Spatio-temporal classification in point patterns under the presence of clutter

Delfín Carot[1,*] and Jorge Mateu[1]

[1] *Department of Mathematics, University Jaume I, Castellón, Spain; carot@uji.es, mateu@uji.es*
[*]*Corresponding author*

Abstract.
We consider the problem of detection of features in the presence of clutter for spatio-temporal point patterns. This problem was previously treated but only in the spatial context. In particular, Byers et al. *(1998) used k-th nearest neighbour distances to classify points between clutter and features. They proposed a mixture of distributions whose parameters were estimated using an EM algorithm.*
This paper extends this methodology to the spatio-temporal context by considering the properties of the spatio-temporal k-th nearest neighbour distances. We make use of several spatio-temporal n-dimensional distances ($n-1$ spatial dimensions, and 1 temporal dimension), that are mixtures of defined distances for the p-norm. We show close forms for the probability distributions of such k-th nearest neighbour distances. We also present an intensive simulation study that covers a wide range of practical scenarios.

Keywords. *Clutter; EM algorithm; Features; Mixtures; Spatio-temporal point patterns.*

1 Methodological setup

We consider spatio-temporal point patterns living in the same region, with at least one pattern defining the feature of interest, and the other(s) defining the clutter or noise that has to be removed. The classification is based on considering a spatio-temporal measure of similarity able to detect structural differences among the locations of the points in the patterns. We thus consider the following spatio-temporal n-dimensional similarity measure ($n-1$ spatial dimensions, and 1 temporal dimension), which is a mixture of defined distances for the p-norm

$$D^{ST}\left((a_1,a_2,\ldots,a_{n-1},t^a),(b_1,b_2,\ldots,b_{n-1},t^b)\right) := \left[\left(D^S\left((a_1,a_2,\ldots,a_{n-1}),(b_1,b_2,\ldots,b_{n-1})\right)\right)^{p_r} + \rho D^T(t^a,t^b)^{p_r}\right]^{\frac{1}{p_r}} \tag{1}$$

where $(a_1,a_2,\ldots,a_{n-1},t^a)$, and $(b_1,b_2,\ldots,b_{n-1},t^b)$ are points of a spatio-temporal homogeneous Poisson process (the first $(n-1)$ coordinates are spatial and the last one is the temporal), ρ is a scaling coefficient for the temporal and spatial scales to be commensurate, $p_r \geq 1$ is a suitable value to describe the relationship between space and time in the data (the changing of the data position over the time), D^S is a spatial $(n-1)$-dimensional distance defined by the p-norm

$$D^S\Big((a_1,a_2,\dots,a_{n-1}),(b_1,b_2,\dots,b_{n-1})\Big) := \left(\sum_{i=1}^{n-1} |a_i - b_i|^{p_s}\right)^{1/p_s} \tag{2}$$

where $p_s \geq 1$ is a suitable value for p to measure the distance in the spatial dimension of the data with a distance defined by the p-norm, and D^T is a temporal one-dimensional distance of the form $D^T(t^a,t^b) = |t^a - t^b|$. Thus, we have

$$D^{ST}\Big((a_1,a_2,\dots,a_{n-1},t^a),(b_1,b_2,\dots,b_{n-1},t^b)\Big) := \left[\left(\sum_{i=1}^{n-1} |a_i - b_i|^{p_s}\right)^{\frac{p_r}{p_s}} + \rho|t^a - t^b|^{p_r}\right]^{\frac{1}{p_r}} \tag{3}$$

Let D_k^{ST} be the k-th nearest-neighbour measure associated to D^{ST}. In particular we make use of some particular cases of this similarity measure, which are indeed distances, as follows:
(a) *The Euclidean distance with a "weight" in the temporal coordinate*. In this case, D^S is the Euclidean distance with $p_s = 2$ and $p_r = 2$. Then,

$$D^{e\rho}\Big((a_1,a_2,\dots,a_{n-1},t^a),(b_1,b_2,\dots,b_{n-1},t^b)\Big) := \left(\sum_{i=1}^{n-1} |a_i - b_i|^2 + \rho|t^a - t^b|^2\right)^{\frac{1}{2}}. \tag{4}$$

