

# Dynamic forecasting of bus arrivals to stations

Hernández-Suárez, Carlos

*Facultad de Ciencias, Universidad de Colima*

*Colima, Colima, 28140, MEXICO*

*E-mail: [carlosmh@mac.com](mailto:carlosmh@mac.com)*

*and CIMAT- Unidad Monterrey*

*Centro de Investigación y Desarrollo en Ciencias de la Salud*

*Monterrey N.L. 66460, MEXICO*

## Introduction

Public transportation is strongly encouraged all over the world. In cities within developed countries that have managed to have exclusive bus lanes, anticipating the arrival of a bus is not a problem, since most of the time there is a schedule followed by the bus drivers, and so, inter-arrival times are more or less uniform. This can be achieved for instance, when buses belong to a public transportation system, that is, they belong to the governments and the system is regulated and maintained by public authorities.

In some countries, buses belong to private owners, and the government acts only as a regulator of fares and routes, but the administration of the system is left to the owners. This autonomy results sometimes in a competition for passengers, resulting in irregular inter-arrival times, due to variation in speed, or even to the fact that some owners decide to send their buses to cover a different route at will. A more efficient service resulted from a self-organization of bus owners, mainly due to the need of avoiding competition among them than to please passengers. Still, the quality of service is limited because most buses have to compete for space with regular traffic, and inter-arrival times are highly irregular.

Passengers would benefit from knowing how long they have to wait for the next bus to arrive for several reasons, including taking the decision of using a different transportation alternative (i.e. taxis). This knowledge also helps to administer the waiting time. In any case, it has been suggested that the knowledge of the arrival time tends to reduce stress.

The problem here is how to provide an accurate forecast even when in general, inter-arrival times are highly variable, due to the aforementioned factors.

There is plenty of literature in the field of bus-arrival forecasts. (See 1-4 and references therein) Instead of building models for bus displacement as a function of parameters as traffic, route, speed, etc, we build our forecast system based on an analog technique.

The idea of seeking analogs is natural to many long-range weather forecast systems, but transportation data is less limited than yearly weather data. The adapted key concept is this: a bus will take the same time in traveling from point A to point B if all the actual conditions repeat. Of course, since not all conditions can be repeated exactly and some conditions are more important than others, we took as the first goal to collect all factors that may be relevant to the forecast. Colima and Villa de Álvarez are two cities in west Mexico forming a metropolitan area of pop. 255,000 where the system was tested. The city has 182 buses covering 26 routes, but it has 110,500 vehicles, for one of the largest ratios of private automobiles per person, thus, bus circulation is continuously challenged.

### Building efficient forecasts

The first task was to ask bus drivers what were the regular factors that they believe were important in the variability of bus arrivals to bus stops. These were catalogued by route. The most common factor detected from the interviews was time of the day followed by day of the week. School day and Holidays were also mentioned. These factors influenced both the amount of traffic and the amount of passengers, all of which affect the speed of a bus. The second task was to classify the reported factors in categories, for instance: day of the week was divided in 7 categories: (SU,MO,TU,WE,TH,FR,SA), time of the day was divided in 15 minutes categories starting at 00:00, weather was classified in no rain, mild rain and severe rain (NR, MR, SR, School day (SY, SN) and Holiday (HY, HN) were also included.

There were historical records of AVL data, available at every two-minutes during the last six months. In addition, the route was divided in 50-m segments and the coordinates were translated to the actual segment the bus was in. The previous classification was added to every record, a typical record line would be:

[40567 07/06/11 11:43:12 19.2877 -103.7380 47 280 21A TR 33 NR SY HN]

Which is read as:

[ID dd/mm/yy hh:mm/ss Latitude Longitude Quarter Segment Route Weekday Rain\_status School\_day\_status Holiday\_status]

The position of a bus in a given day along a specific route is depicted as a position in a cylinder where the basis of the cylinder is the position of the bus (segment) (Figure 1) and its side is time (minutes). So, a bus “climbs up” the cylinder by turning around it giving a “bus signature” (Figure 1)

