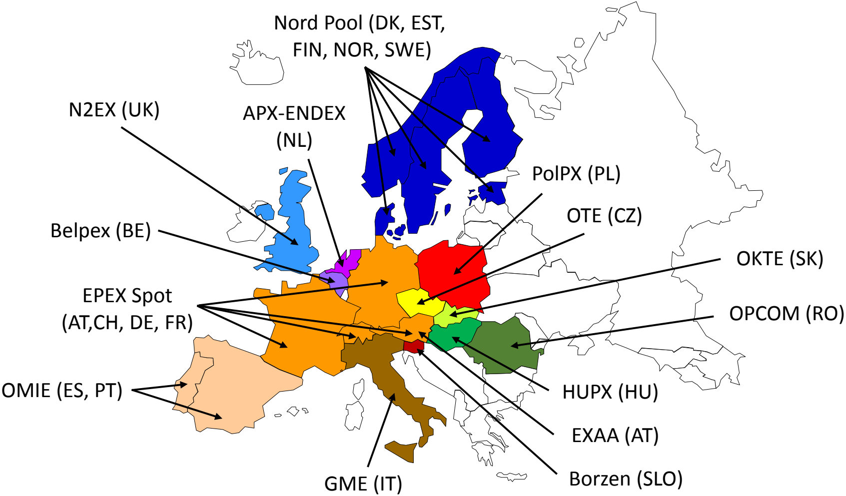
Importance of the long-term seasonal component in day-ahead electricity price forecasting: Regression vs. neural network models[*]

Rafał Weron

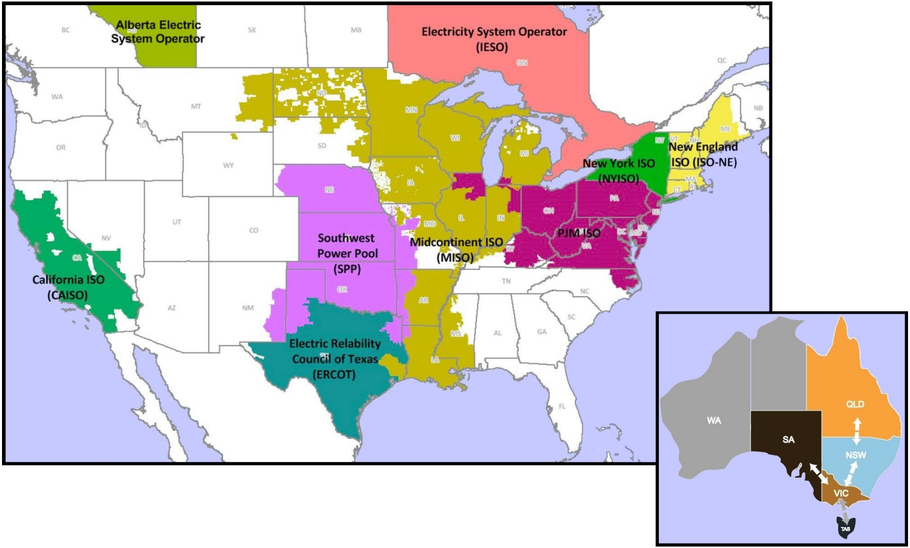
Department of Operations Research
Wrocław University of Science and Technology, Poland

[*] Based on a working paper with Grzegorz Marcjasz and Bartosz Uniejewski, available from RePEc: `https://ideas.repec.org/p/wuu/wpaper/hsc1703.html`

