

# Determinants of spatial intensity of stop locations on cruise passengers tracking data

Nicoletta D'Angelo, Mauro Ferrante, Antonino Abbruzzo, Giada Adelfio

## 1. Introduction

A tourism destination can be seen as a mix of tourist attractions and of tourist supporting elements, such as accommodation, transport and tourist-related services, which make it attractive and accessible and, in turn, determine its value. Various authors have highlighted the importance of managing key locations and of understanding tourist spatial behaviour and its main determinants (Cooper, 1981; Liu et al., 2017; Russo, 2002). Tourist services characteristics and the spatial distributions of attractions represent supply-side factors which have an influence on tourists' spatial behaviour (Zheng et al., 2017). It is acknowledged that the spatial movements of tourists in a destination are also influenced by demand-side factors, such as time budget, motivations, and destination knowledge, to mention but a few (Lew and McKercher, 2006). Moreover, human interactions may have a role in tourists' spatial behaviour, these may include tourists-residents as well as tourist-tourist interactions.

Despite the importance of understanding tourist movements within a destination, collecting data on tourist mobility is not an easy task (Stopher, 2012). Traditional methods are generally based on post-visit questionnaire or trip diaries, which rely on the accurate recall of the places visited and activities made. Moreover, they may introduce a bias on participant's behaviour, who knows which is being observed (East et al., 2017). Nowadays, GPS technology allows to collect information on human mobility at a very high temporal and spatial detail, with no effort required from the participant in recalling the places visited. Since the influential book of Shoval and Isaacson (2009) many studies in tourism field have been conducted by using GPS technology [see Shoval and Ahas (2016) for a review of the first decade].

This paper expands the knowledge of tourists' spatial behaviour within a destination – considering cruise tourists as a case study – by analyzing their stop location pattern in order to highlight the main determinants of spatial intensity of stops at their destination. To this end, a stochastic point process modelling approach on a linear network is proposed. We refer to Baddeley et al. (2020) for a review of spatial point processes on networks.

In this paper, we fit a Gibbs point process model adapted on the network, that takes into account individual-related variables, contextual-level information, and spatial interaction among stop points. From an applied perspective, this allows to determine the attractiveness of various places in the destinations, as well as the influence of destination-related characteristics and of individual-level variables on stop location pattern. Moreover, the use of Gibbs point process approach allows for the analysis of interactions among points, in order to check whether attraction or repulsive relationships exist among tourists' stop location choice. From a more methodological perspective, while most of the recent literature on this topic is concerned with non-parametric intensity estimation, both in space (Moradi et al., 2019) and space-time (Moradi and Mateu, 2020; Mateu et al., 2019), our approach contributes to the framework of point processes on networks by proposing a parametric model.

Nicoletta D'Angelo, University of Palermo, Italy, nicoletta.dangelo@unipa.it, 0000-0002-8878-5986

Mauro Ferrante, University of Palermo, Italy, mauro.ferrante@unipa.it, 0000-0003-1287-5851

Antonino Abbruzzo, University of Palermo, Italy, antonino.abbruzzo@unipa.it, 0000-0003-2196-3570

Giada Adelfio University of Palermo Italy, giada.adelfio@unipa.it 0000-0002-3194-4296

FUP Best Practice in Scholarly Publishing (DOI 10.36253/fup\_best\_practice)

Nicoletta D'Angelo, Mauro Ferrante, Antonino Abbruzzo, Giada Adelfio, *Determinants of spatial intensity of stop locations on cruise passengers tracking data*, pp. 159-164, © 2021 Author(s), CC BY 4.0 International, DOI 10.36253/978-88-5518-304-8.31, in Bruno Bertaccini, Luigi Fabbris, Alessandra Petrucci, *ASA 2021 Statistics and Information Systems for Policy Evaluation. Book of short papers of the opening conference*, © 2021 Author(s), content CC BY 4.0 International, metadata CC0 1.0 Universal, published by Firenze University Press (www.fupress.com), ISSN 2704-5846 (online), ISBN 978-88-5518-304-8 (PDF), DOI 10.36253/978-88-5518-304-8

