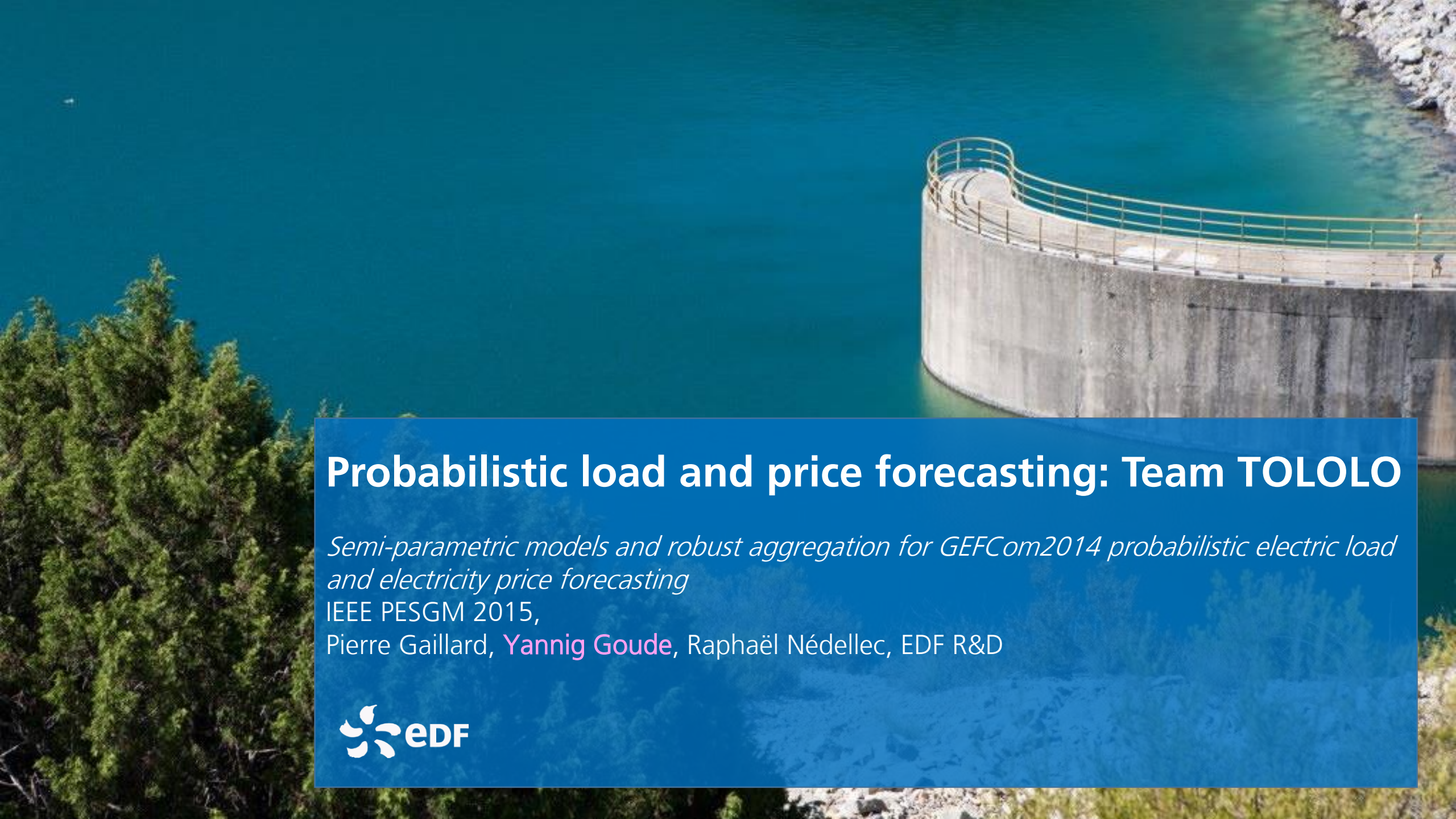


Probabilistic load and price forecasting: Team TOLOLO

Semi-parametric models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting

IEEE PESGM 2015,

Pierre Gaillard, Yannig Goude, Raphaël Nédellec, EDF R&D

Summary

- Generalized Additive Models and Quantile Regression
- QuantGAM
- Application to load and price forecasting
- Other approches: *combining experts* and *quantGLM*
- Conclusion

GAM and quantile regression

Statistical background

Generalized Additive Models

- Generalized Additive Models: generalized version of GLM, these are regression models.
- Introduced by [Hastie and Tibshirani 1990] , improved in [Wood, 2006]
- Well known methods for load forecasting. See [Pierrot and Goude 2011], [Fan and Hyndman 2012], [Goude, Nédellec and Kong 2013]
- Already used in GEFCOM2012 by Team Tololo, see [Nédellec, Cugliari and Goude 2014]
- Easy interpretation, fast computation, adaptive to different datasets

Generalized Additive Models

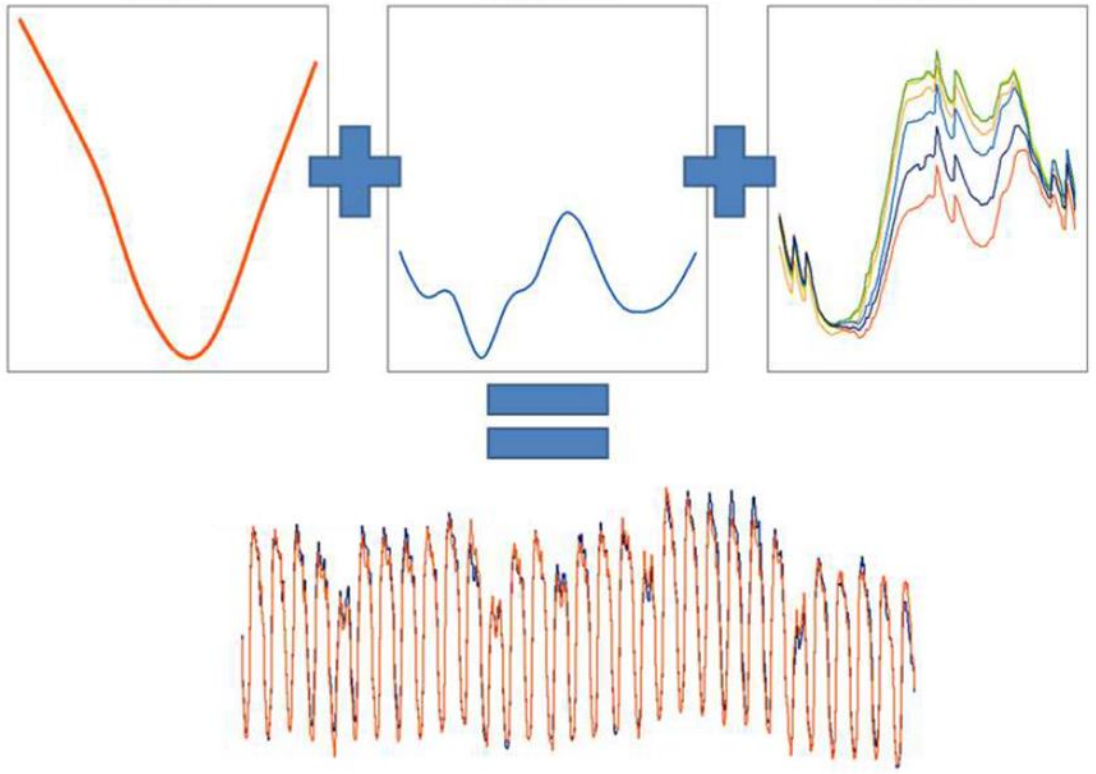
- Regression Models
- Allow non-linearity
- GAMs can be written as:

$$g(\mu_i) = f_1(X_{1,i}) + f_2(X_{2,i}, X_{3,i}) + \dots$$

where $\mu_i \triangleq \mathbb{E}(Y_i)$ and $Y_i \sim$ an exponential family.

