Energy Proceedings
Vol 38, 2024
ISSN 2004-2965

Comparative Analysis of Baseline Models for Rolling Price Forecasts in the German Continuous Intraday Electricity Market

Timothée Hornek[1], Sergio Potenciano Menci[1], Joaquín Delgado Fernández[1], and Ivan Pavić[1]

[1]Interdisciplinary Centre for Security, Reliability and Trust - SnT, University of Luxembourg

ABSTRACT

Short-term electricity trading on intraday markets is crucial for integrating variable renewable energy in the power system. For instance, it allows energy suppliers to adjust their market positions based on updated variable renewable energy and consumption forecasts, reducing their potential imbalances. In the case of Germany, the continuous intraday market allows trading from the day before delivery until several minutes before delivery. However, the complex market design and high price volatility make developing price forecasting models challenging. This paper lays a foundation for price forecasting by comparing baseline models used to benchmark rolling continuous intraday price forecasts. These baselines help develop price forecasting models as they serve as a reference for these models. We also adapt a price normalization approach from the literature to benchmark price forecasts in a volatile market environment. Our baselines include the generalization of two baselines used in literature and one new baseline. We benchmark our baselines throughout 2021 and 2022. Among other baselines, we find that the price average of the last four trades yields the lowest root mean squared error. Moreover, the analysis suggests that baseline errors are independent of the market price development through normalization.

Keywords: machine learning, electricity price forecasting, continuous intraday market

NOMENCLATURE

Abbreviations

CID	continuous intraday
DA	day-ahead
EPEX	European Power Exchange
EPF	electricity price forecasting
MW	megawatt
PV	photovoltaic
SDAT	Single Delivery Area Trading
SIDC	Single Intraday Coupling
VRE	variable renewable energy
XBID	Cross-Border Intraday Market

Symbols

(s, ℓ)	product starting delivery at s with a length of ℓ
$P_k^{s,\ell}$	price of trade k in product (s, ℓ)
$R_{u,\gamma}^{s,\ell}$	regression target for product (s, ℓ) of horizon length γ at forecasting time u
$V_k^{s,\ell}$	volume of trade k in product (s, ℓ)
ℓ	product length
$\mathbb{B}_{u,\omega}$	time interval $[u - \omega, u)$
$\mathcal{S}_{u,\omega}^{s,\ell}$	set of trades in product (s, ℓ) during interval $[u - \omega, u)$
$\mathcal{T}^{s,\ell}$	set of trades in product (s, ℓ)
$\mathcal{U}_{\gamma}^{s,\ell}$	set of forecasting times for product (s, ℓ) and horizon length γ
s	delivery start

1. INTRODUCTION

In recent years, the European Union has introduced continuous intraday (CID) markets. They allow market participants, e.g., energy suppliers, to adjust their potential energy imbalances - the difference between the anticipated and the actual energy required. These imbalances are caused mainly by intermittent variable renewable energy (VRE) production [1]. A CID market in the European context is a continuous trading platform where trading is possible after the day-ahead (DA) auction until several minutes before delivery [2]. For instance, participants in the German CID market can trade electricity up to a few minutes before actual power delivery. In view of their real importance at the technical and economic level in the electricity system, we have noticed that the literature on CID price forecasting is scarce. The existing literature on CID electricity price forecasting (EPF) mainly focuses on the prediction of single aggregate CID price measures, i.e., averages of completed CID trades [3, 4]. Even less literature focuses on continuous price forecasts within trading sessions [5, 6, 7]. Further, to our knowledge, no existing literature explores and compares different baseline models, which

This is a paper for 15th International Conference on Applied Energy (ICAE2023), Dec. 3-7, 2023, Doha, Qatar.

are critical to developing CID EPF models. Baseline models are simple models against which one can compare more complex models. For the CID market, baseline model designs can be based on economic price theory, which states that current prices fully reflect available information in efficient markets [8]. However, it remains unclear how to extract optimal price information from the most recent transactions, i.e., whether the price of the most recent transaction is the best estimate or if including further transaction prices is beneficial. In addition, market volatility, such as that observed in 2021 and 2022 [9], presents an additional difficulty in designing baselines since popular error measures such as the root mean squared error (RMSE) are highly sensitive to volatile data. The availability of baselines that are robust to price volatility is crucial to ensure the quality of machine learning (ML)/artificial intelligence (AI) forecasting models and to enable the comparability of the performance over time and between models from different publications [10]. Normalization techniques can help to reduce data volatility. Existing literature on price data normalization only covers DA markets [11] and cannot be applied to CID markets, as price data differs due to different market designs. While an auction clears the DA market, yielding a single price for every product, the CID market is continuous, yielding a series of transaction prices for every product. Further, in the CID market, price volatility increases along the trading window [12]. In summary, given the identified gaps: 1) the lack of a systematic comparison of different baselines for CID EPF, and 2) the lack of literature on price data normalization for the CID market, we contribute with the following:

1. We explore existing literature to gather currently used baselines and generalize these to obtain two baseline designs.

2. We propose one new baseline design, combining and extending the available baselines in the literature.

3. We compare all baselines by evaluating their regression accuracy at several horizons using different metrics.

4. We adapt and adopt a price normalization approach for CID prices to enable the comparability of baseline models over time in a volatile market environment. The approach adapts to the volatility within a trading session for a single product and across products.

We base our study on completed CID trades of hourly products bought or sold in all German balancing zones during 2021 and 2022 on the European Power Exchange (EPEX) Spot market.

Overall, as a regression model baseline, the average of the four most recently completed trades performs well over the different forecast horizons. Further analysis indicates that the baseline performance is independent of market price increases.

Going forward, we introduce the German wholesale power market in section 2. Afterward, we discuss related work on the topic of EPF in section 3. We explain our normalization approach and baseline models in section 4. We provide a detailed analysis of our empirical comparison results in section 5. Finally, we conclude and provide an outlook in section 6.

