

# Calibration window selection based on change-point detection for forecasting electricity prices\*

Julia Nasiadka<sup>[0000–0003–3031–0117]</sup>, Weronika Nitka<sup>\*\*[0000–0002–3360–4260]</sup>, and  
Rafał Weron<sup>[0000–0003–1619–5239]</sup>

Department of Operations Research and Business Intelligence,  
Wrocław University of Science and Technology, 50-370 Wrocław, Poland  
[weronika.nitka@pwr.edu.pl](mailto:weronika.nitka@pwr.edu.pl)

**Abstract.** We employ a recently proposed change-point detection algorithm, the Narrowest-Over-Threshold (NOT) method, to select subperiods of past observations that are similar to the currently recorded values. Then, contrarily to the traditional time series approach in which the most recent  $\tau$  observations are taken as the calibration sample, we estimate autoregressive models only for data in these subperiods. We illustrate our approach using a challenging dataset – day-ahead electricity prices in the German EPEX SPOT market – and observe a significant improvement in forecasting accuracy compared to commonly used approaches, including the Autoregressive Hybrid Nearest Neighbors (ARHNN) method.

**Keywords:** change-point detection; Narrowest-Over-Threshold method; electricity price forecasting; autoregressive model; calibration window

## 1 Introduction

Electricity price forecasting (EPF) is an extremely challenging task. A number of methods have been developed for this purpose, ranging from linear regression to hybrid deep learning architectures utilizing long-short term memory and/or convolutional neural networks. While most studies focus on improving model structures, selecting input features with more predictive power or implementing more efficient algorithms [3,5,6], the issue of the optimal calibration window is generally overlooked [4].

This work is inspired by a recent article [2], which utilized a relatively simple change-point detection method [9] to split the time series into segments with the ‘same’ price level, and an ICCS 2021 paper [7], which employed the  $k$ -nearest neighbors ( $k$ -NN) algorithm to select the calibration sample based on similarity over a subset of explanatory variables. Here, we utilize a recently proposed change-point detection algorithm – the Narrowest-Over-Threshold (NOT)

---

\* Supported by the Ministry of Education & Science (MEiN, Poland) through Grant No. 0027/DIA/2020/49 (to WN) and the National Science Center (NCN, Poland) through Grant No. 2018/30/A/HS4/00444 (to RW).

\*\* Corresponding author

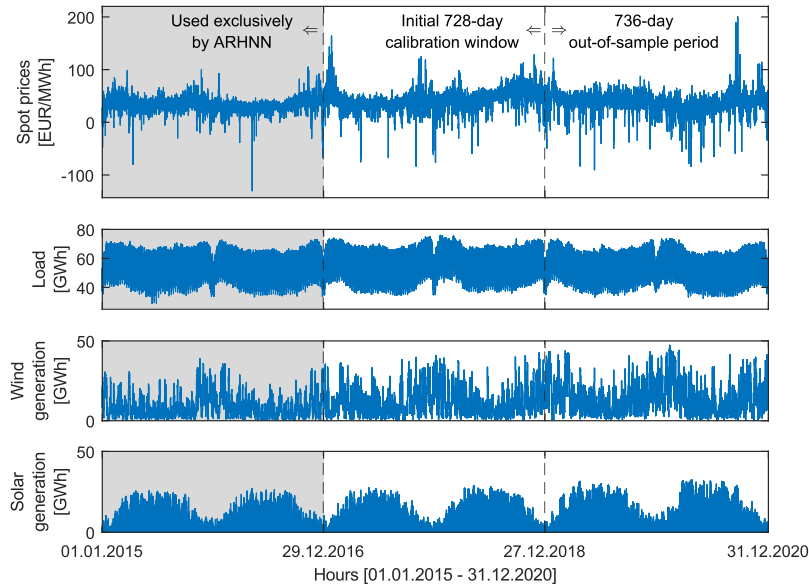


Fig. 1: Electricity spot prices and day-ahead load, wind and solar power generation forecasts in Germany. The last 736 days constitute the test period.

method [1] – to construct an automatic method for detecting subperiods exhibiting different temporal dynamics. Once identified, those not resembling the current behavior are discarded when estimating the predictive model. In what follows, we provide empirical evidence that significant improvement in forecasting accuracy can be achieved compared to commonly used EPF approaches.

