

# Anomaly Detection Algorithm Using a Hybrid Modelling Approach for Energy Consumption Time Series



Florian Rippstein, Steve Lenk, Andre Kummerow, Lucas Richter, Stefan Klaiber, and Peter Bretschneider

**Abstract** Many energy time series captured by real-time systems contain errors or anomalies that prevent accurate forecasts of time series evolution. However, accurate forecasting of load time series and fluctuating renewable energy feed-in as well as subsequent optimisation of the dispatch of controllable generators, storage and loads is crucial to ensure a cost-effective, sustainable and reliable energy supply. Therefore, we investigate methods and approaches for a system solution that automatically detect and replace anomalies in time series to enable accurate forecasts. Here, we introduce a hybrid anomaly detection system for energy consumption time series, which consists of two different neural networks (Seq2Seq and autoencoder) and two more classical approaches (entropy, SVM classification). This network is able to detect different types of anomalies, namely, outliers, zero points, incomplete data, change points and anomalous (parts of) time series. These types are defined for the first time mathematically. Our results show a clear advantage of the hybrid modelling approach for detecting anomalies in previously unknown energy time series compared to the single approaches. In addition, due to the generalisation capability of the hybrid model, our approach allows very good estimation of energy values without requiring a large amount of historical data to train the model.

**Keywords** Anomaly detection · Energy consumption · Time series processing · Seq2Seq · Autoencoder hybrid neural network

---

F. Rippstein (✉) · S. Lenk · A. Kummerow · L. Richter · S. Klaiber · P. Bretschneider  
Fraunhofer-Institut Optronik, Systemtechnik und Bildauswertung (IOSB), Ilmenau, Germany  
e-mail: [Florian.Rippstein@iosb-ast.fraunhofer.de](mailto:Florian.Rippstein@iosb-ast.fraunhofer.de); [Steve.Lenk@iosb-ast.fraunhofer.de](mailto:Steve.Lenk@iosb-ast.fraunhofer.de);  
[Andre.Kummerow@iosb-ast.fraunhofer.de](mailto:Andre.Kummerow@iosb-ast.fraunhofer.de); [Lucas.Richter@iosb-ast.fraunhofer.de](mailto:Lucas.Richter@iosb-ast.fraunhofer.de);  
[Stefan.Klaiber@iosb-ast.fraunhofer.de](mailto:Stefan.Klaiber@iosb-ast.fraunhofer.de); [Peter.Bretschneider@iosb-ast.fraunhofer.de](mailto:Peter.Bretschneider@iosb-ast.fraunhofer.de)

## 1 Introduction

Many energy data sets of real-time systems include errors or anomalies, which hinder an appropriate prediction. However, the prediction and the following optimisation of energy load, generation and storage are crucial to prevent blackouts or brownouts due to unbalanced fluctuations in the energy grid [9]. For critical infrastructures, e.g. the energy sector, new challenges arise due to the increasing amount of data to handle, the increasing automation level and possible threats by cyberattacks. Thus, resilience, i.e. to be prepared for and to prevent threats, to protect systems against them, to respond to threats and to recover from them, became more and more important.

Therefore, we study a system which automatically detects and replaces anomalies in time series to enable accurate predictions.

