Probabilistic forecasting of wholesale electricity prices

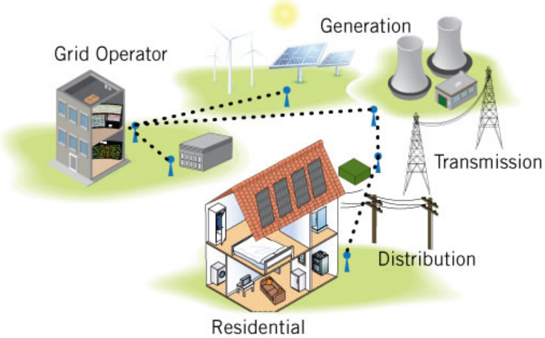
Rafał Weron

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

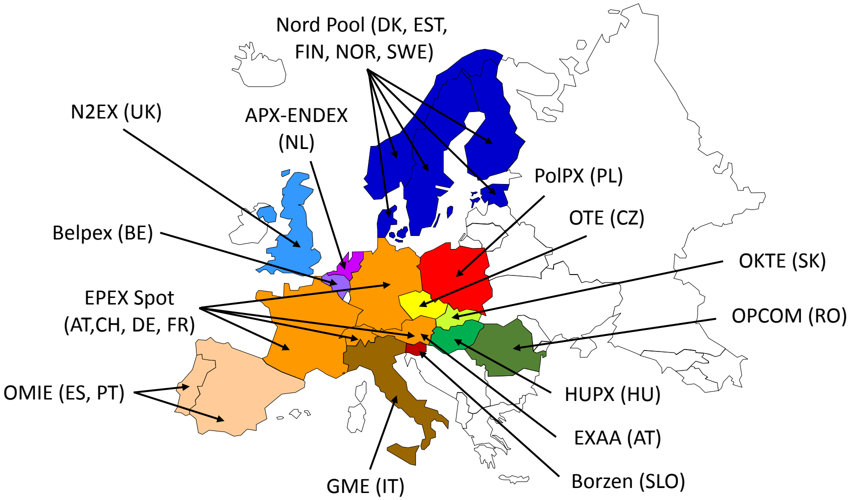
http://kbo.pwr.edu.pl/pracownik/weron

The vocabulary

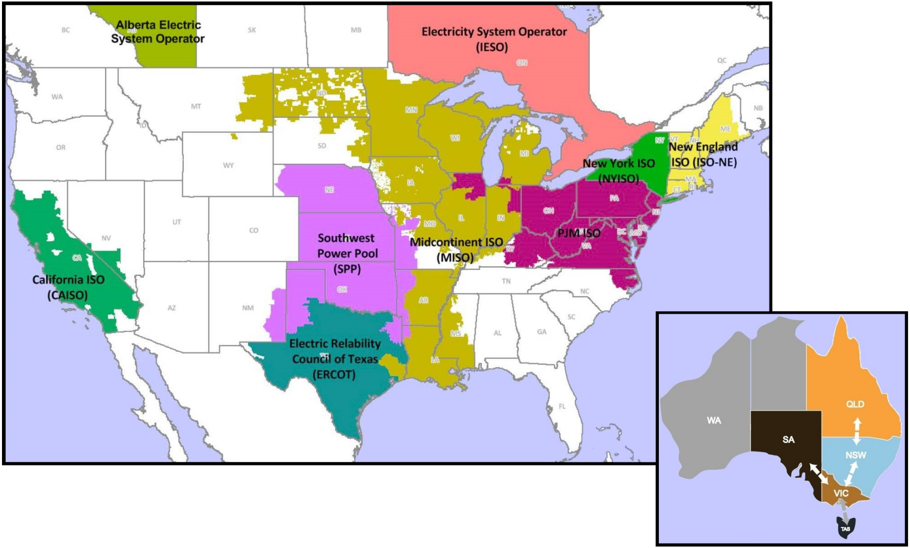
- **Smart grids** (smart meters, appliances, houses, ... cities)
- **Prosumers** = producing consumers
- **Load** = consumption ($\approx$ demand) + losses
- **Non-storability**
- **Power grid/network**
- **Interconnector**
- **Power exchange, power pool**



