Advances in probabilistic forecasting of electricity loads and prices

Rafał Weron

Department of Operations Research
Wrocław University of Technology (PWR), Poland

Why forecast loads and prices?

(Hong, 2015, EnergyBiz Magazine)

- A ballpark estimate of savings from a 1% reduction in MAPE for a utility with 1GW peak load:
  - \$500k/year from long-term load forecasting
  - \$300k/year from short-term load forecasting
  - \$600k/year from short-term load + price forecasting

A look into the future of price forecasting

(Weron, 2014, Int.J.Forecasting)

1. Modeling and forecasting the trend-seasonal components
2. **Beyond point forecasts – probabilistic forecasts**
3. **Combining forecasts**
   - **Point forecasts**
   - **Probabilistic forecasts**
4. Multivariate factor models
5. Guidelines for evaluating forecasts

Reviews and competitions

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

2014

Global Energy Forecasting Competition 2012

Tao Hong[a,*], Pierre Pinson[b], Shu Fan[c]

[a] SAS Institute Inc, United States
[b] Technical University of Denmark, Denmark
[c] Monash University, Australia

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

2014

Review

Electricity price forecasting: A review of the state-of-the-art with a look into the future

Open Access

Rafał Weron

Institute of Organization and Management, Wrocław University of Technology, Wrocław, Poland

ARTICLE INFO

Keywords:
Electricity price forecasting
Day-ahead market
Seasonality
Autoregression
Neural network
Factor model
Forecast combination
Probabilistic forecast

ABSTRACT

A variety of methods and ideas have been tried for *electricity price forecasting* (EPF) over the last 15 years, with varying degrees of success. This review article aims to explain the complexity of available solutions, their strengths a and threats that the forecasting tools offer or that looks ahead and speculates on the directions EPF or so. In particular, it postulates the need for objec (i) the same datasets, (ii) the same robust error eva testing of the significance of one model's outperfor
© 2014 The Author. Published by Elsevier B.V

This is an open access (http://creative

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

2016

Probabilistic Electric Load Forecasting: A Tutorial Review

Tao Hong and Shu Fan

Abstract

Load forecasting is a fundamental business problem established since the inception of the electric power industry. Over the past 100 plus years, both research efforts and industry practices in this area are primarily on point load forecasting. In the recent decade, due to the increased market competition, aging infrastructure and renewable integration requirements, probabilistic load forecasting is becoming more and more important to energy systems planning and operations. This paper offers a tutorial review of probabilistic electric load forecasting, including notable techniques, methodologies, evaluation metrics, common misunderstandings and recommended research directions.

Global Energy Forecasting Competition 2014

Proudly sponsored by IEEE Power & Energy Society

IEEE PES Power & Energy Society®

Join GEFCOM Forum

GEFCOM 2014	GEFCOM 2014	GEFCOM 2014	GEFCOM 2014
Load Forecasting	Price Forecasting	Wind Forecasting	Solar Forecasting

GEFCom2012

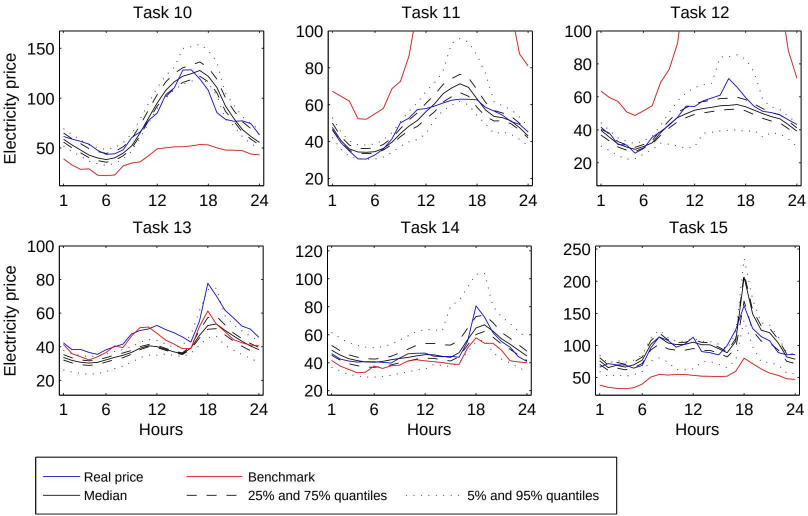
- Two tracks
- Participants
  - 2000+ entries
  - 200+ teams
  - 30+ countries