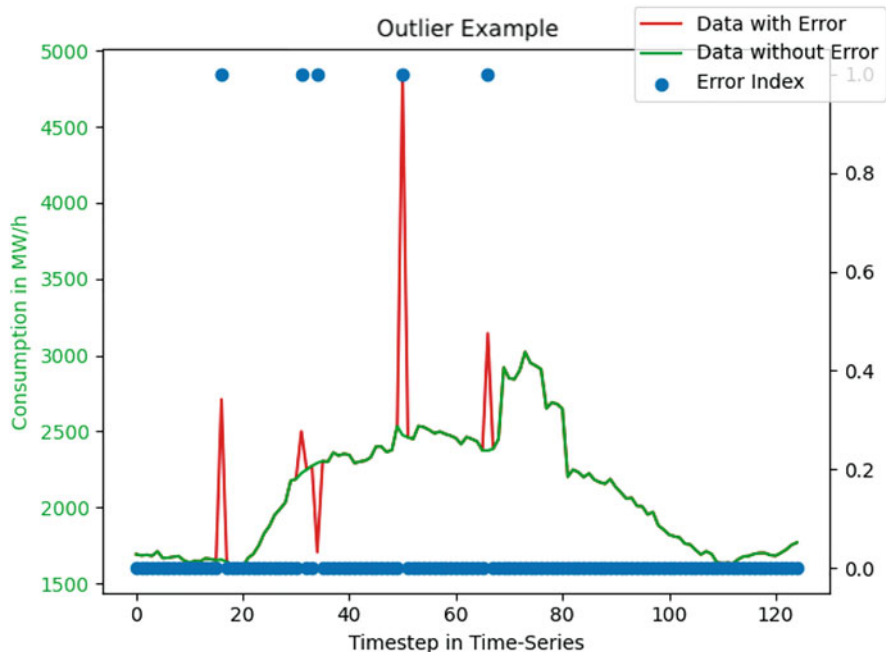
Thereby, we define anomalies as data, which do not belong to the normal characteristics of time series, whereas errors are normal or anomalous parts of time series, which are known to be erroneous due to external information, e.g. information of fallen power pole.

To classify anomalies, we distinguish outliers, zero points, incomplete data, change points and anomalous (part of) time series similarly to [3, 10], but we concentrated their definitions mathematically (see Sect. 2). To study our detection methods, we manipulated real, highly accumulated energy consumption time series, which were manually verified and corrected [1].

An example is shown in Fig. 1 in which a part of such an accumulated energy consumption time series [1] (green) is shown. A classical approach to detect anomalies is to calculate the difference between a prediction and an observation [15]. This difference is called “surprise” by Goldberg et al. [4] and is calculated as the difference between the true and the observed values. Unfortunately, this approach is only applicable if a precise prediction can be calculated, which in case of a regression needs sufficient amount of data. Alternatively, neural networks show good results using unknown data, either by default or by techniques such as domain adaption [16].

Three approaches to detect anomalies in energy data sets were suggested by Zhang et al. [19], namely, using Shannon entropy, classification or a regression approach. For unknown data sets, the regression approach is obviously inadequate since the amount of training data is too small. However, using the well-known Shannon entropy from information theory [12] to measure the surprise or uncertainty of data points in a time series, it is possible to detect anomalous data points in previously unknown time series to a limited amount. The entropy  $H$  is calculated as:

$$H(x) = \sum_{i=1}^n p(x_i) \log_b p(x_i),$$



**Fig. 1** Example of an anomalous time series including outliers with different anomaly delta

where  $p$  is the probability of the energy consumption  $x$ . We also have  $b$  as the base of the logarithm. The two common used bases are 10 or 2 [12]. However, this measured accuracy and precision is not as high as a regression approach.

A neural network approach can be created by using Seq2Seq networks, which are able to predict values of time series [5, 6]. Thus, we can classify by using the surprise.

Autoencoders otherwise show strong in the reconstruction of data in general [14] and also in time series [11]. Hence, it can also be used to evaluate a time series by calculating a surprise based on the reconstruction error. Furthermore, support vector machines (SVM) have a strong theoretical foundation and are fast implementable to classify data. Yet, SVM have some disadvantages, like overfitting and the need for labelled data, which are the common weaknesses of supervised learning. Additionally, SVM needs good kernel (function) to separate between classes [17], i.e. normal data and anomalies.

To overcome the limitations and drawbacks of these approaches, a hybrid model was developed for all defined anomalies.

## 2 Our Definitions

In general, we consider a time series  $X$  as a sequence of  $n$ -tuples:

$$((c_1, t_1), \dots, (c_n, t_n)).$$

The discussed anomalies are defined in the following:

**Definition 1 (Noise Data)** Noise data is incomprehensible for either computers or unstructured data. These can be logical errors or inconsistent data [3], e.g. string in databases, not detected bit flips.

