Copula-based Probabilistic Prediction of Grid Frequency Dynamics

Bolin Liu*
Karlsruhe Institute of Technology (KIT)
Institute of Operations Research, Analytics and Statistics
Karlsruhe, Germany
bolin.liu@kit.edu

Maximilian Koblenz
Ludwigshafen University of Business and Society
Department of Services and Consulting
Ludwigshafen, Germany
maximilian.coblenz@hwg-lu.de

Oliver Grothe
Karlsruhe Institute of Technology (KIT)
Institute of Operations Research, Analytics and Statistics
Karlsruhe, Germany
grothe@kit.edu

Abstract

With the growing share of renewable energy sources in the electricity supply, monitoring the grid stability has become increasingly important. A major challenge is the modeling and prediction of grid frequency as a stability indicator based on external techno-economic features recorded on an hourly basis.

Common models in the literature are based on the assumption of Gaussian distributions in which either no dependency between the points in time is modeled or, if a dependency is taken into account, it is only captured linearly. We present a data-driven approach to modeling grid frequency that constructs a copula-based probabilistic predictor from an existing point predictor that is able to account for the nonlinear dependence between time points. The construction is based on the probabilistic correction of the point predictors by feature space driven error estimation. Models that are corrected using the error distribution perform better on probabilistic evaluation measures than baselines that assume independence. In addition, our best copula-based model performs also better than a Gaussian prediction model that takes dependence by correlation into account. Our simulation study also shows that the conventional models underestimate the uncertainty of the grid frequency evolution and our copula-based predictors can provide a more accurate representation of this uncertainty.

CCS Concepts

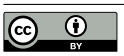
• **Applied computing** → **Forecasting**.

Keywords

data driven modeling, electric power system, power-grid frequency

ACM Reference Format:
Bolin Liu, Maximilian Koblenz, and Oliver Grothe. 2025. Copula-based Probabilistic Prediction of Grid Frequency Dynamics. In *The 16th ACM International Conference on Future and Sustainable Energy Systems (E-ENERGY '25), June 17–20, 2025, Rotterdam, Netherlands.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3679240.3734641

*Corresponding author



E-ENERGY '25, Rotterdam, Netherlands

ACM ISBN 979-8-4007-1125-1/25/06
https://doi.org/10.1145/3679240.3734641

1 Introduction

Power grid frequency reflects the balance between power generation and consumption in a transmission network and characterizes the stability and security of a power grid. Deviations of the grid frequency from the nominal frequency reflect imbalances between power generation and consumption, which can lead to overloads, shutdowns, or, in the worst case, impair critical infrastructures [8]. In the course of integrating renewable energy sources, monitoring and controlling the electricity grid frequency has become a challenge due to the volatile nature of renewable energies [16, 19].

A prediction model of grid frequency can detect grid instability situations early, and the forecast information can be used to efficiently deploy control power to balance grid fluctuations in a timely manner and prevent bottlenecks and failures. A particularly interesting prediction task in this context is to forecast the dynamic development of the grid frequency based on information available at the beginning of an hourly interval [10, 11]. Various approaches exist in the literature. For example, a physics-informed machine learning model has been developed that builds upon physically meaningful model equations which take into account the influence of operating conditions [10]. In another approach, a purely data-driven methodology is presented, where the model architecture is constructed without domain knowledge and the model learns directly from the data [14].

Although the above approaches can effectively model many aspects, they are fundamentally based on the assumption of Gaussian processes, which either assume independent points in time within the hourly interval to be predicted, or can only account for linear dependencies through correlation.

In this paper, we address this problem by transforming an existing point predictor of grid frequency based on historical data into a probabilistic estimator capable of recognizing and flexibly modeling serial dependencies in the power grid frequency dynamics. In particular, we analyze the forecast errors generated by the point estimator on historical data by considering different subspaces of features defined based on the observed errors to identify different error states. The subspace-based decomposition allows us to capture localized error patterns that may be missed in a global analysis. We model the error distribution within each subspace using nonparametric copula estimators. Our methodology is motivated by [7, 22] , and the underlying idea has been successfully applied to day-ahead electricity price forecasting [7]. Here, we extend this approach by incorporating feature-based error correction within the identified subspaces. We demonstrate our methodology

using historical data from the European continental grid and show that by accounting for dependence structures between the time points, more potent multivariate estimators can be constructed. Furthermore, we demonstrate that failing to consider serial dependencies significantly underestimates the uncertainty in power grid frequency forecasts.

The paper is structured as follows. Section 2 introduces the probabilistic estimator construction for power grid frequency through feature-based error correction and the associated sampling process. In Section 3, we conduct a study to implement the methodology for modeling the grid frequency dynamics in continental Europe and evaluate the performance using the energy score and the marginal average continuous ranked probability score. In particular, we demonstrate that the performance of our copula based probablitic forecasting models is significantly better when predicting the average grid frequency deviation. The paper ends with a conclusion. The source code of this work is available at [15].

2 Methodology

Let $\hat{\mathbf{f}} : \mathcal{X} \to \mathbb{R}^d$ be an existing point predictor for the evolution of the grid frequency over d time points within a one-hour interval, i.e., it outputs a deterministic prediction of the grid frequency trajectory based on input features associated with the respective one-hour interval. Since time is considered in discrete steps, the frequency evolution can be represented as a d-dimensional grid frequency vector. Such a point predictor can, e.g., be derived from a probabilistic model by using the expected value function of the Gaussian distribution-based approaches, as employed in the models of [10, 14].