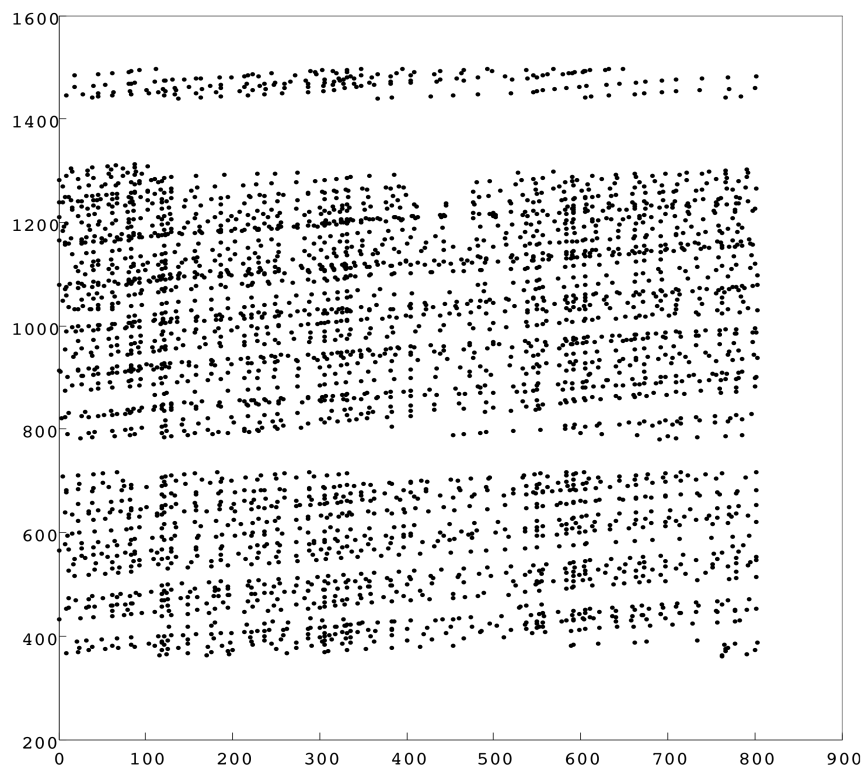


Figure 1. The signature of eleven buses during a whole day. X-axis is the segment the bus is in (1 = an arbitrary but fixed segment) and the Y axis the time in minutes since the beginning of the day.

## In advance forecasting

One of the most challenging problems in bus arrival forecasting, is dealing with the massive amount of information required to make the forecast of every bus to every station, as well as its nearly continuous update. Our approach was to anticipate the queries as follows: a given route is divided in  $S$  segments and is transited by  $N$  buses. For every quarter of the day and for every segment, we calculate using historical data, the average time it take for a bus to arrive to every one of all the following segments for each of the possible combinations of current Weekday, Rain status, School day status, and Holiday status. This is computational time consuming but the table can be constructed days or weeks beforehand. When a bus reports its position, the current condition of the bus is classified according to the actual segment, the quarter, the weekday, the Rain status, School day status, and Holiday status, and the forecast is taken from the table constructed beforehand, avoiding calculations. With this we update a dynamic table, containing the forecast of each bus to each following segment (see Table 1)

Table 1. The dynamic table

		BUS NUMBER						
		33	35	77	18	221	324	42
SEGMENT NUMBER	1	38	22	61	23	5	40	17
	2	39	27	63	30	9	46	24
	3	42	36	7	34	17	49	33
	4	48	38	8	44	25	3	43
	5	58	43	17	45	27	10	49
	6	68	53	27	8	32	17	51
	7	9	61	34	10	37	19	53
	8	19	71	42	12	44	21	56
	9	21	2	50	13	52	26	8
	10	31	12	54	14	60	36	11

The advantage of this strategy is that all calculations can be made beforehand: every time a bus updates its position, this is classified according to the reported position and other variables and the forecast is sought in a database with minor calculations. A person waiting for the bus at segment number 7, for instance, will receive as a result to its query, the numbers of the 7-th row, sorted. We can filter those arrivals exceeding some amount of time, say 30 min. In practice, only those segments that contain a bus station are included in the table.

## Results

At the end, there were no difference in the forecasts between weekdays MO-FR, and there were only three weekday categories: MO-FR, SA and SU. So far, we have an overall

10% error in the forecast with a 95% confidence level, meaning that when the forecast is 30 minutes, the arrival will occur within (27 min -33 min) with a 95% probability, and for a forecast of 10 minutes, within (9 min -11 min). Nevertheless, we lack of data for the rainy season, where the system has not been tested yet.

## REFERENCES

[1] Mei Chen, Xiaobu Liu, Jingxin Xia, Steven I.Chien, "A Dynamic Bus-Arrival Time Prediction Model Based on APC Data", Journal of Computer-Aided Civil and Infrastructure Engineering, United Kingdom, Volume 19, 2004, pg 364-376

[2] Wei-Hua Lin, Jian Zeng, "An Experimental Study On Real Time Bus Arrival Time Prediction With GPS Data", Journal of Center for Transportation Research and Department of Civil and Environmental Engineering, Virginia

[3] Jens Biesterfeld, Elyes Ennigrou, Klaus Jobmann, "Neural Networks for Location Prediction in Mobile Networks", Journal of Institute far Allgemeine Nachrichtentechnik, University Hannover, Germany

[4] Lajos KISGYORGY, Laurence R.RILETT, "Travel Time Prediction By Advanced Neural Network", Journal of Budapest University of Technology and Economics, Texas, 10 October 2001