Applied Energy 277 (2020) 115527

Contents lists available at ScienceDirect

Applied Energy

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

Short-term forecasting of CO_2 emission intensity in power grids by machine learning

Kenneth Leerbeck[a,*], Peder Bacher[a], Rune Grønborg Junker[a], Goran Goranović[a], Olivier Corradi[b], Razgar Ebrahimy[a], Anna Tveit[a], Henrik Madsen[a,c]

[a] Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark
[b] Tomorrow (Tmrow IVS), Njalsgade, 2300 Copenhagen, Denmark
[c] Faculty of Architecture and Design, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway

HIGHLIGHTS

- Short-term CO_2 emission forecast in the power system.
- Emission-based energy flexibility.
- Effect of international power exchange interactions.
- Average versus marginal CO_2 emissions.

ARTICLE INFO

Keywords:
CO_2 emission forecasting
Electrical power grids
Machine learning
Feature selection
LASSO
ARIMA

ABSTRACT

With the aim of enabling effective flexible electricity demand, a machine learning algorithm is developed to forecast the CO_2 emission intensities in European electrical power grids distinguishing between average and marginal emissions. The analysis focuses on Danish bidding zone DK2 and was done on a data set comprised of a large number (473) of explanatory variables such as power production, demand, import, weather conditions etc., collected from selected neighboring zones. The number of variables was reduced to less than 30 using both LASSO (a penalized linear regression analysis) and a forward feature selection algorithm. Three linear regression models that capture different aspects of the data (non-linearities and coupling of variables etc.), were created and combined into a final model using Softmax weighted average. Cross-validation is performed for debiasing and an autoregressive moving average model (ARIMA) implemented to correct the residuals, making the final model an ARIMA with exogenous inputs (ARIMAX). Forecast errors vary between 0.095 and 0.183 (NRMSE) for the average emissions and 0.029–0.160 for the marginals depending on the forecast horizon (1–24 h). The forecasts with the corresponding uncertainties are analyzed and performance on very short (below six hours) and longer horizons are discussed –. One interesting result is that the marginal emissions were shown to be highly independent of any variables in the DK2 zone, suggesting that the marginal generators are located in the neighboring zones. The developed methodology can be applied to any bidding zone in the European electricity network without requiring detailed knowledge about the zone and with very few manual interactions.

1. Introduction

Generation of electricity contributes heavily to CO_2 emissions [1]. The intensity with which it does so depends on the proportion of renewable sources (e.g. solar, wind) vs. nonrenewable sources (e.g. coal, gas, nuclear). The proportion, and hence the CO_2 emission intensity, fluctuates with time based on power market mechanisms and weather conditions. Ideally, in the future, electricity users (the demand) will respond to the renewable power generation in a n attempt to lower the emissions. Proposed solutions include scheduling of storage (e.g. batteries, fuel cells, hydro reservoirs) and flexible demand (e.g. heat pumps, electric cars) [2]. Forecasting of the CO_2 emission intensity is essential for this scheduling and the present paper describes a methodology for carrying out such forecasting with minimum effort – hence it provides an important component to fulfill the goal of lowering the emissions.

* Corresponding author at: Anker Engelunds vej 1, Building 101A, 2800 Kongens Lyngby, Denmark.
E-mail address: kenle@dtu.dk (K. Leerbeck).

https://doi.org/10.1016/j.apenergy.2020.115527
Received 7 November 2019; Received in revised form 24 June 2020; Accepted 14 July 2020
Available online 31 July 2020

Forecasting of conditions in the power grid is essential and already exists in various forms. For renewables, new forecasting methods are developed on a regular basis to increase revenue. Recently in [3], a multi-step ahead method for deterministic forecasting of wind power has been presented (many prior models only did one-step ahead), using complex multi-stage machine learning (kernel-based) algorithms for error correction. The deterministic model, however, cannot include information about forecast uncertainty for the volatile wind speeds. Hence a probabilistic model (distribution forecast) of wind speeds based on robust machine learning algorithms was developed in [4]. The probabilistic models quantify uncertainty of a forecast, crucial for risk management. On the other hand, multi-step forecasts are important in any scheduling application, such as market bids and flexibility. Examples of the combination of the two types of models are the multi-step ahead probabilistic solar power forecasts developed in [5,6] with horizons of 36 and 72 h ahead, respectively. For scheduling and analysis of emission reductions it is very important to distinguish between average and marginal generation. A generator responding to small changes in demand is called a marginal generator and is not weather dependent (must generate on-demand) – hence the category does not include wind turbines and PV-panels. A good estimate of the marginal generator is achieved by using price signals. In Fig. 1, this is exemplified with a merit order on the day-ahead market. The highest generator in the merit order is the one crossing with the demand curve – a coal generator in this case as indicated by the orange color. The **average emission** is the weighted average from all activated generators. The **marginal generator** is the generator that will be activated by moving the demand line (dashed blue line) slightly to the right.

Because price-based scheduling can be both economically and environmentally beneficial, forecasting of the day-ahead spot prices has been proposed in recent papers [7,8], using different methodologies (e.g. ARIMA, Neural Networks, machine learning). However, there is a problem with using spot prices, known as the merit order emission dilemma, illustrated for the German-Austrian power market [9]: The price for coal is low but the emissions are high. A price-based scheduling, therefore, only leads to low emissions if there is a surplus of renewable energy (more renewable energy than needed) – otherwise coal is favored.

Precise estimates and forecasts of the marginal CO_2 emissions are needed to effectively lower emissions using storage and flexible demand incurrent power grids – this will naturally change in a 100% renewable scenario. In discussing CO_2 emissions, two distinct concepts are used; average and marginal emission intensities $\left(\frac{kgCO_2-eq}{MWh}\right)$. Average emissions correspond to the overall, e.g. region-wide, electricity generation including net imports (sum of imports and exports). The marginal intensity reflects the emissions of the marginal generator. The concepts are compared in [10] and the importance of distinguishing between them is highlighted due to their highly opposing patterns.

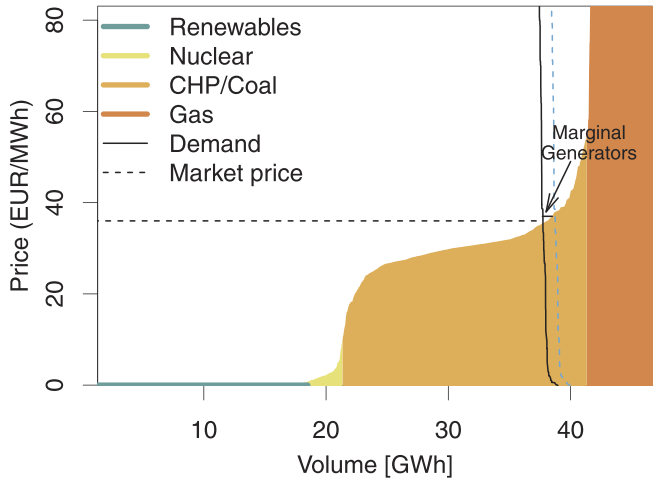

Fig. 1. The merit order illustrated with a supply and demand curve example. On the x-axis is the volume of the accumulated generators in the power system and on the y-axis are their corresponding costs. The color indicates the type of generator and the lines indicate key numbers. Data source: Nord Pool AS. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)(For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Estimating the average emissions is straight forward by summing the power generation including import/export and finding the respective average emissions. Traditionally, the imports have been treated as dummy variables, but in a recent study [11] that focused on the average emissions in the Nordic European countries, they have been treated explicitly. The study revealed the interplay between the imports and the average emissions, and how both vary from one bidding zone to another. Estimating the marginal emissions becomes much more tricky. Early studies of the marginal emissions, e.g.[12],estimate the highest generator in the merit order of the power generation system. However, this is rarely the only generator responding to a change in demand, as addressed in [13],where these early approaches are discussed and a new empirical approach is suggested. By estimating the contribution of all power generators to a specific change in system demand, the power plants were disaggregated to investigate the impact of plant turnovers (switching old power plants for new ones). The approach is using linear regression on historical data, including outputs from major power producers in the UK, to estimate the average response from each generation technology class (e.g. coal, wind, solar etc.) to changes in demand. Incorporating the imports in the marginal emissions, the company Tomorrow (www.tmrow.com) has developed a new empirical approach to assess the impact of a specific generator or load on the power system[14].The approach is using machine learning on historical data to model the chain of imports (the so-called flow tracing, originally introduced in[15,16])– enabling the tracing of marginal generators in regions potentially very far away. The approach is scaled using data from the majority of bidding zones around the world[17].

The above mentioned studies provide methodologies for marginal CO_2 emission *estimation*. Also, the *long-term* (e.g. annual) forecasting of CO_2 emissions is widely conducted for promoting green energy, e.g. [18]. However, the more accurate short-term emission forecasting methodologies, which are needed to implement the flexible demand, are currently unavailable in the literature. In the present study a methodology for short-term (24 h ahead) forecasting with uncertainty margins (95% prediction intervals) of both the average and marginal CO_2 emission intensities, is described. The results of applying it to power generation in bidding zone DK2 (Sealand region, Denmark) are presented and discussed. The forecasts can be used by flexible consumers (with electric cars, heat pumps, etc.) to schedule for optimal electricity usage, i.e. minimizing CO_2 emissions. The methodology can be applied to other bidding zones in the European electricity network without requiring detailed knowledge of response and explanatory variables with a minimum of effort as model selection and tuning is fully automized.

