

Articulo requetebonico que queremos que publiquen

Muchos y bonitos autores, V.M. Rivas^{a,*}, M.G. Arenas^b, P.A.
Castillo-Valdivieso^b

^a*Department of Computer Sciences, University of Jaen (Spain)*

^b*Department of Architecture and Computer Technology. CITIC. University of Granada (Spain)*

Abstract

Blah, blah, blah...

Keywords:

1. Introduction

The development of new systems of traffic conditions information and the affluence of vehicles on roads, seem to be a key point nowadays. With an increasingly informed population, provided with ubiquitous communication devices, which are commonly used by more than 90% of the population, obtaining information about the traffic in any part of the above 165000 kilometres of Spanish roads would mean an optimal management of a communication network fundamental for a high percentage of users.

In this work, a system based on bluetooth (BT) device discovery is proposed, proposing an information system of the traffic status, as well as a valuable dataset suitable to be used in a time-series forecasting process.

Both, the data-collecting and the time-series forecasting processes [REF] are necessary steps to achieve an ultimate goal: to have information about traffic flows that occur or will occur in a certain area, allowing or aiming to optimally manage motion decisions by citizens.

Therefore, from the transport management viewpoint, various issues have been found:

*Campus Las Lagunillas S/N; E23071, Jaen (Spain)

Email address: vrivas@ujaen.es (V.M. Rivas)

URL: <http://geneura.wordpress.com> (V.M. Rivas)

- A versatile, autonomous data collector, and monitoring device.
- Traffic information collecting in real time.
- The adequate processing of the data being collected,
- And finally, a system that allows sharing data and information with those who make decisions about mobility.

In order to fulfil these requirements, the BT device discovery used for this work is able to catch waves emitted by different technological components. The components can be the ones embedded in vehicles (hands-free or GPS), accessories that the users incorporate to their vehicles, as well as mobile phones, tablets or laptops. The main data being collected is the MAC address of the device BT card. This is an unique identifier for each device, so that passing vehicles can be identified. From the user's personal data protection point of view, the intrusiveness is minimum since one-way encryption (or codification) algorithms are used to hide the real MAC addresses which are stored. Moreover it is even safer from the point of view of data privacy, since this piece of information can not be attached to any given person.

The work shows how the large amount of data collected, related to detected BT devices (which pass near the device), are considered as the starting point to compute statistics, to study several indicators about the use of vehicles, and to perform time-series forecasting for the monitored area population.

For this reason, the paper is organized as follows: In Section ?? current technologies to monitor the traffic that passes through a certain area is summarized. Section ?? details the goals of this paper. In Section ??, the Intelify device is presented. In Section ?? several analysis and statistics are reported from the data obtained. Finally, we present some conclusions and future work (Section 5).

2. Background and State of the Art

This section introduces the two techniques used in this work. On the one hand, the methods developed to detect traffic flows; and on the other hand, algorithms and techniques suitable to be used for time-series prediction tasks.

2.1. Traffic detection technologies

Traffic detection technologies can be generally classified into two groups: intrusive and non-intrusive.

Intrusive detection technologies are installed on or within the roadway, requiring lane closures. Using this type of technology is inherently more hazardous and is generally more time consuming, especially for temporary traffic data collection. This technology has a number of drawbacks:

- Installation requires pavement cut.
- Improper installation decreases pavement life.
- Installation and maintenance require lane closure.
- Detection accuracy may decrease when design requires detection of a large variety of vehicle classes.
- Poor pavement condition can dramatically shorten the life span of intrusive sensors.

Non-intrusive technologies are traffic detection sensors that cause minimal disruption to normal traffic operations during installation, operation and maintenance compared to conventional detection methods. They can also be deployed more safely than conventional detection methods, since they are located adjacent to the roadway and require minimal interaction with traffic flow. Non-intrusive technologies can be classified in two big groups: active technologies (microwave radar, ultrasonic and laser radar), and passive technologies (infrared, acoustic and video image processing).

Figure 1 shows a classification of information systems according to the intrusiveness of the technology.

Current technologies most frequently used in traffic monitoring include pneumatic tubes, loop detectors, floating vehicles, and automatic recognition systems, among others.

Manual counts is the most traditional method. In this case trained observers gather traffic data that cannot be efficiently obtained through automated counts; e.g., vehicle occupancy rate, pedestrians and vehicle classifications. The most common equipments used are tally sheet, mechanical count boards, and electronic count board systems.

Passive and active infra-red sensors are based on detecting the presence, speed, and type of vehicles using the infrared energy radiating from the

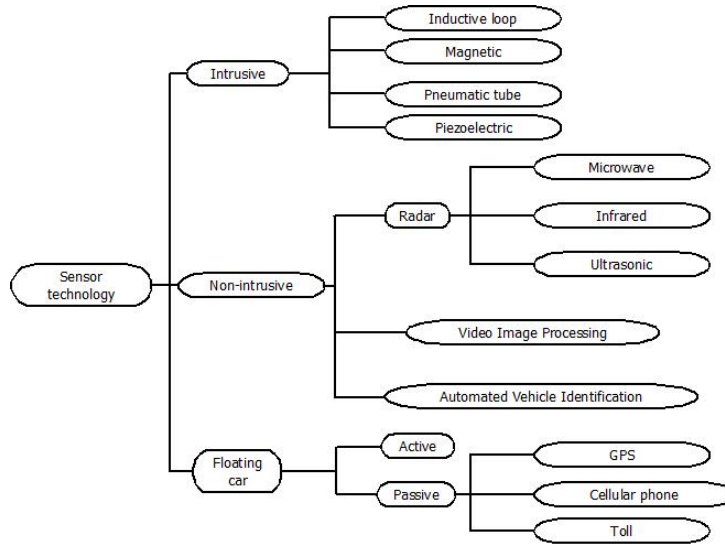


Figure 1: Information systems classification, according to the intrusiveness of the technology.

detection area. The main drawbacks are the performance during bad weather, as well as limited lane coverage.

