



Managing Anomalies in Energy Time Series for Automated Forecasting

Marian Turowski^(✉) , Oliver Neumann , Lisa Mannsperger, Kristof Kraus, Kira Layer, Ralf Mikut , and Veit Hagenmeyer 

Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

marian.turowski@kit.edu

Abstract. The increasing number of recorded energy time series enables the automated operation of smart grid applications such as load analysis, load forecasting, and load management. However, to perform well, these applications usually require clean data that well represents the typical behavior of the underlying system. Unfortunately, recorded time series often contain anomalies that do not reflect the typical behavior of the system and are, thus, problematic for automated smart grid applications such as automated forecasting. While various anomaly management strategies exist, a rigorous comparison is lacking. Therefore, in the present paper, we introduce and compare three different general strategies for managing anomalies in energy time series forecasting, namely the raw, the detection, and the compensation strategy. We compare these strategies using a representative selection of forecasting methods and real-world data with inserted synthetic anomalies. The comparison shows that applying the compensation strategy is generally beneficial for managing anomalies despite requiring additional computational costs because it mostly outperforms the detection and the raw strategy when the input data contains anomalies.

Keywords: Anomalies · Anomaly management · Forecasting · Energy time series

1 Introduction

Since energy systems around the world transition to an increasing share of renewable energy sources in energy supply, the implementation of smart grids supporting this transition also advances. Smart grid implementation implies a growing number of smart meters that record power or energy consumption and generation as time series [4]. These recorded energy time series are characterized by a multi-seasonality, an aggregation-level dependent predictability, and a dependence on exogenous influences such as weather [16]. The increasing number of recorded energy time series enables a wide range of possible applications for this data and the goal of their automated operation. Exemplary applications for the smart grid that support the transition to renewable energy sources include customer profiling, load analysis, load forecasting, and load management [39, 47].

© The Author(s) 2024

B. N. Jørgensen et al. (Eds.): EIA 2023, LNCS 14467, pp. 3–29, 2024.

https://doi.org/10.1007/978-3-031-48649-4_1

However, to perform well, these applications usually require clean data that represents the typical behavior of the underlying system well [30, 47].

Unfortunately, recorded time series are usually not clean, but contain anomalies [13]. Anomalies are patterns that deviate from what is considered normal [10]. They can occur in energy time series for many reasons, including smart meter failures [46], unusual consumption [32, 40], and energy theft [24]. All anomalies have in common that they potentially contain data points or patterns that represent false or misleading information, which can be problematic for any analysis of this data performed by the mentioned applications [47]. For example, anomalies such as positive or negative spikes may strongly deviate from what is considered normal, and a subsequent forecasting method that uses the data as input in an automated manner may generate an incorrect forecast. This forecast could in turn lead to an inappropriate energy schedule and ultimately affect the stability of the energy system in an automated smart grid setting.

Therefore, managing anomalies in energy time series – in the sense of dealing with their presence – is an important issue for applications in the energy system such as billing and forecasting [47]. In energy time series forecasting, the importance of an adequate anomaly management is generally known, e.g. [2, 3, 36]. For this reason, various anomaly management strategies exist, including the use of robust forecasting methods [23, 28], the use of information on detected anomalies [43], and the compensation of detected anomalies [9, 14, 37, 50]. However, it is not clear which strategy is the best for managing anomalies in energy time series forecasting regarding the obtained accuracy and also the associated necessary effort, which is why a rigorous comparison of available strategies is needed.

Therefore, the present paper introduces and compares different general strategies for managing anomalies in energy time series forecasting. For this purpose, we build on the typically used strategies mentioned above and describe three different general strategies based on them, namely the raw, the detection, and the compensation strategy. While the raw strategy applies forecasting methods directly to the data input without any changes, the detection strategy provides information on anomalies detected in the input data to the forecasting method. The compensation strategy cleans the input data by detecting and thereafter compensating anomalies in the input data before applying a forecasting method.

To comparatively evaluate these strategies, we use a representative selection of forecasting methods, including naive, simple statistical, statistical learning, and machine learning methods. We also make use of real-world energy time series with inserted synthetic anomalies derived from real-world data. Given these forecasting methods and data, we compare the obtained forecast accuracy of all proposed strategies and present an example of how these strategies work and perform.

The remainder of the present paper is structured as follows: After describing related work in Sect. 2, Sect. 3 introduces the strategies for managing anomalies in energy time series forecasting. In Sect. 4, we evaluate the presented strategies. Finally, we discuss the results and the strategies in Sect. 5 and conclude the paper in Sect. 6.

2 Related Work

Since anomalies are potentially limiting the performance of any downstream application, dealing with their presence is generally a well-known topic. For example, all kinds of pre-processing methods aim to raise data quality to ensure the validity and reliability of data analysis results, e.g. [5, 19, 37]. Similarly, the influence of the choice of preprocessing methods on the accuracy of forecasting methods is also known, e.g. [1, 3]. In energy time series forecasting, several works also address how to deal with the presence of anomalies. We organize these works along three strategies.

Works of the first strategy focus on the robustness of forecasting methods. These works, for example, develop forecasting methods that are robust against anomalies, e.g. [23, 28, 29, 51, 53], strengthen existing forecasting methods, e.g. [54], or at least investigate the robustness of forecasting methods with respect to anomalous data, e.g. [27, 52]. The second strategy consists of works that make use of information on detected anomalies. In [43], for example, the information on predicted anomalies is used to adapt the energy production. Works of the third strategy detect anomalies and replace the detected anomalies with appropriate values, e.g. [9, 14, 30, 33], or even remove the detected anomalies, e.g. [11].

Despite these works on specific anomaly management strategies, it is not known which strategy is the best for managing anomalies in energy time series forecasting. For this reason, a rigorous comparison of the available strategies – as done in the present paper – is lacking.

3 Strategies for Managing Anomalies in Energy Time Series Forecasting

In this section, we present three general strategies for managing anomalies in energy time series forecasting, which build on the previously described anomaly management strategies in literature.¹ All of these strategies apply a forecasting method $f(\circ)$ to create a forecast for an input power time series $\mathbf{y} = \{y_t\}_{t \in T}$ with T measured values. This forecasting method creates a forecast based on historical values of the input power time series and exogenous features \mathbf{e} such as calendar information or weather forecasts. More specifically, the forecasting method combines the most recent N historical values of the input power time series $\mathbf{y}_t = y_{t-(N-1)}, \dots, y_t$ with the exogenous features $\mathbf{e}_{t+H} = e_{t+1}, \dots, e_{t+H}$ for the forecasting horizon H . Using this combination, the forecasting method then generates a forecast at time point t

$$\hat{\mathbf{y}}_{t+H} = f(\mathbf{y}_t, \mathbf{e}_{t+H}), \quad (1)$$

where $\hat{\mathbf{y}}_{t+H} = \hat{y}_{t+1}, \dots, \hat{y}_{t+H}$ is the forecast value for the input power time series for each time step in the forecast horizon. Nevertheless, the considered strategies

¹ The implementation of the proposed and evaluated strategies is available at <https://github.com/KIT-IAI/ManagingAnomaliesInTimeSeriesForecasting>.

comprise different steps and thus differ in the inputs to the applied forecasting method (see Fig. 1). We thus describe the included steps, the used input, and the underlying assumptions for each strategy in the following.

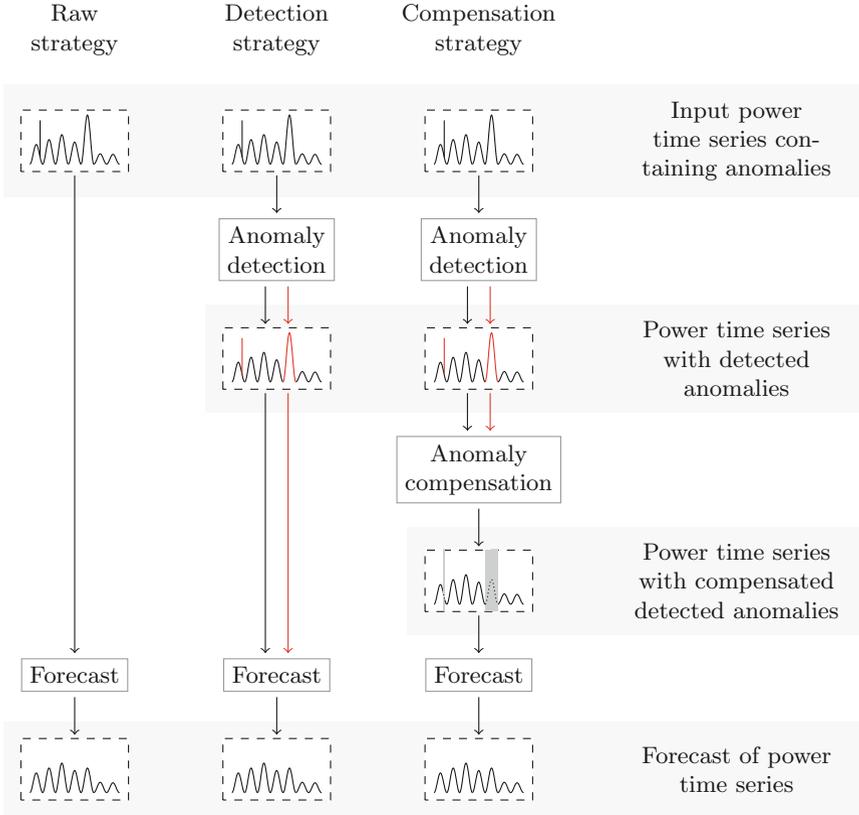


Fig. 1. The three strategies for managing anomalies in energy time series forecasting. The raw strategy directly uses the input power time series to create a forecast. The detection strategy first detects anomalies in the input power time series, before providing a forecast using the information on the detected anomalies from the power time series with detected anomalies. The compensation strategy detects anomalies and additionally compensates the detected anomalies before performing a forecast based on the power time series with compensated detected anomalies.

