Renewable and Sustainable Energy Reviews 206 (2024) 114842

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

journal homepage: www.elsevier.com/locate/rser

Review article

Generating synthetic energy time series: A review☆

M. Turowski [a,1], B. Heidrich [a,1], L. Weingärtner [a,1], L. Springer [a], K. Phipps [a,b,*], B. Schäfer [a], R. Mikut [a], V. Hagenmeyer [a]

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

[b] Karlsruhe Institute of Technology, Scientific Computing Center, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany

ARTICLE INFO

Keywords:
Energy time series
Time series generation
Synthetic time series
Generation approach
Synthetic time series characteristics
Synthetic time series use cases
Synthetic time series evaluation

ABSTRACT

As the energy system transitions to an intelligent smart grid with a mostly renewable energy supply, synthetic energy time series are required to facilitate the development and improvement of methods for smart grid applications. These synthetic energy time series must exhibit characteristics similar to the real energy time series and applicable to specific use cases. Furthermore, evaluation methods must be applied to verify that synthetic energy time series have the desired characteristics. Whilst many methods exist in the literature to generate synthetic energy time series, up until now, no work has focused on analysing and comparing these methods. Therefore, this study provides a structured literature review of generating synthetic energy time series. The review focuses on five aspects: (1) Identifying methods used to generate synthetic energy time series, (2) categorising these methods according to the generation approach taken, (3) analysing the characteristics of these generated synthetic energy time series, (4) identifying the uses cases for which the time series are generated, and (5) considering how the generated synthetic energy time series are evaluated. In total, this study reviews 169 articles focusing on generating synthetic energy time series and identifies several key research gaps leading to multiple open research fields. The most important open research fields include the need for a standardised evaluation, more generation methods for synthetic time series from generation and battery storage systems, and a stronger focus on further use cases.

Word count: 11395

Contents

☆ This project is funded by the Helmholtz Association's Initiative and Networking Fund, Germany through Helmholtz AI, the Helmholtz Association, Germany under the Program "Energy System Design", and the Helmholtz Association, Germany under grant no. VH-NG-1727.

* Corresponding author at: Karlsruhe Institute of Technology, Scientific Computing Center, Hermann-von-Helmholtz-Platz 1, 76344 Eggenstein-Leopoldshafen, Germany.

E-mail address: kaleb.phipps@kit.edu (K. Phipps).

[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.rser.2024.114842
Received 27 December 2023; Received in revised form 16 August 2024; Accepted 18 August 2024
Available online 30 August 2024

1. Introduction

To mitigate climate change, energy systems worldwide shift to an electricity supply predominantly based on renewable energy sources. Since these renewable energy sources are volatile and uncertain, smart solutions must be integrated into the electricity grid to guarantee affordable, reliable, and sustainable electricity supply. These smart solutions result in the so-called smart grid [1], which enhances the physical grid with information and communication technology [2]. These additional layers of communication and information technology record large amounts of data, typically in the form of time series. These recorded time series, often referred to as energy time series, include electricity generation and consumption, heating and cooling demand, State of Charge measurements from battery storage systems, or measured grid frequency. This energy time series data can then be processed and used in smart grid applications such as customer profiling, forecasting electrical load and generation, power quality analysis, and security [3]. Developing and improving methods for these applications typically requires a reasonable amount of representative and privacy-preserving energy time series data [4], especially when applying machine learning and deep learning methods [5].

Unfortunately, large amounts of such energy time series data are usually not openly available [3,6,7]. This lack of open availability may be due to expensive and time-consuming data collection, security or privacy concerns of users, or newly built buildings and generators that lack a long operation history. As a result, effectively developing and improving methods for the mentioned smart grid applications is challenging. One promising solution to this challenge is to generate synthetic energy time series with the same characteristics as real-world data and use them for method development and improvement.

As shown in Fig. 1, such synthetic energy time series are created for existing or yet-to-be-developed methods based on selected input data and a specific generation method that follows a certain generation approach. The resulting energy time series should exhibit characteristics that mirror real energy time series. Some of these characteristics are common to many energy time series, such as daily, weekly, and seasonal periodicities and the dependence on exogenous influences such as the weather [8]. However, other characteristics differ between different energy time series and depend on the specific time series considered. For example, energy consumption time series have differing characteristics depending on the considered aggregation level, i.e. a single appliance, an entire household, a village, or the entire country. Another example of varying characteristics is the shape of generation time series, which is regular and repetitive for dispatchable sources of generation, such as gas or nuclear power plants, but more difficult to predict for volatile sources of generation, such as wind power. In addition to reflecting these characteristics, the synthetic energy time series must be suitable for the chosen use cases. For example, the use case of a generic load profile may require a different temporal resolution than the use case of generating highly accurate short-term time series forecasts. Finally, to ensure their quality, the generated synthetic energy time series must fulfil selected evaluation criteria, such as diversity and fidelity.

As a result, to facilitate the development and improvement of methods for smart grid applications, it is necessary to generate a wide range of synthetic energy time series with varying characteristics applicable to specific use cases and fulfilling different evaluation criteria. Unfortunately, these requirements mean that it is almost impossible to apply general methods for generating time series, such as Generative Adversarial Networks (GANs) (e.g. [9,10]), Variational Auto-Encoders (VAEs) (e.g. [11]), or Invertible Neural Networks (INNs) (e.g. [12]). At the same time, various specific methods for generating energy time series exist in the literature, which follow different generation approaches such as top-down or bottom-up [13,14] and which are designed for specific use cases but cannot necessarily be applied to other use cases. For this reason, a systematic review is needed that identifies existing methods for generating energy time series, including the generation approach, characteristics, use cases, and evaluation of the time series generated with them. Given this review, it would be possible to identify a generation method that can generate a time series for a given use case, desired energy time series characteristics, and evaluation criteria.

Abbreviations

AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
Bayesian GAN	Bayesian GAN
BN	Bayesian Network
cGAN	conditional GAN
CNN	Convolutional Neural Network
cVAE	conditional VAE
DCGAN	Deep Convolutional GAN
DNN	Deep Neural Network
EM	Expectation Maximisation algorithm
EV	Electric Vehicle
FLOP	Floating-Point Operations Per Second
GAN	Generative Adversarial Network
GMM	Gaussian Mixture Model
GMMN	Generative Moment Matching Network
GNN	Graph-based Neural Network
HMM	Hidden Markov model
INN	Invertible Neural Network
IoT	Internet of Things
MA	Moving Average
MCMC	Markov Chain Monte Carlo
NN	Neural Network
PCA	Principal Component Analysis
PMU	Phasor Measurement Unit
PV	Photovoltaics
RCGAN	Recurrent Conditional GAN
RGAN	Recurrent GAN
RNN	Recurrent Neural Network
SOM	Self-Organising Map
t-SNE	t-distributed Stochastic Neighbour Embedding
TSTR	Train on Synthetic and Test on Real
VAE	Variational Auto-Encoder
VARMA	Vector Autoregressive Moving Average
WGAN	Wasserstein GAN
wRNG	weighted Random Number Generator

Unfortunately, to the best of the authors' knowledge, no such structured literature review of existing energy time series generation methods exists in the literature. Therefore, this study aims to fill this gap by providing a structured review of the literature on the generation of synthetic energy time series. More specifically, this review emphasises energy time series from electricity generation, electricity consumption, and battery storage systems and reviews existing literature concerning the following five questions:

1. Which methods exist to generate synthetic energy time series?
2. Which generation approach do these identified generation methods follow?
3. Which characteristics do the synthetic energy time series generated by these methods have?
4. For which use cases are the synthetic energy time series generated?
5. How are the generated synthetic energy time series evaluated?

Based on the answers to these five questions, this study then further discusses the related findings and makes suggestions for future research.

The remainder of the study is structured as follows. First, Section 2 provides an overview of existing literature reviews similar to this study and discusses their limits. Section 3 then describes the methodology used for the literature review, before the five subsequent sections present and discuss the results of the five-step analysis: Section 4 introduces the methods identified in the literature for generating synthetic energy time series. Section 5 determines the generation approach these methods apply. Section 6 describes the characteristics of the synthetic energy time series generated with these methods. Section 7 presents the identified use cases of the generated synthetic energy time series. Section 8 describes the evaluation of generated synthetic energy time series performed in literature. Finally, the study concludes in Section 9.

2. Existing literature reviews

This study presents a structured literature review of generating synthetic energy time series. Therefore, this section motivates this study by considering and distinguishing it from existing literature reviews focusing on similar topics. More specifically, this section identifies four groups of existing literature reviews, highlights their limitations, and presents the research gap addressed with this study.

The first identified group comprises reviews that focus on synthetic data generation for a specific domain such as finance [15] or medicine [16]. These reviews, however, do not focus on energy time series.

Considering purely time series data, the second identified group contains literature reviews that concentrate on certain properties of the generated synthetic time series. These reviews examine generation methods in terms of their ability to generate privacy-preserving synthetic time series that allows valid statistical inference [17] or their publication [18]. In this group, however, the focus is clearly on privacy perseverance as a property of synthetic time series and not on the characteristics of synthetic energy time series.

The third identified group consists of reviews that consider the validation of the generated time series [19] or their evaluation concerning diversity [20], fidelity and diversity [21], or resemblance [22]. Whilst these works review certain evaluation metrics, they all consider only specific evaluation criteria and thus fail to consider evaluation criteria specifically relevant to generated synthetic energy time series.

The fourth identified group of related reviews are reviews focusing on smart grid applications. This group includes reviews that investigate available methods for modelling residential electrical load profiles [13,14,23] and energy consumption [24] as well as simulating smart home activities [25]. Some of these reviews [13,14,24] also categorise the identified methods into bottom-up and top-down generation approaches. Furthermore, another work in this group supports the selection of a suitable time series generation method [26] but is limited to the analysis of power system operation and expansion. Whilst these reviews are closer to this study, they still focus on single use cases and fail to cover the variety of smart grid applications.

Altogether, to the best of the authors' knowledge, no work exists that reviews methods for generating synthetic energy time series and their approach as well as the characteristics, use cases, and evaluation of these generated synthetic time series. Therefore, the analysis presented in the following sections aims to fill this identified research gap.

3. Methodology

To identify relevant literature on generating synthetic energy time series, this review follows three basic steps proposed in [27], namely the identification of major contributions, backward search, and forward search.

First, the review identifies major contributions with a keyword-based search. This search was first conducted in Google Scholar between June and December 2021 and again in May 2023 using the iteratively expanded keywords listed below and considering the first 70 results of each search:

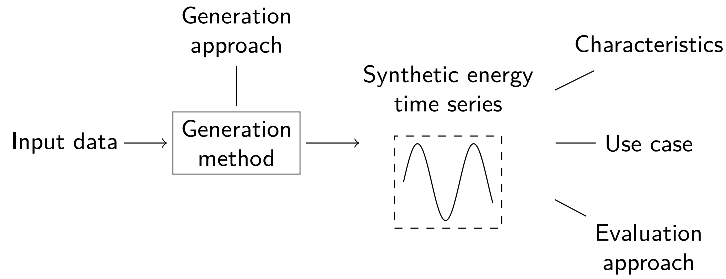

Fig. 1. A generation method following a certain generation approach generates a synthetic energy time series based on input data. The generated synthetic energy time series has certain characteristics, is created for a use case, and is evaluated using an evaluation approach.

- synthetic data generation AND time series
- synthetic time series generation
- time series AND synthetic data
- energy time series AND synthetic data
- "time series" AND "synthetic data" AND "models"
- "synthetic load time series"
- "synthetic load datasets"
- "synthetic load data sets"
- "energy time series" AND "generated data"
- "energy time series" AND "data generation"
- "time-series load data" AND "synthetic data"
- "generation" AND "load profiles"
- "time series load data" AND "synthetic data"
- "generate data" AND "energy time series"

These identified articles are then screened based on inclusion and exclusion criteria to determine the initial pool of articles. To be included, the title and the abstract of an article must indicate that the article proposes a method for generating synthetic energy time series. Additionally, articles that use words similar to energy time series generation such as energy time series creation, augmentation, and synthesis as well as scenario generation and load profiling are included, as long as the article actually proposes a method for generating synthetic energy time series. This review excludes articles not written in English, technical reports, and articles not published in a journal, at a conference, or on arXiv. This step results in a pool of 68 articles proposing a method for generating synthetic time series.

Given this initial pool of articles, backward and forward searches for all articles from this pool are performed. These searches apply the same screening based on inclusion and exclusion criteria as mentioned above. This screening results in 101 additional articles, leading to a total of 169 considered articles published during the past 40 years as shown in Fig. 2. Due to the focus on energy time series from electricity consumption, electricity generation, and battery storage systems, four articles dealing exclusively with generating water time series [28–31] and one article concerned exclusively with generating heat time series [32] are excluded.

All 169 articles finally considered form the basis for the subsequent five-step analysis. This analysis first identifies the generation method used in each article and groups articles proposing similar generation methods. In the case that more than one generation method is used in an article, the article is assigned to all these methods used. In the second step this analysis identifies the generation approach these methods apply. Third, this analysis determines the characteristics of the generated energy time series in each article. Fourth, the use cases of the generated synthetic energy time series are identified. Lastly, the analysis focuses on the evaluation performed in the articles to examine the synthetic energy time series generated with the respective generation method.

4. Methods for generating synthetic energy time series

The first step of the analysis identifies the methods used to generate synthetic energy time series found in the considered literature.

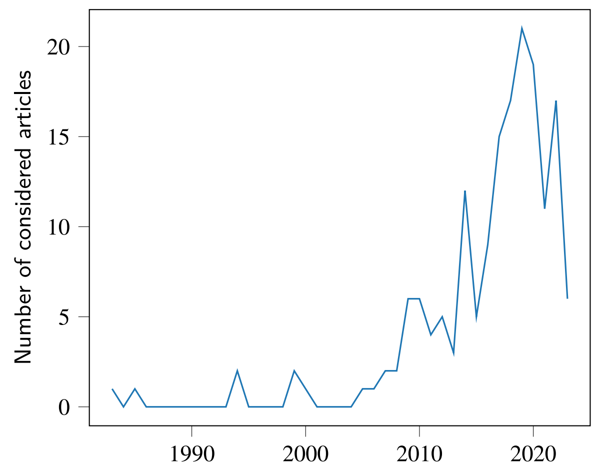

Fig. 2. The number of articles considered by the year of publication.

The analysis identifies thirteen methods that are applied to generate synthetic energy time series. This study focuses on the methods that are proposed in an article whilst methods that are applied as baselines in an article and other types of methods that generally exist in the literature are not considered. As shown in Fig. 3, the thirteen generation methods vary in their frequency of occurrence. Markov models, weighted Random Number Generator (wRNG) methods, and GANs are most often used. These three generation methods taken together are used in more than half of the considered articles, whereas the other ten methods cover the remaining shares.

This section introduces the identified generation methods to provide the basis for further analysis of their application for generating synthetic energy time series. Thereby, this section does not discuss how these methods are applied to generate synthetic energy time series in the identified articles but instead focuses on explaining their basic principles. Furthermore, this section provides literature for each identified generation method that introduces the generation method in question in a comprehensible and detailed manner.

4.1. Markov models

For the generation of synthetic energy time series, a Markov model models a time series as a sequence of different observed states. Each state typically represents a range of values observed in the time series. Transition probabilities between the states define how likely the model is to move from one state to another. In a first-order Markov model, the current state depends on the previous state only; in a N-order Markov model, the current state depends on the N previous states. Thereby, a discrete Markov model with a finite number of states corresponds to a so-called Random Walk model [33]. Given specified states, transition probabilities, and initial probabilities of the states, one can generate a

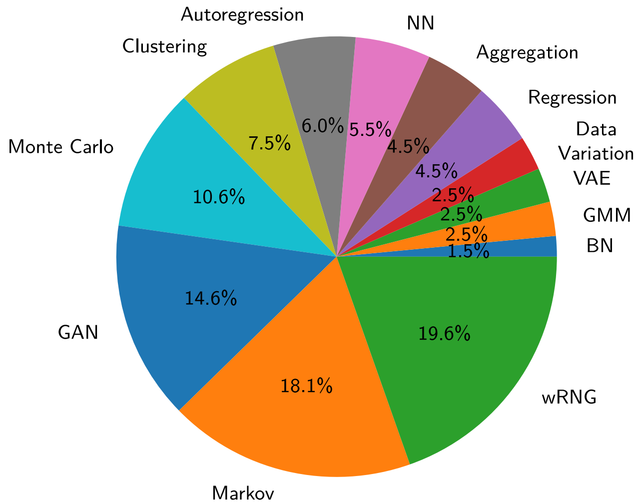

Fig. 3. Shares of the identified methods for generating synthetic energy time series. Note that articles that use more than one generation method count for each method used.

random number at each state, starting with the first state, to create a synthetic time series.

A special type of the Markov model is the Hidden Markov model (HMM). It additionally comprises hidden states representing e.g. values within the value ranges of the observed states and transition probabilities between these hidden states [34].

For more information on Markov models, consider the detailed theoretical introduction in [35] or the overview of different Markov models and their applications in [36]. Additionally, an introduction to Markov models in the context of data generation is given in [37], whilst of the articles identified in this analysis, detailed introductions to the theory behind Markov models can be found in both [38,39].

4.2. Monte Carlo simulations

With a Monte Carlo simulation, synthetic energy time series are generated using random sampling. Although there is no single form of Monte Carlo simulations, many Monte Carlo simulations follow a typical pattern: First, statistical distributions are determined for the values to be generated. Second, random samples are repeatedly drawn from these determined distributions to obtain the desired output. Nevertheless, existing Monte Carlo simulations differ in the methods used to identify statistical distributions [40] and to generate random variables [41].

A special type of Monte Carlo simulation is bootstrapping. In case of unknown or complex underlying distributions, bootstrapped Monte Carlo simulations perform sampling with replacement from a given data set to approximate a statistic's sampling distribution [42,43]. Another type of Monte Carlo simulation is the Markov Chain Monte Carlo (MCMC) method. It combines a Markov model with a Monte Carlo simulation by sampling from a Markov model that has the desired distribution as its equilibrium distribution [44].

A theoretical introduction to Monte Carlo simulations is provided in [40], whilst the practical applications of Monte Carlo simulations are covered in an introductory tutorial in [41]. Additionally, a detailed overview of Monte Carlo simulations and their many applications can be found in [45].

4.3. Weighted random number generators

To generate synthetic energy time series, a weighted Random Number Generator (wRNG) uses random numbers to linearly combine elementary time series. These elementary time series can, for example, be based on the power consumption and generation of devices or on standard load profiles. Similar to a Monte Carlo simulation, a wRNG first determines the required probability distribution for each elementary time series before sampling from them to perform the linear combination.