We denote by $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{f}_i) \mid \mathbf{x}_i \in \mathcal{X},\ \mathbf{f}_i \in \mathbb{R}^d,\ i = 1, \ldots, N\}$ a dataset consisting of N sample points, where each $\mathbf{x}_i$ is a historical feature vector and $\mathbf{f}_i$ is the corresponding observed true frequency vector. For instance, the training data used to train the point predictor $\hat{\mathbf{f}}$ can be utilized as $\mathcal{D}$. Let $\hat{\mathbf{f}}_i \in \mathbb{R}^d$ denote the prediction of the point predictor $\hat{\mathbf{f}}$ at x_i, for $i = 1, \ldots, N$, where $x_i \in \mathcal{D}$.

The methodology presented below aims to develop a prediction model based on the point estimator $\hat{\mathbf{f}}$ and the dataset $\mathcal{D}$. The model is designed to provide a probabilistic prediction of the grid frequency vector for a given hour, represented as a d-dimensional conditional distribution F given a feature vector x.

For this purpose, the feature space $\mathcal{X}$ is partitioned into meaningful feature subspaces derived from the clustering results of the observed feature vectors in the historical dataset. This allows for a more fine-grained analysis of the point predictor's error behavior. The prediction errors observed in $\mathcal{D}$ are analyzed independently across these clusters, and we estimate a separate error distribution for each subspace. The point predictor is then corrected by the error distributions to obtain a probabilistic predictor that can also map non-linear serial dependencies between different points in time in a predicted hour. Our methodology is motivated by [7] and extends the approach by feature space based prediction error considerations. In contrast to the approach in [7], where the error distribution is considered over the entire historical data set, here we attempt to consider the error distribution over different subspaces of the entire feature space to account for variation in error behavior as a function of features. In addition, we propose the use of the empirical Bernstein copula for modeling the dependence of different time points. The complete construction procedure is given in the Algorithm 1.

Algorithm 1 Copula-based probabilistic grid frequency predictor construction

Require: Point predictor for grid frequency $\hat{\mathbf{f}}$, Dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{f}_i), \mathbf{x}_i \in \mathcal{X}, \mathbf{f}_i \in \mathbb{R}^d, i = 1, \ldots, N\}$

Ensure: Probabilistic grid frequency predictor $\mathbf{f}_p$ with distribution P

1: Compute the error statistics from forecasting error time series.

$$\epsilon_i = \hat{\mathbf{f}}(\mathbf{x}_i) - \mathbf{f}_i, i = 1, \cdots, N.$$

2: Compute the mutual information between all error statistics and the feature space.
3: Compute the normalized vectors of the average feature importance by using the computed mutual information.
4: Use the average feature importance vector to perform a k-means method with the scaled distance function to cluster the feature vectors in $\mathcal{D}$ into the clusters $\{\mathcal{D}_1, \cdots, \mathcal{D}_K\}$ and decompose the feature space into K subspaces $\{\mathcal{X}_1, \cdots, \mathcal{X}_K\}$.
5: **for** $k \leftarrow 1$ to K **do**
6: Compute the copula structure $C^{(k)}$ between the d time points using the empirical Bernstein copula. For the copula estimation, the forecast error vectors of the feature vectors in $\mathcal{D}_k$ are used.
7: **for** $j \leftarrow 1$ to d **do**
8: Compute the empirical marginal distribution of the frequency values (in the cluster $\mathcal{D}_k$) with the time point index j: $\hat{F}_j^{(k)}$.
9: **end for**
10: Build the probabilistic predictor P_k on the feature subspace $\mathcal{X}_k$ with df

$$\hat{F}^{(k)}(f_1, \ldots, f_d \mid x) = C^{(k)}\left(\hat{F}_1^{(k)}(f_1 - \hat{f}_1(x)), \ldots, \hat{F}_d^{(k)}(f_d - \hat{f}_d(x))\right)$$

11: **end for**
12: $P = \{P_k, k = 1, ..., K\}$.
13: **return** P

2.1 Feature Space Decomposition

The distribution of the prediction error of the point estimator depends on the features that represent the conditions for the prediction. To account for this feature dependency in the subsequent error correction, the dataset $\mathcal{D}$ is first clustered according to features to obtain different prediction states. It should be noted that the feature space for the construction of the probabilistic predictor can generally be broader or different from the space of features originally used to train the point estimator. This has the advantage that we are able to extend the point predictor to a probabilistic predictor in aspects that are interesting for certain applications.

Additionally, different predictors often use varying amounts of information, meaning they operate on different datasets. A point estimator may initially be trained on a limited feature space, utilizing only a restricted amount of information. Later, it is possible to correct or adjust this estimator by including additional information.

By expanding the data set, the estimator can be recalibrated in this sense, which can then provide better predictions

Multiple features are often used to predict grid frequency. For example, the data-driven model in [14] uses 14 techno-economic features and the ex-post physics-informed machine learning model in [10] uses 51 features.

To identify different feature subspaces for which we want to determine separate prediction error distributions of the point predictors, we use a feature-weighted k-means clustering approach[2, 26]. This approach enhances clustering accuracy by giving higher weights to relevant features while reducing the influence of irrelevant or noisy features [4].

To this end, we systematically analyze the relationship between individual features and prediction errors. For each observed feature point in $\mathcal{D}$, the point predictor is evaluated to obtain the corresponding predicted grid frequency vector, which is then compared to the observed (true) grid frequency vector within the same hourly interval to calculate the prediction error values. This sequence of prediction errors over the associated interval can be represented as a d dimensional error vector

$$\epsilon_i = \hat{\mathbf{f}}_i - \mathbf{f}_i, i = 1, \cdots, N. \tag{1}$$

From a prediction error vector ϵ_i, we can extract various characteristic statistics such as mean, standard deviation, maximum, minimum, and autocorrelation values. To quantify the strength of association between features and these error statistics, we employ mutual information, which measures the amount of shared information between two random variables and is particularly well-suited for detecting nonlinear dependencies [3]. Higher mutual information between an error statistic and a feature indicates that the feature strongly influences this error statistic.

