AStA Advances in Statistical Analysis (2024) 108:477–499
https://doi.org/10.1007/s10182-023-00484-w

ORIGINAL PAPER

GPS data on tourists: a spatial analysis on road networks

Nicoletta D'Angelo[1] · Antonino Abbruzzo[1] · Mauro Ferrante[2] · Giada Adelfio[1] · Marcello Chiodi[1]

Received: 1 February 2023 / Accepted: 12 September 2023 / Published online: 3 November 2023

Abstract

This paper proposes a spatial point process model on a linear network to analyse cruise passengers' stop activities. It identifies and models tourists' stop intensity at the destination as a function of their main determinants. For this purpose, we consider data collected on cruise passengers through the integration of traditional questionnaire-based survey methods and GPS tracking data in two cities, namely Palermo (Italy) and Dubrovnik (Croatia). Firstly, the density-based spatial clustering of applications with noise algorithm is applied to identify stop locations from GPS tracking data. The influence of individual-related variables and itinerary-related characteristics is considered within a framework of a Gibbs point process model. The proposed model describes spatial stop intensity at the destination, accounting for the geometry of the underlying road network, individual-related variables, contextual-level information, and the spatial interaction amongst stop points. The analysis succeeds in quantifying the influence of both individual-related variables and trip-related characteristics on stop intensity. An interaction parameter allows for measuring the degree of dependence amongst cruise passengers in stop location decisions.

Keywords Cruise passengers · Global Positioning System · Gibbs model · Linear network · Stop density

✉ Nicoletta D'Angelo
nicoletta.dangelo@unipa.it

Antonino Abbruzzo
antonino.abbruzzo@unipa.it

Mauro Ferrante
mauro.ferrante@unipa.it

Giada Adelfio
giada.adelfio@unipa.it

Marcello Chiodi
marcello.chiodi@unipa.it

[1] Department of Economics, Business and Statistics, University of Palermo, Palermo, Italy

[2] Department of Culture and Society, University of Palermo, Palermo, Italy

1 Introduction

Several papers discussed the importance of managing key locations and understanding tourist spatial behaviour and its main determinants (Cooper 1981; Russo 2002; Liu et al. 2017). At the micro-level, a better knowledge of tourist intra-destination behaviour is relevant for destination marketing and management (Zoltan and McKercher 2015). Increased understanding of most visited places and how tourists experience the destination may orient policy actions aimed at mitigating overcrowding in certain places, promoting less visited places, and enhancing tourist safety and satisfaction (Li et al. 2019). Despite the importance of understanding tourist movements within a destination, collecting data on visitors' mobility are a challenging task (Stopher 2012). Amongst such complexity sources, we could identify the vast diversity of routes and attractions available for each visitor (Lew and McKercher 2002), which are also influenced by the visitor and visit features and the spatial locations of infrastructures, restaurants, accommodations (Smallwood et al. 2012). Traditional methods are generally based on post-visit questionnaires or trip diaries (East et al. 2017). However, collecting information on mobility through traditional travel diaries can be a complex and labour-intensive process that demands considerable time and attention from the respondent (Stopher 2012). Moreover, relying on diaries or self-reported routes for the analysis of tourist mobility may result in several pitfalls (McKercher and Zoltan 2014). These may include instances in which tourists find the effort required to complete the diary to be arduous and therefore leave it incomplete or omit certain parts. Furthermore, tourists may encounter difficulties in locating the site to be marked on the diary or map (Puczkó et al. 2010). To overcome these problems, in the last decades, there has been an increase in the number of studies that used tracking technologies for collating data on intra-destination mobility behaviour, especially in tourism research (Shoval et al. 2015). Nowadays, Global Positioning System (GPS) technology allows for collecting information on human mobility at great temporal and spatial details, without any effort required to recall the visited places. Given this vast data, many challenges arise in modelling highly detailed spatio-temporal data.

Although various techniques may be implemented to analyse tracking data (Atluri et al. 2018), more research is needed to study stop location points concerning destination-related characteristics and individual-related variables (Grinberger and Shoval 2019). Approaches considering the surrounding elements, like road networks or locations of attractions, are even more scarce. For instance, Smallwood et al. (2012) analysed the movement patterns of visitors travelling on the road network in northwestern Australia.

Due to geometrical complexities, analysing individual data observed along a network of lines are highly challenging. Moreover, the intrinsic lack of homogeneity in a network militates against the traditional methods of spatial statistics (Baddeley et al. 2021).

Point processes theory on linear networks has been recently proposed for analysing events occurring on a network of lines. They were firstly introduced in the

spatial context and then extended to the spatio-temporal case, focusing on the analysis of first- and second-order summary statistics (Ang et al. 2012; McSwiggan et al. 2017; Rakshit et al. 2017, 2019; D'Angelo et al. 2021b, 2022b). Whilst most of the literature about point processes on spatio-temporal first-order intensity on networks is concerned with nonparametric estimation (Moradi et al. 2019; Moradi and Mateu 2020; Mateu et al. 2020), only a recent paper by D'Angelo et al. (2022) has dealt with parametric intensity specification of inhomogeneous first-order intensities on networks. The authors fitted a Gibbs point process model with mixed effects for the purely spatial component and a spatio-temporal log-Gaussian Cox process, adapting both models to the underlying road network. Furthermore, D'Angelo et al. (2022c) have considered the self-exciting behaviour of crime point process data, proposing a spatio-temporal Hawkes model adapted to linear networks, including a parametric estimation of the background based on covariates.

From a methodological perspective, various approaches have been used to model intra-destination tourist movements. Identifying stop locations are essential for summarising the information in tracking data since they may indicate touristic areas of interest for the individuals. Gong et al. (2015) review the main research results on stop location identification. In tourism, stop identification may reveal popular places, such as tourist attractions, restaurants, or shopping centres. After identifying attractions, several studies used graph-based methods to analyse tourist movement patterns (Kurashima et al. 2010; Yang et al. 2017; Hu et al. 2019), intending to reconstruct the network of the relationships amongst the various attractions. Other authors instead focused their attention on trajectory similarity. Shoval et al. (2015), for example, used the sequence alignment method to segment tourists according to their trajectory similarity. Zheng et al. (2019) propose a combination of the dynamic time warping with earth mover distance to compare the trajectories of 56 tourists in the Xiamen campus and Petry et al. (2019) provide a comparative approach amongst various similarity measures for comparing trajectories. Due to the complexity of modelling stop location intensity, most studies undertake descriptive approaches based on visual examination of stop location maps. At the same time, a modelling-based approach may help to reveal essential tourist behaviours.

From an empirical perspective, similar to recent studies (De Cantis et al. 2016; Ferrante et al. 2018; Domènech et al. 2020a, 2020b; Navarro-Ruiz et al. 2020; Shoval et al. 2020; Casado-Díaz et al. 2021), we focus on cruise tourism segments for several reasons. First, the single entry-exit point that characterises cruise visits at the destination makes implementing a GPS-based survey feasible. Second, the limited duration of the visit allows for a complete picture of cruise passengers' experience at the destination. Third, the cruise tourism segment has mainly been debated due to their impact on destinations, determining to overcrowding of places in the face of a lower expenditure than traditional tourists (Brida and Zapata 2010; Larsen et al. 2013). However, the previous studies have not explicitly focused on cruise passengers' stop activities to the best of our knowledge. Moreover, this paper provides a comprehensive picture of individuals' characteristics, integrating a questionnaire-based survey and GPS tracking data in Palermo (Italy) and Dubrovnik (Croatia), based on the assumption that stop locations reveal critical points of interest of

the destination. A Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996) is used to identify the stop locations from the GPS tracking data. The data are analysed using a spatial point process applied to a linear network. Specifically, in this paper, we extend the work presented in D'Angelo et al. (2022) in two directions: (i) by using the DSCAN algorithm, to identify the stops, and (ii) by including external spatial covariates in an application of the Gibbs point process model. Note that the Gibbs point process model is fitted on a road network represented by the configuration of the streets.

All the analyses are carried out through the statistical software R Core Team (2023), and the source codes to reproduce the analysis in the city of Palermo are contained in the GitHub repository https://github.com/nicolettadangelo/gps_data.