In the presented methodology the given response variable (average and marginal CO_2 emission intensity) is modelled using explanatory variables – e.g. power generation, demand, import, weather conditions – which are represented by generalized basis functions (splines). Examples of the explanatory variables and their relation to the response variables are shown in Section 2. The data is divided into two sets: one set available for $\leqslant 6$ hour horizons and another available for horizons >6 hours. The applied machine learning techniques, which include trend extraction (seasonality, nonlinearity, interaction terms), feature selection (LASSO, forward feature selection), residual correction (ARIMA), and cross-validation, are described in Section 3 and 4 – where

three different models are also built and combined with a Softmax weighted average into the final forecasting model. The most significant variables and forecasting results are highlighted in Section 5, and concluding remarks are found in Section 6. A list of all the variables can be reviewed in Appenidx A, where it is indicated which variables are being used for which set of data.

2. Explorative data analysis

The CO_2 emissions in Denmark are interesting when scheduling flexible consumers due to large amounts of wind power generation – wind covered 48% of the total electricity generation in 2017. The country also has a variety of good trading options with its neighboring countries – e.g. Germany (DE), Sweden (SW) and Norway (NO). Indirectly, influence can also come from countries further away, but the scope of this study is limited to selected surrounding bidding zones DE, DK1, DK2, NO2, SW3 and SW4. The focus area, DK2, has direct transmission cables to DK1, SW4 and DE.

First, linear relationships between the response and explanatory variables are explored. The power generation in DK2 and net import in SW4 from SW3 show the highest correlations to the response variables as illustrated in Fig. 2. It can be seen that the average emissions are highly correlated to the power generation in DK2, simply because a higher level of generation will mainly activate coal and gas-fueled generators. Net import in SW4 from SW3 shows the highest correlation to the marginal emissions, this is due to an indirect influence – the marginal generator is often located in SW3 where all Sweden's nuclear power is produced [19]. This is because nuclear power is cheaper than the local options in DK2, e.g. coal and gas. Fig. 2 also shows a yearly seasonality in the local power generation varying by about 1,000 MW – lowest in the summer. The import shows the same yearly seasonality, but less significant because nuclear power serves the baseload.

Next, a linear regression model was fitted onto the average and marginal emissions using the discussed power generation and import respectively to reveal other important variables. The residuals from these models will represent the average and marginal emissions that are independent of these variables. This reveals the non-linear relationships shown in Fig. 3.

The marginal emissions are highly dependent on the net import to DK1 from DE but only when DK1 exports to DE. The higher the export the more domestic generators must serve as marginals. The net import in DK2 from SW4 shows a yearly seasonality being highest in the summer – since demand is generally low, the proportion of nuclear power is large in Sweden, and Denmark can therefore import from Sweden rather than generate locally from gas or coal.

3. Regression models and basis functions

3.1. Linear Regression Models (lm)

The CO_2 emissions are modeled using a multivariate linear regression model

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon, \tag{1}$$

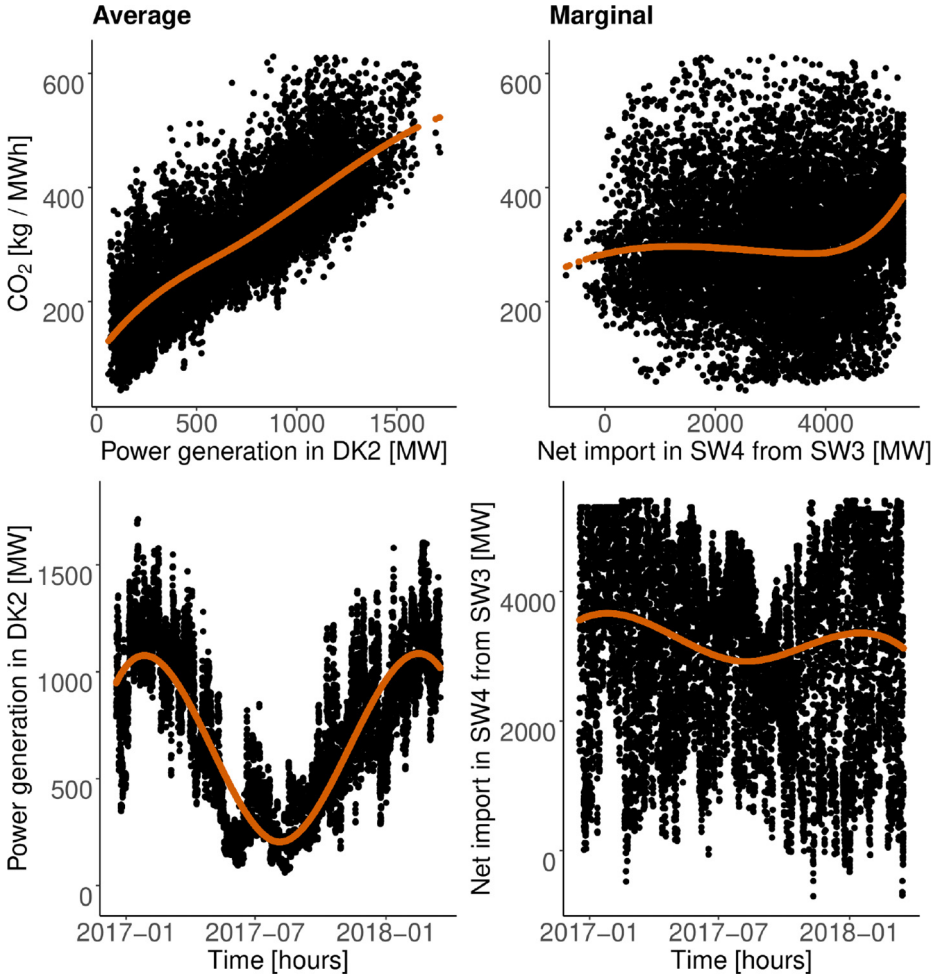

Fig. 2. Top left: The average CO_2 emission intensity vs. the power generation in DK2. Top right: The marginal CO_2 emission intensity vs. the net import in SW3 from SW4. Bottom: The power generation in DK2 and the net import in SW3 from SW4 vs. time (hourly resolution).

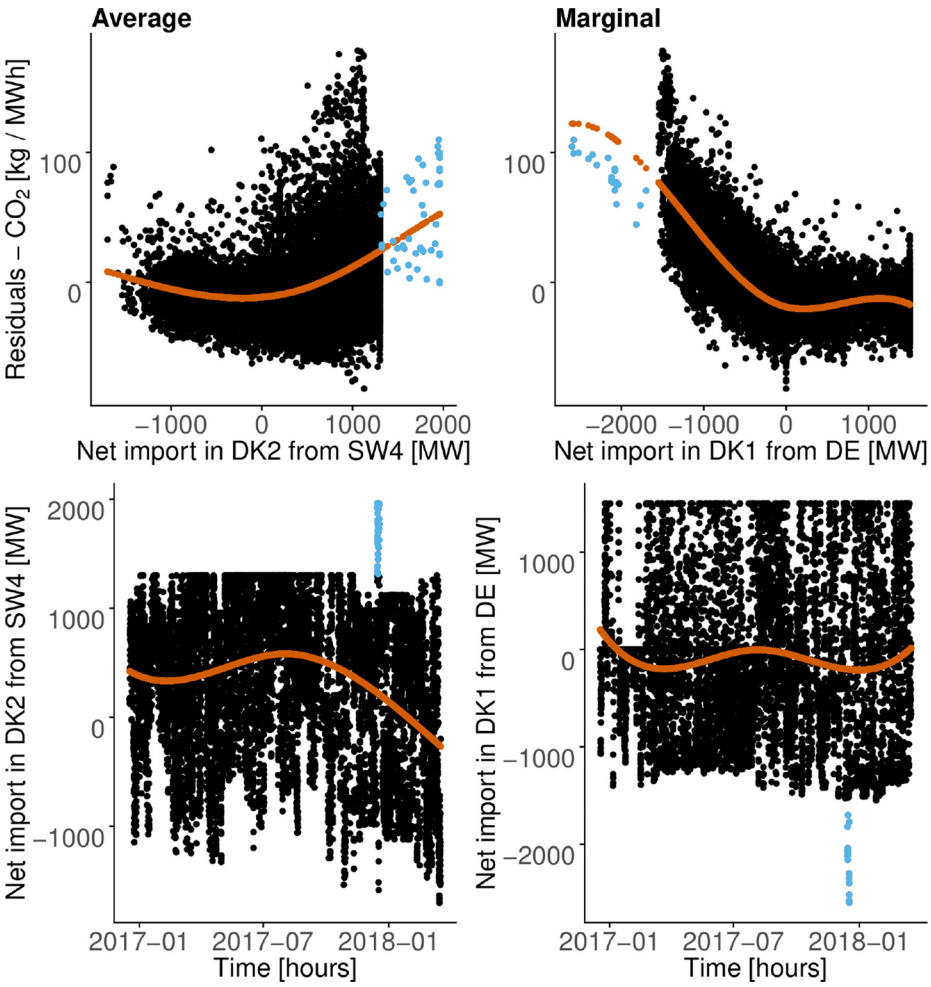

Fig. 3. Top left: The residuals from the linear fit of the domestic power generation (DK2) onto the average CO_2 emission intensity vs the net import in DK2 from SW4. Top right: The residuals from the linear fit of the import in SW4 from SW3 onto the marginal CO_2 emission intensity vs the net import in DK1 from DE. Bottom: The net import in DK2 from SW4 and the net import in DK1 from DE vs time (hourly resolution). Note the outliers (light blue); over three days, the import seemingly overloaded both the transmission cables to Germany and Sweden. For this study, these data points are modified to the maximum capacity (1300 and 1,550 MW in these cases). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for $\epsilon \sim N(0, \sigma^2 I)$,

where $\mathbf{X}$ is the input matrix (explanatory variables), $\mathbf{Y}$ is the output vector (response variables: CO_2 emissions) and β is a vector of regression coefficients to be found. ϵ represents errors which should be independent and identically distributed (i.i.d.) if the model is not underfitted.

