



STATISTICS FOR ENERGY MANGEMENT R&D PROJECT PRESENTATION

Mai 2014

Yannig Goude



SOMMAIRE

- 1. STATISTICS AT EDF R&D**
STATE OF THE ART
PERSPECTIVES
- 2. CONTRIBUTIONS OF THE PROJECT**
FORECASTING
SIMULATION

INDUSTRIAL CHALLENGES

- **Smart grids**

- More and more « real time » data (ex: linky, 1million meter in 2016)
- Demand response (new tariffs, real time pricing...)
- New communication tools with customers (webservice....)

- **Renewables energy development**

- A more and more probabilistic context

- **Opening of the electricity market:**

- Losses/gains of customers

- **Sensors data:**

- Production/consumption sites
- Smart home, internet of things

- **New usages/tariffs:**

- Electric cars
- Heat pumps, smart phones, battery charge, computers, flat screens....
- Demand response

STATISCAL CHALLENGES

■ Big data

- Parallelizing statistical algorithms
- Complex data analysing: heterogenous spatial/temporal sampling, different sources/nature of data
- Sequential data treatment (data flow, CEP)
- Cohérence de données hétérogènes (ex: INSEE, météo, linky)
- Functional data analysis

■ Adaptivity

- Non-parametric models
- Model selection, data driven penalty...

■ Sequential estimation

- Break detection, on-line update
- Aggregation with on-line weighs

■ Spatio-temporal

- Spatial correlation modeling/simulation

■ Multi-scale models

- Multi-horizon models
- Multi level data on the grid

STATISCAL CHALLENGES

- **Data mining for time series**
 - Machine learning for time series
- **Probabilistic forecasts**
 - Density forecast
 - Quantile modelization
 - Functional data analysis
- **Large scale simulations**
 - Simulation platform, parallel processing
 - Complex systems dynamics

CONTRIBUTIONS

- **Big data**

- GAM parallel processing
- With dpt ICAME: connection with HADOOP, teradata data-lab...
- IBM simulation platform

- **Adaptivity**

- GAM models, automatic GAM selection
- functional data analysis (CLR: curve linear regression, KWF)

- **Sequential learning:**

- Adaptiv GAM
- Combining forecasts

- **Spatioa temporal/multi-scale models**

- Jiali Mei: link INSEE data to local electricity consumption (linky, grid data) and meteo data
- IBM simulation platform

CONTRIBUTIONS

- **Data mining for time series**
 - Random forest, boosting
 - Clustering and forecasting functional time series(collaboration with Q. Yao et H. Cho (LSE), and J.M. Poggi et E. Devijver (Orsay))
- **Large scale simulation**

GAM MODELS

- A good trade-off complexity/adaptivity

$$y_t = f_1(x_t^1) + f_2(x_t^2) + \dots + f(x_t^3, x_t^4) + \dots + \varepsilon_t$$

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

- **Publications**

- **Application on load forecasting**

- A. Pierrot and Y. Goude, Short-Term Electricity Load Forecasting With Generalized Additive Models **Proceedings of ISAP power**, pp 593-600, 2011.
 - R. Nédellec, J. Cugliari and Y. Goude, GEFCom2012: Electricity Load Forecasting and Backcasting with Semi-Parametric Models, *International Journal of Forecasting*, 2014, 30, 375 - 381.