(b) *The maximum distance with a "weight" in the temporal coordinate*. In this case, D^S is the maximum distance with $p_s = \infty$ and $p_r = \infty$. Also,

$$D^{m\rho}\Big((a_1,a_2,\dots,a_{n-1},t^a),(b_1,b_2,\dots,b_{n-1},t^b)\Big) := \max\Big(|a_1 - b_1|, |a_2 - b_2|, \dots, |a_{n-1} - b_{n-1}|, \rho|t^a - t^b|\Big). \tag{5}$$

(c) *A fire-type distance*. This distance is described by Peng *et al.* (2005) for evaluating a wildfire hazard index. Now, D^S is the Euclidean distance with $p_s = 2$ and $p_r = 1$. And

$$D^{f}\Big((a_1,a_2,\dots,a_{n-1},t^a),(b_1,b_2,\dots,b_{n-1},t^b)\Big) := \left(\sum_{i=1}^{n-1} |a_i - b_i|^2\right)^{\frac{1}{2}} + \rho|t^a - t^b| \tag{6}$$

Denoting by $D_k^{e\rho}$, $D_k^{m\rho}$ and D_k^f the corresponding k-th nearest-neighbour distances associated to $D^{e\rho}$, $D^{m\rho}$ and D^f, respectively, we can show the close form of their probability distribution functions, under homogeneous spatio-temporal Poisson processes. The general property states that $D_k \sim \Gamma^{\frac{1}{n}}(k, \alpha_n\lambda)$, where α_n is the coefficient in the volume of the ball of radius r in the associated topology, i.e. $Vol(B_r) = \alpha_n r^n$.

In our context, where we have two types of processes to be classified, we model the k-th nearest-neighbour distances through a mixture of the corresponding k-th nearest-neighbour distances coming from the clutter and feature. Thus we assume the following mixture

$$D_k \sim p\Gamma^{\frac{1}{n}}(k, \alpha_n\lambda_1) + (1-p)\Gamma^{\frac{1}{n}}(k, \alpha_n\lambda_2) \tag{7}$$

where λ_1 and λ_2 are the intensities of the two homogeneous Poisson point processes (clutter and feature). The corresponding parameters associated to this mixture are estimated using an EM algorithm, where in the expectation step we use the close form provided by an inverse Gamma distribution.

2 Simulation study and results

We conducted an intensive simulation study to detect features in a three-dimensional region that covers a wide range of practical scenarios, and implemented an $\mathbb{R}$ function called `NNcleanst`. We considered several shapes for the feature, several spatio-temporal distances, several values for ρ, and also several values for k. To chose the optimal value of k we used an entropy-type measure of separation, $-\sum_{i=1}^{N} \delta_i log(\delta_i)$, where N is the sample size and δ_i are the probabilities of a point belonging to the feature. As a way of example, Figure 1 shows a simulated three-dimensional spatio-temporal data (clutter in the unit cube, feature in a much smaller central rectangular region) and the corresponding estimated feature obtained through our procedure with $k = 10$ and the three distances presented in this abstract.

