

The Working Group on Risk - CREAR, with the support of the IDS dpt, Institut des Actuaire, LabEx MME-DII and the group BFA (SFdS), has the pleasure to invite you to the Seminar by:

**Prof. Jean-Michel POGGI**

Univ. Paris Descartes and LMO - Univ. Paris Sud

**Friday, November 9 , 2018, 12:30pm – 1:30pm**  
**EEE - ESSEC La Défense (CNIT) – Amphi 202**

## "Random Forests for Big Data"

Big Data is a major challenge of statistical science and has numerous algorithmic and theoretical consequences. Big Data always involves massive data and often includes online data and data heterogeneity. Recently statistical methods have been adapted to process Big Data, like linear regression models, clustering methods and bootstrapping schemes. Based on decision trees combined with aggregation and bootstrap ideas, random forests (RF) are a powerful nonparametric statistical method allowing to consider in a single and versatile framework regression problems, as well as classification ones. Focusing on classification problems, this talk proposes a review of proposals that deal with scaling random forests to Big Data problems. These proposals rely on parallel environments or on online adaptations of RF. We also describe how the out-of-bag error is addressed in these methods. Then, we formulate various remarks for RF in the Big Data context. Finally, we experiment five variants on two massive datasets, a simulated one and a real-world dataset. These numerical experiments lead to highlight the relative performance of the different variants, as well as some of their limitations.

This talk is related to the paper: R. Genuer et al. (2017), *Random Forests for Big Data*, Big Data Research 9

*Financial support from ESSEC Research Center and LabEx MME-DII is gratefully acknowledged.*



For any information, please contact  
Stéphanie MARTINEZ  
(01 34 43 37 97 / [martinez@essec.edu](mailto:martinez@essec.edu))

<http://crear.essec.edu/working-group-on-risk>



## Prof. Jean-Michel POGGI

Univ. Paris Descartes and LMO - Univ. Paris Sud

**Jean-Michel Poggi** is Professor of Statistics at Paris-Descartes University and Lab. Maths Orsay University. His research interests are in time series, wavelets, tree-based and resampling methods, applied statistics. Research activities combine theoretical and practical contributions together with industrial applications (mainly environment and energy) and software development.

Jean-Michel is Elected Member of the ISI. He is Associate Editor of the *Journal of Statistical Software*, *CSBIGS* (Case Studies in Business, Industry and Government Statistics) and *Journal de la SFdS*. From 2011 to 2013, he was President of the French Statistical Society (SFdS). He is Vice-President of ECAS, member of the Board of Directors of the ERS of IASC and Council Member of the ISI.



ESSEC  
BUSINESS SCHOOL



Labex MME-DII

Modèles Mathématiques et Économiques de la Dynamique, de l'Incertitude et des Interactions



Banque  
Finance  
Assurance

For any information, please contact  
Stéphanie MARTINEZ

(01 34 43 37 97 / [martinez@essec.edu](mailto:martinez@essec.edu))

<http://crear.essec.edu/working-group-on-risk>