Received 29 September 2021; revised 22 January 2022; accepted 9 March 2022.
Date of publication 23 March 2022; date of current version 10 June 2022.

Digital Object Identifier 10.1109/OAJPE.2022.3160933

Smoothed Bernstein Online Aggregation for Short-Term Load Forecasting in IEEE DataPort Competition on Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm

FLORIAN ZIEL

House of Energy Markets and Finance, University of Duisburg-Essen, 45141 Essen, Germany
e-mail: florian.ziel@uni-due.de

This work was supported by the Open Access Publication Fund of the University of Duisburg-Essen.

ABSTRACT We present a winning method of the IEEE DataPort Competition on Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm. The day-ahead load forecasting approach is based on a novel online forecast combination of multiple point prediction models. It contains four steps: i) data cleaning and preprocessing, ii) a new holiday adjustment procedure, iii) training of individual forecasting models, iv) forecast combination by smoothed Bernstein Online Aggregation (BOA). The approach is flexible and can quickly adjust to new energy system situations as they occurred during and after COVID-19 shutdowns. The ensemble of individual prediction models ranges from simple time series models to sophisticated models like generalized additive models (GAMs) and high-dimensional linear models estimated by lasso. They incorporate autoregressive, calendar, and weather effects efficiently. All steps contain novel concepts that contribute to the excellent forecasting performance of the proposed method. It is especially true for the holiday adjustment procedure and the fully adaptive smoothed BOA approach.

INDEX TERMS Load forecasting, forecasting competition, prediction, demand analytics, forecast combination, online learning, aggregation method, ensemble, holiday effects, COVID-19, shutdown, lockdown, post-COVID effects, generalized additive models, GAM, lasso, weather effects.

I. INTRODUCTION

THE COVID-19 pandemic led to lockdowns and shutdowns all over the world in 2020 and 2021 to reduce the spread of the coronavirus SARS-CoV-2 and the resulting COVID-19 disease. Mentioned lockdowns and shutdowns substantially impacted the behavior of the people. Thus, also the consumption of electricity changed dramatic during those periods, [1]–[3]. Electricity load forecasting during lockdowns and shutdown periods is a challenging task, but even months afterward the forecasting task is still complicated. One reason is that is not obvious which of the changed behavioral pattern during the lockdown observed in many countries (e.g. increased remote work, getting up later) will persist months and years after the lockdown. Another problematic aspect is the disruption of annual seasonalities during the lockdown periods.

The IEEE DataPort Competition *Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm* focuses on Post-COVID aspects in electricity load forecasting [4]. The day-ahead load forecasting competition was based on real data and ran over a test period of 30 days. This manuscript describes one of the winning methods that scored 3rd in the competition.[1] The prediction approach is based on smoothed Bernstein Online Aggregation (BOA) that is applied on individual load forecasting models. The complete model flow with the organization within the paper is depicted in Fig. 1.

First, we introduce the data set and the forecasting task in more detail and discuss the initial data preprocessing steps. Afterward, we explain a novel holiday adjustment

[1] According to significance test conducted by the organizers, the top 3 positions where not significantly different from each other.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

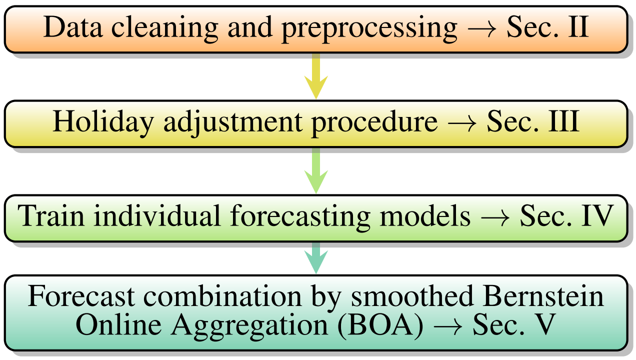

FIGURE 1. **Structure of forecasting approach used for the forecasting competition.**

procedure to deal adequately with holidays in the data based on ideas from [5]. Section IV considers several state-of-the-art individual forecasting models, sometimes referred to as experts or base learners, [6]. Then, we describe a novel online aggregation procedure is used to combine the expert forecasts. The combination method is based on an extension of the fully adaptive Bernstein Online Aggregation (BOA) procedure which satisfies certain optimality guarantees and is utilized in recent load forecasting tasks, see [7], [8]. The BOA is extended by a smoothing component and is implemented in the `R` package `profoc` [9]. It is similarly used in [10] for probabilistic forecasting using CRPS learning.

Next to the considered aggregation procedure, a crucial aspect is the choice of individual forecasting models. The considered models range from simple time series models like autoregressive and exponential smoothing models [11] to more advanced statistical learning procedures. Among the latter one we consider high-dimensional statistical models estimated by lasso and generalized additive models (GAMs) that were also among the winning methods in past load forecasting competitions, [12], [13]. Also, we tested several modern non-linear models like gradient boosting machines (GBM) (using the `R` packages `gbm` and `lightgbm`) and neural networks (using the `R` packages `nnet` and `keras`). But the forecasting accuracy was rather low and did not improve the forecasting performance in the forecasting combination method described in Section V. The reason might be that the major impacts are linear, esp. autoregressive and seasonal effects. Also, another winning group of the competition observed low predictive performance of methods based on decision trees and neural networks, [14]. Still, they contain them in a final aggregation procedure.

The evaluation metric is the mean absolute error (MAE) which is a popular point forecasting metric. More precisely, median forecasts are required to minimize the MAE, see [15].

II. DATA AND PREPROCESSING

The load forecasting competition contains initially hourly load data from 2017-03-18 00:00 to 2021-01-17 07:00, visualized in Fig 2. According the organizers the load data corresponds to one city, but the origin of the load data to predict was disclosed.

The daily forecasting task is to predict the next days hourly load, which corresponds to forecast 24 values 17 to 40 hours ahead. Thus, the first forecasting task was aiming for the hourly load for 2021-01-18 from 00:00 to 23:00. The second task was to predict the load on 2021-01-19. This rolling forecasting procedure was continued over 30 days in the competition. In the bottom chart of Fig (2) you see clearly the structural break due to the COVID-19 lockdown in March 2020. The overall load level dropped and the weekly profile got disturbed dramatically. In the proceeding months we observe some slowly increasing recovery of the electricity consumption. However, even in 2021 we observe that especially the peak hours have a lower load level than the previous years.

Next to the actual load data, also weather input data was provided. This was actual data on humidity, pressure, cloud cover, temperature, wind speed such as day-ahead forecasts of all meteorologic features except humidity were provided, Fig 3 for last years data. The day-ahead weather forecasts were in fact 48-hours ahead forecast. Thus, for the first day, weather forecasts data up to 2021-01-19 07:00 was provided. During the competition the actual load and weather data, and the weather forecast data for the next 24 hours were released, leading to a typical rolling forecasting study design.

The weather data contained some obvious reporting problems which were cleaned using linear interpolation and the `R`-package `tsrobprep`, see [16], [17]. Afterwards, we transformed the wind direction data to the north-south (NS) and east-west (EW) component by evaluating the cosine and sine of the wind direction data. Thus, Fig 3 shows the cleaned data for the available weather forecasts and actuals. For further analysis, we extend the weather data input space by adding rolling daily means of all 7 weather inputs (humidity, pressure, cloud cover, temperature, wind speed, wind direction NS, wind direction EW). Thus, we consider in total $2 \times 7 = 14$ weather inputs.

III. HOLIDAY ADJUSTMENT PROCEDURE

As the origin of the data was disclosed and no holiday calendar was provided a specific solution for dealing with holidays is required. Handling holidays adequately is an important task and may improve the forecasting accuracy substantially even for the non-holidays, see e.g. [5]. Fig. 4 provides a small overview on the considered two model steps.

