OPEN ACCESS
International Journal of Information Technology & Decision Making
(2025) 1–33

DOI: 10.1142/S0219622026500185

Tourist Mobility Forecasting with Region-Based Flows and Hierarchical Spatial Tessellation

Fernando Terroso-Saenz *

Department of Information and Communication Technologies
Technical University of Cartagena, Plaza del Hospital, 30202
Cartagena, Spain
fernando.terroso@upct.es

Juan Morales-García

Department of Software and Computing Systems
University of Alicante, Alicante, Spain

Miguel Puig-Cabrera

Research Centre for Tourism, Sustainability and Well-being (CinTurs)
Universidade do Algarve, Portugal

Ginesa Martinez-del Vas

Faculty of Economics and Business
Universidad Catolica de Murcia (UCAM), Murcia, Spain

Andres Muñoz

Department of Computer Engineering
University of Cádiz, Cadiz, Spain

Received 14 December 2023
Revised 3 October 2025
Accepted 10 November 2025
Published 19 December 2025

This paper introduces a novel approach to tourist mobility prediction based on Graph Neural Networks (GNNs) trained with general human mobility (GMD) data, evaluating its performance through multiple spatial scales. By using the Region of Murcia (Spain) as a case study, we demonstrate that enriching GNNs with GMD flows significantly improves prediction accuracy compared to univariate time-series models and CNNLSTM baselines. Specifically, the results reveal that incorporating the total number

*Corresponding author.

of visitors and overnight tourists in our model significantly improves the mobility prediction accuracy. In contrast, the benefits for including excursionist flows are limited to short-term forecasts only. Moreover, the improvement in tourist flow prediction is more evident at coarser spatial scales compared to finer municipal areas, suggesting that the utility of GMD is dependent on the spatial granularity of the target region. These findings can be leveraged to inform policy-making and large-scale tourism management.

Keywords: Tourism; human mobility; prediction; open data.

1. Introduction

The tourism industry, a significant global economic driver, benefits from advances in Information and Communication Technologies (ICT) that enhance data collection and analysis, improving services and strategic planning. One critical research area focuses on predicting tourist flows,[1] essential for optimizing marketing, infrastructure, and service provision. Despite progress, two challenges persist:

- Many models rely on univariate time-series approaches, limiting the integration of diverse human mobility datasets that could improve forecasting accuracy. While time-based methods predict tourist trips in specific regions,[2,3] incorporating alternative mobility data as *exogenous* variables remains underexplored. These variables, such as commuter flows, enrich forecasts by revealing broader movement patterns beyond historical tourism data. In this regard, public and research institutions increasingly release general mobility datasets,[4–6] yet their use in tourism prediction is limited.
- Most studies operate at fixed spatial scales, using static region divisions that fail to capture dynamic mobility patterns. Practical scenarios often require forecasts across multiple geographic levels, from cities to neighborhoods. Multi-scale forecasting models are essential for adapting infrastructure, marketing, and services, yet few studies assess their accuracy across hierarchical spatial scales.

To tackle these challenges, we integrate multiple mobility datasets into predictive models, evaluating their performance at different spatial granularities. We explore how combining region-based mobility data enhances tourist flow prediction accuracy. This leads to two Research Questions (RQs):

- RQ1: Can general human mobility feeds improve predictions of touristic flows toward a region?
- RQ2: Does this improvement vary with the spatial scale of the target region?

The rationale behind these questions is that general mobility trends may offer complementary insights into tourist behavior. For example, a surge in tourist arrivals could follow a decline in commuter traffic due to a major event. The usefulness of these patterns may differ across neighborhood, city, or national scales.

To address these questions, we use two mobile-phone-based human mobility datasets: one capturing tourist movements within the Region of Murcia (RM), Spain, and another reflecting nationwide mobility patterns. We evaluate univariate and multivariate models predicting tourist arrivals in RM across different spatial

divisions. To effectively handle these multi-scale predictions, we employ Graph Neural Network (GNN) architectures. GNNs are employed to manage multi-scale predictions, effectively modeling spatial dependencies between locations. Unlike traditional models assuming geographic independence, GNNs propagate information across connected regions, enhancing prediction robustness. Our goal is to assess how time-based tourist mobility forecasting is influenced by (1) the type of mobility data used and (2) the granularity of the target regions.

In addition to its methodological contribution, this study provides actionable insights for the formulation of evidence-based tourism policies. Integrating general human mobility data improves the ability of DMOs and regional authorities to generate accurate and scalable forecasts, anticipate pressure points in visitor flows, and align tourism and mobility strategies. This approach supports more effective planning decisions, such as optimizing resource allocation, targeting the activation of specific attractions based on their expected impact on visitor dispersion, and identifying potential congestion hotspots. The proposed framework thus bridges advanced data science and real-world instruments for territorial governance and strategic tourism development.

This paper is arranged as follows. Section 2 summarizes the most relevant current approaches for human mobility prediction in the touristic sector. Then, Sec. 3 describes the use-case setting of our study. In Sec. 4, the most important results of the deployed palette of predictors are described and evaluated. Lastly, Sec. 5 summarizes the main conclusions and potential future research lines motivated by this work.

2. Related Work

Recent studies in tourist flow prediction increasingly rely on machine learning and statistical models to analyze historical data and identify mobility patterns. The accuracy of these models largely depends on the input data, which typically includes tourist flows, such as transportation records, social media activity, geolocation data, and tourist surveys. Below, we categorize key contributions in the field based on the datasets employed in predictive models.

For univariate datasets, the most common input datasets, the most common input is the incoming tourism flow to a specific region. In Ref. 7, ARMA-based models (ARIMA, ARAR, ARFIMA) were used to forecast the number of visitors in the major tourist destinations in the Asia-Pacific region (Hong Kong, Japan, Korea, Taiwan, Singapore, Thailand, the Philippines, Australia, and New Zealand), achieving less than 2% in terms of MAPE.

Another work using a similar approach is Ref. 8, where an ARMA model forecasts the future five-year trend of tourism in China. In Ref. 9, an LSTM addresses the weakness of ARMA-based models in capturing the stochastic and nonlinear nature of tourist flows. Results show that LSTM outperforms ARMA-based methods (ARIMA and BPNN). In Ref. 10, the CIR# model was applied to predict the number of nights spent in tourist accommodations in Italy on a monthly basis,

achieving a MAPE of 1.18%. Moreover, the same model was applied to forecast changes in the arrivals of foreign tourists in Italy, achieving a MAPE of 0.0068 in Ref. 11.

The work in Ref. 12 compares neural networks and time series models for forecasting tourism in Catalonia (Spain) demonstrating that time series models perform better for short-term predictions, whereas neural networks excel in capturing nonlinear trends.

In Ref. 13, a novel approach proposed the conversion of time series data into sequence images as the primary input to a CNNLSTM model. This method enhances the capture of temporal and local patterns of tourist demand. This proposal was tested on two Chinese attractions, Jiuzhaigou and Mount Siguniang, demonstrating improved accuracy even in unusual events such as the COVID-19 outbreak. Finally, Petropoulos and Spiliotis[14] evaluate five different univariate models (Theta, MTA, Bagging, FOSS and MSP) and demonstrate the challenges of univariate forecasting, emphasizing the limitations of time-series models like ETS and ARIMA in accurately predicting tourism demand.

On the other hand, there are several studies that evaluate the use of multivariate datasets. For example, Zhu *et al.*[15] analyzed tourist flows from six countries to predict tourist arrivals in Singapore.

The authors pair the flows from Canada–US, Japan–South Korea, and Malaysia–Thailand using a copula-based model to model dependencies among geographically related countries. Their results demonstrate that pairwise modeling improves MAPE and RMSE values compared to independent flow-based models, particularly for annual predictions.

Yang and Zhang[16] compared spatial and temporal datasets of tourist flows in China against purely spatial mobility data. Their study evaluates univariate ARIMA models versus the spatio-temporal STARMA model, using tourist mobility data from 29 Chinese regions (1987–2016). Results show improved accuracy for STARMA, particularly in neighboring areas with strong spatial associations.

Zhang *et al.*[17] incorporated tourism flow volume and Google Trends search interest indicators (SII) (e.g., "Hong Kong metro", "Hong Kong airport") to improve machine learning and deep learning models for forecasting tourist arrivals over one, three, and six months. Similarly, Law *et al.*[18] applied SII in Macau, China, evaluating various deep learning models with attention mechanisms, which outperform support vector regression models.

Another approach leverages contextual information as a complementary source for tourist flow data. In Ref. 19, data from Points of Interest (POIs) in Verona, Italy (2014–2019) are augmented with weather conditions, holidays, time slots, and special events (e.g., Valentine's Day, Easter). This enhances forecast accuracy using deep learning models such as Random Forest and Deep Neural Networks. An extension of this work in Ref. 20 uses the same contextual features to recommend new POIs, incorporating user preferences into the prediction model.

Building on these previous works, the key novelty of the present study is related to lies in the nature of the data used to enhance tourist mobility prediction.

Unlike prior studies that rely on web-based indicators,[17] spatial data[16] or indirect flows[15] as exogenous variables, this study directly incorporates human mobility data extracted from an open, general-purpose dataset covering the geographical area under analysis. Moreover, our study evaluates the impact of spatial tessellation granularity, considering how different geographic area definitions affect prediction accuracy. To the best of our knowledge, this is the first study to conduct a multi-scale evaluation of tourist mobility prediction.

3. Setting Overview

The use case setting of our study has been the RM, a Spanish autonomous community located in the south-east of the country.

This region has a population of around 1.5 million people and comprises 39 municipalities in an area of 11,313 km^2. From a touristic point of view, this region received more than 800,000 tourists in 2022 and with an average growth of 5.4% for the last years.[21]

From a mobility-related approach, this area was chosen given the lack of an integrated transportation system along the region. Indeed, RM's public transport infrastructure is fragmented and limited, with urban bus lines under separate municipal control and interurban routes split into multiple concessions without a single integrated ticketing or scheduling system. Rail connections within the region are also limited. As a result, the majority of visitors rely on private vehicles for their travel. According to Ref. 22, two-thirds of tourists arrive by or use their own car to get around, making road transport the dominant mode for both incoming tourists and intra-regional visits. This implies that all the analyzed movements of tourists can be better profiled, as private transport becomes the main transport mode of tourists coming from other regions as well as those visiting a location within this territory.

In the following subsections, we describe the different datasets and the spatial hierarchy of the setting.

3.1. *Datasets*

For this study, we have used two different mobility datasets, one comprising the tourist mobility in RM and other comprising the global human mobility in the same region.

3.1.1. *Tourist Mobility Dataset* (*TMD*)

The movement of tourists in RM was collected by means of the Tourist Mobility Dataset (TMD) provided by the Tourism Institute of the RM[a] as part of its Smart Region project. This feed comprises the incoming and outgoing movement of

[a]https://www.itrem.es/nexo/dti/

tourists for each of the 39 cities in RM during a 16-month period from January 1st, 2022 to April 30th, 2023. These flows are based on the *network events* generated by the mobile-phones connected to the Telefonica network, one of the most important carriers in Spain.[b] Next, an anonymization, extrapolation and aggregation step was performed to compute the final touristic flows included in the dataset.

One important detail of this dataset is that it distinguishes, for each city c in RM, its incoming flow of national (living in Spain) and international (coming from other countries) *tourists* ($\mathcal{NT}$, $\mathcal{IT}$) and *excursionists* ($\mathcal{NE}$, $\mathcal{IE}$). The former flow refers to people who stay at least one night in c and the latter refers to *day trippers* who spend the day in the city but eventually return to their homes.

Given the aforementioned flows, we computed three aggregated ones comprising the overall number of tourists $\mathcal{T}$ $(= \mathcal{NT} + \mathcal{IT})$, the overall number of excursionists $\mathcal{E}$ $(= \mathcal{NE} + \mathcal{IE})$ and the overall number of visitors $\mathcal{A}$ $(= \mathcal{T} + \mathcal{E})$.

In terms of format, this dataset defines the flows on a weekly basis. In this manner, the mobility of the wth week of the year y for a city c is defined as a single tuple comprising six flow features:

$$\langle y, w, c, f^m_{\text{work}}, f^a_{\text{work}}, f^n_{\text{work}}, f^m_{\text{end}}, f^a_{\text{end}}, f^n_{\text{end}} \rangle,$$

where $f^m_{\text{work}}, f^a_{\text{work}}$ and f^n_{work} are the *week slices* comprising the overall incoming flows towards c during the morning (m), afternoon (a) and night (n), respectively, considering all the working days of w. Similarly, $f^m_{\text{end}}, f^a_{\text{end}}$ and f^n_{end} provide the same information for weekend days (Saturday and Sunday in Spain). As a result, for each combination of year, week and city (y, w, c), the dataset comprises eight different tuples, one for each type of flow, $\mathcal{NE}$, $\mathcal{NT}$, $\mathcal{IE}$, $\mathcal{IT}$, $\mathcal{T}$, $\mathcal{E}$ and $\mathcal{A}$.