Power markets in Europe

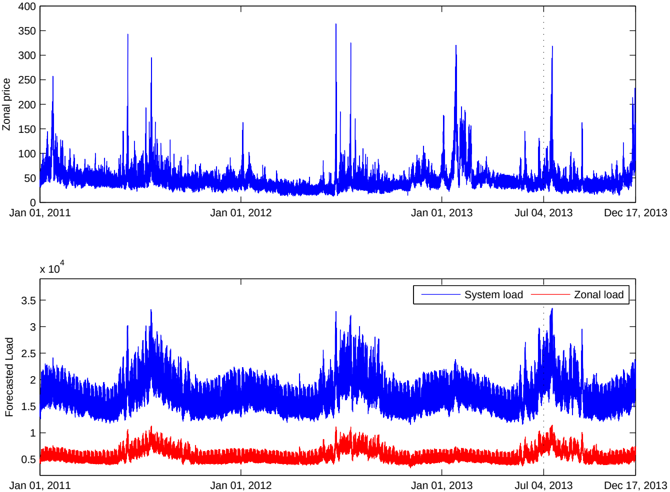


... in North America and Australia



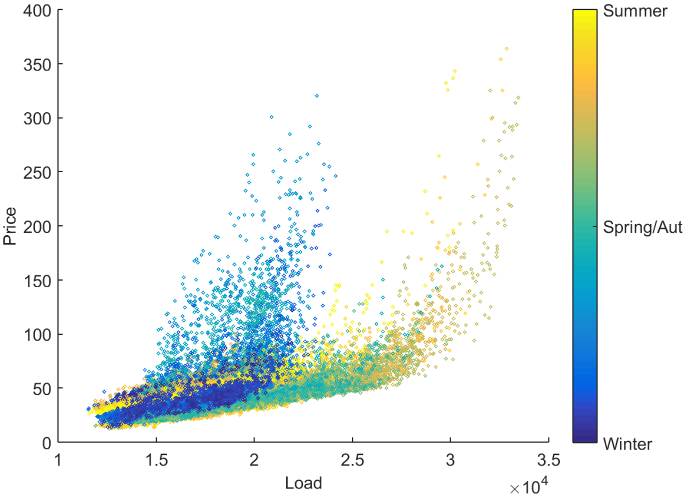
Electricity prices and loads (GEFCom2014)

Seasonality, floor reversion and price spikes

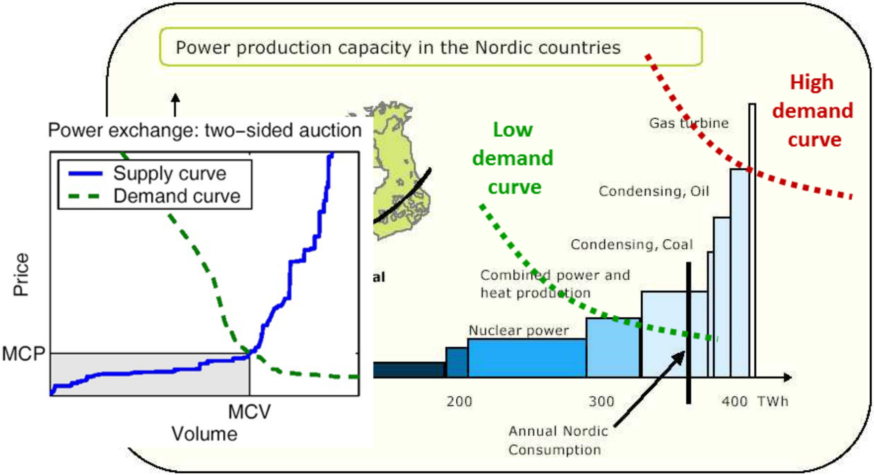


Electricity prices vs. loads (GEFCom2014)