Microwave radar can detect moving vehicles and their speed (Doppler radar). It records count data, speed, and simple vehicle classification. Its behaviour is not affected by weather conditions.

Ultrasonic sensors devices emit sound waves to detect vehicles by measuring the time for the signal to return to the device. They can be affected by temperature or bad weather.

Pneumatic road tubes are placed across the road lanes to detect the pressure changes produced when a vehicle tyre passes over them. The pulse of air that is created is recorded and processed by a counter located on the side of the road. The main drawback of this technology is that it has limited lane coverage and its efficiency is subject to weather, temperature, and traffic conditions. This system may also not be efficient in measuring low speed flows. *Piezoelectric sensors* are very similar to pneumatic road tubes, although the principle is to convert mechanical energy into electrical energy. Indeed, mechanical deformation of the piezoelectric material modifies its surface charge density leading to a potential difference that appears between the electrodes. The amplitude and frequency of the signal is directly proportional

to the degree of deformation. This system can be used to measure weight and speed.

Magnetic loops (inductive, magnetic, or video processing based) may be used temporarily or permanently, the latter being the more usual. It is the most conventional technology used to collect traffic data. The loops are embedded in roadways in a square formation that generates a magnetic field. The information is then transmitted to a counting device placed on the side of the road. This has a generally short life expectancy because it can be damaged by heavy vehicles, but is not affected by bad weather conditions. This technology has been widely deployed over the last decades. However, the implementation and maintenance costs can be expensive.

The use of so-called *floating vehicles* consists on a vehicle provided with sensors to collect information while driving on a predefined route. This active device data collection is one of the most popular among operators of roads, used especially for the collection of travel time and for loop detector calibration. Depending on the level of automation in the data collection, the cost can vary.

In some areas, such as electronic toll or transit systems, *automatic vehicle identification systems* (AVI) are also widely used. These sensors are non-exhaustive data sources to identify tags located in vehicles, such as in payment-systems without stopping. The system detects the pass, and the data is sent to the server where it is processed in order to carry out an event (pay toll, opening in the fence, etc).

The main disadvantage of these systems is that they are unable to identify vehicles detected, in order to obtain origin/destination matrixes. Just the number of vehicles and their type can be calculated, but does not allow to obtain moves flow, nor to determine whether a certain vehicle passes repeatedly. In addition, its high cost makes it unprofitable covering secondary roads with them, so they are often located on major roads.

Finally, the *automatic recognition* technology has experienced an increase in recent years due to its ability to detect individual vehicles without relying on in-vehicle systems. Video image detection is a good example: video cameras record vehicle numbers, type and speed by means of different video techniques, e.g. trip line and tracking. Furthermore, they are used for automatic detection of incidents on the road. That is the main advantage over previous information systems. However, system reliability might not be the best, as the system can be sensitive to meteorological conditions. Moreover, these systems are very costly compared to the previous ones. Finally, from

a privacy point of view, the Spanish Data Protection Agency¹ considers the car license plate as a personal data, so that it would require the user consent.

2.1.1. Commercial products

There are different companies working in the traffic information area using approaches similar to the presented in this work.

- **Bit Carrier** ??: It offers a traffic management system based in BT to count people and commercial routes (pathsolver). Its technology was implanted in highways managed by Abertis for traffic control and monitoring. Actually it has a 150 devices network in Catalonia, so it allows count the traffic times of 200.000 persons each day.
- **Trafficnow** ?: Another BT system product. A pilot experience has been implanted in Vigo.
- **Traffax Inc** ?: It is a company that also has used BT for calculating origin-destination and transport time matrixes.
- **Savari Networks** ?: It offers the commercial product StreetWAVE for traffic monitoring to know in real time the traffic status.
- **TrafficCast** ?: They have developed prediction models in different cities based on different technologies, such as cameras, BT and RFID included in the vehicles.

The proposal presented in this work have some common features with the previous approaches, offering similar functionality but with reduced cost.

2.2. Time series forecasting

Briefly described, a time series is a set of chronologically collected data. Time series forecasting is the task of predicting values of a given series using its own past and present values; the values of any related exogeneous variable can be also used, when available. Time series forecasting is a major field of research, mainly in the areas of statistics ? and operational research ?, as time series can be found in fields like Engineering, Biology, Economy, or Social Sciences, among many others.