For the sake of clarity, Fig. 1 illustrates the seasonal decomposition of the overall number of visitors, $\mathcal{A}$, for the 420 time steps (70 weeks $\times$6 records per week) based on an additive model, capturing potential weekly cyclic patterns. The original series (top panel) reveals regular short-term fluctuations along with several abrupt level shifts, most notably around day 200. The trend component (second panel) reflects the broader structural evolution of the series, with discernible increases and plateaus throughout the period, notably, a sustained upward movement between days 200 and 250, followed by a plateau toward the end of the observation period. The seasonal component (third panel) highlights a highly regular and repeating pattern with a clear 7-day periodicity, indicating a consistent weekly rhythm in trip volume (potentially driven by weekday/weekend effects). The residuals (bottom panel) exhibit stochastic behavior with several noticeable spikes, which may correspond to exogenous events or irregular disruptions not accounted for by the trend or seasonal structure. Overall, the decomposition suggests that a substantial proportion of the series' variability is explained by its trend and seasonal components, underscoring the importance of incorporating these elements into any predictive modeling framework. Furthermore, this decomposition confirms the presence of strong weekly

[b] https://www.telefonica.com/es/

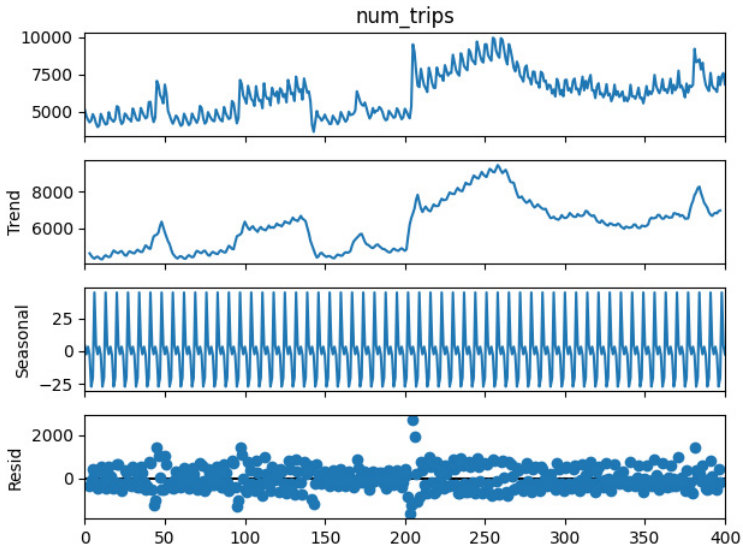


Fig. 1. Seasonal decomposition of the overall number of visitors, $\mathcal{A}$. The first panel shows the original time series, which includes the observed data over time with all inherent fluctuations and patterns. The second panel presents the trend, representing the long-term progression or general direction of the series, smoothing out short-term variations. The third panel depicts the seasonality, capturing the repeating and periodic patterns that occur at regular intervals, such as weekly cycles. Lastly, the fourth panel displays the residuals, which consist of the random variations or noise remaining after removing the trend and seasonal components, highlighting irregularities or unexplained events within the data.

seasonality in the data and supports the use of models that account for short-term periodic behavior in forecasting applications.

3.1.2. *General Human Mobility Dataset* (*GMD*)

This dataset has been retrieved from the nation-wide human mobility report initially released by the Spanish Ministry of Transportation (SMT) in January 2022.[c] It indicates the number of trips among 2735 cities 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 following form:

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

reporting that there was n_{trp} human trips from the city m_{origin} to the city m_{dest} with a_{dest} and whose distance was dist km during the indicated date and hour.

[c]https://www.mitma.es/ministerio/proyectos-singulares/estudios-de-movilidad-con-big-data/estudio-basico-diario

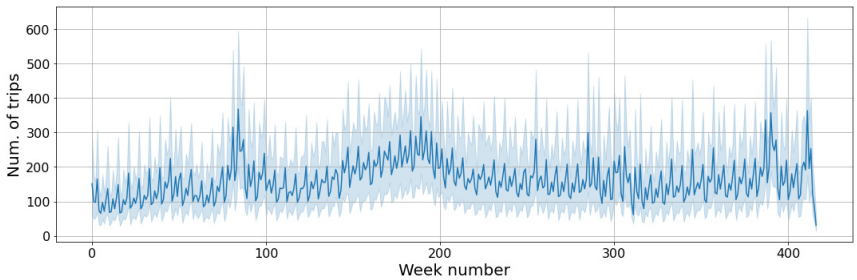


Fig. 2. (Color online) General flow of incoming trips to the RM cities during the period of study considering the general human mobility dataset. The dark blue line indicates the average value of all the stations whereas the bluish area around the line indicates the 95% confidence interval.

According to the official documents,[23] these mobility data were 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 captures the movement of people regardless of the means of transport used for their displacements. Figure 2 shows the overall number of incoming trips to all the cities in RM during the same time period considered for the touristic dataset (2022/01/01–2023/04/30).

As observed, the two datasets provide a region-based view of the setting, however the scope of the human mobility captured by each one is very different. Consequently, they define different points of view of the human mobility in the RM so that they can be used to answer the RQ1 of this work.

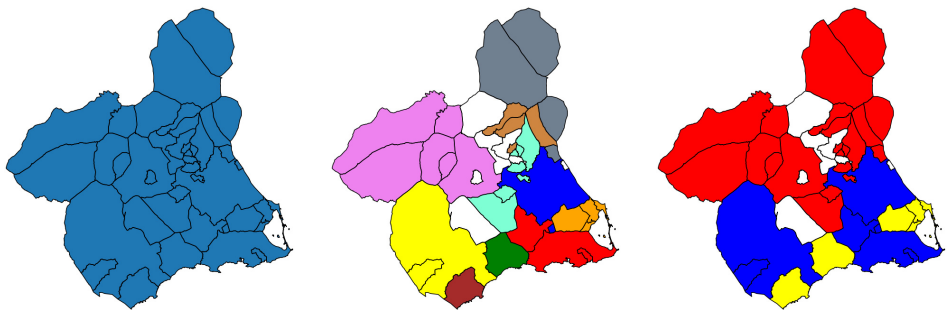
3.2. *Spatial hierarchy*

In this study, we have also used a 3-layer spatial hierarchy, where each layer defined a particular spatial tessellation of RM. Figure 3 shows the spatial distribution of these three levels.

To begin with, layer 1 is composed of the 39 municipalities ($\mathcal{M}$) of RM (Fig. 3(a)). The two mobility feeds described in Sec. 3.1 are defined on top of this first layer as they indicate the incoming and outgoing flows at city level.

The second level is based on the aggregation of municipalities, to compose a new spatial entity coined *destinations* ($\mathcal{D}$). These destinations are defined by the Tourism Local Office (TLO) of RM.[d] This aggregation merges the cities based on their spatial closeness or their similarity with respect to the touristic offer and infrastructure. As observed in Fig. 3(b), this level defines 10 destinations. The white areas indicate municipalities that are not included in any destination. Therefore, they are not considered at this hierarchy level.

[d]https://www.turismoregiondemurcia.es/en/tourism_statistics/?pagina=destinos-y-localidades



(a) Level 1 - Municipalities.

(b) Level 2 - Destinations (aggregation of municipalities). Each destination is shown with a particular color.

(c) Level 3 - Destination types (aggregation of destinations). Each destination type is shown with a particular color.

Fig. 3. Spatial distribution of the three tessellation levels composing the spatial hierarchy of the setting.

Table 1. Average population and size of the areas for each level of the spatial taxonomy.

Level	Population per area	Area size (km^2)	Number of areas
1 - Municipalities ($\mathcal{M}$)	40,238 ($\pm$80, 393)	289.78 ($\pm$370)	39
2 - Destinations ($\mathcal{D}$)	142,806.9 ($\pm$132, 082)	1,029.1 ($\pm$914)	10
3 - Destination types ($\mathcal{DT}$)	476,023.0 ($\pm$ 327,558.8)	3,430.3 ($\pm$ 2,490)	3

Finally, the same TLO further aggregates the aforementioned destinations into 3 *destination types* ($\mathcal{DT}$), each one defining a particular touristic profile within RM, namely *Coast* comprising the destinations in $\mathcal{D}$ whose most important tourist activity comes from sun-and-beach visitors (yellow areas in Fig. 3(c)), *Interior* comprising the destinations with an important rural-tourism sector (red areas in Fig. 3(c)) and *City* which includes the destinations whose primary touristic offer is related to an urban context (blue areas in Fig. 3(c)).

Not that these levels are defined following an administrative and tourist management criteria. The rationale of this approach, instead of a data-driven one, is to use the spatial tessellations that already use public and private stakeholders to analyzed the touristic activity of RM. In operational terms, this allows for a smoother integration of the tourist-mobility predictions of our approach to the existing ecosystem of data-analysis platforms used by the touristic sector in RM. Table 1 shows the average population and size of the areas given the three aforementioned levels.

Given this hierarchy and the TMD (Sec. 3.1.1), we composed a time series for each flow $\mathcal{F} \in \langle \mathcal{NE}, \mathcal{NT}, \mathcal{IE}, \mathcal{IT} \rangle$ and spatial scale $\mathcal{S} \in \langle \mathcal{M}, \mathcal{D}, \mathcal{DT} \rangle$ covering the 420 time steps under study with the following format $\mathcal{F}^{\mathcal{S}}_{\text{TM}} = \langle f^{m,1}_{\text{work}} \rightarrow f^{a,1}_{\text{work}} \rightarrow f^{n,1}_{\text{work}} \rightarrow f^{m,1}_{\text{end}} \rightarrow f^{a,1}_{\text{end},c} \rightarrow f^{n,1}_{\text{end}} \rightarrow f^{m,2}_{\text{work}} \rightarrow f^{a,2}_{\text{work}} \rightarrow f^{n,2}_{\text{work}} \rightarrow f^{m,2}_{\text{end}} \rightarrow f^{a,2}_{\text{end}} \rightarrow f^{n,2}_{\text{end}} \rightarrow \ldots \rightarrow f^{m,420}_{\text{work}} \rightarrow f^{a,420}_{\text{work}} \rightarrow f^{n,420}_{\text{work}} \rightarrow f^{m,420}_{\text{end}} \rightarrow f^{a,420}_{\text{end}} \rightarrow f^{n,420}_{\text{end}} \rangle$ where,

for example, $f^{m,i}_{\text{work}} = \langle f^{m,i}_{\text{work},r_1}, f^{m,i}_{\text{work},r_2}, \ldots, f^{m,i}_{\text{work},r_{|\mathcal{S}|}} \rangle$ is the record comprising the overall number of incoming tourists or excursionists toward each region r_j defined at the $\mathcal{S}$ scale during the morning of the working days of the ith week.

Finally, the definition of this hierarchy allows us to evaluate RQ2. Note that each level is based on an aggregation of the previous one, so that each one provides a different granularity in terms of area size, population density, and touristic profile.

3.3. *Correlation study*

Given the datasets and spatial hierarchy described in the previous section, the next step was to study the similarity between the touristic and general mobility flows. To do so, we first cleaned the mobility flows of GMD by just keeping the records accomplishing the following two conditions: (1) their destination m_{dest} was one of the cities in $\mathcal{M}$ and (2) their distance dist is above a certain threshold δ. The first condition captures the incoming flows of the target cities and the second condition allows to refine such incoming flows based on the distance traveled by visitors to arrive to one of the cities.

Given the fact that the peninsular area of Spain has an approximate radius of 540 km and the average distance traveled by commuters in Spain range from 19 to 34 km,[24] we used three different δ values, 100, 400 and 800 km. By means of these three filters, we firstly generated a subflow comprising short-distance trips with $\delta = 100$ km ($\mathcal{F}^{100}_{\text{GM}}$). Hence, this flow captured some regular and non-touristic trips along with all types of tourist flows toward RM. Next, we also composed a second subflow covering medium-distance trips with $\delta = 400$ km ($\mathcal{F}^{400}_{\text{GM}}$) and a third flow with long-distance travelers given $\delta = 800$ km ($\mathcal{F}^{800}_{\text{GM}}$). Thus, we were able to incrementally removed the proportion of regular and non-touristic trips in each subflow by increasing δ.

Each of the resulting subflows are defined as an ordered sequence of *week-slices* (like the TMD flows). For example, $\mathcal{F}^{100,\mathcal{M}}_{\text{GM}} = \langle f^{m,1}_{\text{work},gm,100} \rightarrow f^{a,1}_{\text{work},gm,100} \rightarrow f^{n,1}_{\text{work},gm,100} \rightarrow f^{m,1}_{\text{end},gm,100} \rightarrow f^{a,1}_{\text{end},c} \rightarrow f^{n,1}_{\text{end},gm,100} \rightarrow \cdots \rightarrow f^{m,420}_{\text{work},gm,100} \rightarrow f^{a,420}_{\text{work},gm,100} \rightarrow f^{n,420}_{\text{work},gm,100} \rightarrow f^{m,420}_{\text{end},gm,100} \rightarrow f^{a,420}_{\text{end},gm,100} \rightarrow f^{n,420}_{\text{end},gm,100} \rangle$ where $f^{m,i}_{\text{work},gm,100} = \langle f^{m,i}_{\text{work},gm,100,r_1}, f^{m,i}_{\text{work},gm,100,r_2}, \ldots, f^{m,i}_{\text{work},gm,100,r_{|\mathcal{S}|}} \rangle$ is the record comprising the overall number of incoming trips covering at least 100-km toward each region r_j defined at $\mathcal{S}$ scale during the morning of the working days of the ith week. Similarly, we composed the same flows at $\mathcal{D}$, and $\mathcal{DT}$ levels.

Next, a correlation study between the GM and TMD flows was performed by means of three different metrics, namely the Pearson's correlation coefficient (PCC), the Mutual Information Score (MIS) and the Granger Causality Test (GCT).

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 is just captures the linear correlation between the variables. For that reason, the MIS was also used as it allows to measure other types of nonlinear correlations.[25] MIS is a non-negative score for which higher values

mean higher dependency. The MIS between two variables X and Y is defined as

$$\mathrm{MIS}(X,Y) = H(X) + H(Y) - H(X,Y),$$

where H stands for the entropy, that is, the expected amount of information held in a variable.