Mutual information can be empirically estimated from the feature and error statistics data [9, 17]. After calculating the mutual information between individual features and a specific error statistic, these mutual information values can be normalized so that their sum equals 1. We perform this procedure for all selected error statistics. For each error statistic, we obtain a mutual information vector, where each component represents the mutual information between the error statistic and a specific feature. To derive a summary feature importance vector, we compute the mean of all mutual information vectors across the different error statistics. We refer to this aggregated vector as the feature importance vector $\boldsymbol{\lambda}$.

Here, the standard k-means distance function is then modified with $\boldsymbol{\lambda}$ to be able to incorporate the varying relevance of features for prediction errors

$$d_{\boldsymbol{\lambda}}(x, y) = \sqrt{\sum_{i=1}^{d} \lambda_i (x_i - y_i)^2}, x, y \in \mathcal{X}. \tag{2}$$

It should be noted that features are already standardized prior to weighting to avoid scale effects.

To determine the optimum number of clusters, the silhouette coefficient can be optimized (for details see [18]).

Assume that the historical feature vectors in the dataset $\mathcal{D}$ are partitioned into disjoint clusters $\mathcal{D}_k$, for $k = 1, \ldots, K$, through the clustering process described above. Each cluster $\mathcal{D}_k$ induces a corresponding feature subspace $\mathcal{X}_k \subset \mathcal{X}$. A feature vector $x \in \mathcal{X}$ is assigned to the feature subspace $\mathcal{X}_k$ if the distance between x and the center z_k of cluster $\mathcal{D}_k$ is minimal among all K clusters, i.e.,

$$k = \arg \min_{1 \le j \le K} d_{\boldsymbol{\lambda}}(x, z_j). \tag{3}$$

2.2 Multivariate Frequency Distribution via Copula Model

Following the procedure described in Section 2.1, the clusters $\mathcal{D}_k$ and the feature subspaces $\mathcal{X}_k, k = 1, ..., K$ are obtained.

For each $\mathcal{D}_k$, we consider the associated set of observed prediction error vectors, defined as

$$\mathcal{E}_k := \left\{ \epsilon_i \in \mathbb{R}^d \;\middle|\; x_i \in \mathcal{D}_k \right\}, \quad k = 1, \ldots, K. \tag{4}$$

For each of these subsets of prediction error vectors $\mathcal{E}_k$, we estimate a multivariate probability distribution P_k that captures the stochastic characteristics of the errors within the respective cluster. A key aspect of this process is modeling the dependency structure of the prediction error vector.

Since the prediction error vector represents the temporal progression of errors within a one-hour interval, the estimated dependency structure reflects how error values at different time points within that interval are interrelated. To flexibly capture these temporal dependencies and represent the hourly stochastic dynamics within each cluster, we employ the copula framework.

While a Gaussian distribution can only model linear dependencies through correlation, a copula model allows capturing complex, nonlinear dependency structures. In general, any multivariate distribution function can be decomposed into a dependence structure, represented by a copula C, and its marginal distributions F_i, as stated by Sklar's theorem [24]:

$$F(z_1, \ldots, z_d) = C\left(F_1(z_1), \ldots, F_d(z_d)\right). \tag{5}$$

If all marginal distributions $F_i, i = 1, ..., d$ are continuous, the copula is unique.

To estimate a copula for the prediction error vector in the feature subspace $\mathcal{X}_k$, the residual data $\mathcal{E}_k$ can be used. Following the approach proposed in [6], we first compute the so-called pseudo-observations and then estimate a copula from these data. This can be done either using a parametric model, such as the Gaussian copula, if one wishes to model a linear dependence between the frequencies at different time points. However, without knowing the underlying dependency structure, we have to estimate the frequency dependence with a nonparametric copula.

Let the elements of $\mathcal{E}_k$ be indexed as $\mathbf{x}_i^{(k)} \in \mathbb{R}^d$, for $i = 1, \ldots, n_k$. The pseudo-observations are defined as

$$\hat{U}_{ij}^{(k)} = \frac{R_{ij}^{(k)}}{n_k + 1}, \tag{6}$$

where $R_{ij}^{(k)}$ denotes the rank of the j-th component of $\mathbf{x}_i^{(k)}$ among the values $\mathbf{x}_{1j}^{(k)}, \mathbf{x}_{2j}^{(k)}, \ldots, \mathbf{x}_{n_k j}^{(k)}$. In the following, we omit the index k for clarity in the representation of the copula formula.

A simple non-parametric estimation method is to use the empirical copula

$$C_n(u_1, \ldots, u_d) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left(\hat{U}_{i1} \leq u_1, \ldots, \hat{U}_{id} \leq u_d\right). \tag{7}$$

In the literature, several approaches exist for nonparametric copula estimation. In this work, we employ the empirical Bernstein copula with a smoothing degree m, typically an integer between 1 and n. It is defined as

$$C_n^B(u_1, \ldots, u_d) = \sum_{v_1=0}^{m} \cdots \sum_{v_d=0}^{m} \prod_{j=1}^{d} \binom{m}{v_j} u_j^{v_j} (1-u_j)^{m-v_j} C_n\left(\frac{v_1}{m}, \ldots, \frac{v_d}{m}\right), \tag{8}$$

which constitutes a smoothed version of the empirical copula [20, 23]. It generalizes families of polynomial copulas [20] and has already been successfully applied in the literature for various modeling purposes [5, 21, 27].