GEFCom2014

GEFCOM 2014 Load Forecasting | GEFCOM 2014 Price Forecasting | GEFCOM 2014 Wind Forecasting | GEFCOM 2014 Solar Forecasting | IEEE PES Power & Energy Society®

- 400+ people, 40+ countries
- Incremental data sets released on weekly basis
- Price Track:
  - 287 contestants
  - Submit 99 quantiles for 24h load periods of the next day

GEFCom2014 Price Track



GEFCom2014 Price Track

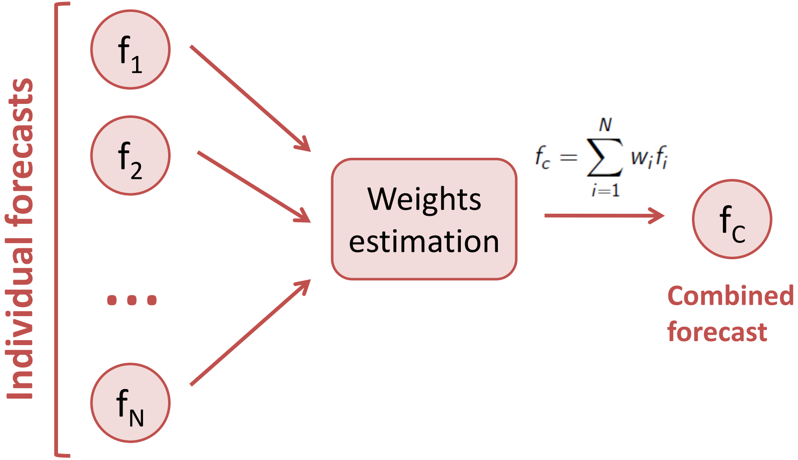
Top winning teams

1. Pierre Gaillard, Yannig Goude, Raphaël Nedellec (EDF R&D, F)
2. Katarzyna Maciejowska, Jakub Nowotarski (Wroclaw UT, PL)
3. Grzegorz Dudek (Czestochowa UT, PL)
4. Zico Kolter, Romain Juban, Henrik Ohlsson, Mehdi Maasoumy (C3 Energy, USA)
5. Frank Lemke (KnowledgeMiner Software, D)

Forecast combinations, forecast/model averaging

- The idea goes back to the 1960s to the seminal papers of Bates and Granger (1969) and Crane and Crotty (1967)
- In electricity markets:
  - Electricity demand or transmission congestion forecasting (Bunn, 1985a; Bunn and Farmer, 1985; Løland et al., 2012; Smith, 1989; Taylor, 2010; Taylor and Majithia, 2000)
  - Only recently applied in the context of electricity price forecasting (EPF): Bordignon et al. (2013), Nowotarski et al. (2014) and Raviv et al. (2013)

Point forecast averaging: The idea



Beyond point forecasts

- Variability of the electricity demand becoming a challenge to the utility industry in the smart grid era
- Extreme variability of electricity prices
- Ability to plan different strategies for the range of possible outcomes indicated by the probabilistic forecast
- Useful in practice → risk management and decision-making

- GEFCom2012 (point) → GEFCom2014 (probabilistic forecasts)