An introduction to wRNGs and random number generation in general is provided in [46]. Essentially, all methods classified as wRNG and identified in this study apply wRNGs to sample from pre-determined probability distributions and combine mathematical models. Although they do not provide a general introduction to wRNGs, a detailed and helpful description of how wRNGs are applied in their approaches can be found in [47,48].

4.4. Gaussian mixture models

For the generation of synthetic energy time series, a Gaussian Mixture Model (GMM) uses a weighted sum of normal distributions to model the distribution of the values to be generated. Given the mixture character of the Gaussian Mixture Model (GMM), the resulting distribution comparatively accurately matches the real distribution of e.g. electrical consumption or generation [49]. To determine the parameters of the GMM, a typical way is to use the Expectation Maximisation algorithm (EM) method [50]. A fully specified GMM can then serve as a distribution to sample values from and thus generate a synthetic time series. Alternatively, GMMs are applied in Markov models to generate the desired values of the states within the Markov process.

For more information on GMMs, consider the concise introduction in [51]. Moreover, see [52] for an overview of numerous applications of GMMs including generating synthetic time series. From the articles identified in this analysis, clear overviews are provided in both [53,54].

4.5. Neural networks

To generate synthetic energy time series, a Neural Network (NN) learns the mapping between an input and an output space using training data. To learn the mapping, a Neural Network (NN) consists of one or multiple layers of interconnected neurons where each connection is associated with a weight. A NN with multiple layers is also known as a Deep Neural Network (DNN). With new input data, the trained NN then generates synthetic energy time series. Since various types of NN exist, the identified types are briefly introduced in the following.

One type of NN is the feed-forward NN where the connections between nodes are not recurrent or circular. This way, the information between the layers flows in only one direction and thus is unidirectional [55]. A type of feed-forward NN with two layers is the Self-Organising Map (SOM) network. It learns to represent high-dimensional data in a low dimension while preserving the topological structure of the data. Unlike most other NNs that apply the error-correction learning, a SOM network uses competitive learning where the neurons compete to respond to a specific input [56]. If each neuron in a layer of a feed-forward NN is connected with all other neurons of the previous layer, it is called a fully-connected NN. Such a network can be combined with convolutional and down-sampling layers, resulting in a Convolutional Neural Network (CNN) [57]. In a Recurrent Neural Network (RNN), another type of NN, the connections between nodes can form cycles. For this reason, the output of a node can affect the subsequent input to the same node, making the information flow between layers bidirectional [58]. Another type of NNs is the Graph-based Neural Network (GNN) that processes data represented as a graph. It uses a series of NNs to convert the input graph into a lower dimensional space while preserving information on nodes, edges, and context [59]. In a further type of NN, the Generative Moment Matching Network (GMMN), a NN creates data similar to the training data and applies a two-sample test to distinguish created data and real data [60]. The Invertible Neural Network (INN) also represents a type of NN. An INN realises a bijective mapping from input to output by implementing a normalising flow with the help of so-called coupling layers such as NICE [61] or RealNVP [62]. To introduce non-linearities into these transformations, each coupling layer contains multiple NN as subnetworks [63].

There is a vast array of research considering NNs and explaining the theory. However, a thorough introduction is presented in [64], whilst the foundations are summarised and specific NNs are explained in more detail in [65]. Finally, see [66] for a complete overview of the theory and application of NNs.

This analysis identified certain NNs that are applied more frequently and warrant specific attention. Therefore, the following section introduces two further types of NNs, namely the GAN and the VAE.

4.6. Generative adversarial networks

To generate synthetic energy time series, a Generative Adversarial Network (GAN), which is another type of NN, uses two interconnected NNs, namely a generator and a discriminator. While the generator creates data similar to the training data by using random noise as an input, the discriminator applies a continuous value to distinguish data created by the generator and real data [67].

A type of GAN is the Wasserstein GAN (WGAN). By using a different loss function, the WGAN improves the learning stability and avoids mode collapse, i.e. very similar or identical samples generated by the generator [68]. The conditional GAN (cGAN) as another type of GAN adds an additional input layer to condition the generator and the discriminator on additional information. This way, a cGAN can generate data under the guidance of known information [69]. The Recurrent GAN (RGAN) is one more type of GAN. In a RGAN, RNNs are used for the generator and the discriminator. RGAN can be extended with a conditional input, resulting in a Recurrent Conditional GAN (RCGAN) [9]. Similarly, a Deep Convolutional GAN (DCGAN) makes use of only CNNs for the generator and the discriminator [70]. Another type of GAN, the Bayesian GAN (Bayesian GAN), includes a Bayesian formulation in the training process. In this way, a set of generators can be found that completely capture distinct modes in the historical data [71]. A GAN can also be combined with a Monte Carlo Simulation such that the Monte Carlo Simulation post-processes the energy time series generated by the GAN.

More information on GANs can be found in the overview in [72] or the detailed discussion presented in [73]. Among the articles identified in this analysis, a concise introduction to GANs is presented in [74], whilst different GANs are briefly explained in [75], and GANs are specifically introduced in the context of time series generation in [76].

4.7. Variational auto-encoders

For the generation of synthetic energy time series, a Variational Auto-Encoder (VAE) applies two NNs, one for the so-called encoder and one for the so-called decoder. While the encoder learns the probability distribution of the input data in the latent space, the decoder reconstructs its input [77].

One type of VAE is the conditional VAE (cVAE). It adds conditional information as inputs to the encoder and decoder, allowing data to be generated given certain information [11]. Additionally, a VAE can be combined with a GAN by merging the decoder of the VAE and the generator of the GAN.

Detailed theoretical information on VAEs can be found in [77], whilst an intuitive tutorial is presented in [78], and key concepts and different types of VAEs are summarised and discussed in [79]. From the articles identified in this analysis, VAEs are clearly and concisely introduced in [80] with a focus on synthetic time series generation.

4.8. Regression methods

For generating synthetic energy time series, a Regression method models the relationship between so-called dependent and independent variables using mathematical functions, including the commonly used linear functions. While a simple Regression method only has one independent variable, a multiple Regression method considers several independent variables. Given a determined relationship between dependent and independent variables, a Regression method creates synthetic energy time series by varying the input data or randomly sampling from a distribution [81]. The Regression method can be used in combination with a Markov model so that the Regression method learns the transition probabilities.

For more information on Regression methods, consider the complete theoretical introduction in [82], whilst the basic principles are summarised and different forms of regression are explained in more detail in both [83] and [84].

4.9. Autoregression methods

To generate synthetic energy time series, an Autoregression method models each value of the energy time series as linearly dependent on its previous values and a stochastic term. Given this autoregressive structure, the Autoregressive (AR) model then iteratively generates new values [85].

An AR model is often combined with a Moving Average (MA) model to create a so-called Autoregressive Moving Average (ARMA) model. The MA model assumes that the value of the energy time series is correlated with an additional random variable. Therefore, the ARMA model assumes that each value of the energy time series is linearly dependent on its previous values, the previous values of an additional random variable, and a stochastic term [85]. Furthermore, it is possible to model multiple energy time series at once by stacking them as a vector and considering a Vector Autoregressive Moving Average (VARMA) model [85]. Finally, ARMA models rely on the assumption of stationarity. If this is not given, the values of the time series can be replaced with the difference between their values, resulting in an integrated ARMA model, i.e. Autoregressive Integrated Moving Average (ARIMA) [85]. In combination with a GAN, Autoregression methods can be used to preprocess features for the GAN. Autoregression methods are also applied to model the dynamics of hidden states in a Hidden Markov model (HMM).

Further information on Autoregression methods can be found in the detailed introduction in [86] or alternatively with a stronger focus on practical applications in [87]. Additionally, generalised Autoregression methods are described in detail in [88].

4.10. Clustering methods

Clustering methods first group similar time series in clusters and then use the properties of the determined clusters to generate synthetic energy time series [89].

One type of Clustering method is the k-means Clustering method. It minimises the distance between all time series of the cluster and the cluster centre [90]. The fuzzy c-means Clustering method applies the same principle as the k-means Clustering method. However, in contrast, it uses a degree of membership such that each time series can belong to more than one cluster [91]. Another type of Clustering method is the hierarchical Clustering method. In hierarchical Clustering, clusters are created in two ways. These clusters are either created by gradually merging individual time series or, alternatively, a single cluster consisting of all time series is gradually split [89]. Spectral Clustering is another type of Clustering method. It models all time series as nodes of a graph and their distances using weighted edges [92]. Based on this representation, clusters are identified as groups of nodes that are considered close according to the weighted edges. The Expectation

Maximisation algorithm (EM) Clustering as another type of Clustering method represents the time series with GMMs and uses the EM algorithm [50] to determine the clusters [93]. A Clustering method can be combined with other methods. It can provide inputs for a NN, GAN, or Markov model. Moreover, a SOM network can first be applied to reduce the dimensionality of the data for the subsequently used Clustering method.

More information on Clustering methods can be found in the detailed introduction in [94] and in the discussion and comparison of multiple Clustering methods in [95]. Moreover, a detailed overview of recent developments in Clustering methods is given in [96]. From the articles identified in this analysis, a short but informative overview of Clustering is presented in [97].

4.11. Aggregation methods

For the generation of synthetic energy time series, an aggregation method makes use of information about the underlying energy system. Based on geographic, topological, usage, or hardware information, energy time series are determined for subunits of the system before all these time series are aggregated into the single energy time series of the considered system.

An Aggregation method can be combined with a Markov model such that the Aggregation method augments the Markov model with usage information. Moreover, a Markov model can be used to generate the energy time series of the subunits to be aggregated by the Aggregation method. Similarly, a Markov model can provide usage information and a Bayesian network hardware information for an Aggregation method.

For more information on aggregation alternatives, see the mathematical theory of aggregation presented in [98]. Moreover, an overview of existing Aggregation methods can be found in [99].

4.12. Bayesian networks

To generate synthetic energy time series, a Bayesian Network (BN) models the system in question using a directed acyclic graph. In this graph, each node represents a unique random variable describing a characteristic such as the size of a household and each edge the conditional dependency between two nodes. The parameters of the BN can be estimated using the EM algorithm [50]. Given the parameters of the BN, one can sample from the determined probability distributions to generate the desired values [55]. A BN can be combined with a SOM that provides the relevant input features or a Markov model that considers the results of the BN to model different states.

More information on BN can be found in the detailed theoretical introduction in [100]. Additionally, practical applications of BNs, namely how to effectively implement and train them, are discussed in [101]. Finally, a concise overview of BNs in the context of time series and wind energy is provided in [102].

4.13. Data variation methods

To generate synthetic energy time series, a Data Variation method varies an existing time series directly or indirectly. In the direct variation, existing measurements for the target energy time series are varied, for example, via scaling or introducing noise. In the indirect variation, the inputs for an existing model that generates the target energy time series are varied. A data variation method can use the results of a Clustering method as a starting point for the variation.

For more information on Data Variation methods, see the overview of similar methods from data augmentation presented in [103] or the introduction to Data Variation methods for time series in [104] or [105].

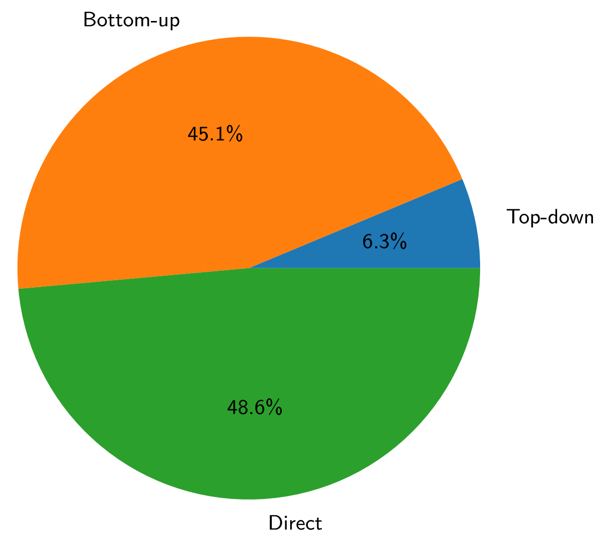

Fig. 4. Shares of the generation approaches used in the considered articles.

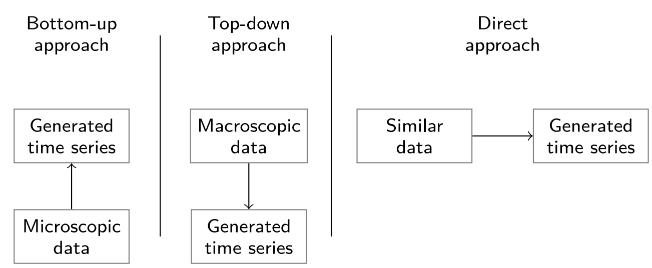

Fig. 5. An overview of the three approaches to generate synthetic time series.

5. Generation approaches of methods for generating synthetic energy time series

This section presents the second analysis step which determines the generation approach with which the identified generation methods are applied. To categorise the approaches, this analysis distinguishes three generation approaches building on [13,24], namely the bottom-up, the top-down, and the direct approach.

As shown in Fig. 4, almost half of the articles use generation methods with the direct approach and 45% with the bottom-up approach. Only 6% of the articles apply generation methods with the top-down approach.

In the following, the three generation approaches illustrated in Fig. 5 are introduced. Table 1 gives an overview of all identified methods for generating synthetic energy time series and the associated generation approach.

5.1. Bottom-up approach

Generation methods that follow the bottom-up approach obtain a synthetic energy time series by using microscopic data. For the considered context, microscopic data from subsystems within a larger system are used to generate synthetic energy time series for the larger system being considered. For example, data on all appliances in a household could be used to generate energy time series for that household. Therefore, the bottom-up approach requires knowledge of the interdependencies and elements comprising the system for which a synthetic energy time series is being generated. Table 1 shows that all generation methods except the GMM, VAE, and Autoregression methods are applied with a bottom-up approach. Especially for Markov models, Monte Carlo methods, wRNG methods, and Aggregation methods, several articles exist that use a bottom-up approach.

Table 1
Overview of the identified methods for generating synthetic energy time series and the associated generation approach. The generation approaches comprise the top-down, the bottom-up, and the direct approach. Note that articles that combine generation approaches count for each generation approach applied. Note also that articles that combine more than one generation method or compare different generation methods are listed in all methods used.

	Bottom-Up	Top-Down	Direct
Markov	[106–123]	[111,124–127]	[39,128–140]
Monte Carlo	[107–109,113,114,117,123,141–144]	[143]	[54,129,135,140,145–150]
wRNG	[47,48,151–182]	[152,153,183]	[184–188]
GMM		[124,127]	[53,54,130,189]
NN	[190,191]	[191]	[137,192–199]
GAN	[75,200,201]		[74,76,145,187,189,201–222]
VAE			[80,215,223–225]
Regression	[114,226]	[127]	[130,131,227–230]
Autoregression			[132,189,206,231–239]
Clustering	[110,240–243]	[124,125]	[97,137,199,217,244–247]
Aggregation	[118,142,248–253]	[253]	[254]
BN	[118,190]		[138]
Data Variation	[241,255–257]		

5.2. Top-down approach

To obtain a synthetic energy time series, generation methods applying the top-down approach use macroscopic data. For a defined context, macroscopic data from a larger system are used to generate energy time series for a subsystem that belongs to this larger system. For example, energy consumption data in a city quarter and data on the shares of residential, commercial, and industrial use could be used to generate typical energy time series of single residential, commercial, and industrial consumers. Therefore, the top-down approach requires knowledge about the structure of the system and how a specific system element depends on the surrounding system to generate synthetic energy time series for that specific element. As observed in Table 1, all generation methods except five generation methods are applied with the top-down approach. While most articles use the top-down approach with Markov models and wRNG methods, no articles use the top-down approach with GANs, VAEs, Autoregression methods, BNs, and Data Variation methods.

5.3. Direct approach

Generation methods following the direct approach obtain synthetic energy time series by using data similar to the energy time series to be generated. This way, energy time series can be generated without any knowledge about the surrounding system, its elements, and their interdependencies. For example, existing energy time series from households can serve as the starting point for generating synthetic energy time for a similar household. Nevertheless, the direct generation approach assumes that the similar data used contain all the information required to generate the desired synthetic energy time series. Observing Table 1 shows that all generation methods except the Data Variation methods are applied with the direct approach. The most articles using the direct approach are found for the GANs and Markov models.

5.4. Discussion & open research fields

Given the classification in Table 1 and the results described above, this analysis leads to two key observations. First, the majority of the identified methods to generate synthetic energy time series use either the bottom-up or the direct approach. In contrast, only fifteen of the 169 identified articles apply the top-down approach. Second, whilst most generation methods are applied with at least two approaches, this often occurs with varying degrees of frequency. For example, wRNGs are applied with the bottom-up approach much more frequently than with the top-down or direct approach. This is logical since wRNGs usually sum up the power consumption or generation of elementary systems to obtain the final synthetic energy time series. However, NN-based methods, including GANs and VAEs, almost exclusively apply a direct approach. This can be explained by the black-box nature of a NN, which is ideal for learning direct relationships to generate synthetic time series.

These observations lead to the identification of two open research fields. First, it would be interesting to compare the different approaches for similar generation methods and quantitatively determine the strengths and weaknesses of each approach, specifically when considering specific scenarios. Second, there is a lack of literature focusing on the top-down approach, therefore, it is important to investigate the top-down approach further to determine with which generation methods and in which situations it may be more beneficial for generating synthetic energy time series.

6. Characteristics of generated synthetic energy time series

This section presents the third analysis step which determines the characteristics of the generated synthetic energy time series. The synthetic time series generated by the previously introduced generation methods can be classified using three attributes, namely the type, the aggregation level, and the use. The following describes the findings with respect to these three attributes of the generated energy time series. Table 2 gives an overview of all methods for generating synthetic energy time series according to their type, aggregation level, and use.

6.1. Type

The type describes the direction of the flow of the electrical energy. To consider different application contexts, the analysis distinguishes the type of energy time series between consumption, generation, and battery storage systems. As shown in Fig. 6, generation methods are used to create consumption time series in more than three quarters of the articles, generation time series in about one fifth of the articles, and time series from battery storage systems in only a few articles.

Matching this observation, Table 2 shows that all identified generation methods are used to generate synthetic consumption time series. All generation methods except the NNs, the Regression methods, and the BN are applied to create synthetic generation time series. Synthetic time series from battery storage systems are only generated with six generation methods, namely Markov models, wRNGs, NNs, GANs, Autoregression methods, and Clustering methods.