A total of K copulas $\{C^{(k)},\ k = 1, \ldots, K\}$ are estimated based on $\mathcal{E}_k, k = 1, ..., K$. In addition to the temporal dependency structure $C^{(k)}$, the marginal distributions need also to be determined for each $\mathcal{E}_k$. This refers to the distributions of the prediction error from the point estimator at specific time points within the predicted hour interval. The empirical distribution function (denoted as $\hat{F}_i^{(k)}, i = 1, ..., d$) and the empirical quantile function (denoted as $\hat{Q}_i^{(k)}, i = 1, ..., d$) can be used for this purpose. Note that this also generalizes the Gaussian model, as it allows us to model skewness, kurtosis, and other properties of the marginal distribution.

The learned copula functions and the empirical marginal distributions of individual time points within an hourly interval can be combined into a probabilistic predictor of grid frequency. For a given feature vector $x \in \mathcal{X}_k$ and a point predictor $\hat{\mathbf{f}}$, this *probabilistic predictor* is represented by the conditional multivariate distribution function:

$$\hat{F}^{(k)}(f_1, \ldots, f_d \mid x) = C^{(k)}\left(\hat{F}_1^{(k)}(f_1 - \hat{f}_1(x)), \ldots, \hat{F}_d^{(k)}(f_d - \hat{f}_d(x))\right), \tag{9}$$

where $\hat{f}_i(x)$ denotes the i-th component of the point prediction $\hat{\mathbf{f}}(x)$.

To generate a concrete prediction from the probabilistic predictor for a feature vector $x \in \mathcal{X}_k$, i.e., to sample a data point from the conditional distribution $\hat{F}^{(k)}$, one first samples a point $(\hat{u}_1, \ldots, \hat{u}_d)$ from the learned copula $C^{(k)}$. Then, for each dimension $i = 1, \ldots, d$, the estimated quantile function $\hat{Q}_i^{(k)}$ is applied individually to obtain the residual term

$$\hat{\varepsilon} := \left(\hat{Q}_1^{(k)}(\hat{u}_1), \ldots, \hat{Q}_d^{(k)}(\hat{u}_d)\right). \tag{10}$$

By correcting the point prediction $\hat{\mathbf{f}}(x)$ with $\hat{\varepsilon}$, the final prediction from the probabilistic predictor is given by

$$\hat{\mathbf{f}}_p(x) = \hat{\mathbf{f}}(x) + \hat{\varepsilon}. \tag{11}$$

3 Results

In the following, we demonstrate how a trained point estimator can be transformed into a probabilistic estimator using the methodology described above using real grid frequency data of continental Europe. We show that the prediction performance with respect to the energy score is thereby improved. Furthermore, through a simulation study on the average frequency deviation in an hourly interval, we show that a Gaussian-based frequency prediction model that does not account for dependencies underestimates the uncertainty of the forecast. The developed copula-based model is better suited in this regard.

Type	Feature	Unit
External	Load Day-Ahead	MW
External	Solar Day-Ahead	MW
External	Offshore Wind Day-Ahead	MW
External	Onshore Wind Day-Ahead	MW
External	Load Ramp Day-Ahead	MW/h
External	Generation Ramp Day-Ahead	MW/h
External	Solar Ramp Day-Ahead	MW/h
External	Offshore Wind Ramp Day-Ahead	MW/h
External	Onshore Wind Ramp Day-Ahead	MW/h
External	Price Day-Ahead	euro/MWh
External	Price Ramp Day-Ahead	euro/MWh/h
Time	$\cos(\pi/12\ \text{Hour})$	-
Time	$\sin(\pi/12\ \text{Hour})$	-
Initial Value	Grid Frequency Value at the Beginning of the Hour to Be Forecast	1/s

Table 1: Included Features in Dataset $\mathcal{D}$ (also see [10, 14]).

3.1 Study Setup

We use the same data and setup as in [10, 14]. In particular, we use the historical dataset $\mathcal{D}$ to construct the probabilistic predictors, which contains historical grid frequency data and external feature data for continental Europe from 2015 to 2018. This is the same data used to train the point estimators described below, which serve as baselines in our evaluation. The performance of the probabilistic predictors is then evaluated using data from 2019, which was completely excluded from the training process for both the probabilistic models and the baseline point estimators, and thus serves as a test set.

The external techno-economic features are recorded hourly, while the associated grid frequency data are available as time series at different timestamps within the corresponding hourly interval. In this work, we analyze minute-by-minute predictions of frequency values within a given hourly interval. The output of a point predictor is a 60-dimensional grid frequency vector, whereas the output of a probabilistic predictor is a 60-dimensional multivariate distribution. An overview of the techno-economic features can be found in Table 1. Details on the collection and preprocessing of the data can be found in [12, 13].

We use selected predictors from [10, 14] as the basis for the point estimation. Table 2 provides an overview of the different point predictors and their properties. While the KNN model from [14] directly provides a point prediction (i.e., a predicted frequency vector for a given hourly interval), the other models are probabilistic in nature. They are based on Gaussian distributions and model the frequencies at several discrete time points within an hourly interval

