Short-term Electricity Price Forecasting Using Generalized Additive Models

Jan-Hendrik Meier[1][0000-0002-3080-2210], Stephan Schneider[1][0000-0003-1810-8813], Chan Le[1]

[1] Kiel University of Applied Sciences, Sokratesplatz 2, 24149 Kiel, Germany
jan-hendrik.meier|stephan.schneider@fh-kiel.de

Abstract. If one examines the spot price series of electrical power over the course of time, it is striking that the electricity price across the day takes a course that is determined by power consumption following a day and night rhythm. The daily course changes in its height and temporal extent in both, the course of the week, as well as with the course of the year. This study deals methodologically with this intra-day and seasonal behaviour. We contribute the usage of Generalized Additive Models (GAM) and apply these models with European data.

Keywords: electricity prices, forecasting, generalized additive models.

1 Introduction

Since the come about of energy deregulation in the 1990s, the electric power industry has undergone significant restructuring, driving the market away from its natural monopoly and opening chances for thriving competition and reduction in prices through privatization. As a result, the last two decades have seen a remarkable rise in importance of electricity price forecasting (EPF). Invaluable inputs are provided in aid of optimal decisions and responses from both producers and retailers in the pool-based market.

Electricity, though conforming to the definition of a commodity [11], is a special case with very distinct characteristics: non-storability of electricity, inelasticity of the short-term demand, wide spectrum of cost, and oligopolistic behavior of the generators [20]. Without any loss-free form of storage, it is crucial that great effort is needed to ensure and maintain the stability of a balanced supply and demand [10]. Hence, there are many challenges in modeling electricity prices.

In comparison to the time series of the electricity load, Aggarwal et al. [1] mentioned that the series of the electricity price oftentimes contains patterns of much greater complexity, including non-constant mean and variance, strong seasonality and various calendar effects. Moreover, EPF models must effectively cope with numerous abrupt large jumps in the course of the time series. This phenomenon is attributable to problems with transmission infrastructure and unforeseeable, non-proportional or inverse fluctuations in demand and supply [8]. Ziel et al. [28] also pointed out that the existence of a universal model for electricity price forecasting is highly improbable due to vast differences among countries, such as their individual political and climatic circumstances. Thus, not all the findings and methodologies successfully employed to one country are applicable to another country or region.

For these reasons, many different approaches to electricity price forecasting have been proposed to various extents of effectiveness and success. Papers published by Weron [24, 26] summarize the current methods of EPF, reviewing their strengths and weaknesses, effectiveness and potential, as well as providing an outlook on this topic over the next decade. For more than 15 years, various solutions and fitting models can be categorized into the following groups of methodology: fundamental/structural methods, reduced-form quantitative, stochastic models, statistical approaches, and computational intelligence; many of which being hybrid of two or more of these groups. These papers also emphasize the importance of appropriate inputs and predictors, along with the possibility of capturing different levels of seasonality in the models. Moreover, the author suggests extensions of the methodology going far beyond point forecasting: interval forecasting, density forecasting, threshold forecasting, and their combinations.

This paper proposes the use of the Generalized Additive Models (GAM) in attempt to improve the quality of the electricity spot price forecasting by applying a non-parametric estimation of multiple seasonal predictors. In the case of multivariate analysis, the key problem is to fit a d-dimensional model to the observed data, which leads to the exponential increase in the model's complexity as more variables or features are added to the dataset [18]. To combat this so-called "curse of dimensionality", a term coined by Bellman [3], the Additive Model method deals with each dimension separately, treating them as individual univariate smooth functions and adding up their approximations. This allows for an interpretable solution in which the marginal impact of a single variable could be explained independently of the other variables. Following this, the GAM method takes a major step forward where the response variable may be derived from any exponential family distribution, thus removing even further constraints and allowing greater flexibility, capturing nonlinear patterns that a classic linear model would otherwise miss [27]. Moreover, with the utilization of tensor product smooth interactions, the degree of smoothness in each direction can be controlled independently, resulting in an overall anisotropic penalty.

A comparable GAM was introduced by Pierrot and Goude [17] based on the hourly electricity load data in France from 2000 to 2005. Twenty-four separated time series regarding the daily observations are considered and fitted by the correspondent models. These models are set up to account for various levels of seasonality: daily, weekly, monthly, and a yearly global trend, so that a summer break (a large downturn in electricity demand during summer holidays) could be incorporated. Additionally, hourly meteorological data is included, e.g. the temperature, the cloud cover and the wind speed. A semi-parametric approach is adopted to these models, comprising a regressive part with explanatory variables and an autoregressive part with lagged loads. In the end, the residuals of the models are examined to detect remaining autocorrelation. The best model selection was conducted based on the comparison of the Generalized Cross Validation (GCV) scores. The forecasting results from this model, measured using the Root Mean Square Errors (RMSE), were significantly better than the unspecified benchmark model used by the authors.

