



**No. 35-2021**

**Erik Heilmann, Janosch Henze and Heike Wetzel**

## **Machine Learning in Energy Forecasts with an Application to High Frequency Electricity Consumption Data**

This paper can be downloaded from:

<https://www.uni-marburg.de/en/fb02/research-groups/economics/macroeconomics/research/magks-joint-discussion-papers-in-economics>

Coordination: Bernd Hayo • Philipps-University Marburg  
School of Business and Economics • Universitätsstraße 24, D-35032 Marburg  
Tel: +49-6421-2823091, Fax: +49-6421-2823088, e-mail: [hayo@wiwi.uni-marburg.de](mailto:hayo@wiwi.uni-marburg.de)

# Machine learning in energy forecasts with an application to high frequency electricity consumption data

Erik Heilmann<sup>a</sup>, Janosch Henze<sup>b</sup>, Heike Wetzel<sup>a</sup>

<sup>a</sup>*Institute of Economics, University of Kassel, Nora-Plattiel-Str. 4, 34109 Kassel, Germany*

<sup>b</sup>*Intelligent Embedded Systems Group, University of Kassel, Wilhelmshöher Allee 73, 34121 Kassel, Germany*

October 11, 2021

---

## Abstract

Forecasting plays an essential role in energy economics. With new challenges and use cases in the energy system, forecasts have to meet more complex requirements, such as increasing temporal and spatial resolution of data. The concept of machine learning can meet these requirements by providing different model approaches and a standardized process of model selection. This paper provides a concise and comprehensible introduction to the topic by discussing the concept of machine learning in the context of energy economics and presenting an exemplary application to electricity load data. For this, we introduce and demonstrate the structured machine learning process containing the preparation, model selection and test of forecast models. This process is intended to serve as a general guideline for energy economists and practitioners who need to apply sophisticated forecast models.

### *Keywords:*

machine learning, electricity consumption forecast, artificial neural network, time series forecast

*JEL classification:* C45, C53, Q47

---

## 1. Introduction

The energy system faces various challenges, most of them related to the overall goal of a reliable, efficient and carbon-free energy supply. For many years, energy forecasts for different time horizons have been an essential part of energy research and industrial tasks. In the past, the energy production of large generating units followed the aggregated demand at the country level. As a result, such highly aggregated energy consumption was often the object of forecasts – in the short term for power plant dispatch (Hahn et al., 2009), and in the long term for energy system design (Carvallo et al., 2018; Suganthi and Samuel, 2012). This situation will evolve in light of the growing share of fluctuating, distributed renewable energy sources and new technologies such as electric vehicles and other power-to-X-technologies. With new challenges arising, such as local energy markets (Mengelkamp et al., 2019) and network operation issues (Hernandez et al., 2014), forecasts have to meet new and more complex requirements.

The field of ‘machine learning’ comprises many different applications, including a spectrum of forecast approaches. Activities in this research area have been growing steadily for more than twenty years and are increasingly applied in energy-related studies. The documentation of machine learning approaches is mainly part of the technically driven and data science literature (see Ahmed and Khalid, 2019; Voyant et al., 2017; El Moursi et al., 2016; Hernandez et al., 2014; Suganthi and Samuel, 2012, as exemplary reviews in the field of renewable production as well as electricity consumption forecasting). Applications of machine learning in economic studies often focus on price forecasts (see, e.g., Kraft et al., 2020a; Keles et al., 2016) or on highly aggregated energy demand (see, e.g., Günay, 2016; Geem, 2011; Hong, 2010; Murat and Ceylan, 2006).

Although economic studies sporadically use machine learning approaches, we state that there is a lack of common knowledge in this field and therefore still a substantial unused potential of sophisticated techniques. Against this background, the contribution of the paper at hand is twofold: (i) we provide an introduction to the concept of machine learning in the context of energy economics and forecasts. Therefore, we discuss a generalized process of machine learning that is intended to serve as a general guideline for energy economists and practitioners who need to apply sophisticated forecast models; (ii) we present the application of three model approaches, containing a classical time series analysis approach, a common machine learning approach and a more sophisticated approach from the field of deep learning, on the basis of two different input datasets. The datasets used for this application are high-frequency electricity load data in connection with weather data – both at a relatively

small level of aggregation – and are thus exemplary for the new requirements of energy forecasting. The Python codes of the application presented here can serve as an introduction for non-experts in the field of machine learning. In combination with its literature review, this paper provides a concise and comprehensible overview of the topic of machine learning in the field of energy forecasting.

The paper is structured as follows: Section 2 discusses the relationship between machine learning and classical econometric approaches and the role of forecasting in energy economics based on a literature review. Section 3 presents a standardized modeling procedure for machine learning forecast models and gives an overview of the applied model approaches. Section 4 describes the application of the three model approaches, based on the intended modeling process, while Section 5 illustrates the results and discusses the application. Finally, Section 6 contains a conclusion and implications for the use of machine learning applications in the context of energy forecasting.

## 2. Machine Learning and Forecasts in Energy Economics

### 2.1. Machine Learning

The term ‘machine learning’ is often used in today’s data-driven science and in connection with new applications such as computer-based recognition of images or language, autonomous driving and many other innovative topics. All these applications have in common the automated processing of large amounts of data. In recent years, machine learning has gained in importance due to significant improvements in data availability, computational power, and algorithms for processing large amounts of data (Rebala et al., 2019). Because of the broad field of application, different definitions of the term ‘machine learning’ exist. Often the superordinate terms ‘artificial intelligence’, ‘machine learning’, and ‘big data’ applications are used synonymously. To reach a more general understanding of the notion, we define the term ‘machine learning’ in line with the technically driven definition by Mitchell (1997): “A computer program is said to learn from experience  $E$  with respect to some class of tasks  $T$  and performance measure  $P$ , if its performance at tasks in  $T$ , as measured by  $P$ , improves with experience  $E$ .” From this definition three main characteristics of a machine learning model can be derived:

- (i) A machine learning model is based on some experience and is therefore data-driven.
- (ii) A machine learning model is trained on a specific task that can be measured by a performance measure (e.g. an error term).

