

SARIMAX as a Reference Model for Emission Intensity Forecasting for HPC Systems

Michael Zent*, Daniel Lübbert[†], Florian Burger[†] and Alexander Kammeyer*
0009-0002-0964-557X, 0000-0003-3852-5665, 0000-0003-4745-5515 and 0000-0002-7858-0354

* *Freie Universität Berlin*, Germany, {m.zent, a.kammeyer}@fu-berlin.de

[†] *Physikalisch-Technische Bundesanstalt*, Berlin, Germany, {daniel.luebbert, florian.burger}@ptb.de

Abstract—The increasing appliance of systems imitating human cognition requires reasonable, in particular transparent and resource-efficient, reference models.

This work demonstrates a course of action in the field of short-term greenhouse gas emission intensity forecasting with statistical Time Series Analysis to plan HPC-system operations in advance. For that, SARIMAX modelling in combination with methods from econometrics and computer science has shown to be qualified for benchmarking by its step-by-step traceability. It also provided high-accuracy rolling day-ahead prognoses on the hourly emission intensity of the German electricity production over a span of half a year. The NMAE of 10.9% and MASE of 73.9% when predicting their absolute values, and NMAE of 3.4% and MASE of 61.7% for the corresponding hourly changes, prove the performance.

Such an elaborated model, and the applied measures, can be used as a reasonable reference base to benchmark various methods in the area of time series forecasting, whether statistical ones or also those imitating human cognition.

Index Terms—Time Series Forecast, Reasonable Reference Model, SARIMAX, HPC, Emission Intensity Forecast, Machine Learning

I. INTRODUCTION

AUTOMATED systems imitating human cognition, colloquially named *Artificial Intelligence* (AI), are used in rapidly increasing scale in almost all areas of economy, science and administration [1]. Besides benefits, their usage not rarely causes overblown energetic, monetary, and environmental costs, including greenhouse gas emissions [2]. Moreover, one has to expect quality and security incidents [3], and the threat of *Model Collapse* by feedback effect [4, 5]. To evaluate the legitimacy to apply techniques of whatever kind, and to preserve the quality, including performance of the solutions, reference methods are required which are considered to be validated standards in their field [6]. Indeed limited objective evidence is available regarding the performance of e.g. *Machine Learning* (ML) models relative to such a reference [7, 8].

The issue with ML approaches essentially is, that despite the relatively simple underlying mathematical operations they nevertheless are computationally expensive as there are up to the trillions of model parameters, the training of which also has to be based on large-scale data sets with ideally no gaps. The resultant models are hard to interpret, because the way they came about is not traceable even for the most versed

experts. That makes the choice of hyperparameters difficult as there is no directly determinable connection to the fitted model parameters. Bearing in mind also the general nonlinear complexity of ML models, it is obvious that there is a need for benchmarking to decide whether the potential advantages, like the out of the box support for multivariate data and joint forecasts on them, justify the increased resource expenditures. See [8–10].

Classical techniques of statistics seem suited to serve for benchmarking due to the fact that all conducted operations including the results are deducible for humans, allowing to adjust the hyperparameters more systematically. The statistical *Autoregressive Integrated Moving Averages* (ARIMA) model family [11], which is continuously developed, well documented and resource-efficient, has been proven to be a reliable and flexible method in the field of Time Series Analysis, being able to handle even relatively small and fragmentary data sets. Over time several extensions have added support for weakly stationary, multivariate, and nonlinear but linearizable processes, widening their scope of application into areas associated with ML. General low algorithmic time complexity also allows frequent experiments to determine a robust baseline model to compare with.

In this study a reference short-term forecast model with high accuracy has been developed during the exploration of a potential usage of ML to forecast the greenhouse gas emission intensity time series data of the German energy production, based on the in many areas successfully employed and performant *Seasonal ARIMA with Exogenous Variables* (SARIMAX) model [11–14]. The predictions made shall contribute to a model of the resource consumption of HPC clusters to plan their operation in advance, inter alia to shift their peak energy demand to periods of low greenhouse gas emissions caused by electricity production, in compliance with metrics of energy usage effectiveness like the *Blue Angel Certificate* [15–17].

At the present time the number of related studies on short-term greenhouse gas emission intensity forecasting is limited. [18, 19] provide insights considering the current state of art, characterised by the usage of a variety of methods without an unified reference basis. Most studies focus on day-ahead forecasting, which corresponds to the daily cycle of bidding on the energy exchange market. Hence it was reasonable for the work presented here to firstly focus

on that forecast horizon, also for the sake of comparability.

In the following, Section II outlines in detail the SARIMAX modelling procedure which was applied to the dataset introduced in Section III in a stepwise traceable manner, as Section IV demonstrates, till the aspired reference forecasting model has been found.

II. METHODOLOGY

To describe the seasonal behavior of the CO_2 Equivalent (CO_2eq) [20] emission intensity data we used the SARIMAX modelling technique. A model shall be denoted with $SARIMAX(p, d, q)(P, D, Q)_s(Y|\vec{X})$, where the first triple of parameters refers to nonseasonal characteristics, and the next triple to the seasonal ones at seasonality s . The time series Y shall be modelled while considering the influence of exogenous variables, represented by a vector of time series $\vec{X}$. The sample value of Y at time t is notated with $y_t = Y(t)$, and the changes in the series between consecutive time steps with Δy_t . Thus the series of step-wise value changes is ΔY .

A. Determining Seasonality

To identify the time series' principal period s , which characterises the model's *Seasonal Component* (S), one firstly plots the series and tries to find patterns. For this purpose one groups the data seasonally and inspects their variation over time via boxplots for Y as well as ΔY . Then a visual inspection can yield several candidates.

To evaluate the first assumptions made, the *Periodogram* representations of the *Discrete Fourier Transforms* (DFT) of Y and ΔY are investigated [21, 22], keeping only those periods which have relevant spikes in both. These are then ordered by their *Squared Magnitude Spectrum*, and the strongest one is picked for s . Further periods can be noted as candidates for potential exogenous variables.

B. Ensuring Stationarity

To employ any kind of ARIMA, it is required that the underlying stochastic process is stationary, i.e. its mean, variance and autocorrelation are time-invariant and only fluctuate around a long-term constant value. If one has to assume that the series is strictly nonstationary because of indications for the existence of *Unit Roots* [23], then an ARIMA model can not be applied in principle. Its properties would have to change over time like the series, and therefore one could not derive future values as a consistent function of past ones. However, when weak stationarity is detected, then a stationary state can be approximated well by differencing the series [24]. A *trend-stationary* series has a deterministic long-term trend to which it always reverts to after random shocks. Differencing stationarizes the series by removing the trend. A *difference-stationary* series has a trend of stochastic nature and will not recover from shocks. Here differencing removes the permanent effects caused by these jumps in the series' mean level.

To look for nonseasonal stationarity one uses a combination of the following *Null Hypothesis Significance Tests* (NHST).

Table I: Interpretation of combined nonseasonal (a) and seasonal (b) stationarity tests.

		(a)	
		KPSS	
ADF/PP	Claim	rejected	unrejected
	rejected	Difference-Stationarity: $d = d + 1$, repeat tests	Stationarity: return d , done
	unrejected	Nonstationarity: ARIMA unapplicable	Trend-Stationarity: $d = d + 1$, repeat tests
		(b)	
		CH	
OCSB	Claim	rejected	unrejected
	rejected	$D = D + 1$, repeat tests	return D , done
	unrejected	$D = D + 1$, repeat tests	$D = D + 1$, repeat tests

- *Augmented Dickey-Fuller* (ADF) and *Phillips-Peron* (PP) claim that the series has a unit root. Both have their advantages in regard to stability in certain situations, so they can be used to double-check each other [25].
- *Kwiatkowski-Phillips-Schmidt-Shin* (KPSS) [26] claims that the series is trend-stationary. The alternative hypothesis is the presence of a unit root.

ADF and PP should yield the same conclusion. If not, one either faces a too small data set, so that the tests are not meaningful enough, or the series might contain *Structural Breaks*. Those occur when one or more of the statistical properties which constitute stationarity change abruptly at relatable *Breakdates* in the series, e.g. the outbreak of a pandemic or war, and do not return to the previous characteristics within a short term. Not considering breaks affects a model's reliability, and may cause large forecasting errors when one for instance ignores that an upward trend has changed into a downward one. A common approach is to split the series at the breakdates and to continue the analysis on the subsamples. An outline on how to identify breaks is given in [27].

ADF/PP and KPSS on the other hand can disagree. The combined results are to be interpreted as in Table I(a). The order of nonseasonal differencing d is the number of applied differencings until both tests let assume strict stationarity.

For checking on the need of seasonal differencing the following NHSTs are used.

- *Osborn-Chui-Smith-Birchenhall* (OCSB) [28] claims a seasonal unit root, analogous to ADF/PP.
- *Canova-Hansen* (CH) [29] claims seasonal stability, analogous to KPSS.