The paper is organised as follows. The data are described in Sect. 2. Section 3 presents the DBSCAN method employed to identify stop activities at the destination. In Sects. 4 and 5, we introduce both the theory of spatial point processes on linear networks and the Gibbs point process model for the intensity specification. Sections 6 and 7 report and discuss the results obtained for the cruise tourism segment in Palermo and Dubrovnik, respectively. Section 8 concludes this paper and indicates future research directions.

2 Data

Data were collected from two surveys: the first was conducted in the Spring of 2014 in Palermo (Italy), and the second in the Summer of 2015 in Dubrovnik (Croatia). Both surveys were implemented by integrating a questionnaire-based survey and GPS tracking devices on cruise passengers disembarking at their destination. Specifically, 303 cruisers were interviewed in Palermo and 51 in Dubrovnik; in both cases, 12 variables were selected for each interviewed cruise passenger, according to the literature on visitors' intra-destination mobility. Data collected in the port of Palermo are initially presented in De Cantis et al. (2016). A description of the Dubrovnik survey can be found in Ferrante et al. (2018), together with the complete description of the collected variables. An opening and a closing questionnaire were administered to collect socio-demographic characteristics and pre-, and post-visit information for each cruise passenger interviewed.

In addition, a GPS tracking device was released to each cruise passenger sampled, recording their position during the visit to the destination. Therefore, the resulting GPS tracking data consist of coordinate points recorded every ten seconds. As an example, Fig. 1 shows the GPS track of a sampled cruise passenger (black points) and the streets (red lines), with the detail of the stops preference and of the time (in minutes) in each stop (identified by the red stars), using Palermo data. Moreover, the top numbers in the circles indicate which stop has been first visited, whilst the bottom numbers indicate the time spent at each stop. The three stops represent the disembarking area, Piazza Pretoria and the Cathedral, the last two being in the list of the main attraction of the municipality of Palermo.

Even though the data quality is very high, we can identify some noise points. Indeed, when analysing the complete trajectory, a set of pre-processing operations

Fig. 1 Example of a GPS track (black points) and stops derived from the DBSCAN algorithm (red stars) for a cruiser in the municipality of Palermo. The top numbers in the circular indicate which stop has been first visited. The circular's bottom numbers indicate the times spent in minutes at each stop (colour figure online)

on the raw GPS data are necessary to remove outliers and regularise the trajectories (see Abbruzzo et al. (2021) for further details). However, in this paper, our main aim is to model the stops according to some covariates and the spatial characteristics, then consider the outliers as noise points.

The analysis of cruise passengers' stop activities, based on the integration of the questionnaire-based survey and GPS tracking data, prevents some typical data collection problems on tourist mobility. Indeed, traditional methods are generally based on post-visit questionnaires or trip diaries, which rely on the accurate recall of the places visited and activities made. However, nowadays, GPS technology allows for collecting information on human mobility at a very great temporal and spatial detail, without any effort required to recall the visited places. In this way, we can evaluate the relationships amongst the number of stops obtained from the DBSCAN algorithm (in Sect. 3), and both individual-related variables, collected through the questionnaire-based survey, and some itinerary-related characteristics, identified from the GPS tracking data on the whole itinerary.

2.1 Covariates

According to the literature on visitors' intra-destination mobility, a set of covariates related to the individual characteristics, information on the itinerary undertaken, and the selected tourist attractions are included in the analysis. First, a set of variables is available from a questionnaire issued to the cruise passengers: *Day of the interview*, *Age group*, *Country of origin*, *Education level* (dichotomized in `low` (High

school diploma or Bachelor degree) and `high` (Master or PhD)), *First visit* (indicating whether the cruise passenger was visiting the city for the first time or not), *Yearly income* (dichotomized in $< 40{,}000$ and $\geq 40{,}000$ euro). The questionnaire is issued to only one person per group if the person travels with the family or a group of friends.

Regarding the information on the undertaken itinerary, a set of synthetic information for the two contexts under analysis was derived from the GPS tracking data and considered as potential determinants of stop activity, namely: *Total length of tour* (dichotomized in < 11 and ≥ 11 km), *Total time of tour* (dichotomized in < 3.5 and ≥ 3.5 hours), *Maximum distance from the port* (dichotomized in < 3.5 and ≥ 3.5 km), *Average duration of the stops* (in minutes). Specific threshold values for some covariates were considered to distinguish between different types of tourists. For instance, the maximum distance from the port discriminates between those who stayed within the historical centre of the city of Palermo and those who ventured outside. As for income, low sample size for some income categories suggested that reducing the number of categories to two would be more appropriate. Additionally, dichotomising certain covariates can facilitate easier interpretation of the groups considered (for example between those undertaking a short tour and those making a longer one) and can mitigate some potential problems stemming from measurement errors.

Considering the restricted area of the historical centre of Dubrovnik, information on the maximum distance from the port was not included in the model selection procedure, being relatively constant for all the units under analysis.

Finally, a set of dummy variables are also derived by the questionnaire issued to cruise passengers, indicating whether the cruise passenger has visited a specific touristic attraction. The tourist attractions of the two destinations have been selected after comparing the top attractions on popular websites, such as Tripadvisor, Google rank, Lonely planet and Skyscanner. These tourist attractions include: the `Norman Palace`, the `Cathedral`, `Massimo Theatre`, `La Martorana`, the `Chiaramonte Palace`, `Quattro Canti`, and `Ballarò Market` in the case of Palermo city, and `Minceta Tower`, `Bokar Tower`, `Stradun of Dubrovnik`, `the Cathedral`, `Loggia Square`, `Fountain of Onofrio`, and `The Rector's Palace and Cultural Historical Museum`, in the case of Dubrovnik.

3 The DBSCAN algorithm

The DBSCAN is a density-based algorithm (Ester et al. 1996) designed to identify arbitrary-shaped clusters, where the clusters are sets of spatial points that fall within a certain distance. Concurrently, the algorithm can identify the *noise* points, which are spatial points not belonging to any cluster. Given $D^{(i)}$ a generic trajectory for a unit i, a *cluster* is defined as a minimum set of spatial points which are sufficiently close to each other. The following definitions are given to clarify the concept of points sufficiently close to each other.

Let $p = (x, y)$ be a point in a trajectory D for a generic unit i. The *ϵ-neighbourhood* of a point p is defined by $ne_\epsilon(p) = \{q \in D : d(p, q) \leq \epsilon \in \mathbb{R}^+\}$, where $d(\cdot, \cdot)$ is a distance function. If the cardinality of an ϵ-neighbourhood of a point p, i.e. card $(ne_\epsilon(p))$, is at least greater than a minimum number ($minpts \in \mathbb{N}$) then p is a *core point*. Moreover, a point p is *directly density-reachable* from q, with respect to ϵ and *minpts*, if $p \in ne_\epsilon(q)$ and card $(ne_\epsilon(q)) \geq minpts$. Finally, let's define density-reachable and density-connected points. A point p is *density-reachable* from the point q with respect to ϵ and *minpts* if there is a chain $p_1, \ldots, p_l$, $p_1 = q$, $p_l = p$ such that p_{i+1} is directly density-reachable from p_i. A point p is *density-connected* to point q, with respect to ϵ and *minpts*, if there is a point o such that both p and q are density-reachable from o with respect to ϵ and *minpts*.

A *cluster* C is a non-empty subset of D satisfying the following requirements:

- $\forall\ p, q \in D$: if $p \in C$ and q is density-reachable from p with respect to ϵ and *minpts*, then, $q \in C$;
- $\forall\ p, q \in C$: p is density-connected to q with respect to ϵ and *minpts*.

Let $C_1, \ldots, C_k$ be the clusters of D with respect to ϵ and *minpts*, then, $p \in D$ is a *noise* point if it does not belong to any cluster C_i. The algorithm starts with the first point p in the database D, retrieving all the neighbours of a point p concerning ϵ and *minpts*. This procedure will yield a cluster if p is a core point. Otherwise, no points will be density-reachable from p, and the DBSCAN algorithm will proceed to consider the next point of the database. Centroids summarise the spatial coordinates of each point belonging to a cluster, and the rest are considered as *noise*. The DBSCAN requires choosing two parameters, namely ϵ and *minpts*, which indicate the search radius and minimum number of points in the search radius for identifying the location as a cluster. Finally, this algorithm does not consider temporal information. For an algorithm extension, which would also incorporate the temporal information, see Birant and Kut (2007).

