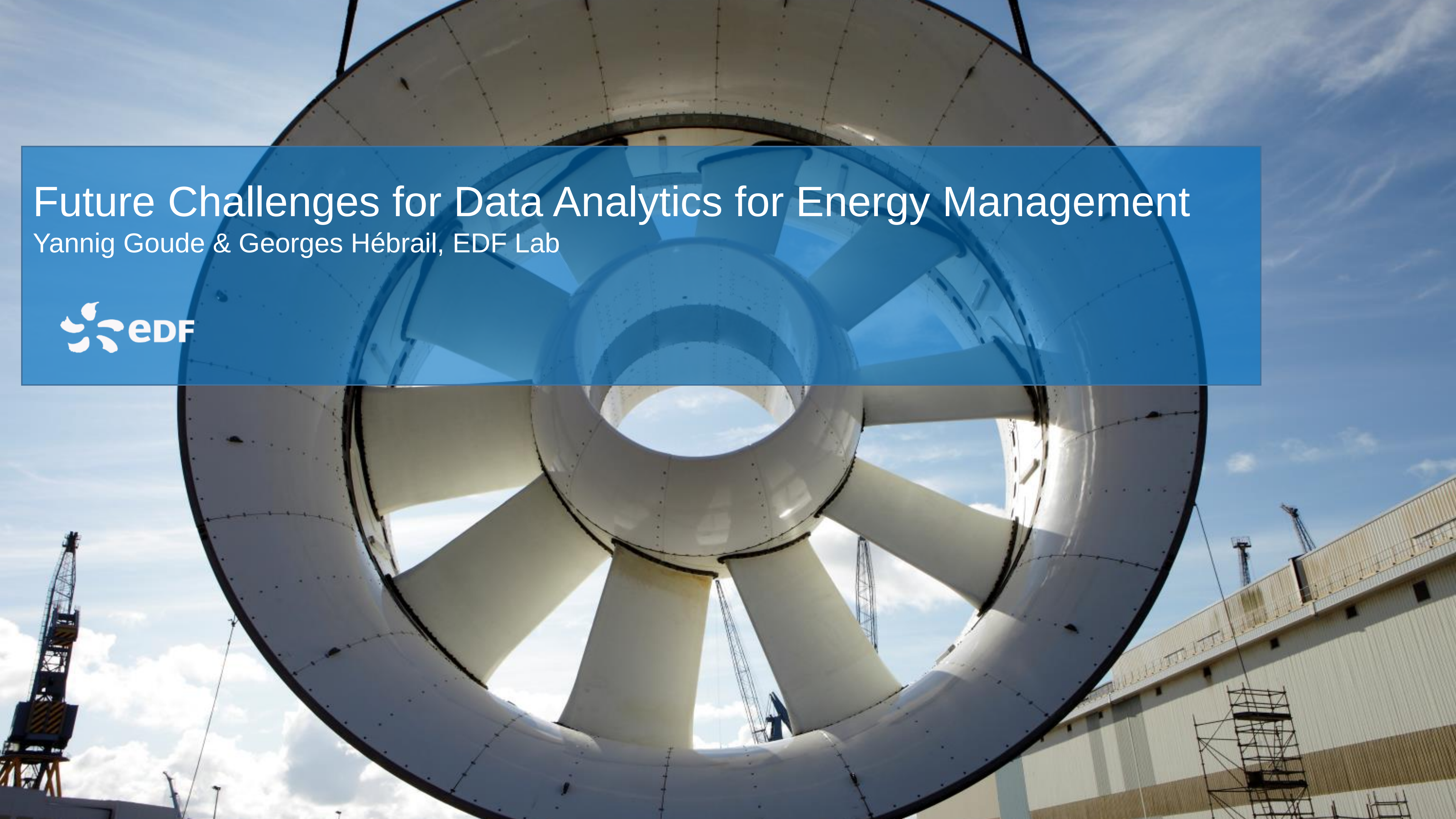


Future Challenges for Data Analytics for Energy Management

Yannig Goude & Georges Hébrail, EDF Lab

Main Data science applications for EDF

- Power Generation
  - Process monitoring and condition-based maintenance from sensors
  - Power generation forecasting for renewables
- Energy management
  - Load forecasting
  - Balancing and optimizing generation and consumption (using smart metering information, including renewables)
- Electrical networks
  - Smart Grid operations (local)
  - Condition-based maintenance
- Customers and sales
  - Customer Relationship Management
  - New services to customers using smart-metering data
  - Smart Homes, Smart Building, Smart Cities operation related to energy

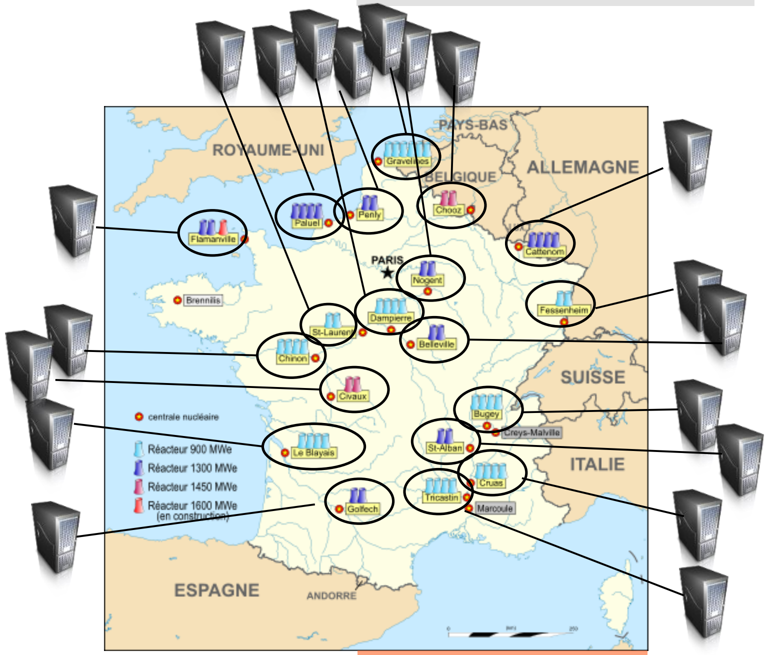
Data science value: *optimization of internal processes*

Data science value: *creation of new services to customers/partners*

Power Generation

BIG DATA SOLUTION FOR Operations and maintenance of the nuclear fleet

One DB by nuclear site, gathering data from sensors. Use of Data Historians.



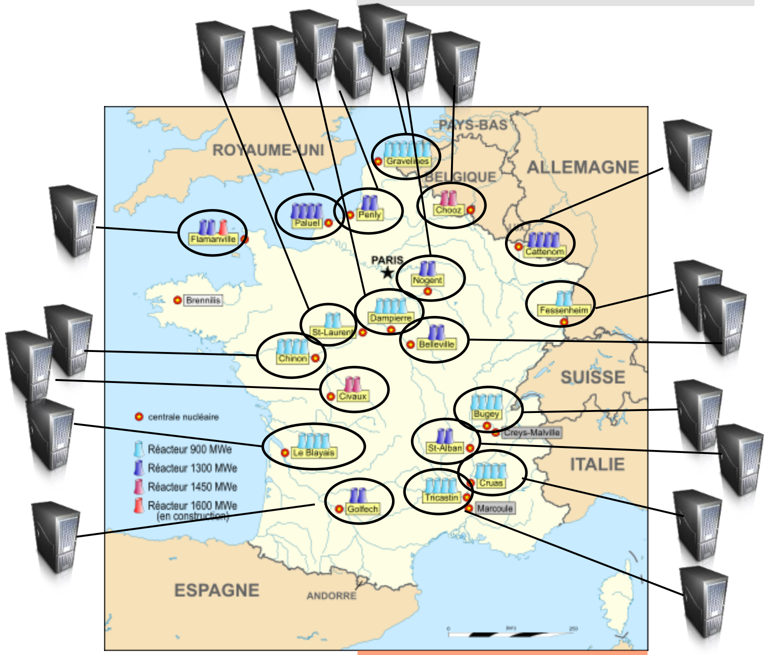
Source : Wikipedia

- **Focus on data:**
  - High volume:
    - data is stored up to 40-60 years (plant lifetime)
    - SCADA data can be sampled every 20 to 40 ms (but mainly a few seconds)
    - Around 10.000 sensors per plant
  - Variety:
    - Data is heterogeneous
    - Time series, images, documents
    - Various data sources

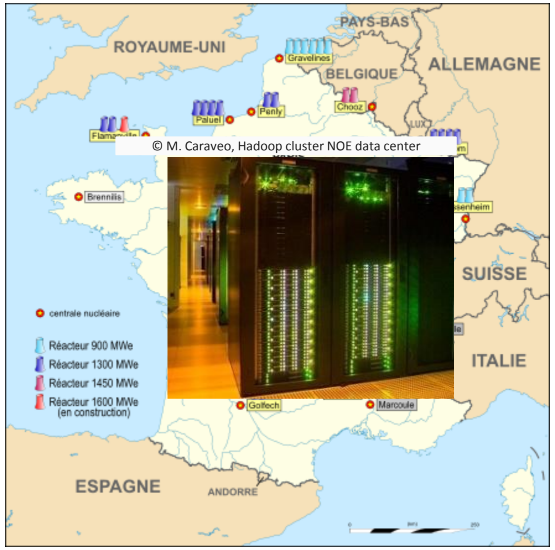
- **Current systems (historians) don't allow too many concurrent access, SLA's are quite low**

A DATA LAKE FOR THE NUCLEAR FLEET

One DB by nuclear site, gathering data from sensors. Use of Data Historians.