The least square regression is performed to minimize

$$S(\beta) = ||\mathbf{Y} - \mathbf{X}\beta||^2$$

and obtain the ordinary least-squares solution

$$\beta = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}. \quad (2)$$

Once β is obtained on training data, the response variables can be forecast on new input data (the test data) by

$$\hat{\mathbf{y}}_{t+h} = [\mathbf{x}^{\mathrm{F}}_{t+h} \quad MA(\mathbf{x}^{\mathrm{RT}}_{t+h})_{24} \quad MA(\mathbf{x}^{\mathrm{RT}}_{t+h})_{48} \quad lag(\mathbf{y})_{t+h}]\beta \quad (3)$$

$$\equiv \mathbf{z}^{*}_{t+h}\beta, \quad (4)$$

where F and RT refer to Forecast and Real-Time, and MA to Moving Average. Note, the model above is not used as stated, but is extended as explained in the following sections. The terms in the brackets (the matrix elements) are provided by Tomorrow (www.tmrow.com): $\mathbf{x}^{\mathrm{F}}\_t + h$ are the input variables forecasted h-hours ahead ($h \leqslant 6$ hours for all explanatory variables except weather data which is also available for $h \geqslant 6$ hours). The moving average is constructed to smooth the real-time input variables by

$$MA(\mathbf{x}^{\mathrm{RT}}\_t + h)\_n = \frac{1}{n}\sum \_i = 0^{n-1}\mathbf{x}^{\mathrm{RT}}\_t - i, \quad (5)$$

where n is the length of the averaging period – either 24 or 48 h. Finally, $lag(\mathbf{y})\_t + h$ in Eq. 4 is just $\mathbf{y}\_t$, the last available observation at time t, thus an Auto-Regressive (AR) term.

Both $\mathbf{X}$ and $\mathbf{Y}$ are given as multivariate time series. $\mathbf{X}$ consists of 418 variables - for $h \leqslant 6$ hours - (related both to weather and power system from six bidding zones; 66 listed in Appenidx A) and $\mathbf{Y}$ of 10,897 observations. The forecasted input variables with horizon h constitute a new input matrix $\mathbf{Z}$ that will be expanded to model periodicities and non-linear relationships, as explained in the next three subsections.

3.2. Periodic variations: Fourier Series

Periodic variations (seasonality) of the average and marginal CO_2 emission intensities are investigated using Fourier Series defined as

$$FS(n, period)\_t = \hat{\mathbf{y}}\_FS(n, period), \; t= \quad (6)$$

$$A\_0 + \sum \_i = 1^{n-\mathrm{series}} A\_i \cdot \sin\left(i\frac{2\pi t}{\mathrm{period}}\right) + B\_i \cdot \cos\left(i\frac{2\pi t}{\mathrm{period}}\right),$$

where $\mathbf{y}$ is the response variable, t is the time, A and B are linear regression coefficient matrices, n is the order of the Fourier Series and 'period' is the length of the seasonality period. n is adjusted to find the best fit. In Fig. 4, the daily, weekly, and yearly patterns are estimated using data ranging from January 2015 to January 2018 with $n = [2, 1, 2]$respectively.

It is observed that average and marginal emissions follow completely different patterns. The yearly pattern shows the largest variations for the average emissions, as already discussed in Section 2. The daily average emission varies very little in comparison to the yearly

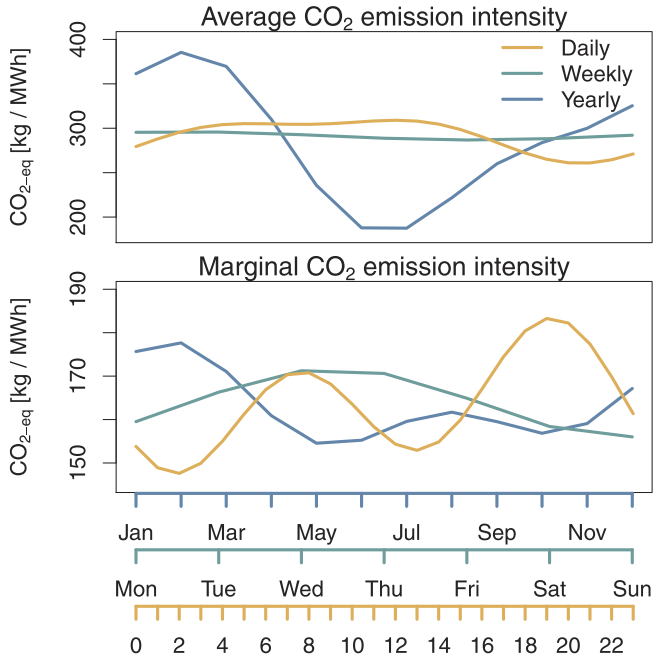

Fig. 4. Fourier Series showing the daily, weekly and yearly patterns for the average and marginal CO_2 emission intensities.

emissions. On the other hand, the daily marginal emissions are the highest when the average emissions are the lowest, illustrating the importance of using the correct emission measure. The weekly pattern is only of importance to the marginal emissions and is lowest on the weekends. The seasonality components are added to the input matrix $\mathbf{Z}$ of Eq. 4, such that

$$\mathbf{z}\_t = [\mathbf{z}^*\_t \quad \exp\mathbf{z}^*\_t \quad FS(2, 24)\_t \quad FS(1, 7)\_t \quad FS(2, 12)\_t], \tag{7}$$

Note that an exponential term of $\mathbf{Z}$ is added, too, which makes Z a (10,897x951) matrix (both $\mathbf{Z}^*$ and $\exp(\mathbf{Z}^*)$ have 474 columns after the clean-up of non-available points and constant variables).

3.3. Non-linearities: splines

Non-linearities are captured by using splines, the local polynomials between specified points called knots [20]. Splines are implemented in R with the built-in functions `bs()` (base splines) and `ns()` (natural splines). The former are basis functions and increasing their number in the expansion of a function improves the fitting procedure at the risk of overfitting. In this study, four base splines with knots located at the quantiles (default settings in R), are used to avoid overfitting

$$bs(\mathbf{z}\_t) = [bs\_0(\mathbf{z}\_t) \quad bs\_1(\mathbf{z}\_t) \quad bs\_2(\mathbf{z}\_t) \quad bs\_3(\mathbf{z}\_t)]^\top. \tag{8}$$

The number is justified by the assumption that the relationships between the explanatory and the response variables are stationary. The least-square coefficients associated with the base splines are labeled by the vector β^{bs}. To refine the fitting, natural splines $ns(\mathbf{z}\_t)$ are also used and represented analogously to Eq. 8.

Fig. 5 shows the estimated mean (orange) of the average CO_2 emissions in DK2 calculated on the basis of the demand, based on four and six splines (green).

3.4. Interaction terms

The interaction [21] between the explanatory variables is modeled as

$$IA(\mathbf{z}\_i, t, \mathbf{z}\_j, t) = [\mathbf{z}\_i, t \quad \mathbf{z}\_j, t \quad \mathbf{z}\_i, t{\cdot}\mathbf{z}\_j, t]^\top, \tag{9}$$

where the product $\mathbf{z}\_i{\cdot}\mathbf{z}\_j$ denotes the mutual interaction of the (i, j) pair. The coefficients are denoted as vector β^{IA}.

Interactions were also represented by splines to refine the non-linearity, i.e.

$$bs\left(IA\left(z\_i, t, z\_j, t\right)\right) = \begin{bmatrix} bs\_0(\mathbf{z}\_i, t) & bs\_0(\mathbf{z}\_j, t) & bs\_0(\mathbf{z}\_i, t{\cdot}\mathbf{z}\_j, t) \\ bs\_1(\mathbf{z}\_i, t) & bs\_1(\mathbf{z}\_j, t) & bs\_1(\mathbf{z}\_i, t{\cdot}\mathbf{z}\_j, t) \\ bs\_2(\mathbf{z}\_i, t) & bs\_2(\mathbf{z}\_j, t) & bs\_2(\mathbf{z}\_i, t{\cdot}\mathbf{z}\_j, t) \\ bs\_3(\mathbf{z}\_i, t) & bs\_3(\mathbf{z}\_j, t) & bs\_3(\mathbf{z}\_i, t{\cdot}\mathbf{z}\_j, t) \end{bmatrix}^\top, \tag{10}$$

with the corresponding coefficient matrix $\beta^{bs(IA)}$.