2. BACKGROUND

The following section introduces the German spot market for electricity, which allows trading during the last hours before physical power delivery. Note that we focus on Germany, as it has the largest CID market in Europe by traded volume. Nevertheless, the market description and findings of the paper are also applicable to similar markets covered by the EPEX due to a similar market design [13]. The spot market is a short-term future market with two stages: an auction stage and a continuous stage, during which price clearing differs. The contracts traded on the spot market are commitments to deliver a specific volume of electricity in megawatt (MW) during a fixed time interval. Contracts with a length of $60min$, i.e., hourly products, and contracts with a length of $15min$, i.e., quarter-hourly products, dominate the market in Germany. Regarding traded volume, hourly products dominate both the auction and the continuous stage in Germany [13]. In our paper, we focus on the continuous trading of hourly products.

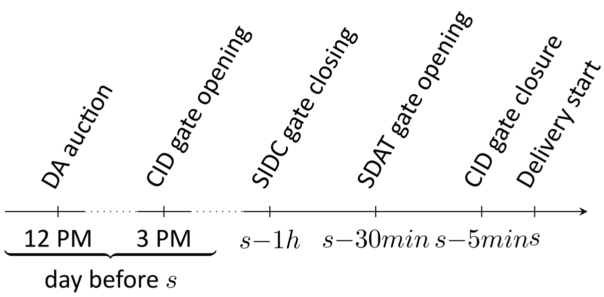

Fig. 1 Overview of CID trading for hourly products in Germany.

Figure 1 provides an overview of the trading process for hourly products in Germany. On the day before delivery at 12 PM, market participants can participate in the "DA auction" for hourly products. Note that another auction, called the "intraday auction", takes place three hours later at 3 PM for quarter-hourly products. The merit order principle used for the auction clearing determines prices in these two auctions. Afterward, the CID market gate opens at 3 PM for hourly products and at 4 PM for quarter-hourly products. The CID market is a continuous market where the participants submit “bids” and “asks” continuously to determine clearing prices. “Bids” are prices at which buyers are willing to buy, while “asks” are prices at which sellers are willing to sell. At market opening time, the Single Intraday Coupling (SIDC) mechanism couples European countries, meaning that participants can buy and sell power from other participating European countries. One hour before delivery, the SIDC gate closure limits transactions to participants within the German-Luxembourgish bidding zone [14, 15]. In other words, markets that are not in the same bidding zone decouple. Note that the frequently used term Cross-Border Intraday Market (XBID) refers to the technical implementation of the SIDC. Afterward, only buyers and sellers within the German-Luxembourgish bidding zone can trade with each other. Half an hour later, the Single Delivery Area Trading (SDAT) gate opening decouples the German-Luxembourgish bidding zone, which restricts trading to within each of the four control areas in Germany until the CID market closes five minutes before delivery. Since trading for all daily CID products of the same length starts simultaneously, but delivery begins at different times for each product, the trading interval lengths differ. For instance, an hourly product with a delivery starting at 3 AM trades one hour longer on the CID as a product with a delivery starting at 2 AM. For a more comprehensive description, readers can refer to more detailed market descriptions [2, 15, 16].

Table 1 collects the market liquidity analysis of hourly products using two different metrics, i.e., the share of volume traded and the percentage of completed trades. Our analysis of 2021 and 2022 CID trading data from EPEX [13] shows that over 65% of volume and 68% of trades occur during the last three hours before delivery. When also excluding the trades within the last $30min$ before delivery, these figures drop to 59% for volume traded and 62% for total trades. The market is most liquid during the last three hours before delivery since most trading occurs during this interval. Consequently, we limit ourselves to the last three hours before delivery for the remainder of the paper. Additionally, we exclude the last $30min$ before delivery because the restriction of trading within control areas does not allow for a single Germany-wide analysis.

Table 1 Liquidity of hourly products during the last three hours before delivery.

Year	Volume share		Trade count share	
	full $3h$	excluding last $30min$	full $3h$	excluding last $30min$
2021	65.73%	59.74%	68.76%	62.82%
2022	66.54%	59.32%	70.83%	64.21%

3. RELATED WORK

A large body of literature exists on the general topic of EPF. Weron [17] reviews the literature on EPF for DA markets and covers forecasting approaches such as fundamental models, statistical models, and computational intelligence models. The author identifies the benefit of fundamental models not in their forecasting accuracy but in their ability to depict market characteristics. The author further concludes that computational intelligence models are better able to handle non-linearity than statistical models. While Weron [17] focuses on point forecasts, other authors consider probabilistic price forecasts for the DA market Nowotarski and Weron [18].

Hong et al. [19] analyze the best solutions for probabilistic EPF submitted to the Global Energy Forecasting Competition 2014 (GEFCom2014). The context of the competition allowed a direct comparison of different forecasting methods. The best-ranked teams either use Neural Networks (NNs) or use quantile regression in combination with other regression methods. The literature review on the topic of probabilistic EPF by Nowotarski and Weron [18] confirms the superior performance of quantile regression-based models. However, it questions NNs’ reliability, since in their study, the NNs underperform for extended test periods.

The expansion of CID electricity markets in Europe during the last decade has opened a new field of study in the context of EPF. Shinde and Amelin [20] review general literature on intraday markets and prices. They find that most literature focuses on European markets and that the expansion of wind power is one of the drivers behind the introduction of CID markets. Literature provides different methodological approaches, such as econometric methods, point forecasting, and continuous forecasting.

Starting with econometric methods, various studies investigate different market properties from the economic point of view. More specifically, Kiesel and Paraschiv [21] analyze the prices of $15min$ products. They find that prices adjust asymmetrically to VRE forecasting errors. In other words, when the forecasted VRE production volumes increase, CID prices tend to fall and vice versa, making VRE production forecasts a potentially valuable feature for price forecasting. Likewise, Kremer, Kiesel, and Paraschiv [22] study $15min$ contracts, for which they build a fundamental model to assess how the trading depends on the slope of their empirically estimated merit order curve of the market. They find that prices of neighboring products, i.e., products with delivery just before or after the current product, have strong explanatory power for the price of the current product and that renewable forecast changes have an asymmetric effect on prices depending on the slope of the merit order curve. Kremer, Kiesel, and Paraschiv [23] study $15min$ products as well, but they focus on the prices of $15min$ contracts during night hours. They conclude that during the night, fundamentals such as VRE forecasts lose importance for EPF because price information drives the market.