The remainder of the paper is structured as follows. In Section 2 we present the dataset and the transformation, which is used to standardize the data. In Section 3 we briefly describe the NOT method and introduce our approach to selecting subperiods for model calibration. Next, in Section 4 we present the forecasting models and in Section 5 the empirical results. Finally, in Section 6 we conclude and discuss future research directions.

## 2 The data

For comparison purposes, we use the same dataset as in [7]. It spans six years (2015-2020) at hourly resolution and includes four series from the German EPEX SPOT market: electricity spot prices  $P_{d,h}$  (more precisely: prices set in the day-ahead auction on day  $d - 1$  for the 24h of day  $d$ ) and day-ahead load  $\hat{L}_{d,h}$ , wind  $\hat{W}_{d,h}$  and solar power generation  $\hat{S}_{d,h}$  forecasts, see Fig. 1. The first two years are exclusively used for estimating the Autoregressive Hybrid Nearest Neighbors (ARHNN) method [7]; the remaining methods require less data for calibration. The last 736 days constitute the out-of-sample test period.

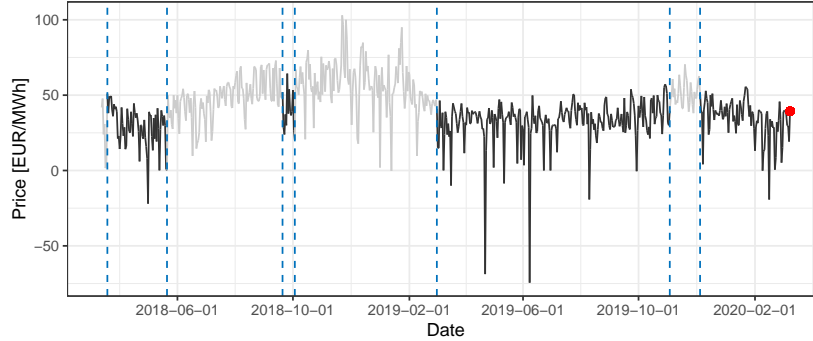


Fig. 2: A sample run of the algorithm introduced in Section 3. The red dot is the target day. Vertical dashed lines indicate the located change-points separating periods with different statistical properties. The discarded prices are in gray.

The most distinct feature of the German power market are frequent spikes and negative prices. Similarly volatile is wind energy generation, while load and solar generation are more predictable. Following [8,10], to cope with this extreme volatility, we transform the electricity prices using the area hyperbolic sine:  $Y_{d,h} = \text{asinh}(\frac{1}{b}\{P_{d,h} - a\})$  with  $\text{asinh}(x) = \log\{x + (x^2 + 1)^{0.5}\}$ , where  $a$  is the median of  $P_{d,h}$  in the calibration window and  $b$  is median absolute deviation. The price forecasts are then obtained by the inverse transformation.

Note that in [7] a different transformation was used. All series, not just  $P_{d,h}$ , were normalized by subtracting the mean and dividing by the standard deviation in each calibration window. We denote models utilizing *asinh*-transformed data with subscript H; the remaining ones use the standard normalization, as in [7].

### 3 Calibration window selection using NOT

The Narrowest-Over-Threshold (NOT) method [1] can detect an unknown number of change-points at unknown locations in one-dimensional time series data. The key feature is its focus on the smallest local sections of the data on which the existence of a change-point is suspected. A change-point is said to occur when the behavior of the series changes significantly [2]. See [www.changepoint.info](http://www.changepoint.info) for an excellent review site and software repository on this topic. Said differently, change-points split the data into stationary subseries, see Fig. 2. This is what makes them interesting for model calibration and forecasting.

Our algorithm for calibration window selection, i.e., identifying periods with similar time series dynamics to the currently observed, is as follows:

1. Set the maximum number  $N_c^{max}$  of change-points to be identified.
2. Use the NOT method to identify  $N_c \in [0, N_c^{max}]$  change-points  $c_i$ ,  $i = 1, \dots, N_c$ , in the initial calibration window  $\mathcal{C}_0$  of length  $\tau$ . Additionally, denote the first observation in  $\mathcal{C}_0$  by  $c_0$ .