(iii) The performance of a machine learning model generally improves with the amount of relevant experience.

Based on the available input data<sup>1</sup> and the task<sup>2</sup> being pursued, machine learning models can be classified into different learning types: unsupervised learning, supervised learning, semi-supervised learning and reinforcement learning (Murphy, 2012; Rebala et al., 2019).

- Unsupervised learning is used for unlabeled data. The most common task in this type of machine learning is clustering. This task groups the input data into similar clusters which were unknown before the model's application. A typical application is the clustering of customers into customer groups with initially unknown properties.
- Supervised learning is used for labeled data. The most common tasks in this type of machine learning are prediction and classification. A prediction model aims to predict a continuous number based on a set of input data, while a classification model aims to predict a discrete category from the input data. Typical applications are predicting a market price or classifying e-mails as spam.
- Semi-supervised learning is used for only partially labeled data. The model task is often similar to supervised learning, but information gaps in the data must be filled, for example by clustering approaches. Therefore, semi-supervised learning combines unsupervised and supervised approaches.
- Reinforcement learning is used for changing data or large state spaces. Often, a reinforcement model interacts with a changing environment and uses its observations as input for the learning process. Examples for the application of reinforcement learning are autonomous driving or learning complex games such as chess.

In general, machine learning models use the learned structure regarding the data to fulfill their tasks for unknown data points. The resulting model often features a highly non-linear relationship between the input data and the model task that is not easily interpretable by humans. Due to this, some literature mentions the non-linearity of machine learning as a distinguishing feature between machine learning models and 'classical' econometric

---

<sup>1</sup>What kind and amount of data needs to be processed? Is the dataset labeled with 'true values' or unlabeled? Is the database static or does it change (for example due to a dynamic environment or a high frequency of new observations)?

<sup>2</sup>Basic tasks are prediction, classification and clustering. More complex tasks (like image recognition) are often based on these basic tasks or they combine them.

models. However, using the aforementioned definition, even a simple linear regression can be interpreted as machine learning. Therefore, the distinction between machine learning models and ‘classical’ econometric models must be sought in the application of these models.

‘Classical’ econometric models in economics are in general designed to explain the relationship between two or more parameters within the observed data (Varian, 2014; Kleinberg et al., 2015; Athey, 2019). The model structure is based on an underlying economic theory and is parametrized using the observed data. In contrast, machine learning models aim at a generalized description of the data structure and the prediction of unknown data points. The model structure is usually learned within the modeling process (Athey, 2019). By splitting the data into training data and validation data and testing many different model structures, the resulting model is not necessarily the best fit to the observed data, but the most appropriate model for predicting new data points. Therefore, the choice between different modeling approaches may lead to a trade-off between transparency and explicability on the one hand and the predictive quality of new data on the other. This is particularly important if the model output is intended to be used for policy- and decision-making (Athey, 2017).

An ongoing task of data-based science is to combine the strengths of both approaches. Mullainathan and Spiess (2017) point out that machine learning models can be used as input to ‘classical’ econometric models, for instance by generating new input data or by setting a baseline to test theories. Due to the increasing importance of empirical analyses and the growing number of technical possibilities, an increasing synthesis of machine learning and ‘classical’ econometric models is to be expected (Athey, 2019; Mullainathan and Spiess, 2017).

## 2.2. Forecasts in Energy Economics

Forecasting is a fundamental tool of decision-making in the field of energy economics. In the past, the forecast of energy demand has been a crucial determinant for the supply side, especially in the electricity sector (Hahn et al., 2009). Important tasks that require as input accurate forecasting of energy demands over different time spans are:

- (i) In the short term:
  - optimal dispatch of energy plants (Hahn et al., 2009),
  - coordination of network operation tasks (Hernandez et al., 2014), and
  - obtaining of price indications for optimal marketing (Hernandez et al., 2014; Hahn et al., 2009).

(ii) In the long term:

- analyzing different energy sectors and peak demands to design future energy supply and networks (Carvallo et al., 2018; Suganthi and Samuel, 2012),
- identifying strategies to increase efficiency and to reduce greenhouse gas emissions (Suganthi and Samuel, 2012), and
- providing a basis for policy decisions (Suganthi and Samuel, 2012).

Due to the increasing share of fluctuating renewable energies, especially in the electricity sector, supply side forecasts continue to gain importance (Kaur et al., 2016). In addition to the short-term optimization of the market (Barthelmie et al., 2008) and network operations (El Moursi et al., 2016), Ahmed and Khalid (2019) mention the calculation of energy storage and reserve capacities as key tasks to be carried out.

In addition, the current trend in the energy system includes the development of smaller autonomous energy sub-systems, such as microgrids and smart buildings (Hernandez et al., 2014). Emerging concepts of regional energy markets take into account the local relationship between supply and demand (Mengelkamp et al., 2019). Therefore, energy forecasts that predict the behavior of relatively small units of the energy system over a short time span will become increasingly important from both an energy engineering and especially an energy economics perspective.

