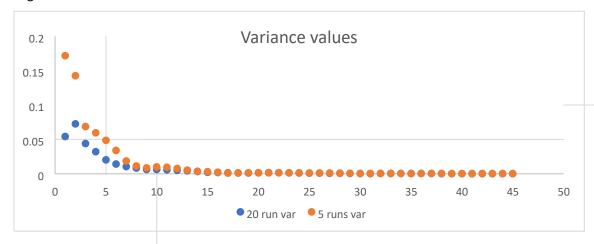
The output of the simulation is basically the average vehicle speed in the highway section every minute during simulation, an input of a 60 min is considered the simulation time with simulation timestep 0.1 sec,

the first 15 minutes is considered as warm up period that did not taken into account for the average speed calculation



Fig(3): The mean speed values for 1 run, an average of 5 runs and an average of 20 runs, in addition of an average of 2

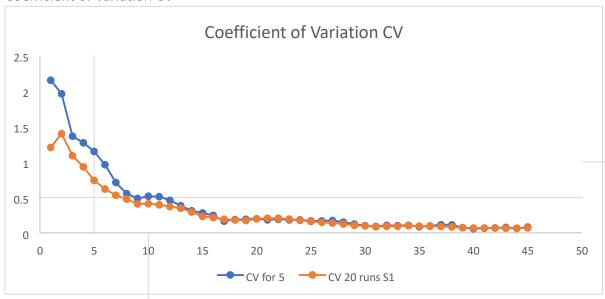
From the graph we can see one major differences between times around 0 and 10 Minutes of the simulation time this can It corresponds to the random loading of the vehicle at the first node, although the 15 min warmup time did not included during the simulation, the vehicles still carried a random speed when emerging into the node, however after the 10 min, the speed continuous to be almost uniform until the end of the simulation time. The values of the speed dropped from its highest value to its lowest after around 10 min, this could be due to the constrained the vehicle faces from the other vehicles, also as mentioned before vehicle random loading into the first node, however by the end of the simulation the value started to increase again and this can be explained as the vehicles started to departure the road segment.



Fig(4): the variance values for the average 5 runs and 20 runs

The fig() above describe the variance values for the average of 5 runs and 20 runs, the Variance is a measurement of the spread between numbers in a data set. The variance measures how far each number in the set is from the mean. In the first node and due to the random loading of the vehicles, the variance value for each simulation time is different for both runs, however and around the 10min of the simulation time, the values for both runs are almost equal.

Coefficient of Variation CV



Fig(5): coefficient of variation

CV %	average runs	
0.378323		5
0.293376		20

Table(): the average values of the CV for 5 and 20 runs

The fig(5) & table(1) compare between the average of 5 runs and 20 runs, personally, I believe the CV is more valid to compare between 2 samples than the variance, each of the runs has its own mean values thus compare between them in variance is not validate, however CV is more particularly useful when comparing results from two different tests that have different measures or values, from the table () it is noticeable that the 5 runs are more spread and has more variance relative to the mean more than the 20 runs.

Correlation Coefficient R:

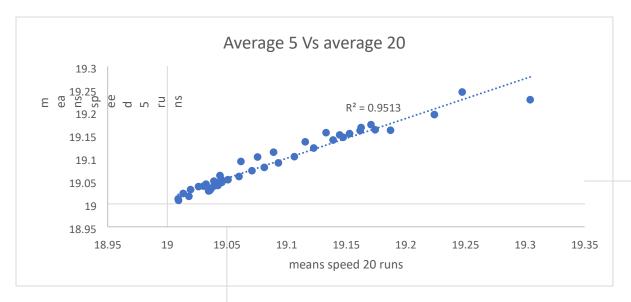
	AVERAGE 20	AVERAGE 5	1 RUN
AVERAGE 20	1		
AVERAGE 5	0,975338697	1	
1 RUN	-0,561687218	-0,46719113	1

Table (1): Correlation Coefficient R

The correlation indicate how much the data from two sample are similar to each other, or how the expected data is described by the observation data, however here it made to see how the numbers of

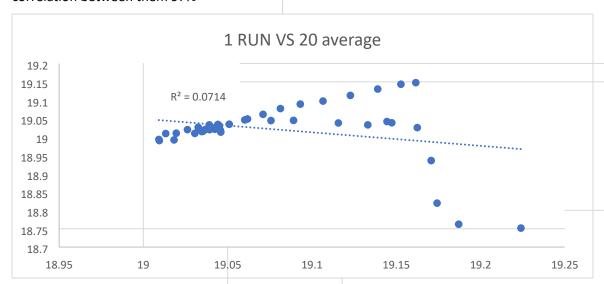
runs are correlated between each other, the single run and the average of 5 runs and 20 rubs have a correlation R equal to 46% and 56% respectively, however the average of 5 and 20 has a higher similarity between them, 97.5%, this can indicate that the average values of 5 and 20 runs are more presentable to the mean speed of the street more than the values of a single runs.