6.2. Aggregation level

The aggregation level is the spatial resolution of the generated energy time series. This analysis identifies the following eight different aggregation levels from large to small: country, state, region, city, district, single unit (e.g. a building or a wind turbine), household, and appliance. If there is no explicit specification, the respective article

Table 2
Overview of the identified methods for generating synthetic energy time series and the characteristics of the synthetic energy time series generated thereby. The characteristics include the type, the aggregation level, and the use associated with the generated synthetic energy time series.

Method	Type	Aggregation Level	Use
Markov	Battery: [137]; Consumption: [106–128,130–134,138]; Generation: [39,129,133,135,136,139,140]	State: [110]; Region: [39,123,129,135,136,139,140]; City: [118]; Single unit: [106,120,137]; Household: [107,108,111,114,121,124–127,130–134,138]; Appliance: [109,112,113,115–119,122,128,134]	Commercial: [110,124]; Electric vehicle: [123]; Industrial: [39,110,129,135–137,139,140]; Renewables: [39,129,135,136,139,140]; Residential: [106–122,124–128,130–134,138]
Monte Carlo	Consumption: [107–109,113,114,117,123,141–144,147,150]; Generation: [54,129,135,140,145,146,148,149]	Country: [146–148]; State: [150]; Region: [54,123,129,135,140,145,149]; District: [142]; Single unit: [144]; Household: [107,108,114,143]; Appliance: [109,113,117,141,143]	Academic: [144]; Electric vehicle: [123,147,150]; Industrial: [54,129,135,140,145,146,148]; Renewables: [54,129,135,140,149]; Residential: [107–109,113,114,117,141–143]
wRNG	Battery: [170]; Consumption: [47,48,151,153–169,171–184,186–188,258]; Generation: [152,185]	Country: [176,187]; Region: [179]; City: [155,166,258]; District: [48,152,154,163]; Single unit: [170,172,178,181,185,186,188]; Household: [48,159,167,175,183]; Appliance: [47,48,151,153,155–158,160–162,164,165,168,169,171,173,174,177,180,182,184]	Academic: [181]; Commercial: [156,176,178,186,187]; Electric vehicle: [170,179,188]; Industrial: [176,187]; Renewables: [185]; Residential: [47,48,151–155,157–169,171–177,180,182–184,187,258]
GMM	Consumption: [53,124,127,130,189]; Generation: [54,189]	Region: [54]; District: [53]; Single unit: [189]; Household: [124,127,130]	Commercial: [124,189]; Industrial: [54]; Renewables: [54]; Residential: [53,124,127,130,189]
NN	Battery: [137,192,195,197]; Consumption: [190,191,193,194,196,198,199]	Country: [198]; State: [199]; City: [194]; District: [190,196]; Single unit: [137,192,193,195,197]; Appliance: [190,191]	Agricultural: [199]; Commercial: [198,199]; Electric vehicle: [195,197]; Industrial: [137,199]; Residential: [190,191,193,194,196,198,199]; Storage: [192]
GAN	Battery: [213]; Consumption: [74,76,187,189,200,202,206–210,212,214–218,220–222]; Generation: [75,145,189,201,203–205,211,215,216,219]	Country: [187,200,210,214]; State: [75,202,204,211,219]; Region: [145,203–205,222]; District: [201]; Single unit: [74,76,189,205,206,213,216,217,220]; Household: [208,212,215,218]; Appliance: [206,209,215]; Not specified: [207,221]	Commercial: [74,76,187,189,202,206,210,214,217,220]; Electric vehicle: [213]; Industrial: [145,187,202–205,214]; Medical: [220]; Renewables: [75,203–205,216]; Residential: [76,187,189,200–202,206,208,209,212,214–216,218]; Not detailed: [207,211,219,221,222]
VAE	Consumption: [80,215,223–225]; Generation: [215]	Country: [223]; City: [224]; Single unit: [80,225]; Household: [215]; Appliance: [215]	Commercial: [224]; Electric vehicle: [225]; Residential: [80,215]; Not detailed: [223]
Regression	Consumption: [114,127,130,131,226–230]	Country: [226]; City: [228]; Single unit: [229]; Household: [114,127,130,131,227]; Not specified: [230]	Agricultural: [230]; Commercial: [226,230]; Industrial: [226,229,230]; Residential: [114,127,130,131,227]; Not detailed: [228]
Autoregression	Battery: [231]; Consumption: [132,189,206,231,232,234–239]; Generation: [189,231,233]	Country: [235,237]; State: [238]; Region: [231,233]; City: [234,239]; Single unit: [189,206,231,236]; Household: [132]; Appliance: [206]; Not specified: [232]	Commercial: [189,206]; Electric vehicle: [234]; Industrial: [231,233,239]; Renewables: [231,233]; Residential: [132,189,206,236,239]; Not detailed: [232,235,237,238]
Clustering	Battery: [137]; Consumption: [97,110,124,125,199,217,240–247]; Generation: [240,241,247]	State: [110,199,240,241,246,247]; District: [243,245]; Single unit: [137,217,242]; Household: [97,124,125,244]	Agricultural: [199]; Commercial: [110,124,199,217,240,241,245–247]; Industrial: [110,137,199,240,241,247]; Residential: [97,110,124,125,199,240–244,247]
Aggregation	Consumption: [118,142,248–254]; Generation: [251]	State: [249,250]; City: [118,248,253]; District: [142,254]; Single unit: [251,252]; Appliance: [118,251,253]	Commercial: [248–250]; Industrial: [249,250]; Renewables: [251]; Residential: [118,142,248–254]
BN	Consumption: [118,138,190]	City: [118]; District: [190]; Household: [138]; Appliance: [118,190]	Residential: [118,138,190]
Data Variation	Consumption: [241,255–257,259]; Generation: [241,257]	State: [241,256]; Region: [259]; Single unit: [255,257]; Household: [257]	Commercial: [241,255,259]; Industrial: [241]; Renewables: [257]; Residential: [241,257]; Not detailed: [256]

is marked as “not specified”. As illustrated in Fig. 7, most energy time series are generated with regards to appliances and single units, with both aggregation levels each making up more than 20% of the identified articles, followed by households accounting for 16% and regions making up 11%. While these four aggregation levels already account for about three quarters of the identified articles, the remaining aggregation levels and “not specified” cover the last quarter.

Moreover, Table 2 shows that the identified generation methods are applied to generate synthetic energy time series of various, but not all, aggregation levels and that the number of covered aggregation levels differs. The Monte Carlo simulations, the wRNGs, the GANs, and the Autoregression methods cover the most aggregation levels with seven aggregation levels, whereas the GMMs, the Regression methods, the Clustering methods, and the BNs cover the least aggregation levels with four aggregation levels. Lastly, there are only a few articles where the aggregation level of the generated energy time series is not clearly specified.

6.3. Use

The use describes which electricity use is represented in the generated energy time series. This analysis distinguishes the following nine uses: residential, industrial, commercial, renewables, electric vehicle, and other (i.e. academic, agricultural, medical, and storage). If time series are generated for more than one use or mixed uses (e.g. at higher aggregation levels), they are counted in each use. When a generation method is applied in an article to generate energy time series of different uses, the article counts in all these uses. In case of a missing clear indication, the respective article is marked with “not detailed”. As shown in Fig. 8, the majority of generation methods are used for generating residential energy time series, followed by industrial and commercial energy time series. Few generation methods are applied to generate energy time series of renewable energies and Electric Vehicles (EVs). The remaining methods either concentrate on other uses or the use is not detailed.

Additionally, Table 2 shows that all generation methods create synthetic energy time series associated with various uses. While the

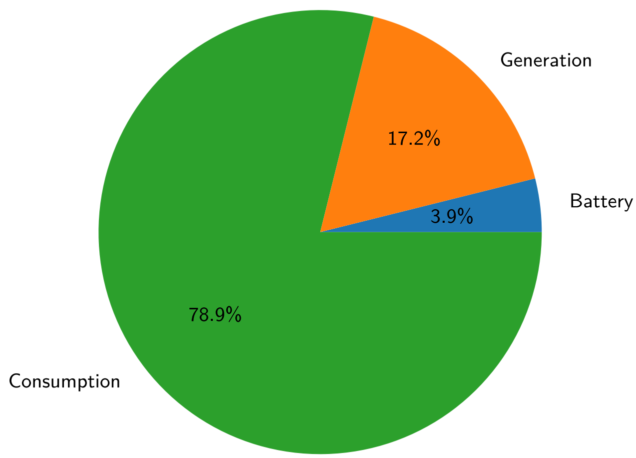

Fig. 6. Shares of the different types of generated synthetic energy time series. Note that articles that consider more than one type count for each type considered.

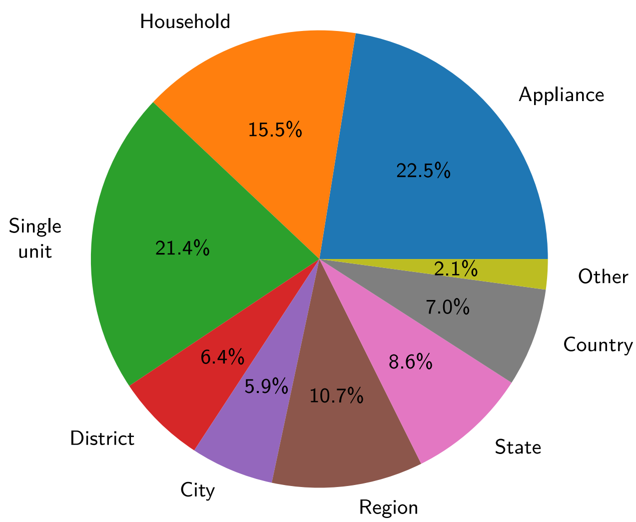

Fig. 7. Shares of the different aggregation levels of generated synthetic energy time series.

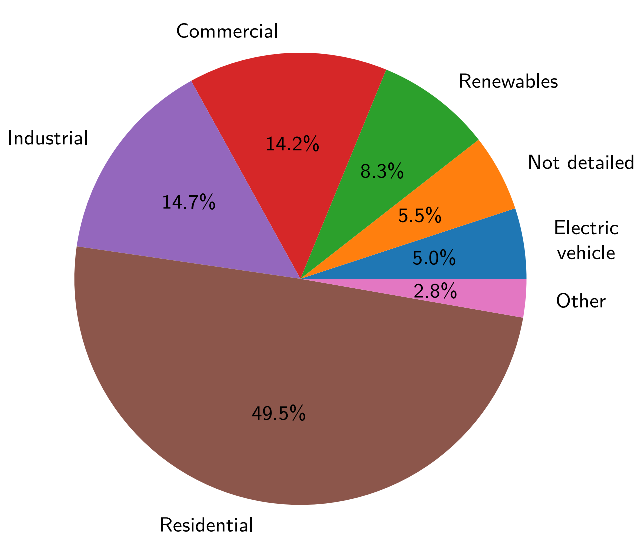

Fig. 8. Shares of the different uses of the generated synthetic energy time series. Other comprises academic, agricultural, medical, and storage. Note that articles that consider more than one use count for each use considered.

wRNGs, the NNs, and the GANs are used for six uses, the majority of generation methods are applied for four or five uses. Only the VAEs and the BNs are associated with three and one use respectively. Furthermore, a certain number of articles are identified, where the use of the generated energy time series is not detailed.

6.4. Discussion & open research fields

Given the previous observations and Table 2, this analysis results in four main observations. First, the vast majority of the identified articles focus on generating synthetic consumption time series. This could be due to the fact that traditional energy systems are based on generation from dispatchable fossil fuel-based sources and rarely include battery storage systems. Therefore, the largest source of uncertainty in such systems is the consumption. However, with the transition to smart grids containing volatile generation from renewable sources, these other types of synthetic energy time series are also gaining importance. Second, low aggregation levels such as appliances, households, and single units are the most studied, accounting for 58.2% of the identified articles. This is mirrored by the fact that residential use is by far the most studied use, with almost 75% of articles being attributed to this use. This may also be due to residents in single households having the most unpredictable consumption pattern when compared to city quarters or larger commercial or industrial companies. Third, both the commercial and industrial uses are studied similarly, although far less than the residential use. Fourth, the number of identified uses is connected to the aggregation level and the type. As the aggregation level rises, the considered uses naturally also increase since energy time series of higher aggregation levels typically comprise the energy consumption or generation of various consumers or producers. Similarly, the type also influences the associated uses of a generated synthetic energy time series.

As a result of these observations, this analysis finds three promising open research fields. First, there should be far more emphasis on generating synthetic energy time series for battery storage systems and generation. Such time series will be vital to enable the transition to a sustainable smart grid. Second, there should be more focus on generating synthetic time series for higher aggregation levels and for commercial and industrial uses. Although these uses may not be as unpredictable as a single residential household, they offer vast potential for load management in a flexible future smart grid, and to reach this potential, synthetic time series are required. Finally, future smart grids will be characterised by the blurring of boundaries between consumption and generation and thus integration of prosumers and flexumers. Therefore, a third research direction would be investigating the generation of synthetic energy time series for combinations of consumption, generation, and battery storage systems while considering the increasing mix of uses.

7. Use cases of generated synthetic energy time series

This section describes the fourth analysis step, where the use cases of the generated synthetic energy time series in the considered articles are identified. To categorise the identified use cases, this analysis uses the ten categories of smart grid data analysis derived in the systematic mapping study in [3]. Almost all of the identified use cases can be assigned to six of the ten categories when adapting the mapping to data generation, namely customer profiling, energy output forecasts, event analysis, load segregation, power loads/consumption, and privacy. Since all remaining identified use cases that do not fit into any of the ten categories deal with analyses on the network level, the final use case presented in this analysis is a new category called network analysis, resulting in seven total categories of use cases.

As shown in Fig. 9, the considered use cases occur with different frequencies. The majority of use cases with 48% forms power load/consumption, followed by customer profiling with 30%, energy

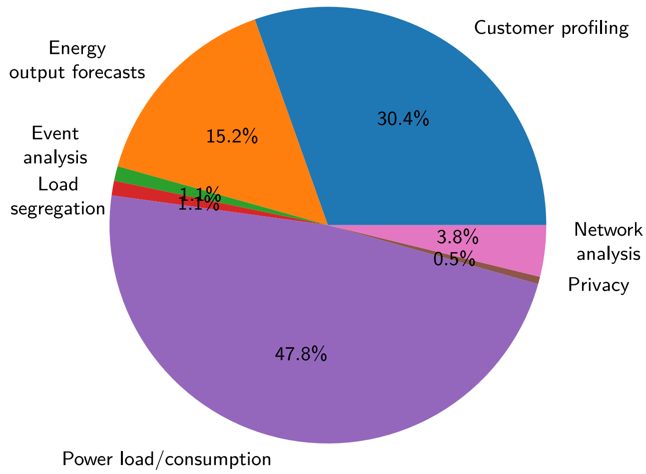

Fig. 9. Shares of the seven categories of use cases for the generated synthetic energy time series.

output forecasts with 15%, and network analysis with 4%. Event analysis, load segregation, and privacy, as the remaining three categories of use cases, each have a share of 1% or less.

Within the identified use cases there is a further difference in the set of use cases with regard to their aim: The sole aim of some use cases is the generation of synthetic data in a certain context, whereas others aim to solve a specific problem with the generated data. Therefore, this analysis additionally distinguishes generic and specific use cases within the seven used categories of use cases.

The following presents the seven categories of use cases and the related identified generic and specific use cases. Table 3 gives an overview of all methods for generating synthetic energy time series and the related generic and specific use cases in the seven categories of use cases.

7.1. Customer profiling

Customer profiling refers to the identification of the typical consumption pattern of users with similar characteristics [3]. Customer profiling is essential to understand the type and behaviour of consumers within a power grid. The main aim of customer profiling is to improve the power system's reliability, management and extension through comprehensive and accurate simulations based on consumption time series generated from the identified customer profiles. As a result, effective customer profiling requires consumer consumption data across a wide range of spatial and temporal resolutions as input, i.e. from single households to entire countries, considering a temporal resolution from one minute up to a day or larger. Due to this diverse data requirement, one of the main challenges in customer profiling is finding reliable data from representative users [124] at the desired spatial and temporal levels [107,131].

In this analysis, the following two generic use cases identified in the considered articles are assigned to customer profiling: the generation of load profiles, which represent the typical power consumption behaviour of user groups or appliances, and the creation of load patterns, which depict typical patterns in electrical power consumption. One specific use case that examines the self-selection bias of households that change their behaviour due to participating in a study is also assigned to customer profiling. Based on these use cases, all but one of the related articles consider one of the generic use cases. Within the two generic use cases, load profile and load pattern, most articles noticeably focus on load profiles.

7.2. Energy output forecasts

Energy output forecasts focus on forecasting the output from renewable energy resources [3]. Renewable energy sources are increasingly integrated into the system as the power grid transitions to reduce its carbon footprint but they are uncertain and cannot be easily controlled. As a result, obtaining forecasts of the expected outputs can help to make reliable decisions in power system operation, scheduling, and planning. However, the considered energy output depends on various conditions such as the temporal and spatial settings, the number of resources, and the weather [233]. Therefore, energy output time series can be generated to capture the desired conditions. One main challenge of energy output forecasts is thus to accurately capture the physical processes [204] and temporal dynamics [133] underlying the considered renewable energy sources.

Energy output forecasts match the following two generic identified use cases: the generation of synthetic power curves that represent the electrical power generation of a given renewable energy power plant and the creation of generation profiles, which represent the typical power generation of a power plant. This category also matches specific use cases, namely the generation of synthetic time series to appropriately size Photovoltaics (PV) power plants or batteries, the creation of synthetic time series to forecast electricity generation, and the creation of synthetic energy time series to predict future electricity generation for certain scenarios.

In this category, most articles address power curves as one of the generic use cases whereas the other generic use case, namely generation profile, only appears once. Among the specific use cases, battery sizing and battery and PV sizing are the most common, while forecasting and scenario generation only occur three and two times, respectively.

7.3. Event analysis

Event analysis concentrates on analysing relevant events in energy time series [3]. Its main aim is to recognise events as early as possible to be able to react to them if necessary. Due to the advancing integration of information and communication technologies into the power grid, event monitoring and analysis are increasingly automated. A variety of events to be monitored and analysed exist, which are also relevant to different stakeholders. For example, critical events such as extreme weather conditions or anomalies resulting in grid instabilities might require an immediate reaction from the grid operators. Conversely, events less critical for the grid stability such as non-technical losses require an action by the respective utility to prevent a large economic loss. All these events have in common that they rarely occur. For this reason, generated energy time series containing these events can support the development of automated systems to detect and analyse such events. Thereby, one main challenge in event analysis is to accurately capture and replicate these rare events.