Note that explanatory variables generally change in time; for example, the power generation in DK2 has a clear daily pattern and its corresponding linear regression coefficient will vary accordingly. This can be expressed as interactions with the time variables (hour, week, month). A separate matrix τ is thus defined to group the periodic as well as nonlinear character of the time variables Appendix B.

4. Statistical selection and refinement of models

4.1. Cross validation strategy

Throughout the study rolling forward cross-validation is used, where data is divided into eight sets each consisting of training, validation and testing data [22]. There are eight rounds of cross-validation, where the training data is increased by one set in each round (the validation set of a previous round becomes part of the training data in the next one), and the validation and testing sets are always new independent data sets.

The cross-validation is done by averaging the Root Mean Squared Error (RMSE) over the eight validation sets. The final comparison of models is made on the test sets. The data range is 15 months, from 19th December 2016 to 18th March 2018. This is a smaller period than the one mentioned in seasonality, Section 3.2. The reason is that many of the explanatory variables have limited historical data and hence cannot be cross-validated further back.

4.2. Feature selection techniques

Feature selection is necessary to remove co-linearity and overfitting in linear regression. Co-linearity happens when two or more explanatory variables are linearly dependent or highly correlated. When this happens the condition number of the matrix $\mathbf{X}$ is lowered, making the determinant of it close to zero, and thus inverting it results in large numerical errors. Overfitting is related to a large number of model parameters used to fit training data, which then causes poor model performance on test data, as seen from the yellow points in Fig. 6. In the extreme, high order polynomials are found to perfectly fit scattered data that in reality follow a simple, say, linear trend.

Two methods are used for feature selection. In the first, Least Absolute Shrinkage and Selection Operator (LASSO) algorithm, a penalty term is added to the objective function S that is to be minimized

$$S(\beta) = ||\mathbf{y} - \mathbf{X}\beta||^2 + \lambda||\beta||\_1, \tag{11}$$

where λ is the penalty parameter and the subscript 1 indicates the L1 norm; the larger the L1 norm of the coefficient β, the larger the penalty [23]. This reduces parameter estimates to zero, hence its name "Shrinkage".

The penalty term λ is tuned using the average RMSE from the validation sets defined in the above eightfold cross-validation strategy. The higher the value of λ the more the β's shrink towards zero and therefore fewer variables will be selected. λ is slowly decreased from an initial high value until the optimal value is reached when model performance stops improving. If the performance starts to decrease, the model is over-fitted.

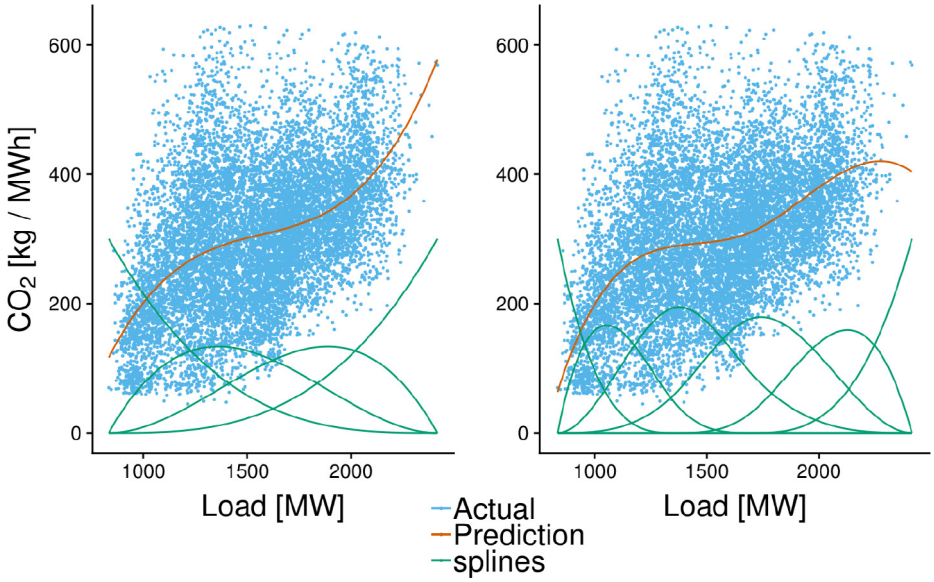

Fig. 5. Estimation of the average CO_2 emission intensity (orange) based on four and six splines, respectively. The splines are scaled for illustration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

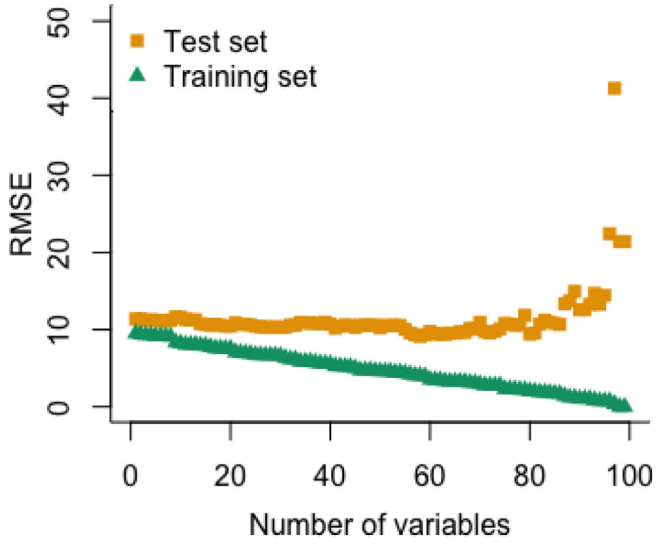

Fig. 6. RMSE vs number of variables. Generated using 200 observations from 100 independently simulated explanatory variables, all with a certain degree of correlation to a simulated response variable.

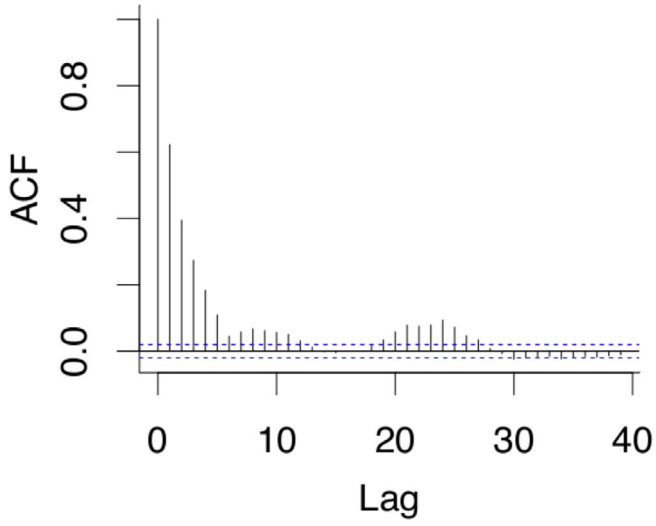

Fig. 7. Auto Correlation Function of the residuals from the combined model – **y** is the average CO_2 emission intensity and $h = 6$ hours – on the training set from CV set 8.

Before applying LASSO, highly correlated variables are removed manually to reduce the computation time of the LASSO regression.

The second method is the forward feature selection algorithm in which new variables are added to the best models and tested for improvement, Appenidx C. These methods are used for this study in the following order:

1. Highly correlated variables ($\rho > 0.99$) are removed.
2. LASSO regression is applied.
3. The forward selection algorithm is applied.
4. Step two and three repeated with updated variables until convergence (number of variables does not decrease anymore).

4.3. Residual correction

When the model does not accurately predict the test data, residuals are non-random and become auto-correlated (correlated to lagged versions of themselves). Thus corrections to the model need to be made to account for the correlation.

Residual auto-correlations are checked with the ACF (Auto Correlation Function) that reveals any linear dependencies in the residuals. In Fig. 7, residuals of the compound model (see later, Section 4.5) of the average CO_2 emission intensity on a six-hour horizon is shown. There are high correlations up until the lag of six hours, and also a smaller correlation around 24 h due to seasonality.

The residual was modeled independently with an Auto Regressive Moving Average (ARIMA) model [24]

$$\mathbf{Y}\_t = \epsilon\_t + \psi\_1 \epsilon\_t - 1 + \cdots + \psi\_q \epsilon\_t - q - \phi\_1 \mathbf{Y}\_t - 1 - \cdots - \phi\_p \mathbf{Y}\_t - p, \tag{12}$$

containing p lagged values (AR part) and q errors of previous observation of the moving average (MA). Considering an AR model the prediction errors are obtained as

$$\mathbf{Y}\_t + h - \widehat{\mathbf{Y}}\_t + h|t = \epsilon\_t + h + \psi\_1 \epsilon\_t - 1 + h + \cdots + \psi\_h - 1 \epsilon\_t + 1. \tag{13}$$

The models of this type are denoted $ARIMA_{p,d,q}$ process, where parameters p, d, q refer to the AR, I and MA part, respectively – I is an integrating term used to make the data stationary (mean and variance are constant over time). In this study an extension is used, $ARIMA\_(p, d, q)(P, D, Q)\_M$, where P, D and Q are the seasonal parameters and M is the length of the season. The parameters are fitted using information from the ACF and using the built-in function `auto.arima()` in R, which automatically selects the model with the best fit. In this case, it is found to be seasonal $ARIMA\_(3, 0, 0)(0, 1, 2)\_24$ and $ARIMA\_(5, 1, 0)(2, 0, 0)\_24$ models that remove all significant correlations for the average and marginal emission models with $h = 6$ hours.

