



Delft University of Technology

Document Version

Final published version

Licence

CC BY

Citation (APA)

Bölat, K., Palensky, P., & Tindemans, S. H. (2027). The living forecast: Evolving day-ahead predictions into intraday reality. *Electric Power Systems Research*, 262, Article 113532. <https://doi.org/10.1016/j.epsr.2026.113532>

Important note

To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright

In case the licence states "Dutch Copyright Act (Article 25fa)", this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

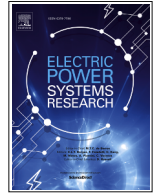
Sharing and reuse

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.



The living forecast: Evolving day-ahead predictions into intraday reality^{*}

Kutay Bölat^{ID*}, Peter Palensky^{ID}, Simon H. Tindemans^{ID}

Department of Electrical Sustainable Energy, Delft University of Technology, Delft, The Netherlands

ARTICLE INFO

Keywords:

Bayesian update
Gaussian mixture models
Intraday forecasting
Probabilistic forecasting
Variational autoencoders

ABSTRACT

Accurate intraday forecasts are essential for power system operations, complementing day-ahead forecasts that gradually lose relevance as new information becomes available. This paper introduces a Bayesian updating mechanism that converts fully probabilistic day-ahead forecasts into intraday forecasts without retraining or re-inference. The approach conditions the Gaussian mixture output of a conditional variational autoencoder-based forecaster on observed measurements, yielding an updated distribution for the remaining horizon that preserves its probabilistic structure. This enables consistent point, quantile, and ensemble forecasts while remaining computationally efficient and suitable for real-time applications. Experiments on household electricity consumption and photovoltaic generation datasets demonstrate that the proposed method improves forecast accuracy up to 25% across likelihood-, sample-, quantile-, and point-based metrics. The largest gains occur in time steps with strong temporal correlation to observed data, and the use of pattern dictionary-based covariance structures further enhances performance. The results highlight a theoretically grounded framework for intraday forecasting in modern power systems.

1. Introduction

Power systems rely heavily on accurate short-term forecasts to operate efficiently under uncertainty. While day-ahead forecasts offer a global view of the upcoming day, they become progressively outdated as the day progresses and new information becomes available. Intraday forecasting addresses this gap by refining the forecast for the remaining hours of the day, leveraging partially observed measurements and updated exogenous inputs [1]. This additional forecasting layer is critical for real-time decision-making tasks such as rescheduling generation, updating market positions, and managing reserves [2,3]. As renewable penetration and demand-side flexibility increase, the role of intraday forecasting in maintaining operational reliability becomes vital [4].

Given this operational role of intraday forecasting, the ability to quantify uncertainty is essential. While point forecasts provide a single-valued estimate of future outcomes, they fail to capture the uncertainty inherent in power system operations. Probabilistic forecasts address this by offering richer information about forecast uncertainty, enabling risk-aware decisions in scheduling, reserve allocation, and market bidding [5,6]. For example, quantile forecasts are widely used in electricity market bidding strategies, where upper and lower quantiles help balance profit maximization with the risk of imbalance penalties [3,7]. However,

these forecasts are typically produced independently for each time step, which means they provide information only about marginal distributions, not the temporal joint distribution. As a result, they do not support sampling of realistic trajectories and are limited in applications such as scenario generation or time-coupled optimization. Fully probabilistic forecasters overcome these limitations by modelling a joint probability distribution over the entire forecast horizon, conditioned on available inputs [8]. From this distribution, one can directly obtain point, quantile, and ensemble forecasts, enabling consistent and flexible use across a range of decision-making tasks. However, as the day progresses and new observations become available, these forecasts must be updated to preserve accuracy and relevance in intraday operations.

A variety of methods have been explored for intraday forecast refinement, yet each presents important limitations. One common approach is to train separate models for intraday forecasting conditioned on partial observations [3]. While effective, this strategy increases training and maintenance effort and may lead to inconsistencies with the original day-ahead forecast. Adaptive schemes based on sliding-window autoregressive training update predictions using recent data, but their data requirements grow with temporal resolution, and they may struggle to preserve performance over longer horizons. Heuristic corrections, such as bias adjustment or smoothing [1], are computationally efficient

^{*} This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956433.

^{*} Corresponding author.

E-mail addresses: K.Bolat@tudelft.nl (K. Bölat), P.Palensky@tudelft.nl (P. Palensky), S.H.Tindemans@tudelft.nl (S.H. Tindemans).

<https://doi.org/10.1016/j.epsr.2026.113532>

Received 14 October 2025; Received in revised form 13 March 2026; Accepted 20 June 2026

Available online 29 June 2026

0378-7796/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

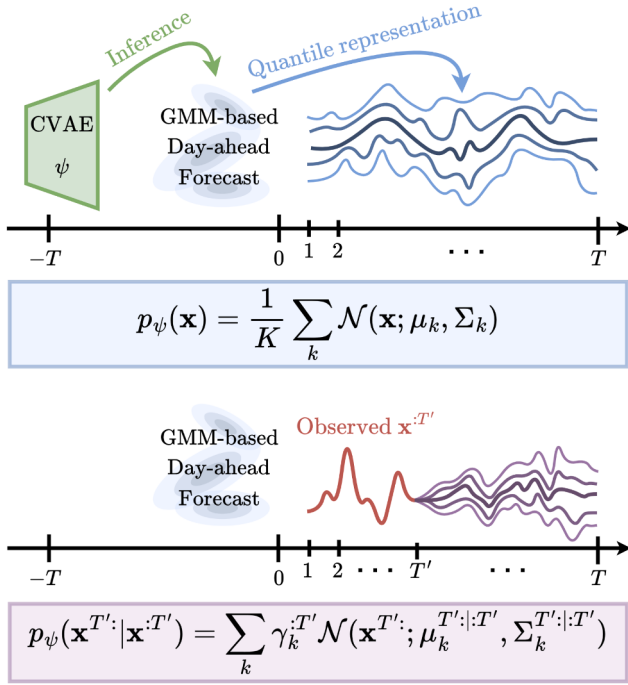


Fig. 1. Overview of the proposed Bayesian intraday update mechanism. A fully probabilistic day-ahead forecast, represented as a Gaussian mixture model, is obtained from a conditional variational autoencoder (top). Given partial intraday observations, the model applies a closed-form Bayesian update to adjust the remaining-horizon distribution (bottom), enabling consistent quantile, ensemble, and point forecasts without retraining or re-inference.

but operate primarily at the marginal level, lacking a coherent probabilistic structure. Copula-based approaches explicitly model temporal dependencies while preserving marginal distributions [9]. However, separating marginal and dependency modelling can introduce structural inconsistencies.

