


Human Mobility Prediction with Region-based Flows and Road Traffic Data


Fernando Terroso-Saenz

(Universidad Católica de Murcia (UCAM), Murcia, Spain

 <https://orcid.org/0000-0002-1921-1137>, fterroso@ucam.edu)

Andres Muñoz

(University of Cadiz, Cadiz, Spain

 <https://orcid.org/0000-0002-8491-4592>, andres.munoz@uca.es)

Abstract: Predicting human mobility is a key element in the development of intelligent transport systems. Current digital technologies enable capturing a wealth of data on mobility flows between geographic areas, which are then used to train machine learning models to predict these flows. However, most works have only considered a single data source for building these models or different sources but covering the same spatial area. In this paper we propose to augment a macro open-data mobility study based on cellular phones with data from a road traffic sensor located within a specific motorway of one of the mobility areas in the study. The results show that models trained with the fusion of both types of data, especially long short-term memory (LSTM) and Gated Recurrent Unit (GRU) neural networks, provide a more reliable prediction than models based only on the open data source. These results show that it is possible to predict the traffic entering a particular city in the next 30 minutes with an absolute error less than 10%. Thus, this work is a further step towards improving the prediction of human mobility in interurban areas by fusing open data with data from IoT systems.

Keywords: Human mobility, Machine Learning, open data, road traffic, inductive loop sensor.

Categories: H.3.1, H.3.2, H.3.3, H.3.7, H.5.1

DOI: 10.3897/jucs.94514

1 Introduction

At the dawn of the digital transformation era, modern societies now face the fact that most of the habitual objects and artefacts that we wear or use on a daily basis, from watches to vehicles, are equipped with location-enabling technologies like GPS, Wi-Fi or Bluetooth, capable of locating these objects in physical locations in the real world.

As a side effect of this development, the *human-mobility mining* discipline has emerged within the Artificial Intelligence field in an attempt to extract meaningful knowledge about human movement behaviours at different scales [Solmaz and Turgut, 2019]. One of the most relevant findings in this discipline is that human mobility is quite predictable at some extent [Guo et al., 2020]. As a result, the prediction about people's displacements is an instrumental tool in domains like healthcare [Xi et al., 2020], urban services [Kałużny and Filipowska, 2018] and transportation management [Castrogiovanni et al., 2020].

An important factor in developing these forecasting methods is the location data that will feed the system. Thus, it is possible to find related proposals based on GPS

traces [Liu et al., 2021], Call Detail Records (CDRs) [Batra et al., 2018] or Online Social Network (OSN) posts [Ullah et al., 2020] covering different temporal and spatial scenarios. These solutions basically rely on the raw spatio-temporal trajectories generated by different moving *objects* like taxis or individuals reporting their current locations with high frequency. Nonetheless, this type of high-quality individual-based location data are, in most occasions, rather inaccessible due to some privacy and economic policies defined by data providers and operators [von Mörner, 2017].

On the other hand, public institutions at different levels have promoted the release of open data for the research community (e.g. see the European Commission's principle *As open as possible, as closed as necessary*¹). This approach has given rise to an increasing number of human-mobility datasets [Chan et al., 2020, Barlacchi et al., 2015, Chang et al., 2020]. However, such open availability comes at the cost of data filtering and aggregation stages before the dataset is released to comply with several restrictions. As a result, the development of predictors leveraging such coarse-grained mobility data is still scarce in the mobility mining domain.

In addition, the IoT paradigm has played a key role in the actual realization of solutions in the field of intelligent transportation [Germanaitė et al., 2020, Chen et al., 2022]. The deployment of different roadside sensors to measure numerous road-traffic features is now a reality in modern cities [Nellore and Hancke, 2016]. Several works now propose spatio-temporal predictors to anticipate traffic volume in a particular region or road ecosystem from these infrastructure-based sensors [Frez et al., 2019, Lv et al., 2021].

In this context, this paper proposes a mechanism to *predict human mobility in an urban area* utilising two different types of human flow data. Firstly, we use a region-based open mobility dataset that defines human flows at a large spatial scale, rather than using the more typical high-resolution mobility datasets. Secondly, this proposal also incorporates the traffic volume of a main road close to the urban area of interest through an *exogenous* input provided by an inductive loop sensor. The rationale for using this second source is based on studies that indicate that vehicles remain the most important means of transport in most urban settlements [Organisation Internationale des Constructeurs d'Automobiles, 2020].

Therefore, the key novelty of our work is that we compose a human mobility predictor integrating an exogenous variable that operates at a quite different spatial scale than the primary one. While the region-based mobility data defines the mobility of the target urban region at a macro-scale level, the data reported by the traffic sensor focus on a particular section of the road infrastructure of such a region. In that sense, some solutions for mobility prediction have already been proposed for each of these two types of data separately, but not in a combined manner [Nagy and Simon, 2018]. To the best of the authors' knowledge, this is one of the first attempts to combine such heterogeneous mobility sources to compose a functional predictive service.

The contribution of this paper is threefold: First, by using human movement data in an aggregated manner, it is possible to accurately capture human flows in regions that do not have a dedicated sensing infrastructure for this purpose. Moreover, as a side effect, it can be regarded as a cost-effective mechanism, since it reuses data from already deployed sensors and open data sources. Finally, the anonymization of the data thanks to the aggregation process will also relieve the privacy concerns among end-users.

The remainder of the paper is structured as follows. Section 2 reviews existing trends

¹ https://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm

in predicting human mobility and the use of open and sensor data for this purpose. Then, section 3 describes the use case in which our solution has been deployed. In section 4, the proposed predictor is described and evaluated. Lastly, section 5 summarizes the main conclusions and potential future research lines motivated by this work.

2 Related Work

This section provides an overview on the usage of road traffic data for building human mobility predictors.

Many works have already used different roadside sensors as primary data sources for developing a palette of road-traffic prediction frameworks. For example, dual-loop sensor traffic data from freeways in California and Portland (USA) were used to evaluate a modification of the space–time autoregressive integrated moving average (STARIMA) model [Duan et al., 2019]. Besides, roadside cameras of the freeway network of Los Angeles city (USA) have been also used to develop a macroscopic traffic flow model able to anticipate potential traffic jams [Chu et al., 2019]. Magnetic sensors have been also used to estimate not only the speed but also the category of vehicles in a motorway [Feng et al., 2022]. In addition to that, Bluetooth sensors have been deployed in certain parts of urban road networks to perform forecasting tasks by training Artificial Neural Networks (ANNs) to estimate the next intersection that a vehicle is going to visit based on its previously visited ones [Choi et al., 2019]. A similar approach has been proposed in [Ikidid et al., 2021] but, in this case, by means of fuzzy logic and a multi-agent system.

In a different approach, mobile location sensors embedded in several types of vehicles or personal devices have become an instrumental datasource so as to develop forecasting frameworks in the human mobility field. For instance, GPS trajectories extracted from a bikes and taxis are used in [Wang et al., 2020] to forecast urban traffic conditions in peak hours using external factors such as weather conditions by means of a bidirectional Long short-term memory (LSTM) network. The same architecture is used in [Zhao et al., 2018] to develop a destination predictor based on high-resolution GPS trajectories that has been tested with datasets from two Chinese cities, Beijing and Chengdu. In addition to that, the framework in [Lee et al., 2021] aggregates GPS data from cabs in Seoul (South Korea) as input to a feed-forward ANN to predict traffic speed in 170 segments of the city’s road architecture. Finally, semantically-enriched GPS trajectories are also used in [Zhao et al., 2019] to estimate the vehicle speed in a motorway network within the Chinese province of Jiangxi.

Other works have fused GPS trajectories with geo-tagged documents coming from Online Social Networks. An example of this fusion is proposed in [Miyazawa et al., 2020] where OSNs and aggregated GPS trajectories are use together to provide predictions about individual movements in the city of Tokyo. The authors fed an LSTM network with the embeddings generated from the spatio-temporal and textual data extracted from the OSN and GPS traces.

Apart from raw GPS data, Location-Based services (LBS), usually installed in personal devices, have become another interesting mobile data source for human mobility exploration. As a matter of fact, the work in [Kong and Wu, 2018] made use of a spatio-temporal LSTM model (ST-LSTM) to predict the next area of interest visited by a particular user based on his/her previous trace of visits to other city’s areas. The proposal was successfully tested by using location data extracted from an LBS in Beijing.

At a broader scope, the fusion of infrastructure and fine-grained mobile traffic data has been also extensively considered in the literature. For instance, a Graph Neural Network

(GNN) has been proposed in [Zhao et al., 2020] to predict the traffic conditions of the cities of Shenzhen and Los Angeles through latent input graphs extracted from inductive loop detectors and aggregated GPS trajectories. The same type of data sources are used in the proposal introduced in [Cui et al., 2020] to also predict the traffic conditions in the city of Seattle by considering the free-flow traffic dependencies among road networks.

Finally, region-based human flows have also been evaluated as a suitable source to anticipate human movements. In the case of the work in [Terroso-Sáenz and Muñoz, 2021], authors proposed a GNN to predict the nation-wide mobility of Spain by modelling the latent connections among regions by means of the gravity model. Such system was feed with an open mobility dataset covering the whole Spanish territory. The same dataset is used in [Terroso-Sáenz et al., 2021] to develop a region-based urban mobility predictor. In this case, authors also made use the water consumption of a residential area as the exogenous variable to improve the accuracy of the proposal.

