

Real Data for a Drone-and-Bus Delivery System Scenario

Julien Ars

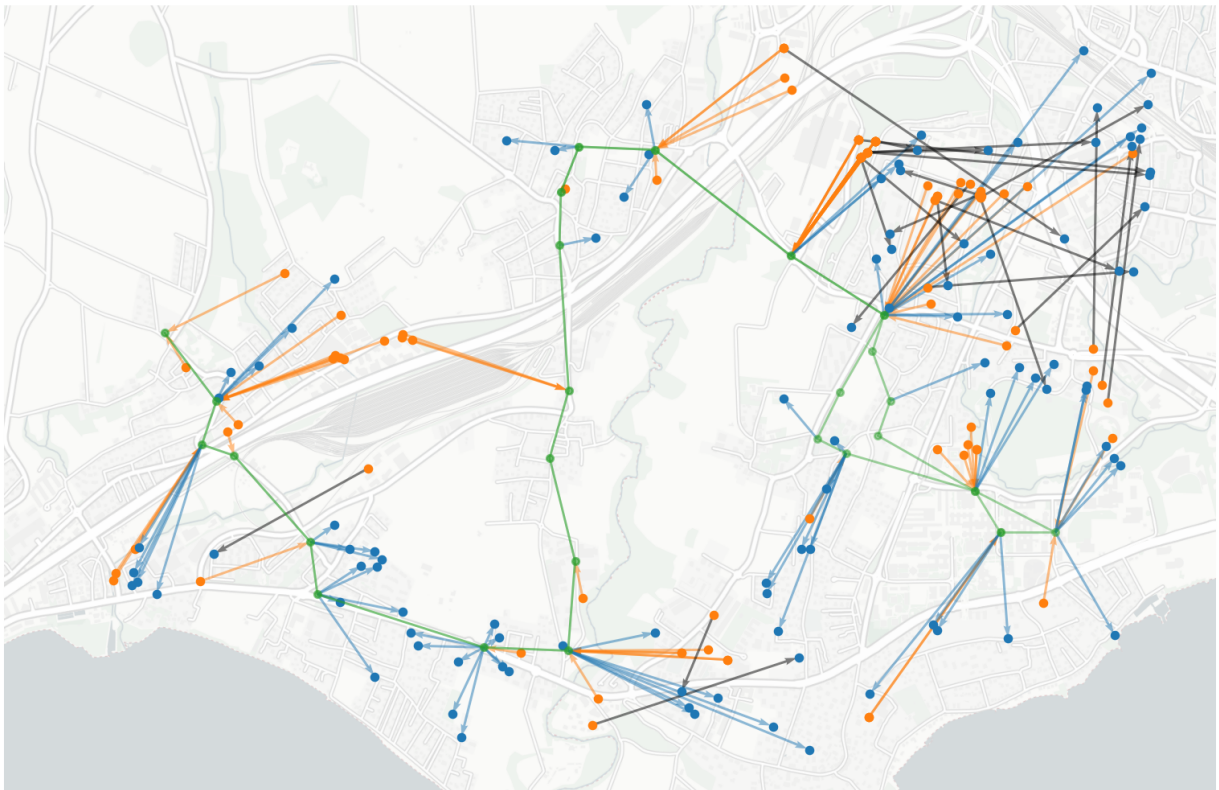


Figure 1: Bus line 705 (MBC), example result

Introduction

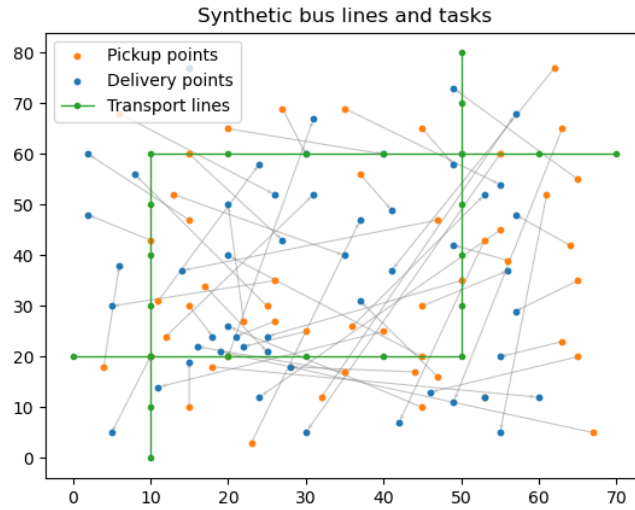


Figure 2: A synthetic bus network with synthetic drone tasks

The increasing demand for efficient urban logistics has led to the exploration of innovative delivery solutions. One such approach is the exploration of drone-based deliveries system. One interesting idea to research in that field is the integration of public transport systems to improve the performance of drone-based systems. Indeed, in many cases, the delivery system would operate alongside regular public transport service, which it could use to reduce the distance drones have to share. An example could be using a drone to transport the package from the pickup point to a bus station, put the package in a bus, then have another drone pick the package at another bus stop for the last mile delivery to the delivery point.

In researching the pertinence of such a system, one could use synthetic data like randomised tasks and geometric bus lines, an example of which is found in Figure 2. Although a good starting point, it is an interesting question to ask ourselves to what extend results obtained with this method correspond to the reality.

The goal of this project is to provide the data that will allow to solve this question, by anchoring the experiment to the real world. Using available data about public transport in Switzerland, we export bus lines trajectories as well as timetables. Then, using statistical data about the population density and the location of shops, we generate tasks that match the surrounding geography. This will enable a comparison with synthetic data and provide a validation framework for previously developed assumptions. Finally, we implement a basic metrix (total drone distance) to compare the impact a drone-and-bus delivery system could have (i) around a suburban bus route; (ii) in a full agglomeration and (iii) in the synthetic environment shown in Figure 2.

Contents

1	Methodology	1
1.1	Objectives	1
1.2	Data sources	1
1.3	Task generation	1
1.4	Bus data processing	2
1.5	Evaluation	3
2	Examples	3
3	Données de base	4
4	Conclusion	4
	Bibliography	5

1 Methodology

1.1 Objectives

In this project, we aim to challenge and refine prior assumptions about drone and bus environments, by introducing real-world data in two different aspects. With real bus data, it will be possible to take into account at first the location of bus stations, their mutual distance and the routes of the different bus lines. One could also imagine taking into account real travel times, variations in travel time, delays and unpredictability. By generating tasks based on statistical data, models could take into account the relationship between the location of customers and shops with the location of bus stops, as well as their concentration.