In addition to point forecasting, Serinaldi [20] introduced the GAM for Location, Scale and Parameter (GAMLSS) for short-term price forecasting, based on the work of Stasinopoulos and Rigby [22, 23]. The aim of this paper is to reduce uncertainty of EPF by explicitly incorporating a wide range of distribution functions into the model, where the parameters of these distribution functions change dynamically in the course of a day, week, and year. According to this paper, the use of a position parameter, reflecting daily and weekly periodicity, a scale parameter, encompassing daily price standard deviation, and a shape parameter in form of a constant value is emphasized. The GAMLSS performance was put to test against many statistical benchmarks, from the naïve method [15], the classical linear Autoregressive model (AR) and Generalized Autoregressive Conditional Heteroskedastic model (GARCH) [13], to the Threshold Autoregressive (TAR) models [15]. In some instances, the performance of GAMLSS outstood the reference models and proved to be a reliable method for the comparison among different forecasting procedures.

Fan and Hyndman [7] took a semi-parametric additive approach with the aim of developing short-term forecasting models for regions in the National Electricity Market (NEM) of Australia from 1997 to 2009. In order to predict half-hourly demand loads, 48 sets of model parameters were estimated for each half-hour slice. For the point forecasting, the proposed additive regression model framework allowed non-linear and non-parametric terms to be accounted for the fit of the electricity load. Within the model setup, three mains effects were determined. Calendar effects include annual, weekly and daily seasonality, with public holidays also being recorded. Temperature effects from two sites are considered, whose average temperature and the differences between the daily maximums and minimums were incorporated into the model. Lagged demand effects were added to capture the autocorrelations within the demand time series, as well as its variance throughout the time. Prior to execution, a piece-wise backwards variable selection process was implemented to identify the best model, using the Mean Average Percentage Error (MAPE) as the selection criterion. In addition to the point forecasting, the forecasting outcome distribution was also estimated, providing a further indication of the forecast accuracy. Since the parametric method of delivering the forecasting distribution and prediction intervals would assume an i.i.d. error with zero mean and finite variance, the alternative of using bootstrapping as a non-parametric approach is encouraged, which is robust against violations of the normality assumption. Due to heavy computational tasks, a modified bootstrap method was conducted, constructing the empirical prediction intervals by centering the simulated forecast residuals around the original predicted point values.

The remainder of the paper is organized as follow. Section 2 provides a brief exploratory analysis to the data used in this study. Section 3 introduces GAM, as well as the model setup. Section 4 shortly introduces the structure and setup of the benchmark models. Season 5 evaluates the forecasting results. Finally, conclusion closes the study.

2 Sample and Methodology

This study focuses on the course of the hourly day-ahead spot price of the EEX Phelix-DE contract at the EPEX SPOT market of the European Power Exchange (EEX). This day-ahead spot contract is considered as a benchmark contract for European electricity. The exchange operates, among other trading activities, the power spot market for Germany, Austria, Luxembourg, France, the United Kingdom, the Netherlands, Belgium, and Switzerland. Purchase and sale orders are placed hourly for power which will be delivered the following day. The daily cycle ends at 12:00 pm, at which time the EPEX SPOT calculates the market clearing price. The visualization of the data used in this study can be seen in Figure 1.

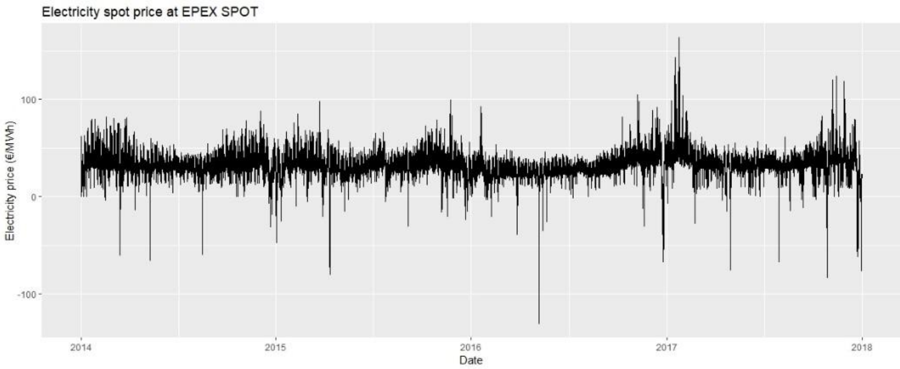

Fig. 1. Electricity spot price at EPEX SPOT

Figure 2 shows the average daily electricity price trend for four exemplary months in 2017, separately for weekdays and weekends. An overall M-shaped daily pattern throughout all months is evidently recognizable. The price is comparatively low for the first five hours of the day before rising to its first peak around 9 am, followed by a local minimum around 3 pm, peaking again around 8 pm before decreasing back to night level. Furthermore, the graphs also show a weekly pattern, as the weekend spot prices are constantly below those of the weekdays. Monthly seasonality also plays a part in determining the spot prices. Spring and summer time see a steeper mid-day gradient, while during fall and winter the price declines more constantly without retaining its peak.