- f_1, f_2, are estimated using penalized splines

$$y_t = f_1(T_t) + f_2(I_t) + f_3(H_t) + \varepsilon_t$$

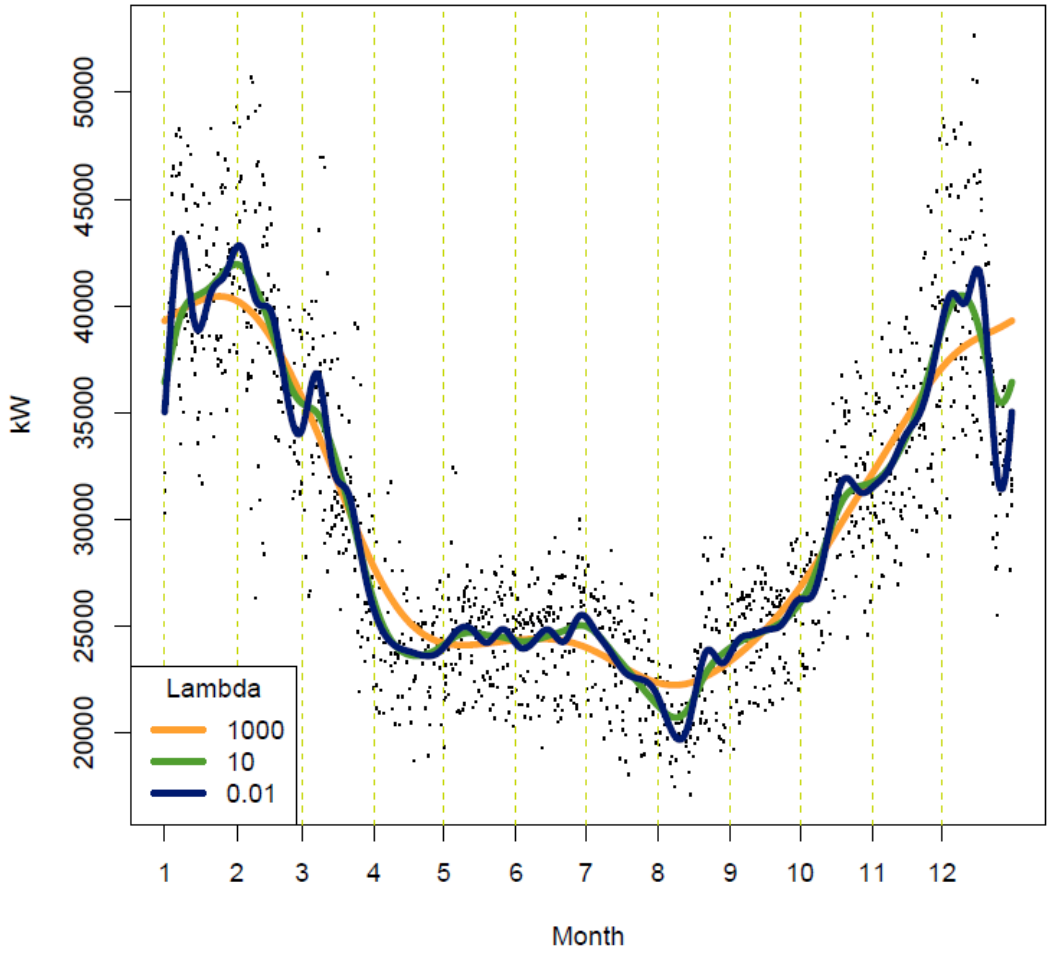


Generalized Additive Models

trade-off between fitting to your data and smoothness of the functions

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

Estimated using numerical optimisation of the GCV criteria



Quantile Regression

- Introduced by [Koenker et al., 1978] in a linear setting,

well described in [Koenker, 2005]

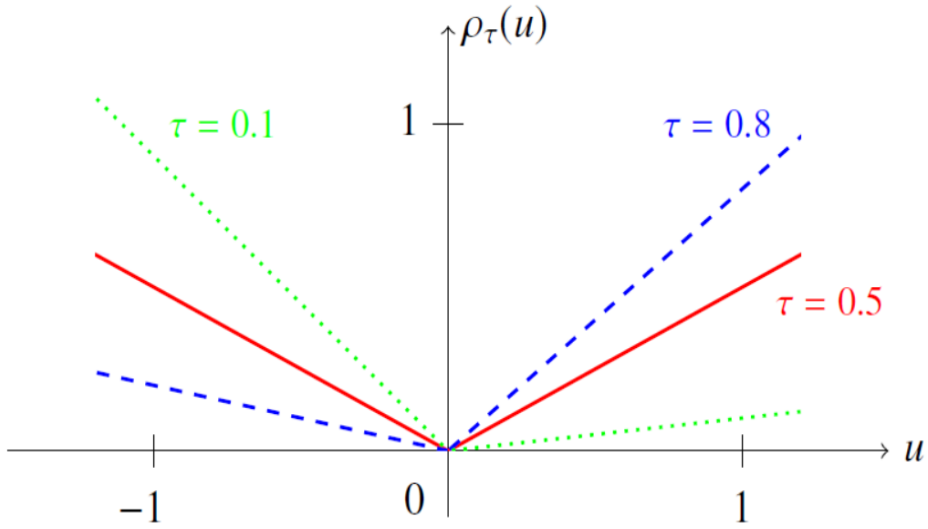
$$q_\tau(Y|X) = F^{-1}_{Y|X}(\tau) = \inf\{y \in \mathbb{R}, F_{Y|X}(y) \geq \tau\}$$

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

$$p_\tau(u) = \tau(u - \mathbb{I}(y < 0))$$

Ex. here

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



With quantreg package, we faced numerical issues in a non-linear regression setup, we propose our own solution: QuantGAM

