Statistics applied to energy sector

Thi Thu Huong Hoang
Electricity of France (EDF) R&D

02/06/2023

Outline

1. Context of supply-demand balance

1. Streamflow simulation for mid-term management

2. Demand forecast for short-term management

1

Context of supply-demand balance

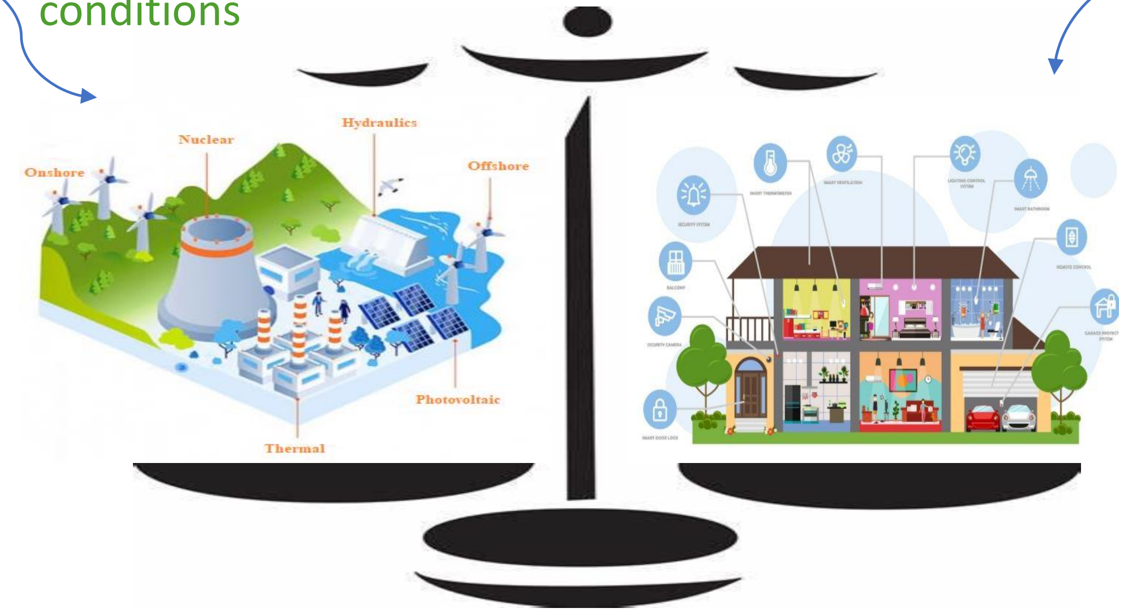
Supply – demand balance

Since electricity cannot be stored on a large scale, it must be consumed as soon as it is produced. The proper functioning of the electricity system is therefore based on the constant and real-time balance between production and consumption.

Weather conditions

Crisis (health, economic, energy)

Unavaibility of power plants



customer behaviors

Portfolio Management Process

Long Term

Investment decisions.

Marginal cost and price estimation

.

Several Years

Nuclear outages scheduling.

Quantification of failure risks

Yearly

Water & stock value determination.

Market hedging.

Weekly

Trials settlement.

Thermal outages scheduling.

Market hedging.

Daily & Intraday

Demand forecast and

generation dispatch plan

GOAL

To ensure minimised cost for generation-demand balancing under physical and financial constraints

SHORT TERM

MID TERM

LONG TERM

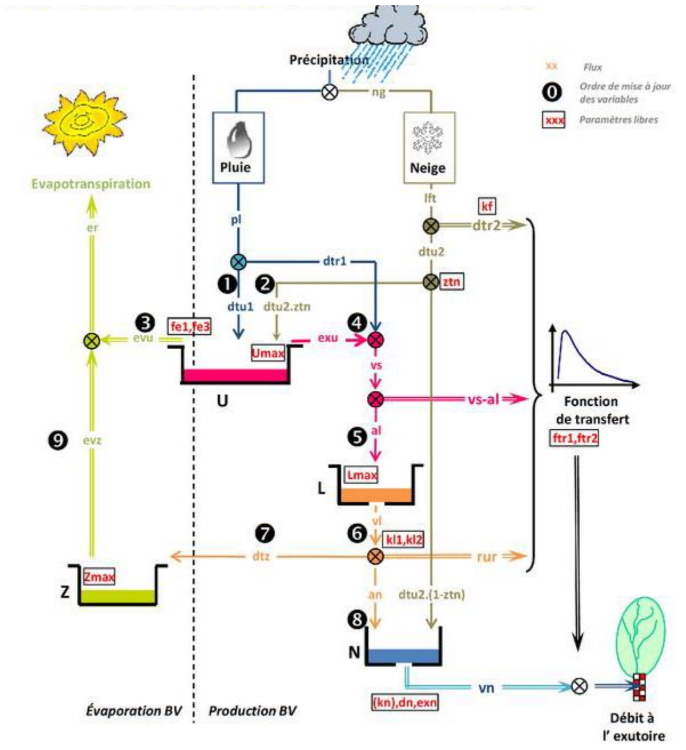
2

Streamflow simulation

For mid-term management

Context

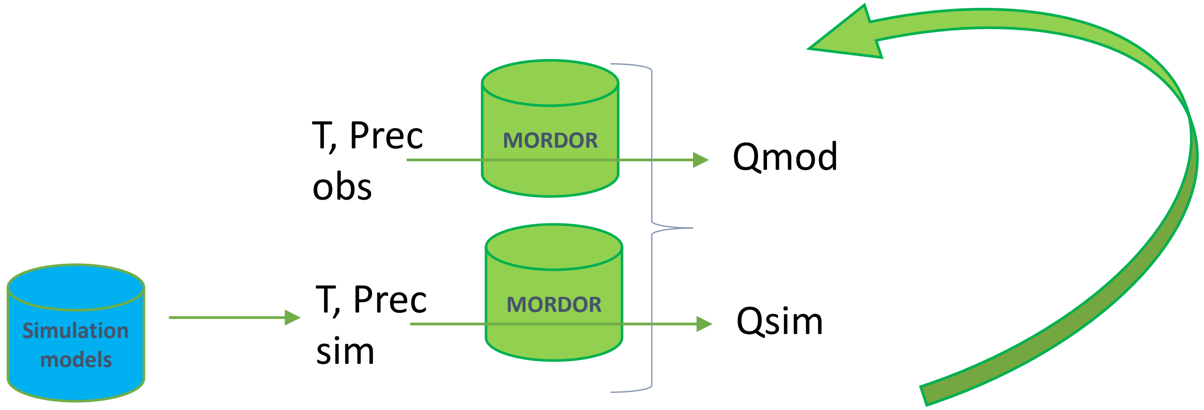
Objectif: Streamflow estimation to assess the hydraulic production