By eyeballing, it is easy to spot some obvious date-based public holidays in the data (12Jan, 17Apr, 1Aug, 18Sep, 11Dec, 18Dec). But there are also a couple days which behave like holidays but the pattern of occurrence seems to be different. We consider a holiday adjustment procedure to take into account the holiday impact appropriately. The procedure is based on a high-dimensional time series model, similarly used in the GEFCom2014 (Global Energy Forecasting Competition 2014), see [12]. The result of the considered procedure is illustrated for the period from October to

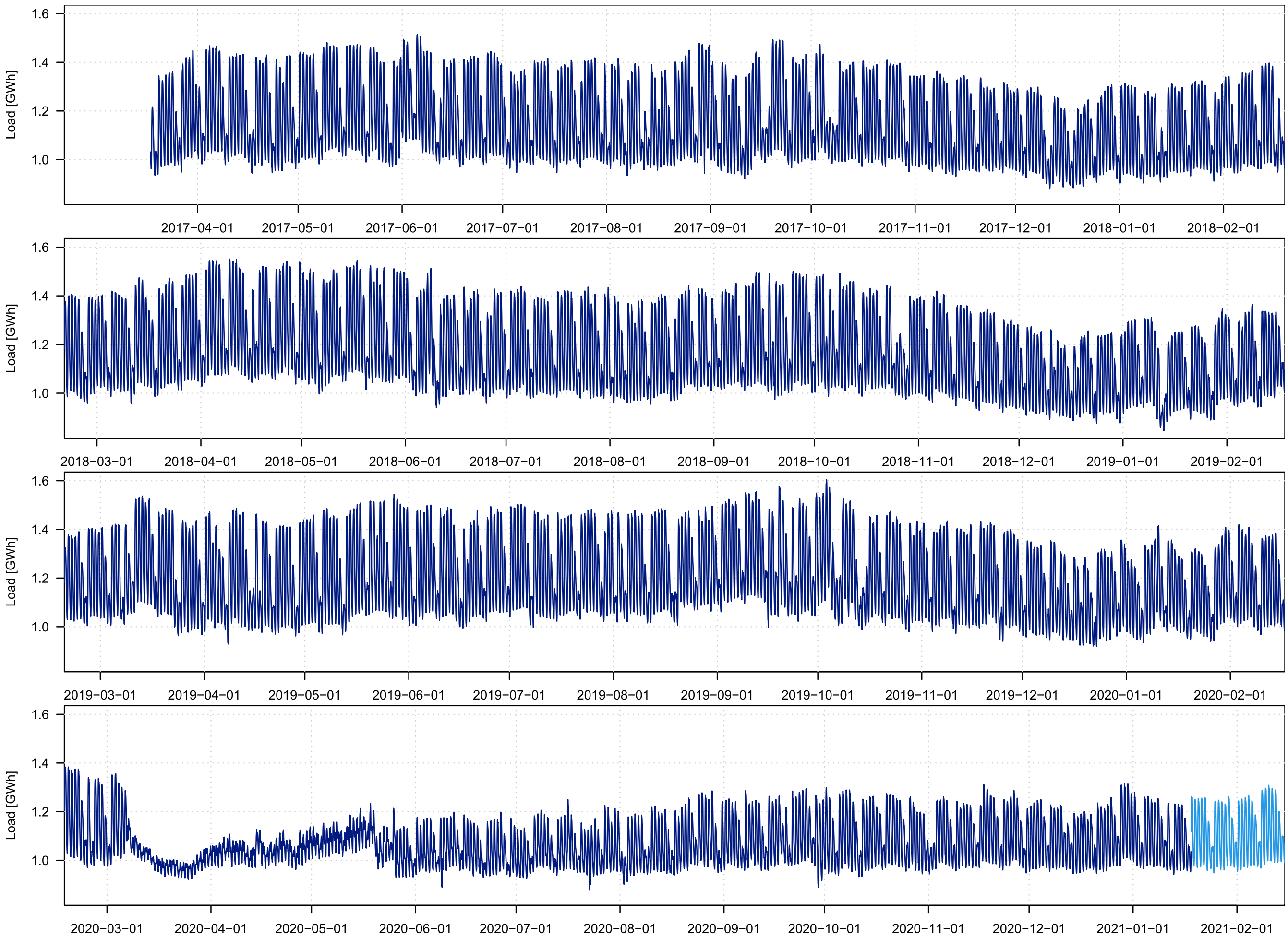

FIGURE 2. Available load data for day-ahead for the competition. The test data is highlighted in light blue.

December in Fig 5. There, we see clearly the holidays on 11 and 18 December, while the holidays pattern in October and November follow complex pattern.

To introduce the holiday adjustment procedure formally, we require some notations. Denote $\ell_t = \log(L_t)$ the logarithm of the load L_t at time point t. Let T be the number of observations that is currently available for model training. The considered model is a high-dimensional linear model for ℓ_t:

$$\ell_t = \sum_i \boldsymbol{\beta}_i x_{t,i} + \varepsilon_t = \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \tag{1}$$

Given data for $t = 1, \ldots, T$ the input matrix $(\boldsymbol{x}_1', \ldots, \boldsymbol{x}_T)'$ contains the following components:

- **i)** lagged log-load ℓ_{t+k} values, for $k \in \mathcal{I}_{\text{pos}} \cup \mathcal{I}_{\text{neg}}$ where $\mathcal{I}_{\text{pos}} = \{168, 169, \ldots, 510\}$ and $\mathcal{I}_{\text{neg}} = \{-168, -169, \ldots, -510\}$
- **ii)** p-quantile ReLU-transformed (Rectified Linear Unit transformed) of all available weather data on quantile grid of probabilities $\mathcal{P} = \{0, 0.1, \ldots, .9\}$. In detail we compute $x_t^{p\text{-ReLU}} = \max\{x_t - q_p(\boldsymbol{x}), 0\}$ with $q_p(\boldsymbol{x})$ for $p \in \mathcal{P}$ as p-quantile of $\boldsymbol{x}$ for weather input feature $\boldsymbol{x} = (x_1, \ldots, x_T)'$.
- **iii)** All weather data interactions, i.e. $x_{x_t,y_t,t}^{\text{inter}} = x_t y_t$ for inputs x_t and y_t. x_t and y_t are chosen from the 14 weather inputs which yields in total $14 \times 15/2 = 105$ weather interactions.
- **iv)** Daily and weekly deterministic effects. The daily and weekly effects are modeled by standard and cumulative dummies:

$$x_{k,t}^{\text{day}} = \mathbb{1}\{\text{HoD}_k(t) = k\} \text{ for } k \in \{1, \ldots, 24\} \tag{2}$$
$$x_{k,t}^{\text{cday}} = \mathbb{1}\{\text{HoD}_k(t) \leq k\} \text{ for } k \in \{1, \ldots, 24\} \tag{3}$$
$$x_{k,t}^{\text{week}} = \mathbb{1}\{\text{HoW}_k(t) = k\} \text{ for } k \in \{1, \ldots, 168\} \tag{4}$$
$$x_{k,t}^{\text{cweek}} = \mathbb{1}\{\text{HoW}_k(t) \leq k\} \text{ for } k \in \{1, \ldots, 168\} \tag{5}$$

  where $\text{HoD}_k(t)$ and $\text{HoW}_k(t)$ are the hour-of-the-day and hour-of-the-week dummies taking values in 0 and 1. They are 1 if and only if k is the k's hour of the day/week.
- **v)** Annual deterministic effects described by periodic cubic B-splines with annual periodicities ($A = 24 \times 365.24$ hours). Precisely, we consider 12 basis functions on a equidistant grid on $[0, A)$.