Recent probabilistic forecasting studies further emphasise the importance of joint uncertainty modelling in intraday power system applications. Ensemble-based mixture density networks have been applied to intraday PV forecasting to provide flexible distributional outputs [10], and Bayesian hierarchical frameworks have been proposed for continuous intraday electricity markets [2]. Multivariate update models for renewable infeed explicitly account for temporal correlations [1], while hybrid approaches combining Kalman post-processing with machine learning aim to refine intraday uncertainty estimates sequentially [11]. Advanced deep learning architectures integrating transformer models with Gaussian process approximations further demonstrate the potential of expressive probabilistic modelling [12].

Despite these advances, most existing approaches either require retraining, repeated neural network inference, or architectural modifications to incorporate new observations during the day. As a result, maintaining structural consistency with the original day-ahead joint distribution while ensuring computational scalability remains challenging.

In this work, we propose a simple and theoretically grounded approach for intraday forecast updating that directly builds on the existing day-ahead probabilistic forecast. Rather than retraining models, modifying architectures, or generating new scenarios from scratch, our method applies a Bayesian update to the original forecast distribution, conditioned on newly available intraday observations. The proposed update process is illustrated in Fig. 1: starting from the fully probabilistic day-ahead forecast distribution, the method conditions on the observed portion of the day to produce an updated distribution for the

remaining hours. The underlying day-ahead forecast is modelled using a conditional variational autoencoder with a Gaussian mixture output, enabling rich and flexible representation of uncertainty across time. This structure enables the update to be performed directly in the distribution space, without requiring retraining or re-running the neural network. As a result, the updated forecast remains fully probabilistic, supports sampling and quantile extraction, and preserves consistency with the original model. The approach avoids retraining and additional neural network inference during intraday operation, relying instead on closed-form linear algebraic updates whose complexity scales with the number of observed time steps. This makes it suitable for integration into real-time forecasting pipelines in practical power system applications.

2. Problem setting

We consider the problem of obtaining fully probabilistic intraday forecasts by leveraging the day-ahead forecasting distribution and observed intraday measurements. Specifically, we aim to model the conditional probability density function (pdf) $p(\mathbf{x}^{T'} | \mathbf{x}^{T'}, \mathbf{c})$ representing the intraday forecast using the day-ahead forecast pdf $p(\mathbf{x} | \mathbf{c})$. Here, $\mathbf{x} = [\mathbf{x}_t]_{t=1}^T \in \mathbb{R}^T$ represents the daily observation vector, consisting of two intra-day partitions defined as $\mathbf{x} = [\mathbf{x}^{T'}, \mathbf{x}^{T'}:]$ where $\mathbf{x}^{T'} = [\mathbf{x}_t]_{t=1}^{T'}$ corresponds to observations already obtained and $\mathbf{x}^{T'} = [\mathbf{x}_t]_{t=T'+1}^T$ represents future intraday observations. The vector $\mathbf{c} \in \mathbb{R}^C$ denotes exogenous inputs, including historical measurements, calendar-related variables (e.g., holidays, weekdays), and external factors such as weather forecasts (e.g., temperature, wind speed).

In practical terms, the objective is to “update” the day-ahead forecast $p(\mathbf{x} | \mathbf{c})$ with new incoming measurements observed up to time T' and convert it efficiently into an updated intraday forecast $p(\mathbf{x}^{T'} | \mathbf{x}^{T'}, \mathbf{c})$. This updating process is illustrated in Fig. 1. Note that the fully probabilistic formulation of the forecast inherently facilitates seamless extraction of various predictive outputs, including point, ensemble, and quantile forecasts [8].

Formally, the intraday forecast pdf $p(\mathbf{x}^{T'} | \mathbf{x}^{T'}, \mathbf{c})$ is obtained from the day-ahead forecast $p(\mathbf{x} | \mathbf{c})$ using Bayes’ theorem:

$$p(\mathbf{x}^{T'} | \mathbf{x}^{T'}, \mathbf{c}) = \frac{p(\mathbf{x} | \mathbf{c})}{p(\mathbf{x}^{T'} | \mathbf{c})}. \quad (1)$$

This relationship emerges directly from Bayesian inference, where the denominator $p(\mathbf{x}^{T'} | \mathbf{c})$ can typically be computed by marginalizing over the original forecast distribution. However, the resulting conditional pdf does not generally remain within the same family as the original distribution, particularly when complex neural network-based distributions are employed. Consequently, our goal is to develop a day-ahead forecast distribution model whose structure is preserved under Bayesian updating, ensuring computational scalability and structural consistency in generating intraday probabilistic forecasts.

3. Methodology

3.1. Day-ahead forecasting model

We employ a conditional variational autoencoder (CVAE) [8,13] for modelling the fully probabilistic day-ahead forecasting model as

$$p_\psi(\mathbf{x} | \mathbf{c}) = \mathbb{E}_{p(\mathbf{z})}[p_\psi(\mathbf{x} | \mathbf{z}, \mathbf{c})] \approx \frac{1}{K} \sum_k p_\psi(\mathbf{x} | \mathbf{z}_k, \mathbf{c}) =: \tilde{p}_\psi(\mathbf{x} | \mathbf{c}) \quad (2)$$

where $\mathbf{z} \in \mathbb{R}^L$ is the latent variable distributed as $p(\mathbf{z})$ and $\mathbf{z}_k \sim p(\mathbf{z})$ are the Monte Carlo samples used to approximate the intractable expectation. This results in a K -component mixture distribution $\tilde{p}_{\psi, \mathbf{U}}(\mathbf{x} | \mathbf{c})$,