The results are to be interpreted as in Table I(b). The order of seasonal differencing D is the number of applied differencings until both tests do not indicate the need for further seasonal differencing. More than one order of seasonal differencing, or more than two orders of total differencing should not be used because of the risk of overdifferencing, leading to a degenerated model [30].

Such, the model's *Integrated Component* (I) is determined by the differencing orders, i.e. taking the difference between the present data value y_t and a past value. For nonseasonal differencing of order d consecutive timesteps are used as seen in (1). At seasonal differencing of order D the difference to the value of s timesteps ago is taken, as shown in (2).

$$\Delta^d y_t = y_t^{(d)} = y_t^{(d-1)} - y_{t-1}^{(d-1)} \quad (1)$$

$$\Delta_s^D y_t = y_t^{(D;s)} = y_t^{(D-1;s)} - y_{t-s}^{(D-1;s)} \quad (2)$$

$$\Delta_s^{d,D} y_t = \Delta_s^D \Delta^d y_t \quad (3)$$

Both kinds of differencing are executed in a row as in (3). The resulting values $\Delta_s^{d,D} y_t$ now represent a stationarized time series $\Delta_s^{d,D} Y$. For the case $d = 1, D = 0$ we write Δy_t respectively ΔY . All further modelling steps pertain to the stationarized time series, ergo also any forecast will be made on $\Delta_s^{d,D} Y$, instead of Y . The future values of latter can be predicted by calculating back the applied orders of differencing from the forecast for $\Delta_s^{d,D} Y$.

C. Autoregression

The *Autoregressive Component* (AR) terms describe the regression of a stationarized process against itself.

A nonseasonal AR-term of order p expresses the linear dependency of the present value y_t on the past p values y_{t-i} , weighted by coefficients φ_i which have to be determined during model fitting. Analogously, a seasonal AR-term of order P relates y_t to P past values y_{t-si} , weighted by coefficients Φ_i . Considering also the integrated component, the overall *Seasonal Autoregressive Integrated* model $SARI(p, d)(P, D)_s$ is formulated with (4).

$$\Delta_s^{d,D} y_t = c + \varepsilon_t + \sum_{i=1}^p \varphi_i \Delta_s^{d,D} y_{t-i} + \sum_{i=1}^P \Phi_i \Delta_s^{d,D} y_{t-si} \quad (4)$$

The residual between the AR-terms and the actual value of $\Delta_s^{d,D} y_t$ consists of some process-inherent constant c , and a variable error ε_t which should be a random shock, i.e. a single realization of white noise, if the model would describe the time series ideally.

D. Moving Average

The *Moving Average Component* (MA) terms describe the impact of random-shock errors on a stationarized process.

A nonseasonal MA-term of order q expresses the linear dependency of the present value y_t on the q past errors ε_{t-i} , which are assumed to be mutually independent and normally distributed like white noise. The magnitude of the impact is weighted by the coefficients ϑ_i which are determined during model fitting. Correspondingly, a seasonal MA-term of order Q relates y_t to P past errors ε_{t-si} , weighted by coefficients Θ_i . Considering also the integrated component, the overall *Seasonal Integrated Moving Average* model $SIMA(d, q)(D, Q)_s$ is formulated with (5).

$$\Delta_s^{d,D} y_t = \mu + \varepsilon_t + \sum_{i=1}^q \vartheta_i \Delta_s^{d,D} \varepsilon_{t-i} + \sum_{i=1}^Q \Theta_i \Delta_s^{d,D} \varepsilon_{t-si} \quad (5)$$

Thereupon the residual between the MA-terms and the actual value of $\Delta_s^{d,D} y_t$ consists of the mean μ of $\Delta_s^{d,D} Y$, and a present random-shock error ε_t .

E. SARIMA

A *Seasonal Autoregressive Integrated Moving Average* (SARIMA) model is the combination of the SARI and SIMA models. The corresponding constants c and μ are usually summarized into a single constant, in (6) denoted as C within the $SARIMA(p, d, q)(P, D, Q)_s$ model.

$$\begin{aligned} \Delta_s^{d,D} y_t = C + \varepsilon_t &+ \sum_{i=1}^p \varphi_i \Delta_s^{d,D} y_{t-i} + \sum_{i=1}^P \Phi_i \Delta_s^{d,D} y_{t-si} \\ &+ \sum_{i=1}^q \vartheta_i \Delta_s^{d,D} \varepsilon_{t-i} + \sum_{i=1}^Q \Theta_i \Delta_s^{d,D} \varepsilon_{t-si} \end{aligned} \quad (6)$$

All coefficients are tested on the null hypothesis of being insignificantly different of zero. As recommended we will use the results with care when selecting model variables, and apply a higher p -value threshold for rough orientation, namely 50% instead of the common 5% [31].

F. Adding Exogenous Factors

Exogenous Factors (X) [32] are determined by forces outside the model, and can impact the regarded time series Y , such giving additional information to explain its behavior. The process behind Y either does not influence the exogenous factors, or that backward influence is not taken into account for the sake of model simplicity.

The $SARIMAX(p, d, q)(P, D, Q)_s(Y|\vec{X})$ model adds a linear combination of n exogenous variables X_i with by coefficients β_i weighted values $x_{i;t}$ as in (7). The X_i have to be differenced by the same orders as Y .

$$\begin{aligned} \Delta_s^{d,D} y_t = \text{SARIMA}((p, d, q)(P, D, Q)_s; y_t) \\ + \sum_{i=1}^n \beta_i \Delta_s^{d,D} x_{i;t} \end{aligned} \quad (7)$$

Candidates for exogenous variables are additional periodicities representing further seasonal influences, as well as variables describing processes with high correlation to the targeted series. To identify latter one selects another time series from the same causal context, assesses the degree of linear correlation using the *Pearson Coefficient* [33], and double checks for possible nonlinear relation via the *Distance Correlation* measure [34].

SARIMAX allows to use the exogenous data till time t to forecast the stationarized time series for $t + 1$, i.e. a single step ahead. To predict for larger time spans ahead one has to forecast the exogenous variables, too. Periodicities could be modelled either to follow a corresponding sinusoid, or to continue the course provided by a *Seasonal Decomposition* [35]. For an accompanying time series chosen by correlation a prediction till the forecast horizon has to be self-made or be provided from a third-party source.

G. Prediction Error Assessment

The difference between the values $\hat{y}_t$ which a model predicts for a process to have over time t , and its actually observed ones y_t can be quantified in various ways.

Error Metrics directly compare $\hat{y}_t$ against y_t without considering model complexity. The often used *Mean Absolute Error* (MAE) has been shown to be a natural measure of average error [36], but is only meaningful for a concrete data set as it is scale-dependent. Normalizing as in (8), with T being the total number of observations, turns it into NMAE, a scale-independent percentage error measure useful for comparing across data sets.

$$\text{NMAE} = \frac{\sum_{t=1}^T |\hat{y}_t - y_t|}{T(y_{\max} - y_{\min})} \quad (8)$$

However, the NMAE does not indicate how much improvement has been achieved compared to a naive baseline model. For this the *Mean Absolute Scaled Error* (MASE) has been introduced [37], defined in (9) as the MAE for the values predicted by the model in question, divided by the MAE of a *Naive Seasonal Forecast* model which claims that $y_t = y_{t-s}$ at seasonality s . The lower the MASE is, the better the model, with MASE = 1 for an equal performance.

$$\text{MASE} = \frac{\sum_t |\hat{y}_t - y_t|}{\sum_t |y_{t-s} - y_t|} \quad (9)$$

In this work it has been forgone to use the *Mean Absolute Percentage Error* (MAPE) for being biased in favour of systematic underestimation and having issues with data containing zero or close-to-zero values, and the *Root Mean Square Error* (RMSE) for being too sensible to disturbances in the data like outliers. Also the *Coefficient of Determination* (R^2) has been excluded for overlooking model bias and tending to indicate correlations in randomness. Such we followed the recommendations in [36, 38].

Information Criteria take into account the complexity of the model, i.e. for SARIMAX the number of AR- and MA-terms. While error metrics on the test set measure predictive accuracy, information criteria help to compare models based on their goodness-of-fit penalized by complexity. A model of a data-generating process will almost never be exact, therefore some information will be lost when using the model to represent the process. Information criteria estimate the relative amount of information lost by a given model. The less information is lost, the higher the model quality. In this work we use the *Akaike Information Criterion* (AIC) [39] as it is asymptotically optimal when selecting a model with the least mean squared error [40], and complementary the *Bayesian Information Criterion* (BIC) [41] for penalizing model complexity harder.

H. Residual Analysis

The *Residual* is the time series created by taking the difference between the observed series values and the predicted ones. The more the forecasting model has been able to catch of the information patterns in the underlying process, the more

the residual will resemble white noise, which would contain no information anymore. The residual should show the following characteristics.

No Autocorrelation, i.e. the residuals at different time steps should be independent from each other. Commonly the *Ljung-Box* (LB) [42] portmanteau test with the null hypothesis of no autocorrelation on certain lag-values is used for diagnosing models of the ARIMA family. The test statistics should be as low as possible and the hypothesis not rejected. Often this check is also done visually via autocorrelograms.

