PhD scholarship at Université de Lorraine (Nancy, France): Stochastic seismic structural interpretation of geological faults

Context

The RING Team is seeking outstanding PhD candidates to address research questions in integrative numerical geology. The PhD topic outlined below, can be tailored to the interests and experience of the successful candidate. This full-time position is for a three-year term and shall start in fall 2020.

RING PhD scholarships are sponsored by an international consortium of 12 companies and 139 research institutes. The successful candidates will work in the RING Team[1], a pluridisciplinary group of 12-15 researchers and graduate students working at the interface of geoscience, computer science and applied mathematics. The team is part of *École Nationale Supérieure de Géologie* (ENSG) in the *GeoRessources*[2] laboratory, a research lab of *Université de Lorraine* and CNRS. The research team is driven by passion for developing computer-based methods and theories for geological modeling, serving the geoscience community to address scientific and natural resource managements challenges. It has a strong industry partnership culture.

The PhD candidate will also be affiliated to the mathematics research center *Institut Elie Cartan de Lorraine*[3] (IECL). Finally, the Phd candidate will benefit from a collaboration with the Computational Interpretation Group[4] at USTC in Hefei (China).

PhD duration is three years, and includes teaching opportunities at ENSG.

Location: Nancy, France. Nancy is a UNESCO World Heritage city with a vibrant student life and a rich cultural agenda, only 90 minutes away from Paris, Luxembourg and Strasbourg. Opportunities for research visits to USTC will also be considered.

Candidate profile

The ideal candidate is strongly motivated by the topic, passionate about science, has a solid background in applied mathematics, statistics and physics, and has good scientific writing skills. **A proven experience and taste for computer programming is required**. A good command of statistical modeling and/or and computational geometry is a plus. A background or interest in geoscience is appreciated, but not mandatory.

Candidates should hold a MSc in Computer Science, Geophysics or Physics, Geostatistics, (quantitative) Earth Sciences, Porous Media, Applied Mathematics, Engineering or related fields.

A strong command of English language is required. French language is preferable, but not necessary.

How to apply:

Application files must be sent to jobs@ring-team.org before July 19th and must include:

- A cover letter or email,
- A CV, including contact information for two or more referees
- A research outcome (Master's thesis or paper) written by the candidate
- A transcript of grades

Incomplete application files will not be considered.

---

[1] http://ring.georessources.univ-lorraine.fr
[2] http://georessources.univ-lorraine.fr/
[3] http://www.iecl.univ-lorraine.fr/
[4] http://cig.ustc.edu.cn/

Description

Keywords: Image analysis, Machine learning, Faults

Advisors: Guillaume Caumon (GeoRessources, Université de Lorraine), Radu Stoica (IECL, Université de Lorraine), Xinming Wu (CIG, USTC Hefei)

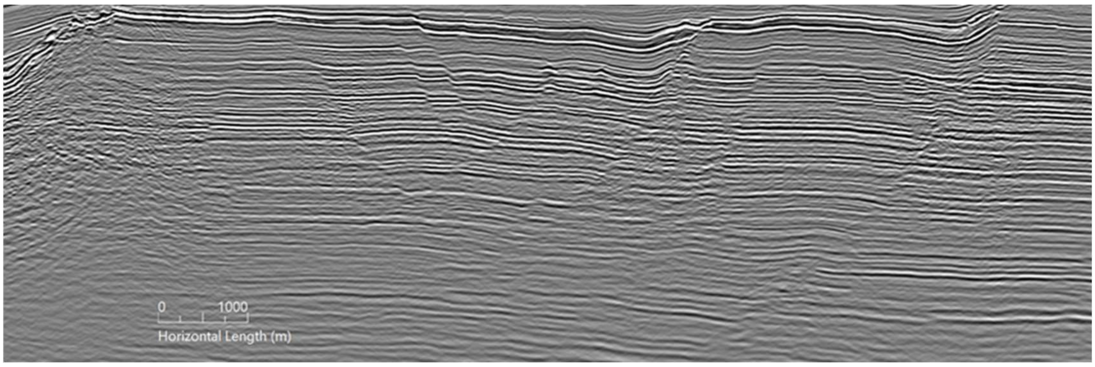

Figure 1: An example of seismic image (Carnavon Basin, Australia. Courtesy of GA)

Seismic interpretation aims at identifying geological features on seismic images (Figure 1). To help interpreters, automatic seismic interpretation has made tremendous progresses in recent years, thanks to advances in signal processing and machine learning methods. For example, very convincing results from a convolutional neural network (CNN) trained with synthetic models have been presented on seismic images of variable quality (Wu et al., 2019). However, this method, like classical expert-based seismic interpretation, only aims at producing the best possible structural interpretation, whereas limited seismic bandwidth and imaging uncertainties can raise interpretation ambiguities. This uncertainty can be consequential when running analysis from the obtained interpretation, e.g., trying to understand reservoir compartmentalization consistently with reservoir production data (Islam & Manzocchi, 2019; Jolley et al., 2007; Julio et al., 2015).