Source : Wikipedia

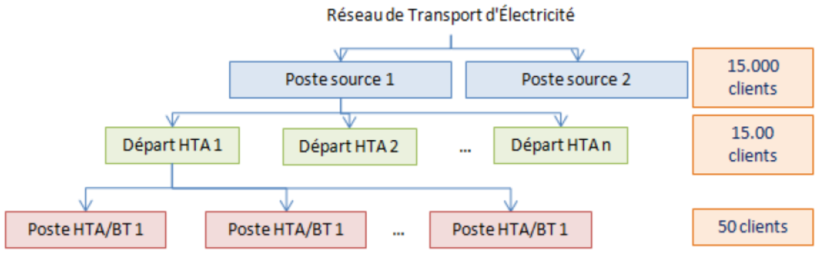


ESPADON : the Data Lake for the nuclear fleet

A Data Lake for the nuclear fleet

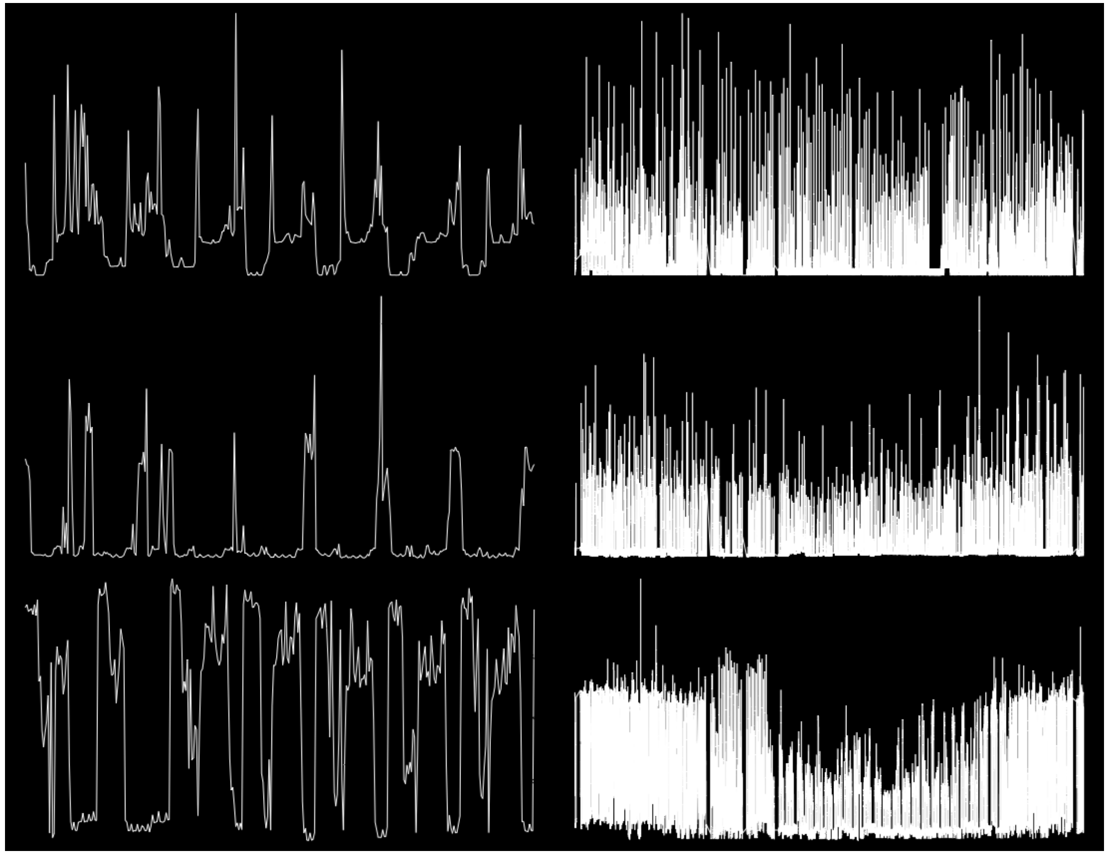
Industrial Motivation (1)

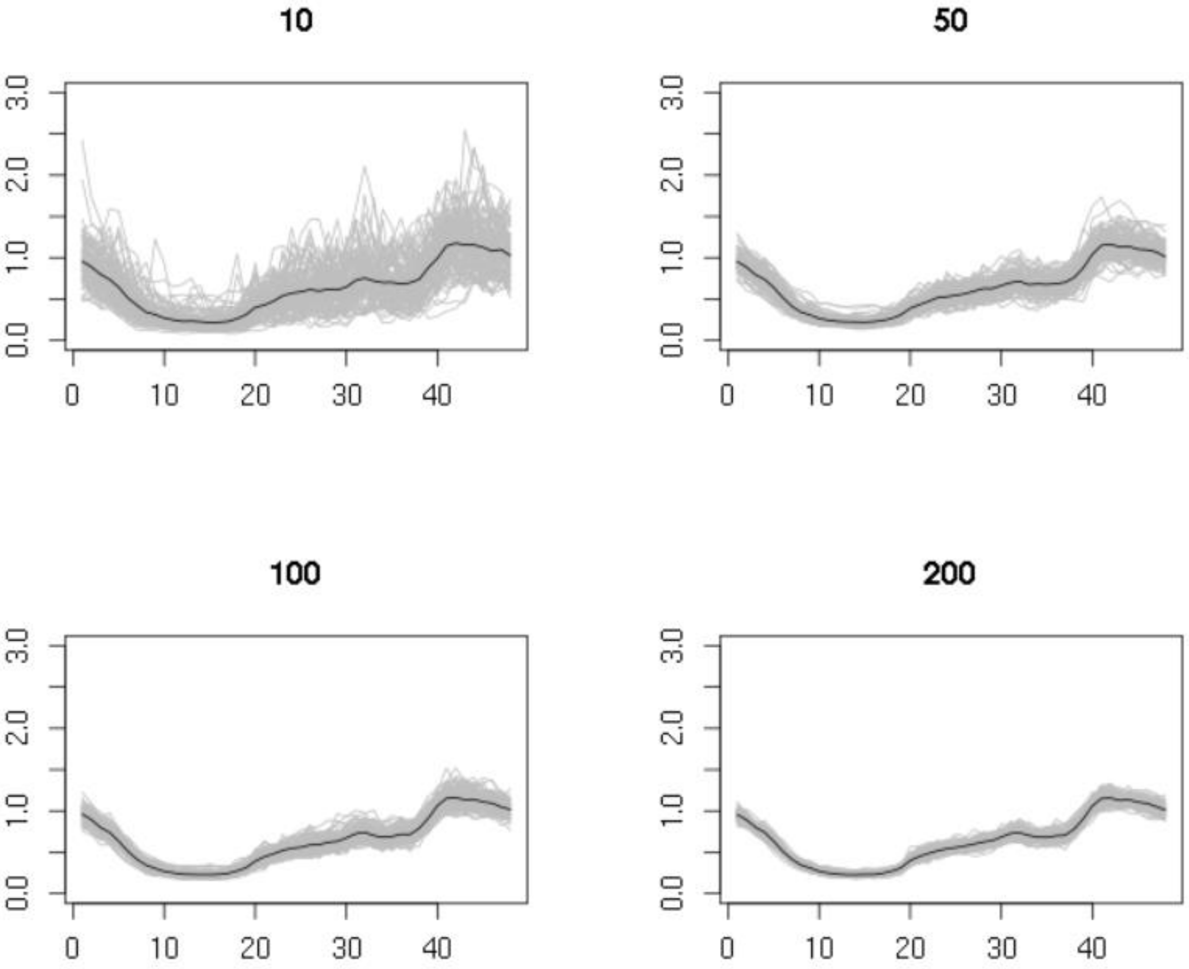
- Forecasting at a low spatial resolution level for the grid management



Industrial Motivation (2)

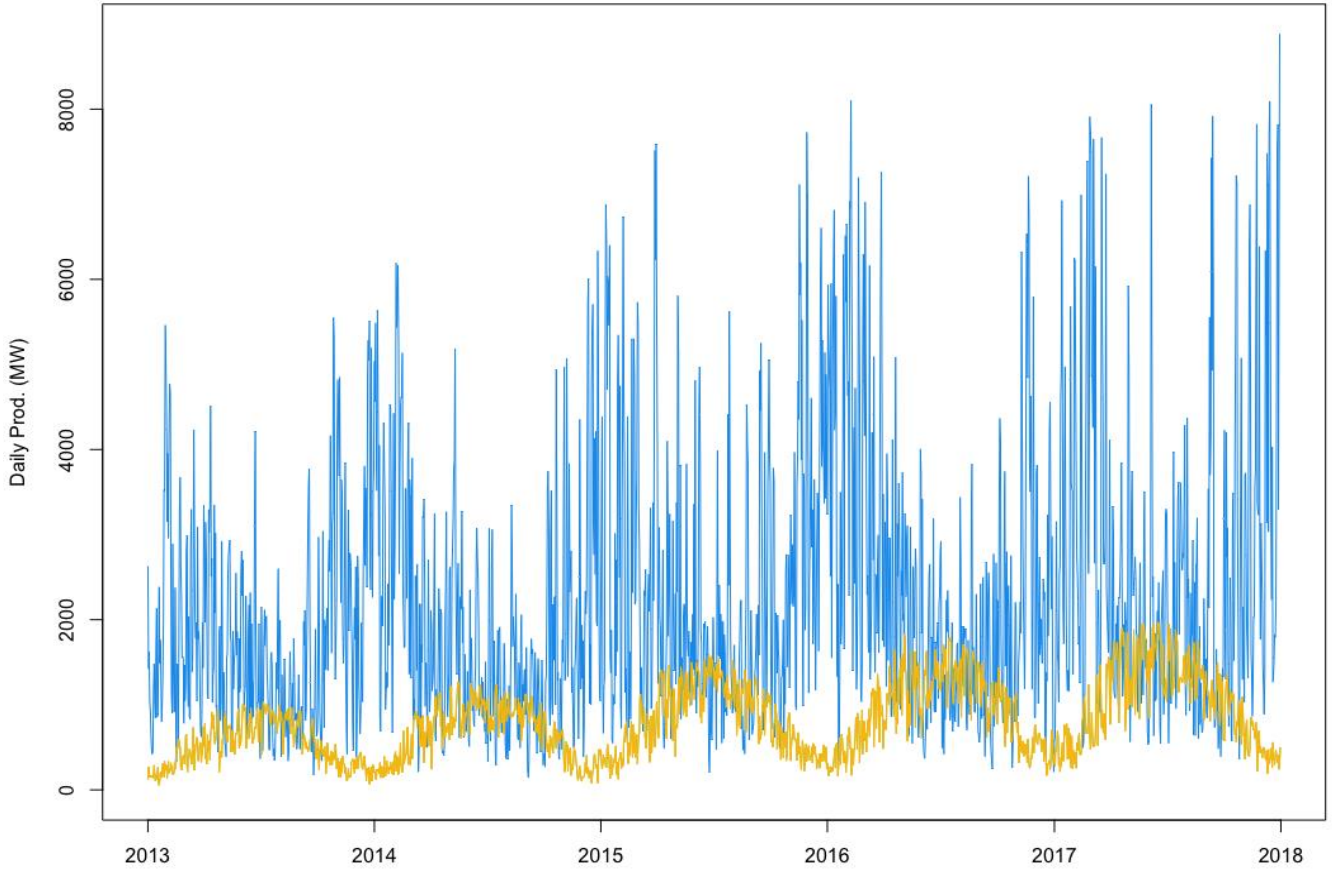
- Integrate individual metered data in our (global) forecasts