Raw Strategy. The first strategy is the so-called raw strategy. It directly uses a power time series containing anomalies as input to a forecasting method. Given this input, the applied forecasting method thus provides a forecast of the input power time series. Formally, the raw strategy thus creates a forecast at time point t

$$\hat{\mathbf{y}}_{t+H}^{\text{raw}} = f(\mathbf{y}_t, \mathbf{e}_{t+H}), \quad (2)$$

where \mathbf{y}_t are the historical values of the input power time series containing anomalies and H , \mathbf{e}_{t+H} , and $\hat{\mathbf{y}}_{t+H}$ are defined as above.

The raw strategy assumes that the anomalies contained in the input time series do not strongly affect the forecast of the applied forecasting method or that the applied forecasting method is robust against anomalies. Therefore, the applied forecasting method is assumed to still achieve an accurate forecast.

Detection Strategy. The second strategy is the so-called detection strategy. This strategy first applies an anomaly detection method to the power time series containing anomalies to detect contained anomalies whereby the anomaly detection method can be supervised or unsupervised. The resulting power time series with detected anomalies serves as input to the forecasting method that then provides the forecast of the power time series. Formally, the detection strategy, therefore, results in a forecast at time point t

$$\hat{\mathbf{y}}_{t+H}^{\text{detection}} = f(\mathbf{y}_t, \mathbf{d}_{t+H}, \mathbf{e}_{t+H}), \quad (3)$$

where $\mathbf{d}_{t+H} = d_{t+1}, \dots, d_{t+H}$ are the labels of the detected anomalies for the forecasting horizon H and \mathbf{y}_t , \mathbf{e}_{t+H} , and $\hat{\mathbf{y}}_{t+H}$ are defined as above.

The assumption of the detection strategy is that the applied forecasting method can incorporate information about detected anomalies in its model so that the consideration of detected anomalies leads to an accurate forecast.

Compensation Strategy. The third strategy is the so-called compensation strategy. It also first applies a supervised or unsupervised anomaly detection method to the power time series containing anomalies to identify the contained anomalies. However, this strategy then uses the power time series with detected anomalies as input to an anomaly compensation method $c(\circ)$ that replaces the detected anomalies with realistic values, i.e.,

$$\tilde{\mathbf{y}}_{t+H} = c(\mathbf{y}_t, \mathbf{d}_{t+H}, \circ), \quad (4)$$

where $\tilde{\mathbf{y}}_{t+H}$ is the power time series with compensated detected anomalies and \circ are additional parameters of the compensation method. This power time series with compensated detected anomalies $\tilde{\mathbf{y}}_{t+H}$ serves as input to the forecasting method that provides the forecast of the power time series. Formally, we describe the forecast of the compensation strategy at time point t with

$$\hat{\mathbf{y}}_{t+H}^{\text{compensation}} = f(\tilde{\mathbf{y}}_t, \mathbf{e}_{t+H}), \quad (5)$$

where $\tilde{\mathbf{y}}_t$ are the historical values of the input power time series with compensated detected anomalies and H , \mathbf{e}_{t+H} , and $\hat{\mathbf{y}}_{t+H}$ are defined as above.

The compensation strategy assumes that anomalies have to be detected and compensated in order to enable the applied forecasting method to provide an accurate forecast.

4 Evaluation

To evaluate the proposed strategies for managing anomalies in energy time series forecasting, we compare the forecasting accuracy of all strategies using different

forecasting methods. Before presenting the results, we detail the performed evaluation: We introduce the used data and the inserted synthetic anomalies, the anomaly detection methods applied in the detection and compensation strategies, and the anomaly compensation method applied in the compensation strategy. We also describe the used forecasting methods and the experimental setting.

4.1 Data and Inserted Synthetic Anomalies

For the evaluation, we use real-world data in which we insert synthetic anomalies. The chosen data set is the “ElectricityLoadDiagrams20112014 Data Set”² from the UCI Machine Learning Repository [18]. It includes electrical power time series from 370 clients with different consumption patterns [38]. The 370 time series are available in a quarter-hourly resolution for a period of up to four years, namely from the beginning of 2011 until the end of 2014. We choose the power time series MT_200 for the evaluation to cover the entire four-year period, to account for the electrical load of a typical client, and to consider a time series that is anomaly-free compared to other time series in the data set (see Fig. 2).

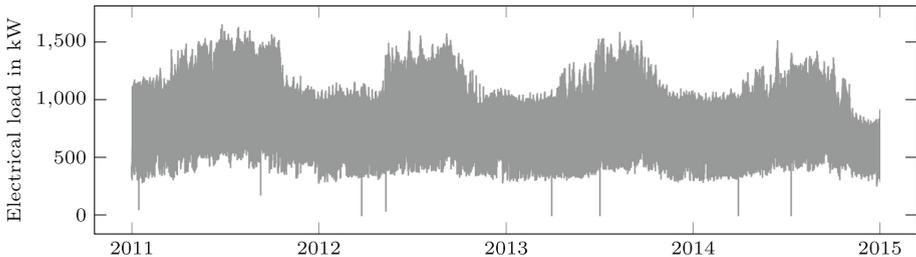


Fig. 2. Overview of the data used for the evaluation.

Since the chosen time series does not include labeled anomalies and thus do not allow for controlled experimental conditions, we insert synthetic anomalies in the complete chosen time series. For this, we consider the two anomaly groups used in [44], namely technical faults in the metering infrastructure and unusual consumption. Using the corresponding available anomaly generation method³, we insert four types of anomalies from each group: Anomalies of types 1 to 4 are from the group of technical faults and based on anomalies identified in real-world power time series in [45]. These anomalies violate the underlying distribution corresponding to normal behavior. Anomalies of types 5 to 8 are from the group of unusual consumption and represent unusual behavior as described in [44]. These anomalies are characterized by unusually low or high power consumption. We give formulas and examples for all types of anomalies in Appendix 1.

² <https://archive.ics.uci.edu/ml/datasets/ElectricityLoadDiagrams20112014>.

³ <https://github.com/KIT-IAI/EnhancingAnomalyDetectionMethods>.

For the evaluation, we insert once 20 anomalies of types 1 to 4 each from the group of technical faults and once 20 anomalies of types 5 to 8 each from the group of unusual consumption into the selected time series. We insert 20 anomalies per type to consider a reasonable number of anomalies and we insert all four types of anomalies from a group at once to consider their simultaneous occurrence [44, 45]. The inserted anomalies correspond to 5% of the data for the technical faults and 11% of the data for the unusual consumption.

4.2 Applied Anomaly Detection Methods

For the evaluation of the detection and compensation strategies, we choose anomaly detection methods based on the evaluation results in [44], where a variety of anomaly detection methods is already evaluated on the selected data. More specifically, we choose the method from the evaluated supervised and unsupervised anomaly detection methods that overall performs best for the considered groups of anomalies. For both groups of anomalies, the best-performing method is an unsupervised anomaly detection method, namely the Variational Autoencoder (VAE) for the technical faults and the Local Outlier Factor (LOF) for the unusual consumption. We briefly introduce both chosen anomaly detection methods, before we describe their application.

The Variational Autoencoder (VAE) learns to map its input to its output using the probability distribution of ideally anomaly-free data in the latent space, so it is trained to only reconstruct non-anomalous data [26]. The Local Outlier Factor (LOF) estimates the local density of a sample by the distance to its k -nearest neighbors and uses low local densities compared to its neighbors to determine anomalies [8].