MORDOR Hydrological model
(see Garavaglia & Le-Lay 2017)

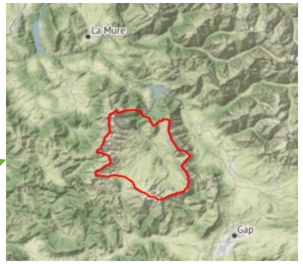
- Input: temperature, precipitation
- Output: streamflow

IDEA: Simulate temperature and precipitation in order to produce a large sample for streamflow covering different possibilities

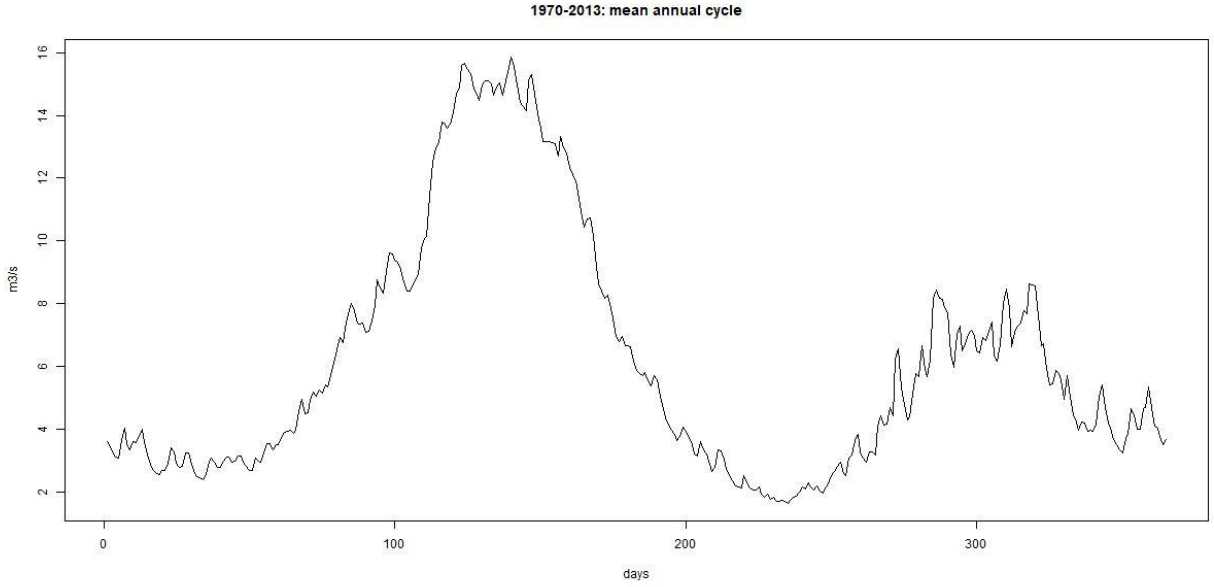


Used data

- Variables: Temperature, precipitation, streamflow
- Location: Souloise-Infernet watershed in French Alpes
- Period: 1970-2013
- Characteristics: surface 214km², latitude [850 – 2700] m
- Annual cycle streamflow: high in spring, low in winter and late summer



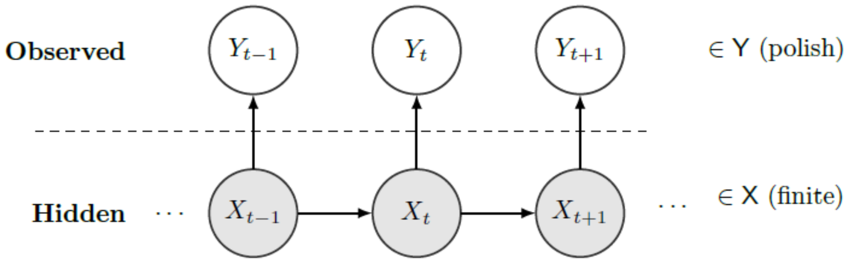
1970-2013: annual cycle streamflow



Souloise-Infernet watershed

Bivariate stochastic weather generator

- **Bivariate stochastic weather generator: temperature and precipitation**
  - *From the thesis of Augustin Touron 2019 (co-supervised by Elisabeth Gassiat)*
  - *Convergence of MLE for seasonal HMM with trend is proven in Touron 2019*
  - *Dynamic of Hidden Markov Model*



The dynamic of a hidden Markov model.

  - *Conditional distribution of Y(t) given X(t) = k $\in$ {1, 2, …, K}*

$$\nu_{k,t} = \underbrace{\sum_{m=1}^{M_1} \underbrace{p_{km}\delta_0}_{\text{Precipitation}} \otimes \underbrace{\mathcal{N}\left(T_k(t)+S_k(t)+\mu_{km},\sigma^2_{km}\right)}_{\text{Temperature}}}_{\text{Dry days}} + \underbrace{\sum_{m=M_1+1}^{M} \underbrace{p_{km}\mathcal{E}\left(\frac{\lambda_{km}}{1+\sigma_k(t)}\right)}_{\text{Precipitation}} \otimes \underbrace{\mathcal{N}\left(T_k(t)+S_k(t)+\mu_{km},\sigma^2_{km}\right)}_{\text{Temperature}}}_{\text{Wet days}}$$

Bivariate stochastic weather generator

Hyper-parameters

The model requires to specify several hyper-parameters:

- K the number of hidden states
- d the degree of the trigonometric polynomials, which sets the complexity of the seasonality,
- M and M_1 which correspond to the complexity of the emission distributions.

By BIC criterion, K = 7
By experience on univariate models, we select d = 2, M = 4 and M_1 = 2.