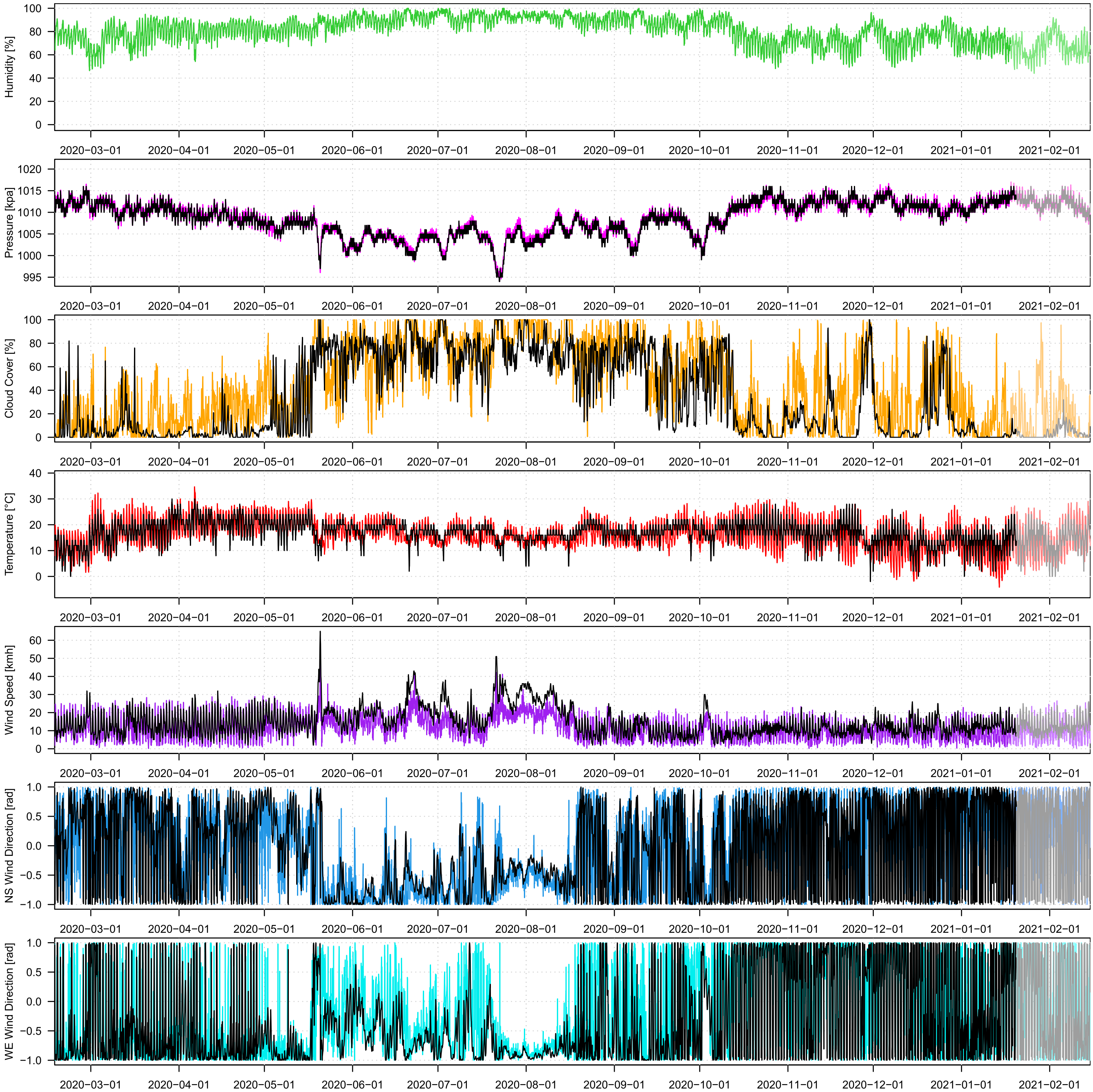

FIGURE 3. External input weather data available for the competition for the last year. Colored data correspond to actuals and black ones to day-ahead forecasts. The test data is indicated by lighter colors.

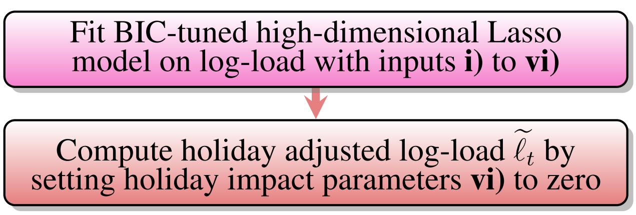

FIGURE 4. Flowchart of the holiday adjustment procedure.

vi) Impact-adjusted holiday dummies on days which were identified in advance as potential holidays.

The lagged log load in **i)** describes the autoregressive impact on a specific day for the surrounding 3 weeks of information without using nearby information of the surrounding week, to exclude any impact from bridging effects. As predicting the holiday effects is not a forecasting tasks it is totally legible to use positive lags to improve predictive accuracy.

Note that the ReLU-transformed weather input in **ii)** is relevant to capture non-linear weather impacts. However, for $p = 0$ the linear effect is modelled. Component **iii)** is motivated from the second order Taylor approximation. Considering all weather data interactions allows us to capture relevant

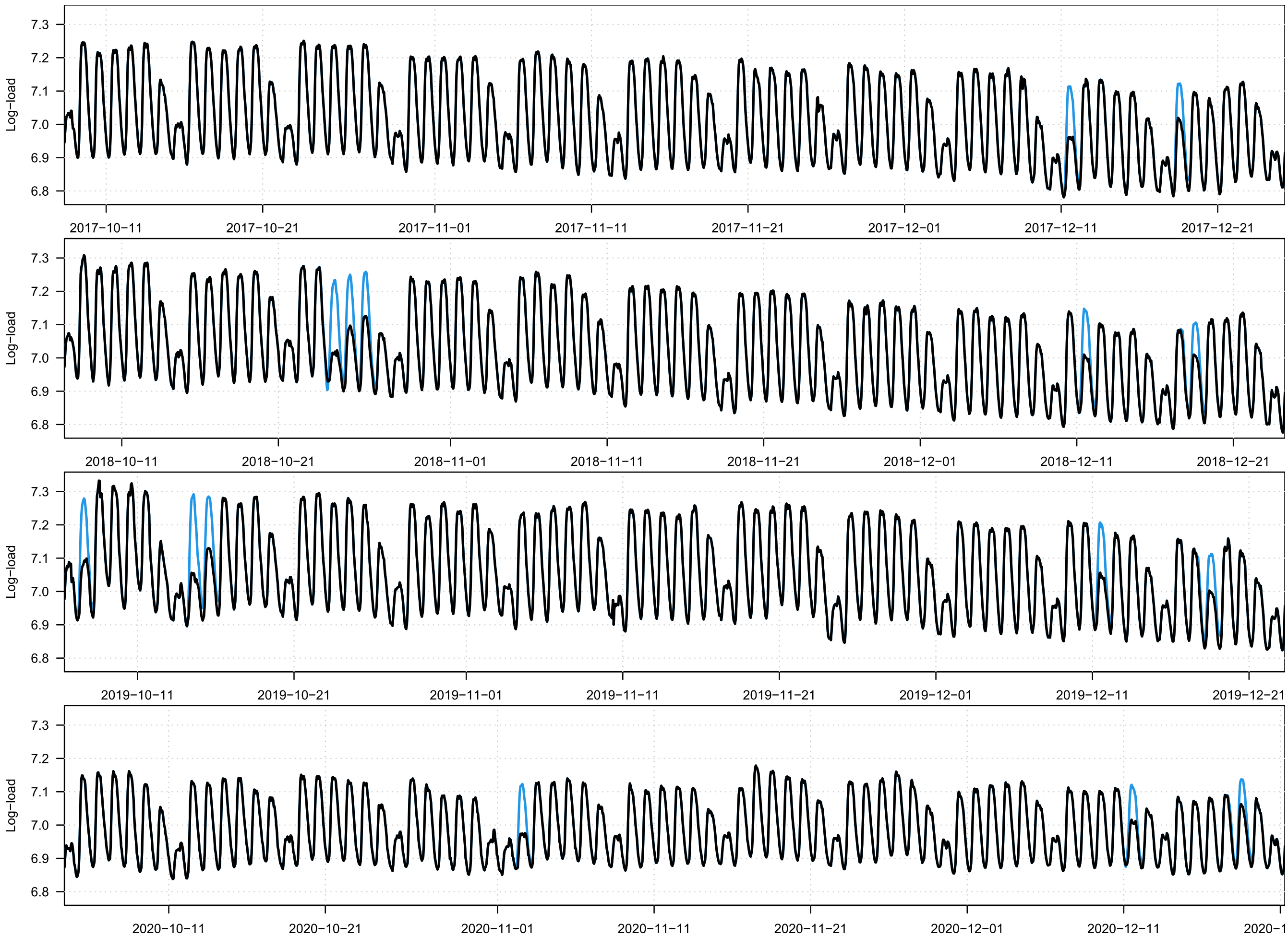

FIGURE 5. Log-load data in October to December in considered years 2017-2020 (black) with holiday adjustment of the proposed procedure (blue).

non-linear information. In fact, components **ii)** and **iii)** may be regarded as a manual application of the kernel trick to the input data to enlarge the feature space.