There is a broad spectrum of literature covering energy forecast methods, which is mainly technically driven. Several reviews provide a comprehensive overview of forecast methods for wind energy production (see for example, Ahmed and Khalid, 2019; Chang, 2014; Giebel et al., 2011; Lei et al., 2009), solar energy production (see, Voyant et al., 2017; Antonanzas et al., 2016; Law et al., 2014; Inman et al., 2013), and demand (see, Hernandez et al., 2014; Suganthi and Samuel, 2012; Hahn et al., 2009).

Beyond the forecast object (demand, supply or prices), there are different options for classifying forecast methods, such as by setting the forecast horizon, by characterizing the input and output parameters, and by describing the forecast model. Classifying forecast models is not a uniform approach in the literature. However, examples of clusters that can be identified are physical bottom-up models, statistical models, spatial correlation models, probabilistic models, and ‘intelligent’ models (see for instance, Henze et al., 2020b; Ahmed and Khalid, 2019; Jung and Broadwater, 2014; Lei et al., 2009). ‘Intelligent’ models often include non-linear approaches, such as artificial neural networks (ANN), support vector machines (SVM), and other related model types (see, Henze et al., 2020a; Ahmed and

Khalid, 2019; Gensler et al., 2016; Hernandez et al., 2014; Hahn et al., 2009). Alongside these, hybrid models and ensemble methods are sometimes referred to as model clusters (see, Ren et al., 2015; Tascikaraoglu and Uzunoglu, 2014; Foley et al., 2012).

Reviewing the economically-oriented literature, some common area of investigation can be identified. The first is country-level modeling of annual energy demand (see for instance, Carvallo et al., 2018; Günay, 2016; Hussain et al., 2016; Limanond et al., 2011; Neto et al., 2011; Hong, 2010; Bhattacharyya and Timilsina, 2010; Azadeh et al., 2008). These applications often aim at long-term energy forecasts of up to 20 years. The second object of investigation is the modeling of high-frequency demand data at the country level (for examples, see García-Ascanio and Maté, 2010; Darbellay and Slama, 2000). This research is closely related to short-term forecast models of energy prices (for examples, see Fraunholz et al., 2021; Karabiber and Xydis, 2019; Ward et al., 2019; Lago et al., 2018; Keles et al., 2016; Ziel et al., 2015; Weron, 2014; Papadimitriou et al., 2014; Keles et al., 2013) or electricity reserve balancing power prices (Kraft et al., 2020b). These studies are based on different types of model approaches, containing a diversity of linear and non-linear statistical models, including econometric models as well as ‘intelligent’ approaches, and in specific cases, bottom-up models.

The situation of forecasts in the context of energy economics can be summarized by two findings:

- (i) Methods of machine learning, especially ANN and SVM, are already part of energy economics studies. However, documentation of the diversity of forecast methods is most pronounced in the technically driven literature, and thus also in renewable production forecasts.
- (ii) Energy economics studies often aim at low-frequency and highly aggregated data. However, given the current development of the energy system, high-frequency forecasts in a relatively small local context will also become increasingly important in this field.

Based on these two findings, we contribute two points to the current development: (i) we introduce a structured framework for the model selection of machine learning approaches in general and in particular for forecast models; (ii) we provide a case study and apply three different forecast model approaches on local electricity demand data. The three selected approaches are exemplary in that they focus on a wide range of possible approaches and contain one classical time series analysis approach (ARIMAX), one common machine learning approach (ANN), and one advanced deep-learning approach (Auto-LSTM). Applying and comparing these approaches demonstrates the usability of machine learning in the field



of high-frequency energy data and illustrates potential differences in how the various models perform.

### 3. Methodology

#### 3.1. Application of Forecast Models

Out of the broad field of forecast techniques, we focus on data-driven statistical models. The related modeling problem is a classical regression problem, whereby modeling approaches can differ regarding their linearity and the complexity of data pre-processing. Therefore, we follow a standardized machine learning procedure to implement forecast models independent of the model approach chosen. Figure 1 summarizes the procedure and divides it into three fundamental phases: preparation (I), model selection (II), and testing (III). These phases feature several connected steps, beginning with a dataset (1) and ending with the evaluation of the implemented model (6).

The initial dataset (1) for our purpose is a set of energy data, such as a time series of energy consumption or production, in connection with certain exogenous parameters, such as a time stamp, weather parameters, or other explanatory variables. In terms of machine learning approaches this is a labeled dataset, where the exogenous variables are the input data and the related electricity consumption data the target data (Rebala et al., 2019).