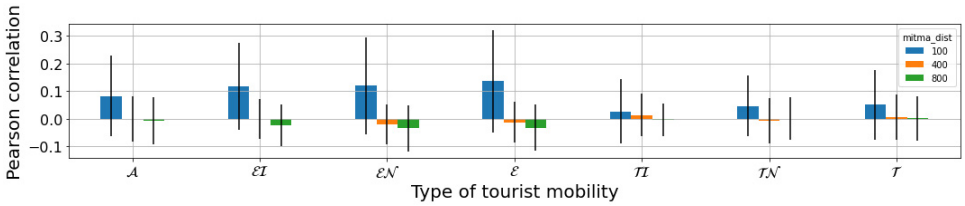
Eventually, GCT is a statistical test to determine if one time series (x) is helpful for predicting another (y).[26] It defines the null hypothesis that the coefficients of past values of x in the regression equation to predict y are zero. Hence, GCT compares an *unrestricted model*, in which the time series y is explained by the lags of y and the lags of an additional series of observations from x (both lags up to the same fixed order), to a *restricted model*, in which y is only explained by the lags of y. In our case, we used as metric the number of lags of the GMD flows that can be used to explain the TMD flows. Hence, the larger the number lags, the more able is the GMD flow to explain a particular TMD flow.

Figures 4–6 show the scores of these metrics for each flow type and spatial hierarchy. For the sake of completeness, Appendix A contains the tables with the values depicted in the bar plots.

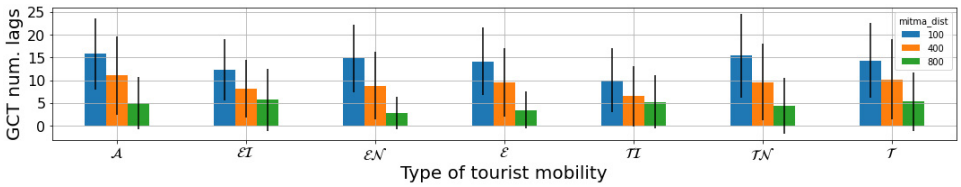
Regarding the $\mathcal{M}$ level, Fig. 4 shows that the highest PC, number of GCT lags and MI were achieved given the $\mathcal{F}_{\mathrm{GM}}^{100}$ flow. However, it is true that Fig. 4(a) shows that the PC scores were rather low when we considered the touristic-related flows ($\mathcal{T}$, $\mathcal{TN}$ and $\mathcal{TI}$) with values very close to 0 showing no linear correlation. The fact that the best results of $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{M}}$ were obtained when it was compared with the excursion-related flows actually makes sense as excursionists are expected to travel to short distances to spend a day in a particular spot. This also made flows $\mathcal{F}_{\mathrm{GM}}^{400,\mathcal{M}}$ and $\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{M}}$ to achieve poorest results as they cover longer-distance trips. Indeed, the other two metrics, MI and GCT lags (Figs. 4(b) and 4(c)), followed a similar approach where the best results are obtained with $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{M}}$ in almost all cases.

Concerning the $\mathcal{D}$ scale, Fig. 5 shows a slight increment of the correlation scores in all the configurations with respect to the previous level. Indeed, the PC of $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$ with respect to the global flow $\mathcal{A}$ moved from 0.083 at $\mathcal{M}$ level to 0.154. Again, $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$ achieved its highest PC scores given the excursionist flows. The flows with more distant trips, $\mathcal{F}_{\mathrm{GM}}^{400,\mathcal{D}}$ and $\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{D}}$ achieved higher PC scores given the tourist flows ($\mathcal{T}$, $\mathcal{TN}$ and $\mathcal{TI}$). This is also consistent with the idea that excursionists tend to travel shorter trips than tourists bearing in mind the definition of both terms in the TMD context. However, the MI values did not show such patterns as $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$ achieved the highest values for all the flows but $\mathcal{TN}$. This reflects that $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$ seem to better capture nonlinear relationships with the TMD flows than the other two counterparts at this level.

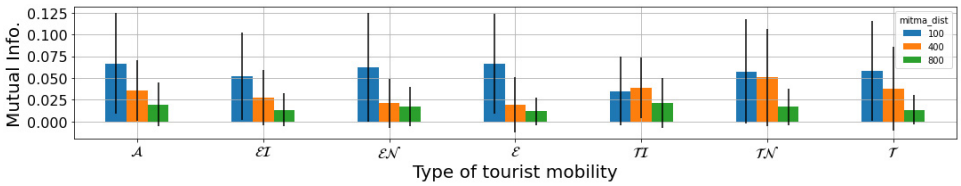
As far as the $\mathcal{DT}$ tessellation is concerned, Fig. 6 shows the same four patterns observed in the previous layer, (1) a general increment of the correlation scores (e.g., the PC between $\mathcal{F}_{\mathrm{GM},\mathcal{DT}}^{800}$ and $\mathcal{A}$ increased from 0.200 at $\mathcal{D}$ level to 0.276), (2) the highest PC scores are obtained with $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{DT}}$ and the excursionist flows, (3) the tourist flows are more correlated with $\mathcal{F}_{\mathrm{GM}}^{400,\mathcal{DT}}$ and $\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{DT}}$ (Fig. 6(a)) and (4)



(a) Pearson Correlation.



(b) Number of lags in the GCT.



(c) MIS.

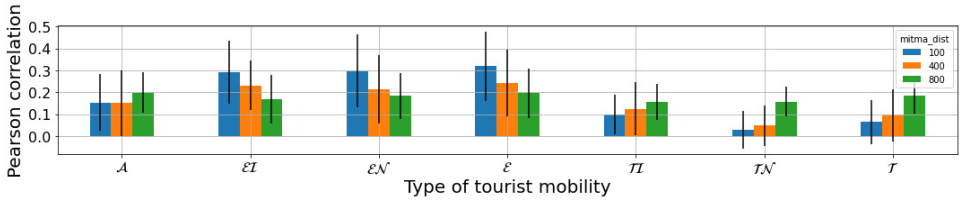
Fig. 4. Comparison scores between the TMD and GMD flows at $\mathcal{M}$ level.

a decrement of MI and the GCT lags as long as we move from $\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{DT}}$ to higher distance trips ($\mathcal{F}_{\mathrm{GM}}^{400,\mathcal{DT}}$ and $\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{DT}}$).

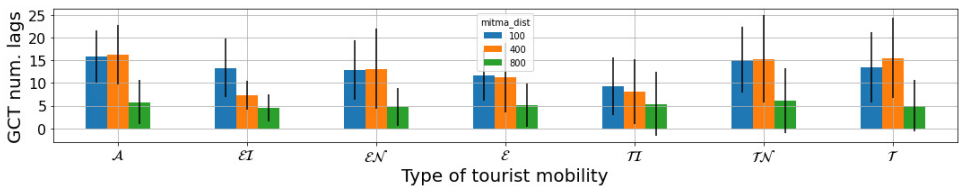
Bearing all these results in mind, we eventually kept with the pairs of TMD and GMD flows that achieved the highest correlation scores in each of the three spatial levels. The idea was to use such tuples to develop tourist predictors that integrate general human flows showing a promising behavior to improve the model's accuracy. These pairs are shown in Table 2 along with the reason of its selection. Each of these tuples gave rise to a particular touristic-flow predictor as indicated in Sec. 4. For the sake of completeness, Tables A.1–A.3 in Appendix A report the values of the three correlation metrics used in this study across the three target spatial levels.

4. Design of a Tourist Mobility Prediction for Multiple Spatial Scales

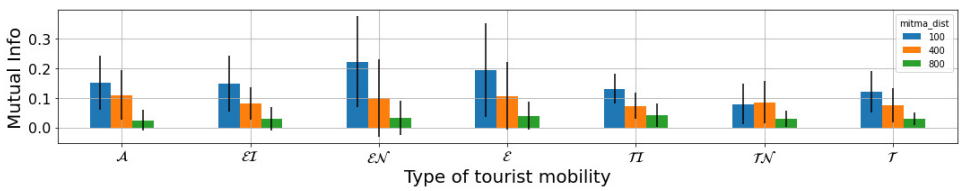
In this section, we describe the approach used to develop the different tourist flow predictors derived from the selection of human flows made in the previous section.



(a) Pearson Correlation.



(b) Number of lags in the GCT.



(c) MIS.

Fig. 5. Comparison scores between the TMD and GMD flows at $\mathcal{D}$ level.

4.1. *Problem formulation*

The tourist mobility prediction problem can be formulated as a temporal regression problem where the objective is to estimate future incoming tourist and excursionist flows based on historical mobility data. Given a temporal window w, the available past observations for a spatial region at a given scale S are defined as

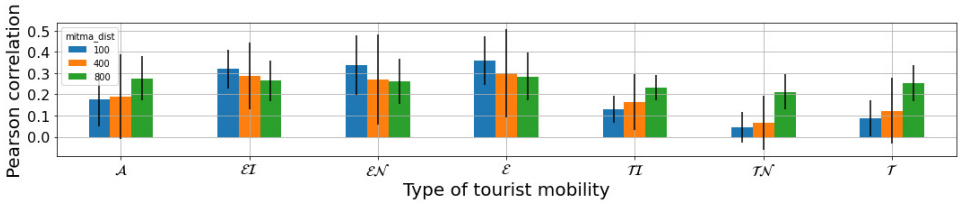
$$F_{\mathrm{TM}}^{w,S} = \{f_{\mathrm{TM}}^{w,S}, f_{\mathrm{TM}}^{w-1,S}, \ldots, f_{\mathrm{TM}}^{w-w_{\mathrm{prev}},S}\}, \tag{1}$$

where $F_{\mathrm{TM}}^{w,S}$ represents the time series of historical tourist mobility data at scale S, with w_{prev} being the number of previous time steps considered as input. Similarly, general human mobility data (GMD) for the same time window are denoted as

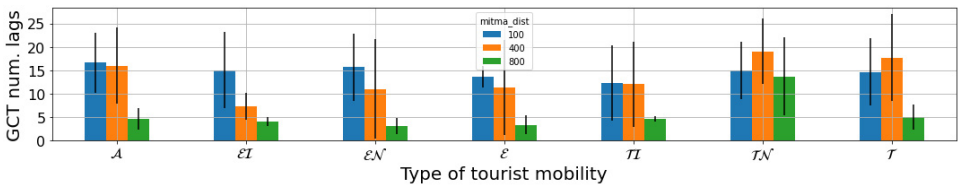
$$F_{\mathrm{GM}}^{w,S,\delta} = \{f_{\mathrm{GM}}^{w,S,\delta}, f_{\mathrm{GM}}^{w-1,S,\delta}, \ldots, f_{\mathrm{GM}}^{w-w_{\mathrm{prev}},S,\delta}\}, \tag{2}$$

where δ represents a distance threshold used to filter human mobility trips to remove local short-distance displacements.

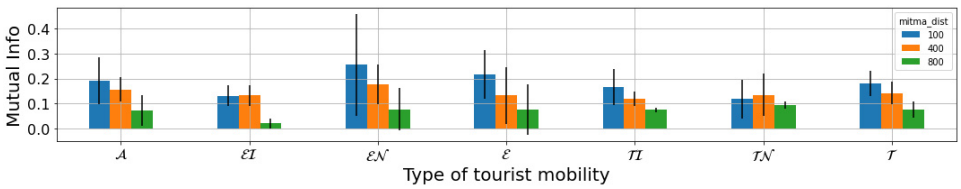
The objective of the model is to learn a function P that maps the historical tourist mobility data and general human mobility data to the future tourist mobility



(a) Pearson Correlation.



(b) Number of lags in the GCT.



(c) MIS.

Fig. 6. Comparison scores between the TMD and GMD flows at $\mathcal{DT}$ level.

for a given prediction horizon T:

$$P(F_{\mathrm{TM}}^{w,S}, F_{\mathrm{GM}}^{w,S,\delta}) \rightarrow F_{\mathrm{TM}}^{w+T,S}, \tag{3}$$

where $F_{\mathrm{TM}}^{w+T,S}$ represents the estimated number of tourists and excursionists arriving at each region in S at the future time step $w+T$.

By structuring the problem in this way, we provide a unified formulation that integrates multiple data sources and spatial hierarchies into a single predictive framework, leveraging GNNs to model spatial dependencies among regions.

4.2. *Definition of the latent graph*

In order to develop the tourist-mobility predictor, we opted for a GNN approach. This allows to define a particular spatial scale $\mathcal{S}$, given a week slice w and a δ value to filter the GMD flow, as a graph $\mathcal{G}_{w,\delta}^{\mathcal{S}} = \langle \mathcal{N}_{\mathcal{S}}^{w}, \mathcal{E}_{\mathcal{S}}, \mathcal{W}_{\mathcal{S}} \rangle$ where $\mathcal{N}_{\mathcal{S}}^{w} = \langle \mathcal{F}_{\mathrm{TM}}^{w,\mathcal{S}}, \mathcal{F}_{\mathrm{GM}}^{\delta,\mathcal{S},w} \rangle$ are the node features defined as the GM and TM flows at w, $\mathcal{E}_{\mathcal{S}} \in \mathbb{N}^{|\mathcal{S}|\times|\mathcal{S}|}$ is the adjacency matrix defining the edges connecting each region in $\mathcal{S}$ with all the other

Table 2. Tuples of TMD and GMD flows selected to develop the tourist predictors.

