Traffic Engineering (CIVIL-349)
Nikolas Geroliminis
Lab Assignment 1
Studying traffic flow properties through real trajectory datasets collected by a swarm of drones

Fall 2025

École Polytechnique Fédérale de Lausanne
School of Architecture, Civil & Environmental Engineering

Introduction

In this laboratory assignment, you will use a real dataset of vehicle trajectories in order to produce and extract useful information from a set of traffic descriptive graphs, namely (i) Time-Space Diagrams, (ii) Fundamental Diagrams and (iii) Input-Output Diagrams, and you will study the formation and the characteristics of shockwaves. The objective of this assignment is for students to understand and familiarize themselves with the concept and the basic characteristics of these important graphical tools, so that they are able to extract traffic related information from real traffic datasets.

To achieve this, you will work together with your lab group to fill out missing sections in the notebook `lab1_notebook.ipynb`, and produce a lab report to answer the questions.

Figure 1: Using a fleet of drones to collect high-quality traffic data in large-scale networks

The Experiment

Transportation engineers have been collecting traffic data mainly with fixed-location sensors or mobile sensors (GPS or cellphones) which may be inefficient when it comes to studying large scale networks. Lately, the use of Unmanned Aerial Systems (UAS - commonly known as drones) has been proposed as a state-of-the-art tool.

Figure 2: Heavy congestion and complex vehicle interactions, seen in a snapshot of the video recorded from a drone during this large-scale experiment in Athens, Greece. Car trajectories used for this lab assignment come from this specific arterial road.

We recently conducted a first of its kind experiment, deploying a swarm of drones over the busy city centre of Athens, Greece, in order to record traffic streams for multiple days. The total study area covers a 1.3 km^2 area with more than 100 km-lanes of road network, consisting of low, medium and high volume arterials, around 100 busy intersections (signalized or not), many bus stops and close to half a million trajectories. The experiment took place during the morning peak hour (8:00 am-10:30 am) for 5 working days.

The aim of the experiment is to record traffic streams in a multi-modal congested environment over an urban setting by using drones. This unique dataset, consisting of detailed vehicle trajectories, can allow for deep investigation of critical traffic phenomena in an-order-of-magnitude higher than what was available till now.

Available Dataset

The dataset that is used for the completion of this lab assignment, collected from the arterial road seen in Figure 2, consists of two groups of files:

- A set of .csv files of vehicle trajectories that may expand in **all lanes** of the road segment (folder "all lanes")
- A set of .csv files of vehicle trajectories that remain in the **central lane** of the road segment (folder "central lane")

Each file is named 'trajectoryX.csv', where X is an integer number denoting the unique ID of a specific trajectory (which refers to a specific vehicle). The file contains the following information:

- The unique vehicle ID number ('Tracked Vehicle'),
- The type of vehicle, e.g. 'car', 'bus', etc. ('Type')
- The point in time ('Time') counted in [seconds],
- The position of the vehicle ('Latitude' and 'Longitude') in the WGS-84 system of coordinates [degrees], at the corresponding time,
- The speed of the vehicle in [km/h] ('Speed') at the corresponding time,
- The distance of the vehicle from the entry point of the road segment in [metres] ('Distance'), at the corresponding time

Task 1: Time-Space Diagrams

In the first task, you are asked to create the Time-Space diagram referring to the road segment of interest from $x = 0$ m to $x = 360$ m, where the trajectories of all the vehicles observed will be plotted between $t = 0$ sec and $t = 900$ sec. You will then use this data to perform some analysis on headway, spacing, and queuing characteristics. Open Jupyter Notebook `lab1_notebook.ipynb` which contains your main code file for this lab.

Go through the functions related to this task, try to understand what each one does and how they can be used to create Time-Space diagrams of vehicle trajectories, calculate headway, spacing, etc.

1.1 Time-space diagram for all lanes

For this task, use the dataset referring to **all lanes** of the arterial road studied.

a. Select the trajectory data for the vehicles with ID numbers 25 and 32. Using their velocity profiles, plot the instantaneous vehicle accelerations as a function of time, assuming that it is constant in each time interval between two consecutive velocity measurements. Create two well annotated graphs with a clear, descriptive title and properly labeled axes with physical units. Discuss the shape of each acceleration curve and comment on what the acceleration patterns suggest about the driving behavior of each vehicle (e.g., smooth driving, frequent braking/acceleration). Reflect on whether it is possible to conclude that an overtaking maneuver occurred, based **only on the acceleration graphs**.

b. Plot the trajectories of all the vehicles provided in the dataset in a single Time-Space diagram. Create a well annotated graph with a clear, descriptive title and properly labeled axes with physical units.

c. Based on the produced Time-Space diagram, give a qualitative description of the traffic conditions in this road segment during the time of observation. Can you identify any particular traffic phenomena taking place (e.g., vehicle overtakes, density and speed changes, queue formation etc.)? What information is difficult to extract due to the fact that all lanes are plotted in the same diagram? Give a reasoning for each of your observations.