Additionally, we also had personal goals. The first one was to develop a deeper personal understanding of transport and geographical data available in Switzerland, and the second one to develop a reusable and user-friendly tool that can generate bus and task data on demand, for a given area and a given date. The methodology used was designed to stay scalable and adaptable so that it can be applied across different regions rather than being limited to a single bus line or geographical area.

1.2 Data sources

To ensure scalability of its results, the project relies on publicly available national data. The primary source for public transport data is the [Open data platform mobility Switzerland](#) managed by SBB CFF FFS, which provides comprehensive real-time and historical records of public transport movements, including bus, train, and metro schedules. In particular, we will use the [Actual data](#) dataset for bus information and the [Services points](#) dataset for stop informations. These comprehensive datasets include details such as stop locations, bus schedules, as well as actual departure and arrival time, from which actual bus routes will be reconstructed.

In addition to transport data, the study incorporates demographic and commercial activity data from the Swiss Federal Office of Statistics, a dataset known as GEOSTAT. Specifically, the STATPOP data provides population density metrics, household sizes, and demographic distributions, which help to determine where potential delivery demand is highest. The STATENT data offers insight into business distributions, particularly retail locations, enabling the identification of key shops from where packages would have to be taken. By combining transport, demographic, and commercial data, we provide insight into real-world demand patterns.

1.3 Task generation

Task generation is a crucial component of this study, as it determines the origin and destination points for package deliveries. As previously stated, we used the GEOSTAT dataset for this part. This dataset uses grid cells of 100m x 100m to aggregate the data, however, to gain in realism we introduced random offsets for each customer and shop. The method however

varies for customers and shops.

For customers, we used the STATPOP data that provides, in particular, total population within each cell. For each task generated, we will firstly select a cell within the study area, using the total population in each cell as weights. Then, we add random offsets from -50 m to 50 m to the center of the cell and consider this point to be the delivery point of the task. By using population data to generate customer points, we ensure that task distribution reflects actual residential density. Random offsets within the grid cells add variability, preventing clustering in a single point and better simulating realistic delivery demand.