Level	TMD	GMD	Reason
$\mathcal{M}$	$\mathcal{A}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{M}}$	Highest MI at $\mathcal{M}$ level
	$\mathcal{E}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{M}}$	Highest PC at $\mathcal{M}$ level
	$\mathcal{IE}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{M}}$	Second highest PC at $\mathcal{M}$ level
$\mathcal{D}$	$\mathcal{T}$	$\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{D}}$	Highest PC among the touristic flows at $\mathcal{D}$ level
	$\mathcal{E}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$	Highest PC at $\mathcal{D}$ level
	$\mathcal{NE}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$	Second highest PC at $\mathcal{D}$ level
	$\mathcal{IE}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{D}}$	Third highest PC at $\mathcal{D}$ level
$\mathcal{DT}$	$\mathcal{A}$	$\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{DT}}$	Highest PC with $\mathcal{A}$ in the 3 levels.
	$\mathcal{T}$	$\mathcal{F}_{\mathrm{GM}}^{800,\mathcal{DT}}$	Highest PC among the touristic flows at $\mathcal{DT}$ level
	$\mathcal{E}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{DT}}$	Highest PC at $\mathcal{DT}$ level
	$\mathcal{NE}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{DT}}$	Second highest PC at $\mathcal{DT}$ level
	$\mathcal{IE}$	$\mathcal{F}_{\mathrm{GM}}^{100,\mathcal{DT}}$	Third highest PC at $\mathcal{DT}$ level

regions in the same level. Hence, $\mathcal{G}_{\mathcal{S}}$ takes the form of a fully-connected graph. Last, $\mathcal{W}_{\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times|\mathcal{S}|}$ defines the weights associated the edges $\mathcal{E}$. In that sense, given the edge $e_{i,j} \in \mathcal{E}_{\mathcal{S}}$ connecting regions r_i and r_j, its associated weight $w_{i,j} \in \mathcal{W}_{\mathcal{S}}$ is defined as the Hervasine distance between both regions.

4.3. *GNN architecture*

Regarding the particular GNN used to develop the predictor mechanism, we opted for the attention temporal graph convolutional network (A3T-GCN) model.[27] This model is able to capture the temporal dependencies among the data by means of Gated Recurrent Units (GRU) cells and the spatial ones by a spectrum-based graph convolutional network[28] along with a soft attention mechanism. In the original A3T-GCN architecture, the attention mechanism was used to learn the importance of traffic information at every moment. Due to the fact that our solution was focused on anticipating tourist flows at spatial scales, this attention mechanism was instrumental in capturing the global variation trends of the tourist flows in the regions at a particular scale.

The pipeline of the GNN used in this study is illustrated in Fig. 7. The model processes a graph, denoted as $\mathcal{G}_{\mathcal{S}}^{w}$, as input to predict future mobility flows. Each graph represents a specific spatial tessellation $\mathcal{S}$ at a given time w. The architecture follows a structured approach:

- Spatial Modeling: The GNN first uses graph-convolutional (GC) layers to capture relationships between regions (nodes) in $\mathcal{S}$. For example, neighboring regions with strong mobility connections will influence each other's predictions.

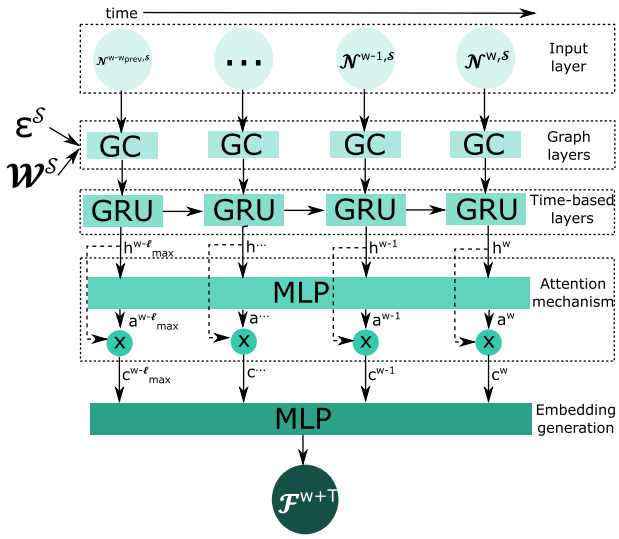


Fig. 7. Architecture of the GNN used for the link prediction task. The daily input graphs are first processed through GC layers to capture spatial dependencies. GRU cells then model temporal changes. The attention mechanism identifies key temporal features, and the final prediction is generated by fully connected layers.

- Temporal Modelling: A stack of GRU, specialized in handling sequential data, models changes in mobility over time.
- Attention Mechanism: A Multilayer Perceptron (MLP) adjusts the importance of temporal features generated by the GRU, focusing on the most relevant information.
- Output Layer: Finally, the adjusted embeddings, x_i, pass through fully connected layers to produce predictions for the desired time horizon, $w+T$, represented as $\mathcal{F}_{\text{TM}}^{w+T,\mathcal{S}}$.

Figure 7 summarizes the data flow in the architecture, showing how information is sequentially processed across the different layers. This architecture involves a total of 10,10178 trainable parameters. We provide more technical details about the GC and GRU operators next.

On the one hand, Graph Convolution (GC) is the cornerstone of spatial modeling. It aggregates information from neighboring nodes in the graph to update each node's features. This process is defined mathematically as

$$H^{(l)} = \text{ReLU}(\tilde{D}^{-\frac{1}{2}}\hat{\mathcal{E}}\tilde{D}^{-\frac{1}{2}}H^{(l-1)}\theta^{(l-1)}),$$

where

- $\hat{\mathcal{E}}$: The adjacency matrix (with added self-loops for stability).
- $\tilde{D}$: Degree matrix of the graph, indicating the number of edges connected to each node.

- $H^{(l-1)}$: Node features from the previous layer.
- $\theta^{(l-1)}$: Trainable weights of the previous layer.
- ReLU: A function ensuring non-negative outputs, enhancing training stability.

The GC operation propagates and aggregates information from each node's immediate neighbors, progressively capturing more contextual information as it passes through multiple layers.

On the other hand, GRUs are used to capture temporal patterns in the data. They utilize three gates to process sequential information effectively:

- Update Gate (z_t): It decides how much information from the past should be carried forward.
- Reset Gate (r_t): It controls how much past information to forget.
- Memory Content (g_t): It combines the current input and past information into a new memory state.

These computations are as follows:

$$
\begin{aligned}
z_t &= \sigma(W_z * [\mathrm{GC}(\mathcal{N}_{\mathcal{S}}^{w}, \mathcal{E}_{\mathcal{S}}, \mathcal{W}_{\mathcal{S}}), h_{t-1}] + b_z), \\
r_t &= \sigma(W_r * [\mathrm{GC}(\mathcal{N}_{\mathcal{S}}^{w}, \mathcal{E}_{\mathcal{S}}, \mathcal{W}_{\mathcal{S}}), h_{t-1}] + b_r), \\
g_t &= \tanh(W_g * [\mathrm{GC}(\mathcal{N}_{\mathcal{S}}^{w}, \mathcal{E}_{\mathcal{S}}, \mathcal{W}_{\mathcal{S}}), r_t * h_{t-1}] + b_g), \\
h_t &= z_t * h_{t-1} + (1 - z_t) * g_t,
\end{aligned}
$$

where σ represents the sigmoid activation function, $W_{\{z,r,g\}}$ are trainable weight matrices, and $b_{\{z,r,g\}}$ are bias terms. This mechanism allows the model to effectively learn from sequences, adjusting the importance of past data as needed.

Finally, the aforementioned GNN architecture returns a matrix $\mathcal{F}_{\mathrm{TM}}^{w+T,\mathcal{S}} \in \mathbb{R}^{|\mathcal{S}|\times 1}$ with the estimated target touristic flow for each region in $\mathcal{S}$ at the $w+T$ week slice. In that sense, we have developed 12 different instances of this GNN, one for each configuration of TMD and GMD flows included in Table 2.

4.4. *Computational complexity analysis*

The proposed approach, based on GNN, presents specific computational demands in terms of time and space.

Regarding the model's time complexity, it is driven by the operations of the GC layers and the temporal GRU layers. For a graph with n nodes, each having a feature vector of dimension d, the computational complexity of a single graph convolution operation is $\mathcal{O}(n^2 d)$. This complexity arises from the pairwise computations between nodes, dictated by the adjacency matrix. Since the model employs stacked graph-convolution layers, followed by temporal operations integrated with GRU cells, the overall complexity grows linearly with the number of layers L and the number of

temporal steps T:

$$\mathcal{O}(L \cdot T \cdot n^2 d).$$

Concerning the space complexity, the memory usage of the model is influenced by the storage of node feature vectors and adjacency matrix information. Specifically:

- Node Features: Storing feature vectors for all nodes requires $\mathcal{O}(n \cdot d)$.
- Adjacency Matrix: For dense graphs, the adjacency matrix storage complexity is $\mathcal{O}(n^2)$.
- Intermediate States: The GRU layers necessitate additional memory for storing intermediate states during temporal processing, further contributing to $\mathcal{O}(T \cdot n \cdot d)$.

Therefore, the model's space complexity can be defined as follows:,

$$\mathcal{O}(n \cdot d + n^2 + T \cdot n \cdot d).$$

5. Evaluation of the Predictors

In this section, we evaluate the accuracy of the GNN predictors described in the previous section.

5.1. *Baselines*

In order to properly evaluate our proposal, we have also used one alternative GNN model, $\mathcal{G}^{\mathcal{S}}_{\mathrm{TM}}$. This graph model only takes as input the incoming TMD flow. Bearing in mind the problem definition in Sec. 4.1, it can be defined as a univariate mapping function $\mathcal{P}(\mathcal{F}^{w,\mathcal{S}}_{\mathrm{TM}}) \rightarrow \mathcal{F}^{w+T,\mathcal{S}}_{\mathrm{TM}}$. By means of this univariate baseline, we can assess the suitability of our multivariate approach fusing the touristic and a general mobility flow to improve the final predictive outcome.

We have also developed a Convolutional Neural Network + Long Short-Term Memory (CNNLSTM) model that integrates the strengths of both CNN and LSTM architectures. This model is capable of extracting the most relevant information through its CNN layer and learning temporal patterns via its LSTM layer. Finally, a fully connected layer is responsible for making the prediction.

Furthermore, the Cox–Ingersoll–Ross (CIR) model was also used as baseline.[29] In brief, the CIR model is a stochastic process originally developed to describe the evolution of interest rates. It captures mean-reverting behavior and enforces positivity of the predicted variable, making it suitable for time series that are strictly non-negative and exhibit dynamic variability. The model is defined by the stochastic differential equation:

$$dX_t = a(b - X_t)dt + \sigma\sqrt{X_t}dW_t, \tag{4}$$

where X_t is the value of the process at time t, a is the speed of mean reversion, b is the long-term mean, σ is the volatility, and W_t is a Wiener process. In our baseline

implementation, we fit the CIR parameters using maximum likelihood estimation and forecast future values via Monte Carlo simulation.

Last, we used the CIR# model as fourth baseline.[10,11,30] In brief, this model extends the classical CIR framework by preserving its analytical tractability while overcoming key limitations such as the inability to capture jumps, clustered volatility, and near-zero or negative values. Unlike the original CIR, CIR# employs ARIMA-based Gaussian residuals instead of Brownian motion and incorporates data partitioning and shifting, which enhance its forecasting accuracy in both erratic and cyclical time series.

It is important to note that the GNN approach allowed us to create a single instance for each spatial scale $\mathcal{S}$. This is because the architecture predicts the incoming trips of all regions within the target scale simultaneously. Regarding the CNNLSTM baseline, we also built a single instance per spatial scale, as it follows a multivariate approach. In contrast, we trained a separate CIR model for each region, since it followed a univariate approach. In this way, each CIR instance focused on forecasting the incoming trips of a specific region within a given scale. As a result, we implemented 39 CIR and CIR# models at the $\mathcal{M}$ scale, 10 at the $\mathcal{D}$ scale, and 3 at the $\mathcal{DT}$ scale according to the number of regions in each level (see Table 1).

5.2. *Metrics*

Regarding the metric to evaluate the accuracy of the models, we have used four commonly used metrics are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Variation of RMSE (CVRMSE). The formulae of each of them are described next where y_i is the real number of tourist arrivals, $\hat{y_i}$ is the predicted number and n the total number of observations.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i,)^2}, \tag{5}$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \tag{6}$$

$$\text{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{7}$$

$$\text{CVRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \times 100. \tag{8}$$

To ensure the statistical significance of the differences observed among the models, we perform a paired t-test analysis.[31] The t-Student test is applied to compare the aforementioned metrics of GNNs enriched with GMD against their closest competitors ($\mathcal{G}_{\text{TM}}$ or CNNLSTM). This statistical test evaluates whether the observed

Table 3. GNN parameters for the experiments.

Type	Parameter	Value
Training	Training rate	0.95
	Loss	Mean Squared Error (MSE)
	Learning factor	0.01
	Weight decay	0.0005
	Optimizer	Adam
	Early Stop	Patience: 5, min_delta: 0.001
	Num. of epochs	1,000
GC	Output size	256
	Num. cells	2
GRU	Input size	256
	Output size	256
Dense - layer 3	Num. of heads	7
	Output size	256
Dense - layer 4	Input size	256
	Output size	1

Table 4. CNNLSTM parameters for the experiments.

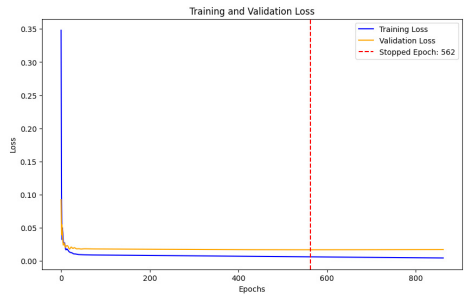
Type	Parameter	Value
Training	Training rate	0.003 (+ *ReduceLROnPlateau*)
	Loss	Mean Squared Error (MSE)
	Learning rate	0.003 (+ *ReduceLROnPlateau*)
	Optimizer	Adam
	Epochs	15000 (+ *EarlyStopping*)
	Batch size	2880
	Activation function	Tanh
CNNLSTM	Units	70
	Filters	64
	Kernel size	2
	Strides	1

differences are statistically significant, providing confidence in the superiority of certain models over others. We report the corresponding t-values and p-values, considering a significance level of $\alpha = 0.05$.