To address this problem, the goal of this project is to propose a stochastic modeling approach exploiting seismic amplitudes and prior knowledge to generate possible structural models. As compared to classical stochastic structural modeling methods, which all require some interpretation picks or surfaces (see Sect. 4 of Wellmann & Caumon, 2018, and references therein), this method would directly start from the seismic image and do not necessarily impose interpretation picks.

FaultNet3D (Wu et al., 2019) could be an excellent starting point for such a method. Indeed, its CNN computes the probability and orientation of faults at every single location in a given seismic image. A key question is, therefore, to sample realistic fault networks from such probabilities. This problem has recently been formulated as a weighted graph decomposition problem (Godefroy et al., 2019, 2020). Stochastic structural simulations in this framework raise several practical and fundamental questions, which could be addressed in the course of this PhD project:

- In a typical seismic image, the number of locations with non-zero fault existence probability is likely in the range of thousands to millions. This yields an enormous search space to sample from. Efficient strategies will need to be designed to reduce this complexity.
- Fault lateral extension, connectivity and segmentation can be highly uncertain, even in good quality seismic data sets (Figure 1). What would be a good mathematical model to represent these specific types of uncertainties? More generally, what model could appropriately deal with interactions between neighboring faults, and reproduce for instance sensible relay zones? How could parameter inference be efficiently achieved for the proposed mathematical model? Is it possible to account for ancillary data such as reservoir compartmentalization / connectivity?

- What is a good structural sampler? Classically, geostatistical simulation algorithms generate independent realizations. However, the ability to control the degree of redundancy of realizations can be very interesting for sampling the search space efficiently, either to maximize the exploration ability of the sampler or, conversely, to look for samples close to an interesting scenario. Sequential graph-based sampling can, in principle, directly generate structural scenarios regrouped by affinity (Godefroy et al., 2020), but this has yet to be verified and measured.
- The notion of similarity between realizations also raises the questions about the criteria used to define such a similarity. Geometrical or topological criteria already exist, but research is certainly needed to handle a variable number of faults and find geological distances which are also relevant in terms of physics.
- Although seismic imaging has dramatically improved during the last decades, the velocity model used to generate the seismic image may also be affected by uncertainties. Exploring the impact of these velocity uncertainties on the sample space could also be considered.

Methodology

The mathematical methodology to be developed to tackle the above questions will adopt the stochastic geometry framework. The seismic fault network can be seen as the realization of a random geometrical set. Two mathematical frameworks allow to deal with such random sets: random graphs and object point processes(Chiu et al., 2013).

The aim of this thesis is to build an object point process able to describe the statistical and the morphological characteristics of the fault network. The main idea is to approximate the 3D seismic network by a network of 2D random interacting rectangles that evolve in a 3D space. The rectangles have random length and width and random 3D orientation. They interact, align and connect in order to form a network. The distribution of these rectangles is controlled by a Gibbsian probability distribution made of two terms. The first term is called data term and is similar to a local likelihood term. It manages to locate the fault surfaces in the regions where the data indicate the highest probabilities of a fault presence. The second term manages the interactions among rectangles. In order to form a fault, the rectangles should align and connect, not cluster in a “noisy” way. The construction of the model follows similar principles to the ones applied to detect galactic filaments and clusters in cosmological catalogues (R. S. Stoica et al., 2010; E. Tempel et al., 2014; Elmo Tempel et al., 2018).

The simulation of the model will be done via MCMC methodology. In order to be numerically efficient, the proposed algorithms should be tailored to the model. This means that the usually “naive” updates of such an algorithm should be replaced by tailored proposals that help the random rectangles to connect and to align in order to form the fault network, while guaranteeing the necessary convergence properties.

Once in the possession of the model and the algorithm, statistical inference can be performed. The rectangles approximating the fault network should maximize the Gibbsian probability distribution built. This maximization can be achieved via a simulated annealing procedure guided by the previously mentioned Monte Carlo dynamics.