**Definition 2 (Outlier)** A time series  $X^*$  with outlier can be created by modifying tuples of  $X$  by multiplying  $c_i$  with factor  $o_i \in \mathbb{R}_0^+ \setminus [0.9, \dots, 1.1]$  to the left elements of the chosen tuples were the predecessor and successor of the single tuples are not modified, i.e.

$$(o_i * c_i, t_i), \text{ where as } i \in \{2, \dots, n - 1\}$$

Then the modified tuple is an outlier.

**Definition 3 (Zero Point)** Based on Definition 2, an outlier is called zero point if the modifying factor  $o_i$  is 0 instead.

**Definition 4 (Change Point)** For given time series  $X$  is  $2 \leq m \leq n - 2$ . Then a time series  $X^*$  with change points can be created by replacing a consecutive  $m$ -sub-sequence of  $X$  by  $o_i \in \mathbb{R}_0^+$ . Additionally, the first modifier  $o_j$  of the sub-sequence has to satisfy  $o_j \notin [0.9, \dots, 1.1]$ , to the left elements of the chosen tuples were the predecessor and successor of this  $m$ -sub-sequence are not modified, i.e.

$$(o_i * c_i, t_i), \text{ where as } i \in \{j, \dots, j + m - 1\}, |o_i - 1| > |o_{i+1} - 1| \text{ and:}$$

$$o_i > 1 \text{ and } o_{i+1} > 1 \text{ or}$$

$$o_i < 1 \text{ and } o_{i+1} < 1, \quad \forall i \in \{j, \dots, j + m - 1\}.$$

The points of this consecutive  $m$ -sub-sequence are called change points.

**Definition 5 (Incomplete Data)** For given time series  $X$  is  $2 \leq m \leq n - 2$ . A time series  $X^*$  with incomplete data can be created by replacing a consecutive  $m$ -sub-sequence of  $X$  by using factors  $o_i \in \mathbb{R}_0^+ \setminus [0.9, \dots, 1.1]$ , with  $o_j$  being the first modifier of the  $m$ -sub-sequence and  $o_j = o_i$ , where  $i \in \{j, \dots, j + m - 1\}$ ,

to the left elements of the chosen tuples were the predecessor and successor of this  $m$ -sub-sequence are not modified, i.e.

$$(o_i * c_i, t_i), \text{ where as } i \in \{j, \dots, j + m - 1\}$$

The points of this consecutive  $m$ -sub-sequence are called incomplete data.

**Definition 6 (Anomalous Time Series/Outlier Type B)** For given time series  $X$  is  $2 \leq m \leq n - 2$ . An anomalous time series  $X^*$  can be created by replacing a consecutive  $m$ -sub-sequence of the  $n$ -sequence  $X$  by multiplying factors  $o_i \in \mathbb{R}$ , with  $o_j$  being the first modifier of the  $m$ -sub-sequence and  $o_i \neq 1$ , where  $i \in \{j, \dots, j + m - 1\}$ , to the left elements of the chosen tuples were the predecessor and successor of this  $m$ -sub-sequence are not modified, and where the sub-sequence is either incomplete data or change point, i.e.

$$(o_i * c_i, t_i), \text{ where as } i \in \{j, \dots, j - m - 1\}$$

The points of this consecutive  $m$ -sub-sequence are called incomplete data.

*Information: Anomalous time series are similar to a set of outliers; therefore, we decided to use the name outlier type B.*

### 3 Our Hybrid Model

Our developed architecture is shown in Fig. 2.

It contains of the two previously mentioned neural networks, an autoencoder and a Seq2Seq networks, and the Shannon entropy and SVM as more classical approaches.

Autoencoder is able to reconstruct time series to find anomalous data points [2]. Thus, autoencoder can be trained to reconstruct a time series, and such a reconstructed time series can be compared with the original time series using the mean squared error (MSE) or alternatives like RMSE to classify them.

