


An Analysis of Synthetic Timeseries as an Enabler to Improve Region-based Human Mobility Forecasting


Juan Morales-García *

(Universidad de Alicante, Alicante, Spain)

 <https://orcid.org/0000-0003-0008-4825>, juan.morales@ua.es)


Fernando Terroso-Sáenz

(Universidad Politécnica de Cartagena, Murcia, Spain)

 <https://orcid.org/0000-0002-1921-1137>, fernando.terroso@upct.es)


Andrés Bueno-Crespo

(Universidad Católica de Murcia, Murcia, Spain)

 <https://orcid.org/0000-0003-1734-6852>, abueno@ucam.edu)

José M. Cecilia

(Universitat Politècnica de València, Valencia, Spain)

 <https://orcid.org/0000-0001-5648-214X>, jmcecilia@disca.upv.es)

Abstract: Motivated by the large number of wearables offering geolocation, human mobility mining has emerged as a novel research field within AI. The study of mobility creates increasingly predictable models in which it is easy to find patterns of behaviour. However, this data is not publicly available and access to it is restricted to large telecommunications operators. In this context, this paper aims to solve one of the main problems of human mobility databases, i.e. the scarcity of data for the generation of human mobility models. For this purpose, Generative adversarial networks (GANs) have been proposed to generate synthetic time-series mobility data. Moreover, several neural network models are proposed to assess the impact of synthetic data generation on the prediction of human mobility. Our results show that the use of synthetic data improves predictions of human mobility compared to models based on available measured data. Specifically, the reinforcement learning with synthetic data benchmark, when compared to using only ground truth data, achieved a 1.22% improvement in R^2 , a 0.70% reduction in RMSE, a 2.97% decrease in MAE, a 27.07% reduction in MAPE, and an 18.18% improvement in CVRMSE, demonstrating its effectiveness in enhancing predictive accuracy.

Keywords: Deep Learning, Synthetic time series data generation, Generative Adversary Networks, Graph Neural Networks, Time series forecasting, People mobility forecasting

Categories: I.2, I.2.0, I.2.1

DOI: 10.3897/jucs.135198

1 Introduction

At the dawn of the era of digital transformation, modern societies are now faced with the fact that almost all objects and artefacts that citizens carry or use on a daily basis,

* Corresponding author.

from watches to vehicles, are equipped with location technologies such as GPS, Wi-Fi or Bluetooth, capable of pinpointing these objects to physical locations in the real world.

As a side effect of this development, the *human-mobility mining* discipline has emerged within the Artificial Intelligence (AI) 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 to 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 [Wang et al., 2017], and transportation management [Castrogiovanni et al., 2020].

In order to develop such mobility predictors, a foremost approach focuses on the time dimension to forecast the sheer number of incoming or outgoing human trips of a spatial region (e.g. a city or administrative area) [Li et al., 2020, Wang et al., 2021a]. To do so, this type of solution usually relies on time-series forecasting algorithms to generate the desired outcome.

Moreover, public institutions at different levels have promoted open data policies to enhance innovation and research in communities (e.g. see the European Commission's principle *As open as possible, as closed as necessary*^{*}). 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, these open feeds usually cover the human displacements that occurred in a limited time period. As a matter of fact, each of the open feeds proposed in [Chan et al., 2020, Barlacchi et al., 2015, Chang et al., 2020, Secretaría de Estado de Infraestructuras, Transporte y Vivienda, 2019] covers a 2-month time period and other larger datasets only comprise records of about 24 months [Mathieu et al., 2020, Secretaría de Estado de Transportes, Movilidad y Agenda Urbana, 2021]. This limitation seriously hampers the development of the aforementioned time-based human mobility predictors as it is rather complex to compose data *corpus* large enough to train and evaluate the algorithms properly.

In this context, Generative Adversarial Networks (GANs) have emerged as deep artificial neural networks capable of generating artificial data of equal or higher quality than the original one [Yoon et al., 2019]. Moreover, they have obtained very interesting results in different setting such as image processing [Sorin et al., 2020] or speech recognition [Chatziagapi et al., 2019]. Within the human-mobility field, GANs have been mostly applied to generate synthetic spatio-temporal trajectories [Smolyak et al., 2020, Belhadi et al., 2020, Ouyang et al., 2018] but their usage to compose realistic human flows have limited to specific urban settings [Bao et al., 2020]. Thus, to the best of the authors' knowledge, synthetic time-series generation has not been fully applied within the mobility field for human-flow forecasting in spite of the aforementioned limitation of unrestricted mobility datasets related to their time dimension.

In this paper, we fill this gap by making use of synthetic human-flows as a suitable input to increase the accuracy of such predictors. To do so, we evaluate a palette of deep-learning algorithms for time-series prediction and define several training strategies to fuse real and synthetic data coming from a nationwide open human-mobility feed. This way, we are able to improve the prediction of human displacements among different administrative areas.

All in all, the salient contributions of the present work are the following,

- Creation of synthetic human-flows as time-series using GAN techniques considering

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

a nationwide human mobility dataset.

- Study of the best prediction technique using neural networks to predict the incoming flows of a set of administrative areas.
- Analysis and comparison of the different models created with synthetic and original data in an isolated manner, as well as with the fusion of both types of data.

The rest of the paper is organized as follows. Section 2 shows a review of some work on the synthetic creation of time series datasets and mobility studies and forecasting. Section 3 describes the proposed GAN technique for creating synthetic time series data, as well as the techniques used for the evaluation of such synthetic data, and benchmarks followed. Section 4 the results obtained are shown, analyzed and discussed. Finally, Section 5 presents the main conclusions and discusses future works.

2 Related works

The present work focuses on the usage of time-series forecasting algorithms to predict human mobility flows. In that sense, these algorithms can be categorized into parametric and non-parametric approaches [Van Lint and Van Hinsbergen, 2012]. The former relates to proposals that focus on generating a model based on a palette of theoretical or physical foundations and tuning its set of parameters. Examples of these approaches are ARIMA [Shahriari et al., 2020] and linear regression models [Kwak and Geroliminis, 2020]. Nonetheless, these studies do not offer a reliable performance under abnormal mobility situations due to, for example, hazardous weather events. Regarding non-parametric methods, Deep Learning (DL) algorithms have been widely used for such a forecasting task. In that sense, LSTM [Miyazawa et al., 2020], Conv-LSTM [Yang et al., 2018] and GRU [Fan et al., 2018] networks have been some of the most relevant Recurrent Neural Networks used in this setting. More recently, some works have profited from the graph-based structure of most human-mobility settings (e.g. road networks) to make use of Graph Neural Networks (GNNs) for flow prediction [Zhao et al., 2020, Terroso-Sáenz and Muñoz, 2022].

The study of human mobility has been greatly enhanced by deep learning models, which help address challenges in predicting user behavior and mobility patterns. Existing frameworks, such as the Personalized Behavior Modeling Network (PBMN) introduced by [Wu et al., 2024], offer significant improvements in capturing complex user preferences and contextual information for accurate mobility predictions. Moreover, frameworks like the Context-Directional SpatioTemporal Graph Network (CD-STGNet), as shown in [Zhou et al., 2023], tackle spatiotemporal heterogeneity in mobility data and improve prediction performance by integrating context-aware learning. Additionally, uncertainty quantification methods, such as the one proposed by [Zhou et al., 2021], enhance model robustness by addressing spatiotemporal uncertainty and optimizing prediction accuracy. These frameworks provide a comprehensive foundation for evaluating mobility prediction models and can be applied to future studies to further refine predictions in human mobility tasks [Luca et al., 2021].