- **GAM parallel**

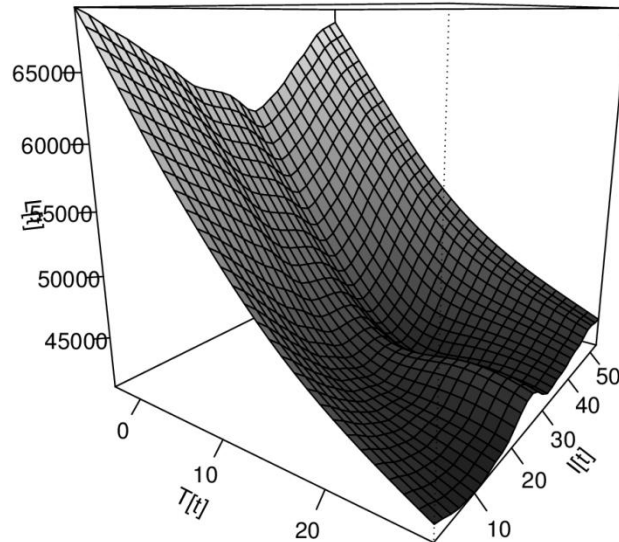
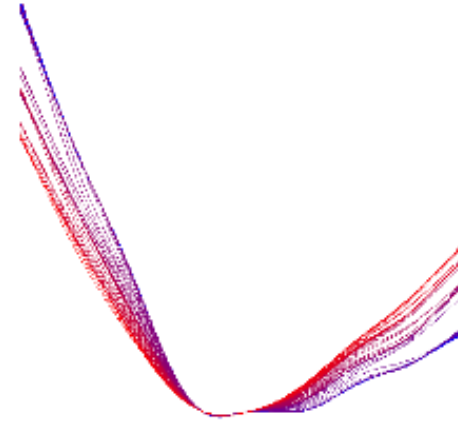
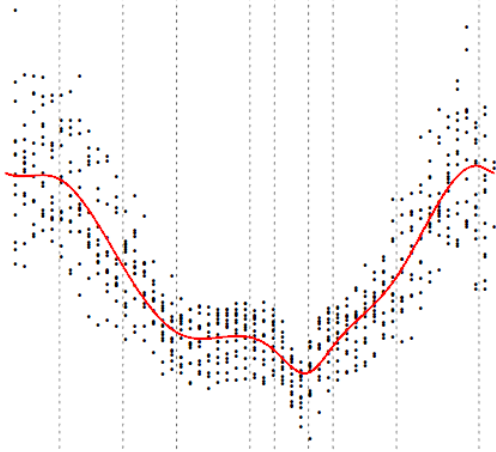
- S.N. Wood, Goude, Y. and S. Shaw, Generalized additive models for large datasets, to appear in **Journal of Royal Statistical Society-C**.

- **Adaptive GAM (forgetting factor)**

- A. Ba, M. Sinn, Y. Goude and P. Pompey, Adaptive Learning of Smoothing Functions: Application to Electricity Load Forecasting Advances in **Neural Information Processing Systems** 25, 2012, 2519-2527.

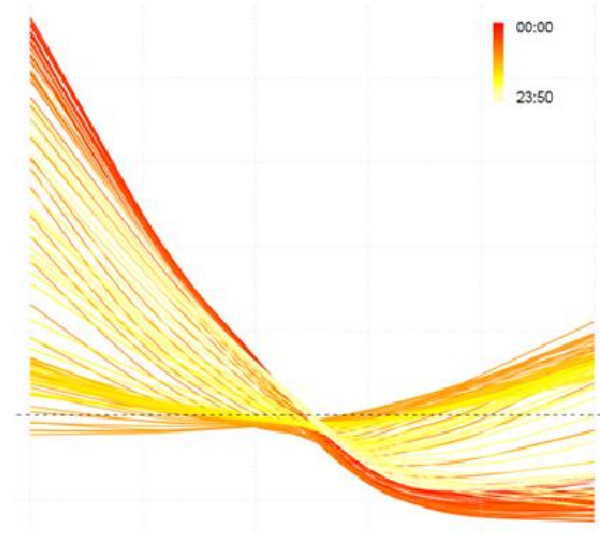
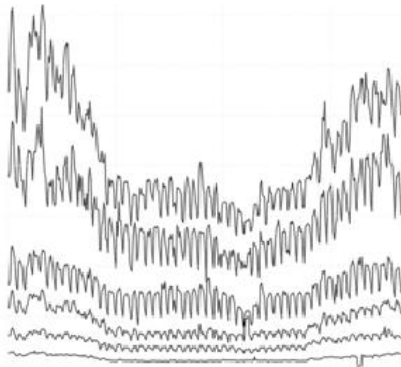
GAM MODELS

SOME ADDITIVE EFFECTS



GAM MODELS

APPLICATION ON ERDF SUBSTATIONS



Y.Goude et G. Adde, Modélisation et prévision des données de consommation électriques de postes sources, note H-R39-2012-00750-fr, 2012.

Goude, Y., Nédellec, R. and Kong, N., Local Short and Middle term Electricity Load Forecasting with semi-parametric additive models to appear in IEEE transactions on smart grid, 2013, 5 ,

Issue: 1, 440 - 446

COMBINING FORECASTS

$$\hat{y}_t = \sum_{i=1}^N \hat{p}_{i,t} x_{i,t} \quad \underbrace{\sum_{t=1}^T (\hat{y}_t - y_t)^2}_{\text{Notre perte}} = \underbrace{\min_{q \in \Delta_N} \sum_{t=1}^T (q^\top x_t - y_t)^2}_{\text{Perte de la meilleure combinaison convexe}} + \underbrace{R_T}_{\text{Erreur d'estimation}}$$

■ Publications

- Y. Goude : Mélange de prédicteurs et application à la prévision de consommation électrique. Thèse de doctorat, Université Paris-Sud XI, 2008.
 - M. Devaine, P. Gaillard, Y. Goude & G. Stoltz, Forecasting electricity consumption by aggregating specialized experts - A review of the sequential aggregation of specialized experts, with an application to Slovakian and French country-wide one-day-ahead (half-)hourly predictions **Machine Learning**, 2013, 90, 231-260.
- **Recent work on designing experts for combining**
- Gaillard, P. & Goude, Y., Forecasting electricity consumption by aggregating experts; how to design a good set of experts *top appear in* **Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension**, 2013.