Homoscedasticity, i.e. constant variance over time. The corresponding null hypothesis is checked with a *Goldfeld-Quandt* (GQ) test on whether the sum-of-squares in the first third of the sample is not significantly different than the sum-of-squares in the last third of the sample. The value of the test statistic should be close to 1, and the hypothesis unrejected [43].

It is desirable, but not critical, that the residual is *Normally Distributed*, being an indicator for white noise. Hence, as close as possible should the mean and skewness be to 0, the standard deviation to 1, and the kurtosis to 3. The statistic of the corresponding *Jarque-Bera* (JB) test should be as small as possible, and the null hypothesis of a normal distribution not be rejected [44].

I. Finding Candidates for the AR- and MA-Term Orders

The naive approach to find p , P , q and Q would be to execute a grid-search over all combinations of these parameters within some arbitrary fixed range, measure the performance and residual quality of the corresponding models, and choose the one with the best results. As this would be computationally intensive, one should instead predetermine a set of parameter candidates. For that purpose we have to analyse the autocorrelative behavior of the stationarized time series.

The *Autocorrelation Function* (ACF) measures the Pearson correlation of the stationarized series $\Delta_s^{d,D} y_t$ with a by k time steps lagged version of itself $\Delta_s^{d,D} y_{t-k}$. It provides information about the order of the MA-terms in our model, as we will see below. We can not use it for the AR-terms though, because the ACF value for lag k is influenced by the correlated lagged values i in between, i.e. $0 < i < k$. The *Partial Autocorrelation Function* (PACF) is the result of removing these *Confounding Variables* and measure solely the impact of $\Delta_s^{d,D} y_{t-k}$ on $\Delta_s^{d,D} y_t$. See also [45].

The predetermination of the AR- and MA-term orders requires a visual examination of the ACF and PACF *Autocorrelogram* plots, which also include *Significance Bounds* along a 95% confidence interval of $\pm 2/\sqrt{N}$ along the time-axis, where N is the number of samples in the considered series [46].

Only spikes outside the significance bounds are regarded as significantly different from zero, and only their corresponding lag values are of interest. The ACF and PACF plots are then examined on whether they cut off after a certain lag value k , i.e. the (partial) autocorrelation is significant for the first k

Table II: Derivation of minimum values for the nonseasonal AR- and MA-term orders from the ACF and PACF correlograms' behavior.

ACF	Gradual decay	Cutoff after lag k	Gradual decay
PACF	Cutoff after lag k	Gradual decay	Gradual decay
Conclusion	$p \geq k, q = 0$	$p = 0, q \geq k$	$p \geq 1, q \geq 1$

lags and then drops abruptly to statistical insignificance. Otherwise they decay gradually, i.e. the (partial) autocorrelation diminishes as the lag increases.

Then the nonseasonal parameters p and q can be expected to be at least as in Table II by inspecting the subsequent lags k for $k \geq 0$. The same procedure holds for the seasonal parameters P and Q , however here the behavior of the seasonal multiples of lags k are considered, i.e. the lags at every $s \times k$, with s being the seasonality.

One should take into account, that the series has to be stationarized, nonseasonally as well as seasonally, by differencing according to the already determined parameters of differential order d and D . Typical for a nonstationary series is a very slow decay in the ACF plot.

Also real data are noisy, i.e. there might be spikes that are a bit outside the bounds just by pure chance. These may point at candidates for the corresponding parameters which may refine the model, but they can also just be caused by noise. The analyst has to subjectively assess whether it is plausible that such questionable spikes could be part of the process. This circumstance is an issue when trying to automatise the interpretation of autocorrelograms for noisy data.

If both an AR- and a MA-process are detected, i.e. $p, q \geq 1$ and/or $P, Q \geq 1$, then one can expect that the values of the corresponding pair of orders will rarely exceed 2 in the end.

The assumptions made can be checked against the results of *auto_arima* from the Python library *Pmdarima* [47], which applies an automatic parameter determination algorithm. However, it may come to other conclusions than the visual examination suggest. Then the model proposed by *auto_arima* should be added as a candidate model for further diagnosis.

J. Model Selection

Having determined candidate models by visual autocorrelogram analysis, and perhaps additionally *auto_arima*, one now has to choose the model which captures the patterns in the data best.

Firstly one has to assess how much autocorrelation the residual still contains. Relevant spikes in the autocorrelogram at nonseasonal and seasonal lags indicate that too much structure in the data has been left out.

The lower the NMAE, MASE and AIC values are, the better does the model perform. In connection with this one should avoid computationally expensive models of high complexity. Considering the BIC and the p -values for the model coefficients may help to reduce orders of the AR- and MA-terms.

The closer the characteristics of the residual time series are to a normal distribution, the better, even if normality is not

reached. A mean value which significantly differs from zero can suggest a bias in the model.

The closer the numerical result of the GQ test on homoscedasticity to 1 is, the better, even if the null hypothesis has been rejected. In that case however one might want to consider to adjust the original data e.g. by *Box-Cox Transformation* [48].

III. DATA

This work utilized the quarter-hourly German electricity production's CO₂eq emission intensity data in kgCO₂eq/MWh as provided by [16], which in its turn are based on the *Strommarktdaten* (SMARD) of the *Bundesnetzagentur* (BNetzA) [49] about the Germany-wide power generation, in total as well as per energy source. Also current and historical 24h-forecast data have been included, though available in less detail considering the involved energy sources.

Before being used here, the data have been aggregated to hourly data by taking the mean of four consecutive data points within the same hour.

IV. RESULTS

The hourly aggregated data for the years 2020 till 2024 with 51864 records are used to analyse the characteristics of the corresponding time series to determine the parameters of a SARIMA model that would describe it properly, and to find exogenous factors that would suit best as a complementary to the CO₂eq emission intensity data we are interested in, formulating the final SARIMAX model. The data for the first two quarters of 2025 with 4343 records serve for the validation of the predictions made.

In the following we will denote with Y the series of CO₂eq emission intensity, our target feature, and with ΔY its in first order differenced series. We are interested to forecast the absolute values, as well as their changes.

A. Naive Baseline

We will refer to the naive seasonal forecast model by applying the MASE prediction error measure on the models for Y and ΔY . Hence it is not considered necessary to explicitly assess the naive baseline as such.

Table III: Results of tests on nonseasonal stationarity of the CO₂eq emission intensity series Y , and its differenced series ΔY , for the years 2020-2024.

Test	Y		ΔY	
	p -Value	Result	p -Value	Result
ADF	2.4×10^{-15}	rejected	0.0	rejected
PP	1.9×10^{-20}	rejected	0.0	rejected
KPSS	0.9×10^{-3}	rejected	0.998	unrejected

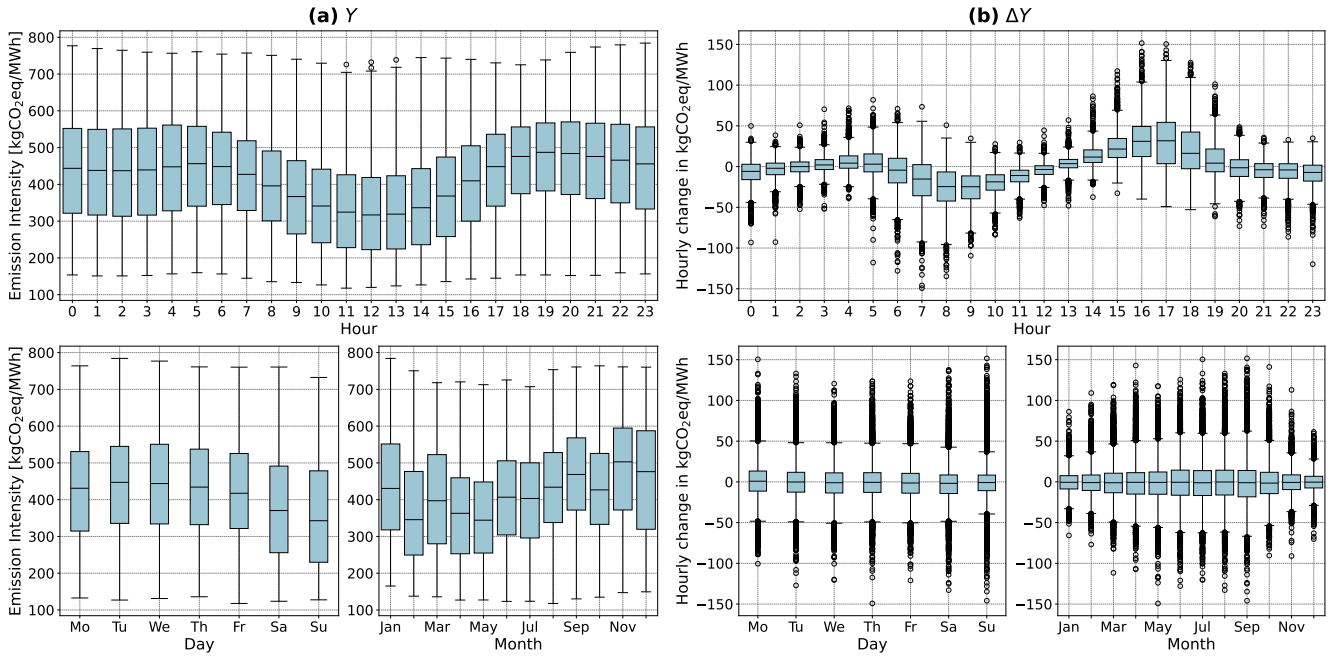


Figure 1: Daily (on top), weekly and monthly (below) variation of the per-hour CO₂eq emission intensity series Y for the German electricity production (a) of the years 2020-2024, as well as of its per-hour differenced series ΔY (b).