To enhance the detection performance of the selected anomaly detection methods, we apply them to the latent space representation of the selected data as suggested in [44] and visualized in Fig. 8 in Appendix 2. We choose the generative method to create these latent space representations for each selected anomaly detection method based on the evaluation results in [44]: We create the latent space data representations for the vae with a conditional Invertible Neural Network (cINN) [6] and the latent space data representation for the LOF with a conditional Variational Autoencoder (cVAE) [41]. We detail the architecture and training of the used cINN and cVAE in Appendix 2. Given the created latent space representation, we apply the selected anomaly detection methods to the entire selected time series of the chosen data as in [44].

4.3 Applied Anomaly Compensation Method

For the anomaly compensation in the evaluation of the proposed compensation strategy, we use a Prophet-based imputation method because of its superior imputation performance for power time series determined in [48]. The Prophet-based imputation method [48] is built on the forecasting method Prophet which is capable of estimating a time series model on irregularly spaced data [42].

Prophet uses a modular regression model that considers trend, seasonality, and holidays as key components. It can be described as

$$y(t) = g(t) + s(t) + h(t) + \varepsilon_t, \quad (6)$$

where g models the trend, s the seasonality, h the holidays, and ε_t all other changes not represented in the model. The Prophet-based imputation method trains the regression model using all values available in the power time series. Given the trained regression model, the Prophet-based imputation method considers all anomalies in the power time series as missing values and imputes them with the corresponding values from the trained regression model.

4.4 Anomaly-Free Baseline Strategy

In the evaluation, we examine the proposed raw, detection, and compensation strategies all based on the selected data containing inserted synthetic anomalies. For the evaluation of these strategies, we additionally provide an anomaly-free baseline. This baseline strategy comprises forecasts that are calculated on that selected data but without any inserted anomalies (see Fig. 3).

Anomaly-free baseline strategy

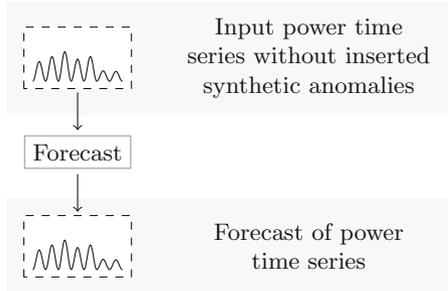


Fig. 3. For evaluating the proposed strategies on the data with inserted synthetic anomalies, we use the forecast calculated on the input power time series without inserted anomalies as an anomaly-free baseline strategy.

4.5 Applied Forecasting Methods

For the evaluation of the proposed strategies, we consider a multi-step 24 h-ahead forecast with a multiple output strategy for which we apply a representative selection of forecasting methods to the selected data. Due to the quarter-hourly resolution of the selected data, the forecast comprises 96 values. For forecasting methods with hyperparameters, we use hyperparameters that we initially choose

based on best practices and then verify. We first present the selected forecasting methods and their input data for the raw and compensation strategies, before we describe them for the detection strategy and the anomaly-free baseline strategy. We lastly present the used train-test split.

Methods Applied in Raw and Compensation Strategies. To examine the raw and compensation strategies comprehensively, we consider methods with different learning assumptions. We apply eight forecasting methods, namely two naive and six advanced methods. The advanced methods comprise a simple statistical method, a simple and two more complex machine learning methods, and two statistical learning methods.

The first naive method is the Last Day Forecast. It uses the values of the previous 24 h for the values to be predicted, i.e.,

$$\hat{y}_{t,h} = y_{t-96+h}, \quad (7)$$

where $\hat{y}_{t,h}$ is the forecast value of the electrical load for the forecast horizon h at time t and y_t is the electrical load at time t .

The second naive method is the Last Week Forecast. It takes the corresponding values of the last week as the forecast values, i.e.,

$$\hat{y}_{t,h} = y_{t-672+h}, \quad (8)$$

where $\hat{y}_{t,h}$ is the forecast value of the electrical load for the forecast horizon h at time t and y_{t-672} is the electrical load one week ago at time $t - 672$.

The first advanced method is the Linear Regression (LinR). As a statistical method, it models the forecast values as a linear relationship between the historical load values and calendar information and determines the corresponding parameters using ordinary least squares. It is defined as

$$\hat{y}_{t,h} = c_h + \sum_j \beta_{h,j} \cdot y_{t-j} + \sum_k \gamma_{h,k} \cdot C_{t,k} + \varepsilon, \quad (9)$$

where c is a constant, index j iterates over the lagged load features y_{t-j} , index k iterates over the calendar information $C_{t,k}$, and ε is the error.

The second advanced method is a commonly applied simple machine learning method, namely a Neural Network (NN). It organizes a network of interconnected nodes in input, hidden, and output layers to apply different functions that activate the corresponding nodes to learn the relationship between input and output (e.g., [31, 49]). The implementation of the used NN is detailed in Table 6 in Appendix 3. For its training, we use a batch size of 64, the Adam optimizer [25] with default parameters, and a maximum of 50 epochs.

The third advanced method is the Profile Neural Network (PNN) [21] as a state-of-the-art and more complex machine learning method. It combines statistical information in the form of standard load profiles with convolutional neural networks (CNNs) to improve the forecasting accuracy. For this, it decomposes a power time series into a standard load profile module, a trend module, and a colorful noise module, before aggregating their outputs to obtain the forecast

[21]. For the training, the PNN uses a batch size of 512, the Adam optimizer [25], and a maximum of 50 epochs.

The fourth advanced method is the Random Forest (RF) Regressor representing a statistical learning method. It creates several randomly drawn regression trees and takes the mean of each individual tree’s forecast as forecast [7], i.e.,

$$\hat{y}_{t,h} = \frac{1}{B} \sum_{b=1}^B t_{b,h}(x), \quad (10)$$

where B is the number of bootstrap samples of the training set, t_b is an individual fitted tree, and x are the values from the test set. For the evaluation, we use $B = 100$.

The fifth advanced method is the Support Vector Regression (SVR) and represents another statistical learning method. It determines a regression plane with the smallest distance to all data points used for the training. The data points closest to the regression plane on both sides are the so-called support vectors [17]. We apply the SVR with a linear kernel, $C = 1.0$, and $\epsilon = 1.0$.

The sixth advanced method is the XGBoost Regressor, which represents a more complex machine learning method. It iteratively creates regression trees and uses gradient descent to minimize a regularized objective function [12].

All introduced forecasting methods use the historical values of the selected power time series that contains inserted synthetic anomalies. The advanced methods also consider calendar information as input (see Table 5 in Appendix 3 for more details). While the naive methods directly use the mentioned historical load values, all other methods obtain the normalized load of the last 24 h and the calendar information for the first value to be predicted.

Methods Applied in Detection Strategy. For the detection strategy that can use the information on the detected anomalies for the forecast, we apply the forecasting methods introduced for the raw and compensation strategies. This way, we also evaluate the detection method using forecasting methods with different learning assumptions. However, we adapt the previously introduced methods as follows: We change the Last Day Forecast so that it uses the value from a week ago in the case of a detected anomaly. Similarly, we modify the Last Week Forecast so that it uses the corresponding value of the second to last week as the forecast value if the value to be predicted is a detected anomaly. In accordance with the detection strategy, all other forecasting methods obtain the information on the detected anomalies of the last 24 h as additional features.

Methods Applied in Anomaly-Free Baseline Strategy. To calculate the anomaly-free baseline strategy for the data containing synthetic anomalies, we apply all forecasting methods described for the raw and compensation strategies to the same data but without inserted synthetic anomalies. These forecasting methods obtain the inputs in the way described for the raw and compensation strategies.

Train-Test Split. Regardless of the considered strategy, we use the same train-test split for all evaluated forecasting methods. Each forecasting method is trained on

80% of the available data and tested on the remaining 20%. For all strategies, the available data is the selected time series without the first 96 data points. When calculating the anomaly-free baseline strategy for this data, we use the same period of time, i.e., all values except the first 96 data points.

4.6 Experimental Setting

For evaluation, we use evaluation metrics in a defined hard- and software setting.

Metrics. In order to evaluate the proposed strategies for managing anomalies in energy time series forecasting, we examine the accuracy of the obtained forecasts compared to the data without inserted synthetic anomalies using two metrics.

The first metric is the commonly used root mean squared error (RMSE). Given N data points to be predicted, it is defined as

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=1}^N (y_t - \hat{y}_t)^2}, \quad (11)$$

with the actual value y_t of the anomaly-free time series and the forecast value \hat{y}_t . Due to the squared differences considered, the RMSE is sensitive to outliers.