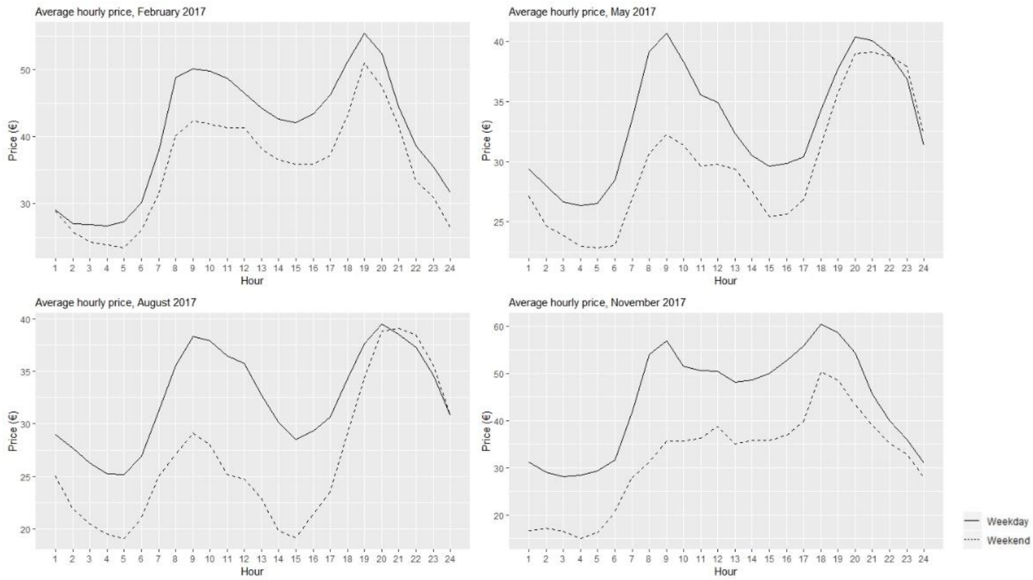

Fig. 2. Average hourly price for selected months in 2017, weekdays vs. weekend

3 The Generalized Additive Model

3.1 The GAM theory

Generalized Additive Model (GAM) [9, 27] is a non-parametric extension of the Generalized Linear Model (GLM), in which the relationship between the response and predictors are expressed by several smooth functions in order to capture the non-linearities underlying the data. The GAM can be formally expressed as:

$$g(E(y_i)) = \beta_0 + f_1(x_{i1}) + \cdots + f_p(x_{ip}) + \varepsilon_i \quad (1)$$

where i = 1,…,N, g is a link function (identical, logarithmic or inverse, etc.), y is a response variable, $x_1,\ldots,x_p$ are independent variables, β_0 is an intercept, $f_1(x_{i1}),\ldots,f_p(x_{ip})$ are unknown non-parametric smooth functions, and ε is an i.i.d. random error.

One way of determining these smooth functions is through the use of smoothing splines [9, 27]. These piecewise polynomial functions join many polynomials to generate a smooth curve through a set of points. The polynomials connect at certain points, called knots. At these knots, the joint polynomials share the same derivatives up to several degrees. The level of model smoothness depends on the degree of the polynomials, the number of knots, and their location. The locations of these knots are typically evenly-spaced. In this case, the smooth function is estimated by minimizing the penalized sum of squares:

$$min(RSS) \rightarrow \left(\sum_{i=1}^{n} [y_i - f(x_i)]^2 + \lambda \int_{x_{min}}^{x_{max}} [f''(x)]^2 dx \right)^* \qquad (2)$$

The first half of the function, $\sum_{i=1}^{n}[y_i - f(x_i)]^2$, is the standard residual sum of squares, representing how closely the fitted values are in alignment with the observed values, whereas the second half, $\int_{x_{min}}^{x_{max}}[f''(x)]^2 dx$, penalizes the "roughness", or the "wiggliness" of the data. Minimizing the integrated square of the second derivative would smooth out the data towards linearity. The key here is the smoothness parameter λ, which controls the trade-off between model fit and model smoothness. Wood [25] postulates, that the natural cubic splines are the smoothest interpolators, making the cubic smoothing splines (a natural cubic spline with knots at every data point) the best choice regarding the polynomial degree of the smooth term. However, this procedure has one major disadvantage: if the number of knots is approximately equal to the number of data records *n*, this will lead to model overfitting, and furthermore to a computational waste. Since λ, in most cases, obviously shrinks down the roughness at many knots, this will result in a spline that is much smoother than *n* degrees of freedom.

Another alternative to the presentation of the smooth functions is the penalized regression spline [27]. It can be expressed as a linear combination of a family of basis functions:

$$s(x_j) = B_0(x_j)\beta_0 + B_1(x_j)\beta_1 + \cdots + B_q(x_j)\beta_q = \boldsymbol{B}'\boldsymbol{\beta}, \forall j = 1 \dots p \qquad (3)$$

where $B_0(.), \dots, B_q(.)$ are the basis functions, $\beta_0, \dots, \beta_q$ are the associated coefficients with the basis dimension q, so that a linear relationship between the predictor and the smooth function is formed through the basis functions, with $\boldsymbol{B}$ being the model matrix of the basis functions, and $\boldsymbol{\beta}$ being the vector of regression coefficients. These coefficients applied to the basis functions act as amplifiers of the curvature of the spline. Like in the case of the above-mentioned smoothing spline, it is also possible to apply a penalty in the course of estimating the basis function coefficients of the regression spline to produce smoothness. Hence, in lieu of solving for the estimated $\hat{\beta}$ with a standard linear model, the penalized sum of squares can be minimized:

$$\min_{\beta}(RSS) \rightarrow (\|\boldsymbol{y} - \boldsymbol{B}'\boldsymbol{\beta}\| + \boldsymbol{\beta}'\boldsymbol{P}\boldsymbol{\beta})^* \qquad (4)$$

where $\boldsymbol{P}$ is the penalty matrix, imposing smoothness by directly penalizing the difference among the adjacent coefficients. This method is called the Penalized Iteratively Reweighted Least Squares method (P-IRLS), that for any given λ, the regression coefficients $\hat{\beta}$ can be obtained.

Hence, the problem has shifted from measuring the degree of smoothness for the model to determining the smoothing parameter λ. Since there is a trade-off between overfitting and oversmoothing the data, one option of determining the optimal degree of smoothness is by implementing backwards selection. This method is rather computationally expensive and can also result in relatively poor model accuracy due to uneven knot spacing. Instead, the smoothing parameter λ can be estimated using either the Generalized Cross Validation criteria (GCV) or the mixed model approach via Restricted Maximum Likelihood (REML).

With regard to the available choices of regression splines, GAM offers a wide range of smoothing bases, including cubic regression splines, cyclic regression splines, thin plate regression splines, P-splines, etc. These models differ in the choice of number of knots, the spacing of the knots, the level of rank and order, as well as the number of predictors in the model. Moreover, the interactions among the predictors play a critical role in the regression model. The inclusion of interactions extends from the most basic form of multiplication to the tensor product, allowing the possibility of implementing different smoothing bases for variables while applying penalization in different ways, resulting in an anisotropic penalty. In this paper, the use of tensor product smooth and the choice of cyclic penalized cubic thin plates regression spline are emphasized through the model setup below.

3.2 Model setup

Aggarwal et al. [1] classified the factors that have possible impact on the electricity prices in five different categories: market characteristics, nonstrategic uncertainties, other stochastic uncertainties, behavioral aspects, and temporal effects. As shown in the data analysis, there are three main seasonal patterns: the daily effect, weekly effect, and yearly effects which are represented by the dichotomous explanatory variables hour of the day, day of the week, and month of the year.

The goal of this model is to produce short-term forecasts for 12 randomly chosen weeks (one in each month) within the year 2017. For this setup, each model receives 260 weeks (approximately five years) of training data prior to the forecasted week. We begin setting up the model structure by determining the smooth function components for the daily, weekly and yearly pattern separately. Thus, in model M1 the individual effects form three different univariate smooths additively:

$$\mathbf{M1}: P_t = f_1(Daily) + f_2(Weekly) + f_3(Yearly) + C + \varepsilon_i \qquad (5)$$

Cubic regression splines were applied for all individual components. The number of knots is equal to the number of unique values in each predictor, in this case 24, 7, and 12, respectively. This initial model treats the three predictors individually, assuming that all effects are independent. This assumption is not realistic, since in the exploratory data analysis it could be observed that the effects are mutually dependent.

To account for the interaction among the predictors, thin plate regression splines are recommended by the extant literature [27]. Here, a truncated version of the thin plate splines is applied, using the thin plate spline penalty to acquire a low-rank smoother that has far fewer coefficients than there is data to smooth. Moreover, it can deal with any number of predictors and tends to give the best MSE performance [27]. Accordingly, the same isotropic smoothing base is used for all three predictors in one smooth function:

$$\mathbf{M2}: P_t = f(Daily, Weekly, Yearly) + C + \varepsilon_i \qquad (6)$$

In this case, only one single value of the smoothing parameter λ is applied in all directions. The problem with this isotropic penalty is, that its result is only reliable when the predictors are approximately on the same scale. In other words, the discrepancy among the different units of the different explanatory variables could result in a false integration of the second derivative due to their disproportional contribution to the overall integration. Hence, the use of tensor product smooths is proposed [27].

Tensor product smoothing is a type of multivariate smoothing base that derives the multivariate bases from individual univariate marginal bases. In other words, the non-separable smooth function $f(Daily, Weekly, Yearly)$ can instead be approximated by the tensor product of its component, $f(Daily)$, $f(Weekly)$ and $f(Yearly)$. Each of the basic functions is smoothed in its corresponding dimensions individually, so that the correspondent coefficient matrix is obtained. Then the tensor product ($\otimes$) of the three matrices is computed, as shown in model M3:

$$\mathbf{M3}: P_t = f_1(Daily) \otimes f_2(Weekly) \otimes f_3(Monthly) + C + \varepsilon_i \quad (7)$$

As a result, each component represents a unique combination of the three marginal basis functions. This allows for an overall anisotropic smoothing penalty, with the possibility of using different smoothing bases for every predictor and penalize it in many different ways. Each smoothing parameter λ_{Daily}, λ_{Weekly} and λ_{Yearly} is individually determined through the same method as the single smoothing parameter for the univariate smoothing, which results in an overall tensor product smooth that is indifferent to the rescaling of its independent variables.