We improved this approach by calculating the (squared) difference of every single data point and using this as input for a convolutional neural network (CNN), which is trained together with the autoencoder. The training process utilises loss weight to comply with the fact that a good classification is more important than a good reconstruction. To evaluate a whole time series, we used a rolling window (standard size 24 time stamps) to evaluate each single data point with the single autoencoder.

Additionally, we created a Seq2Seq prediction network similar to the network by Hwang et al. [6]. Seq2Seq networks are well known for their strong capabilities in the field of natural language processing [8].

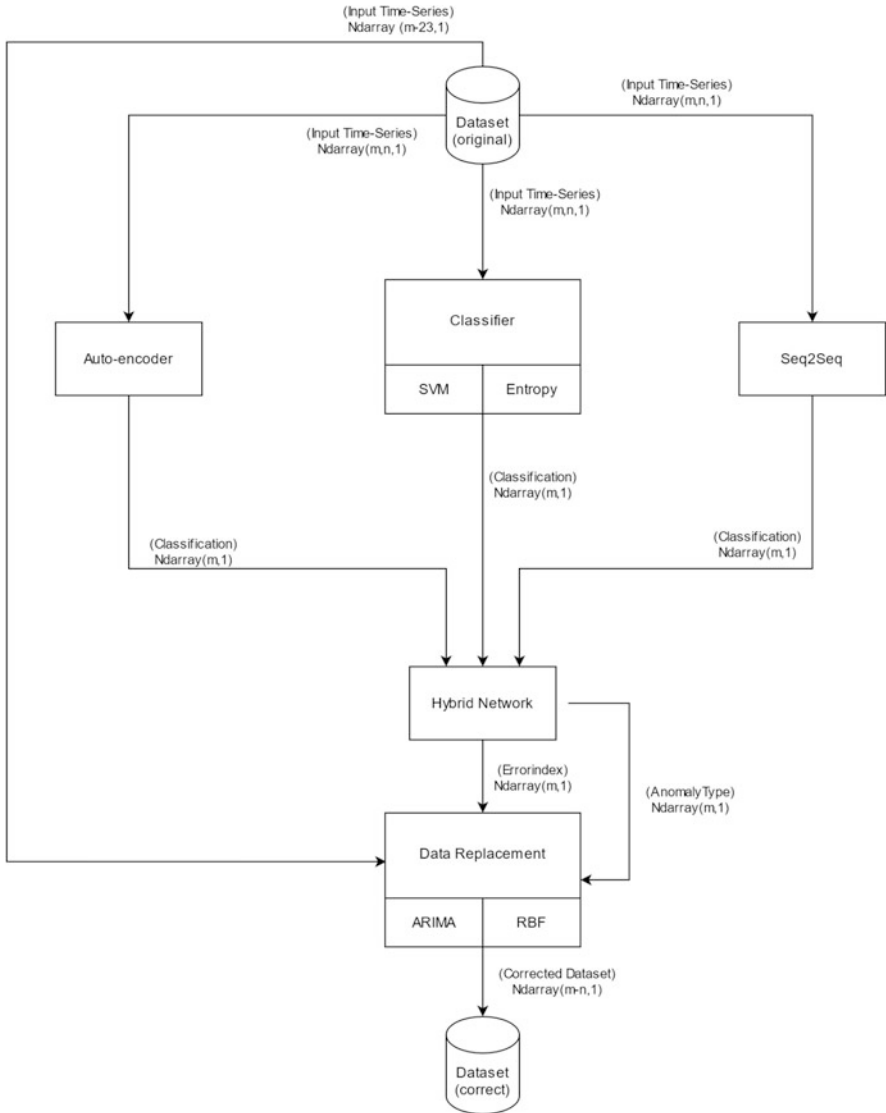


Fig. 2 Our solution

The Seq2Seq networks use the unrolling properties of RNN [13] to evaluate an input. Again, a full time-set evaluation was be done using a rolling window.