Estimation method

EM (Expectation – Maximization) algorithm: Find the MLE of the marginal likelihood by **iteratively** applying these two steps

Let X = (X_1, …, X_n) and Y = (Y_1, …, Y_n) and recall that X is not observed. The likelihood function with initial distribution π

$$L_{n,\pi}[\theta;Y] = \sum_{\mathbf{x}\in\mathsf{X}^T} \pi_{x_1} f^{\theta_Y}_{x_1,1}(Y_1) \prod_{t=2}^{n} Q_{x_{t-1}x_t}(t-1) f^{\theta_Y}_{x_t,t}(Y_t),$$

- **Expectation step**: Define $Q\left(\theta|\theta^{(q)}\right)$ as the expected value of the log likelihood function of θ, with respect to the current conditional distribution of Y and the current estimates of the parameters $\theta^{(q)}$

$$\mathbf{Q}\left[(\theta,\pi),\left(\theta^{(q)},\pi^{(q)}\right)\right] := \mathbb{E}^{\pi^{(q)},\theta^{(q)}}\left[\log L_{n,\pi}\left(\theta;(X,Y)\right) \mid Y\right]$$

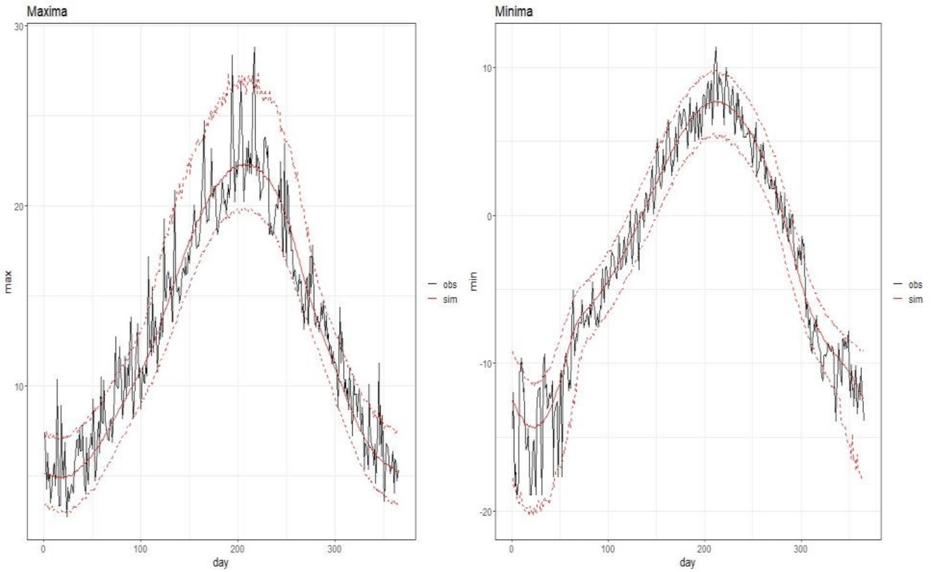
- **Maximization step**: Find the parameters that maximize this quantity:

$$\left(\theta^{(q+1)},\pi^{(q+1)}\right) = \arg\max Q\left(\theta^{(q)},\pi^{(q)}\right)$$

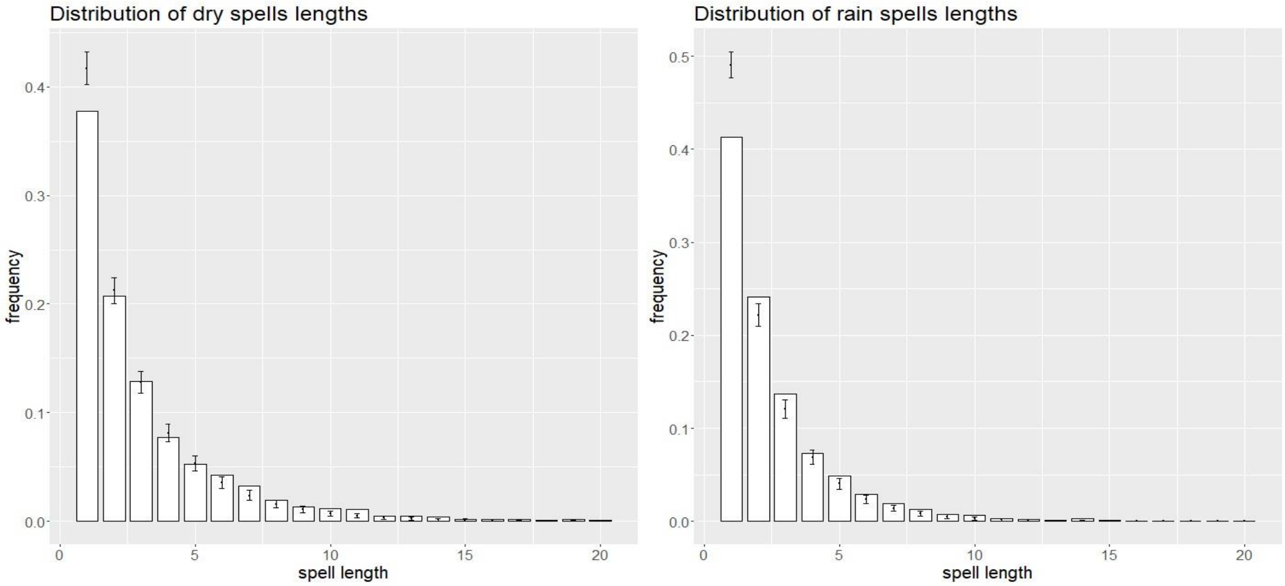
***Remark**: EM depends strongly on initial parameters, we thus execute a big number of EM with different initial parameters and we retain which one with highest log likelihood*

Quality of 1000 scenarios of 1970-2013 time series of temperature and precipitation