1.2 Time-space diagram for central lane only

For this question, use the dataset referring only to the **central lane** of the road.

a. Plot the trajectories of all the vehicles provided in the dataset in a single Time-Space diagram in the same way as in Task 1.1. Create a well annotated graph with a clear, descriptive title and properly labeled axes with physical units.

b. Similar to question 1.1.c, provide a short description of the various traffic phenomena that you can identify from the produced Time-Space diagram. What are the differences that you observe in this graph with respect to the Time-Space diagram referring to all the lanes of the road? Justify your answers.

c. Take the sequence of vehicle trajectories $[12, 13, 14, 15, 16, 17]$ in the Time-Space diagram of question 1.2.a, plot them separately in the same plot, and estimate the following:

1. The average spacing $\bar{s}$ between them **every second**, for a period of **100 seconds**, with start time $t = 260$ sec. Provide a graph to visualize the evolution of the average spacing between the selected group of vehicles over time for the selected period, and calculate the mean and the standard deviation of the average spacing.

2. The average headway $\bar{h}$ between them as it is observed in **10 different points** (stationary observers) of the road, that are located in a row, every 25 m of distance, with the first one being at the position $x = 50$ m from the beginning of the road segment. Provide a graph to visualize the evolution of the average headway between the selected group of vehicles over the length of the selected distance (defined by the 10 points of the stationary observers). Calculate the mean and the standard deviation of the average headway for the selected length of the road.

d. Describe in words how the average spacing and headway of this group of vehicles change over time and space, respectively, and comment on the plots you created. Why do these changes appear? Give an intuitive explanation based on what you observe in the graphs.

1.3 Queueing analysis for all lanes

For this question, you will use the dataset for **all lanes** of the road. Please answer the following questions related to the arrival of vehicles and the service rate downstream of the road segment located between $x_0 = 160$ m and $x_1 = 290$ m, for the entire duration between 0 sec and 900 sec.

Suppose that the road segment between x_0 and x_1 is normally a signalized intersection with the stop line at x_1; however, the traffic lights are currently out of service. As a result, the traffic is operating as if all-way stop signs, such as the one shown in Figure 3, are present; in other words, it is as if there were a stop sign placed at each approach. The traffic lights at other neighbouring intersections are unaffected and operate as usual.

Figure 3: All-way stop sign at the intersection.

a. In your own words, describe qualitatively how your time-space diagram in Task 1.1 would look differently in this scenario. You should comment on the speed and the behaviour of traffic leaving the intersection (you can refer to the appearance of a time-space diagram or input-output diagram). To simplify the problem, you can assume that traffic arrivals along the studied segment and in opposing approaches are both Poisson processes with a similar distribution.

Once vehicles arrive at the intersection, they proceed in a first-in-first-out manner. Assume that the arrival and departure of vehicles at this intersection can be described as an M/G/1 queuing system with infinite capacity, i.e., a single-server system with exponentially distributed arrivals (to the stop line at x_1) and departures (through x_1) that follow some general distribution. The average vehicle arrival rate is λ [time^{-1}], and the average service rate is μ [time^{-1}]. The service time corresponds to the time that the first vehicle in line takes to cross the stop line at x_1 and the waiting time in queue corresponds to the extra delay that vehicles experience before they become the first vehicle at the stop line.

b. We want to estimate the arrival rate λ using the available vehicle trajectory data. To this end, you will fit an exponential distribution to the empirical data describing the interarrival times between vehicles by following the steps below:

1. Firstly, plot the histogram of the interarrival times between the vehicles and assess the goodness-of-fit of a Poisson distribution assumption for the arrival rate.

2. To estimate λ, you will first create a suitable dataset $\mathcal{D}$ for fitting a continuous exponential distribution function. Represent each histogram bin i as a triplet $(t_{\text{start},i}, t_{\text{end},i}, N_i)$, where N_i is the number of interarrival time observations in that bin. Then, approximate the location of each bin by its midpoint

$$x_i = \frac{t_{\text{start},i} + t_{\text{end},i}}{2},$$

   and compute the corresponding empirical probability density

$$p_i = \frac{N_i}{N(t_{\text{end},i} - t_{\text{start},i})},$$

   where $N = \sum_i N_i$ is the total number of samples. Finally, collect the pairs (x_i, p_i) to form the dataset $\mathcal{D}$.