Two of the identified specific use cases are assigned to this category, namely, the generation of energy time series for electricity theft detection and the determination of network losses. In line with the small share of this category, the use cases only appear in one article each.

7.4. Load segregation

Load segregation is concerned with non-intrusive load monitoring [3]. Non-intrusive load monitoring focuses on analysing the changes in voltage and current passing through a building to deduce which appliances are active [260]. Such load segregation is important for increasing the building's energy efficiency, improving demand response measures, and identifying inefficient and malfunctioning appliances. However, for effective load segregation, data without missing or corrupt values and containing many appliances and varying consumption patterns over long measurement periods is required. Such data, however, is not commonly available due to privacy concerns and the effort

Table 3
Overview of the identified methods for generating synthetic energy time series and the use cases where the synthetic energy time series generated by these methods are applied. The use cases are categorised into seven categories and comprise generic and specific use cases.

Method	Customer Profiling	Energy Output Forecasts	Event Analysis	Load Segregation	Power Load/ Consumption	Privacy	Network Analysis
Markov	Generic: Load profile: [107,118,124,126,127,130,131,133]	Generic: Generation profile: [133]; Power curve: [39,129,135,136,139,140] Specific: Battery sizing: [137]			Generic: Load curve: [106,108,110–114,116,117,119–123,125–128,132,134,138]; Load modelling: [109] Specific: Demand side management: [115,138]		Specific: Distribution network design: [118]
Monte Carlo	Generic: Load profile: [107]	Generic: Power curve: [54,129,135,140,146,149] Specific: Forecasting: [145]; Scenario generation: [148]	Specific: Network loss determination: [143]		Generic: Load curve: [108,113,114,117,123,141,142,144,147,150]; Load modelling: [109]		
wRNG	Generic: Load profile: [47,48,151,153–156,158,160,161,163,165,166,168–172,174,175,178,179,181,183,184,258]	Generic: Power curve: [185] Specific: Battery sizing: [179,188]		Generic: Load monitoring: [167]	Generic: Load curve: [157,162,164,168,173,176,177,182,186,187] Specific: Forecasting: [180]; Demand side management: [163]		Specific: Impact analysis of new technologies: [152,159]
GMM	Generic: Load profile: [124,127,130]	Generic: Power curve: [54] Specific: PV and battery sizing: [189]			Generic: Load curve: [53,127]		
NN	Generic: Load profile: [190,193,198,199]	Specific: Battery sizing: [137,192,195,197]			Generic: Load curve: [191,194] Specific: Scenario generation: [196]		Specific: Distribution network design: [190]
GAN	Generic: Load profile: [200,208,220,222]; Load pattern: [200,209,212]	Generic: Power curve: [201,203,204,216,219] Specific: PV and battery sizing: [189]; Battery sizing: [213]; Forecasting: [145]; Scenario generation: [205]		Generic: IoT data generation: [75]	Generic: Load curve: [76,187,201,210,214,216,217,221]; Load modelling: [202,215,218] Specific: Forecasting: [74,206,207]		Specific: Power flow sample generation: [211]
VAE	Generic: Load profile: [225]; Load pattern: [223]		Specific: Electricity theft detection: [224]		Generic: Load modelling: [215] Specific: Forecasting: [80]		
Regression	Generic: Load profile: [127,130,131,230] Specific: Self selection: [227]				Generic: Load curve: [114,127,229] Specific: Forecasting: [226,228]		
Autoregression	Generic: Load profile: [236]	Specific: PV and battery sizing: [189]; Forecasting: [233]			Generic: Load curve: [132,231,232,234,238,239] Specific: Forecasting: [206,233–235,237]		
Clustering	Generic: Load profile: [124,199,240]	Specific: Battery sizing: [137]			Generic: Load curve: [110,125,217,244–247]; Load modelling: [242,243]	Generic: Anonymisation: [97]	Generic: PMU data generation: [241]
Aggregation	Generic: Load profile: [118,251,252]				Generic: Load curve: [142,250,253,254]; Load modelling: [249] Specific: Demand side management: [248]		Specific: Distribution network design: [118]
BN	Generic: Load profile: [118,190]				Generic: Load curve: [138] Specific: Demand side management: [138]		Specific: Distribution network design: [118,190]
Data Variation	Generic: Load profile: [255,259]				Generic: Load curve: [257]		Generic: PMU data generation: [241,256]

required for associated measurement campaigns. Generated energy time series can be an alternative. One challenge with load segregation, however, is to have correct appliances' models and generated energy time series that represent the desired use of the appliances.

Load segregation matches two identified generic use cases in this analysis: the generation of synthetic energy time series for non-intrusive load monitoring and the creation of Internet of Things (IoT) data to support the dynamic management of the energy system. Consistent with the small share of the category, there is only one article for each use case.

7.5. Power loads/consumption

Power loads/consumption comprises analyses of the power consumption [3]. Similar to energy output forecasts, accurate power loads/consumption data is crucial for power system operation, scheduling, and extension planning. However, the availability of power consumption data is limited at low aggregation levels due to privacy concerns. Moreover, the power consumption evolves over time with the increasing integration of EVs, heat pumps, and renewable energies such as PV plants on a household level. For this reason, generated power consumption data can serve as a data source and can additionally be used to account for future changes, such as the inclusion of electric vehicles [53] and an increasing share of renewable energy sources [127]. Additionally, this generated power consumption data can be used to simulate future scenarios [125] and help develop demand side management strategies for these scenarios [116]. One of the biggest challenges in electricity loads/consumption is to accurately represent the desired use in the generated consumption time series.

Of the use cases identified, the following two generic use cases fit this category: the creation of load curves representing the electrical consumption over a certain period of time and load modelling using mathematical models that can be used to generate energy time series. Additionally, the following specific use cases match this category: the generation of energy time series to support demand side management analyses, the creation of energy time series to improve forecasts, and the generation of energy time series to predict future electricity consumption for certain scenarios. In this category, load curve is the most frequently occurring generic use case while only few articles address load modelling as the other generic case. In the specific use cases, many articles refer to forecasting, few to demand side management, and only one to scenario generation.

7.6. Privacy

Privacy deals with different aspects of avoiding the disclosure of private information when using smart grid data [3]. Privacy is one of the largest challenges facing smart grids. For smart grids to operate effectively, they continuously require and use data. However, making consumer data directly available raises multiple privacy concerns, as it is possible to derive confidential and personal information from it. Therefore, data anonymisation techniques are required to protect personal artefacts in the data. One promising approach is to generate privacy-preserving data from original data. However, it is associated with the challenge of keeping as many properties of the original data available for other smart grid applications without revealing any private information [261].

This category matches one identified generic use case in this analysis where load time series are anonymised to generate private synthetic load time series consistent with the original data for further analyses. In line with the small share of the category, only one article addressing this use case is identified.

7.7. Network analysis

Network analysis deals with analyses on the network level. The power network plays a central role in the power system as it, for example, allows energy supply and demand to be balanced. The associated analyses, therefore, cover various aspects of the network, including its design, its operation, and its behaviour under certain conditions. Since available data is limited and cannot necessarily cover all hypothetical or future considerations about the network addressed in the analyses, generated data can support the analyses with purposefully set properties. Nevertheless, the main challenge of network analysis is to ensure that the generated data sufficiently represents the network to be analysed.

Several generic and specific use cases are appropriate for the identified use cases in this analysis. The appropriate generic use cases comprise the generation of synthetic time series to simulate the network under different circumstances and the creation of Phasor Measurement Unit (PMU) data to facilitate the grid operation. The appropriate specific use cases are the generation of synthetic time series for the design of the distribution network, the creation of synthetic time series to analyse the impact of new technologies, and the creation of power flow samples for network analysis. In this category, only two articles address PMU data generation as the generic use case while distribution network design as a specific use case appears in most articles of this category.

7.8. Discussion & open research fields

There are several points worth discussing concerning the use cases of generated synthetic energy time series. Overall, the majority of identified articles focus on generic use cases. This could be useful since these generated synthetic time series can then be used to conduct research with different focuses as long as the generic use case is appropriate. However, this focus on generic use cases also means that only a small portion of the identified articles concentrate on a specific use case. This may be due to the fact that generic synthetic time series are sufficient for such use cases. However, it may also suggest a lack of focus on generating synthetic time series for specific use cases. This lack of focus could unnecessarily limit the benefits of using synthetic energy time series although they are very valuable for research such as the generation of scenarios for the design of the future energy system. Furthermore, there is an asymmetric distribution across the seven considered categories of use cases. Specifically, hardly any of the identified articles focus on event analysis, load segregation, or privacy. Finally, whilst certain categories such as customer profiling include representatives of all methods identified to generate synthetic time series, other categories such as energy output forecasts only contain a subset of these methods.

In light of these findings, three open research fields can be identified. First, it would be interesting to investigate further specific use cases. Additionally, it would be worth examining whether synthetic energy time series generated for a generic use case are sufficient for use in such applications or whether specific synthetic energy time series are required. Such use cases may include generating energy time series for specific network devices and evaluating the quality of this energy time series. Second, the usefulness of generating synthetic time series for the underrepresented categories, namely event analysis, load segregation, and privacy, should be examined in more detail. In particular, the use of synthetic energy time series to investigate very unlikely events could be useful, for example, in increasing the robustness of the energy system. Finally, it would be interesting to investigate why certain methods are not applied for particular use cases — especially with regard to energy output forecasts.

8. Evaluation of generated synthetic energy time series

This section presents the fifth analysis step, in which the evaluation performed on the generated synthetic energy time series is considered. This analysis distinguishes the determined evaluations by their aim. For this classification, the three aims applied in [10] are considered, namely diversity, fidelity, and usefulness. Since not all identified evaluations pursue these aims, computational performance is considered as an additional aim.

As shown in Fig. 10, more than three quarters of all evaluations follow the aim of fidelity. Rather small shares of evaluations pursue the aims of diversity, usefulness, and computational performance, or are not specified in more detail. The following presents the identified evaluations and applied methods along these four aims. Table 4 gives an overview of all methods for generating synthetic energy time series and the identified performed evaluations.

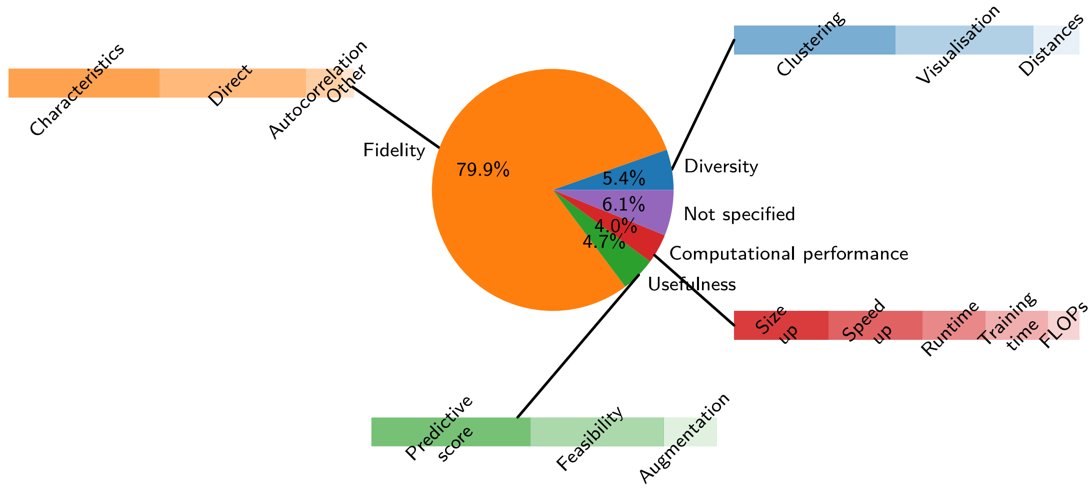

Fig. 10. Shares of the different evaluation aims of generated synthetic energy time series and the applied methods. Other evaluation methods of fidelity comprise regression and discriminative score.

Table 4
Overview of the identified methods for generating synthetic energy time series and the performed evaluations of the synthetic energy time series generated thereby. The performed evaluations are structured along their aims and list the applied methods.

Method	Diversity	Fidelity	Usefulness	Computational Performance
Markov	Clustering: [124,125]	Autocorrelation: [39,107,112,124,125,129,135,136,139,140]; Characteristics: [106,108–110,112,113,116–122,124,127–134,136,138,140]; Direct: [107–110,113,114,116,117,120,121,123,124,126–130,132,135,140]; Regression: [117,130]	Feasibility: [136]	Size up: [132]; Speed up: [132]
Monte Carlo		Autocorrelation: [54,107,129,135,140,149]; Characteristics: [54,108,109,113,117,129,140,142,144–146,148]; Direct: [107–109,113,114,117,123,129,135,140–142,145,146,148]; Regression: [117]		Training time: [145]
wRNG		Autocorrelation: [183]; Characteristics: [47,48,155,157–159,164,167,171–173,175–178,181–184,186,187,258]; Direct: [47,48,151–154,157–159,162–164,166–168,171–174,179–181,183,185,187,258]	Feasibility: [176]; Predictive score: [187]	
GMM	Clustering: [124]	Autocorrelation: [54,124]; Characteristics: [54,124,127,130]; Direct: [53,124,127,130,189]; Regression: [130]	Feasibility: [189]	
NN	Visualisation: [193]	Autocorrelation: [194,196]; Characteristics: [190,194,199]; Direct: [190–192,194–198]; Regression: [198]	Feasibility: [193]	
GAN	Clustering: [205,217]; Visualisation: [200,202,211–213]; Distances: [220]	Autocorrelation: [204,205]; Characteristics: [76,145,187,201–208,214,217,218,220,222]; Direct: [75,76,145,187,189,200,204,206,209,210,212–216,219,221]; Discriminative score: [202,208,209]	Augmentation: [74]; Feasibility: [189,214]; Predictive score: [187,200,202,213,214]	FLOPs: [75]; Runtime: [219]; Training time: [145,219]
VAE		Characteristics: [80,223,224]; Direct: [80,215,224,225]	Augmentation: [224]; Predictive score: [223]	
Regression		Autocorrelation: [226]; Characteristics: [127,130,131,227–230]; Direct: [114,127,130]; Regression: [130,228]		
Autoregression	Clustering: [232]	Autocorrelation: [234,239]; Characteristics: [132,206,231,233–239]; Direct: [132,189,206,232,233]; Regression: [237]	Feasibility: [189]	Size up: [132,232]; Speed up: [132,232]
Clustering	Clustering: [124,125,217,246]	Autocorrelation: [124,125]; Characteristics: [97,110,124,199,217,243,244]; Direct: [97,110,124,240,242,244,245]		Runtime: [245]
Aggregation		Autocorrelation: [250]; Characteristics: [118,142,249,250,253]; Direct: [142,248–252,254]		Size up: [253]; Speed up: [253]
BN		Characteristics: [118,138,190]; Direct: [190]		
Data Variation		Characteristics: [255,256,259]; Direct: [257]		

8.1. Diversity

Diversity aims to ensure that synthetic time series cover all values of real time series. Thus, it is necessary to generate all types of time series from the original data set so that not only a subset of the original data is generated. This aim can be evaluated using three different evaluation methods.

The first evaluation method is visualisation, which is a qualitative evaluation. For this, sections of the synthetic and real time series are created and then visualised using a two-dimensional t-distributed Stochastic Neighbour Embedding (t-SNE) [262] or Principal Component Analysis (PCA) [263]. A synthetic time series performs well in terms of diversity when the sections of the synthetic and real time series are similarly distributed on the projection plane. Similarly, a heat map can be used to visualise which specific power values at given periods of time are covered by the generated synthetic time series. The second evaluation method is clustering. It assumes that there are different clusters in the real time series. For the evaluation, both the synthetic and the real time series are clustered and the corresponding clusters are compared. The diversity is high when these clusters are similar for both time series. As the third evaluation method for diversity, distances can be calculated. These distances can be determined between generated time series or between generated and real time series using, for example, the Euclidean distance.

Table 4 shows that only six of the thirteen generation methods are evaluated considering diversity. For most of them, clustering and visualisation are applied to determine the diversity of the generated energy time series. For only one generation method, distances are calculated concerning the diversity.

8.2. Fidelity

Fidelity aims for synthetic time series that are similar to and indistinguishable from real time series. Five different evaluation methods are applied to evaluate this aim.

The first evaluation method compares synthetic and real time series characteristics such as statistical moments and hallmark characteristics. Once derived, these characteristics can be compared using distance metrics or visualisations to assess their similarity. The second evaluation method is the direct comparison of the synthetic and real time series. This method directly compares both time series or the distributions of both time series without extracting or deriving any characteristics. This comparison can be done either qualitatively using visualisations or quantitatively using metrics. The third evaluation method is the calculation of a discriminative score. Similar to the direct comparison, the discriminative score also aims to compare the similarity without deriving any features. For this purpose, a classifier is trained to distinguish the synthetic from the real time series. Given the classifier's performance, the discriminative score can be derived. The fourth evaluation method is the autocorrelation. It calculates and compares the autocorrelative structure of the synthetic and the real time series to assess whether the temporal structure of both time series is similar. The fifth evaluation method is a regression, which assesses whether the synthetic and real time series are similarly influenced by exogenous variables. To determine this influence, the correlation between the time series and exogenous variables is analysed.

Consistent with the large share of this evaluation, Table 4 shows that the energy time series generated by all methods are evaluated regarding their fidelity. More specifically, the comparison of characteristics and the direct comparison are both used for all generation methods. In comparison, noticeably fewer articles apply autocorrelation and only few articles the discriminative score or regression to determine an appropriate fidelity.

8.3. Usefulness

Usefulness measures synthetic time series with regard to their suitability for downstream applications. Ideally, a synthetic time series should be at least as useful for downstream applications as real time series. Thereby, a downstream application can be a forecaster, a classifier, or a more specific domain-related application such as a battery storage system. For the evaluation of this aim, three evaluation methods are applied.

The first evaluation method is the calculation of a predictive score. Based on Train on Synthetic and Test on Real (TSTR) method as proposed in [9], a downstream application is trained with the synthetic time series and evaluated with the real time series. The better the performance of the downstream application on the real time series is, the more useful is the synthetic time series. Augmentation as the second evaluation method is strongly related to the TSTR method. It trains a downstream application once on the real time series only and once on the combination of real and synthetic time series, before it compares their results. If the combination provides better results than just the real time series, then the synthetic time series is useful. The third evaluation method is a feasibility assessment that determines whether the synthetic time series is usable by a downstream application. If, for example, the synthetic time series corresponds to a dispatch schedule, the feasibility checks whether this dispatch schedule can be realised.