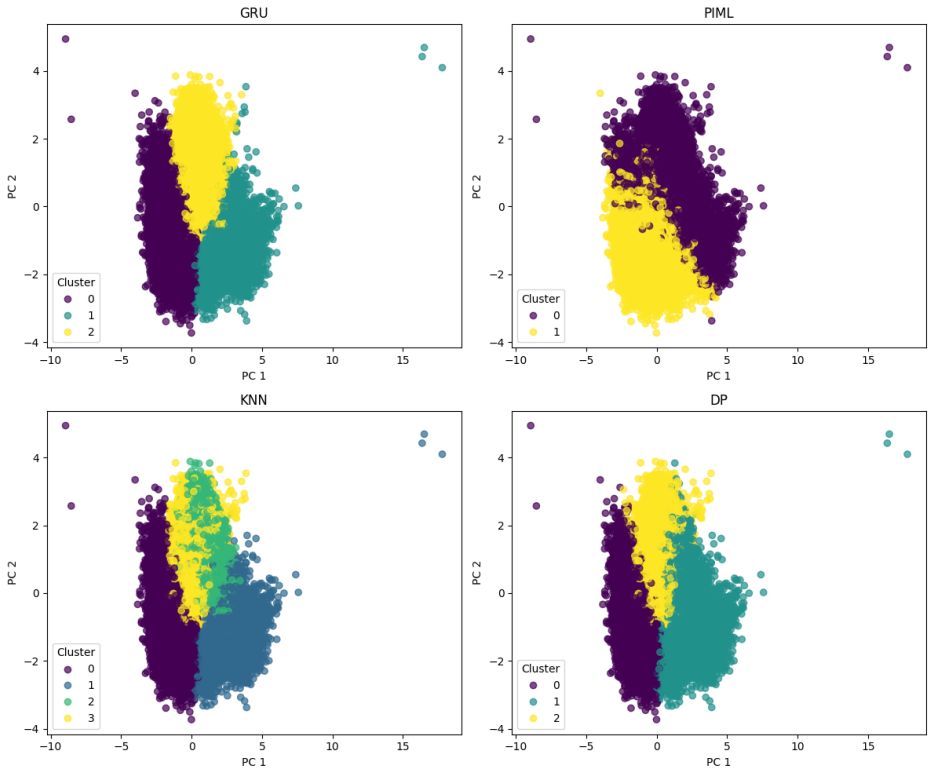


Figure 1: Error based feature space clustering results of different point predictors. The 2D coordinates obtained from the PCA transformation of the clustered points are shown.

either as independent or correlated. Given a set of input features, they predict a mean frequency vector along with either a vector of standard deviations (assuming independence between time points) or a full covariance matrix (in the case of correlated time points). To derive point predictions from these probabilistic models, we use the predicted mean vector as a deterministic estimate.

Point Predictor	Description	Features Used for Training
GRU Independent Gaussian	The mean frequency value prediction of the independent Gaussian process model based on GRU	Same as in $\mathcal{D}$
Day-Ahead PIML	The mean frequency value prediction of the day-ahead physics-informed machine learning model	Same as in $\mathcal{D}$
KNN	Prediction based on the feature distance of historical data	Same as in $\mathcal{D}$
Daily Profile	The mean of all training data at this time of day	Only temporal features (time of day)

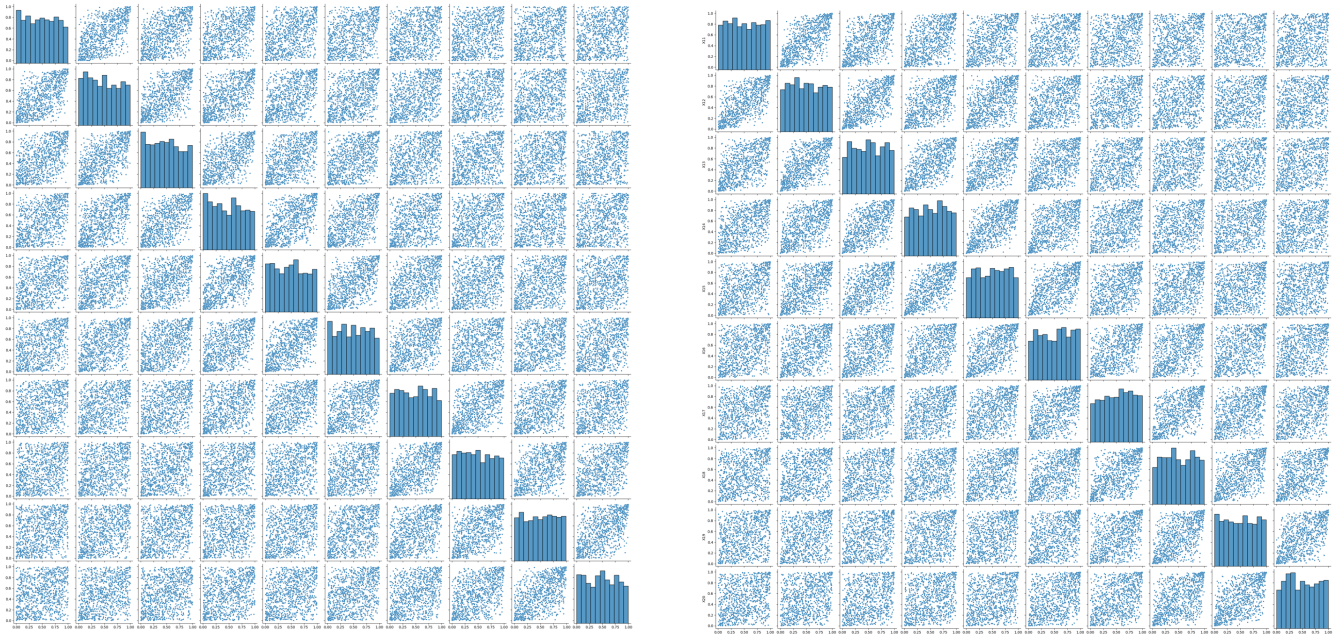
Table 2: Overview of point predictors and their feature usage.