Focusing on the generation of synthetic data for time series data with GANs, they have been widely used to detect anomalies both in univariate [Lee et al., 2021, Jiang et al., 2019, Niu et al., 2020] and multivariate models [Li et al., 2019]. Besides, it is possible to find in the literature some *domain-agnostic* GANs. For instance, the work in [Ramponi et al., 2018], where authors propose a conditional GAN to create synthetic data from

temporal serials, where both the generator and the discriminator are conditional on the sampling time, in order to obtain the hidden pattern between the data and the timestamps when samples are irregular. Yoon et al. proposed a framework for generating synthetic time series data, where supervised and unsupervised techniques are combined [Yoon et al., 2019]. Specifically, the authors propose an unsupervised GAN with supervised training using autoregressive models.

The role of bias in synthetic data generation is a critical concern, especially in machine learning applications. Several studies have explored how synthetic data can help mitigate bias. The authors in [Shahul Hameed et al., 2024] review various methods for using synthetic data to reduce bias across different fields, emphasizing the importance of fairness in AI models. Similarly, [Barbierato et al., 2022] presents a methodology to control both bias and fairness in synthetic data generation, focusing on structural equation modeling to adjust bias levels deliberately. Moreover, the authors in [Liang et al., 2023] demonstrate how synthetic data can be employed to benchmark and correct algorithmic bias in face recognition systems, showing its potential for identifying and addressing biases in predictive models. Finally, [Xu et al., 2022] explores random graph generation techniques for synthetic data, which help improve dataset diversity and mitigate biases in machine learning models. These studies collectively highlight the importance of designing synthetic data generation techniques that enhance fairness and reduce biases in AI systems.

We can also find some interesting GAN models in the human mobility field so as to generate synthetic spatio-temporal trajectories. For example, Smolyak et al. propose an architecture combining Gaussian Mixture Models and bidirectional GANs to generate realistic synthetic datasets of individual trajectories for anomaly detection based on the well-known Geolife trajectory dataset [Smolyak et al., 2020]. A similar approach using CNN-based GANs was introduced by Ouyang et al. in [Ouyang et al., 2018]. Moreover, Belhadi et al. put forward a GAN model to simulate taxi trajectories so as to compose benchmarks to identify frauds in this type of public means of transport [Barlacchi et al., 2015]. In the same vehicle-based movement field, Wang et al. proposed a GAN architecture that embeds road network information to generate realistic GPS trajectories fitting a particular map setting [Wang et al., 2021b]. Finally, another interesting GAN proposal is the one stated in [Zhang et al., 2021] where convolutional GANs are used to generate synthetic travel times between different locations in a geographical area.

In this scope, our work goes beyond the generation of synthetic human-flow data and it actually focuses on studying the utility of such data to compose reliable predictors. This way, we provide a novel evaluation of synthetic human mobility data from a more *application-oriented* perspective. To do so, we rely on well-known DL architectures in the human-mobility field as LSTM, Conv-LSTM and GNN models.

3 Materials and Methods

This section introduces (1) the datasets used as well as their characteristics; (2) the synthetic data generation model; (3) the artificial intelligence models used to validate the effectiveness of the synthetic data; and (4) the different benchmarks followed to achieve the above objective.

3.1 Human Mobility Dataset

The target *ground truth* dataset has been retrieved from the nationwide human mobility report released by the Spanish Ministry of Transportation (SMT) in December

2020[Secretaría de Estado de Transportes, Movilidad y Agenda Urbana, 2021]. It covers a 15-month period from February 29th, 2020 to May 4th, 2021, indicating the number of trips among 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 a 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, Movilidad y Agenda Urbana, 2021], this mobility data was collected through Call Detail Records (CDRs) from 13 million users of an unspecified mobile-phone carrier. Once anonymized, 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 performed two major data-curation steps. First, we only considered the tuples from 2020-04-01 to 2021-04-30. The rationale for removing the first 2 months of the dataset was to avoid the effect that the COVID-19 pandemic had on the nationwide displacements in Spain. Secondly, we aggregated the flows from a MA granularity to the province level. In that sense, Spain is divided into 52 different provinces each one with an average population of about 904,000 people ($\pm 120,000$). As Figure 1 shows, each MA geographically fits into a single province, then a province can be regarded as an aggregation of MAs providing a coarse-grained representation of the mobility flows. As a result, we finally generate a multivariate time-series comprising 492,856 records with 52 features each. The reason for this aggregation was to provide a time series' granularity with a trade-off between its spatial granularity and the explainability of the results. In that sense, using the MA-based dataset would give rise to a multivariate time-series with more than 2,000 features (one feature per MA) that would be too large to perform a proper region-based analysis of the results.

In summary, the dataset employed in this study begins on 2021-05-01, and comprises a total of 492,856 records and 52 features, with each feature representing a time-series corresponding to a specific MA.

3.2 Multivariate Synthetic data generation using Generative Adversarial Networks (GANs)

A Generative Adversarial Network (GAN) comprises two neural networks—the generator and the discriminator—that are trained simultaneously in a competitive framework. The generator aims to create synthetic data from random noise inputs, striving to make these data indistinguishable from real data. Conversely, the discriminator is tasked with evaluating both real and synthetic data to accurately differentiate between the two. Throughout the training process, the generator seeks to improve its ability to produce realistic data capable of deceiving the discriminator, while the discriminator enhances its capacity to distinguish between genuine and generated data. Over time, this adversarial training enables the generator to synthesize data that closely resemble real data, fulfilling its objective of realistic data generation.

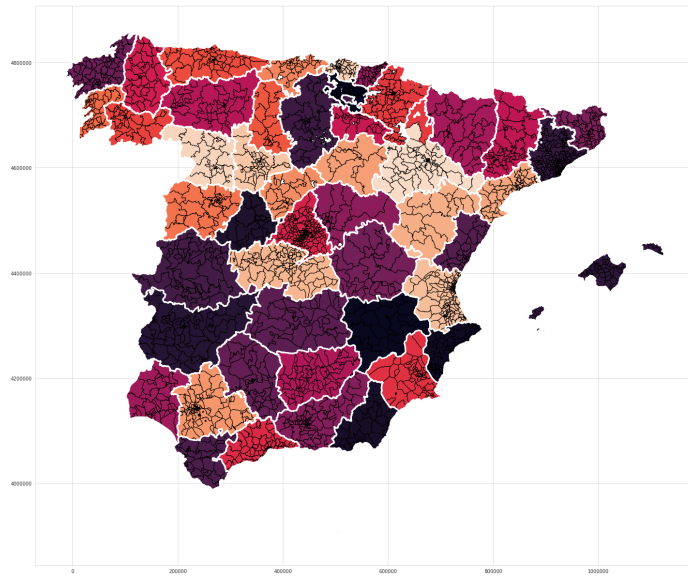


Figure 1: Spanish provinces. Each province is limited by a white-line boundary. The boundaries of the MAs are depicted as black lines

For time series data, the sequential nature introduces unique challenges due to temporal correlations and dependencies between values. To address these complexities, the generator must ensure temporal consistency in the synthesized data. This is often achieved by employing architectures such as recurrent neural networks (RNNs) in both the generator and discriminator, facilitating the effective modeling of temporal dynamics and dependencies inherent to time-series data.

This study utilizes *DoppelGANger*, a GAN architecture specifically designed for sequential data, as proposed in [Lin et al., 2020], for the generation of synthetic data. The GAN architecture employed, illustrated in Figure 2, builds upon the canonical structure of strawman GANs tailored for time series generation.