Other publications focus on EPF beyond econometric studies. In the context of the CID market, two lines of research emerge.

First, the literature covers point forecasts for the CID, where the authors forecast a single price average for every product. One such average is the ID_3 price, which one can compute for every contract by averaging all trades between $3h$ and $30min$ before delivery [24]. The advantage of averaging is that by calculating a single price value for every contract, one obtains a time series of prices at the frequency of the contracts. One can then apply traditional time series forecasting methods for EPF. For instance, Kath and Ziel [25] forecast DA and full averages of CID prices and estimate the monetary benefit of using forecasts. Further, Uniejewski, Marcjasz, and Weron [26] apply the least absolute shrinkage and selection operator (LASSO) method to select explanatory variables for CID ID_3 prices. The most important variable in their model is the most recent intraday price. They also find clear economic benefits when applying simple trading strategies to their forecasting approach, exhibiting market inefficiency. The studies by Narajewski and Ziel [3] and by Marcjasz, Uniejewski, and Weron [4] forecast the ID_3 price. Narajewski and Ziel [3] study $60min$ and $15min$ contracts and find indications of weak-form market efficiency. In contrast, Marcjasz, Uniejewski, and Weron [4] study $60min$ contracts and dispute the claims of market efficiency.

Second, studies on continuous or path forecasts for CID prices exist. They consider trades in every contract as separate time series. Forecasting, therefore, focuses on the price changes while trading a single product, as opposed to price differences between different products. For instance, Scholz et al. [5] propose a model for CID EPF based on a rolling window approach. They train a NN and a gradient-boosting algorithm to forecast the future price trajectory. Similarly, Narajewski and Ziel [6] apply ensemble forecasting to simulate price trajectories of the CID market. The difference is that the authors obtain a probabilistic price forecast by simulating many price trajectories. Serafin, Marcjasz, and Weron [7] also propose a short-term probabilistic path forecasts model. Instead of simulating different paths of the CID price, they predict the price distribution by combining path and probabilistic forecasts.

In summary, the topic of CID EPF has considerable study potential since, so far, most studies focus on forecasting aggregate price measures such as the ID_3. Only one study applies a rolling window approach for CID EPF [5], two others propose short-term probabilistic forecasts [6, 7]. No study addresses the recent market volatility in the context of EPF. Further, literature does not yet address baseline selection for CID EPF. In our study, we contribute to bridging these gaps by comparing baselines for CID price forecast benchmarking in combination with a price normalization approach to address market volatility.

4. METHODS AND MODELS

The upcoming section covers our data processing approach, which leads to the computation of baselines and the configuration of ML/AI regression models. Section 4.1 introduces the basic notation used throughout the paper. Section 4.2 explains our data normalization approach, which is the first step for data processing. Afterward, we compute the targets in section 4.3 and the baselines in section 4.4.

4.1 General Notation

We introduce a notation to describe and aggregate CID prices, which we later rely on to define the baselines. Our notation expands the notation introduced by Narajewski and Ziel [3]. Starting with the definition of products, tet the tuple (s, ℓ) denote a product starting delivery at s with a length of ℓ. Products are contracts for the delivery of electricity during a specific time interval, such as the hourly products that we study. The delivery interval of the product (s, ℓ) is $[s, s + \ell)$, where $\ell = 1h$ for hourly products and s is the start of an arbitrary full hour. Furthermore, let the tuple $\mathcal{T}^{s,\ell}$ denote the trades in the product (s, ℓ). Within the tuple $\mathcal{T}^{s,\ell}$, the trades are ordered according to their execution time, i.e., from oldest to most recent. For any trade $k \in \mathcal{T}^{s,\ell}$, $V_k^{s,\ell}$ and $P_k^{s,\ell}$ denote the volume and price of the respective trade k. The unit of the volume $V_k^{s,\ell}$ is MWh, while that of the prices $P_k^{s,\ell}$ is €$/MWh$. Let $\mathcal{S}^{s,\ell} \subseteq \mathcal{T}^{s,\ell}$ be an arbitrary subset of trades of a particular product (s, ℓ).

Based on the notation introduced above, one can define volume-weighted price averages of trades in an arbitrary tuple of trades $\mathcal{S}^{s,\ell}$ as follows [3]:

$$m(\mathcal{S}^{s,\ell}) := \frac{1}{\sum_{k \in \mathcal{S}^{s,\ell}} V_k^{s,\ell}} \sum_{k \in \mathcal{S}^{s,\ell}} V_k^{s,\ell} P_k^{s,\ell}. \quad (1)$$

We can formulate well-known price indices with the above notation for any product (s, ℓ) [24]:

$$ID_1 = m(\mathcal{T}^{s,\ell} \cap [s - 1h, s - 30min)), \text{ and}$$
$$ID_3 = m(\mathcal{T}^{s,\ell} \cap [s - 3h, s - 30min)), \quad (2)$$

where trades within the last $30min$ of trading are not included in the averages. As motivated in the background section (see section 2), we proceed similarly, studying the last three hours before delivery and ignoring the trades within the last $30min$ before delivery. Within the remaining trading window, we repeat our computations every minute. We obtain the set of all forecasting times

$$\mathcal{U}_\gamma^{s,\ell} := \{s - 3h, s - 3h + 1min, \\ s - 3h + 2min, ..., s - 30min - \gamma\}, \quad (3)$$

where γ is the length of the forecasting horizon. Longer forecasting horizons lead to smaller sets of forecasting times $\mathcal{U}_\gamma^{s,\ell}$ to exclude the trades executed within the last $30min$ before delivery from the forecasting horizon. More concretely, assuming a forecasting horizon of $\gamma = 15min$ for a product (s, ℓ) starting at s, we make the earliest forecast at $s - 3h$ and the last forecast at $s - 45min$. The earliest corresponding forecast would cover the interval $[s - 3h, s - 2h45min)$ and the latest forecast would cover $[s - 45min, s - 30min)$, where the end of the last interval coincides with the end of the considered trading period. In our paper, we chose the forecasting horizons $\gamma \in \{1min, 5min, 10min, 15min, 30min, 60min\}$. We chose the horizons $15min$, $30min$, and $60min$ considering the study by [5]. We added shorter horizons to obtain results for the short-term price development.