In Table 4, identified articles of six generation methods are evaluated concerning their usefulness. While most articles determine a predictive score or apply a feasibility assessment, only two articles perform an augmentation.

8.4. Computational performance

Computational performance aims to generate synthetic time series with as little effort as possible and thus focuses on the generation itself. To evaluate different aspects of computational performance, five different evaluation methods are used.

The first evaluation method focuses on the effort needed to generate a time series by measuring the runtime. It is the time required by the generation method to generate a time series. Another evaluation method concentrates on the preparation of a generation method and measures the training time. It refers to the time that is necessary to train a generation method before it can be used to generate time series. The third evaluation method considers the complexity of the applied methods by measuring the required Floating-Point Operations Per Seconds (FLOPs) that represent basic machine operations. The focus of the fourth evaluation method lies on the scalability of the generation method and examines its size-up behaviour. The size up determines the effect of adding additional computing resources on the number of time series generated in a fixed period of time. The fifth evaluation method also concentrates on the scalability and analyses the speed-up behaviour of a generation method. The speed up measures the effect of adding additional computing resources on the required time when generating a fixed number of synthetic time series.

Table 4 shows that again articles of only six generation methods evaluate the computational performance of the generation itself. Of the available evaluation methods, all are used roughly with the same frequency.

8.5. Discussion & open research fields

Given the previous observations, including Table 4 and Fig. 10, this analysis leads to four main observations. First, the vast majority of identified articles focus on fidelity without considering usefulness for a specific use case. However, there is little point in generating a synthetic energy time series if this time series is not useful for a given scenario. Second, whilst some articles consider multiple aims such as both diversity and fidelity, none of them considers all four identified aims of evaluating synthetic energy time series. As a result, it is challenging to identify state-of-the-art generation methods since different evaluations are applied, and it is thus actually hard to compare the results. Third, even if some articles evaluate synthetic energy time series with the same aim, the methods used to evaluate this aim also vary, which makes a comparison of the results even more difficult. Finally, many of the available evaluation methods focus on directly comparing generated time series and real time series or comparing their derived characteristics. This approach is beneficial given the assumption that synthetic time series should resemble observed characteristics. However, this assumption does not hold for use cases where potentially observable but not yet observed characteristics are important. For these use cases, simply comparing statistical properties and derived characteristics may result in misleading evaluation results.

Therefore, three important open research fields can be identified. First, it is imperative that a gold standard for the evaluation of synthetic energy time series is established. This standard should consider all four of the identified aims and be applicable to a wide variety of generation methods. Only once such a standard is established can generated synthetic time series be accurately compared and benchmarked. Second, the usefulness of generated synthetic time series should be given higher priority. Simply generating an energy time series with statistically similar properties is only useful if this generated time series is also useful for the considered use case. Therefore, criteria defining when a time series is useful should be developed and investigated concerning multiple use cases. Finally, many of the proposed evaluation methods are qualitative in nature. Therefore, it would be interesting to investigate whether further quantitative methods can be found, specifically for diversity.

9. Conclusion

To support the transition to smart grids from predominately renewable sources, energy time series are required for developing and improving methods for smart grid applications. Since large amounts of energy time series are not openly available, generating synthetic energy time series is often a promising solution. However, these generated synthetic energy time series must exhibit characteristics similar to real energy time series and applicable to specific use cases. Furthermore, these characteristics of the synthetic time series must be verified by evaluation methods. Although a wide range of methods exist in the literature for generating synthetic energy time series, up until now, no structured literature review exists that identifies existing generation methods, including the applied generation approach as well as the characteristics, use cases, and evaluation of the generated energy time series. Therefore, this study presents a systematic literature review of methods for generating synthetic energy time series from electricity generation, electricity consumption, and battery storage systems. The review focuses on five key aspects: (1) Identifying which methods are used to generate synthetic energy time series, (2) categorising these methods according to the applied generation approach, (3) identifying the characteristics exhibited by the generated time series, (4) discovering which use cases the generated time series are applied to, and (5) describing how the synthetic energy time series are evaluated.

This structured literature review considers 169 articles and provides five key insights regarding the generation of synthetic energy time series. First, a large range of methods are used to generate synthetic energy time series. Second, these methods are applied with different generation approaches, namely the bottom-up, top-down, and direct approaches. Furthermore, whilst all three generation approaches are applied, the majority of the identified articles use either the bottom-up or direct approach. Third, the majority of the generated energy time series are consumption time series of residential use at a household, appliance, or single unit aggregation level. Only a few of the identified articles focus on generation energy time series, and hardly any consider time series of battery storage systems. Fourth, whilst many of the generated synthetic energy time series are applied for the use cases of customer profiling, power/load consumption, and energy output forecasts, hardly any are considered for the use cases of event analysis, load segregation, and privacy. Moreover, the majority of the identified use cases are generic in nature, with only a small subset of the considered articles focusing on specific use cases. Finally, the vast majority of the identified articles focus on evaluating the fidelity of the generated synthetic energy time series, whilst most ignore their usefulness. Furthermore, none of the considered articles evaluates all four aims of synthetic time series, namely diversity, usefulness, fidelity, and computational performance.

Based on this analysis, several open research fields can be identified, four of which are outlined in the following. First, an important open research field is to correctly evaluate generated synthetic time series. A standard evaluation of generated synthetic time series must be established that considers all evaluation aims and that can be applied to benchmark a wide range of methods. Importantly, such a standard evaluation should specifically consider the usefulness of the generated synthetic energy time series and open up to use cases that have other requirements than that the statistical properties are similar to real data. Second, future work should consider exploring the top-down approach for generating synthetic energy time series in more detail. Third, to help facilitate the transition to a smart grid from mostly renewable energy sources, far more work should focus on generating synthetic energy time series for generation and battery storage systems, specifically for residential and commercial contexts. Finally, generating synthetic energy time series with a focus on event analysis, load segregation, and privacy as use cases is also a promising open research field.

CRediT authorship contribution statement

M. Turowski: Conceptualisation, Data Curation, Investigation, Methodology, Writing – original draft, Visualisation, Writing – review & editing. **B. Heidrich:** Conceptualisation, Investigation, Visualisation, Writing – original draft. **L. Weingärtner:** Conceptualisation, Investigation, Validation, Visualisation, Writing – original draft. **L. Springer:** Data Curation, Investigation, Writing – original draft. **K. Phipps:** Conceptualisation, Writing – original draft, Writing – review & editing. **B. Schäfer:** Writing – review & editing. **R. Mikut:** Funding acquisition, Supervision, Writing – review & editing. **V. Hagenmeyer:** Funding acquisition, Supervision, Writing – review & editing.

Declaration of competing interest

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

Data availability

No data was used for the research described in the article.

References

[1] Farhangi Hassan. The path of the smart grid. IEEE Power Energy Mag 2010;8(1):18–28. http://dx.doi.org/10.1109/MPE.2009.934876.
[2] Li Fangxing, Qiao Wei, Sun Hongbin, Wan Hui, Wang Jianhui, Xia Yan, et al. Smart transmission grid: Vision and framework. IEEE Trans Smart Grid 2010;1(2):168–77. http://dx.doi.org/10.1109/TSG.2010.2053726.
[3] Rossi Bruno, Chren Stanislav. Smart grids data analysis: A systematic mapping study. IEEE Trans Ind Inf 2020;16(6):3619–39. http://dx.doi.org/10.1109/TII.2019.2954098.
[4] Hastie Trevor, Tibshirani Robert, Friedman Jerome. The elements of statistical learning. 2nd ed. New York: Springer; 2009, http://dx.doi.org/10.1007/978-0-387-84858-7.
[5] Hippert Henrique Steinherz, Pedreira Carlos Eduardo, Souza Reinaldo Castro. Neural networks for short-term load forecasting: A review and evaluation. IEEE Trans Power Syst 2001;16(1):44–55. http://dx.doi.org/10.1109/59.910780.
[6] Pfenninger Stefan, DeCarolis Joseph, Hirth Lion, Quoilin Sylvain, Staffell Iain. The importance of open data and software: Is energy research lagging behind? Energy Policy 2017;101:211–5. http://dx.doi.org/10.1016/j.enpol.2016.11.046.
[7] Rydin Gorjão Leonardo, Jumar Richard, Maass Heiko, Hagenmeyer Veit, Yalcin G Cigdem, Kruse Johannes, et al. Open database analysis of scaling and spatio-temporal properties of power grid frequencies. Nature Commun 2020;11(1):6362. http://dx.doi.org/10.1038/s41467-020-19732-7.
[8] Dannecker Lars. Energy time series forecasting: Efficient and accurate forecasting of evolving time series from the energy domain. Wiesbaden: Springer Fachmedien; 2015, http://dx.doi.org/10.1007/978-3-658-11039-0.
[9] Esteban Cristóbal, Hyland Stephanie L, Rätsch Gunnar. Real-valued (medical) time series generation with recurrent conditional GANs. 2017.
[10] Yoon Jinsung, Jarrett Daniel, van der Schaar Mihaela. Time-series generative adversarial networks. In: Wallach H, Larochelle H, Beygelzimer A, D'Alché-Buc F, Fox E, Garnett R, editors. In: Advances in neural information processing systems, vol. 32, Curran Associates, Inc.; 2019, p. 5508–18.
[11] Sohn Kihyuk, Yan Xinchen, Lee Honglak. Learning structured output representation using deep conditional generative models. In: Cortes C, Lawrence N, Lee D, Sugiyama M, Garnett R, editors. In: Advances in neural information processing systems, vol. 28, Curran Associates, Inc.; 2015, p. 3483–91.

[12] Heidrich Benedikt, Turowski Marian, Phipps Kaleb, Schmieder Kai, Süß Wolfgang, Mikut Ralf, et al. Controlling non-stationarity and periodicities in time series generation using conditional invertible neural networks. Appl Intell 2023;53:8826–43. http://dx.doi.org/10.1007/s10489-022-03742-7.

[13] Grandjean A, Adnot J, Binet G. A review and an analysis of the residential electric load curve models. Renew Sustain Energy Rev 2012;16(9):6539–65. http://dx.doi.org/10.1016/j.rser.2012.08.013.

[14] Proedrou Elisavet. A comprehensive review of residential electricity load profile models. IEEE Access 2021;9:12114–33. http://dx.doi.org/10.1109/ACCESS.2021.3050074.

[15] Assefa Samuel A, Dervovic Danial, Mahfouz Mahmoud, Tillman Robert E, Reddy Prashant, Veloso Manuela. Generating synthetic data in finance: opportunities, challenges and pitfalls. In: Proceedings of the first ACM international conference on AI in finance. ACM; 2020, http://dx.doi.org/10.1145/3383455.3422554.

[16] Goncalves Andre, Ray Priyadip, Soper Braden, Stevens Jennifer, Coyle Linda, Sales Ana Paula. Generation and evaluation of synthetic patient data. BMC Med Res Methodol 2020;20(108). http://dx.doi.org/10.1186/s12874-020-00977-1.

[17] Raghunathan Trivellore E. Synthetic data. Annu Rev Stat Appl 2021;8:129–40. http://dx.doi.org/10.1146/annurev-statistics-040720-031848.

[18] Surendra H, Mohan HS. A review of synthetic data generation methods for privacy preserving data publishing. International Journal of Scientific & Technology Research 2017;6(5):95–101.

[19] Cramer Eike, Gorjão Leonardo Rydin, Mitsos Alexander, Schäfer Benjamin, Witthaut Dirk, Dahmen Manuel. Validation methods for energy time series scenarios from deep generative models. IEEE Access 2022;10:8194–207. http://dx.doi.org/10.1109/ACCESS.2022.3141875.

[20] Bahrpeyma Fouad, Roantree Mark, Cappellari Paolo, Scriney Michael, McCarren Andrew. A methodology for validating diversity in synthetic time series generation. MethodsX 2021;8:101459. http://dx.doi.org/10.1016/j.mex.2021.101459.

[21] Thorve Swapna, Vullikanti Anil, Mortveit Henning S, Swarup Samarth, Marathe Madhav V. Fidelity and diversity metrics for validating hierarchical synthetic data: Application to residential energy demand. In: 2022 IEEE international conference on big data. IEEE; 2022, p. 1377–82. http://dx.doi.org/10.1109/BigData55660.2022.10020837.

[22] Bhanot Karan, Pedersen Joseph, Guyon Isabelle, Bennett Kristin P. Investigating synthetic medical time-series resemblance. Neurocomputing 2022;494:368–78. http://dx.doi.org/10.1016/j.neucom.2022.04.097.

[23] McLoughlin Fintan, Duffy Aidan, Conlon Michael. Evaluation of time series techniques to characterise domestic electricity demand. Energy 2013;50:120–30. http://dx.doi.org/10.1016/j.energy.2012.11.048.

[24] Swan Lukas G, Ugursal V Ismet. Modeling of end-use energy consumption in the residential sector: A review of modeling techniques. Renew Sustain Energy Rev 2009;13(8):1819–35. http://dx.doi.org/10.1016/j.rser.2008.09.033.

[25] Synnott Jonathan, Nugent Chris, Jeffers Paul. Simulation of smart home activity datasets. Sensors 2015;15(6):14162–79. http://dx.doi.org/10.3390/s150614162.

[26] Palma-Behnke Rodrigo, Vega-Herrera Jorge, Valencia Felipe, Núñez-Mata Oscar. Synthetic time series generation model for analysis of power system operation and expansion with high renewable energy penetration. J Mod Power Syst Clean Energy 2021;9(4):849–58. http://dx.doi.org/10.35833/MPCE.2020.000747.

[27] Webster Jane, Watson Richard T. Analyzing the past to prepare for the future: Writing a literature review. MIS Q 2002;26(2):xiii–xiii.

[28] Alvisi S, Ansaloni N, Franchini M. Comparison of parametric and nonparametric disaggregation models for the top-down generation of water demand time series. Civ Eng Environ Syst 2016;33(1):3–21. http://dx.doi.org/10.1080/10286608.2015.1126823.

[29] Ansaloni N, Alvisi S, Franchini M. Generation of synthetic cross-correlated water demand time series. Water Supply 2013;13(4):977–86. http://dx.doi.org/10.2166/ws.2013.048.

[30] Bakhshipour Amin E, Koochali Alireza, Dittmer Ulrich, Haghighi Ali, Ahmad S, Dengel A. A Bayesian generative adversarial network (GAN) to generate synthetic time-series data, application in combined sewer flow prediction. In: 2nd international joint conference on water distribution systems analysis & computing and control in the water industry. Universitat Politécnica de València; 2022, http://dx.doi.org/10.4995/WDSA-CCWI2022.2022.14699.

[31] Wu Jianxun, Reddy T Agami, Claridge David. Statistical modeling of daily energy consumption in commercial buildings using multiple regression and principal component analysis. In: Eighth symposium on improving building systems in hot and humid climates. 1992, p. 155–64.

[32] Menapace Andrea, Santopietro Simone, Gargano Rudy, Righetti Maurizio. Stochastic generation of district heat load. Energies 2021;14(17):5344. http://dx.doi.org/10.3390/en14175344.

[33] Levin David A, Peres Yuval. Markov chains and mixing times. 2nd ed. Providence: American Mathematical Society; 2017.

[34] Baum Leonard E, Petrie Ted. Statistical inference for probabilistic functions of finite state Markov chains. Ann Math Stat 1966;37(6):1554–63. http://dx.doi.org/10.1214/aoms/1177699147.

[35] Rabiner LR, Juang BH. An introduction to hidden Markov models. IEEE ASSP Mag 1986;3(1):4–16. http://dx.doi.org/10.1109/MASSP.1986.1165342.

[36] Mor Bhavya, Garhwal Sunita, Kumar Ajay. A systematic review of hidden Markov models and their applications. Arch Comput Methods Eng 2021;28(3):1429–48. http://dx.doi.org/10.1007/s11831-020-09422-4.

[37] Messina Enza, Toscani Daniele. Hidden Markov models for scenario generation. IMA J Manag Math 2008;19(4):379–401. http://dx.doi.org/10.1093/imaman/dpm026.

[38] Andersson Maria. Modeling electricity load curves with hidden Markov models for demand-side management status estimation. Int Trans Electr Energy Syst 2017;27(3):e2265. http://dx.doi.org/10.1002/etep.2265.

[39] Chen Peiyuan, Berthelsen Kasper Klitgaard, Bak-Jensen Birgitte, Chen Zhe. Markov model of wind power time series using Bayesian inference of transition matrix. In: IECON proceedings (Industrial electronics conference). IEEE; 2009, p. 627–32. http://dx.doi.org/10.1109/IECON.2009.5414993.

[40] Harrison Robert L. Introduction to Monte Carlo simulation. AIP Conf Proc 2010;1204(1):17–21. http://dx.doi.org/10.1063/1.3295638.

[41] Raychaudhuri Samik. Introduction to Monte Carlo simulation. In: 2008 winter simulation conference. IEEE; 2008, p. 91–100. http://dx.doi.org/10.1109/WSC.2008.4736059.

[42] Beasley William Howard, Rodgers Joseph Lee. Bootstrapping and Monte Carlo methods. In: Cooper Harris, Camic Paul M, Long Debra L, Panter AT, Rindskopf David, Sher Kenneth J, editors. APA handbook of research methods in psychology. American Psychological Association; 2012, p. 407–25. http://dx.doi.org/10.1037/13620-022.

[43] Efron Bradley, Tibshirani Robert J. An introduction to bootstrap. New York: Chapman and Hall/CRC; 1994, http://dx.doi.org/10.1201/9780429246593.

[44] Gilks WR, Richardson S, Spiegelhalter David, editors. Markov chain Monte Carlo in practice. New York: Chapman and Hall/CRC; 1995, http://dx.doi.org/10.1201/b14835.

[45] Mooney Christopher Z. Monte Carlo simulation. Thousand Oaks: SAGE Publications; 1997.

[46] L'Ecuyer Pierre. Random number generation. In: Gentle James E, Härdle Wolfgang Karl, Mori Yuichi, editors. Handbook of computational statistics: Concepts and methods. 2nd ed. Berlin, Heidelberg: Springer; 2012, http://dx.doi.org/10.1007/978-3-642-21551-3.

[47] Foteinaki Kyriaki, Li Rongling, Rode Carsten, Andersen Rune Korsholm. Modelling household electricity load profiles based on danish time-use survey data. Energy Build 2019;202:109355. http://dx.doi.org/10.1016/j.enbuild.2019.109355.

