Data mining and real systems modeling

J. Chaskalovic

Institut Jean le Rond d'Alembert
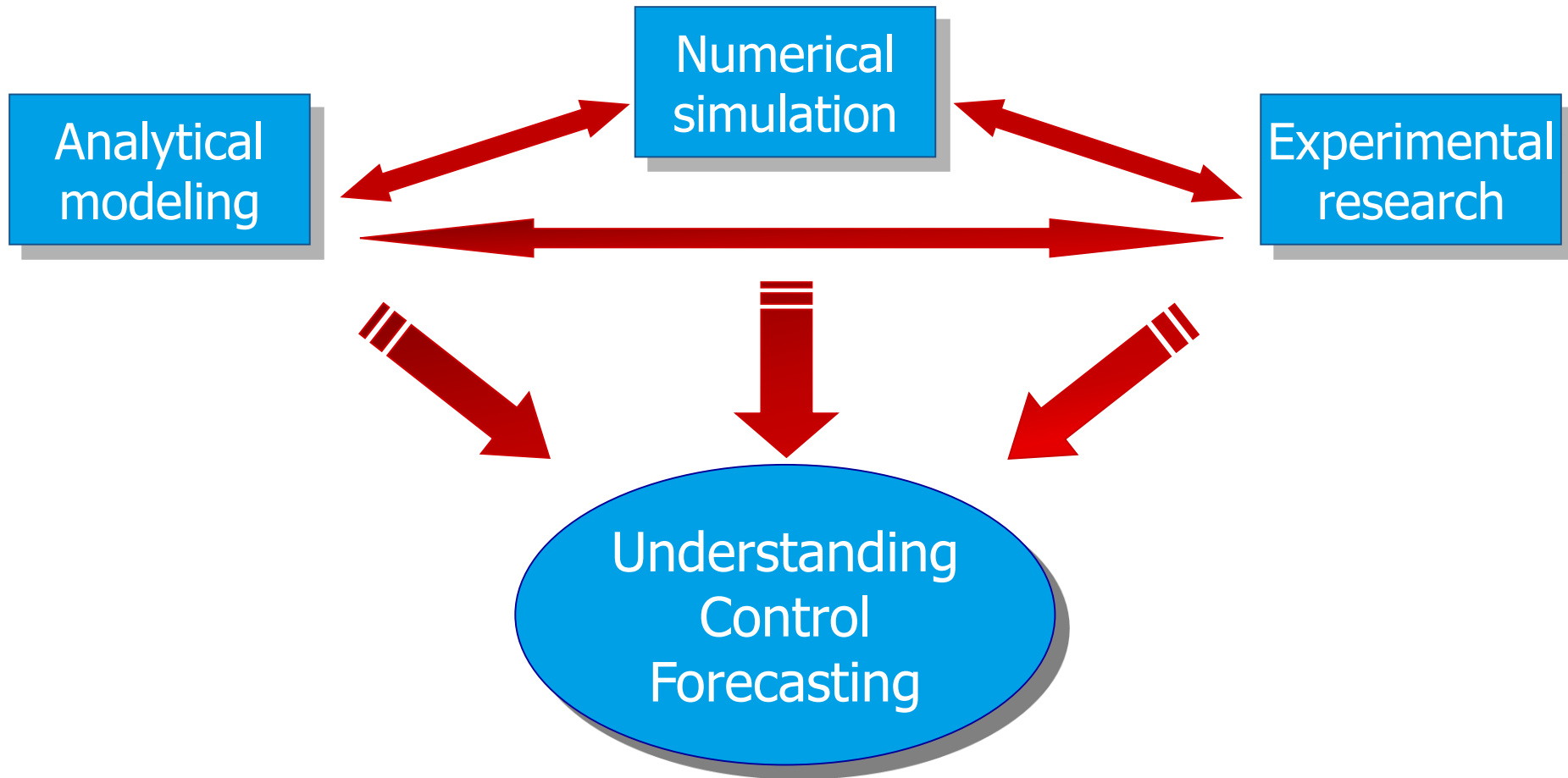*University Pierre and Marie Curie*