## 2. Data

The cruise tourism segment was selected for the analysis in consideration of the single exit/entry point and the relatively brief visiting time, which characterize cruise passengers' experience at their destination. These features make the use of GPS technology particularly suitable for the analysis of such a relevant phenomenon (Shoval, 2008). Data have been collected in Spring 2014 in the city of Palermo through an integration of questionnaire-based survey and GPS technology [see Ferrante et al. (2018) for details on data collection procedures]. For the purposes of the present study, due to computational reasons, only two days of survey have been considered, referred to cruise passengers visiting the City after disembarking from the cruise ship. After pre-processing of GPS tracking data, stop locations were derived through the implementation of the *dbscan* algorithm on individual trajectories, according to the procedure described in Abbruzzo et al. (2020).

The final spatial point pattern considered consists of 429 stops made by 58 visitors, stopping 7 times on average during their visit in the downtown of Palermo city on the 27<sup>th</sup> and 28<sup>th</sup> April 2014. In order to properly account for the constrained structure of the space support, the road network of selected area was considered, providing a linear network  $L$  with 4473 vertices and 5399 lines. Other information have been derived both from destination-related characteristics, questionnaire-based survey, whereas synthetic information on cruise passengers' spatial mobility at the destination have been derived from individual trajectories. As for destination-related characteristics, beyond the geographical configuration of the destination, determined by the road network, also the shortest-path distance of each stop location from the nearest tourist attraction was computed. In Figure 1, the locations of stop locations are displayed in red, along with the main attractions considered, displayed in green. Among socio-demographic characteristics, according to the literature on tourist mobility, age, education level, and income are supposed to be the main potential determinants of the spatial studied phenomenon. In addition, synthetic information derived from individual trajectories includes: total length of tour, total duration of the visit, maximum distance from the port location, and average speed.



Figure 1: In red: the spatial point pattern. In green: the location of the touristic attractions.

### 3. Model proposal

We here introduce a novel modelling approach for describing the spatial behaviour of the visitors. In detail, we fit a parametric model to the visitors' stops accounting for both the underlying network and the individual tourists' choices by introducing a random subject-specific effect. At this aim we refer to the Gibbs point process models with mixed effects (Illian and Hendrichsen, 2010), conforming the procedure to the linear networks context. Let  $M$  be the number of visitors on a linear network  $L$ , each generating the point patterns  $\mathbf{x}_1, \dots, \mathbf{x}_M$  that can be thought as the individual pattern of stops. This flexible procedure allows to account for the individual information both by suitable random and fixed factors, and by external covariates. We therefore assume, for each  $\mathbf{x}_m$  with  $m = 1, \dots, M$ , a pairwise interaction process (Van Lieshout, 2000) with conditional intensity (Kallenberg, 1984) given by:

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

where  $n(\mathbf{x}_m)$  is the number of points in  $\mathbf{x}_m$ , that is, the number of stops per visitor,  $b_{\theta, \phi_m}(u)$  and  $h_{\theta, \phi_m}(u, v)$  are two functions that model the intensity and the interaction, respectively. For estimation purposes, the Berman-Turner device for maximum pseudolikelihood is considered. The final quadrature scheme used for model fitting consists of the analysed 429 data points, representing the visitors' stops, and of 10798 dummy points, obtained generating the quadrature scheme on the analysed network. This leads to a dataset of 651166 quadrature points, that is equal to the number of data points plus the number of dummy points, all replicated for the number of marks  $M$ . In this paper, we fit the proposed model to these new quadrature points, in order to enable the inclusion of random effects and subject-specific covariates. We denote by  $u_{im}$  the location of the new set of points.