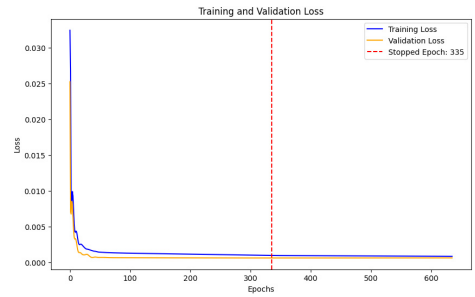
5.3. *Evaluation parameters*

Concerning the hyperparameters used in the evaluation of the proposal, Table 3 comprises the most relevant ones related to the training strategy and each of the 4 layers of the GNN according to sec. 4.3. Furthermore, we tested our models with 6 different time horizons T, namely, 6, 12, 18, 24, 32 and 48 week slices.

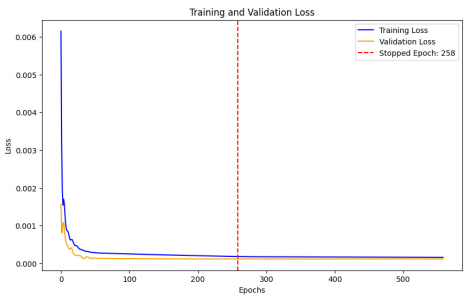
Similarly, Table 4 presents the hyperparameters used for the previously described CNNLSTM model.



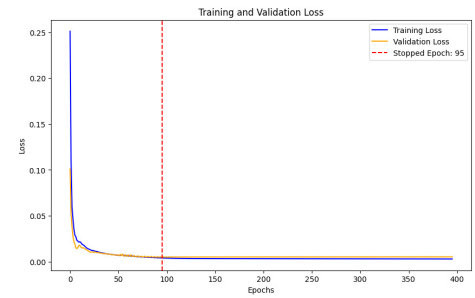
(a) $\mathcal{G}^{\mathcal{M}}_{100}$.



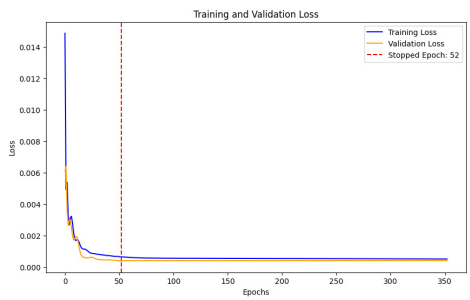
(b) $\mathcal{G}^{\mathcal{D}}_{100}$.



(c) $\mathcal{G}^{\mathcal{D}}_{800}$.



(d) $\mathcal{G}^{\mathcal{DT}}_{100}$.



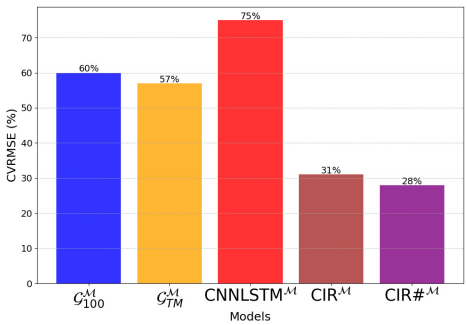
(e) $\mathcal{G}^{\mathcal{DT}}_{800}$.

Fig. 8. Model loss evolution of the GNNs during the training stage.

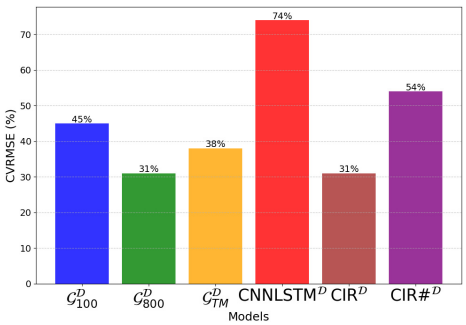
For the sake of completeness, Fig. 8 shows the loss evolution during the training of the GNNs used in the evaluation of the approach. As we can see, the usage of the early stopping policy allowed to avoid the overfitting of the models.

5.4. *Discussion*

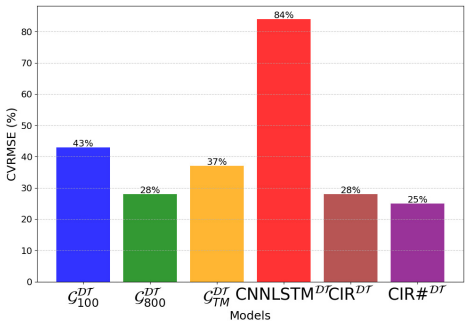
Regarding the results obtained by the predictors, Fig. 9 shows the average CVRMSE of the GNN models in each of the spatial hierarchies under study. As observed, the



(a) Spatial Level $\mathcal{M}$.



(b) Spatial Level $\mathcal{D}$.



(c) Spatial Level $\mathcal{DT}$.

Fig. 9. Average CVRMSE of the different GNNs and baselines per hierarchy level.

multi-variate model with a 800-km threshold, $\mathcal{G}^{\mathcal{DT}}_{800}$, obtained the same CVRMSE than the CIR, which was the lowest one at $\mathcal{D}$ scale as Fig. 9(b) ($t-\text{value} = -5.40, p < 0.005$)($t-\text{value} = -8.70, p < 0.001$) shows. Besides, the differences between $\mathcal{G}_{800}$ and the GNN solely relying on the TDM to perform the prediction, $\mathcal{G}_{\text{TM}}$, increased from $\mathcal{D}$ to $\mathcal{DT}$, 31% versus 38% at $\mathcal{D}$ level and 28% versus 37% at $\mathcal{DT}$ level. This reflects that the multi-variate approach works better when the complexity in terms of number of nodes of the latent graph decreases. Regarding the $\mathcal{M}$ and $\mathcal{DT}$ levels, Figs. 9(a) and 9(c) show that the ensemble of univariate CIR# models clearly obtained the lowest CVRMSE.

Concerning the MAPE metric, Table 5 reports the average values obtained by the different models across the various spatial levels. As observed, the proposed GNN obtained a promising accuracy in order to predict the incoming flows of excursionists. More in detail, it consistently achieved the lowest MAPE values for most incoming flows related to international excursionists ($\mathcal{IE}$) across the three spatial hierarchies. For instance, the $\mathcal{G}^{\mathcal{D}}_{800}$ model achieved an average MAPE of 8.71 for predicting the incoming flow of international excursionists ($\mathcal{IE}$) at the $\mathcal{D}$ level, outperforming the best baseline for that flow, CIR#, which obtained a MAPE of

Table 5. Average MAPE of the different GNNs and baselines per hierarchy level and type of flow. The best values for each (Flow, T) pair are shown in bold.

Flow		Trips Range	Model	T						Avg.
				12	18	24	32	48	60	
$\mathcal{M}$	$\mathcal{A}$	1266.92 ± 446.44	$\mathcal{G}^{\mathcal{M}}_{100}$	182.68	180.15	178.41	183.79	177.71	182.52	180.88
			$\mathcal{G}^{\mathcal{M}}_{\text{TM}}$	180.07	182.19	181.37	181.43	180.60	185.94	181.93
			CNNLSTM$^{\mathcal{M}}$	234.38	223.10	218.06	212.92	213.05	253.57	225.85
			CIR$^{\mathcal{M}}$	**36.16**	12.97	17.58	26.37	38.61	36.55	28.04
			CIR#$^{\mathcal{M}}$	**23.83**	**11.28**	**9.6**	**11.38**	**33.89**	**18.05**	**18.01**
	$\mathcal{E}$	614.16 ± 278.14	$\mathcal{G}^{\mathcal{M}}_{\text{TM}}$	57.85	57.24	91.14	59.62	65.50	75.85	67.87
			$\mathcal{G}^{\mathcal{M}}_{100}$	94.47	83.86	69.40	69.86	78.38	78.69	79.11
			CNNLSTM$^{\mathcal{M}}$	86.75	86.67	89.71	90.22	94.59	87.74	89.28
			CIR$^{\mathcal{M}}$	43.93	47.75	26.41	**32.01**	39.94	27.58	36.27
			CIR#$^{\mathcal{M}}$	**43.83**	**25.76**	**24.90**	**47.83**	**32.68**	**24.44**	**33.24**
	$\mathcal{IE}$	426.12 ± 197.76	$\mathcal{G}^{\mathcal{M}}_{100}$	**20.04**	**19.95**	**19.88**	21.61	18.38	**19.52**	19.90
			$\mathcal{G}^{\mathcal{M}}_{\text{TM}}$	21.71	22.00	21.36	21.53	17.58	21.63	20.97
			CNNLSTM$^{\mathcal{M}}$	40.32	35.26	33.68	35.22	38.14	47.41	38.34
			CIR$^{\mathcal{M}}$	25.98	25.31	21.88	17.40	32.32	27.44	25.06
			CIR#$^{\mathcal{M}}$	**25.62**	**26.89**	**13.30**	**14.17**	**18.24**	**20.84**	**19.18**
$\mathcal{D}$	$\mathcal{E}$	4939.30 ± 2223.36	$\mathcal{G}^{\mathcal{D}}_{800}$	**12.72**	**15.81**	**15.96**	15.72	**20.58**	**16.82**	**16.27**
			$\mathcal{G}^{\mathcal{D}}_{\text{TM}}$	25.46	18.51	24.07	23.86	23.92	24.05	23.31
			CNNLSTM$^{\mathcal{D}}$	67.73	70.28	73.93	86.52	74.36	73.95	74.46
			CIR$^{\mathcal{D}}$	24.11	23.15	18.49	**11.97**	30.61	31.68	23.34
			CIR#$^{\mathcal{D}}$	**17.55**	**17.38**	**17.62**	**17.49**	**18.35**	**17.75**	**17.69**
	$\mathcal{T}$	15738.00 ± 5294.57	$\mathcal{G}^{\mathcal{D}}_{800}$	103.88	102.61	101.62	102.16	100.02	102.67	102.16
			$\mathcal{G}^{\mathcal{D}}_{\text{TM}}$	117.44	116.80	118.17	116.48	120.03	117.33	117.71
			CNNLSTM$^{\mathcal{D}}$	155.14	154.42	160.83	175.94	164.53	158.35	161.53
			CIR$^{\mathcal{D}}$	28.89	35.44	20.1	**17.40**	50.07	38.06	31.66
			CIR#$^{\mathcal{D}}$	**10.00**	**34.85**	**18.53**	**17.75**	**20.01**	**20.74**	**20.65**
	$\mathcal{NE}$	3210.56 ± 1543.21	$\mathcal{G}^{\mathcal{D}}_{800}$	**18.62**	19.45	19.38	28.69	**33.35**	**27.62**	25.37
			$\mathcal{G}^{\mathcal{D}}_{\text{TM}}$	33.10	32.85	32.70	34.63	38.45	33.48	33.35
			CNNLSTM$^{\mathcal{D}}$	78.27	75.85	77.51	79.57	85.07	71.28	77.93
			CIR$^{\mathcal{D}}$	21.77	17.07	19.49	23.04	36.65	33.61	25.27
			CIR#$^{\mathcal{D}}$	**19.54**	**14.52**	**17.70**	**10.54**	**34.68**	**31.54**	**21.09**
	$\mathcal{IE}$	1726.75±811.22	$\mathcal{G}^{\mathcal{D}}_{800}$	**9.45**	6.93	8.90	**9.54**	**8.15**	**9.27**	**8.71**
			$\mathcal{G}^{\mathcal{D}}_{\text{TM}}$	10.97	10.91	10.85	10.69	8.44	10.76	10.44
			CNNLSTM$^{\mathcal{D}}$	43.57	37.37	35.29	35.33	39.23	55.70	41.08
			CIR$^{\mathcal{D}}$	16.27	15.9	22.57	30.23	26.03	26.49	22.92
			CIR#$^{\mathcal{D}}$	**10.96**	**6.58**	**7.71**	**9.27**	**9.43**	**9.80**	**9.29**
$\mathcal{DT}$	$\mathcal{A}$	56862.60±19402.59	$\mathcal{G}^{\mathcal{DT}}_{800}$	115.58	116.84	114.46	115.08	111.72	107.38	113.51
			$\mathcal{G}^{\mathcal{DT}}_{\text{TM}}$	127.02	129.04	127.63	128.6	128.88	126.18	127.89
			CNNLSTM$^{\mathcal{DT}}$	229.36	233.27	229.44	243.03	270.52	223.77	238.23
			CIR$^{\mathcal{DT}}$	17.14	**12.65**	**10.49**	9.99	24.64	24.21	16.52
			CIR#$^{\mathcal{DT}}$	**12.79**	**16.72**	**15.05**	**5.03**	**16.28**	**16.05**	**13.32**
	$\mathcal{E}$	13583.09±6114.17	$\mathcal{G}^{\mathcal{DT}}_{800}$	24.19	42.4	41.77	42.52	42.18	42.52	39.26
			$\mathcal{G}^{\mathcal{DT}}_{\text{TM}}$	51.61	51.5	51.32	51.64	51.2	51.69	51.49
			CNNLSTM$^{\mathcal{DT}}$	97.08	95.05	92.76	93.15	93.82	93.56	94.24
			CIR$^{\mathcal{DT}}$	20.03	16.39	15.32	16.43	32.25	30.52	21.82
			CIR#$^{\mathcal{DT}}$	**20.21**	**14.36**	**10.93**	**12.87**	**14.57**	**23.04**	**15.99**

Table 5. Average MAPE of the different GNNs and baselines per hierarchy level and type of flow. The best values for each (Flow, T) pair are shown in bold.

