Thèse – DEEPSURF – 2019

TITLE: Simultaneous model and pattern learning in spatial data. Application to the statistical analysis of the mixing dynamics of geological fluids.

Board :

Director:
Radu Stoica
Affiliation: IECL, Université de Lorraine
contact: radu-stefan.stoica@univ-lorraine.fr
web https://sites.google.com/site/radustefanstoica/
Co-Director:
Madalina Deaconu
Affiliation: Inria/IECL
contact: madalina.deaconu@inria.fr
web: http://www.iecl.univ-lorraine.fr/~Madalina.Deaconu/
Co-Advisor:
Antonin Richard
Affiliation: GeoRessources, Université de Lorraine
contact: antonin.richard@univ-lorraine.fr
web: http://georessources.univ-lorraine.fr/fr/content/richard

PhD's subject

The aim of this thesis is to derive an algorithm able to solve simultaneously pattern and parameter learning. This problem was already tackled in the literature for the case of Markov random fields. Here, the purpose is to continue on this direction, to develop and to extend it towards new modelling framework such as: marked point processes, random sets and random graphs.

Parameter estimation and pattern detection using point processes are both rather well explored and possessing well known open problems, while regarded separately. For instance, for parameter estimation, the still not solved challenge consists of the computation of the ratio of the normalising constants for a given model. While, for pattern detection, such a point would be, the convergence diagnosis of the implemented optimisation algorithm. The theoretical framework and the algorithm proposed by (Stoica et al., 2017) allow to tackle these two problems from a new perspective.

Application to geology

The data to be processed during this thesis will be the chemical composition of some geological fluids (e.g. river waters, groundwaters, hydrothermal fluids). Geological fluids often result from the mixing between various sources. Tracing the sources of geological fluids and analysing the dynamics of fluid mixing is crucial to solve key problems like identifying unknown sources of pollutants, understanding how geothermal systems can be exploited, and deciphering how mineral deposits form (Richard et al., 2010). The variation of concentrations of the different elements composing a fluid are seen as points (mixing terms) in a measure 2D space (e.g. Na vs Cl concentration). The variations of elements concentrations may be explained by the presence of several sources (mixing end-members) not all observed. The spatial distribution of the mixing terms may reveal the presence of unknown sources that need to be detected. In order to get a better understanding of the studied phenomenon, the parameters of the model controlling the spatial distribution of the mixing terms, should be also estimated.

Working plan

1 – state of the art study regarding the methods of statistical inference for spatial data: ABC (Approximate Bayesian Computation), stochastic gradient, classical Monte Carlo methods. (Estimated time 3 months.)

2 – conditionally on the observed mixing end-members, explain the spatial distribution of the mixing terms: implement methods of parameter estimation on real and simulated data. Expected results: spatial modelling of the mixing terms distribution, statistical significance of the end-members contribution, statistical significance of the possible interactions among the mixing terms. (Estimated time 4 months.)

3 - communication of the obtained results: submitting a paper in a peer-reviewed international journal in applied statistics or geosciences. (Estimated time 2 months.)

4 – conditionally on the observed mixing terms, detect the number and the position of the mixing end-members: implement methods of pattern detection on real and simulated data. Expected results: spatial modelling of the mixing end-members distribution, detection of the mixing end-members under spatial constraints, introducing uncertainties regarding the mixing terms. (Estimated time 4 months.)

5 - communication of the obtained results: submitting a paper in a peer-reviewed international journal in applied statistics or geosciences. (Estimated time 2 months.)

6 – simultaneous mixing end-members detection and model parameter estimation. (Estimated time 12 months.)

7 - communication of the obtained results: submitting a paper in a peer-reviewed international journal in applied statistics or geosciences. (Estimated time 2 months.)

8 – submission of the PhD manuscript and preparation of the PhD defence. (Estimated time 4 months.)

PhD Candidate

The duration of the PhD thesis is three years. The candidate should possess a solid background in applied mathematics, more particularly, in probability and statistics. The candidates with a Master's degree in mathematics or applied mathematics, or a corresponding degree of engineering studies in applied sciences or numerical geoscience, are strongly encouraged to postulate to this position.

Bibliography :

R. S. Stoica, M. Deaconu, A. Philippe and L. Hurtado. Simulated annealing through ABC Shadow dynamics. In preparation.

E. Tempel, M. Kruuse, R. Kipper, T. Tuvikene, J. G. Sorce and R. S. Stoica. Bayesian group finder based on marked point processes. Method and application to the 2MRS data set. *Astronomy and Astrophysics*, 618 (A61) : 1-18, 2018.

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(5) : 1225-1238, 2017.

A. Richard , T. Pettke, M. Cathelineau, M.C. Boiron, J. Mercadier, M. Cuney, D. Derome (2010). Brine-rock interaction in the Athabasca basement (McArthur River U deposit, Canada): Consequences for fluid chemistry and uranium *uptake.* Terra Nova *22*, 303-308.