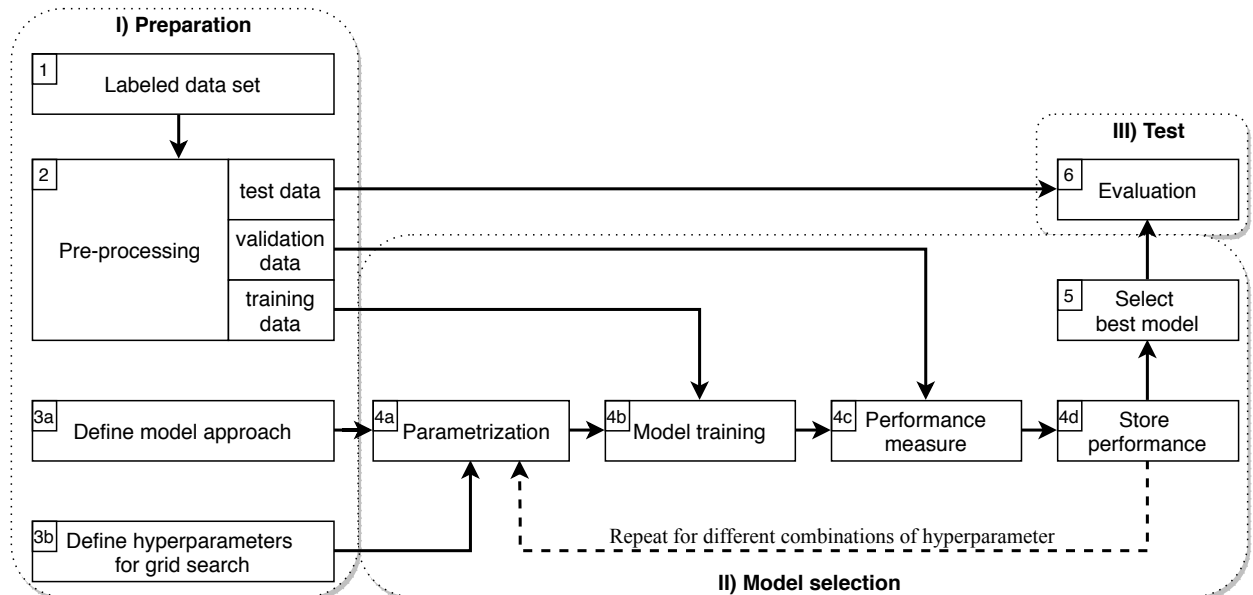


Figure 1: Modeling process of forecast models.

The original dataset must be pre-processed (2). The necessary preparation of the data depends on the selected model approach (3a). As an obligatory pre-process for the further steps, the original data must be divided into three sub-datasets: training data, validation data, and test data. Each of these sets is used for a specific subtask to find the best model for an out-of-sample forecast. Notably, the share of each of the three subsets is not necessarily fixed. In general, the training set is the biggest, making up around 60% of the dataset, while the validation and test sets are smaller, at around 20% of the dataset.

Based on the dataset, a suitable model approach is chosen (3a). In general, each model approach contains various hyperparameters that describe the structure of the model and the various aspects of the learning process. The machine learning approach involves selecting the best model structure (Athey, 2019). Therefore, the search space of possible models must be defined as a grid search (3b), containing a set of model configurations with regard to the selected hyperparameters.

Each model configuration out of the grid search is then parametrized (4a) and fitted with the data of the training set (4b), which is called model training. By using the validation data, the performance of each model configuration can be measured (4c) and stored (4d). Such a performance measure may include various indicators for the discrepancy between the predicted values and the actual values of the validation set and is usually denoted as validation or training error. The steps of parametrization (4a), model training (4b), performance measure (4c), and storage of the performance (4d) are repeated for each model configuration in the grid search.

The division of training data and validation data in these steps is obligatory, as an unbiased statement about the out-of-sample performance of a fitted model can only be made with data that were not part of the learning process. However, to avoid random influences of unfavorable chosen distribution between the training and the validation set, the validation process is often modified. In so-called k-fold cross-validation, different distributions between training and validation data are used for the training and performance measurement of a model (James et al., 2017).

The result of the whole process is a model configuration (5) that – given the chosen model approach (3a) – is the most appropriate for predicting future observations for the analyzed dataset. Therefore, typically the model configuration with the lowest validation error gets selected. To subsequently compare the actual performance of different model configurations, an evaluation (6) is performed on test data that is not itself part of the modeling process

(Rebala et al., 2019). The results of the evaluation, usually denoted as test error, can then be used to compare the different model approaches.

In the following, we give a brief introduction to the model approaches and the forecast horizons we compare in our application. We focus on general characteristics and the most relevant aspects. Detailed documentation for the individual model approaches can be found in the corresponding literature.

### 3.2. ARIMAX

Autoregressive (AR) integrated (I) moving average (MA) models with exogenous input variables (X) are part of the ARMA model family, which is well documented in the economic literature (see for example, Paoletta, 2019; Kirchgässner et al., 2013; Hamilton, 1994). ARMA models are a common approach used for time series analysis in economic applications in general and for energy-related economic forecasts in particular.

The general idea of ARMA models is to explain the current or future values of a time series with their past values. An ARMA model therefore contains  $p$  AR terms expressing the autoregressive influence of  $p$  past values, and  $q$  MA terms expressing the influence of  $q$  past error terms. The number of past values and error terms used can vary and is described by the  $p$ - and  $q$ -order of the model. Furthermore, since an ARMA model should only be used for stationary time series, it is usually extended to an ARIMA model in which the  $d$ -order represents the order of integration or the number of differences needed to make the time series stationary. Finally, if data from outside the time series is inserted into the model, an ARIMAX model is obtained in which X stands for ‘exogenous’. For short-term energy forecasts, such exogenous data can be the temperature or other weather variables.

Another possible expansion of an ARIMA model is the aspect of seasonality. The assumption of a recurring behavior of a system can be modeled by means of cyclic dependencies. Since in our application the exogenous data such as temperature have their own seasonal structure, which could be strongly correlated with the seasonal structure of a seasonal ARIMA model, we omit seasonal model elements and here use only exogenous data.

The following aspects should be kept in mind when using ARIMA/ARIMAX-models for short term energy forecasts:

- ARIMA models assume linear relationships between the target variable (energy) and the past values of the target variable.

- Exogenous input variables (X) can complement an ARIMA model, also assuming a linear relationship. With cyclical exogenous variables, there is no need to model seasonality with seasonal ARIMA model approaches.
- The structure of an ARIMA/ARIMAX-model can be described by the  $(p, d, q)$ -order. These three model parameters are the most important adjustment options of the model and should be considered in the grid search.