3. Fit a continuous exponential distribution to the dataset $\mathcal{D}$ and report the estimated value λ in [vehicles/second] rounded to three decimal places. In the same figure, plot the **empirical pdf** by normalizing the histogram in part **b.1** and the one obtained by the

fitting procedure. Create a well annotated graph and comment on the obtained results.

c. Assume now that the service times follow the Erlang distribution. For the arrival rate λ found in **b.3**, and using the theoretical results seen in class, plot the corresponding service rate μ in [vehicles/second] as a function of the average queue length $L_Q \in (0, \infty)$ and comment qualitatively on the obtained shape of the curve. Which service rate yields the average queue length $L_Q = 2$ veh and what is the corresponding theoretical average queuing time w_Q?

d. Assume now that the queuing is modeled as M/M/1/20 system, meaning that the vehicle occupancy of the road segment cannot exceed 20 vehicles. For the arrival rate λ determined in **b.3**, plot the average queue length as a function of the service rate μ and comment qualitatively on the obtained shape of the curve. Determine approximately which service rate yields the average queue length $L_Q = 2$ veh in this case.

Task 2: Fundamental Diagram

In this task, you will estimate and plot the Fundamental Diagram (FD) between flow q and density k for two segments of the given road, based on the given trajectory datasets. The FD can be created by estimating pairs of flow q and density k at several time periods and within a specific road segment, as described below. The generated points will be plotted on a common graph, with the density k on the horizontal axis and the flow q on the vertical axis. For this task, use the dataset of trajectories referring to **all lanes** of the road.

We will describe the process of creating the FD based on the Time-Space diagram that you already produced in question 1.1. As illustrated in Figure 4, consider that the time horizon of this diagram (900 sec) is split in periods of $dt = 10$ seconds. You may think of this process as drawing vertical lines passing through times 0 sec, 10 sec, 20 sec, 30 sec, ..., and so on. Also, define two points of reference (where the imaginary stationary observers would be):

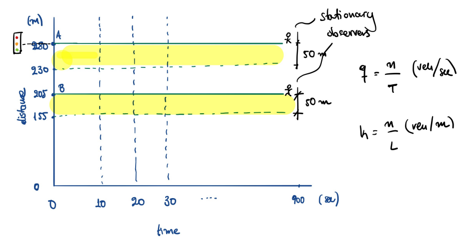

Figure 4: Schematic of Time-Space diagram to visualize the collection of $(q,\ k)$ pairs of values.

- point **A**, which is located at $x = 280$ m, where the traffic light is, and
- point **B**, which is located 75 m upstream of point A.

a. You can visualize what the two stationary observers would see over time by drawing two horizontal lines in the Time-Space diagram, originating from points A and B, respectively. Please follow the explanations below and plot two sets of pairs of (q, k) values (from points A and B) in a common graph. Use a different color for each set of points. Add a legend on the graph to identify each set and provide a title, as well as names and units for the two axes.

1. Estimate the flow q, in veh/min, at points A and B (as if there was a stationary observer at each of the two points) for every period dt of 10 sec.

   The flow can be estimated by the relationship $q = n/T$, where n is the number of vehicles passing from the point of the observer (A or B) during a period T.

2. For every value of q, estimate the density k in veh/m for the road segment starting at the reference point (A or B) and expanding for 50 m upstream.

   For each time slot, the respective value of density can be found by the relationship $k = x/L$, where x is the number of vehicles inside the road segment of interest (of length $L = 50$ m) at the moment that corresponds to the midpoint of each time slot, e.g., for the time slot $[t_i, t_i + 10]$ sec you should estimate k for $t = t_i + 5$ sec.

b. Comment on the form of the FD that you plotted in question a. Does it correspond to the expected form? What could be the sources of scatter observed in your graph?

c. In this question we will try to formulate a mathematical way to describe the Fundamental Diagram and identify its basic properties like critical (maximum) flow q_{crit} and density k_{crit}, jam density k_{jam}. There are several acceptable methods for this. One of these is the method of four points, in which we find the four measured points (k, q) that are closest to the corners of a rectangle formed by the following points:

- Point 1: $(0, 0)$
- Point 2: $(0, q'_{\text{max}})$
- Point 3: $(k'_{\text{max}}, q'_{\text{max}})$
- Point 4: $(k_{\text{max}'}, 0)$

where q'_{max} and k'_{max} refer to the maximum measured values for flow and density, respectively. Then we can approximate the FD by fitting these points to a 3rd-degree polynomial function.