All in all, it is observed that fine-grained movement data have been extensively used and fused with other sources in the road-traffic analysis field. However, the usage of more coarse-grained human mobility data and their fusion with other complementary sources have not achieved the same level of development. While we have reviewed some works that merge region-based flows with exogenous variables defined at different spatial scales [Terroso-Sáenz et al., 2021], it is worth noting that such a work relies on a latent human behaviour (the presence of people at home), that is completely different to the one used in our proposal, namely the count of vehicles moving across the target area.

3 Setting Overview

The feasibility of our proposal has been evaluated in a geographical area comprising three cities in the center of Spain, namely, Valladolid (VAL), Palencia (PAL) and Burgos (BUR). In 2021, these cities had a population of 297,775, 77,090 and 174,051 people, respectively. Fig. 1 depicts the boundaries of this area whose bounding box is defined by the latitude-longitude coordinates $((42.40, -4, 70), (41.56, -3, 56))$.

The three cities are interconnected by the A-62 national motorway (see Fig. 1). Table 1 shows the distance matrix of the cities. Last, for this spatial area it is extracted the two target datasets, namely the human mobility open dataset and the traffic data from the sensor located in the motorway. Both of them cover the same 2-week period from January, 1st to January, 14th, 2021.

	VAL	PAL	BUR	Sensor
VAL	-	52,9 (45m)	134 (90m)	30 (28m)
PAL	52,9 (45m)	-	92 (60m)	26 (19m)
BUR	134 (90m)	92 (60m)	-	102 (63m)
Sensor	30 (28m)	26 (19m)	102 (63m)	-

Table 1: Distances among the target cities and the sensor location in Km. In brackets the Estimated Travel Time by car in minutes according to the Google Maps Navigator.

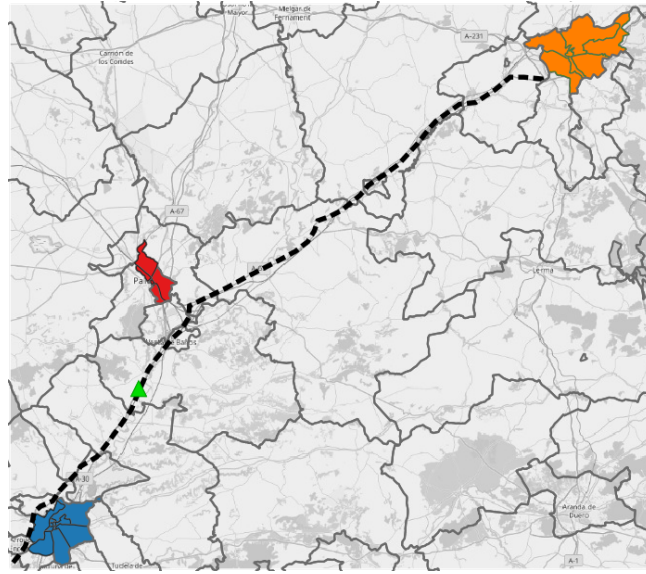


Figure 1: Geographical area under study. The three colored polygons are the three cities under consideration: Valladolid is colored in blue, Palencia in red and Burgos in orange, respectively. The grey lines define the boundaries of the mobility areas (MAs) defined in the open mobility dataset. The green triangle indicates the location of the traffic sensor and the dashed black line indicates the path of the A-62 motorway connecting the three cities.

3.1 Traffic-Volume data

The traffic volume data have been extracted from an inductive loop sensor installed in the A-62 national roadway at the 98-kilometer point. This sensor reports the number of vehicles traversing that point in each of the four lanes of the motorway (two per direction) every 15 minutes.

As a result, three different time series were composed from the extracted data, \mathcal{T}_n , \mathcal{T}_s and \mathcal{T}_{total} . The first series $\mathcal{T}_n = \langle t_1^n, t_2^n, \dots, t_{H-1}^n, t_H^n \rangle$ reports the sheer number of vehicles moving north (from VAL to PAL or BUR) every 15 minutes whereas \mathcal{T}_s just mirrors the previous one reporting the number of vehicles moving south. Lastly, \mathcal{T}_{total} reports the total number of vehicles traversing the sensor's location regardless their direction.

3.2 Human mobility data

This dataset has been retrieved from the nation-wide human mobility report released by the Spanish Ministry of Transportation (SMT) in December 2020². It covers a 15-month period from February 29th, 2020 to May 10th, 2021, indicating the number of trips among

² <https://www.mitma.es/ministerio/covid-19/evolucion-movilidad-big-data/pendata-movilidad>

3216 ad-hoc administrative areas (hereby *Mobility Areas, MA*) per hour in Spain both in its peninsular and insular extension. A *single trip* stands for the spatial displacement of an individual with distance above 500 meters. Consequently, this dataset could be regarded as a set of tuples where each one takes the form

$$\langle date, hour, m_{origin}, m_{dest}, n_{trp} \rangle$$

reporting that there is n_{trp} human trips from the MA m_{origin} to the MA m_{dest} during the indicated *date* and *hour*.

According to the official documents [Secretaría de Estado de Transportes, 2020], these mobility data was collected through Call Detail Records (CDRs) from 13 million users of an unspecified mobile-phone carrier. Once anonymised, this dataset was used to infer representative mobility statistics at the nation-level of the population of Spain and made publicly available as open data. Note that this dataset captured the movement of people regardless of the means of transport used for their displacements.

Given this dataset, we first needed to filter the trips related to the target cities. To do so, we focused on the MAs (out of the 3216 of the study) that spatially covered each of the three cities. In particular, VAL was covered by 11 MAs, PAL by 6 and BUR by 7. This gave rise to the set of MAs, \mathcal{M}_V , \mathcal{M}_P and \mathcal{M}_B respectively.

From this dataset, 12 different human flows were extracted for the MAs reflecting different human-mobility behaviours. Each flow captures a different type of human displacement for a particular combination of origin or destination MA. This extraction was done by filtering the tuples of the dataset with the selection criteria defined in Table 2.

Mobility Behaviour	Flow	m_o	m_d	Avg. num. of trips
Incoming trips	\mathcal{F}_V^i	*	\mathcal{M}_V	6,309.57 ($\pm 3,751.71$)
	\mathcal{F}_P^i	*	\mathcal{M}_P	2,022.23 ($\pm 1,213.87$)
	\mathcal{F}_B^i	*	\mathcal{M}_B	1,890.39 ($\pm 1,283.43$)
Outgoing trips	\mathcal{F}_V^o	\mathcal{M}_V	*	6,265.62 ($\pm 3,974.00$)
	\mathcal{F}_P^o	\mathcal{M}_P	*	2,023.76 ($\pm 1,192.03$)
	\mathcal{F}_B^o	\mathcal{M}_B	*	1,871.83 ($\pm 1,254.14$)
Inter-city trips	\mathcal{F}_{V-P}	\mathcal{M}_V	\mathcal{M}_P	89.68 (± 77.83)
	\mathcal{F}_{P-V}	\mathcal{M}_P	\mathcal{M}_V	91.87 (± 83.14)
	\mathcal{F}_{P-B}	\mathcal{M}_P	\mathcal{M}_B	16.11 (± 9.72)
	\mathcal{F}_{B-P}	\mathcal{M}_B	\mathcal{M}_P	14.90 (± 8.63)
	\mathcal{F}_{B-V}	\mathcal{M}_B	\mathcal{M}_V	22.91 (± 16.31)
	\mathcal{F}_{V-B}	\mathcal{M}_V	\mathcal{M}_B	22.61 (± 16.41)

Table 2: Criteria followed to extract the mobility behaviours and human flows from the SMT mobility survey. The symbol * stands for any value. The rightmost column shows the average number of trips every 15 minutes of each flow and its standard deviation.

More in detail, each flow is structured as time series with the number of trips every 15 minutes accomplishing the flow criterion. For example, according to Table 2, flow

\mathcal{F}_V^i represents the incoming flow to Valladolid and it is defined as $\langle f_{(V,1)}^i, f_{(V,2)}^i, \dots, f_{(V,H-1)}^i, f_{(V,H)}^i \rangle$ where $f_{(V,j)}^i$ is the number of human trips that arrive to any of the MAs converging in VAL (\mathcal{M}_V) from any other MA at the j-th time interval.

3.3 Correlation Study

Once we defined the human mobility flows indicated in Table 2, the next step was to study whether the traffic time series $\langle \mathcal{T}_n, \mathcal{T}_s$ and $\mathcal{T}_{total} \rangle$ described in sec. 3.1 were a suitable input to develop a forecasting method for any of such flows. To do so, a correlation study among these sources was performed by means of two different metrics, namely the Pearson’s correlation coefficient (PCC) and the Mutual Information Score (MIS). In short, PCC is a number between -1 and 1 that describes a negative or positive linear correlation, respectively. A value of zero indicates no linear correlation.

One limitation of the PCC is that it just captures the linear correlation between the variables, if any. For that reason, the MIS metric was also used as it allows measuring other types of non-linear correlations [Ross, 2014]. In brief, MIS is a non-negative score where higher values mean higher dependency. The MIS between two variables X and Y is defined as

$$MIS(X, Y) = H(X) + H(Y) - H(X, Y)$$

where H stands for the entropy, that is, the expected amount of information held in a variable. Table 3 shows the values of these scores for each human mobility flow \mathcal{F} and traffic series $\mathcal{T}_{(n,s,total)}$.