¹Agencia de Proteccion de Datos

Time series forecasting is usually tackled trying to find out an underlying model that describes the series behaviour. For this reason, there exist a wide variety of methods to perform forecast using both linear and nonlinear models. The group of linear methods comprises the exponential smoothing methods ??, simple exponential smoothing (SES), or State space models ?, among many others. Nevertheless, the most well-known, widely used linear methods are the ARIMA ones ?. ARIMA methods can be summarized as iterative cycles in which: a) the time series is classified as belonging to a pre-established class; b) according to that class, a set of parameters is estimated; and c) the obtained model is verified in order to accept it, or search for another one returning to first step. The models provided by ARIMA integrate in their equations autoregressive (AR) and moving average (MA) components to have into account past values as well as previous forecasting errors.

Despite their ease of use, linear models have shown not to be accurate for many real applications, being this the main reason why new (but also more complex and difficult to be used) nonlinear methods have been developed. Nonlinear models include regime-switching models, which comprise the wide variety of existing threshold autoregressive (TAR) models ?. Clements ? exposed main drawbacks of nonlinear methods, mainly the excessively complex models they provide, the lack of robust performance, and, worst of all, the difficulty to use. De Gooijer ? also concludes that future research on nonlinear models should include, among others, the search for easy to use software.

Forecasting values for time series has been also faced with soft computing approaches; for instance, the ones by Samanta ? and Zhu ?, who developed methods based on cooperative particle swarm optimization. In ? and ? fuzzy time series models are proposed in order to predict new values. Similarly, Yu and Huarng ? applied ANNs for training and forecasting in their fuzzy time series model. Models such as support vector regression ? and fuzzy expert system ? were proposed for the electricity demand forecasting, among others.

As can be seen, ANNs have been applied to time series and they are currently recognized as an important tool for forecasting.

Tang ? concluded that ANNs could provide better long-term forecasting; moreover, ANNs can perform better than ARIMA models with short series of input data. Furthermore, contrary to the traditional linear and nonlinear time series models, ANNs are nonlinear data-driven approaches with more flexibility and effectiveness in modeling for forecasting ?. Jain and Kumar determined in their work ? that the ANN models were able to produce more

accurate forecasts than traditional models because they do not presuppose any functional form of the model to be developed and they do not depend on the assumptions of linearity.

There exist numerous works of different application areas where ANNs are used to forecast time series. The work by Arizmendi [?] obtained accurate predictions of the airborne pollen concentrations using ANNs. Zhang and Hu [?] employed ANNs, and Rivas et al. [?] RBFNs, for forecasting British pound and US dollar exchange rates. Bezerianos et al. [?] employed RBFNs for the assessment and prediction of the heart rate variability.

Specifically inside the ANNs, the use of RBFs as activation functions for them and its application to time series forecasting were firstly considered by Broomhead and Lowe in 1988 [?]. Afterwards, new works by Carse and Fogarty [?], and Whitehead and Choate [?] focused on the prediction of time series.

In later works, Harpham and Dawson [?] studied the effect of different basis functions on an RBFN for time series prediction. Moreover, Du [?] used time series with an encoding scheme for training RBFNs by GAs. Both the architecture (numbers and selections of nodes and inputs) and the parameters (centers and widths) of the RBFNs were represented in one chromosome and evolved simultaneously by GAs so that the selection of nodes and inputs could be automatically achieved.

Previous works found in literature can also be classified according to the prediction horizon. Thus, forecasting can be divided into short-term, medium-term, and long-term. Generally, forecasting is trended to short-term prediction such as one-step ahead prediction, since longer period prediction (medium-term or long-term) is more difficult, and sometimes may not be reliant because of the error propagation [?]. Thus, neural network models have been traditionally applied in short-term forecasting [?]. For instance, the work by Perez-Godoy [?] applied a hybrid evolutionary cooperative-competitive algorithm for the design of RBFNs to the short-term and even medium-term forecasting of the extra-virgin olive oil price.

As mentioned, another problem that emerges working with time series is the correct choice of the lags considered for representing the series. The relationship involving time series historical data defines a d -dimensional space where d is the minimum dimension capable of representing such a relationship. Takens' theorem [?] establishes that if d is sufficiently large is possible to build a state space using the correct time lags and if this space is correctly rebuilt also guarantees that the dynamics of this space is topologically

identical to the dynamics of the real systems state space.

In order to tackle the lags selection problem, an evolutionary method that performs a search for the minimum number of dimensions, Time-delay Added Evolutionary Forecasting (TAEF), is presented in ?. The methodology is inspired in Takens' theorem [***CITA***] and consists of an iterative hybrid model composed of an ANN combined with a genetic algorithm (GA). In ? the evolutionary selection of lags is divided into two stages: first, the optimal dimension of the reconstructed phase space is determined by the false nearest neighbors algorithm and then a near-optimal set of time lags is found with a genetic algorithm for a fuzzy inference system.

There are some methods that carry out an automatic search of the relevant lags. QIEHI algorithm ?, for instance, is an evolutionary hybrid intelligent method which is composed of an ANN and a modified evolutionary algorithm to search the minimum dimension to determine the characteristic phase for time series. Another hybrid methodology composed of a modular morphological ANN with an evolutionary algorithm that searches for the best time lags is described in ?. In ? a study on the selection not only of the lags but also of the exogenous features with classical feature selection algorithms as pre-processing stage is performed. The lag selection is performed as a post-processing stage in ? with a sensitivity computation of the output to each time lag.