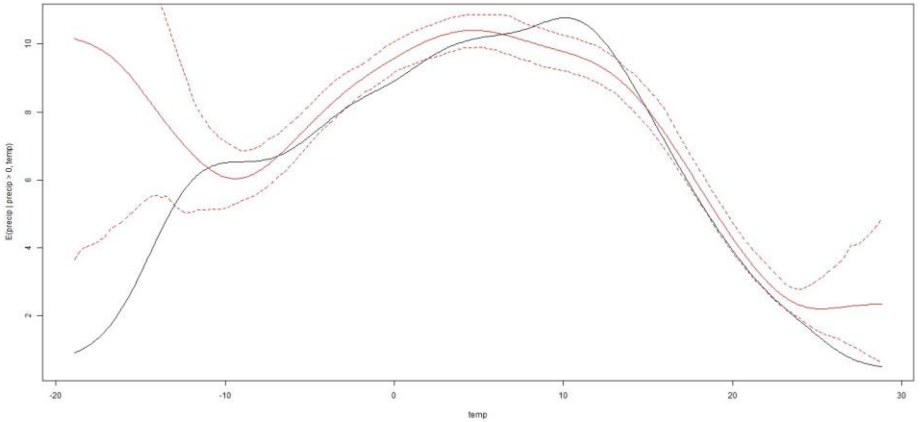
Temperature minimum and maximum



Precipitation dry and wet spells



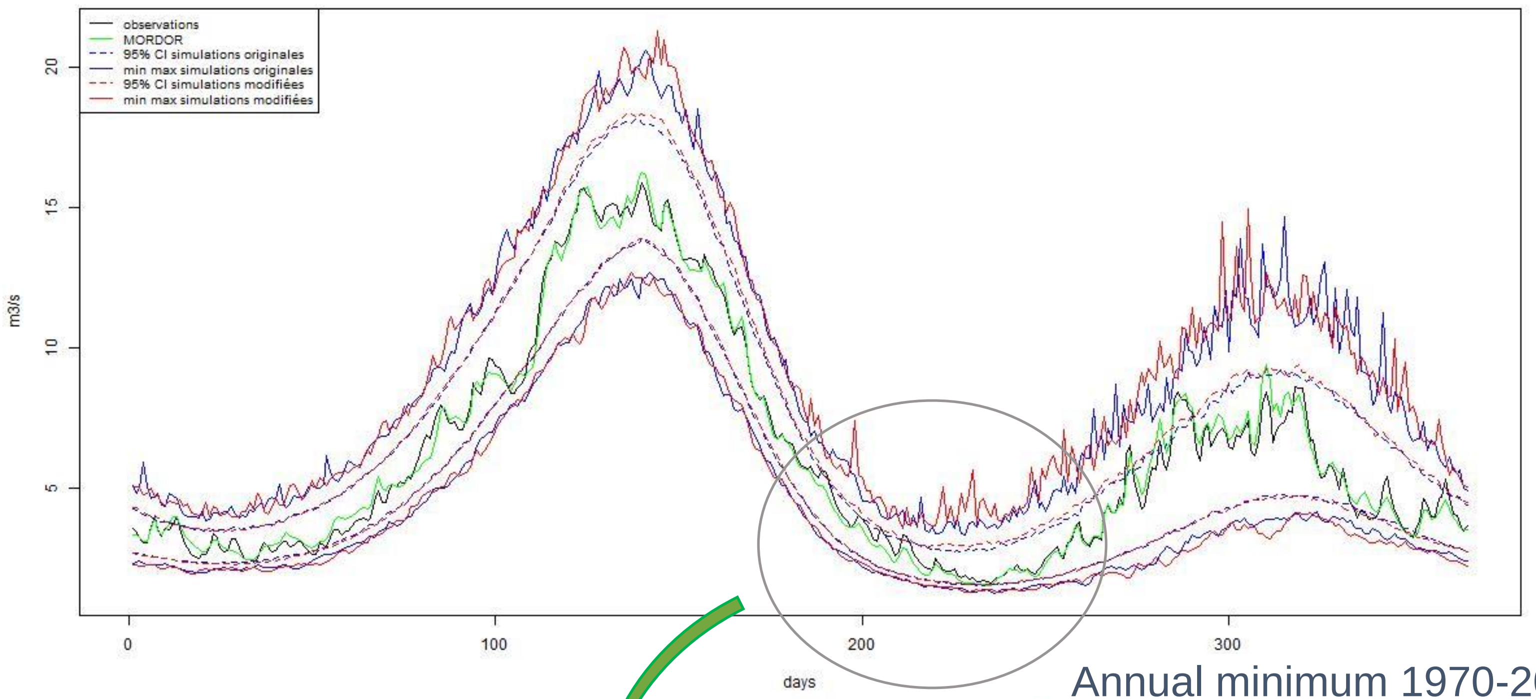
Conditional expectation Precip | Temp



Default of the model in the representation of rainy spells

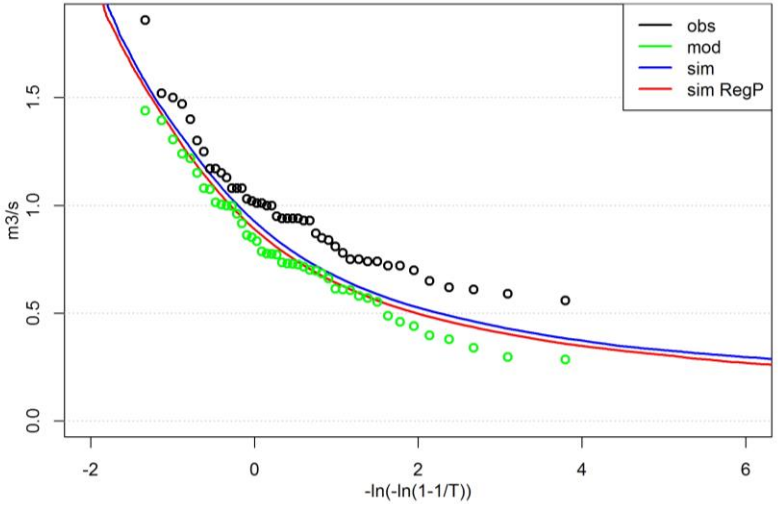
Validation of simulated streamflows

1970-2013: mean annual regime



Zoom on low flows (annual minimum)