Interval forecast averaging

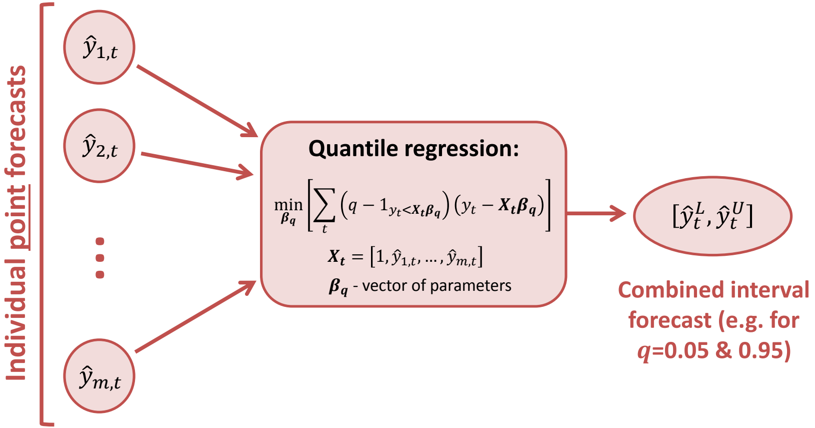
- For point forecasts: $f_c = \sum_{i=1}^{N} w_i f_i$ (e.g. a linear regression model)
- For interval forecasts the above formula does not hold
- A linear combination of q-th quantiles is **not** the q-th quantile of a linear combination of random variables

$$x_c^q \neq \sum_{i=1}^{N} w_i x_i^q$$

- $\rightarrow$ Need for development of new approaches

Quantile Regression Averaging

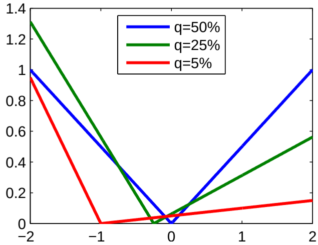
(Nowotarski & Weron, 2015, Computational Statistics)



How does the score function look like?

For vector $\boldsymbol{X_t} = [1, \hat{y}_{1,t}, ..., \hat{y}_{m,t}]$ of point forecasts, i.e. explanatory variables, weights $\boldsymbol{\beta_q}$ are estimated by minimizing:

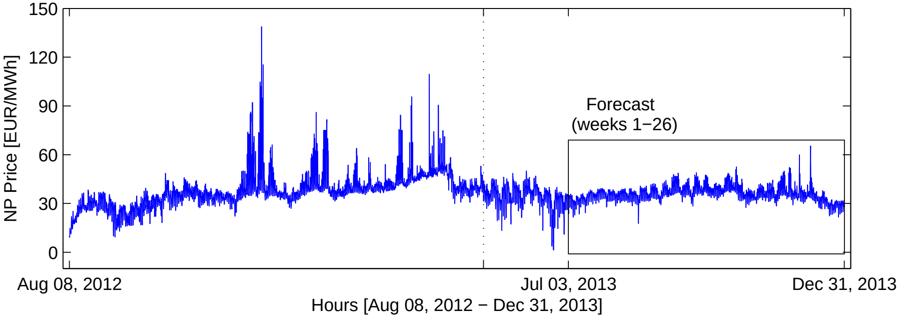
$$\min_{\boldsymbol{\beta_q}} \left[\sum_{\{t: y_t \geq \boldsymbol{X_t}\boldsymbol{\beta_q}\}} q|y_t - \boldsymbol{X_t}\boldsymbol{\beta_q}| + \sum_{\{t: y_t < \boldsymbol{X_t}\boldsymbol{\beta_q}\}} (1-q)|y_t - \boldsymbol{X_t}\boldsymbol{\beta_q}| \right]$$



Case study I: Combining individual price forecasts
(Nowotarski & Weron, 2014, EEM14)

- Six individual point forecast models:
  - Autoregression (AR)
  - Threshold AR (TAR)
  - Semi-parametric AR (SNAR)
  - Mean-reverting jump diffusion (MRJD)
  - Non-linear AR neural network (NAR)
  - Factor model (FM)

The data



- **Seven** months for calibration of individual models
- **Four** weeks for calibration of quantile regression
- **26** weeks for evaluation of interval forecasts