Further, in **iv)** the standard dummies with '='-sign in the definition (see (2) and (4)), have the job of detecting demand effects that happen only at the day or week period (e.g. if the load is high only at a certain hour of the day). In contrast, the cumulative dummies (see (3) and (5)) have the purpose to describe effects that persists over multiple hours in the day or week period. Similarly, feature **v)** captures the annual seasonal behavior. In comparison, annual effects tend to vary slowly over time. Thus considering a few parameters is usually sufficient to capture all relevant effects. Note that B-splines are local basis functions, in contrast to Fourier approximations by sine and cosine functions which are global ones. An important consequence for local basis functions is that anomaly will only impact the seasonal neighborhood. For instance, an outlier in winter will not affect the fit in summer. For more details on periodic cubic B-splines in energy forecasting see [18].

The component **vi)** models the holiday effect and is crucial for the holiday adjustment procedure. Its design corresponds to the holiday modeling approach used in see [12]. However, next to the impact multiplication also a scaling of the impact. Precisely it is scaled by the difference of rolling quantiles at probabilities 90% and 37% of the previous week. The idea is that the upper quantile is an estimate standard activity in a working week and the lower quantile and estimate for the Sunday peak. This adjustment procedure is required to deal with the strong structural breaks during the COVID-19 shut-down. This, effect can be seen in Fig 5 as well. We observe that the absolute holiday impact of 11th December is smaller in 2020 than the years before.

The model for the log-load ℓ_t with all inputs **i)** to **vi)** is estimated using lasso (least absolute shrinkage and selection operator) on scaled input data. The tuning parameter is chosen by minimizing the Bayesian information criterion (BIC), see e.g. [19]. Now, we take the fitted parameter vector $\widehat{\boldsymbol{\beta}}$ and set all estimated parameters which correspond to the holiday impacts vi) to zero, to receive $\widehat{\boldsymbol{\beta}}^{\text{hldadj}}$. The fitted values with respect to $\widehat{\boldsymbol{\beta}}^{\text{hldadj}}$ is the holiday-adjusted log-load time series $\widetilde{\ell}_t$, as illustrated in Fig 5 in blue.

Note that for the inital and final three weeks (exactly 510 hours as the maximum in $\mathcal{I}_{\text{pos}}$) the procedure cannot be applied as ℓ_{t+k} is not available all the time. Therefore, we train for the inital three weeks the same model without $\mathcal{I}_{\text{neg}}$ and for the last three weeks the model without $\mathcal{I}_{\text{pos}}$.

The complete lasso training procedure including tuning parameter selection on the full data set takes around half a minute on the authors laptop using `glmnet` of `R` on a single core, [20]. However, it is important to use sparse matrix support to reduce computation time.

IV. TRAINING OF INDIVIDUAL FORECASTING MODELS

Given the holiday adjusted log-load $\widetilde{l}_t$ and the resulting holiday adjusted load $\widetilde{L}_t$ we train many forecasting models to create a big ensemble of forecasters (or experts). The considered models, can be categorised into four types:

A) STL-decomposed exponential smoothing → Sec. IV-A
B) AR(P) models → Sec. IV-B
C) Generalized additive models (GAMs) → Sec. IV-C
D) Lasso estimated high-dimensional linear regression models → Sec. IV-D

Only the advanced models **C)** and **D)** evaluate provided meteorologic data. The lasso model **D)** had best individual prediction accuracy. Further, all models are applied to the holiday adjusted log-load $\widetilde{l}_t$ and the holiday adjusted load $\widetilde{L}_t$. For convenience, we introduce the notation $Y_t \in \{\widetilde{l}_t, \widetilde{L}_t\}$. When considering a log-load model, the exponential function is applied to the point forecasts $\widehat{\widetilde{\ell}}_{T+h}$ for the forecasting horizon $h \in \mathcal{H} = \{h_{\min}, \dots, h_{\max}\} = \{17, 18, \dots, 40\}$ to predict the load at $T+h$.

All models were estimated using a calibration window size of $C \in \{28, 56, 77, 119, 210, 393, 758, 1123\}$ days minus 16 hours (as the last available data point was at 8am). The general idea behind this is quite simple, models with short calibration windows (e.g. 4, 8, 12 weeks) shall adjust better to more recent data, models with larger windows have more data to learn better about rare event like the annual effects. Moreover, several forecasting studies in energy forecasting have shown that combining short and long calibration windows, may lead to substantial gain in forecasting performance, see e.g. [21], [22].

The described forecasting procedure was applied in a rolling forecasting study to all days starting from 1st June 2020 as first day to predict. This date was chosen by manual inspection the historic data, as the hard COVID-19 shutdown effects seem to be vanished.

A. STL DECOMPOSITION WITH EXPONENTIAL SMOOTHING

This approach applies first an STL decomposition on Y_t. STL acronym represents the decomposition into to trend, seasonal and remainder components by loess (locally weighted scatterplot smoothing).

On the remainder component an additive exponential smoothing model is fitted. This is done using the `stlf` function of the `forecast` package in `R`, [23]. The seasonality of the time series are set to 168. Forecasting is done recursively for forecasting horizon up to $h_{\max}$, and report $h_{\min}, \dots, h_{\max}$.

B. AR(P) TIME SERIES MODEL

Here, Y_t is modeled by a simple autoregressive process ($AR(P)$) where P, sometimes used in energy forecasting [24], [25]. The only tuning parameter P is selected by minimizing the Akaike information criterion (AIC) with $P_{\max} = 24 \times 22 = 528$ (3 weeks plus 1 day). This done using the `R` function `ar` of the `stats` package in `R`, see [26]. Again, the forecasting is done recursively to $h_{\max}$, and report $h_{\min}, \dots, h_{\max}$.

C. GENERALISED ADDITIVE MODELS (GAMs)

This procedure utilized generalised additive models which are popular in load forecasting, see e.g. the winning method of the Global Energy Forecasting Competition 2014 in the load track [13]. In general, a GAM is an additive non-linear model that may utilize elements of high-dimensional linear models. A general GAM with two-way interaction depth is given by

$$Y_t = \sum_i \beta_{1,i} f_i(X_{t,i}) + \sum_i \sum_{j \neq i} \beta_{2,i,j} f_{i,j}(X_{t,i}, X_{t,j}) \qquad (6)$$

where the first term represent a sum of non-linear impacts of X_i and the second term the two-way interactions between X_i and X_j. The non-linear relationships are specified by the functions f_i and $f_{i,j}$, which allow a high degree of flexibility. Historically, f_i and $f_{i,j}$ where mainly modeled by splines, e.g. B-splines and P-splines. Nowadays, also complex non-linear model structures, like gradient boosting machines and (deep) neural networks are considered as well.