Annual minimum 1970-2013

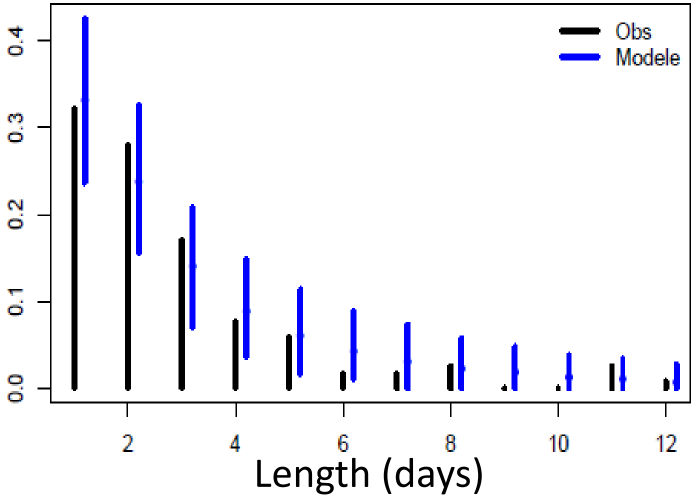


Good representation of simulated streamflows for annual minima

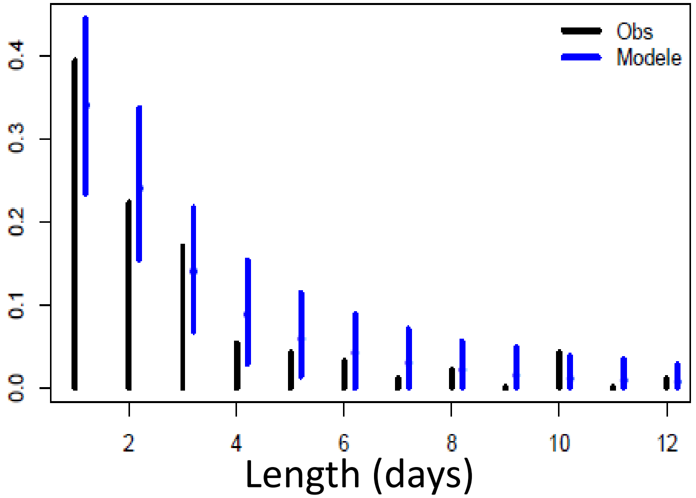
Validation of simulated streamflows

Low flows sequences

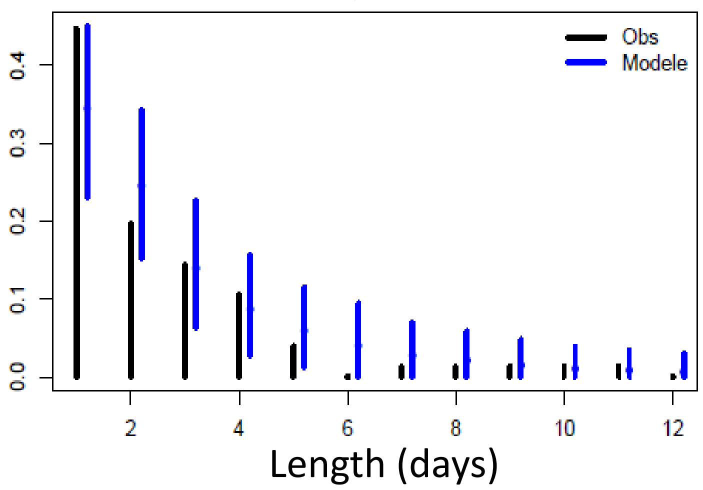
Q < 9th percentile



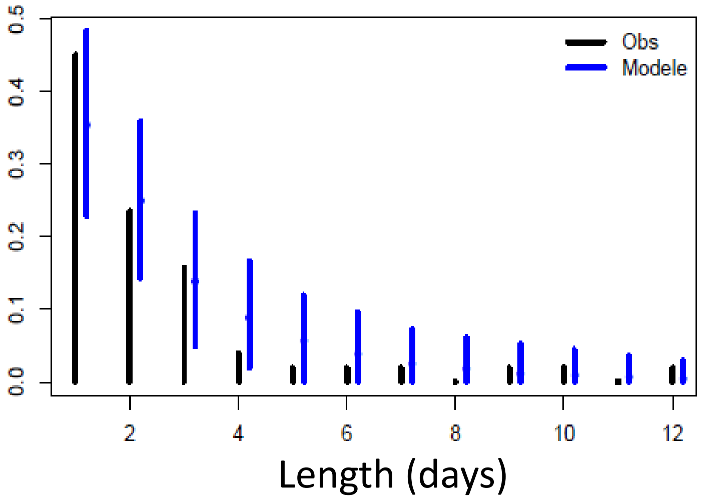
Q < 7th percentile



Q < 5th percentile



Q < 3rd percentile



Good representation of simulated streamflows for low flows sequences

3

Demand forecast

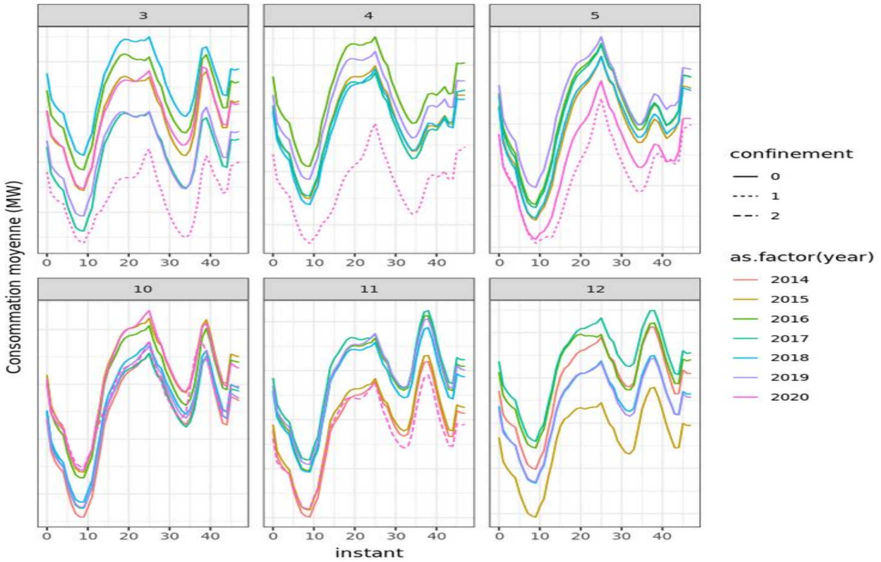
for short-term management

Context

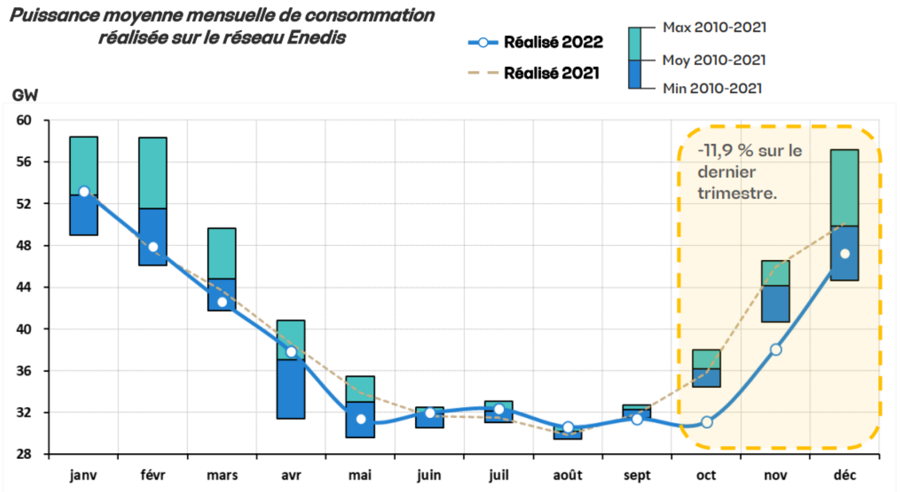
- At EDF, forecast of electricity demand over **several horizons**:
  - **Long term** : clarification for investment choices
  - **Mid term**: quantification of failure risks, storage management (nuclear, hydraulic), purchases
  - **Short term** :
    - ✓ ***Weekly*** : load shedding, storage management, purchases
    - ✓ ***Daily/Intraday***: power generation planning

