Spatio-Temporal Linear Network Point Processes for GPS Data Analysis

Nicoletta D'Angelo[1,*], Giada Adelfio[1], Antonino Abbruzzo[1], Jorge Mateu[2], Mauro Ferrante[3]

1. Department of Economics, Business and Statistics, University of Palermo, Italy;
2. Department of Mathematics, University Jaume I, Castellon, Spain;
3. Department of Culture and Society, University of Palermo, Italy;

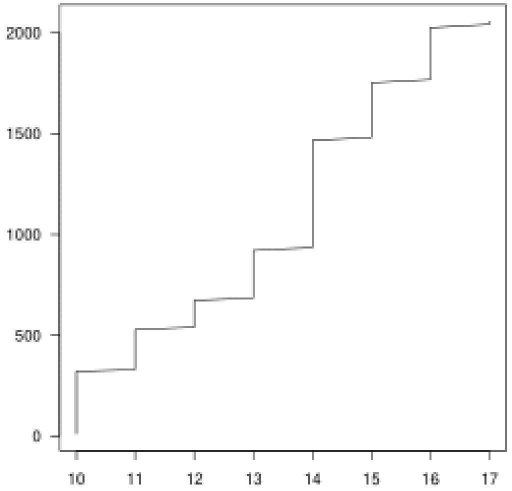
*nicoletta.dangelo@unipa.it

Aim

We propose, for the first time in the context of **spatio-temporal point processes occurring on linear networks**, parametric estimates of the spatial and temporal components following a dependence model with mixed effects approach, to analyse the spatio-temporal distribution of visitors' stops by touristic attractions in Palermo (Italy).

Data and motivating problem

GPS data are collected in Palermo in 2014 through a survey carried out on cruiser passengers disembarking, recording their space-time location.



We assume that the stops identify the spatio-temporal point pattern of interest, consisting of 159 stops of 44 visitors, stopping 4 times on average during the day.

Model proposal 1: parametric separable intensity

The spatial intensity is given by a **Gibbs point process model with random effects**

$$\log \hat{\lambda}_{\theta,\phi_m}(\mathbf{u}_{im}) = \hat{\theta}_1 + \hat{\phi}_{1m} B_2(\mathbf{u}_{im}) + \hat{\theta}_2 v_{im} + \hat{\theta}_3 B_3(\mathbf{u}_{im}) + B_4(\mathbf{u}_{im})$$

where:

- u_{im} is the set of data points generated applying the Berman-Turner device,
- $\hat{\theta}_1$ is the common intercept,
- $\hat{\theta}_2$ is the effect of an interaction term depending on shortest-path distances,
- $\hat{\theta}_3$ is the effect of the *distance from the nearest point of attraction*,
- $\hat{\phi}_{1m}$ is the random effect of the ID.

The chosen model for the purely temporal component is a **Poisson harmonic model**

$$\log \hat{\lambda}(t) = \hat{\delta} + \hat{\alpha}_1 \cos(\omega t) + \hat{\beta}_1 \sin(\omega t) + \hat{\alpha}_2 \cos(2\omega t) + \hat{\beta}_2 \sin(2\omega t) + \hat{\alpha}_3 \cos(4\omega t) + \hat{\beta}_3 \sin(4\omega t).$$

The spatio-temporal intensity is obtained through an estimator for spatio-temporal linear network point processes, based on the first-order separability assumption:

$$\tilde{\lambda}(\mathbf{u}, t) = \frac{\hat{\lambda}(\mathbf{u})\hat{\lambda}(t)}{\int_{L \times T} \hat{\lambda}(\mathbf{u}, t) d_2(\mathbf{u}, t)}$$

where $\hat{\lambda}(\mathbf{u}, t) = \hat{\lambda}(\mathbf{u})\hat{\lambda}(t)/n$ is an unbiased estimator for the expected number of observed points. The purely spatial and temporal intensities, given the network, are usually estimated by non-parametric approaches mostly based on kernel estimators.