Flow	PCC(\mathcal{T}_n)	MIS(\mathcal{T}_n)	PCC(\mathcal{T}_s)	MIS(\mathcal{T}_s)	PCC(\mathcal{T}_{total})	MIS(\mathcal{T}_{total})
\mathcal{F}_V^i	0.846	0.863	0.831	0.696	0.863	0.864
\mathcal{F}_V^o	0.812	0.807	0.745	0.717	0.802	0.809
\mathcal{F}_P^i	0.856	0.897	0.821	0.783	0.863	0.901
\mathcal{F}_P^o	0.775	0.746	0.746	0.680	0.783	0.795
\mathcal{F}_B^i	0.823	0.719	0.795	0.641	0.833	0.772
\mathcal{F}_B^o	0.770	0.610	0.705	0.592	0.760	0.608
\mathcal{F}_{V-P}	0.750	0.560	0.685	0.505	0.738	0.558
\mathcal{F}_{P-V}	0.723	0.555	0.694	0.521	0.730	0.574
\mathcal{F}_{P-B}	0.383	0.181	0.341	0.242	0.369	0.242
\mathcal{F}_{B-P}	0.348	0.140	0.317	0.218	0.343	0.239
\mathcal{F}_{B-V}	0.689	0.435	0.631	0.429	0.681	0.441
\mathcal{F}_{V-B}	0.384	0.138	0.423	0.144	0.415	0.158

Table 3: Correlation values between each extracted human flow and the traffic series \mathcal{T}_θ . The highest value of each column is shown in bold.

From this table it is observed four interesting findings. First, the highest correlation for the northwards traffic flow (\mathcal{T}_n) is obtained by the incoming flows of PAL (\mathcal{F}_P^i), 0.856 PCC. This is reasonable as it is the closest city to the traffic sensor in its north side

(see Fig. 1). Therefore, the volume of traffic captured in the northwards lanes is likely to be very related to the actual incoming flows to the city of Palencia.

Secondly, the highest correlation for the southwards road traffic (\mathcal{T}_s) is achieved by the incoming flow of VAL (0.831 PCC). Again, this also makes sense as Valladolid is the closest city of the sensor in its south side.

A third interesting finding shows that the correlation between each incoming flow of the three cities (\mathcal{F}^i) is higher when considering the total traffic flow reported by the sensor (\mathcal{T}_{total}) than when relying solely on a particular direction. For example, Table 3 shows that the MIS for the incoming flow of PAL (\mathcal{F}_P^i) with the total traffic volume (0.901) is higher than with the northwards (0.897) or the southwards (0.783) traffic. It reveals a latent relationship between the sheer number of vehicles in a motorway close to a city (regardless of its direction) and the incoming human displacements of that city. This can be due to the fact that cities usually have multiple *entry points* and certain vehicles may enter the city by making a detour.

Finally, the pairwise flows in Table 3 show a lower correlation than the incoming and outgoing city-level flows. This is specially remarkable for the mobility between PAL and BUR cities (\mathcal{F}_{P-B} and \mathcal{F}_{B-P}) whose PCC is below 0.4 for the three traffic flows. This is because the road traffic between these two cities mainly occurs in the segment of the A-62 motorway that is above the location of the sensor. Besides, this correlation decrease might be also due to the existence of an inter-city railway that connects the three cities. Hence, travellers may opt to use this means of transport instead of road vehicles to move among cities. As a result, the road sensor is not able to capture this type of flow between each pair of cities.

After analyzing these results, we eventually decided to focus on providing a predictive solution for the city flow \mathcal{F}_V^i , representing the incoming flow of people to VAL, and for \mathcal{F}_{P-V} representing the trips from PAL to VAL. These are the city and inter-city flows with the highest correlation scores with the total traffic flow \mathcal{T}_{total} . Furthermore, these flows represent the mobility behaviour related to incoming and inter-city trips with the highest number of displacements (see Table 2). Hence, they represent the most challenging scenario among all the evaluated urban flows. For the sake of completeness, Fig. 2 shows the time series of both human flows along with the traffic one.

4 Designing a Mobility Predictor with Traffic Data and Region-based Flows

This section describes our approach to develop a human mobility predictor based on the time series selected in sec. 3.3.

4.1 Problem formulation

The human mobility prediction problem could be formulated following two approaches. The first one can be expressed as the following regression problem:

Given the time instant $d \in \langle 1, \dots, D \rangle$, the number of inner trips of VAL during the last d_{prev} minutes, $\mathcal{F}_V^{i,d} = \langle f_{(V,d)}^i, f_{(V,d-1)}^i, \dots, f_{(V,d-d_{prev})}^i \rangle$ and the total number of vehicles measured by the traffic sensor during the same minutes $\mathcal{T}_{total}^d = \langle t_d, t_{d-1}, \dots, t_{d-d_{prev}} \rangle$, **Find** a mapping function \mathcal{P} ,

$$\mathcal{P}(\mathcal{F}_V^{d,i}, \mathcal{T}_{total}^d) \rightarrow f_{(V,d+T)}^i$$

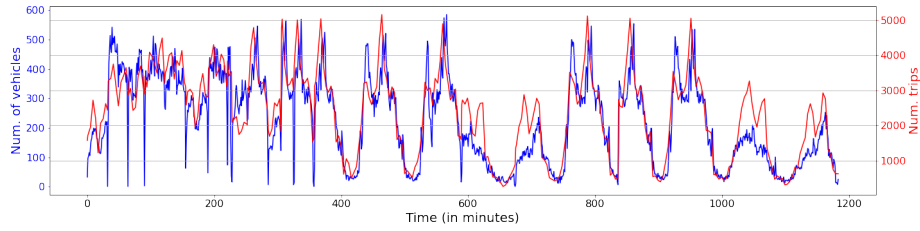
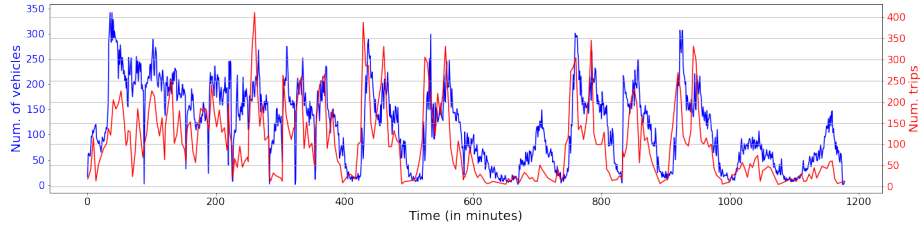
(a) \mathcal{F}_V^i (in blue) and \mathcal{T}_{total} (in red) time series.(b) \mathcal{F}_{P-V}^i (in blue) and \mathcal{T}_{total} (in red) time series.

Figure 2: Time series of the selected flows for developing the predictor.

where $f_{(V,d+T)}^i$ is the sheer number of inner trips at the $d + T$ instant in VAL being T the time horizon of the prediction ($T \geq 1$). As can be seen, the key goal of this work is to develop a predictive model enriched with traffic data instead of solely relying on the SMT data source.

The second approach to define the human mobility predictor can be formulated as the previous one by just replacing $\mathcal{F}_V^{i,d}$ with \mathcal{F}_{P-V}^d , namely the number of trips from PAL to VAL at time instant d during the last d_{prev} minutes.

4.2 Causality Test

After formally defining the target problems, we leveraged the Granger Causality Test (GCT) [Spirites et al., 2000] to determine if one time series (x) is helpful for predicting another (y). To do so, GCT tests the null hypothesis that the coefficients of past values of x in the regression equation to predict y are zero. Hence, GCT compares an *unrestricted model*, in which the time series y is explained by the lags of y and the lags of an additional series of observations from x (both lags up to a same fixed order), to a *restricted model*, in which y is only explained by the lags of y . Table 4 shows the p-value of this test for different number of lags (i.e., past values) for the time series \mathcal{F}_V^i and \mathcal{F}_{P-V} (acting as y) and \mathcal{T}_{total} (x).

As observed, all the p-values are below the significance level of 0.05, thus rejecting the null hypothesis that the coefficients of past values of \mathcal{T}_{total} in a regression equation to predict either \mathcal{F}_V^i or \mathcal{F}_{P-V} are zero. As a result, the sensor time series seems a promising exogenous variable to be integrated in the predictive model.

4.3 Candidate Models

Regarding the methods to address this regression problem, 4 candidate models have been considered: a Long short-term Memory (LSTM) and a Gated Recurrent Unit (GRU) neural network and 2 algorithms for time series forecasting, namely the Vector autoregression (VAR) and the AutoRegressive Integrated Moving Average (ARIMA) models. A brief description of each of them is provided next.

x	Num. of lags							
	1	2	3	4	5	6	7	8
\mathcal{F}_V^i	0.0000	0.0082	0.0062	0.0022	0.0049	0.0067	0.0004	0.0017
\mathcal{F}_{P-V}	0.0000	0.0076	0.0038	0.0057	0.0343	0.0357	0.0120	0.0204

Table 4: P-values of the GCT for different number of lags related to the human flows \mathcal{F}_V^i and \mathcal{F}_{P-V} and the sensor traffic \mathcal{T}_{total} .

4.3.1 LSTM model

It is a principal variant of Recurrent Neural Networks (RNN) [Smagulova and James, 2019]. In brief, LSTM models are able to learn short-term and long term patterns in sequences of data. Fig. 3a depicts a general schema of this model comprising 3 layers. Moreover, Fig. 3b shows the inner structure of a LSTM cell where $h_{(d-1)}$ indicates the short-term state at time instant (day in the present setting) d and t^d is the number of trips at time instant d . Concerning the cell outputs, $\hat{t}_{(d)} \in \mathbb{R}$ is the predicted number of trips at day d whereas $c_{(d)}$ is the long-term state that traverses the network from left to right.