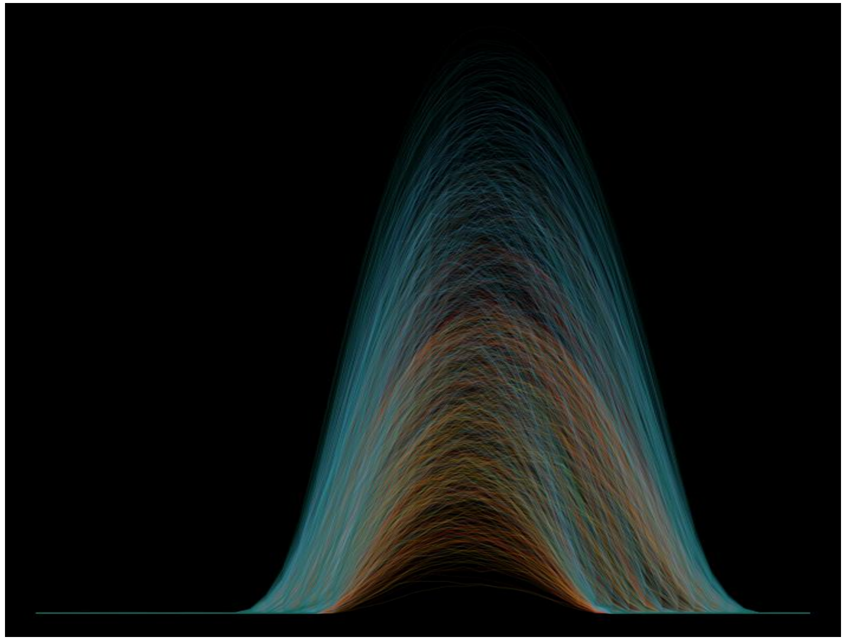


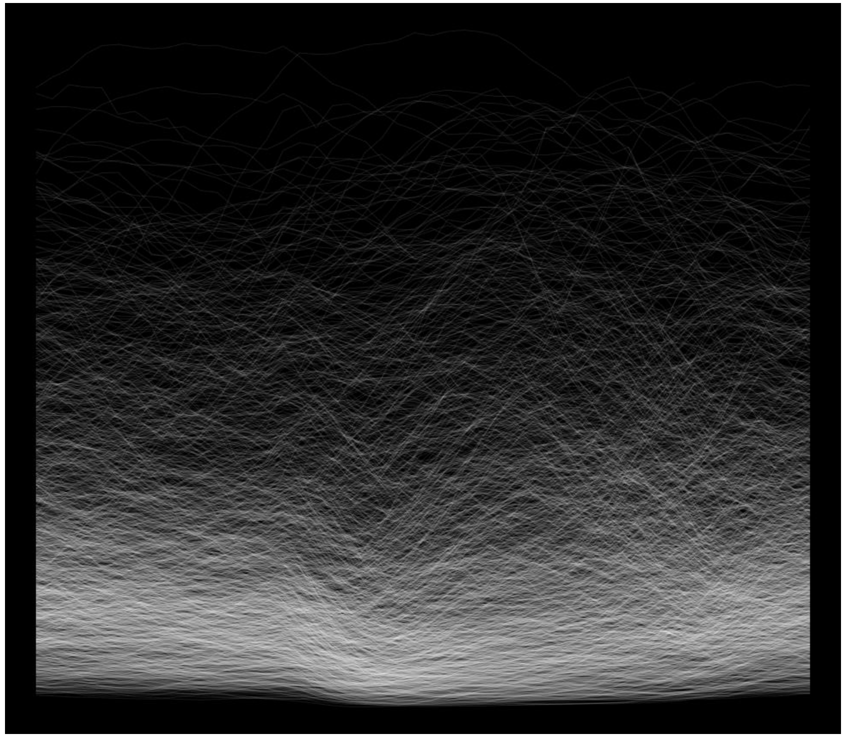


Industrial Motivation (3)

- Probabilistic forecasts

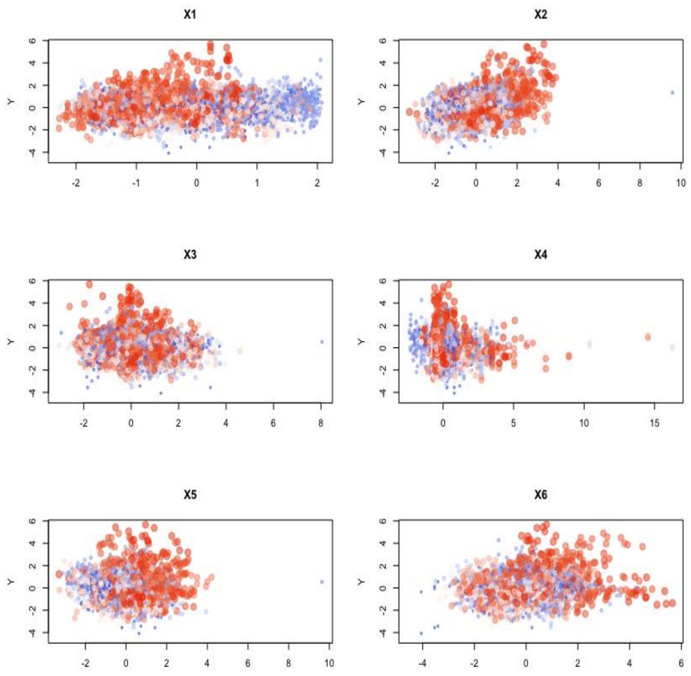


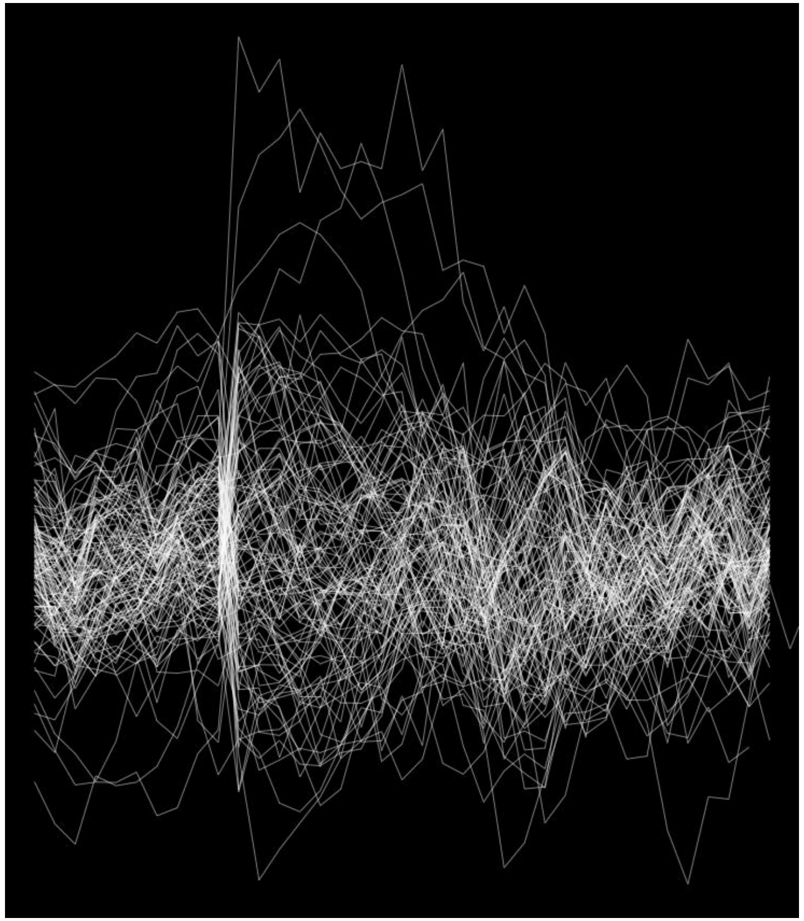
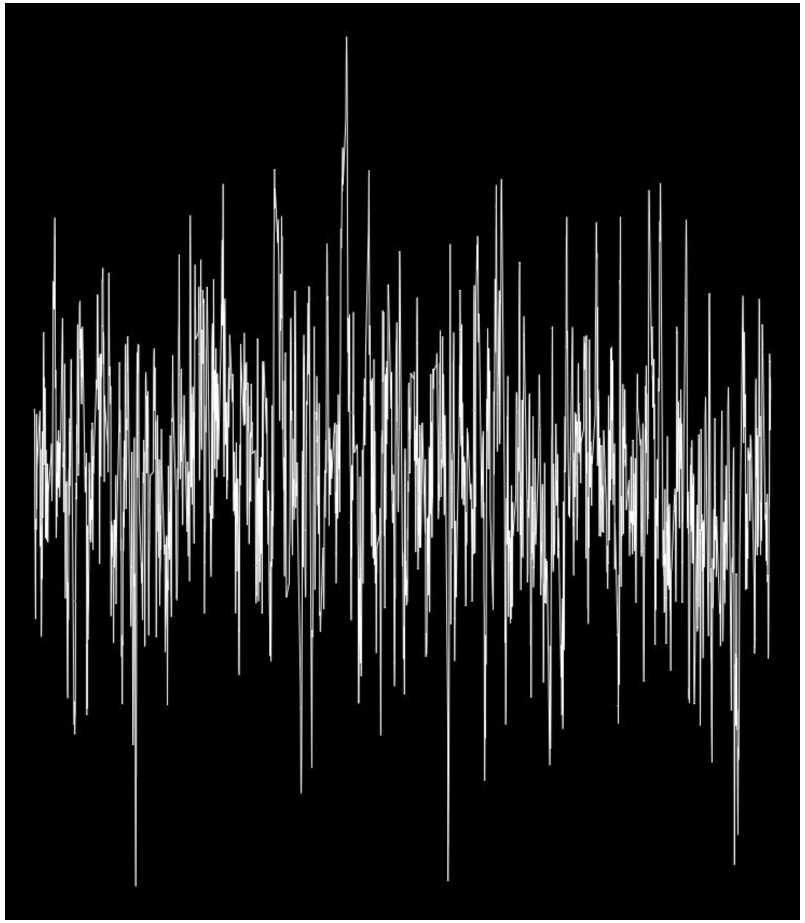




Industrial Motivation (4)

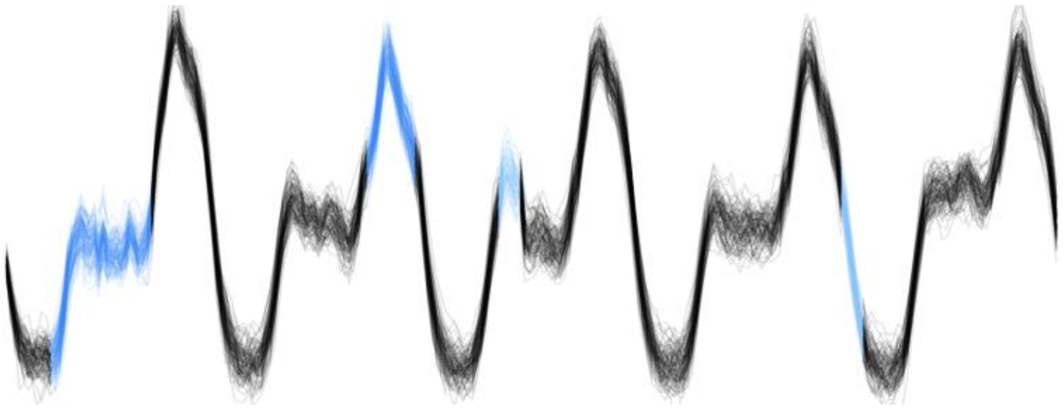
- Online learning for energy markets

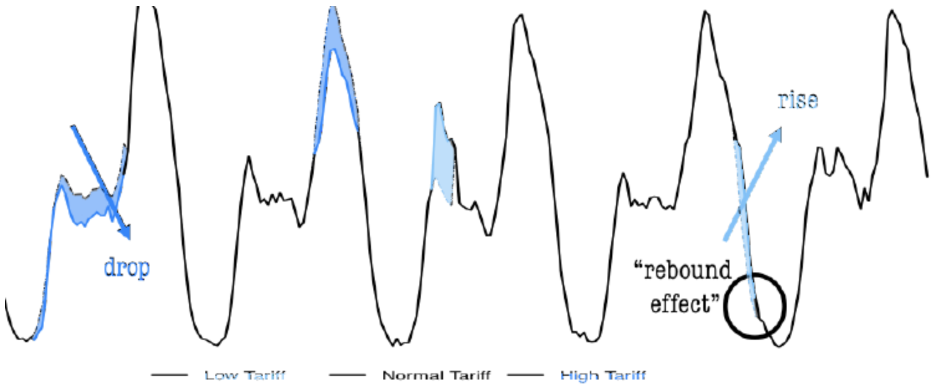


Industrial Motivation (5)