Markets for electricity in Europe

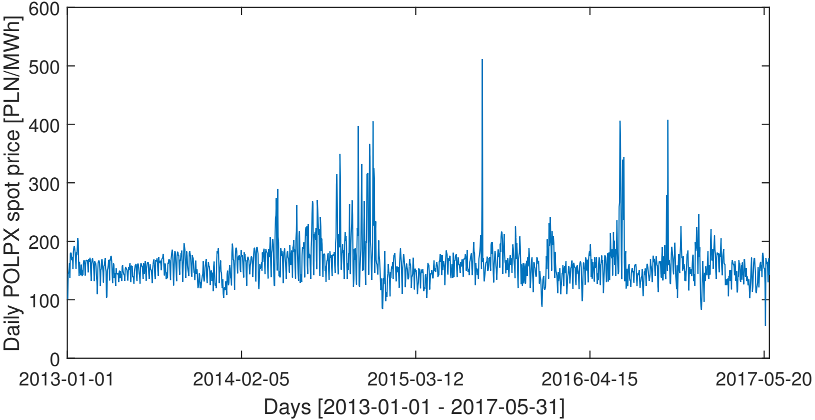


… in North America and Australia

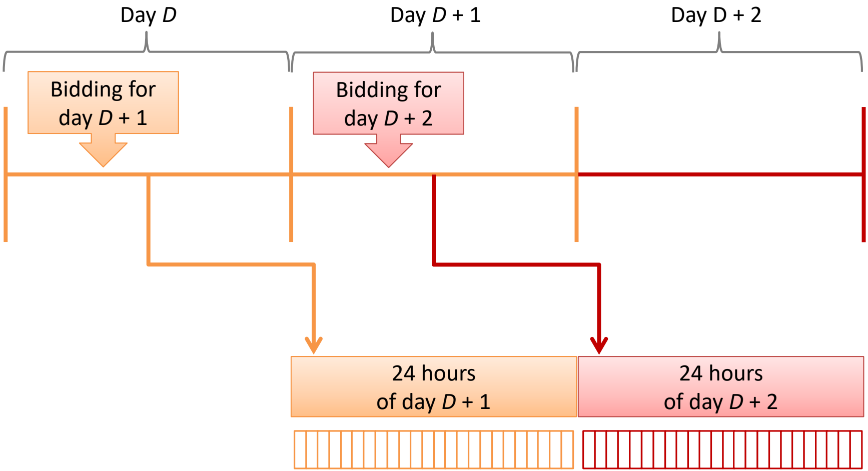


Electricity price time series

Seasonality, mean-reversion and price spikes



The electricity 'spot' (day-ahead) price



Day D

Day $D + 1$

Day D + 2

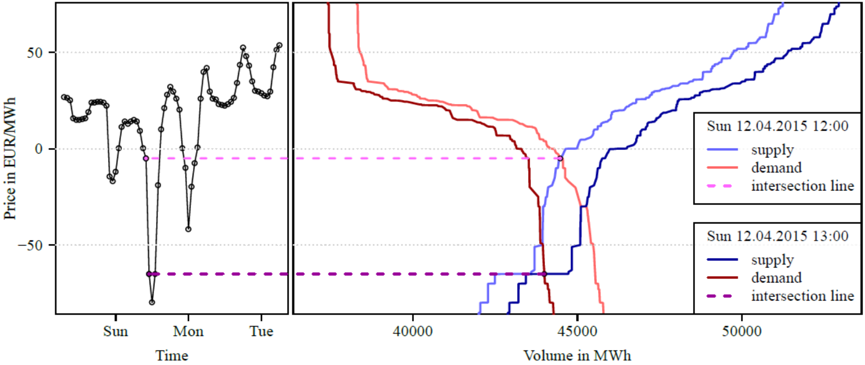
Bidding for day $D + 1$

Bidding for day $D + 2$

24 hours of day $D + 1$

24 hours of day $D + 2$

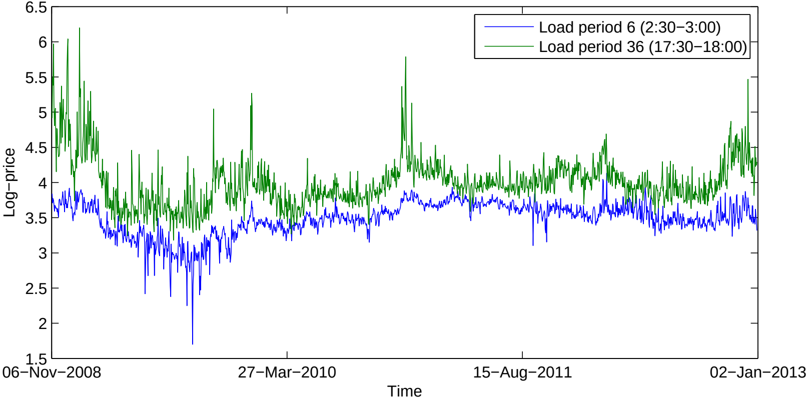
Supply and demand, renewables and negative prices



Source: Ziel & Steinert (2016)

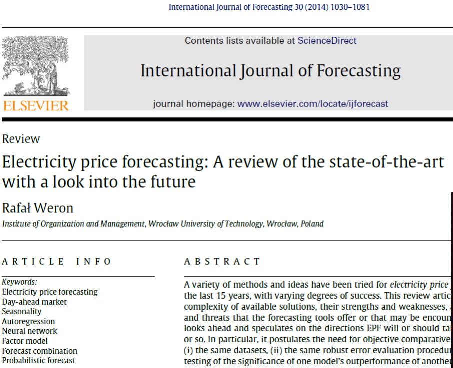
Prices for different load periods

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



First read on electricity price forecasting (EPF)

R.Hyndman: “this paper alone is responsible for 0.7 of the current IF$_{2Y}$=2.642” ;-)



International Journal of Forecasting 30 (2014) 1030–1081

Contents lists available at ScienceDirect

International Journal of Forecasting

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

Review

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

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*
the last 15 years, with varying degrees of success. This review artic
complexity of available solutions, their strengths and weaknesses,
and threats that the forecasting tools offer or that may be encoun
looks ahead and speculates on the directions EPF will or should ta
or so. In particular, it postulates the need for objective comparative
(i) the same datasets, (ii) the same robust error evaluation procedu
testing of the significance of one model's outperformance of another

international journal of forecasting

IJF Hong award for
Energy Forecasting 2013-2014

Rafal 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.