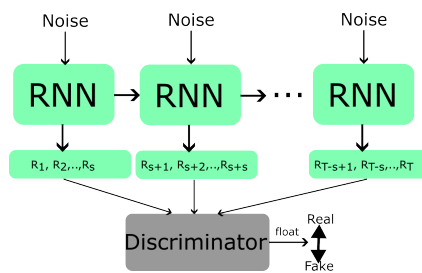


Figure 2: Architecture of the DoppelGAN used for the synthetic human flows generation

It relies on Recurrent Neural Networks (RNNs) to generate synthetic time series data. Particularly, the generative part of *DoppelGANger* relies on a layer of LSTM cells with 100 units, following a batch generation strategy. Hence, the model generates, in each pass, S consecutive records of the synthetic time series data, instead of a single one as most of the traditional approaches (e.g., R_1, R_2, \dots, R_S in Figure2). According to the authors, this allows us to better capture the temporal correlation of long series and reduce the number of passes required by the model to generate the synthetic data. Furthermore, the GAN also includes a *normalization* mechanism for each input time-series to tackle the well-known model-collapse problem of many GAN models. Then, the discriminator, which takes the form of a Multilayer perceptron (MLP) with up to five layers of two hundred neurons each followed by a ReLU activation function, uses the Wasserstein loss to report the differences between the ground truth and the fake data.

3.3 Multivariate time-series forecasting using Artificial Intelligence (AI) models

In order to carry out an evaluation of the impact of synthetic data on the prediction of time-series mobility data, 3 different Artificial Intelligence (AI) models are proposed.

3.3.1 MultiLayer Perceptron (MLP)

This model is a group of multiple perceptrons in each layer forming a multilayer perceptron. A perceptron is an artificial neuron, which can be understood as a logistic regression function, such that a set of them can simulate the functioning of a set of biological neurons which enables learning of any non-linear function by creating a model between the input and the output. In this work, the MLP is used as a regression model presenting three layers: the input layer, the hidden layer, which is responsible for processing the inputs, and finally producing an output through the third layer called the output layer [Bishop, 2006, Haykin and Lippmann, 1998].

3.3.2 Graph Neural Network (GNN)

The GNN mode used in this work is based on the one proposed in [Terroso-Sáenz and Muñoz, 2022]. The model relies on a graph $\mathcal{G} = \langle \mathcal{P}, E \rangle$ where the nodes \mathcal{P} are the Spanish Provinces, $|\mathcal{P}| = N = 52$ and the edges E represent the connections among them. Such connections are defined by an adjacency matrix, \mathcal{A} , based on the gravity model for human mobility [Pappalardo et al., 2016]. Briefly, this model establishes that the magnitude of displacements $MD_{a,b}$ between two regions r_a, r_b can be computed with the formula $MD_{a,b} = \frac{P_a \times P_b}{Dist_{a,b}}$ where $P_{\{a,b\}}$ is the population of each region and $Dist_{a,b}$ its distance. This way, this weighted adjacency matrix is populated with such a model so that a cell $a_{ij} \in \mathcal{A}$ is set to $MD_{i,j}$.

Moreover, a set of feature matrices $\mathcal{OT}_t \forall t \in \mathcal{T}$ comprising the last h_{prev} incoming trips of the provinces at time t is also generated.

Last, the model uses as an input the matrices \mathcal{A} and \mathcal{OT} to perform the prediction task. Then, the spatial dependencies among MAs are modeled with a set of graph-convolutional layers whereas the temporal ones are processed by a stack of LSTM layers. Last, a Multilayer Perceptron (MLP) processes the sequences from the LSTM layers.

3.3.3 Convolutional Neuronal Network + Long Short-Term Memory (CNNLSTM)

This model is a combination of a convolutional network (CNN) and a long short-term memory (LSTM) network. In a first stage, the CNN network automatically extracts the input features as it is formed by filter blocks to identify the relevant features from the input by convolution operations, obtaining the necessary and most relevant features from the input data [Li et al., 2021]. Unlike a traditional CNN model, this model exchanges the classical MLP in the output layer for an LSTM regression network. This network is a type of recurrent neural architecture with state memory and multilayer cell structure that differs from a classical recurrent network in that it does not overwrite its content at each time step, but is able to decide whether or not to keep the memory, allowing long-term dependencies to be detected [Hochreiter and Schmidhuber, 1997].

3.4 Benchmarks performed

In order to carry out a correct evaluation of the impact of synthetic data on the prediction of time-series mobility data, four different training-testing strategies are proposed:

1. **Ground truth benchmark (GT):** This strategy consists of dividing the dataset into two parts: (a) the training set, consisting of all the ground truth data except the last day; and (b) the test set, consisting of the last day of the ground truth data. Once the two datasets are available, the training set is trained with the training set and tested with the test set.
2. **Synthetic benchmark (S):** This strategy consists of dividing the dataset into two parts: (a) the training set, formed by all the synthetic data except the last day; and (b) the test set, formed by the last day of the ground truth data. Once the two datasets are available, the training set is trained with the training set and tested with the test set.
3. **Synthetic + Ground truth benchmark (RS):** This strategy consists of dividing the dataset into two parts: (a) the training set, formed by a combination of all the synthetic data and the ground truth data except the last day; and (b) the test set, formed by the last day of the ground truth data. Once the two data sets are available, the training set is trained with the training set and tested with the test set.
4. **Synthetic + Ground truth with Reinforcement Learning benchmark (RL):** This strategy consists of dividing the dataset into three parts: (a) the first training set, consisting of all synthetic data except the last day; (b) the second training set, consisting of all ground truth data except the last day; and (c) the test set, consisting of the last day of ground truth data. Once the three data sets are available, we train with the first training set, then we train with the second training set, and finally, tested with the test set.

4 Evaluation and Discussion

The dimension of the problem targeted is twofold: (1) the use of GANs for synthetic data generation (in this particular case, time-series data applied to human mobility) and (2) the impact on the accuracy of artificial intelligence models depending on whether ground truth data, synthetic data, or a combination between ground truth and synthetic is used.

4.1 Multivariate synthetic data generation

For the correct analysis of both the ground truth data and the synthetic data, an aggregation has been made on both. Specifically, the aggregation performed is the sum of each of the values of each province for a given timestamp, as follows:

$$\sum_{i=1}^{52} x_i = x_1 + x_2 + x_3 + \dots + x_{52}$$

In this way, a univariate dataset is obtained that gathers the characteristics of all the provinces analyzed, allowing the creation of two-time series (ground truth and synthetic) that can be compared. This new aggregated dataset has the following characteristics: (1) the ground truth time series is made up of a total of 9478 records and (2) the synthetic time series is made up of a total of 26280 records.

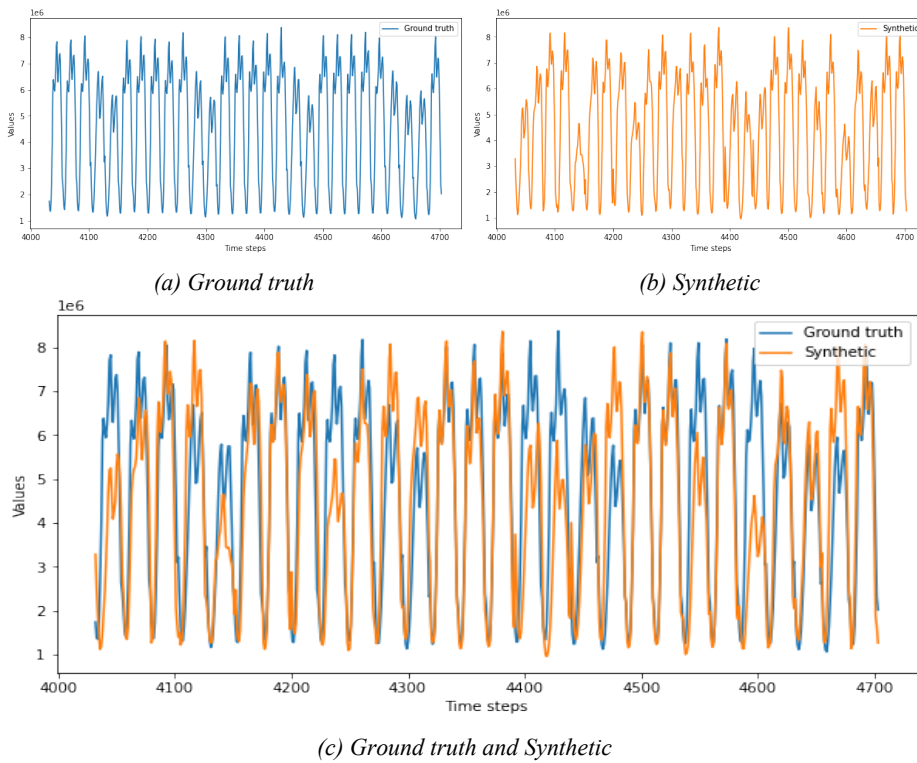


Figure 3: Line chart showing a brief 7-day example of the ground truth and synthetic time series

A 7-day sample of the dataset, either ground truth (see Figure 3a), synthetic (see Figure 3b), or both at the same time (see Figure 3c) can be seen in Figure 3.