Although this method proves to yield significantly better results, it also becomes significantly more computationally expensive as the dimensionality of the tensor product increases by the introduction of more predictors. Within the framework of this paper, this issue is addressed by using the pairwise bivariate tensor product smooths for the three predictors, resulting in model M4:

$$\begin{gathered}\mathbf{M4}: P_t = f_1(Daily) \otimes f_2(Weekly) + f_1(Daily) \otimes f_3(Yearly) + \\ f_1(Weekly) \otimes f_3(Yearly) + C + \varepsilon_i\end{gathered} \quad (8)$$

Finally, the combination of the three individual effects and their three mutual interactions enables the decomposition of the model, analyzing to what extent each individual predictor influences the response individually, as well as each of the pairwise interactions. Accordingly, the ultimate model M5 can be annotated as follows:

$$\begin{gathered}\mathbf{M5}: P_t = f_1(Daily) + f_2(Weekly) + f_3(Yearly) + \\ f_1(Daily) \otimes f_2(Weekly) + f_1(Daily) \otimes f_3(Yearly) + \\ f_1(Weekly) \otimes f_3(Yearly) + C + \varepsilon_i\end{gathered} \quad (9)$$

In the extant literature a variety of model accuracy measures are discussed. The trade-off between model accuracy and model complexity is often in the focus of the consideration. Accuracy measures that penalize for model complexity are proposed by Akaike (Akaike Information Criterion, AIC) [2] and Schwarz (Bayes Information Criterion, BIC) [19]. However, the AIC and the BIC are typical in-sample accuracy measures. Since this study deals with forecasting accuracy and not with model fitting, an out-of-sample / forecasting accuracy measure needs to be applied. A popular choice among the forecasting accuracy measures, the Mean Absolute Percentage Error (MAPE) fails in the context of price forecasting, since the spot prices for electricity are oftentimes negative, which leads to a possible erroneous interpretation. Moreover, when the prices are high, MAPE is rather indifferent to a considerable absolute change, whereas it would scale up drastically to the same price difference, when the prices are close to zero. In line with extant literature, the weekly Root Mean Square Errors (RMSE) is used for the evaluation of the forecasting accuracy here [26].

4 Statistical benchmark models

4.1 Autoregressive Integrated Moving Average model with external regressors (ARIMAX) with seasonality

The benchmark ARIMAX model in this paper, as derived by Meier et al. [14], is an extension of the classical ARIMA model [4]. The X-term of the model comprises the external regressors, accounting for various level of seasonality in form of dummy-coded variables, including hour of the day, day of the week, and month of the year. The Hyndman-Khandakar algorithm [12, 13] is utilized to achieve the optimal ARIMAX parameterization. This step includes the determination of the number differentiations (d) needed to achieve stationary using the KPSS tests as well as the simultaneous determination of the number of lags for the autoregressive (p) and the moving average (q) term, applying Akaike Information Criterion (AIC). Since the data sample is identical with Meier et al. [14] in both analyses, the original ARIMAX (3,1,3) model with 40 dummy variables is adopted as the benchmark model for this paper.

4.2 Naïve forecasts

The similar-day method estimates the electricity price of a certain day on the basis of the electricity price of the same weekday of the previous week [21, 24]. Further adaptations of this method match characteristics like the hour of the day, the day of the week, the month of the year by applying linear combinations or regression procedures. One of the variations of the similar-day method, is the naïve method. Here the forecast is based on the previous day, with the exception of the Saturdays, Sundays, and Mondays. These are forecasted by looking back to values of the previous week [16, 24]. Despite its simplicity, this "naïve test" proves its effectiveness in identifying inept forecasting models, thus turning it into one of the most popular benchmark models in EPF [5, 6, 16].

5 Assessment of the model performance

Figure 3 represents three seasonal smoothed effects of the electricity price time series which originate from model M5: daily, weekly and yearly. As seen in the data analysis, these plots confirm that there is a difference in price throughout the course of a day, throughout the course of a week, and throughout the course of a year. The first daily peak around 10:00 am could be due to the morning working routines, and the second one around 8:00 pm accounts for the heating and lighting needs in winter, as well as extra activities in summer, where there is a longer period of daylight. The electricity price is fairly stable at a higher level from Tuesday to Friday, and sinks at the weekend to rise again at the Monday, confirming the higher need for electricity on working days. Regarding the yearly pattern, the prices in fall and winter are higher than in the other two seasons, emphasizing the heating and lighting demands.