B. Seasonality and Stationarity

Firstly Y has to be checked on stationarity to verify that ARIMA-models are applicable. The ADF, PP and KPSS tests are employed and, as listed in Table III, all their p -values fall below the set significance level of 5%. As ADF and PP do not contradict each other in their indication of stationarity, while KPSS indicates nonstationarity, we conclude that the time series is difference-stationary. After differencing the series once, yielding ΔY , and applying all three tests again, all tests agree that ΔY is nonseasonally stationary.

Hence, one can indeed employ ARIMA models, and we set the nonseasonal differencing order of Y to $d_Y = 1$. The nonseasonal order of ΔY can therefore be set to $d_{\Delta Y} = 0$, as nonseasonal differencing has already been applied here.

A glance on Figure 1(a) lets assume that Y might have a daily, weekly and annual periodicity. As the data are hourly, that corresponds to seasonality values of 24, 168 and 8766. For ΔY one can suspect from Figure 1(b) a daily periodicity. The corresponding periodograms Figure 2 on Y and ΔY confirm this observation and designate the daily periodicity to be the principal one. As we have to difference the series by one order to reach stationarity, the weekly and annual periodicities can be put aside according to Figure 2(b), as they disappear after differencing and will therefore play no role in further SARIMAX modelling. On the other hand both periodograms designate also a semidaily periodicity which is even seemingly intensified in the differenced series, together with other smaller subdaily periodicities. However, consulting Figure 1(b) again,

no subdaily periodicity is recognizable. The amplification of the semidaily periodicity could be explained by a sign-loss during the spectral power calculation due to squaring, causing the valley in the first half of the differenced hourly timeline to be interpreted as a hill instead.

The parameter for the principal seasonality is hereby set to $s = 24$ for Y as well as ΔY . No further seasonalities are left to be considered as exogenous factors.

Having determined s , we check on the need for seasonal differencing. Using `nsdiffs()` from the `Pmdarima` library [50], OCSB indicates that there is no seasonal unit root, while according to the CH test one order of seasonal differencing is necessary to reach stability. After applying seasonal differencing the tests are repeated, and both indicate that seasonal stationarity is reached. Therefore, the order of seasonal differencing is set to $D = 1$ for Y and ΔY likewise.

C. Identifying Exogenous Factors

The results of the correlation analysis are depicted in Figure 3. The absolute Pearson Correlation values $|\rho|$ on the degree of linear correlation, and complementary the Distance Correlation values $dCor$ on nonlinear one, suggest to use the time series on the production of renewable energy as an exogenous factor X . The causal connection to the emission intensity series Y is obvious.

For renewable energy sources only a limited 24h-forecast from the BNetzA on the production at wind and solar plants is available, which represent about 75% of electricity production from low-emission sources. However, the correlation of the

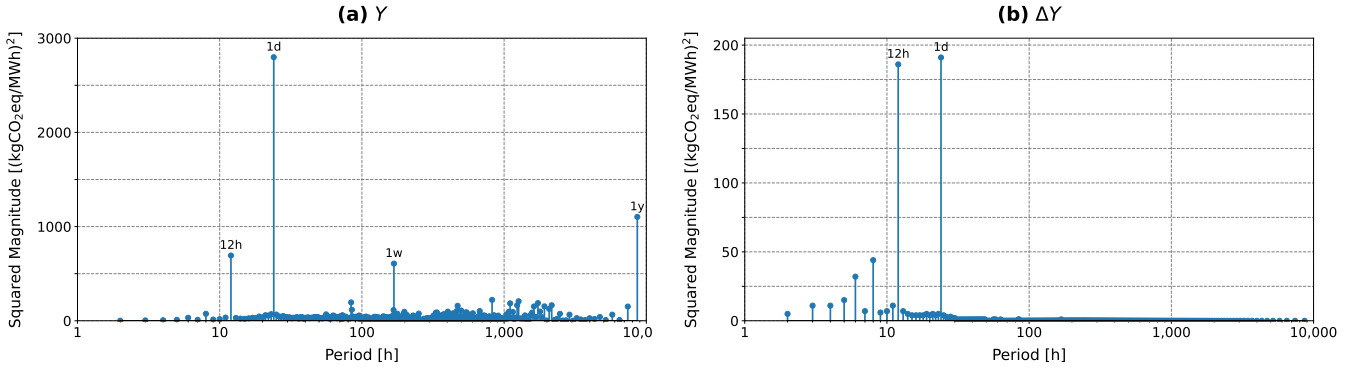


Figure 2: Periodogram of the per-hour CO₂eq emission intensity series Y (a), resp. the series of its hourly changes ΔY (b), for the German electricity production of the years 2020-2024.

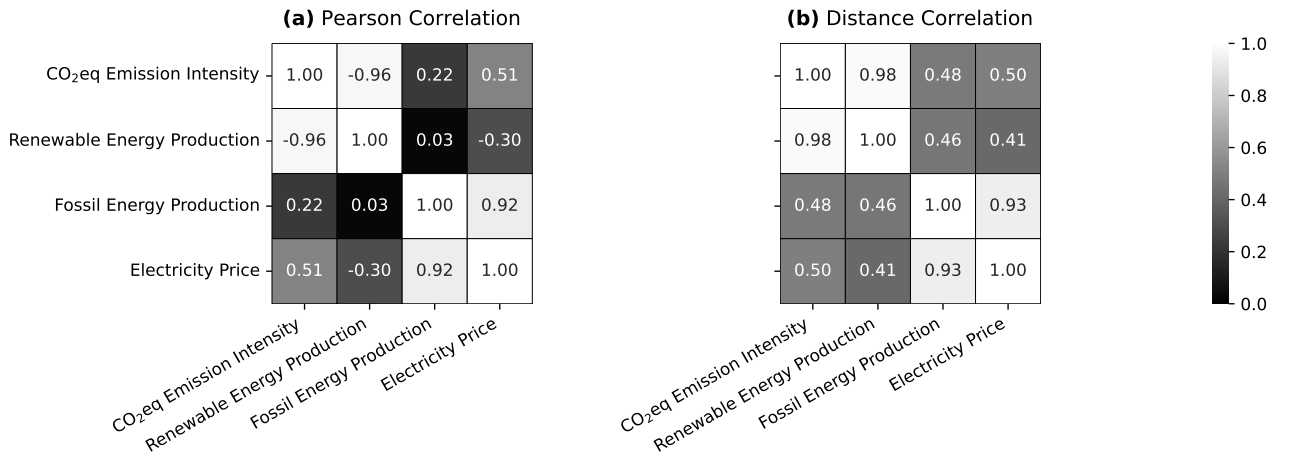


Figure 3: Heatmaps of the Pearson Correlation ρ and Distance Correlation $dCor$ between the features derived from the data set on German electricity production for the years 2020-2024 [16]. The grayscale mapping of the correlation values $|\rho|$ and $dCor$ is depicted at the right.

past forecasts for those to Y is the same as of the renewable energy production to Y .

Hence we choose to use the actual renewable energy production as an exogenous variable for model parametrization and fitting, denoted as X , and the 24h-forecast on wind and solar energy production as an exogenous variable for forecasting purposes, denoted as X_f . As the correlational connections are unaffected by differencing we analogously choose ΔX and ΔX_f as exogenous variables for fitting resp. forecasting ΔY .

D. Model Identification

Having found the primary seasonality $s = 24$, set the necessary orders of differencing to $d_Y = 1$, $d_{\Delta Y} = 0$ and $D = 1$, and decided on the exogenous factor X to be considered, now proper nonseasonal and seasonal orders p resp. P of the AR-terms, and q resp. Q of the MA-terms have to be found. We will focus on the year 2024, as annual events like the publishing of new emission intensity reports or new legislation acts can influence the pattern.