For delivery origins, the STATENT dataset provides business density metrics, which are filtered by sector. As such, we selected the retail sector. Here, we use two variables per cell: the number of retail locations and the number of retail jobs in jobs equivalent. The methodology is a little different than to the customer points. We firstly will generate, inside each cell, one random point for each shop and assign him a weight equal to the average number of jobs per shop in that cell, that is the number of jobs in the cell divided by the number of shops in the cell. Then, for each task generated, we will pick a random point from this set using those weights. This point will be the pickup point of the task. This method allows to represent the reality that individual shops are likely to be the pickup point of many drones deliveries.

1.4 Bus data processing

Processing the raw bus data in a format that correspond to our study proved more challenging. Indeed, the original data is a table with a new row for every time a public transport reached a stop in Switzerland on the given day. The first step therefore involves filtering the dataset to keep only the rows concerning the bus lines we are studying. The next step consists of reconstructing bus routes based on the records. One key challenge here is handling bus lines that have multiple routes. Some bus lines follow different routes depending on the direction, others have different variants, some will desserve some stops depending of the time (for example, one could stop at a school when classes start and finish only, or some only desserve their full route on peak hour). For each individual bus, we must determine which bus stops he stops at and the order in which he goes through them. Then, we group buses that follow the same route, and name those routes. We then select the most used route and order, and use it to base our order of stops. We then add the stops that are not in this order by interpolation on the next routes. This process allows us to define a fictional **distance** variable that place all stops of a given line on a single axis, making our final results much more readable. The third step is cleaning the data. Indeed, we noticed two irregularities in the data that we wanted to correct : (i) Some bus lines had strange routes that share only one or two stops with the 'main' route. We automatically filter those routes out and remove them from our results (by applying a default threshold of 5, that is we remove routes that share less than 5 stops with the route most used), although the threshold can be adjusted by the user. (ii) Some buses seemed to have skipped one stop and then come back to it later. Assuming a technical error, we corrected the time data automatically as well, by keeping the maximum value registered (that is, if a bus

goes backwards, we change the time of the next stops and make it equal to the previous time). Finally, we export the results. To do so, we produce 4 types of data for each bus line:

1. Timetable data, that contains, for each bus, its scheduled and real arrival and departure time at each stop. This data is then further separated into scheduled times and real times, for easier readability. The output is both human- and machine-readable
2. Stops data, that contains for each stop its name and id, its position, its **distance** value and whether it is on each route
3. Route data, that contains for each route, whether each stop is a part of it.
4. Journey data, that contains for each individual bus journey, its ID, its route, number of stops and direction, and its start and end stops as well as the time at which it started and ended.

A python object that contains this data is also returned, for further use.

These step ensures that the returned dataset accurately represents real-world transit operations, capturing the real routes but also route differences, delays and variabilities accross time and space, while staying sufficiently clear that it can easily be incorporated into an existing or future simulation.

1.5 Evaluation

In order to summarly evaluate the impact of real world data into a drone-and-bus delivery simulation, we decided to focus on the total drone distance, that is the sum, for every task, of the distance from the pickup point and the delivery point to their respective nearest bus stop, comparing it with what would have been the direct distance without using the bus system. In the cases with different bus lines, we assumed that each package could only use one bus line and therefore used the line which would give him the smallest distance.

For visualisation purposes, we represented on the maps the packages for which the improvement would be negative as not using the bus network and instead going directly to the delivery point.

2 Examples

A comparative study was conducted to analyze the impact of using real data on delivery simulation metrics across three different cases. The first case utilized synthetic bus lines and tasks as a baseline scenario. The second case focused on the area surrounding Bus Line 705, while the third case analyzed task generation in the Lausanne metropolitan area. The findings indicate that real data significantly enhances the probability that bus-based deliveries reduce drone travel distances. Moreover, task distributions generated using real data more accurately

reflect real-world delivery demands, demonstrating the advantages of this approach over synthetic modeling. We now introduce 2 examples of using our code to generate real data in two different contexts :

3 Données de base

4 Conclusion

References