Ecole doctorale IAEM, Université de Lorraine

Inference of kinetics of large multi-scale chemical reaction networks

Context. Many chemists have been confronted since the groundbreaking experiment by Miller and Urey in the 50es, or even before, with the difficulty of dealing with large chemical networks comprising hundreds of molecules or more. Such networks arise in particular in a prebiotic context. The famous 1952 Miller-Urey experiment demonstrated the synthesis of a large diversity of organic molecules from inorganic components, and set out a large research program aiming at understanding physico-chemical conditions and processes having led to the emergence of life on the early Earth through a sequence of by-and-large unknown evolution steps. With the multiplication of observations of exoplanets since 2004, the interest has broadened to the discussion of possible "biosignatures" attesting to the presence of life elsewhere in the Universe.

Depending on the groups, the emphasis has been put either on "RNA-first" or "metabolism-first" scenarios. In both cases, the key element to be demonstrated is an evolution mechanism leading to more complex molecules or molecule networks, and that could possibly be extrapolated to extant biological systems. In this respect, autocatalytic processes, characterized by linear instabilities of the underlying equations, are expected to play a prominent role.

Challenges. Detection of compounds proceeds through complex GC-MS (gas chromatography/mass spectroscopy) or LC/MS (liquid chromatography/mass spectroscopy) techniques. The mass spectrum of a compound is in the form of a series of peaks. Databases provide only a tiny fraction of these signatures. For samples with a large diversity, only raw formulas are readily accessible ; thus, it is impossible to write down a closed list of chemical compounds. Experiments clearly show several time phases, in which new compounds may appear, and then disappear, which are very difficult to interpret.

From the mathematical side, huge progress has been made very recently towards a general characterization of autocatalysis, and beyond that, a semi-quantitative description in terms of "hierarchical models" of the time behavior of generic chemical reaction networks under a scale-separation hypothesis.

Project. The main aim of the thesis is to fit mass spectroscopy data derived by chemists interested in the origin of life with the family of "hierarchical models". An adequate statistical method will be built which can be used to infer the parameters of the hierarchical models, which can be interpreted as "proxy" kinetic rates. A Bayesian prior on the kinetic parameters k of potentially all mechanistically simple chemical reactions has been made available by recent work in computational chemistry, complementing chemical expertise, and allowing an inference of the network itself. Because only raw formulas are accessible through measurements, the general inference framework is that of HMM (hidden Markov models), for which a large panel of techniques have been developed, including expectation maximization (EM) and variational methods.

Outcomes. Using the fitted model will make it possible to numerically investigate at a very low computational cost a large variety of experimental set-ups, and hopefully give access to

time-scales beyond the time-duration of the experiments, providing some insights about likely chemical evolution processes of organic matter found on asteroids and planets in their early days after their formation.

Candidate. We are looking for a highly motivated mathematician or theoretical physicist with a background in statistical inference and/or statistical physics, and strong interest in applications and interference with scientists with very different backgrounds. Some proficiency in algorithmic programming in Python is required.

Situation. The project will be hosted at IECL (Institut Elie Cartan de Lorraine), which is the mathematics laboratory of Université de Lorraine, Nancy, France. Salary around xxx €. Starting in Fall 2025 or beginning 2026, at the convenience of the selected candidate. The candidate is expected to interact strongly with experimental partner teams in Marseille (Université Aix-Marseille, PIIM lab, équipe Astro, Astrochimie, Spectroscopie, Théorie, Réactivité, Origines) and Poitiers (IC2MP, Institut de Chimie des Milieux et Matériaux de Poitiers), and with mixed theoretical/experimental close collaborators at ESPCI (Ecole Supérieure de Physique et Chimie Industrielle, Paris).

Contact. Send resume and cover letter to jeremie.unterberger@univ-lorraine.fr and ulysse.herbach@univ-lorraine.fr. Do not hesitate to contact us for further information

References

[NgheUnt1] Nghe P., Unterberger J. (2022). *Stoechiometric and dynamical autocatalysis for diluted chemical reaction networks*, J. Math. Biol. **85**.

[NanNgheStuyUnt] Nandan P., Nghe P., Stuyver T., Unterberger J. *A parametrization of kinetics of organic chemistry reaction mechanisms*, work in progress.

[Parikh2014] Parikh N. (2014). *Proximal algorithms*, Foundations and Trends in Optimization **1**.

[Rob] Robinson W. E., Daines E., van Duppen P., de Jong T., Huck W. T. S. (2022). *Environmental conditions drive self-organization of reaction pathways in a prebiotic reaction network*, Nature Chemistry **14**.

[Unt4] Unterberger J. *Optimal multi-time-scale estimates for diluted autocatalytic chemical networks* (preprint).