This study uses a value of 20 ($\sim$ 3 min) for the *minpts* tuning parameter and 40 (meters) for the distance value ϵ. The distance between consecutive points is calculated as Euclidean since it is approximately the same as the one on the linear network when one considers spatial points measured every few seconds. Note that we collect data every ten seconds in the case of Palermo and Dubrovnik.

Figure 1 shows an example of an application for the Palermo cruise passengers' GPS tracks. The DBSCAN reveals three stops in which the cruise passenger has spent 8.38 (Stop 1), 18.23 (Stop 2) and 16.23 (Stop 3) minutes. The shortest-path distances, which consider the road network, between these stops are around 2300 m, $(1 \rightarrow 2)$ and $(2 \rightarrow 3)$, and 764 m. On the contrary, the euclidean distances between the stops are around 1500 $(1 \rightarrow 2)$ and 600 $(2 \rightarrow 3)$ meters, respectively. This simple example motivates spatial point processes on a linear network.

4 Spatial point processes on a linear network

A point process model assumes that a point pattern $\mathbf{x}$ is a realisation of a finite point process X in the bounded region $D \subset \mathbb{R}^2$ without multiple points. A spatial point pattern is an unordered set $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of points $\mathbf{x}_i$ where $n(\mathbf{x}) = n$ denotes the number of points, not fixed in advance (Cressie 2015). In this context, $N(A)$ denotes the number of points of a set $A \cap X$, where $A \subseteq D$.

We denote a point location in the plane as $\mathbf{u}$, specified by Cartesian coordinates $\mathbf{u} = (u_1, u_2)$ without explicitly mentioning the coordinates. The intensity function describes the first-order property of the point process and is defined as follows:

$$\lambda(\mathbf{u}) = \lim_{|d\mathbf{u}| \to 0} \frac{\mathbb{E}[N(d\mathbf{u})]}{|d\mathbf{u}|},$$

where $d\mathbf{u}$ represents an infinitesimal region containing the point $\mathbf{u} \in D$, $|d\mathbf{u}|$ is the area of $d\mathbf{u}$, and $\mathbb{E}[N(d\mathbf{u})]$ is the expected number of events in $d\mathbf{u}$. If the intensity is constant, the process is called homogeneous. In the case of inhomogeneity, the intensity varies within the study area and may depend on the coordinates of points. A point process model, assuming independence, is fully described by its conditional intensity function (Daley and Vere-Jones 2007). For a comprehensive treatment of spatial point processes in Euclidean space, please refer to Baddeley et al. (2015).

In a formal sense, a linear network $L = \cup_{i=1}^{n} \mathbf{l}_i \subset \mathbb{R}^2$ is typically represented by a finite union of line segments in two-dimensional space (Ang et al. 2012). These line segments, denoted as $\mathbf{l}_i$, have positive lengths and together form the linear network L ($L = \cup_{i=1}^{n} \mathbf{l}_i \subset \mathbb{R}^2$). The endpoints of these line segments are referred to as *nodes*, and the *degree* of a node corresponds to the number of line segments that share that particular node (Okabe and Sugihara 2012). A line segment, $\mathbf{l}_i$, is defined as $[\mathbf{u}_i, \mathbf{v}_i] = \{k\mathbf{u}_i + (1-k)\mathbf{v}_i : 0 \le k \le 1\}$, where $\mathbf{u}_i$ and $\mathbf{v}_i$ represent the endpoints of $\mathbf{l}_i$ in two-dimensional space. It is worth noting that the intersection of any two line segments, $\mathbf{l}_i$ and $\mathbf{l}_j$ (where $i \neq j$), can either be empty or consist of an endpoint shared by both segments. The total length of all line segments in the linear network L is denoted by $|L|$. When determining the distance between two locations, $\mathbf{u}$ and $\mathbf{v}$, within the network L, the most commonly used method is the shortest-path distance, denoted as $d_L(\mathbf{u}, \mathbf{v})$. This distance corresponds to the minimum length amongst all possible paths connecting $\mathbf{u}$ and $\mathbf{v}$ in the network. However, other alternative distance measures have also been discussed in Rakshit et al. (2017).

Similar to planar point processes, a point process on a linear network L, denoted as X, involves a random countable subset of $\mathbb{R}^2$ without overlapping points. In practical scenarios, we observe $n > 0$ events, $\mathbf{x} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, of the point process X occurring within a linear network $L \subset \mathbb{R}^2$. We use $N(L)$ to represent the number of events that take place on L. It is important to note that these events are not limited to occurring solely at the nodes of the network. Just like in the planar case, we define the intensity function $\lambda(\cdot)$ as the point process equivalent of the mean function for a real-valued stochastic process. For further information about the statistical properties of the intensity function, refer to Moradi et al. (2018).

A point process N on a linear network L is considered homogeneous if it exhibits a constant intensity, that is, $\lambda(\mathbf{u}) = \lambda$ for all $\mathbf{u} \in L$. In general, $\lambda(\mathbf{u})$ is interpreted as the expected number of events per unit length of the linear network L in the vicinity of $\mathbf{u}$. However, a more realistic and general scenario are the inhomogeneous case, where the constant intensity λ of the Poisson process is replaced by a spatially varying intensity function over the linear network L. Recent research on statistical methods for analysing spatial patterns of points on a network of lines, such as road accident locations along a road network, has been reviewed by Baddeley et al. (2021).

5 Modelling the spatial stop intensity

In this section, we introduce a modelling approach useful for describing the spatial behaviour of the visitors. The presented methodology is developed through a stochastic point process modelling approach on a linear network originally proposed in D'Angelo et al. (2022). Specifically, we incorporate a random subject-specific effect and employ a parametric model to capture the visitors' stops, taking into account both the underlying network structure and the individual tourists' choices. To accomplish this, we adopt the Gibbs point process models with mixed effects, as outlined by (Illian and Hendrichsen 2010), adapting the procedure to the specific context of linear networks. This flexible procedure enables the incorporation of individual information through appropriate random and fixed factors, as well as external covariates.

Consider a scenario where there are M visitors on a linear network L, and each visitor generates a point pattern of stops represented by $\mathbf{x}_1, \dots, \mathbf{x}_M$. In this context, we make the assumption that each $\mathbf{x}_m$ (where $m = 1, \dots, M$) follows a pairwise interaction process (Van Lieshout 2000), characterised by a conditional intensity (Kallenberg 1984), which is given by:

$$\lambda_{\theta,\phi_m}(\mathbf{u};\mathbf{x}_m) = b_{\theta,\phi_m}(\mathbf{u}) \prod_{i=1,\mathbf{x}_{im}\neq\mathbf{u}}^{n(\mathbf{x}_m)} h_{\theta,\phi_m}(\mathbf{u}, \mathbf{x}_{im}),$$

where the parameter θ is a vector of fixed effects and the random effects ϕ_m are assumed to come from $\Phi \sim N(0, \sigma_\phi^2 I)$. Then $n(\mathbf{x}_m)$ is the number of points in $\mathbf{x}_m$, that is, the number of stops per visitor, $b_{\theta,\phi_m}(\mathbf{u})$ and $h_{\theta,\phi_m}(\mathbf{u}, \mathbf{v})$ are two functions that model the intensity and the interaction, respectively.

5.1 Model estimation through pseudolikelihood

Inference is carried out through the maximisation of the pseudolikelihood over the subset $A \subseteq L$

$$\mathrm{PL}_A\left(\theta,\phi;\mathbf{x}_1,\ldots,\mathbf{x}_M\right)=\prod_{m=1}^{M}\left(\prod_{\mathbf{x}_{im}\in A}\lambda_{\theta,\phi_m}\left(\mathbf{x}_{im};\mathbf{x}_m\right)\exp\left(-\int_A\lambda_{\theta,\phi_m}(\mathbf{u};\mathbf{x}_m)d\mathbf{u}\right)\right)$$

that is,

$$\begin{aligned}\mathrm{PL}_A(\theta,\phi;\mathbf{x}_1,\ldots,\mathbf{x}_M)=&\prod_{m=1}^{M}\left(\prod_{i=1}^{n(\mathbf{x}_m)}b_{\theta,\phi_m}(\mathbf{x}_{im})\prod_{i\neq j}^{n(\mathbf{x}_m)}h_{\theta,\phi_m}\left(\mathbf{x}_{im},\mathbf{x}_{mj}\right)\right)\\&\times\exp\left\{-\int_A b_{\theta,\phi_m}(\mathbf{u})\prod_{i=1}^{n(\mathbf{x}_m)}h_{\theta,\phi_m}\left(\mathbf{u},\mathbf{x}_{im}\right)d\mathbf{u}\right\},\end{aligned}\tag{1}$$

where $\phi=(\phi_1,\ldots,\phi_M)$.