To select plausible candidates, we take the ACF and PACF of the final, i.e. seasonally and nonseasonally differenced, series $\Delta_{24}^{1,1}Y$, as seen in Figure 4. The ACF tails off gradually from the first lag on, while the PACF cuts off sharply after lag 2. That indicates a nonseasonal AR-process with $p = 2$ and $q = 0$. It may still be worth to check on $p = 3$, too, as the corresponding spike may still be significant. The spikes from lag 4 on are ambiguous and could be induced by noise. Further, the ACF shows seasonal spikes which cut off after lag $s = 24$, however the spikes at lag $2s$ could be significant enough to improve the model. The PACF shows seasonal spikes which tail off. So we can assume a seasonal MA-process with $P = 0$ and $Q = 1$, where it might be worth to also check on $Q = 2$. Hence, autocorrelation analysis suggests the model $\text{SARIMAX}(2, 1, 0)(0, 1, 1)_{24}(Y|X)$.

The *auto_arima* function from the Python package *Pmdarima* has also been employed in an attempt to automatically determine the model parameters. However, *auto_arima* rejects

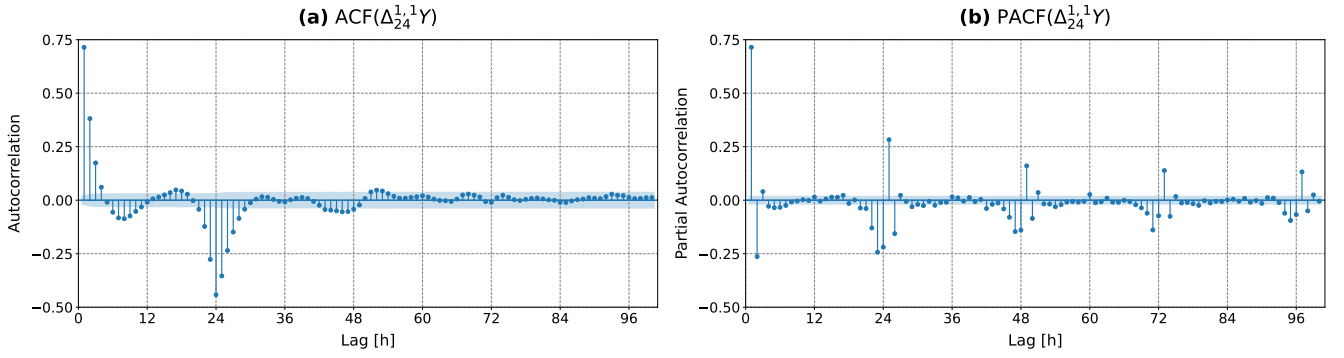


Figure 4: ACF (left) and PACF (right) of the fully stationarized CO_2eq emission intensity series $\Delta_{24}^{1,1}Y$. At lag 0 the autocorrelation always equals 1, hence that lag is omitted here.

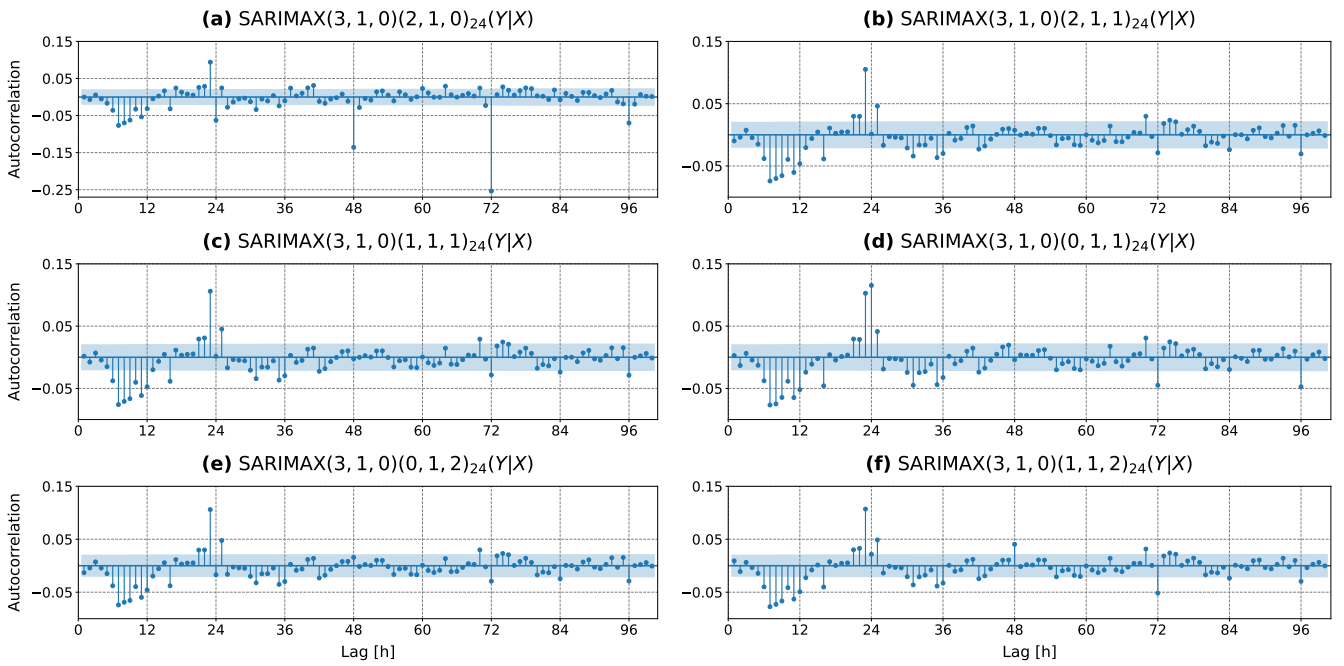


Figure 5: ACFs of the candidate model residuals. At lag 0 an ACF always equals 1, hence that lag is omitted.

the assumed model with a seasonal MA-process, and instead suggests a seasonal AR-process with $P = 2$ and $Q = 0$. Besides, it increases the nonseasonal AR-order even beyond lag 4 when orienting on the yielded AIC values. When using the BIC instead to penalize model complexity harder, *auto_arima* stops at $\text{SARIMAX}(3, 1, 0)(2, 1, 0)_{24}(Y|X)$. The such found seasonal AR-process seems to contradict the handmade autocorrelation analysis, but is added to the candidate list alongside the model with a seasonal MA-process.

E. Diagnostics and Selection

When a model has successfully captured the non-seasonal and seasonal processes, the residuals of the model should be uncorrelated. However, the ACF plot for $\text{SARIMAX}(3, 1, 0)(2, 1, 0)_{24}(Y|X)$, suggested by *auto-*

arima, still shows large spikes at the seasonal lags, as one can see in Figure 5(a). We can rule out this model and increment the seasonal MA-order Q according to our first findings. Correspondingly we then also try to reduce the seasonal AR-order P to zero.

Though the resulting $\text{SARIMAX}(3, 1, 0)(2, 1, 1)_{24}(Y|X)$ model solves the issue, as Figure 5(b) shows, it became computationally expensive as listed in Table V. The relatively high p -value for the second seasonal AR-term Φ_{48} in Table IV suggests to remove it, and indeed $\text{SARIMAX}(3, 1, 0)(1, 1, 1)_{24}(Y|X)$ is five times cheaper at almost the same results considering the autocorrelation according to Figure 5(c) and the NMAE value, though the GQ test values on homoscedasticity and the JB ones on normality

Table IV: Coefficients of the SARIMAX candidate models fitted on the total values of the CO₂eq emission intensity series for the year 2024, as well as their p -values resulting from the null hypothesis test on whether a coefficient is only insignificantly different from zero. The finally chosen model is highlighted.

SARIMAX	(3, 1, 0)(2, 1, 0) ₂₄		(3, 1, 0)(2, 1, 1) ₂₄		(3, 1, 0)(1, 1, 1) ₂₄		(3, 1, 0)(0, 1, 1) ₂₄		(3, 1, 0)(0, 1, 2) ₂₄		(3, 1, 0)(1, 1, 2) ₂₄		(2, 1, 0)(1, 1, 1) ₂₄	
Coefficient	Value	p-Val	Value	p-Val	Value	p-Val	Value	p-Val	Value	p-Val	Value	p-Val	Value	p-Val
Exogen. β	-0.0326	0.000	-0.0331	0.000	-0.0333	0.000	-0.0332	0.000	-0.0333	0.000	-0.0334	0.000	-0.0333	0.000
AR														
φ_1	0.7659	0.000	0.7891	0.000	0.7772	0.000	0.7928	0.000	0.7889	0.000	0.7705	0.000	0.7806	0.000
φ_2	-0.2520	0.000	-0.2568	0.000	-0.2442	0.000	-0.2491	0.000	-0.2564	0.000	-0.2351	0.000	-0.2450	0.000
φ_3	0.0053	0.615	0.0084	0.435	0.0026	0.811	0.0061	0.568	0.0072	0.501	-0.0008	0.938	–	–
Φ_{24}	-0.5499	0.000	0.1259	0.000	0.1304	0.000	–	–	–	–	-0.3658	0.000	0.1314	0.000
Φ_{48}	-0.2833	0.000	-0.0115	0.186	–	–	–	–	–	–	–	–	–	–
MA														
Θ_{24}	–	–	-1.0450	0.000	-0.9592	0.000	-0.9452	0.000	-0.8129	0.000	-0.4821	0.000	-0.9593	0.000
Θ_{48}	–	–	–	–	–	–	–	–	-0.1394	0.000	-0.4476	0.000	–	–

Table V: Diagnostics for the SARIMAX candidate models on CO₂eq emission intensity from fitting onto the data of 2024, along with the computational costs for the fitting process using the Powell method for noisy data [51], measured in CPU instructions using the Python library *Cirron* [52]. The LB autocorrelation test results on the lags 1 to 4 and 1s to 4s are not included due to lack of space. The finally chosen model is highlighted.