- **Challenges for short-term forecast**:
  - **Evolving context**: new uses, health crisis, energy crisis, sobriety, etc.
  - Delay of availability of data
  - Weather forecast

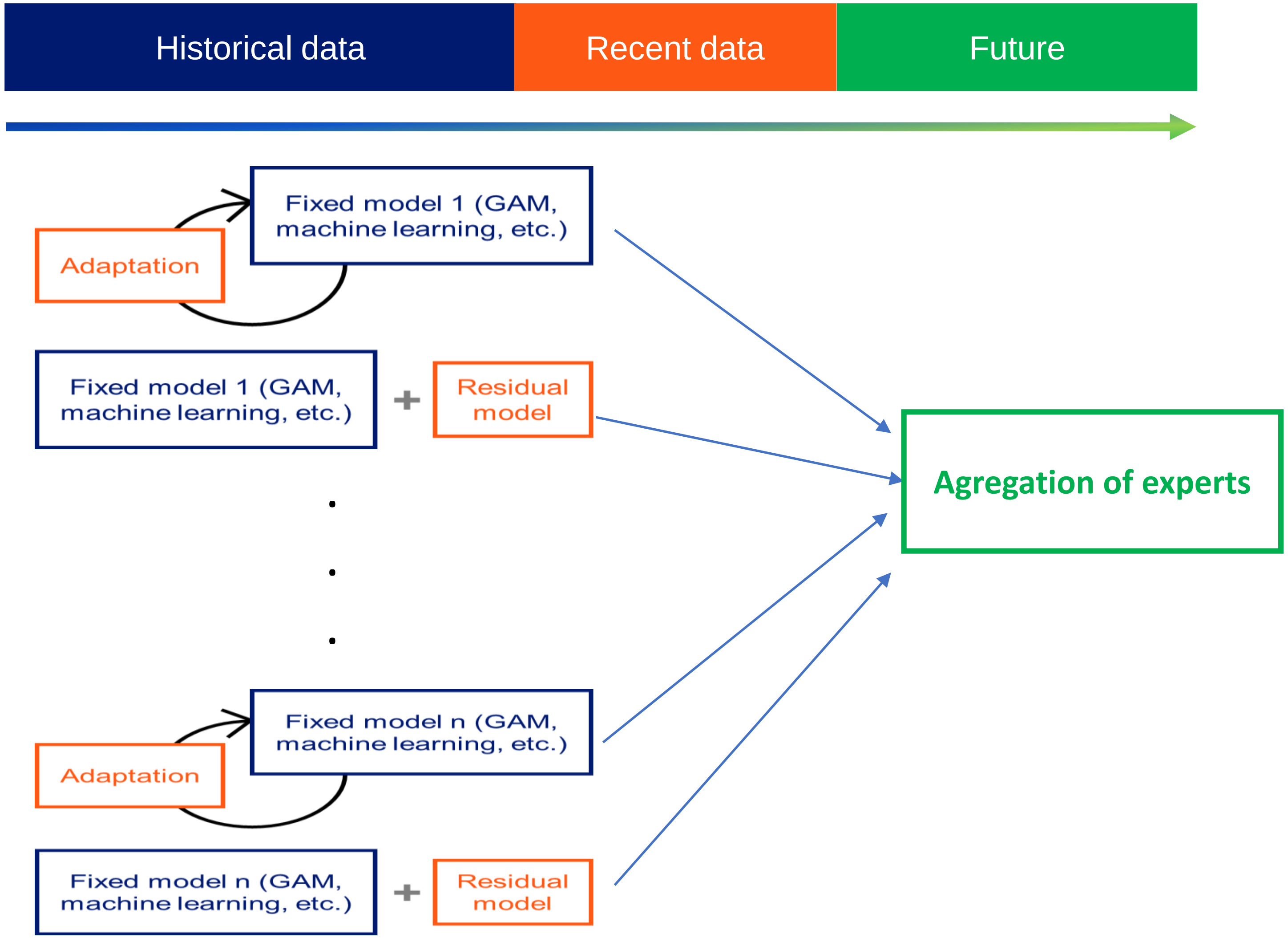
Impact of COVID



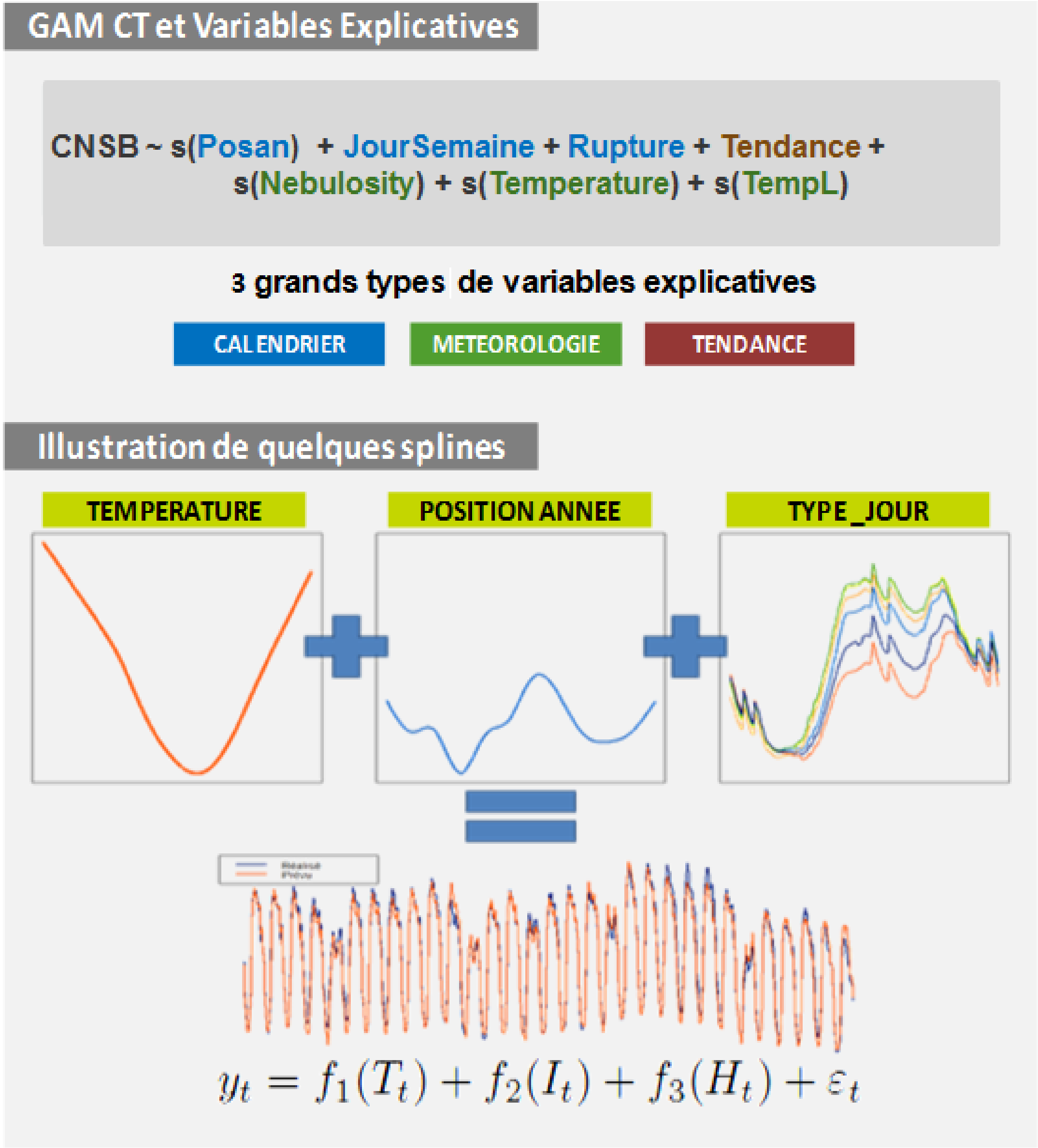
Impact of sobriety **(source Bilan Electrique – Enedis)**



Consumption forecast methodology



Load forecast modelling



- Generalized Additive Model (GAM)

$$y_i = X_i\beta + f_1(x_{1,i}) + f_2(x_{2,i}) + f_3(x_{3,i}, x_{4,i}) + \ldots + \varepsilon_i$$

- Parameters estimated by maximizing penalized log-likelihood

$$min_{\beta,f_j} ||y - X\beta - f_1(x_1) - f_2(x_2) + \ldots||^2 + \lambda_1 \int f_1''(x)^2 dx + \lambda_2 \int f_2''(x)^2 dx + \ldots$$

- With GAM, we separate explicative variables in 3 types: non-thermosensitive or calendar effects, thermosensitive effects and trend

- Performance:
  - before COVID, less than 2% of MAPE
  - with last COVID crisis and sobriety, about 3% of MAPE

Kalman filter

Goal : Using of Kalman filter (thesis of Joseph Moullart de Vilmarest) to ajust GAM's effects:

- Start with a GAM model $y_t = \beta_0 + \sum_{j=1}^{d} f_j(x_{t,j}) + \varepsilon_t$
- Define $f(x_t) = \left(1, \overline{f_1}(x_{t,1}), \dots, \overline{f_d}(x_{t,d})\right)^T$