In order to avoid the computation of integrals in the pseudolikelihood, we employ the Berman–Turner device (Baddeley and Turner 2000), which utilises a quadrature rule. This approach involves approximating the integral in (1) for each m by a finite sum over a set of points $\mathbf{u}_{mj}$, where $j=1,\ldots,l_m$. These points include all the data points and allow for a more manageable computation. This approach is advantageous as it can be shown that the log-pseudolikelihood is formally equivalent to the log-likelihood of independent weighted Poisson variables, i.e.

$$\log\mathrm{PL}_A\left(\theta,\phi;\mathbf{x}_1,\ldots,\mathbf{x}_M\right)\approx\sum_{m=1}^{M}\sum_{j=1}^{l_m}\left(y_{mj}\log\lambda_{\theta,\phi_m}\left(\mathbf{u}_{mj}\right)-\lambda_{\theta,\phi_m}\left(\mathbf{u}_{mj}\right)\right)w_{mj}$$

where w_{mj} is the quadrature weights, $y_{mj}=\frac{z_{mj}}{w_{mj}}$ and

$$z_{mj}=\begin{cases}1 & \text{if} \quad \mathbf{u}_{mj}\in\{\mathbf{x}_{m1},\ldots,\mathbf{x}_{mn(\mathbf{x}_m)}\}\\ 0 & \text{if} \quad \mathbf{u}_{mj}\notin\{\mathbf{x}_{m1},\ldots,\mathbf{x}_{mn(\mathbf{x}_m)}\}\end{cases}.\tag{2}$$

The initial step in fitting the proposed spatial model involves creating a quadrature scheme on the linear network. Subsequently, we replicate dummy points for each potential mark, which in this case, correspond to the ID of the visitor. These replicated dummy points are generated at the same locations as the data points, but with distinct marks. If the original point pattern includes additional external information, such as individual covariates, their values remain associated with the generated dummy marked points. This is important as it allows for the inclusion of individual-specific covariates in the model specification, denoted as $\mathbf{Z}(\mathbf{u}_{mj})$, which is replicated for all the quadrature points. Note that the suggested approach basically requires the generation of new dummy points for each level of the marks. Indeed, the procedure follows that of simulating a multitype point pattern. For this reason, our proposal does not directly extend to the case of individual covariates defined on a continuous scale.

Finally, the two sets of dummy points and the data points are superimposed, and the quadrature weights and the indicators (2) are computed.

In this paper, we apply the proposed models to these newly generated quadrature points, allowing for the incorporation of random effects and subject-specific

covariates. We denote the location of these new sets of points as $\mathbf{u}_{im}$. The model can be then fitted through the R Core Team (2023) package `mgcv` (Wood 2017).

For the intensity function $b_{\theta,\phi_m}(\mathbf{u}_{im})$, we assume an additive structure on the log-scale, that is:

$$b_{\theta,\phi_m}(\mathbf{u}_{im}) = \exp\left(\theta_1 B_1\left(\mathbf{u}_{im}\right) + \phi_{1m} B_2\left(\mathbf{u}_{im}\right) + \theta_3 B_3\left(\mathbf{u}_{im}\right) + B_4\left(\mathbf{u}_{im}\right)\right).$$

In detail, we assume that $B_1(\mathbf{u}_{im}) = \mathbf{1}$ and $B_3(\mathbf{u}_{im})$ is the *distance from the nearest attraction*, computed as the shortest-path distance on the road network of the city. (see Fig. 5). This spatial covariate replaces the dummy variables of the model in D'Angelo et al. (2021), which indicated the visited specific touristic attraction by the cruise passenger, resulting from the questionnaire issued to the cruise passengers when disembarking. For this reason, the dummy variables considered in D'Angelo et al. (2021) could suffer from the participants' recall bias. This bias is overcome by the spatial covariate $B_3(\mathbf{u}_{im})$, which accounts for the spatial displacement of the considered touristic attractions by the GPS tracking data. In addition, $B_2(\mathbf{u}_{im})$ denotes the `ID` of the visitor, included as a random effect. $B_4(\mathbf{u}_{im})$ is a nonparametric function for the spatial location $\mathbf{u}_{im} \in L$, which might be estimated through kernel, splines or other nonparametric techniques. Further external covariates, whose values are obtained for each point of the quadrature scheme when generating their new location as $\mathbf{Z}(\mathbf{u}_{mj})$, can be included in the specification of the model.

To describe the interaction function $h_{\theta,\phi_m}(\mathbf{u}_{im}, \mathbf{v}_{im})$, we consider:

$$h_{\theta,\phi_m}(\mathbf{u}_{im}, \mathbf{v}_{im}) = \exp(\theta_2 H(\mathbf{u}_{im}, \mathbf{v}_{im}) + \phi_{2m} H(\mathbf{u}_{im}, \mathbf{v}_{im})).$$

We propose a smooth interaction function $H(\cdot,\cdot)$, which is assumed to be dependent only on the shortest-path distance between any pair of points, i.e. the length of the shortest path between the location of the two points on the network.

For two points located on the network, denoted as $\mathbf{u}$ and $\mathbf{v}$, we define the interaction function as follows:

$$H(\mathbf{u}, \mathbf{v}) = \begin{cases} \left(1 - \left(\frac{d(\mathbf{u},\mathbf{v})}{R}\right)^2\right)^2 & \text{if} \quad 0 < d(\mathbf{u}, \mathbf{v}) \leq R \\ 0 & \text{else} \end{cases}, \tag{3}$$

where $d(\mathbf{u}, \mathbf{v})$ represents the shortest-path distance, and $R \geq 0$ defines the radius of interaction.

In this particular application, we set the interaction radius to $R = 100\,m$, which is considered a reasonable distance threshold that captures potential interactions amongst visitors' choices of stop locations.

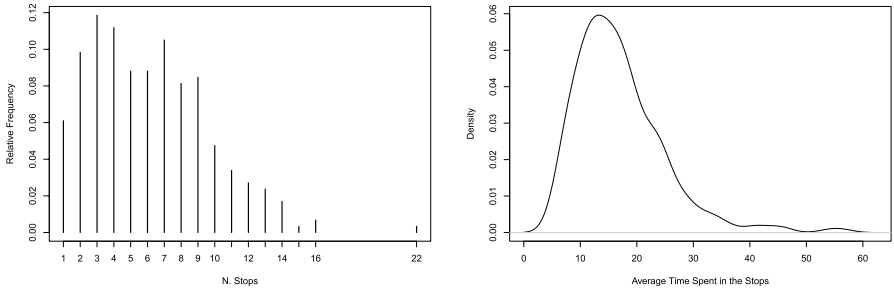

Fig. 2 On the left: Relative frequency distribution of the number of cruisers stops in Palermo. On the right: Distribution of the average time spent in each stop, derived from the DBSCAN algorithm

6 Discussion of results in Palermo city

6.1 Stop activity identification

The DBSCAN algorithm summarises the information on the GPS tourist tracks by counting the number of stops and the time spent at each stop. The application of the DBSCAN to the 303 cruise passengers sampled during six days of surveys