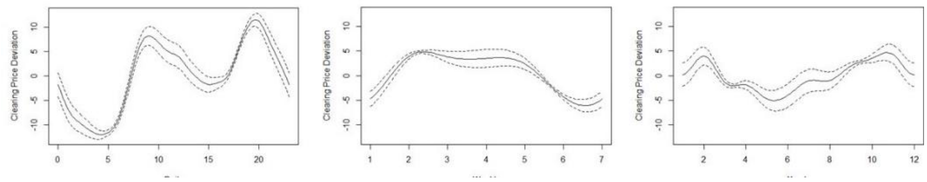

Fig. 3. The electricity price as a mean deviation with forecasting intervals

Figure 4 shows the tensor product smooths of the effects in pairs, so that the interaction among the effects are easier to spot. It can be observed from the daily and weekly smoothing, that the daily peaks around 10 am and 8 pm are still prominent throughout the week, although at a remarkably lower level at the weekends. The middle graphs show the relationship between the weekly and the yearly effect: the daily peaks are now smoothed along different months, with the prices in summer lower than in winter, showing peaks at the morning and evening time in December and January. Lastly, the tensor product between the weekly and yearly effect showcases a minimum price on Sundays in May, as opposed to the maximum on Mondays in January.

These figures demonstrate one of the most decisive advantages of GAM in comparison to other methods: interpretability with visualization. GAM takes on the nature of an additive regression model, in which the interpretation of the marginal impact of a singular variable, the partial derivative, is not contingent on the values of the other variables in the model. Looking at Figure 3, one could intuitively draw conclusion on the effects the temporal predictors have on the electricity prices, each of which is accounted for separately by an individual smoothed function; so that the daily peaks, the weekend cutback, and the decrease of prices in summer months are appointed to the right temporal effects accordingly. Moreover, GAM is able to isolate the individual effects from the predictors alone from the intercorrelated influences among them upon the response variable; for instance, in our final model, the influence of the hourly variable alone, the interaction between the hourly and the weekly variable, as well as the one between the hourly and the monthly variable, are all accounted for separately. Figure 4 shows the interactions being plotted, so that the original patterns could be revealed, even though the dataset at hand may suggest a noisier relationship. Hence, by simply taking a glance at the output and its visualization of the model, one can make intuitive statements about the effects of the predictors which is comprehensible to a nontechnical person.

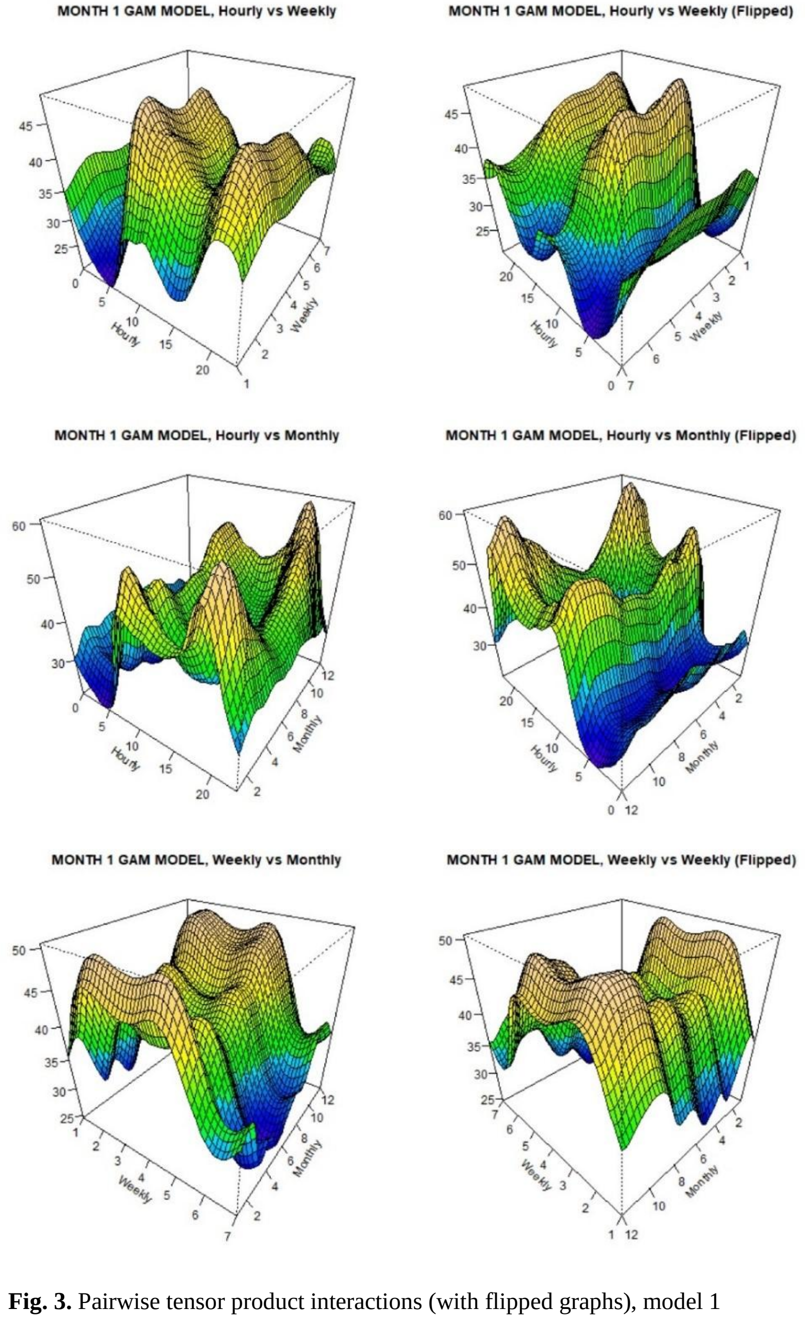

Fig. 3. Pairwise tensor product interactions (with flipped graphs), model 1

Figure 5 illustrates the forecasting accuracies of the GAM, ARIMAX and naïve models applied using the months of March and April of the 2017 forecast period as examples.