Coefficient of determination R²

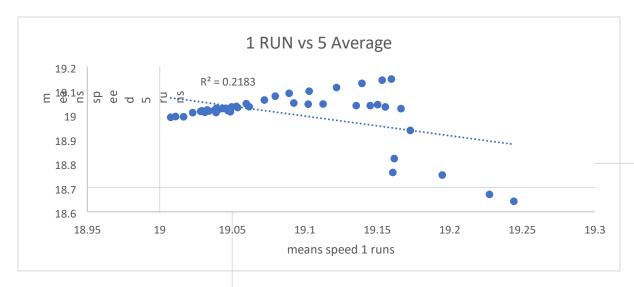


Fig(6): Coefficient of determination R² between the average 5 and 20 runs

There is a significantly similarity on mean speed values between the two runs (both increase the same way with almost the same values). Also, the R² value is equal to 0.9513 which can be interpreted as the strength of the linear relationship between the two simulation outputs is equal to 95%, and the correlation between them 97%



Fig(7): Coefficient of determination R² between the average 20 and 1 run

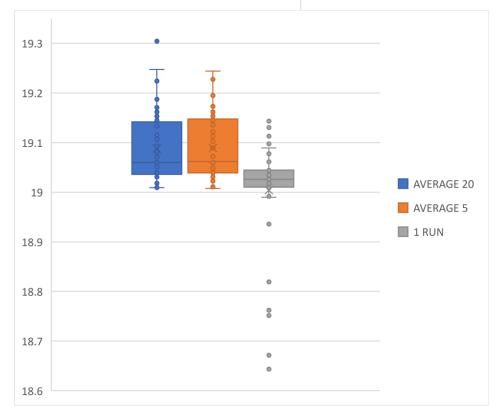


Fig(8): Coefficient of determination R² between the average 5 and 1 run

However, by checking the 1 run vs the 5 average and 20 average runs the determination Coefficient R² gives a value equal to that 21% and 7% respectively, and the correlation between them R -46% and - 56% respectively. This can be due to the random loading of the vehicles at the first node during the 10 min period as previously explained.

Outliers deduction:

To deduct the outliers a box-plot diagram and identify potential outliers in the data



However, personally, I believe this outliers should be included in the further calculation, this values are random but repeatable in each runs at the first node such, further analysis such as regression analysis in case i.e. a Robust regression which used when there are outliers in the data and its important to have them in the data points.

AVERAGE	20	AVER	4 <i>GE 5</i>		1 RUN
Mean	19,08810976	Mean	19,08913198	Mean	19,00399662
Standard Error	0,010401579	Standard Error	0,009275269	Standard Error	0,016102704
Median	19,06017768	Median	19,06169045	Median	19,02602961
Mode	#N/A	Mode	#N/A	Mode	#N/A
		Standard		Standard	
Standard Deviation	0,069775916	Deviation	0,062220394	Deviation	0,108020224
		Sample			
Sample Variance	0,004868678	Variance	0,003871377	Sample Variance	0,011668369
Kurtosis	0,899030832	Kurtosis	-0,530141334	Kurtosis	4,277384577
Skewness	1,149800388	Skewness	0,702329798	Skewness	-2,052801784
Range	0,295192425	Range	0,2366187	Range	0,504528679
Minimum	19,00915626	Minimum	19,00758841	Minimum	18,64337184
Maximum	19,30434869	Maximum	19,24420711	Maximum	19,14790052
Sum	858,9649394	Sum	859,010939	Sum	855,1798477
Count	45	Count	45	Count	45
Confidence		Confidence		Confidence	
Level(95,0%)	0,020963006	Level(95,0%)	0,018693076	Level(95,0%)	0,032452868