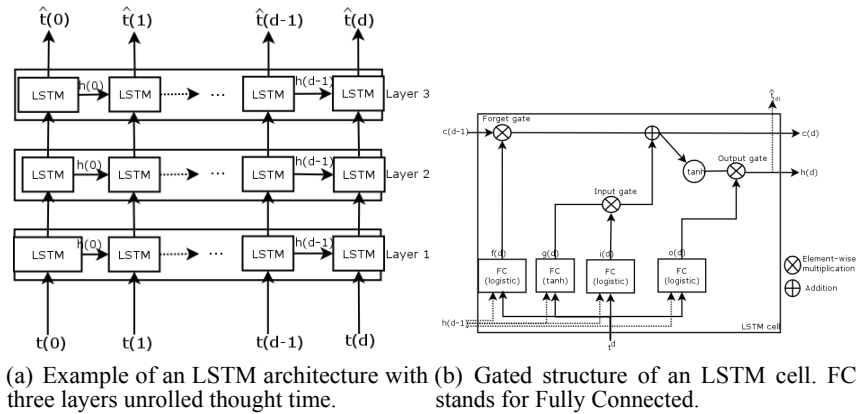


Figure 3: Structure of the LSTM network.

4.3.2 GRU model

This is a major variant of Recurrent Neural Networks (RNN) [Cho et al., 2014]. Unlike other RNN models like LSTM, a GRU model has a slightly simpler structure that makes it faster to train [Chung et al., 2014]. Fig. 4 depicts the inner structure of its cells. As observed, a GRU cell makes use of a gating mechanism to memorize long-term patterns in the target sequence. Thus, a cell receives as input the current input vector $x(t)$ and the previous state vector $h(t - 1)$. Then, the cell generates the associated output $y(t)$, which is also the state vector $h(t)$ of the next cell.

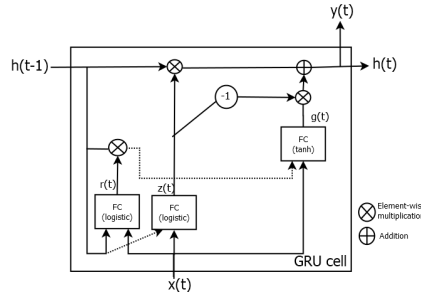


Figure 4: Gated structure of a GRU cell. FC stands for Fully Connected.

More in detail, the cell comprises three different gates, the update $z(t)$, the reset $r(t)$ and memory-content $g(t)$. The computations of each gate are as follows

$$z(t) = \sigma(W_z * x_t + U_z h_{(t-1)} + b_z) \tag{1}$$

$$r(t) = \sigma(W_r x_t + U_t h_{(t-1)} + b_t) \tag{2}$$

$$g(t) = \tanh(W_g x_t + U_g (r(t) \otimes h_{(t-1)}) + b_g) \tag{3}$$

$$y(t) = h(t) = z(t) \otimes h_{(t-1)} + (1 - z(t)) \otimes g(t) \tag{4}$$

where $W_{\{z,r,g\}}$ are the weight matrices for the input x_t , $U_{\{z,r,g\}}$ are the weight matrices for the connections to the previous short-term state $h(t - 1)$ and $b_{\{z,r,g\}}$ are the bias terms of each layer.

4.3.3 VAR model

The VAR model is a well-known statistical method to capture changes in multi-variate time series. In brief, given a k -dimensional time series $y_t = (y_{1t}, \dots, y_{kt})$, a VAR model of order p can be formulated as follows:

$$y_t = \delta + \Theta_1 y_{t-1} + \dots + \Theta_p y_{t-p} + \epsilon_t$$

where ϵ_t is a white-noise vector, $\delta = (\delta_1, \dots, \delta_k)$ is a constant vector and Θ_t is a $k \times k$ matrix. It is a very stable model that ensures the generation of stationary time series with invariant means, variance and covariance structure [Pfaff et al., 2008]. It is worth mentioning that the model must be fed with a stationary time series so as to generate reliable predictions.

4.3.4 Autoregressive Integrated Moving Average model (ARIMA)

It is one of the foremost predictor within the time series analysis domain [Shahriari et al., 2020]. In brief, it fits a regressive model to a time series to predict new values.

4.4 Configuration of the models

Table 5 shows the hyper-parameters of the LSTM and GRU models for the two prediction tasks at hand. Note that the models being fed solely with human flow data will be used for comparative purposes.

	LSTM(\mathcal{F}_V^i)	LSTM($\mathcal{F}_V^i, \mathcal{T}_{total}$)	LSTM(\mathcal{F}_{P-V})	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	GRU(\mathcal{F}_V^i)	GRU($\mathcal{F}_V^i, \mathcal{T}_{total}$)	GRU(\mathcal{F}_{P-V})	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)
Input seq. length	7 (105 min.)							
Training rate	70%							
Loss function	Mean Squared Error (MSE)							
Activation function	Hyperbolic tangent function							
Batch size	32							
Learning factor	0.001	0.01	0.001	0.01	0.001	0.01	0.001	0.01
Optimizer	Adam							
Num. of layers	3	4	3	4	4	4	4	4
Num of cells per layer	50	100	40	100	50	50	50	50
Num. of epochs	80	60	90	60	100	80	100	80

Table 5: Hyper-parameters of the LSTM and GRU models evaluated in the study. The input datasets of each model are shown in brackets.

Regarding the ARIMA and VAR models, Table 6 shows their configuration parameters. In order to set the p and q parameters, it was considered the autocorrelation and partial autocorrelation of the \mathcal{F}_V^i and \mathcal{F}_{P-V} time series. As observed in Fig. 5, there is a strong partial autocorrelation in both time series in the first two lags whereas it slightly decreases in both time series for lags 12 and 17, respectively.

	ARIMA(\mathcal{F}_V^i)	VAR($\mathcal{F}_V^i, \mathcal{T}_{total}$)	ARIMA(\mathcal{F}_{P-V})	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)
Num. of lags (p)	2	7	2	7
Mov. avg. window size (q)	17	-	12	-
Num. of differentiations (d)	0	-	0	-

Table 6: Hyper-parameters of the VAR and ARIMA models evaluated in the study. The input datasets of each model are shown in brackets.

Regarding the number of differentiations (d), it was necessary to check whether \mathcal{F}_V^i , \mathcal{F}_{P-V} and \mathcal{T}_{total} were stationary or not. To do so, we performed the Augmented Dickey-Fuller Test [Cheung and Lai, 1995]. The null hypothesis of this test states that the time series has a unit root and thus it is non-stationary. The alternative hypothesis is that the time series does not have a unit root, thus being stationary. As Table 7 shows, the statistics obtained for both raw time series indicate a strong evidence to reject the null hypothesis. For example, the statistic value for \mathcal{F}_V^i was -5.064, lower than the critical value at 5% (-2.864), so the null hypothesis was rejected.

Given the results of Table 7, the three time series proved to be stationary. Therefore, the required number of differentiations was set to 0.

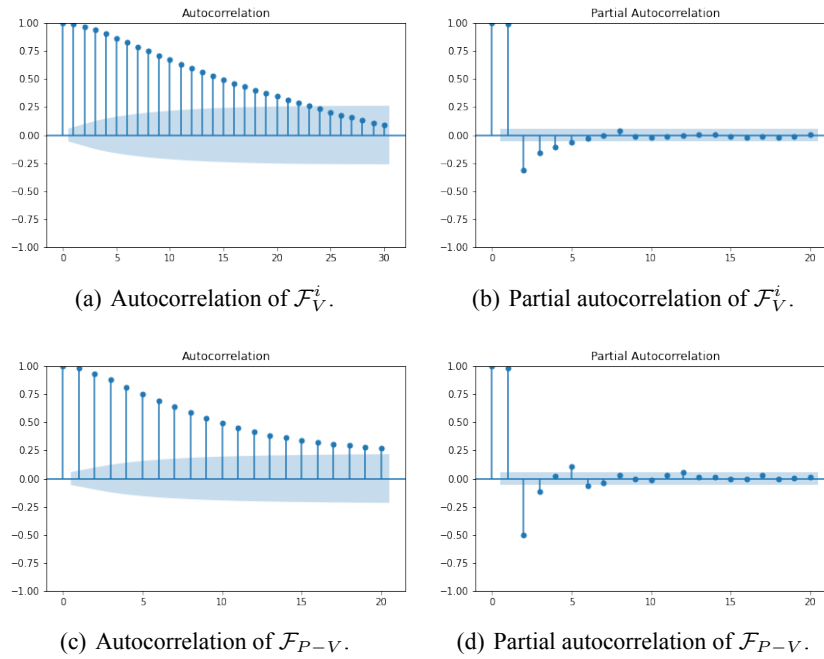


Figure 5: Autocorrelation plots of the two target human flows. The 95% confidence intervals are shown as a cone in the plots. Values outside of these cones most likely represent a correlation.

Model	\mathcal{F}_V^i	\mathcal{F}_{P-V}	\mathcal{T}_{total}
Significance level	0.05		
Test statistic	-5.064	-4.536	-4.741
Critical value 1%	-3.436	-3.436	-3.436
Critical value 5%	-2.864	-2.864	-2.864
Critical value 10%	-2.568	-2.568	-2.568

Table 7: Results of the Augmented Dickey-Fuller Test.