As can be observed, the approaches in the literature consider the lags selection as a pre- or post-processing or as a part of the learning process but, instead of together, in hybrid processes with two or three stages. On the contrary, our goal is to address the selection of the lags which represent the series (with any type of correlation) jointly with the design process.

3. Data-collecting

Identifying the MAC address of BT devices passing by the roads has being achieved using Intelify (see Figure 2), a hardware solution with a low power consumption and a high detection range.

Intelify is a small autonomous computer that can be installed in any area to be monitored. It is an autonomous unit, provided by several sensors so that it can discover what is happening in its surroundings (like the flow of people and vehicles). At the same time the environment is been scanned, Intelify is able to send the information to a central server for further processing and interpretation. Table 1 shows main features of the device.

Dimensions	113x163x30mm
LEDs	Power 3G activity Ethernet activity
Networking	Ethernet Wireless Bluetooth 3G
USB ports	City Analytics Antenna 3G dongle
Other ports	RS-232 VGA
Power	18v - 1200mA external jack - 5.5mm internal jack - 2.1mm
Network connections	Ethernet RJ45 3G USB Modem
Antennas	City Analytics USB antenna wireless antenna
Microphone	noise sensor
Temperature	main board temperature sensor with extrapolation
Box	1.5mm aluminum box external use possible
Operating System	Debian 6.0 Squeeze

Table 1: Main features of the Intelify device.



Figure 2: Intelify device with a connected USB 3G dongle.

This hardware device is based on technology developed by Ciudad 2020 ??, and the services it offers are based on a net of Intelify devices. Using such an structure, it has the capacity to discover information about the physical environment and help with decision making to any kind of organization based on people flow and behavior.

Valuable information about tourism, trade and mobility can be gathered through the deployment of autonomous devices around a city. A specific example is the service offered in ?. It offers information about foot traffic through Cordoba city center.

The cost of this solution is 1000 euros per device, including maintenance of remote computer, communications using a 3G telephony service and storage and data management.

Data accuracy is very representative, compared against other technologies. In ? it was obtained an a priori error estimation of 8.5% of detections.

??????Descripcin de los datos recopilados

4. Experiments and results

????? TODO ESTO HAY QUE CAMBIARLO PARA LOS DATOS REALES CON LSO QUE AL FINAL HEMOS TRABAJADO

In this section the analysis of collected data during the monitoring period (November 8 to December 9) to obtain statistics and so study the use of vehicles is carried out.

Specifically, the following subsections report information about the total number of vehicles detected by each node, on weekdays or holidays, infor-

Node Id.	N. of devices detected
1	31408
2	45032
3	33165
4	358494
5	297874
6	7872

Table 2: Number of BT devices detected by each node.

mation on traffic density by time range on individual movements, and the average speed on a section delimited by two consecutive nodes.

Finally, since November 14 2012 a general strike was held, we will study how the strike affected traffic in the metropolitan area of Granada (Spain), by comparing total number of devices detected that day (November 14), and the following day (November 15).

4.1. Total number of vehicles detected (weekdays and holidays)

The first analysis consisted in calculating the number of devices detected by each node.

In total, 773,845 BT devices have been detected by the six nodes. As shown in Table 2, nodes located in the Sierra Nevada Highway (A44, nodes 4 and 5) have collected a higher number of data, while the node located in a side street (node 6) has detected the smallest number of devices.

4.2. Total vehicles detected on non-working days

To compare the traffic intensity between working and non-working days, the number of pass on holidays and non-working days have been obtained.

Table 3 shows how the number of detected devices lowers by all nodes on non-working days, compared to the number of detections on weekdays. Nodes located in the Sierra Nevada Highway still collected much more data than the remainder, due to the traffic this road supports on holidays.

4.3. Traffic density on the road by time range

Traffic density can be calculated taking into account the total number of detected devices by time range.

Figure 3 shows higher density on all nodes, at peak times or out of work and school.

Node Id.	N. of devices detected
1	2149
2	2804
3	2832
4	32182
5	24166
6	1269

Table 3: Total number of BT devices detected by each node (only on non-working days).

4.4. Total detections by time range

Additionally we can calculate for each node, the number of detected devices by time range, without differentiating whether the device is the same or not (repeated passes). Thus, repeated passes of the same vehicle are counted.

As in the previous case, a greater traffic density can be observed on all nodes, at peak times or out of work and school. (see Figure 4).

4.5. Number of individual vehicles detections

We can take advantage of the proposed system’s ability to identify BT devices. Thus, it can be detected whether vehicles pass by different nodes.

Figure 5 shows a large number of vehicles that pass repeated times (up to 10 times) by some of the nodes (mainly those located in the A44). Even it can be seen that nodes 4 and 5 detect about 1,000 vehicles passing more than 25 times repeated. On the other nodes, over 25 repetitions of the same device have been detected only around 120 times.

4.6. Complexity of displacement

To study the complexity of displacements, the number of vehicles that have passed through two nodes, 3 nodes and up to 6 nodes were calculated. Table 4 also shows the average number of times that vehicles have passed through 2, 3, 4, 5 or 6 nodes.