### 3.3. Artificial Neural Networks

Artificial Neural Networks (ANN) are among the most popular model approaches when it comes to machine learning or artificial intelligence. The original idea of ANN is to mimic a human brain structure through a mathematical model. Artificial neurons that process information in a defined way are connected with each other. Such a structure allows highly non-linear relationships between input and output data to be modeled. The topic of ANN has been part of academic discussion for more than 20 years, but has gained importance in the last five due to the development of computing power and available algorithms (see for instance, [Rebala et al., 2019](#); [Goodfellow et al., 2016](#); [Murphy, 2012](#); [Bishop, 2009](#)).

As shown in Figure 2, an ANN is organized in different layers. The input layer contains all kinds of input data. From a modeling perspective, there is no differentiation between exogenous and endogenous input parameters. The output layer contains one or more variables to be predicted by the ANN. Between the input and the output layer, at least one hidden layer with at least one node, also called a neuron, must be defined. Such a neuron

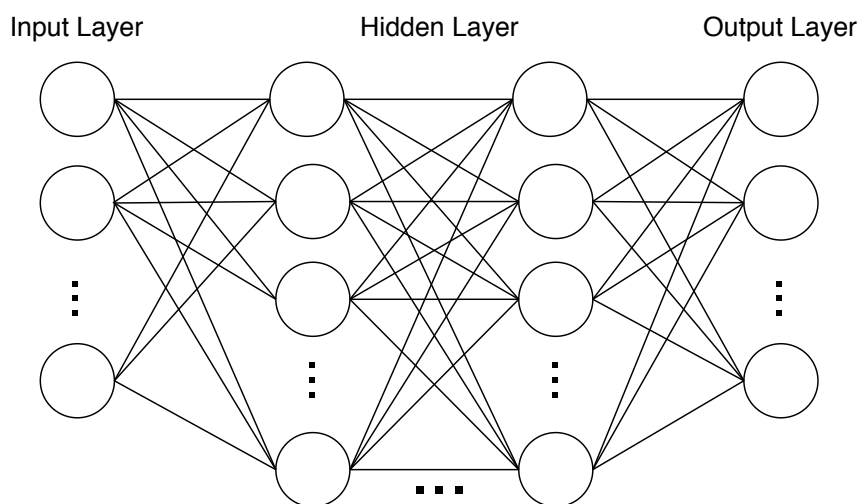


Figure 2: ANN structure

processes an input signal with an activation function in a defined way to an output signal and is then called a perceptron. The perceptrons of two neighboring layers are connected by weights. These weights are the parameters of an ANN. The weights are updated in an iterative back-propagation algorithm to minimize the cost function between the predicted values and the actual values in the training process. In machine learning, the cost function, also often called the loss function or error function, is the function that is minimized via the learning process. It measures how far the predicted values are from the actual values.

A trained ANN can be viewed as a highly complex function between the defined input and output that is not easily interpretable by humans. Note that due to high complexity and non-linearity, the training process finds only a local optimum of the cost function. Therefore, the training process itself can influence the model performance and should be taken into account when identifying the optimal model. A detailed description of ANN modeling and improvements can be found, for example, in [Rebala et al. \(2019\)](#), [Goodfellow et al. \(2016\)](#) and [Bishop \(2009\)](#).

The following aspects should be kept in mind when using ANN models for short-term energy forecasts:

- ANN models assume non-linear relationships between the target variable (energy) and a set of input variables, which can be exogenous or endogenous.
- To obtain a robust modeling process and reduce computation time, all data should be scaled.
- ANN structure can be described as  $n$  layers with  $m$  neurons in each layer and the connections between the layers. In addition to this fundamental structure, the most important design parameter of an ANN is the activation function<sup>3</sup> used in the neurons. These aspects should be part of the grid search.
- The performance of ANN is not only related to the model structure, but also to the learning process. Therefore, some hyperparameters should also be considered in the grid search. There are a number of hyperparameters that can influence the learning

---

<sup>3</sup>The activation function is in general a non-linear function, such as 'ReLU', 'Sigmoid' or 'Tanh' ([Rebala et al., 2019](#)).

process of an ANN. The most important hyperparameters are the initial learning rate<sup>4</sup> and the regularization term.<sup>5</sup>

### 3.4. Auto-LSTM

An Auto-LSTM network is a combination of an AutoEncoder (AE) neural network and a Long Short Term Memory (LSTM) network (Gensler et al., 2016). These neural networks can be accredited to deep learning, an emerging field of new types and methods for neural networks. Deep learning has gained traction in recent years as it has solved problems that have hindered the initial advance of neural networks, like the exploding or vanishing of gradients during the learning process. By applying new paradigms such as layer-wise pre-training, stochastic gradient descent, or rectified linear units, deep learning has led to a resurgence in the potential of neural networks (Goodfellow et al., 2016).

---

<sup>4</sup>The learning rate determines how much the weights change at each iteration (Rebala et al., 2019). Too small a learning rate can lead to a very long training process that can also be terminated. A learning rate that is too high, on the other hand, can lead to suboptimal weights and thus to a suboptimal training process. The initialization of the learning rate is especially important for learning processes with adaptable learning rates. Otherwise, a fixed learning rate is defined.

<sup>5</sup>The regularization term is part of the cost function and prevents an overfitting of the model. The higher the regularization term, the less likely an overfitting will ensue.

