Modelling, Forecasting for Renewable Energy Production and Statistical Inference

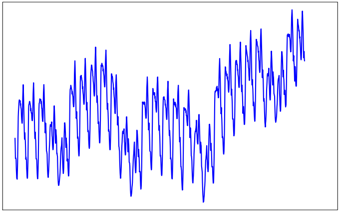
.

Statistical and machine learning methods to model and forecast Energy

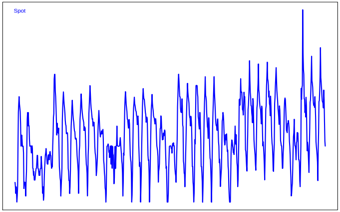
Mathilde Mougeot

ENSIIE & LPSM

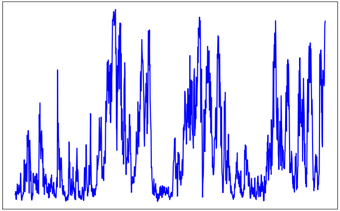
Thematic semester on Statistics for Energy markets
Dourdan, June 2018



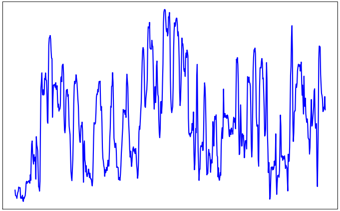
French electrical consumption



Energy Spot prices



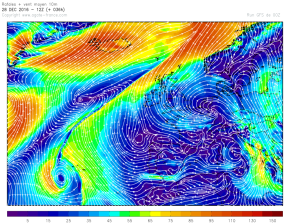
Wind turbine power



Industrial equipment

Physical Modelling vs Statistical Modelling

- Physical Modelling mosly based on physical equations :
  - Simple Direct equation
    $PV = nRT$
  - Partial differential equations
    $\rightarrow$ Need of a numerical model to study the dynamic and the evolution of a model. Ex : Navier-Stokes equation



$\rightarrow$ Need to well understand the underlaying phenomena

Statistical Modelling

- In a supervized setting (Y, X) :

  $Y \in \mathbb{R}$, $X \in \mathbb{R}^p$ $\quad \boxed{Y = \mathcal{M}(X) + \epsilon}$

  - Parametric models.

    $\mathcal{M}_1(X) : Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \epsilon$

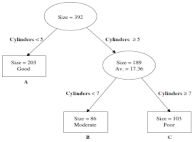
    $\mathcal{M}_2(X) : Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + ... + \beta_p X_p + \epsilon$

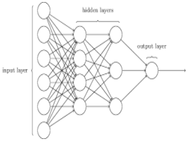
    $\rightarrow$ Probabilistic or not, depending on the law assumption on ϵ.

    Data Set : n observations : (y_i, x_i)

    $\rightarrow$ to estimate (to compute) the parameters $\hat{\beta}$ of the model

  - Non Parametric models No analytical (explicit) expression



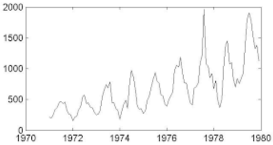


$\rightarrow$ Models : Kernels, decision trees, neural networks(black boxes),...

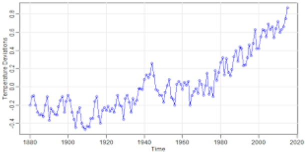
Crucial question : How to find the correct model ?

Time Series Modelling

- Time dependency and Patterns



- Without (iid ?) :



$$\boxed{Y_t = \mathcal{M}(X_t) + \epsilon_t}$$

How to find the best model ?

- Bunch of p unitary, calibrated models
  $\mathcal{M}_1, \dots, \mathcal{M}_p$: Parametric, Non Parametric models...
  $\rightarrow \hat{\mathcal{M}}_1, \dots, \hat{\mathcal{M}}_p$: estimated model using Train data $\{(y_i, x_i) 1 \leq i \leq n\}$.

- Selection vs Aggregation

  1. Model selection
     $\mathcal{M}_{opt} = \arg\min_{k \in p} E_{\mathcal{M}_k}$, $E_{\mathcal{M}_k}$ is a cost function to evaluate $\mathcal{M}_k$
     $\hat{\mathcal{M}}_{opt} = \arg\min_{k \in p} E_{\hat{\mathcal{M}}_k}$
  2. Aggregation
     $\hat{\mathcal{M}}_{opt}(X) = \sum_{k=1}^{p} \hat{w}_k \hat{\mathcal{M}}_k(X)$ Exponential weight aggregation

     $\hat{\mathcal{M}}_{opt}(X) = \sum_{i=1}^{n} \hat{w}_{n,i}^{\hat{\mathcal{M}}_{1..p}(X_i)}(X) Y_i^{Train}$ Consensuel aggregation

Electricity Framework

Production	Consumption
Nuclear power plants	industrial plants
coal fired power plants	home and heating
wind farms	building heating
photovoltaic farms	...

→ Electricity can hardly be stored. There is a need to :
Balance between electrical production and consumption
Forecast consumption and production

Statistical & ML models for Energy in several directions.

1. Consumption
   - High dimensional regression models to Forecast the French National Consumption with D. Picard, K. Tribouley, V. Lefieux, JRSSB, AADA.
   - Energy Savings. Over consumption monitoring and diagnostic for industrial equipments with O. Cadet, C. Harper.

2. Production
   - Modeling the Wind Farm production using machine learning tools with A. Fischer, L. Montuelle, D. Picard, Wind Energy 2017.
   - From Numerical Weather Prediction outputs to accurate local surface Wind speed : statistical modelling and forecasts with B. Alonzo, R. Plougonven, A. Fischer, A. Dupre, & P. Drobinski.

3. Meteorological inputs

   Homogeneous climate regions using learning algorithms.

   with D. Picard, V. Lefieux, M. Marchand.

Outline of the lessons

1. Functional models to forecast electrical consumption.
   Statistical models.
   Exponential weight Aggregation

2. Machine learning models to model wind turbine production.
   Machine learning models.
   Consensuel Aggregation.

3. Segmentation of multidimentional functional data.
   Unsupervized learning.
   Spectral clustering.
   Aggregation of Similar clustering instances.

The never ending story...to answer to real data problems