- Demand response
- Sensors data
- Smart meters

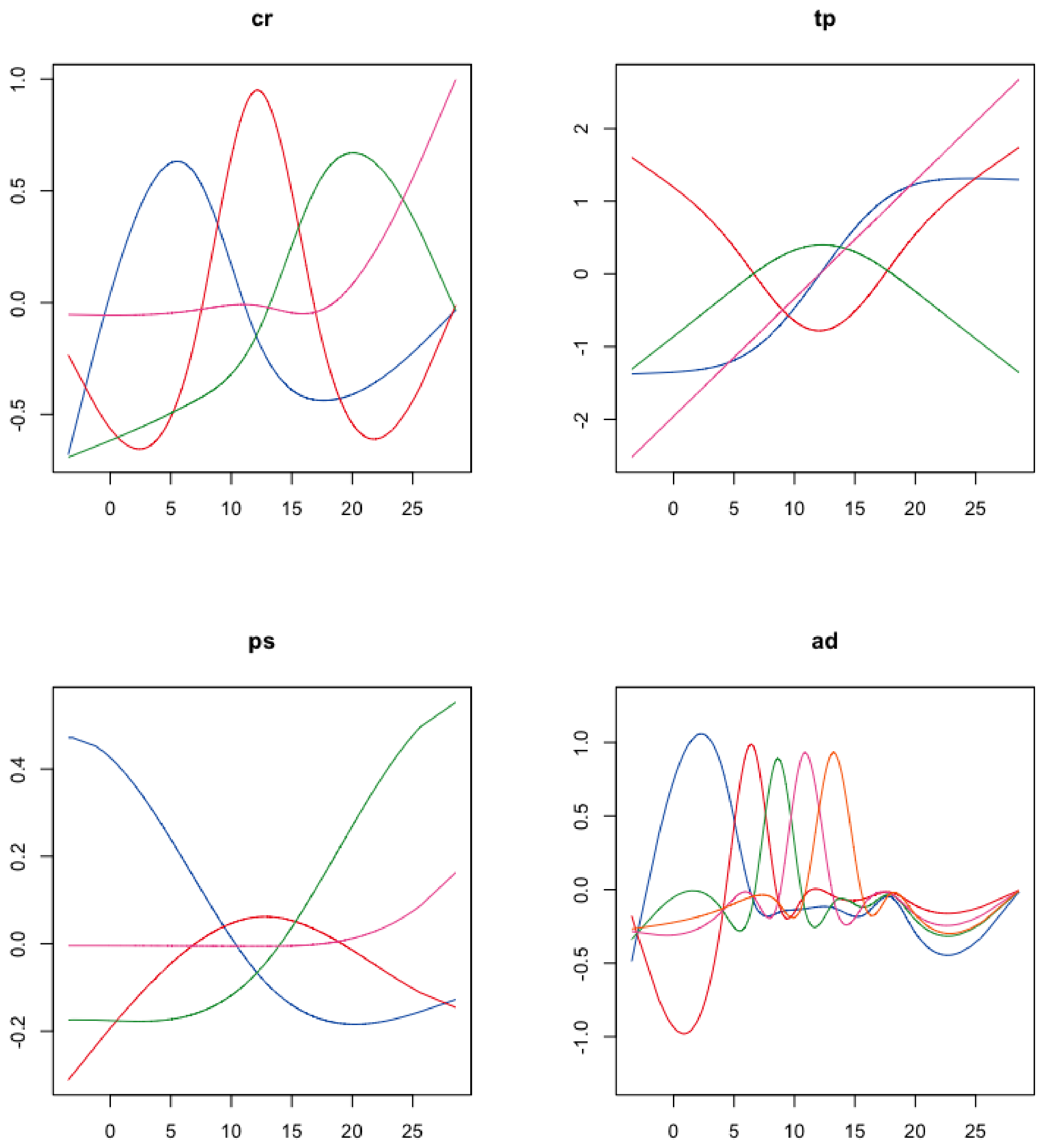




GAMs (1)

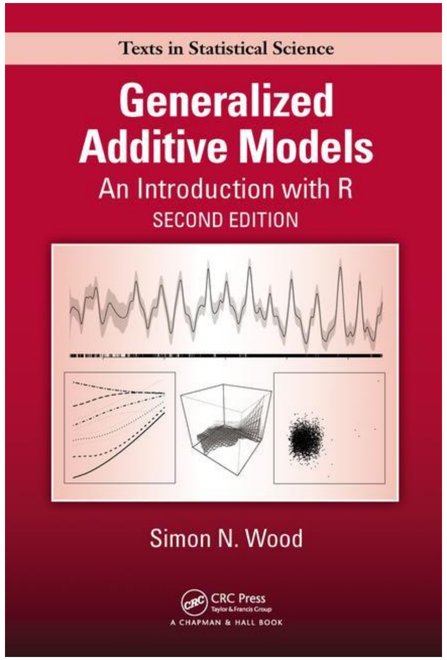
$$Y_i = \beta_0 + f_1(X_{1,i}) + \cdots + f_d(X_{d,i}) + \varepsilon_i$$

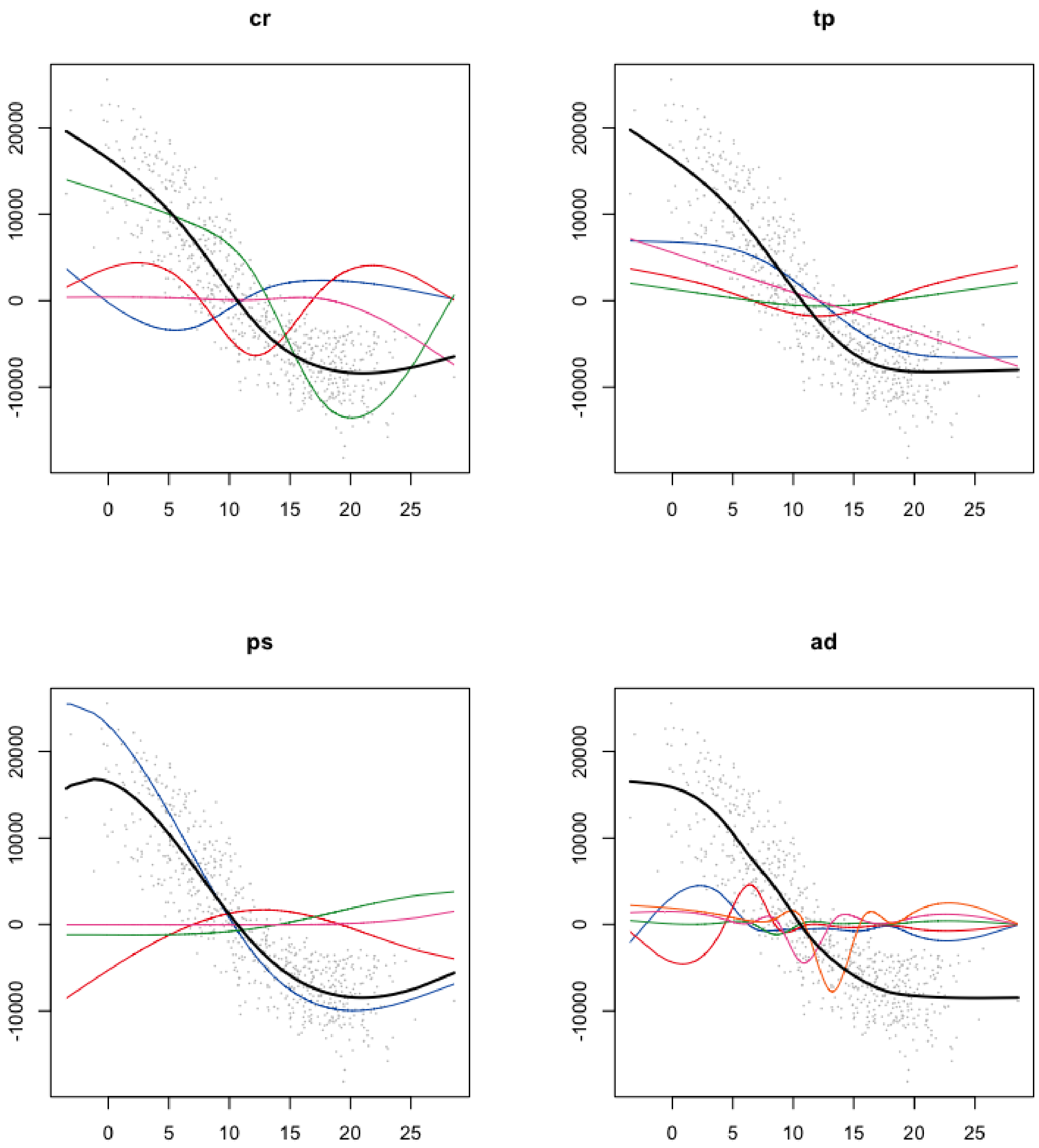
$$f_j(x_j) = \sum_{i=1}^{k} \beta_{ji} b_{ji}(x_j).$$



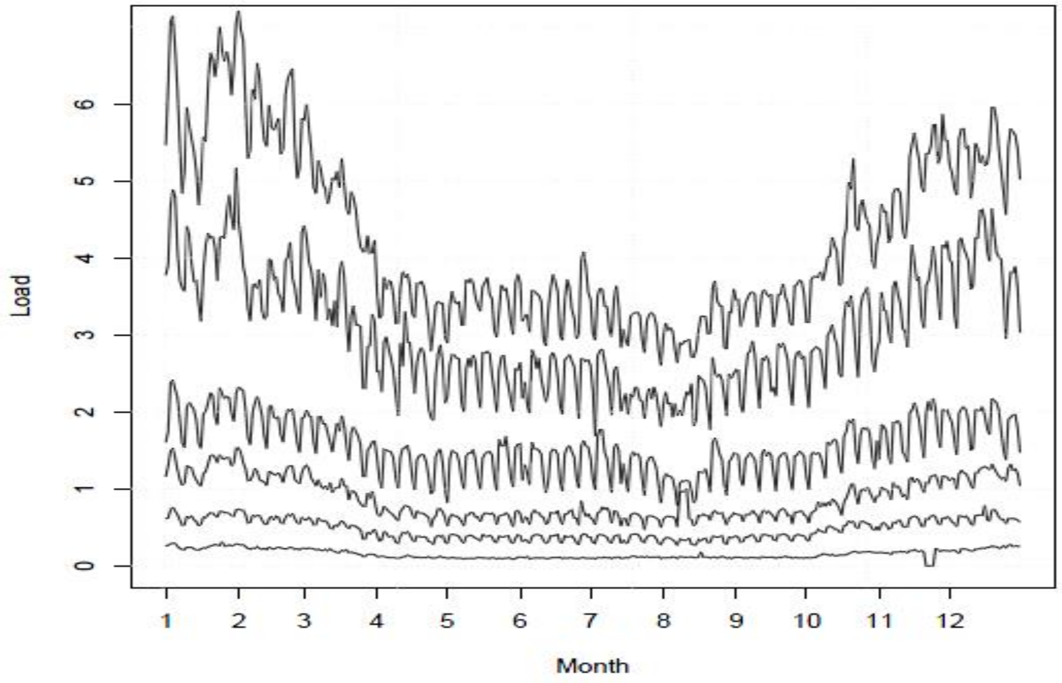
GAMs (2)

$$\sum_{i=1}^{n}(y_i - \beta_0 \mathbf{X}_i^0 - \sum_{q=1}^{p} f_q(x_i))^2 + \sum_{q=1}^{p} \lambda_q \int |||f_q^{''}(x)|||^2 dx$$





GAMs (3)



Goude, Y.; Nedellec, R. & Kong, N. Local Short and Middle term Electricity Load Forecasting with semi-parametric additive models I*EEE transactions on smart grid,* **2013***, 5 , Issue: 1*, 440 – 446.