The prediction of the residual model is added to Eq. 4 to obtain the final prediction

$$\mathbf{y}\_t + h|t = \hat{\mathbf{y}}\_t + h + ARIMA(\mathbf{y}\_1{:}\ t - \hat{\mathbf{y}}\_1{:}\ t, h) + \epsilon\_t + h|t \tag{14}$$

where $\mathbf{y}\_t + h|t$ and $\epsilon\_t + h|t$ are the response variable and error at time $t + h$ given (residual) information at t, and $\hat{\mathbf{y}}\_t + h$ is the prediction from the linear model. $ARIMA(\mathbf{y}\_1{:}\ t - \hat{\mathbf{y}}\_1{:}\ t, h)$ is the ARIMA predicted error in line with Eq. 13, however with an extended version. Note that once obtained, the ARIMA model is used for predicting the residuals at the horizon h. Finally, the uncertainty of the model is evaluated by applying the 95% prediction interval. This is in accordance to the equations defined in [24], and is applied easily in R through built-in options in `lm()` and `arima()`.

The described residual correction can improve both the forecast for the specific horizon and the forecast for the lower horizons at the same time. Besides this, it also gives more consistent results since the cross-validation sets do not stand out (the variance of the errors becomes smaller).

4.4. Base models

The formalism of Sections 3, 4.1 and 4.2 is used to assemble different models of increasing complexity, listed below. The starting point is the model

- **M0**

$$\mathbf{y}\_t + h = \beta\_0 + \sum\_i = 1^{951}\beta\_i(\mathbf{z}\_i, t + h) + \epsilon\_t + h,$$

where the 951 refer to the columns in $\mathbf{Z}$.

- **M1**

$$\mathbf{y}\_t + h = \beta\_0 + \sum\_i = 1^{951}\beta\_i(\mathbf{z}\_i, t + h) + \sum\_i = 1^{15}\beta^{\tau}\_i(\tau\_i, t + h) + \sum\_i = 1^{951}\beta^{bs}\_i \cdot bs(\mathbf{z}\_i, t + h) + \sum\_i = 1^{951}\beta^{ns}\_i \cdot ns(\mathbf{z}\_i, t + h) + \epsilon\_t + h$$

includes time variables τ and non-linearities (splines).

- **M2**

$$\mathbf{y}\_t + h = \beta\_0 + \sum\_i = 1^{50}\sum\_j = 1 \wedge j \neq i^{50}\beta^{IA}\_i, j, k \cdot IA(\mathbf{z}\_i, t + h, \mathbf{z}\_j, t + h)\_k + \sum\_i = 1^{50}\sum\_k = 1^{15}\beta^{IA}\_i, k, l \cdot IA(\mathbf{z}\_i, t + h, \tau\_k, t + h)\_l + \epsilon\_t + h$$

is based on the reduced number of features (maximum 50, obtained from the LASSO regression of **M0** and ranked based on the size of their linear regression coefficients), and their interactions defined by the vector IA, Eq. 9.

- **M3**

$$\mathbf{y}\_t + h = \beta\_0 + \sum\_i = 1^{50}\sum\_j = 1 \wedge j \neq i^{50}\beta^{bs(IA)}\_i, j, l, m \cdot bs(IA(\mathbf{z}\_i, t + h, \mathbf{z}\_j, t + h))\_l, m + \sum\_i = 1^{50}\sum\_k = 1^{21}\beta^{bs(IA)}\_i, k, l, m \cdot bs(IA(\mathbf{z}\_i, t + h, \tau\_k, t + h))\_l, m + \epsilon\_t + h,$$

is also based on the reduced features, but with the interactions defined via matrix $bs(IA)$ of Eq. 10.

Table 1
Root-Mean-Squared Error [RMSE] of the average and marginal emissions for models **M1**-**M3**, with a prediction horizon of 6 h. The weights are calculated using the RMSE values from the validation sets. $\mathbf{M_{WA}}$ is the resulting weighted average model. Units: $\left[\frac{\text{kgCO\_2-eq}}{\text{MWh}}\right]$.

Average	M1	M2	M3	$\mathbf{M_{WA}}$ (*Softmax/Flat*)
RMSE $\left[\frac{\text{kgCO\_2-eq}}{\text{MWh}}\right]$:				
-Validation	39.6	39.0	39.2	**37.86/37.88**
-Test	40.6	39.8	39.4	**38.49/38.55**
Weights:				
-Softmax	**0.22**	**0.43**	**0.35**	
-Flat	**0.33**	**0.34**	**0.33**	
LASSO penalty (λ)	*1.6e−1*	*2.5e−4*	*1.6e−1*	
Marginal				
RMSE $\left[\frac{\text{kgCO\_2-eq}}{\text{MWh}}\right]$:				
-Validation	11.1	8.8	10.0	**8.6/8.9**
-Test	11.9	9.9	10.8	**9.63/9.85**
Weights:				
-Softmax	**0.07**	**0.73**	**0.20**	
-Flat	**0.30**	**0.38**	**0.33**	
LASSO penalty (λ)	*1.0e−3*	*2.5e−3*	*1.6e−5*	

The feature selection procedure defined in Section 4.2 is applied to **M1**, **M2** and **M3** and reduces the number of variables to 10–30 depending on the horizon.

4.5. Weighted average model

The final model was the weighted average of the above Models 1–3. The weights were based on the performance of models on eight validation sets. The Softmax function was used

$$w\_i = \left(\frac{\exp x\_i}{\sum\_j = 1^n \exp x\_j}\right),$$

where w is the weight vector, x is a vector representing the average *RMSE* scores of the included models on the validation sets and n is the number of models included. Compared to the flat weight, $w\_i = \left(\frac{x\_i^{-1}}{\sum\_j = 1^n x\_j^{-1}}\right)$, the Softmax function gives more weight to the good models and almost neglects the bad ones due to the exponential term. The results clearly state this as superior to the flat weights (Table 1).

Model **M2** with the linear interaction terms is the best model for both the average and marginal emissions implying the importance of variable interactions. The marginal emissions have lower errors compared to the average emissions, suggesting that the marginal value is easier to predict – it is less influenced by highly uncertain (weather dependent) variables as already remarked.

The weighted average model M_WA is constructed by combining the models with the Softmax weights: the RMSE in the *test* set becomes 38.49 and 9.63 $\left[\frac{\text{kgCO\_2-eq}}{\text{MWh}}\right]$ for the average and marginal emissions, respectively.

4.6. The compound model

The pure M_*WA* model of the previous section – i.e., without the residual correction of Section 4.3 – was used for forecasts on each individual horizon $h = 1, 2, ..., 24$ hours (denoted M_*WA*1...). Trials showed that the ARIMA model performed on $h = 6$ hours corrects the residuals on all earlier horizons; however, individually **M**_WA1, 2

Table 2
Performance of the final best models for particular horizons on test set (measured in normalized RMSE).

	Average			Marginal	
	$\mathbf{M}_{WA1,2}$	$\mathbf{M}_{ARIMA6}$	$\mathbf{M}_{WA7-24}$	$\mathbf{M}_{ARIMA6}$	$\mathbf{M}_{WA7-24}$
Horizons [h]	1–2	3–6	7–24	1–6	7–24
NRMSE [–]	0.095–0.118	0.127–0.132	0.176–0.183	0.029–0.053	0.148–0.160

outperforms the **M**_ARIMA6 on their corresponding horizons for the average emissions. **M**_WA7 − 24 showed to outperform **M**_ARIMA24 for all corresponding horizons for both the average and marginal emissions. This is because individual models designed for specific horizons may perform better, and also ARIMA predictions converge towards the average as the prediction horizon increases, thus being less suitable for longer horizons. The models for different horizons with the corresponding NRMSEs for the average and marginal emissions are summarized in Table 2.

Note from the table that NRMSE during 7–24 h becomes almost stationary. The reason is that on longer horizons, current information, say, on power generation data (available through short-term forecasts, Appenidx A) affects the predictions and associated uncertainties much less than the available long-term information of e.g. weather. On shorter horizons, 0-6 h, RMSE depends on short-term data and gradually increases in time until reaching the stationary value. The features can be seen in Fig. 8.

The final compound model **M**_24 used for the 24-h forecast of the emissions is the sum of the corresponding triplets of Table 2. When the short-term forecast data is included in **M**_24, the RMSE are naturally smaller than when it is excluded (yellow vs. green line, Fig. 8). The improvement is by 31% (within the 0-6 h horizon, for which short-term forecasts are available).