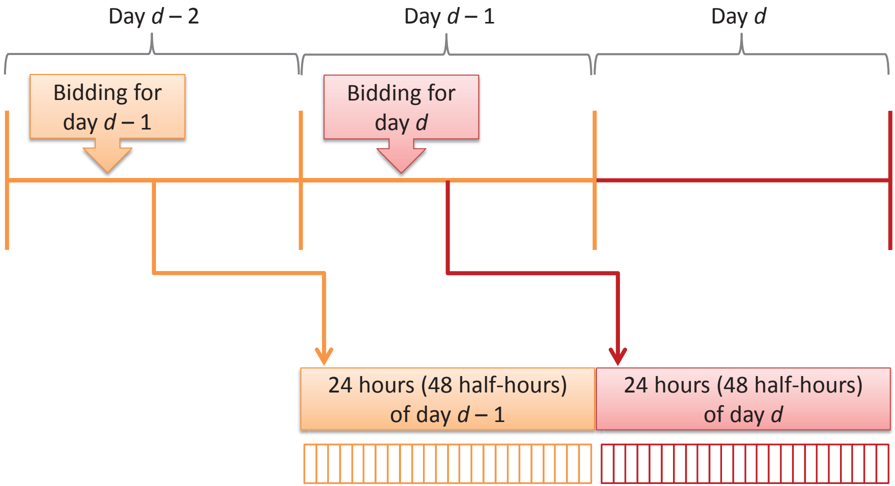
Non-linear, time-varying dependence



Supply stack and price formation

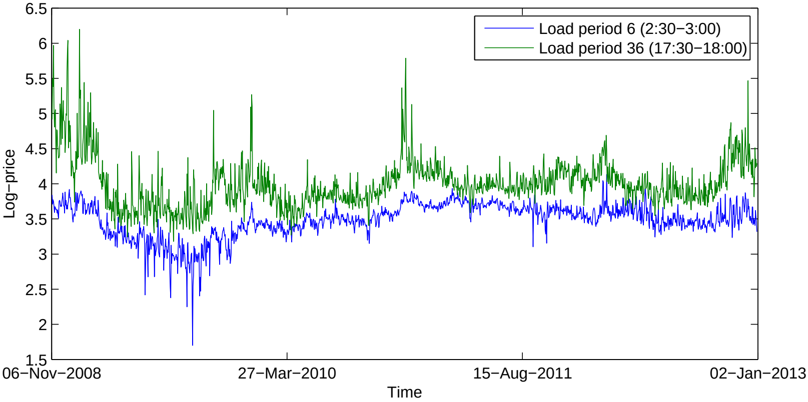


The electricity ‘spot’ price



Prices for different load periods

Strongly correlated but seem to follow different *data generating processes* (DGPs)



A commodity ... but a very special one

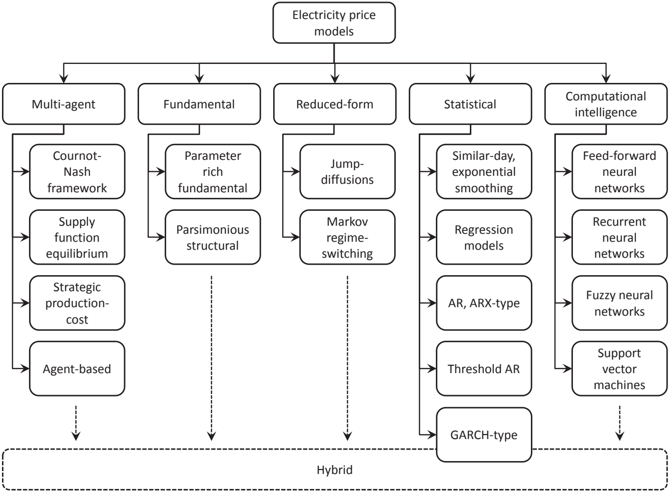
- Not storable (economically)
- Time consuming shut-down/start-up procedures for some technologies
- Extreme price changes → spikes
- Possible negative prices
- Pronounced daily and weekly cycles, annual seasonality
- Mean (floor) reversion
- Highly volatile