SARIMAX	(3, 1, 0)(2, 1, 0) ₂₄	(3, 1, 0)(2, 1, 1) ₂₄	(3, 1, 0)(1, 1, 1) ₂₄	(3, 1, 0)(0, 1, 1) ₂₄	(3, 1, 0)(0, 1, 2) ₂₄	(3, 1, 0)(1, 1, 2) ₂₄	(2, 1, 0)(1, 1, 1) ₂₄
CPU Instr.	529×10^9	1023×10^9	211×10^9	125×10^9	669×10^9	565×10^9	168×10^9
AIC	66016.2	63986.4	64142.2	64286.1	63995.6	64015.8	64147.0
NMAE	3.44%	3.04%	3.04%	3.07%	3.04%	3.04%	3.04%
GQ statistics	0.94	1.00	1.01	1.01	1.00	1.00	1.01
p -value	0.10	0.95	0.81	0.74	0.96	0.92	0.80
JB statistics	2313.2	2994.2	3040.1	2998.8	3001.93	3117.4	3033.3
p -value	0.00	0.00	0.00	0.00	0.00	0.00	0.00

worsen a bit. Reducing P further to zero by removing Φ_{24} however reintroduces significant spikes in the autocorrelogram Figure 5(d) at lags 24, 72 and 96. Therefore the model SARIMAX(3, 1, 0)(0, 1, 1)₂₄($Y|X$) has to be ruled out, despite the relevantly reduced computation costs, as it misses too much of the process pattern. The worsening NMAE, AIC and GQ homoscedasticity test values agree with that.

Increasing Q by one via adding Θ_{48} , yielding SARIMAX(3, 1, 0)(0, 1, 2)₂₄($Y|X$), rectifies these issues considering autocorrelation as shown in Figure 5(e) and the model measures in Table V, but come at high computational costs without achieving better results than SARIMAX(3, 1, 0)(1, 1, 1)₂₄($Y|X$). Increasing P by re-adding Φ_{24} for SARIMAX(3, 1, 0)(1, 1, 2)₂₄($Y|X$) reduces these costs a bit, but reintroduces autocorrelation spikes at lag 24, 48 and 72 as visible in Figure 5(f).

So the best compromise can be found in SARIMAX(3, 1, 0)(1, 1, 1)₂₄($Y|X$), uniting the initial assumptions made by visual inspection and with *auto_arima*. However we have to note that Table IV shows for all discussed models high p -values for the third nonseasonal coefficient φ_3 , close to or above the threshold of 0.5 we have decided on before. Removing it results in almost the same model performance but at 20% lower computational costs. The autocorrelogram is barely distinguishable from

Figure 5(c) and is hence not explicitly depicted here.

Ergo we decide on **SARIMAX(2, 1, 0)(1, 1, 1)₂₄($Y|X$)** to model the emission intensity time series Y , and correspondingly **SARIMAX(2, 0, 0)(1, 1, 1)₂₄($\Delta Y|\Delta X$)** to model the differenced series ΔY .

All assessments were supported by corresponding results of LB tests on the nonseasonal lags 1 to 4 and the seasonal lags 1s to 4s, which alas did not find a place in this article due to limited space.

F. Forecasting

The SARIMAX(2, 1, 0)(1, 1, 1)₂₄($Y|X_f$) model was applied to forecast the development of Y in a rolling forecasts approach. The model was fitted on the last 12 weeks of data, used to generate a forecast for 24h ahead. The choice of 12 weeks was made after experiments in a search for a suitable preceding time span, i.e. the series should not be nonstationary over this span for any day in 2024.

Then the rolling window stepped forward by 24h, refitted on the last 12 weeks including now the actual data of the newly regarded day, forecasts the next day and so on. In a real-world application this corresponds to a 24h forecast horizon.

An exemplary comparison between forecasted and observed data is shown for January 2025 in Figure 6(a) as the month with the lowest, and May 2025 in Figure 7(a) as the one

Table VI: Result metrics of rolling 24h validation forecasts on the 1st and 2nd quarter 2025 for the hourly CO₂eq emission intensity series Y , as well as the differenced series ΔY , with SARIMAX(2, 1, 0)(1, 1, 1)₂₄($Y|X_f$) resp. SARIMAX(2, 0, 0)(1, 1, 1)₂₄($\Delta Y|\Delta X_f$). The months with lowest and highest MASE are highlighted.

Month	Y		ΔY	
	NMAE	MASE	NMAE	MASE
Jan	10.60%	52.47%	5.40%	52.18%
Feb	12.55%	70.49%	4.22%	58.31%
Mar	13.51%	80.03%	3.63%	62.44%
Apr	13.71%	74.34%	4.58%	61.91%
May	11.78%	91.85%	3.70%	71.64%
Jun	13.47%	84.33%	4.94%	64.97%
Jan-Jun	10.87%	73.86%	3.36%	61.73%

with the highest MASE value. The same was done to forecast ΔY , however the differencing had to be considered, so with $d_{\Delta Y} = 0$, and ΔX as exogenous factor at fitting resp. ΔX_f at forecasting. The corresponding forecasting model therefore has been SARIMAX(2, 0, 0)(1, 1, 1)₂₄($\Delta Y|\Delta X_f$). Again, January and May 2025 serve as examples in Figure 6(b) and Figure 7(b).

The month-wise prediction error metrics on forecasting Y and ΔY for the validation set in the first two quarters 2025 are listed in Table VI, by NMAE and MASE.

It is notable that the forecast on ΔY is significantly better than on Y by all metrics. This is explainable, as we fitted the model on the stationarized series $\Delta_s^{1,1}Y$ which is by differencing order closer to ΔY than Y . Further, tracing only the hourly changes in the data helps to focus on the seasonal pattern as trends have been removed. Also the values range has been reduced, and centred around zero, while still carrying the same information about the underlying process considering its seasonality, allowing to fit the model into the curve of actual data with less effort. This observation comes in handy, when algorithms as used in [16] for planning HPC cluster operations in advance, use the changes in the CO₂eq emission intensity data instead of the total values.

Further, January, February and April 2025 have shown a less regular daily pattern in Y , than the other three months. Therefore in these months the SARIMAX model is clearly superior to the naive seasonal forecast, while it performs in May better by only 9% according to MASE. However, when looking at ΔY , the SARIMAX model stays highly performant across all months. Though differencing emphasizes the regular seasonal pattern, it also via the reduced values range weights smaller daily changes stronger, which are harder to catch for a naive model than for a specially fitted model with more grades of freedom.

V. CONCLUSION

As a reference for a potential application of ML for scheduling jobs on HPC clusters in advance, in this work a

statistical SARIMAX model was developed and employed to forecast the hourly CO₂eq emission intensity of the German electricity production, and also its hourly changes, for a rolling 24h-horizon. To support the model, the concomitantly available data on the production of energy from low-emission sources were used as an exogenous factor, which has been proven as highly correlated, and for which suitable daily day-ahead forecast data by the BNetzA are available. Additional seasonalities in the data, which initially seemed promising for inclusion into the model as further exogenous factors, turned out to be not relevant for the model.

The suitability as a reference model has been shown by the accuracy of its predictions especially at pronounced swings, i.e. it satisfies its immediate purpose of forecasting and is already usable in corresponding productive systems. Further, the method's traceability allowed us to find a model in between the models suggested by conventional plot-interpretation and automatized parameter-search, which combines useful properties of both and provides better results.

However, the forecast horizon of SARIMAX is confined by the outreach of the exogenous variables' prediction. In this work for that the solely available day-ahead prognosis data of the Bundesnetzagentur [49] was used. Though the result is good for a pilot study, also with regard to the accomplished accuracy, and 24h of hourly prognosis data is the usual span in the field of CO₂eq emission intensity forecasting [18, 19], nevertheless for efficient HPC job scheduling this time span would possibly not be enough. One might want to take into account *Seasonal Cointegrated VARMAX*, the vector-based multivariate sibling of SARIMAX, which would in addition to the consideration of exogenous factors also allow joint predictions on temporally related variables [53].

With the reference base created in this work also further statistical methods can be evaluated like the *Generalised Additive Model* (GAM) or *Autoregressive Conditional Heteroskedasticity* (ARCH), ML-approaches like *Long Short-Term Memory* (LSTM) or *Foundation Models*, as well as hybrids of both types.