Therefore, we also consider a second commonly used metric, the mean absolute error (MAE), which is robust to outliers. It is defined as

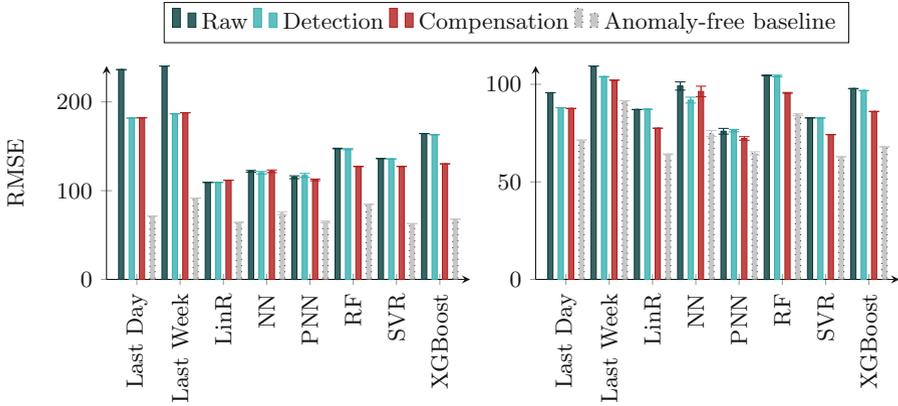
$$\text{MAE} = \frac{1}{N} \sum_{t=1}^N |y_t - \hat{y}_t| \quad (12)$$

with N data points to be forecast, the actual value y_t of the anomaly-free time series, and the forecast value \hat{y}_t .

Hard- and Software. In order to obtain comparable results, we use the same hardware throughout the evaluation and implement all evaluated strategies and used anomaly detection, anomaly compensation, and forecasting methods in Python (see Appendix 4 for more details).

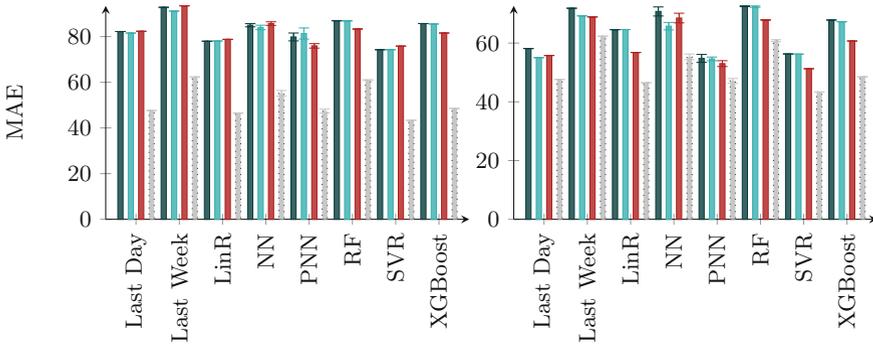
4.7 Results

To examine the presented strategies, we compare their accuracy on the selected time series with the described inserted synthetic anomalies and using the described anomaly detection, anomaly compensation, and forecasting methods. After presenting the results of this comparison for the technical faults and unusual consumption, we show a part of the selected time series as an example of how the different strategies work and perform.



(a) Technical faults (RMSE).

(b) Unusual consumption (RMSE).



(c) Technical faults (MAE).

(d) Unusual consumption (MAE).

Fig. 4. The accuracy of the eight forecasting methods applied to the data with 20 synthetic anomalies of each type from the technical faults and unusual consumption. For each forecasting method introduced in Sect. 4.5, the bars indicate the average RMSE or MAE for the raw strategy, detection strategy, compensation strategy, and anomaly-free baseline strategy. The error bars show the observed standard deviation across all runs on the test data set. Note that the anomaly-free baseline strategy generally performs best because it uses data that does not contain inserted synthetic anomalies.

Comparison. For the comparison, we apply all proposed strategies to the selected data with synthetic anomalies from the technical faults and unusual consumption. For both groups of anomalies, we insert 20 anomalies of each type belonging to this group. Figure 4a and 4c show the resulting RMSE and MAE for the technical faults and Fig. 4b and 4d for the unusual consumption. For each considered forecasting method, the bars indicate the average RMSE or MAE for the raw strategy, the detection strategy, the compensation strategy, and the

anomaly-free baseline strategy. The error bars show the observed standard deviation across all runs on the test data set.

Technical Faults. Regarding the technical faults, all considered forecasting methods except the Last Day Forecast, the Last Week Forecast, the LinR, and the NN have both the lowest RMSE and MAE when using the compensation strategy. The SVR has only the lowest RMSE with the compensation strategy but the lowest MAE with the raw strategy. Even though the difference to the compensation strategy is only small, the Last Day Forecast, the Last Week Forecast, and the NN achieve their lowest RMSE and MAE using the detection strategy and the LinR with the raw strategy. Moreover, the difference between the RMSE when using the compensation strategy and the RMSE using the second best strategy is largest for the XGBoost Regressor, the RF Regressor, and the SVR. Similarly, the difference between the MAE when using the compensation strategy and the MAE using the second best strategy is largest for the XGBoost Regressor, the PNN, and the RF Regressor. Additionally, we see the largest difference between the RMSEs in the use of the raw, detection, and compensation strategies for the Last Day Forecast and the Last Week Forecast, followed by the XGBoost Regressor. With respect to the MAE, we observe the largest differences for the PNN, the XGBoost Regressor, and RF Regressor.

Compared to the anomaly-free baseline strategy, the RMSE of all forecasting methods, especially of the Last Day Forecast, the Last Week Forecast, the SVR, and the XGBoost Regressor, is also noticeably greater for all three strategies. Concerning the MAE, we also see large differences between the anomaly-free baseline strategy and the three other strategies for all forecasting methods but especially the Last Day Forecast, the XGBoost Regressor, the LinR, and the SVR. Considering the actual accuracy, the LinR, the PNN, and the NN form the group of forecasting methods that achieve the lowest RMSE and the SVR, the PNN, and the LinR the group with the lowest MAE.

Unusual Consumption. For the unusual consumption, all considered forecasting methods except the NN achieve both the lowest RMSE and MAE using the compensation strategy. The NN has its lowest RMSE with the detection strategy. The Last Day Forecast also has its lowest RMSE using the compensation strategy but its lowest MAE using the detection strategy. The difference in the RMSE and MAE between using the compensation strategy and using the second best strategy is large for the XGBoost Regressor, the LinR, the RF Regressor, the SVR, and small for the NN, the PNN, the Last Day Forecast, and the Last Week Forecast. Moreover, we observe the largest differences between the RMSEs for using the raw, detection, and compensation strategies for the LinR, the RF Regressor, and the SVR. The largest observed differences in the MAE of these three strategies are for the LinR, the XGBoost Regressor, and the NN.

In comparison to the anomaly-free baseline strategy, the RMSE and MAE of all forecasting methods is clearly larger for all three strategies. With regard to their actual accuracy, the PNN achieves the lowest RMSE, followed by the SVR and the LinR. Considering the accuracy in terms of the MAE, the SVR, the PNN, and the Last Day Forecast achieve the lowest MAE.

Example. To demonstrate how the different strategies work and perform, we finally look at a part of the time series used for the evaluation in more detail. Using three days of this time series from June 2014, Fig. 5 illustrates an inserted synthetic anomaly, how this anomaly is detected, and what the resulting forecasts of the different strategies look like.

More specifically, Fig. 5a illustrates the selected original time series, which we assume to be anomaly-free, and the time series with inserted synthetic anomalies. In the latter, we observe an anomaly of type 8 which increases the load values for about one of the three days. In addition to these two time series, Fig. 5b shows the data points of the time series with inserted synthetic anomalies that are detected as anomalous by the LOF, which is the applied anomaly detection method. We observe that the LOF, detects various but not all data points of the inserted synthetic anomaly as anomalous. Figure 5c then illustrates how these detected anomalous data points are compensated using the Prophet-based imputation method. With regard to the compensated detected anomalous data points, our observation is that compensated values are all close to the original anomaly-free time series.

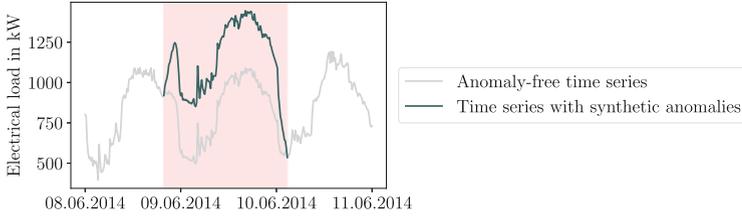
Finally, Fig. 5d additionally shows the multi-step 24 h-ahead forecasts of the four different strategies using the PNN and given the previously described information. We observe that all strategies result in different forecasts: The forecast of the compensation strategy, that is based on the time series with compensated synthetic anomalies introduced in Fig. 5c, is closest to the forecast of the anomaly-free baseline strategy. Moreover, the forecast of the detection strategy, that uses the information on the detected anomalous data points introduced in Fig. 5b, is closer to the forecast of the anomaly-free baseline strategy than the forecast of the raw strategy.

