

# Stage Master 2 :

Modélisation des dynamiques multi échelles d'un système biologique par apprentissage de données expérimentales biologiques et données expertes. Graphes dynamiques Bayésiens et arbres de décision flous

---

Reconstruction of dynamics of biological systems by automatic learning of a real data base and expertise. Comparison of two formalism: Dynamic Bayesian Networks and fuzzy decision trees.

## Contexte

Traditional mechanistic models of food processes can reproduce their original dynamics quite effectively, but often lack the precious information available in the form of human experts' implicit know-how. In order to overcome these limitations, machine learning techniques are now being applied to food processes, with the aim to create new models integrating both trends emerging from raw data and expertise gathered from skilled practitioners in the field.

We are particularly concerned by microbial ecosystems which are of major importance in fermented food products and represent an important part of our diet. The composition, evolution, and control of fermented food microbiota is an important issue not only for their influence on flavour and organoleptic properties, but also due to their role in food shelf-life and safety. Assessment of microbial ecosystems is a challenging issue for several reasons: multi-scale phenomena, uncertainty in the data, out-of-equilibrium systems, complex quantitative and qualitative information to process. Moreover, many of the revealed secondary metabolites and their derivatives, such as flavor compounds pathways, are unknown, because non-essential in metabolism. (Liu et al, 2014). More generally, the emergence of macroscopic properties from microbial ecosystems is complex, and the modeling of these living systems is challenging, especially when applied to the intestinal microbiote. Emergence of secondary metabolites and macroscopic properties could modify the human global behaviour, and not surprisingly a new field of research now revolves around these issues. Microbial food ecosystems can be seen as a microbiote at a reduced size (Mounier et al., 2008). It allows to test computing methodologies and generate methodological approaches useful for a big challenge in life science. Here lies a huge opportunity for the computing science community developing research for help in the prediction of the missing links through scale reconstruction and connection.

Several machine learning approaches have been used so far in food science, ranging from expert systems with fuzzy rule sets (Marsala, 2013), to dynamic Bayesian networks (Baudrit et al., 2010), to evolutionary algorithms (Tonda et al., 2013). Following this stimulating research line, the candidate student shall build on existing techniques, and apply them to a real-world dataset representing microbiological data in an microbial process.

The candidate will first integrate and use tools developed by the team LFI, LIP6 and MALICES, INRA, to create a fuzzy decision tree and a dynamic Bayesian network model from the same dataset and expert knowledge; the two models will then be compared on their ability to represent the phenomenon, their generality, and their prediction capabilities.

The ideal candidate possess good programming skills, an interest in machine learning and computational intelligence, and an open-minded approach to science.

The internship will last 6 months, and its reimbursement amounts to 479.65 €

## Buts du stage

Test, adapt and compare on a real database of a life science application, automatic learning of dynamic networks and fuzzy decision trees

## Prérequis

- Bonnes compétences en programmation (java, C, C++, Matlab, R), apprentissage, et développement d'interfaces.
- Un intérêt pour la recherche (ce stage pourra éventuellement déboucher sur une thèse).
- Une connaissance des méthodes d'apprentissage automatique ; des reseaux bayésiens dynamiques ou des arbres de décision flous sera un plus.
- Il n'est pas nécessaire d'avoir des connaissances en biologie pour aborder ce sujet, mais l'étudiant aura à collaborer avec des biologistes.

## Lieu de travail

UMR GMPA, Thiverval-Grignon.

## Encadrement

Nathalie Perrot et Alberto Tonda, *Equipe MALICES*.

Christophe Marsala, *Equipe LFI, LIP6*

Françoise Irlinger, Dominique Swennen et Eric Dugat-Bony, *Equipe ECOMIC*.

## Pour candidater

Envoyez CV et lettre de motivation à

[nathalie.perrot@grignon.inra.fr](mailto:nathalie.perrot@grignon.inra.fr), [alberto.tonda@grignon.inra.fr](mailto:alberto.tonda@grignon.inra.fr)

## Quelques références pour commencer

Baudrit, C., Sicard, M., Wuillemin, P. H., & Perrot, N. (2010). Towards a global modelling of the Camembert-type cheese ripening process by coupling heterogeneous knowledge with dynamic Bayesian networks. *Journal of Food Engineering*, 98(3), 283-293.

Liu, M., Bienfait, B., Sacher, O., Gasteiger, J., Siezen, R.J., Nauta, A., Geurts, J.M.W. 2014. Combining chemoinformatics with bioinformatics: in silico prediction of bacterial flavor-forming pathways by a chemical systems biology approach “reverse pathway engineering. *PLOS one*, 9(1).

Marsala, C. Fuzzy Decision Trees for Dynamic Data. in Proc. of the IEEE Symposium on Evolving and Adaptive Intelligent Systems - EAIS'2013, IEEE Symposium Series on Computational Intelligence, Singapore, pp. 17-24, avril 2013.

Mounier J, Monnet C, Vallaes T, Arditi R, Sarthou AS, Hélias A, Irlinger F. 2008. Microbial interactions within a cheese microbial community. *Appl Environ Microbiol.*;74(1):172-81

Tonda, A., Lutton, E., Squillero, G., & Wuillemin, P. H. (2013). A memetic approach to bayesian network structure learning (pp. 102-111). Springer Berlin Heidelberg.

Van Mil, H.G.J. , Foegeding, A.E. , Windhab, E.J. , Perrot, N., Van der Linden, E. (2014 to appear). Using a complex system approach to address world challenges in Food and Agriculture. Submitted to *Trends in Food Science and Technology* and accepted after revisions. July 2014. Revue rang 1, excellente pour all Sc. exceptionnelle pour le food science and technology (2/118 in Food science et technology) (2 years IF 4.135). (article de positionnement avec partenaires Européens).