By combining the two classical approaches (entropy and SVM) and the two neural networks (autoencoder and Seq2Seq), a hybrid model was built (as shown in Fig. 2), which takes advantage of each of the single approaches. The hybrid network in Fig. 2 itself is a SVM, which evaluates the different results and computes a more

precise final decision. Decision trees or a neural network could be used as well. These approaches have shown similar or even better scores in other tasks [7]. The next step was used to substitute all detected anomalies by using either interpolation, extrapolation or an autoencoder, depending which of those replacement algorithms is suited best for a given time series.

## 4 Results

Before we show the hybrid results, we explain some benefits of our hybrid solution.

In Fig. 3, we plotted the MSE of anomalies and of normal data after reconstruction by the autoencoder as orange and blue lines, respectively. Here, anomalies have a MSE of approx. 1.0, whereas for normal data, it fluctuates around 0.1. A classification based on the plotted MSE was done by using, e.g. 0.4 as the limit for normal data. This approach yields F1-scores around 0.8, but some data points are wrongly classified.

Here, we developed a different approach based on CNN as described in Sect. 3. Instead of using the MSE, we used the squared error in a CNN for each single data point which improved the F1-score. However, the reconstruction result of

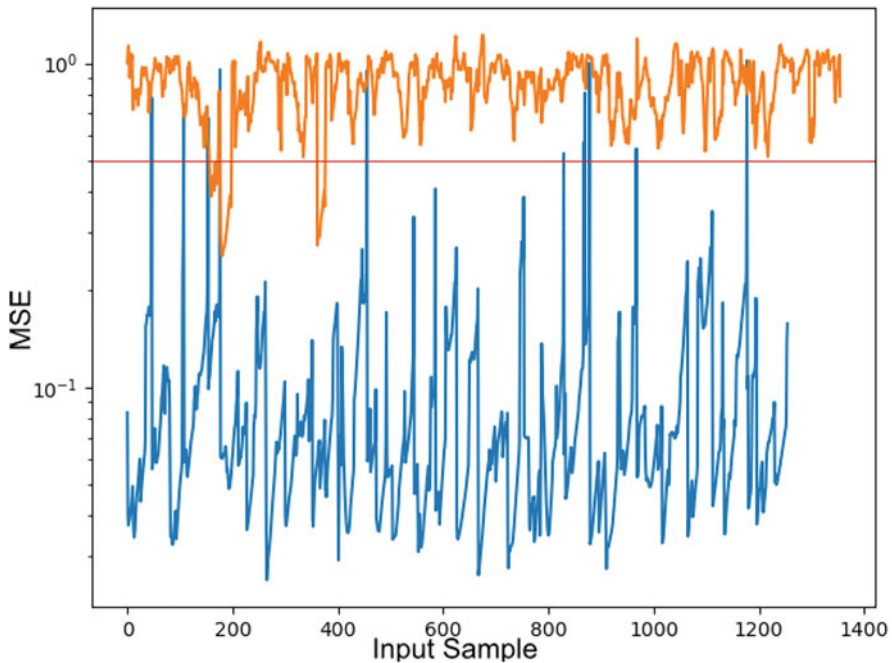
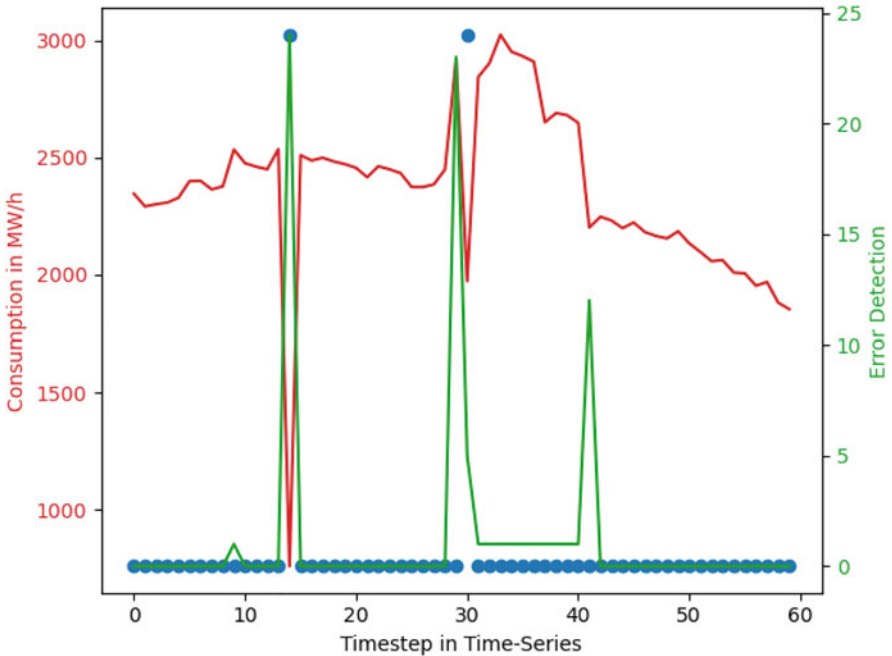