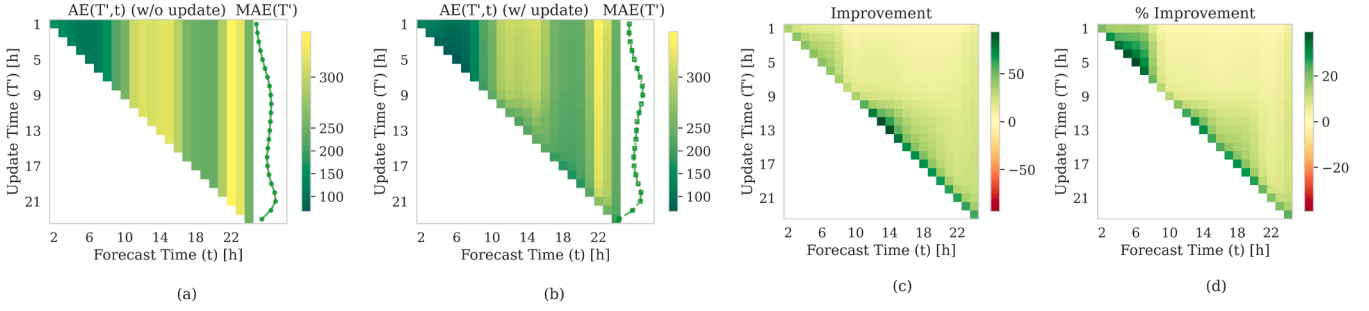


Fig. 2. The sample-based metrics $AE(T', t)$ and $MAE(T')$ for the Gipuzkoa SM dataset obtained by (a) the baseline non-updated model and (b) the intraday model with the proposed update mechanism. The (c) absolute and (d) percentage differences in the metric $AE(T', t)$ between the models reflect the varying improvements that come with the updates.

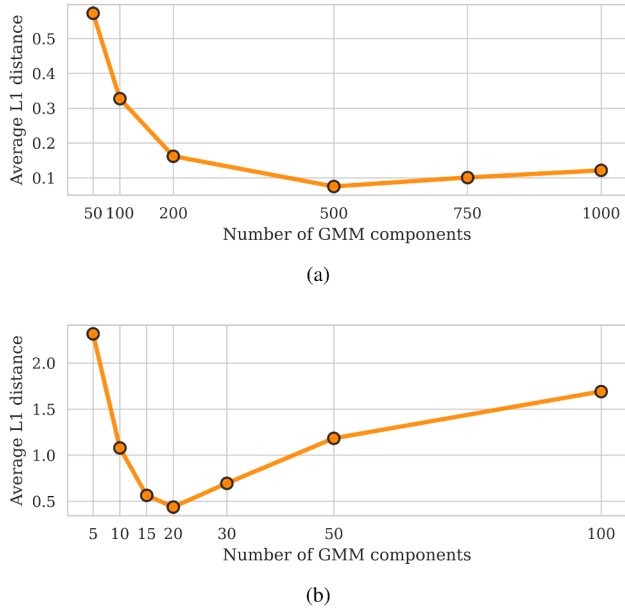


Fig. 3. Tuning of the number of GMM components for the (a) Gipuzkoa SM and (b) Utrecht PV datasets.

parameterized by a deep neural network ψ that maps latent variables and conditions to distribution parameters.

In this work, we model each component as a multivariate normal (MVN) distribution:

$$p_{\psi}(\mathbf{x}|\mathbf{z}_k, \mathbf{c}) = \mathcal{N}(\mathbf{x}; \mu = f_{\psi}^{\mu}(\mathbf{z}_k, \mathbf{c}), \Sigma = f_{\psi}^{\Sigma}(\mathbf{z}_k, \mathbf{c})) \quad (3)$$

where $f_{\psi}^{\mu}(\cdot) : \mathbb{R}^{L+C} \rightarrow \mathbb{R}^T$ and $f_{\psi}^{\Sigma}(\cdot) : \mathbb{R}^{L+C} \rightarrow \mathbb{R}^{T \times T}$ are two distinct outputs of a neural network that parameterize the mean and covariance of the MVN. This structure effectively turns the CVAE into a conditional Gaussian mixture model (GMM), with the number of components K freely chosen during inference, offering flexibility and expressiveness.

A key challenge in this setup is ensuring that the covariance matrix Σ remains positive-definite. A common solution is to restrict Σ to be diagonal:

$$f_{\psi}^{\Sigma}(\mathbf{z}, \mathbf{c}) = \text{diag}(f_{\psi}^{\sigma}(\mathbf{z}, \mathbf{c}))^2, \quad (4)$$

where $f_{\psi}^{\sigma}(\mathbf{z}, \mathbf{c}) : \mathbb{R}^{L+C} \rightarrow \mathbb{R}_+^T$ outputs the marginal standard deviations. To overcome the limitations of diagonal covariance and enable richer representations, we employ Pattern Dictionary-based Covariance Composition (PDCC) [14]:

$$f_{\psi}^{\Sigma}(\mathbf{z}, \mathbf{c}) = \mathbf{U} \text{diag}(f_{\psi}^{\sigma}(\mathbf{z}, \mathbf{c}))^2 \mathbf{U}^T + \xi \mathbf{I}, \quad (5)$$

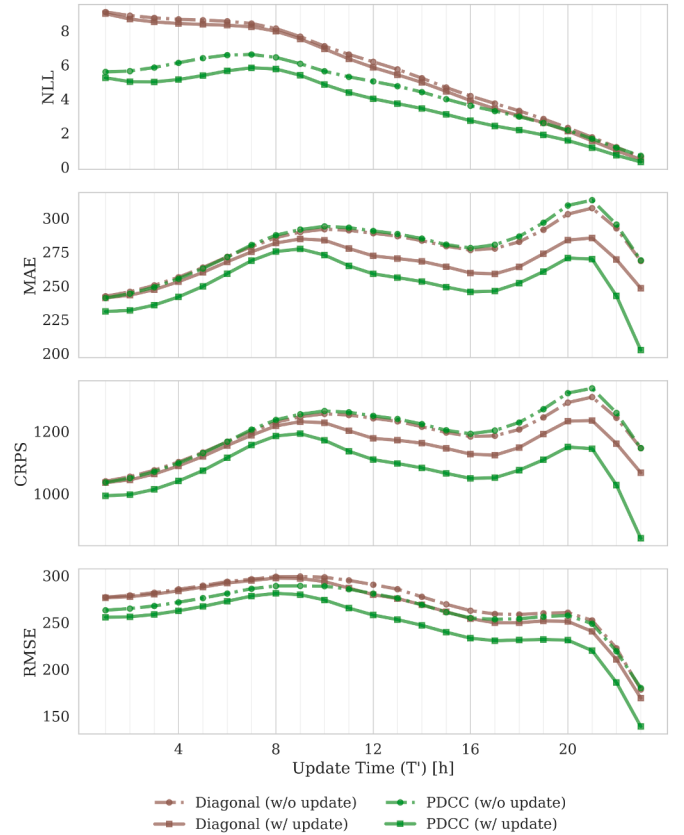


Fig. 4. Performance traces over the Gipuzkoa SM synthetic test set. Comparing models with full covariance matrices (PDCC) against diagonal ones.

where $\mathbf{U} \in \mathbb{R}^{T \times V}$ is a learnable pattern dictionary (with $V \geq T$), $f_{\psi}^{\sigma}(\mathbf{z}, \mathbf{c}) : \mathbb{R}^{L+C} \rightarrow \mathbb{R}_+^V$ outputs auxiliary standard deviations and ξ is a small constant for numerical stability. Importantly, $\mathbf{U}$ is shared across inputs and does not depend on $\mathbf{z}$ and $\mathbf{c}$, enabling the model to learn common covariance patterns across data points.