QuantGAM

Combine GAM and linear quantile regression

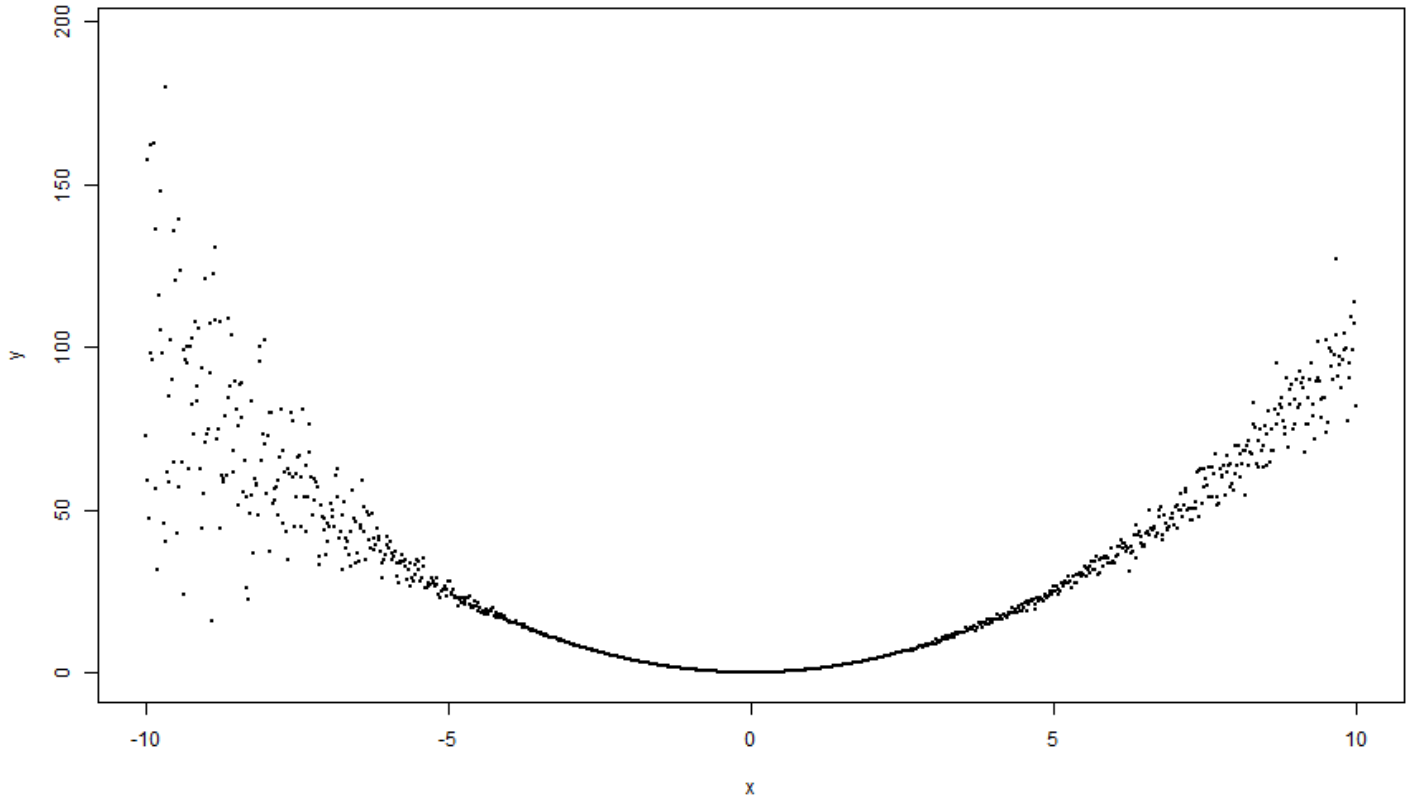
QuantGAM: a two-step approach

- Relationship is non-linear between y and x
- Dispertion depends on x

$$y = f(x) + \sigma(x) * \varepsilon$$

$$\varepsilon \sim N(0, 1)$$

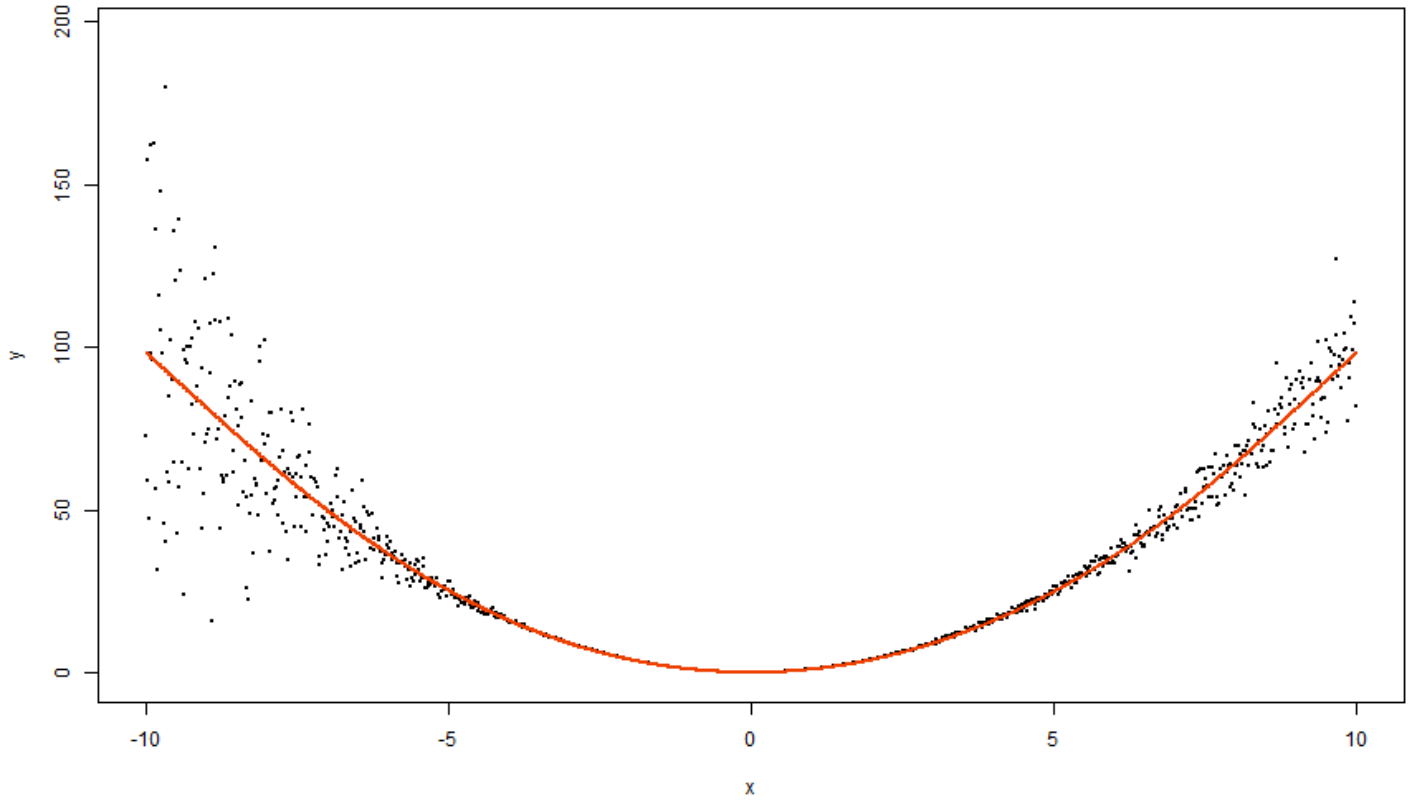
$$\sigma(x) = \begin{cases} x^3/10, & \text{if } x > 0 \\ -x^5/200, & \text{if } x \leq 0 \end{cases}$$



QuantGAM, step 1

- Conditional mean estimation with GAM.

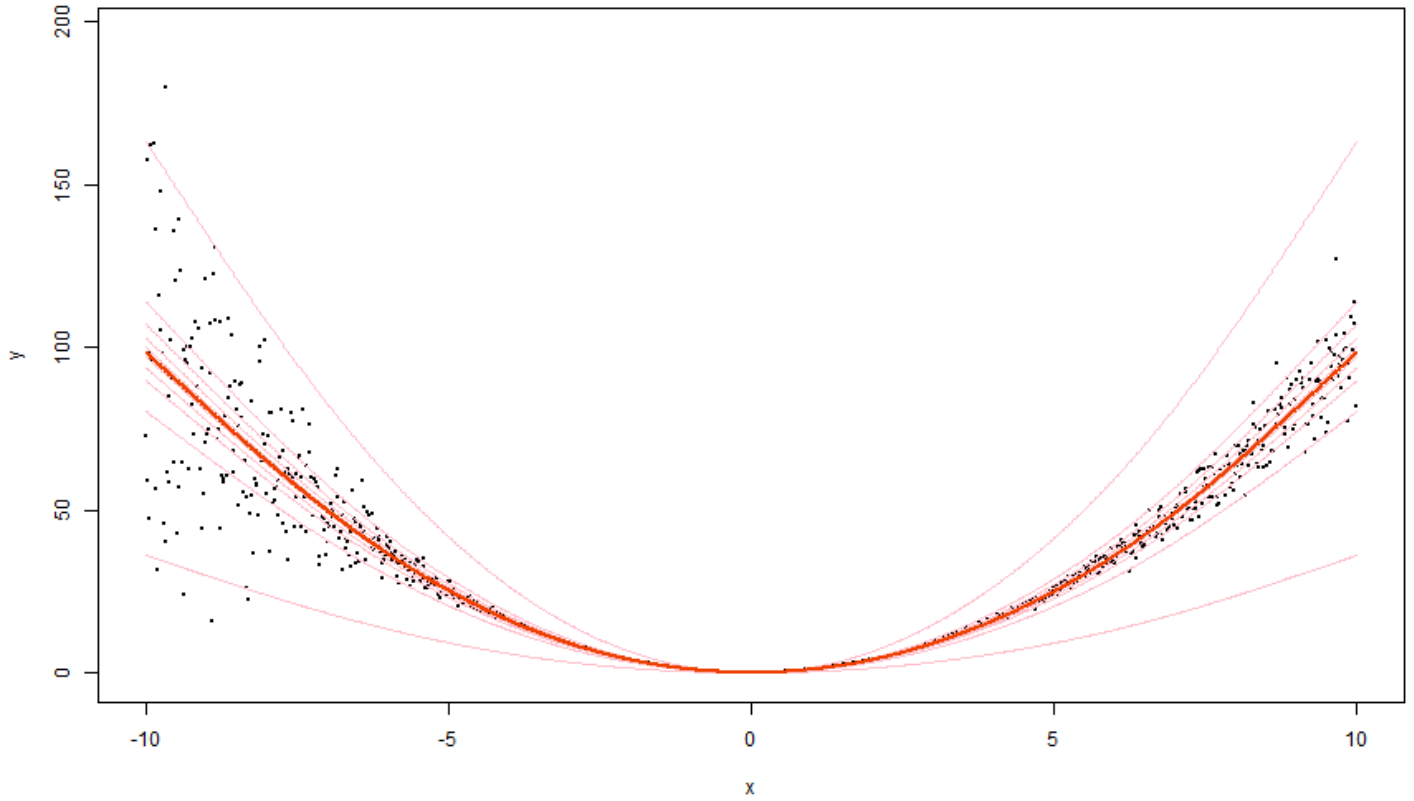
$$\hat{f} = \text{argmin}\, ||y - f(x)||^2 + \lambda\, \text{Pen}(f)$$



QuantGAM, step2

- Quantiles estimation with linear quantile regression.

$$\hat{f}_\tau = \underset{a}{\operatorname{argmin}}\, \rho_\tau(y - a\hat{f}(x))$$