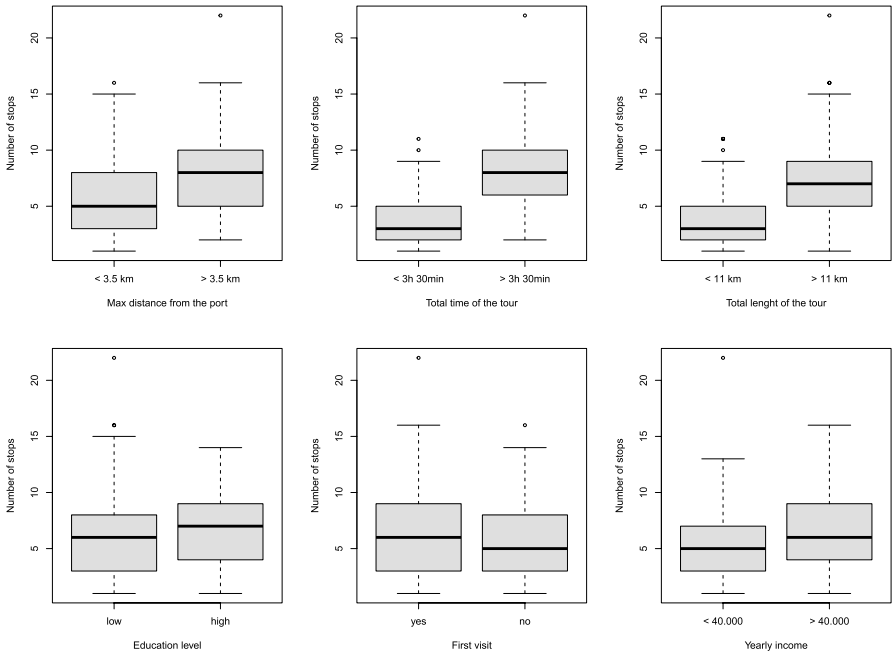

Fig. 3 Boxplots of the number of visitors' stops in Palermo for the considered categorical covariates

identified 1809 stops, with an average number of stops per tourist equal to 6, with a minimum of 1 and a maximum of 22. The duration of stops ranges from 5 to 56 min, with a mean of 17 and a median of 15 min. Figure 2 show the relative frequency distribution of the cruise passengers' stops (on the left) and the average time spent at each stop (on the right), respectively.

Both distributions are positively skewed; around 11% of the cruise passengers make eight stops, and 50% of the sample stops between 7 and 13 times.

6.2 Determinants of stop number

In Fig. 3, bivariate plots of the number of stops according to the set of considered categorical covariates are reported.

As for the exploratory analysis, by looking at the plots in Fig. 3, the number of stops generally appears higher as the mobility behaviour, measured by some synthetic characteristics, increases. Namely, the higher the distance from the port, the higher the number of stops made. Similar considerations hold for the total duration of the tour and the total length of the tour. That is, those who tend to explore more the destination also tend to stop more. As for socio-demographic characteristics, the degree of association is less clear. Nonetheless, the median value of the stops is slightly higher for those with higher education and income levels. On the other hand, this value is slightly lower for repeated visitors, compared to those who visit the destination for the first time, which then seem to prefer the destination's attractions (Wang 2004).

These results are confirmed in D'Angelo et al. (2021), which describes the joint effect of the considered variables on the number of stops by fitting a Poisson regression model, to evaluate the influence of the individual-related variables and itinerary-related characteristics on cruise passengers' stop behaviour, and identify the potential determinants of the stop number.

The results obtained from the previous study are rather reasonable. Regarding socio-demographic characteristics, both education and income are slightly significantly associated with the total stop number. More in detail, the association with the income may be explained with the potential expenditure associated with the stop activities (Thrane 2012), as well as education may be associated with the visit to

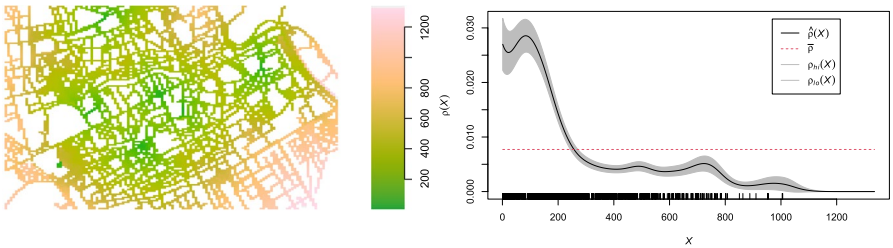

Fig. 4 *Left panel:* Values of the variables *distance from the nearest touristic attraction* in Palermo. *Right panel:* Nonparametric estimate of the intensity of the analysed point process, as a function of the *distance from the nearest touristic attraction*

museums or other types of attractions (Parola et al. 2014). In terms of the visited places, only the Cathedral visit is slightly associated with the total stop number. In contrast, the other considered places of interest are not significantly associated with the total stop number.

Note that in this paper, we aim at accounting for the spatial displacement of points and, therefore, the underlying road network, and we expect that people stop more often near well-known touristic attractions. For these reasons, we compute the spatial variable *distance from the nearest touristic attraction*, where the metric used was the shortest-path distance on the network from the spatial locations $\mathbf{u}$ of the known attractions.

The resulting variable comes in the left panel of Fig. 4. Its marginal effect on the intensity of the process is assessed by plotting the marginal smoothed intensity function. Let $\lambda(\mathbf{u})$ be the intensity of the process under stud and $Z(\mathbf{u})$ the considered covariate. Then, assuming $\lambda(\mathbf{u}) = f(Z(\mathbf{u}))$ where f is a nonparametric estimate of the intensity of the analysed point process, we wonder if the observed intensity depends on the spatial covariate. Smooth estimators of $f(Z(\mathbf{u}))$ were proposed by Baddeley et al. (2012), and the smoothing procedure we consider is based on fixed-bandwidth kernel density estimation. The Poisson confidence bands are also computed.

Looking at the smoothed functions in the right panel of Fig. 4, it is evident that the effect of the spatial covariate varies as a function of the scale. In detail, the intensity exponentially decreases moving away from the nearest tourist attraction, whilst it becomes pretty constant after 200 ms.

6.3 Results of the spatial modelling of the stops on the road network

For the present study, for computational reasons, only two days of the survey are taken into account, referring to the cruise passengers visiting the city after disembarking from the cruise ship. Note that computational times for fitting point process models increase with the number of points. In our particular setting, the computational burden mostly depends on the number of IDs considered as random effects, as this, in turn, affects the quadrature scheme and, consequently, the dummy points generated. The spatial point pattern consists of 278 stops made by 72 visitors, stopping four times on average during their visit to downtown Palermo city on the 25th and 28th April 2014. To properly account for the constrained structure of the space support, the road network of the selected area is considered, providing a linear network L with 4473 vertices and 5399 lines. The quadrature scheme used for model fitting consists of the analysed 278 data points, representing the visitors' stops, and 10,798 dummy points obtained generating the quadrature scheme on the analysed network. This leads to a dataset of 797,472 quadrature points, which is equal to the number of data points plus the number of dummy points, all replicated for the number of marks M.

In Fig. 5, the locations of the observed stops (in red) and the main considered attractions (in green) are displayed.

The motivation for including the socio-economic characteristics and synthetic information on the itinerary undertaken as covariates is threefold: to explain the

Fig. 5 In red: the spatial point pattern of visitors' stops in the downtown of Palermo city on the 25th and 28th April 2014. In green: the location of the tourist attractions (colour figure online)

spatial inhomogeneity, to consider the characteristics of the visit, and to follow the literature on intra-destination behaviour.

In particular, starting from the fitted model in D'Angelo et al. (2021), based on the AIC values, none of the individual-related variables resulted significant.

Moreover, amongst other destination-related considered characteristics, both the geographical configuration of the destination determined by the road network and the shortest-path distance of each stop location from the nearest tourist attraction is taken into account.

Thus, we model the spatial intensity as follows:

Table 1 Parameters' estimates and approximate significance of smooth terms of the Gibbs model (4) for the downtown of Palermo city

	Estimate	Std. error	z value	Pr(> $\|z\|$)
$\hat{\theta}_1$ Intercept	− 9.009	0.924	− 9.749	0.000
$\hat{\theta}_2$ Interaction	0.231	0.013	17.696	0.000
$\hat{\theta}_3$ Distance from the nearest attraction	− 0.0040	0.001	− 2.849	0.004
	edf	Ref.df	Chi.sq	p-value
s(lat,long)	22.312	25.890	143.400	0.000
$\hat{\phi}_{1m}$ s(id)	0.000	71.000	0.000	0.000

$$\log \lambda_{\theta,\phi_m}(\mathbf{u}_{im};\mathbf{x}_m) = \theta_1 + \phi_{1m}B_2(\mathbf{u}_{im}) + \theta_2 v_{im} + \theta_3 Z(\mathbf{u}_{im}) + B_4(\mathbf{u}_{im}) \tag{4}$$

where: $v_{im} = \sum_{j=1}^{n(\mathbf{x}_m)} H(\mathbf{u}_{im}, \mathbf{x}_{jm})$; θ_2 is the fixed effect of the smooth function in Eq. (3); θ_3 is the fixed effect of the *distance from the nearest attraction*, computed as the shortest-path distance on the network from the location of these known attractions; ϕ_{1m} is the random effect of the `ID`; the nonparametric function $B_4(\mathbf{u}_{im})$ is estimated through thin plate regression splines with a chosen number of 29 knots for our analysis.