For the competition, we consider two separate GAM model designs due to the limited accessibility of the Y_{t-24} for forecasting horizons $h \in \mathcal{H}$. For hour the first 8 horizons $h \in \{17, \dots, 24\}$ the GAM model is

$$\begin{aligned} Y_t \sim & \underbrace{\sum_{k \in \{24,168\}} s(Y_{t-k}) + \sum_{k \in \mathcal{J}} Y_{t-k}}_{\text{autoregressive effects (non-linear + linear)}} + \underbrace{\text{te(hour, week)}}_{\text{weekly profile term}} \\ & + \underbrace{s(f_{\text{temp}}) + s(f_{\text{temp}}, \text{week}) + s(f_{\text{cc}}) + s(f_{\text{cc}}, \text{week})}_{\text{non-linear temperature and cloud cover effects depending on weekday}} \\ & + \underbrace{s(f_{\text{temp}}, f_{\text{cc}})}_{\text{temperature and cloud cover interaction}} \\ & + \underbrace{s(f_{\text{pres}}) + s(f_{\text{wind}}) + s(f_{\text{dircos}}) + s(f_{\text{dirsin}})}_{\text{non-linear effects from pressure, wind speed and direction}} \end{aligned}$$

for index set $\mathcal{J} = \mathcal{J}_{\text{short}} \cup \mathcal{J}_{\text{long}}$ with $\mathcal{J}_{\text{short}} = 24 \cdot \{2, 3, 8, 14, 21, 28, 35, 42\}$ and $\mathcal{J}_{\text{long}} = 24 \cdot \{350, 357, 364, 371, 378, 385\}$. Here, s represents regression smoothing terms and te tensor products, f_* represent the forecasts of the daily rolling averages of the meteorologic components. The inputs hour and week take values 1,..., 24 and 1,...,7 depending on the corresponding time t. For horizons $h > 24$, the term $s(Y_{t-24})$ in the model is dropped. The smoothing spline s is a reduced rank versions of the thin plate splines which use the thin plate spline penalty [27].

The autoregressive terms capture the dependency structure of the past for the corresponding hour. Note that yesterday's load Y_{t-24} and previous week's load Y_{t-168} is regarded as very important and therefore non-linear effects are considered. Preliminary analysis showed that the weather variables temperature and cloud cover are more relevant to explain the load behavior than other weather variables. There, we included next plain non-linear effects on each individual variable which potentially varies over the week also interaction effects. The remaining weather variables enter with non-linear smoothing effects.

The models are trained by considering only the data of the corresponding target hours. The implementation uses the `gam` function of the `R`-package `mgcv`, see [28]. The computation time for the two models with the longest considered calibration window is below a minute using a single core.

D. LASSO BASED HIGH-DIMENSIONAL REGRESSION MODELS

The lasso based models are very similar to the model used for the holiday adjustment in Section IV-D. Therefore, we only highlight the differences which concerns the autoregressive design and details on the estimation procedure.

The high-dimensional linear models are trained for each forecasting horizons $h \in \mathcal{H}$ separately. Additionally, the lag sets $\mathcal{I}_h$ in component **i)** are adjusted to $\mathcal{I}_h = \mathcal{I}_{h,\text{day}} \cup \mathcal{I}_{h,\text{week}} \cup \mathcal{I}_{h,\text{year}}$ with $\mathcal{I}_{h,\text{day}} = \{h, \ldots, 24 \cdot 15 + h\} - h$, $\mathcal{I}_{h,\text{week}} = 24 \cdot \{21, 28, \ldots, 56\} - h$ and $\mathcal{I}_{h,\text{year}} = 24 \cdot \{350, 357, 364, 371\} - h$, for $h \in \mathcal{H}$ to incorporate daily, weekly and annual autoregressive effects. The high-dimensional regression model is trained by lasso on an exponential tuning parameter grid of size 20. In detail the grid for the regularization parameter α is $2^{\mathcal{L}}$ where $\mathcal{L}$ is an equidistant grid form 6 to -1 of size 20.

V. FORECAST COMBINATION BY SMOOTHED BERNSTEIN ONLINE AGGREGATION (BOA)

The smoothed BOA algorithm is an online aggregation method that combines multivariate forecasts. Thus, it provides a weighted ensemble of forecasts where the weights change over time. The weights get updated as soon new information enters the procedure such that better-performing forecasts receive more weight. The smoothed BOA extends the standard BOA by adding a smoothing procedure to the updating schema of the multivariate forecast vectors. In the competition context, these are 24-dimensional forecasts across the prediction horizon.

A. FORMAL DESCRIPTION OF THE ALGORITHM

To introduce the smoothed BOA formally, we require some further notations. Denote $\widehat{L}_{d,h,k}$ the available load forecasts for forecast issue day d, prediction horizon h and forecasting model k. If current forecast is for day d, then we are looking for optimal combination weights $w_{d,h,k}$. This is used to combine the predictions linearly so that

$$L_{d,h}^{\text{comb}} = \sum_k w_{d,h,k} \widehat{L}_{d,h,k} \tag{7}$$

is the forecast aggregation to report. Moreover, denote $\text{AD}(x, y) = |y - x|$ the absolute deviation (also known as ℓ_1-loss) which is a strictly proper score for median predictions and MAE-minimization, see [15]. Additionally, let $\text{AD}^{\nabla}(x, y) = \text{AD}'(L_{d,h}^{\text{comb}}, y)x$ where AD' is the (sub)gradient of AD with respect to x evaluated at forecast combination $L_{d,h}^{\text{comb}}$. We require AD^{∇} to apply the so called gradient trick to enable optimal convergence rates in the BOA, see [7], [10].

The smoothed fully adaptive BOA with gradient trick and forgetting has the five update steps. In every update step we update the instantaneous regret $r_{d,h,k}$, the range $E_{d,h,k}$, the learning rate $\eta_{d,h,k}$, the regret $R_{d,h,k}$, and the combination weights $w_{d,h,k}$ for forecasting horizon h and forecaster k:

$$r_{d,h,k} = \text{AD}^{\nabla}(L_{d,h}^{\text{comb}}, L_t) - \text{AD}^{\nabla}(\widehat{L}_{d,h,k}, L_t) \tag{8}$$

$$E_{d,h,k} = \max(E_{d-1,h,k}, |r_{d,h,k}|) \tag{9}$$

$$\eta_{d,h,k} = \min\left(\frac{E_{d,h,k}}{2}, \sqrt{\frac{\log(K)}{\sum_{i=1}^{t} r_{i,k}^2}}\right) \tag{10}$$

$$\begin{aligned} R_{d,h,k} = {} & R_{t-1,k} + r_{d,h,k}\left(\eta_{d,h,k} r_{d,h,k} - 1\right)/2 \\ & + E_{d,h,k} \mathbb{1}\{-2\eta_{d,h,k} r_{d,h,k} > 1\} \end{aligned} \tag{11}$$

$$w_{d,h,k} = \frac{\eta_{d,h,k} \exp\left(-\eta_{d,h,k} R_{d,h,k}\right) w_{0,h,k}}{\frac{1}{K}\sum_{k=1}^{K} \eta_{d,h,k} \exp\left(-\eta_{d,h,k} R_{d,h,k}\right)} \tag{12}$$

with inital values $w_{0,h,k} = 1/K$, $R_{0,h,k} = 0$ and $E_{0,h,k} = 0$.

As it can be seen in equation (12) the BOA considers an exponential updating schema as the popular exponential weighted averaging (EWA), see [29]. The BOA will lead always to a convex combination of the forecasters, as the EWA. Further, is well known that the EWA in combination with the gradient trick can achieve optimal convergence rates, if the considered updating loss is exp-concave, see [29], [30]. Unfortunately, the required absolute deviation AD is not exp-concave. Therefore, the BOA uses a second order refinement in the weight update to achieve better convergence rates under weaker regularity conditions on the considered loss. In fact, the mentioned gradient trick and the second order refinement allow the BOA to achieve almost optimal convergence rates for the selection problem and convex aggregation problem. [7] and [31] prove that the BOA considered for absolute deviation loss has almost linear convergence with respect to the prediction performance of the best individual expert and a almost (standard) square root convergence with respect to the optimal convex combination. Both convergence rates are only almost optimal as there is an additional log(log) term in both convergence rates which is due to the online calibration of the learning rate.

Now, we motivate the smoothing extension of the BOA: The described BOA algorithm applies the forecast combination to each target hour h individually. However, it could be a reasonable assumption that the weights $w_{d,h,k}$ are constant across all $h \in \mathcal{H}$. This restriction reduces the estimation risk in the algorithm for sacrificing theoretical optimality. Hence, we want to find solution between those two extreme situations which finds the optimal trade-off. Therefore, we are considering smoothing splines, applied to the weights $w_{d,h,k}$. As suggested by [10] we consider cubic P-splines on an equidistant grid of knots of size 24. The smoothed weights $\widetilde{w}_{d,h,k}$ are computed by

$$\widetilde{w}_{d,h,k} = B(B'B + \lambda\Delta'\Delta)^{-1}B'w_{d,h,k} \tag{13}$$

where $\lambda \geq 0$ is a smoothing parameter, B is the matrix of cubic B-splines and Δ is the difference matrix where the difference operator is applied to the identity. Note that we difference only once, as this implies smoothing towards a constant function if $\lambda \to \infty$, see [10]. The tuning parameter λ has to be determined.