4.2 Data Normalization

We use transaction prices of public trades from EPEX for the German CID market for 2021 and 2022, during which the wholesale electricity market was very volatile, which manifests itself in large price movements beyond historical price ranges [9]. The price volatility underlines the necessity for normalization. We propose an expanding window approach for normalization to make contracts statistically comparable. The general idea behind our normalization approach is that all trades for a specific product belong to a distribution unknown until trading stops. During the trading of the product, only past trades are available. Therefore, we estimate the unknown distribution's mean and standard deviation based on all known trades. As time passes and new trades come in, we get a new distribution based on which we update the respective estimates to include the latest trades.

In more technical terms, we adopt the Z-score normalization. For an arbitrary product (s, ℓ) and at any forecasting time $u \in \mathcal{U}_\gamma^{s,\ell}$, we compute the volume-weighted mean $\mu_u^{s,\ell}$ and standard deviation $\sigma_u^{s,\ell}$ of prices of all trades completed in the past, i.e., of all trades in $\mathcal{T}^{s,\ell} \cap (u - \infty, u)$. The set $\mathcal{T}^{s,\ell}$ contains all trades in product (s, ℓ). Using the estimated mean and standard deviation, we can normalize a value x as follows:

$$y := \frac{x - \mu_u^{s,\ell}}{\sigma_u^{s,\ell}}, \quad (4)$$

where y is the normalized value. We apply the same normalization to the baselines and their respective regression target.

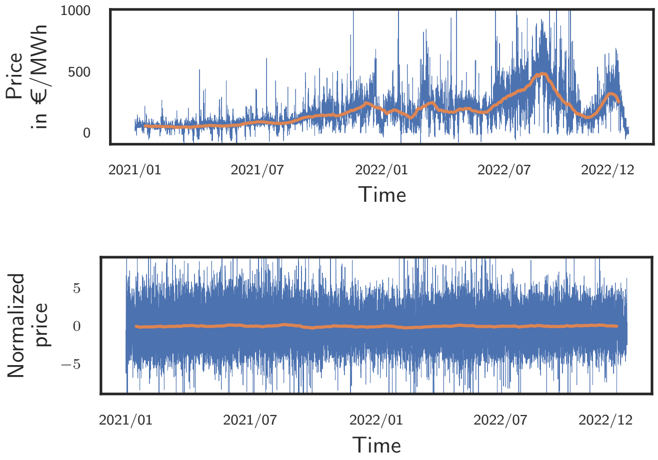

Fig. 2 One-minute price averages before and after normalization, and 30-day moving average.

Figure 2 contains the one-minute price trajectories before and after normalization. Additionally, the plot contains a 30-day moving average to visualize the price trends. Applying the normalization removes the visible trend in the price data and centers the data around zero. Consequently, with our normalization, we can compare the forecasting errors of our baselines in the time domain.

4.3 Targets

Regression targets are the values we aim to forecast with our baselines. We define the regression targets based on the price averaging introduced in Equation 1. For an arbitrary product (s, ℓ) and a given forecasting horizon γ, we obtain the regression targets $R_{u,\gamma}^{s,\ell}$ by computing

$$R_{u,\gamma}^{s,\ell} := m\left(\mathcal{T}^{s,\ell} \cap [u, u + \gamma)\right) \text{ for all } u \in \mathcal{U}_\gamma^{s,\ell}. \quad (5)$$

In other words, we compute the volume-weighted price average of the trades in $\mathcal{T}^{s,\ell} \cap [u, u + \gamma)$. The interval $[u, u + \gamma)$ covers the forecasting interval of length γ starting at forecasting time u.

4.4 Baselines

We consider baselines as simple, i.e., easy to understand and to compute, market-independent forecasting models that one can use to benchmark more complicated models. We consider different regression baselines to approximate future price averages, i.e., regression targets $R_{u,\gamma}^{s,\ell}$. The general idea for the computation of the baselines is that the most recent trades contain the most accurate market information. Computation-wise, the baselines select and average different numbers of the most recent trades. The following subsection introduces the tested baselines in more detail.

4.4.1 Baseline Definition

In general, we take a two-step approach to defining regression baselines:

1. We fix a base time interval in the past, which ends at the current forecasting time $u \in \mathcal{U}_\gamma^{s,\ell}$. We explore different base interval lengths ω, which extend to the *past*. The past interval length ω should not be confused with the forecasting horizon γ, which spans into the *future*. For a given base interval length ω at forecasting time $u \in \mathcal{U}_\gamma^{s,\ell}$, we obtain the base interval $\mathbb{B}_{u,\omega} := [u - \omega, u)$.

2. We average a subset of trades completed within the base interval, i.e., trades within $\mathcal{S}_{u,\omega}^{s,\ell} := \mathcal{T}^{s,\ell} \cap \mathbb{B}_{u,\omega}$. The base interval limits the age of the trades, i.e., trades within the interval can be no older than ω. The actual baselines differ in how they choose trades to average within the base interval.

We present three baselines with increasing complexity, starting with the simplest one, which averages all trades of a product within a base interval $\mathbb{B}_{u,\omega}$

$$m_\omega(\mathcal{S}_{u,\omega}^{s,\ell}) := m(\mathcal{S}_{u,\omega}^{s,\ell}). \quad (6)$$

Our formulation is a generalization of the baselines used by Scholz et al. [5]. The time interval limits the trades used for the baseline calculation according to age, which ensures that the baseline only averages recent trades.