In Table 1, the estimates of the fixed effects and the summary of the random effects of the selected model (4) are reported.

When $\exp(\hat{\theta}_1)$ is multiplied by the length of the network, the estimated stops for each individual are 14.57, higher than the original average stops. This finding is likely due to the sparsity of the original points in some areas of the network.

The positive interaction parameter $\exp(\hat{\theta}_2) = 1.26$ indicates that overall the visitors' stops attract each other. Therefore, visitors tend to stop in the same spots. Furthermore, $\exp(\hat{\theta}_3) = 0.996$, being less than 1, indicates that moving away from any tourist attraction slightly decreases the probability of visitors stopping. From the significant random effects, we notice that the intensity varies only amongst visitors ($\hat{\phi}_{1m}$), contrary to the interaction ($\hat{\phi}_{2m}$) which is not significant and therefore excluded from the model. The variance for the visitors' random effect is estimated as 0.0003. Overall, this conclusion opens new research perspectives on modelling human behaviour and applying ecological theories (Meekan et al. 2017). Finally, the inclusion of the smooth term $B_4(\mathbf{u}_{im})$ accounting for the spatial coordinates improves the fitting of the model significantly.

To make the estimator unbiased, that is, given the expected number of points $\mathbb{E}[\int_L \lambda(\mathbf{u})d(\mathbf{u})] = n$, the intensity obtained by the Equation (4) has been normalised $\hat{\lambda}(\mathbf{u}) = \frac{n\hat{\lambda}(\mathbf{u})}{\int_L \hat{\lambda}(\mathbf{u})d(\mathbf{u})}$. This is a peculiar problem of models fitted to linear networks, as the resulting intensity is interpreted as the number of points per unit length of the network (and not per unit area as in typical planar point pattern analysis). Therefore, it is often the case that the intensity resulting from a fitted

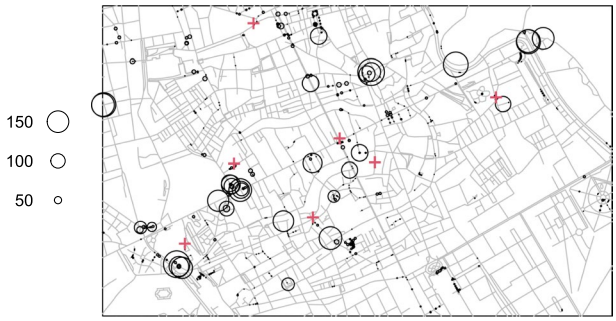

Fig. 6 Estimated pointwise intensities for Palermo, above the 95th percentile: The bigger the circle, the higher the intensity. The intensity has been normalised to obtain the expected number of stops for each location. Locations of the tourist attractions are displayed in red (colour figure online)

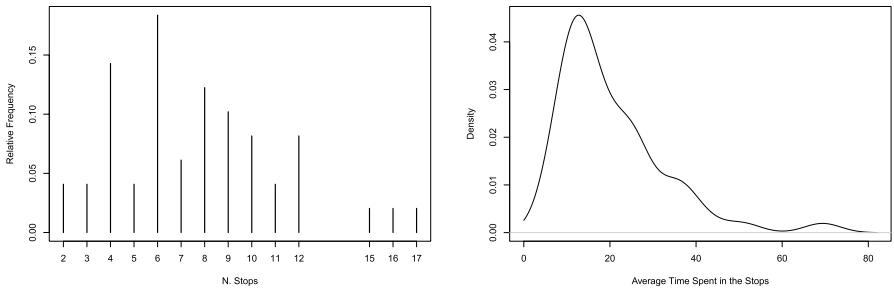

Fig. 7 On the left: Relative frequency distribution of the number of cruisers stops in Dubrovnik. On the right: Distribution of the average time spent in each stop, derived from the DBSCAN algorithm

Fig. 8 *Left panel:* the spatial point pattern of visitors' stops in the downtown of Dubrovnik city (in red) on the 3rd and 5th September 2015, and the location of the tourist attractions (in green). *Right panel:* Values of the variables *distance from the nearest touristic attraction* in Dubrovnik (colour figure online)

model on a linear network does not integrate directly with the number of original points in the analysed point pattern. Therefore, Fig. 6 shows the estimated intensity, displaying the expected number of stops for each location of the quadrature on the network. To make easier the reading and to highlight the regions where visitors are most likely to stop, only the estimated intensities that are higher than the 95th percentile are reported.

7 Discussion of results in Dubrovnik city

7.1 Stop activity identification

The application of the DBSCAN to the 51 cruise passengers sampled during six days of surveys identified 374 stops, with an average number of stops per tourist equal to 7.63, with a minimum of 2 and a maximum of 17. The stop duration ranges from 5.73 to 69.38 min, with a mean of 20.36 and a median of 16.46 min.

Figures show the relative frequency distribution of the cruiser stops (on the left) and the average time spent at each stop (on the right). Both distributions are positively skewed (Fig. 7).

7.2 Results of the spatial modelling of stop locations in Dubrovnik

In the case of Dubrovnik, the spatial point pattern consists of 374 stops made by 49 visitors, stopping seven times on average during their visit in the city downtown, on 3rd and 5th September 2015. The linear network L comes with 1178 vertices and 1245 lines. The quadrature scheme used for model fitting consists 140,434 quadrature points (374 data points and 2492 dummy points).

In the left panel of Fig. 8, the locations of the observed stops (in red) and the main considered attractions (in green) are displayed. As for the Italian case, we model the spatial intensity as in Eq. (4). Indeed, starting from the full model considering all the individual-related variables, the best model for the Croatian data is the same. The spatial covariate obtained as the shortest-path distance from the nearest point of attraction comes in the right panel of Fig. 8.

In Table 2, the estimates of the fixed effects and the summary of the random effects of the final selected model are reported.

The model estimates the stops for each individual, which are 6.18. Some comments on the main differences with respect to the Palermo spatial intensity come as follows. The interaction parameter is estimated as positive also for the Dubrovnik data, indicating that also here, stops tend to attract each other. The variance for the visitors' random effect is estimated as 0.275.

As in the case of Palermo, the exponential of the parameter representing the distance from the nearest point of attraction is negative, even though very close to 1. This suggests that the probability of stopping decreases moving away from the points of interest, and this is indeed confirmed by the clusters found in Fig. 8, quite close to the considered points of interest. Furthermore, only the random effect of the `ID` of the tourist is significant, whilst the interaction amongst their stops is not.

Table 2 Parameters' estimates and approximate significance of smooth terms of the Gibbs model (4) in Dubrovnik

	Estimate	Std. error	z value	Pr(> $\|z\|$)
$\hat{\theta}_1$ Intercept	− 7.773	0.233	− 33.367	0.00
$\hat{\theta}_2$ Interaction	0.031	0.006	4.956	0.000
$\hat{\theta}_3$ Distance from the nearest attraction	− 0.002	0.001	− 2.227	0.026
	edf	Ref.df	Chi.sq	p-value
s(lat,long)	23.960	26.920	189.530	0.000
$\hat{\phi}_{1m}$ s(id)	17.050	48.000	27.310	0.005

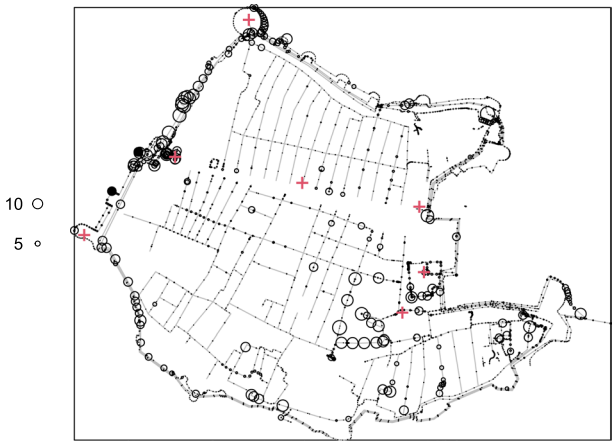

Fig. 9 Estimated pointwise intensities for Dubrovnik. The bigger the circle, the higher the intensity. The intensity has been normalised to obtain the expected number of stops for each location. Locations of the tourist attractions are displayed in red (colour figure online)

Finally, Fig. 9 displays the obtained estimated intensity, showing the expected number of stops for each location of the quadrature on the network.