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.

S.N. Wood, Goude, Y. and S. Shaw, Generalized additive models for large datasets, Journal of Royal Statistical Society-C, Volume 64, Issue 1, pages 139–155, January 2015.

Covariate selection with GAMs

Algorithm

1. **First step: subset selection (Group LASSO)**
   For each $\lambda_i \in \Lambda_{GrpL}$

   - Solve

   $$\hat{\boldsymbol{\beta}}^{\lambda_i} = \arg\min\{Q^{OLS}(\boldsymbol{\beta}) + \lambda_i \sum_{j=1}^{d} \sqrt{m_j} ||\boldsymbol{\beta}_j||_2\}$$

   - Denote $S^{\lambda_i} = \{j | \hat{\boldsymbol{\beta}}_j^{\lambda_i} \neq 0\}$

2. **Second step: Estimation of the additive model (by OLS)**
   For each support set $S^{\lambda_s} \in \{S^{\lambda_{min}}, \ldots, S^{\lambda_{max}}\}$

   - Compute

   $$Q_{S_{\lambda_s}}^{OLS}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(Y_i - \beta_0 - \sum_{j \in S_{\lambda_s}} C_{ij}(\boldsymbol{\beta}_j) \right)^2$$

   - Solve

   $$\tilde{\boldsymbol{\beta}}^{S_{\lambda_s}} = \arg\min\{Q_{S_{\lambda_s}}^{OLS}(\boldsymbol{\beta})\},$$

   - Compute the BIC (see Eq. (5)) for each $\tilde{\boldsymbol{\beta}}^{S_{\lambda_s}}$

3. **Third step: Selection of the final model** Select $\tilde{\boldsymbol{\beta}}^{S_{\lambda_b}}$ which minimizes the BIC

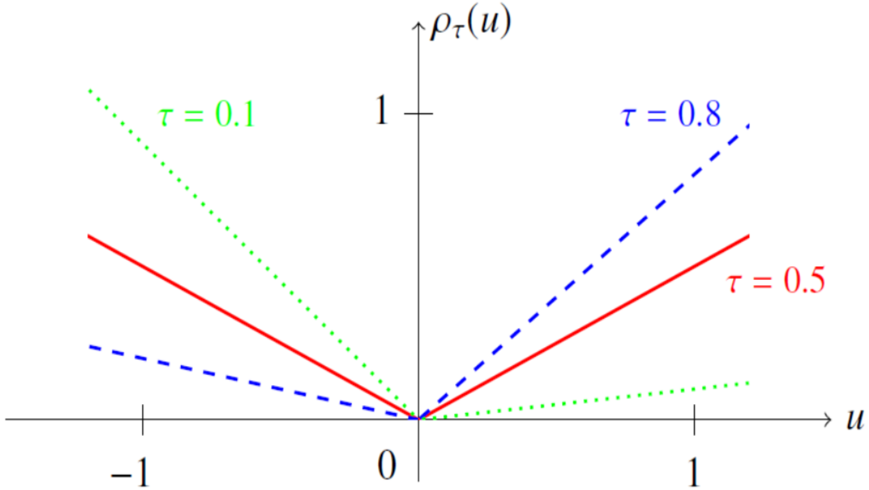
Criterion	MAPE	RMSE
Post2Bic>0.2	**1.12**	**645**
Post2Gcv>0.3	1.15	648
Post2Aic	1.17	663
Post2Gcv	1.17	667
EDF model	1.16	667
Post2Bic	1.24	730
BenchMT1	2.00	1173

- Automatic calibration and selection of GAMs
- Perform as an expert calibrated model on EDF data

Thouvenot, V.; Pichavant, A.; Goude, Y.; Antoniadis, A. & Poggi, J.-M. Electricity Forecasting Using Multi-Stage Estimators of Nonlinear Additive Models Power Systems, IEEE Transactions on, 2015, PP, 1-9

PhD thesis of Vincent Thouvenot (UPSUD-EDF R&D) Estimation et sélection pour des modèles additifs et application à la prévision de la consommation électrique.

qGAM (1)

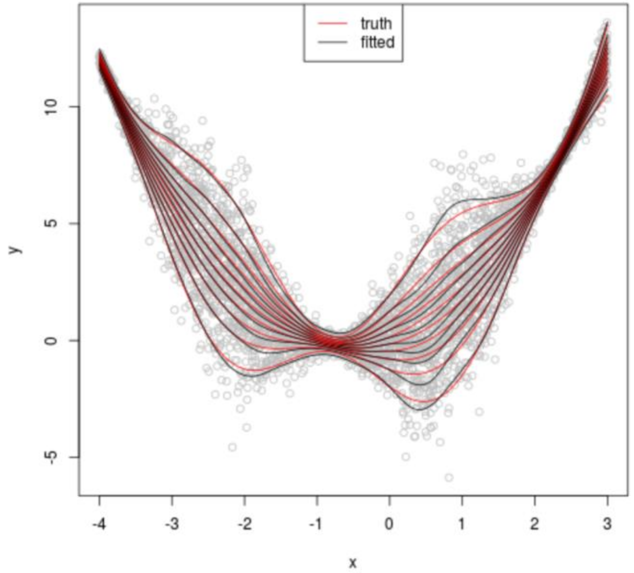


$$q_\tau(Y|\boldsymbol{X}) = F^{-1}_{Y|\boldsymbol{X}}(\tau) = \inf\{y \in \mathbb{R}, F_{Y|\boldsymbol{X}}(y) \geq \tau\}$$

$$q_\tau(Y|\boldsymbol{X}) \in \arg\min_g \mathbb{E}[\rho_\tau(Y - g(\boldsymbol{X}))|\boldsymbol{X}]$$

Ex. here

$$f_1(X_{t,1}) + f_2(X_{t,2}) + f_3(X_{t,3}, X_{t,4}) + \ldots$$

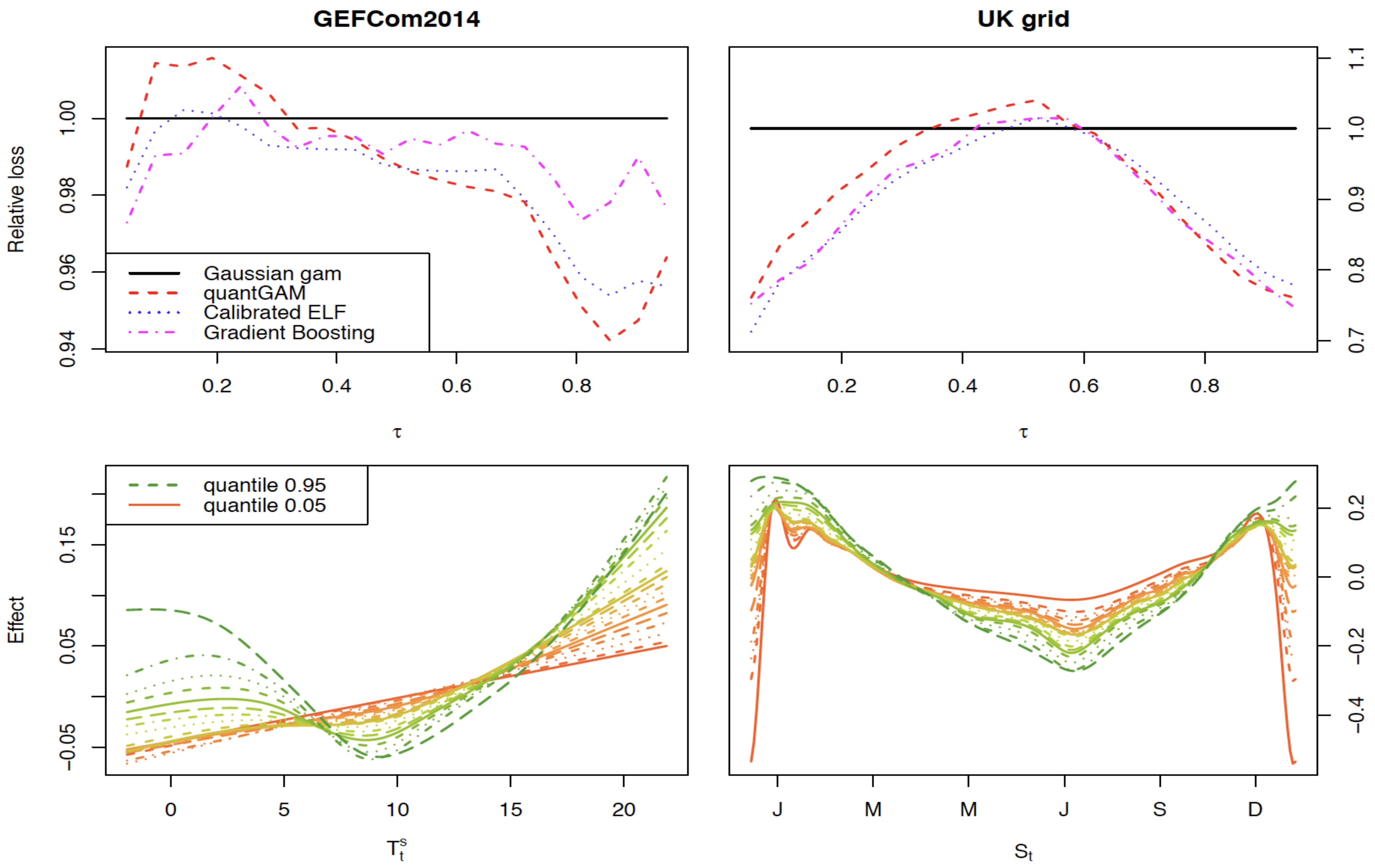


Fasiolo, M., Goude, Y., Nedellec, R. and Wood, S. N. (2016). Fast calibrated additive quantile regression. Available at https://arxiv.org/abs/1707.03307

Gaillard, P., Goude, Y. and Nedellec, R. (2016). Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting. International Journal of Forecasting, 32, 3, 1038-1050.