[48] Mahmood Imran, Quair-tul-ain, Nasir Hasan Arshad, Javed Fahad, Aguado José A. A hierarchical multi-resolution agent-based modeling and simulation framework for household electricity demand profile. Simulation 2020;96(8):655–78. http://dx.doi.org/10.1177/0037549720923401.

[49] Singh Ravindra, Pal Bikash C, Jabr Rabih A. Statistical representation of distribution system loads using Gaussian mixture model. IEEE Trans Power Syst 2010;25(1):29–37. http://dx.doi.org/10.1109/TPWRS.2009.2030271.

[50] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Ser B Stat Methodol 1977;39(1):1–38. http://dx.doi.org/10.1111/j.2517-6161.1977.tb01600.x.

[51] Reynolds Douglas. Gaussian mixture models. In: Li Stan Z, Jain Anil K, editors. Encyclopedia of biometrics. 2nd ed. New York: Springer Science & Business Media; 2015, p. 827–32. http://dx.doi.org/10.1007/978-1-4899-7488-4_196.

[52] Bouguila Nizar, Fan Wentao, editors. Mixture models and applications. The exponential distribution. Springer Nature Switzerland; 2020, http://dx.doi.org/10.1007/978-3-030-23876-6.

[53] Bernards Raoul, Morren Johan, Slootweg Han. Statistical modelling of load profiles incorporating correlations using copula. In: 2017 IEEE PES innovative smart grid technologies conference Europe. IEEE; 2017, http://dx.doi.org/10.1109/ISGTEurope.2017.8260181.

[54] Li Yufei, Hu Bo, Niu Tao, Gao Shengpu, Yan Jiahao, Xie Kaigui, Ren Zhouyang. GMM-HMM-based medium- and long-term multi-wind farm correlated power output time series generation method. IEEE Access 2021;9:90255–67. http://dx.doi.org/10.1109/ACCESS.2021.3091460.

[55] Zhou Zhi-Hua. Machine learning. Singapore: Springer; 2021, http://dx.doi.org/10.1007/978-981-15-1967-3.

[56] Kohonen Teuvo. Self-organized formation of topologically correct feature maps. Biol Cybernet 1982;43(1):59–69. http://dx.doi.org/10.1007/BF00337288.

[57] LeCun Yann, Bengio Yoshua. Convolutional networks for images, speech, and time-series. In: The handbook of brain theory and neural networks. MIT Press; 1995, p. 255–8.

[58] Salehinejad Hojjat, Sankar Sharan, Barfett Joseph, Colak Errol, Valaee Shahrokh. Recent advances in recurrent neural networks. 2018.

[59] Scarselli Franco, Gori Marco, Tsoi Ah Chung, Hagenbuchner Markus, Monfardini Gabriele. The graph neural network model. IEEE Trans Neural Netw 2009;20(1):61–80. http://dx.doi.org/10.1109/TNN.2008.2005605.

[60] Li Yujia, Swersky Kevin, Zemel Richard. Generative moment matching networks yujia. In: 32nd international conference on machine learning, vol. 3. 2015, p. 1718–27.

[61] Dinh Laurent, Krueger David, Bengio Yoshua. NICE: Non-linear independent components estimation. In: 3rd international conference on learning representations. 2015.

[62] Dinh Laurent, Sohl-Dickstein Jascha, Bengio Samy. Density estimation using real NVP. In: 5th international conference on learning representations. 2017.

[63] Ardizzone Lynton, Kruse Jakob, Wirkert Sebastian, Rahner Daniel, Pellegrini Eric W, Klessen Ralf S, et al. Analyzing inverse problems with invertible neural networks. In: 7th international conference on learning representations. 2019.

[64] Gurney Kevin. An introduction to neural networks. London: CRC Press; 1997, http://dx.doi.org/10.1201/9781315273570.

[65] LeCun Yann, Bengio Yoshua, Hinton Geoffrey. Deep learning. Nature 2015;521:436–44. http://dx.doi.org/10.1038/nature14539.

[66] Goodfellow Ian, Bengio Yoshua, Courville Aaron. Deep learning. Cambridge: MIT Press; 2016.

[67] Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, et al. Generative adversarial nets. In: Ghahramani Z, Welling M, Cortes C, Lawrence N, Weinberger KQ, editors. In: Advances in neural information processing systems, vol. 27, Curran Associates, Inc.; 2014, p. 4089–99.

[68] Arjovsky Martin, Chintala Soumith, Bottou Léon. Wasserstein generative adversarial networks. In: Precup Doina, Teh Yee Whye, editors. Proceedings of the 34th international conference on machine learning, vol. 70. PMLR; 2017, p. 214–23.

[69] Mirza Mehdi, Osindero Simon. Conditional generative adversarial nets. 2014.

[70] Radford Alec, Metz Luke, Chintala Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks. In: 4th international conference on learning representations. 2016.

[71] Saatchi Yunus, Wilson Andrew Gordon. Bayesian GAN. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R, editors. In: Advances in neural information processing systems, vol. 30, Curran Associates, Inc.; 2017, p. 3622–31.

[72] Goodfellow Ian, Pouget-Abadie Jean, Mirza Mehdi, Xu Bing, Warde-Farley David, Ozair Sherjil, et al. Generative adversarial networks. Commun ACM 2020;63(11):139–44. http://dx.doi.org/10.1145/3422622.

[73] Creswell Antonia, White Tom, Dumoulin Vincent, Arulkumaran Kai, Sengupta Biswa, Bharath Anil A. Generative adversarial networks: An overview. IEEE Signal Process Mag 2018;35(1):53–65. http://dx.doi.org/10.1109/MSP.2017.2765202.

[74] Tian Chenlu, Li Chengdong, Zhang Guiqing, Lv Yisheng. Data driven parallel prediction of building energy consumption using generative adversarial nets. Energy Build 2019;186:230–43. http://dx.doi.org/10.1016/j.enbuild.2019.01.034.

[75] Ma Zhengjing, Mei Gang, Piccialli Francesco. An attention based cycle-consistent generative adversarial network for IoT data generation and its application in smart energy systems. IEEE Trans Ind Inf 2023;19(4):6170–81. http://dx.doi.org/10.1109/TII.2022.3204282.

[76] Baasch Gaby, Rousseau Guillaume, Evins Ralph. A conditional generative adversarial network for energy use in multiple buildings using scarce data. Energy AI 2021;5:100087. http://dx.doi.org/10.1016/j.egyai.2021.100087.

[77] Kingma Diederik P, Welling Max. Auto-encoding variational Bayes. 2014.

[78] Doersch Carl. Tutorial on variational autoencoders. 2016.

[79] Pinheiro Cinelli Lucas, Araújo Marins Matheus, da Silva Eduardo Antúnio, Lima Netto Sérgio. Variational autoencoder. In: Variational methods for machine learning with applications to deep networks. Cham: Springer; 2021, p. 111–49. http://dx.doi.org/10.1007/978-3-030-70679-1_5.

[80] Fan Cheng, Chen Meiling, Tang Rui, Wang Jiayuan. A novel deep generative modeling-based data augmentation strategy for improving short-term building energy predictions. Build Simul 2022;15(2):197–211. http://dx.doi.org/10.1007/s12273-021-0807-6.

[81] Montgomery Douglas C, Peck Elizabeth A, Vining G Geoffrey. Introduction to linear regression analysis. 6th ed. Hoboken: John Wiley & Sons; 2021.

[82] Draper Norman R, Smith Harry. Applied regression analysis. 3rd ed. John Wiley & Sons; 1998, http://dx.doi.org/10.1002/9781118625590.

[83] Ryan Thomas P. Modern regression methods. 2nd ed. Hoboken: John Wiley & Sons; 2008, http://dx.doi.org/10.1002/9780470382806.

[84] Lewis-Beck Colin, Lewis-Beck Michael. Applied regression: An introduction. Thousand Oaks: Sage Publications; 2016.

[85] Box George EP, Jenkins Gwilym M, Reinsel Gregory C, Ljung Great M. Time series analysis. 5th ed. New Jersey: John Wiley & Sons Hoboken; 2016.

[86] Hamilton James D. Time series analysis. Princeton University Press; 1994, http://dx.doi.org/10.1515/9780691218632.

[87] Cryer Jonathan D, Chan Kung-Sik. Time series analysis. 2nd ed. New York: Springer; 2008, http://dx.doi.org/10.1007/978-0-387-75959-3.

[88] Benjamin Michael A, Rigby Robert A, Stasinopoulos D Mikis. Generalized autoregressive moving average models. J Amer Statist Assoc 2003;98(461):214–23. http://dx.doi.org/10.1198/016214503388619238.

[89] Aghabozorgi Saeed, Seyed Shirkhorshidi Ali, Ying Wah Teh. Time-series clustering - A decade review. Inf Syst 2015;53:16–38. http://dx.doi.org/10.1016/j.is.2015.04.007.

[90] MacQueen James. Some methods for classification and analysis of multivariate observations. In: Le Cam Lucien, Neyman Jerzey, editors. Proceedings of the fifth Berkeley symposium on mathematica statistics and probability. University of California Press; 1967, p. 281–97.

[91] Bezdek James C. Pattern recognition with fuzzy objective function algorithms. New York: Plenum Press; 1981.

[92] von Luxburg Ulrike. A tutorial on spectral clustering. Stat Comput 2007;17:395–416. http://dx.doi.org/10.1007/s11222-007-9033-z.

[93] Witten Ian H, Frank Eibe. Data mining: Practical machine learning tools and techniques. 2nd ed. Morgan Kaufmann Publishers; 2005, p. 150.

[94] Rokach Lior, Maimon Oded. Clustering methods. In: Maimon Oded, Rokach Lior, editors. Data mining and knowledge discovery handbook. Boston: Springer; 2005, p. 321–52. http://dx.doi.org/10.1007/0-387-25465-X_15.

[95] Xu Rui, Wunsch Donald. Survey of clustering algorithms. IEEE Trans Neural Netw 2005;16(3):645–78. http://dx.doi.org/10.1109/TNN.2005.845141.

[96] Saxena Amit, Prasad Mukesh, Gupta Akshansh, Bharill Neha, Patel Om Prakash, Tiwari Aruna, et al. A review of clustering techniques and developments. Neurocomputing 2017;267:664–81. http://dx.doi.org/10.1016/j.neucom.2017.06.053.

[97] Ravi Nikhil, Scaglione Anna, Kadam Sachin, Gentz Reinhard, Peisert Sean, Lunghino Brent, et al. Differentially private K-means clustering applied to meter data analysis and synthesis. IEEE Trans Smart Grid 2022;13(6):4801–14. http://dx.doi.org/10.1109/TSG.2022.3184252.

[98] Grabisch Michel, Marichal Jean-Luc, Mesiar Radko, Pap Endre. Aggregation functions. Cambridge University Press; 2009.

[99] Hoffmann Maximilian, Kotzur Leander, Stolten Detlef, Robinius Martin. A review on time series aggregation methods for energy system models. Energies 2020;13(3):641. http://dx.doi.org/10.3390/en13030641.

[100] Stephenson Todd Andrew. An introduction to Bayesian network theory and usage. IDIAP research report, 2000.

[101] Heckerman David. A tutorial on learning with Bayesian networks. In: Jordan Michael I, editor. Learning in graphical models. Dordrecht: Springer Science & Business Media; 1998, p. 301–54. http://dx.doi.org/10.1007/978-94-011-5014-9_11.

[102] Adedipe Tosin, Shafiee Mahmood, Zio Enrico. Bayesian network modelling for the wind energy industry: An overview. Reliab Eng Syst Saf 2020;202:107053. http://dx.doi.org/10.1016/j.ress.2020.107053.

[103] Mumuni Alhassan, Mumuni Fuseini. Data augmentation: A comprehensive survey of modern approaches. Array 2022;16:100258. http://dx.doi.org/10.1016/j.array.2022.100258.

[104] Wen Qingsong, Sun Liang, Yang Fan, Song Xiaomin, Gao Jingkun, Wang Xue, et al. Time series data augmentation for deep learning: A survey. In: Proceedings of the thirtieth international joint conference on artificial intelligence. 2021, p. 4653–60. http://dx.doi.org/10.24963/ijcai.2021/631.

[105] Iglesias Guillermo, Talavera Edgar, González-Prieto Ángel, Mozo Alberto, Gómez-Canaval Sandra. Data augmentation techniques in time series domain: a survey and taxonomy. Neural Comput Appl 2023;35:10123–45. http://dx.doi.org/10.1007/s00521-023-08459-3.

[106] Ardakanian Omid, Keshav Srinivasan, Rosenberg Catherine. Markovian models for home electricity consumption. In: Proceedings of the 2nd ACM SIGCOMM workshop on green networking. ACM; 2011, p. 31–6. http://dx.doi.org/10.1145/2018536.2018544.

[107] Bottaccioli Lorenzo, Di Cataldo Santa, Acquaviva Andrea, Patti Edoardo. Realistic multi-scale modeling of household electricity behaviors. IEEE Access 2019;7:2467–89. http://dx.doi.org/10.1109/ACCESS.2018.2886201.

[108] Bouvenot Jean-Baptiste, Latour Benjamin, Flament Bernard, Siroux Monica. High resolution stochastic generator of European household specific electricity demand load curves for decentralized power self-production applications. Energy Build 2020;229:110480. http://dx.doi.org/10.1016/j.enbuild.2020.110480.

[109] Collin Adam J, Tsagarakis George, Kiprakis Aristides E, McLaughlin Stephen. Development of low-voltage load models for the residential load sector. IEEE Trans Power Syst 2014;29(5):2180–8. http://dx.doi.org/10.1109/TPWRS.2014.2301949.

[110] Huang Yunyou, Zhan Jianfeng, Luo Chunjie, Wang Lei, Wang Nana, Zheng Daoyi, et al. An electricity consumption model for synthesizing scalable electricity load curves. Energy 2019;169:674–83. http://dx.doi.org/10.1016/j.energy.2018.12.050.

[111] Kadavil Rahul, Hansen Timothy M, Suryanarayanan Siddharth. An algorithmic approach for creating diverse stochastic feeder datasets for power systems co-simulations. In: 2016 IEEE power and energy society general meeting. IEEE; 2016, http://dx.doi.org/10.1109/PESGM.2016.7741708.

[112] McLoughlin Fintan J, Duffy Aidan, Conlon Michael. The generation of domestic electricity load profiles through Markov chain modelling. Euro-Asian J Sustain Energy Dev Policy 2010;3.

[113] Nijhuis M, Gibescu M, Cobben JFG. Bottom-up Markov chain Monte Carlo approach for scenario based residential load modelling with publicly available data. Energy Build 2016;112:121–9. http://dx.doi.org/10.1016/j.enbuild.2015.12.004.