CURVE LINEAR REGRESSION

- **Regressing curves on curves**

- Dimension reduction, SVD of $\text{cov}(Y,X)$, selection with penalised model selection
- Scalability for big data set

$$Y_i(u) = \mu_Y(u) + \int_{\mathcal{I}_2} \{X_i(v) - \mu_X(v)\} \beta(u, v) dv + \varepsilon_i(u)$$

- **Publications**

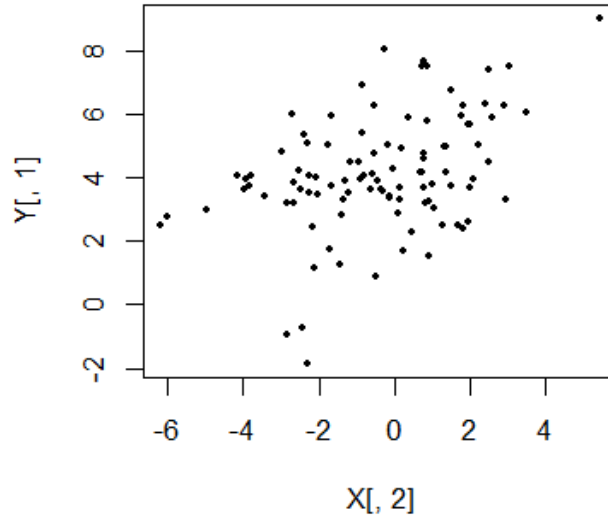
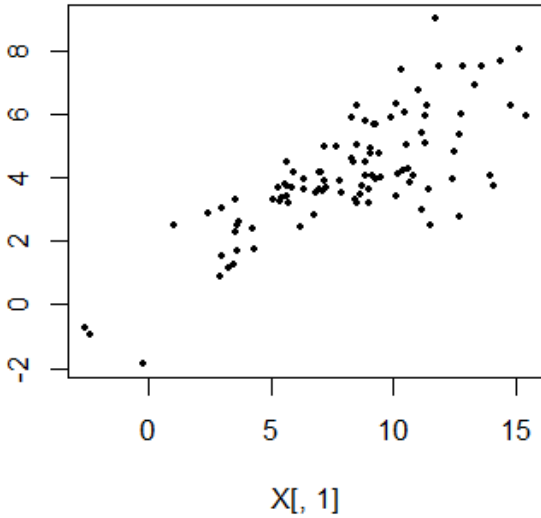
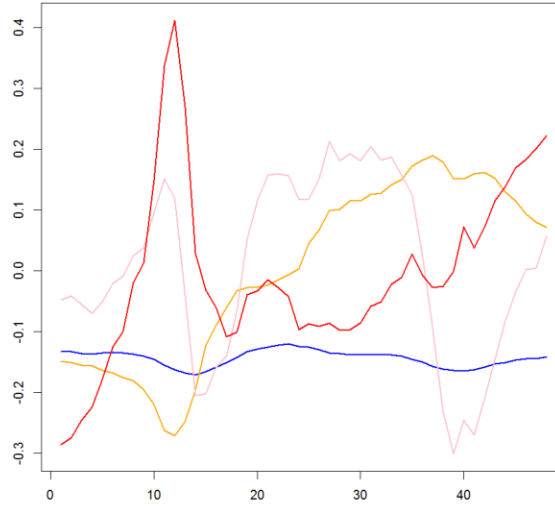
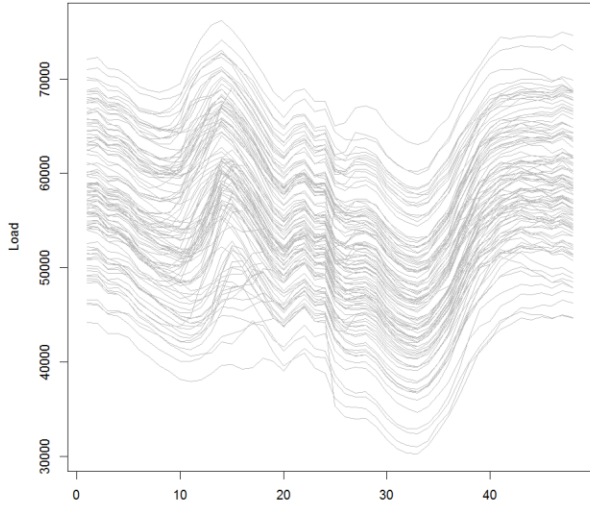
- **Application on electricity load forecasting**