3.2 Clustering and Copula Results

For each considered predictor, we perform Feature-Weighted K-Means Clustering. The corresponding feature importance weights are first determined and then used in the respective clustering process. To compute the error statistics within a one-hour interval, we analyze the time series of prediction errors. We consider several metrics: First, we calculate the mean and standard deviation of the errors to quantify the average error level and its variability. Second, we assess the autocorrelation of the error time series to capture the temporal dependency structure within the interval. Specifically, we focus on the average autocorrelation over the first 15 lags (equivalent to 15 minutes).

We then compute the mutual information between each dimension of the feature space and the error statistics on $\mathcal{D}$ and then, as described in Section 2.1, the feature importance vector as the weights for the distance function. A clear dependency can be observed between temporal features and error behavior across all the models considered. Interestingly, while the GRU-based baseline and the physics-informed machine learning model exhibit similar performance [14], their prediction errors are based on different features, apart from the temporal ones (see Table 3).

Now, the feature weighted k-means algorithm can be performed using the calculated feature importance weights. The optimal number of clusters is determined by minimizing the silhouette coefficient. Figure 1 illustrates the clustering results concerning different



(a) GRU-based Point Estimator

(b) Daily Profile Point Estimator

Figure 2: Pairwise plots of simulation data (1,000 points) generated from the respective learned empirical Bernstein copula of the error distribution within one of the respective clusters. The plots illustrate the pairwise dependencies between time points within a 10-minute interval, from the 10th to the 19th minute (10-dimensional).

Table 3: Average Mutual Information Between Error Statistics and Feature Space

Feature	GRU	PIML	KNN	DP
Solar Day-Ahead	0.0510	0.0597	**0.1283**	**0.0816**
Wind On Day-Ahead	0.0760	0.0375	0.0184	0.0654
Wind Off Day-Ahead	0.0686	**0.1332**	0.0601	0.0106
Prices Day-Ahead	0.0940	**0.1808**	0.0338	0.0436
Load Day-Ahead	0.0447	**0.1538**	0.0732	0.0546
Load Ramp Day-Ahead	**0.1080**	0.0000	0.0714	0.0562
Total Gen Ramp Day-Ahead	0.0770	0.0877	**0.1209**	0.0602
Wind Off Ramp Day-Ahead	0.0159	0.0352	0.0084	0.0141
Wind On Ramp Day-Ahead	0.0000	0.0765	0.0158	0.0000
Solar Ramp Day-Ahead	0.0167	0.0447	0.0710	0.0751
Price Ramp Day-Ahead	**0.1040**	0.0126	0.0887	0.0703
Hour (sin)	**0.1148**	0.0005	**0.1445**	**0.1838**
Hour (cos)	**0.1585**	**0.1069**	**0.1291**	**0.1217**
Initial Value	0.0707	0.0709	0.0362	**0.1627**

error behaviors of the point predictors. To visualize the clustering results, we first performed a principal component analysis (PCA) on the feature space and then plotted the clustering results using the 2D coordinates obtained from the PCA transformation of the clustered points.

With these cluster results, we calculate the respective error distributions for each point predictor for all their clusters, as described in Section 2.2, using the empirical Bernstein copula. For practical estimation, we use the Python package OpenTURNS [1]. For each point predictor, a separate copula function is estimated for each cluster to model the temporal dependence between the deviations from the corresponding point predictions at the 60 time steps within a one-hour interval.

We illustrate some examples of the learned dependency structures. First, we sample 1000 60-dimensional vectors from a trained empirical Bernstein copula. The resulting pair plots visualize the dependency structure between two dimensions, i.e., between two time points within a one-hour interval. Due to space limitations, we only present the learned dependency structure for 10 of the 60 dimensions (see Figure 2a-2b). Here, one can observe that the greater the distance between two time points in an hourly interval, the less dependence is present. Although many pairwise dependencies have a Gaussian copula-like shape, there are many pairwise dependencies that are not similar to a Gaussian copula.

3.3 Prediction Performance

We evaluate the obtained probabilistic predictors in comparison to the original probabilistic predictors, from which the point predictors are derived, using data from 2019 as the test set. The evaluation is based on energy scores and marginal CRPS scores (i.e., the average of the CRPS scores of the marginal predictors). Additionally, we consider a probabilistic GRU-based Gaussian process model with a rational quadratic kernel from [14], which accounts for the