B. APPLICATION, PARAMETER TUNING AND FORECASTING RESULTS

As explained in the introduction the competition was conducted in a rolling window framework and maps realistic settings. However, for illustration purpose, we concentrate one forecasting task, this is to forecast the 1st February 2021 from 0:00 to 23:00 where the last available observation is on 31st January 2021 7:00. Fig. 6 depicts the a short overview on the considered BOA approach.

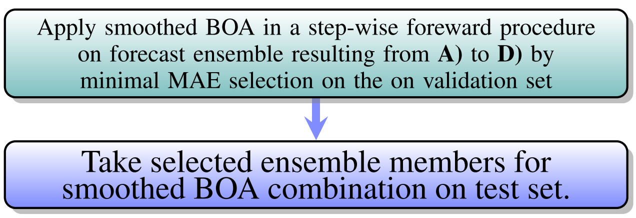

FIGURE 6. Flowchart of the smoothed BOA procedure.

We decided to utilize a stepwise forward approach to determine which forecasts to combine using the BOA with its smoothing component extension and implementation in the R package `profoc` [9] which is mainly implemented in C++. Therefore, we consider a burn-in period of 30 days (to allow local convergence of the BOA) and keep the last 60 days of available data for calibration. The final number of models M to combine was determined by evaluating the MAE of the $M_{\max} = 40$ combination procedures on the calibration data set. The results for the validation MAE across all forecasting horizons are shown in Fig (7). Additionally, we label the selected models for the optimal number of models to combine, which is 5 in this situation. The symbols represent different model types, and the colors of the five selected models correspond to those in later Figures. We see that the best individual model is a lasso model and has a validation score of about 12.5MW. Interestingly, also the other winning method [14] report best individual score for a linear model.

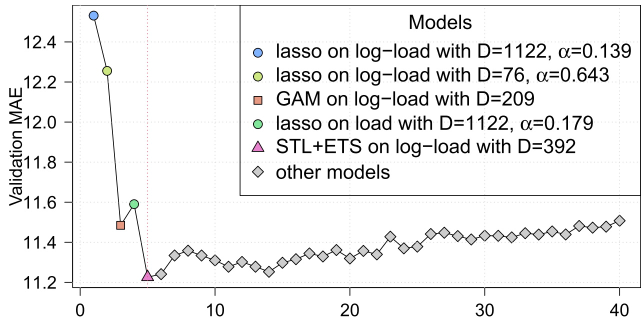

FIGURE 7. MAE (in MW) on the validation data set with highlighted optimal number of forecasting models. For the 5 optimal models we show the corresponding calibration window length D and tuning parameter λ.

Moreover, we observe that especially the first few models contribute substantially to the MAE reduction which is about 10% compared to the best individual model. It is interesting to see that the selected 5 models are quite diverse. Those are three lasso based models, a GAM model and an STL+ETS model. From the selected lasso models, two use a long history of about 3 years of data and one just a very short history of about 3 months. Also the GAM model considers a relatively short history of 7 months.

After selecting the forecasters to combine we run a BOA algorithm on an exponential λ-grid. We choose always the λ-value which performs best in the past to predict the next day. More precisely, we chose the λ-value so that the exponentially discounted MAE with a forgetting parameter $\rho = 0.01$ is minimized. Note that this forget corresponds to an effective sample size of $1/\rho$ which is 100, so about 3 months. As this combination procedure is an online aggregation method, the computation time for such combination procedure is very low. It is below a second in our setting. Fig (8) shows the results for the selected values for the smoothing parameter λ on the considered training and validation set.

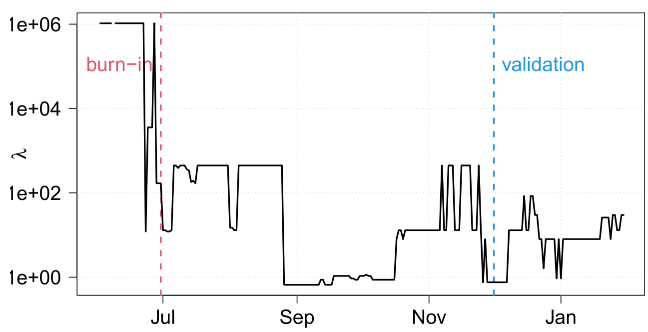

FIGURE 8. Selected λ on the test and validation set with highlighted burn-in and validation period.

We observe that the selected smoothing parameter clearly varies over time. It is also interesting to see that in the burn-in phase very high λ values where selected. This correspond to

TABLE 1. MAE values on test set (30 days) of five individual models highlighted in validation Fig. 7 with naive benchmark and performance of smoothed Bernstein Online Aggregation Method (BOA) method as used in the competition.

	$\text{lasso}^{\widetilde{\ell}}_{D=1112,\alpha=0.139}$	$\text{lasso}^{\widetilde{\ell}}_{D=76,\alpha=0.643}$	$\text{GAM}^{\widetilde{\ell}}_{D=209}$	$\text{lasso}^{\widetilde{L}}_{D1112,\alpha=0.179}$	$\text{STLF}^{\widetilde{\ell}}_{D=392}$	Naive	smoothed BOA
MAE [MW]	12.50	16.36	17.51	12.64	12.45	15.52	**11.89**
Improvement [%]	19.40	-5.40	-12.80	18.50	19.70	0.00	**23.40**

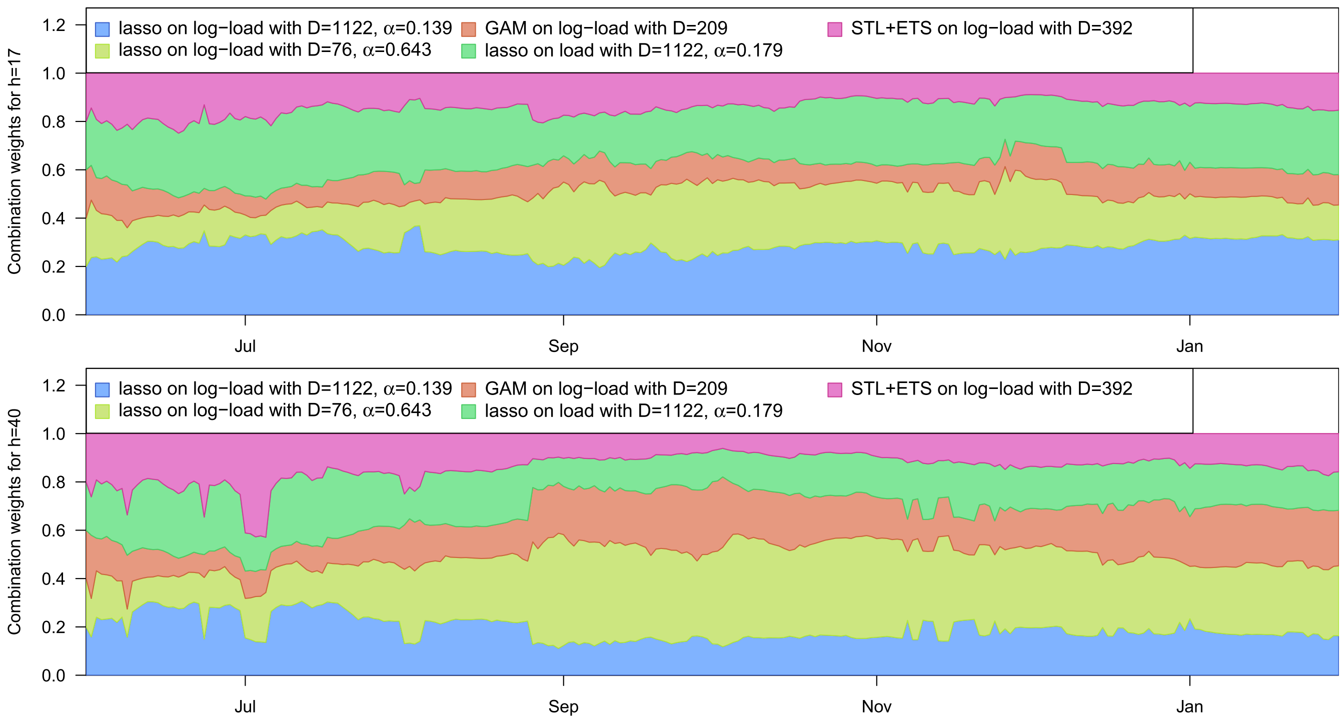

FIGURE 9. Evolution of combination weights for forecasting horizon $h = 17$ and $h = 40$, predicting the load at 0:00 and 23:00.

a conservative selection with low estimation risk. This selection is plausible, as the amount of information to evaluate is low in the burn-in period.