Forecasting horizons

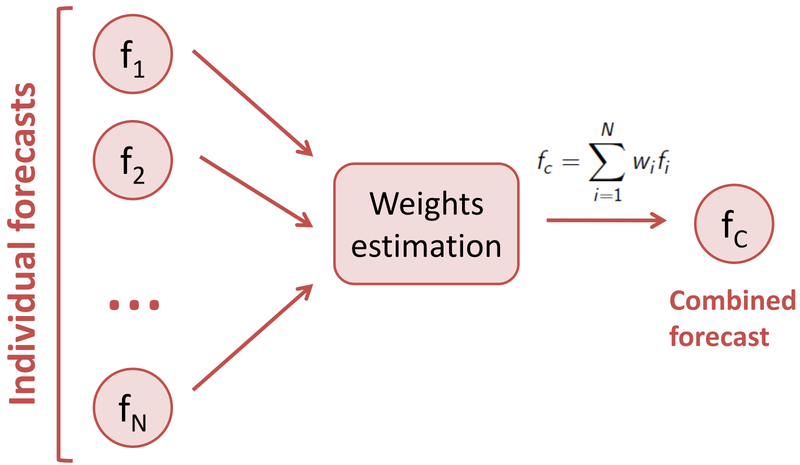
- Short-term
  - From a few minutes up to a few days ahead
  - Of prime importance in day-to-day market operations
- Medium-term
  - From a few days to a few months ahead
  - Balance sheet calculations, risk management, derivatives pricing
  - Inflow of 'finance solutions'
- Long-term
  - Lead times measured in months, quarters or even in years
  - Investment profitability analysis and planning
  - Beyond the scope of this review

A taxonomy of (price) modeling approaches

(Weron, 2014, Int. J. Forecasting)



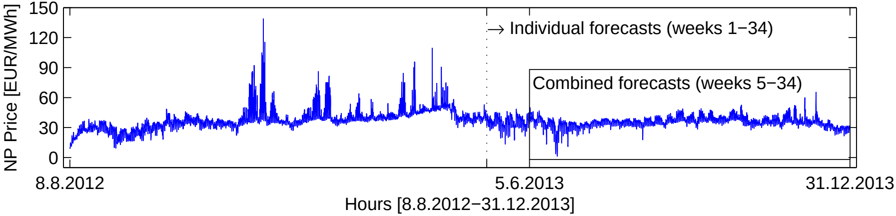
Point forecast averaging: The idea

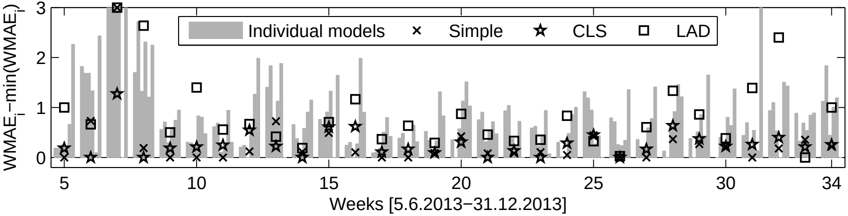


Forecast combinations, forecast/model averaging

- 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)

Case study I: Combining price forecasts
(Weron, 2014, Int.J.Forecasting)





Case study I: Combining price forecasts

Summary statistics for 6 individual and 3 averaging methods: $\overline{\text{WMAE}}$ is the mean value of WMAE for a given model (with standard deviation in parentheses), $\#$ *best* is the number of weeks a given averaging method performs best in terms of WMAE, and finally *m.d.f.b.* is the mean deviation from the best model in each week. The out-of-sample test period covers 30 weeks (5.6.2013–31.12.2013).

	Individual models						Forecast combinations		
	AR	TAR	SNAR	MRJD	NAR	FM	Simple	CLS	LAD
$\overline{\text{WMAE}}$	5.03	5.07	4.77	4.98	4.88	5.36	4.47	**4.29**	4.92
	(3.40)	(3.53)	(3.26)	(3.17)	(1.62)	(3.17)	(2.87)	(1.88)	(2.41)
$\#$ *best*	1	3	4	1	2	4	**8**	6	1
m.d.f.b.	1.01	1.05	0.75	0.96	0.86	1.34	0.45	**0.27**	0.89

In the ‘AI world’ ...

- *Committee machines*, *ensemble averaging*:
  - 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 !