Table(2): descriptive analysis for 1 run and average of 5 and 20 runs

The table above, give details about the descriptive analysis of the mean speed for the given simulation time, for 3 different runs,

hypothesis testing:

 μ = 19,1685 s/m (null hypothesis that the average of 20 runs has no effect on the average mean speed of a single run 1) the mean speed of is still will equal to 19,003

H₁: $\mu \neq$ 19,003 s/m (the alternative hypothesis that the the ACC has effect on the average mean speed) Assumptions:

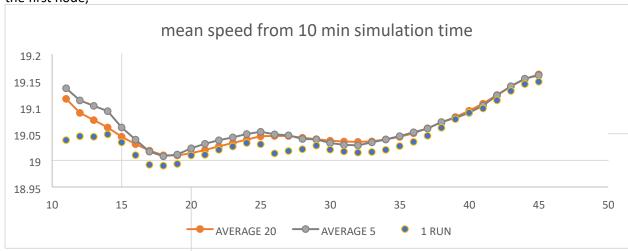
- 1. Level of confidence equal to 95% for a level of confidence 95% the $t\alpha/2$ equal to 1,96.
- 2. Although the sample size is 45 the t-distribution will be used instead of the z-distribution since we don't know the population variance.
- 3. Regarding the variance, the sample variance will be used

for that the mean speed values are similar at 5% level of confidence

 $^{= \}frac{19.088-19.003}{\frac{0.0860,0697799}{0.0860,0697799}} = 8.172528798 > 1.96 \text{ we reject the null hypotheses There is not sufficient evidence}$

A trail to element the first node:

The fig() below, illustrating the value of the mean speed after the elimination the 10 first values in at the first node,



Fig(10): mean speed after eliminating the 10 mins of simulation

AVERAGE 20		<u>AVERAG</u>	<u>E 5</u>		1 RUN	
Mean	19.05912	Mean Standard	19.06376	Mean Standard		19.04164
Standard Deviation	0.040522	Deviation	0.041578	Deviation Sample		0.041496
Sample Variance	0.001642	Sample Variance	0.001729	Variance		0.001722

table(3): descriptive analysis after eliminating the 10 mins

AVERAGE	20	AVERAG	GE 5		1 RUN
Mean	19,08810976	Mean	19,08913198	Mean	19,00399662
		Standard		Standard	
Standard Deviation	0,069775916	Deviation	0,062220394	Deviation	0,108020224
		Sample			
Sample Variance	0,004868678	Variance	0,003871377	Sample Variance	0,011668369

Table (4): descriptive analysis for all the 45 simulation time

	AVERAGE 20	AVERAGE 5	1 RUN
AVERAGE 20	1		
AVERAGE 5	0,973811599	1	
1 RUN	0,943985234	0,894589845	1

Table(5):

The table above shows the correlation between the different runs,

the mean is the most sensitive measurement when it comes to outliers because it takes the "value" of each data and this may cause a clustering around a certain data just because it has a large value while the other data share a slight difference in their values, in our case the mean dropped from 165 secs to 159 sec. standard deviation as well affected by the outliers because it measures how the data are spread around the mean.

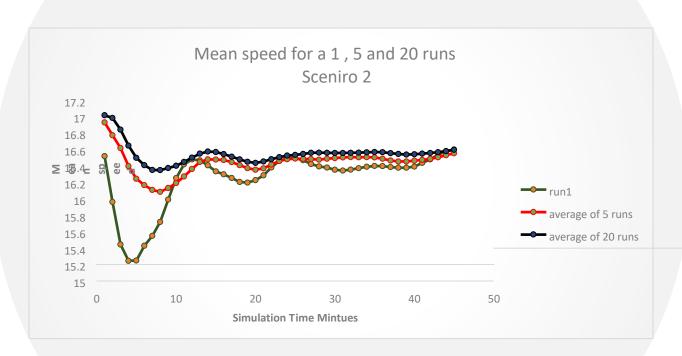
3.3 Alternative scenario (scenario 2)

The last edge, edge 5 is changed to a two lane instead of 3 lanes, due to some construction events and the speed limit is set to 40 km/h for the s