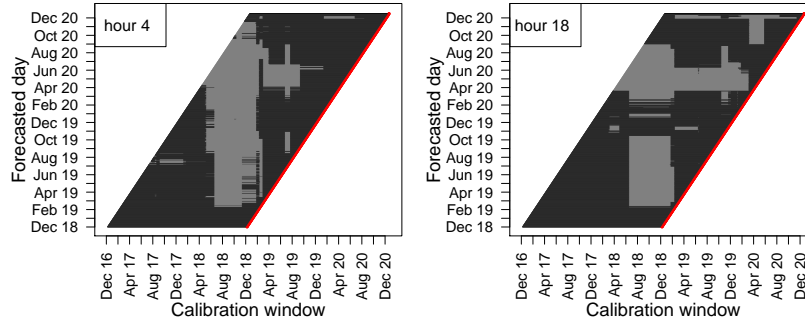


Fig. 3: Overview of NOT-selected (black) and discarded (gray) periods in the two-year calibration window (dates on the  $x$  axis) for hours 4 (*left panel*) and 18 (*right panel*). The red line indicates the forecasted day (date on the  $y$  axis).

3. If  $N_c = 0$  return calibration sample  $\mathbf{C} = \mathbf{C}_0$ . Otherwise, compute the empirical quantiles  $q_{low}$  and  $q_{high}$  of the observations within the period between the most recent change-point found and the last observation in  $\mathbf{C}_0$ .
4. For every interval between two subsequent change-points  $c_{i-1}$  and  $c_i$  compute the median  $m_i$  of its observations,  $i = 1, \dots, N_c - 1$ .
5. Set  $\mathbf{C} = \bigcup_{m_i \in (q_{low}, q_{high})} [c_{i-1}, c_i]$ .

Based on a limited simulation study, we set the maximum number of change-points  $N_c^{max} = 12$  and the order of quantiles  $(q_{low}, q_{high}) = (q_{0.025}, q_{0.975})$ . We also use the least constraining form of NOT, i.e., we assume that the data have piecewise continuous variance and piecewise continuous mean. Any deviations from this are treated as a breach of stationarity.

A sample run of the algorithm is presented in Fig. 2. In the plot, the period closest to the forecasted day (red dot) is characterized by relatively stable, low prices with a small variance. The algorithm discards the light gray subperiods, when either the prices or their variance are significantly higher. In Figure 3 we illustrate the results for two sample hours and the whole test window. We use a rolling scheme, i.e., once forecasts for the 24 hours of the first day in the test sample are computed, the calibration window is moved forward by one day and forecasts for the 2nd day in the test sample are calculated. A clear pattern of vertical gray stripes emerges, meaning that for a range of windows the change-points are consistently detected on the same or neighboring days. Comparing these plots with the price trajectory in Fig. 1, we can observe that much fewer observations are selected by NOT when the prices tend to be more spiky, as can be seen in Spring 2020 (Apr 20 – Jun 20 on the  $y$  axis, esp. for hour 18).

## 4 Forecasting models

For comparison purposes, the underlying model we use is the same as in [7]. It is an autoregressive structure with exogenous variables dubbed ARX. Since the

prices  $P_{d,h}$  are set in the day-ahead auction on day  $d - 1$  independently for the 24h of day  $d$ , it is customary in the EPF literature [4,6] to treat every hour as a separate time series. Hence, we consider 24 ARX models of the form:

$$\begin{aligned}
 P_{d,h} = & \underbrace{\alpha_h \mathbf{D}_d}_{\text{Dummies}} + \underbrace{\sum_{p \in \{1,2,7\}} \beta_{h,p} P_{d-p,h}}_{\text{AR component}} + \underbrace{\theta_{h,1} P_{d-1,min} + \theta_{h,2} P_{d-1,max}}_{\text{Yesterday's price range}} \\
 & + \underbrace{\theta_{h,3} P_{d-1,24}}_{\text{Last known price}} + \underbrace{\theta_{h,4} \hat{L}_{d,h} + \theta_{h,5} \hat{W}_{d,h} + \theta_{h,6} \hat{S}_{d,h}}_{\text{Exogenous variables}} + \underbrace{\varepsilon_{d,h}}_{\text{Noise}}.
 \end{aligned} \tag{1}$$

The autoregressive (AR) dynamics are captured by the lagged prices from the same hour yesterday, two and seven days ago. Following [10], yesterday’s minimum  $P_{d-1,min}$ , maximum  $P_{d-1,max}$  and the last known price  $P_{d-1,24}$ , as well as day-ahead predictions of the three exogenous variables are included. Finally, a  $1 \times 7$  vector of dummy variables  $\mathbf{D}_d$  is used to represent the weekly seasonality and the uncertainty is represented by white noise.