5. Results and discussion

5.1. The interaction coefficients β^{IA}

Of the three base models, **M2** that includes the interaction terms performs the best (Table 1). The parameters and results of **M2** are discussed here for both the average and marginal emissions. The analysis also applies to the compound model, since the weights often favours **M2**. Co-linearity can affect the model coefficients and make the analysis of individual coefficients incorrect. From a forecasting perspective, this is not crucial, because the model only considers the summed coefficient of two correlated variables while the individual coefficients can be completely opposite (false negative/positive). However, a Variance Inflation Factor (VIF) test has been applied and confirmed no further co-linearity, using a maximum VIF of 10.

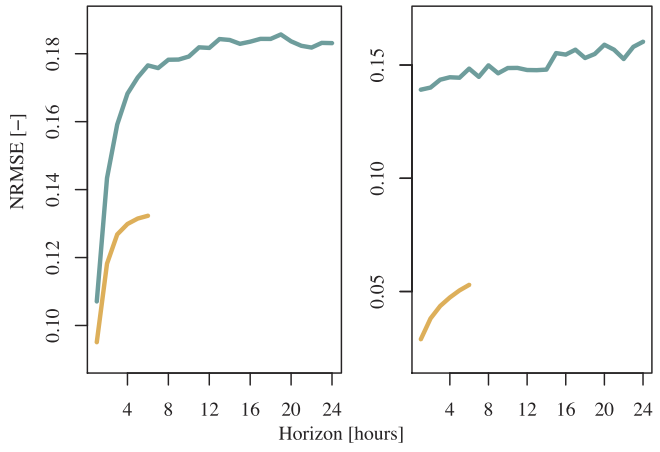

Fig. 8. Normalized RMSE (NRMSE) of CO_2 emission predictions vs. prediction horizon, using full data (orange line) and data w/o short term forecast (green line). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The five largest β coefficients of **M2** are featured in Table 3, both for the average and the marginal emissions. β^{IA}_i, β^{IA}_j and β^{IA}_i, j from the table refer to the first, second and third coefficient of the interaction vector β^{IA} in Eq. 9. The minus sign indicates opposing trends. To understand the table properly please consider the following for interaction terms:

$$\beta^{IA}\_i\mathbf{z}\_i + \beta^{IA}\_j\mathbf{z}\_j + \beta^{IA}\_i, j\mathbf{z}\_i, j \tag{15}$$

$$= (\beta^{IA}\_i + \beta^{IA}\_i, j\mathbf{z}\_j)\cdot\mathbf{z}\_i + \beta^{IA}\_j\mathbf{z}\_j \tag{16}$$

$$= (\beta^{IA}\_j + \beta^{IA}\_i, j\mathbf{z}\_i)\cdot\mathbf{z}\_j + \beta^{IA}\_i\mathbf{z}\_i \tag{17}$$

This means the final coefficient for e.g. $\mathbf{z}\_i$ becomes $(\beta^{IA}\_i + \beta^{IA}\_i, j\mathbf{z}\_j)$. The column for β^{IA}_i is thus the coefficient for $\mathbf{z}\_i$ if $\mathbf{z}\_j$ is zero and vice versa. The column for β^{IA}_i, j is the linear coefficient that explains the change in the final coefficient for both $\mathbf{z}\_i$ and $\mathbf{z}\_j$.

In the **average** emission model, the power generation in DK2 has the largest coefficient, 57.26: more generation → higher emissions, as also concluded in Fig. 2. This generation is interacting with the daily periodicities: the midday spline (row one in Table 3, see Appendix B), and the daily pattern from Fig. 4 (row five). The negative interaction coefficients β^{IA}_i, j, −4.7 and −2.3, via Eq. 16, mean the impact from the power generation decays and reaches a minimum at around noon. This is because other factors start to influence the emissions more.

The midday spline is again interacting with the net export from DK1 to DE (row three) that has a negative coefficient −11.4; here, β^{IA}_i, j is positive, 3.8, and suggests the least impact during the midday period, concluded from Eq. 16.

The wind speed in DK2 (row four) has a negative coefficient: the average emissions will decrease as more wind power enters the grid. The wind speed is positively interacting with net export from DK2 to DE (4.2) hence the final wind speed coefficient becomes larger, approaching zero, because the wind power is being exported rather than being used to decrease the emissions in DK2.

The offshore wind power generation in DK2 (row five) also has a negative coefficient, −16.7, and is interacting with the net export from SW3 to SW4 – this is primarily a one way inter-connector explicitly exporting to SW4. The emissions in DK2 decrease as exports increase – this is expected because SW3 is the location of all Sweden's nuclear power [19], and that is exported further into DK2. β^{IA}_i, j is positive and the final wind power coefficient will approach zero as the export increase, Eq. 16: the nuclear power takes part in the overall emissions in DK2 making the wind power contribution account for relatively less.

In the **marginal** emission model, note that none of the listed variables describe the grid in DK2. All are external influences from neighboring bidding zones and include net exports in all interactions. They are created because the net exports often only impact the emissions in one trading direction. Recall from Fig. 3, only the exports from DK1 to DE have an impact on the emissions. In this model, the coefficient of that trading pair is positive (row six) fitting the exports. To force the coefficient closer to zero during import, it is negatively interacting with the wind speed in DE, $\beta^{IA}\_i, j = -1.9$. DK1 will import if DE has wind

Table 3
Strongest interactions in Model 2 for the average and marginal emissions. z, i and j correspond to the terms in Eq. 9. Note the original variables in Appenidx A use "net import" rather than "net export". These terms are switched here, to make it easier to relate to.

z_i	z_j	β^{IA}_i	β^{IA}_j	β^{IA}_i, j
Average				
Power generation in DK2	Spline (midday)	57.3	1.7	−4.7
Power generation in DK2	Daily pattern	57.3	3.3	−2.3
Net export from DK1 to DE	Spline (midday).	−11.4	1.7	3.8
Wind speed in DK2	Net export from DK2 to DE	−11.8	−8.2	4.2
Offshore wind in DK2	Net export from SW3 to SW4	−16.7	−14.1	2.6
Marginal				
Net export to DE from DK1 - exp	Wind speed in DE	10.9	−6.1	−1.9
solar radiation in DE	Net export from DK1 to NO2 - exp	−1.7	−1.3	1.8
solar radiation in DE	Net export from DK1 to SW3 - exp	−1.7	−3.7	1.4
Net export from DK1 to SW3	Net export from SW3 to SW4 - exp	−8.0	0.54	−1.5
Net export from SW3 to SW4	Demand in SW4	11.4	2.7	−0.9

power to offer and the final coefficient for the net export decreases.

The solar radiation in DE is both interacting with net exports from DK1 to NO2 and net exports DK1 to SW3 ($\beta^{IA}\_i, j = 1.8$ and 1.4 respectively). The two interactions explain the same phenomenon: the emissions decrease as DK1 exports to SW3 and NO2 because $\beta\_j = -1.3$ is negative. However, when solar radiation in DE is high DK1 will again import from DE and the final coefficients for the net exports increase approaching zero because of the positive interaction coefficients.

The net export from DK1 to SW3 (row 9) is interacting with the net export from SW3 to SW4 (exponential term) too with $\beta^{IA}\_i, j = -1.5$: export from SW3 to SW4 decreases the impact from the net export from DK1 to SW3. That is because the marginal emissions in SW3 increase as their total export increases. The net export from SW3 to SW4 is the most significant variable and is negatively interacting with the demand in SW4 $\beta^{IA}\_i, j = -0.9$: the final coefficient for the net import in SW4 from SW3 decreases as the demand increases, most likely because in this case power is consumed in SW4 rather than exported further into DK2.

5.2. CO_2 emissions: the forecast of M_24

A few examples for the average and marginal emission forecasts by model **M**_24 are shown in Fig. 9 and Fig. 10 to illustrate the performance, where a 24-h horizon forecast is released at midnight. Note, the dates in the plots for the average and marginal emissions are identical.

Comparing the plots (**average** emissions) to the daily pattern in Fig. 4, plots B and C fit best with the highest emissions during the day. Plot B peaks in the morning slightly higher than the predictions and in plot C there is a lower than expected decrease observed in the evening. To a certain degree, daily patterns are often expected, so when the real observations differ too much, the accuracy decreases: in plot D, the emissions had a downward going trend all day, but it was expected to peak at around noon and then decrease. Plot A illustrates a day with irregularities too where the trend is captured to an adequate degree.

The **marginal** emissions have a much slimmer confidence interval than the average emissions due to the higher accuracy. Plot F differs the least from the average daily pattern and the prediction shows this too. The predictions in Plot G had a low accuracy because of irregular and small spikes. Plots E and H are good examples for a control mechanism: in plot A, the emissions are expected to increase in the evening, so it is encouraged to schedule flexible demand as early as possible before the emissions increase. In plot D, the opposite is seen: here, the demand should be shifted to the late evening where the emissions decreased.

The underlying relationship between the explanatory variables and the emissions are only focused to a limited analytical extend. Since the point of the paper is automated selection requiring no further knowledge about the regional grid dynamics for the researcher.