Agenda

- ✓ Modeling and real systems

- ✓ What's Data Mining
  - ○ *Why Mine Data? Scientific Viewpoint*
  - ○ *Why Mine Data? Commercial Viewpoint*

- ✓ Live Data Mining: a medical application

Modeling
&
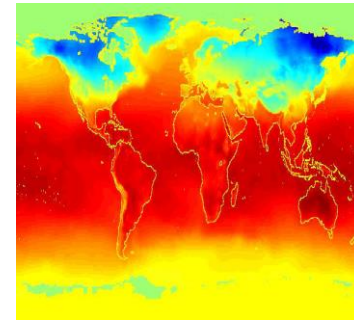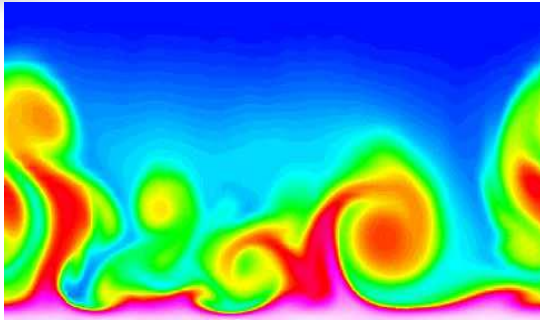# real systems

Real systems modeling

Too much independant !

What's Data Mining ?

✓ **Data mining is the knowledge discovery process in databases:**

- Efficient techniques which lead to the identification of interesting information or patterns from data usually in large databases.

Why Mine Data? Scientific Viewpoint

✓ Data collected and stored at enormous speeds (GB/hour)

- Remote sensors on a satellite.
- Telescopes scanning the skies.
- Scientific simulations generating terabytes of data.

✓ Traditional techniques infeasible for raw data

✓ Data mining may help scientists

- In classifying and segmenting data.
- In assumptions modelling (*analytical modeling* and *experiments*).

Why Mine Data? Commercial Viewpoint

✓ Lots of data is being collected and warehoused

- Web data, e-commerce.
- Purchases at supermarkets or big department stores.
- Bank / Credit Card.

✓ Computers have become cheaper and more powerful

✓ Competitive Pressure is Strong

- Provide better, customized edge services (e.g. in Customer Relationship Management).

It's coming from…

Customer Relationship Management

Interactivity devoted to develop Clients Relationship

✓ Optimization consumers' relationships during the customer life cycle in several Marketing topics :