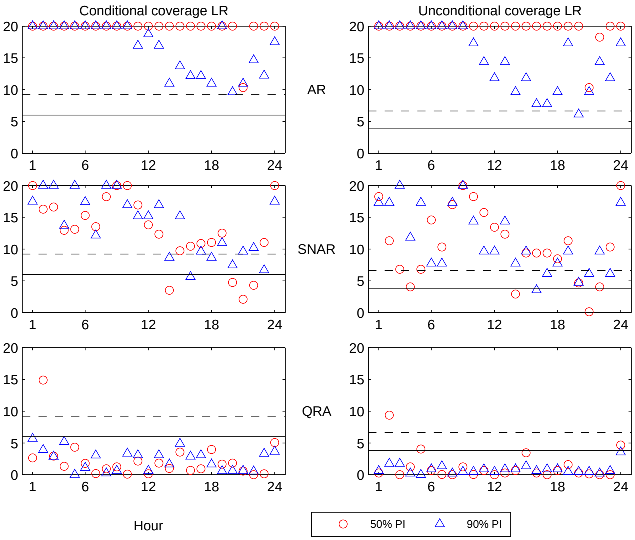
Evaluation of forecasts

- 50% and 90% two-sided day-ahead prediction intervals
- Two benchmark models: AR and SNAR
- Christoffersen's (1998) test for unconditional and conditional coverage
- The focus on the sequence: $I_t = \begin{cases} 1 & y_t \in [\hat{y}_t^L, \hat{y}_t^U] \\ 0 & y_t \notin [\hat{y}_t^L, \hat{y}_t^U] \end{cases}$
Conditional Coverage test	Unconditional Coverage test
(UC + independece)	
Asymptotically $\chi^2(2)$	Asymptotically $\chi^2(1)$

Results: Unconditional coverage

PI	AR	SNAR	**QRA**
	Unconditional coverage		
50%	77.50	61.93	**49.77**
90%	97.53	96.41	**89.33**
	Mean width (STD of interval width)		
50%	4.55 (1.34)	2.76 (0.61)	**2.23 (0.81)**
90%	11.14 (3.31)	9.33 (2.45)	**6.78 (2.20)**

Results: Christoffersen's test

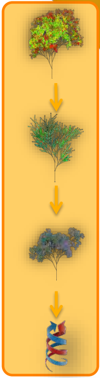


GEFCom2014 Price Track: (1st and) 2nd place for QRA!

In the 'AI world' ...

- *Committee machines*, *ensemble averaging*, *expert aggregation*:
  - Guo and Luh (2004) combine a RBF network (23 inputs and six clusters) and a MLP (55 inputs and eight hidden neurons) to compute daily average on-peak electricity price for New England
  - Forecast combinations and committee machines seem to *evolve independently*, with researchers from both groups not being aware of the parallel developments !

GEFCom2014 Solar and Wind Tracks: 2nd place



GRADIENT BOOSTING REGRESSOR

1. FRIEDMAN, 1999
2. ROBUST COMBINATION OF WEAK LEARNERS
3. RESISTANT TO OVERFITTING, FEW TUNING PARAMETERS, PROVIDES VARIABLE IMPORTANCE
4. QUANTILE REGRESSION CAPABLE!
5. EMPLOYS BOOSTING

