On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks

Grzegorz Marcjasz\textsuperscript{1,2}, Bartosz Uniejewski\textsuperscript{1,2} & Rafał Weron\textsuperscript{1}

\textsuperscript{1}Department of Operations Research
and
\textsuperscript{2}Faculty of Pure and Applied Mathematics
Wrocław University of Science and Technology, Poland

*Partially supported by the National Science Center (NCN, PL) through grant no. 2015/17/B/HS4/00334
LTSC and short-term price forecasting

Does removing the long-term seasonal component (LTSC) improve short-term (day-ahead) electricity price forecasts?
Yes, it does!

On the importance of the long-term seasonal component in day-ahead electricity price forecasting

Jakub Nowotarski, Rafał Weron*

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

**ABSTRACT**

In day-ahead electricity price forecasting (EPF) the daily and weekly seasonalties are always taken into account, but the long-term seasonal component (LTSC) is believed to add unnecessary complexity to the already parameter-rich models and is generally ignored. Conducting an extensive empirical study involving state-of-the-art time series models we show that (i) decomposing a series of electricity prices into a LTSC and a stochastic component; (ii) modeling them independently and (iii) combining their forecasts can bring – contrary to a common belief – an accuracy gain compared to an approach in which a given time series model is calibrated to the prices themselves.
Yes, it does!

Percentage (no.) of hours for which a SCARX model significantly outperforms the ARX benchmark (#better) and vice versa (#worse):

<table>
<thead>
<tr>
<th></th>
<th>GEFCom2014</th>
<th>Nord Pool</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCARX-S_{12}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#better</td>
<td>46% (11)</td>
<td>29% (7)</td>
</tr>
<tr>
<td>#worse</td>
<td>0% (0)</td>
<td>17% (4)</td>
</tr>
<tr>
<td>mSCARX-S_{12}</td>
<td></td>
<td></td>
</tr>
<tr>
<td>#better</td>
<td>92% (22)</td>
<td>42% (10)</td>
</tr>
<tr>
<td>#worse</td>
<td>0% (0)</td>
<td>0% (0)</td>
</tr>
</tbody>
</table>

- Only Seasonal Component ARX vs. ARX models tested
- But is this phenomenon more general? E.g., for ANNs?
This study is based on a forthcoming paper:

*On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks*

Grzegorz Marcjasz\textsuperscript{a,b}, Bartosz Uniejewski\textsuperscript{a,b}, Rafał Weron\textsuperscript{a}

\textsuperscript{a}Department of Operations Research, Wrocław University of Science and Technology, Wrocław, Poland
\textsuperscript{b}Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland

**Abstract**

In day-ahead electricity price forecasting the daily and weekly seasonalties are always taken into account, but the long-term seasonal component was believed to add unnecessary complexity and in most studies ignored. The recent introduction of the *Seasonal Component AutoRegressive* (SCAR) modeling framework has changed this viewpoint. However, the latter is based on linear models estimated using Ordinary Least Squares. Here we show that considering non-linear autoregressive (NARX) neural network-type models with the same inputs as the corresponding SCAR-type model
Study setup
The same as in Nowotarski & Weron (2016)
18 wavelet and HP-filter based LTSCs

- **Wavelet filters** \((-S_J)\): \(S_5, S_6, \ldots, S_{14}\), ranging from ‘daily’ smoothing \((S_5 \rightarrow 2^5\) hours) up to ‘biannual’ \((S_{14} \rightarrow 2^{14}\) hours)

- **HP-filters** \((-\text{HP}_\lambda)\): with \(\lambda = 10^8, 5 \cdot 10^8, 10^9, \ldots, 5 \cdot 10^{11}\)
The **ARX** model

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

\[
p_{d,h} = \beta_{h,1}p_{d-1,h} + \beta_{h,2}p_{d-2,h} + \beta_{h,3}p_{d-7,h} + \beta_{h,4}p_{d-1,\text{min}} + \beta_{h,5}z_t + \sum_{i=1}^{3} \beta_{h,i+5}D_i + \epsilon_{d,h}
\]

- \( p_{d-1,\text{min}} \) 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; Uniejewski, Marcjasz & Weron, 2017, WP)

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

1. (a) Decompose the log-price 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., $\hat{P}_{d+1,h} = \exp(\hat{p}_{d+1,h})$
ANNs in other EPF studies

- A variety of ANN implementations
- Different datasets and inputs → impossible to compare with published studies that use regression models
- A few papers acknowledge the need for deseasonalizing data before fitting 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
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

- The training function finds only local minima and initial weights are random
- Every forecast yields slightly different results \(\Rightarrow\) two ‘models’ are considered:
  - \(\text{ANN}_1\) – the ‘expected’ result for a single ANN network, an average of error scores across separate runs
  - \(\text{ANN}_5\) – a forecast average of five runs (hour-by-hour) with identical parameters, a so-called committee machine
Committee machines of (SC)ANN networks