- ❑ Recruitment (acquiring customers)
- ❑ Loyalty (retaining good customers)
- ❑ Sales (increasing the value of customers)
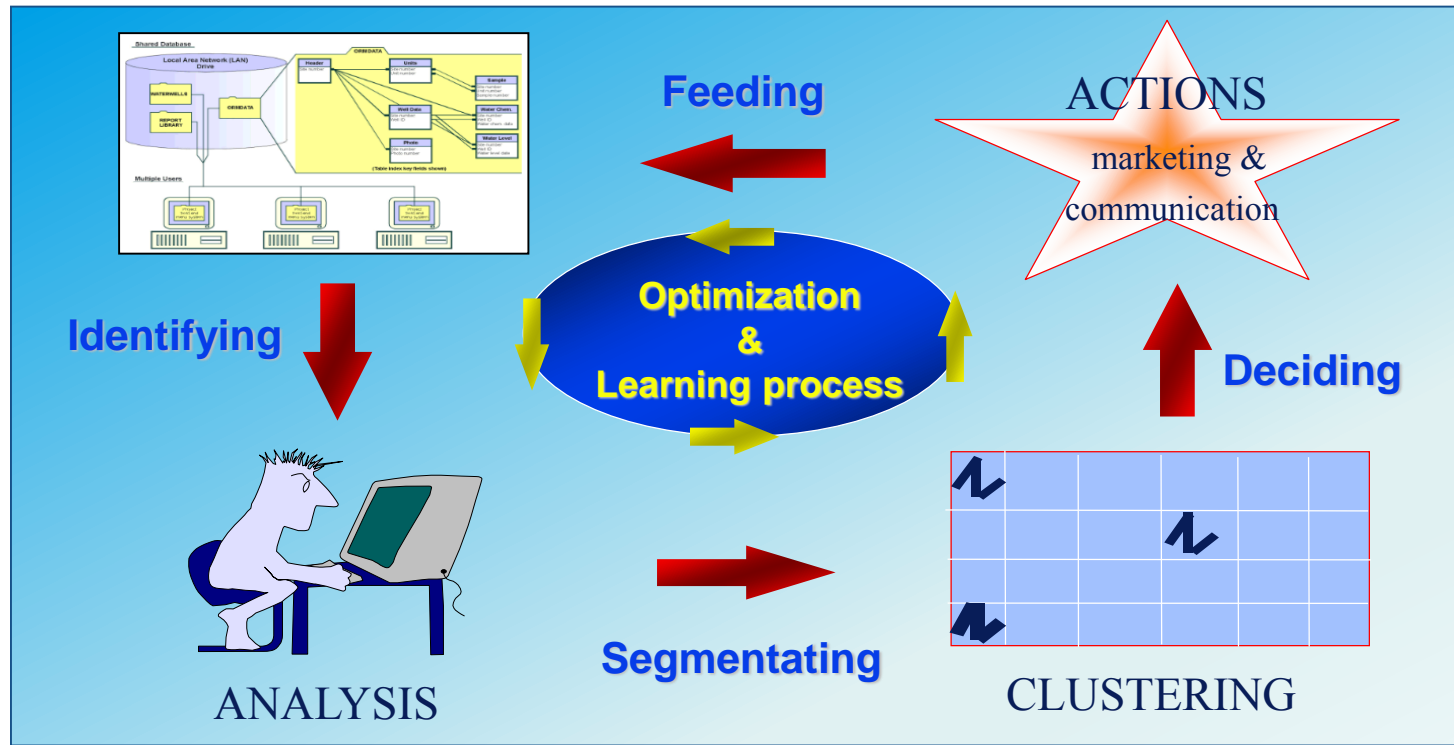
✓ Better knowledge of Clients and Prospects :

Which product
For which client
At which moment
From which media
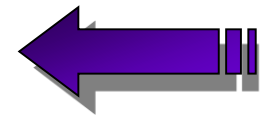
Process of CRM & Data Mining

Step 1 : The Brief

- **Step 2 : Collect and Data Management**
- **Step 3 : Data Analysis and Modeling**
- **Step 4 : Segmentation of the Database**
- **Step 5 : Optimization & Data campaign**

Data Mining Principles

✓ <u>*Supervised Data Mining*</u> : One or more *target* variables must be explained in terms of a set of *predictor* variables.

$$y = \varphi(x_1, x_2, \ldots, x_n)$$

- Segmentation by Decision Tree, Neural Networks, etc.

✓ <u>*Non supervised Data Mining*</u> : No variable to be explained, all available variables are considered to create groups of individuals with homogeneous behavior.

$$\Psi(x_1, x_2, \ldots, x_n) = 0$$

- Typology by Kohonen's cards, Clustering, etc.

Data Mining and not Data Analysis

The Data Mining is a discovery process

✓ Data Scan : inventory of potential and explicative variables.

✓ Data Management : collection, arrangement and presentation of the Data in the right way for mining.

✓ Data Analysis :

- ❑ Learning
- ❑ Modeling
- ❑ Forecasting

Live Data Mining

Medical application

Thank you very much !