ELSEVIER

International Institute of Forecasters

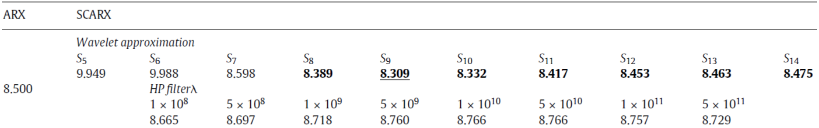
A look into the future of EPF

EPF directions in the next decade (according to Weron, 2014, IJF):

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

Role of the long-term seasonal component (LTSC) for short-term EPF

- Significant prediction accuracy gains possible for linear regression models (Nowotarski & Weron, 2016, ENEECO):



ARX	SCARX									
8.500	*Wavelet approximation*									
	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
	9.949	9.988	8.598	**8.389**	**8.309**	**8.332**	**8.417**	**8.453**	**8.463**	**8.475**
		HP filterλ								
		1×10^8	5×10^8	1×10^9	5×10^9	1×10^{10}	5×10^{10}	1×10^{11}	5×10^{11}	
		8.665	8.697	8.718	8.760	8.766	8.766	8.757	8.729	

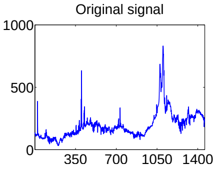
- Unknown effects for non-linear (e.g., ANN) models
- **Is this phenomenon more general?**

Agenda

- Introduction
  - Electricity markets and prices
  - Motivation
- Trend-seasonal components
  - Wavelets
  - The Hodrick-Prescott (HP) filter
- Case study
  - Datasets and LTSCs
  - ARX and SCARX models
  - ANNs in EPF
  - Committee machines of (SC)ANN networks
  - Results and conclusions