First merge all observed data points (from observers A and B) for k and q in a single dataset. Then, apply the four-points method to approximate the FD by fitting a 3rd degree polynomial to the four resulting points. Report the mathematical expression of the FD and plot it in the same graph from question (a).

d. Based on the FD that you found in the previous question, try to estimate the values of the critical (maximum) flow q_{crit}, the critical density k_{crit}, the jam density k_{jam}, the free flow speed v_{ff} and the speed of the shockwave w. Describe the process you follow based on the theory of the Fundamental Diagram.

Task 3: Input-Output Diagrams

In this task, you will create an Input-Output diagram for the given road segment over time over all lanes. The Input curve represents the cumulative number of vehicles that have entered the road segment of interest over time while the Output curve represents the cumulative number of vehicles that have exited the road segment over time. The vertical distance between Input and Output curves at a specific point in time represents the number of vehicles that are inside the road segment at this particular time, while the horizontal distance between the two curves represents the time that a specific vehicle spends inside the road segment.

Consider as entry gate of the road segment of interest the point $x = 230$ m from the beginning of the road segment, which we name C, and as exit gate the point $x = 280$ m from the beginning of the road segment, which we name D. Divide the time horizon into periods of $dt = 5$ sec.

a. At the end of each time step dt, calculate the cumulative number of vehicles that have passed points C and D, based on the relationship:

$$N_C(t) = N_C(t - dt) + n_C(t - dt,\ t) \quad \forall t >= dt \tag{1}$$

where, $n_C(t - dt,\ t)$ denotes the number of vehicles that pass point C between times $t - dt$ and t, and $N_C(t)$ denotes the cumulative number of vehicles that have already passed point C at time t (same for point D). The Input curve will consist of the set of points $(t,\ N_C)$ while the Output curve will consist of the set of points $(t,\ N_D)$.

Plot the two sets of points as continuous lines in a common graph in different colors. Add a legend on the graph to identify each set and provide a title, as well as names and units for the two axes.

b. Comment on the form of the I/O diagram that you created. What can you infer about the traffic conditions in this road segment?

c. Calculate the total time spent in the road segment by all vehicles during the time period considered for the I/O diagram (in veh $\cdot$ min).

d. Using the trajectory data, estimate the minimum, maximum and the mean time that vehicles spend inside this road segment. Plot the distribution of the time spent in the form of a histogram. Comment on the result.

e. Using the data that you plotted in the I/O diagram, estimate the minimum, maximum and mean number of vehicles that are present in the road segment over time. Note that you should begin your analysis at the time which the first vehicle(s) enter the road segment. Plot the distribution of the number of vehicles inside in the form of a histogram. Comment on the results.

Task 4: Shockwave Analysis

In this section, you will use the Time-Space diagram of the central lane that you created in Task 1.2 in order to study the formation of shockwaves and their characteristics.

a. By studying the Time-Space diagram, identify three of the most characteristic cases of shockwaves observed. Report the time periods that they are observed and draw lines on the diagram to highlight their formation.

b. By using graphical methods, estimate the speed of shockwave formation.

c. What is the maximum length of the queue of vehicles that is created for each case of recognised shockwave (maximum distance from the starting point)? In how much time did the tail of the queue reach this point?

d. By using graphical methods, estimate the speed of shockwave dissolution. In how much time was the shockwave completely dissolved counting from the moment that the queue starts to dissolve?

e. Based on the principles of the LWR theory that you were taught in class, what would you say about the relationship that you observe between the decreasing part of the FD that you created in Task 2 and the speed of shockwave dissolution? What was expected in theory and what do you observe in the dataset? Justify your answers.

Required Material & Administration

Students are expected to form groups of 5 and work together to complete all tasks and prepare a high-quality report. Most tasks will require the use of Python to handle the data and produce the required graphs. A base code is provided to you, which you are expected to further develop, in order to complete the required tasks.

The report should include a cover page with the title of the report, the group name (e.g. Group A, B, etc.), the names of the students and the date of submission. The report should consist of the following structural elements:

- Table of contents
- Introduction (topic, objectives, brief background on the experiment)
- One chapter per task
- Conclusions (general remarks, knowledge acquired)
- References (cite every source utilized, e.g. books, scientific articles, web pages etc.)

The report must be compiled in the form of a single .pdf file with `civil349_groupX_lab1_report.pdf` as the naming convention, where X is replaced by your group name. Your completed Jupyter Notebook, renamed `civil349_groupX_lab1_notebook.ipynb` where X is your group name, must be submitted as well; it should contain comments in the code whenever appropriate, and we should be able to run your notebook to reproduce your results.

Please save your report and notebook altogether under the name `civil349_groupX_lab1.zip`, and submit it once for the whole group through the Moodle page.

This lab assignment counts as 50% of the overall lab grade, which counts towards 30% of your total grade for the CIVIL-349 course.