	Ground truth	Synthetic	Diff	Diff (%)
count	9478.000	26280.000	16802.000	1.773
mean	4303623.730	4314695.155	11071.425	0.003
std	2070657.346	2070401.992	-255.354	-0.000
min	871631.384	880322.202	8690.818	0.010
25%	2218994.312	2270672.589	51678.277	0.023
50%	4553507.953	4623365.782	69857.829	0.015
75%	6120366.941	6116349.300	-4017.641	-0.001
max	8632575.354	8437420.398	-195154.956	-0.023

Table 1: Statistical study of ground truth and synthetic time series, showing the count, mean, standard deviation (std), quartiles (25%, 50%, 75%) min and max values of each ground truth and synthetic dataset. The difference between these statistics in absolute values (Diff) and in percentages (Diff (%)) is also shown

The complete statistical study performed on the real and synthetic datasets, as well as the differences observed in both, is shown in Table 1. It is important to note that the differences observed between both datasets are minimal, even having generated 3 years of synthetic data from 1 year of real data, very similar values are being obtained in terms of mean, standard deviation, quartiles, maximum and minimum values. Specifically, the percentage of variation of all the previously mentioned aspects is less than 0.023.

A graphical representation of the statistical data analyzed in Table 1 is shown in Figure 4. Specifically, a box plot, a histogram of the dataset, and two density plots, one total and the other by minute sequences are shown. These graphical representations verify the statistical values previously described in Table 1.

		Pettit	SNHT	Buishand Q	Buishand Range	Buishand Likelihood Ratio	Buishand U
Ground truth	h	True	True	True	True	True	True
	p	0.000	0.000	0.000	0.000	0.000	0.000
	cp	1472	1135	1471	1471	1471	1471
	mu1	3217993.139	2949741.065	3217128.945	3217128.945	3217128.945	3217128.945
	mu2	4503230.054	4487808.894	4503228.305	4503228.305	4503228.305	4503228.305
Synthetic	h	False	True	False	False	True	False
	p	0.109	0.049	0.108	0.148	0.044	0.249
	cp	17781	191	17781	17781	17781	17781
	mu1	4337439.097	4831195.378	4337439.097	4337439.097	4337439.097	4337439.097
	mu2	4267111.906	4310913.808	4267111.906	4267111.906	4267111.906	4267111.906

Table 2: Homogeneity test performed on the ground truth and synthetic data. The rows show the homogeneity value (h), the p-value (p), the possible cut-off point (cp) and the before (mu1) and after (mu2) means of the cp for both, ground truth and synthetic data. The columns show the six statistical tests performed

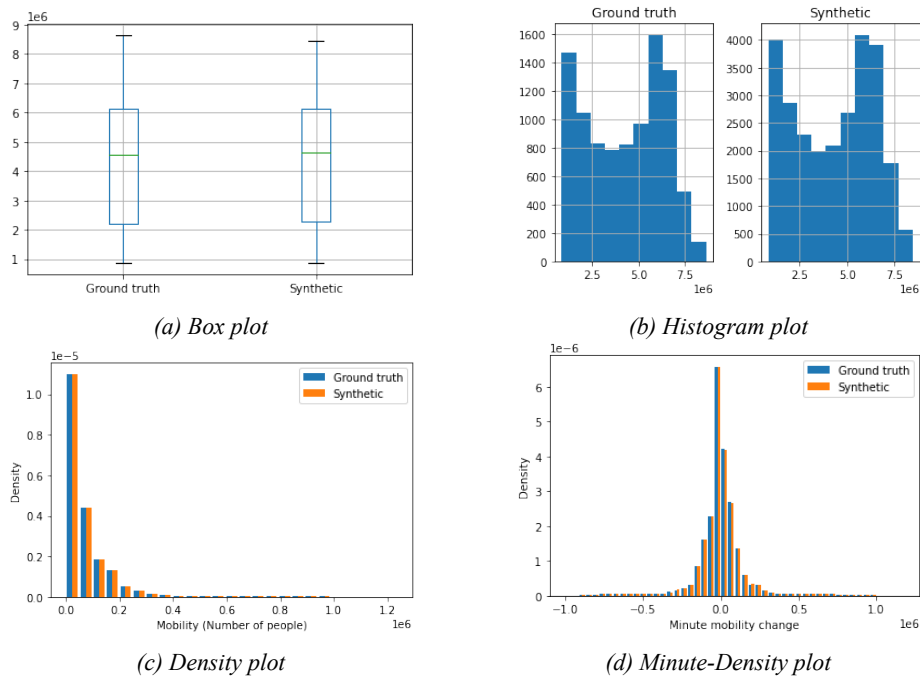


Figure 4: Graphs of the statistical study of ground truth and synthetic time series

Table 2 shows all the results of homogeneity tests performed for both, ground truth and synthetic datasets. In particular, six different statistical tests have been used to test the validity of the statistics obtained: (1) Pettit test [Pettit, 1979]; (2) SNHT (Standard Normal Homogeneity Test) test [Alexanderson, 1986]; (3) Buishand Q test [Buishand, 1982]; (4) Buishand Range test [Buishand, 1982]; (5) Buishand Q test [Buishand, 1982]; and (6) Buishand U test [Buishand, 1982]. These results show that, for the vast majority of tests, they identify both time series as non-homogeneous (a value of True at the h-value indicates that a time series is not homogeneous, while a value of False at the h-value indicates that the time series is homogeneous). These results show that all tests identify the ground truth time series as non-homogeneous, while the vast majority of tests indicate that the synthetic data are homogeneous (a value of True at the h-value indicates that a time series is not homogeneous, while a value of False at the h-value indicates that the time series is homogeneous). This is because the beginning of the ground truth time series has a series of unstable values that make the mean of the values drop significantly, stabilizing at the end of the time series, while the synthetic data are stable throughout the time series.

4.2 Multivariate time-series forecasting

In order to measure quality of each model described in section 3.3, we have used three different metrics to perform such an evaluation, the Coefficient of determination (R^2), the Root Mean Squared Error (RMSE), the Mean Absolute Error (MAE), the Mean

Absolute Percentage Error (MAPE) and the Coefficient of the Variation of the Root Mean Square Error (CVRMSE). These are some of the most common metrics used to measure accuracy for continuous variables. Their definition is as follows:

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

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

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

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \frac{|e_i|}{y_i}$$

$$CVRMSE = \frac{RMSE}{\bar{y}}$$

where, for our experiment, y_i is the real (ground truth) value, \hat{y}_i is the forecasted value, e_i^2 is the error term and n is the number of observations.

In Table 3, all the hyperparameters used in each of the models are presented.