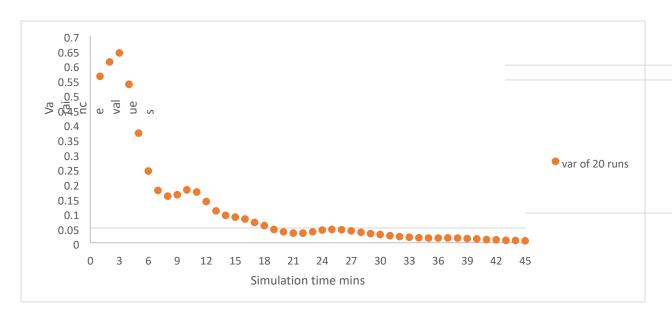
Same step^c
were re^c

file editing

Sim

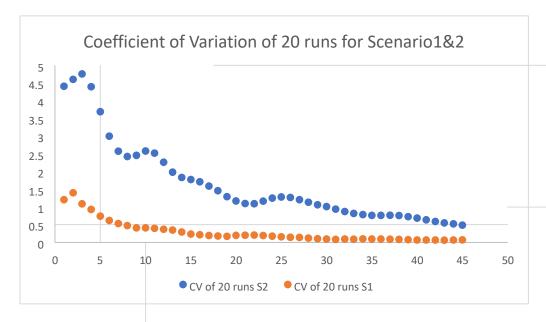


Fig

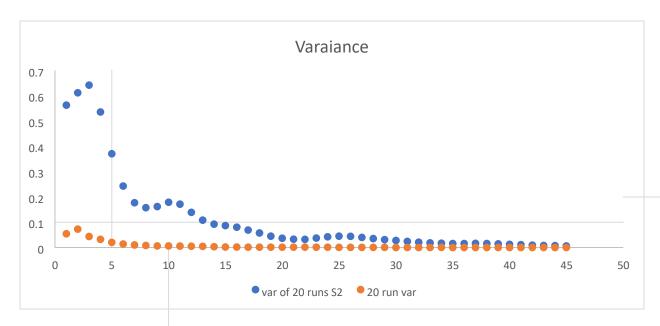


Fig(12): the Variance values for S2 average of 20 runs

Compare between the two scenarios:



Fig(13): CD for S1 and S2

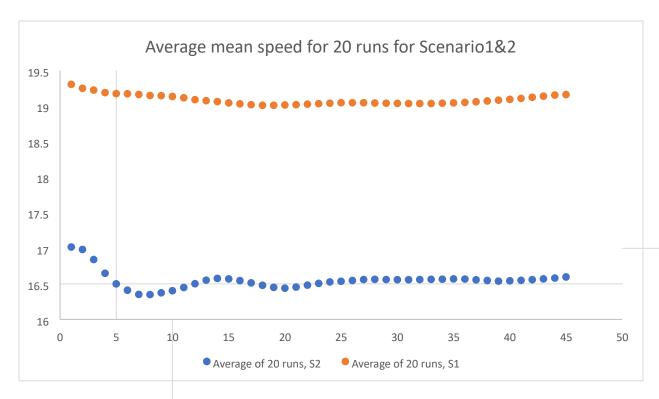


Fig(14): variance for S2 and S1

It is noticeable that there is a large variation and spread of the speed in S2 than S1, and this is due to the change in speed limits from 70 to 40 km/hr at the last lane and also to reducing in the number of lanes from 3 to 2.

Average of 20 runs, S2		Average of 20 runs, S1	
Mean	16.54824	l Mean	19.08811
Median	16.54993	B Median	19.06018
Standard Deviation	0.127924	Standard Deviation	0.069776
Sample Variance	0.016364	Sample Variance	0.004869
Range	0.668018	3 Range	0.295192
Minimum	16.34986	5 Minimum	19.00916
Maximum	17.01788	3 Maximum	19.30435
Sum	744.6707	7 Sum	858.9649
Count	45	5 Count	45
Confidence Level(95.0%)	0.038433	Confidence Level(95.0%)	0.020963

Table(6): descriptive analysis for both Scenario 1 and 2



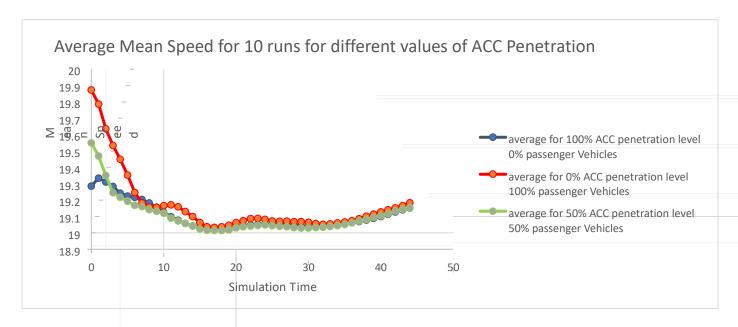
Fig(15): average mean speed for S2,S1

From the table (6) and fig (15) The average mean speed dropped in the Scenario 2 which is expected due to the reduce of the number of the lanes in that lanes and the speed reduction.