Fig (9) visualizes the evolution of the combination weights of the BOA algorithm over time for the forecasting horizons $h = 17$ and $h = 40$. We observe significant differences, especially the models with short calibration windows (lasso model with $D = 76$ and GAM with $D = 209$) have more weight for $h = 40$.

The same finding can be seen in Fig (10). Here, we illustrate the smoothing across the forecasting horizon for the 24 hours in the forecasting horizon. We added limiting cases with constant weights ($\lambda \to \infty$) and pointwise optimized weights ($\lambda = 0$) to illustrate the effect of smoothing. The forecast of the smoothed BOA approach is illustrated in Fig (11). There we see that the GAM model tends to underestimate and the STL+ETS model overestimated the load for the considered forecasting horizon. Thus, they can be regarded as bias correcting models.

Finally, Table 1 shows the MAE values of the 30 days test data of the individual models highlighted in Fig. 7 along with the naive competition benchmark and the proposed smoothed BOA combination. Additionally, it also provides

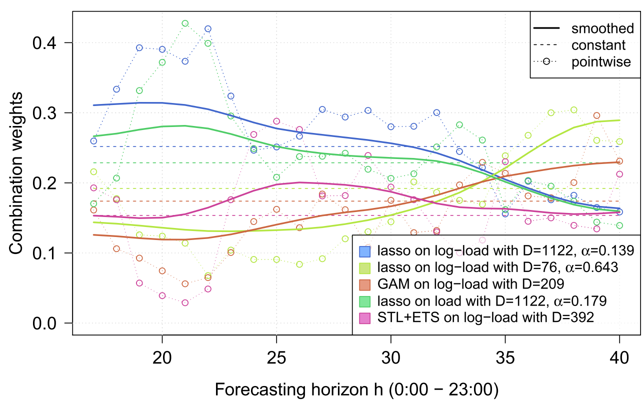

FIGURE 10. Smoothed combination weights for all forecasting horizons $h = 17, \ldots, 40$ (0:00 to 23:00) on 1st February 2021. Additionally, we show the limiting constant ($\lambda \to \infty$) and pointwise cases ($\lambda = 0$).

the corresponding improvements over the benchmark. We see that BOA method has a test MAE below 12 MW which corresponds to an improvement of 23.4%. Noticeably, this is an additional 4% improvement over the individual best model.

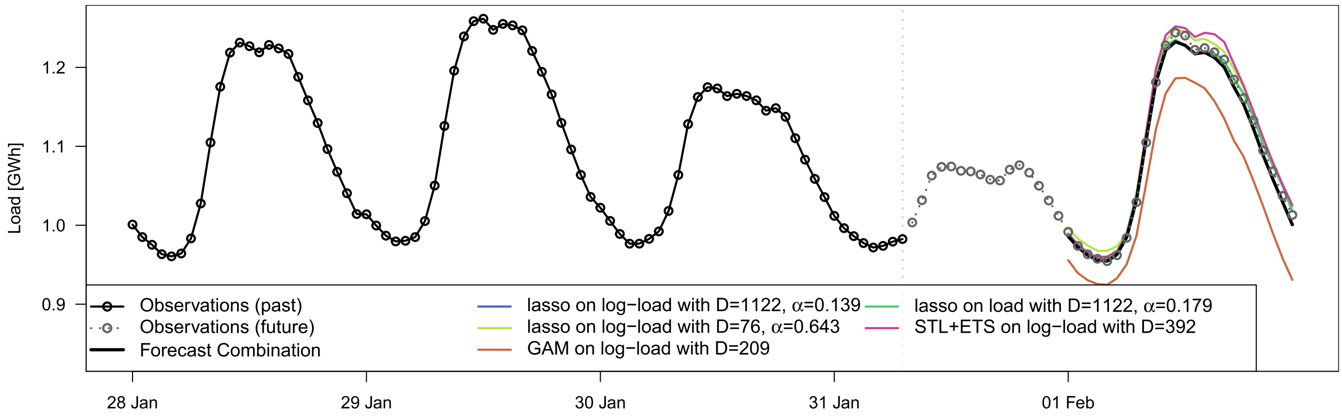

FIGURE 11. **Individual forecasts and forecast combination with observations for the 1st February 2021.**

Moreover, we see that test MAE is similar to the validation MAE of $\text{lasso}^{\widetilde{\ell}}_{D=1112,\alpha=0.139}$ is similarly as for the validation sample (about 12.5 MW). This highlights the robustness of the validation schema and indicates that the model is not overfitted.

VI. CONCLUSION

In this manuscript, we present one of the winning methods the IEEE DataPort Competition on Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm. It utilizes a novel and sophisticated holiday adjustment procedure and a new forecast combination method based on smoothed Bernstein online aggregation (BOA). The approach is flexible and can quickly adapt to new energy system situations.

Slightly better results may be obtained by a more advanced tuning parameter selection design. For example, some parameter optimization decisions are ad hoc (e.g. forgetting rate for tuning parameter selection of $\rho = 0.01$, validation period of 60 days), which could be optimized. In addition, other BOA extensions discussed in [10] like fixed share or regret forgetting could also be used. Moreover, the ensemble of individual forecasting models could also be enriched. Even though linear models designed with some degree of expert knowledge performed best adding high-depth non-linear models like gradient boosting machines or artificial neural networks may improve the results further. However, the analysis showed that the main features for this short-term load forecasting task are linear, especially the autoregressive and seasonal effects. Hence, no substantial improvement should be expected by integrating mentioned models, and some degree of expert knowledge for suitable feature engineering will be required.

APPENDIX A
NOMENCLATURE

A	Length of a meteorologic year in hours
AD, AD$^{\nabla}$	Absolute deviation, and linearized version of AD
α	Lasso tuning parameter
B	Smoothing matrix of cubic B-splines
D	Calibration window length (in days)
Δ	Difference matrix for penalized smoothing
$E_{d,h,k}$	Range estimate at d and h of expert k
$\eta_{d,h,k}$	Learning rate at d and h of expert k
f_*	Weather forecasts (f_{temp}, f_{cc}, f_{pres}, f_{wind}, f_{dircos}, f_{dirsin})
$\mathcal{I}_*$	Lag index sets ($\mathcal{I}_{\text{pos}}$, $\mathcal{I}_{\text{neg}}$, $\mathcal{I}_h$, $\mathcal{I}_{h,\text{day}}$, $\mathcal{I}_{h,\text{week}}$, $\mathcal{I}_{h,\text{year}}$)
$\mathcal{J}_*$	Lag index sets ($\mathcal{J}_{\text{short}}$, $\mathcal{J}_{\text{long}}$)
$\mathcal{H}$	Forecasting horizon index set
$h_{\min}$, $h_{\max}$	Minimum and maximum forecasting horizon
HoD$_k$, HoW$_k$	Hour-of-the-day and hour-of-the-week dummies
L_t, $L_{d,h}$	Load at time t or at day d and hour h
ℓ_t	Log-load $\ell_t = \log(L_t)$ at time t
$\widetilde{L}_t$, $\widetilde{\ell}_t$	Holiday adjusted load ad log-load at time t
$\widehat{L}_{d,h,k}$	Load forecast at day d and hour h of expert k
$L^{\text{comb}}_{d,h}$	Load forecast combination at d and h of expert k
λ	BOA P-spline shrinkage tuning parameter
M, $M_{\max}$	(Maximum) number of models to combine
$\mathcal{P}$	Grid of probabilities, $\mathcal{P} = \{0, 0.1, \ldots, 0.9\}$
q_p	p-quantile
$r_{d,h,k}$	Instantaneous regret at d and h of expert k
$R_{d,h,k}$	Regret at d and h of expert k in the BOA algorithm
ρ	Forget parameter in BOA algorithm
s	Smoothing spline
T	Calibration window length (in hours)
te	Tensor product
$w_{d,h,k}$	Combination weights at d and h of expert k
$\widetilde{w}_{d,h,k}$	Smoothed comb. weights at d and h of expert k
x^*_t	Various external regressors
Y_t	Target variable at time t, either $\widetilde{L}_t$ or $\widetilde{\ell}_t$

REFERENCES

[1] M. Narajewski and F. Ziel, "Changes in electricity demand pattern in Europe due to COVID-19 shutdowns," in *Proc. IAEE Energy Forum.*, 2020, pp. 44–47.