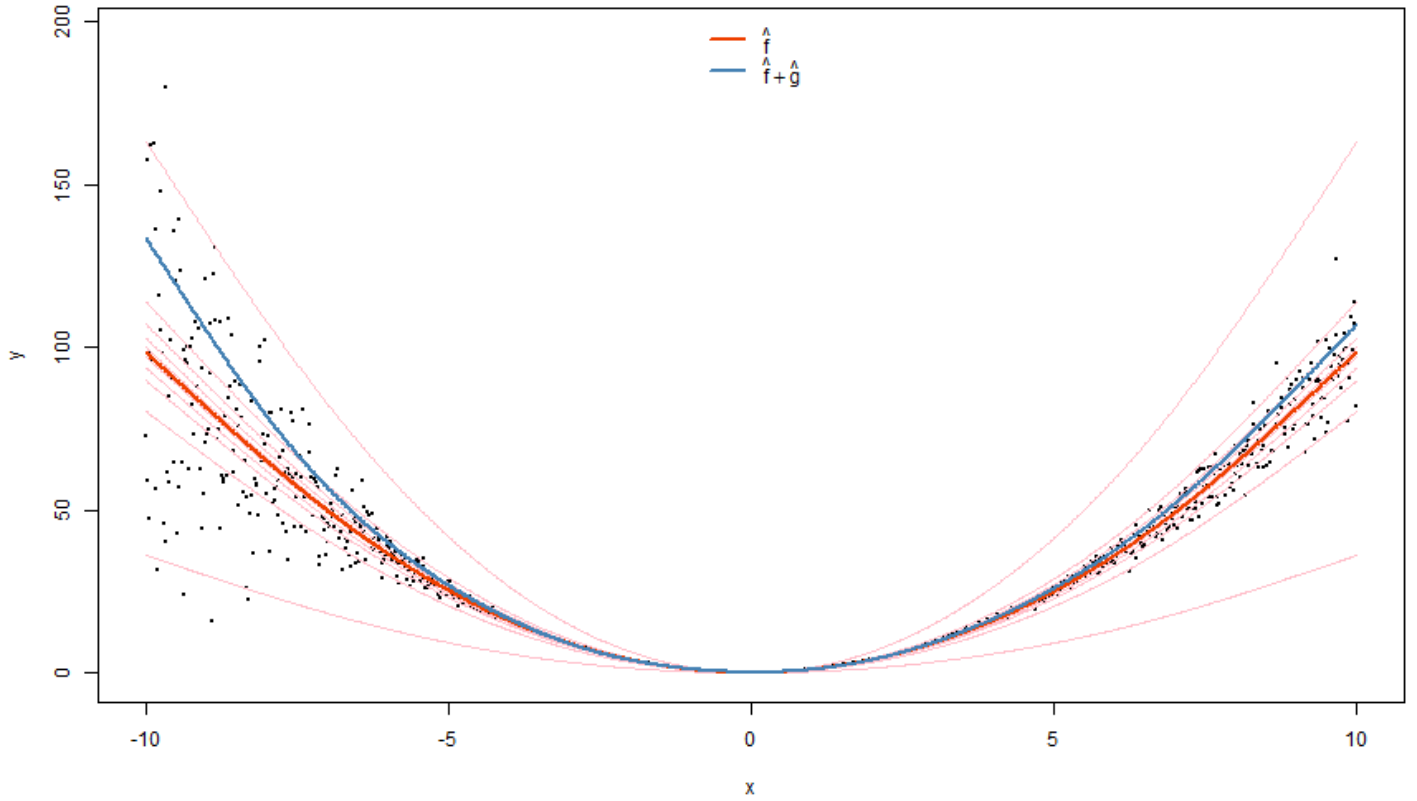


QuantGAM, 2nd order improvement, 1

- Fit a gam on the square residuals to estimate the conditional variance

$$\hat{\varepsilon} = y - \hat{f}(x)$$

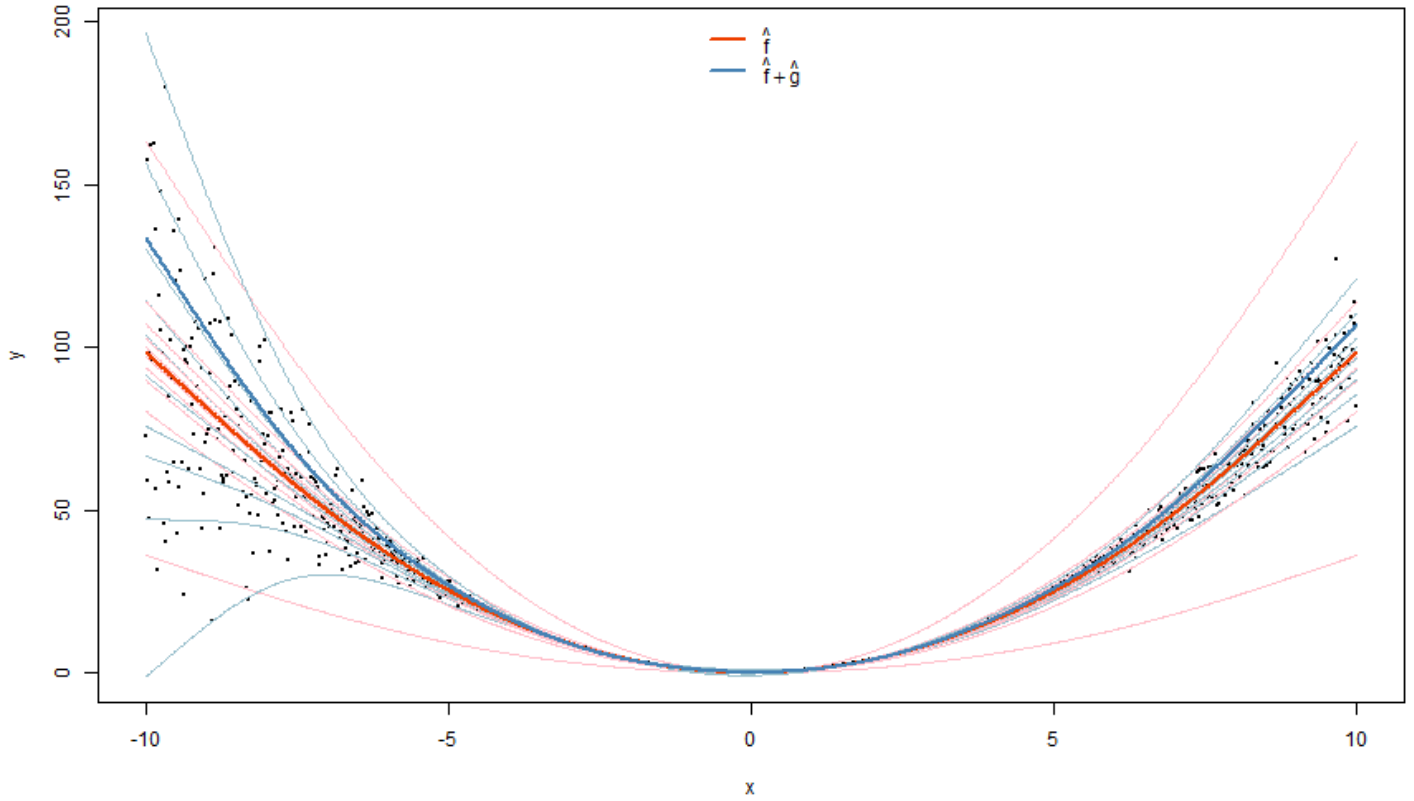
$$\hat{g} = \text{argmin}\, ||\hat{\varepsilon} - g(x)||^2 + \lambda\, \text{Pen}(g)$$



QuantGAM, 2nd order improvement, 2

- Quantiles estimation with linear quantile regression on the conditional mean and variance features.

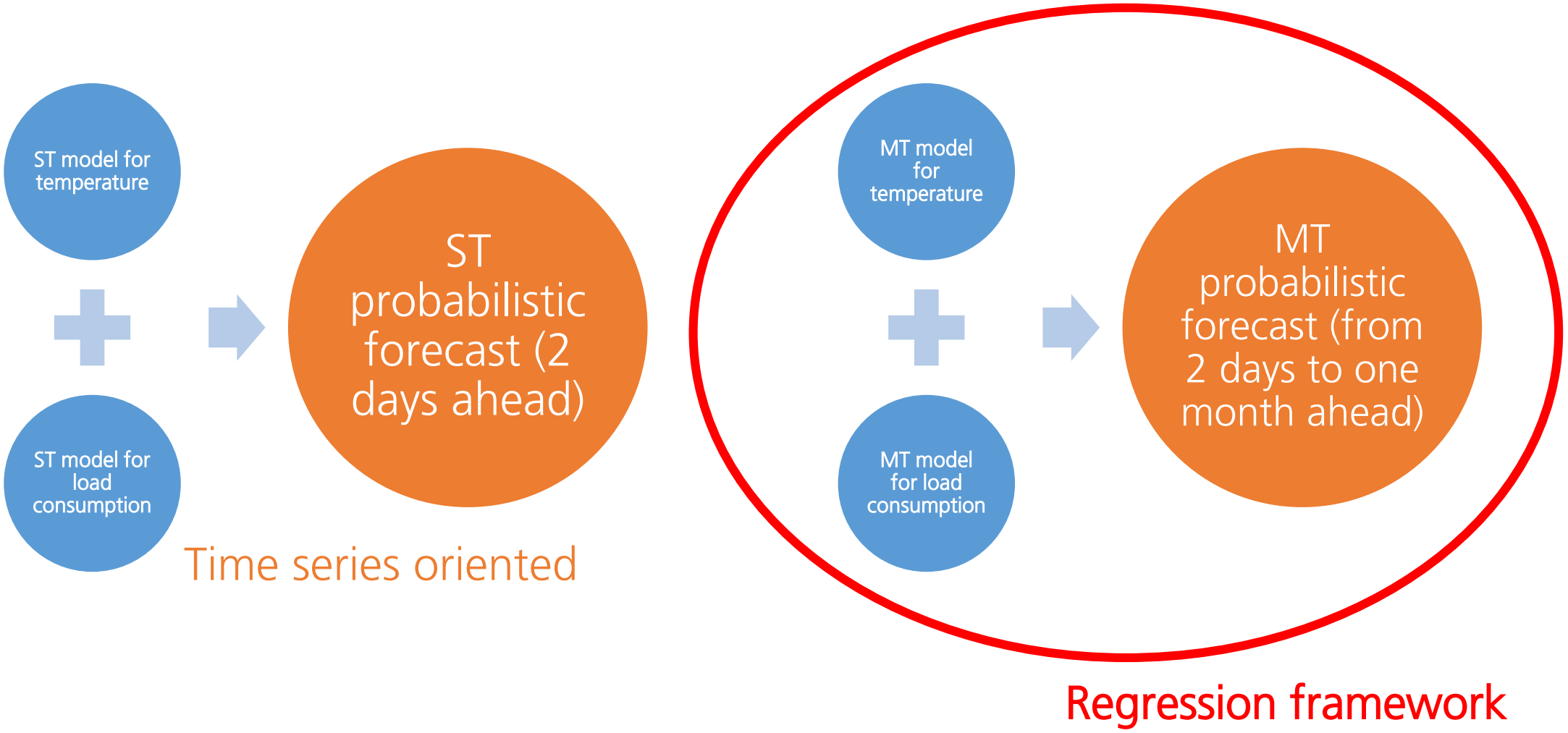
$$\hat{g}_\tau = \underset{a,b}{\operatorname{argmin}}\, \rho_\tau(y - a\hat{f}(x) - b\hat{g}(x))$$



