Big Data

Course title – Intitulé du cours	Big data – Part 3
Level / Semester – Niveau /semestre	M2 / second semester
School – Composante	Ecole d'Economie de Toulouse
Teacher – Enseignant responsable	Rémi Perrichon
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Other teacher(s) – Autre(s) enseignant(s)	
Lecture Hours – Volume Horaire CM	12
TA Hours – Volume horaire TD	
TP Hours – Volume horaire TP	
Course Language – Langue du cours	English / Anglais
TA and/or TP Language – Langue des TD et/ou TP	English / Anglais

Teaching staff contacts – Coordonnées de l'équipe pédagogique :

Email : remi.perrichon@enac.fr

Course's Objectives - Objectifs du cours :

Thanks to an increasing computational power and advances in technology, data collected by practitioners and scientists have become more and more complex to analyze. In many fields, relying on a data matrix and applying traditional methods from multivariate statistics is outdated. Noticing this paradigme shift, some authors have called for the use of Object Oriented Data Analysis (OODA) [Wang and Marron, (2007)]. This general framework is useful in the analysis of complex data, that is to say when data are not easily represented as an unconstrained matrix of numbers (tree-structured data, sounds, images …).

Part of OODA, a fashionable area in statistics is Functional Data Analysis (FDA) where the goal is to analyse a sample of curves. A good introduction to this research area can be found in [Ramsay and Silverman, (2002)] and [Ramsay and Silverman, (2005)]. The goal of the course is to provide an introduction to FDA and to work on real data.

First, functional data are introduced as new statistical objects belonging to infinite dimensional spaces. Famous datasets of the literature are presented as well as more complex functional data samples for students to be aware of the vast scope of the subject (functional data with temporal or/and

spatial correlation, multivariate functional data, sparse functional data). Yet, the course focuses on i.i.d., univariate, dense functional data. Traditional steps to represent functional data are detailed. Indeed, as functional data are observed as a set of discrete measured values, the first task is to convert these values to a function (the so-called *smoothing or* interpolation step). An emphasize is then made on *data registration / feature alignment* to understand what is expected from functional descriptive statistics. *Functional Principal Components Analysis* (FPCA) is presented as a technique to highlight substantial modes of variation in the data and but also to deal with the curse of dimensionality. If time allows, one or two statistical problems are developed, for instance functional data clustering.

• Wang, H. and Marron, J.S. (2007). Object oriented data analysis: Sets of trees. *Ann. Statist.*, 35(5):1849-1873.

Course outline :

1. Introduction
2. Smoothing, interpolation
3. Registration
4. Functional Principal Components Analysis (FPCA)
5. Clustering

Prerequisites - Pré requis :

It is assumed that students have a working knowledge of statistics and R.

Grading system - Modalités d'évaluation :

Final project, R (80%)

Practical sessions, R (20%)

Bibliography/references - Bibliographie/références :

• Ramsay, J. O. and Silverman, B. W. (2002) *Applied Functional Data Analysis: Methods and Case Studies*. Springer-Verlag New York Inc.

• Ramsay, J. O. and Silverman, B. W. (2005) *Functional data analysis*. Springer series in statistics. Springer-Verlag New York Inc., 2nd edn.