Hyperparameter	Description	CNNLSTM	MLP	GNN
Units	Number of neurons in hidden layers	-	70	128
Filters	Number of filters used to detect features	64	-	-
Kernel size	Size of the filter matrix used for feature extraction	2	-	-
Strides	Number of pixel shifts applied to the input matrix	1	-	-
Activation function	Function that determines whether a neuron is activated	Tanh		
Batch size	Number of samples in each training/forecasting batch	2880	2880	12
Epochs	Number of epochs for model training	15000 (+ <i>EarlyStopping</i>)	15000 (+ <i>EarlyStopping</i>)	900 (+ <i>EarlyStopping</i>)
Optimizer	Algorithm used to optimize model learning	Adam		
Loss function	Metric used to evaluate the model's error during training	MSE		
Learning rate	Rate at which the model's weights are updated	0.003 (+ <i>ReduceLROnPlateau</i>)	0.003 (+ <i>ReduceLROnPlateau</i>)	0.001

Table 3: Hyperparameters of the CNNLSTM, MLP, and GNN models

Additionally, the graphs depicting the loss and learning rate for each of the studied models are presented in Figure 5. Specifically, for each model, the first graph illustrates the training loss (blue line) and validation loss (orange line), along with the epoch at which training stopped (red perpendicular line). On the other hand, the second graph displays the evolution of the learning rate throughout the training process (purple line), as well as the epoch at which training stopped (red perpendicular line).

Table 4 shows all the results obtained in the evaluation for each benchmark and model. In order to generally evaluate the results obtained, similar to what was done with the Exploratory Data Analysis (EDA) in section 4.1, the metrics of each province have been aggregated, in this case, by averaging the values of each metric analyzed, as follows: $\frac{1}{n} \sum_{i=1}^n x_i$. These results show that the best model is the CNNLSTM while the worst model is the MLP, with the GNN being an intermediate model that improves

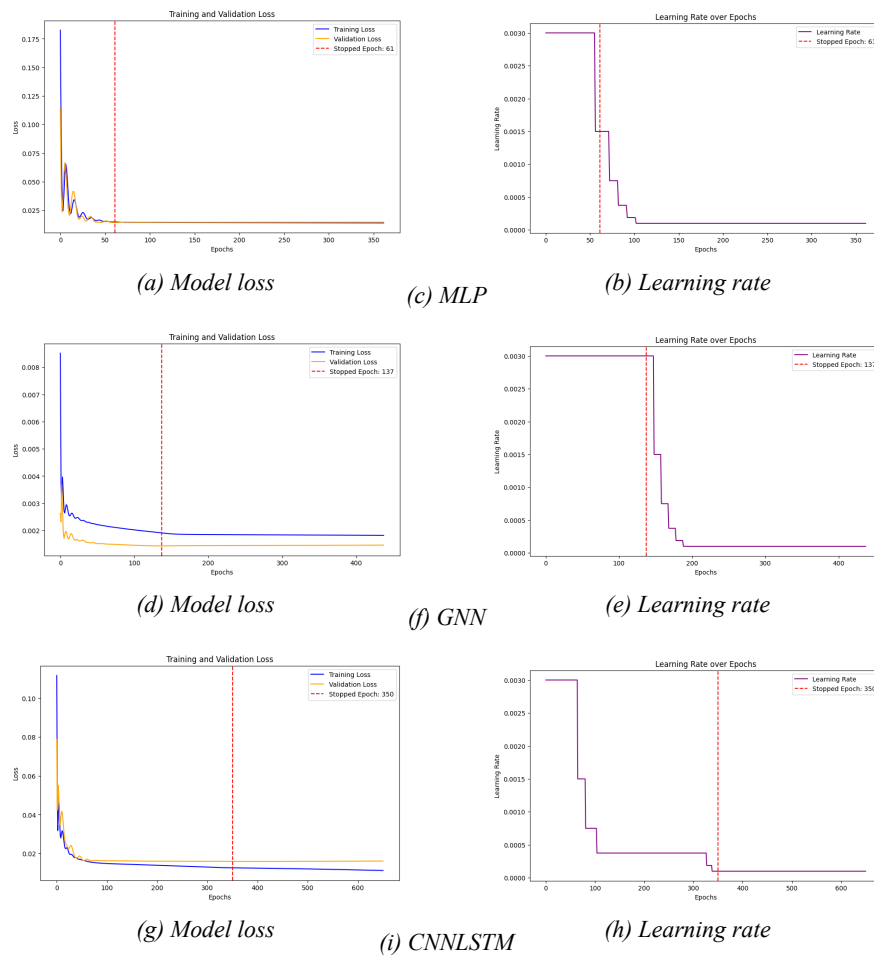


Figure 5: Model loss and learning rate plots for each of the models studied

the MLP but is not as good as the CNNLSTM. It is also shown that, within all the benchmarks performed, the worst is the one that uses only synthetic data while the one that outperforms the rest is the RL approach obtaining (as the most remarkable metrics for the CNNLSTM model) an R2 of 0.97 and a MAPE of 10.16%, which means that it is a model whose prediction maintains the same linearity and trend as the ground-truth time series with a very low error percentage, that is, with hardly any error between the real and predicted time series.

Figure 6 shows the CVRMSE for each province and training strategy given the CNNLSTM model. An important finding, in this case, is that we observed a similar spatial distribution of the error among provinces regardless of the training strategy where the higher error is obtained in the inner provinces of Spain. However, such error is slightly lower when the RL approach is used (see Figure 6d).

Model	Training	R2	RMSE	MAE	MAPE	CVRMSE
CNNLSTM	GT	0.9591	6,324.6979	5,167.8582	0.1393	0.1181
	S	0.9422	19,115.9421	16,118.3923	0.1909	0.2069
	RS	0.9595	8,075.4235	6,590.8826	0.1404	0.1136
	RL	0.9708	6,280.4270	5,014.1337	0.1016	0.0921
MLP	GT	0.7815	12,841.5730	10,379.7768	0.3698	0.2853
	S	0.6957	24,392.1727	20,392.2185	0.4862	0.4198
	RS	0.8107	16,276.2321	13,276.9280	0.3245	0.2741
	RL	0.8500	12,887.0919	10,419.6757	0.2758	0.2318
GNN	GT	0.9193	9,921.2796	7,587.2496	0.1750	0.2130
	S	0.8133	20,373.8268	16,254.7354	0.4738	0.5062
	RS	0.7659	15,623.3347	11,822.6768	0.2814	0.3240
	RL	0.9280	8,971.5182	6,881.5508	0.1496	0.1751

Table 4: Evaluation metrics for all the models and training strategies under consideration. The best value for each metric is shown in bold

Moreover, Figure 7 shows the spatial distribution of the MAPE metric for the 3 target models. As the lighter colors in Figure 7c show, the CNNLSTM instance achieves the most accurate predictions for almost all the provinces. Again, we also observed that the spatial distribution of this error is similar in all the models with a more accurate prediction in the coastal provinces with respect to the inner ones.

4.3 Comparison between daily and weekly forecasting

Finally, we analyzed the impact of temporal aggregation on human mobility flows. To this end, we aggregated both the ground-truth and synthetic flows on a weekly basis, considering a Monday-to-Sunday week. Table 5 presents the evaluation metrics obtained by the models when trained and tested under this aggregation scheme. In this context, it is important to note that the CNNLSTM model achieved the best results in this evaluation as well.

As observed, the models achieved slightly improved values for R^2 , MAPE, and CVRMSE. This improvement aligns with the expectation that these metrics benefit from higher-level aggregation, as daily fluctuations are smoothed, leading to better alignment between predictions and the aggregated data.

Conversely, RMSE and MAE values worsened, showing an increase of approximately 2–4%. This degradation is attributable to the nature of weekly aggregation: both absolute and squared errors increase in magnitude due to the cumulative nature of the data over seven days.

Furthermore, Figure 8 shows the CVRMSE spatial distribution of the CNNLSTM model for the 4 benchmarks. As we can see, such a distribution follows similar spatial patterns than the one observed with daily data with inner provinces exhibiting higher errors than the coastal ones.