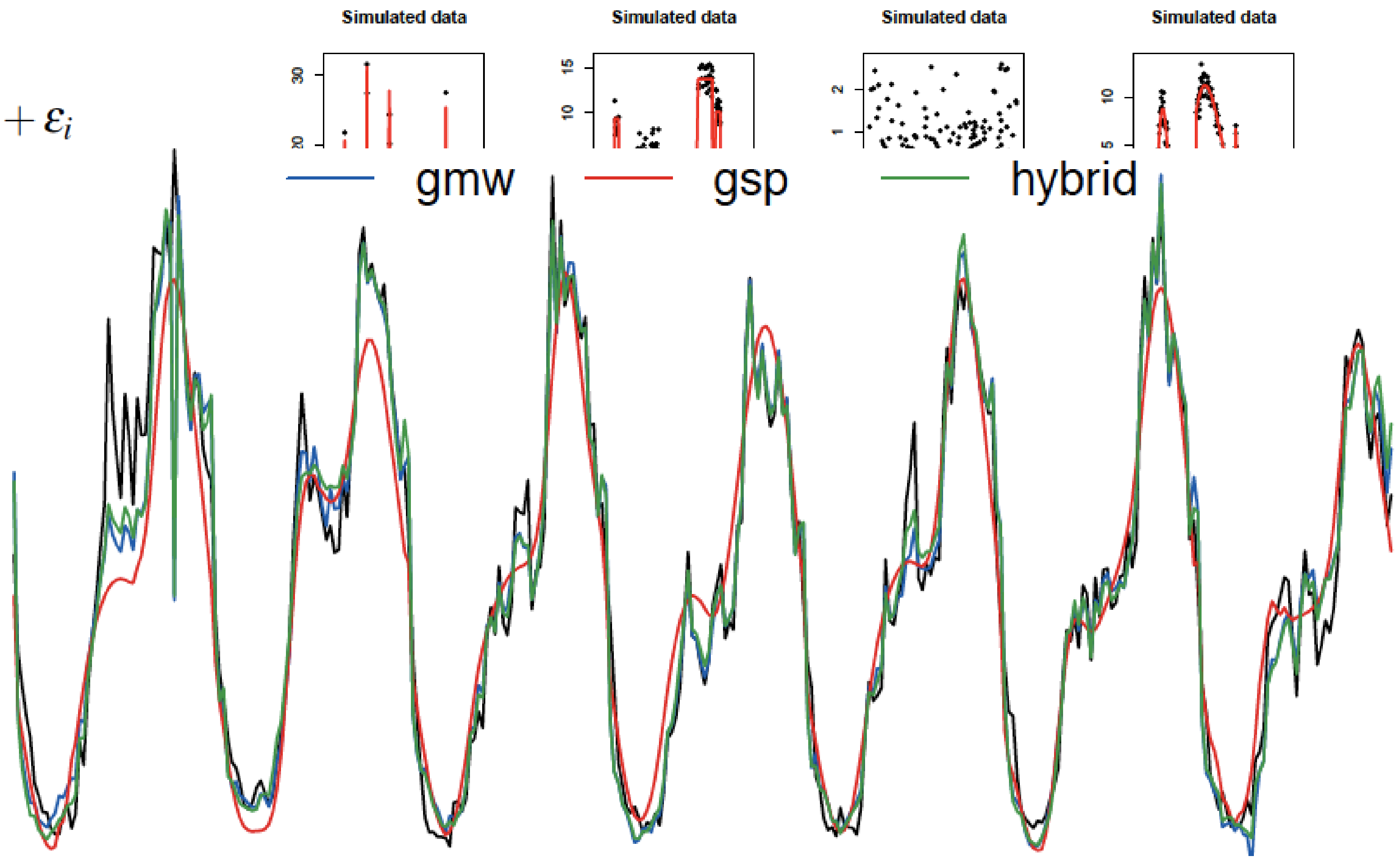
https://cran.r-project.org/web/packages/qgam/index.html

qGAM (2)


GEFCom2014
UK grid
Relative loss
Gaussian gam
quantGAM
Calibrated ELF
Gradient Boosting
1.00
0.98
0.96
0.94
1.1
1.0
0.9
0.8
0.7
0.2
0.4
0.6
0.8
τ
quantile 0.95
quantile 0.05
Effect
0.15
0.05
−0.05
0.2
0.0
−0.2
−0.4
0
5
10
15
20
T^s_t
J
M
M
J
S
D
S_t

Hybrid PLAM (Wavelets and splines)

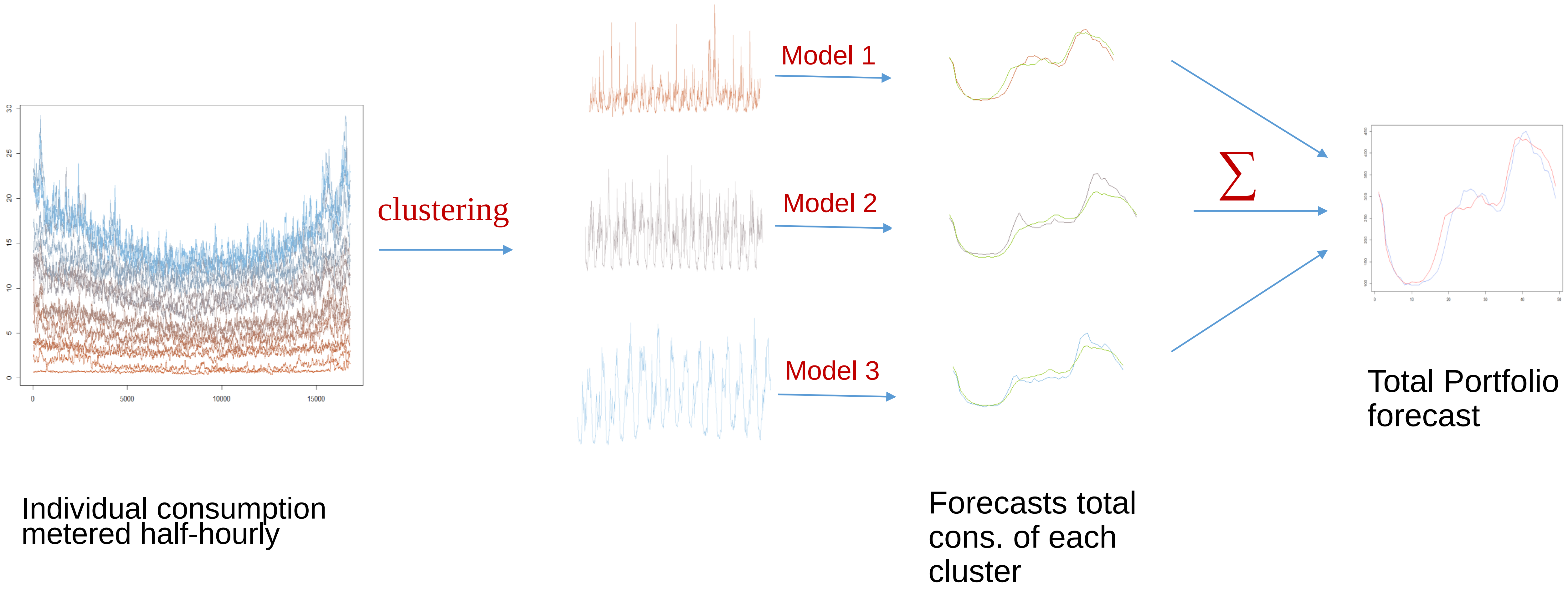
$$Y_i = \mathbf{X}_i^T \beta + \sum_{j=1}^{q_s} f_j^{(1)}(T_{ij}^{(1)}) + \sum_{j=1}^{q_w} f_j^{(2)}(T_{ij}^{(2)}) + \varepsilon_i$$



- estimation of unsmooth components at low cost
- Tarif effects, peaks

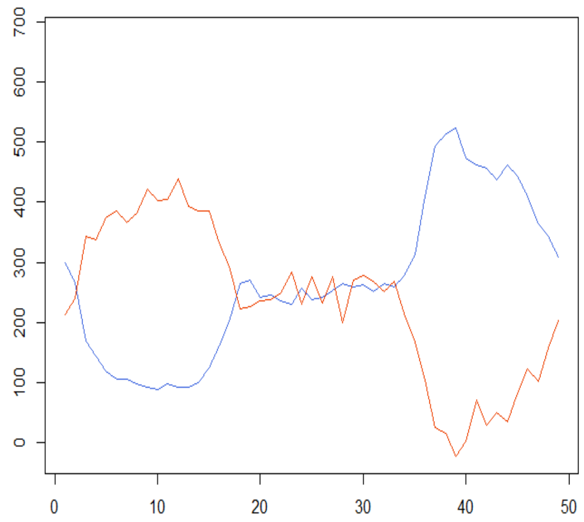
Estimation and group variable selection for additive partial linear models with wavelets and splines
Authors: Umberto Amato, Anestis Antoniadis, Italia De Feis and Yannig Goude, South African Statistical Journal 51, pp 235 –272 (2017)

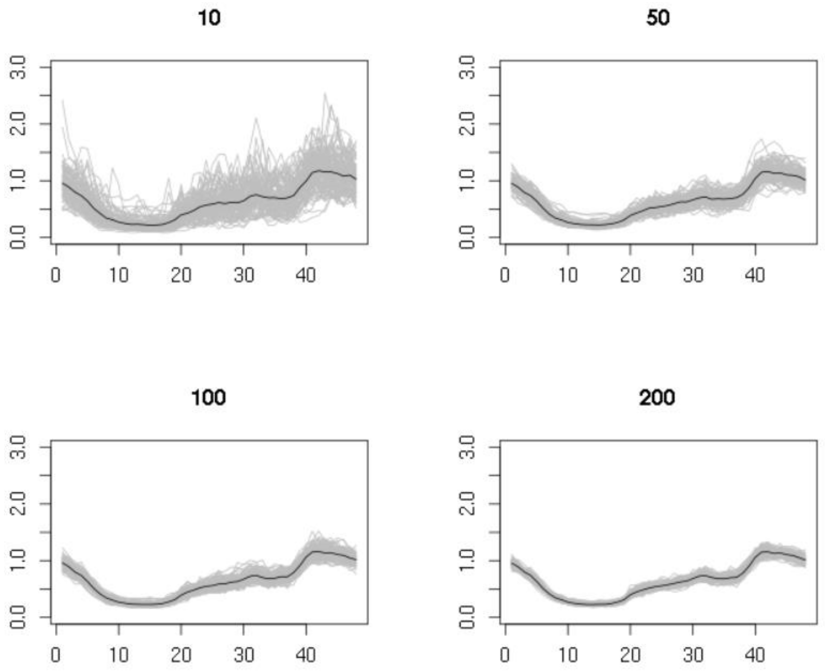
Forecasting total consomption of a set of customers (1)