Test for spatio-temporal clustering

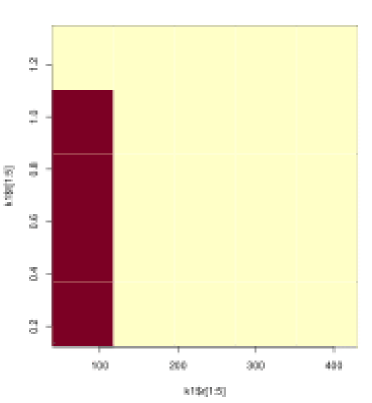
Simulating $Q > 1$ realisations of spatio-temporal point patterns $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_Q$, a test for spatio-temporal clustering can be performed, based on the inhomogeneous spatio-temporal K-functions on linear networks $\hat{K}_q(r, h)$

$$T_q = \int_{r_0}^{r_{max}} \int_{h_0}^{h_{max}} \frac{\hat{K}_q(r, h) - E_K(r, h)}{\sqrt{V_K(r, h)}}.$$

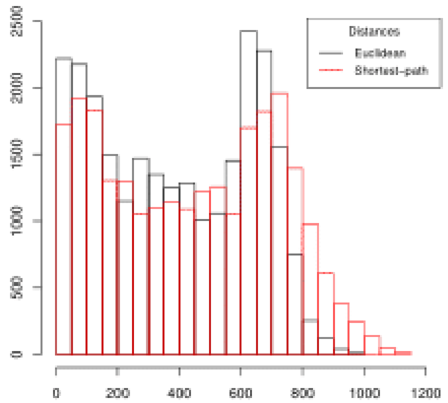
After the same test statistic is computed also for the empirical point pattern T^*, if the obtained p-value $(1 + \sum_{q=1}^{Q} I(T_q > T^*))/(Q + 1)$ is smaller than a significance level α there is evidence against the null hypothesis, that is, the analysed spatio-temporal point process still presents some clustering behaviour.

Having chosen $\alpha = 0.95$, a borderline p-value of 0.05 of the test of clustering suggests that the Model 1 cannot completely catch the features of the observed pattern.

Model proposal 2: Log-Gaussian Cox process



[a]



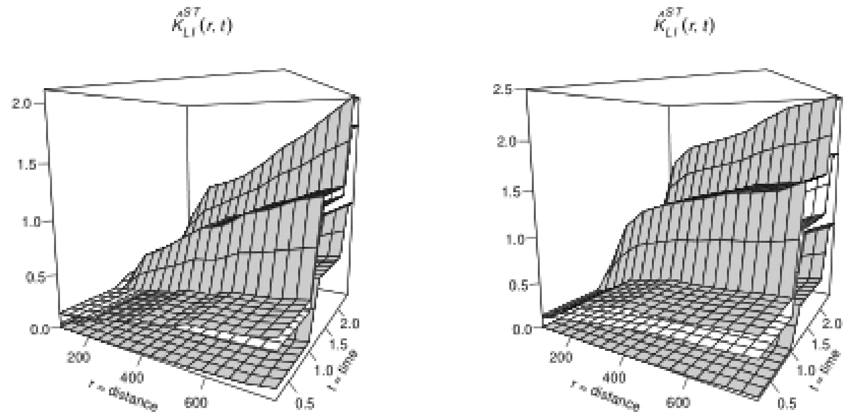
[b]

The presence of ranges where the observed K-function lays above the envelopes (Figure (a)) suggests that a **Log-Gaussian Cox process** (LGCP) model could be further used.

We choose to consider the Euclidean distances to fit the LGCP model (Figure (b)).

Diagnostics

The K-functions $\hat{K}_q(r, h)$ can also be used to obtain upper and lower envelopes at a chosen significance level α, and so to visually assess the possible residual clustered structure of the analysed point pattern, unexplained by the proposed model.



Having obtained a p-value of 0.19 of the test of clustering and knowing that the envelopes of the processes simulated from the fitted LGCP model include the empirical K-function, we conclude that a LGCP achieves a much better fit compared to inhomogeneous Poisson case.

Results

- When $\exp(\hat{\theta}_1)$ is multiplied by the length of the network, the model estimates 8.082637 stops for each individual, higher than the original average stops.
- The positive interaction parameter $\exp(\hat{\theta}_2) = 1.096828$ indicates that overall the visitors' stops attract each other, and therefore, visitors tend to stop in the same spots.
- The estimated paramter $\exp(\hat{\theta}_3) = 0.9877283$ indicates that moving away from any touristic attraction decreases the probability of visitor stopping.
- By the inspection of random effects, we notice that only the intensity varies among visitors, and not the interaction.

Summary and conclusions

- We have described GPS referred to tourists visiting the downtown of Palermo, one of the most touristic cities in Italy, with its many touristic attractions, belonging to the UNESCO-World Heritage listing.
- The generation of the Gaussian Random Field on the network remains an open question.

Thank you for your attention

nicoletta.dangelo@unipa.it

D'Angelo, N., Adelfio, G., Abbruzzo, A., Mateu, J. (2021). Inhomogeneous spatio-temporal point processes on linear networks for visitors' stops data. *The annals of applied statistics*, Forthcoming.