Finally, the parameter p of the VAR model was selected by means of a grid search based on the Akaike Information Criterion (AIC) [Hu, 2007]. According to the results in Table 8, this led to the value of 7 as the most promising one.

5 Evaluation of the Predictor

This section analyses the evaluation of the different models described in the previous section.

5.1 Evaluation metrics

Regarding the metrics to evaluate the previous models, the Mean Absolute Error (MAE), the Mean Squared Error (MSE) and the Root Mean Squared Error (RMSE) [Willmott and Matsuura, 2005] are three of the most common metrics used to measure accuracy for continuous variables. They are suitable for model comparisons as they express average model prediction error in the units of the variable of interest. Their definition is as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$RMSE = \sqrt{MSE},$$

where, for our experiment, y_i is the real number of nationwide daily trips, \hat{y}_i is the predicted number of trips at this same scale and n is the number of observations.

Num. of lags	AIC	Num. of lags	AIC
1	18.244752	8	18.064981
2	18.117103	9	18.070551
3	18.080703	10	18.075614
4	18.066120	11	18.080955
5	18.068602	12	18.084146
6	18.072590	13	18.087857
7	18.062935	14	18.088350

Table 8: AIC score of the VAR model for different number of lags (p parameter). The minimum value is shown in bold.

Furthermore, we complement these metrics with the coefficient of variance of the RMSE (CVRMSE), a non-dimensional measure calculated by dividing the RMSE of the predicted number of trips by the mean value of the actual number of trips. For example, a CVRMSE value of 5% would indicate that the mean variation in the actual number of trips which is not explained by the prediction model is 5% [Reddy et al., 1997]. Similarly, the Mean Average Prediction Error (MAPE) metric expresses the average absolute error as a percentage. These two additional metrics are calculated as follows:

$$CVRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{y}} \times 100$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100.$$

5.2 Evaluation results

In this section, we discuss and explain the main results obtained from the evaluation of the models given the inner-trips and inter-city settings.

5.2.1 Inner trips evaluation

Regarding the results obtained to predict the overall inner trips of Valladolid (VAL), Table 9 shows the scores of each metric and candidate model for different time horizons (T). As observed, the multi-variate models ($\text{VAR}(\mathcal{F}_V^i, \mathcal{T}_{total})$, $\text{LSTM}(\mathcal{F}_V^i, \mathcal{T}_{total})$ and $\text{GRU}(\mathcal{F}_V^i, \mathcal{T}_{total})$) outperformed the uni-variate models that solely relied on the human flow ($\text{ARIMA}(\mathcal{F}_V^i)$, $\text{LSTM}(\mathcal{F}_V^i)$ and $\text{GRU}(\mathcal{F}_V^i)$) for most of the time horizons and metrics. Besides, $\text{GRU}(\mathcal{F}_V^i, \mathcal{T}_{total})$ achieved the best average scores for all the metrics (see the rightmost column of the table). For example, the average CVRMSE of this model was 16.8 whereas the GRU model solely fed with open data got a 20.4 score.

More in detail, Table 9 shows that the VAR model achieved the highest accuracy for time horizons up to 30 minutes. For example, enriching the model with traffic data allowed the VAR model to get a MAPE score of 4.88 instead of 5.40 as in the uni-variate ARIMA model for a 15-minute horizon. For time horizons beyond 30 minutes, the multi-variate GRU model achieved the best scores for almost all the metrics. This is consistent with the fact that RNN models are usually more capable of detecting complex patterns within the input sequence.

Table 9 also shows that the larger improvement of the models enriched with traffic data occurred when the target prediction horizon was set over 30 minutes. For example, the reduction of the MAPE score of $\text{GRU}(\mathcal{F}_V^i, \mathcal{T}_{total})$ with respect to $\text{GRU}(\mathcal{F}_V^i)$ was 4.57% (13.99 vs 14.63), 22% (15.54 vs 18.99) and 26% (18.3 vs 25.0) for time horizons of 30, 45 and 75 minutes, respectively. However, the uni-variate version of the model outperformed the multi-variate one for $T=15\text{m}$ (11.74 vs 11.0). This is consistent with the location of the traffic sensor. Indeed, the distance between the city center of VAL and the location of the sensor is roughly 33 km. This implies an Estimated Travel Time (ETT) of 30 minutes between both locations by car (see Table 1). Consequently, it is reasonable to think that the latent information provided by the sensor for predicting inner trips to VAL will mainly cover time ranges over such a value.

Another interesting finding in Table 9 is that the uni-variate version of the LSTM model ($\text{LSTM}(\mathcal{F}_V^i)$) outperformed its GRU counterpart ($\text{GRU}(\mathcal{F}_V^i)$) for all the metrics. This is completely different from the results obtained in the multi-variate versions, in which the GRU network clearly improved the LSTM alternative. This shows the capability of the GRU network to better capture the latent mobility patterns from the multi-variate timeseries with respect to the LSTM network given the same training *corpus*. This is mainly due to its lighter architecture as we have discussed in sec. 4.3.2. On the contrary, the more complex gated structure of the LSTM allowed to better fit the uni-variate training data.

For the sake of completeness, Fig. 6 depicts the aforementioned evaluation metrics as bar plots. It clearly shows that $\text{GRU}(\mathcal{F}_V^i, \mathcal{T}_{total})$ achieved the best results in almost all the metrics for time horizons over 15 minutes.

5.2.2 Inter-urban trips evaluation

Concerning the inter-urban trips from Palencia (PAL) to VAL, Table 10 shows the scores of all the models for different time horizons. In this case, it is observed that the results

exhibited a different pattern than the one obtained for the sheer incoming trips of VAL. In particular, the multi-variate models provided a more accurate prediction for time horizons beyond 60 minutes for the LSTM model and 30 minutes for the GRU versions, respectively. For example, LSTM(\mathcal{F}_{P-V} , \mathcal{T}_{total}) obtained a CVRMSE of 56.36 which is remarkably lower than the score of its uni-variate counterpart LSTM(\mathcal{F}_{P-V}), 64.68, for a time horizon of 90 minutes.

At this point, we should recall that the sensor data involve the overall traffic volume of the motorway. Hence, this sensor captures both the traffic moving from PAL to VAL and from VAL to PAL. Besides, the LSTM and GRU models considered the traffic and human flows of the last 105 minutes as the number of lags were set to 7 (each lag involved a 15-minute period). Bearing in mind the ETT between PAL and VAL (45 minutes according to Table 1), 105 minutes is a time window large enough so that the models consider the latent go-and-return traffic patterns between both cities in the prediction loop. However, this type of patterns cover a larger period of time (90 minutes in total to go from VAL to PAL and return again). As a side effect, the improvement of the accuracy of the multi-variate LSTM and GRU models is higher as long as the prediction horizon exceeds the 90-minute threshold.

Furthermore, it is worth mentioning that the relative scores (CVRMSE and MAPE) of Table 9 and Table 10 show that the multi-variate predictors achieved much better results to estimate the \mathcal{F}_{V}^i flow than \mathcal{F}_{P-V} . As a matter of fact, the average CVRMSE of VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total}) was 22.30 and the same score for VAR(\mathcal{F}_{P-V} , \mathcal{T}_{total}) was much