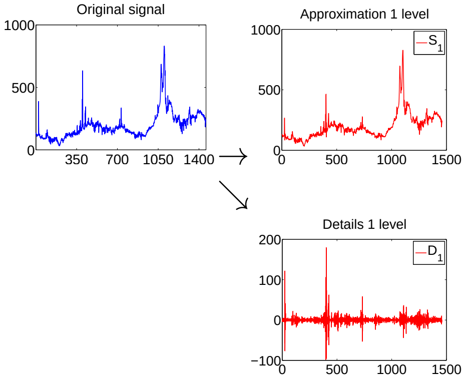
Wavelets

Decomposition of a signal



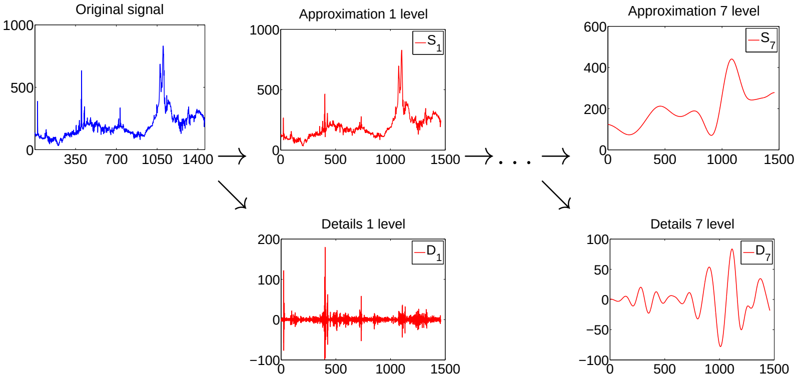
Wavelets

Decomposition of a signal



Wavelets

Decomposition of a signal



Original signal

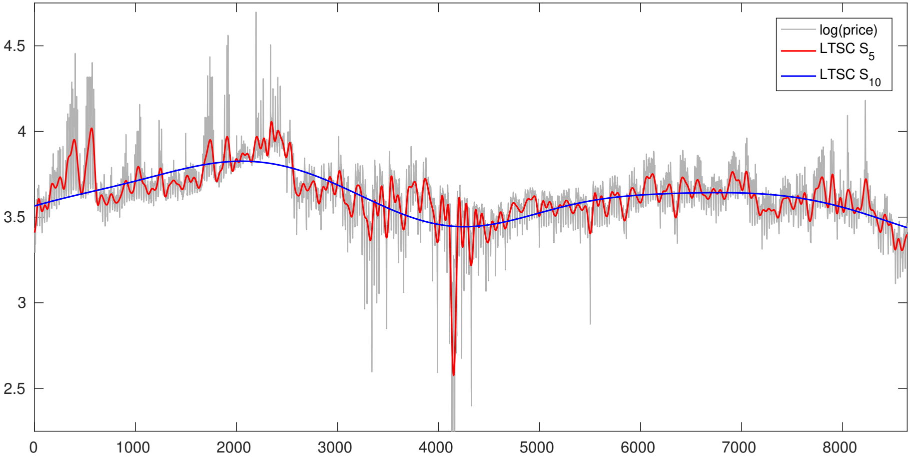
Approximation 1 level

Approximation 7 level

Details 1 level

Details 7 level

Sample fits to Nord Pool data



The Hodrick-Prescott (1980, 1997) filter

A simple alternative to wavelets

- Originally proposed for decomposing GDP into a long-term growth component and a cyclical component
- Returns a smoothed series τ_t for a noisy input series y_t:

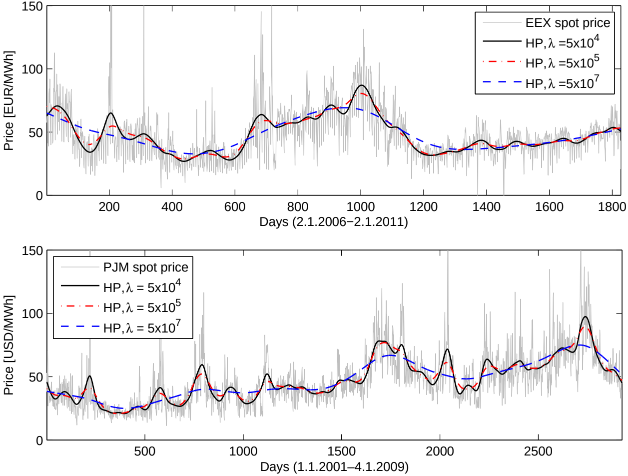
$$\min_{\tau_t}\left\{\sum_{t=1}^{T}(y_t-\tau_t)^2+\lambda\sum_{t=2}^{T-1}\left[(\tau_{t+1}-\tau_t)-(\tau_t-\tau_{t-1})\right]^2\right\},$$

Punish for:

  - deviating from the original series
  - roughness of the smoothed series

Sample fits to EEX and PJM data

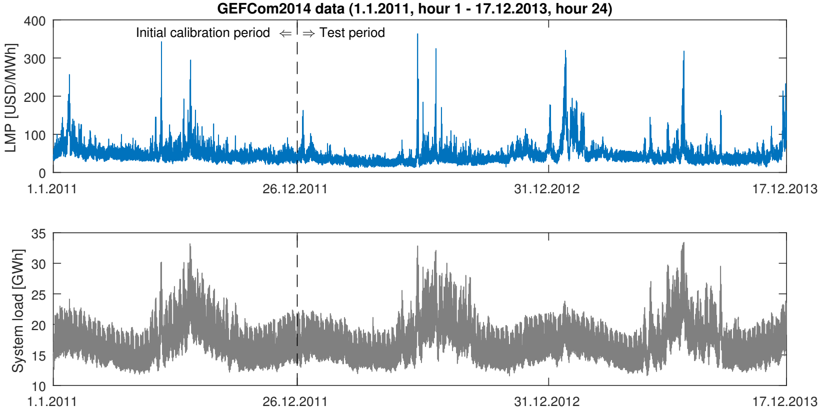
(Weron & Zator, 2015, ENEECO)



Agenda

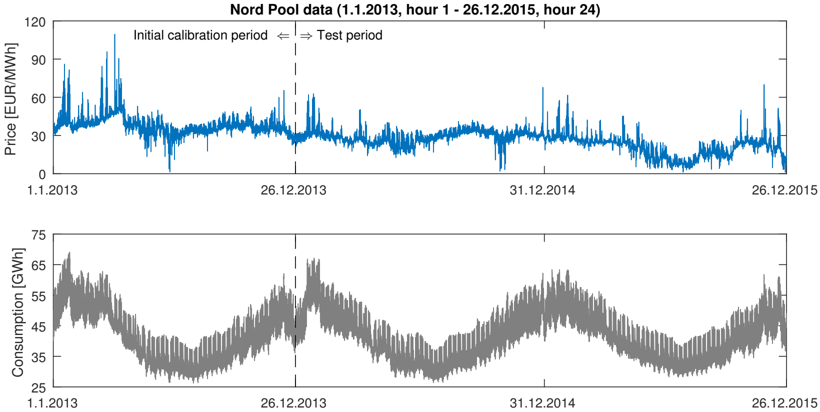
- Introduction
  - Electricity markets and prices
  - Motivation
- Trend-seasonal components
  - Wavelets
  - The Hodrick-Prescott (HP) filter
- Case study
  - Datasets and LTSCs
  - ARX and SCARX models
  - ANNs in EPF
  - Committee machines of (SC)ANN networks
  - Results and conclusions