Marcjasz et al. (Wrocław, PL)
Seasonal Component (SCANN) models 15.12.2017, EFC17, Kraków
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...
... however, the time needed may be substantial, e.g., for generating forecasts for the next 24 hours:

<table>
<thead>
<tr>
<th>Model</th>
<th>ARX</th>
<th>SCARX-HP_{10^8}</th>
<th>SCARX-S_9</th>
<th>ANN_1</th>
<th>ANN_5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>8.6ms</td>
<td>13.5ms</td>
<td>37.3ms</td>
<td>7.6s</td>
<td>38.2s</td>
</tr>
</tbody>
</table>

SCANN times are omitted here, because LTSC computation is negligible compared to training the ANN.
Results
Weekly-weighted Mean Absolute Error (WMAE)

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

\[
WMAE_w = \frac{1}{\bar{P}_{168}} \text{MAE}_w = \frac{1}{168 \cdot \bar{P}_{168}} \sum_{d=\text{Mon}}^{\text{Sun}} \sum_{h=1}^{24} \left| P_{d,h} - \hat{P}_{d,h} \right|
\]

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

\[
\bar{WMAE} = \frac{1}{w_{\text{max}}} \sum_{w=1}^{w_{\text{max}}} WMAE_w
\]

where \( w_{\text{max}} = 103 \) for GEFCom and 104 for Nord Pool
Aggregate results of SCANN performance

Note: Step 1(b) is important (green vs. yellow)!
Testing for significance: Diebold-Mariano

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

\[ D_d = L(\varepsilon_{d}^{model_X}) - L(\varepsilon_{d}^{model_Y}) \]

- \( H_0: E(D_d) \leq 0 \), forecasts of \( model_X \) outperform those of \( model_Y \)
- \( H_0^R: E(D_d) \geq 0 \), i.e., the reverse hypothesis
Diebold-Mariano test: $p$-values

The closer are the $p$-values to zero (dark green) the more significant is the difference between the forecasts of a model on the X-axis ($\rightarrow$ better) and those of a model on the Y-axis ($\rightarrow$ worse).
Alternative networks and training methods
Alternative training methods: MATLAB’s trainbr

- Better results for a single run than with trainlm: (SC)ANN₁
- Lower gains from averaging: (SC)ANN₅

![Graphs showing comparison between trainlm, trainbr, and SCARX for GEFCom HP₁₀⁹, GEFCom S₉, NordPool S₁₀, and NordPool HP₅·₁₀⁹]
Alternative training methods: PythonFANN

- Python interface to the Fast Artificial Neural Network Library
- Unmatched performance for some data periods ... 
- ... while on average similar to MATLAB’s `trainlm`
- Training algorithms and execution time
- Default training parameters $\Rightarrow$ terrible forecasting performance
Radial Basis Function (RBF) networks

- Potentially good interpolation with many radial basis functions
- Iteratively adds hidden neurons $\Rightarrow$ high computational cost
(Layer) Recurrent Neural Networks (RNNs)

- Dynamic temporal behavior $\Rightarrow$ prediction largely based on the input sequence
- Various types, e.g., Long Short-Term Memory (LSTM) networks
Conclusions
Conclusions

- Using Seasonal Component ANN (SCANN) models can yield statistically significant improvement over the ANN benchmark
  - \( \text{SCANN}_5 \) returns 0.72–0.99% lower WMAE than \( \text{ANN}_5 \)

- The accuracy gains from using LTSC are greater in ANN models than in regression models
  - \( \text{SCARX} \) models yield only a 0.35–0.80% improvement in WMAE vs. the \( \text{ARX} \) benchmark
Conclusions cont.

- Forecast averaging is crucial in outperforming the **SCARX** model
  - $\text{SCANN}_5$ yields 0.21–0.36% lower WMAE than corresponding **SCARX** models ...
  - whereas $\text{SCANN}_1$ returns 0.22–1.02% higher WMAE than **SCARX**
On the importance of the long-term seasonal component in day-ahead electricity price forecasting with NARX neural networks

Grzegorz Marcjasz\textsuperscript{a,b}, Bartosz Uniejewski\textsuperscript{a,b}, Rafał Weron\textsuperscript{a}

\textsuperscript{a}Department of Operations Research, Wrocław University of Science and Technology, Wrocław, Poland
\textsuperscript{b}Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland

Abstract

In day-ahead electricity price forecasting the daily and weekly seasonalities are always taken into account, but the long-term seasonal component was believed to add unnecessary complexity and in most studies ignored. The recent introduction of the Seasonal Component AutoRegressive (SCAR) modeling framework has changed this viewpoint. However, the latter is based on linear models estimated using Ordinary Least Squares. Here we show that considering non-linear autoregressive (NARX) neural network-type models with the same inputs as the corresponding SCAR-type model