Reviews and competitions

Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast
ELSEVIER
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
ELSEVIER
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

Contents lists available at ScienceDirect
International Journal of Forecasting
journal homepage: www.elsevier.com/locate/ijforecast
ELSEVIER
2015

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

Interval forecast averaging

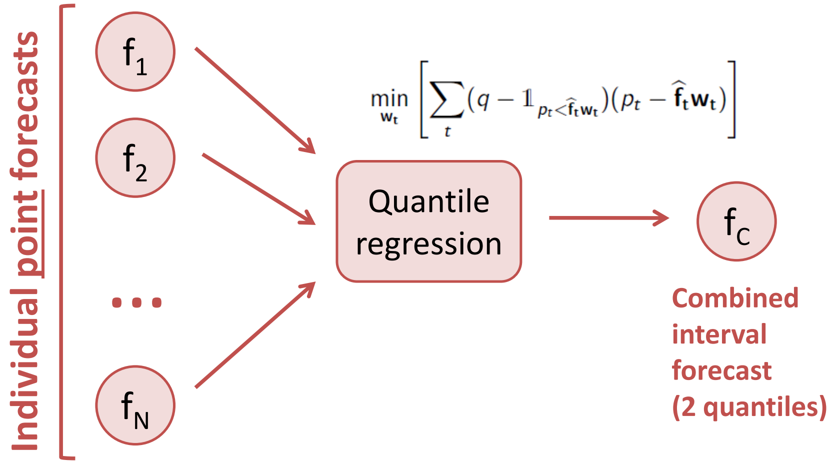
- For point forecasts: $f_c = \sum_{i=1}^{M} w_i f_i$ (e.g. a linear regression model)
- For interval forecasts the above formula does not hold
- A linear combination of α-th quantiles is **not** the α-th quantile of a linear combination of random variables

$$q_c^\alpha \neq \sum_{i=1}^{M} w_i q_i^\alpha$$

- $\rightarrow$ Need for development of new approaches

Quantile Regression Averaging

(Nowotarski & Weron, 2015, Computational Statistics)



Quantile Regression Averaging cont.

The averaging problem is given by:

$$Q_p(q|\widehat{\boldsymbol{p}}_{\boldsymbol{t}}) = \widehat{\boldsymbol{p}}_{\boldsymbol{t}}\boldsymbol{w}_{\boldsymbol{q}}$$

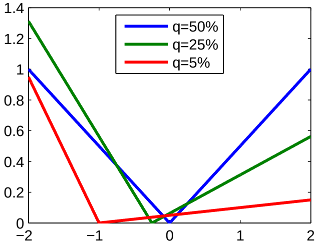
- $Q_p(q|\cdot)$ is the conditional q-th quantile of the electricity spot price distribution,
- $\widehat{\boldsymbol{p}}_{\boldsymbol{t}}$ are the regressors (explanatory variables)
- $\boldsymbol{w}_{\boldsymbol{q}}$ is a vector of parameters for a given q-th quantile

Quantile Regression Averaging cont.

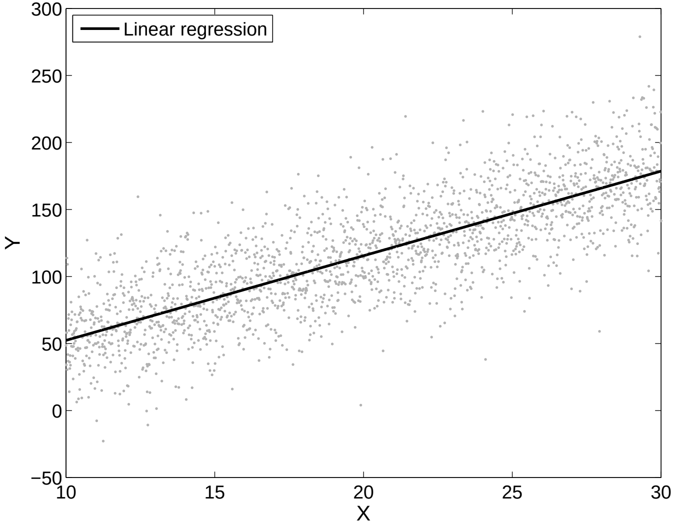
The weights are estimated by minimizing:

$$\min_{\boldsymbol{w_t}} \left[\sum_{\{t: p_t \geq \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}\}} q|p_t - \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}| + \sum_{\{t: p_t < \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}\}} (1-q)|p_t - \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}| \right] =$$