5 Discussion

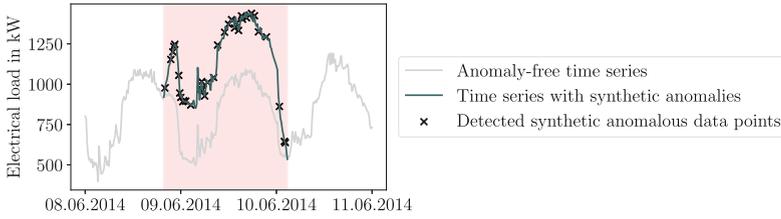
In this section, we first discuss the results from the evaluation of the proposed strategies for managing anomalies in energy time series forecasting, before reviewing the evaluation regarding its limitations and insights.

In the comparison of the accuracy of the proposed strategies, we observe that using the compensation strategy yields the lowest RMSE and MAE for most forecasting methods and both groups of anomalies. However, while the results are generally consistent across both accuracy metrics, some forecasting methods benefit from the two other strategies with respect to the RMSE, the MAE, or both: The NN and the Last Day Forecast perform best using the detection strategy for the technical faults and unusual consumption, the Last Week Forecast using the detection strategy for only the technical faults, and the LinR and the SVR using the raw strategy for the technical faults. However, it is worth noting that the compensation strategy is often the second-best strategy in these cases with similar accuracy, so it could serve as a default strategy.

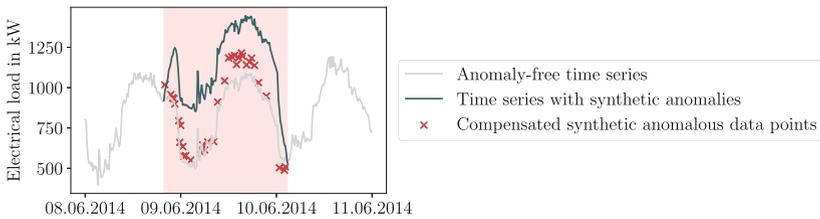
Nevertheless, using the compensation and also the detection strategy is associated with additional computational costs because of the necessary anomaly



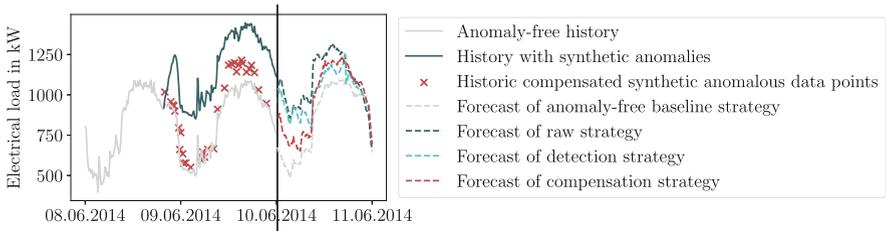
(a) Anomaly-free time series and time series with inserted synthetic anomalies.



(b) Anomaly-free time series, time series with inserted synthetic anomalies, and the anomalous data points of the synthetic anomaly that are detected using the LOF.



(c) Anomaly-free time series, time series with inserted synthetic anomalies, and the detected anomalous data points of the synthetic anomaly that are compensated using the Prophet-based imputation method.



(d) Forecasts of four all strategies using the PNN, the anomaly-free history, the history with synthetic anomalies, and the historic compensated synthetic anomalous data points.

Fig. 5. Three exemplary days of the power time series used for the evaluation, where a synthetic anomaly of type 8 is inserted, detected, and compensated. Finally, the inserted anomaly is dealt with differently in the forecast depending on the strategy.

detection and anomaly compensation. In case of the anomaly detection, the computational costs also include the creation of the latent space representation that we use to enhance the detection performance. Whether the improvement in accuracy over the raw strategy of using the data essentially as-is justifies this additional computational cost depends on the forecasting method and the anomalies contained in the data and requires careful consideration. From these results, one could infer that applying strategies that actively handle anomalies, namely the compensation and the detection strategies, is generally beneficial and, more specifically, that using the compensation strategy is mostly beneficial. Based on this inference, a best practice could be to apply the compensation strategy in energy time series forecasting. Additionally, given the nature of the data used for the evaluation, we assume that the gained insights also apply to similar periodic data, for example, from the areas of solar power generation, mobility, and sales.

Moreover, the comparison of the accuracy of all strategies shows that there is no clearly best-performing forecasting method for technical faults and unusual consumption. Instead, there are rather groups of similarly well-performing forecasting methods, for example, the LinR and the PNN for the technical faults and the PNN and the SVR for the unusual consumption. Additionally, regarding their actual accuracy, we observe that even naive forecasting methods can provide reasonable forecasts, which can serve as a computationally light baseline when looking for competitive forecasts.

Furthermore, the example of three days from the time series used for the evaluation illustrates how the strategies differ. By showing an inserted synthetic anomaly, the data points of this anomaly that are detected as anomalous, and how the detected anomalous data points are compensated, the example presents the inputs for the raw, detection, and compensation strategies. Additionally, the example includes the resulting forecasts of all strategies for the next day. Thereby, the influence of the different inputs on the forecast accuracy of the different strategies becomes comprehensible. Considering the results of the example, we also observe that the compensation strategy provides the comparatively best forecast although it is dependent on the only partly detected and compensated inserted synthetic anomaly. With regard to the detection performance of the detection method, however, it should be noted that anomaly types of unusual consumption are difficult to be detected from experience.

Nevertheless, we note that these results are associated with certain limitations. One limitation is that we only evaluate the proposed strategies with the selected anomaly detection, anomaly compensation, and forecasting methods since the performance of the proposed strategies highly depends on these methods. For example, forecasting methods vary by design in their sensitivity to anomalies and detection methods may not detect all anomalous data points. While we believe that the different selected methods are a representative sample of existing methods, it would be interesting to extend the evaluation to further anomaly detection, anomaly compensation, and forecasting methods. The performance of the selected methods additionally depends on the used hyperpa-

rameters. Although the hyperparameters used are carefully selected, their optimal choice could be investigated. Moreover, the reported results are based on the selected data. Although we perceive the selected time series based on our domain knowledge as comparatively anomaly-free, it could contain anomalies that influence the results. For example, the contained anomalies could worsen the results of the raw strategy and improve the results of the detection and compensation strategies. However, in this case, the relative comparison of the strategies would remain the same. Nevertheless, future work could examine more closely whether anomalies are contained and affect the results. In addition to contained anomalies, the results also depend on the inserted anomalies and might change with different numbers and types of inserted anomalies. Future work could thus also examine the influence of the inserted anomalies on the results. Furthermore, the data used for the evaluation represents the electrical power consumption on a client level. In future work, it might, therefore, be interesting to use other data to investigate how the aggregation level of the data influences the results.

Overall, we conclude from the performed evaluation that the compensation strategy is generally beneficial as it mostly allows for better or at least similar forecasting results as the other evaluated strategies when the input data contains anomalies. By favoring precise forecasts, the compensation strategy provides a means for appropriately managing anomalies in forecasts using energy time series, which could also be beneficial for automated machine learning forecasting.

6 Conclusion

In the present paper, we evaluate three general strategies for managing anomalies in automated energy time series forecasting, namely the raw, the detection, and the compensation strategy. For the evaluation, we apply a representative selection of forecasting methods to real-world data containing inserted synthetic anomalies in order to compare these strategies regarding the obtained forecast accuracy. We also present an example of how these strategies work and perform.

Despite requiring additional computational costs, the compensation strategy is generally beneficial as it mostly outperforms the detection and the raw strategy when the input data contains anomalies.

Given the proposed strategies for managing anomalies in energy time series, future work could address several follow-up questions. For example, future work could verify the results by applying other data including labeled data, anomaly detection methods, and anomaly compensation methods. Similarly, future work could evaluate the proposed strategies with further forecasting methods. Furthermore, future work could integrate the proposed strategies into existing approaches for automated machine learning to include them in the optimization problem of finding the best forecast for a given data set.

Acknowledgments. This project is funded by the Helmholtz Association’s Initiative and Networking Fund through Helmholtz AI and the Helmholtz Association under the Program “Energy System Design”. The authors thank Nicole Ludwig, Benedikt Heidrich, and Kaleb Phipps for their valuable input during the work on this project.

Appendix 1: Inserted Synthetic Anomalies

For the evaluation, we insert anomalies of the two anomaly groups used in [44], namely technical faults in the metering infrastructure and unusual consumption. For both groups, we consider four types of anomalies each. Each anomaly $\hat{p}_{j,i}$ of type j has a start index i and is inserted in the given power time series $P = p_1, p_2, \dots, p_N$ with length N .