Applications

Load and price forecasting tracks

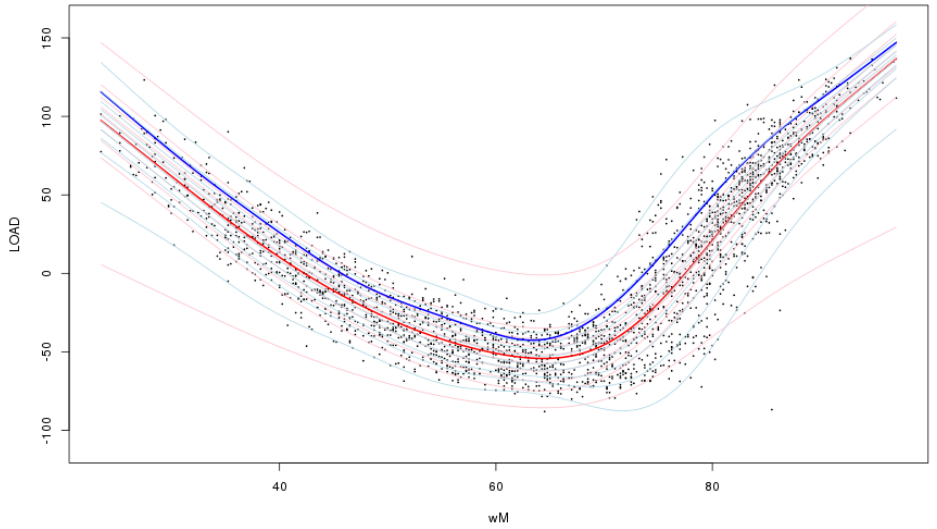
Probabilistic load forecasting workflow



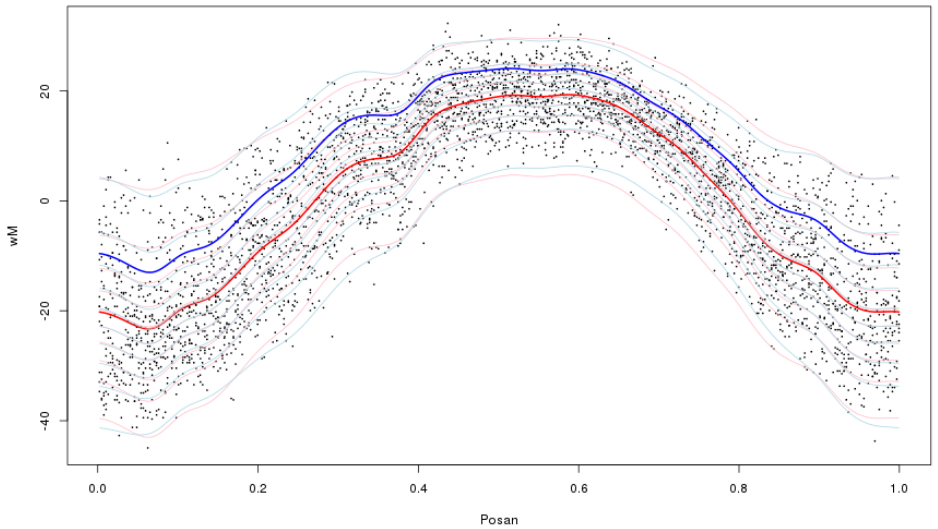
Focus on MT probabilistic forecast

Application of quantGAM for MT forecasting:
For temperature and load forecasts.