Overall, we compare seven types of approaches that all use ARX as the underlying model. The first three are the same as in [7]: (i) **Win**( $\tau$ ) – the ARX model estimated using a window of  $\tau$  days, with  $\tau \in [56, 57, \dots, 728]$ , (ii) **Av(Win)** – the arithmetic average of six forecasts of the ARX model for three short ( $\tau = 56, 84, 112$ ) and three long windows ( $\tau = 714, 721, 728$ ), and (iii) the **ARHNN** model. The next four include: (iv) **Win<sub>H</sub>**( $\tau$ ) – the same as Win( $\tau$ ) but calibrated to *asinh*-transformed prices, (v) **NOT<sub>H</sub>(728)** – the ARX model calibrated to *asinh*-transformed prices in NOT-selected subperiods from the 728-day window, (vi) **Av(Win<sub>H</sub>)** – the same as Av(Win) but calibrated to *asinh*-transformed prices, and (vii) **Av(NOT<sub>H</sub>)** – the same as Av(Win<sub>H</sub>) but with the forecasts for the three long windows ( $\tau = 714, 721, 728$ ) replaced by NOT<sub>H</sub>(728). The rationale behind the latter averaging scheme is that NOT<sub>H</sub>( $\tau$ ) performs best for long calibration windows and offers little or even no gain for  $\tau < 1$  year.

## 5 Results

We evaluate the forecasting performance of the seven approaches presented in Section 4 in terms of the *root mean squared error* (RMSE; results for the mean absolute error are similar and available from the authors upon request). The RMSE values reported in Fig. 4 are aggregated (averaged) across all hours in the 736-day test sample, see Fig. 1. Additionally, to test the significance of differences in forecasting accuracy, for each pair of models we employ the multivariate variant of the Diebold-Mariano (DM) test, as proposed in [10].

Several conclusions can be drawn. Firstly, changing the preprocessing method from normalization [7] to *asinh* transformation [8] generally reduces the RMSE. Even the worst performing out of the latter approaches, Win<sub>H</sub>(728), improves on ARHNN, the most accurate method in [7]. Secondly, NOT-selection yields further improvement, although not statistically significant if considered on its own. Compare NOT<sub>H</sub>(728) with Win<sub>H</sub>(728) and Av(NOT<sub>H</sub>) with Av(Win<sub>H</sub>).

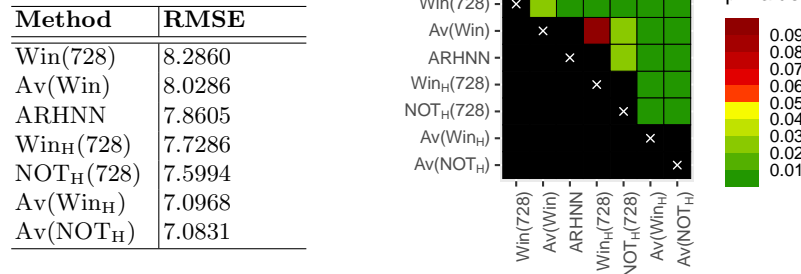


Fig. 4: RMSE errors in the out-of-sample test period (*left panel*). A heatmap of the  $p$ -values for the multivariate Diebold-Mariano test [10] for each pair of methods (*right panel*). The smaller the  $p$ -values, the more significant is the difference between the forecasts of a model on the  $x$ -axis (better) and the forecasts of a model on the  $y$ -axis (worse). Black color indicates  $p$ -values in excess of 0.1.

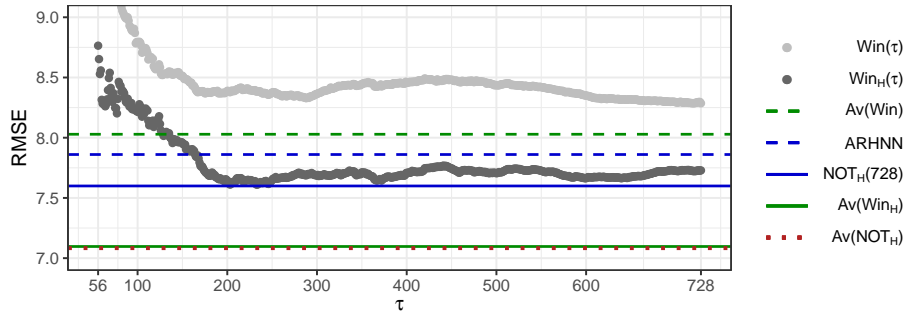


Fig. 5: RMSE values for all considered models;  $\tau$  is the calibration window length.