The above information is complemented with Figure 6, that shows how many cars pass by only one node, two nodes, three nodes, etc.

As expected, most of the BT devices rarely passed by all nodes, while most of devices pass only by one or two of nodes (their displacements are focused on a small part of the monitored area).

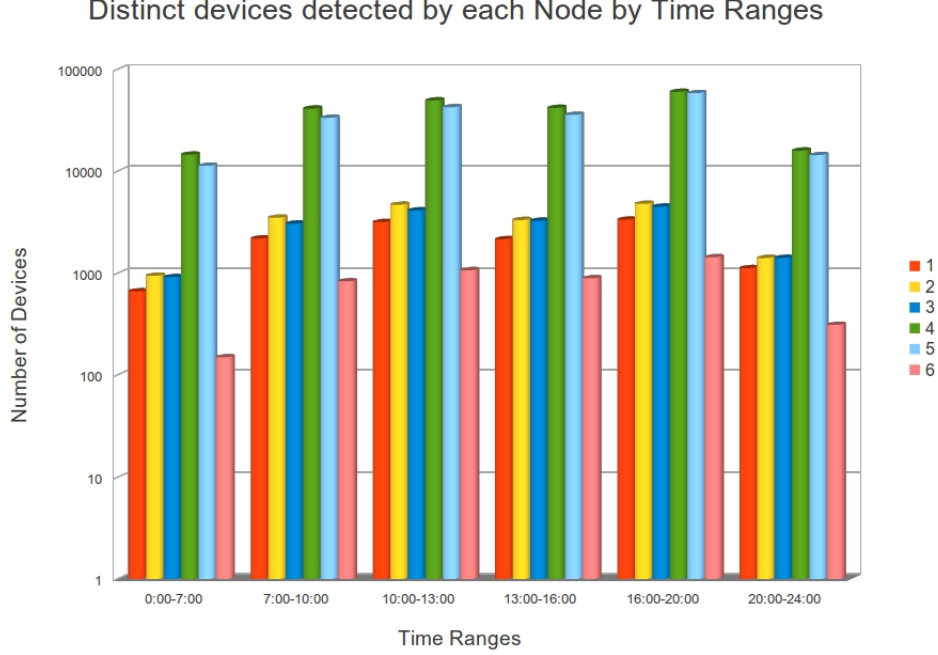


Figure 3: For each node, the total number of different detected devices by time range is shown. Figure is shown in logarithmic scale.

4.7. Effect of Nov-14 strike on zone traffic

Right in the middle of the monitoring period in Spain was held a day of general strike (November 14, 2012), which has been reflected in the number of detected devices (cars) on the nodes.

The effect of the general strike in the traffic of the monitored area has been analyzed and shown in Table 5 as the number of detected devices on November 14 and the very next day.

Table 8 shows a lowest number of detected vehicles the day of the strike that on the following day (working day in which the activity should be normal in the area).

4.8. Analysis of vehicles speed between two consecutive nodes

Finally, taking two consecutive nodes, located on the A44 highway, average speeds in the section bounded by nodes 4 (located at km 119.550) and

No. of nodes	No. of devices	Total number of passes	Mean \pm std. dev.
1	72989	165033	2.26 ± 31.16
2	53947	425667	7.89 ± 11.48
3	8125	131570	16.19 ± 24.71
4	1359	39241	28.88 ± 140.82
5	254	8603	33.87 ± 59.51
6	61	3731	61.16 ± 94.78

Table 4: Total number of vehicles that have passed through two nodes, 3 nodes and up to 6 nodes, and average number of times that vehicles have passed through 2, 3, 4, 5 or 6 nodes. In some cases the deviations are high because some devices have a very high number of occurrences for some nodes.

Node	Total number of passes (nov-14)	Total number of passes (nov-15)
2	1841	2722
3	891	1169
4	10807	16942
5	831	4017
6	946	1419

Table 5: Comparison of the number of passes for each node between the general strike day (November 14) and the very next day. Node 1 results are not reported because the hardware device suffered a power supply problem for a couple of days at that time.

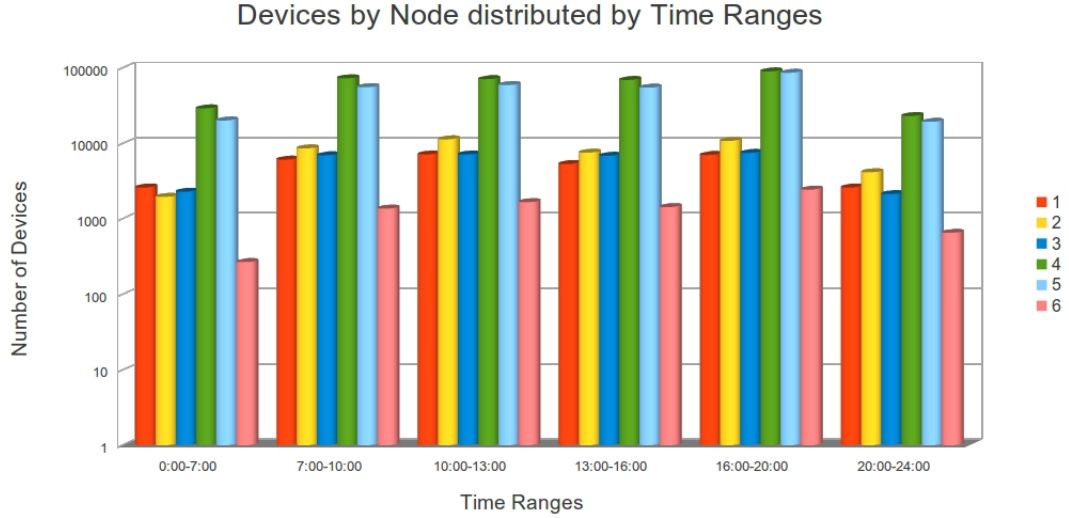


Figure 4: For each of the six nodes, the total number of detected devices by time range is shown. Figure is shown in logarithmic scale.

