

Raphaëlle Momal

APPLIED MATHEMATICS · POSTDOC

☎ (+33) 6-27-60-75-79 | ✉ raphaelle.momal@inrae.fr | 🏠 Rmomal.github.io | 📷 Rmomal | 📺 MomalRaphaelle | 🐦 @MomalRaphaelle

Current Position

PostDoc fellow

Bât. 325, INRAE Jouy-en-Josas (78)

METAGENOPOLIS, INRAE

Dec 2020 - Present

Under the supervision of Magali Berland (Research Engineer, MGP), this project aims both the inference and the analysis of the human gut microbiota network, from shotgun metagenomic data.

- Develop a robust inference method with clear shotgun data pre-processing protocol accounting for the sampling effects and the data high sparsity level. The inference focus on direct dependency relationships, and should be able to handle the available clinical covariates.
- Use clustering methods to identify microbial guilds and keystone species, and build a guild catalogue.
- Participate in working groups "Artificial Intelligence for the microbiota", and "Microbial guilds".

Education

Doctoral degree in Applied mathematics

Faculté des Sciences d'Orsay, France

UNIVERSITÉ PARIS-SACLAY, WITHIN THE DOCTORAL SCHOOL N°574 - HADAMARD GRADUATE SCHOOL IN MATHEMATICS

2017 - 2020

Research M.S. in Mathematics applied to Statistics

Rue du Thabor, Rennes (35), France

RENNES 1 UNIVERSITY

2016 - 2017

M.S. Biostatistical Engineer

Rue Blaise Pascal, Bruz (35), France

ENSAI (NATIONAL SCHOOL FOR STATISTICS AND DATA ANALYSIS)

2014 - 2017

CPGE MPSI/MP

44 Bd Saint-Michel, Paris (75), France

LYCÉE SAINT-LOUIS

2012 - 2014

Experience

PhD student in Mathematics applied to Biology and Ecology

16 rue Claude Bernard, Paris (75)

MIA-PARIS (APPLIED MATHEMATICS AND IT), AGROPARIS TECH, INRAE, PARIS-SACLAY UNIVERSITY

Oct 2017 - Nov 2020

- **Supervisors:** Stéphane Robin (MIA-Paris, INRAE/AgroParis Tech), Christophe Ambroise (LaMME, Évry)
- **Title:** Network inference from incomplete abundance data.
- **Keywords:** Variational Expectation-Maximization algorithm, Graphical models, Spanning trees dependency structures, Matrix Tree Theorem, Poisson log-Normal distribution
- **Abstract:** Networks are tools used to represent species relationships in microbiology and ecology. Gaussian Graphical Models provide with a mathematical framework for the inference of conditional dependency networks, which allow for a clear separation of direct and indirect effects. However, observed data are often discrete counts and the inference cannot be directly performed with this model. This work develops a methodology for network inference from species observed abundances. The method relies on specific algebraic properties of spanning tree structures to perform an efficient and complete exploration of the space of spanning trees. The inference takes place in a latent space of the observed counts. Then, observed abundances are likely to depend on unmeasured actors (e.g. species or covariate). This results in spurious edges in the marginal network between the species linked to the latter in the complete network, causing inaccurate further analysis. The second objective of this work is to account for missing actors during network inference. To do so, we adopt a variational approach yielding valuable insights regarding the missing actors.
- **Defense:** November 12th 2020, in front of the jury composed of Chi Tran (Professor, Gustave-Eiffel University, President), Florence Forbes (Research Director, Inria Grenoble, Reviewer), Otso Ovaskainen (Professor, Helsinki University, Reviewer), Camille Coron (Assistant Professor, Paris-Saclay University, Examiner), and Stéphane Dray (Research Director, Lyon 1 University, Examiner).

PELAGIS Observatory - UMS 3462

5 Allée de l'Océan, La Rochelle (17)

SCIENTIFIC STAY - GRANT FROM THE ECOSTAT RESEARCH GROUP.