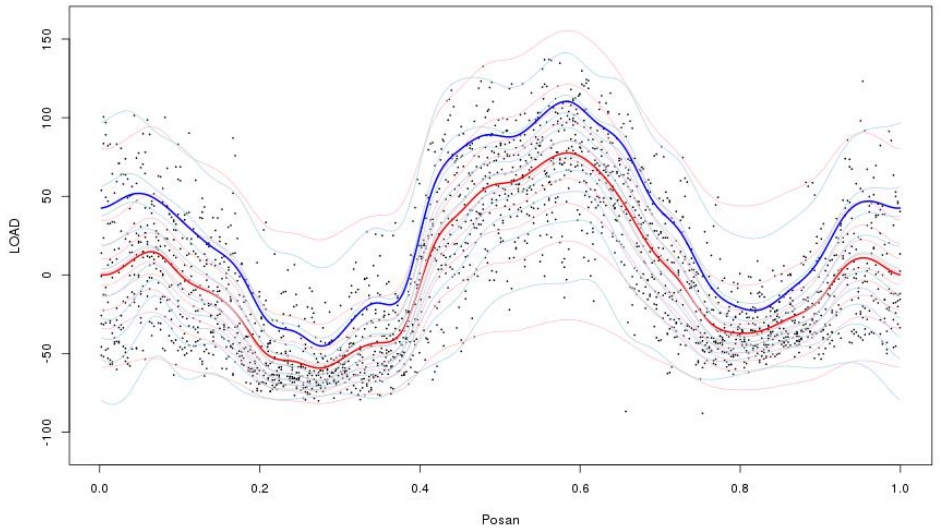
Load: Temperature effect



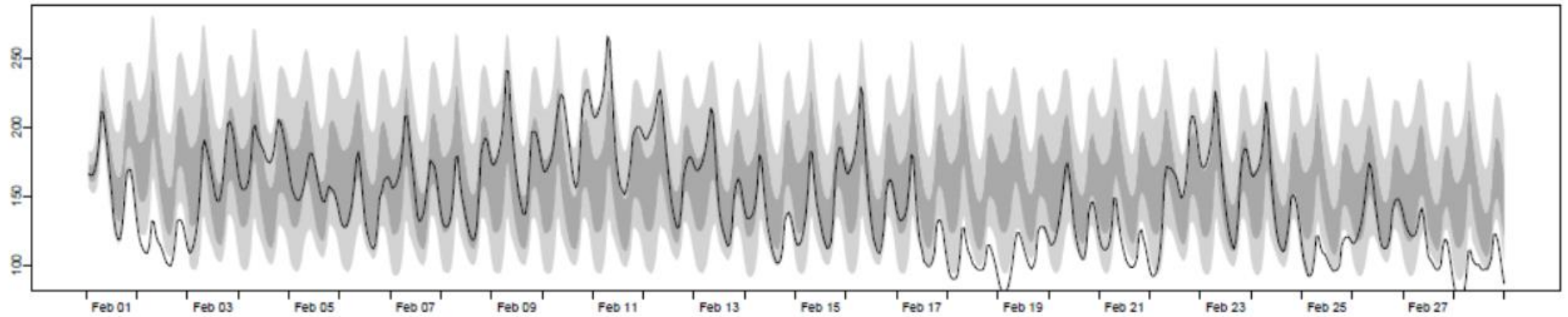
Temperature: Toy effect



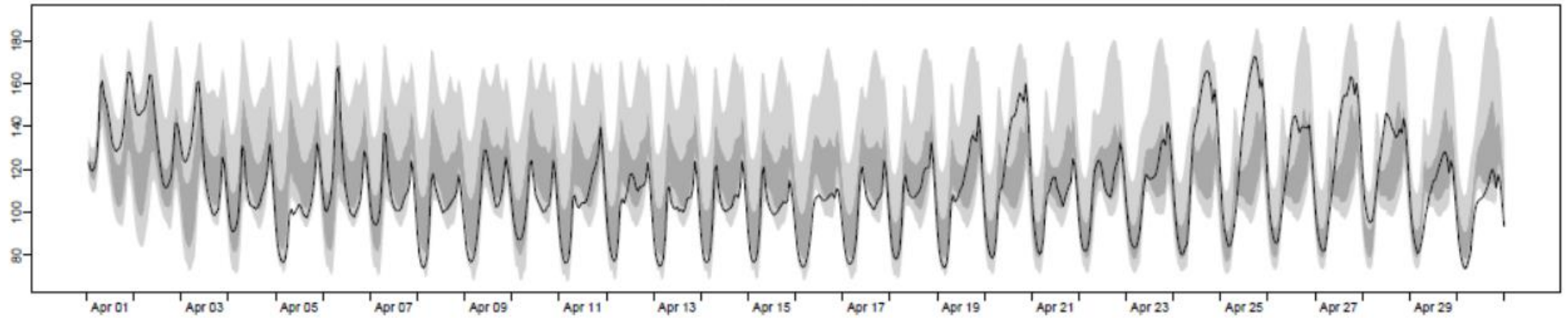
Load: Toy effect



Some results

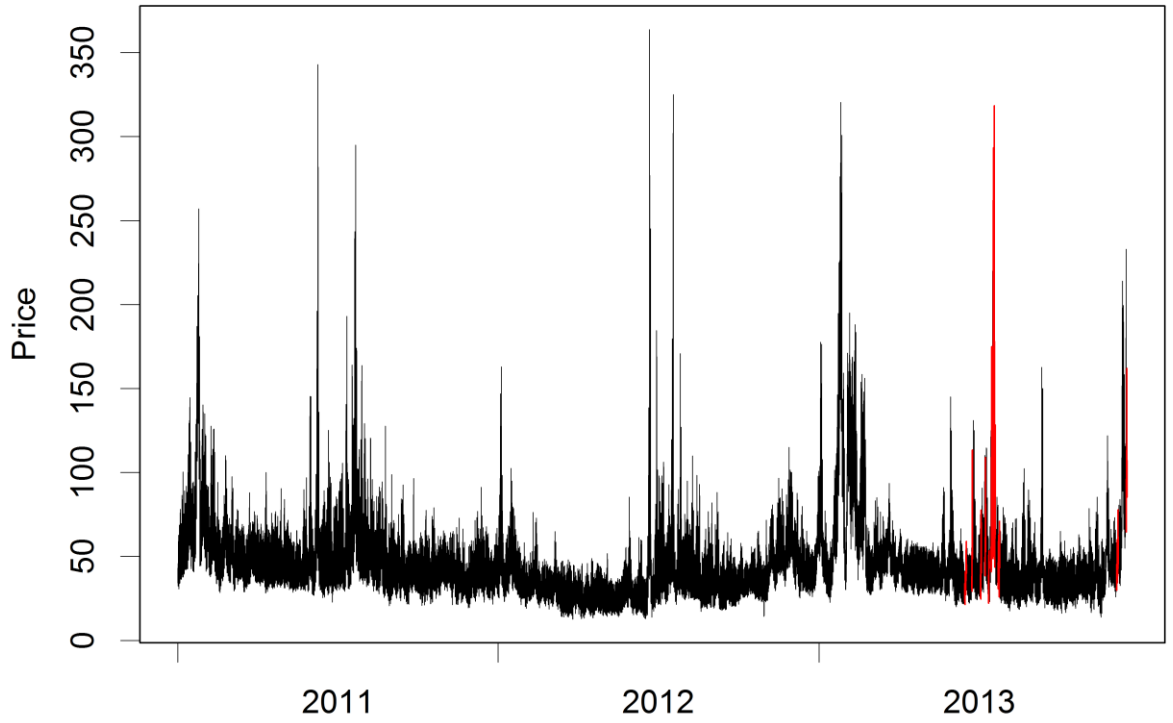


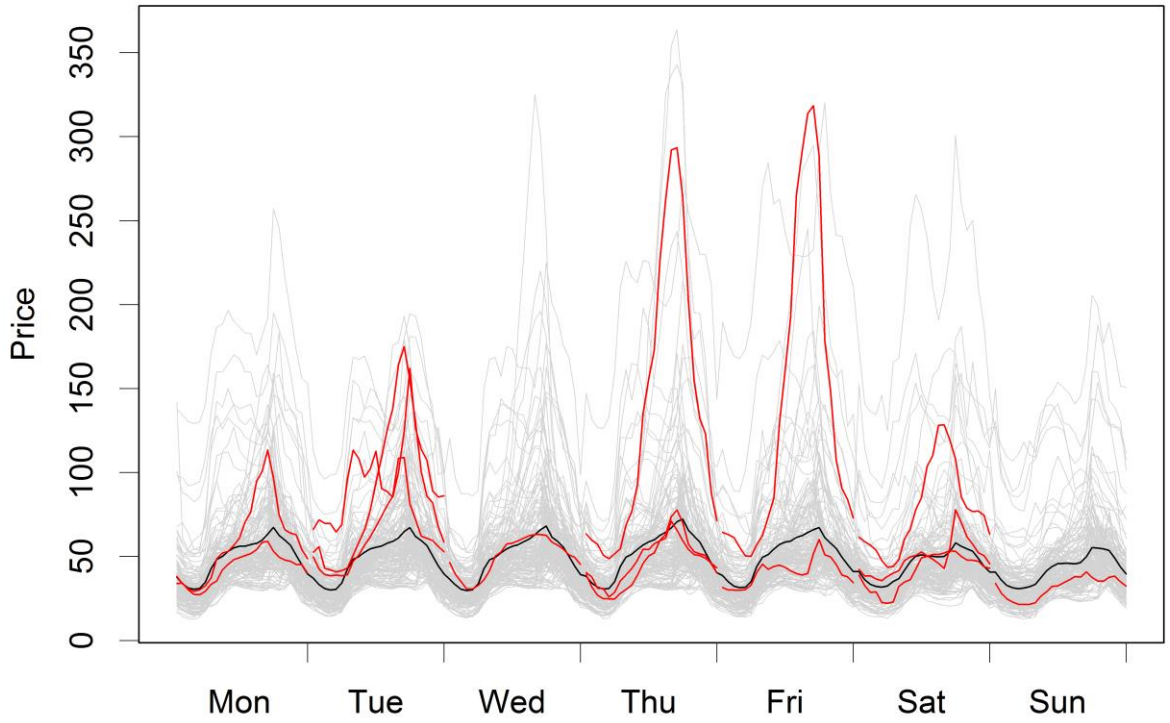
(a) February



(b) April

Probabilistic price forecasting



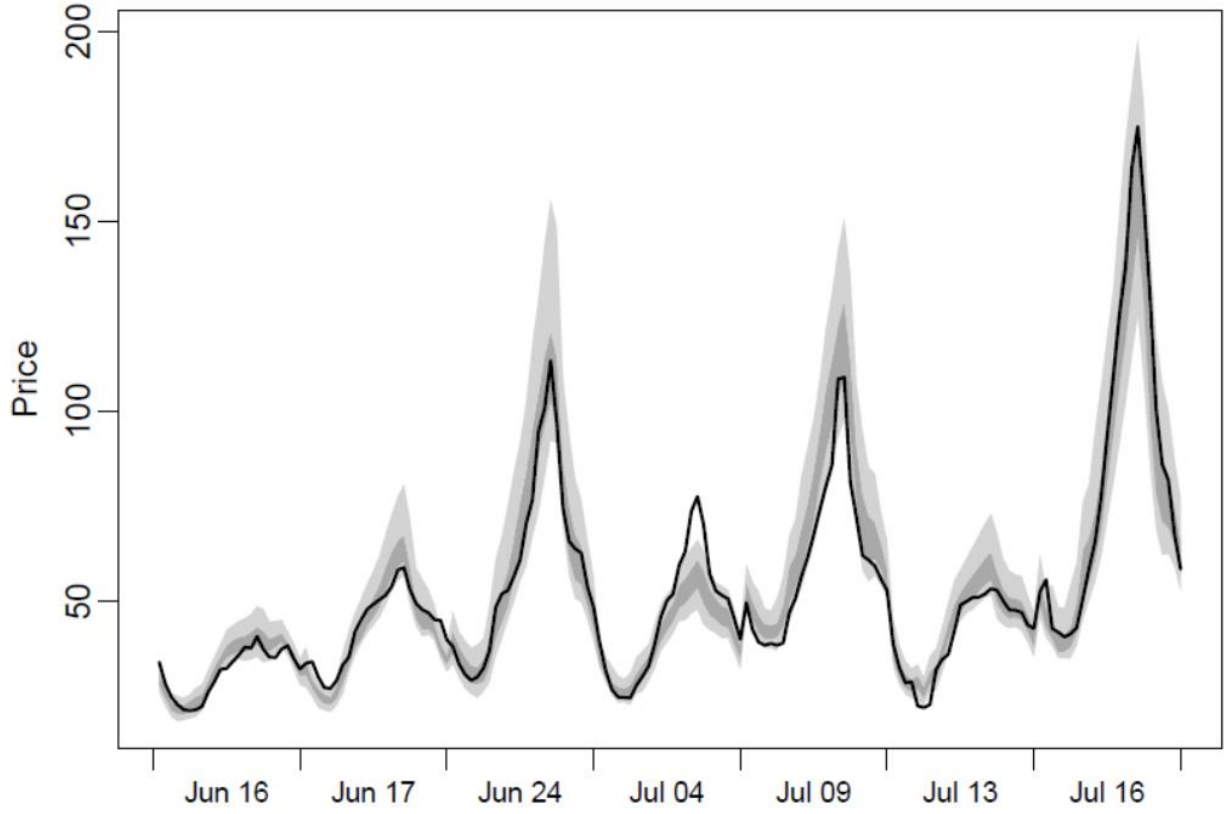


Probabilistic price forecasting

Model maximum, then price with quantGAM

$$\log(P_t^{(\max)}) = \alpha_1 \log(P_{t-24}^{(\max)}) + \alpha_2 \log(P_{t-48}^{(\max)}) + \alpha_3 \log(P_{t-24}^{(\text{mean})}) + \alpha_4 \log(P_{t-48}^{(\text{mean})}) + f_1(\log(FTL_{t-24}^{(\text{mean})})) + f_2(\log(FTL_t^{(\text{mean})})) + f_3(\log(FZL_t^{(\max)})) + f_4(FZL_{t-24}^{(\max)}) + \varepsilon_t ,$$

$$\boldsymbol{Z}_t = \Big(\log(P_t^{(\text{last})}),\ \log(P_t^{(\max)}),\ \log(P_{t-24}),\ \log(P_{t-48}),\ \log(P_{t-168}),\ \log(P_{t-24}^{(\min)}),\ DayType_t,\ FZL_t^{(0.95)},\ FTL_t^{(0.95)},\ FZL_t^{(0.8)},\ FTL_t^{(0.8)}\Big).$$