8 Summary and conclusions

This paper proposes an approach to derive meaningful information on a tourist visit at the destination, starting from the Global Positioning System (GPS) tracking data and the questionnaire-based information. The complex structure of the GPS data requires methods able to synthesise the vast amount of data collected to extract relevant information on the visitors' behaviour. Amongst the various types of information which can be derived from the GPS tracking data, in this contribution, we focused our attention on stop activities as an essential element for destination management, not only because stop locations may highlight business opportunity (Adongo et al. 2017), but also because they may indicate crowded places, sustainability issues at attractions and traffic congestion (Chiou and Hsieh 2020).

To identify cruise passengers' stop activities, a spatial clustering algorithm, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) (Ester et al. 1996), has been applied to the GPS tracking data collected at an individual level. Moreover, thanks to the integration of stop activities, socio-demographic characteristics and other itinerary-related information, it was possible to identify some of the potential determinants of stop activity at the destination. Determining the number of stops and analysing their main determinants is fundamental for service management since the stop locations may identify places where most of the expenditure is concentrated (Thrane and Farstad 2012).

Moreover, a novel model has been proposed to analyse the main determinants of spatial intensity of cruise passengers' stop locations during their visit. The proposed

model considers the linear network determined by the street configuration of the destination under analysis. The results do show that socio-demographic characteristics do not have a significant influence on stop location patterns, and the trip-related characteristics are the most influencing factors, including the distance from the main attractions and the potential interaction amongst cruise passengers in stop location decisions. Thus, attraction locations and interaction amongst cruise passengers seem to be the main determinants in stop location patterns. Nonetheless, these results could be influenced by the particular segment under study, the cruise passengers, characterised by a limited time for visiting the destination and a tendency towards group behaviour (Kriwoken and Hardy 2018).

From a methodological perspective, this paper contributes to the analysis and synthesis of the GPS tracking data, particularly for the modelling of spatial point processes on a linear network. Moreover, the Gibbs point process approach allows for analysing interactions amongst points to check whether attraction or repulsive relationships exist amongst tourists' stop location choices. Whilst most of the recent literature on this topic is concerned with nonparametric intensity estimation, both in space and space-time, our approach contributes to the framework of point processes on networks by proposing a parametric model. A drawback of the present study is that it does not account for the temporal component. Indeed, an exciting line of future research is analysing the spatial point pattern of the individual trajectories (Moradi et al. 2018a) marked by time spent in each stop location.

From an applied perspective, improved knowledge of tourists' spatial behaviour has relevant implications for destination management, given that a better knowledge of the determinants of spatial intensity of visitors' stop locations in urban contexts may orient destination management policy. Nonetheless, another limitation of our study concerns the analysis of stops made by cruise passengers, which may be influenced by various factors that require further investigation. For instance, the decision to take organised tours or the behaviour of passengers in groups may be influenced by prior information provided before visiting the destination or by the accompanying tour guide. A more detailed investigation of the determinants of stops made by cruise passengers is necessary to gain a better understanding of these factors and their potential impact on our results. Also, the analysis is here focused on a restricted destination area. A wider study area would better account for other potential determinants of the stop activity related to individual trajectories and socio-demographic characteristics.

Funding Open access funding provided by Università degli Studi di Palermo within the CRUI-CARE Agreement. The work was supported by the "SMART CULTURE IN SICILY - SCuSi" project (code: 2017-NAZ-0221-DSEAS-D16), azione 1.1.5 PO FESR Sicilia 2014/2020, by the Targeted Research Funds 2023 (FFR 2023) of the University of Palermo (Italy), and by the European Union - NextGenerationEU, in the framework of the GRINS—Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP C93C22005270001). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Declarations

Ethical approval Not applicable.

References

Abbruzzo, A., Ferrante, M., De Cantis, S.: A pre-processing and network analysis of GPS tracking data. Spat. Econ. Anal. **16**(2), 217–240 (2021)

Adongo, C.A., Badu-Baiden, F., Boakye, K.A.A.: The tourism experience-led length of stay hypothesis. J. Outdoor Recreat. Tour. **18**, 65–74 (2017)

Ang, Q.W., Baddeley, A., Nair, G.: Geometrically corrected second order analysis of events on a linear network, with applications to ecology and criminology. Scand. J. Stat. **39**(4), 591–617 (2012)

Atluri, G., Karpatne, A., Kumar, V.: Spatio-temporal data mining: A survey of problems and methods. ACM Comput. Surv. (CSUR) **51**(4), 1–41 (2018)

Baddeley, A., Chang, Y.-M., Song, Y., Turner, R.: Nonparametric estimation of the dependence of a spatial point process on spatial covariates. Stat. Interface **5**(2), 221–236 (2012)

Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., Davies, T.M.: Analysing point patterns on networks-a review. Spat. Stat. **42**, 100435 (2021)

Baddeley, A., Rubak, E., Turner, R.: Spatial Point Patterns: Methodology and Applications with R. Chapman and Hall, Boca Raton (2015)

Baddeley, A., Turner, R.: Practical maximum pseudolikelihood for spatial point patterns: (with discussion). Aust. N. Z. J. Stat. **42**(3), 283–322 (2000)

Birant, D., Kut, A.: ST-DBSCAN: An algorithm for clustering spatial-temporal data. Data Knowl. Eng. **60**(1), 208–221 (2007)

Brida, J.G., Zapata, S.: Economic impacts of cruise tourism: The case of Costa Rica. Anatolia **21**(2), 322–338 (2010)

Casado-Díaz, A.B., Navarro-Ruiz, S., Nicolau, J.L., Ivars-Baidal, J.: Expanding our understanding of cruise visitors' expenditure at destinations: The role of spatial patterns, onshore visit choice and cruise category. Tour. Manag. **83**, 104199 (2021)

Chiou, Y.-C., Hsieh, C.-W.: Determinants of tourists' length of stay at various tourist attractions based on cellular data. Transportmetrica A: Transp. Sci. **16**(3), 716–733 (2020)

Cooper, C.: Spatial and temporal patterns of tourist behaviour. Reg. Stud. **15**(5), 359–371 (1981)

Cressie, N.: Statistics for Spatial Data. John Wiley & Sons, Hoboken (2015)

Daley, D.J., Vere-Jones, D.: An Introduction to the Theory of Point Processes. Volume II: General Theory and Structure, 2nd edn. Springer-Verlag, New York (2007)

D'Angelo, N., Adelfio, Giada, amd Abbruzzo, A., and Ferrante, M.: Identification and modeling of stop activities at the destination from GPS tracking data. In: Book of short papers - SIS 2021, pp. 811–816 (2021)

D'Angelo, N., Adelfio, G., Abbruzzo, A., Mateu, J.: Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. Ann. Appl. Stat. **16**(2), 791–815 (2022)

D'Angelo, N., Adelfio, G., Mateu, J.: Assessing local differences between the spatio-temporal second-order structure of two point patterns occurring on the same linear network. Spat. Stat. **45**, 100534 (2021)

D'Angelo, N., Adelfio, G., Mateu, J.: Local inhomogeneous second-order characteristics for spatio-temporal point processes occurring on linear networks. Stat. Pap. (2022). https://doi.org/10.1007/s00362-022-01338-4

D'Angelo, N., Payares, D., Adelfio, G., Mateu, J.: Self-exciting point process modelling of crimes on linear networks. Stat. Model. (2022). https://doi.org/10.1177/1471082X221094146

De Cantis, S., Ferrante, M., Kahani, A., Shoval, N.: Cruise passengers' behavior at the destination: Investigation using GPS technology. Tour. Manag. **52**, 133–150 (2016)

Domènech, A., Gutiérrez, A., Anton Clavé, S.: Cruise passengers' spatial behaviour and expenditure levels at destination. Tour. Plan. Dev. **17**(1), 17–36 (2020)