Datasets: GEFCom 2014



Datasets are the same as in Nowotarski & Weron (2016, ENEECO)

Datasets: Nord Pool



Datasets are the same as in Nowotarski & Weron (2016, ENEECO)

Long-Term Seasonal Components (LTSCs)

Like in Nowotarski & Weron (2016, ENEECO), we consider 18 LTSCs from two categories:

- **Wavelet filters** $S_5, S_6, \ldots, S_{14}$, ranging from 'daily' smoothing ($S_5 \rightarrow 2^5$ hours) up to 'biannual' ($S_{14} \rightarrow 2^{14}$ hours)
  - Models with wavelet filters are denoted by suffixes **-S**$_J$
- **HP-filters** with $\lambda = 10^8, 5 \cdot 10^8, 10^9, \ldots, 5 \cdot 10^{11}$, also ranging from 'daily' up to 'biannual' smoothing
  - Models with HP filters are denoted by suffixes **-HP**$_\lambda$

Benchmark: The **ARX** model

For the log-price, i.e., $p_{d,h} = log(P_{d,h})$, the model is given by:

$$p_{d,h} = \underbrace{\beta_{h,1} p_{d-1,h} + \beta_{h,2} p_{d-2,h} + \beta_{h,3} p_{d-7,h}}_{\text{autoregressive effects}} + \underbrace{\beta_{h,4} p_{d-1,\min}}_{\text{non-linear effect}} + \underbrace{\beta_{h,5} z_t}_{\text{load}} + \underbrace{\sum_{i=1}^{3} \beta_{h,i+5} D_i}_{\text{weekday dummies}} + \varepsilon_{d,h} \qquad (1)$$

- $p_{d-1,min}$ is yesterday's minimum hourly price
- z_t is the logarithm of system load/consumption
- Dummy variables D_1, D_2 and D_3 refer to Monday, Saturday and Sunday, respectively

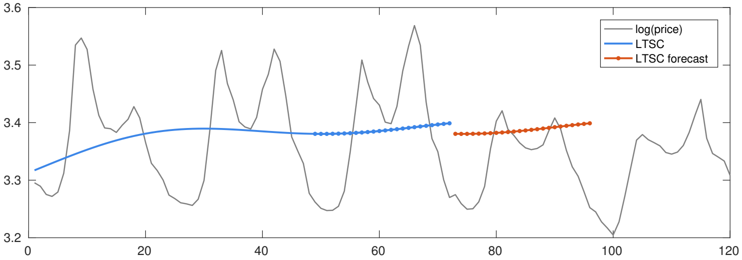
The SCAR modeling framework

(Nowotarski & Weron, 2016, ENEECO)

The Seasonal Component AutoRegressive (SCAR) modeling framework consists of the following steps:

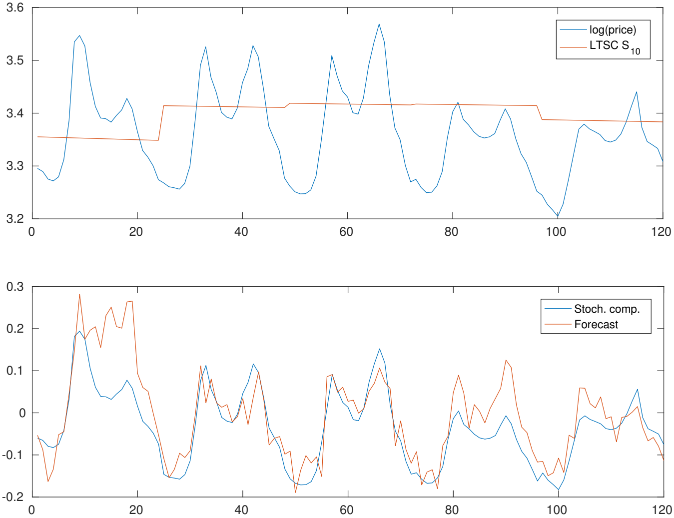
1. (a) Decompose the series in the calibration window into the LTSC $T_{d,h}$ and the stochastic component $q_{d,h}$
   (b) Decompose the exogenous series in the calibration window using the same type of LTSC as for prices
2. Calibrate the **ARX** model to q_t and compute forecasts for the 24 hours of the next day (24 separate series)

The SCAR modeling framework cont.