Fig. 3 MSE output of the autoencoder

the autoencoder is no longer usable for replacing the abnormal data, since both networks, autoencoder and CNN, are trained together focusing on MSE for classification. Thus, it will yield a large difference between MSE of normal and abnormal data points but not necessarily anomalies will have a larger MSE.

The Seq2Seq network used the introduced surprise calculating approach. Therefore, the network classifies data by building an internal confidence window [18]. Additionally, we used a similar CNN-based approach as for the autoencoder. This approach showed that the prediction accuracy of a Seq2Seq network depends on the placement of the data point within the sample window, i.e. the closer to the window borders, the worse the prediction accuracy. For better classification results, we combine the different anomaly detection results for a single data point, i.e. 24 decisions for each data point due to a standard rolling window size of 24. The result of an (part of an) energy consumption time series is shown in Fig. 4 as green line. In this figure, the time series is shown as red line and the time stamp of generated anomalies (as a Boolean index in the (not-shown) range between 0 and 1). It is observable that abrupt changes in the time series result in an increased detection rate by the Seq2Seq network as desired. Thus, points with a higher surprise are detected more often than normal data. This Seq2Seq worked to a certain degree as seen in Fig. 4. Here, the network detected a normal spike on data point 30 as outlier, but detected the real outlier only six times. This behaviour is explainable because



**Fig. 4** Seq2Seq output



the network learned that outliers are always single points, and, thus, it is not capable to distinguish correctly between the two data points with high surprise. After adding change points or incomplete data to our train set, this behaviour was not observed anymore. Unfortunately, Seq2Seq networks, trained only with long anomalies, are always detected at least 3 points as anomalies in test with single-point outliers. Our approach of deciding upon majority votes can be used to decrease the amount of false positive or false negative. The hybrid solution is trained on using a higher or lower limit depending on the Seq2Seq networks.

It is notable that the capability of the Seq2Seq network to generalise is not as high as in case of the autoencoder. Therefore, only inter-domain tests can be well detected by the Seq2Seq network. A domain transfer approach is highly recommended to get Seq2Seq networks, which can be usable for a larger variety of data.

So far, we have shown two approaches for detecting anomalies separately, yielding reasonable results, but still improvable ones.