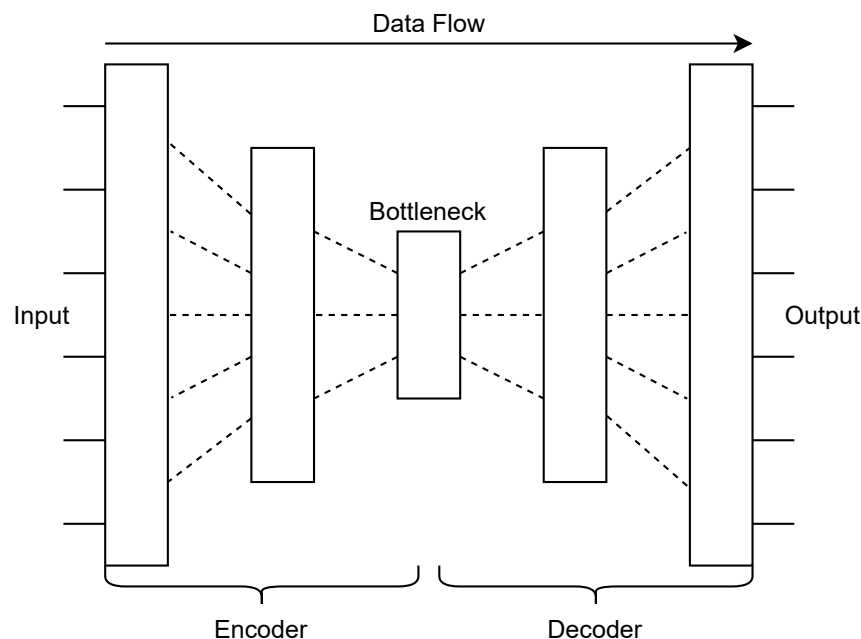


Figure 3: Simplified AutoEncoder with encoder, decoder, and bottleneck

An AutoEncoder learns a representation of data by creating so-called feature hierarchies. A schematic AE is shown in Figure 3. The size of those feature hierarchies, or layers, will become smaller until they reach a user-defined bottleneck layer. After the bottleneck layer, the layer size increases again until the input size is achieved. This neural network architecture allows the AE to obtain a feature representation at the bottleneck. An AE is trained in an unsupervised environment, as the AE only requires input data to obtain the feature representation.

LSTMs belong to the group of recurrent neural networks (RNN) that have a temporal connection. They do this by creating a direct connection between the output of the RNN and the input of the RNN. This connection allows them to take into account previously computed output in the next calculation of the output. Alongside this temporal connection, LSTMs have memory cells that enable them to store information for later use. This boosts their performance in forecasting, as they have both the knowledge of the previous evaluation of the neural network and the knowledge they have stored in their memory cell (Hochreiter and Schmidhuber, 1997).

The Auto-LSTM, as shown in Figure 4, combines both of these neural network types, the AE and the LSTM. Starting from the AE, the Auto-LSTM uses the pretrained encoder part of the neural network. The encoder acts like a pre-processing neural network for the LSTM and allows the LSTM to use the information stored in the latent variables of the learned feature representation. During the training, the AE and the Auto-LSTM are trained separately. First, the complete AE is trained using only the input data. Then the AE is split at the bottleneck so that only the learned feature representation is accessed. The output of the bottleneck is then used as input to the LSTM layer. Before training the LSTM, the weights of the encoder are fixed so that the learned representation does not

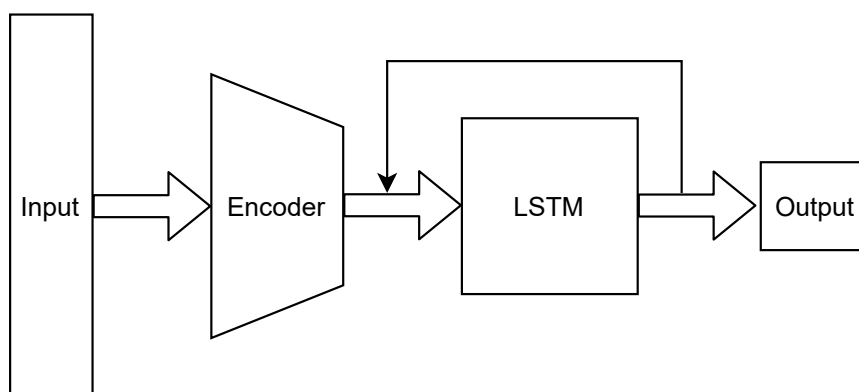


Figure 4: Simplified Auto-LSTM that combines the encoder part of the AE with an LSTM network

change. In the last training step, only the weights of the LSTM are trained in a supervised training (Gensler et al., 2016).

### 3.5. *Static vs. Dynamic Forecasts*

Even with the focus on short-term energy forecasts, there is still a range of possible time scales from only a few seconds or minutes (so called ultra-short-term forecasts) to hours or even one or two days (El Moursi et al., 2016; Chang, 2014). An ultra-short-term forecast is often only relevant for special technical considerations. From an energy economics point of view, the minimum relevant forecast horizon appears to be a quarter of an hour or even an hour. In addition, the minimum forecast horizon is determined by the temporal resolution of the data basis. The maximum relevant forecast horizon on the other hand depends on the application and can represent any time scale within the defined range.

The targeted forecast horizon must be linked to the forecast method. If the implemented model uses information about past output data (endogenous variables), two types of forecast can be distinguished. A static forecast predicts the output data in the next time step based on complete information about the input data. This includes the exogenous variables for the next time step and the real output data from the previous time steps as endogenous variables. In contrast, a dynamic forecast predicts more than just the next time step. For each additional time step, the exogenous variables for this time step and the predicted output data of the previous time steps are used. In consequence, dynamic forecasts tend to produce greater forecast errors than static forecasts. Therefore, the choice of an appropriate forecast horizon is often a trade-off between a long forecast horizon and high accuracy.