3. Add stochastic component forecasts $\hat{q}_{d+1,h}$ to persistent forecasts $\hat{T}_{d+1,h}$ of the LTSC to yield log-price forecasts $\hat{p}_{d+1,h}$
4. Convert them into price forecasts of the **SCARX** model, i.e., $\hat{P}_{d+1,h} = \exp\left(\hat{p}_{d+1,h}\right)$

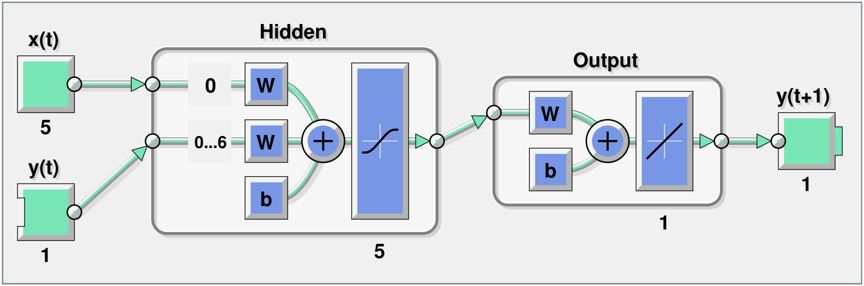
Sample LTSC and stochastic component forecasts



ANNs in other EPF studies

- Variety of ANN implementations, as well as considered inputs, making it impossible to compare with commonly used methods based on linear regression
- Several studies that acknowledge the need of removing seasonal components from time series for neural network models:
    - Andrawis et al. (2011)
    - Zhang and Qi (2005)
    - Keles et al. (2016), the only one in the context of EPF

ANN: Based on Matlab's NARXnet



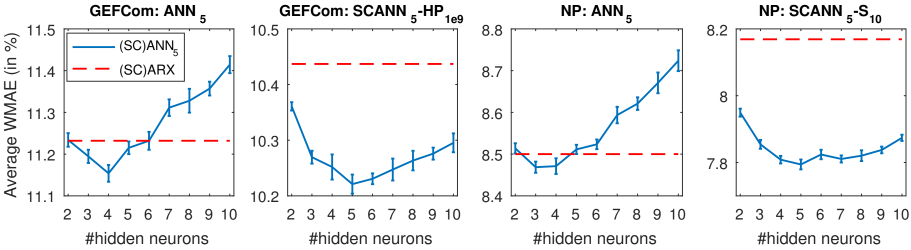
- One hidden layer with 5 neurons and sigmoid activation functions
- Inputs identical as in the **ARX** model
- Trained using Matlab's `trainlm` function, utilizing the Levenberg-Marquardt algorithm for supervised learning

Seasonal Component ANN (SCANN)

The SCANN modeling framework is a generalization of the **ANN** model, analogous to the SCAR framework for the **ARX** model:

1. (a) Decompose the series in the calibration window into the LTSC $T_{d,h}$ and the stochastic component $q_{d,h}$
   (b) Decompose the exogenous series in the calibration window using the same type of LTSC as for prices
2. Calibrate the **ANN** model to q_t and compute forecasts for the 24 hours of the next day (24 separate series)
3. Add stochastic component forecasts $\hat{q}_{d+1,h}$ to persistent forecasts $\hat{T}_{d+1,h}$ of the LTSC to yield log-price forecasts $\hat{p}_{d+1,h}$
4. Convert them into price forecasts of the **SCANN** model, i.e., $\hat{P}_{d+1,h} = \exp\left(\hat{p}_{d+1,h}\right)$

Number of hidden neurons

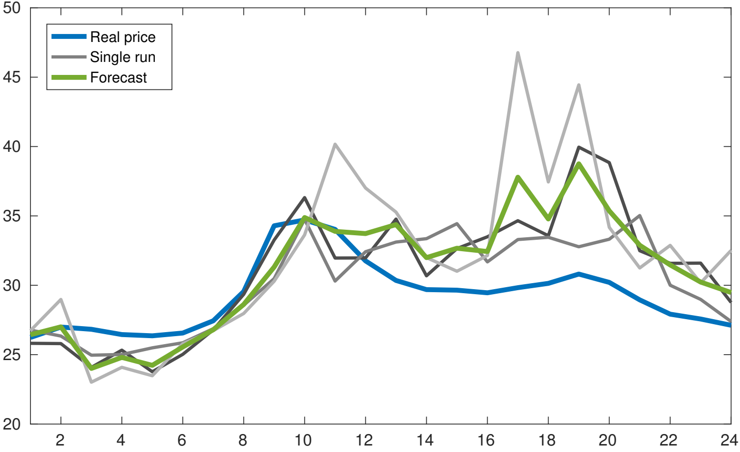


There is no universally optimal number, but the errors are smallest for 4 to 6 neurons in the hidden layer

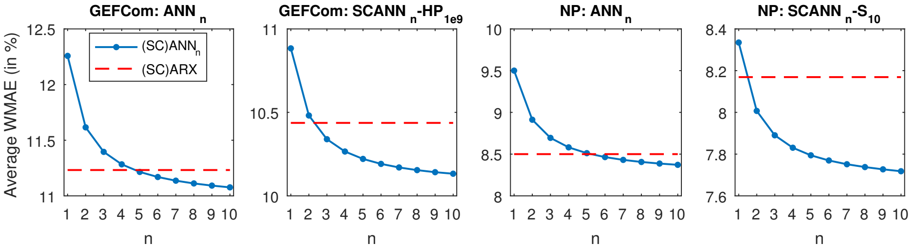
Committee machines of **(SC)ANN** networks