Metric	Model	T (minutes)								Avg.
		15	30	45	60	75	90	105	120	
MSE	ARIMA(\mathcal{F}_{V}^i)	199,989.33	516,350.53	942,091.99	1.50×10^6	2.17×10^6	2.93×10^6	3.75×10^6	4.60×10^6	2.07×10^6
	VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total})	190,396.44	478,504.51	868,231.12	1.38×10^6	2.01×10^6	2.73×10^6	3.51×10^6	4.33×10^6	1.94×10^6
	LSTM(\mathcal{F}_{V}^i)	441,394.25	560,520.88	688,906.12	944,690.62	1.25×10^6	1.68×10^6	2.09×10^6	2.67×10^6	1.29×10^6
	LSTM(\mathcal{F}_{V}^i , \mathcal{T}_{total})	333,386.12	577,700.25	751,174.56	1.01×10^6	1.27×10^6	1.71×10^6	2.11×10^6	2.69×10^6	1.31×10^6
	GRU(\mathcal{F}_{V}^i)	391,390.66	648,561.44	872,213.44	1.16×10^6	1.52×10^6	2.01×10^6	2.46×10^6	3.23×10^6	1.54×10^6
	GRU(\mathcal{F}_{V}^i , \mathcal{T}_{total})	269,591.66	452,458.53	536,525.88	7.10×10^5	9.70×10^5	1.35×10^6	1.81×10^6	2.29×10^6	1.05×10^6
RMSE	ARIMA(\mathcal{F}_{V}^i)	447.20	718.58	970.61	1.22×10^3	1.47×10^3	1.71×10^3	1.94×10^3	2.14×10^3	1.33×10^3
	VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total})	436.34	691.74	931.79	1.17×10^3	1.41×10^3	1.65×10^3	1.87×10^3	2.07×10^3	1.28×10^3
	LSTM(\mathcal{F}_{V}^i)	664.38	748.68	830.00	971.95	1.12×10^3	1.29×10^3	1.45×10^3	1.63×10^3	1.09×10^3
	LSTM(\mathcal{F}_{V}^i , \mathcal{T}_{total})	577.40	760.07	866.70	1.00×10^3	1.13×10^3	1.31×10^3	1.45×10^3	1.64×10^3	1.09×10^3
	GRU(\mathcal{F}_{V}^i)	625.61	805.33	933.92	1.08×10^3	1.23×10^3	1.42×10^3	1.57×10^3	1.80×10^3	1.18×10^3
	GRU(\mathcal{F}_{V}^i , \mathcal{T}_{total})	519.22	672.65	732.48	8.42×10^2	9.85×10^2	1.16×10^3	1.34×10^3	1.51×10^3	9.72×10^2
CVRMSE	ARIMA(\mathcal{F}_{V}^i)	7.78	12.49	16.88	2.13×10^1	2.56×10^1	2.98×10^1	3.37×10^1	3.73×10^1	2.31×10^1
	VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total})	7.60	12.05	16.23	2.05×10^1	2.47×10^1	2.88×10^1	3.27×10^1	3.62×10^1	2.23×10^1
	LSTM(\mathcal{F}_{V}^i)	11.37	12.81	14.25	1.67×10^1	1.93×10^1	2.24×10^1	2.51×10^1	2.85×10^1	1.88×10^1
	LSTM(\mathcal{F}_{V}^i , \mathcal{T}_{total})	9.88	13.01	14.88	1.73×10^1	1.95×10^1	2.27×10^1	2.52×10^1	2.86×10^1	1.89×10^1
	GRU(\mathcal{F}_{V}^i)	10.71	13.78	16.03	1.85×10^1	2.13×10^1	2.46×10^1	2.73×10^1	3.14×10^1	2.04×10^1
	GRU(\mathcal{F}_{V}^i , \mathcal{T}_{total})	8.89	11.51	12.58	1.45×10^1	1.70×10^1	2.01×10^1	2.34×10^1	2.64×10^1	1.68 $\times 10^1$
MAE	ARIMA(\mathcal{F}_{V}^i)	214.58	435.64	663.38	8.92×10^2	1.12×10^3	1.35×10^3	1.56×10^3	1.76×10^3	9.99×10^2
	VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total})	202.10	412.14	627.18	8.46×10^2	1.07×10^3	1.29×10^3	1.51×10^3	1.71×10^3	9.60×10^2
	LSTM(\mathcal{F}_{V}^i)	468.93	472.28	559.87	6.92×10^2	7.98×10^2	9.43×10^2	1.06×10^3	1.21×10^3	7.76×10^2
	LSTM(\mathcal{F}_{V}^i , \mathcal{T}_{total})	343.55	462.86	570.53	6.90×10^2	8.04×10^2	9.57×10^2	1.09×10^3	1.27×10^3	7.73×10^2
	GRU(\mathcal{F}_{V}^i)	402.88	524.09	665.89	7.78×10^2	8.98×10^2	1.05×10^3	1.16×10^3	1.31×10^3	8.49×10^2
	GRU(\mathcal{F}_{V}^i , \mathcal{T}_{total})	329.71	431.02	507.11	5.96 $\times 10^2$	7.08 $\times 10^2$	8.60 $\times 10^2$	1.00 $\times 10^3$	1.14 $\times 10^3$	6.97 $\times 10^2$
MAPE	ARIMA(\mathcal{F}_{V}^i)	5.40	11.51	18.20	2.54×10^1	3.28×10^1	4.03×10^1	4.79×10^1	5.50×10^1	2.96×10^1
	VAR(\mathcal{F}_{V}^i , \mathcal{T}_{total})	4.88	10.63	16.95	2.39×10^1	3.13×10^1	3.87×10^1	4.62×10^1	5.34×10^1	2.82×10^1
	LSTM(\mathcal{F}_{V}^i)	11.08	11.25	13.94	1.82×10^1	1.94×10^1	2.33×10^1	2.61 $\times 10^1$	3.09×10^1	1.93 $\times 10^1$
	LSTM(\mathcal{F}_{V}^i , \mathcal{T}_{total})	8.08	11.55	14.02	1.73×10^1	2.10×10^1	2.57×10^1	3.11×10^1	3.81×10^1	2.09×10^1
	GRU(\mathcal{F}_{V}^i)	11.00	14.63	18.99	2.15×10^1	2.50×10^1	3.01×10^1	3.32×10^1	3.90×10^1	2.42×10^1
	GRU(\mathcal{F}_{V}^i , \mathcal{T}_{total})	11.74	13.99	15.54	1.70 $\times 10^1$	1.83 $\times 10^1$	2.28 $\times 10^1$	2.63×10^1	2.93 $\times 10^1$	1.94×10^1

Table 9: Metric values of the models for different time horizon values (T) for the inner urban flows \mathcal{F}_{V}^i . The best metric value for each time horizon is shown in bold.

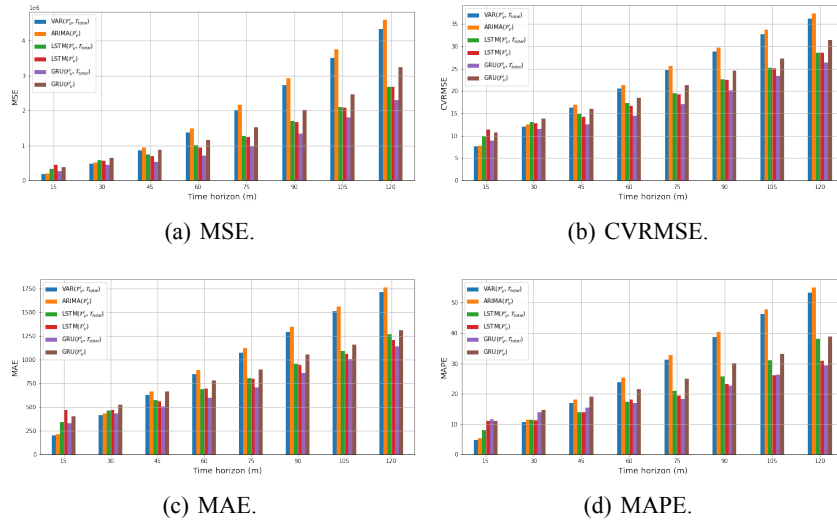


Figure 6: Plots showing the evaluation metrics for different time horizons for the inner urban flows \mathcal{F}_V^i given all the considered models.

higher, up to 48.89. Similarly, the MAPE of $GRU(\mathcal{F}_V^i, \mathcal{T}_{total})$, 19.4, was much lower than the one achieved by $GRU(\mathcal{F}_{P-V}, \mathcal{T}_{total})$, 42.03. This reveals that the human mobility between PAL and VAL is not strongly biased by road traffic but by other means of transport. At the same time, this might also involve that the higher prediction accuracy to estimate the overall inner flows of VAL is mostly due to the road displacements coming from more distant cities.

For the sake of completeness, Fig. 7 comprises the bar plots describing the evaluation metrics of the inter-city setting. Observe that the multi-variate LSTM model achieved the best scores for time horizons above 60 minutes. It is also observed the performance degradation of the VAR and ARIMA models for predictions beyond 30 minutes.

Finally, Fig. 8 depicts the raw human flows data along with the prediction of the multi-variate LSTM models for the training and evaluation sets.

6 Conclusions and Future Work

Research on human mobility is undergoing a turning point thanks to the myriad of different types of data that can be captured with today’s technologies. Data from mobile phones, GPS or IoT systems such as road traffic sensors provide a more in-depth understanding of human mobility patterns at different scales. This information is a valuable resource for the development of intelligent transport systems capable of predicting mobility flows, with different applications ranging from forecasting the spread of an epidemic to the efficient management of public transport in situations where a high demand for these services is expected.

Most of the work to date on predicting human mobility between urban areas has focused on the analysis of data sources with the same spatial level (i.e. always covering

the same area of analysis). In this context, the main contribution of this paper is that we proposed an alternative approach by combining two data sources with different spatial dimension aiming to demonstrate that such data fusion improves mobility prediction compared to using a single source. On the one hand, we use an open macro-level mobility dataset covering an interurban mobility area, and on the other hand, a dataset provided by a road traffic sensor located at a specific point on a motorway within the previous mobility area. Note that the novelty of this work does not reside in the algorithmic part of the research, but in the detailed study of how two different, complementary datasources can be integrated together to feed well-known predictive machine learning models.