There are no studies with comparable results, thus this is a novel methodology from which future studies can extract inspiration.

5.3. Method discussion

The biggest challenge is to downsize the data to avoid overfitting and multi co-linearity, thus the methods to do so require considerate decisions. Feature selection exists in many forms, e.g. stepwise regression (forward and backward variable selection), generic algorithms and information theory (GA&IT)[25]. These methods are tested and discussed in [26], concluding that stepwise regression is the best option, hence also used in this paper. Backward selection can be a good choice when dealing with relatively few variables as proposed in [7],but due to its large computation complexity and little effect on the studied data, it is not considered in this paper. Another approach is the Variance Inflation Factor (VIF), a variant of stepwise regression used to detect multi co-linearity. However, VIF becomes computationally infeasible with too many explanatory variables due to R^2 approaching 1. Therefore, VIF is only used to confirm the final models are without co-linearity.

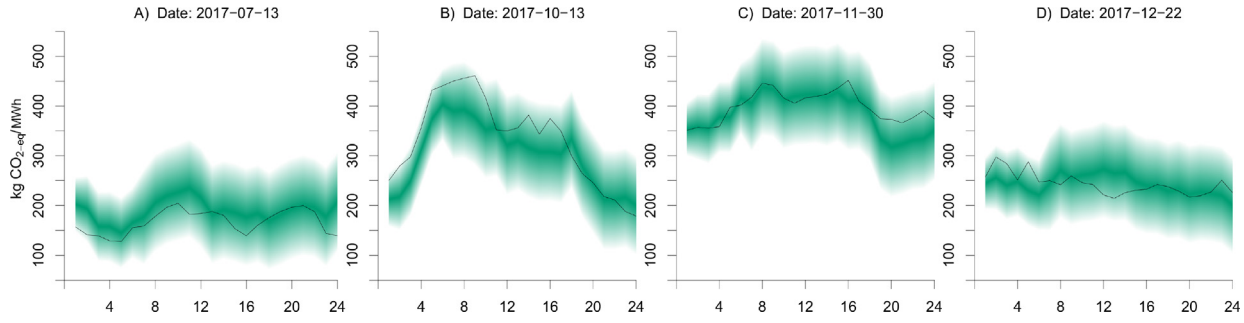

Fig. 9. Average CO_2 emissions: The final forecast submitted at midnight with a 24 h horizon for 8 different days. The real CO_2 emission intensity is the thick line, and the colored areas are bounding the 95% confidence interval and the point prediction.

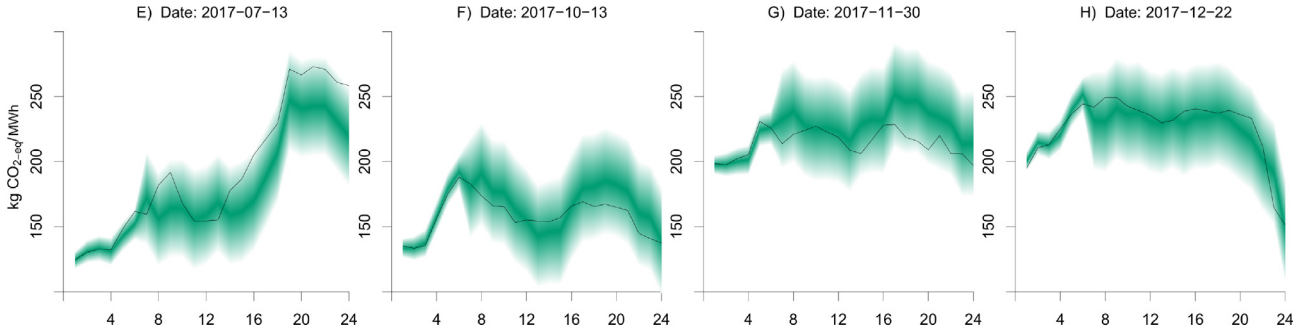

Fig. 10. Marginal CO_2 emissions: The final forecast submitted at midnight with a 24 h horizon for 8 different days. The real CO_2 emission intensity is the thick line, and the colored areas are bounding the 95% confidence interval and the point prediction.

5.4. Research limitations and future work

A forecast model is only as good as the provided data, but fortunately, both the quality and quantity of available data is increasing, raising the potential for future studies. Furthermore, the majority of the considered variables mainly explain the generation side. The variables explaining the demand side, besides weather conditions, could be interesting, but their use requires a better understanding of consumer behavior that could be probed via e.g. TV guides and big event schedules.

6. Conclusion

A novel machine learning methodology has been developed for forecasting the average and marginal CO_2 emissions in any European bidding zone.

The company Tomorrow (www.tmrow.com) supplied 473 variables (weather and electrical grid data), as well as the estimates of the marginal and average emissions.

Systematic downsizing, i.e. the selection of the 30 salient variables for forecasting, was done by forward feature selection in combination with LASSO regression. The approach prevented overfitting and co-linearity. Linear regression models that account for periodic trends, non-linearities and pairwise interactions in selected data sets were combined into a single model with a Softmax averaging. Finally, a residual correction was performed using an ARIMA model.

The study aims to provide forecasts that can help electricity consumers schedule their load to minimize CO_2 emissions. The forecasts of emissions are up to 24 h ahead and can thus be used as a basis for load scheduling. The average and marginal emissions were found to follow different temporal patterns: during a day, the average emissions peak in the afternoon (following roughly the power generation trend) and the marginal emissions peak in the morning and evening (following the consumption i.e. electricity prices). The opposing trends can be exploited for different applications. Since only the marginal CO_2 emissions are valid for small changes in demand, they are the signal which must be used when scheduling home appliances, such as heat pumps. The average emissions are useful for evaluating total electricity system emissions, but should not be used as a control signal.

The analysis of pairwise interactions between the selected explanatory variables gives insights into the interplay in the Nordic European electricity grid. For the average emissions, local power generation is pronounced, however, the least around midday as other factors come more into play. Interestingly, none of the most important variables related to the marginal emissions in DK2 were local (DK2) variables – all contributions came from neighboring bidding zones (DK1, DE, SW3 (indirect) and SW4). This suggests that the marginal generator is effectively supplied from the import.

To evaluate the usefulness of the marginal emission forecast, testing on various flexible applications, such as heat pumps and electric cars, should be conducted in the future. Results from this can indicate whether there is a need for model improvements. Further studies are needed to incorporate the weather dependent generators as marginals to fully understand the concept.

CRediT authorship contribution statement

Kenneth Leerbeck: Methodology, Software, Validation, Writing - original draft, Visualization, Formal analysis, Data curation. **Peder Bacher:** Conceptualization, Resources, Writing - review & editing, Supervision, Funding acquisition, Project administration. **Rune Grønborg Junker:** Conceptualization, Resources, Writing - review & editing, Supervision, Conceptualization, Resources, Writing - review & editing, Supervision. **Goran Goranović:** Writing - review & editing, Supervision. **Olivier Corradi:** Data curation. **Razgar Ebrahimy:** Software, Writing - review & editing. **Anna Tveit:** Writing - review & editing. **Henrik Madsen:** Resources, Writing - review & editing, Funding acquisition, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement & Funding

We are thankful for Tomorrow (www.tmrow.com) who has provided the data used in this study (including emission calculations for the bidding zone DK2). The work is supported through the project “Smart Cities Accelerator 2016–2020” funded by the EU program Interreg resund-Kattegat-Skagerrak, the European Regional Development Fond and the CITIES project (DSF1305-00027B).

Appendix A. Explanatory variables

Here, all explanatory variables are listed by data set for bidding zone DK2. For each variable in the data sets, it is indicated whether the variable is used to create models for $h \leqslant 6$ hours or $h > 6$ hours.

Short Term Forecasts	$h \leqslant 6$ hours	$h > 6$ hours
dewpoint	X	

precipitation	X
solar	X
temperature	X
price	X
generation	X
consumption	X
wind_speed	X
wind_direction_x	X
wind_direction_y	X
power_net_import_DK-DK1	X
power_net_import_DE	X
power_net_import_SE-SE4	X
power_net_import_SE	X

6 hours Forecasts (updated every sixth hour) provided by Tmrow IVS. Imports are originally coming from ENTSO-E who collect the data from individual Transmission System Operators (TSO's).

Weather Forecasts		
	$h \leqslant 6$ hours	$h > 6$ hours
dewpoint_mean_value		X
precipitation_mean_value		X
solar_mean_value		X
temperature_mean_value		X
wind_mean_value		X

52 hours Forecasts (updated every sixth hour) provided by Tomorrow. Originally created by GFS - Global Forecasting System

Market Data (Nordpool)		
	$h \leqslant 6$ hours	$h > 6$ hours
solar_power	X	X
wind_power_offshore	X	X
wind_power_onshore	X	X
generation		X
consumption		X
spot_price		X

Published at 6pm CET covering the whole next day. Reported in at 12pm CET, and technically available at that time.