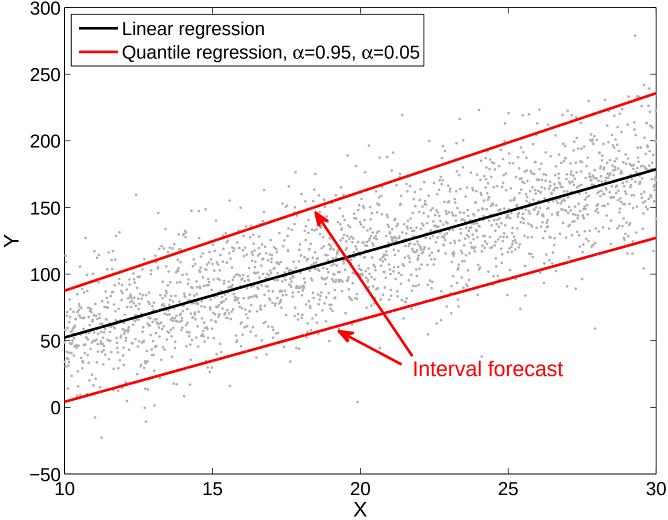
$$\min_{\boldsymbol{w_t}} \left[\sum_t (q - \mathbb{1}_{p_t < \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}})(p_t - \widehat{\boldsymbol{p}}_t \boldsymbol{w_t}) \right]$$



Quantile regression



Quantile regression

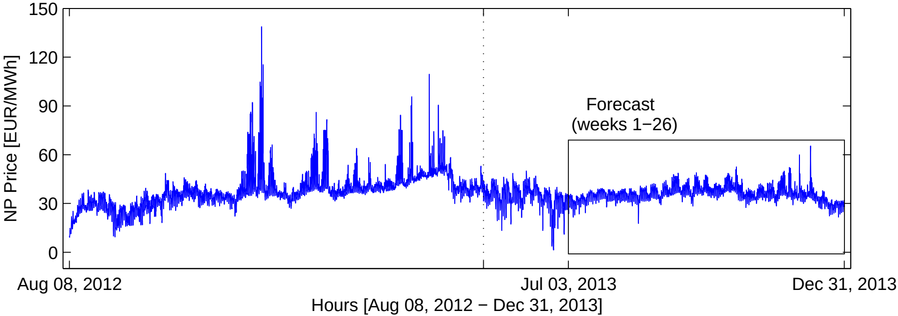


Case study II: Combining individual price forecasts

(Nowotarski & Weron, 2015, Computational Statistics)

- 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

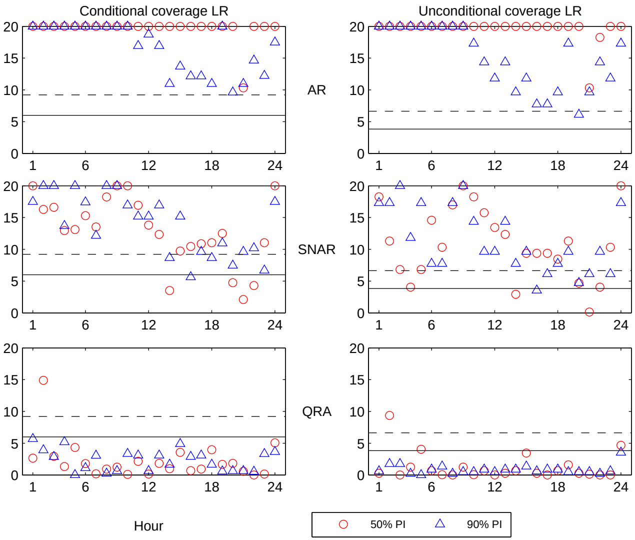
- 50% and 90% two-sided day-ahead prediction intervals
- Two benchmark models: AR and SNAR
- Christoffersen's (1998) test for unconditional and conditional coverage
- The focus on the sequence: $I_t = \begin{cases} 1 & p_t \in [L_{t|t-1}, U_{t|t-1}] \\ 0 & p_t \notin [L_{t|t-1}, U_{t|t-1}] \end{cases}$
-

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!