As for the intensity function  $b_{\theta, \phi_m}(u)$ , we set  $B_1(u_{im}) = \mathbf{1}$ , with  $\mathbf{1}$  the identity function and  $B_3(u_{im})$  is the *distance from the nearest attraction* (see Figure 1). In addition,  $B_2(u_{im})$  denotes the ID of the tourist, included as a random effect.  $B_4(u_{im})$  is a non-parametric function for  $u_{im} \in L$ , estimated through thin plate regression splines with a chosen number of 29 knots for our analysis. Therefore, for the intensity function we have:

$$b_{\theta, \phi_m}(u_{im}) = \exp(\theta_1 + \phi_{1m}B_2(u_{im}) + \theta_3B_3(u_{im}) + B_4(u_{im})).$$

To describe the interaction function  $h_{\theta, \phi_m}(u, v)$ , we propose a smooth interaction function  $H(\cdot, \cdot)$  which is assumed dependent only from the shortest-path distance between any pairs of points, i.e. the length of the shortest path between the location of the two points on the network. For two points occurring on the network, with location  $u$  and  $v$ , we define:

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

where  $d(u, v)$  is computed as the shortest-path distance, and  $R \geq 0$  defines the radius of interaction. Therefore, for the interaction function we have:

$$h_{\theta, \phi_m}(u_{im}, v_{im}) = \exp(\theta_2 H(u_{im}, v_{im}) + \phi_{2m} H(u_{im}, v_{im})).$$

In this application, the interaction radius is set to  $R = 100$  meters, as a reasonable threshold of distance up to which we assume that there may be interaction among visitors' stop location choice.

In order to explain the spatial inhomogeneity and to consider the characteristics of the visit, socio-economic characteristics and synthetic information on the itinerary undertaken are included as covariates. These are:

- *income*: *yearly income*, dichotomized in  $< 40000$  and  $\geq 40000$  euro;
- *education*: *education level*, dichotomized in *low* (High school diploma or Bachelor degree) and *high* (Master or Ph.d.);
- *visit*: *independent visit*, indicating whether the visitor is travelling independently (*yes*) or by an organized visit (*no*);
- *dist*: *maximum distance from the port*, dichotomized in  $> 3.5$  and  $\leq 3.5$  km.

Thus, we propose to model the spatial intensity as:

$$\log \hat{\lambda}_{\theta, \phi_m}(u_{im}) = \hat{\theta}_1 + \hat{\phi}_{1m} B_2(u_{im}) + \hat{\theta}_2 v_{im} + \hat{\theta}_3 Z(u_{im}) + B_4(u_{im}) + \hat{\phi}_{2m} v_{im} + \hat{\theta}_4 \text{income} + \hat{\theta}_5 \text{education} + \hat{\theta}_6 \text{visit} + \hat{\theta}_7 \text{dist} \quad (2)$$

where:  $v_{im} = \sum_{j=1}^{n(\mathbf{x}_m)} H(u_{im}, x_{jm})$ ;  $\theta_2$  is the fixed effect of the smooth function in (1);  $\theta_3$  is the fixed effect of the *distance from the nearest attraction*;  $\phi_{1m}$  is the random effect of the ID; and  $\phi_{2m}$  represents the random effects for the interaction smooth function.

## 4. Results

In Table 1 the estimates of the fixed effects and the summary of the random effects of the final selected model are reported.

Table 1: Model coefficients and approximate significance of smooth terms of the Gibbs model

	Estimate	Std. Error	z value	Pr(>  z )	
$\hat{\theta}_1$ Intercept	-10.808	0.519	-20.834	0.000	***
$\hat{\theta}_2$ Interaction	0.152	0.010	15.305	0.000	***
$\hat{\theta}_3$ Distance from the nearest attraction	-0.005	0.001	-3.369	0.001	***
$\hat{\theta}_4$ Income (<40000)	0.412	0.148	2.775	0.006	**
$\hat{\theta}_5$ Education (low)	0.296	0.144	2.057	0.040	*
$\hat{\theta}_6$ Independent visit (yes)	0.796	0.266	2.993	0.002	**
$\hat{\theta}_7$ Max distance from the port (> 3.5 km)	0.627	0.190	3.304	0.001	***
	edf	Ref.df	Chi.sq	p-value	
s(lat,long)	21.036	24.91	119.22	0.000	***
$\hat{\phi}_{1m}$ s(id)	0.000	44.00	0.00	0.000	***
$\hat{\phi}_{2m}$ s(v,id)	19.269	45.00	40.16	0.000	***