[114] Palacios-Garcia EJ, Moreno-Munoz A, Santiago I, Flores-Arias JM, Bellido-Outeirino FJ, Moreno-Garcia IM. A stochastic modelling and simulation approach to heating and cooling electricity consumption in the residential sector. Energy 2018;144:1080–91. http://dx.doi.org/10.1016/j.energy.2017.12.082.
[115] Richardson Ian, Hodgson Graeme, Thomson Murray, Infield David, Delahunty Alice. Simulation of high-resolution domestic electricity demand based on a building occupancy model and its applicability to the study of demand side management. In: Bertoldi Paolo, Werle Rita, editors. Energy efficiency in domestic appliances and lighting: Proceedings of the 5th international conference EEDAL'09. Publications Office of the European Union; 2010, p. 97–105. http://dx.doi.org/10.2788/54535.
[116] Richardson Ian, Thomson Murray, Infield David, Clifford Conor. Domestic electricity use: A high-resolution energy demand model. Energy Build 2010;42(10):1878–87. http://dx.doi.org/10.1016/j.enbuild.2010.05.023.
[117] Rouleau Jean, Ramallo-González Alfonso P, Gosselin Louis, Blanchet Pierre, Natarajan Sukumar. A unified probabilistic model for predicting occupancy, domestic hot water use and electricity use in residential buildings. Energy Build 2019;202:109375. http://dx.doi.org/10.1016/j.enbuild.2019.109375.
[118] Sansregret Simon, Lavigne Karine, Le Lostec Brice, Laurencelle François, Guay Frédéric. High-resolution bottom-up residential electrical model for distribution networks planning. In: Corrado V, Fabrizio E, Gasparella A, Patuzzi F, editors. Proceedings of building simulation 2019: 16th conference of IBPSA, vol. 5. IBPSA; 2019, p. 3540–7. http://dx.doi.org/10.26868/25222708.2019.210716.
[119] Widén Joakim, Lundh Magdalena, Vassileva Iana, Dahlquist Erik, Ellegård Kajsa, Wäckelgård Ewa. Constructing load profiles for household electricity and hot water from time-use data-Modelling approach and validation. Energy Build 2009;41(7):753–68. http://dx.doi.org/10.1016/j.enbuild.2009.02.013.
[120] Widén Joakim, Nilsson Annica M, Wäckelgård Ewa. A combined Markov-chain and bottom-up approach to modelling of domestic lighting demand. Energy Build 2009;41(10):1001–12. http://dx.doi.org/10.1016/j.enbuild.2009.05.002.
[121] Widén Joakim, Wäckelgård Ewa. A high-resolution stochastic model of domestic activity patterns and electricity demand. Appl Energy 2010;87(6):1880–92. http://dx.doi.org/10.1016/j.apenergy.2009.11.006.
[122] Wills AD, Beausoleil-Morrison I, Ugursal VI. Adaptation and validation of an existing bottom-up model for simulating temporal and inter-dwelling variations of residential appliance and lighting demands. J Build Perform Simul 2018;11(3):350–68. http://dx.doi.org/10.1080/19401493.2017.1369570.
[123] Xia Yun, Hu Bo, Xie Kaigui, Tang Junjie, Tai Heng-Ming. An EV charging demand model for the distribution system using traffic property. IEEE Access 2019;7:28089–99. http://dx.doi.org/10.1109/ACCESS.2019.2901857.
[124] Dalla Maria Enrico, Secchi Mattia, Macii David. A flexible top-down data-driven stochastic model for synthetic load profiles generation. Energies 2022;15(1):269. http://dx.doi.org/10.3390/en15010269.
[125] Labeeuw Wouter, Deconinck Geert. Residential electrical load model based on mixture model clustering and Markov models. IEEE Trans Ind Inf 2013;9(3):1561–9. http://dx.doi.org/10.1109/TII.2013.2240309.
[126] Uhrig Martin, Mueller Robin, Leibfried Thomas. Statistical consumer modelling based on smart meter measurement data. In: 2014 international conference on probabilistic methods applied to power systems. IEEE; 2014, http://dx.doi.org/10.1109/PMAPS.2014.6960656.
[127] Zufferey Thierry, Toffanin Damiano, Toprak Diren, Ulbig Andreas, Hug Gabriela. Generating stochastic residential load profiles from smart meter data for an optimal power matching at an aggregate level. In: 2018 power systems computation conference. Power systems computation conference, 2018, http://dx.doi.org/10.23919/PSCC.2018.8442470.
[128] Casella IRS, Sanches BCS, Filho AJ Sguarezi, Capovilla CE. A dynamic residential load model based on a non-homogeneous Poisson process. J Control Autom Electr Syst 2016;27:670–9. http://dx.doi.org/10.1007/s40313-016-0269-8.
[129] Denaxas Evangelos A, Bandyopadhyay Rubenka, Patiño-Echeverri Dalia, Pitsianis Nikos. SynTiSe: A modified multi-regime MCMC approach for generation of wind power synthetic time series. In: 2015 annual IEEE systems conference. IEEE; 2015, p. 668–74. http://dx.doi.org/10.1109/SYSCON.2015.7116827.
[130] Groß Daniel, Wiest Pascal, Rudion Krzysztof. Comparison of stochastic load profile modeling approaches for low voltage residential consumers. In: 2017 IEEE manchester powerTech. IEEE; 2017, http://dx.doi.org/10.1109/PTC.2017.7981272.
[131] Groß Daniel, Wiest Pascal, Rudion Krzysztof, Probst Alexander. Parametrization of stochastic load profile modeling approaches for smart grid simulations. In: 2017 IEEE PES innovative smart grid technologies conference Europe. IEEE; 2017, http://dx.doi.org/10.1109/ISGTEurope.2017.8260180.
[132] Liu Xiufeng, Iftikhar Nadeem, Huo Huan, Li Rongling, Nielsen Per Sieverts. Two approaches for synthesising scalable residential energy consumption data. Future Gener Comput Syst 2019;95:586–600. http://dx.doi.org/10.1016/j.future.2019.01.045.
[133] Machado João AC, Carvalho Pedro MS, Ferreira Luís AFM. Building stochastic non-stationary daily load/generation profiles for distribution planning studies. IEEE Trans Power Syst 2018;33(1):911–20. http://dx.doi.org/10.1109/tpwrs.2017.2687624.
[134] Meiser Michael, Duppe Benjamin, Zinnikus Ingo. SynTiSeD – Synthetic time series data generator. In: 2023 11th workshop on modelling and simulation of cyber-physical energy systems. IEEE; 2023, http://dx.doi.org/10.1109/MSCPES58582.2023.10123429.
[135] Papaefthymiou George, Klöckl Bernd. MCMC for wind power simulation. IEEE Trans Energy Convers 2008;23(1):234–40. http://dx.doi.org/10.1109/TEC.2007.914174.
[136] Pesch T, Schröders S, Allelein HJ, Hake JF. A new Markov-chain-related statistical approach for modelling synthetic wind power time series. New J Phys 2015;17(5):055001. http://dx.doi.org/10.1088/1367-2630/17/5/055001.
[137] Pyne Moinak, Yurkovich Benjamin J, Yurkovich Stephen. Generation of synthetic battery data with capacity variation. In: 2019 IEEE conference on control technology and applications. IEEE; 2019, p. 476–80. http://dx.doi.org/10.1109/CCTA.2019.8920488.
[138] Roje Tomislav, Marín Luis G, Sáez Doris, Orchard Marcos, Jiménez-Estévez Guillermo. Consumption modeling based on Markov chains and Bayesian networks for a demand side management design of isolated microgrids. Int J Energy Res 2017;41(3):365–76. http://dx.doi.org/10.1002/er.3607.
[139] Weber Juliane, Zachow Christopher, Witthaut Dirk. Modeling long correlation times using additive binary Markov chains: Applications to wind generation time series. Phys Rev E 2018;97(3):032138. http://dx.doi.org/10.1103/PhysRevE.97.032138.
[140] Wu Tong, Ai Xiaomeng, Lin Weixing, Wen Jinyu, Luo Weihua. Markov chain Monte Carlo method for the modeling of wind power time series. In: 2012 IEEE innovative smart grid technologies Asia. IEEE; 2012, http://dx.doi.org/10.1109/ISGT-Asia.2012.6303304.
[141] Capasso A, Grattieri W, Lamedica R, Prudenzi A. A bottom-up approach to residential load modeling. IEEE Trans Power Syst 1994;9(2):957–64. http://dx.doi.org/10.1109/59.317650.
[142] Oliveira Panão Marta JN, Brito Miguel C. Modelling aggregate hourly electricity consumption based on bottom-up building stock. Energy Build 2018;170:170–82. http://dx.doi.org/10.1016/j.enbuild.2018.04.010.
[143] Raisz David, Dán András. A stochastic residential load model and its application to the unified loss theory. Eur Trans Electr Power 2009;19(8):1118–30. http://dx.doi.org/10.1002/etep.288.
[144] Ward Rebecca, Choudhary Ruchi, Heo Yeonsook, Guillas Serge. Data driven bottom-up approach for modelling internal loads in building energy simulation using functional principal components. In: Proceedings of the 3rd IBPSA-England conference building simulation and optimization. 2016.
[145] Dong Ming, Xie Kaigui, Li Wenyuan. Generating long-term continuous multi-type generation profiles using generative adversarial network. 2020.
[146] Kimbrough Steven O, Umitcan Yilmaz Hasan. A symmetric block resampling method to generate energy time series data. In: 2021 22nd IEEE international conference on industrial technology. IEEE; 2021, p. 546–51. http://dx.doi.org/10.1109/ICIT46573.2021.9453485.
[147] Lojowska Alicja, Kurowicka Dorota, Papaefthymiou Georgios, van der Sluis Lou. Stochastic modeling of power demand due to EVs using copula. IEEE Trans Power Syst 2012;27(4):1960–8. http://dx.doi.org/10.1109/TPWRS.2012.2192139.
[148] Ma Xi-Yuan, Sun Yuan-Zhang, Fang Hua-Liang. Scenario generation ofwind power based on statistical uncertainty and variability. IEEE Trans Sustain Energy 2013;4(4):894–904. http://dx.doi.org/10.1109/TSTE.2013.2256807.
[149] Usaola Julio. Synthesis of hourly wind power series using the moving block bootstrap method. In: 2014 international conference on probabilistic methods applied to power systems. IEEE; 2014, http://dx.doi.org/10.1109/PMAPS.2014.6960602.
[150] Zhou Niancheng, Xiong Xicong, Wang Qianggang. Probability model and simulation method of electric vehicle charging load on distribution network. Electr Power Compon Syst 2014;42(9):879–88. http://dx.doi.org/10.1080/15325008.2014.903537.
[151] Armstrong Marianne M, Swinton Mike C, Ribberink Hajo, Beausoleil-Morrison Ian, Millette Jocelyn. Synthetically derived profiles for representing occupant-driven electric loads in Canadian housing. J Build Perform Simul 2009;2(1):15–30. http://dx.doi.org/10.1080/19401490802706653.
[152] Asare-Bediako B, Kling WL, Ribeiro PF. Future residential load profiles: Scenario-based analysis of high penetration of heavy loads and distributed generation. Energy Build 2014;75:228–38. http://dx.doi.org/10.1016/j.enbuild.2014.02.025.
[153] Bucher Christof, Andersson Göran. Generation of domestic load profiles - an adaptive top-down approach. In: 12th international conference on probabilistic methods applied to power systems. 2012, p. 436–41.
[154] Chen CS, Kang MS, Hwang JC, Huang CW. Synthesis of power system load profiles by class load study. Int J Electr Power Energy Syst 2000;22(5):325–30. http://dx.doi.org/10.1016/S0142-0615(99)00046-0.
[155] Chuan Luo, Ukil Abhisek. Modeling and validation of electrical load profiling in residential buildings in Singapore. IEEE Trans Power Syst 2015;30(5):2800–9. http://dx.doi.org/10.1109/TPWRS.2014.2367509.
[156] Clemente Shelly, Beauchêne Solène, Nefzaoui Elyes. Generation of aggregated plug load profiles in office buildings. Energy Build 2021;252:111398. http://dx.doi.org/10.1016/j.enbuild.2021.111398.

[157] Dickert J, Schegner P. Curve model for residential customers. In: 2011 IEEE trondheim PowerTech. IEEE; 2011, http://dx.doi.org/10.1109/PTC.2011.6019365.

[158] Fischer David, Härtl Andreas, Wille-Haussmann Bernhard. Model for electric load profiles with high time resolution for german households. Energy Build 2015;92:170–9. http://dx.doi.org/10.1016/j.enbuild.2015.01.058.

[159] Fischer David, Surmann Arne, Biener Wolfgang, Selinger-Lutz Oliver. From residential electric load profiles to flexibility profiles – A stochastic bottom-up approach. Energy Build 2020;224:110133. http://dx.doi.org/10.1016/j.enbuild.2020.110133.

[160] Gruber Jorn K, Prodanovic Milan. Residential energy load profile generation using a probabilistic approach. In: 2012 sixth UKSim/AMSS European symposium on computer modeling and simulation. IEEE; 2012, p. 317–22. http://dx.doi.org/10.1109/EMS.2012.30.

[161] Gruber JK, Jahromizadeh S, Prodanović M, Rakočević V. Application-oriented modelling of domestic energy demand. Int J Electr Power Energy Syst 2014;61:656–64. http://dx.doi.org/10.1016/j.ijepes.2014.04.008.

[162] Hartvigsson Elias K, Ehnberg Jimmy, Ahlgren Erik O, Molander Sverker. Using system dynamics for long term bottom-up electric load modeling in rural electrification. In: 34th international conference of the system dynamics society 2016. Curran Associates; 2016, p. 1060–72.

[163] Hoogsteen Gerwin, Molderink Albert, Hurink Johann L, Smit Gerard JM. Generation of flexible domestic load profiles to evaluate demand side management approaches. In: 2016 IEEE international energy conference. IEEE; 2016, http://dx.doi.org/10.1109/ENERGYCON.2016.7513873.

[164] Hosseini S, Kelouwani S, Agbossou K, Cardenas A, Henao N. A semi-synthetic dataset development tool for household energy consumption analysis. In: 2017 IEEE international conference on industrial technology. IEEE; 2017, p. 564–9. http://dx.doi.org/10.1109/ICIT.2017.7915420.

[165] Jeyaranjani J, Devaraj D. Load profile generation for DR program. In: 2019 IEEE international conference on intelligent techniques in control, optimization and signal processing. IEEE; 2019, http://dx.doi.org/10.1109/INCOS45849.2019.8951357.

[166] Kewo Angreine, Manembu Pinrolinvic DK, Nielsen Per Sieverts. Synthesising residential electricity load profiles at the city level using a weighted proportion (wepro) model. Energies 2020;13(14):3543. http://dx.doi.org/10.3390/en13143543.

[167] Klemenjak Christoph, Kovatsch Christoph, Herold Manuel, Elmenreich Wilfried. A synthetic energy dataset for non-intrusive load monitoring in households. Sci Data 2020;7(1):108. http://dx.doi.org/10.1038/s41597-020-0434-6.

[168] Kong Weicong, Dong ZY, Chen Guo, Jia Youwei. A rule based domestic load profile generator for future smart grid. In: 2014 Australasian universities power engineering conference. IEEE; 2014, http://dx.doi.org/10.1109/AUPEC.2014.6966579.

[169] Lan Yu, Wu Jiang, Tang Zhe. Generation of domestic load profiles using appliances' activating moments. In: 2014 IEEE PES innovative smart grid technologies conference. IEEE; 2014, http://dx.doi.org/10.1109/ISGT.2014.6816467.

[170] Litzlbauer Markus. Generation of stochastic load profiles for mobile energy storages. World Electr Veh J 2010;4(4):705–9. http://dx.doi.org/10.3390/wevj4040705.

[171] Marszal-Pomianowska Anna, Heiselberg Per, Kalyanova Larsen Olena. Household electricity demand profiles - A high-resolution load model to facilitate modelling of energy flexible buildings. Energy 2016;103:487–501. http://dx.doi.org/10.1016/j.energy.2016.02.159.

[172] McKenna R, Hofmann L, Kleinebrahm M, Fichtner W. A stochastic multi-energy simulation model for UK residential buildings. Energy Build 2018;168:470–89. http://dx.doi.org/10.1016/j.enbuild.2018.02.051.

[173] Ortiz Joana, Guarino Francesco, Salom Jaume, Corchero Cristina, Cellura Maurizio. Stochastic model for electrical loads in mediterranean residential buildings: Validation and applications. Energy Build 2014;80:23–36. http://dx.doi.org/10.1016/j.enbuild.2014.04.053.

[174] Paatero Jukka V, Lund Peter D. A model for generating household electricity load profiles. Int J Energy Res 2006;30(5):273–90. http://dx.doi.org/10.1002/er.1136.

[175] Pflugradt Noah, Muntwyler Urs. Synthesizing residential load profiles using behavior simulation. Energy Procedia 2017;122:655–60. http://dx.doi.org/10.1016/j.egypro.2017.07.365.

[176] Pinceti Andrea, Kosut Oliver, Sankar Lalitha. Data-driven generation of synthetic load datasets preserving spatio-temporal features. In: 2019 IEEE power & energy society general meeting. 2019, http://dx.doi.org/10.1109/PESGM40551.2019.8973532.

[177] Richardson Ian, Thomson Murray, Infield David, Delahunty Alice. Domestic lighting: A high-resolution energy demand model. Energy Build 2009;41(7):781–9. http://dx.doi.org/10.1016/j.enbuild.2009.02.010.

[178] Sandels C, Brodén D, Widén J, Nordström L, Andersson E. Modeling office building consumer load with a combined physical and behavioral approach: Simulation and validation. Appl Energy 2016;162:472–85. http://dx.doi.org/10.1016/j.apenergy.2015.10.141.

[179] Schäuble Johannes, Kaschub Thomas, Ensslen Axel, Jochem Patrick, Fichtner Wolf. Generating electric vehicle load profiles from empirical data of three EV fleets in Southwest Germany. J Clean Prod 2017;150:253–66. http://dx.doi.org/10.1016/j.jclepro.2017.02.150.

[180] Walker CF, Pokoski JL. Residential load shape modeling based on customer behavior. IEEE Trans Power Appar Syst 1985;PAS-104(7):1703–11. http://dx.doi.org/10.1109/TPAS.1985.319202.

[181] Ward RM, Choudhary R, Heo Y, Aston JAD. A data-centric bottom-up model for generation of stochastic internal load profiles based on space-use type. J Build Perform Simul 2019;12(5):620–36. http://dx.doi.org/10.1080/19401493.2019.1583287.

[182] Yao Runming, Steemers Koen. A method of formulating energy load profile for domestic buildings in the UK. Energy Build 2005;37(6):663–71. http://dx.doi.org/10.1016/j.enbuild.2004.09.007.

[183] Duque Edgar Mauricio Salazar, Vergara Pedro P, Nguyen Phuong H, van der Molen Anne, Slootweg JG. Conditional multivariate elliptical copulas to model residential load profiles from smart meter data. IEEE Trans Smart Grid 2021;12(5):4280–94. http://dx.doi.org/10.1109/TSG.2021.3078394.

[184] Bouderraoui Houda, Chami Mouhcine. SGSim: Load profile generator for smart grid applications. In: 2018 renewable energies, power systems & green inclusive economy. IEEE; 2018, http://dx.doi.org/10.1109/REPSGIE.2018.8488843.

[185] Carreño Ignacio Losada, Ramakrishna Raksha, Scaglione Anna, Arnold Daniel, Roberts Ciaran, Ngo Sy-Toan, et al. Soda: An irradiance-based synthetic solar data generation tool. In: 2020 IEEE international conference on communications, control, and computing technologies for smart grids. IEEE; 2020, http://dx.doi.org/10.1109/SmartGridComm47815.2020.9302941.

[186] Henriet Simon, Simsekli Umut, Richard Gaël, Fuentes Benoit. Synthetic dataset generation for non-intrusive load monitoring in commercial buildings. In: Proceedings of the 4th ACM international conference on systems for energy-efficient built environments. ACM; 2017, http://dx.doi.org/10.1145/3137133.3141458.

[187] Pinceti Andrea, Sankar Lalitha, Kosut Oliver. Generation of synthetic multi-resolution time series load data. IET Smart Grid 2023;6(5):492–502. http://dx.doi.org/10.1049/stg2.12116.

[188] Uimonen Semen, Lehtonen Matti. Simulation of electric vehicle charging stations load profiles in office buildings based on occupancy data. Energies 2020;13(21):5700. http://dx.doi.org/10.3390/en13215700.

[189] Sun Sun S, Kazhamiaka Fiodar, Keshav Srinivasan, Rosenberg Catherine. Using synthetic traces for robust energy system sizing. In: Proceedings of the 10th ACM international conference on future energy systems. ACM; 2019, p. 251–62. http://dx.doi.org/10.1145/3307772.3328306.

[190] Caquilpan Victor, Sáez Doris, Hernández Roberto, Llanos Jacqueline, Roje Tomislav, Nuñez Alfredo. Load estimation based on self-organizing maps and Bayesian networks for microgrids design in rural zones. In: 2017 IEEE PES innovative smart grid technologies conference Latin America. IEEE; 2017, http://dx.doi.org/10.1109/ISGT-LA.2017.8126709.

[191] Xu Fang Yuan, Wang Xue, Lai Loi Lei, Lai Chun Sing. Agent-based modeling and neural network for residential customer demand response. In: 2013 IEEE international conference on systems, man, and cybernetics. IEEE; 2013, p. 1312–6. http://dx.doi.org/10.1109/SMC.2013.227.

[192] Channegowda Janamejaya, Lachappa Yugendra Gowda, Jagwani Shefali. A graph-based lithium-ion battery parameter estimation approach to produce diverse synthetic data. Adv Theory Simul 2022;5(8):2200128. http://dx.doi.org/10.1002/adts.202200128.

[193] Förderer Kevin, Schmeck Hartmut. State-based load profile generation for modeling energetic flexibility. Energy Inform 2019;2(1). http://dx.doi.org/10.1186/s42162-019-0077-z.

[194] Ge Leijiao, Liao Wenlong, Wang Shouxiang, Bak-Jensen Birgitte, Pillai Jayakrishnan Radhakrishna. Modeling daily load profiles of distribution network for scenario generation using flow-based generative network. IEEE Access 2020;8:77587–97. http://dx.doi.org/10.1109/ACCESS.2020.2989350.