  - y_t , x_t : domestic demand and explicative variable at t
  - f_j : linear and non linear effects
  - $\overline{f_j}$: standardized effects

- We seek to estimate the coefficients θ_t in adaptative way so that : $E[y_t|x_t] = \theta_t^T f(x_t)$

Principal

$$\begin{cases} y_t = \theta_t^T f(x_t) + \varepsilon_t \ \ avec \ \ \varepsilon_t \sim N(0, \sigma^2) & \text{Space equation} \\ \theta_{t+1} = \theta_t + \eta_t \ \ avec \ \ \eta_t \sim N(0, Q) & \text{State equation} \end{cases}$$

wher Q is diagonal covariance matrix.

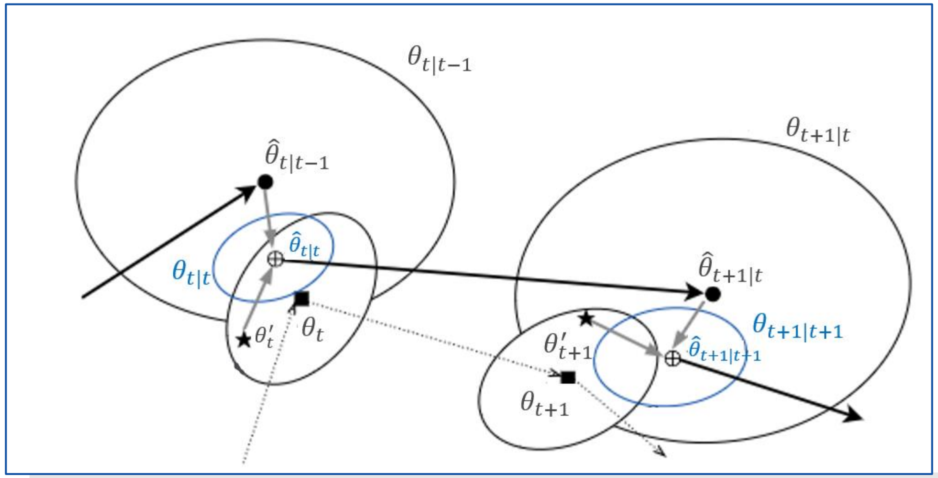


Illustration of fonct of Kalman filter

- $\boldsymbol{\theta_{t+1}}$ real value of coefficients at *t+1*
- $\boldsymbol{\hat{\theta}_{t+1|t}}$ estimation at *t* of θ_{t+1} obtained by state equation
- $\boldsymbol{\theta'_{t+1}}$ estimation at t+1 of θ_{t+1} obtained indirectly by the knowledge of y_t (space equation)
- $\boldsymbol{\hat{\theta}_{t+1|t+1}}$ estimation at t+1 of θ_{t+1} by compromise between θ'_{t+1} and $\hat{\theta}_{t+1|\,t}$

Algorithme

Algorithm 1: Kalman Filter

Initialization: the prior $\theta_1 \sim \mathcal{N}(\hat{\theta}_1, P_1)$ where $P_1 \in \mathbb{R}^{d \times d}$ is positive definite and $\hat{\theta}_1 \in \mathbb{R}^d$.