Jun. 2019

- Worked together with Research Engineer Matthieu Authier to extend our inference method of species interaction networks to macro-ecology.
- Data analysis of 2016 marine mega-fauna survey from New Caledonia.

INRAE - GABI - PSGen

Allée de Vilvert, Jouy-en-Josas (78)

INTERNSHIP - BIOSTATISTICS RESEARCH ENGINEER

Apr. 2017 - Oct. 2017

- Worked under the supervision of Dr. Andrea Rau on statistical integration of OMICS data (RNA-seq, ATAC-seq and HiC data), part of the FR-AgENCODE project (functional annotation of livestock genomes).
- Used the mixed-model paradigm to assess the importance of taking both into account the DNA 3D structure and the chromatin accessibility when explaining the genome-wide expression.

Soladis

94 Rue Saint-Lazare, Paris

INTERNSHIP - STATISTICAL ENGINEER

Jun. 2016- Sept. 2016

- Worked in supervised autonomy on several biotech problems.
- Used SAS software to implement logistic regression models, ANOVAs, Bayesian estimation MCMC procedures.

Skills

Programming R (expert), C++ (basic), Python (basic), SAS (basic)

Desktop tools Latex, R Markdown, git, Linux

Languages French (mother tongue), English (C1, almost native), German (B2, intermediate)

Teaching Experience

Statistical data analysis with R - IT Sessions

Paris-Saclay University (LMO)

L3 LEVEL

Jan. 2020 - present

- Introduce R programming to HEC students and Paris-Saclay students in IT and mathematics (3rd year).
- Teach both the programming language, and peculiar statistical functions dedicated to statistical data analysis.

Mathematical Modelling - Exercises Sessions

Paris-Saclay University (bat. 336)

L1 LEVEL

Oct. 2019 - Jan 2020

- First year students in biology (BCPST).
- Function analysis (one and several variables), Probabilities computation, random variables (discrete and continuous), Laws approximation.

MISS (House for Science Initiation and Awareness) - scientific mediation

Paris-Saclay University (bat. 224)

SCHOOL PUPILS FROM 8 TO 15 YEARS OLD

Oct. 2017 - Oct 2019

- History of numbers: how did they count in Mesopotamia without our current numerary system?
- Mathematics and Botany: to the discovery of fractals in nature.
- Bubbles and foam: how can a foam be solid and liquid both at the same time.

The aim of MISS is to bring the lab into the classroom. Teachers would come with their class at MISS and experiment on a specific subject during an entire day. These scientific workshops are animated by junior scientists (often PhD students), which facilitate exchanges with students.

Scientific production

Articles

Accounting for missing actors in interaction network inference from abundance data

arXiv preprint

R. Momal, S. Robin, C. Ambroise

PREPRINT ARXIV:2007.14299

2020

- Extends the inference method to account for missing actors.
- A simulation study shows the method's efficiency, and two examples using field data illustrates its usage.

Tree-based Inference of Species Interaction Networks from Abundance Data

R. Momal, S. Robin, C. Ambroise

Methods in Ecology and Evolution

DOI:10.1111/2041-210X.13380

2020

- Explains the inference method developed in the PhD project, compares it to existing methods and illustrates its application with examples using field data.
- This article is written towards the community of ecology, by bringing together concepts from the fields of mathematics, statistics and ecology.

Multi-species annotation of transcriptome and chromatin structure in domesticated animals.

BMC Biology vol.17(1), Art. n°108

Sylvain Foissac, Sarah Djebali, Kylie Munyard, Nathalie Vialaneix, Andrea Rau and 31 others

DOI:10.1186/S12915-019-0726-5

2019

IT materials:

nestor (<https://github.com/Rmomal/EMtree>)

R PACKAGE

- Network inference from Species counts with missing actors: nestor is an R package implementing the extended inference method accounting for missing actors.