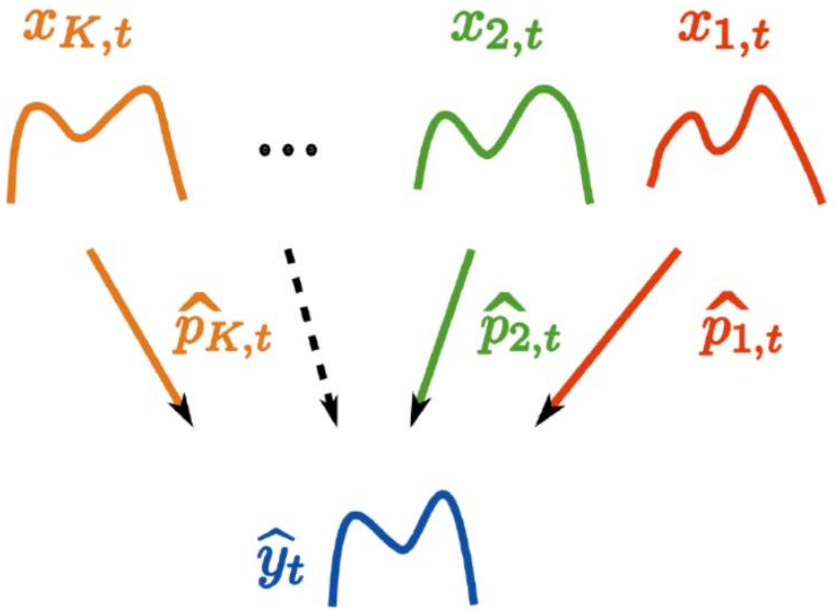


Other methods

Others methods have been used, especially for price forecasting:

- Robust expert aggregation
- *quantGLM*

See [Gaillard, Goude, Nédellec]



Robust Sequential Expert Aggregation

- More and more models/tools
  - R: 207 regression models in caret (boosting, neural nets, trees, bagging, random forest, kernel regression, splines, deep neural networks…)
  - Python: scikit-learn (ridge, elastic net, lasso, kernel ridge regression, orth. matching pursuit, support vector regression, stochastic gradient descent, nearest neightbors, ensemble methods, feature selection…)

  …..

- More and more data scientist/ statisticiens

- A few examples:

  - Netflix price

Sill, J.; Takács, Gá.; Mackey, L. & Lin, D. Feature-Weighted Linear Stacking, CoRR (Computing Research Repository), 2009, abs/0911.0460

Jahrer, M.; Töscher, A. & Legenstein, R. Combining Predictions for Accurate Recommender Systems Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 2010, 693-702

Töscher, A.; Jahrer, M. & Bell, R. M. The BigChaos Solution to the Netflix Grand Prize 2009

*"During the nearly 3 years of the Netifx competition, there were two main factors which improved the overall accuracy: The quality of the **individual algorithms** and the **ensemble idea**."*

  - Kaggle competitions

*"This is how you win ML competitions: you take other peoples' work and **ensemble** them together."* Vitaly Kuznetsov NIPS2014

EDF

Robust Sequential Expert Aggregation

Exponentially Weighted Aggregation

- Parameters $\eta > 0$ $\quad \hat{\mathbf{p}}_0 = (1/K, \ldots, 1/K)$
- Weights update

$$\hat{p}_{k,t} = \frac{\exp\left(-\eta \sum_{s=1}^{t-1} \ell_{k,t}\right)}{\sum_{j=1}^{K} \exp\left(-\eta \sum_{s=1}^{t-1} \ell_{j,t}\right)}$$

- Oracle bounds

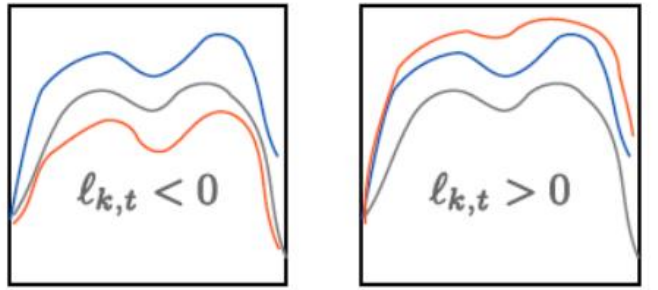
$$\mathrm{Reg}_T \overset{\mathrm{def}}{=} \frac{1}{T}\sum_{t=1}^{T} \hat{\ell}_t - \min_k \frac{1}{T}\sum_{t=1}^{T} \ell_{k,t} \leqslant \square\sqrt{\frac{\log K}{T}} \to 0$$

EDF

Robust Sequential Expert Aggregation

Exponential Gradient

- Parameters $\eta > 0$ $\quad \hat{\mathbf{p}}_0 = (1/K, \ldots, 1/K)$



- update $\quad \hat{p}_{i,t} \propto exp\left(-\eta \sum_{s=1}^{t-1} \ell_{i,s}\right) \qquad \ell_{i,s} = 2(\hat{y}_s - y_s)x_{i,s}$

- Oracle bounds

$$\underbrace{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}_{\text{Our loss}} \leqslant \underbrace{\min_{q \in \Delta_N} \sum_{t=1}^{T}(q \cdot x_t - y_t)^2}_{\text{Loss of the best convexe combination}} + \underbrace{\square\sqrt{T \log N}}_{\text{Estimation error}}$$

Robust Sequential Expert Aggregation

ML-poly

- Parameters

$$\hat{\mathbf{p}}_0 = (1/K, \ldots, 1/K)$$

- More weights to an expert with high regret
- Focus on worst cases for the aggregation: cases with positive regret
- Confidency to « stable » expert (low error variance)

- update

$$\hat{p}_{k,t} = \frac{\eta_{k,t}\left(\sum_{s=1}^{t-1}(\hat{\ell}_s - \ell_{k,s})\right)_+}{\sum_{j=1}^{K}\eta_{j,t}\left(\sum_{s=1}^{t-1}(\hat{\ell}_s - \ell_{j,s})\right)_+}$$

$$\eta_{k,t} = \left(1 + \sum_{s=1}^{t-1}(\hat{\ell}_t - \ell_{k,t})^2\right)^{-1}$$

- Oracle bounds

$$\frac{1}{T}\sum_{t=1}^{T}\hat{\ell}_t \leqslant \min_{k=1,\ldots,K}\left\{\frac{1}{T}\sum_{t=1}^{T}\ell_{k,t} + \frac{\square}{T}\sqrt{K\sum_{t=1}^{T}(\hat{\ell}_t - \ell_{k,t})^2(\log\log T)}\right\}$$

Experts

$$\log(P_t) = \alpha_1 \log(P_{t-24}) + \alpha_2 \log(P_{t-48}) + \alpha_3 \log(P_{t-168}) + \alpha_4 \log(P_{t-24}^{(\min)}) + h(DayType_t) + \varepsilon_t ,$$

AR

$$\log(P_t) = \alpha_1 \log(P_{t-24}) + \alpha_2 \log(P_{t-48}) + \alpha_3 \log(P_{t-168}) + \alpha_4 \log(P_{t-24}^{(\min)}) + \alpha_5 \log(FTL_t) + \alpha_6 \log(FZL_t) + h(DayType_t) + \varepsilon_t .$$

ARX

$$\log(P_t) = f_1(\log(P_{t-24})) + f_2(\log(P_t^{(\max)})) + f_3(\log(P_{t-24}^{(\mathrm{mean})})) + f_4(FTL_t) + f_5(FZL_t) + f_6(ToY_t) + h(DayType_t) + \varepsilon_t .$$

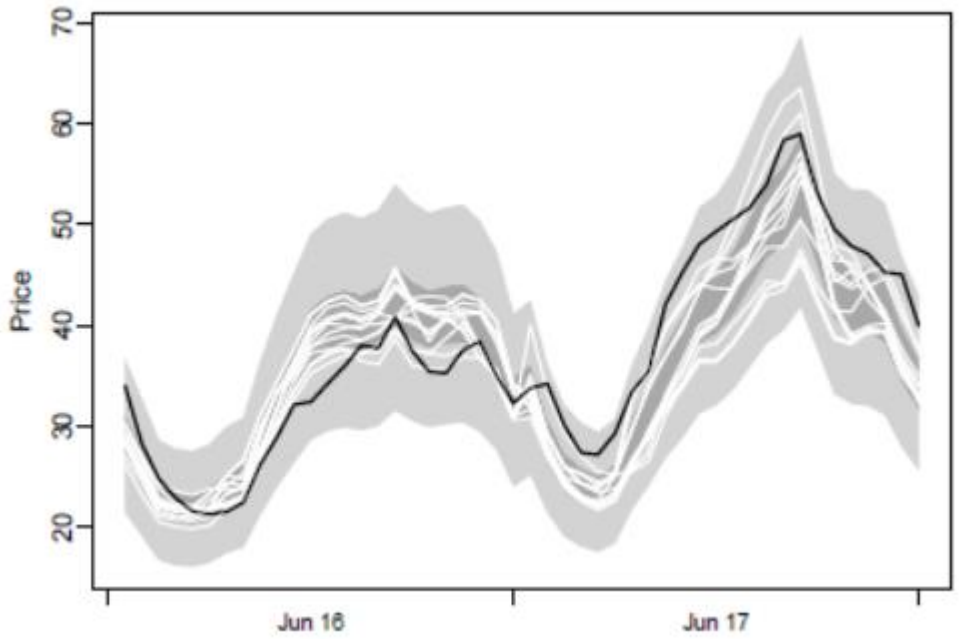
GAM

But also:

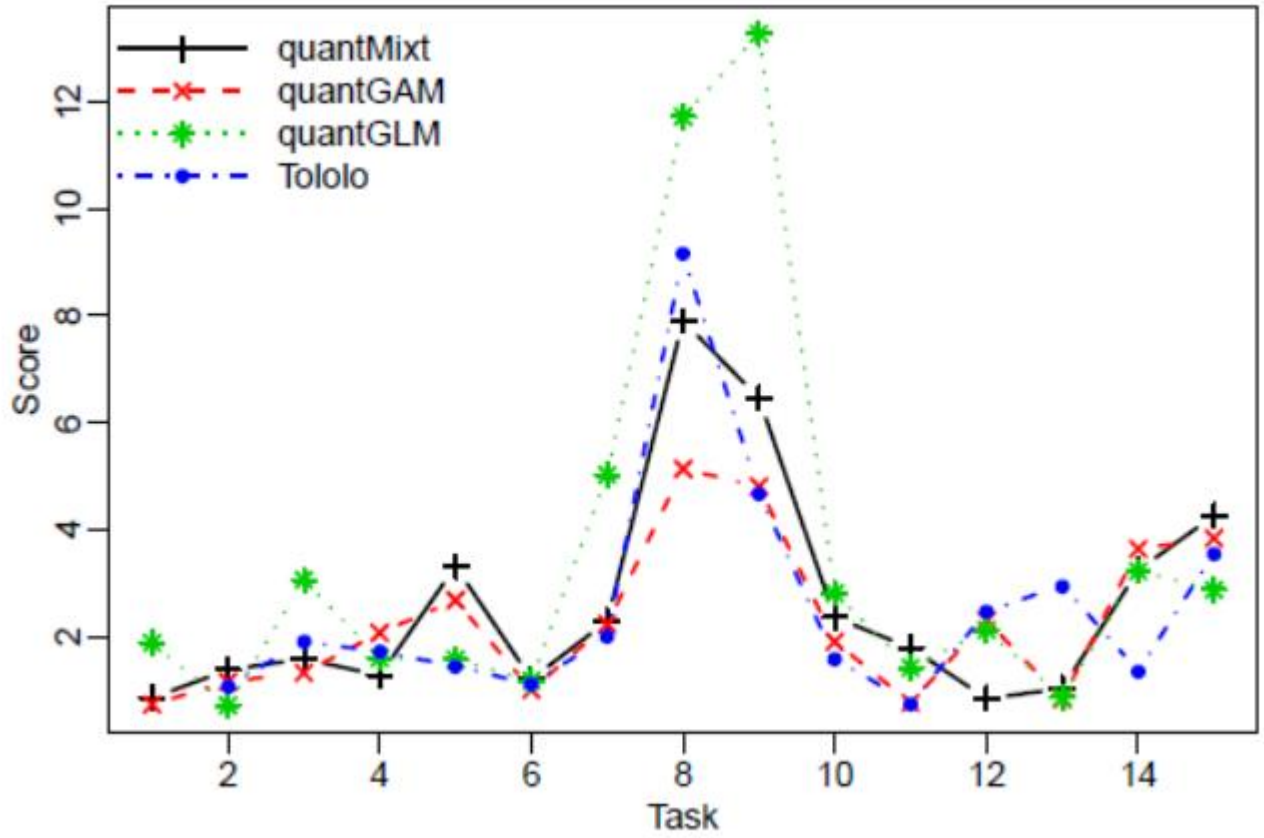
- PAR: spike preprocess AR (Weron, R., Misiorek, A., 2008)
- TARX
- PARX
- Random Forests (Breiman, 2001)
- Gradient boosting machines (Friedman, 1999)
- Specialized by season

Loss function: $\sum_{t \in E} \rho_\tau(P_t - \boldsymbol{\theta}^\top \boldsymbol{x}_t)$

Results



Results


quantMixt
quantGAM
quantGLM
Tololo
Score
2
4
6
8
10
12
Task
2
4
6
8
10
12
14

Implementation and computing tools

- All the methods are implemented using the R for statistics.
- Several packages have been used : *mgcv*, *quantreg*, *randomForest*, *gbm*, *glmnet*, for the modeling, *parallel* and *doMC* for parallel computing,
- All the computations were run on a Z820 desktop workstation on linux debian.

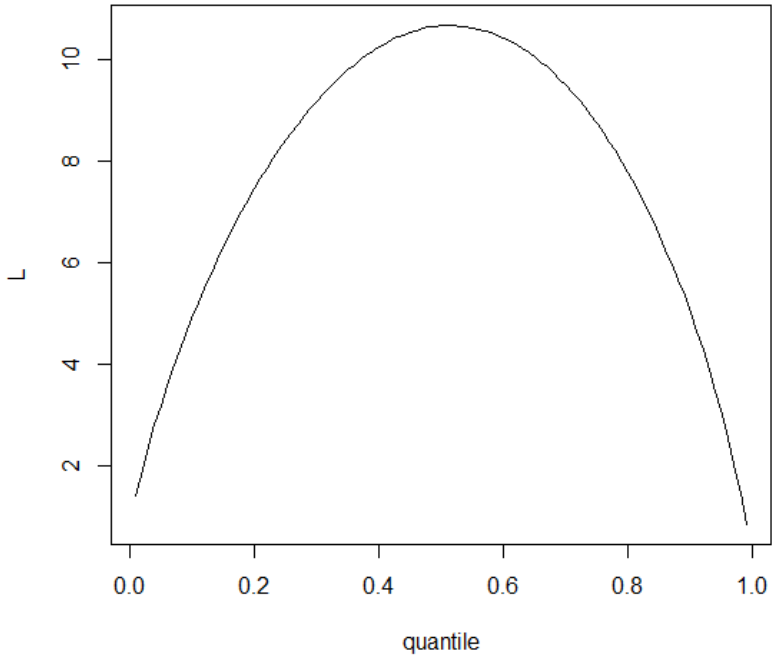
Conclusion

Fascinating competition, thanks again to the organizers

GAM quantile is highly adaptive, works well for (probabilistic) load and price forecasting

Still open questions/improvements:

- how do we measure the quality of a probabilistic forecast?
- quantGAM could be generalized to higher moment feature generation (not only mean and variance)
- identifiability constraint (positivity…) in GAM, useful for 2nd order moment modelling



Thanks !

Bibliography

[Hastie and Tibshirani 1990] Hastie, T. J. & Tibshirani, R. J. Generalized Additive Models Chapman & Hall/CRC, 1990

[Koenker 2005] Koenker, R. Quantile Regression Cambridge University Press, 2005

[Wood 2006] Wood, S. Chapman & Hall (Eds.) Generalized Additive Models, An Introduction with R 2006

[Pierrot and Goude 2011] Pierrot, A. & Goude, Y. Short-Term Electricity Load Forecasting With Generalized Additive Models Proceedings of ISAP power, pp 593-600, 2011

[Fan and Hyndman 2012] Fan, S. & Hyndman, R. J. Short-term load forecasting based on a semi-parametric additive model IEEE Transactions on Power Systems, 2012, 27(1), 134-141

[Goude, Nédellec and Kong 2013] 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

[Nédellec, Cugliari and Goude 2014] Nedellec, R.; Cugliari, J. & Goude, Y. GEFCom2012: Electric load forecasting and backcasting with semi-parametric models International Journal of Forecasting , 2014, 30, 375 – 381

[Gaillard, Goude, Nédellec *to appear*] Gaillard, P.; Goude, Y. & Nédellec, R. Semi-parametric models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting, to appear in International Journal of Forecasting

Software

R Team, R. C. R: A Language and Environment for Statistical Computing 2014

quantreg Koenker, R. quantreg: Quantile Regression 2013

mgcv Wood, S. mgcv:GAMs and Generalized Ridge Regression for R. R News 1(2) R News, 2001, 1(2), 20-25

randomForest Liaw, A. & Wiener, M. Classification and Regression by randomForest R News, 2002, 2, 18-22

gbm Ridgeway, G. & contributions from others gbm: Generalized Boosted Regression Models 2015

doMC RevolutionAnalytics doMC: Foreach parallel adaptor for the multicore package 2014