Recursion: at each time step $t = 1, 2, \dots$

1) Prediction:

$$\mathbb{E}\left[y_t \mid (x_s, y_s)_{s<t}, x_t\right] = \hat{\theta}_t^\top f(x_t),$$
$$Var\left[y_t \mid (x_s, y_s)_{s<t}, x_t\right] = \sigma^2 + f(x_t)^\top P_t f(x_t).$$

2) Estimation:

$$\hat{\theta}_{t+1} = \hat{\theta}_t + \frac{P_t f(x_t)}{f(x_t)^\top P_t f(x_t) + \sigma^2}\left(y_t - \hat{\theta}_t^\top f(x_t)\right),$$
$$P_{t+1} = P_t - \frac{P_t f(x_t) f(x_t)^\top P_t}{f(x_t)^\top P_t f(x_t) + \sigma^2} + Q.$$

Different approaches

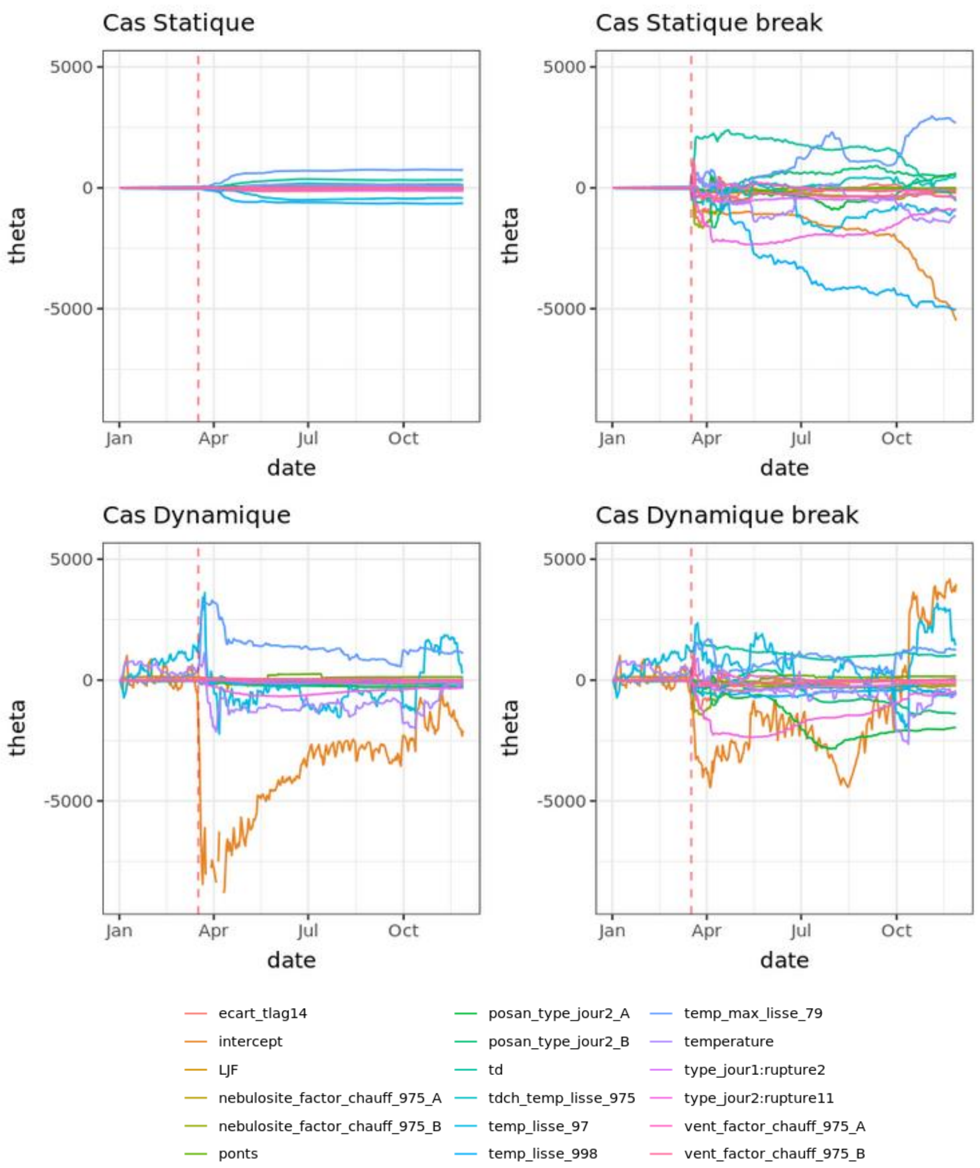
Different cases in the fonction of the matrix Q:

- **Static case**: $Q = 0, \quad \sigma^2 = 1, \quad \theta_1 = 1, \quad P_1 = diag(d)$ where d is the number of effects **=> the coefficients θ vary slightly.**

- **Dynamic case :**
  - σ^2, θ_1 initialized by an analytical formula (maximum likelihood)
  - Q initialized by grid search on a stable period (for ex. 01/09/2014 - 31/08/2019). We seek for each effect, a value of $Q^* = \frac{Q}{\sigma^2}$. **We guarantee then $\mathbf{Q < \sigma^2}$ to avoid the coefficients varying too much with the demand.**

- **Break :** In order to take into account the break linked to the first lockdown, we add a « break » in modelling => We take $Q = \sigma^2 \times Diag(d)$ then $Q^* = Diag(d)$ only the day before of the first lockdown, then we take previous value of Q^* .

Evolution of coefficients θ for each case



Conclusion

- Modelling in electricity field is both rich and complex. We live in a changing world then we have to improve constantly our methods and our models. Combining theory and practice is constantly required.

- Machine learning and deep learning are emerging but the capacity to interpret the models is important for operational purposes.

- Data is the nerve of the war. Data in electricity field is difficult to measure exactly. Data processing is then needed to obtain clean data. Furthermore, the data availability delay reduces the performance of adaptative methods.

Thank you