Case study III: Combining sister load forecasts

(Liu, Nowotarski, Hong & Weron, 2015, IEEE Transactions on Smart Grid)

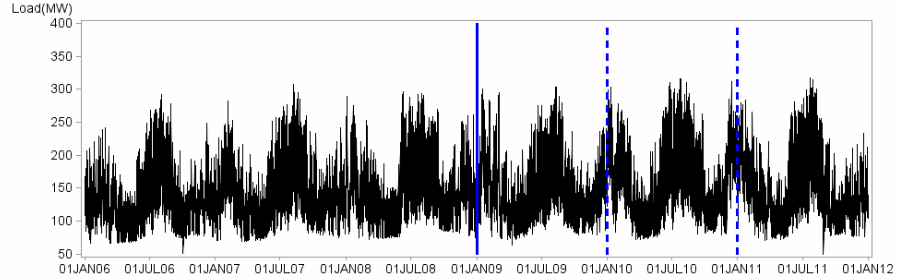
- Variable selection may be difficult in load forecasting
- **Sister models** – constructed by different subsets of variables with overlapping components
  - Here: 2 or 3 years for calibration and 4 ways of partitioning training and validation periods

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

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

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)

$$Pinball_t = \begin{cases} (1-q)(\hat{p}_t^q - p_t), & p_t < \hat{p}_t^q \\ q(p_t - \hat{p}_t^q), & p_t \geq \hat{p}_t^q \end{cases}$$

- Winkler score for central $(1-\alpha) \times 100\%, \alpha = 0.5, 0.9$, two-sided day-ahead PI:

$$W_t = \begin{cases} \delta_t & \text{for} \quad p_t \in [L_{t|t-1}, U_{t|t-1}], \\ \delta_t + \frac{2}{\alpha}(L_{t|t-1} - p_t) & \text{for} \quad p_t < L_{t|t-1}, \\ \delta_t + \frac{2}{\alpha}(p_t - U_{t|t-1}) & \text{for} \quad p_t > U_{t|t-1}, \end{cases}$$

where $\delta_t = U_{t|t-1} - L_{t|t-1}$ is the PI width

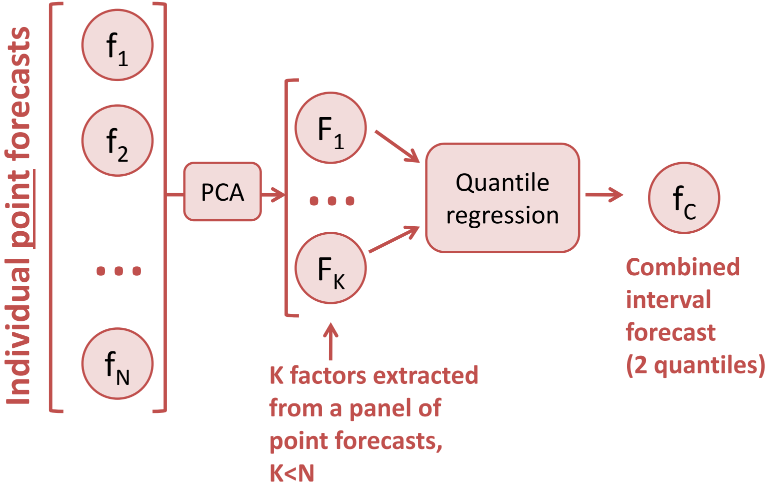
Results: Test period

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

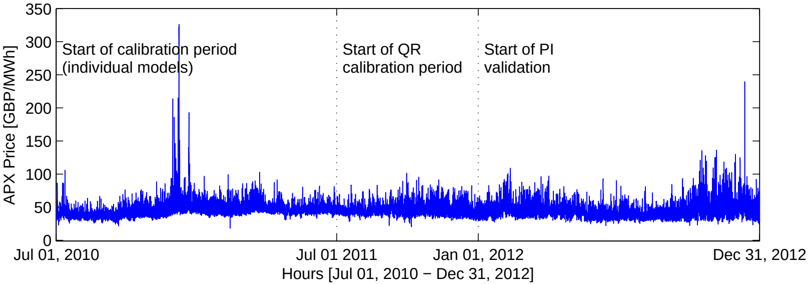
Extension: A large number of predictors

(Maciejowska, Nowotarski & Weron, 2015, Int.J.Forecasting)



Case study IV

(Maciejowska, Nowotarski & Weron, 2015, Int.J.Forecasting)