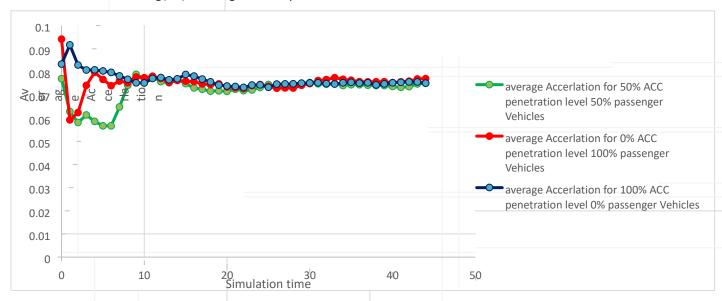
Effects of car-following models

DM model has been used to represent vehicles with adaptive cruise control (ACC) system given that no driver reaction time is considered. the default Krauss model is used to represent passenger vehicle whereas IDM model is used to represent ACC controlled vehicles in traffic simulation.

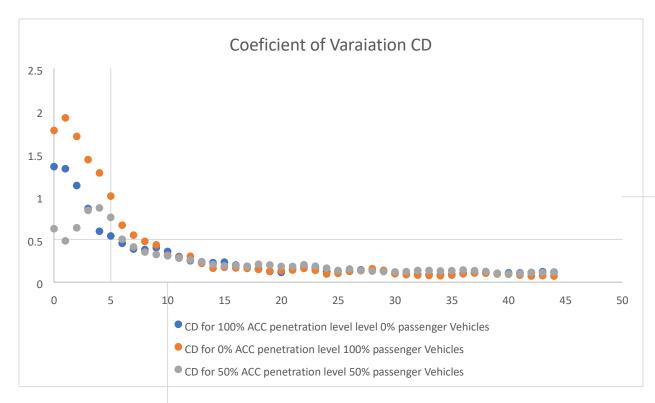
Consider that the same highway scenarios 1 but with two types vehicles in traffic: passenger vehicles and ACC control vehicles. The effects of ACC penetration levels (0%, 50% and 100%) is compared below



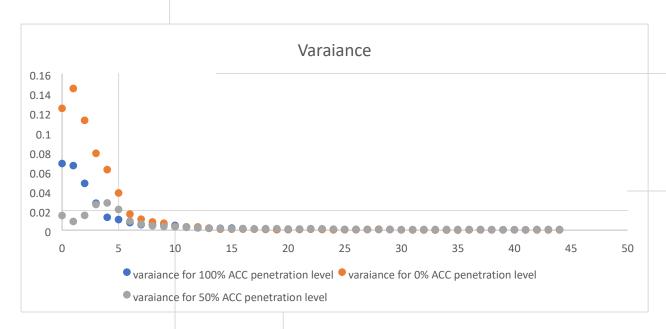
Fig(16): average mean speed for 10 runs for the 3 scenarios



The fig(17) shows the acceleration on the 3 scenarios, Acceleration is defined as the rate of change of velocity with time



Fig(18): the CD for 3 sceneries



Fig(19): The Variance for the 3 sceneries

average Accerlation for 50% ACC penetration level 50% passenger Vehicles		average Accerlation for 0% ACC penetration level 100% passenger Vehicles		average Accerlation for 100% ACC penetration level 0% passenger Vehicles	
Mean	0.07211	Mean	0.075266	Mean	0.076599
Standard Deviation	0.005644	Standard Deviation	0.004485	Standard Deviation	0.003323
Sample Variance	3.19E-05	Sample Variance	2.01E-05	Sample Variance	1.1E-05
Range	0.022234	Range	0.034931	Range	0.01841
Minimum	0.056684	Minimum	0.059215	Minimum	0.073235
Maximum	0.078918	Maximum	0.094145	Maximum	0.091645
Sum	3.244931	Sum	3.386964	Sum	3.446954
Count	45	Count	45	Count	45
Confidence Level(95.0%)	0.001696	Confidence Level(95.0%)	0.001348	Confidence Level(95.0%)	0.000998