3.2. Intraday updating mechanism

Once the day-ahead forecast $\tilde{p}_{\psi}(\mathbf{x}|\mathbf{c})$ is obtained, we aim to reuse it for intraday forecasting without any re-inference¹ or retraining. That is, we adapt the GMM output of the CVAE directly, as illustrated in Fig. 1.

¹ We refer to re-inference as running another forward pass on the neural network.

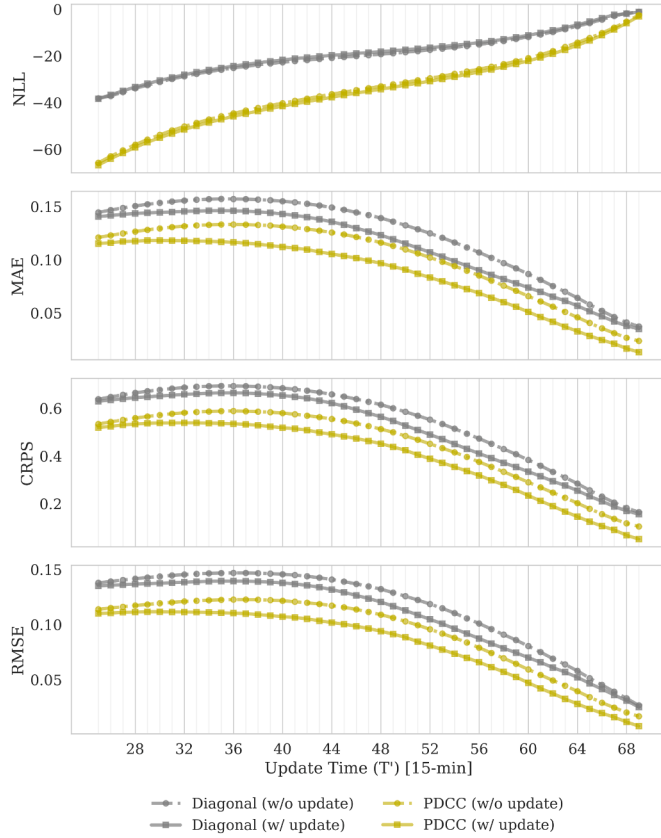


Fig. 5. Performance traces over the Utrecht PV synthetic test set. Comparing models with full covariance matrices (PDCC) against diagonal ones.

Applying the Bayes' rule in (1) to the day-ahead forecast distribution in (2), the updated intraday forecast becomes:

$$\begin{aligned} \tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{x} : T', \mathbf{c}) \\ = \sum_k \underbrace{\frac{\tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{z}_k, \mathbf{c})}{\sum_{k'} \tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{z}_{k'}, \mathbf{c})}}_{\gamma_k^{T'}} \tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{x} : T', \mathbf{z}_k, \mathbf{c}) \end{aligned} \quad (6)$$

which forms a $T - T'$ -dimensional mixture model with weights $\gamma^{T'} = [\gamma_k^{T'}]_k$ lies in the $(K-1)$ -dimensional standard simplex. For the MVN components of the day-ahead forecasting distribution $\tilde{p}_\psi(\mathbf{x} | \mathbf{c})$, let us define

$$\mu_k = \begin{bmatrix} \mu_k^{T'} \\ \mu_k \end{bmatrix} = f_\psi^{\mu}(\mathbf{z}_k, \mathbf{c})$$

and

$$\Sigma_k = \begin{bmatrix} \Sigma_k^{T'} & \Sigma_k^{\times T'} \\ \Sigma_k^{\times} & \Sigma_k \end{bmatrix} = f_\psi^{\Sigma}(\mathbf{z}_k, \mathbf{c}).$$

By definition, the marginal distribution up to time T' becomes

$$\tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{z}_k, \mathbf{c}) = \mathcal{N}(\mathbf{x}^{T'} : | \mu_k^{T'}, \Sigma_k^{T'}). \quad (7)$$

Since each day-ahead mixture component $\tilde{p}_\psi(\mathbf{x} | \mathbf{z}_k, \mathbf{c})$ is an MVN, intraday components $\tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{x} : T', \mathbf{z}_k, \mathbf{c})$ preserve the MVN structure as [15]

$$\tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{x} : T', \mathbf{z}_k, \mathbf{c}) = \mathcal{N}(\mathbf{x}^{T'} : | \mu_k^{T' : | : T'}, \Sigma_k^{T' : | : T'}) \quad (8)$$

where

$$\mu_k^{T' : | : T'} = \mu_k^{T'} + \Sigma_k^{\times} \Sigma_k^{T'-1} (\mathbf{x} : T' - \mu_k^{T'}) \quad (9)$$

and

$$\Sigma_k^{T' : | : T'} = \Sigma_k^{T'} - \Sigma_k^{\times} \Sigma_k^{T'-1} \Sigma_k^{\times T'}. \quad (10)$$