Lastly, Figure 9 presents the MAPE of the three models under consideration based on weekly aggregation, with Reinforcement Learning used as the benchmark. The results

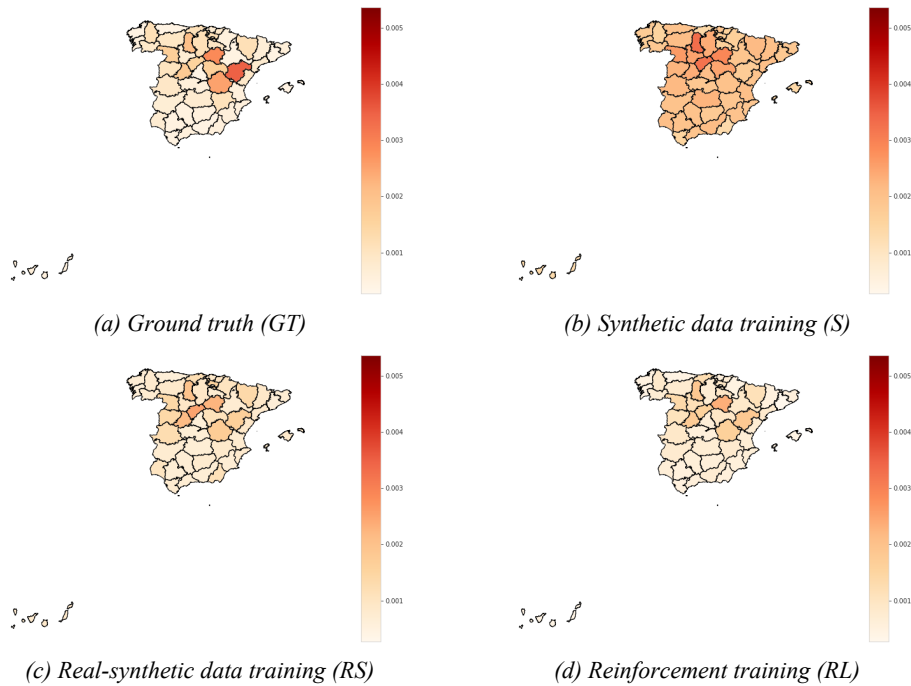


Figure 6: Map showing the spatial distribution of the CVRMSE per province for the 4 benchmarks given the CNNLSTM model

once again exhibit a spatial distribution of the error similar to that observed with daily mobility data (Figure 7), showing a slightly improved performance of the model in the coastal provinces.

5 Conclusion and Future Work

With the advent of Machine Learning models, the use of synthetic data for those scenarios where insufficient data is available is gaining acceptance and growing interest within the field of artificial intelligence. However, this approach has not been widely applied to extend time-series databases for human mobility problems.

In this area, this paper demonstrates the importance of the use of synthetic data in mobility problems due to the size limitations that these data sets usually have. For this purpose, a GAN model applied to time series has been used to enlarge a region-based human mobility dataset at a national scale. In that sense, it has been demonstrated, by means of statistical analysis, that these synthetic data have a similar quality to the original ones. To rule out that the predictions could be too model-dependent, we have evaluated 3 deep learning models (MLP, GNN, and CNNLSTM) and different training strategies combining the synthetic and the original time-series. It has been shown that the best-fitting model is the CNNLSTM based on 5 different metrics. Moreover, the evaluation has also shown very similar spatial distributions of the prediction error across regions regardless of the model or training strategy.

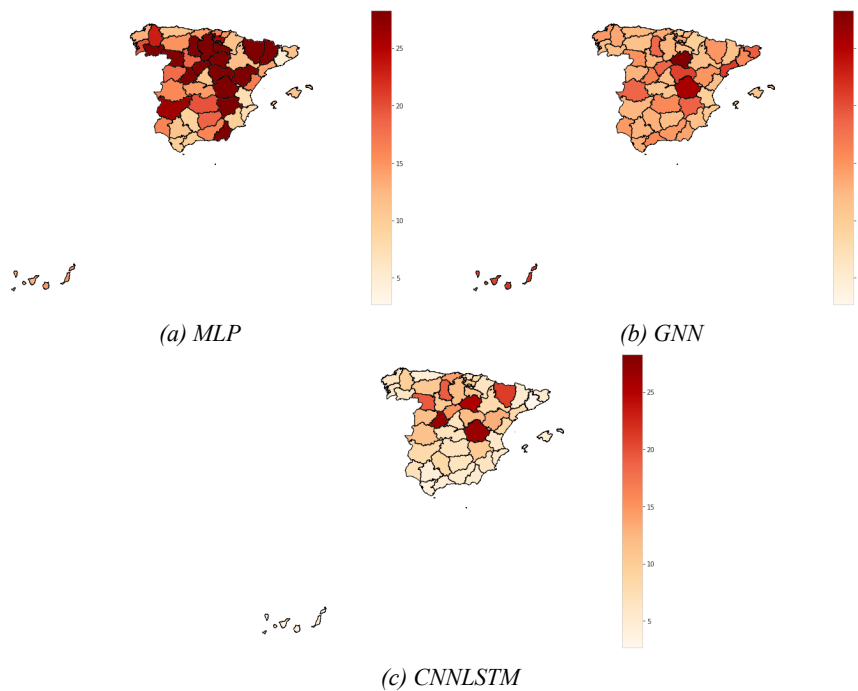


Figure 7: Map showing the spatial distribution of the MAPE per province for the 3 models when they are benchmarked with the Reinforcement Learning approach

When comparing the performance of the CNNLSTM model trained exclusively with real data (GT) against the reinforcement learning (RL) approach, notable improvements were observed. The RL strategy demonstrated a 1.22% enhancement in the coefficient of determination (R^2), reflecting its superior ability to capture the variability in human mobility flows. Additionally, the root mean squared error (RMSE) was reduced by 0.70%, while the mean absolute error (MAE) decreased by 2.97%. Most significantly, the mean absolute percentage error (MAPE) decreased by 27.07%, and the coefficient of variation of the root mean square error (CVRMSE) was reduced by 18.18%. These findings emphasize the effectiveness of reinforcement learning in enhancing the predictive accuracy of deep learning models, particularly when dealing with complex, multivariate time-series data such as human mobility flows.

Future research will focus on three key objectives: improving the GAN model to enhance data quality, testing additional AI models applied to mobility time series, and expanding the scope of this study to domains beyond mobility. Additionally, future work could explore various synthetic data generation strategies using alternative mobility datasets, such as PFLOW and OpenPFLOW, while also investigating other generative models like Variational Autoencoders (VAE) to generate synthetic data of higher quality and yield better results. Finally, scalability issues, including the impact of regional factors on computational time and resource consumption, should be examined. The temporal dynamics of mobility patterns, especially in response to events like the COVID-19 pandemic, also warrant further exploration to address the challenges and limitations of

Model	Training	R2	RMSE	MAE	MAPE	CVRMSE
CNNLSTM	GT	0.9625	6,577.6859	5,374.5725	0.1254	0.1056
	S	0.9470	19,689.4204	16,602.9440	0.1722	0.1924
	RS	0.9629	8,317.6862	6,799.6091	0.1263	0.1012
	RL	0.9740	6,531.6431	5,214.6990	0.8647	0.0809
MLP	GT	0.7894	13,256.9202	10,795.9679	0.3328	0.2630
	S	0.7048	25,032.9778	20,911.0240	0.4375	0.3865
	RS	0.8198	16,846.5014	13,675.2358	0.2920	0.2589
	RL	0.8592	13,402.5756	10,836.4627	0.2483	0.2134
GNN	GT	0.9250	10,218.9170	7,815.8671	0.1575	0.1923
	S	0.8214	21,096.4192	16,742.3774	0.4264	0.4602
	RS	0.7745	16,091.0347	12,079.1303	0.2533	0.2967
	RL	0.9345	9,240.6637	7,088.9975	0.1346	0.1524

Table 5: Evaluation metrics for all the models and training strategies under consideration given a weekly aggregation of the flows. The best value for each metric is shown in bold

synthetic data generation in dynamic real-world contexts.