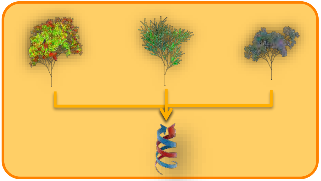
STACKING

GBR + QRF

RF

ENSEMBLING

G. BARTA, S3 SEMINAR @ PWR, WROCŁAW, 15 DEC 2015



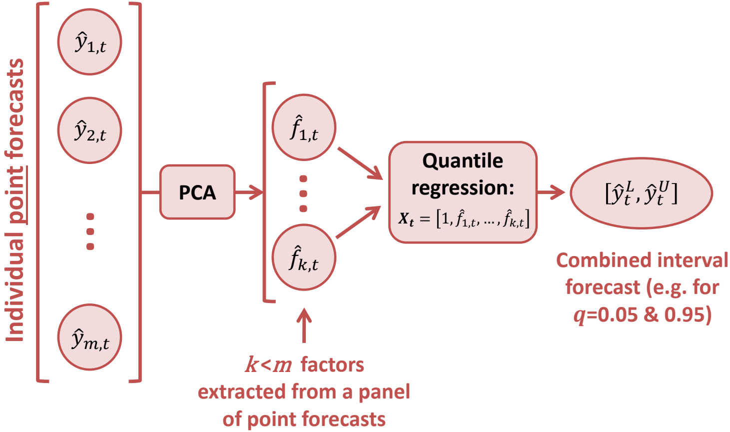
RANDOM FOREST REGRESSOR

1. BREIMAN, 2001
2. ROBUST COMBINATION OF WEAK LEARNERS
3. RESISTANT TO OVERFITTING, FEW TUNING PARAMETERS, PROVIDES VARIABLE IMPORTANCE
4. QUANTILE REGRESSION CAPABLE!
5. EMPLOYS BAGGING

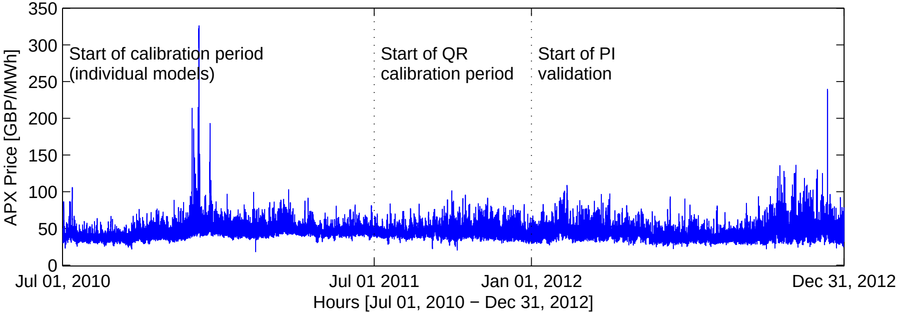
energyforecasting.io

Extension: A large number of predictors

(Maciejowska, Nowotarski & Weron, 2016, Int.J.Forecasting)



Case study II



- **32** individual forecasting models
- **One year** for calibration of individual models
- **Half a year** for calibration of quantile regression
- **One year** for evaluation of interval forecasts

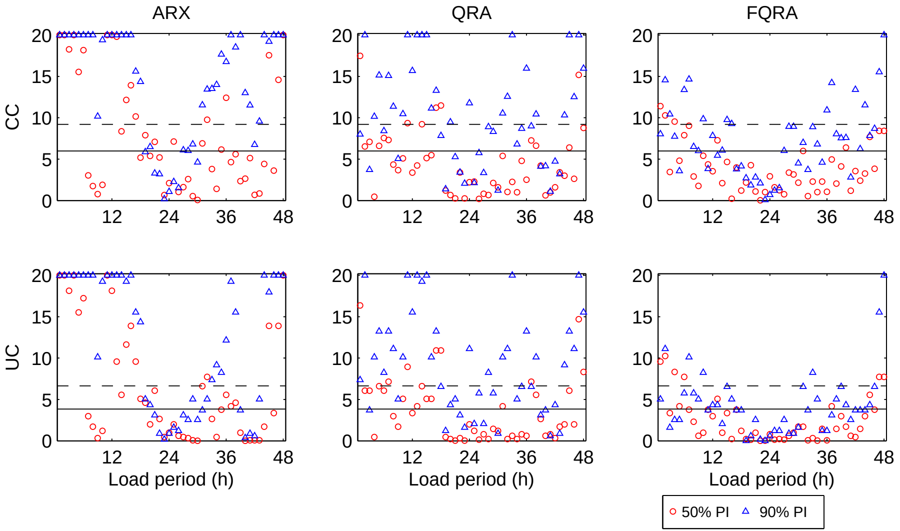
Evaluation of forecasts

- 50% and 90% two-sided day-ahead prediction intervals
- Three methods: QRA, FQRA and ARX (benchmark)
- Christoffersen's (1998) test for unconditional and conditional coverage
- Winkler score for a symmetric $(1-\alpha)\times 100\%$ prediction interval:

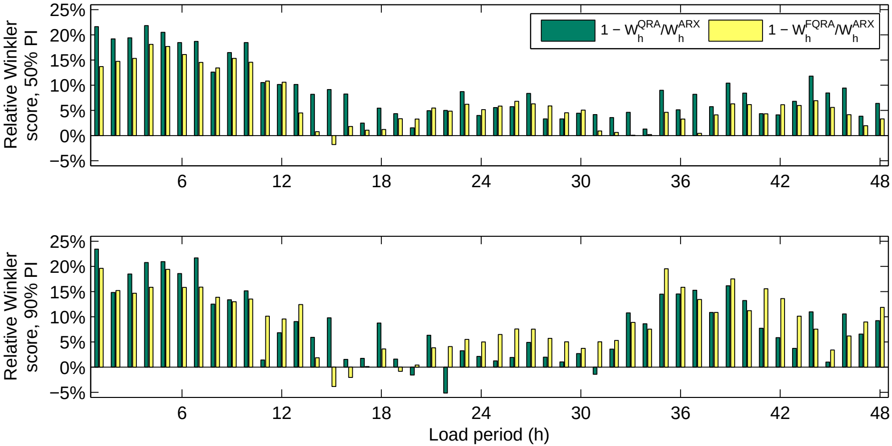
$$W_t = \begin{cases} \delta_t & \text{dla} \quad y_t \in [\hat{y}_t^L, \hat{y}_t^U], \\ \delta_t + \frac{2}{\alpha}(\hat{y}_t^L - y_t) & \text{dla} \quad y_t < \hat{y}_t^L, \\ \delta_t + \frac{2}{\alpha}(y_t - \hat{y}_t^U) & \text{dla} \quad y_t > \hat{y}_t^U, \end{cases}$$

where $\delta_t = \hat{y}_t^U - \hat{y}_t^L$ is the interval's width

Results: Christoffersen's test



Results: Winkler score



Case study III: Combining sister load forecasts

(Liu, Nowotarski, Hong & Weron, 2016, IEEE Transactions on Smart Grid)

- Variable selection may be difficult in load forecasting
- **Sister models** – constructed by different subsets of variables with overlapping components
  - Here: 2 or 3 years for calibration and 4 ways of partitioning training and validation periods

$$\hat{y}_t = \beta_0 + \beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t + f(T_t) + \\ + \sum_d f(\tilde{T}_{t,d}) + \sum_{lag} f(T_{t-lag}),$$

- **Sister forecasts** are generated from sister models

Sister models

$$\hat{y}_t = \beta_0 + \overbrace{\beta_1 M_t + \beta_2 W_t + \beta_3 H_t + \beta_4 W_t H_t}^{\text{calendar effects}} + \overbrace{f(T_t)}^{\text{temp. dependence}} + \\ + \underbrace{\sum_d f(\tilde{T}_{t,d}) + \sum_{lag} f(T_{t-lag})}_{\text{recency effect}},$$

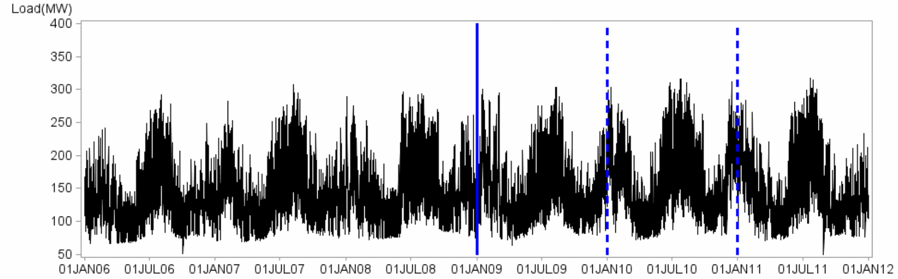
where:

$$f(T_t) = \beta_5 T_t + \beta_6 T_t^2 + \beta_7 T_t^3 + \beta_8 T_t M_t + \beta_9 T_t^2 M_t + \\ + \beta_{10} T_t^3 M_t + \beta_{11} T_t H_t + \beta_{12} T_t^2 H_t + \beta_{13} T_t^3 H_t$$

$$\tilde{T}_{t,d} = \frac{1}{24} \sum_{lag=24d-23}^{24d} T_{t-lag}$$

The data

(from the load forecasting track of GEFCom2014)