- Every forecast yields slightly different results ⇒ two ‘model categories’ are considered:
  - $\overline{\mathbf{ANN}}_1$ – the ‘expected’ result for a single **ANN** network, an average of error scores across separate runs
  - $\mathbf{ANN}_5$ – a forecast average of 5 runs (hour-by-hour) with identical parameters, a so-called committee machine
- Analogously:
  - $\overline{\mathbf{SCANN}}_1$ – the ‘expected’ result for a single **SCANN** network
  - $\mathbf{SCANN}_5$ – a committee machine of 5 SCANNs

Committee machines of **(SC)ANN** networks



Sample gains from using committee machines



- Forecast errors roughly scale as a power-law function of the number of networks in a committee machine
- We should use as large committee machines as we can ...

Sample gains cont.

- … however, the time needed may be substantial, e.g., for generating forecasts for the next 24 hours:

Model	ARX	SCARX-HP$_{10^8}$	SCARX-S$_9$	ANN$_1$	ANN$_5$
Time	8.6ms	13.5ms	37.3ms	7.6s	38.2s

- SCANN times are omitted here, because LTSC computation is negligible compared to training the ANN

Weekly-weighted Mean Absolute Error (WMAE)

- Following Conejo et al. (2005), Weron & Misiorek (2008) and Nowotarski et al. (2014), among others, we use:

$$\mathrm{WMAE}_w = \frac{1}{\bar{P}_{168}} \mathrm{MAE}_w = \frac{1}{168 \cdot \bar{P}_{168}} \sum_{d=Mon}^{Sun} \sum_{h=1}^{24} \left| P_{d,h} - \hat{P}_{d,h} \right|$$

- where $\bar{P}_{168} = \frac{1}{168} \sum_{d=Mon}^{Sun} \sum_{h=1}^{24} P_{d,h}$

$$\overline{\mathrm{WMAE}} = \frac{1}{w_{max}} \sum_{w=1}^{w_{max}} \mathrm{WMAE}_w$$

- where $w_{max} = 103$ for GEFCom and 104 for Nord Pool

Average WMAE for GEFCom2014

Table 1: Average WMAE in percent for all 103 weeks of the GEFCom2014 out-of-sample test period (*upper half*) or all 104 weeks of the Nord Pool out-of-sample test period (*lower half*). Results for the best performing model in each row are emphasized in bold. Note, that results for the **SCARX** models are the same as in Uniejewski et al. (2017).

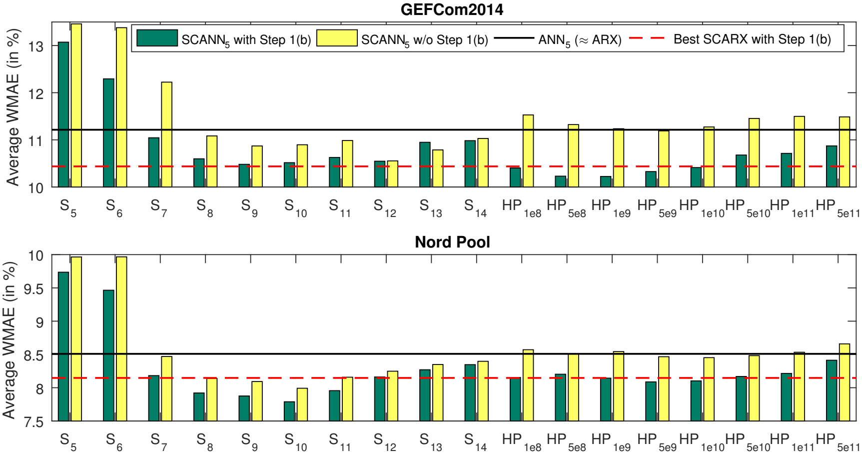
GEFCom2014										
					Benchmarks					
			Naïve	**ARX**	**$\overline{\text{ANN}}_1$**	**ANN_5**				
			14.716	11.232	12.256	**11.214**				
					SCARX / SCANN with wavelet approximation of price and load					
	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
SCARX	12.917	12.226	11.106	10.849	**10.732**	10.776	10.843	10.824	11.100	11.072
$\overline{\text{SCANN}}_1$	13.249	12.555	11.438	**11.066**	11.085	11.216	11.363	11.322	11.784	11.838
SCANN_5	13.072	12.294	11.044	10.598	**10.481**	10.516	10.627	10.547	10.948	10.983
					SCARX / SCANN with HP filter on price and load (λ)					
	10^8	$5 \cdot 10^8$	10^9	$5 \cdot 10^9$	10^{10}	$5 \cdot 10^{10}$	10^{11}	$5 \cdot 10^{11}$		
SCARX	10.519	10.447	**10.437**	10.495	10.559	10.798	10.897	11.060		
$\overline{\text{SCANN}}_1$	10.957	**10.859**	10.893	11.044	11.159	11.534	11.581	11.896		
SCANN_5	10.403	10.230	**10.224**	10.327	10.412	10.678	10.713	10.872		