[195] Gowda Hruday, Channegowda Janamejaya. Contrastive learning for practical battery synthetic data generation using seasonal and trend representations. Int J Energy Res 2022;46(15):24602–10. http://dx.doi.org/10.1002/er.8651.

[196] Liao Wenlong, Wang Yusen, Wang Yuelong, Powell Kody, Liu Qi, Yang Zhe. Scenario generation for cooling, heating, and power loads using generative moment matching networks. CSEE J Power Energy Syst 2022;8(6):1730–40. http://dx.doi.org/10.17775/CSEEJPES.2021.00680.

[197] Naaz Falak, Channegowda Janamejaya. A probabilistic forecasting approach towards generation of synthetic battery parameters to resolve limited data challenges. Energy Storage 2022;4(4):e297. http://dx.doi.org/10.1002/est2.297.

[198] Pillai Gobind G, Putrus Ghanim A, Pearsall Nicola M. Generation of synthetic benchmark electrical load profiles using publicly available load and weather data. Int J Electr Power Energy Syst 2014;61:1–10. http://dx.doi.org/10.1016/j.ijepes.2014.03.005.

[199] Räsänen Teemu, Voukantsis Dimitrios, Niska Harri, Karatzas Kostas, Kolehmainen Mikko. Data-based method for creating electricity use load profiles using large amount of customer-specific hourly measured electricity use data. Appl Energy 2010;87(11):3538–45. http://dx.doi.org/10.1016/j.apenergy.2010.05.015.

[200] Li Jianbin, Chen Zhiqiang, Cheng Long, Liu Xiufeng. Energy data generation with wasserstein deep convolutional generative adversarial networks. Energy 2022;257:124694. http://dx.doi.org/10.1016/j.energy.2022.124694.

[201] Zhang Chi, Kuppannagari Sanmukh R, Kannan Rajgopal, Prasanna Viktor K. Generative adversarial network for synthetic time series data generation in smart grids. In: 2018 IEEE international conference on communications, control, and computing technologies for smart grids. IEEE; 2018, http://dx.doi.org/10.1109/SmartGridComm.2018.8587464.

[202] Asre Shashank, Anwar Adnan. Synthetic energy data generation using time variant generative adversarial network. Electronics 2022;11(3):355. http://dx.doi.org/10.3390/electronics11030355.

[203] Chen Yize, Li Pan, Zhang Baosen. Bayesian renewables scenario generation via deep generative networks. In: 2018 52nd annual conference on information sciences and systems. IEEE; 2018, http://dx.doi.org/10.1109/CISS.2018.8362314.

[204] Chen Yize, Wang Yishen, Kirschen Daniel, Zhang Baosen. Model-free renewable scenario generation using generative adversarial networks. IEEE Trans Power Syst 2018;33(3):3265–75. http://dx.doi.org/10.1109/TPWRS.2018.2794541.

[205] Dong Wei, Chen Xianqing, Yang Qiang. Data-driven scenario generation of renewable energy production based on controllable generative adversarial networks with interpretability. Appl Energy 2022;308:118387. http://dx.doi.org/10.1016/j.apenergy.2021.118387.

[206] Fekri Mohammad Navid, Ghosh Ananda Mohon, Grolinger Katarina. Generating energy data for machine learning with recurrent generative adversarial networks. Energies 2020;13(1):130. http://dx.doi.org/10.3390/en13010130.

[207] Hazra Debapriya, Shafqat Wafa, Byun Yung-Cheol. Generating synthetic data to reduce prediction error of energy consumption. Comput Mater Continua 2022;70(2):3151–67. http://dx.doi.org/10.32604/cmc.2022.020143.

[208] Hu Yi, Li Yiyan, Song Lidong, Lee Han Pyo, Rehm PJ, Makdad Matthew, et al. MultiLoad-GAN: A GAN-based synthetic load group generation method considering spatial-temporal correlations. IEEE Trans Smart Grid 2023. http://dx.doi.org/10.1109/TSG.2023.3302192.

[209] Kababji Samer El, Srikantha Pirathayini. A data-driven approach for generating synthetic load patterns and usage habits. IEEE Trans Smart Grid 2020;11(6):4984–95. http://dx.doi.org/10.1109/TSG.2020.3007984.

[210] Lan Jian, Guo Qinglai, Sun Hongbin. Demand side data generating based on conditional generative. Energy Procedia 2018;152:1188–93. http://dx.doi.org/10.1016/j.egypro.2018.09.157.

[211] Lan Jian, Guo Qinglai, Zhou Yanzhen, Sun Hongbin. Generation of large-scale convergent power flow samples through a data-driven approach. In: 2020 IEEE 4th conference on energy internet and energy system integration. IEEE; 2020, p. 722–6. http://dx.doi.org/10.1109/EI250167.2020.9346606.

[212] Liang Xinyu, Wang Hao. Synthesis of realistic load data: Adversarial networks for learning and generating residential load patterns. In: NeurIPS 2022 workshop on tackling climate change with machine learning. 2022.

[213] Naaz Falak, Herle Aniruddh, Channegowda Janamejaya, Raj Aditya, Lakshminarayanan Meenakshi. A generative adversarial network-based synthetic data augmentation technique for battery condition evaluation. Int J Energy Res 2021;45(13):19120–35. http://dx.doi.org/10.1002/er.7013.

[214] Pinceti Andrea, Sankar Lalitha, Kosut Oliver. Synthetic time-series load data via conditional generative adversarial networks. In: 2021 IEEE power & energy society general meeting. IEEE; 2021, http://dx.doi.org/10.1109/PESGM46819.2021.9637821.

[215] Razghandi Mina, Zhou Hao, Erol-Kantarci Melike, Turgut Damla. Smart home energy management: VAE-GAN synthetic dataset generator and Q-learning. 2023.

[216] Tang Rui, Leong Philip HW, Dore Jonathon, Vassallo Anthony. Generating residential PV production and electricity consumption scenarios via generative adversarial networks. In: Asia Pacific solar research conference 2018. 2018.

[217] Wang Zhe, Hong Tianzhen. Generating realistic building electrical load profiles through the Generative Adversarial Network (GAN). Energy Build 2020;224:110299. http://dx.doi.org/10.1016/j.enbuild.2020.110299.

[218] Yilmaz Bilgi, Korn Ralf. Synthetic demand data generation for individual electricity consumers: Generative adversarial networks (GANs). Energy AI 2022;9:100161. http://dx.doi.org/10.1016/j.egyai.2022.100161.

[219] Yin Linfei, Zhang Bin. Time series generative adversarial network controller for long-term smart generation control of microgrids. Appl Energy 2021;281:116069. http://dx.doi.org/10.1016/j.apenergy.2020.116069.

[220] Zhang Yuhang, Deng Xiangtian, Zhang Yi, Zhany Yi. Generation of sub-item load profiles for public buildings based on the conditional generative adversarial network and moving average method. Energy Build 2022;268:112185. http://dx.doi.org/10.1016/j.enbuild.2022.112185.

[221] Zheng Xiangtian, Wang Bin, Xie Le. Synthetic dynamic PMU data generation: A generative adversarial network approach. In: 2019 international conference on smart grid synchronized measurements and analytics. 2019, http://dx.doi.org/10.1109/SGSMA.2019.8784681.

[222] Zheng Xiangtian, Pinceti Andrea, Sankar Lalitha, Xie Le. Synthetic PMU data creation based on generative adversarial network under time-varying load conditions. J Mod Power Syst Clean Energy 2023;11(1):234–42. http://dx.doi.org/10.35833/MPCE.2021.000783.

[223] Dos Santos Ricardo, Aguilar Jose, R-Moreno Maria D. A synthetic data generator for smart grids based on the variational-autoencoder technique and linked data paradigm. In: 2022 XVLIII Latin American computer conference. IEEE; 2022, http://dx.doi.org/10.1109/CLEI56649.2022.9959918.

[224] Gong Xuejiao, Tang Bo, Zhu Ruijin, Liao Wenlong, Song Like. Data augmentation for electricity theft detection using conditional variational auto-encoder. Energies 2020;13(17):4291. http://dx.doi.org/10.3390/en13174291.

[225] Pan Zhixin, Wang Jianming, Liao Wenlong, Chen Haiwen, Yuan Dong, Zhu Weiping, et al. Data-driven EV load profiles generation using a variational auto-encoder. Energies 2019;12(5):849. http://dx.doi.org/10.3390/en12050849.

[226] Lindberg KB, Bakker SJ, Sartori I. Modelling electric and heat load profiles of non-residential buildings for use in long-term aggregate load forecasts. Util Policy 2019;58:63–88. http://dx.doi.org/10.1016/j.jup.2019.03.004.

[227] Hirst Eric, Trimble John, Goeltz Richard, Cardell N Scott. Use of synthetic data in dealing with self-selection: Improving conservation program energy savings estimates. Eval Rev 1983;7(6):807–30. http://dx.doi.org/10.1177/0193841X8300700606.

[228] Ranjan Manish, Jain VK. Modelling of electrical energy consumption in Delhi. Energy 1999;24(4):351–61. http://dx.doi.org/10.1016/S0360-5442(98)00087-5.

[229] Reddy TA, Claridge DE. Using synthetic data to evaluate multiple regression and principal component analyses for statistical modeling of daily building energy consumption. Energy Build 1994;21(1):35–44. http://dx.doi.org/10.1016/0378-7788(94)90014-0.

[230] Wiest Pascal, Contreras Daniel, Groß Daniel, Rudion Krzysztof. Synthetic load profiles of various customer types for smart grid simulations. In: Conference on sustainable energy supply and energy storage systems. VDE; 2018, p. 273–8.

[231] Chen Jun, Kim Jong Suk, Rabiti Cristian. Probabilistic analysis of hybrid energy systems using synthetic renewable and load data. In: 2017 American control conference. IEEE; 2017, p. 4723–8. http://dx.doi.org/10.23919/ACC.2017.7963685.

[232] Iftikhar Nadeem, Liu Xiufeng, Danalachi Sergiu, Nordbjerg Finn Ebertsen, Vollesen Jens Henrik. A scalable smart meter data generator using spark. In: Panetto Hervé, Debruyne Christophe, Gaaloul Walid, Papazoglou Mike, Paschke Adrian, Ardagna Claudio Agostino, Meersman Robert, editors. On the move to meaningful internet systems. OTM 2017 conferences. Springer; 2017, p. 21–36. http://dx.doi.org/10.1007/978-3-319-69462-7_2.

[233] Lee Duehee, Baldick Ross. Load and wind power scenario generation through the generalized dynamic factor model. IEEE Trans Power Syst 2017;32(1):400–10. http://dx.doi.org/10.1109/TPWRS.2016.2562718.

[234] Louie Henry M. Time-series modeling of aggregated electric vehicle charging station load. Electr Power Compon Syst 2017;45(14):1498–511. http://dx.doi.org/10.1080/15325008.2017.1336583.

[235] Magnano L, Boland JW. Generation of synthetic sequences of electricity demand: Application in south Australia. Energy 2007;32(11):2230–43. http://dx.doi.org/10.1016/j.energy.2007.04.001.

[236] Patidar Sandhya, Jenkins David P, Simpson Sophie A. Stochastic modelling techniques for generating synthetic energy demand profiles. Int J Energy Stat 2016;4(3):1650014. http://dx.doi.org/10.1142/S2335680416500149.

[237] Soares Lacir J, Medeiros Marcelo C. Modeling and forecasting short-term electricity load: A comparison of methods with an application to Brazilian data. Int J Forecast 2008;24(4):630–44. http://dx.doi.org/10.1016/j.ijforecast.2008.08.003.

[238] Talbot Paul W, Rabiti Cristian, Alfonsi Andrea, Krome Cameron, Kunz M Ross, Epiney Aaron, et al. Correlated synthetic time series generation for energy system simulations using Fourier and ARMA signal processing. Int J Energy Res 2020;44(10):8144–55. http://dx.doi.org/10.1002/er.5115.

[239] Vernetti Dos Santos João Carlos, Kleinkauf Werner. A methodology for modelling synthetic daily sequences of hourly power demand for villages and small towns, based on stochastic processes. Sol Energy 1999;66(6):459–67. http://dx.doi.org/10.1016/S0038-092X(99)00049-3.

[240] Della Giustina Davide, Rinaldi Stefano, Robustelli Stefano, Angioni Andrea. Massive generation of customer load profiles for large scale state estimation deployment: An approach to exploit AMI limited data. Energies 2021;14(5):1277. http://dx.doi.org/10.3390/en14051277.

[241] Xu Ti, Li Hanyue, Birchfield Adam B, Overbye Thomas J. Synthesize phasor measurement unit data using large-scale electric network models. In: 2019 North American power symposium. 2019, http://dx.doi.org/10.1109/NAPS46351.2019.9000196.

[242] Zhu Xiangqi, Mather Barry. Data-driven load diversity and variability modeling for quasi-static time-series simulation on distribution feeders. In: 2019 IEEE power & energy society general meeting. 2019, http://dx.doi.org/10.1109/PESGM40551.2019.8973929.

[243] Zhu Xiangqi, Mather Barry. Data-driven distribution system load modeling for quasi-static time-series simulation. IEEE Trans Smart Grid 2020;11(2):1556–65. http://dx.doi.org/10.1109/TSG.2019.2940084.

[244] Chamana Manohar, Mather Barry. Variability extraction and synthesis via multi-resolution analysis using distribution transformer high-speed power data. In: 2017 19th international conference on intelligent system application to power systems. IEEE; 2017, http://dx.doi.org/10.1109/ISAP.2017.8071389.

[245] Kim Young-Il, Ko Jong-Min, Choi Seung-Hwan. Methods for generating TLPs (typical load profiles) for smart grid-based energy programs. In: 2011 IEEE symposium on computational intelligence applications in smart grid. IEEE; 2011, http://dx.doi.org/10.1109/CIASG.2011.5953331.
[246] Kim Young-Il, Kang Shin-Jae, Ko Jong-Min, Choi Seung-Hwan. A study for clustering method to generate typical load profiles for smart grid. In: 8th international conference on power electronics - ECCE Asia. IEEE; 2011, http://dx.doi.org/10.1109/ICPE.2011.5944675.
[247] Yan Rong, Yuan Yuxuan, Wang Zhaoyu, Geng Guangchao, Jiang Quanyuan. Active distribution system synthesis via unbalanced graph generative adversarial network. IEEE Trans Power Syst 2022. http://dx.doi.org/10.1109/TPWRS.2022.3212029.
[248] Berthou Thomas, Duplessis Bruno, Rivière Philippe, Stabat Pascal, Casetta Damien, Marchio Dominique. Smart-E: A tool for energy demand simulation and optimization at the city scale. In: 14th conference of International Building Performance Simulation Association. 2015, p. 1782–9. http://dx.doi.org/10.26868/25222708.2015.2155.
[249] Li Hanyue, Bornsheuer Ashly L, Xu Ti, Birchfield Adam B, Overbye Thomas J. Load modeling in synthetic electric grids. In: 2018 IEEE Texas power and energy conference. IEEE; 2018, http://dx.doi.org/10.1109/TPEC.2018.8312059.
[250] Li Hanyue, Yeo Ju Hee, Bornsheuer Ashly L, Overbye Thomas J. The creation and validation of load time series for synthetic electric power systems. IEEE Trans Power Syst 2021;36(2):961–9. http://dx.doi.org/10.1109/TPWRS.2020.3018936.
[251] López Juan Miguel Gonzalez, Pouresmaeil Edris, Cañizares Claudio A, Bhattacharya Kankar, Mosaddegh Abolfazl, Solanki Bharatkumar V. Smart residential load simulator for energy management in smart grids. IEEE Trans Ind Electron 2019;66(2):1443–52. http://dx.doi.org/10.1109/TIE.2018.2818666.
[252] Omar Nadhirah, Mohd Ariff Mohd Aifaa, Mohd Shah Ainnur Farah Izzati, Mustaza Muhammad Syafiq Anwar. Estimating synthetic load profile based on student behavior using fuzzy inference system for demand side management application. Turk J Electr Eng Comput Sci 2020;28(6):3193–207. http://dx.doi.org/10.3906/ELK-1912-101.
[253] dos Reis Fernando B, Tonkoski Reinaldo, Hansen Timothy M. Synthetic residential load models for smart city energy management simulations. IET Smart Grid 2020;3(3):352–64. http://dx.doi.org/10.1049/iet-stg.2019.0296.
[254] Zhu Xiangqi, Mather Barry. DWT-based aggregated load modeling and evaluation for quasi-static time-series simulation on distribution feeders. In: 2018 IEEE power & energy society general meeting. IEEE; 2018, http://dx.doi.org/10.1109/PESGM.2018.8585535.
[255] Huamani MM, Orlando AF. Methodology for generating thermal and electric load profiles for designing a cogeneration system. Energy Build 2007;39(9):1003–10. http://dx.doi.org/10.1016/j.enbuild.2006.11.009.
[256] Idehen Ikponmwosa, Jang Wonhyeok, Overbye Thomas J. Large-scale generation and validation of synthetic PMU data. IEEE Trans Smart Grid 2020;11(5):4290–8. http://dx.doi.org/10.1109/TSG.2020.2977349.
[257] Steffen Tom, Wiegel Bela, Babazadeh Davood, Youssfi Amine, Becker Christian, Turau Volker. Generation of realistic smart meter data from prosumers for future energy system scenarios. In: Schulz Detlef, editor. NEIS 2022: Conference on sustainable energy supply and energy storage system. VDE Verlag; 2022, p. 183–8.
[258] Lombardi Francesco, Balderrama Sergio, Quoilin Sylvain, Colombo Emanuela. Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. Energy 2019;177:433–44. http://dx.doi.org/10.1016/j.energy.2019.04.097.
[259] Orlando AF, Málaga MP, Huamani MM. Methodology for generating electric load profiles for sizing an electric energy generation system. Energy Build 2012;52:161–7. http://dx.doi.org/10.1016/j.enbuild.2012.04.014.
[260] Hart George W. Nonintrusive appliance load monitoring. Proc IEEE 1992;80(12):1870–91. http://dx.doi.org/10.1109/5.192069.
[261] Asghar Muhammad Rizwan, Dán György, Miorandi Daniele, Chlamtac Imrich. Smart meter data privacy: A survey. IEEE Commun Surv Tutor 2017;19(4):2820–35. http://dx.doi.org/10.1109/COMST.2017.2720195.
[262] van der Maaten Laurens, Hinton Geoffrey. Visualizing data using t-SNE. J Mach Learn Res 2008;9(86):2579–605.
[263] Hotelling Harold. Analysis of a complex of statistical variables into principal components. J Educ Psychol 1933;24(6):417–41. http://dx.doi.org/10.1037/h0071325.