- **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

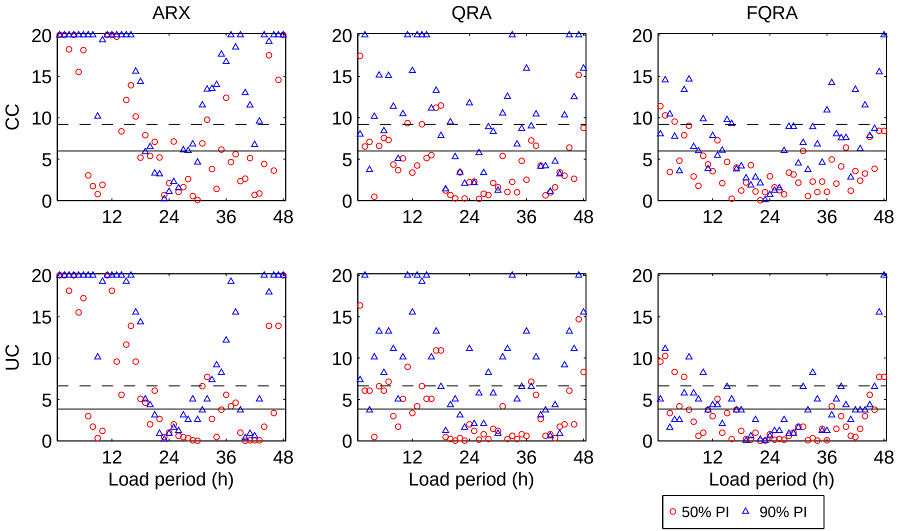
Evaluation of forecasts

- 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:

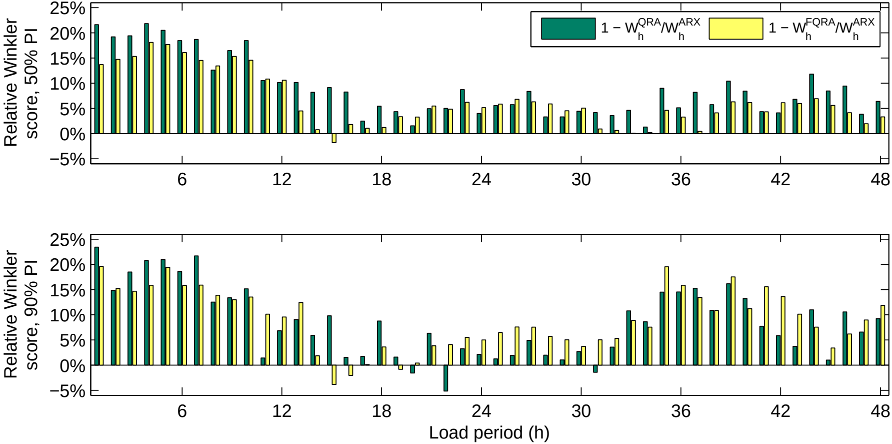
$$W_t = \begin{cases} \delta_t & \text{for} \quad p_t \in [L_{t|t-1}, U_{t|t-1}], \\ \delta_t + \frac{2}{\alpha}(L_{t|t-1} - p_t) & \text{for} \quad p_t < L_{t|t-1}, \\ \delta_t + \frac{2}{\alpha}(p_t - U_{t|t-1}) & \text{for} \quad p_t > U_{t|t-1}, \end{cases}$$

where $\delta_t = U_{t|t-1} - L_{t|t-1}$ is the interval's width

Results: Christoffersen's test



Results: Winkler score



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
- QRA is potentially useful for VaR calculations
- Forecast evaluation is a critical issue

Evaluating probabilistic forecasts

- For interval forecasts
  - The pinball function, as in GEFCom2014
  - The *interval* or *Winkler score*, see e.g. Maciejowska et al. (2015)
- For density forecasts
  - The *Continuous Ranked Probability Score* (CRPS), see e.g. Gneiting and Raftery (2007)
- Statistical tests
  - The *conditional coverage* test of Christoffersen (1998); for extensions and alternatives see Berkowitz et al. (2011)
  - The Berkowitz (2001) approach to the evaluation of density forecasts ($\rightarrow$ VaR backtesting)