Average WMAE for Nord Pool

Table 1: Average WMAE in percent for all 103 weeks of the GEFCom2014 out-of-sample test period (*upper half*) or all 104 weeks of the Nord Pool out-of-sample test period (*lower half*). Results for the best performing model in each row are emphasized in bold. Note, that results for the **SCARX** models are the same as in Uniejewski et al. (2017).

					Nord Pool					
					Benchmarks					
			Naïve	**ARX**	**$\overline{\text{ANN}}_1$**	**ANN$_5$**				
			9.661	**8.500**	9.517	8.509				
					SCARX / SCANN with wavelet approximation of price and load					
	S_5	S_6	S_7	S_8	S_9	S_{10}	S_{11}	S_{12}	S_{13}	S_{14}
SCARX	9.834	9.761	8.411	8.205	**8.147**	8.169	8.319	8.351	8.484	8.389
$\overline{\text{SCANN}}_1$	10.004	9.750	8.597	8.342	8.359	**8.323**	8.570	8.849	9.035	9.185
SCANN$_5$	9.736	9.465	8.182	7.921	7.876	**7.789**	7.956	8.162	8.270	8.347
					SCARX / SCANN with HP filter on price and load (λ)					
	10^8	$5 \cdot 10^8$	10^9	$5 \cdot 10^9$	10^{10}	$5 \cdot 10^{10}$	10^{11}	$5 \cdot 10^{11}$		
SCARX	**8.475**	8.512	8.536	8.601	8.621	8.655	8.663	8.670		
$\overline{\text{SCANN}}_1$	**8.575**	8.682	8.667	8.613	8.645	8.800	8.907	9.180		
SCANN$_5$	8.154	8.203	8.144	**8.088**	8.103	8.169	8.215	8.413		

Aggregate results of SCANN performance



Note: Step 1(b) is important (green vs. yellow)!

The Diebold-Mariano test (1995)

- We define the error function as

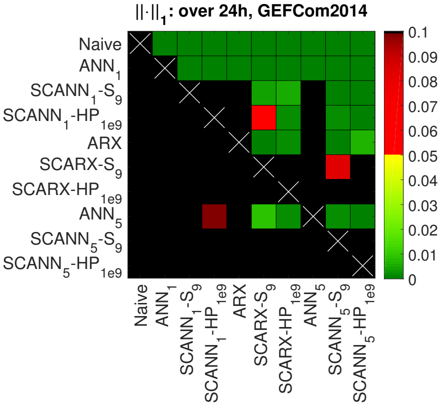
$$L(\varepsilon_d) = ||\varepsilon_d||_1 = \sum_{h=1}^{24} |P_{d,h} - \hat{P}_{d,h}|$$

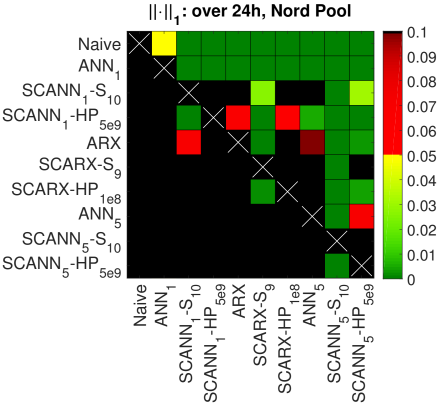
- For each pair of models we compute the loss differential

$$D_d = L(\varepsilon_d^{model_X}) - L(\varepsilon_d^{model_Y})$$

- Hypothesis H_0:$E(D_d) \leq 0$, $model_X$ outperforms $model_Y$
- Reversed hypothesis H_0^R:$E(D_d) \geq 0$, $model_Y$ outperforms $model_X$

Diebold-Mariano test results





Conclusions

- Using Seasonal Component ANN (SCANN) models can yield statistically significant improvement over the ANN benchmark
    - **SCANN$_5$** returns 0.72–0.99% lower WMAE than **ANN$_5$**
- The accuracy gains from using LTSC are greater in ANN models than in regression models
    - **SCARX** models yield only a 0.35–0.80% improvement in WMAE vs. the **ARX** benchmark
- Forecast averaging is crucial in outperforming the **SCARX** model
    - **SCANN$_5$** yields 0.21–0.36% lower WMAE