A second approach is to average the n most recent trades in the base interval $\mathbb{B}_{u,\omega}$, which allows the baseline to adapt to changing market liquidity. Assuming a fixed number of trades n, the age of the oldest trade is higher when liquidity is low and lower when liquidity is high. In other words, the frequency of trades is higher when liquidity is high. Assuming a tuple of trades sorted by the time of execution $\mathcal{S}_{u,\omega}^{s,\ell} = (k_1, k_2, ..., k_{|\mathcal{S}_{u,\omega}^{s,\ell}|})$, we get the baseline

$$l_{\omega,n}(\mathcal{S}_{u,\omega}^{s,\ell}) := m\left(\left(k_{|\mathcal{S}_{u,\omega}^{s,\ell}|-n+1}, ..., k_{|\mathcal{S}_{u,\omega}^{s,\ell}|}\right)\right) \quad (7)$$

for $1 \leq n \in \mathbb{N}$. When $n > |\mathcal{S}_{u,\omega}^{s,\ell}|$, i.e., the number of trades is lower than n, we ignore n and apply the averaging procedure to all trades in $\mathcal{S}_{u,\omega}^{s,\ell}$. Narajewski and Ziel [3] use one particular case of the baseline, where $n = 1$. In contrast, our baseline definition is more general, allowing for different parameterizations. The baselines $m_\omega(\mathcal{S}_{u,\omega}^{s,\ell})$ and $l_{\omega,n}(\mathcal{S}_{u,\omega}^{s,\ell})$ are undefined in case $|\mathcal{S}_{u,\omega}^{s,\ell}| = 0$, i.e., the set of considered trades is empty. Therefore, we fill the respective entries by forward filling from the closest valid baseline value in a previous interval from the same product.

The third approach extends the second baseline $l_{\omega,n}(\cdot)$. Instead of deciding on a fixed number of trades to consider, we choose the number of trades as a share of the total number of trades in a specific time interval ω, i.e., $n_p := \left\lceil p \cdot |\mathcal{S}_{u,\omega}^{s,\ell}| \right\rceil$ for $p \in (0, 1]$:

$$h_{\omega,p}(\mathcal{S}_{u,\omega}^{s,\ell}) := l_{\omega,n_p}(\mathcal{S}_{u,\omega}^{s,\ell}). \quad (8)$$

Thus, the baseline $h_{\omega,p}$ combines the age limit for trades and the adaptability to market liquidity since the baseline covers a fixed interval and takes a share of trades. To our knowledge, such a baseline has not yet been used for CID price forecasting.

4.4.2 Baseline Parameterization

The generic definition of baselines excludes concrete parametrizations. Nevertheless, a concrete set of parameters is necessary to compute and compare our baselines.

Table 2 contains the parameter sets we use for the different baselines. In the first step, we choose the parameters considering existing literature. Scholz et al. [5] use m_{1min} and m_{15min}, which average the trades in the last $1min$ and $15min$ respectively. The study by Narajewski and Ziel [3] uses $l_{\infty,1}$, the price of the last trade, as a benchmark. In the second step, we expand the parametrizations experimentally to improve individual baseline performance and to cover larger parts of the parameter space. We use five parametrizations for m_ω, ten for $l_{\omega,n}$, and forty for $h_{\omega,p}$. Existing literature covers only three of those concrete baseline parametrizations: m_{1min}, m_{15min}, and $l_{\infty,1}$. Note that for $h_{\omega,p}(\cdot)$, we do not consider a base interval of $1min$ since such intervals contain too few trades to extract single-digit percentages of trades.

4.4.3 Baseline Comparison

Given our baseline definition, which requires baselines to be independent of the market, we formulate two criteria for baseline comparison that must apply beyond forecasting accuracy:

1. Baselines should be independent of *long-term* wholesale market prices, i.e., prices in adjacent months and years so that benchmarks are comparable for contracts with very different delivery times.

2. Baselines should be independent of *short-term* market prices, i.e., price movements in a single contract during adjacent minutes and hours so that benchmarks are comparable over different distances to delivery.

We independently assess these properties for every forecasting horizon γ. First, we divide the data by the month of the product considered and by the distance to

Table 2 Parametrization of Baselines.

Baseline	Base interval length	Other parameter
$m_{\omega}(\cdot)$	$\omega \in \{1min, 5min, 10min, 15min, 30min\}$	-
$l_{\omega,n}(\cdot)$	$\omega \in \{\infty\}$	$n \in \{1, 2, ..., 10\}$
$h_{\omega,p}(\cdot)$	$\omega \in \{5min, 10min, 15min, 30min\}$	$p \in \{1\%, 2\%, ..., 10\%\}$

delivery and obtain a dataset for every pair of month and distance to delivery. Second, we compute well-known error measures on these datasets, such as the RMSE and the mean absolute error (MAE). Note that we do not use the mean absolute percentage error (MAPE) since our data can contain entries with the value zero. The regression errors are comparable in scale due to previously applied normalization. Third, we average the computed RMSE and MAE along different dimensions:

1. We average over *all months and distances to delivery*, which yields a single average error for every baseline. The average errors reflect the two desired properties for every baseline since errors at different times and distances to delivery get equal weighting.

2. We average over *all months*, which yields an average error for every distance to delivery. The averages enable an analysis of baseline errors throughout the trading window.

3. We average over *all distances to delivery*, which yields an average error for every month in the dataset. The averages allow an analysis of baseline performance throughout the different months.

We use the different error averages to compare the baselines and assess the desired properties' fulfillment.

5. RESULTS AND DISCUSSION

We discuss our empirical results in the following section. We start by comparing all the baselines and discussing our error scores based on the error measures widely used in ML and AI, namely the RMSE and the MAE. As a result of the comparison, we select the best baseline to perform the remaining analysis. More specifically, we analyze the baseline accuracy over two years and at different distances to delivery.

