Machine Learning Models for Energy Forecasting
(included probabilistic forecasts)
Yannig Goude
Joint work with...
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
• Demand response
• Sensors data
• Smart meters
GAMs (1)

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

\[ f_j(x_j) = \sum_{i=1}^{k} \beta_{ji} b_{ji}(x_j). \]
\[
\sum_{i=1}^{n} (y_i - \beta_0 x_i^0 - \sum_{q=1}^{p} f_q(x_i))^2 + \sum_{q=1}^{p} \lambda_q \int ||f_q''(x)||^2 dx
\]
**GAMs (3)**

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


Covariate selection with GAMs

Algorithm

1. **First step: subset selection (Group LASSO)**
   For each $\lambda_i \in \Lambda_{GrpL}$
   - Solve
     \[
     \hat{\beta}^{S_{\lambda_i}} = \text{arg min} \{Q^{OLS}(\beta) + \lambda_i \sum_{j=1}^{d} \sqrt{m_j} ||\beta_j||_2 \}
     \]
   - Denote $S^{S_{\lambda_i}} = \{j | \hat{\beta}_j^{S_{\lambda_i}} \neq 0\}$

2. **Second step: Estimation of the additive model (by OLS)**
   For each support set $S_{\lambda_i} \in \{S_{\lambda_{\text{min}}}, \ldots, S_{\lambda_{\text{max}}} \}$
   - Compute
     \[
     Q^{OLS}_{S_{\lambda_i}}(\beta) = \sum_{i=1}^{n} \left( Y_i - \beta_0 - \sum_{j \in S_{\lambda_i}} C_{ij} (\beta_j) \right)^2
     \]
   - Solve
     \[
     \hat{\beta}^{S_{\lambda_i}} = \text{arg min} \{Q^{OLS}_{S_{\lambda_i}}(\beta)\}
     \]
   - Compute the BIC (see Eq. (5)) for each $\hat{\beta}^{S_{\lambda_i}}$

3. **Third step: Selection of the final model** Select $\hat{\beta}^{S_{\lambda_{\text{opt}}}}$ which minimizes the BIC

<table>
<thead>
<tr>
<th>Criterion</th>
<th>MAPE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Post2Bic&gt;0.2</td>
<td>1.12</td>
<td>645</td>
</tr>
<tr>
<td>Post2Gcv&gt;0.3</td>
<td>1.15</td>
<td>648</td>
</tr>
<tr>
<td>Post2Aic</td>
<td>1.17</td>
<td>663</td>
</tr>
<tr>
<td>Post2Gcv</td>
<td>1.17</td>
<td>667</td>
</tr>
<tr>
<td>EDF model</td>
<td>1.16</td>
<td>667</td>
</tr>
<tr>
<td>Post2Bic</td>
<td>1.24</td>
<td>730</td>
</tr>
<tr>
<td>BenchMT1</td>
<td>2.00</td>
<td>1173</td>
</tr>
</tbody>
</table>

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


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)

\[ q_\tau(Y|X) = F_{Y|X}^{-1}(\tau) = \inf \{ y \in \mathbb{R}, F_{Y|X}(y) \geq \tau \} \]

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

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


https://cran.r-project.org/web/packages/qgam/index.html
qGAM (2)

The plots show the performance of different models on the GEFCom2014 and UK grid datasets.

The Relative loss plots compare the performance of Gaussian gam, quantGAM, Calibrated ELF, and Gradient Boosting. The y-axis represents the relative loss, and the x-axis represents the quantile \( \tau \).

The Effect plots illustrate the effect of different quantiles on the models. The y-axis represents the effect, and the x-axis represents the time period (\( T_t^a \) or \( S_t \)). The plots show how the effect changes with different quantiles.
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
Forecasting total consumption of a set of customers (1)

Individual consumption metered half-hourly

Forecasts total cons. of each cluster

Total Portfolio forecast

Disaggregated Electricity Forecasting using Wavelet-Based Clustering of Individual Consumers, Proceedings of IEEE EnergyCon, 2016, Jairo Cugliari, Yannig Goude, Jean-Michel Poggi
Forecasting total consumption of a set of customers (2)

What type of customers in each cluster?
Do they behave similarly?
Are they complementary?

How many (at least) customers in each cluster?

Which forecasting model, clustering algorithm?
Are they related in any sense?


How many clusters?
Forecasting total consumption 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 consumption of a set of customers (4)

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

- 2nd stage: (Ward) ascendant hierarchical clustering of the $K'$ super customers with WER (wavelet coherence) distance *coherent with the forecasting algorithm, computer intensive*
Forecasting total consumption 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 observations \( y_1, y_2, \ldots, y_T \)

- Observations and predictions are made in a sequential fashion
  - predictions of \( y_t \) \( \ldots \)
    \( \ldots \) are based on past observations/predictions \( y_1, y_2, \ldots, y_{t-1} \)

- No stochastic assumptions
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

- **Regression tree:**
  - Gbm, blackboost, ctree, ctree2, rpart1SE, rpart2, treebag, xgbTree

- **Kernel**
  - Kernelpls, svmLinear, svmPoly, svmRadial, svmRadialSigma, svmRadialCost, knn, kknn
Online robust aggregation algorithms (3)

• Parameters

\[ \eta > 0 \quad p_0 = \left( \frac{1}{N}, \ldots, \frac{1}{N} \right) \]

• Weights update

\[ p_{j,t} = \frac{\exp(-\eta \sum_{i=1}^{t-1} l_{i,j})}{C} \]

• Oracle bounds

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

Loss of the expert j at time i
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 (>= 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/dralliaag/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
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 algorithms for optimizing demand response (with M. Brégère, P. Gaillard and G. Stoltz)
- Hierarchical GAMs (with M. Fasiolo, R. Néedellec 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/
- RE-Europe, a large-scale dataset for modeling a highly renewable European electricity system Tue V. Jensen & Pierre Pinson, Scientific Data volume 4, Article number: 170175 (2017), https://www.nature.com/articles/sdata2017175