

Model learning and validation from spatio-temporal data

General context Point processes are a versatile modelling tool able to catch morphological and statistical features in spatio-temporal data. Cosmology is one application domain were these models are naturally used. In the figure below it can be observed a sample from a galactic catalogue. The blue points represent the galaxies positions and the green linear structure shows the detected filaments structure formed by the galaxy distribution. The filaments were detected first [5] and the distribution of the galaxies was analysed conditionally on the filaments pattern [2].



Figure 1: Data set : galaxy positions with spines (filaments main axes)

Internship research project The work during this internship consists of first getting familiarised with point processes [1] and then developing and integrating the produced code in DRLib a C++ library on purpose designed for the modelling, simulation and inference of marked point processes:

https://gitlab.univ-lorraine.fr/labos/iecl/drlib.

Actually, the library uses Bayesian inference to fit point process models to real 2d data [4, 3]. The aim of the internship is to study and to program model choice and validation procedures using the results of the previously mentioned inference methods. Following [1] it is possible to define a residual measure which naively can be summarised by the formula *residual=observed-modelled* containing information on the divergence between the model and the data. For instance, consider a point process with estimated intensity $\hat{\lambda}(u)$, the predicted number of points falling in any region B is $\int_B \hat{\lambda}(u) du$. Hence, one possible residual measure of a region B may be defined as the difference between the number of observed points and the number of expected points falling in B:

$$R(B) = n(x \cap B) - \int_{B} \hat{\lambda}(u) du$$

The aim of the internship is to extend and to program the construction and the analysis of different residual measures in the spirit of [1, 2].

Supervision This internship will take place at the Institut Elie Cartan de Lorraine (IECL campus de Nancy) within the Probability and Statistics team and at the Inria Nancy within the PASTA research team.

Radu S. Stoica Professor Université de Lorraine, IECL, Inria Nancy radu-stefan.stoica@univ-lorraine.fr

Didier Gemmerlé Researcher Engineer CNRS, IECL didier.gemmerle@univ-lorraine.fr

References

- [1] A. Baddeley, E. Rubak, and R. Turner. Spatial Point Patterns. Methodology and Applications with R. Chapman and Hall/CRC, 2016.
- [2] L. Hurtado-Gil, R. S. Stoica, V. J. Martínez, and P. Arnalte-Mur. Morpho-statistical characterisation of the spatial galaxy distribution through gibbs point processes. *Monthly Notices of the Royal Astronomical Society*, 507(2):1710–1722, 2021.
- [3] R. S. Stoica, M. Deaconu, A. Philippe, and L. Hurtado-Gil. Shadow simulated annealing: a new algorithm for approximate bayesian inference of gibbs point processes. *Spatial Statistics*, 43, 2021.
- [4] R. S. Stoica, A. Philippe, P. Gregori, and J. Mateu. Abc shadow algorithm: a tool for statistical analysis of spatial patterns. *Statistics and Computing*, 27:1225–1238, 2017.
- [5] E. Tempel, R. S. Stoica, R. Kipper, and E. Saar. Bisous model detecting filamentary pattern in point processes. Astronomy and Computing, 16:17–25, 2016.