Flow	Trips Range	Model	T						Avg.
			12	18	24	32	48	60	
$\mathcal{T}$	43279.51±14562	$\mathcal{G}_{800}^{\mathcal{DT}}$	88.85	89.04	88.81	88.47	89.04	86.27	88.41
		$\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$	105.69	105.14	103.19	99.17	89.93	87.1	98.37
		$\mathrm{CNNLSTM}^{\mathcal{DT}}$	214.13	237.8	256.21	286.18	325.15	203.17	253.77
		$\mathrm{CIR}^{\mathcal{DT}}$	39.39	24.18	30.78	17.99	39.05	**18.83**	28.37
		$\mathrm{CIR}\#^{\mathcal{DT}}$	**17.53**	**16.52**	**19.33**	**17.64**	**14.90**	**14.23**	**16.69**
$\mathcal{NE}$	8829.03±4243.85	$\mathcal{G}_{800}^{\mathcal{DT}}$	35.93	35.92	36.01	36.23	36.22	36.17	36.08
		$\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$	44.10	44.04	43.82	44.21	44.10	44.01	44.05
		$\mathrm{CNNLSTM}^{\mathcal{DT}}$	48.63	53.03	57.45	64.22	67.87	43.81	55.84
		$\mathrm{CIR}^{\mathcal{DT}}$	22.19	22.03	21.04	26.12	36.22	33.44	26.84
		$\mathrm{CIR}\#^{\mathcal{DT}}$	**19.73**	**23.4**	**15.90**	**18.96**	**24.15**	**19.22**	**20.56**
$\mathcal{IE}$	4754.05±2230.86	$\mathcal{G}_{800}^{\mathcal{DT}}$	**12.29**	13.05	**12.41**	**13.07**	**12.97**	**13.02**	**12.80**
		$\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$	15.85	15.96	16.31	15.18	15.69	15.76	15.79
		$\mathrm{CNNLSTM}^{\mathcal{DT}}$	19.1	17.76	17.58	17.93	19.22	20.06	18.61
		$\mathrm{CIR}^{\mathcal{DT}}$	19.56	**10.73**	18.88	37.33	27.04	23.87	22.9
		$\mathrm{CIR}\#^{\mathcal{DT}}$	**19.50**	**13.87**	**18.37**	**19.63**	**15.09**	**17.60**	**17.51**

9.29. Furthermore, our approach also obtained the lowest MAPE in order to forecast the $\mathcal{E}$ flow at $\mathcal{D}$ scale.

Moreover, Table 6 shows the MAE range of each model for across the target forecasting horizons T and flows. Overall, the $\mathcal{G}_{100}^{\mathcal{M}}$ model achieved the best performance in several cases. Specifically, for the $\mathcal{A}$ flow at granularity $\mathcal{M}$, $\mathcal{G}_{100}^{\mathcal{M}}$ outperformed other models for the largest time horizons with the lowest MAE scores, 358.87 at T = 48 and 369.47 at T=60 ($t = 4.77, p < 0.001$). Similarly, in the $\mathcal{IE}$ flow at $\mathcal{M}$ level, $\mathcal{G}_{100}^{\mathcal{M}}$ provided competitive results at longer time horizons, such as T= 32(42.79) and T= 48(34.93) ($t = 2.26, p < 0.05$), outperforming the other models.

At the $\mathcal{D}$ level, the $\mathcal{G}_{800}^{\mathcal{D}}$ model demonstrated superior performance to anticipate the $\mathcal{T}$ flow. More in detail, it achieved the lowest MAE scores across all time horizons ($t = 2.43, p < 0.05$) except T= 60, where $\mathcal{G}_{\mathrm{TM}}^{\mathcal{D}}$ performed slightly better.

At the $\mathcal{DT}$ level, $\mathcal{G}_{800}^{\mathcal{DT}}$ emerged as the best performing model for the flows $\mathcal{A}$ ($t = -9.43, p < 0.001$) and $\mathcal{T}$ ($t = -6.90, p < 0.001$), particularly in long-term forecasts, where it maintains the lowest MAE values. Similarly, $\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$ achieved the best accuracy in $\mathcal{E}$ and $\mathcal{NE}$ flows.

It is also worth noting that the GNN models enriched with general mobility data generally achieved the lowest MAE when predicting the $\mathcal{IE}$ flow for time horizons longer than 32 weeks. For example, the $\mathcal{G}_{100}$ and $\mathcal{G}_{800}$ models yielded the best MAE values for this flow at $T = 48$ across all three spatial levels. These findings are consistent with the previously discussed MAPE evaluation, which demonstrated the superior performance of our approach in forecasting incoming trips of excursionists.

Another notable trend in Table 6 is the consistent underperformance of the CNNLSTM model across all flows and categories. Its MAE scores are significantly

Table 6. Average MAE of the different GNNs and baselines per hierarchy level and type of flow. The best values for each (Flow, T) pair are shown in bold.

Flow		Trips Range	Model	T						Avg.
				12	18	24	32	48	60	
$\mathcal{M}$	$\mathcal{A}$	1266.92 ± 446.44	$\mathcal{G}^{\mathcal{M}}_{100}$	357.82	362.02	360.4	360.51	**358.87**	**369.47**	361.35
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	378.15	372.91	369.3	380.43	367.86	377.81	374.41
			$\mathrm{CNNLSTM}^{\mathcal{M}}$	513.23	488.53	477.5	466.23	466.52	555.25	494.54
			$\mathrm{CIR}^{\mathcal{M}}$	187.93	288.17	304.36	394.13	685.08	522.43	397.02
			$\mathrm{CIR}\#^{\mathcal{M}}$	**136.71**	**198.31**	**199.97**	**339.89**	**392.87**	**464.91**	**288.78**
	$\mathcal{E}$	614.16±278.14	$\mathcal{G}^{\mathcal{M}}_{100}$	187.73	166.64	**137.91**	138.81	155.74	**156.36**	157.20
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	**119.74**	118.48	188.66	**123.42**	**135.59**	157.01	**140.15**
			$\mathrm{CNNLSTM}^{\mathcal{M}}$	189.96	189.79	196.44	197.55	207.12	192.12	194.99
			$\mathrm{CIR}^{\mathcal{M}}$	213.34	146.88	298.02	467.05	207.96	263.05	265.05
			$\mathrm{CIR}\#^{\mathcal{M}}$	**127.39**	**115.44**	**196.20**	**189.20**	**152.15**	**216.26**	**166.44**
	$\mathcal{IE}$	426.12 ± 197.76	$\mathcal{G}^{\mathcal{M}}_{100}$	43.14	43.72	42.45	**42.79**	**34.93**	42.98	41.67
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	41.49	41.3	**41.16**	44.74	38.04	**40.41**	**41.19**
			$\mathrm{CNNLSTM}^{\mathcal{M}}$	88.3	77.2	73.75	77.13	83.51	103.82	83.95
			$\mathrm{CIR}^{\mathcal{M}}$	175.67	85.74	76.18	87.57	225.91	159.59	135.11
			$\mathrm{CIR}\#^{\mathcal{M}}$	**21.76**	**30.02**	**50.52**	**55.79**	**35.14**	**99.37**	**48.10**
$\mathcal{D}$	$\mathcal{E}$	4939.30±2223.36	$\mathcal{G}^{\mathcal{D}}_{800}$	223.83	201.92	146.82	190.91	189.23	**189.72**	190.74
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	**186.39**	**117.88**	**146.48**	**147.89**	**145.61**	190.68	**155.82**
			$\mathrm{CNNLSTM}^{\mathcal{D}}$	582.11	550.65	571.36	600.99	703.37	604.53	601.17
			$\mathrm{CIR}^{\mathcal{D}}$	1248.56	1068.14	1020.22	1283.41	2197.64	2066.34	1477.72
			$\mathrm{CIR}\#^{\mathcal{D}}$	**586.23**	**654.19**	**370.78**	**963.14**	**598.64**	**1103.84**	**712.80**
	$\mathcal{T}$	15738.00±5294.57	$\mathcal{G}^{\mathcal{D}}_{800}$	**914.55**	**931.43**	**926.36**	**937.24**	**923.83**	951.94	**930.56**
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	980.47	962.52	950.79	941.57	946.57	**926.77**	951.28
			$\mathrm{CNNLSTM}^{\mathcal{D}}$	1134.08	1261.2	1255.4	1307.47	1430.31	1337.53	1287.33
			$\mathrm{CIR}^{\mathcal{D}}$	7730.61	4252.88	2626.98	5580.41	8251.49	5952.56	5732.99
			$\mathrm{CIR}\#^{\mathcal{D}}$	**1319.94**	**2817.85**	**1348.36**	**1506.11**	**3590.59**	**3698.24**	**2346.52**
	$\mathcal{NE}$	3210.56±1543.21	$\mathcal{G}^{\mathcal{D}}_{800}$	155.29	154.09	153.42	162.46	156.45	157.07	156.13
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	**99.88**	104.34	**103.97**	153.91	**153.93**	153.91	128.66
			$\mathrm{CNNLSTM}^{\mathcal{D}}$	427.61	414.37	423.46	434.72	464.75	389.41	425.72
			$\mathrm{CIR}^{\mathcal{D}}$	108.73	102.07	107.69	128.15	172.66	142.35	127.11
			$\mathrm{CIR}\#^{\mathcal{D}}$	**65.41**	**20.39**	**231.64**	**142.23**	**168.21**	**115.57**	**123.91**
	$\mathcal{IE}$	1726.75±811.22	$\mathcal{G}^{\mathcal{D}}_{800}$	51.46	51.18	50.92	**50.14**	**39.6**	50.46	47.62
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	50.71	**37.19**	**47.74**	51.16	43.7	**49.76**	**46.41**
			$\mathrm{CNNLSTM}^{\mathcal{D}}$	238.03	204.16	192.77	193.01	214.33	304.27	223.76
			$\mathrm{CIR}^{\mathcal{D}}$	40.14	76.76	64.83	54.76	84.75	53.89	62.86
			$\mathrm{CIR}\#^{\mathcal{D}}$	**36.46**	**38.91**	**53.94**	**53.61**	**53.09**	**58.96**	**49.66**
$\mathcal{DT}$	$\mathcal{A}$	56862.60±19402.59	$\mathcal{G}^{\mathcal{DT}}_{800}$	**4351.26**	**4420.38**	**4372.27**	**4405.46**	**4415.13**	**4322.65**	**4364.19**
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	4794.53	4846.61	4747.9	4773.73	4634.45	4454.22	4741.41
			$\mathrm{CNNLSTM}^{\mathcal{DT}}$	6366.98	6475.71	6369.27	6746.39	7509.69	6211.79	6446.80
			$\mathrm{CIR}^{\mathcal{DT}}$	12370.08	9510.54	7243.54	7071.18	19575.57	16752.79	12020.28
			$\mathrm{CIR}\#^{\mathcal{DT}}$	**7506.97**	**7818.81**	**5928.86**	**6697.28**	**5341.45**	**10317.57**	**7248.99**
	$\mathcal{E}$	13583.09±6114.17	$\mathcal{G}^{\mathcal{DT}}_{800}$	1767.86	1764.15	1758.02	1768.97	1753.86	1770.69	1760.59
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	**1003.31**	**1758.79**	**1732.63**	**1763.69**	**1749.78**	**1763.85**	**1628.75**
			$\mathrm{CNNLSTM}^{\mathcal{DT}}$	2694.88	2638.54	2575.03	2585.77	2604.56	2597.29	2599.68
			$\mathrm{CIR}^{\mathcal{DT}}$	4144.57	3157.91	3405.37	3597.98	6459.24	5397.15	4193.37
			$\mathrm{CIR}\#^{\mathcal{DT}}$	**2282.59**	**2093.87**	**1866.74**	**2272.33**	**2635.77**	**2074.41**	**2204.62**
	$\mathcal{T}$	43279.51 ± 14562	$\mathcal{G}^{\mathcal{DT}}_{800}$	**3620.58**	**3601.96**	**3534.95**	**3397.2**	**3080.75**	**2983.95**	**3369.40**
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	3685.53	3693.48	3684.07	3669.74	3693.48	3578.68	3667.99
			$\mathrm{CNNLSTM}^{\mathcal{DT}}$	5944.17	6601.4	7112.41	7944.48	9026.3	5640.08	7044.81
			$\mathrm{CIR}^{\mathcal{DT}}$	13384.67	20056.88	17294.73	13884.78	21569.74	16635.87	17137.44
			$\mathrm{CIR}\#^{\mathcal{DT}}$	**10331.37**	**7967.83**	**6152.96**	**4857.38**	**5921.27**	**6588.17**	**6893.33**

Table 6. (*Continued*)

Flow	Trips Range	Model	T						Avg.
			12	18	24	32	48	60	
$\mathcal{NE}$	8829.03±4243.85	$\mathcal{G}_{800}^{\mathcal{DT}}$	1510.6	1508.71	1501.17	1514.37	1510.6	1507.77	1508.54
		$\mathcal{G}_{\text{TM}}^{\mathcal{DT}}$	1490.33	1489.86	**1493.63**	**1503.06**	**1502.58**	**1500.23**	**1496.61**
		CNNLSTM$^{\mathcal{DT}}$	**1349.85**	**1472.21**	1594.76	1782.72	1884.2	1216.08	1550.47
		CIR$^{\mathcal{DT}}$	2734.13	2500.21	3161.77	3153.00	4755.37	3750.72	3342.70
		CIR#$^{\mathcal{DT}}$	**1699.41**	**1591.95**	**1464.83**	**1747.66**	**1545.82**	**1559.06**	**1601.62**
$\mathcal{IE}$	4754.05±2230.86	$\mathcal{G}_{800}^{\mathcal{DT}}$	542.9	546.83	558.63	520.17	**537.65**	**539.84**	540.37
		$\mathcal{G}_{\text{TM}}^{\mathcal{DT}}$	**509.68**	541.15	514.93	542.03	538.09	540.06	537.99
		CNNLSTM$^{\mathcal{DT}}$	530.29	**493.1**	**487.9**	**497.68**	533.56	556.92	**516.92**
		CIR$^{\mathcal{DT}}$	1098.33	941.04	968.05	1642.16	2069.22	1719.19	1399.67
		CIR#$^{\mathcal{DT}}$	**570.69**	**808.76**	**929.41**	**593.77**	**1327.06**	**695.93**	**820.27**

higher compared to both GNN-based models, confirming the superiority of graph-based architectures in this predictive task. Besides, the CIR baseline also achieved quite high MAE when it came to predict the $\mathcal{E}$ and $\mathcal{T}$ flows at $\mathcal{D}$ and $\mathcal{DT}$ scales. It is also important to remark that the CIR# model achieved the lowest MAE to predict the $\mathcal{A}$ flow at $\mathcal{M}$ scale.