The RMSE values for all considered approaches are presented graphically in Figure 5. It clearly shows the significant improvement from using the *as-inh* – compare between  $\text{Win}(\tau)$  with  $\text{Win}_H(\tau)$  for all  $\tau$ 's. While  $\text{Win}_H(728)$  is not the best performing of all  $\text{Win}_H(\tau)$  models, the differences in performance are relatively minor for  $\tau \geq 200$  days. Even the best ex-post known model,  $\text{Win}_H(233)$ , is slightly worse than  $\text{NOT}_H(728)$ . The averaged forecasts  $\text{Av}(\text{Win}_H)$  and  $\text{Av}(\text{NOT}_H)$  are further able to improve on the accuracy, although the differences between them are not significant.

## 6 Conclusions and discussion

In this paper we propose a novel method for selecting calibration subperiods based on Narrowest-Over-Threshold (NOT) change-point detection [1]. Contrarily to the traditional time series approach in which the most recent observations are taken as the calibration sample, we propose to estimate the predictive mod-

els only using data in the selected subperiods. We evaluate our approach using German electricity market data and seven variants of autoregressive models tailored for electricity price forecasting (EPF). We provide empirical evidence that significant improvement in forecasting accuracy can be achieved compared to commonly used EPF approaches, including the recently proposed ARHNN [7]. In addition to calibration sample selection, our results also emphasize the importance of using transformations like the *asinh*, in line with [8,10].

The roughly sixfold increase in computational time of the NOT-based methods – 4.97s for 24 hourly forecasts using NOT<sub>H</sub>(728) vs. 0.83s using Win<sub>H</sub>(728), running R ver. 3.6.3 on an i7-9750H processor – can be seen as a drawback, especially compared to less complex ways of improving forecast accuracy, like calibration window averaging [4]. However, the automation of the forecasting process may make the trade-off worthwhile. If this is the case for more complex models than the autoregressive ones considered here or the shallow neural network in [2], e.g., LASSO-estimated AR (LEAR) and deep neural networks [6], is left for future work.

## References

1. Baranowski, R., Chen, Y., Fryzlewicz, P.: Narrowest-over-threshold detection of multiple change-points and change-point-like features. *Journal of the Royal Statistical Society B* **81**(3), 649–672 (2019)
2. De Marcos, R., Bunn, D., Bello, A., Reneses, J.: Short-term electricity price forecasting with recurrent regimes and structural breaks. *Energies* **13**(20), 5452 (2020)
3. Heijden, T., Lago, J., Palensky, P., Abraham, E.: Electricity price forecasting in European day ahead markets: A greedy consideration of market integration. *IEEE Access* **9**, 119954–119966 (2021)
4. Hubicka, K., Marcjasz, G., Weron, R.: A Note on Averaging Day-Ahead Electricity Price Forecasts Across Calibration Windows. *IEEE Transactions on Sustainable Energy* **10**(1), 321–323 (2019)
5. Jahangir, H., Tayarani, H., Baghali, S., et al.: A novel electricity price forecasting approach based on dimension reduction strategy and rough artificial neural networks. *IEEE Transactions on Industrial Informatics* **16**(4), 2369–2381 (2020)
6. Lago, J., Marcjasz, G., Schutter, B.D., Weron, R.: Forecasting day-ahead electricity prices: A review of state-of-the-art algorithms, best practices and an open-access benchmark. *Applied Energy* **293**, 116983 (2021)
7. Nitka, W., Serafin, T., Sotiros, D.: Forecasting Electricity Prices: Autoregressive Hybrid Nearest Neighbors (ARHNN) Method. *LNCS* **12745**, 312–325 (2021)
8. Uniejewski, B., Weron, R., Ziel, F.: Variance stabilizing transformations for electricity spot price forecasting. *IEEE Transactions on Power Systems* **33**(2), 2219–2229 (2018)
9. Zeileis, A., Kleiber, C., Walter, K., Hornik, K.: Testing and dating of structural changes in practice. *Computational Statistics & Data Analysis* **44**, 109–123 (2003)
10. Ziel, F., Weron, R.: Day-ahead electricity price forecasting with high-dimensional structures: Univariate vs. multivariate modeling frameworks. *Energy Economics* **70**, 396–420 (2018)