5.1 Baseline Comparison

In the first step, we compare the baselines and their overall performance. We assess whether the last price $l_{\infty,1}$ is the most accurate baseline. Figure 3 depicts the relative change in error over the $l_{\infty,1}$ baseline. Note that although we computed the results for all parameters presented in Table 2, we only show the results for a selection of baselines due to space restrictions. The selection includes the best baselines and the baselines used in literature. Negative values point to a decrease, and positive numbers to an increase in error over the $l_{\infty,1}$ baseline. Hence, all entries for $l_{\infty,1}$ are zero. We highlight the three best baselines for every horizon length γ, i.e., those with the lowest average error. For the RMSE, at every horizon length γ, the baseline $l_{\infty,4}$, which averages the prices of the last four trades, performs best at all horizon lengths γ. For the MAE, the best baselines differ for the different horizon lengths γ. The baseline $h_{5min,2\%}$, which averages the last 2% of trades in the preceding $5min$ interval, performs best for $\gamma = 1min$. For $\gamma \in \{5min, 10min, 15min, 30min\}$, the baseline $l_{\infty,3\%}$ beats all the other baselines. For the longest forecasting horizon, i.e., for $\gamma = 60min$, the baseline $l_{\infty,2}$ yields the lowest error. For both error metrics and all horizon lengths, the last price baseline $l_{\infty,1}$ is never among the best-performing baselines. Further, the performance difference between the best and worse performing baselines decreases significantly for longer forecasting horizon lengths.

The reduction in error difference between the baselines when increasing the forecasting horizon means that baseline choice is critical for short-term forecasts. At the same time, the baseline choice matters less for long-term forecasts due to more similar performance between the baselines. The identified baselines yield lower RMSEs and MAEs than the baselines found in literature, i.e., m_{1min}, m_{15min} [5], and $l_{\infty,1}$ [3]. Moreover, only relying on the price of the last transaction is insufficient to capture the most recent price information optimally since the best baselines all consider more than a single price. Choosing a single baseline depends on the error metric and the forecasting horizon. For regression problems in ML and AI, it is common to minimize the RMSE. We, therefore, select the $l_{\infty,4}$ baseline, which averages the prices of the last four completed trades, because its RMSE is the lowest for all forecasting horizons γ. We limit the subsequent results and analysis to the baseline $l_{\infty,4}$.

5.2 Distance to Delivery Analysis

We analyze the baseline behavior over different distances to delivery next. Figure 4 presents the development of the forecasting error of the baseline $l_{\infty,4}$ throughout the considered trading window, which spans from $3h$ until $30min$ before delivery. The plot contains averages over 2021 and 2022 for every distance

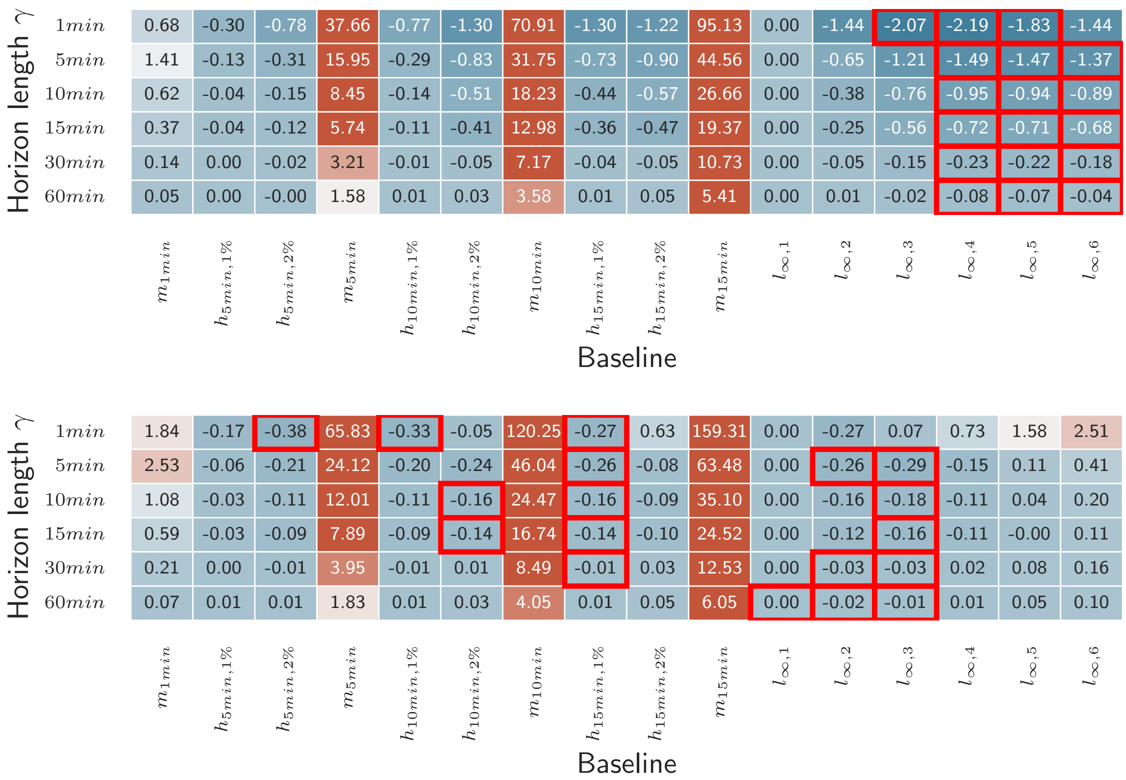

Fig. 3 RMSE (upper) and MAE (lower) changes in % over last price baseline $l_{\infty,1}$.

to delivery. Due to price data normalization, the absolute error values have no unit and are not directly interpretable. Note that for a forecasting horizon γ, we compute the last forecasting error at $\gamma + 30min$ before delivery. Consequently, depending on the forecasting horizon γ, the plot's error series stop at different distances to delivery. Over the complete interval, the errors for longer forecasting horizons are higher. The errors increase around full hours for all forecasting horizons γ. The rise is especially high for the RMSEs at short forecasting horizons, particularly at $60min$ before delivery. Apart from the heightened errors around full hours, the error is stable between $180min$ and $60min$ before delivery. After that, errors settle at a higher level.