Table 7 presents the RMSE values for each time horizon T, target flow, model, and spatial scale. In line with the MAPE and MAE analyses discussed previously, it is noteworthy that the GNN models trained with both touristic and general mobility data also achieved the lowest RMSE when forecasting the general flow of incoming tourists ($\mathcal{T}$) and international excursionists ($\mathcal{IE}$) for $T \geq 48$ at the $\mathcal{D}$ and $\mathcal{DT}$ levels.

These results are consistent with the MAE values reported in Table 6.

5.5. *Lessons learned*

After analyzing the evaluation results stated in the previous section, we can answer the two RQs formulated at the beginning of the paper.

RQ1: Can general human mobility feeds, representing daily and regular trips, improve the prediction of certain type of touristic flows towards a region?

The results have shown that the enrichment of the GNN with TMD flows has consistently improved the accuracy of the predictions for the number of incoming international excursionists $\mathcal{IE}$ across the three spatial hierarchies, based on the considered baselines. This type of flow is particularly interesting to predict due to the continuous interest of the Spanish tourism industry in attracting foreign visitors, given their typically high income level.

Furthermore, the analysis has also revealed that our approach allowed us to improve the MAE score for $\mathcal{A}$ and $\mathcal{T}$ at the higher spatial levels. Again, these are the two types of touristic mobility that are more profitable for the industry, as $\mathcal{A}$ represents the total number of visitors to a region, and $\mathcal{T}$ indicates the total number of tourists who are going to spend at least one night in the target region. Concerning the other excursionist flows, $\mathcal{E}$ and $\mathcal{NE}$, the evaluation has shown that

Table 7. Average RMSE of the different GNNs and baselines per hierarchy level and type of flow. The best values for each (Flow, T) pair are shown in bold.

Flow		Trips Range	Model	T						Avg.
				12	18	24	32	48	60	
$\mathcal{M}$	$\mathcal{A}$	1266.92 ± 446.44	$\mathcal{G}^{\mathcal{M}}_{100}$	1955.16	1982.99	1974.11	1975.71	1961.16	2020.07	1978.20
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	2069.93	2037.68	2019.82	2079.05	2008.54	2065.65	2046.78
			CNNLSTM$^{\mathcal{M}}$	2884.36	2687.59	2625.98	2602.87	2547.66	2542.68	2648.52
			CIR$^{\mathcal{M}}$	221.32	444.77	**386.58**	427.02	547.29	452.08	413.18
			CIR#$^{\mathcal{M}}$	**112.99**	**271.99**	**255.5**	**354.10**	**469.84**	**298.97**	**277.73**
	$\mathcal{E}$	614.16 ± 278.14	$\mathcal{G}^{\mathcal{M}}_{100}$	655.39	648.16	1030.97	675.25	740.02	858.98	768.13
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	1025.34	910.68	751.91	755.96	851.06	855.61	858.43
			CNNLSTM$^{\mathcal{M}}$	11036.21	1084.31	1095.29	1023.62	1026.60	1082.09	1058.02
			CIR$^{\mathcal{M}}$	303.92	406.97	268.76	202.24	287.01	222.66	281.93
			CIR#$^{\mathcal{M}}$	**244.19**	**277.39**	**228.63**	**175.87**	**275.19**	**123.85**	**220.19**
	$\mathcal{IE}$	426.12 ± 197.76	$\mathcal{G}^{\mathcal{M}}_{100}$	231.72	230.97	229.26	249.08	**191.72**	225.53	226.38
			$\mathcal{G}^{\mathcal{M}}_{\mathrm{TM}}$	236.06	239.80	232.27	234.29	221.12	235.22	233.13
			CNNLSTM$^{\mathcal{M}}$	554.20	498.23	541.35	412.40	438.05	485.78	488.34
			CIR$^{\mathcal{M}}$	144.47	156.54	107.05	110.09	253.39	**178.88**	158.40
			CIR#$^{\mathcal{M}}$	**123.34**	**146.24**	**105.49**	**107.93**	**198.85**	**264.92**	**157.63**
$\mathcal{D}$	$\mathcal{E}$	4939.30±2223.36	$\mathcal{G}^{\mathcal{D}}_{800}$	2550.91	1610.66	2011.37	2020.89	**1990.42**	**2061.85**	2041.02
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	3050.46	2750.85	2010.22	2610.25	2590.47	2595.08	2601.22
			CNNLSTM$^{\mathcal{D}}$	8042.08	8026.82	7648.79	8013.40	8398.87	9633.38	8293.89
			CIR$^{\mathcal{D}}$	1500.85	1448.64	1277.38	1336.84	2529.74	2187.7	1713.52
			CIR#$^{\mathcal{D}}$	**619.46**	**724.63**	**421.33**	**1179.4**	**2009.71**	**2063.87**	**1096.73**
	$\mathcal{T}$	15738.00 ± 5294.57	$\mathcal{G}^{\mathcal{D}}_{800}$	5650.18	5750.31	5730.34	5790.97	**5710.38**	**5880.18**	**5752.06**
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	6229.51	6120.93	6040.27	5990.27	6010.93	5890.21	6047.02
			CNNLSTM$^{\mathcal{D}}$	8560.72	6255.18	7996.39	7857.80	7958.60	9948.61	8096.22
			CIR$^{\mathcal{D}}$	4430.36	4979.12	5969.02	4250.02	7870.43	9092.02	6098.50
			CIR#$^{\mathcal{D}}$	**1817.98**	**4425.44**	**3166.66**	**42937.41**	**6192.97**	**5608.64**	**11008.85**
	$\mathcal{NE}$	3210.56 ± 1543.21	$\mathcal{G}^{\mathcal{D}}_{800}$	1350.78	1410.35	1405.34	2070.46	2075.50	2078.15	1731.76
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	2100.43	2095.05	2080.08	2200.51	2110.95	2130.26	2119.55
			CNNLSTM$^{\mathcal{D}}$	5084.12	5941.61	5642.76	5983.96	6004.18	6236.25	5815.48
			CIR$^{\mathcal{D}}$	1264.17	1383.10	1159.98	1488.28	1751.00	1333.40	1396.66
			CIR#$^{\mathcal{D}}$	**1183.48**	**1376.58**	**1125.90**	**1316.22**	**1699.37**	**1128.97**	**1305.09**
	$\mathcal{IE}$	1726.75 ± 811.22	$\mathcal{G}^{\mathcal{D}}_{800}$	749.07	549.14	705.89	**740.39**	**545.86**	**735.29**	670.94
			$\mathcal{G}^{\mathcal{D}}_{\mathrm{TM}}$	760.27	756.07	752.59	755.64	685.02	745.42	742.50
			CNNLSTM$^{\mathcal{D}}$	4068.92	3401.42	3021.37	2815.13	2658.27	3260.34	3204.24
			CIR$^{\mathcal{D}}$	438.34	405.14	311.90	759.81	981.75	829.16	621.02
			CIR#$^{\mathcal{D}}$	**425.25**	**385.55**	**125.84**	**846.40**	**618.43**	**742.34**	**514.97**
$\mathcal{DT}$	$\mathcal{A}$	56862.60±19402.59	$\mathcal{G}^{\mathcal{DT}}_{800}$	22800.57	23020.39	22600.66	22700.15	**22010.49**	21020.22	22358.75
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	20750.19	21050.72	20800.79	20990.66	21000.76	20550.43	20857.26
			CNNLSTM$^{\mathcal{DT}}$	30140.55	31856.91	3256.64	32149.71	32597.69	37414.63	27919.02
			CIR$^{\mathcal{DT}}$	14977.74	13275.75	11046.33	9982.00	25060.94	21046.87	15981.94
			CIR#$^{\mathcal{DT}}$	**13452.26**	**12090.13**	**10545.27**	**8554.50**	**24955.38**	**19981.11**	**14979.11**
	$\mathcal{E}$	13583.09±6114.17	$\mathcal{G}^{\mathcal{DT}}_{800}$	4550.19	7980.27	7900.36	8015.33	7940.64	8025.98	7402.13
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	7995.24	8000.93	7985.67	8030.55	7950.62	8035.39	7999.73
			CNNLSTM$^{\mathcal{DT}}$	11563.78	11873.94	11720.49	11698.26	11990.84	12336.51	11897.64
			CIR$^{\mathcal{DT}}$	4667.21	4061.21	3470.04	3633.50	6852.32	5464.7	4691.83
			CIR#$^{\mathcal{DT}}$	**2549.92**	**3093.62**	**2266.19**	**2201.69**	**3490.77**	**5990.98**	**3265.36**
	$\mathcal{T}$	43279.51 ± 14562	$\mathcal{G}^{\mathcal{DT}}_{800}$	19450.76	19350.55	18990.78	18250.95	**16550.29**	**16030.76**	18103.85
			$\mathcal{G}^{\mathcal{DT}}_{\mathrm{TM}}$	16450.89	16500.17	16400.57	16350.22	16682.66	16250.84	16439.56
			CNNLSTM$^{\mathcal{DT}}$	25060.17	25489.55	30115.86	32875.95	35278.03	41346.47	31694.34
			CIR$^{\mathcal{DT}}$	7931.14	13550.01	17056.62	10088.74	21602.49	25597.15	15904.36
			CIR#$^{\mathcal{DT}}$	**3310.35**	**11721.58**	**16698.30**	**9293.6**	**1712.87**	**19821.68**	**9060.73**

Table 7. (*Continued*)

Flow	Trips Range	Model	T						Avg.
			12	18	24	32	48	60	
$\mathcal{NE}$	8829.03 ± 4243.85	$\mathcal{G}_{800}^{\mathcal{DT}}$	6280.29	6270.89	6350.10	6550.78	6540.85	6490.92	6413.30
		$\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$	6710.86	6670.07	6510.91	6790.47	6710.31	6650.62	6673.87
		$\mathrm{CNNLSTM}^{\mathcal{DT}}$	6021.08	6604.92	7337.41	7815.52	8621.93	8816.10	7536.50
		$\mathrm{CIR}^{\mathcal{DT}}$	3550.54	3408.21	3549.07	3559.64	5055.86	3872.79	3832.69
		$\mathrm{CIR}\#^{\mathcal{DT}}$	**2634.05**	**2854.19**	**3326.44**	**3273.21**	**4964.23**	**2931.51**	**3323.27**
$\mathcal{IE}$	4754.05 ± 2230.86	$\mathcal{G}_{800}^{\mathcal{DT}}$	1560.84	2280.13	1680.77	2300.56	**2200.25**	**2250.38**	2045.82
		$\mathcal{G}_{\mathrm{TM}}^{\mathcal{DT}}$	2320.41	2410.62	2680.17	1800.92	2210.48	2255.74	2286.72
		$\mathrm{CNNLSTM}^{\mathcal{DT}}$	3487.12	2825.64	2720.51	2754.66	2786.93	3255.18	2971.67
		$\mathrm{CIR}^{\mathcal{DT}}$	900.19	940.65	917.81	1704.09	2639.50	2308.61	**1401.81**
		$\mathrm{CIR}\#^{\mathcal{DT}}$	**504.62**	**240.58**	**519.75**	**1560.88**	**2491.56**	**3900.62**	**1703.17**

the impact of introducing the TMD flow as an exogenous variable is much more limited than in other flows.

RQ2: If such an improvement exists, does it depend on the spatial scale of the target region?

The evaluation of the proposal has shown that the improvement provided by the TMD flows are much more remarkable in the upper spatial scales of the hierarchy, $\mathcal{D}$ and $\mathcal{DT}$, than at $\mathcal{M}$ level. This indicates that the benefit of enriching the predictor with TMD flows is affected by the granularity of the spatial tessellation when a GNN model is used. One possible reason of this dependency is that tourist movement prediction at fine-grained spatial scales require of very specific flow information that is not captured by GMD flows. In that sense, region-based general mobility feeds are only able to capture the latent touristic flow across regions only up to a certain spatial scale. Given the nature of the human mobility datasets used in this evaluation, this might be due to multiple factors such as the data collection or curation processes made to the raw location data before being released as open data. Therefore, the spatial *threshold* to successfully integrate general human-mobility data for touristic forecasting will likely vary depending on the particular spatial context and nature of the data.

5.6. *Generalization of findings*

The findings of this study highlight key considerations for the generalization of our approach to other geographic and operational settings.