Technical Faults

The considered technical faults comprise the anomalies of types 1 to 4 taken from [45]. Figure 6 shows an example of each of these types, which we define in the following.

Anomaly Type 1

$$\hat{p}_{1,i+n} = \begin{cases} -1 \cdot \text{mean}(P) + r_s \cdot \text{std}(P), & n = 0 \\ 0, & 0 < n < l - 1 \\ \sum_{t=1}^{i+l-1} p_t, & n = l - 1, \end{cases} \quad (13)$$

where the length $l \sim \mathcal{U}_{[5,24]}$ and the random scaling factor $r_s = 2 + r \cdot 3$ with $r \sim \mathcal{U}_{[0,1]}$.

Anomaly Type 2

$$\hat{p}_{2,i+n} = \begin{cases} 0, & 0 \leq n < l - 1 \\ \sum_{t=i}^{i+l-1} p_t, & n = l - 1, \end{cases} \quad (14)$$

where the length $l \sim \mathcal{U}_{[5,24]}$.

Anomaly Type 3

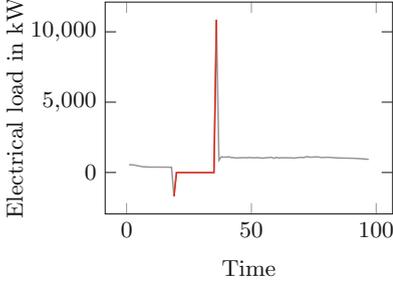
$$\hat{p}_{3,i} = -r_s \cdot \text{mean}(P), \quad (15)$$

where the random scaling factor $r_s = 0.01 + r \cdot 3.99$ with $r \sim \mathcal{U}_{[0,1]}$.

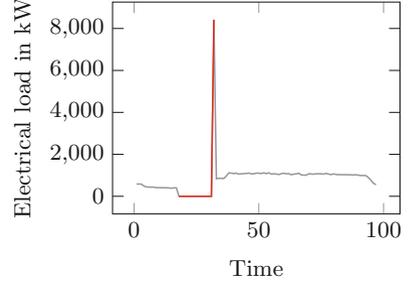
Anomaly Type 4

$$\hat{p}_{4,i} = r \cdot \text{mean}(P), \quad (16)$$

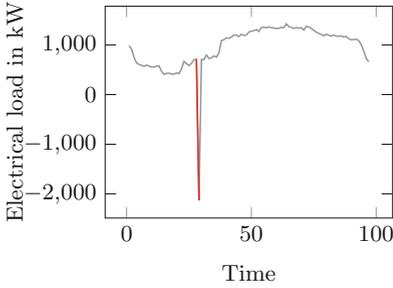
where the random scaling factor $r_s = 3 + r \cdot 5$ with $r \sim \mathcal{U}_{[0,1]}$.



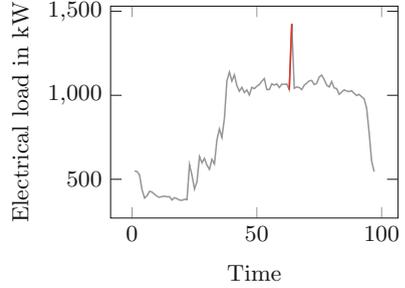
(a) Anomaly type 1: negative power spike followed by zero values and positive spike.



(b) Anomaly type 2: zero power values followed by a positive spike.



(c) Anomaly type 3: negative power spike.



(d) Anomaly type 4: positive power spike.

Fig. 6. Examples of the anomaly types 1 to 4 from the technical faults taken from [45] that we insert as synthetic anomalies into the selected data. The anomalies are plotted in red. Note that the anomalies of types 3 and 4 actually have a length of one but are marked together with their previous value to be recognizable. (Color figure online)

Unusual Consumption

The considered unusual consumption comprise the anomalies of types 5 to 8 taken from [44]. Figure 7 shows an example of each of these types, which we define in the following.

Anomaly Type 5

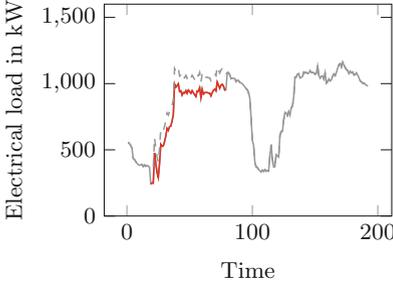
$$\hat{p}_{5,i+n} = p_i - r \cdot p_{\min}, \quad 0 < n < l - 1, \quad (17)$$

where the length $l \sim \mathcal{U}_{[48,144]}$, the random scaling factor $r \sim \mathcal{U}_{[0.3,0.8]}$, and $p_{\min} = \min\{p_i, p_{i+1}, \dots, p_{i+l-1}\}$.

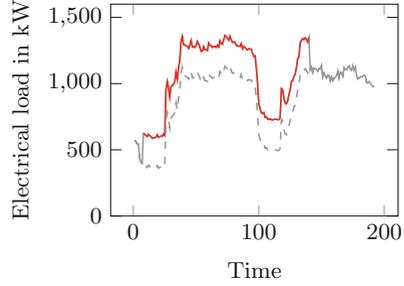
Anomaly Type 6

$$\hat{p}_{6,i+n} = p_i + r \cdot p_{\min}, \quad 0 < n < l - 1, \quad (18)$$

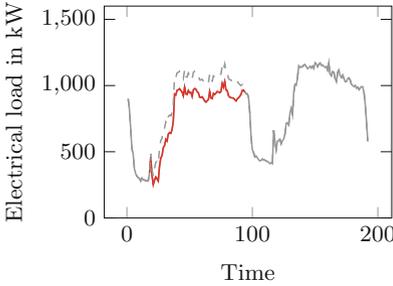
where the length $l \sim \mathcal{U}_{[48,144]}$, the random scaling factor $r \sim \mathcal{U}_{[0.5,1]}$, and $p_{\min} = \min\{p_i, p_{i+1}, \dots, p_{i+l-1}\}$.



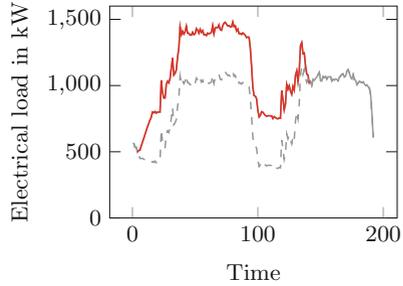
(a) Anomaly type 5: abrupt small temporary reduction in the power values.



(b) Anomaly type 6: abrupt small temporary increase in the power values.



(c) Anomaly type 7: small temporary reduction in the power values with a gradual start and end.



(d) Anomaly type 8: small temporary increase in the power values with a gradual start and end.

Fig. 7. Examples of the anomaly types 5 to 8 from the unusual consumption taken from [44] that we insert as synthetic anomalies into the selected data. The anomalies are plotted in red. (Color figure online)

Anomaly Type 7

$$\hat{p}_{7,i} = \begin{cases} p_i - r \cdot p_{\min} \cdot \frac{l}{10} \cdot i, & 0 < n < \frac{l}{10} \\ p_i - r \cdot p_{\min}, & \frac{l}{10} \leq n \leq 1 - \frac{l}{10} \\ p_i - r \cdot p_{\min} \cdot \frac{l}{10} \cdot (1 - i), & 1 - \frac{l}{10} < n < l - 1, \end{cases} \quad (19)$$

where the length $l \sim \mathcal{U}_{[48,144]}$, the random scaling factor $r \sim \mathcal{U}_{[0.3,0.8]}$, and $p_{\min} = \min\{p_i, p_{i+1}, \dots, p_{i+l-1}\}$.

Anomaly Type 8

$$\hat{p}_{8,i} = \begin{cases} p_i + r \cdot p_{\min} \cdot \frac{l}{10} \cdot i, & 0 < n < \frac{l}{10} \\ p_i + r \cdot p_{\min}, & \frac{l}{10} \leq n \leq 1 - \frac{l}{10} \\ p_i + r \cdot p_{\min} \cdot \frac{l}{10} \cdot (1 - i), & 1 - \frac{l}{10} < n < l - 1, \end{cases} \quad (20)$$

where the length $l \sim \mathcal{U}_{[48,144]}$, the random scaling factor $r \sim \mathcal{U}_{[0.5,1]}$, and $p_{\min} = \min\{p_i, p_{i+1}, \dots, p_{i+l-1}\}$.

Appendix 2: Applied Anomaly Detection

Applied Latent Space-Based Anomaly Detection. To enhance their detection performance, we apply the selected anomaly detection methods to the latent space representation of the selected data which we create by a trained generative method (see Fig. 8).