To evaluate our proposal, we have defined four different models to predict the traffic flow in an area of north-eastern Spain between three main cities. The first two models are based on the use of long short-term memory (LSTM) and gated recurrent unit (GRU) neural networks, while the other two rely on regression models based on time series such as Vector AutoRegression (VAR) and AutoRegressive Integrated Moving Average (ARIMA). The results show that the two deep-learning models outperform the other ones in the reliability of the predictions in almost all the proposed situations, especially when the LSTM is fed with the two mobility flows considered for the inter-city flows

Metric	Model	T (minutes)								Avg.
		15	30	45	60	75	90	105	120	
MSE	ARIMA(\mathcal{F}_{P-V})	70.76	267.09	601.46	1,058.79	1,529.99	2,005.18	2,472.73	2,919.05	1,365.63
	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	71.48	267.06	598.31	1,047.37	1,500.86	1,948.15	2,378.73	2,785.91	1,324.73
	LSTM(\mathcal{F}_{P-V})	148.91	259.16	511.85	970.20	1,520.26	1,953.88	2,372.50	2,760.23	1,312.12
	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	292.08	478.82	713.60	1,039.02	1,306.36	1,483.51	1,673.37	1,819.58	1,100.79
	GRU(\mathcal{F}_{P-V})	184.25	337.60	860.73	1,876.76	3,133.32	4,768.63	6,972.61	8,853.70	3,373.45
	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	235.66	383.09	643.52	1,011.28	1,397.22	1,678.32	2,013.78	2,270.55	1,204.18
RMSE	ARIMA(\mathcal{F}_{P-V})	8.41	16.34	24.52	32.54	39.12	44.78	49.73	54.03	33.68
	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	8.45	16.34	24.46	32.36	38.74	44.14	48.77	52.78	33.26
	LSTM(\mathcal{F}_{P-V})	12.20	16.10	22.62	31.15	38.99	44.20	48.71	52.54	33.31
	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	17.09	21.88	26.71	32.23	36.14	38.52	40.91	42.66	32.02
	GRU(\mathcal{F}_{P-V})	13.57	18.37	29.34	43.32	55.98	69.06	83.50	94.09	50.90
	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	15.35	19.57	25.37	31.80	37.38	40.97	44.88	47.65	32.87
CVRMSE	ARIMA(\mathcal{F}_{P-V})	12.40	24.09	36.16	48.00	57.73	66.13	73.47	79.86	49.73
	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	12.40	23.98	35.91	47.54	56.94	64.91	71.75	77.68	48.89
	LSTM(\mathcal{F}_{P-V})	17.69	23.35	32.83	45.34	56.92	64.68	71.47	77.32	48.70
	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	24.78	31.74	38.76	46.92	52.76	56.36	60.02	62.78	46.77
	GRU(\mathcal{F}_{P-V})	19.68	26.65	42.57	63.06	81.71	101.05	122.53	138.48	74.47
	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	22.26	28.39	36.81	46.29	54.56	59.95	65.85	70.13	48.03
MAE	ARIMA(\mathcal{F}_{P-V})	5.36	11.52	18.03	24.74	30.72	36.12	40.81	44.95	26.53
	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	5.25	11.17	17.46	23.77	29.40	34.37	38.64	42.35	25.30
	LSTM(\mathcal{F}_{P-V})	8.32	10.06	13.67	18.79	23.92	27.90	31.21	33.37	20.91
	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	12.66	15.17	17.83	20.51	22.20	22.85	23.80	24.76	19.97
	GRU(\mathcal{F}_{P-V})	9.27	11.76	19.45	27.99	35.83	43.20	50.81	58.70	32.13
	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	10.96	12.71	15.87	18.92	21.69	23.45	26.87	29.38	19.98
MAPE	ARIMA(\mathcal{F}_{P-V})	15.81	36.74	61.68	88.99	113.99	137.58	159.86	180.40	99.38
	VAR($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	14.84	33.93	57.05	82.08	105.03	126.13	145.87	164.22	91.14
	LSTM(\mathcal{F}_{P-V})	23.37	28.72	33.56	49.81	69.35	88.60	103.79	115.71	64.11
	LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	25.61	30.14	37.18	43.97	44.96	46.77	47.65	49.64	40.74
	GRU(\mathcal{F}_{P-V})	19.27	24.43	43.70	66.35	93.20	121.22	159.55	215.81	92.94
	GRU($\mathcal{F}_{P-V}, \mathcal{T}_{total}$)	23.47	28.04	35.24	45.89	43.39	48.47	53.61	58.09	42.03

Table 10: Metric values of the models for different time horizon values (T) using the inter-urban flows \mathcal{F}_{P-V} . The best metric value for each time horizon is shown in bold.

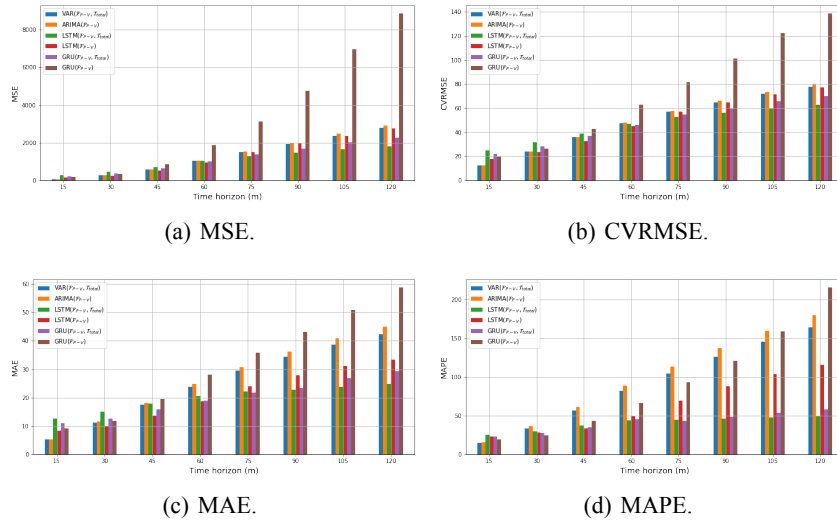
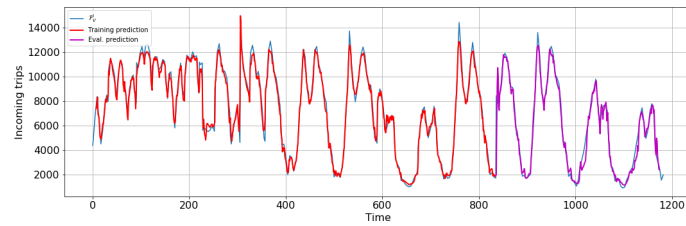
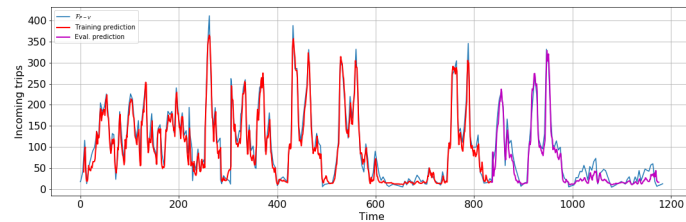


Figure 7: Plots showing the evaluation metrics for different time horizons for the inter-urban flows \mathcal{F}_{P-V}^i given all the considered models.



(a) \mathcal{F}_V^i (in blue) and the prediction of the LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$) model for the training (in red) and test set (in purple).



(b) \mathcal{F}_{P-V}^i (in blue) and the prediction of the LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$) model for the training (in red) and test set (in purple).

Figure 8: Time series of two target urban flows along with the associated prediction of the LSTM($\mathcal{F}_{P-V}, \mathcal{T}_{total}$) model for the training and test sets.

and when the multi-variate GRU model is used to predict the inner trips of the target city. In particular, the prediction of traffic entering one of the cities in the next 30 minutes following the GRU model obtains a mean average prediction error (MAPE) below 15%. However, this same model presents a greater difficulty in capturing the prediction of traffic volume between two of the cities in the study, mainly because the number of trips available in both data sources is not large enough for training the model.

The obtained results support the conclusion that the combination of heterogeneous data sources at different spatial levels, including open data sources, is a promising line for the improvement of human mobility prediction, and that those models that can take advantage of the fusion of such sources offer a higher reliability than those based on a single source. However, new factors such as commuting time between cities and travel data from other means of transport need to be added to this prediction to further improve this research line.

As future work, we are exploring the addition of sensor data from the actual vehicles to move towards real-time mobility prediction models, as well as predicting possible travel hazards such as snowfall. Another source of data to be considered in the future is official public transport data, in order to complement the low level of traffic data quantity in certain mobility areas.

Acknowledgements

Financial support for this research has been provided under grant PID2020-112827GB-I00 funded by MCIN/AEI/10.13039/501100011033. It is also partially granted by the “EMERGIA” programme, funded by the Junta de Andalucía through the grant EMC21_004171.

Code Availability

The source code of the project is available at https://github.com/fterroso/open_sensor_data_predictor

References

- [Barlacchi et al., 2015] Barlacchi, G., De Nadai, M., Larcher, R., Casella, A., Chitic, C., Torrisi, G., Antonelli, F., Vespignani, A., Pentland, A., and Lepri, B. (2015). A multi-source dataset of urban life in the city of milan and the province of trentino. *Scientific data*, 2(1):1–15.
- [Batan et al., 2018] Batan, M., Mejia, M. G., Kanasugi, H., Sekimoto, Y., and Shibasaki, R. (2018). Inferencing human spatiotemporal mobility in greater maputo via mobile phone big data mining. *ISPRS International Journal of Geo-Information*, 7(7):259.
- [Castrogiovanni et al., 2020] Castrogiovanni, P., Fadda, E., Perboli, G., and Rizzo, A. (2020). Smartphone data classification technique for detecting the usage of public or private transportation modes. *IEEE Access*, 8:58377–58391.
- [Chan et al., 2020] Chan, H. F., Skali, A., Torgler, B., et al. (2020). A global dataset of human mobility. Technical report, Center for Research in Economics, Management and the Arts (CREMA).
- [Chang et al., 2020] Chang, S., Pierson, E., Koh, P. W., Gerardin, J., Redbird, B., Grusky, D., and Leskovec, J. (2020). Mobility network models of covid-19 explain inequities and inform reopening. *Nature*, pages 1–6.