With respect to scalability to different regions, the proposed methodology is adaptable to other regions where general human mobility datasets are available. However, the spatial granularity at which the GMD is effective may vary depending on the urban structure and transportation networks of the target region.

We should mention the potential for additional contextual features. While GMD provides useful exogenous information for tourist flow prediction, it should be considered incorporating additional features such as weather conditions, public events, and economic indicators to further enhance accuracy.

6. Conclusions, Policy Implications and Future Work

The prediction of touristic flows has become an important tool for developing proactive actions in the tourist marketing and service field. However, most of these approaches only rely on tourist-specific mobility data to generate their outcomes given a fixed geographical distribution. In this context, the major contribution of the present work is to offer an evaluation of the actual impact of enriching touristic predictors with general-purpose data considering different aggregations of spatial regions.

The study has been carried out considering human mobility in the RM (Spain). Results revealed that the human flows extracted from a general-purpose mobility feed have a positive effect on the prediction accuracy of specific touristic flows, such as the total number of visitors or the tourists having at least one overnight in a region. Besides, the granularity of the spatial tessellation was also revealed as an important factor for enhancing the accuracy of the predictor. In that sense, the improvement provided by the general-flow enrichment was higher when the spatial scale of the target geographical regions increased. On the basis of these results, the key benefit of this work is that it opens a new course of action within the tourist-mobility discipline to enrich predictors with general-purpose mobility data when certain conditions in the geographical setting and the target touristic flow are met.

Beyond methodological innovation, this study also contributes to evidence-based tourism governance. Accurate and scalable predictions of visitor flows allow destination management organizations and regional authorities to anticipate demand surges, detect potential congestion hotspots, and plan infrastructure development more effectively. In addition, the model can support strategic decisions such as activating or moderating the promotion of specific attractions based on their expected impact on flow distribution. These forecasting capabilities are particularly valuable in regions with seasonal peaks or fragmented transport networks, helping transition from reactive to proactive policy-making. Moreover, the integration of human mobility intelligence into tourism analytics provides a strategic opportunity to advance smarter and more resilient destination management. The proposed model enables public administrations to embed mobility forecasting within broader digital governance platforms, enhancing coordination between tourism, transport, and spatial planning. Thanks to its adaptability and scalability, the approach can be replicated in other regions with comparable data availability, supporting the alignment of planning practices across different territories. In this way, the framework contributes to ongoing efforts aimed at achieving sustainable tourism development, fostering regional cohesion, and strengthening data-driven innovation in public policy.

Some limitations are identified in this work. First, our GNN-based approach entails significant computational overhead due to the complexity of graph convolution operations, which may hinder scalability for large datasets. Our approach also depends on the quality and preprocessing of the exogenous mobility data, which can affect the accuracy of tourist flow prediction. Besides, the

adaptability of our model to other regions may require recalibration of spatial thresholds due to differences in their urban shapes and transport features. Finally, since no open-source implementation of the CIR# model is available, a specific version was developed for this study in accordance with the guidelines provided by its authors. Consequently, minor differences may exist between the original model proposed by the authors and the implementation used in this work.

Two lines of research can be explored in the next stages of this work. First, incorporating additional contextual data, such as weather conditions, events, and holidays, could enhance the predictive accuracy of the tourist flow model. Second, the extension of this study to multiple regions would allow to evaluate whether the impact of general-mobility flows based on the target spatial tessellation also occurs in other geographical settings.

Acknowledgments

Financial support for this research has been provided by the Instituto de Turismo de la Región de Murcia through project CAT/TU/47-22 "Cátedra Internacional de Inteligencia Turística Región de Murcia UCAM-ITREM".

Appendix A. Tables with the Values of the Correlation Metrics

Table A.1. Correlation metrics for the target flows at $\mathcal{M}$ level.

	GMD flow ($\mathcal{F}_{\mathrm{GM}}^{\delta,\mathcal{M}}$)	PC	MI	GCT lags
$\mathcal{A}$	100	0.083	0.067	15.842
	400	0.000	0.035	11.079
	800	−0.008	0.018	4.947
$\mathcal{IE}$	100	0.117	0.052	12.351
	400	−0.001	0.028	8.162
	800	−0.024	0.010	5.730
$\mathcal{NE}$	100	0.120	0.062	14.865
	400	−0.020	0.021	8.865
	800	−0.034	0.018	2.865
$\mathcal{E}$	100	0.136	0.067	14.189
	400	−0.012	0.019	9.568
	800	−0.032	0.011	3.486
$\mathcal{IT}$	100	0.027	0.035	10.027
	400	0.014	0.038	6.514
	800	−0.002	0.021	5.297
$\mathcal{NT}$	100	0.047	0.057	15.421
	400	−0.007	0.051	9.658
	800	0.001	0.018	4.342
$\mathcal{T}$	100	0.051	0.058	14.395
	400	0.006	0.037	10.237
	800	0.002	0.013	5.316

Table A.2. Correlation metrics for the target flows at $\mathcal{D}$ level.

	GMD flow ($\mathcal{F}_{\mathrm{GM}}^{\delta,\mathcal{D}}$)	PC	MI	GCT lags
$\mathcal{A}$	100	0.154	0.150	15.9
	400	0.151	0.109	16.3
	800	0.200	0.024	5.8
$\mathcal{IE}$	100	0.293	0.148	13.3
	400	0.231	0.082	7.3
	800	0.169	0.030	4.5
$\mathcal{NE}$	100	0.297	0.222	12.9
	400	0.215	0.099	13.1
	800	0.184	0.033	4.8
$\mathcal{E}$	100	0.320	0.194	11.6
	400	0.243	0.106	11.2
	800	0.197	0.039	5.1
$\mathcal{IT}$	100	0.100	0.130	9.3
	400	0.126	0.073	8.1
	800	0.155	0.042	5.4
$\mathcal{NT}$	100	0.030	0.079	15.1
	400	0.048	0.085	15.3
	800	0.157	0.029	6.1
$\mathcal{T}$	100	0.065	0.121	13.5
	400	0.096	0.076	15.5
	800	0.185	0.030	5.0

Table A.3. Correlation metrics for the target flows at $\mathcal{DT}$ level.

	GMD flow ($\mathcal{F}_{\mathrm{GM}}^{\delta,,\mathcal{DT}}$)	PC	MI	GCT lags
$\mathcal{A}$	100	0.176	0.191	16.667
	400	0.189	0.157	16.000
	800	0.276	0.073	4.667
$\mathcal{IE}$	100	0.320	0.131	15.000
	400	0.287	0.132	7.333
	800	0.265	0.022	4.000
$\mathcal{NE}$	100	0.338	0.255	15.667
	400	0.271	0.177	11.000
	800	0.263	0.077	3.000
$\mathcal{E}$	100	0.359	0.218	13.667
	400	0.300	0.132	11.333
	800	0.284	0.075	3.333
$\mathcal{IT}$	100	0.130	0.168	12.333
	400	0.164	0.118	12.000
	800	0.231	0.075	4.667
$\mathcal{NT}$	100	0.043	0.117	15.000
	400	0.066	0.135	19.000
	800	0.211	0.093	13.667
$\mathcal{T}$	100	0.086	0.181	14.667
	400	0.123	0.142	17.667
	800	0.253	0.076	5.000

ORCID

Fernando Terroso-Saenz https://orcid.org/0000-0002-1921-1137
Juan Morales-García https://orcid.org/0000-0003-0008-4825
Miguel Puig-Cabrera https://orcid.org/0000-0003-4524-9830
Ginesa Martinez-del Vas https://orcid.org/0000-0002-2713-0636
Andres Muñoz https://orcid.org/0000-0002-8491-4592

References

1. A. Chantre-Astaiza, L. Fuentes-Moraleda, A. Muñoz-Mazón and G. Ramirez-Gonzalez, Science mapping of tourist mobility 1980–2019. Technological advancements in the collection of the data for tourist traceability, *Sustainability* **11**(17) (2019) 4738.
2. C. Li, P. Ge, Z. Liu, and W. Zheng, Forecasting tourist arrivals using denoising and potential factors, *Annals of Tourism Research* **83** (2020) 102943.
3. W. Wang, J. Chen, Y. Zhang, Z. Gong, N. Kumar and W. Wei, A multi-graph convolutional network framework for tourist flow prediction, *ACM Transactions on Internet Technology* **21**(4) (2021) 1–13.
4. H. F. Chan *et al.*, A global dataset of human mobility, Technical Report, Center for Research in Economics, Management and the Arts (CREMA) (2020).
5. G. Barlacchi, M. De Nadai, R. Larcher, A. Casella, C. Chitic, G. Torrisi, F. Antonelli, A. Vespignani, A. Pentland and B. Lepri, A multi-source dataset of urban life in the city of Milan and the province of Trentino, *Scientific Data* **2**(1) (2015) 1–15.
6. S. Chang, E. Pierson, P. W. Koh, J. Gerardin, B. Redbird, D. Grusky and J. Leskovec, Mobility network models of COVID-19 explain inequities and inform reopening, *Nature* **589** (2006) 1–6.
7. F.-L. Chu, Forecasting tourism demand with ARMA-based methods, *Tourism Management* **30**(5) (2009) 740–751.
8. L. Yin, Prediction and analysis of China's tourism labor force based on univariate model based on ARMA, in *2nd Artificial Intelligence and Complex Systems Conference* (Bangkok, Thailand, 2021) pp. 202–206.
9. Y. Li and H. Cao, Prediction for tourism flow based on LSTM neural network, *Procedia Computer Science* **129** (2018) 277–283.
10. M. Bufalo and G. Orlando, Improved tourism demand forecasting with cir# model: a case study of disrupted data patterns in Italy, *Tourism Review* **79**(2) (2024) 445–464.
11. G. Orlando and M. Bufalo, Time series forecasting with the cir# model: from hectic markets sentiments to regular seasonal tourism, *Technological and Economic Development of Economy* **29**(4) (2023) 1216–1238.
12. O. Claveria and S. Torra, Forecasting tourism demand to catalonia: Neural networks vs. time series models, *Economic Modelling* **36** (2014) 220–228.
13. J.-W. Bi, H. Li and Z.-P. Fan, Tourism demand forecasting with time series imaging: A deep learning model, *Annals of Tourism Research* **90** (2021) 103255.
14. F. Petropoulos and E. Spiliotis, The wisdom of the data: Getting the most out of univariate time series forecasting, *Forecasting* **3**(3) (2021) 478–497.
15. L. Zhu, C. Lim, W. Xie and Y. Wu, Modelling tourist flow association for tourism demand forecasting, *Current Issues in Tourism* **21**(8) (2018) 902–916.
16. Y. Yang and H. Zhang, Spatial-temporal forecasting of tourism demand, *Annals of Tourism Research* **75** (2019) 106–119.

17. Y. Zhang, G. Li, B. Muskat and R. Law, Tourism demand forecasting: A decomposed deep learning approach, *Journal of Travel Research* **60**(5) (2021) 981–997.
18. R. Law, G. Li, D. K. C. Fong and X. Han, Tourism demand forecasting: A deep learning approach, *Annals of Tourism Research* **75** (2019) 410–423.
19. A. Belussi, A. Cinelli, A. D. Vecchia, S. Migliorini, M. Quaresmini and E. Quintarelli, Forecasting POI occupation with contextual machine learning, in *26th European Conference Advances in Databases and Information Systems*, *Turin*, *Italy* (Springer International Publishing, 2022), pp. 361–376.
20. S. Migliorini, A. D. Vecchia, A. Belussi and E. Quintarelli, ARTEMIS: a context-aware recommendation system with crowding forecaster for the touristic domain, *Information Systems Frontiers* (2024) DOI:10.1007/s10796-024-10512-y.
21. G. de Murcia, Tourism strategic framework of Murcia 2022-2032, Technical Report, Comunidad Autónoma de la Región de Murcia (2022).
22. A. C. Abellán and R. G. Marín, Del turismo de salud a la diversificación turística cultura en el sureste de España, *Estudios Turísticos* **201** (2014) 47–72.
23. Secretaría de Estado de Transportes, Movilidad y Agenda Urbana. (2020). Analysis of mobility in Spain using Big Data technology during the state of alarm for the management of the COVID-19 crisis: Methodological report (Version 0.12). Ministry of Transport, Mobility and Urban Agenda.
24. G. Pasaoglu, D. Fiorello, A. Martino, L. Zani, A. Zubaryeva and C. Thiel, Travel patterns and the potential use of electric cars–results from a direct survey in six European countries, *Technological Forecasting and Social Change* **87** (2014) 51–59.
25. B. C. Ross, Mutual information between discrete and continuous data sets, *PloS one* **9**(2) (2014) e87357.
26. P. Spirtes, C. N. Glymour, R. Scheines and D. Heckerman, *Causation*, *Prediction*, *and Search* (MIT Press, 2000).
27. J. Bai, J. Zhu, Y. Song, L. Zhao, Z. Hou, R. Du and H. Li, A3T-GCN: Attention temporal graph convolutional network for traffic forecasting, *ISPRS International Journal of Geo-Information* **10**(7) (2021) 485.
28. J. Bruna, W. Zaremba, A. Szlam and Y. LeCun, Spectral networks and deep locally connected networks on graphs, in *2nd Int. Conf. Learning Representations* (Banff, AB, Canada, 2014).
29. J. C. Cox *et al.*, A theory of the term structure of interest rates, *Econometrica* **53**(2) (1985) 385–407.
30. G. Orlando and M. Bufalo, Interest rates forecasting: Between hull and white and the cir#—how to make a single-factor model work, *Journal of Forecasting* **40**(8) (2021) 1566–1580.
31. Student, The probable error of a mean, *Biometrika* **6** (1908) 1–25.