Declarations

Ethical Approval

Not applicable

Conflict of interest

The authors declare they do not have any conflict of interest.

Authors' contributions

Conceptualization, J.M.G. and F.T.S.; methodology, J.M.G. and F.T.S.; software, J.M.G. and F.T.S.; validation, A.B.C. and J.M.C.; formal analysis, J.M.G. and F.T.S.; investigation, J.M.G. and F.T.S.; writing—original draft preparation, J.M.G. and F.T.S.; writing—review and editing, A.B.C. and J.M.C.; visualization, J.M.G. and F.T.S.; supervision, A.B.C. and J.M.C.; project administration, A.B.C. and J.M.C.; funding acquisition, A.B.C. and J.M.C.

Consent to publish

All authors have read and agreed to the published version of the manuscript.

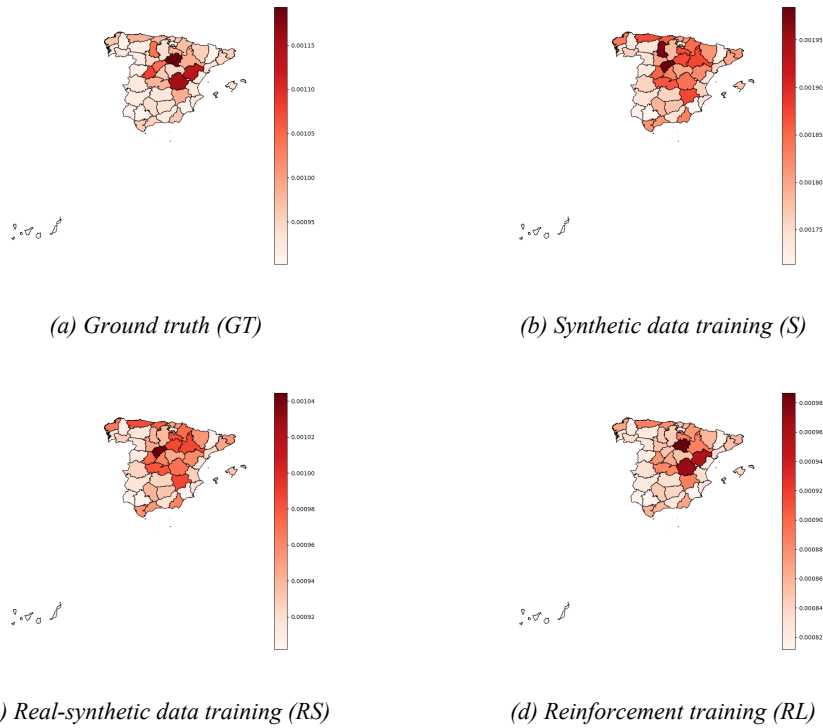


Figure 8: Map showing the spatial distribution of the CVRMSE per province for the 4 benchmarks given the CNLSTM model for the weekly aggregation of flows

Acknowledgements

This work is derived from R&D projects RTC2019-007159-5, as well as the Ramon y Cajal Grant RYC2018-025580-I, funded by MCIN/AEI/10.13039/501100011033, “FSE invest in your future” and “ERDF A way of making Europe” and the grant PID2020-112827GB-I00 funded by MCIN/AEI/10.13039/501100011033.

Availability of data and materials

All data and materials are available on request from the authors of this paper.

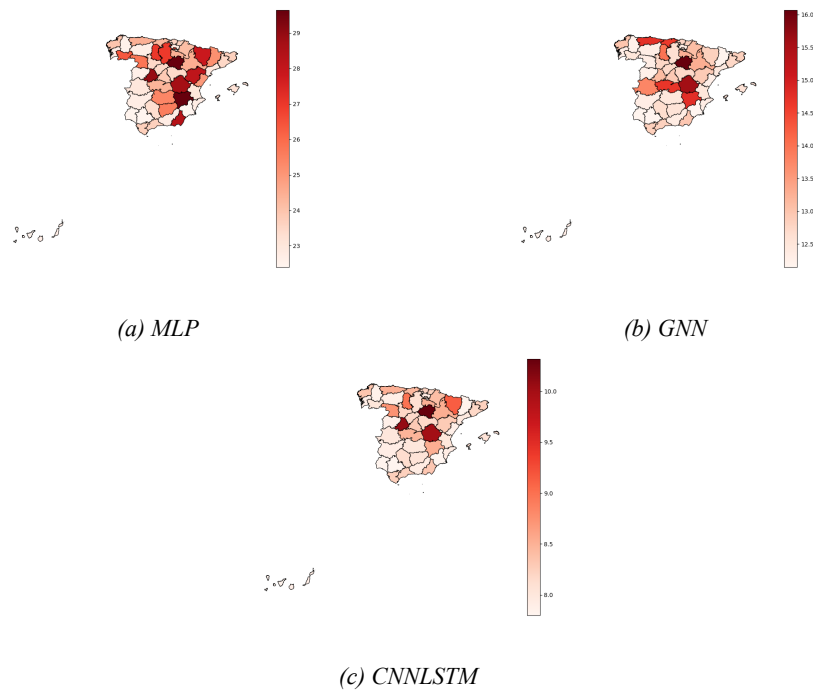


Figure 9: Map showing the spatial distribution of the MAPE per province for the 3 models when they are benchmarked with the Reinforcement Learning approach given the weekly aggregation of the flows

References

- [Alexandersson, 1986] Alexandersson, H. (1986). A homogeneity test applied to precipitation data. *Journal of climatology*, 6(6):661–675.
- [Bao et al., 2020] Bao, H., Zhou, X., Zhang, Y., Li, Y., and Xie, Y. (2020). Covid-gan: Estimating human mobility responses to covid-19 pandemic through spatio-temporal conditional generative adversarial networks. In *Proceedings of the 28th International Conference on Advances in Geographic Information Systems, SIGSPATIAL '20*, page 273–282, New York, NY, USA. Association for Computing Machinery.
- [Barbierato et al., 2022] Barbierato, E., Vedova, M. L. D., Tessera, D., Toti, D., and Vanoli, N. (2022). A methodology for controlling bias and fairness in synthetic data generation. *Applied Sciences*, 12(9):4619.
- [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.
- [Belhadi et al., 2020] Belhadi, A., Djenouri, Y., Srivastava, G., Djenouri, D., Cano, A., and Lin, J. C.-W. (2020). A two-phase anomaly detection model for secure intelligent transportation ride-hailing trajectories. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4496–4506.