5 (located at km 123.250) can be calculated. This highway section where the study takes place is of 3700 meters long. Actually, we can calculate the average speed in the global section, not the speed at which each vehicle has at each instant within the section.

In that section, the speed is limited to 100 km/h. However, although most of the vehicles respect this limit, a lot of cars exceed this limitation.

4.9. Flow traffic prediction by means of time-series forecasting

As far as data is chronologically collected, methods coming from the area of time-series forecasting can be used in order to predict future behaviour of traffic flow.

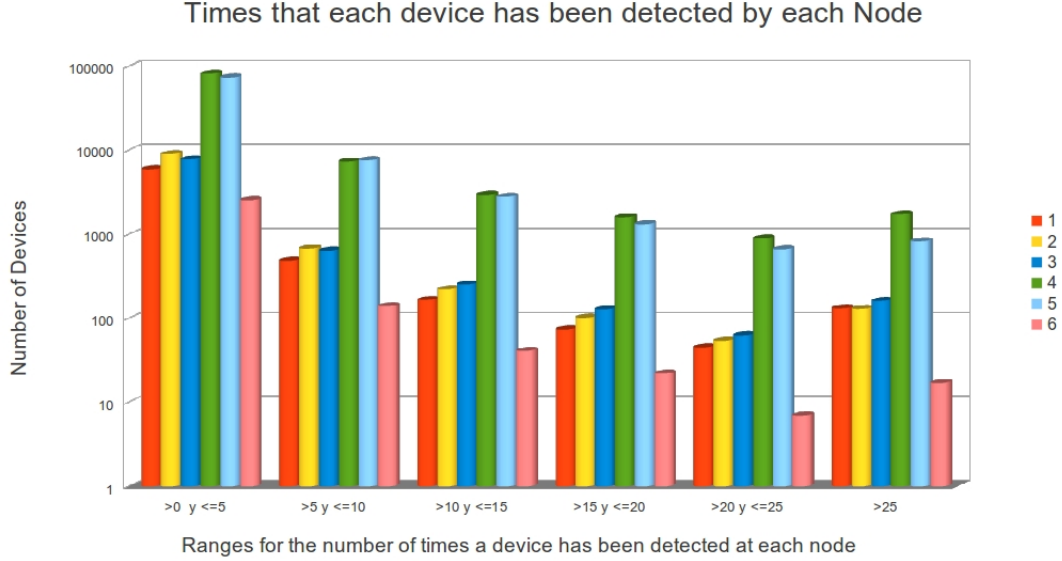


Figure 5: For each of the six nodes, the total number of detected devices N times (repeated occurrences of the same device) are shown. Figure is shown in logarithmic scale.

From the big variety of existing methods, 6 different ones have been selected in order to compare the validity of their predictions. The methods being used are the followings: ARIMA, Croston, Theta, Spline, L-Co-R, and as control algorithm, the mean value. Methods ARIMA, Croston, Theta, Spline and Mean have been executed using the R software, by means of the Forecast package (cited). In the other hand, L-co-R is a co-evolutionary algorithm of our own, based on radial basis function neural networks. In order to compare the methods, a statistical study is included allowing us to draw some conclusions.

The experimentation has been carried out using the data collected for five nodes over 62 days. The first 55 days have been used to train the methods, and obtain their associated models, and the remaining 7 days have been used to test their generalization capabilities. The prediction has been for an horizon of 1, so that known data until day n have been used to forecast day $n + 1$.

For each method and each node, a set of measures have been computed in order to determine the accuracy of the prediction. As most textbooks

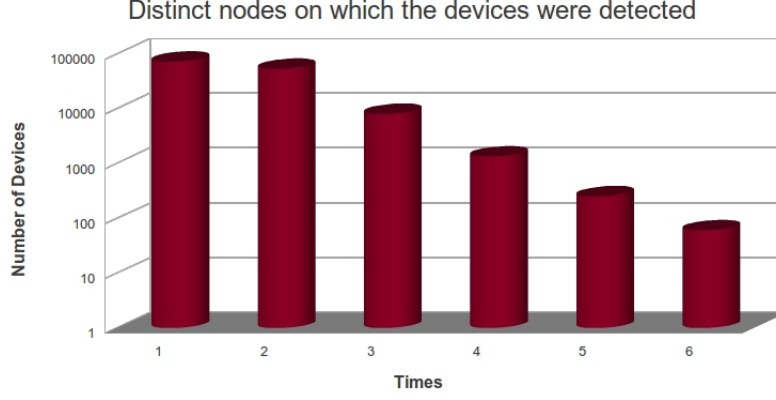


Figure 6: Figure shows how many cars pass by only one node, two nodes, three nodes, etc. Figure is shown in logarithmic scale.