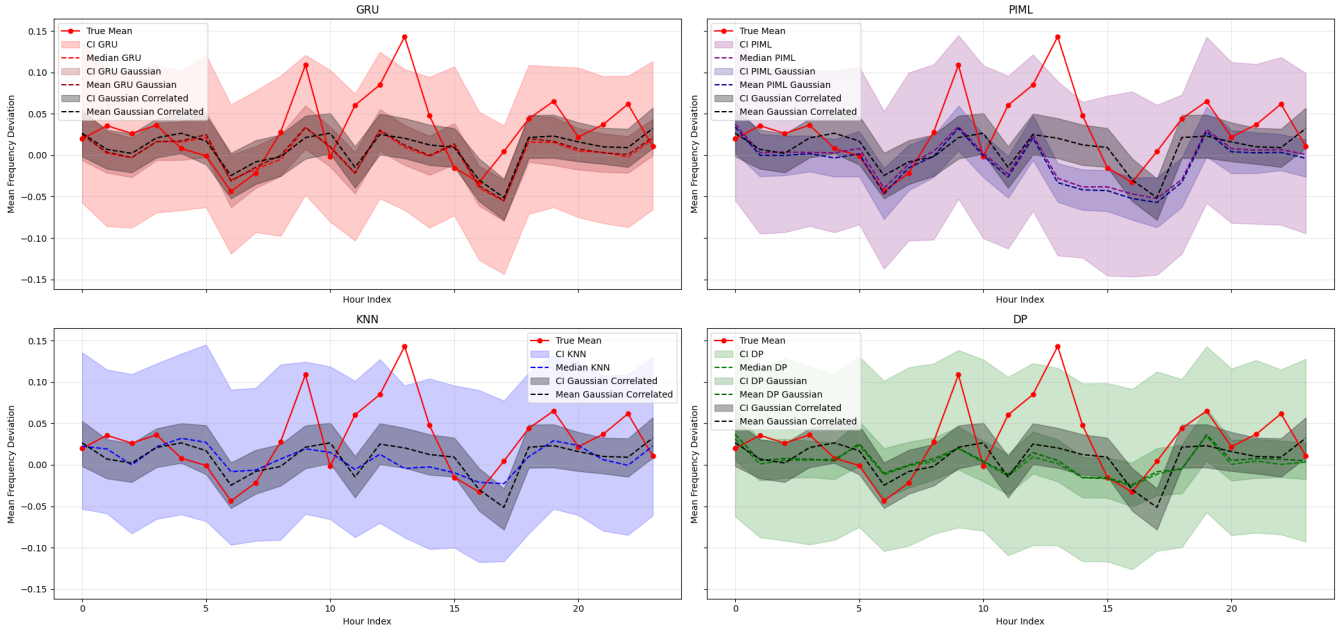


Figure 3: Confidence intervals for the hourly average grid angular frequency deviation. For each hour, 1000 time series are sampled from the respective probabilistic model. From these time series, the mean frequency deviations are calculated and the confidence intervals are derived. The red line indicates the true progression of the hourly average frequency deviation in an example day.

Method	Energy Score	Marginal Avg. CRPS	p-value vs Gaussian Baseline (ES)	p-value vs Corr. Gaussian (ES)	p-value vs Gaussian Baseline (CRPS)	p-value vs Corr. Gaussian (CRPS)
Copula-based						
GRU	**0.4774**	**0.0497**	< 0.0001	< 0.0001	0.0252	< 0.0001
PIML	0.4936	0.0509	< 0.0001	< 0.0001	< 0.0001	< 0.0001
KNN	0.4868	0.0507	–	< 0.0001	–	< 0.0001
DP	0.5108	0.0532	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Independent Gaussian						
GRU	0.4823	0.0499	–	< 0.0001	–	< 0.0001
PIML	0.5001	0.0515	–	< 0.0001	–	< 0.0001
DP	0.5940	0.0612	–	< 0.0001	–	< 0.0001
Correlated Gaussian						
Rational Quadratic	0.4827	0.0503	–	–	–	–

Table 4: Median values of energy score and marginal average CRPS for copula-based, independent Gaussian, and correlated Gaussian models, evaluated on the full test dataset. p-values from the Wilcoxon test for comparisons to Gaussian baselines and the correlated Gaussian model are presented. "–" indicates that no test was performed, either because it would be a comparison with itself or because no baseline is available (as in the case of KNN).

correlation between different time points in the considered hour interval.

We evaluate the scores on the test data using *scoringrules*, a Python library for assessing probabilistic forecasts [28], and present the median values in Table 4. It is evident that the copula-based probabilistic methods improve the energy score of their corresponding baseline models. Furthermore, the result of the copula-GRU model is also better than the considered correlated Gaussian process model. Additionally, Wilcoxon signed-rank tests [25] were conducted to assess the statistical significance of differences between the score values. As shown in Table 4, the differences in

model performance are statistically significant at the 0.01% level for nearly all comparisons (with the exception of the comparison of marginal CRPS values between the copula-based GRU model and the Gaussian GRU model).

3.4 Forecast of Hourly Aggregated Grid Frequency Deviation

To highlight the importance of accurately accounting for dependencies across different time points, we consider the hourly average frequency deviation, which is defined as the average of the frequency deviations across all 60 discrete time points within a one-hour interval. A prediction for this quantity can be directly derived from probabilistic predictors of the grid frequency vector. To do so, a prediction of the grid frequency vector is first generated. Then, the average of the predicted grid frequency values over all time points within the hour is computed, and the deviation of this average from the nominal grid frequency is taken.

To illustrate the performance of different probabilistic models, we use the learned probabilistic predictors to generate synthetic grid frequency vectors for each of the 24 hourly intervals of a given example day in the test dataset, based on the corresponding input features. For each interval, we perform a Monte Carlo simulation by generating 1,000 synthetic grid frequency vectors, resulting in 1,000 corresponding predictions of the average frequency deviation. Based on these simulated predictions, we estimate a confidence interval for the mean frequency deviation for each hourly interval.

The confidence level is set between the quantile level 0.02275 and the quantile level 0.97725, which corresponds to an uncertainty range of $\mu \pm 2\sigma$ in a normal distribution. We compare the results of copula-based probabilitic models with independent Gaussian models and the GRU-based correlated Gaussian model. In Figure 3, one can clearly see that the mean frequency deviation is very often outside the 2 σ range of the independent Gaussian models or the confidence interval of the correlated Gaussian model, while the uncertainty regions of the copula-based predictors cover the fluctuations of the mean frequency deviation much better. Interestingly, none of the models are able to accurately predict the mean frequency deviation around midday on the example day.

4 Conclusion

In this paper, we presented a grid frequency prediction methodology where an existing point predictor can be transformed into a copula-based probabilistic predictor using feature space-based probabilistic error correction. The created probabilistic models are able to capture the dependencies between different time points in a prediction hour interval and show better performance in terms of energy scores compared to their independent baseline counterparts. Additionally, by simulating the mean grid frequency deviation from different models, we have shown that both temporally independent models such as independent Gaussian process-based models as well as correlated Gaussian process models cannot account for enough uncertainty, while the copula-based models perform much better with respect to this. The feature based error correction technique can also be used for other energy data prediction tasks.