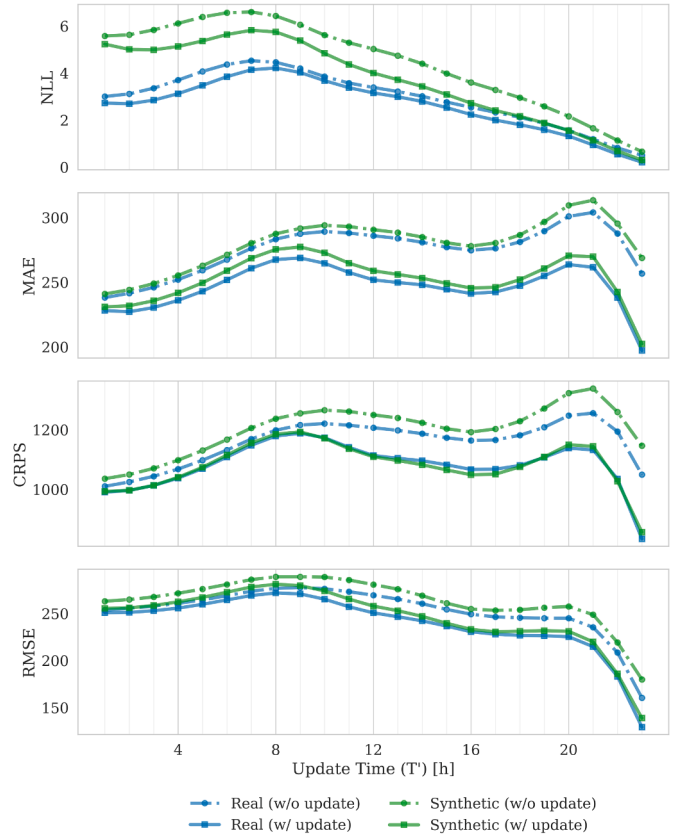


Fig. 6. Performance traces of the models with full covariance matrices (PDCC) for the Gipuzkoa SM dataset. Comparing the synthetic test set against the real one.

Intuitively, the update process can be seen as favouring the day-ahead mixture components $\tilde{p}_\psi(\mathbf{x} | \mathbf{z}_k, \mathbf{c})$ that best “guessed” the observed measurements $\mathbf{x}^{T'}$, by using $\gamma^{T'}$ to increase the weight of their updated distributions $\tilde{p}_\psi(\mathbf{x}^{T'} : | \mathbf{x} : T', \mathbf{z}_k, \mathbf{c})$ relative to the other components for the rest of the day ($\mathbf{x}^{T' :}$).

Efficient intraday sampling

The updated distribution can be used to form an ensemble forecast $\{\hat{\mathbf{x}}_s^{T'}\}_{s=1}^S$ by sampling S equiprobable traces from it. This sampling can be done in a hierarchical manner:

1. Randomly select a component: $\hat{k}_s \leftarrow \text{Categorical}(\gamma^{T'})$
2. Calculate and store $\mu_{\hat{k}_s}^{T' : | : T'}$ and $\Sigma_{\hat{k}_s}^{T' : | : T'}$ using (9) and (10) if not stored before
3. Sample from the selected and updated component: $\hat{\mathbf{x}}_s^{T'} \leftarrow \mathcal{N}(\mathbf{x}^{T'} : | \mu_{\hat{k}_s}^{T' : | : T'}, \Sigma_{\hat{k}_s}^{T' : | : T'})$

Note that the storing in the second step can significantly reduce the computational cost as (9) and (10) are computed only once for a sampled $\hat{k}_s$ and re-used later if it gets sampled again. This is particularly effective when the number of samples S is large or the weights in $\gamma^{T'}$ are sparse, i.e. most entries are near zero.

Computational considerations

The intraday update requires only closed-form linear-algebraic operations and no additional neural network inference. For each mixture component, the dominant cost stems from inverting $\Sigma_k^{T'} \in \mathbb{R}^{T' \times T'}$, resulting in a per-component complexity of $\mathcal{O}(T'^3)$, where T' denotes the number of observed time steps. The remaining matrix multiplications scale as $\mathcal{O}((T - T')T'^2)$. Since T' increases gradually during the day and

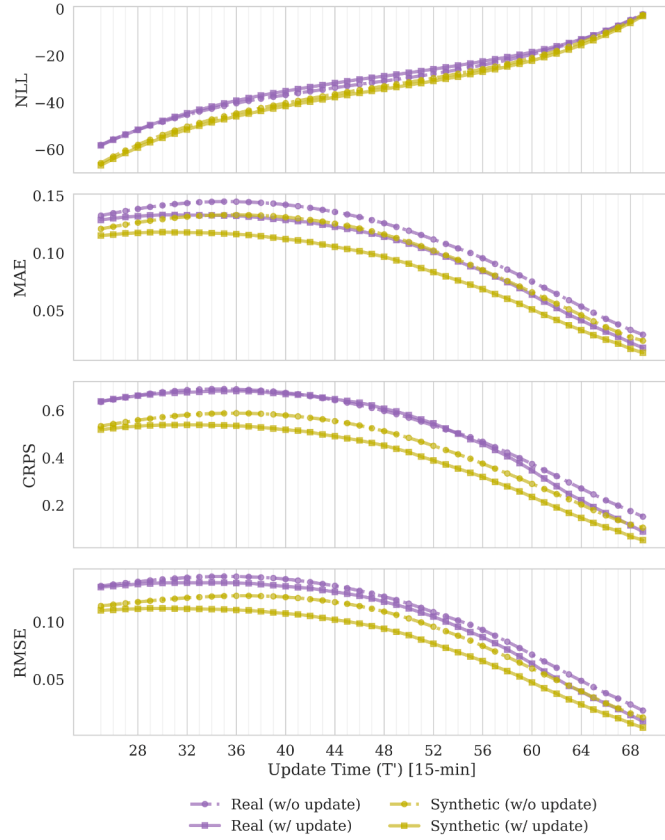


Fig. 7. Performance traces of the models with full covariance matrices (PDCC) for the Utrecht PV dataset. Comparing the synthetic test set against the real one.

remains moderate in practical settings, the update remains computationally tractable. Moreover, conditional parameters are computed once per component and cached during sampling, and sparsity in mixture weights further reduces effective run time.

4. Experiments

4.1. Datasets

4.1.1. Gipuzkoa SM

We employed a smart meter dataset capturing hourly electricity consumption in kWh for 25,559 residential, office, and business customers across 47 Spanish provinces from November 2014 to June 2022 [16]. Experiments were conducted on a subset of the dataset, consisting of records from the data-dense Gipuzkoa region and a one-year period from June 2021 to May 2022. The final dataset includes 6830 customers with one full year of continuous data, resulting in approximately 2.5 million records (365 days per customer, $T = 24$ hourly readings per day). Data preprocessing involved excluding energy-producing and consistently zero-consumption users, applying zero-preserved log-normalization with hyperparameters from [14], and partitioning the dataset into training, validation, and testing sets (60%, 20%, and 20%, respectively). As in [8], day-before consumption profiles and cyclically encoded (sin-cos) months (Jan, Feb, etc.), days of the week (Mon, Tue, etc.) were used as the exogenous inputs.