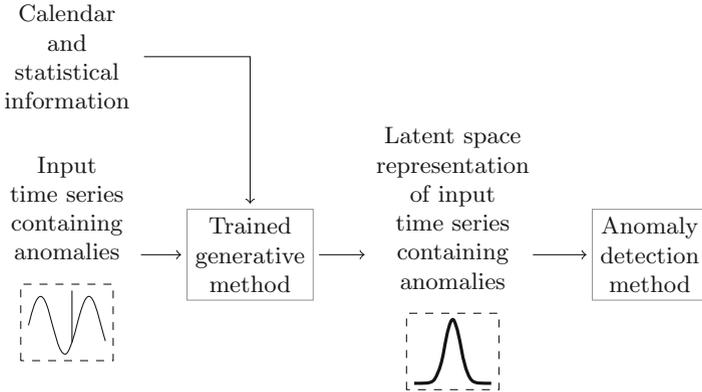


Fig. 8. According to the selected anomaly detection approach from [44], a trained generative method creates the latent space data representation of an input time series containing anomalies. The latent space data representation then serves as input to an anomaly detection method.

Architecture of Generative Methods. For our evaluation, we use the conditional Invertible Neural Network (cINN) [6] and the conditional Variational Autoencoder (cVAE) [41] as described in [22, 44] and as detailed in Tables 1 and 4.

Table 1. Architecture of the used cINN [44].

Element	Description
Number of blocks	10
Layers per block	Glow coupling layer and random permutation
Subnetwork in each block	Fully connected NN (see Table 2)
Conditioning network	Fully connected NN (see Table 3)

Training of Generative Methods. The training of the used cINN and cVAE follows the training described in [44]: We apply the unsupervised cINN and cVAE to the data with inserted synthetic anomalies under the assumption of 10% of the data points are anomalies by setting the contamination parameter of the unsupervised cINN and cVAE to 0.1. Both generative methods obtain standardized data points of the selected time series as samples with a size of 96. Both generative methods also use the mean of the considered time series sample as statistical information as well as the hour of the day, the month of the year, and the weekday as calendar information.

Table 2. Details of the subnetwork in the used cINN [44].

Layer	Description
Input	[Output of previous coupling layer, conditional information]
1	Dense 32 neurons; activation: tanh
2	Dense horizon neurons; activation: linear

Table 3. Details of the conditioning network in the used cINN [44].

Layer	Description
Input	[Calendar information, statistical information]
1	Dense 8 neurons; activation: tanh
2	Dense 4 neurons; activation: linear

Table 4. Architecture of the encoder and decoder in the used cVAE [44].

(a) Encoder	
Layer	Description
Input	[Normal data, conditional information]
1	Dense 64 neurons; activation: tanh
2	Dense 32 neurons; activation: tanh
3	μ : dense latent dimension; activation: linear
4	σ : dense latent dimension; activation: linear
(b) Decoder	
Layer	Description
Input	[Latent data, conditional information]
1	Dense 32 neurons; activation: tanh
2	Dense 64 neurons; activation: tanh
3	Dense horizon neurons; activation: linear

Appendix 3: Applied Forecasting Methods

Table 5. Overview of the used calendar information.

Calendar information	Implementation
Weekday	Boolean
Workdays (Monday to Friday)	Boolean
Hour of the day	$\sin(2 \cdot \pi \cdot \text{hour}/24)$ and $\cos(2 \cdot \pi \cdot \text{hour}/24)$
Day of the month	$\sin(2 \cdot \pi \cdot \text{day}/\text{days of the month})$
Month of the year	$\sin(2 \cdot \pi \cdot \text{month}/12)$ and $\cos(2 \cdot \pi \cdot \text{month}/12)$

Table 6. Details of the applied NN.

Layer	Description
Input	[Load data, encoded calendar information]
1	Dense 256 neurons; activation: relu
2	Dense 128 neurons; activation: relu
Output	Dense 96 neuron; activation: linear

Appendix 4: Hard- and Software

Hardware. The used hardware is an HPC system with two Intel Xeon Gold 5118 CPUs and with 256 GB RAM.

Software. For the anomaly detection using the respective latent space data representation created by the selected cINN or cVAE, we apply the implementation described in [44]. It uses FrEIA⁴ and PyTorch⁵ [34] for the cINN, PyTorch [34] for the cVAE, Keras⁶ [15] for the VAE, and scikit-learn⁷ [35] for the LOF.

For the anomaly compensation, we apply the implementation of the Prophet-based method described in [48] that is based on the available Prophet implementation⁸ [42].

For the forecasting methods, we use Keras for the NN and scikit-learn for the LinR, SVR, and RF Regressor. Additionally, we apply the available implementation⁹ [12] for the XGBoost Regressor, and adapt the available implementation of the PNN¹⁰ [21] to work without weather data.

We finally use pyWATTS¹¹ [20] to implement the proposed strategies and to automate their evaluation.

References

1. Ahmed, N.K., Atiya, A.F., El Gayar, N., El-Shishiny, H.: An empirical comparison of machine learning models for time series forecasting. *Economet. Rev.* **29**(5–6), 594–621 (2010). <https://doi.org/10.1080/07474938.2010.481556>
2. Akouemo, H.N., Povinelli, R.J.: Probabilistic anomaly detection in natural gas time series data. *Int. J. Forecast.* **32**(3), 948–956 (2016). <https://doi.org/10.1016/j.ijforecast.2015.06.001>

⁴ <https://github.com/VLL-HD/FrEIA>.

⁵ <https://pytorch.org/>.

⁶ <https://keras.io/>.

⁷ <https://scikit-learn.org/>.

⁸ <https://facebook.github.io/prophet/>.

⁹ <https://xgboost.ai/>.

¹⁰ <https://github.com/benHeid/Profile-Neural-Network>.

¹¹ <https://github.com/KIT-IAI/pyWATTS>.

3. Akouemo, H.N., Povinelli, R.J.: Data improving in time series using ARX and ANN models. *IEEE Trans. Power Syst.* **32**(5), 3352–3359 (2017). <https://doi.org/10.1109/TPWRS.2017.2656939>
4. Alahakoon, D., Yu, X.: Smart electricity meter data intelligence for future energy systems: a survey. *IEEE Trans. Industr. Inf.* **12**(1), 425–436 (2016). <https://doi.org/10.1109/TII.2015.2414355>
5. Alasadi, S.A., Bhaya, W.S.: Review of data preprocessing techniques in data mining. *J. Eng. Appl. Sci.* **12**(16), 4102–4107 (2017). <https://doi.org/10.36478/jeasci.2017.4102.4107>
6. Ardizzone, L., Lüth, C., Kruse, J., Rother, C., Köthe, U.: Guided image generation with conditional invertible neural networks. [arXiv:1907.02392](https://arxiv.org/abs/1907.02392) (2019)
7. Breiman, L.: Random forests. *Mach. Learn.* **45**, 5–32 (2001). <https://doi.org/10.1023/A:1010933404324>
8. Breunig, M.M., Kriegel, H.P., Ng, R.T., Sander, J.: LOF: identifying density-based local outliers. In: *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, pp. 93–104. ACM (2000). <https://doi.org/10.1145/342009.335388>
9. Chakhchoukh, Y., Panciatici, P., Mili, L.: Electric load forecasting based on statistical robust methods. *IEEE Trans. Power Syst.* **26**(3), 982–991 (2011). <https://doi.org/10.1109/TPWRS.2010.2080325>
10. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. *ACM Comput. Surv.* **41**(3), 15:1–15:58 (2009). <https://doi.org/10.1145/1541880.1541882>
11. Charlton, N., Singleton, C.: A refined parametric model for short term load forecasting. *Int. J. Forecast.* **30**(2), 364–368 (2014). <https://doi.org/10.1016/j.ijforecast.2013.07.003>
12. Chen, T., Guestrin, C.: XGBoost: a scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794. ACM (2016). <https://doi.org/10.1145/2939672.2939785>
13. Chen, W., Zhou, K., Yang, S., Wu, C.: Data quality of electricity consumption data in a smart grid environment. *Renew. Sustain. Energy Rev.* **75**, 98–105 (2017). <https://doi.org/10.1016/j.rser.2016.10.054>
14. Chen, X., Kang, C., Tong, X., Xia, Q., Yang, J.: Improving the accuracy of bus load forecasting by a two-stage bad data identification method. *IEEE Trans. Power Syst.* **29**(4), 1634–1641 (2014). <https://doi.org/10.1109/TPWRS.2014.2298463>
15. Chollet, F., et al.: *Keras* (2015). <https://keras.io>
16. Dannecker, L.: *Energy Time Series Forecasting*. Springer Fachmedien, Wiesbaden (2015). <https://doi.org/10.1007/978-3-658-11039-0>
17. Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support vector regression machines. In: *Mozer, M.C., Jordan, M., Petsche, T. (eds.) Advances in Neural Information Processing Systems*, vol. 9. MIT Press, Cambridge (1996)
18. Dua, D., Graff, C.: *UCI machine learning repository* (2019). <https://archive.ics.uci.edu/ml>
19. Fan, C., Chen, M., Wang, X., Wang, J., Huang, B.: A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data. *Front. Energy Res.* **9**, 652801 (2021). <https://doi.org/10.3389/fenrg.2021.652801>
20. Heidrich, B., et al.: *pyWATTS: Python workflow automation tool for time series*. [arXiv:2106.10157](https://arxiv.org/abs/2106.10157) (2021)