The increase in errors for longer forecasting horizons reflects the higher forecasting uncertainty in the distant future. The rise of errors around full hours hints at increased short-term dispersion of errors, resulting from the market area change one hour before delivery. For the other hours, the explanation is more complicated. One possible explanation could be that some market participants try to close positions, disregarding current market prices. In contrast, the overall error levels align with the market structure outside the full hours. Until $60min$ before delivery, the low error level coincides with the larger market area. After that, the number of market participants shrinks as only domestic participants in the German-Luxembourgish bidding zone [15] are allowed trade with each other, reducing liquidity and

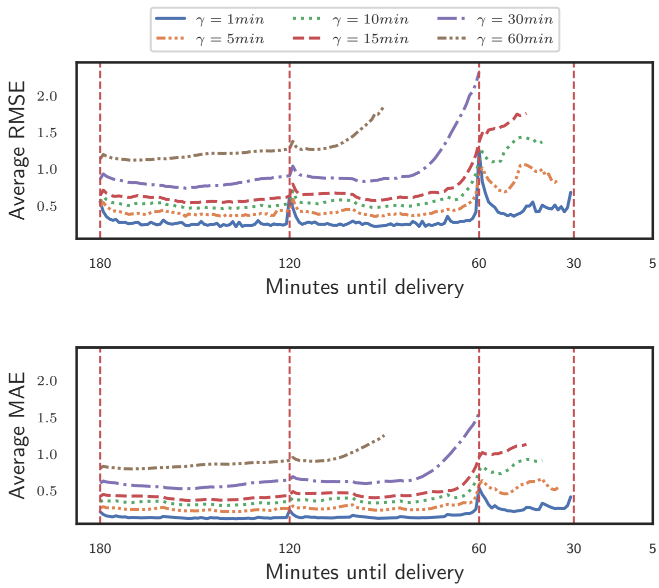

Fig. 4 Averages of RMSEs and MAEs at different distances to delivery.

increasing volatility, making price forecasting more difficult.

5.3 Monthly Analysis

Next, we evaluate the error levels throughout 2021 and 2022 to assess baseline performance in the volatile market environment. Figure 5 depicts the error development in monthly frequency. The errors are averages over all distances to delivery. Again, the errors increase for higher forecasting horizons γ. Besides small fluctuations, the errors are stable throughout the two years. During both years, the errors decrease slightly towards the end of the year and increase again at the start of the new year. Nonetheless, the errors are not correlated to market prices and market volatility.

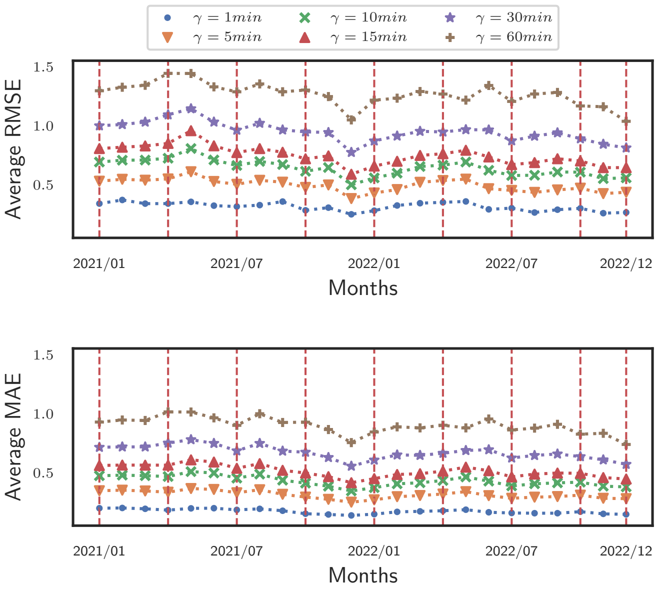

Fig. 5 Monthly averages of RMSEs and MAEs during 2021 and 2022.

Considering monthly errors is especially important in the context of market volatility. We observe that the errors are almost stable throughout the timeframe, confirming that our data normalization approach works. In other words, the different products are comparable in scale and volatility over time following normalization. It is, therefore, possible to use price data from the volatile market environment of 2021 and 2022 for model building. The seasonal fluctuations where lower errors occur in the winter could result from lower variable photovoltaic (PV) generation, leading to lower forecasting errors and, therefore, less market uncertainty.

5.4 Findings Summary

In summary, we can extract four important findings from our result analysis and discussion. First, baselines benefit from including the prices of several transactions in their calculations. Second, among the forecasting horizon we consider, which range from $1min$ to $60min$, the baseline forecasting errors increase for longer forecasting horizons, reflecting the higher uncertainty further into the future. Third, the market design has an impact on market volatility. Our analysis across different distances to delivery suggests that markets become less predictable when the trading area is reduced to the German-Luxembourgish bidding zone, leading to fewer market participants and hence lower market liquidity. Fourth, we have empirically demonstrated that the errors stay steady throughout the two-year interval in our dataset, which means that our normalization successfully makes prices comparable throughout the volatile market environment.

6. CONCLUSION AND OUTLOOK

In our paper, we study the topic of CID price forecasting. In the first part, we consider different baselines, which use different rules to select and average the most recent transaction prices. We find that the average of the last four transactions is a good baseline for price forecasts of different horizons from $1min$ to $60min$. Beside higher errors around full hours, the baseline performs well throughout the CID. We propose a normalization approach to handle changing market volatility and price levels. The steady errors of the baseline throughout 2021 and 2022 confirm that the normalization performs as intended. However, the limited time period examined remains a limitation of our study. It is imperative to continuously validate the findings with updated market data.

Different topics are relevant for future study of the CID market. For instance, one area for investigation is the market behavior around full hours. Further, the lower errors in December 2021 and 2022 deserve more attention. Regarding CID EPF, future regression model development can rely on the baselines compared in this paper for benchmarking. Further, multi-step EPF models are crucial in trading, as traders can derive a market direction from such forecasts. Model development should consider features reflecting fundamental market properties, such as VRE production forecasts, prices of neighboring products, and imbalance measures. Beyond features, different ML algorithms and NN architectures deserve further investigation.