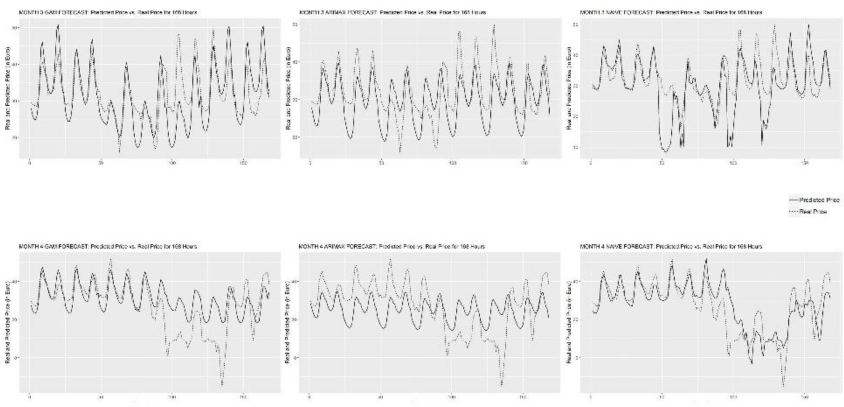

Fig. 4. Forecast vs. real time series

Table 1 documents the forecasting accuracy of the GAM against the other two benchmark models. The GAM proves to be more accurate in the overall testing and less prone to price peaks and troughs. In roughly 75% of the cases GAM shows better forecasting accuracies than the benchmark models.

Table 1. Forecasting performance of GAM in comparison with the benchmark models

Training Phase		Test Phase		Comparison (RMSE of Real vs Predicted Price)		
Begin	**End**	**Begin**	**End**	**GAM**	**ARIMAX**	**Naïve**
10.01.2012	02.01.2017	03.01.2017	09.01.2017	16.78	17.331	16.816
10.02.2012	02.02.2017	03.02.2017	09.02.2017	19.185	12.713	12.668
28.03.2012	21.03.2017	22.03.2017	28.03.2017	5.796	5.95	7.386
23.04.2012	16.04.2017	17.04.2017	23.04.2017	10.487	12.494	8.565
08.05.2012	01.05.2017	02.05.2017	08.05.2017	6.639	18.588	27.218
02.07.2012	25.06.2017	26.06.2017	02.07.2017	5.638	6.765	6.87
13.07.2012	06.07.2017	07.07.2017	13.07.2017	6.2	4.93	6.218
27.08.2012	20.08.2017	21.08.2017	27.08.2017	6.427	11.116	10.042
23.09.2012	16.09.2017	17.09.2017	23.09.2017	5.817	8.65	10.121
08.10.2012	01.10.2017	02.10.2017	08.10.2017	18.791	13.901	20.018
06.12.2012	29.11.2017	30.11.2017	06.12.2017	12.073	13.384	8.776
15.12.2012	08.12.2017	09.12.2017	15.12.2017	12.454	12.884	13.799

For checking the robustness of the presented GAM models, the outliers were identified and substituted by applying the seasonal and trend decomposition method Loess (Locally Weighted Least Squares Regression). Loess smoothing calculates an average of the data around the vicinity, giving more weight to data near the vicinity and less weight to data further away from the vicinity. Given the identical model set up, the

GAM model fitting process shows little difference in results when fed with the original or the modified input data. Accordingly, the GAM model is robust towards outliers. Nevertheless, the identified outliers are not measurement inaccuracies but real clearing prices and reflect the stark fluctuation of the electricity price time series. Thus, they should be included in the model.

Furthermore, it was examined whether the length of the training data time series has an influence on the forecasting accuracy. Hypothetically, the quality of the model would monotonically rise as the number of training data records increases. We were not able to find an optimum length of the training time series that could be applied for all months. This indicates that a large number of structural breaks make a perfect adaptation of the model impossible. These structural breaks are mainly due to the strong promotion of renewable energies in Germany, which over time are accompanied by a strong increase in volatility and are predominantly politically driven.

6 Conclusion

In this study, the use of Generalized Additive Model (GAM) [9, 27] is proposed as an alternative stochastic method to conduct one-week ahead forecasting of electricity market prices. Overall, GAM is an extension of the Generalized Linear Model, demonstrating its superiority in terms of flexibility, in which the relationships between the predictor and the response variables are assumed to be non-linear. A model using isolated additive smoothing components according to our model M1 could therefore not exploit the advantages of GAM, since the interactions between the dimensions are not taken into account. The complete consideration of all interactions of the predictors in only one smoothing function lead to the best prediction accuracies, but is so computationally intensive that, so that its practical applicability is rather limited. A significant improvement was the use of tensor smoothing function in our model M3, where the single smoothing functions were connected via the tensor product. In order to be able to work out the interactions between the dimensions even better without running the risk of achieving high computing capacities again, we developed the models M4 and M5. These models combine the tensor products between the smoothing functions pairwise, so that an excessive computing load is avoided and the interaction effects can still be reproduced with sufficient accuracy.