Table(7): descriptive analysis for average accerlation for the 3 penetration scenarios

average mean speed for 100% ACC penetration level 0% passenger Vehicles		average mean speed for 0% ACC penetration level 100% passenger Vehicles		average mean speed for 50% ACC penetration level 50% passenger Vehicles	
Mean	19.10261	Mean	19.1685	Mean	19.10433
Standard Deviation	0.086995	Standard Deviation	0.192005	Standard Deviation	0.113436
Sample Variance	0.007568	Sample Variance	0.036866	Sample Variance	0.012868
Range	0.318987	Range	0.839611	Range	0.534971
Minimum	19.01607	Minimum	19.03363	Minimum	19.01503
Maximum	19.33505	Maximum	19.87324	Maximum	19.55
Sum	859.6176	Sum	862.5826	Sum	859.6947
Count	45	Count	45	Count	45
		Confidence		Confidence	
Confidence Level(95.0%)	0.026136	Level(95.0%)	0.057685	Level(95.0%)	0.03408

Table(8): descriptive analysis for average mean speed for the 3 penetration scenarios

To compare between the 3 penetration scenarios , in scenario 1 with 100% of passenger vehicle, from fig(16 and 17) also table 8& 7as the speed is rapidly decreasing from 19.9 to 19.2 at 10 mins , and the accelation which defined as the rate of changing velocity with time decreased to a 36% of its values within the first minute at simulation time, and then the vehicle accerlation stared to increase until it reached a the 5^{th} min of the simulation time were drop of 15% from its peak values is recorded , however the value continue to be uniform and almost within an equal rate along the simulation time after around the 10 min of simulation

In the case of scenario 2^d , which introduce 50% of each vehicle types, the average accertation has it lowest values among all the other scenarios, this could be due to the affect that the driver reaction time on the passenger vehicle has on the ACC which has zero reaction time.

In the 3 rdscenario with the 100% of ACC, both mean speed and average accerlation values seemed to have almost uniform distribuation and the values within an fixed rate along the road expect for some peak on the average accerlation on the start of the simulation process.

However all the scenarios and as the simulation time increased the values of the average mean speed and the average of accerlation reminded almost the same value after the 10 min of the simulation time. same explanation goes for the variance values and the CD as well

Hypothesis testing

 μ = 19,1685 s/m (null hypothesis that the ACC has no effect on the average mean speed) the mean speed of is still will equal to 19,1685 even within ACC

H₁: $\mu \neq$ 19,16850283 s/m (the alternative hypothesis that the the ACC has effect on the average mean speed) Assumptions:

- 1. Level of confidence equal to 95% for a level of confidence 95% the $t\alpha/2$ equal to 1,96.
- 2. Although the sample size is 45 the t-distribution will be used instead of the z-distribution since we don't know the population variance.
- 3. Regarding the variance, the sample variance will be used

Test statistic

$$t = (\bar{x} - \mu_0) / (s / \sqrt{n})$$

where:

- x = the sample mean
- μ₀ = the hypothesized population mean
- s = the sample standard deviation
- n = the sample size

The mean of having 100 % penetration of the ACC vehicles = 19,10261

average for 100% ACC penetration level 0% passenger Vehicles				
Mean	19,10261353			
Standard Deviation	0,086995477			
Sample Variance	0,007568213			
Count	45			
Confidence Level(95,0%)	0,026136335			

$$\frac{19.1026-19.1685}{0.08699}$$
 = -5,081535868 ,, 5,081535868 < -1,96 we regent the null hypothesis

There is not sufficient evidence for that the mean speed values are similar at 5% level of confidence

Parallel traffic simulations

The parallel experiments on parallelization of traffic simulation runs were carried out in KTH PDC , the number of processes (max_process) on a computer node on Tegner is around 42 maximum process while a normal laptop has a total of 4 limits,

To check the performance the time it took the computer to do the simulation is taken , The PDC values for a parallel simulation with 5,10 and 100 seeds

5:0m10.66s

10: 0m30.405s

100: 1m44.95