When  $\exp(\hat{\theta}_1)$  is multiplied by the length of the network, the estimated stops for each individual are 2.4, lower than the original average stops. This is likely due to the sparsity of the original points in certain regions of the network. Regarding the fixed part of the model, among socio-demographic characteristics, cruise passengers with higher level of education and higher income tend to stop more. This is in line with expectations, by considering both a more detailed enjoyment of cultural attractions for people with a higher education level, and a potential association of stops with spending activities, such as purchasing of food and beverage, visit to museums, etc. Also being and independent cruise passengers increases the stop intensity, compared to organized cruise passengers. This is likely due to the fixed scheduling of activities of the organized tour. Still, maximum distance from the port has been considered as a *proxy* of the degree of exploration of the destination (Jaakson, 2004), and it resulted positively associated also with stop intensity. The positive interaction parameter  $\exp(\hat{\theta}_2) = 1.164$  indicates that

overall the visitors' stops attract each other. Therefore, visitors tend to stop in the same spots. Furthermore,  $\exp(\hat{\theta}_3) = 0.995$  indicates that moving away from any tourist attraction slightly decreases the probability of visitor stopping. From the significant random effects, we notice that not only the intensity varies among visitors ( $\hat{\phi}_{1m}$ ), but also the interaction ( $\hat{\phi}_{2m}$ ). This opens new research perspectives on the modeling of human behaviour, and on the application of ecological theories (Meekan et al., 2017). Finally, the inclusion of the smooth term  $B_4(u_{im})$  accounting for the spatial coordinates improves significantly the fitting of the model.

In order to make the estimator unbiased, that is, given the expected number of points  $\mathbb{E}[\int_L \lambda(u)d(u)] = n$ , the intensity obtained by (2) has been normalized  $\hat{\lambda}(u) = \frac{n\hat{\lambda}(u)}{\int_L \hat{\lambda}(u)d(u)}$ . Therefore, in Figure 2 the estimated intensity is shown, displaying the expected number of stops for each location. We report only those estimated intensities higher than the 99<sup>th</sup> percentile, to facilitate reading and to highlight the regions where visitors are most likely to stop.

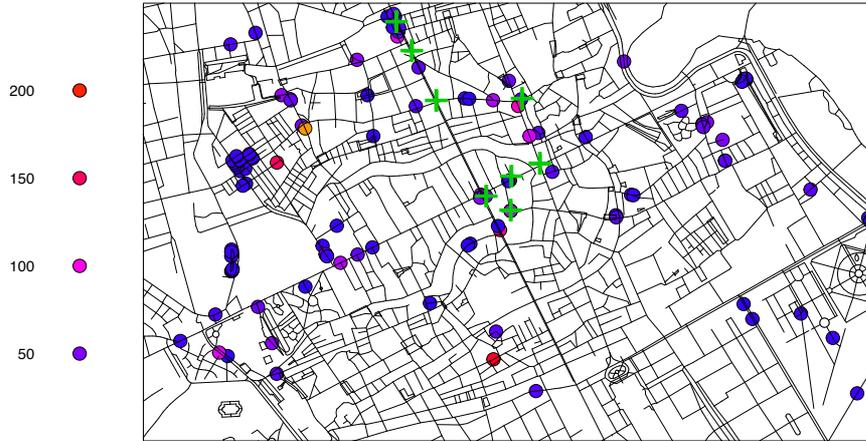


Figure 2: Estimated pointwise intensities above the 99<sup>th</sup> percentile: the lighter the colour the higher the intensity. The intensity has been normalized in order to obtain the expected number of stops for each location. Location of the tourist attractions are displayed in green.

## 5. Conclusion

In this paper, we have proposed a novel model to analyze the main determinants of spatial intensity of cruise passengers' stop locations during their visit. The proposed model allows taking into account the linear network determined by the street configuration of the destination under analysis. The results show an influence of both socio-demographic and trip-related characteristics on the stop location patterns, as well as the relevance of distance from the main attractions, and potential interactions among cruise passengers in stop configuration. The proposed approach represents an improvement both from the methodological perspective, related to the modelling of spatial point process on a linear network, and from the applied perspective, given that a better knowledge of the determinants of spatial intensity of visitors' stop locations in urban contexts may orient destination management policy. A limit of the present study is not accounting for the temporal component. Also, the analysis is here focused in a restricted area of the destination. Considering a wider study area would allow to better account for covariates related to the individuals trajectories. Indeed, the total length of the tour, as well as the duration of the visit, represent useful information that could influence visitor's stop location choice.