4.1.2. Utrecht PV

The other dataset used is a 4-year collection of household PV power production measurements in kW from 175 households in Utrecht, Netherlands [17]. The 1-min resolution measurements were aggregated to 15-min values for this study ($T = 96$), and daily profiles were extracted. The last 28 time steps (measurements after 17:00) of the daily

profiles were excluded from performance assessments, as they are dominated by zeros and bias the representativeness of the metrics. Each household's measurement is normalized by its estimated AC capacity (peak output power), and the dataset is partitioned with the same ratios as the Gipuzkoa SM dataset. The exogenous inputs consisted of 24-h weather forecasts (surface solar radiation, clear sky irradiance, medium and total cloud coverages) for Utrecht, in addition to the day-before generation profiles.

4.2. Performance metrics

The intraday forecaster produces fully probabilistic outputs, enabling performance evaluation using likelihood-, sample-, quantile-, and point-based metrics. Let the test set be $\mathcal{X} = \{(\mathbf{x}_n, \mathbf{c}_n)\}_{n=1}^N$, where $\mathbf{x}_n = [x_{t,n}]_{t=1}^T = [\mathbf{x}_n^{T'}, \mathbf{x}_n^{T''}]$. For all metrics, lower values indicate better performance.

4.2.1. Likelihood-based

Using the closed-form intraday forecast in (6), we compute the negative loglikelihood (NLL) of the unobserved part $\mathbf{x}_n^{T''}$:

$$\text{NLL}(T') = -\frac{1}{N} \sum_n \log \tilde{p}_\psi(\mathbf{x}_n^{T''} = \mathbf{x}_n^{T''} | \mathbf{x}_n^{T'}, \mathbf{c}_n). \quad (11)$$

Note that $\tilde{p}_\psi(\mathbf{x}|\mathbf{c})$ depends on the sampled CVAE latent variables $\{\mathbf{z}_k\}$, so different samples may yield different NLL values for the same trained CVAE model, $p_\psi(\mathbf{x}|\mathbf{c})$.

4.2.2. Sample-based

From the intraday distribution, we form ensembles $\hat{\mathbf{x}}_{n,s}^{T''} \sim \tilde{p}(\mathbf{x}_n^{T''} | \mathbf{x}_n^{T'}, \mathbf{c}_n)$ of size S .² Then we compute the absolute error per time step as

$$\text{AE}(T', t) = \frac{1}{NS} \sum_{n,s} |x_{t,n}^{T''} - \hat{x}_{t,n,s}^{T''}|, \quad (12)$$

with the mean absolute error (MAE) across the remaining horizon

$$\text{MAE}(T') = \frac{1}{T - T'} \sum_{t=T'+1}^T \text{AE}(T', t). \quad (13)$$

4.2.3. Quantile-based

We use the same ensembles to form quantile forecasts $y_{t,n,i}^{T''}$: representing the q_i th quantile of the ensemble set $\{x_{t,n,s}^{T''}\}_{s=1}^S$. Here, $\{q_i\}_{i=1}^Q \subset [0, 1]^Q$ is the collection of Q quantile levels with $q_i < q_{i+1}$, $\forall i$. After obtaining the quantile forecasts for each test day, the pinball loss function is used as an approximation of the well-known continuous ranked probability score (CRPS): [18]

$$\zeta(T', t) = \frac{2}{N} \sum_{n,i} \max(q_i(x_{t,n}^{T''} - y_{t,n,i}^{T''}), (1 - q_i)(y_{t,n,i}^{T''} - x_{t,n}^{T''})), \quad (14)$$

$$\text{CRPS}(T') = \frac{1}{T - T'} \sum_{t=T'+1}^T \zeta(T', t). \quad (15)$$

4.2.4. Point-based

Lastly, we extract the expected values of the forecasting distributions as the point forecast: $\bar{\mathbf{x}}_n^{T''} = \mathbb{E}_{\tilde{p}_\psi(\mathbf{x}_n^{T''} | \mathbf{x}_n^{T'}, \mathbf{c}_n)}[\mathbf{x}_n^{T''}] = \sum_k \gamma_{k,n}^{T'} \mu_{k,n}^{T''} |_{t=T'}$. We assess the performance of these point forecasts via the root mean squared error (RMSE)

$$\text{RMSE}(T') = \frac{1}{N} \sum_n \sqrt{\frac{1}{T - T'} \sum_{t=T'+1}^T (x_{t,n}^{T''} - \bar{x}_{t,n}^{T''})^2}. \quad (16)$$

The progressive nature of the updating mechanism within the day makes it cumbersome to summarize the performance with a single

² In this study, we set $S = K$.

number. Instead, we evaluate *performance traces* of the metrics (11), (12), (15), and (16) as a function of the update time T' . Another way to conceptualize these traces is by considering the single-step performance scores for each combination (t, T') of forecast and update time during the day, as shown in Fig. 2. For the MAE and CRPS metrics, the performance trace essentially averages these scores over the remaining forecast time.

4.3. Baselines

The performance of the proposed intraday forecaster is heavily tied to the modelling accuracy of the day-ahead forecasting distribution. To isolate the effects of the updating mechanism on the day-ahead forecaster, we define a baseline model and a synthetic test set. Note that the purpose of this baseline is to isolate the effect of the Bayesian updating operator while keeping the day-ahead distribution fixed, rather than to compare performance with alternative retraining-based forecasting pipelines.

The baseline model ignores the information provided by the observations at $t = 1, \dots, T'$. It is defined as the non-updated (NU) marginal of the day-ahead distribution starting from time T' up to the end of the day:

$$\begin{aligned} \tilde{p}_{\psi}^{\text{NU}}(\mathbf{x}^{T'} : |\mathbf{x}^{1:T'}, \mathbf{c}) &:= \tilde{p}_{\psi}(\mathbf{x}^{T'} : |\mathbf{c}) \\ &= \frac{1}{K} \sum_k \mathcal{N}(\mathbf{x}_{T'}^{T'}; \mu = \mu_k^{T'}, \Sigma = \Sigma_k^{T'}). \end{aligned} \quad (17)$$

Note that evaluations using real-world test data also depend on the day-ahead distribution's modelling performance, which can be conflated with the performance of the forecast update process. Thus, we propose a synthetic test set $\mathcal{X}^{\text{Synth}} = \{(\hat{\mathbf{x}}_n, \mathbf{c}_n)\}_{n=1}^N$ consisting of random samples from the trained model $\hat{\mathbf{x}}_n \sim p_{\psi}(\mathbf{x}_n | \mathbf{c}_n)$. Since it is guaranteed that the sampled data points in this generated test set are “in-distribution”, we ensure that the performance metrics will not be affected by overfitting and the comparison between the two models will reflect only the sample approximation and update effects.