- [Chen et al., 2022] Chen, K.-W., Xie, M.-R., Chen, Y.-M., Chu, T.-T., and Lin, Y.-B. (2022). Dronetalk: An internet-of-things-based drone system for last-mile drone delivery. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–14.
- [Cheung and Lai, 1995] Cheung, Y.-W. and Lai, K. S. (1995). Lag order and critical values of the augmented dickey–fuller test. *Journal of Business & Economic Statistics*, 13(3):277–280.
- [Cho et al., 2014] Cho, K., van Merriënboer, B., Gulcehre, C., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. In *Conference on Empirical Methods in Natural Language Processing (EMNLP 2014)*.
- [Choi et al., 2019] Choi, S., Kim, J., Yu, H., and Yeo, H. (2019). Real-time prediction of arterial vehicle trajectories: An application to predictive route guidance for an emergency vehicle. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, pages 4030–4036.
- [Chu et al., 2019] Chu, K.-C., Saigal, R., and Saitou, K. (2019). Real-time traffic prediction and probing strategy for lagrangian traffic data. *IEEE Transactions on Intelligent Transportation Systems*, 20(2):497–506.
- [Chung et al., 2014] Chung, J., Gulcehre, C., Cho, K., and Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning, December 2014*.
- [Cui et al., 2020] Cui, Z., Henrickson, K., Ke, R., and Wang, Y. (2020). Traffic graph convolutional recurrent neural network: A deep learning framework for network-scale traffic learning and forecasting. *IEEE Transactions on Intelligent Transportation Systems*, 21(11):4883–4894.
- [Duan et al., 2019] Duan, P., Mao, G., Liang, W., and Zhang, D. (2019). A unified spatio-temporal model for short-term traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 20(9):3212–3223.
- [Feng et al., 2022] Feng, Y., Mao, G., Chen, B., Li, C., Hui, Y., Xu, Z., and Chen, J. (2022). Magmonitor: Vehicle speed estimation and vehicle classification through a magnetic sensor. *IEEE Transactions on Intelligent Transportation Systems*, 23(2):1311–1322.
- [Frez et al., 2019] Frez, J., Baloian, N., A. Pino, J., Zurita, G., and Basso, F. (2019). Planning of urban public transportation networks in a smart city. *Journal of Universal Computer Science*, 25(8):946–966.
- [Germanaitė et al., 2020] Germanaitė, I. E., Zaleckis, K., Butleris, R., and Jarmalavičienė, K. (2020). Case study of spatial pattern description, identification and application methodology. *Journal of universal computer science*, 26(6):649–670.
- [Guo et al., 2020] Guo, J., Zhang, S., Zhu, J., and Ni, R. (2020). Measuring the gap between the maximum predictability and prediction accuracy of human mobility. *IEEE Access*, 8:131859–131869.
- [Hu, 2007] Hu, S. (2007). Akaike information criterion. *Center for Research in Scientific Computation*, 93.
- [Ikidid et al., 2021] Ikidid, A., Fazziki, A. E., and Sadgal, M. (2021). A fuzzy logic supported multi-agent system for urban traffic and priority link control. *Journal of Universal Computer Science*, 27(10):1026–1045.
- [Kałużny and Filipowska, 2018] Kałużny, P. and Filipowska, A. (2018). Large scale mobility-based behavioral biometrics on the example of the trajectory-based model for anomaly detection. *Journal of Universal Computer Science*, 24(4).
- [Kong and Wu, 2018] Kong, D. and Wu, F. (2018). HST-LSTM: A Hierarchical Spatial-Temporal Long-Short Term Memory Network for Location Prediction. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, volume 18(7), pages 2341–2347.

- [Lee et al., 2021] Lee, Y., Jeon, H., and Sohn, K. (2021). Predicting short-term traffic speed using a deep neural network to accommodate citywide spatio-temporal correlations. *IEEE Transactions on Intelligent Transportation Systems*, 22(3):1435–1448.
- [Liu et al., 2021] Liu, Q., Zheng, X., Stanley, H. E., Xiao, F., and Liu, W. (2021). A spatio-temporal co-clustering framework for discovering mobility patterns: A study of manhattan taxi data. *IEEE Access*, 9:34338–34351.
- [Lv et al., 2021] Lv, M., Hong, Z., Chen, L., Chen, T., Zhu, T., and Ji, S. (2021). Temporal multi-graph convolutional network for traffic flow prediction. *IEEE Transactions on Intelligent Transportation Systems*, 22(6):3337–3348.
- [Miyazawa et al., 2020] Miyazawa, S., Song, X., Jiang, R., Fan, Z., Shibasaki, R., and Sato, T. (2020). City-Scale Human Mobility Prediction Model by Integrating GNSS Trajectories and SNS Data Using Long Short-Term Memory. *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, 5(4).
- [Nagy and Simon, 2018] Nagy, A. M. and Simon, V. (2018). Survey on traffic prediction in smart cities. *Pervasive and Mobile Computing*, 50:148–163.
- [Nellore and Hancke, 2016] Nellore, K. and Hancke, G. P. (2016). A survey on urban traffic management system using wireless sensor networks. *Sensors*, 16(2).
- [Organisation Internationale des Constructeurs d’Automobiles, 2020] Organisation Internationale des Constructeurs d’Automobiles (2020). Motorization rate 2015. Technical report.
- [Pfaff et al., 2008] Pfaff, B. et al. (2008). Var, svar and svec models: Implementation within r package vars. *Journal of statistical software*, 27(4):1–32.
- [Reddy et al., 1997] Reddy, T. A., Saman, N. F., Claridge, D. E., Haberl, J. S., Turner, W. D., and Chalifoux, A. T. (1997). Baseline methodology for facility-level monthly energy use-part 1: Theoretical aspects. In *ASHRAE transactions*, pages 336–347. ASHRAE.
- [Ross, 2014] Ross, B. C. (2014). Mutual information between discrete and continuous data sets. *PloS one*, 9(2):e87357.
- [Secretaría de Estado de Transportes, 2020] Secretaría de Estado de Transportes, M. y. A. U. (2020). Análisis de la movilidad en España con tecnología Big Data durante el estado de alarma para la gestión de la crisis del COVID-19. Technical report, Ministerio de Transportes, Movilidad y Agenda Urbana.
- [Shahriari et al., 2020] Shahriari, S., Ghasri, M., Sisson, S. A., and Rashidi, T. (2020). Ensemble of ARIMA: combining parametric and bootstrapping technique for traffic flow prediction. *Transportmetrica A: Transport Science*, 16(3):1552–1573.
- [Smagulova and James, 2019] Smagulova, K. and James, A. P. (2019). A survey on lstm memristive neural network architectures and applications. *The European Physical Journal Special Topics*, 228(10):2313–2324.
- [Solmaz and Turgut, 2019] Solmaz, G. and Turgut, D. (2019). A survey of human mobility models. *IEEE Access*, 7:125711–125731.
- [Spirtes et al., 2000] Spirtes, P., Glymour, C. N., Scheines, R., and Heckerman, D. (2000). *Causation, prediction, and search*. MIT press.
- [Terroso-Sáenz and Muñoz, 2021] Terroso-Sáenz, F. and Muñoz, A. (2021). Nation-wide human mobility prediction based on graph neural networks. *Applied Intelligence*, pages 1–17.
- [Terroso-Sáenz et al., 2021] Terroso-Sáenz, F., Muñoz, A., Fernández-Pedauey, J., and Cecilia, J. M. (2021). Human mobility prediction with region-based flows and water consumption. *IEEE Access*, 9:88651–88663.
- [Ullah et al., 2020] Ullah, H., Wan, W., Haidery, S. A., Khan, N. U., Ebrahimpour, Z., and Muzahid, A. A. M. (2020). Spatiotemporal patterns of visitors in urban green parks by mining social media big data based upon who reports. *IEEE Access*, 8:39197–39211.

- [von Mörner, 2017] von Mörner, M. (2017). Application of call detail records - chances and obstacles. *Transportation Research Procedia*, 25:2233 – 2241. World Conference on Transport Research - WCTR 2016 Shanghai. 10-15 July 2016.
- [Wang et al., 2020] Wang, J., Zhu, W., Sun, Y., and Tian, C. (2020). An effective dynamic spatiotemporal framework with external features information for traffic prediction. *Applied Intelligence*, pages 1–15.
- [Willmott and Matsuura, 2005] Willmott, C. J. and Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1):79–82. cited By (since 1996)149.
- [Xi et al., 2020] Xi, W., Pei, T., Liu, Q., Song, C., Liu, Y., Chen, X., Ma, J., and Zhang, Z. (2020). Quantifying the time-lag effects of human mobility on the covid-19 transmission: A multi-city study in china. *IEEE Access*, 8:216752–216761.
- [Zhao et al., 2019] Zhao, J., Gao, Y., Bai, Z., Wang, H., and Lu, S. (2019). Traffic speed prediction under non-recurrent congestion: Based on lstm method and beidou navigation satellite system data. *IEEE Intelligent Transportation Systems Magazine*, 11(2):70–81.
- [Zhao et al., 2018] Zhao, J., Xu, J., Zhou, R., Zhao, P., Liu, C., and Zhu, F. (2018). On Prediction of User Destination by Sub-Trajectory Understanding: A Deep Learning Based Approach. In *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*, CIKM '18, page 1413–1422, New York, NY, USA. Association for Computing Machinery.
- [Zhao et al., 2020] Zhao, L., Song, Y., Zhang, C., Liu, Y., Wang, P., Lin, T., Deng, M., and Li, H. (2020). T-GCN: A Temporal Graph Convolutional Network for Traffic Prediction. *IEEE Transactions on Intelligent Transportation Systems*, 21(9):3848–3858.