

# Mascot-Num Workshop

## Handling categorical & continuous data

Sébastien Da Veiga, Snecma

[sebastien.daveiga@snecma.fr](mailto:sebastien.daveiga@snecma.fr)

16/05/2014 IHP

# INTRODUCTION

→ In the GDR community, very often we focus on practical applications with **continuous variables**

- Sensitivity analysis
- Optimization
- Surrogate models

→ However, many applications now involve both **categorical and continuous variables**

- Scenarios (economy, production plans, unpredictable events, ...)
- « Type » (patient, gender, configuration, material, ...)
- ...

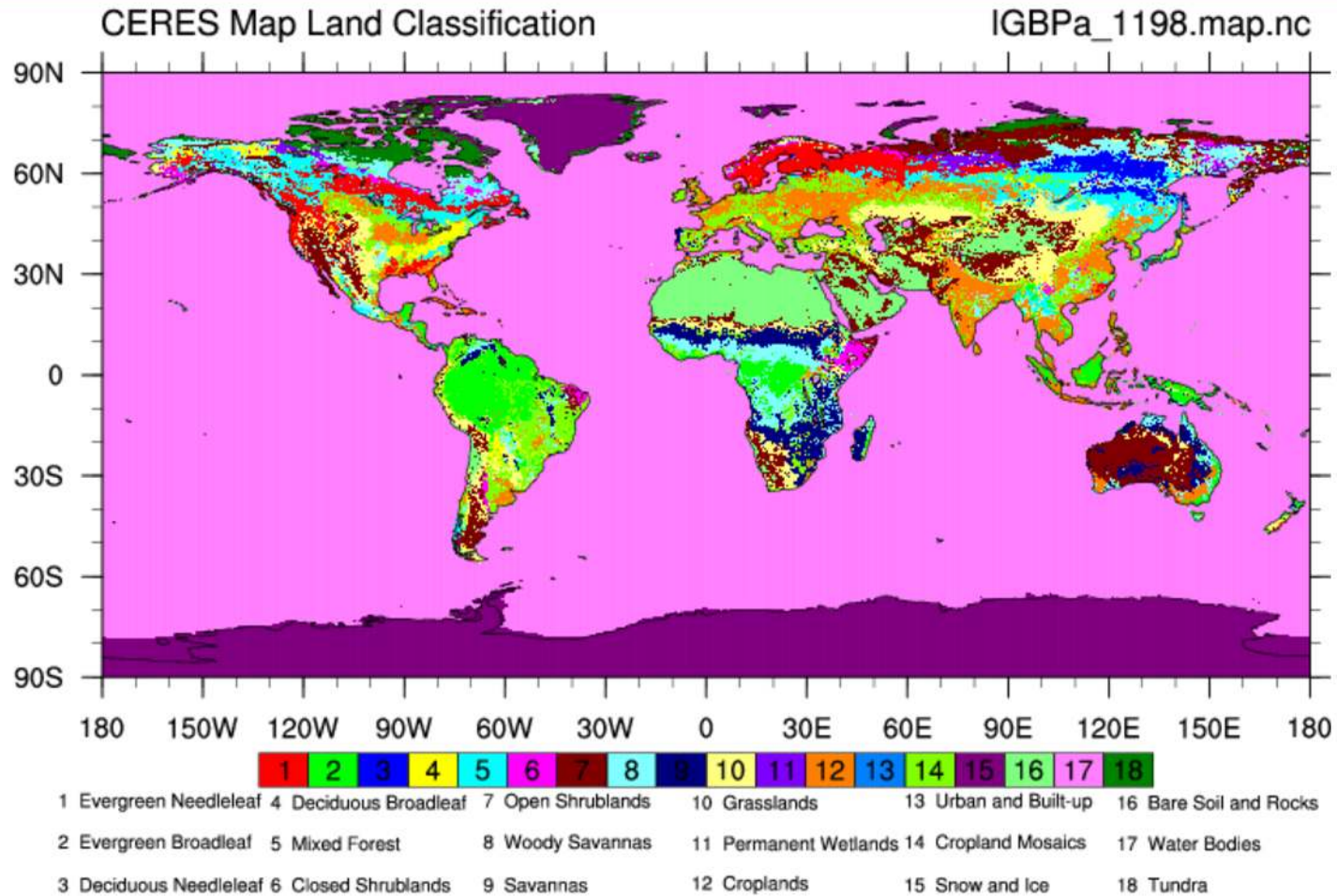
# INTRODUCTION

	Occupational Level				Total
	1	2	3	4	
Scenario A					
White	10	20	30	40	100
Black	20	20	10	0	50
Hispanic	10	10	5	0	25
<b>Total</b>	<b>40</b>	<b>50</b>	<b>45</b>	<b>40</b>	<b>175</b>
Scenario B					
White	10	20	30	40	100
Minority	30	30	15	0	75
<b>Total</b>	<b>40</b>	<b>50</b>	<b>45</b>	<b>40</b>	<b>175</b>

	Occupational Level				Total
	1	2	3	4	
Scenario A					
Male	10	20	30	40	100
Female	40	30	20	10	100
<b>Total</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>200</b>
Scenario B					
Male	10	21	29	40	100
Female	40	29	21	10	100
<b>Total</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>200</b>
Scenario C					
Male	11	20	29	40	100
Female	39	30	21	10	100
<b>Total</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>50</b>	<b>200</b>

Segregation w.r.t. gender, origin, ...  
Reardon 2009

# INTRODUCTION



Classification according to vegetation  
NCL examples

# INTRODUCTION

→ **Dedicated tools exist to handle such variables, but are usually goal-oriented**

- Clustering
- Optimization
- Regression

→ **The goal of the workshop is to give an overview of the available methods**

- Three « sessions » on each of these topics
- Focus on software possibilities
- Exchange between the communities Clustering/Optimization/Regression

# PROGRAM

9:45 – 11:00 : Clustering I – **Marie Chavent** (Université Bordeaux 2 & INRIA)

## ***Variable Clustering and Mixed Data. The ClustOfVar package***

11:00 – 11:15 : Coffee Break

11:15 – 12:30 : Clustering II – **Matthieu Marbac** (Université Lille 1 & INRIA)

## ***Model-based clustering of Gaussian copulas for mixed data***

12:30-14:00: Lunch break

14:00 – 14:45 Optimization I – **Claudia D'Ambrosio & Luca Mencarelli** (LIX Ecole Polytechnique)

## ***Introduction to Mixed Integer Nonlinear Programming***

14:45 – 15:30: Optimization II – **Claire Lizon** (IFPEN & LIX Ecole Polytechnique)

## ***Mixed Integer Nonlinear Optimization for Well Placement***

15:30 – 15:45: Coffee Break

15:45 - 17:00: Regression – **Claire Cannamela** (CEA)

## ***Metamodels involving both categorical and continuous inputs variables - application to a spectrometry gamma problem***