4.4. Tuning the number of GMM components

The final crucial step in our experimentation is to determine the appropriate number of GMM components, K . Theoretically, using more components results in a better approximation of the CVAE model (see (2)). However, as K increases, so does the computational burden, while the marginal gain diminishes. To find a suitable trade-off, we propose a tuning method that does not rely on any (real) test data.

Our tuning method requires a *best-case* synthetic test set $\mathcal{X}^{\text{Best-case}} = \{(\hat{\mathbf{x}}_n^0, \mathbf{c}_n)\}_{n=1}^N$ that consists of samples from the established day-ahead forecasting distribution in (2), i.e. $\hat{\mathbf{x}}_n^0 \sim \tilde{p}_{\psi}(\mathbf{x} | \mathbf{c}_n)$. Note the difference to the $\mathcal{X}^{\text{Synth}}$, which requires a new set of latent samples since the samples must come from the CVAE distribution in (2) that has no closed-form expression. Instead, $\mathcal{X}^{\text{Best-case}}$ utilizes the same set of latent samples as the day-ahead forecaster approximation and guarantees that the test samples are not only in-distribution w.r.t. CVAE, but also the day-ahead GMM forecaster. Intuitively, we supply such samples to the forecaster that the update mechanism can correctly converge to the component k' that the sample of interest came from by assigning a higher weight $\gamma_{k'}^{T'}$ as time proceeds and observations accumulate.

After obtaining the two test sets for varying K , we extract the likelihood-based performance traces for each as $\text{NLL}^{\text{Synth}, K}(T')$ and $\text{NLL}^{\text{Best-case}, K}(T')$ which are obtained by replacing $\mathbf{x}_n^{T'} : (\mathbf{x}_n^{1:T'})$ by $\hat{\mathbf{x}}_n^{T'} : (\hat{\mathbf{x}}_n^{1:T'})$ and $\hat{\mathbf{x}}_n^{0:T'} : (\hat{\mathbf{x}}_n^{0:1})$, respectively. Then, we search for the K that minimizes the average L_1 distance between two traces as

$$K^* = \min_K \frac{1}{T-1} \sum_{T'=1}^{T-1} |\text{NLL}^{\text{Best-case}, K}(T') - \text{NLL}^{\text{Synth}, K}(T')|. \quad (18)$$

Intuitively, this optimal K^* results in an approximate distribution $\tilde{p}_{\psi}(\mathbf{x} | \mathbf{c}_n)$ that covers the distribution $p_{\psi}(\mathbf{x} | \mathbf{c}_n)$ well enough that the model performs similarly with regard to forecast updates.

4.5. Implementation details

The CVAE encoder and decoder were implemented as fully connected feed-forward networks with three hidden layers of 1000 neurons each, using ReLU activations. Dropout with probability 0.2 and batch normalisation were applied after each hidden layer. The latent dimensions were set equal to the profile length $L = T$. The pattern dictionary size in the PDCC covariance model is chosen heuristically as $V = 100$ for the Gipuzkoa SM dataset and $V = 200$ for the Utrecht PV dataset. The models were trained using the Adam optimiser with default parameters in PyTorch 2.4.0 and a batch size of 1024. An early stopping mechanism and a learning rate scheduler were employed based on validation performance. Further implementation details can be found in [14].

5. Results and discussion

We begin by presenting the component tuning results for the Gipuzkoa SM and Utrecht PV datasets, shown in Fig. 3a and b, respectively. The tuning procedure suggests employing $K = 500$ GMM components for the Gipuzkoa SM dataset and $K = 20$ for the Utrecht PV dataset. All subsequent results are based on these settings.

The first general observation from Figs. 4–7 is that the Bayesian updating mechanism consistently outperforms the baseline non-updated forecasts across all evaluation metrics. As expected, incorporating observed measurements improves predictive accuracy for the remaining horizon. However, the performance gain is not uniform across update times T' .

This can be examined more closely in Fig. 2, which shows that improvements concentrate along the diagonal (near-term improvements), reflecting strong correlations between consecutive time steps. The most recent observations provide the greatest benefit for subsequent predictions, while later hours ($t \gg T'$) are less affected. Localized regions of improvement ($t \in [2, 6]$ and $t \in [10, 18]$) further indicate that the effectiveness of updates depends also on temporal correlation structures between unobserved steps. Specifically, for the Gipuzkoa SM dataset, the timings of these localized regions are sensible, as the outside regions correspond to the times of leaving and returning home, whereas the preceding measurements generally convey little information regarding those periods. Lastly, note the difference between the absolute and percentage improvements in Fig. 2(c) and (d), respectively. The discrepancy between them arises from the fact that some regions are already predicted well in the day-ahead. For instance, the $t \in [2, 6]$ period for the Gipuzkoa SM dataset is relatively easy to forecast, as household consumption profiles tend to have less uncertainty during early morning periods. Similarly, for the Utrecht PV dataset, Figs. 5 and 7 suggest that the biggest improvements stem from updating around noon. This is mainly because early-morning measurements are hardly informative for refining the forecast.

We also observe the effect of employing full covariance matrices for modelling via PDCC against the conventional diagonal structure in Figs. 4 and 5. For both datasets, it can be seen that PDCC outperforms its diagonal counterpart both in terms of intraday forecast performance (the position of the performance trace with update) and update performance improvement (the difference between the performance traces with and without update). This showcases the higher predictive power that comes with PDCC, as well as its better suitability for the update mechanism, thanks to the explicit modelling of complex temporal correlations. We observed that this relative performance also holds for real data (data not shown).