Real Time Data		
	$h \leqslant 6$ hours	$h > 6$ hours
carbon_intensity	X	X
carbon_intensity_generation	X	X
carbon_intensity_import	X	X
carbon_rate	X	X
total_generation	X	X
total_storage	X	X
total_discharge	X	X
total_import	X	X
total_export	X	X
total_consumption	X	X
power_origin_%_fossil	X	X
power_origin_%_renewable	X	X
power_generation_biomass	X	X
power_generation_coal	X	X
power_generation_gas	X	X
power_generation_hydro	X	X
power_generation_nuclear	X	X
power_generation_oil	X	X
power_generation_solar	X	X
power_generation_wind	X	X
power_generation_geo	X	X
power_generation_unknown	X	X
power_origin_%_biomass	X	X
power_origin_%_coal	X	X
power_origin_%_gas	X	X
power_origin_%_hydro	X	X
power_origin_%_nuclear	X	X
power_origin_%_oil	X	X
power_origin_%_solar	X	X
power_origin_%_wind	X	X
power_origin_%_geo	X	X
power_origin_%_unknown	X	X
power_origin_%_hydro	X	X
carbon_origin_%_biomass	X	X
carbon_origin_%_coal	X	X
carbon_origin_%_gas	X	X
carbon_origin_%_hydro	X	X

carbon_origin_%_nuclear	X	X
carbon_origin_%_oil	X	X
carbon_origin_%_solar	X	X
carbon_origin_%_wind	X	X
carbon_origin_%_geo	X	X
carbon_origin_%_unknown	X	X
carbon_origin_%_hydro	X	X
power_net_discharge_hydro	X	X
power_net_import_DK-DK1	X	X
power_net_import_DE	X	X
power_net_import_SE-SE4	X	X
power_net_import_SE	X	X

Provided by Tomorrow and originally created by GFS.

Appendix B. Periodic time variables

Time variable matrix τ is defined as:

$$\tau\_\text{hour} = t \bmod 24$$
$$\tau\_\text{w} = \text{weekday}(t)$$
$$\tau\_\text{m} = \text{month}(t)$$
$$\tau\_\text{sin,hour} = \sin(\tau\_\text{hour})$$
$$\tau\_\text{sin,w} = \sin(\tau\_\text{w})$$
$$\tau\_\text{sin,m} = \sin(\tau\_\text{m})$$
$$\tau\_bs, \text{hour}, t = [bs\_0(\tau\_\text{hour}, t) \;\; bs\_1(\tau\_\text{hour}, t) \;\; \ldots \;\; bs\_n, \text{df} - 1(\tau\_\text{hour}, t)]$$
$$\tau\_bs, \text{w}, t = [bs\_0(\tau\_\text{w}, t) \;\; bs\_1(\tau\_\text{w}, t) \;\; \ldots \;\; bs\_n, \text{df} - 1(\tau\_\text{w}, t)]$$
$$\tau\_bs, \text{m}, t = [bs\_0(\tau\_\text{m}, t) \;\; bs\_1(\tau\_\text{m}, t) \;\; \ldots \;\; bs\_n, \text{df} - 1(\tau\_\text{m}, t)],$$

where hour, w and m denote the hour, weekday and month of the datetime t, respectively. τ_bs. hour, τ_bs, w and τ_bs, m each represent five columns corresponding to their underlying splines. n_df is the number of splines which is set to 5 in this case. The periodic splines are illustrated in Fig. B.11.

Appendix C. Forward feature selection

The forward selection algorithm selects the best variables for a model and requires a good cross validation strategy to avoid overfitting.

Algorithm 1. Forward feature selection

1. 1) Find the variable that best describes the response variable. This can be done with any best fit criteria (BIC, AIC or RMSE). This study relies on the RMSE value calculated on the validation sets of Section 4.1. Call this $\text{Model}_{\text{best}}$.
2. 2) Add a new variable $\mathbf{x}\_i$

$$\mathbf{y}\_t + h = \beta\_0 + \beta\_1 \mathbf{x}\_best + \beta\_2 \mathbf{x}\_i. \quad \text{(C. 1)}$$

Call this $\text{Model}_{\text{new}}$.

3. 3) Evaluate the model. If $\text{Model}_{\text{new}}$ is better than $\text{Model}_{\text{best}}$, keep the newly added variable and update: Model_best = Model_new.
4. 4) Repeat step 2 and 3 until all variables have been tested.

Appendix D. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.apenergy.2020.115527.

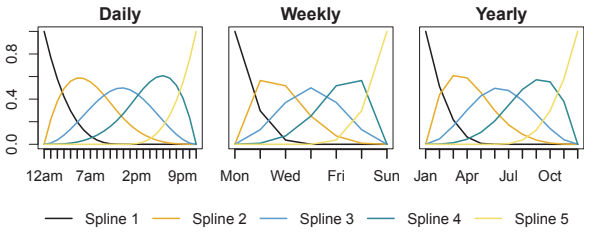

Fig. B.11. Periodic splines – Daily, weekly and yearly. Spine 3 in Daily is referred to as the midday spline. The outer splines (1 and 5) are not included in the model because they can be misleading.

References

[1] Global co2 emissions by sector, IEA, Paris, [Online; accessed 6-Nov-2019]; 2017. https://www.iea.org/data-and-statistics/charts/global-co2-emissions-by-sector-2017.

[2] Status of power system transformation 2019, IEA, Paris, [Online; accessed 6-Nov-2019]; 2019. https://www.iea.org/reports/status-of-power-system-transformation-2019.

[3] Liang Z, Liang J, Wang C, Dong X, Miao X. Short-term wind power combined forecasting based on error forecast correction. Appl Energy 2016;119:215–26.

[4] Wang J, Niu T, Lu H, Guo Z, Yang W, Du P. An analysis-forecast system for uncertainty modeling of wind speed: a case study of large-scale wind farms. Appl Energy 2018;211:492–512.

[5] Bacher P, Madsen H, Nielsen HA. An analog ensemble for short-term probabilistic solar power forecast. Sol Energy 2009;10:1772–83.

[6] Alessandrini S, Monache LD, Sperati S, Cervone G. An analog ensemble for short-term probabilistic solar power forecast. Appl Energy 2015;157:95–110.

[7] Keles D, Scelle J, Paraschiv F, Fichtner W. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. Appl Energy 2016;162:218–30.

[8] Yang Z, Li Lian LC. Electricity price forecasting by a hybrid model, combining wavelet transform, arma and kernel-based extreme learning machine methods. Appl Energy 2017;190:291–305.

[9] Wagner U, Mauch W, von Roon S. Das merit-order-dilemma der emissionen, Tech. rep., Forschungsstelle für Energiewirtschaft e.V; 2002.

[10] Regett A, Baing F, Conrad J. Emission assessment of electricity: Mix vs. marginal power plant method. 15th International Conference on the European Energy Market (EEM). 2018.

[11] Clauß J, Stinner S, Solli C, Lindberg KB, Madsen H, Georges L. Evaluation method for the hourly average co2-eq intensity of the electricity mix and its application to the demand response of residential heating. Energies 2019; 12.

[12] Bettle R, Pout C, Hitchin E. Interactions between electricity-saving measures and carbon emissions from power generation in england and wales. Energy Policy 2006;34:3434–46.

[13] Hawkes A. Estimating marginal co2 emissions rates for national electricity systems. Energy Policy 2010;38:5977–87.

[14] Corradi O. Tomorrow; 2018. https://medium.com/electricitymap/using-machine-learning-to-estimate-the-hourly-marginal-carbon-intensity-of-electricity-49eade43b421.

[15] Bialek J. Tracing the flow of electricity. vol. 143. IEEE; 1996.

[16] Kirschen D, Allan R, Strbac G. Contributions of individual generators to loads and flows. Trans Power Syst 1997;12:52–60.

[17] Tomorrow; 2018. https://www.electricitymap.org/ [accessed 1-Nov-2019].

[18] Heydari A, Garcia DA, Keynia F, Bisegna F, Santoli LD. Renewable energies generation and carbon dioxide emission forecasting in microgrids and national grids using grnn-gwo methodology. Appl Energy 2019;159:154–9.

[19] Elområde 1-4 (sn1-4) - statistik per månad 2017), Svenska Kraftnät; 2017. https://www.svk.se/aktorsportalen/elmarknad/kraftsystemdata/elstatistik/ [accessed 1-Nov-2019].

[20] Trevor Hastie RT, Friedman J. The elements of statistical learning. Springer series in statistics; 2017.

[21] Porter A, West S. Multiple regression: testing and interpreting interactions. J Roy Stat Soc: Ser D (Stat) 1994;43:453.

[22] Hu MY, Zhang G, Jiang CX, Patuwo BE. A cross-validation analysis of neural network out-of-sample performance in exchange rate forecasting. Decis Sci 1999;30:197–215.

[23] Tibshirani R. Regression shrinkage and selection via the lasso: a retrospective. J Roy Stat Soc Ser B (Stat Methodol) 2011;73:273–82.

[24] Madsen H. Time series analysis. Chapman & Hall/CRC - Taylor & Francis Group; 2007.

[25] Ludwig O, Nunes U. Novel maximum-margin training algorithms for supervised neural networks. Trans Neural Netw 2010;21:972–84.

[26] Ceperic E, Zikovic S, Ceperic V. Short-term forecasting of natural gas prices using machine learning and feature selection algorithms. Appl Energy 2017;140:893–900.