- H. Cho, Y. Goude, X. Brossat & Q. Yao, Modeling and Forecasting Daily Electricity Load Curves: A Hybrid Approach **Journal of the American Statistical Association**, 2013, 108, 7-21.
- Cho, H.; Goude, Y.; Brossat, X. & Yao, Q, Modelling and forecasting daily electricity load using curve linear regression submitted to **Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension**.

- **Clustering functional data**

- H. Cho, Y. Goude, X. Brossat & Q. Yao, Clustering for curve linear regression, technical report, 2013.

CURVE LINEAR REGRESSION (2)



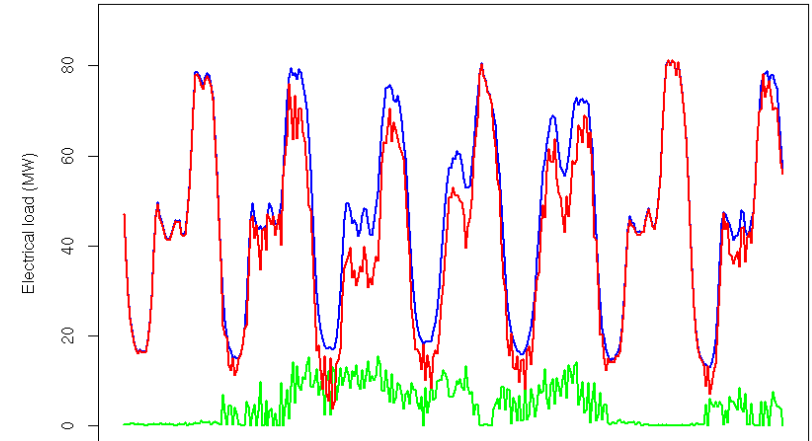
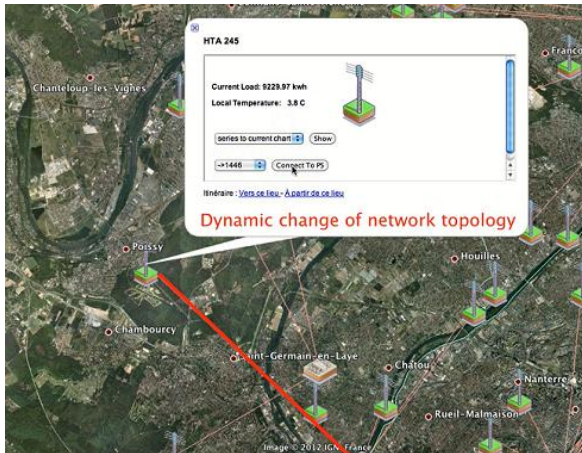
From functional regression to Linear regression

SPATIO-TEMPORAL/COMPLEX DATA STRUCTURE

- **Interpolation of local electricity data, use of INSEE data: Jiali Mei internship (with G. Hebrail)**
 - INSEE data at IRIS geographical definition, electricity data on the grid, local meteo data (model outputs, station data)
 - temporal definition: INSEE yearly, load 10 min, meteo hourly
 - Application: downscaling electricity data
- **Maquette EDF R&D/IBM**
- **Publications:**
 - **Intensive simulation of local load curves :**
 - Pompey, P.; Bondu, A.; Goude, Y.; Sinn, M. & Picard, M.-L. GAM model based large scale electrical load simulation for smart grids submitted to **Lecture Notes in Statistics: Modeling and Stochastic Learning for Forecasting in High Dimension**, 2013.
 - Mei, J.; Hebrail, G. & Goude, Y. Spatial Interpolation of Electricity Consumption Data Using Socio-demographic Information, to be submitted, 2014

LARGE SCALE SIMULATIONS

- **Project SIGMA2: collaboration EDF R&D/IBM (First Of A Kind)**
 - Massive simulation (linky scale: 10^6 meters)
 - On-line forecasting, on-line GAM
 - What if scenario:
 - usages
 - Tariffs
 - Grid topology



Simulation: Week August 11-17

INTERPOLATION OF LOCAL ELECTRICITY DATA

- Reconstitution of electricity consumption at the HTA/BT:
 - Meteo data
 - Intensity metered at the HTA nodes
 - INSEE data