AKNOWLEDGEMENT

This research was funded in part by the Luxembourg National Research Fund (FNR) and PayPal, PEARL grant reference 13342933/Gilbert Fridgen. For the purpose of open access, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission.

Supported by Enovos.

REFERENCES

[1] Koch C and Hirth L. Short-term electricity trading for system balancing: An empirical analysis of the role of intraday trading in balancing Germany's electricity system. en. Renewable and Sustainable Energy Reviews 2019 Oct; 113:109275. DOI: 10.1016/j.rser.2019.109275

[2] Directorate-General for Energy (European Commission) and Frontier Economics. METIS technical note T4: overview of European electricity markets. eng. LU: Publications Office of the European Union, 2019

[3] Narajewski M and Ziel F. Econometric modelling and forecasting of intraday electricity prices. en. Journal of Commodity Markets 2020 Sep; 19:100107. DOI: 10.1016/j.jcomm.2019.100107
[4] Marcjasz G, Uniejewski B, and Weron R. Beating the Naive–Combining LASSO with Naive Intraday Electricity Price Forecasts. en. Energies 2020 Jan; 13. Number: 7 Publisher: Multidisciplinary Digital Publishing Institute:1667. DOI: 10.3390/en13071667
[5] Scholz C, Lehna M, Brauns K, and Baier A. Towards the Prediction of Electricity Prices at the Intraday Market Using Shallow and Deep-Learning Methods. en. Mining Data for Financial Applications. Ed. by Bitetta V, Bordino I, Ferretti A, Gullo F, Ponti G, and Severini L. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2021 :101–18. DOI: 10.1007/978-3-030-66981-2_9
[6] Narajewski M and Ziel F. Ensemble forecasting for intraday electricity prices: Simulating trajectories. en. Applied Energy 2020 Dec; 279:115801. DOI: 10.1016/j.apenergy.2020.115801
[7] Serafin T, Marcjasz G, and Weron R. Trading on short-term path forecasts of intraday electricity prices. en. Energy Economics 2022 Aug; 112:106125. DOI: 10.1016/j.eneco.2022.106125
[8] Fama EF. Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance 1970; 25. Publisher: [American Finance Association, Wiley]:383–417. DOI: 10.2307/2325486
[9] ACER. ACER's Final Assessment of the EU Wholesale Electricity Market Design. 2022 Apr
[10] Jędrzejewski A, Lago J, Marcjasz G, and Weron R. Electricity Price Forecasting: The Dawn of Machine Learning. IEEE Power and Energy Magazine 2022 May; 20. Conference Name: IEEE Power and Energy Magazine:24–31. DOI: 10.1109/MPE.2022.3150809
[11] Uniejewski B, Weron R, and Ziel F. Variance Stabilizing Transformations for Electricity Spot Price Forecasting. IEEE Transactions on Power Systems 2018 Mar; 33. Conference Name: IEEE Transactions on Power Systems:2219–29. DOI: 10.1109/TPWRS.2017.2734563
[12] Baule R and Naumann M. Volatility and Dispersion of Hourly Electricity Contracts on the German Continuous Intraday Market. en. Energies 2021 Jan; 14. Number: 22 Publisher: Multidisciplinary Digital Publishing Institute:7531. DOI: 10.3390/en14227531
[13] EPEX SPOT. EPEX SPOT Market Data. 2023
[14] European Commission. Commission Regulation (EU) 2015/1222 of 24 July 2015 establishing a guideline on capacity allocation and congestion management. en. Legislative Body: OP_DATPRO. 2021 Mar
[15] NEMO Committee. Single Intraday Coupling (XBID) Information Package. 2021
[16] Ehrenmann A, Henneaux P, Küpper G, Bruce J, Klasman B, and Schumacher L. The future electricity intraday market design. en. 2019 Feb
[17] Weron R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. en. International Journal of Forecasting 2014 Oct; 30:1030–81. DOI: 10.1016/j.ijforecast.2014.08.008
[18] Nowotarski J and Weron R. Recent advances in electricity price forecasting: A review of probabilistic forecasting. en. Renewable and Sustainable Energy Reviews 2018 Jan; 81:1548–68. DOI: 10.1016/j.rser.2017.05.234
[19] Hong T, Pinson P, Fan S, Zareipour H, Troccoli A, and Hyndman RJ. Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond. en. International Journal of Forecasting 2016 Jul; 32:896–913. DOI: 10.1016/j.ijforecast.2016.02.001
[20] Shinde P and Amelin M. A Literature Review of Intraday Electricity Markets and Prices. 2019 IEEE Milan PowerTech. 2019 Jun :1–6. DOI: 10.1109/PTC.2019.8810752
[21] Kiesel R and Paraschiv F. Econometric analysis of 15-minute intraday electricity prices. en. Energy Economics 2017 May; 64:77–90. DOI: 10.1016/j.eneco.2017.03.002
[22] Kremer M, Kiesel R, and Paraschiv F. A Fundamental Model for Continuous Intraday Electricity Trading. eng. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 2020. Accepted: 2021-03-18T08:40:51Z Publisher: The Royal Society. DOI: 10.2139/ssrn.3489214
[23] Kremer M, Kiesel R, and Paraschiv F. Intraday Electricity Pricing of Night Contracts. en. Energies 2020 Jan; 13. Number: 17 Publisher: Multidisciplinary Digital Publishing Institute:4501. DOI: 10.3390/en13174501
[24] EPEX. Description of EPEX SPOT market indices. 2023 Feb
[25] Kath C and Ziel F. The value of forecasts: Quantifying the economic gains of accurate quarter-hourly electricity price forecasts. en. Energy Economics 2018 Oct; 76:411–23. DOI: 10.1016/j.eneco.2018.10.005
[26] Uniejewski B, Marcjasz G, and Weron R. Understanding intraday electricity markets: Variable selection and very short-term price forecasting using LASSO. en. International Journal of Forecasting 2019 Oct; 35:1533–47. DOI: 10.1016/j.ijforecast.2019.02.001