Further quantities can be computed and tests can be implemented once in the possession of the approximated fault network. Among them, let us mention the level sets that lead to the local computation of confidence intervals and parametric inference that yields the global characterization of the network (Heinrich et al., 2012; Radu S. Stoica et al., 2017).

Clearly, seismic fault simulation is possible if knowledge of parameters is available. The statistical inference is possible if the fault network is known. None of these quantities is known beforehand. Bayesian modeling is the strategy allowing to overcome these drawbacks, at least partially. Still, even as an always open problem, it deserves to be point out that the biggest mathematical challenge of this thesis is to perform fault detection and related parametric inference simultaneously.

To test and validate the methods, a first phase will consider synthetic data to check for the absence of bias and verify the convergence of the sampler. Real data sets will then be considered, including the Volve field data (North Sea) and some data sets provided by RING Consortium partners and by Geoscience Australia.

The candidate will be advised by Guillaume Caumon (RING, Université de Lorraine), Radu Stoica (IECL, Université de Lorraine) and Xinming Wu (USTC Hefei). The successful candidate will also benefit from interactions with the LOOP project partners.

References

Chiu, S. N., Stoyan, D., Kendall, W. S., & Mecke, J. (2013). *Stochastic geometry and its applications* (Third edition). Chichester, West Sussex, United Kingdom: John Wiley & Sons Inc.

Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2019). Structural interpretation of sparse fault data using graph theory and geological rules. *Mathematical Geosciences*, *51*(8), 1091–1107. https://doi.org/10.1007/s11004-019-09800-0

Godefroy, G., Caumon, G., Laurent, G., & Bonneau, F. (2020). Multi-scenario interpretations from sparse fault evidence using graph theory and geological rules. *Journal of Geophysical Research: Solid Earth (Submitted)*. Retrieved from https://hal.univ-lorraine.fr/hal-02562611

Heinrich, P., Stoica, R. S., & Tran, V. C. (2012). Level sets estimation and Vorob'ev expectation of random compact sets. *Spatial Statistics*, *2*, 47–61. https://doi.org/10.1016/j.spasta.2012.10.001

Islam, M. S., & Manzocchi, T. (2019). A novel flow-based geometrical upscaling method to represent three-dimensional complex sub-seismic fault zone structures into a dynamic reservoir model. *Scientific Reports*, *9*(1), 5294. https://doi.org/10.1038/s41598-019-41723-y

Jolley, S. J., Barr, D., Walsh, J. J., & Knipe, R. J. (2007). Structurally complex reservoirs: an introduction. *Geological Society, London, Special Publications*, *292*(1), 1–24.

Julio, C., Caumon, G., & Ford, M. (2015). Impact of the en echelon fault connectivity on reservoir flow simulations. *Interpretation*, *3*(4), SAC23–SAC34. https://doi.org/10.1190/INT-2015-0060.1

Stoica, R. S., Martínez, V. J., & Saar, E. (2010). Filaments in observed and mock galaxy catalogues. *Astronomy and Astrophysics*, *510*, A38. https://doi.org/10.1051/0004-6361/200912823

Stoica, Radu S., Philippe, A., Gregori, P., & Mateu, J. (2017). ABC Shadow algorithm: a tool for statistical analysis of spatial patterns. *Statistics and Computing*, *27*(5), 1225–1238. https://doi.org/10.1007/s11222-016-9682-x

Tempel, E., Stoica, R. S., Martínez, V. J., Liivamägi, L. J., Castellan, G., & Saar, E. (2014). Detecting filamentary pattern in the cosmic web: a catalogue of filaments for the SDSS. *Monthly Notices of the Royal Astronomical Society*, *438*(4), 3465–3482. https://doi.org/10.1093/mnras/stt2454

Tempel, Elmo, Kruuse, M., Kipper, R., Tuvikene, T., Sorce, J. G., & Stoica, R. S. (2018). Bayesian group finder based on marked point processes: Method and feasibility study using the 2MRS data set. *Astronomy & Astrophysics*, *618*, A81. https://doi.org/10.1051/0004-6361/201833217

Wellmann, F., & Caumon, G. (2018). 3-D Structural geological models: Concepts, methods, and uncertainties. *Advances in Geophysics*, *59*, 1–121. https://doi.org/10.1016/bs.agph.2018.09.001

Wu, X., Shi, Y., Fomel, S., Liang, L., Zhang, Q., & Yusifov, A. Z. (2019). FaultNet3D: Predicting Fault Probabilities, Strikes, and Dips With a Single Convolutional Neural Network. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(11), 9138–9155. https://doi.org/10.1109/TGRS.2019.2925003