- [Bishop, 2006] Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- [Buishand, 1982] Buishand, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of hydrology*, 58(1-2):11–27.
- [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.
- [Chatziagapi et al., 2019] Chatziagapi, A., Paraskevopoulos, G., Sgouropoulos, D., Pantazopoulos, G., Nikandrou, M., Giannakopoulos, T., Katsamanis, A., Potamianos, A., and Narayanan, S. (2019). Data augmentation using gans for speech emotion recognition. In *Interspeech*, pages 171–175.
- [Fan et al., 2018] Fan, Z., Song, X., Xia, T., Jiang, R., Shibasaki, R., and Sakuramachi, R. (2018). Online Deep Ensemble Learning for Predicting Citywide Human Mobility. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 2(3).
- [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.
- [Haykin and Lippmann, 1998] Haykin, S. and Lippmann, R. (1998). *Neural Networks: A Comprehensive Foundation*. Prentice Hall PTR, 2nd edition.
- [Hochreiter and Schmidhuber, 1997] Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- [Jiang et al., 2019] Jiang, W., Hong, Y., Zhou, B., He, X., and Cheng, C. (2019). A gan-based anomaly detection approach for imbalanced industrial time series. *IEEE Access*, 7:143608–143619.
- [Kwak and Geroliminis, 2020] Kwak, S. and Geroliminis, N. (2020). Travel time prediction for congested freeways with a dynamic linear model. *IEEE Transactions on Intelligent Transportation Systems*, pages 1–11.
- [Lee et al., 2021] Lee, C.-K., Cheon, Y.-J., and Hwang, W.-Y. (2021). Studies on the gan-based anomaly detection methods for the time series data. *IEEE Access*, 9:73201–73215.
- [Li et al., 2020] Li, C., Ge, P., Liu, Z., and Zheng, W. (2020). Forecasting tourist arrivals using denoising and potential factors. *Annals of Tourism Research*, 83:102943.
- [Li et al., 2019] Li, D., Chen, D., Jin, B., Shi, L., Goh, J., and Ng, S.-K. (2019). Mad-gan: Multivariate anomaly detection for time series data with generative adversarial networks. In *International conference on artificial neural networks*, pages 703–716. Springer.
- [Li et al., 2021] Li, Z., Liu, F., Yang, W., Peng, S., and Zhou, J. (2021). A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE Transactions on Neural Networks and Learning Systems*.
- [Liang et al., 2023] Liang, H., Perona, P., and Balakrishnan, G. (2023). Benchmarking algorithmic bias in face recognition: An experimental approach using synthetic faces and human evaluation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4977–4987.
- [Lin et al., 2020] Lin, Z., Jain, A., Wang, C., Fanti, G., and Sekar, V. (2020). Using gans for sharing networked time series data: Challenges, initial promise, and open questions. In *Proceedings of the ACM Internet Measurement Conference, IMC '20*, page 464–483, New York, NY, USA. Association for Computing Machinery.

- [Luca et al., 2021] Luca, M., Barlacchi, G., Lepri, B., and Pappalardo, L. (2021). A survey on deep learning for human mobility. *ACM Computing Surveys (CSUR)*, 55(1):1–44.
- [Mathieu et al., 2020] Mathieu, E., Ritchie, H., Rodés-Guirao, L., Appel, C., Giattino, C., Hasell, J., Macdonald, B., Dattani, S., Beltekian, D., Ortiz-Ospina, E., and Roser, M. (2020). Coronavirus pandemic (covid-19). *Our World in Data*. <https://ourworldindata.org/coronavirus>.
- [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).
- [Niu et al., 2020] Niu, Z., Yu, K., and Wu, X. (2020). Lstm-based vae-gan for time-series anomaly detection. *Sensors*, 20(13):3738.
- [Ouyang et al., 2018] Ouyang, K., Shokri, R., Rosenblum, D. S., and Yang, W. (2018). A non-parametric generative model for human trajectories. In *IJCAI*, volume 18, pages 3812–3817.
- [Pappalardo et al., 2016] Pappalardo, L., Rinzivillo, S., and Simini, F. (2016). Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934–939.
- [Pettit, 1979] Pettit, A. (1979). A non-parametric approach to the change-point problem. *Applied statistics*, 28(2):126–135.
- [Ramponi et al., 2018] Ramponi, G., Protopoulos, P., Brambilla, M., and Janssen, R. (2018). T-cgan: Conditional generative adversarial network for data augmentation in noisy time series with irregular sampling. *arXiv preprint arXiv:1811.08295*.
- [Secretaría de Estado de Infraestructuras, Transporte y Vivienda, 2019] Secretaría de Estado de Infraestructuras, Transporte y Vivienda (2019). Estudio de la movilidad interprovincial de viajeros aplicando la tecnología Big Data - proyecto piloto - informe metodológico. Technical report, Ministerio de Fomento.
- [Secretaría de Estado de Transportes, Movilidad y Agenda Urbana, 2021] Secretaría de Estado de Transportes, Movilidad y Agenda Urbana (2021). 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.
- [Shahul Hameed et al., 2024] Shahul Hameed, M. A., Qureshi, A. M., and Kaushik, A. (2024). Bias mitigation via synthetic data generation: A review. *Electronics (2079-9292)*, 13(19).
- [Smolyak et al., 2020] Smolyak, D., Gray, K., Badirli, S., and Mohler, G. (2020). Coupled igmm-gans with applications to anomaly detection in human mobility data. *ACM Trans. Spatial Algorithms Syst.*, 6(4).
- [Solmaz and Turgut, 2019] Solmaz, G. and Turgut, D. (2019). A survey of human mobility models. *IEEE Access*, 7:125711–125731.
- [Sorin et al., 2020] Sorin, V., Barash, Y., Konen, E., and Klang, E. (2020). Creating artificial images for radiology applications using generative adversarial networks (gans)—a systematic review. *Academic radiology*, 27(8):1175–1185.
- [Terroso-Sáenz and Muñoz, 2022] Terroso-Sáenz, F. and Muñoz, A. (2022). Nation-wide human mobility prediction based on graph neural networks. *Applied Intelligence*, 52(4):4144–4160.
- [Van Lint and Van Hinsbergen, 2012] Van Lint, J. and Van Hinsbergen, C. (2012). Short-term traffic and travel time prediction models. *Artificial Intelligence Applications to Critical Transportation Issues*, 22(1):22–41.

- [Wang et al., 2017] Wang, J., Kong, X., Rahim, A., Xia, F., Tolba, A., and Al-Makhadmeh, Z. (2017). Is2fun: Identification of subway station functions using massive urban data. *IEEE Access*, 5:27103–27113.
- [Wang et al., 2021a] Wang, W., Chen, J., Zhang, Y., Gong, Z., Kumar, N., and Wei, W. (2021a). A multi-graph convolutional network framework for tourist flow prediction. *ACM Transactions on Internet Technology (TOIT)*, 21(4):1–13.
- [Wang et al., 2021b] Wang, X., Liu, X., Lu, Z., and Yang, H. (2021b). Large scale gps trajectory generation using map based on two stage gan. *Journal of Data Science*, 19(1):126–141.
- [Wu et al., 2024] Wu, X., Zhang, Z., Wan, W., and Yao, S. (2024). Personalized behavior modeling network for human mobility prediction. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13.
- [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.
- [Xu et al., 2022] Xu, K., Ganev, G., Joubert, E., Davison, R., Van Acker, O., and Robinson, L. (2022). Synthetic data generation of many-to-many datasets via random graph generation. In *The Eleventh International Conference on Learning Representations*.
- [Yang et al., 2018] Yang, G., Wang, Y., Yu, H., Ren, Y., and Xie, J. (2018). Short-term traffic state prediction based on the spatiotemporal features of critical road sections. *Sensors*, 18(7).
- [Yoon et al., 2019] Yoon, J., Jarrett, D., and Van der Schaar, M. (2019). Time-series generative adversarial networks. *Advances in neural information processing systems*, 32.
- [Zhang et al., 2021] Zhang, K., He, Z., Zheng, L., Zhao, L., and Wu, L. (2021). A generative adversarial network for travel times imputation using trajectory data. *Computer-Aided Civil and Infrastructure Engineering*, 36(2):197–212.
- [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.
- [Zhou et al., 2021] Zhou, Z., Wang, Y., Xie, X., Qiao, L., and Li, Y. (2021). Stuanet: Understanding uncertainty in spatiotemporal collective human mobility. In *Proceedings of the Web Conference 2021*, pages 1868–1879.
- [Zhou et al., 2023] Zhou, Z., Yang, K., Liang, Y., Wang, B., Chen, H., and Wang, Y. (2023). Predicting collective human mobility via countering spatiotemporal heterogeneity. *IEEE Transactions on Mobile Computing*.