Domènech, A., Gutiérrez, A., Clavé, S.A.: Built environment and urban cruise tourists' mobility. Ann. Tour. Res. **81**, 102–189 (2020)

East, D., Osborne, P., Kemp, S., Woodfine, T.: Combining GPS & survey data improves understanding of visitor behaviour. Tour. Manag. **61**, 307–320 (2017)

Ester, M., Kriegel, H.-P., Jorg, S., and Xu, X.: A density-based clustering algorithms for discovering clusters. In: KDD-96 Proceedings, vol. 2, pp. 226–231 (1996)

Ferrante, M., De Cantis, S., Shoval, N.: A general framework for collecting and analysing the tracking data of cruise passengers at the destination. Curr. Issue Tour. **21**(12), 1426–1451 (2018)

Gong, L., Sato, H., Yamamoto, T., Miwa, T., Morikawa, T.: Identification of activity stop locations in GPS trajectories by density-based clustering method combined with support vector machines. J. Mod. Transport. **23**(3), 202–213 (2015)

Grinberger, A.Y., Shoval, N.: Spatiotemporal contingencies in tourists' intradiurnal mobility patterns. J. Travel Res. **58**(3), 512–530 (2019)

Hu, F., Li, Z., Yang, C., Jiang, Y.: A graph-based approach to detecting tourist movement patterns using social media data. Cartogr. Geogr. Inf. Sci. **46**(4), 368–382 (2019)

Illian, J.B., Hendrichsen, D.K.: Gibbs point process models with mixed effects. Environ.: off. J. Int. Environ. Soc. **21**(3–4), 341–353 (2010)

Kallenberg, O.: An informal guide to the theory of conditioning in point processes. Int. Stat. Rev. **52**(2), 151–164 (1984)

Kriwoken, L., Hardy, A.: Neo-tribes and antarctic expedition cruise ship tourists. Ann. Leis. Res. **21**(2), 161–177 (2018)

Kurashima, T., Iwata, T., Irie, G., and Fujimura, K.: Travel route recommendation using geotags in photo sharing sites. In: Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 579–588 (2010)

Larsen, S., Wolff, K., Marnburg, E., Øgaard, T.: Belly full, purse closed: Cruise line passengers' expenditures. Tour. Manag. Perspect. **6**, 142–148 (2013)

Lew, A.A., McKercher, B.: Trip destinations, gateways and itineraries: The example of Hong Kong. Tour. Manag. **23**(6), 609–621 (2002)

Li, Y., Yang, L., Shen, H., Wu, Z.: Modeling intra-destination travel behavior of tourists through spatio-temporal analysis. J. Destination Mark. Manag. **11**, 260–269 (2019)

Liu, B., Huang, S.S., Fu, H.: An application of network analysis on tourist attractions: The case of Xinjiang, china. Tour. Manag. **58**, 132–141 (2017)

Mateu, J., Moradi, M., Cronie, O.: Spatio-temporal point patterns on linear networks: Pseudo-separable intensity estimation. Spat. Stat. **37**, 100400 (2020)

McKercher, B., Zoltan, J.: Tourist flows and spatial behavior. In: Lew, A.A., Hall, M.C., Williams, A.M. (eds.) The Wiley Blackwell Companion to Tourism, pp. 33–44. Wiley, Malden (2014)

McSwiggan, G., Baddeley, A., Nair, G.: Kernel density estimation on a linear network. Scand. J. Stat. **44**(2), 324–345 (2017)

Meekan, M.G., Duarte, C.M., Fernández-Gracia, J., Thums, M., Sequeira, A.M., Harcourt, R., Eguíluz, V.M.: The ecology of human mobility. Trends Ecol. Evol. **32**(3), 198–210 (2017)

Moradi, M.M., Cronie, O., Rubak, E., Lachieze-Rey, R., Mateu, J., Baddeley, A.: Resample-smoothing of Voronoi intensity estimators. Stat. Comput. **29**(5), 995–1010 (2019)

Moradi, M.M., Mateu, J.: First-and second-order characteristics of spatio-temporal point processes on linear networks. J. Comput. Graph. Stat. **29**(3), 432–443 (2020)

Moradi, M.M., Pebesma, E., Mateu, J.: trajectories: Classes and methods for trajectory data. J. Stat. Softw. Retrieved from (2018a) https://cran.r-project.org/web/packages/trajectories/vignettes/article.pdf

Moradi, M.M., Rodríguez-Cortés, F.J., Mateu, J.: On kernel-based intensity estimation of spatial point patterns on linear networks. J. Comput. Graph. Stat. **27**(2), 302–311 (2018)

Navarro-Ruiz, S., Casado-Díaz, A.B., Ivars-Baidal, J.: Modelling the intra-destination behaviour of cruise visitors based on a three-dimensional approach. J. Destination Mark. Manag. **18**, 100470 (2020)
Okabe, A., Sugihara, K.: Spatial Analysis Along Networks: Statistical and Computational Methods. John Wiley & Sons, Hoboken (2012)
Parola, F., Satta, G., Penco, L., Persico, L.: Destination satisfaction and cruiser behaviour: The moderating effect of excursion package. Res. Transp. Bus. Manag. **13**, 53–64 (2014)
Petry, L.M., Ferrero, C.A., Alvares, L.O., Renso, C., Bogorny, V.: Towards semantic-aware multiple-aspect trajectory similarity measuring. Trans. GIS **23**(5), 960–975 (2019)
Puczkó, L., Bárd, E., and Füzi, J.: Methodological triangulation: the study of visitor behaviour at the hungarian open air museum. In: Cultural tourism research methods, pp. 61–74. CABI Wallingford UK (2010)
R Core Team.: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2023)
Rakshit, S., Baddeley, A., Nair, G.: Efficient code for second order analysis of events on a linear network. J. Stat. Softw. **90**(1), 1–37 (2019)
Rakshit, S., Nair, G., Baddeley, A.: Second-order analysis of point patterns on a network using any distance metric. Spat. Stat. **22**, 129–154 (2017)
Russo, A.P.: The "vicious circle" of tourism development in heritage cities. Ann. Tour. Res. **29**(1), 165–182 (2002)
Shoval, N., Kahani, A., De Cantis, S., Ferrante, M.: Impact of incentives on tourist activity in space-time. Ann. Tour. Res. **80**, 102846 (2020)
Shoval, N., McKercher, B., Birenboim, A., Ng, E.: The application of a sequence alignment method to the creation of typologies of tourist activity in time and space. Environ. Plann. B. Plann. Des. **42**(1), 76–94 (2015)
Smallwood, C.B., Beckley, L.E., Moore, S.A.: An analysis of visitor movement patterns using travel networks in a large marine park, north-western Australia. Tour. Manag. **33**(3), 517–528 (2012)
Stopher, P.: Collecting, Managing, and Assessing Data Using Sample Surveys. Cambridge University Press, Cambridge (2012)
Thrane, C.: Analyzing tourists' length of stay at destinations with survival models: A constructive critique based on a case study. Tour. Manag. **33**(1), 126–132 (2012)
Thrane, C., Farstad, E.: Nationality as a segmentation criterion in tourism research: The case of international tourists' expenditures while on trips in Norway. Tour. Econ. **18**(1), 203–217 (2012)
Van Lieshout, M.: Markov point processes and their applications. World Scientific, Singapore (2000)
Wang, D.: Tourist behaviour and repeat visitation to Hong Kong. Tour. Geogr. **6**(1), 99–118 (2004)
Wood, S.: Generalized Additive Models: An Introduction with R, 2nd edn. Chapman and Hall, Boca Raton (2017)
Yang, L., Wu, L., Liu, Y., Kang, C.: Quantifying tourist behavior patterns by travel motifs and geo-tagged photos from Flickr. ISPRS Int. J. Geo Inf. **6**(11), 345 (2017)
Zheng, W., Zhou, R., Zhang, Z., Zhong, Y., Wang, S., Wei, Z., Ji, H.: Understanding the tourist mobility using GPS: How similar are the tourists? Tour. Manag. **71**, 54–66 (2019)
Zoltan, J., McKercher, B.: Analysing intra-destination movements and activity participation of tourists through destination card consumption. Tour. Geogr. **17**(1), 19–35 (2015)