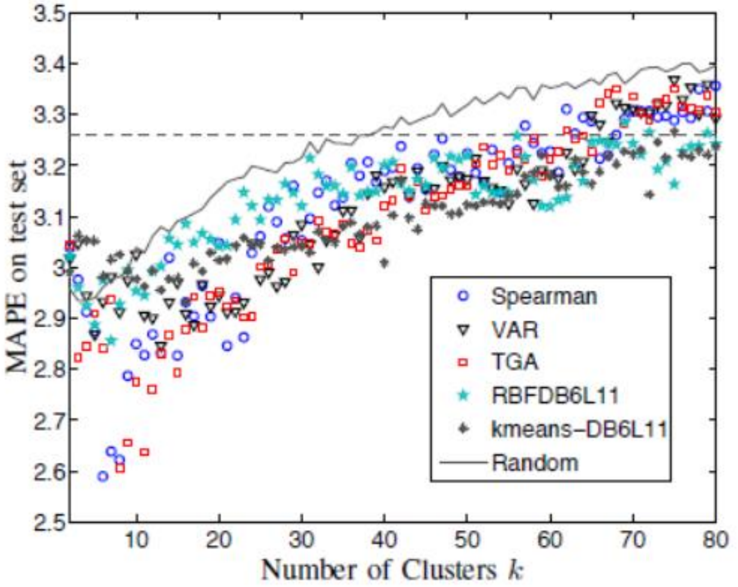


Disaggregated Electricity Forecasting using Wavelet-Based Clustering of Individual Consumer
Proccedings of IEEE Energycon, 2016, Jairo Cugliari, Yannig Goude, Jean-Michel Poggi

Forecasting total consomption of a set of customers (2)







From: C. Alzate and M. Sinn, “Improved electricity load forecasting via kernel spectral clustering of smartmeter”, International Conference on Data Mining, vol. 948, pp. 943 – 948, 2013.

What type of customers in each cluster?

Do they behave similarly?

Are they complementary?

How many (at least) customers in each cluster?

How many clusters?

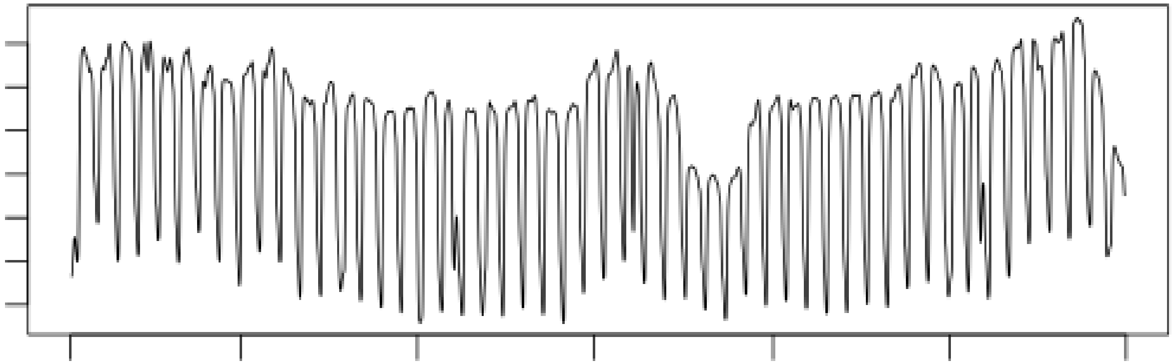
Which forecasting model, clustering algorithm?

Are they related in any sense?

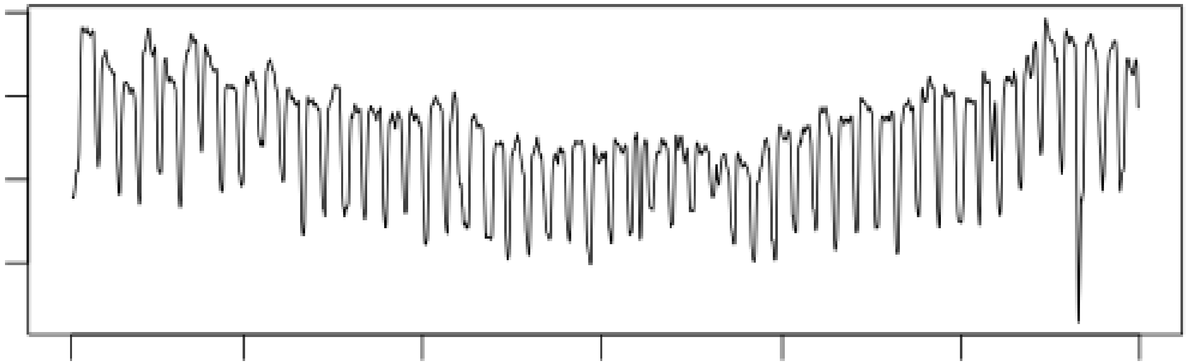
Forecasting total consomption of a set of customers (3)

- Data set of 25011 professional customers
- Sampling period: 30 minutes
- Period: 2009, 2010 and 2011 (only 6 months)
- 1 year = 25011*17520= 438 millions of samples = 3.25 Go

Total consumption 2010



Individual consumption 2010



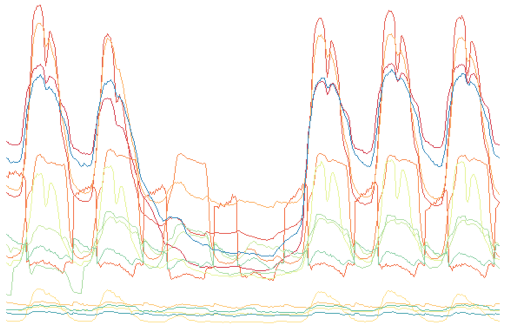
Forecasting total consomption of a set of customers (4)

- 1st stage: create a large number of *K'* = 200 super customers *fast and scalable*

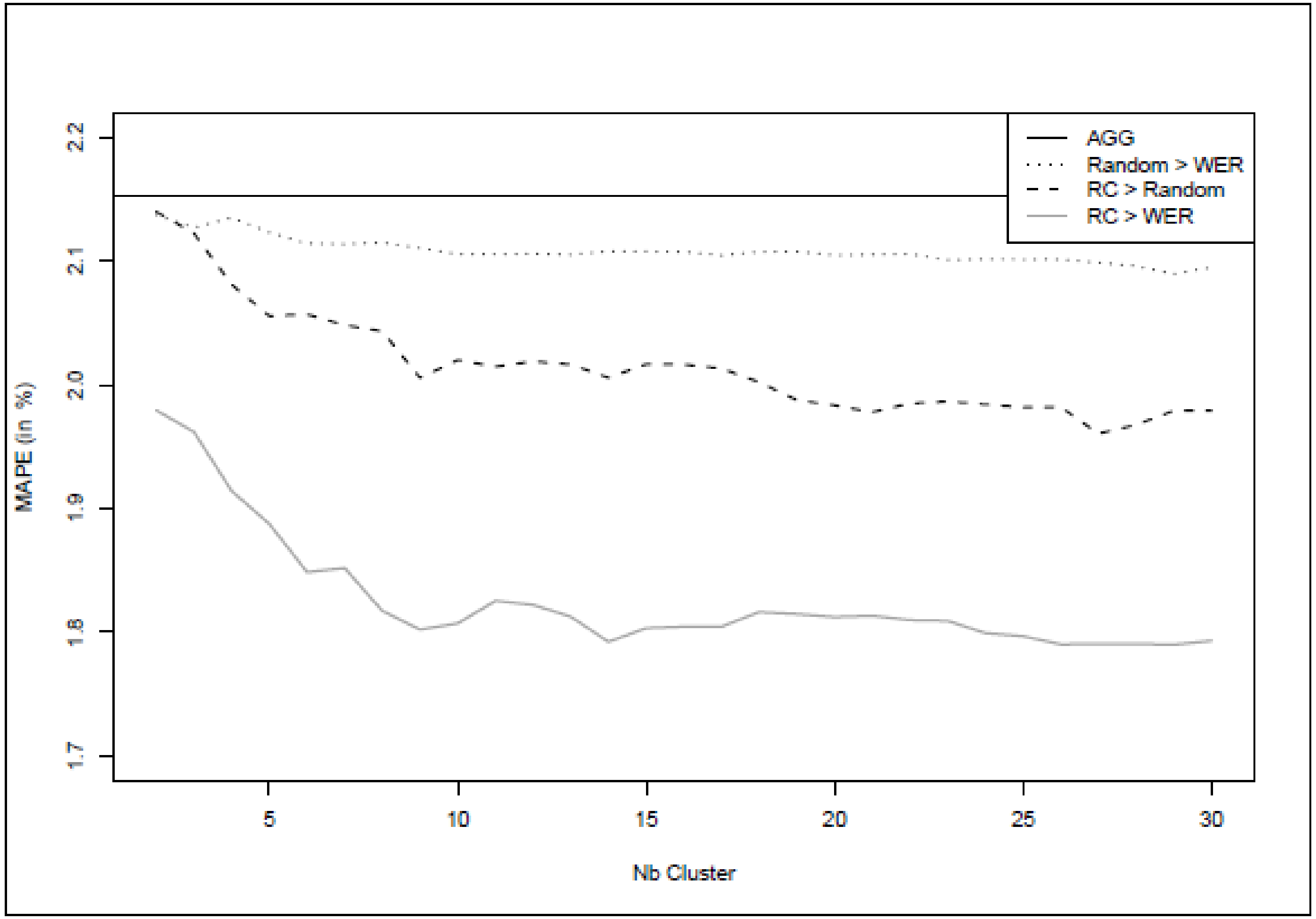
25k × 17k → (• wavelet dec. • **R**el.energy **C**ontribution) → 15 → features selection → 6 → CLARA-PAM (Clustering LaRge Applications Partitioning Arround Medoid) → *K'* = 200 super customers

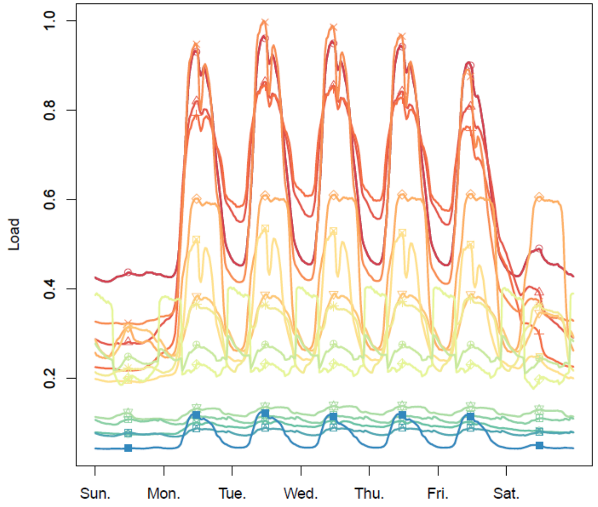
- 2nd stage: (Ward) ascendant hierarchical clustering of the K' super customers with WER (wavelet coherence) distance *coherent with the forecasting algorithm, computer intensive*