REFERENCES

- [1] Y. K. Dwivedi *et al.*, "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy," *Int. J. Inf. Manage.*, vol. 57, 2021. doi: 10.1016/j.ijinfomgt.2019.08.002
- [2] L. Cozzi, "World Energy Outlook Special Report: Energy and AI," International Energy Agency, Tech. Rep., 2025.
- [3] D. Stetskov, "The Great Software Quality Collapse: How We Normalized Catastrophe," *From the Trenches*, 2025. [Online]. Available: <https://techtrenches.substack.com/p/the-great-software-quality-collapse>
- [4] I. Shumailov *et al.*, "AI models collapse when trained on recursively generated data," *Nature*, vol. 631, pp. 755–759, 2024. doi: 10.1038/s41586-024-07566-y
- [5] P. Villalobos *et al.*, "Will we run out of data? Limits of LLM scaling based on human-generated data," 2024. doi: 10.48550/arXiv.2211.04325
- [6] J. Tesch *et al.*, "Metrology and AI: PTB's AI Strategy," *PTB-Mitteilungen*, vol. 132, no. 1, pp. 40–79, 2022. doi: 10.7795/310.20220199
- [7] S. Makridakis *et al.*, "Statistical and Machine Learning forecasting methods: Concerns and ways forward," *PLoS One*, vol. 13, 2018. doi: 10.1371/journal.pone.0194889

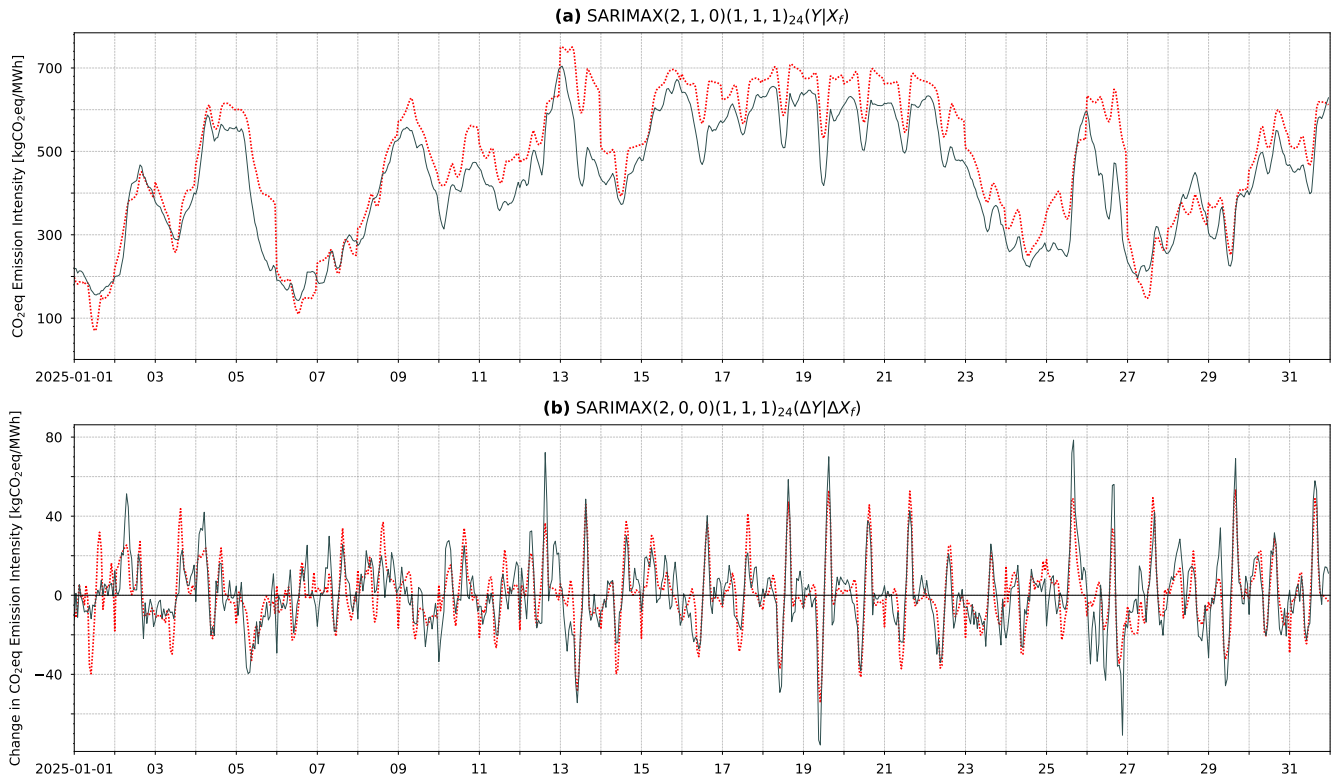


Figure 6: Rolling daily forecast for 24h ahead (dotted red) against the actually observed values (solid black) for the CO₂eq emission intensity series Y (a) resp. their hourly differenced series ΔY (b), both for Jan 2025.

- [8] V. I. Kontopoulou *et al.*, “A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks,” *Future Internet*, vol. 15, no. 8, 2023. doi: 10.3390/fi15080255
- [9] N. Thompson *et al.*, “The Computational Limits of Deep Learning,” in *9th Comput. within Limits. LIMITS*, 2023. doi: 10.21428/bf6fb269.1f033948
- [10] B. Xu and G. Yang, “Interpretability research of deep learning: A literature survey,” *Inf. Fusion*, vol. 115, p. 102721, 2025. doi: 10.1016/j.inffus.2024.102721
- [11] G. E. P. Box *et al.*, *Time Series Analysis: Forecasting and Control*, 5th ed. Hoboken, NJ, US: Wiley, 2015.
- [12] A. Tarsitano and I. L. Amerise, “Short-term load forecasting using a two-stage SARIMAX model,” *Energy*, vol. 133, pp. 108–114, 2017. doi: 10.1016/j.energy.2018.09.157
- [13] N. Elamin and M. Fukushima, “Modeling and forecasting hourly electricity demand by SARIMAX with interactions,” *Energy*, vol. 165, pp. 257–268, 2018. doi: 10.1016/j.energy.2018.09.157
- [14] A. M. Karki *et al.*, “Electric load prediction using SARIMAX,” *Aust. J. Electr. Electron. Eng.*, vol. 22, no. 3, pp. 410–423, 2025. doi: 10.1080/1448837X.2024.2405438
- [15] A. Kammeyer *et al.*, “HPC operation with time-dependent cluster-wide power capping,” in *Proc. 19th Conf. Comput. Sci. Intell. Syst. FedCSIS*, 2024. doi: 10.15439/2024F1066 pp. 385–393.
- [16] A. Kammeyer, “Digital Twin for Energy-Aware High-Performance Computing,” Doctoral Dissertation, Freie Universität Berlin, DE, 2025.
- [17] J. Gröger *et al.*, *Umweltzeichen Blauer Engel für Rechenzentren – Hintergrundbericht zur Erarbeitung der Vergabekriterien DE-UZ 228*. Dessau-Roßlau, DE: Umweltbundesamt, 2024.
- [18] P. Dutillo and F. Lisi, “Short-term CO₂ emissions forecasting: Insight from the Italian electricity market,” 2025. doi: 10.48550/arXiv.2507.12992
- [19] Y. Tian *et al.*, “Carbon dioxide emission forecast: A review of existing models and future challenges,” *Sustainability*, vol. 17, no. 4, 2025. doi: 10.3390/su17041471
- [20] L. K. Gohar and K. P. Shine, “Equivalent CO₂ and its use in understanding the climate effects of increased greenhouse gas concentrations,” *Weather*, vol. 62, no. 11, pp. 307–311, 2007. doi: 10.1002/wea.103
- [21] B. Leiner, *Spektralanalyse ökonomischer Zeitreihen*, 2nd ed. Wiesbaden, DE: Gabler, 1978.
- [22] R. H. Shumway and D. S. Stoffer, “Spectral Analysis and Filtering,” in *Time Series Analysis and Its Applications*. Cham, CH: Springer, 2025, ch. 4.
- [23] W. H. Greene, “Nonstationary Processes and Unit Roots,” in *Econometric Analysis*. New York, NY, US: Pearson, 2018, ch. 21.2.
- [24] R. J. Hyndman and G. Athanasopoulos, “Stationarity and differencing,” in *Forecasting: Principles and Practice*. Melbourne, AU: OTexts, 2021, ch. 9.1. [Online]. Available: <https://otexts.com/fpp3/stationarity.html>
- [25] J. D. Hamilton, “Asymptotic Properties of a p th-Order Autoregression and the Augmented Dickey-Fuller Tests for Unit Roots,” in *Time Series Analysis*. NJ, US: Princeton Univ. Press, 1994, ch. 17.7.
- [26] D. Kwiatkowski *et al.*, “Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?” *J. Econom.*, vol. 54, no. 1, pp. 159–178, 1992. doi: 10.1016/0304-4076(92)90104-Y
- [27] B. E. Hansen, “The New Econometrics of Structural Change: Dating Breaks in U.S. Labour Productivity,” *J. Econ. Perspect.*, vol. 15, no. 4, pp. 117–128, 2001. doi: 10.1257/jep.15.4.117
- [28] D. R. Osborn *et al.*, “Seasonality and the Order of Integration for Consumption,” *Oxford Bull. Econ. Stat.*, vol. 50, no. 4, pp. 361–377, 1988. doi: 10.1111/j.1468-0084.1988.mp50004002.x
- [29] F. Canova and B. E. Hansen, “Are Seasonal Patterns Constant Over Time? A Test for Seasonal Stability,” *J. Bus. Econ. Stat.*, vol. 13, no. 3, pp. 237–252, 1995. doi: 10.1080/07350015.1995.10524598
- [30] R. Nau, “ARIMA models for time series forecasting,” in *Statistical Forecasting: Notes on Regression and Time Series Analysis*. Durham, NC, US: Fuqua School of Business, 2020, ch. 5. [Online]. Available: <https://people.duke.edu/~rnau/411home.htm>
- [31] F. E. Harrell, “Variable Selection,” in *Regression Modeling Strategies*.