Finally, we compare performance over real and synthetic test sets in Figs. 6 and 7. For the Gipuzkoa SM dataset, performance traces are

largely consistent with each other, except for the NLL metric, which suggests that the real test set contains more typical, and therefore easier-to-predict, samples. For the Utrecht PV dataset, on the other hand, the model performs better on the synthetic test set, as expected. Despite this discrepancy, we observe that the update mechanism still provides a performance gain on the real test set in all metrics. The comparatively modest improvement observed for real PV data is likely attributable to the dominant influence of exogenous weather forecast uncertainty and rapid cloud dynamics, which limit the predictive value of early-day measurements for later hours. In addition, the day-ahead model already captures much of the deterministic diurnal solar generation pattern, leaving less room for improvement through conditioning. Consequently, intraday updates yield smaller but still consistent gains across metrics.

6. Conclusion

This work proposes a Bayesian updating mechanism to transform day-ahead probabilistic forecasts into intraday forecasts without requiring retraining or re-inference. By conditioning the Gaussian mixture distribution of a CVAE-based day-ahead forecaster on newly observed measurements, the method produces an updated distribution for the remaining hours that preserves its full probabilistic structure. This modelling enables point, quantile, and ensemble forecasts, while being computationally efficient and directly compatible with existing forecasting pipelines.

The experimental results on both electricity consumption and PV generation datasets confirm that the updating mechanism consistently improves forecast accuracy across likelihood-, sample-, quantile-, and point-based metrics. The improvements are most noticeable for time steps with strong temporal correlation to the observed data, including correlations that extend over longer ranges. This highlights the method's ability to effectively incorporate real-time information. Moreover, the use of pattern dictionary-based covariance structures enhances both the baseline modelling capability and the quality of intraday updates, indicating that intraday conditioning directly benefits from more accurate joint covariance estimation.

Overall, the proposed approach provides a practical and theoretically grounded solution for intraday forecasting in power systems, offering improved adaptability while maintaining probabilistic consistency. Future work will focus on extending the framework with richer exogenous variables, such as enhanced weather forecasts, and exploring its integration into real-time market operations and system balancing tasks.

CRediT authorship contribution statement

Kutay Bölät: Writing – original draft, Visualization, Software, Resources, Methodology, Investigation, Formal analysis, Data curation, Conceptualization; **Peter Palensky:** Writing – review & editing, Supervision, Funding acquisition; **Simon H. Tindemans:** Writing – review & editing, Visualization, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

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

Data availability

Data will be made available on request.

References

- [1] S. Kolkman, L. Ostmeier, C. Weber, Modeling multivariate intraday forecast update processes for wind power, *Energy Econ.* 139 (2024) 107890.
- [2] D. Nickelsen, G. Müller, Bayesian hierarchical probabilistic forecasting of intraday electricity prices, *Appl. Energy* 380 (2025) 124975.
- [3] L.R. Visser, T.A. AlSkaif, A. Khurram, J. Kleissl, W. van Sark, Probabilistic solar power forecasting: an economic and technical evaluation of an optimal market bidding strategy, *Appl. Energy* 370 (2024) 123573.
- [4] L. Zhang, Y. Chu, Y. Xu, W. Guo, Quantitative assessment and optimization strategy of flexibility supply and demand based on renewable energy high-penetration power system, *Energy Inf.* 7 (1) (2024) 117.
- [5] Q. Wang, A. Tuohy, M. Ortega-Vazquez, M. Bello, E. Ela, D. Kirk-Davidoff, W.B. Hobbs, D.J. Ault, R. Philbrick, Quantifying the value of probabilistic forecasting for power system operation planning, *Appl. Energy* 343 (2023) 121254.
- [6] N. Costilla-Enriquez, M.A. Ortega-Vazquez, A. Tuohy, A. Motley, R. Webb, Operating dynamic reserve dimensioning using probabilistic forecasts, *IEEE Trans. Power Syst.* 38 (1) (2022) 603–616.
- [7] H. Zhang, R. Zandehshahvar, M. Tanneau, P. Van Hentenryck, Weather-informed probabilistic forecasting and scenario generation in power systems, *Appl. Energy* 384 (2025) 125369.
- [8] K. Bölät, S. Tindemans, One model to forecast them all and in entity distributions bind them, *arXiv:2501.15499*. (2025).
- [9] P.F. Austnes, C. García-Pareja, F. Nobile, M. Paolone, Probabilistic load forecasting of distribution power systems based on empirical copulas, *Sustain. Energy Grids Netw.* 42 (2025) 101708.
- [10] O. Doelle, N. Klinkenberg, A. Amthor, C. Ament, Probabilistic intraday PV power forecast using ensembles of deep gaussian mixture density networks, *Energies* 16 (2) (2023) 646.
- [11] J. Thaker, R. Höller, Hybrid model for intra-day probabilistic PV power forecast, *Renew. Energy* 232 (2024) 121057.
- [12] B. Xiong, Y. Chen, D. Chen, J. Fu, D. Zhang, Deep probabilistic solar power forecasting with transformer and Gaussian process approximation, *Appl. Energy* 382 (2025) 125294.
- [13] K. Sohn, H. Lee, X. Yan, Learning structured output representation using deep conditional generative models, in: Cortes, C., Lawrence, N., Lee, D., Sugiyama, M., Garnett, R., (Eds.), *Advances in Neural Information Processing Systems*, Curran Associates, Inc. Vol. 28 (2015). https://proceedings.neurips.cc/paper_files/paper/2015/file/8d55a249e6baa5c06772297520da2051-Paper.pdf
- [14] K. Bölät, S. Tindemans, GUIDE-VAE: advancing data generation with user information and pattern dictionaries, *arXiv:2411.03936*. (2024).
- [15] T.J. Page, Jr, Multivariate statistics: a vector space approach, *J. Mark. Res.* 21 (2) (1984) 236.
- [16] C. Quesada, L. Astigarraga, C. Merveille, C.E. Borges, An electricity smart meter dataset of Spanish households: insights into consumption patterns, *Sci. Data* 11 (1) (2024) 59.
- [17] L.R. Visser, B. Elsinga, T.A. AlSkaif, W.G. Van Sark, Open-source quality control routine and multi-year power generation data of 175 PV systems, *J. Renew. Sustain. Energy* 14 (4) (2022). <https://doi.org/10.1063/5.0100939>
- [18] R. Fakoor, T. Kim, J. Mueller, A.J. Smola, R.J. Tibshirani, Flexible model aggregation for quantile regression, *J. Mach. Learn. Res.* 24 (162) (2023) 1–45.