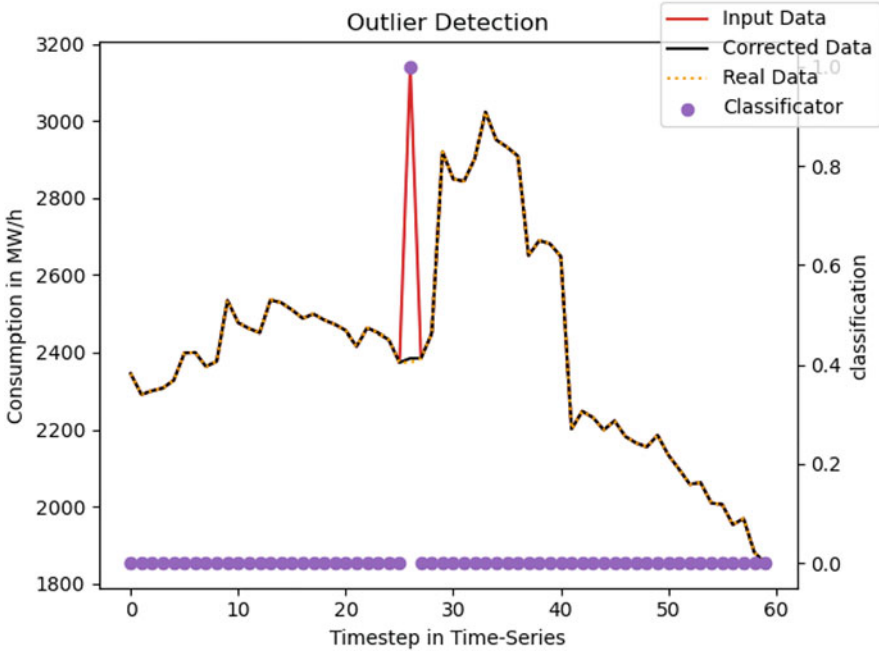
In consequence, this leads to our hybrid network, which combines both approaches. Before presenting the results, we want to emphasise that for the achieved results, our hybrid model was trained with manipulated energy consumption data from Germany and tested it with manipulated consumption data from Austria. So, the evaluation was done with unknown data. The results for the Germany consumption test set showed slightly better results. An example of the F1-score for our networks and the hybrid network can be found in Table 1 and in Fig. 5. Here, we were able to reach F1-scores for outliers above 0.99. Additionally, we studied the influence of the ratio between normal and abnormal values, here called anomaly delta. As shown in Table 1, even anomalies with a deviation of only 5% are detectable by the presented hybrid model. The accuracy for the substitution of outliers is already satisfying as seen in Fig. 5 by comparing the real (broken yellow line) and corrected data (black solid line). The substitution was done with an RBF Interpolation.

If domain adaption techniques were used, the F1-Score of the hybrid solution was decreased by 0.01.

Also the results of the other anomaly types are shown in Table 2.

**Table 1** F1-scores for different anomaly deltas

Anomaly delta	Hybrid result	F1-score	Type
10%	0.9976	0.848	Autoencoder
		0.899	Seq2Seq
7.5%	0.9948	0.748	Autoencoder
		0.812	Seq2Seq
6%	0.9917	0.564	Autoencoder
		0.845	Seq2Seq
5%	0.9908	0.567	Autoencoder
		0.904	Seq2Seq



**Fig. 5** Example of the hybrid solution with anomaly delta of 10%

**Table 2** Comparison of different anomaly types for a delta of 10%

Anomaly type	Autoencoder	Seq2Seq	Hybrid network
Outlier	0.5715	0.8403	0.9941
Incomplete data	0.7970	0.6209	0.8805
Change points	0.8170	0.7623	0.9622

## 5 Summary

We presented a hybrid model approach that uses two classical mathematical approaches and neural networks to detect anomalies and substitute them with an appropriate algorithm. The results showed clear advantages of the hybrid model for detecting anomalies in previously unknown energy time series compared to the single approaches for outliers, but also for other types of anomalies. In addition, due to the generalisation capability of the hybrid model, this approach allows very good estimation of energy values without requiring a large amount of historical data to train the model.

Our anomaly definitions were defined mathematically based on examples of anomalies and will be adapted to better reflect statistical properties of time series and their anomalies in future studies.

**Acknowledgments** The work was financially supported by BMBF (Bundesministeriums für Bildung und Forschung) under the project “reDesigN - Resilience By Design for IoT Platforms in Distributed Energy Management” [1] (support code 01IS18074D) and Fraunhofer Cluster of Excellence Integrated Energy Systems (CINES). The authors want to acknowledge Prof Mäder and M. Sc. Martin Rabe (TU Ilmenau) for their supervision of the master thesis “Automatic energy data processing based on machine learning algorithms” of one of us (F.R.). Additionally, we thank B. Sc. Jonathan Schäfer (FSU Jena) for the fruitful discussions about the mathematical definition of the anomaly types.