21. Heidrich, B., Turowski, M., Ludwig, N., Mikut, R., Hagenmeyer, V.: Forecasting energy time series with profile neural networks. In: The Eleventh ACM International Conference on Future Energy Systems (e-Energy '20), pp. 220–230 (2020). <https://doi.org/10.1145/3396851.3397683>
22. Heidrich, B., et al.: Controlling non-stationarity and periodicities in time series generation using conditional invertible neural networks. *Appl. Intell.* **53**, 8826–8843 (2023). <https://doi.org/10.1007/s10489-022-03742-7>
23. Jiao, J., Tang, Z., Zhang, P., Yue, M., Yan, J.: Cyberattack-resilient load forecasting with adaptive robust regression. *Int. J. Forecast.* **38**(3), 910–919 (2022). <https://doi.org/10.1016/j.ijforecast.2021.06.009>
24. Jokar, P., Arianpoo, N., Leung, V.C.M.: Electricity theft detection in AMI using customers' consumption patterns. *IEEE Trans. Smart Grid* **7**(1), 216–226 (2016). <https://doi.org/10.1109/TSG.2015.2425222>
25. Kingma, D.P., Ba, J.L.: Adam: a method for stochastic optimization. In: 3rd International Conference on Learning Representations (ICLR 2015) (2015)
26. Kingma, D.P., Welling, M.: Auto-encoding variational Bayes. [arXiv:1312.6114v10](https://arxiv.org/abs/1312.6114v10) (2014)
27. Luo, J., Hong, T., Fang, S.C.: Benchmarking robustness of load forecasting models under data integrity attacks. *Int. J. Forecast.* **34**(1), 89–104 (2018). <https://doi.org/10.1016/j.ijforecast.2017.08.004>
28. Luo, J., Hong, T., Fang, S.C.: Robust regression models for load forecasting. *IEEE Trans. Smart Grid* **10**(5), 5397–5404 (2018). <https://doi.org/10.1109/TSG.2018.2881562>
29. Luo, J., Hong, T., Gao, Z., Fang, S.C.: A robust support vector regression model for electric load forecasting. *Int. J. Forecast.* **39**(2), 1005–1020 (2023). <https://doi.org/10.1016/j.ijforecast.2022.04.001>
30. Luo, J., Hong, T., Yue, M.: Real-time anomaly detection for very short-term load forecasting. *J. Mod. Power Syst. Clean Energy* **6**(2), 235–243 (2018). <https://doi.org/10.1007/s40565-017-0351-7>
31. Mitchell, T.M.: *Machine Learning*. McGraw-Hill, New York (1997)
32. Nordahl, C., Persson, M., Grahn, H.: Detection of residents' abnormal behaviour by analysing energy consumption of individual households. In: 2017 IEEE International Conference on Data Mining Workshops (ICDMW), pp. 729–738. IEEE (2017). <https://doi.org/10.1109/ICDMW.2017.101>
33. Park, S., Jung, S., Jung, S., Rho, S., Hwang, E.: Sliding window-based LightGBM model for electric load forecasting using anomaly repair. *J. Supercomput.* **77**(11), 12857–12878 (2021). <https://doi.org/10.1007/s11227-021-03787-4>
34. Paszke, A., et al.: PyTorch: an imperative style, high-performance deep learning library. In: Wallach, H., Larochelle, H., Beygelzimer, A., D'Alché-Buc, F., Fox, E., Garnett, R. (eds.) *Advances in Neural Information Processing Systems*, vol. 32 (2019)
35. Pedregosa, F., et al.: Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011)
36. Quintana, M., Stoeckmann, T., Park, J.Y., Turowski, M., Hagenmeyer, V., Miller, C.: ALDI++: automatic and parameter-less discord and outlier detection for building energy load profiles. *Energy Build.* **265**, 112096 (2022). <https://doi.org/10.1016/j.enbuild.2022.112096>
37. Ranjan, K.G., Prusty, B.R., Jena, D.: Review of preprocessing methods for univariate volatile time-series in power system applications. *Electr. Power Syst. Res.* **191**, 106885 (2021). <https://doi.org/10.1016/j.epsr.2020.106885>

38. Rodrigues, F., Trindade, A.: Load forecasting through functional clustering and ensemble learning. *Knowl. Inf. Syst.* **57**(1), 229–244 (2018). <https://doi.org/10.1007/s10115-018-1169-y>
39. Rossi, B., Chren, S.: Smart grids data analysis: a systematic mapping study. *IEEE Trans. Industr. Inf.* **16**(6), 3619–3639 (2020). <https://doi.org/10.1109/TII.2019.2954098>
40. Seem, J.E.: Using intelligent data analysis to detect abnormal energy consumption in buildings. *Energy Build.* **39**(1), 52–58 (2007). <https://doi.org/10.1016/j.enbuild.2006.03.033>
41. Sohn, K., Yan, X., Lee, H.: 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*, vol. 28, pp. 3483–3491 (2015)
42. Taylor, S.J., Letham, B.: Forecasting at scale. *Am. Stat.* **72**(1), 37–45 (2018). <https://doi.org/10.1080/00031305.2017.1380080>
43. Teng, S.Y., Máša, V., Touš, M., Vondra, M., Lam, H.L., Stehlík, P.: Waste-to-energy forecasting and real-time optimization: an anomaly-aware approach. *Renew. Energy* **181**, 142–155 (2022). <https://doi.org/10.1016/j.renene.2021.09.026>
44. Turowski, M., et al.: Enhancing anomaly detection methods for energy time series using latent space data representations. In: *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*, pp. 208–227. ACM (2022). <https://doi.org/10.1145/3538637.3538851>
45. Turowski, M., et al.: Modeling and generating synthetic anomalies for energy and power time series. In: *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22)*, pp. 471–484. ACM (2022). <https://doi.org/10.1145/3538637.3539760>
46. Wang, L., et al.: Point and contextual anomaly detection in building load profiles of a university campus. In: *2020 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*, pp. 11–15 (2020). <https://doi.org/10.1109/ISGT-Europe47291.2020.9248792>
47. Wang, Y., Chen, Q., Hong, T., Kang, C.: Review of smart meter data analytics: applications, methodologies, and challenges. *IEEE Trans. Smart Grid* **10**(3), 3125–3148 (2019). <https://doi.org/10.1109/TSG.2018.2818167>
48. Weber, M., Turowski, M., Çakmak, H.K., Mikut, R., Kühnapfel, U., Hagenmeyer, V.: Data-driven copy-paste imputation for energy time series. *IEEE Trans. Smart Grid* **12**(6), 5409–5419 (2021). <https://doi.org/10.1109/TSG.2021.3101831>
49. Werbos, P.J.: *Beyond regression: new tools for prediction and analysis in the behavioral sciences*. Ph.D. thesis, Harvard University (1974)
50. Xie, J., Hong, T.: GEFCom2014 probabilistic electric load forecasting: an integrated solution with forecast combination and residual simulation. *Int. J. Forecast.* **32**(3), 1012–1016 (2016). <https://doi.org/10.1016/j.ijforecast.2015.11.005>
51. Yue, M., Hong, T., Wang, J.: Descriptive analytics-based anomaly detection for cybersecure load forecasting. *IEEE Trans. Smart Grid* **10**(6), 5964–5974 (2019). <https://doi.org/10.1109/TSG.2019.2894334>
52. Zhang, Y., Lin, F., Wang, K.: Robustness of short-term wind power forecasting against false data injection attacks. *Energies* **13**(15) (2020). <https://doi.org/10.3390/en13153780>
53. Zheng, R., Gu, J., Jin, Z., Peng, H., Zhu, Y.: Load forecasting under data corruption based on anomaly detection and combined robust regression. *Int. Trans. Electr. Energy Syst.* **30**(7), e12103 (2020). <https://doi.org/10.1002/2050-7038.12103>

54. Zhou, Y., Ding, Z., Wen, Q., Wang, Y.: Robust load forecasting towards adversarial attacks via Bayesian learning. *IEEE Trans. Power Syst.* **8950** (2022). <https://doi.org/10.1109/TPWRS.2022.3175252>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