## References

- Abbruzzo, A., Ferrante, M., & De Cantis, S. (2020). A pre-processing and network analysis of GPS tracking data. *Spatial Economic Analysis*, pp. 1–24, doi: 10.1080/17421772.2020.1769170.
- Baddeley, A., Nair, G., Rakshit, S., McSwiggan, G., & Davies, T.M. (2020). Analysing point patterns on networks – a review. *Spatial Statistics*, 100435.
- Cooper, C. P. (1981). Spatial and temporal patterns of tourist behaviour. *Regional Studies*, **15**(5), pp. 359–371.
- East, D., Osborne, P., Kemp, S., & Woodfine, T. (2017). Combining GPS & survey data improves understanding of visitor behaviour. *Tourism Management*, **61**, pp. 307–320.
- Ferrante, M., De Cantis, S., & Shoval, N. (2018). A general framework for collecting and analysing the tracking data of cruise passengers at the destination. *Current Issues in Tourism*, **21**(12), pp. 1426–1451.
- Illian, J.B., & Hendrichsen, D. K. (2010). Gibbs point process models with mixed effects. *Environmetrics*, **21**, pp.341–353.
- Jaakson, R. (2004). Beyond the tourist bubble?: cruiseship passengers in port. *Annals of tourism research*, **31**(1), pp. 44–60.
- Kallenberg, O. (1984). An informal guide to the theory of conditioning in point processes. *International Statistical Review/Revue Internationale de Statistique*, pp. 151–164.
- Lew, A., & McKercher, B. (2006). Modeling tourist movements: A local destination analysis. *Annals of tourism research*, **33**(2), pp. 403–423.
- Liu, B., Huang, S. S., & Fu, H. (2017). An application of network analysis on tourist attractions: The case of Xinjiang, China. *Tourism Management*, **58**, pp. 132–141.
- Mateu, J., Moradi, M., & Cronie, O. (2019). Spatio-temporal point patterns on linear networks: Pseudo-separable intensity estimation. *Spatial Statistics*, 100400
- Meekan, M. G., Duarte, C. M., Fernández-Gracia, J., Thums, M., Sequeira, A. M., Harcourt, R., & Eguíluz, V. M. (2017). The ecology of human mobility. *Trends in Ecology Evolution*, **32**(3), pp. 198–210.
- Moradi, M.M., Cronie, O., Rubak, E., Lachize-Rey, R., Mateu, J., & Baddeley, A. (2019). Resample-smoothing of voronoi intensity estimators. *Statistics and computing*, **29**, pp. 995–1010.
- Moradi, M.M., & Mateu, J. (2020). First- and second-order characteristics of spatio-temporal point processes on linear networks. *Journal of Computational and Graphical Statistics*, **29**(3), pp. 432–443.
- Russo, A. P. (2002). The ‘vicious circle’ of tourism development in heritage cities. *Annals of tourism research*, **29**(1), pp. 165–182.
- Shoval, N. (2008). Tracking technologies and urban analysis. *Cities*, **25**(1), pp. 21–28.
- Shoval, N., & Ahas, R. (2016). The use of tracking technologies in tourism research: the first decade. *Tourism Geographies*, **18**(5), pp. 587–606.
- Shoval, N., & Isaacson, M. (2009). *Tourist mobility and advanced tracking technologies*. Routledge, London.
- Stopher, P. (2012). *Collecting, managing, and assessing data using sample surveys*. Cambridge University Press, Cambridge.
- Van Lieshout, M. N. M. (2000). *Markov point processes and their applications*. World Scientific
- Zheng, W., Huang, X., & Li, Y. (2017). Understanding the tourist mobility using GPS: Where is the next place?. *Tourism Management*, **59**, pp. 267–280.