References

1. Aggarwal, S. K., Saini, L. M., Kumar, A. Electricity price forecasting in deregulated markets: a review and evaluation. International Journal of Electrical Power & Energy Systems, 31(1), 13-22 (2019).
2. Akaike, H. Information theory and an extension of the maximum likelihood principle. Second International Symposium on Information Theory, 267–281 (1973).
3. Bellman R.E. Adaptive Control Processes. Princeton University Press, Princeton, NJ (1961).

4. Box, G. E. P., Jenkins, G. M. Time series analysis: forecasting and control. Holden-Day, San Francisco (1971).
5. Conejo, A., J., Contreras, J., Espínola, R., Plazas, M. A. Forecasting electricity prices for a day-ahead pool-based electric energy market. International Journal of Forecasting, 21(3), 435-462 (2005).
6. Contreras, J., Espínola, R., Nogales, F. J., Conejo, A. J. ARIMA models to predict next-day electricity prices. IEEE Transactions on Power Systems, 18(3), 2003, 1014-1020.
7. Fan, S., Hyndman, R. J. Short-Term Load Forecasting Based on a Semi-Parametric Additive Model. IEEE Transactions on Power Systems, 27(1), 134-141 (2012).
8. Geman, H., Roncoroni, A. Understanding the fine structure of electricity prices. Journal of Business, 79(3), 1225-1261 (2006).
9. Hastie, T., Tibshirani, R. Generalized Additive Models. Chapman & Hall, London (1990).
10. Hong, T. Energy Forecasting: Past, Present, and Future. Foresight: The International Journal of Applied Forecasting, 32, 43-48 (2014).
11. Hope, E., Rud, L., Singh, B. Electricity futures market. In Lesourd, J.-B., Percebois, J., Valette, F. (eds.), Models for Energy Policy. Routledge, London, New York, 238-249 (1996).
12. Hyndman, R. J.: R Package "forecast": Forecasting Functions for Time Series and Linear Models, 2018.
13. Hyndman, R. J., Khandakar, Y. Automatic Time Series Forecasting: The forecast package for R. Journal of Statistical Software, 27(3), 2008.
14. Meier, J.-H., Schneider, S., Schönfeldt, T., Schüller, P., Wanke, B. Electricity Price Forecasting: A methodological ANN-based Approach with special Consideration of Time Series Properties. ICTERI Conference Kiev. Proceedings of the 14th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume II: Workshops. Part I: 6th International Workshop on Information Technologies in Economic Research (ITER), Paper 5, 2018.
15. Misiorek, A., Trueck, S., Weron, R. Point and interval forecasting of spot electricity prices: linear vs. non-linear time series models. Studies in Nonlinear Dynamics & Econometrics, 10(2), 2006.
16. Nogales, F. J., Contreras, X., Conejo, A. J., Espínola, R. Forecasting next-day electricity prices by time series models. IEEE Transactions on Power Systems 17, 342-348 (2002).
17. Pierre, G., Goude, Y., Nedellec, R. Semi-Parametric Models and Robust Aggregation for GEFCom2014 Probabilistic Electric Load and Electricity Price Forecasting. International Journal of Forecasting, 32(3), 1038-1050 (2015).
18. Schimek, M. G., Turlach, B. A. Additive and generalized additive models: A survey. SFB 373 Discussion Papers, 1998.
19. Schwarz, G. Estimating the dimension of a model. Annals of Statistics 6, 46-464 (1978).
20. Serinaldi, F. Distributional Modeling and Short-Term Forecasting of Electricity Prices by Generalized Additive Models for Location, Scale and Shape. Energy Economics, 33(6), 1216-1226 (2011).
21. Shahidehpour, M., Yamin, H., Li, Z. Market operations in electric power systems: forecasting, scheduling, and risk management. Wiley, New Jersey (2002).
22. Stasinopoulos, M. D., Rigby, R. A. Generalized additive models for location scale and shape. Journal of the Royal Statistical Society, 54(3), 507-554 (2005).
23. Stasinopoulos, M. D., Rigby, R. A. Generalized additive models for location scale and shape (GAMLSS) in R. Journal of Statistical Software, 23(7), 744-763 (2007).

24. Weron, R., Misiorek, A. Short-term electricity price forecasting with time series models: A review and evaluation. In Mielczarski, W. (Ed.), Complex electricity markets, 231-254 (2006).
25. Weron, R., Misiorek, A. Forecasting spot electricity prices: A comparison of parametric and semiparametric time series models. International Journal of Forecasting, 24, 744-763 (2008).
26. Weron, R. 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 (2014).
27. Wood, S. N. Generalized Additive Models: An Introduction with R. 2nd ed. Chapman & Hall, London (2017).
28. Ziel, F., Steinert, R., Husmann, S. Efficient modeling and forecasting of electricity spot prices. Energy Economics, 47, 99 (2015).