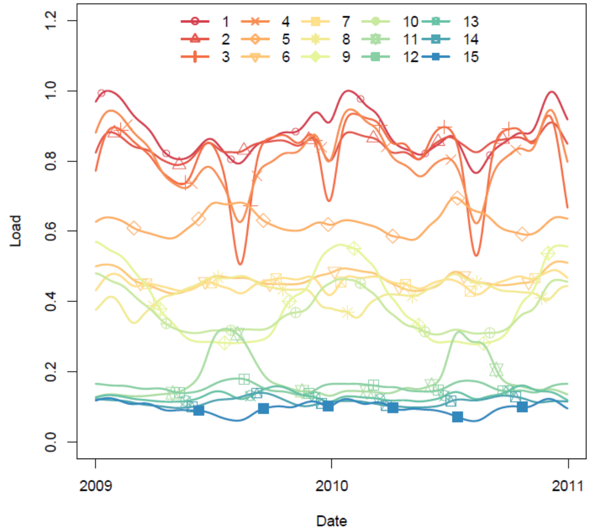
K' × 17k → WER → K' → PAM → K<K' clusters



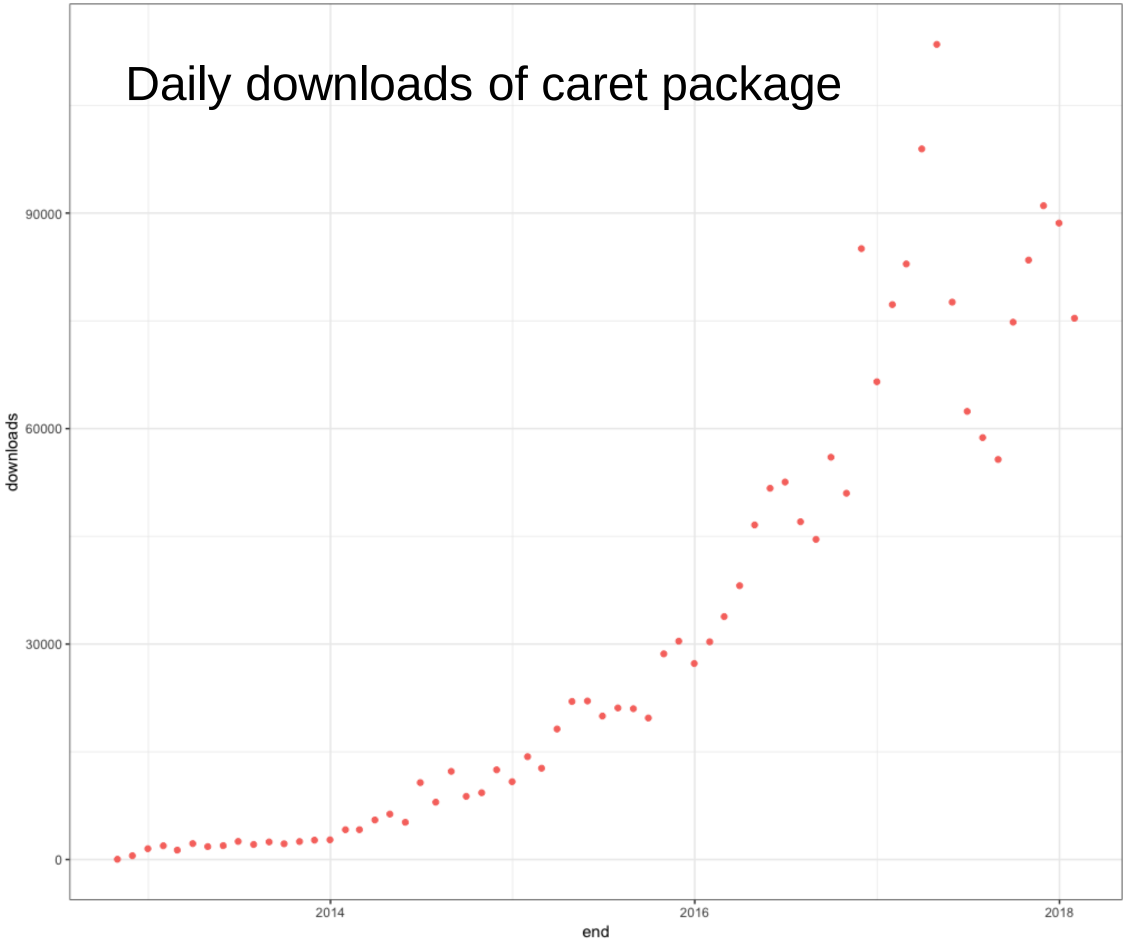
Forecasting total consomption of a set of customers (5)







Automatic calibration of machine learning algorithms



- a need for automatic calibration
- optimising both prediction performance and calculation time (*smart & data driven* grid search)

current work with Charles de Lastic Saint Jaal

Online robust aggregation algorithms (1)

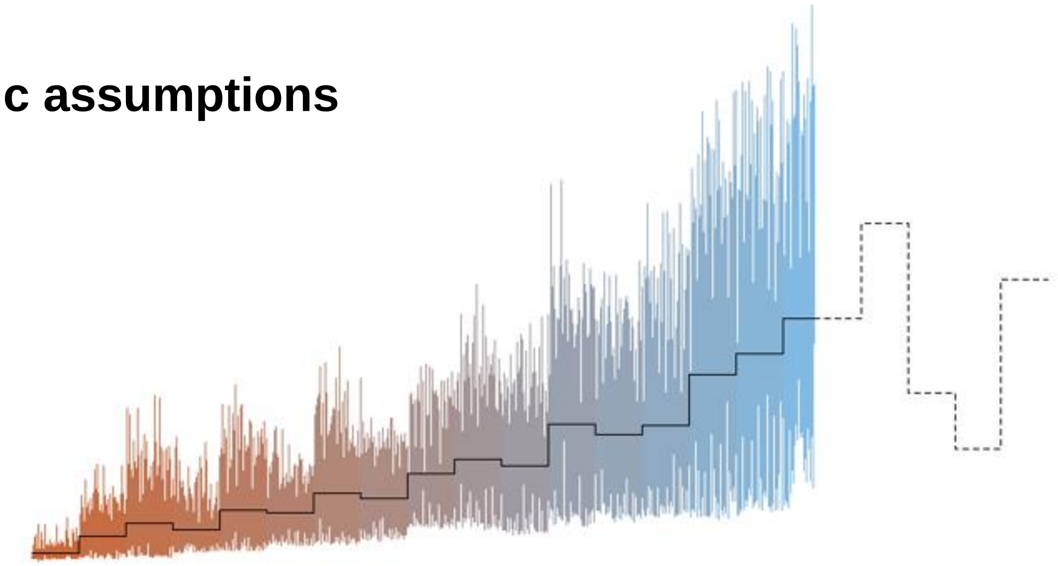
- We want to forecast a sequence of observati $y_1, y_2, \dots, y_T$

- Observations and predictions are made in a sequential fashion

  - predictions o y_t …

  … are based on past observations/predict $y_1, y_2, \dots, y_{t-1}$

- **No stochastic assumptions**



Join work with Pierre Gaillard (during his PhD at EDF R&D/Université Paris-Sud), Gilles Stoltz (CNRS-HEC Paris), Marie Devaine (Ecole Normale Supérieure, Paris, France)

Online robust aggregation algorithms (2)

- Linear

lasso, lars2, lars, enet, foba, icr, leapBackward, leapForward, leapSeq, lm, lmStepAIC, spikeslab, glm, BstLm, glm, glmboost, glmnet, glmStepAIC

- Generalised Additive Models

bagEarth, bagEarthGCV, bstTree, earth, gamLoess, gamSpline, gcvEarth

- Projection based

pcr, ppr, pls, plsRglm, simpls

```
modelList<-c("earth","ppr","gbm","xgbTree")
trControl<-trainControl("repeatedcv", repeats=1, number=5)
k<-1
train(x, y, method = modelList[[k]], trControl = trControl)
```

- Regression tree:

Gbm, blackboost, ctree, ctree2, rpart1SE, rpart2, treebag, xgbTree

- Kernel

Kernelpls, svmLinear, svmPoly, svmRadial, svmRadialSigma, svmRadialCost, knn, kknn

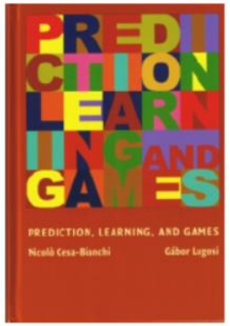
Online robust aggregation algorithms (3)

Loss of the expert j at time i

- **Parameters** $\eta > 0 \quad p_0 = (\frac{1}{N}, \ldots, \frac{1}{N})$

- **Weights update** $$p_{j,t} = \frac{\exp(-\eta \sum_{i=1}^{t-1} l_{i,j})}{C}$$

- **Oracle bounds** $$\frac{1}{T}\sum_{t=1}^{T} \hat{l}_t - \min_k \frac{1}{T}\sum_{t=1}^{T} \hat{l}_{t,k} \leq \square \sqrt{\frac{\log(N)}{T}}$$



Prediction, Learning, and Games
Nicolò Cesa-Bianchi et Gábor Lugosi

Online robust aggregation algorithms (4)

https://cran.rstudio.com/web/packages/opera/index.html

opera: Online Prediction by Expert Aggregation

Misc methods to form online predictions, for regression-oriented time-series, by combining a finite set of forecasts provided by the user.

Version:	1.0
Depends:	R ($\geq$ 3.1.0)
Imports:	quadprog, quantreg, RColorBrewer
Suggests:	testthat, splines, caret, mgcv, survival, knitr, gbm
Published:	2016-08-17
Author:	Pierre Gaillard [cre, aut], Yannig Goude [aut]
Maintainer:	Pierre Gaillard <pierre at gaillard.me>
BugReports:	https://github.com/dralliag/opera/issues
License:	LGPL-2 | LGPL-2.1 | LGPL-3 [expanded from: LGPL]
Copyright:	EDF R&D 2012-2015
URL:	http://pierre.gaillard.me/opera.html

DEMO

Perspectives

- Deep learning for forecasting (with D. Obst, S. Claudel, J. Cugliari and B. Ghattas)
- Random forest for time dependant data (with B. Goerhy, P. Massart and J.M. Poggi)
- Bandit algortihms for optimizing demand response (with M. Brégère, P. Gaillard and G. Stoltz)
- Hierarchical GAMs (with M. Fasiolo, R. Nédellec and S. Wood)
- Hierarchical Deep Learning Models for Forecasting (with M. Huard and G. Stoltz)

A few interesting data sets to test your model:

- Irish individual consumption data, http://www.ucd.ie/issda/data/commissionforenergyregulationcer/
- UK individual consumption data https://data.london.gov.uk/dataset/low-carbon-generators or https://www.kaggle.com/jeanmidev/smart-meters-in-london
- RE-Europe, a large-scale dataset for modeling a highly renewable European electricity systemTue V. Jensen & Pierre Pinson, Scientific Data volume 4, Article number: 170175 (2017), https://www.nature.com/articles/sdata2017175
- gefcom12 &14 https://www.kaggle.com/c/global-energy-forecasting-competition-2012-load-forecasting/data