recommend, one of the chosen measures is the Mean Absolute Percentage Error (MAPE) ?, which was used in the first M-competition ?. Median Absolute Percentage Error (MdAPE) ?? is also used, as a normalized version of the previous one. More recently, the description of many others measures can be found in ? and ?. Among all of them, in this work we use the followings:

- Mean Absolute Error (MAE):

$$MAE = mean(| e_t |) \quad (1)$$

- Mean Absolute Percentage Error (MAPE):

$$MAPE = mean(| p_t |) \quad (2)$$

- Median Absolute Percentage Error (MdAPE):

$$MdAPE = median(| p_t |) \quad (3)$$

- Mean Absolute Scaled Error (MASE):

$$MASE = mean(| q_t |) \quad (4)$$

Speed range (km/h)	No. of passes
$v \leq 60.0$	1495
$60.0 \leq v \leq 70.0$	2585
$70.0 \leq v \leq 80.0$	7421
$80.0 \leq v \leq 90.0$	16339
$90.0 \leq v \leq 100.0$	20144
$100.0 \leq v \leq 120.0$	14384
$120.0 \leq v \leq 140.0$	5434
$v \geq 140.0$	1326

Table 6: Average speeds (globally) in the section bounded by nodes 4 and 5.

- Mean Squared Error (MSE):

$$MSE = \frac{1}{n} \sum_{i=1}^n e_t^2 \quad (5)$$

where Y_t is the observation at time $t = 1, \dots, n$; F_t is the forecast of Y_t ; e_t is the forecast error (i.e. $e_t = Y_t - F_t$); $p_t = 100e_t/Y_t$ is the percentage error, and $q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$

Given the stochastic nature of evolutionary algorithms, and in order to remove biases due to random generator, L-co-R has been running 30 times per node. Thus, the error attached to this algorithm correspond to the average error values of these 30 executions.

Table 7 show the different error values for every method and node. Bold-faces highlight the lowest (best) values per algorithm, node and measure. As table shows, best algorithms tends to be L-Co-R and ARIMA, since L-Co-R obtains the lowest results for measures MAE, MAPE, MdAPE and MASE (except for node 2), but there is not a clear winner in MSE and SMAPE.

The results yielded by every method have been compared using two non-parametric test: the Friedman test and the Iman-Davenport test. Using this two test, we are able to determine whether there exist significant differences between the methods or not, and we can establish a ranking of methods according to their ability to perform forecast. These test have been carried out

Node 1						
	<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
<i>ARIMA</i>	514, 83	40, 42	1, 14	30, 76	424860, 5	36, 83
<i>CROSTON</i>	509, 91	39, 28	1, 13	33, 8	387403, 9	36, 46
<i>THETA</i>	528, 59	41, 16	1, 17	32, 78	443898, 5	37, 87
<i>SPLINE</i>	550, 02	43, 87	1, 21	30, 1	488504, 9	39, 04
<i>MEAN</i>	456, 95	38, 57	1, 01	26, 11	318623, 5	32, 48
<i>L-Co-R</i>	372,77	23,66	0,81	18,46	231731,8	28,87
Node 2						
	<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
<i>ARIMA</i>	433, 52	26, 64	1,03	22, 38	310105,3	25, 15
<i>CROSTON</i>	532, 87	27, 98	1, 26	27, 89	398979, 9	30, 5
<i>THETA</i>	435, 58	26, 77	1, 03	22, 55	312496, 1	25, 28
<i>SPLINE</i>	535, 86	34, 53	1, 27	24, 7	447557, 4	31, 17
<i>MEAN</i>	489, 57	24, 44	1, 16	28, 64	361070, 5	27, 52
<i>L-Co-R</i>	429,78	11,41	1, 04	11,98	311561, 7	23,5
Node 3						
	<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
<i>ARIMA</i>	373, 17	35, 11	1, 32	27, 54	150293,6	33,45
<i>CROSTON</i>	377, 42	37, 03	1, 34	27, 89	160970, 9	33, 78
<i>THETA</i>	394, 95	38, 11	1, 4	29, 41	175710, 7	35, 34
<i>SPLINE</i>	384, 68	38, 86	1, 36	29, 36	184807, 2	34, 36
<i>MEAN</i>	373, 18	35, 11	1, 32	27, 53	150307, 8	33, 45
<i>L-Co-R</i>	363,3	10,56	0,84	17,01	182469, 9	35, 32
Node 4						
	<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
<i>ARIMA</i>	3168, 36	40, 62	1, 04	18, 99	21792093	37, 44
<i>CROSTON</i>	3789, 4	56, 83	1, 25	54, 33	25037291	49, 12
<i>THETA</i>	3085, 3	38, 14	1, 01	25, 42	21510757	37, 48
<i>SPLINE</i>	3758, 06	39, 45	1, 24	43, 53	28745309	48, 79
<i>MEAN</i>	4545, 17	92, 73	1, 5	40, 07	26612429	55, 57
<i>L-Co-R</i>	540,59	6,26	0,07	4,14	1120317,2	8,17
Node 5						
	<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
<i>ARIMA</i>	5012, 95	43, 8	1, 13	20, 19	29974781	33,85
<i>CROSTON</i>	5583, 14	46, 32	1, 26	26, 12	35082013	37, 51
<i>THETA</i>	5779, 37	45, 78	1, 3	28, 25	35859271	38, 94
<i>SPLINE</i>	5878, 35	46, 26	1, 32	28, 69	37288518	39, 64
<i>MEAN</i>	5013, 48	43, 81	1, 13	20, 19	29980578	33, 85
<i>L-Co-R</i>	4978,08	24,59	0,87	9,37	35875329	35, 8