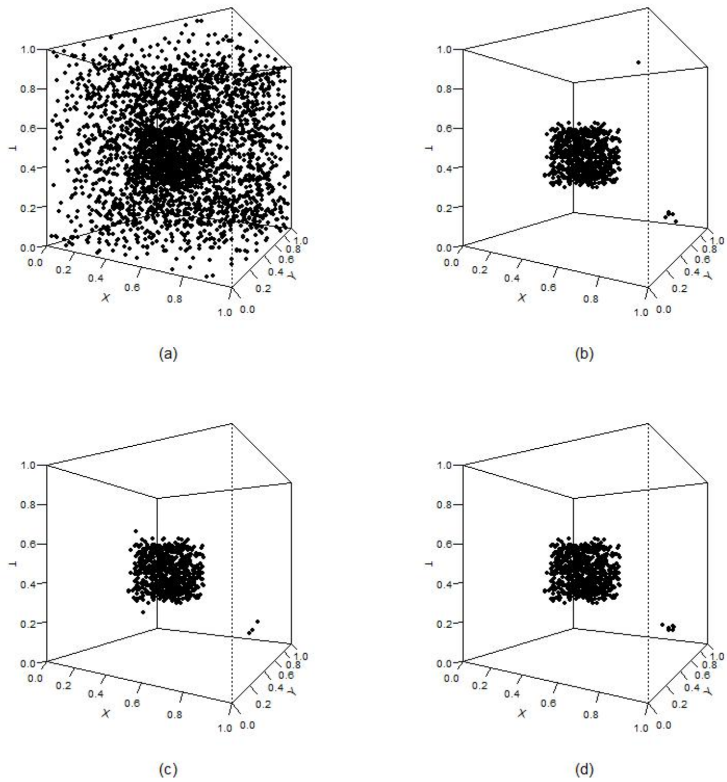

Figure 1: (a) A simulated three-dimensional spatio-temporal data (clutter in the unit cube, feature in a much smaller central rectangular region). Estimated feature obtained through our procedure with $k = 10$ and three distances: *Euclidean with a "weight" in the temporal coordinate* (b), *Maximum with a "weight" in the temporal coordinate* (c), and *Fire-type distance* (d).

Figure 2 shows the corresponding estimated mixtures under these three distances. Finally, Table 1 shows detection and false positive rates for these cases. We can note that our procedure works well for this particular scenario, and is able to detect feature points in this spatio-temporal pattern data.

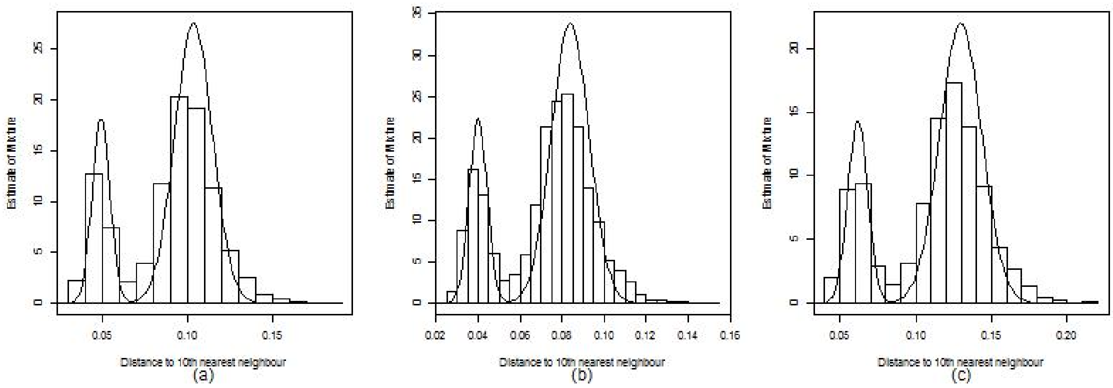

Figure 2: Estimated mixtures under the three distances used in Figure 1: (a) *Euclidean with a "weight" in the temporal coordinate*, (b) *Maximum with a "weight" in the temporal coordinate*, (c) *Fire-type distance*.

Distance	Detection	False positive
Euclid	99.25	3.72
Max	99.25	4.00
Fire	99.25	3.96

Table 1: Detection and false-positive rates (%).

3 Acknowledgements

Acknowledgements are due to the Forest Fire Prevention Section from Generalitat Valenciana for providing the data, and to Pau Aragó to help with GIS data interoperability issues.
Work partially funded by grants PROMETEOII/2014/062 and Gerónimo Forteza 13I106.07 from Generalitat Valenciana.

References

[1] Simon Byers and Adrian E. Raftery (1998). Nearest neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association* **93**, 577–584.

[2] Roger D. Peng, Frederic Paik Schoenberg and James A. Woods (2005). A space-time conditional intensity model for evaluating a wildfire hazard index. *Journal of the American Statistical Association* **100**, 26–35.