- **2 or 3 years** for calibration of sister (individual) models
- **1 year** for validation of sister (individual) models (variable selection)
- **1 year** for validation of probabilistic forecasts (best models selection)
- **1 year** for testing probabilistic forecasts

Benchmarks

- Two naive benchmarks
  - Scenario generation from historical weather data, no recency effect (Vanilla)
  - Quantiles interpolated from 8 individual forecasts (Direct)
- Benchmarks from individual models
  - 8 individual models (Ind) with residuals' distribution
  - Best Individual (BI) individual model according to MAE

Evaluation of forecasts

- Pinball loss function for 99 percentiles (as in GEFCom2014)

$$P_t = \begin{cases} (1-q)(\hat{y}_t^q - y_t), & y_t < \hat{y}_t^q \\ q(y_t - \hat{y}_t^q), & y_t \geq \hat{y}_t^q \end{cases}$$

- Winkler score for 50% and 90% two-sided day-ahead PI:

$$W_t = \begin{cases} \delta_t & \text{dla} \quad y_t \in [\hat{y}_t^L, \hat{y}_t^U], \\ \delta_t + \frac{2}{\alpha}(\hat{y}_t^L - y_t) & \text{dla} \quad y_t < \hat{y}_t^L, \\ \delta_t + \frac{2}{\alpha}(y_t - \hat{y}_t^U) & \text{dla} \quad y_t > \hat{y}_t^U, \end{cases}$$

where $\delta_t = \hat{y}_t^U - \hat{y}_t^L$ is the interval's width

Results: Test period

Model class	Pinball	Winkler (50%)	Winkler (90%)
QRA(8,183)	**2.85**	**25.04**	**55.85**
Ind(1,91)	3.22	26.35	56.38
BI(-,365)	3.00	26.38	57.17
Direct	3.19	26.62	94.27
Vanilla	8.00	70.51	150.0

- Sister forecasts easy to generate
- No need for independent expert forecasts
- Simple way to leverage from point to probabilistic forecasts

Take-home message(s)

- Combining point forecasts is a robust technique, generally improving the performance
- The new trend is probabilistic forecasting
- Combining interval (or density) forecasts is more tricky than combining point forecasts
- QRA is a simple way to leverage from point to probabilistic forecasts
- ... forecast evaluation is a critical issue

See you next year ... in Essen for EFC16

EFC14
(St.Gallen,
11-12.12.2014)

References

- T. Hong (2015) *Crystal ball lessons in predictive analytics*, EnergyBiz, Spring 35-37
- B. Liu, J. Nowotarski, T. Hong, R. Weron (2016) *Probabilistic load forecasting via Quantile Regression Averaging on sister forecasts*, IEEE Transactions on Smart Grid (doi: 10.1109/TSG.2015.2437877)
- K. Maciejowska, J. Nowotarski, R. Weron (2016) *Probabilistic forecasting of electricity spot prices using Factor Quantile Regression Averaging*, International Journal of Forecasting 32, 957-965
- J. Nowotarski, R. Weron (2014) *Merging quantile regression with forecast averaging to obtain more accurate interval forecasts of Nord Pool spot prices*, IEEE Conference Proceedings – EEM'14 (doi: 10.1109/EEM.2014.6861285)
- J. Nowotarski, R. Weron (2015) *Computing electricity spot price prediction intervals using quantile regression and forecast averaging, Computational Statistics 30(3), 791-803*
- R. Weron (2014) *Electricity price forecasting: A review of the state-of-the-art with a look into the future*, International Journal of Forecasting 30(4), 1030-1081