Table 7: Values for the six error measures (MAE, MAPE, MASE, MdAPE, MSE, and SMAPE) yielded for any of the six methods (ARIMA, Croston, Theta, Spline, Mean and L-Co-R) when forecasting time series of nodes 1 to 5. Boldfaces stress best (lowest) values.

<i>Error</i>	<i>Friedman P-value</i>	<i>Iman-Davenport P-value</i>
<i>MAE</i>	0.0052	0.0003
<i>MAPE</i>	0.0080	0.0008
<i>MASE</i>	0.0141	0.0030
<i>MdAPE</i>	0.0063	0.0005
<i>MSE</i>	0.0213	0.0066
<i>SMAPE</i>	0.0440	0.0233

Table 8: *P-value* yielded by Friedman and Iman-Davenport non-parametrics tests for every error measure. A *P-value* lower than 0.05 (all of them in this case) indicates that there exist significant differences between the errors computed for every method.

using ??? SOFTRWAE DE SALVA. Both, Friedman and Iman-Davenport tests, reveal that results can be considered significantly different when the computed *P-value* is less than 0.05. Table 8 shows that any of the *P-value* computed by the tests are under 0.05, and this occurs in any of the six error measures being considered.

Futhermore, Friedman test results also provide a mechanism to establish a ranking between the methods, determining which one can be considered the best. Once the method has been found out, it can be compared against the rest methods using the Holm’s procedure (???CITA). This method computes the probability that two methods can be considered to have a similar behaviour, being this the null hypothesis. The null hypothesis is rejected when this probability is lower than 0.05, so that methods can be considered to be significantly different. For instance, table 9 shows the probabilities computed by the Holm’s procedure for error MAE; L-Co-R does not appear in the table since it results to be the best method with respect to this error, and it is taken as control algorithm. According to the values shown, there are not significant differences between L-Co-R and ARIMA, but there exist between L-Co-R and the rest of methods.

Holm’s procedure has been applied to any of the six error measure being considered. Table ?? summarizes the rankings of methods, as well as the existence or absence of significant differences. Thus, this statistic study shows that L-Co-R is the best algorithm in four of six error measures; nevertheless, it can not be said to be better than ARIMA in any case but, for most measures, it overcomes the rest of methods.

5. Conclusions and Future Work

<i>Ranking</i>	<i>Method</i>	<i>p</i>	<i>Null Hypothesis</i>
1	ARIMA	0.176	Accepted
2	MEAN	0.028	Rejected
3	THETA	0.007	Rejected
4	CROSTON	0.007	Rejected
5	SPLINE	2.003E-4	Rejected

Table 9: Results of the Holm’s procedure applied to error MAE. L-co-R does not appear in the table as it is the best, and consequently, control algorithm. Values of p under 0.05 lead to reject the null hypothesis, i.e., the forecasting provided by L-Co-R and the corresponding algorithm can be considered significantly different.

<i>Error</i>	<i>Best algorithm</i>	<i>No significant differences</i>	<i>Significant differences</i>
<i>MAE</i>	(1) L-Co-R	(2) ARIMA, (3) Mean	(4) Theta, (5) Croston, (6) Spline
<i>MAPE</i>	(1) L-Co-R	(2) ARIMA, (3) Mean, (4) Theta	(5) Croston, (6) Spline
<i>MASE</i>	L-Co-R	Arima, Mean, Theta, Croston	Spline
<i>MdAPE</i>	L-Co-R	Arima, Mean	theta, Spline, Croston
<i>MSE</i>	(1) ARIMA	(2) L-Co-R, (3) MEAN, (4) Croston, (5) Theta	(6) Spline
<i>SMAPE</i>	(1) ARIMA	(2) L-Co-R, (3) MEAN, (4) Croston, (5) Theta	(6) Spline

Table 10: ???COSAS.

<i>MAE</i>	<i>MAPE</i>	<i>MASE</i>	<i>MdAPE</i>	<i>MSE</i>	<i>SMAPE</i>
L-Co-R	L-Co-R	L-Co-R	L-Co-R	ARIMA	ARIMA
No diff.	L-Co-R	L-Co-R	L-Co-R	ARIMA	ARIMA
ARIMA	ARIMA	ARIMA	ARIMA	L-Co-R	L-Co-R
Mean	Mean	Mean	Mean	Mean	Mean
Theta	Theta	Theta	Theta	Croston	Croston
Croston	Croston	Croston	Spline	Theta	Theta
Spline	Spline	Spline	Spline	Croston	Spline

Table 11: ???COSAS.