Acknowledgments

Maximilian Coblenz and Bolin Liu gratefully acknowledge support from Forschungsstart project P2022-13-009 of the Carl-Zeiss-Stiftung. We thank the anonymous reviewers for their careful reading and insightful comments.

References

[1] Michaël Baudin, Anne Dutfoy, Bertrand Iooss, and Anne-Laure Popelin. 2016. *OpenTURNS: An Industrial Software for Uncertainty Quantification in Simulation.* Springer International Publishing, Cham, 1–38. doi:10.1007/978-3-319-11259-6_64-1

[2] Mikhail Bilenko, Sugato Basu, and Raymond J Mooney. 2004. Integrating constraints and metric learning in semi-supervised clustering. In *Proceedings of the twenty-first international conference on Machine learning*. 11.

[3] Thomas M Cover. 1999. *Elements of information theory.* John Wiley & Sons.

[4] Renato Cordeiro De Amorim. 2011. Learning feature weights for K-Means clustering using the Minkowski metric. *Department of Computer Science and Information Systems Birkbeck, University of London* (2011).

[5] Dorothea Diers, Martin Eling, and Sebastian D Marek. 2012. Dependence modeling in non-life insurance using the Bernstein copula. *Insurance: Mathematics and Economics* 50, 3 (2012), 430–436.

[6] Christian Genest and Anne-Catherine Favre. 2007. Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of hydrologic engineering* 12, 4 (2007), 347–368.

[7] Oliver Grothe, Fabian Kächele, and Fabian Krüger. 2023. From point forecasts to multivariate probabilistic forecasts: The Schaake shuffle for day-ahead electricity price forecasting. *Energy Economics* 120 (2023), 106602.

[8] Brendan J Kirby. 2003. *Frequency control concerns in the North American electric power system.* Technical Report. Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).

[9] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. 2004. Estimating mutual information. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* 69, 6 (2004), 066138.

[10] Johannes Kruse, Eike Cramer, Benjamin Schäfer, and Dirk Witthaut. 2023. Physics-informed machine learning for power grid frequency modeling. *PRX energy* 2, 4 (2023), 043003.

[11] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2020. Predictability of power grid frequency. *IEEE access* 8 (2020), 149435–149446.

[12] Johannes Kruse, Benjamin Schäfer, and Dirk Witthaut. 2021. Revealing drivers and risks for power grid frequency stability with explainable AI. *Patterns* 2, 11 (2021). doi:10.1016/j.patter.2021.100373

[13] J. Kruse, D. Witthaut, and B. Schäfer. 2021. Supplementary data: "Revealing drivers and risks for power grid frequency stability with explainable AI". https://doi.org/10.5281/zenodo.5118352.

[14] Bolin Liu, Maximilian Coblenz, and Oliver Grothe. 2024. Predicting grid frequency short-term dynamics with Gaussian processes and sequence modeling. In *Proceedings of the 15th ACM International Conference on Future and Sustainable Energy Systems*. 535–550.

[15] Bolin Liu, Maximilian Coblenz, and Oliver Grothe. 2025. Supplementary Code for the Paper "Copula-based Probabilistic Prediction of Grid Frequency Dynamics". https://github.com/bolin-liu/copula_grid_frequency_prediction. Accessed: 2025-03-28.

[16] Vivek Prakash, Priyanka Kushwaha, Kailash Chand Sharma, and Rohit Bhakar. 2022. Frequency response support assessment from uncertain wind generation. *International Journal of Electrical Power & Energy Systems* 134 (2022), 107465.

[17] Brian C Ross. 2014. Mutual information between discrete and continuous data sets. *PloS one* 9, 2 (2014), e87357.

[18] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[19] Sajeeb Saha, MI Saleem, and TK Roy. 2023. Impact of high penetration of renewable energy sources on grid frequency behaviour. *International Journal of Electrical Power & Energy Systems* 145 (2023), 108701.

[20] Alessio Sancetta and Stephen Satchell. 2004. The Bernstein copula and its applications to modeling and approximations of multivariate distributions. *Econometric theory* 20, 3 (2004), 535–562.

[21] Carlos Sarmiento, Carlos Valencia, and Raha Akhavan-Tabatabaei. 2018. Copula autoregressive methodology for the simulation of wind speed and direction time series. *Journal of Wind Engineering and Industrial Aerodynamics* 174 (2018), 188–199.

[22] Roman Schefzik, Thordis L Thorarinsdottir, and Tilmann Gneiting. 2013. Uncertainty quantification in complex simulation models using ensemble copula coupling. (2013).

[23] Johan Segers, Masaaki Sibuya, and Hideatsu Tsukahara. 2017. The empirical beta copula. *Journal of Multivariate Analysis* 155 (2017), 35–51.

[24] M Sklar. 1959. Fonctions de répartition à n dimensions et leurs marges. In *Annales de l'ISUP*, Vol. 8. 229–231.
[25] Robert F Woolson. 2005. Wilcoxon signed-rank test. *Encyclopedia of biostatistics* 8 (2005).
[26] Eric Xing, Michael Jordan, Stuart J Russell, and Andrew Ng. 2002. Distance metric learning with application to clustering with side-information. *Advances in neural information processing systems* 15 (2002).
[27] Tolga Yamut and Burcu Hudaverdi. 2023. Classification with Bernstein copula as discrimination function. *Communications in Statistics-Simulation and Computation* (2023), 1–17.
[28] Francesco Zanetta and Sam Allen. 2024. *Scoringrules: a python library for probabilistic forecast evaluation.* https://github.com/frazane/scoringrules