## 4. Application

Based on the methodology presented in Section 3, we compare the performance of different model approaches in predicting hourly local energy demand.<sup>6</sup> In the following, we describe the dataset used, the pre-processing of the data, the selection of the grid search parameters, and finally the best fitting model selection.

### 4.1. *Data*

Our dataset consists of one year of data from a small regional power grid in central Germany. It consists of two types of data:

---

<sup>6</sup>The application was implemented in Python. The ARIMAX models are based on the Python module ‘statsmodels’ (Seabold and Perktold, 2010), the ANN models on ‘scikit learn’ (Pedregosa et al., 2011), and the Auto-LSTM models on the authors’ own developments (following Gensler et al., 2016) The set of used Python code can be found in Heilmann (2021).



- numerical weather prediction (NWP) data, and
- electricity data from generators and consumers.

The NWP raw data consists of historical weather data with a temporal resolution of six hours. 27 different weather features are available at each time step, such as temperature, wind speed, or solar radiation.

The complete electricity dataset contains a mix of 11 power generators, 45 metered small business consumers, 44 residential areas, and the aggregated residual load at the transformer. We focus on electricity consumption data and use only time series with completely metered data – that is, the data used consists of the load of 42 small- to medium-sized businesses and the residual load measured at the transformer in the medium- voltage network area. The temporal resolution of the data is 15 minutes.

#### *4.2. Pre-processing of Data*

The pre-processing of the data is made up of three essential steps to customize the original data into an interpretable data format for the different model approaches.<sup>7</sup> In a first step, we build an input dataset, containing the weather parameters, and a target dataset, containing the load data to be predicted. Therefore, we only use the measured load data of the original dataset and do not use the generation data and the data from the residential areas. To obtain a uniform resolution, we adjust the input and target data to an hourly frequency by interpolation and averaging. Besides this, we extend the input data with categorical data describing the time stamp, which are the month of the year, the day of the month, the day of the week, and the hour of the day. To obtain a continuous signal of the time parameters, we encode them with a sin-cos transformation.

In a second step, we define two variations of the input data: a complete set with a total 35 variables and a reduced set with only five variables, containing the encoded signals of the day of the week and the hour of the day, as well as the temperature. The third step is to divide the complete and the reduced dataset into training data, validation data, and test data. In both cases, we declare 60% of the dataset as the training set and 20% each as the validation and test set.

---

<sup>7</sup>Note that additional pre-processing steps may be necessary to meet the requirements of the model approaches used. These include, for example, the scaling of all input and target data, which is particularly important for the ANN learning process (Rebala et al., 2019)). Such minor pre-processing steps are not documented in detail in this section, but appear in the source code.

### 4.3. Grid Search Parameters

For each of the selected model approaches – that is, ARIMAX, ANN, and Auto-LSTM – a search space of model hyperparameters for the grid search must be defined. Table 1 summarizes the grid search parameters for all model approaches. For the ARIMAX approach, only the p-, d- and q-order were varied. Overall, 75 ARIMAX configurations were tested.

For the ANN approach, more hyperparameters were varied. The number of layers and the number of neurons in each layer characterize the structure of the ANN. The activation function defines the kind of processing in each neuron of the ANN. The regularization parameter is part of the cost function and therefore affects the training result of the fitted model. The initial learning rate influences the training process itself. Note that additional hyperparameters, such as the solver used and other fine-tuning parameters of the learning process, can be defined explicitly or can also be part of the grid search.<sup>8</sup> Overall, 648 ANN-configurations were tested.

The grid search of the Auto-LSTM is similar to that of the ANN. As described earlier, the Auto-LSTM follows a multi-step approach during training, where the two parts are trained separately. Therefore, during the grid search, the objective is to first train a good representation of the input data using the AE and then to train an LSTM using the previously obtained feature representation.

The hyperparameters for the AE are fixed. The initial learning rate of 0.001 is adjusted during training. The layout of the AE is customized based on the dataset used as input. The AE structure is (35, 30, 20, 10) for the complete input and (5, 10, 7, 4) for the reduced input. The AE uses Leaky Rectified Linear Units (Leaky ReLU) as activations for each layer. The trained AE is split at the bottleneck and the weights are fixed, so that they are not changed during the training of the LSTM layer. For the LSTM training, the initial learning rate is adjusted in the same manner as for the AE. Furthermore, we varied the number of LSTM cells and the size of their hidden layers. This varies the amount of time-dependent information each LSTM cell retains (hidden layers) and the amount of previous data considered during a forecast (expressed as number of LSTM cells). The initial weight values of the LSTM cells are either randomly initialized or have been initialized with a constant. In addition to training the LSTM cells, we also permitted training on the last layer of the encoder, to allow slight adjustments of the learned representation to the forecast task.

---

<sup>8</sup>The more hyperparameters considered, the larger the grid search and the more calculation effort would be necessary for a comprehensive grid search. Additional information about all possible hyperparameters can be found in the documentation of the Python modules used (Pedregosa et al., 2011).