## References

1. Bundesnetzagentur.: SMARD | SMARD - Strommarktdaten, Stromhandel und Stromerzeugung in Deutschland, Mar 2021. [Online; accessed 12. Mar. 2021]
2. Chaitanya, C.R.A., Kaplanyan, A.S., Schied, C., Salvi, M., Lefohn, A., Nowrouzehrai, D., Aila, T.: Interactive reconstruction of monte carlo image sequences using a recurrent denoising autoencoder. *ACM Trans. Graph. (TOG)* **36**(4), 1–12 (2017)
3. Chen, W., Zhou, K., Yang, S., Wu, C.: Data quality of electricity consumption data in a smart grid environment. *Renewable and Sustainable Energy Reviews* **75**, 98–105 (2017)
4. Goldberg, D., Shan, Y.: The importance of features for statistical anomaly detection. In: 7th {USENIX} Workshop on Hot Topics in Cloud Computing (HotCloud 15) (2015)
5. Gong, G., An, X., Mahato, N. K., Sun, S., Chen, S., Wen, Y.: Research on short-term load prediction based on seq2seq model. *Energies* **12**(16), 3199 (2019)
6. Hwang, S., Jeon, G., Jeong, J., Lee, J.: A novel time series based seq2seq model for temperature prediction in firing furnace process. *Procedia Comput. Sci.* **155**, 19–26 (2019)
7. Kirkos, E., Spathis, C., Manolopoulos, Y.: Support vector machines, decision trees and neural networks for auditor selection. *J. Comput. Methods Sci. Eng.* **8**(3), 213–224 (2008)
8. Klein, G., Kim, Y., Deng, Y., Snellart, J., Rush, A.: OpenNMT: Open-source toolkit for neural machine translation. In: Proceedings of ACL 2017, System Demonstrations (July 2017), pp. 67–72
9. Kummerow, A., Klaiber, S., Nicolai, S., Bretschneider, P., System, A.: Recursive analysis and forecast of superimposed generation and load time series. In: International ETG Congress 2015; Die Energiewende - Blueprints for the New Energy Age, pp. 1–6 (2015)
10. Laptev, N., Amizadeh, S., Flint, I.: Generic and scalable framework for automated time-series anomaly detection. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 1939–1947 (2015)
11. Liguori, A., Markovic, R., Dam, T. T. H., Frisch, J., van Treeck, C., Causone, F.: Indoor environment data time-series reconstruction using autoencoder neural networks. Preprint (2020). arXiv:2009.08155
12. Shannon, C.E.: A mathematical theory of communication. *Bell Syst. Tech. J.* **27**(3), 379–423 (1948)
13. Sherstinsky, A.: Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Phys. D Nonlinear Phenomena* **404**, 132306 (2020)
14. Tewari, A., Zollhofer, M., Kim, H., Garrido, P., Bernard, F., Perez, P., Theobalt, C.: Mofa: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1274–1283 (2017)
15. von Werra, L., Tunstall, L., Hofer, S.: Unsupervised anomaly detection for seasonal time series. In: 2019 6th Swiss Conference on Data Science (SDS), pp. 136–137 (2019)
16. Wang, M., Deng, W.: Deep visual domain adaptation: A survey. *Neurocomputing* **312**, 135–153 (2018)

17. Xu, D., Tian, Y.: A comprehensive survey of clustering algorithms. *Ann. Data Sci.* **2**(2), 165–193 (2015)
18. Yu, Y., Zhu, Y., Li, S., Wan, D.: Time series outlier detection based on sliding window prediction. *Math. Prob. Eng.* **2014**, Article ID 879736 (2014). <https://doi.org/10.1155/2014/879736>
19. Zhang, Y., Chen, W., Black, J.: Anomaly detection in premise energy consumption data. In: 2011 IEEE Power and Energy Society General Meeting, pp. 1–8 (07 2011)