[2] A. M. Tudose, I. I. Picioroaga, D. O. Sidea, C. Bulac, and V. A. Boicea, "Short-term load forecasting using convolutional neural networks in COVID-19 context: The Romanian case study," *Energies*, vol. 14, no. 13, p. 4046, Jul. 2021.

[3] D. Obst, J. de Vilmarest, and Y. Goude, "Adaptive methods for short-term electricity load forecasting during COVID-19 lockdown in France," *IEEE Trans. Power Syst.*, vol. 36, no. 5, pp. 4754–4763, Sep. 2021.

[4] M. Farrokhabadi. (2020). *Day-Ahead Electricity Demand Forecasting: Post-COVID Paradigm*. [Online]. Available: https://dx.doi.org/10.21227/67vy-bs34

[5] F. Ziel, "Modeling public holidays in load forecasting: A German case study," *J. Modern Power Syst. Clean Energy*, vol. 6, no. 2, pp. 191–207, Mar. 2018.

[6] F. Petropoulos *et al.* "Forecasting: Theory and practice," *Int. J. Forecasting*, Jan. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0169207021001758

[7] O. Wintenberger, "Optimal learning with Bernstein online aggregation," *Mach. Learn.*, vol. 106, no. 1, pp. 119–141, Jan. 2017.

[8] E. Adjakossa, Y. Goude, and O. Wintenberger, "Kalman recursions aggregated online," 2020, *arXiv:2002.12173*.

[9] J. Berrisch and F. Ziel. (2021). *PROFOC: Probabilistic Forecast Combination Using CRPS Learning*. [Online]. Available: https://profoc.berrisch.biz/

[10] J. Berrisch and F. Ziel, "CRPS learning," *J. Econometrics*, 2021.

[11] S. Haben, G. Giasemidis, F. Ziel, and S. Arora, "Short term load forecasting and the effect of temperature at the low voltage level," *Int. J. Forecasting*, vol. 35, no. 4, pp. 1469–1484, Oct. 2019.

[12] F. Ziel and B. Liu, "Lasso estimation for GEFCom2014 probabilistic electric load forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1029–1037, Jul. 2016.

[13] P. Gaillard, Y. Goude, and R. Nedellec, "Additive models and robust aggregation for GEFCom2014 probabilistic electric load and electricity price forecasting," *Int. J. Forecasting*, vol. 32, no. 3, pp. 1038–1050, 2016.

[14] J. de Vilmarest and Y. Goude, "State-space models win the IEEE DataPort competition on post-covid day-ahead electricity load forecasting," 2021, *arXiv:2110.00334*.

[15] T. Gneiting, "Making and evaluating point forecasts," *J. Amer. Stat. Assoc.*, vol. 106, no. 494, pp. 746–762, Jun. 2011.

[16] M. Narajewski, J. Kley-Holsteg, and F. Ziel, "tsrobprep—An R package for robust preprocessing of time series data," *SoftwareX*, vol. 16, p. 100809, 2021.

[17] M. Narajewski and F. Ziel, *TsroBprep: Robust Preprocessing Time Series Data*, 2021. [Online]. Available: https://cran.r-project.org/web/packages/tsrobprep

[18] F. Ziel, C. Croonenbroeck, and D. Ambach, "Forecasting wind power–modeling periodic and non-linear effects under conditional heteroscedasticity," *Appl. Energy*, vol. 177, pp. 285–297, 2016.

[19] H. Zou *et al.*, "On the 'degrees of freedom' of the lasso," *The Ann. Statist.*, vol. 35, no. 5, pp. 2173–2192, 2007.

[20] J. Friedman, T. Hastie, and R. Tibshirani, "Regularization paths for generalized linear models via coordinate descent," *J. Statist. Softw.*, vol. 33, no. 1, pp. 1–22, 2010.

[21] K. Hubicka, G. Marcjasz, and R. Weron, "A note on averaging day-ahead electricity price forecasts across calibration Windows," *IEEE Trans. Sustain. Energy*, vol. 10, no. 1, pp. 321–323, Jan. 2019.

[22] K. Maciejowska, B. Uniejewski, and T. Serafin, "PCA forecast averaging—Predicting day-ahead and intraday electricity prices," *Energies*, vol. 13, no. 14, p. 3530, Jul. 2020.

[23] R. Hyndman. (2020). *Forecast: Forecasting Functions for Time Series and Linear Models*. [Online]. Available: http://pkg.robjhyndman.com/forecast

[24] F. Ziel, "Iteratively reweighted adaptive lasso for conditional heteroscedastic time series with applications to ar–arch type processes," *Comput. Statist. Data Anal.*, vol. 100, pp. 773–793, 2016.

[25] R. Steinert and F. Ziel, "Short- to mid-term day-ahead electricity price forecasting using futures," *Energy J.*, vol. 40, no. 1, Jan. 2019.

[26] R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. [Online]. Available: https://www.R-project.org/

[27] S. N. Wood, "Thin plate regression splines," *J. Roy. Stat. Soc., B Stat. Methodol.*, vol. 65, no. 1, pp. 95–114, Feb. 2003.

[28] S. Wood, *Generalized Additive Models: An Introduction with R*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2017.

[29] N. Cesa-Bianchi and G. Lugosi, *Predictive, Learning, Games*. Cambridge, U.K.: Cambridge Univ. Press, 2006.

[30] P. L. M. Gonzalez, D. J. Brayshaw, and F. Ziel, "A new approach to extended-range multimodel forecasting: Sequential learning algorithms," *Quart. J. Roy. Meteorol. Soc.*, vol. 147, no. 741, pp. 4269–4282, Oct. 2021.

[31] P. Gaillard and O. Wintenberger, "Efficient online algorithms for fast-rate regret bounds under sparsity," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 7026–7036.

FLORIAN ZIEL received the M.Sc. degree in statistics from University College Dublin, Ireland, in 2012, the Diploma degree in mathematics from the Dresden University of Technology, Germany, in 2013, and the Ph.D. degree in forecasting in energy markets from European University Viadrina Frankfurt (Oder), Germany, in 2016. He is currently an Assistant Professor of environmental economics at the House of Energy Markets and Finance, University of Duisburg-Essen, Germany. He is the first author of various peer-reviewed journal articles, most notably in top-tier IEEE TRANSACTIONS ON POWER SYSTEMS, *Applied Energy*, *Energy Economics*, *Renewable and Sustainable Energy Reviews*, and the *International Journal of Forecasting*. His research interest includes data analytics with application to energy markets and systems.

DuEPublico

Duisburg-Essen Publications online

UNIVERSITÄT DUISBURG ESSEN

Offen im Denken

ub universitäts bibliothek

This text is made available via DuEPublico, the institutional repository of the University of Duisburg-Essen. This version may eventually differ from another version distributed by a commercial publisher.

DOI: 10.1109/OAJPE.2022.3160933
URN: urn:nbn:de:hbz:465-20220812-103506-2

This work may be used under a Creative Commons Attribution 4.0 License (CC BY 4.0).