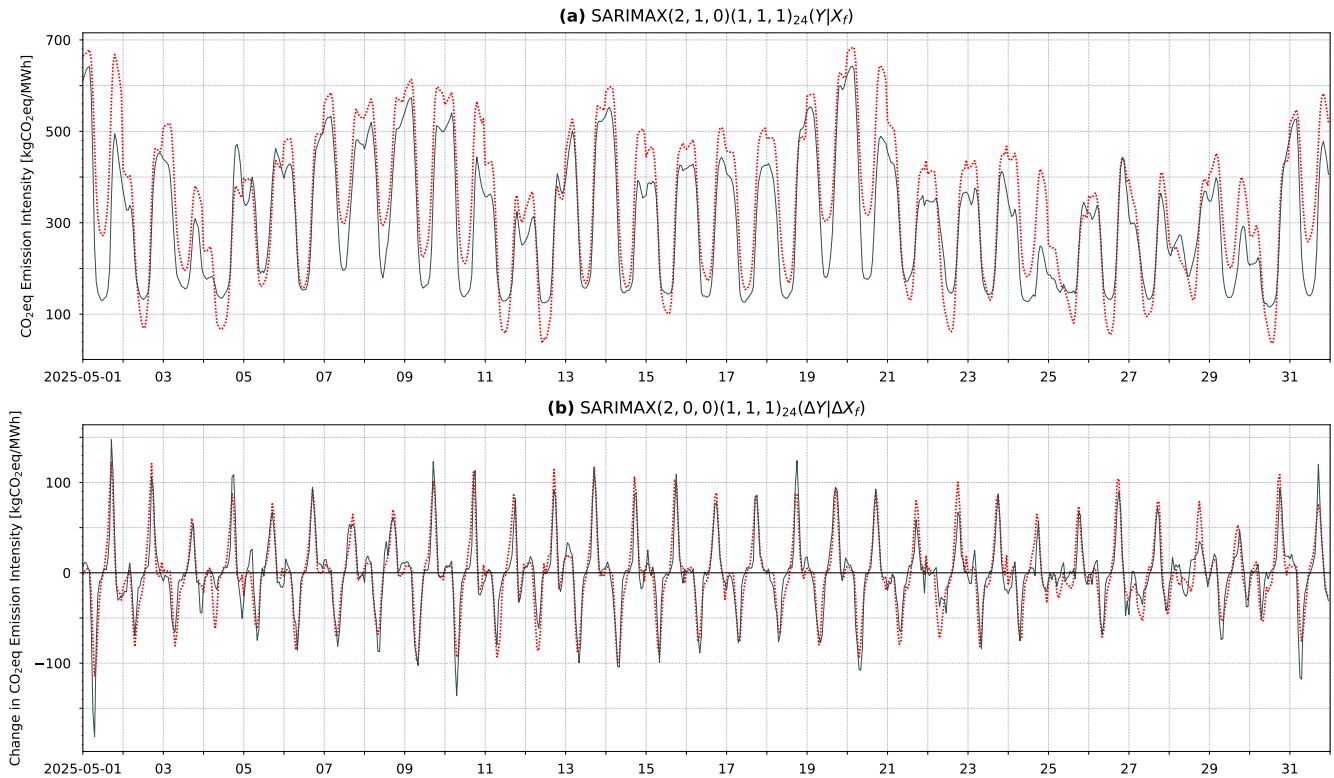


Figure 7: Rolling daily forecast for 24h ahead (dotted red) against the actually observed values (solid black) for the CO₂eq emission intensity series Y (a) resp. their hourly differenced series ΔY (b), both for May 2025.

- Cham, CH: Springer, 2015, ch. 4.3.
- [32] R. F. Engle *et al.*, "Exogeneity," *Econometrica*, vol. 51, no. 2, pp. 277–304, 1983. doi: 10.2307/1911990
- [33] K. Pearson, "Note on regression and inheritance in the case of two parents," *Proceedings of the Royal Society of London*, vol. 58, no. 347–352, pp. 240–242, 1895. doi: 10.1098/rspl.1895.0041
- [34] G. J. Székely *et al.*, "Measuring and testing dependence by correlation of distances," *Ann. Stat.*, vol. 35, no. 6, pp. 2769–2794, 2007. doi: 10.1214/009053607000000505
- [35] R. J. Hyndman and G. Athanasopoulos, "Time series decomposition," in *Forecasting: Principles and Practice*. Melbourne, AU: OTexts, 2021, ch. 3. [Online]. Available: <https://otexts.com/fpp3/decomposition.html>
- [36] C. J. Willmott and K. Matsuura, "On the use of dimensioned measures of error to evaluate the performance of spatial interpolators," *Int. J. Geogr. Inf. Sci.*, vol. 20, no. 1, pp. 89–102, 2006. doi: 10.1080/13658810500286976
- [37] R. J. Hyndman and A. B. Koehler, "Another look at measures of forecast accuracy," *Int. J. Forecasting*, vol. 22, no. 4, pp. 679–688, 2006. doi: 10.1016/j.ijforecast.2006.03.001
- [38] J. S. Armstrong, "Evaluating Forecasting Methods," in *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Dordrecht, NL: Kluwer, 2001, ch. 14.
- [39] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Autom. Control*, vol. 19, no. 6, pp. 716–723, 1974. doi: 10.1109/TAC.1974.1100705
- [40] Y. Yang, "Can the strengths of AIC and BIC be shared? A conflict between model identification and regression estimation," *Biometrika*, vol. 92, no. 4, pp. 937–950, 2005. doi: 10.1093/biomet/92.4.937
- [41] G. Schwarz, "Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978. doi: 10.1214/aos/1176344136
- [42] G. M. Ljung and G. E. P. Box, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 1978. doi: 10.1093/biomet/65.2.297
- [43] A. C. Harvey, "Testing and model selection," in *Forecasting, structural time series models and the Kalman filter*. GB: C.U.P., 1990, ch. 5.
- [44] C. M. Jarque and A. K. Bera, "A Test for Normality of Observations and Regression Residuals," *Int. Stat. Rev.*, vol. 55, no. 2, pp. 163–172, 1987.
- [45] D. Asteriou and S. G. Hall, "ARIMA Models and the Box-Jenkins Methodology," in *Applied Econometrics*. London, GB: Macmillan, 2021, ch. 13.
- [46] C. Chatfield, "Estimating Autocovariance and Autocorrelation Functions," in *The Analysis of Time Series: An Introduction*. Boca Raton, FL, US: Chapman & Hall/CRC, 2004, ch. 4.1.
- [47] T. G. Smith *et al.*, "pmdarima API Reference: auto_arima." [Online]. Available: https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.auto_arima.html
- [48] R. J. Hyndman and G. Athanasopoulos, "Transformations and adjustments," in *Forecasting: Principles and Practice*. Melbourne, AU: OTexts, 2021, ch. 3.1. [Online]. Available: <https://otexts.com/fpp3/transformations.html>
- [49] Bundesnetzagentur (BNetzA), "SMARD-Strommarktdaten, Stromhandel und Stromerzeugung in Deutschland." [Online]. Available: <https://www.smard.de/home/marktdaten>
- [50] T. G. Smith *et al.*, "pmdarima API Reference: nsdiffs." [Online]. Available: <https://alkaline-ml.com/pmdarima/modules/generated/pmdarima.arima.nsdiffs.html>
- [51] K. J. Millman *et al.*, "Mathematical Optimization," in *Scientific Python Lectures*. [Online]. Available: https://scipy-lectures.org/advanced/mathematical_optimization
- [52] M. Stuchlik, "Cirron." [Online]. Available: <https://github.com/s7nfo/Cirron>
- [53] M.-C. Düker *et al.*, "Vector AutoRegressive Moving Average Models: A Review," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 17, 2025. doi: 10.1002/wics.70009