EMtree (<https://github.com/Rmomal/EMtree>)

R PACKAGE

- EMtree is an R package implementing the inference method developed in my PhD project.
- EMtree was developed as a tool for experts in biology and ecology who wishes to infer a direct dependency network from abundance data they collected.
- The method is appreciated as it efficiently handles large datasets, environmental covariates and experimental offsets.

Fancy plotting with ggraph, ganimate (<https://stateofther.github.io/post/fancy-plotting/GG.html>)

STATEOFTHER WORKSHOP

- Introduction to ggplot framework and ganimate, examples and exercises
- Networks with ggraph and tidygraph, examples and exercises
- Edge selection on a simulated ecological dataset

Scientific groups

StateOfTheR (<https://stateofther.github.io/>)

SELF-STUDY GROUP

- A group of researchers and engineers who meet to deepen their know-how, improve the dissemination of their statistical methods and exchange views on the latest innovations in the field of statistics.
- Participate to monthly workshops and annual bootcamps
- Co-designed and led a workshop with *Marie Perrot-Dockes* on how to use ggplot and R's newest packages to draw fancy graphs, and networks in particular

EcoStat (<https://sites.google.com/site/gdrecostat/>)

RESEARCH GROUP

- Development and application of statistical methods and tools for ecology and evolutionary biology
- Benefited from a grant from the group for a stay in an ecology laboratory (PELAGIS Observatory, La Rochelle, France).

EcoNet (https://cmatias.perso.math.cnrs.fr/ANR_EcoNet.html)

ANR RESEARCH PROJECT

- Advanced statistical modelling of ecological networks

Conferences

With talk:

- Feb. 2021 **Rouen Statistical Seminar**, Laboratoire de Mathématiques Raphaël Salem *Web conference*
- Feb. 2021 **Poitiers Statistical Seminar**, Applied Mathematics Laboratory *Web conference*
- Dec. 2020 **NETBio**, Networks inference in Biology *Web conference*
- Mar. 2020 **EcoStat GdR**, Conference of the EcoStat Research Group *Rennes, France*
- Sep. 2019 **WiMLDS**, Women in Machine Learning and Data Science *Paris, France*
- Jan. 2019 **SMPGD**, Statistical Methods for Post Genomic Data *Barcelona, Spain*
- Dec. 2018 **NETBIO**, Networks inference in Biology *Montpellier, France*
- Jul. 2018 **JOBIM**, Open Days of Biology, IT and Mathematics *Marseilles, France*

Attended:

- Oct. 2019 **NETBIO**, Networks inference in Biology *Saclay, France*
- Sep. 2019 **networks2019**, 4th Symposium on Ecological Networks *Paris, France*
- Jul. 2019 **MMEE**, Mathematical Models in Ecology and Evolution conference *Lyon, France*
- May 2019 **JDS**, Statistical Days *Nancy, France*
- Feb. 2019 **SatRday**, International conference on R *Paris, France*
- Oct. 2018 **EcoStat GdR**, Conference of the EcoStat Research Group *Grenoble, France*
- May 2018 **JDS**, Statistical Days *Saclay, France*

Research Schools

With talk:

- Aug. 2019 **FINIST'R 19**, R and GitHub summer bootcamp organised by StateOfTheR. *Roscoff, France*
- Feb. 2019 **COSTNET**, European early career training - Networks. Course by **Tom A.B. Snijders** (University of Groningen, University of Oxford) and **Tiago de Paula Peixoto** (University of Bath). *Munich, Germany*
- Aug. 2018 **FINIST'R 18**, R and GitHub summer bootcamp organised by StateOfTheR. *Beg Meil, France*
- May 2018 **Chaire MMB**, Spring Research School - Mathematical Modeling and Biodiversity. Course by **S. Robin** and **J. Josse**. *Aussois, France*
- Mar. 2018 **Rochebrune Seminar**, Winter Research School - Statistical Seminar. Course by **Julyan Arbel**. *Rochebrune, France*

Attended:

- Sep. 2019 **StatMathAppli'19**, Summer Research School. Course by **P. Bühlmann** and **M. Lerasles** *Fréjus, France*