Table 1: Grid search ranges of all model approaches

Model approach	Hyperparameter	Range	Variants
ARIMAX	$p$ - and $q$ -order	[0, 1, 2, 3, 4]	75
	$d$ -order	[0, 1, 2]	
ANN	Number of layers	[2, 3]	648
	Neurons in each layer	[5, 10, 15]	
	Activation functions	[ReLU, tanh]	
	Regularization parameter	[0.001, 0.01, 0.1]	
	Initial learning rate	[0.001, 0.01, 0.1]	
Auto-LSTM	Number of LSTM cells	[1, 2, 4]	18
	Hidden size of LSTM cells	[1, 5, 10]	
	Random initialization of LSTM weights	[True, False]	

#### 4.4. Model Selection

The model selection phase aims at finding the best fitting model for each model configuration. A model configuration contains a specific target dataset (one of the load time series), an input dataset (‘complete’ or ‘reduced’), and one of the three model approaches. Therefore, for each of the 43 target datasets, we implement six different forecast models. Each model undergoes a grid search for possible model hyperparameters by training the model on the training dataset for each model configuration. As a performance measure for each trained model, we use the root mean squared error (RMSE) between the statistical model output and the validation dataset. We select the model with the best performance measure for each model configuration – that is the model with the lowest RMSE.

## 5. Results and Discussion

With the procedure described in Section 4, we obtain a total of 258 models (43 load time series x 3 model approaches x 2 input variations). We evaluate the RMSE of two forecast horizons, a one-step static forecast and a 24-hour dynamic forecast, as described in Section 3.5. Note that the Auto-LSTM models only provide a forecast over the entire horizon and therefore only provide dynamic forecasts in this case. To make the RMSE comparable between the different load datasets, we normalize each RMSE with the maximum value of the load in the training set. For evaluating the forecast performance of the different models,

primarily the test errors are essential. However, for the analysis of the learning process, training errors can also be helpful.

Figure 5 shows the distribution of training and test errors of the selected best-fitting ARIMAX and ANN models for the one-step static forecast. The left-hand side shows the results for the models with the complete input data while the right-hand side shows the results for the models with the reduced input data.

Overall, three main aspects for the static forecast horizon can be derived from this: (i) the ANN models have, on average, lower errors than the ARIMAX models and thus perform better on average; (ii) the training errors of the ANN models are lower than the test errors. This phenomenon is typical for machine learning modeling processes, but cannot be observed for the ARIMAX models; (iii) there is no clear difference between the performance of the models with complete input data and the models with reduced input data.

The results for the 24-hour dynamic forecast horizon are shown in Figure 6. Here, in addition to the selected best-fitting ARIMAX and ANN models, the selected best-fitting Auto-LSTM models are also included.

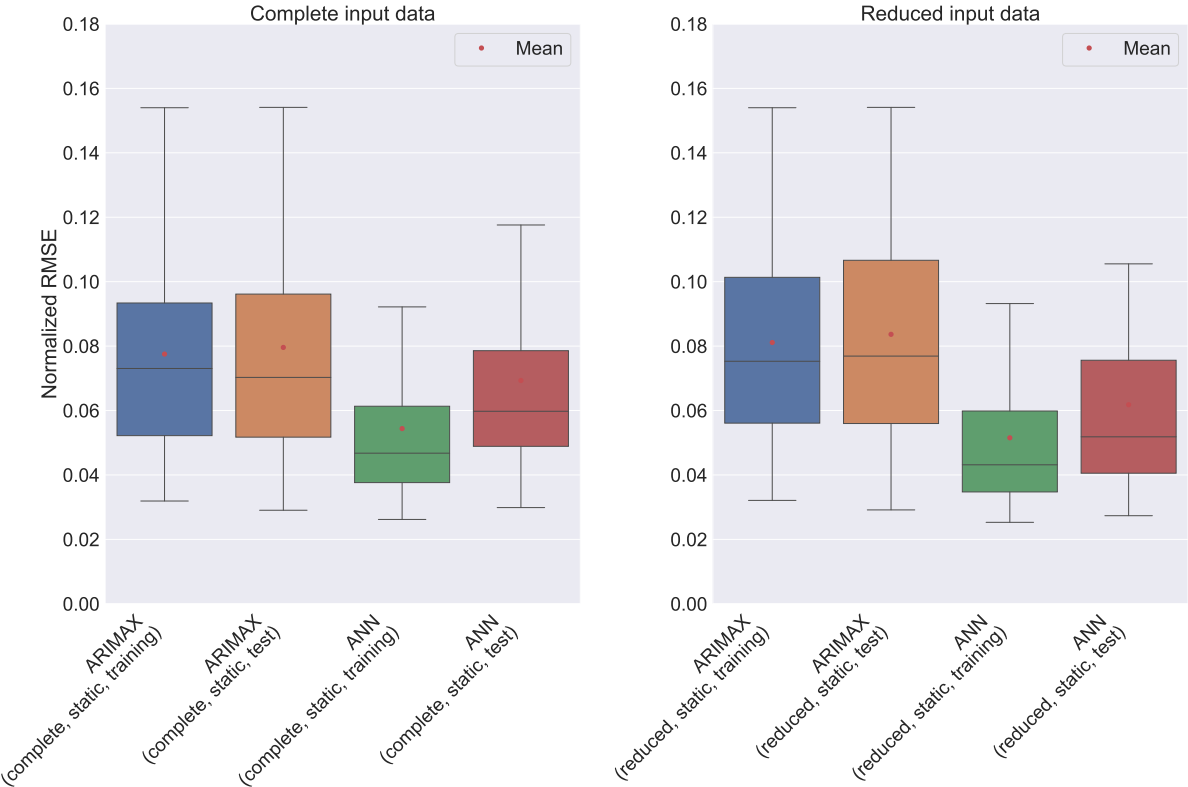


Figure 5: Distribution of training and test errors for the one-step static forecast

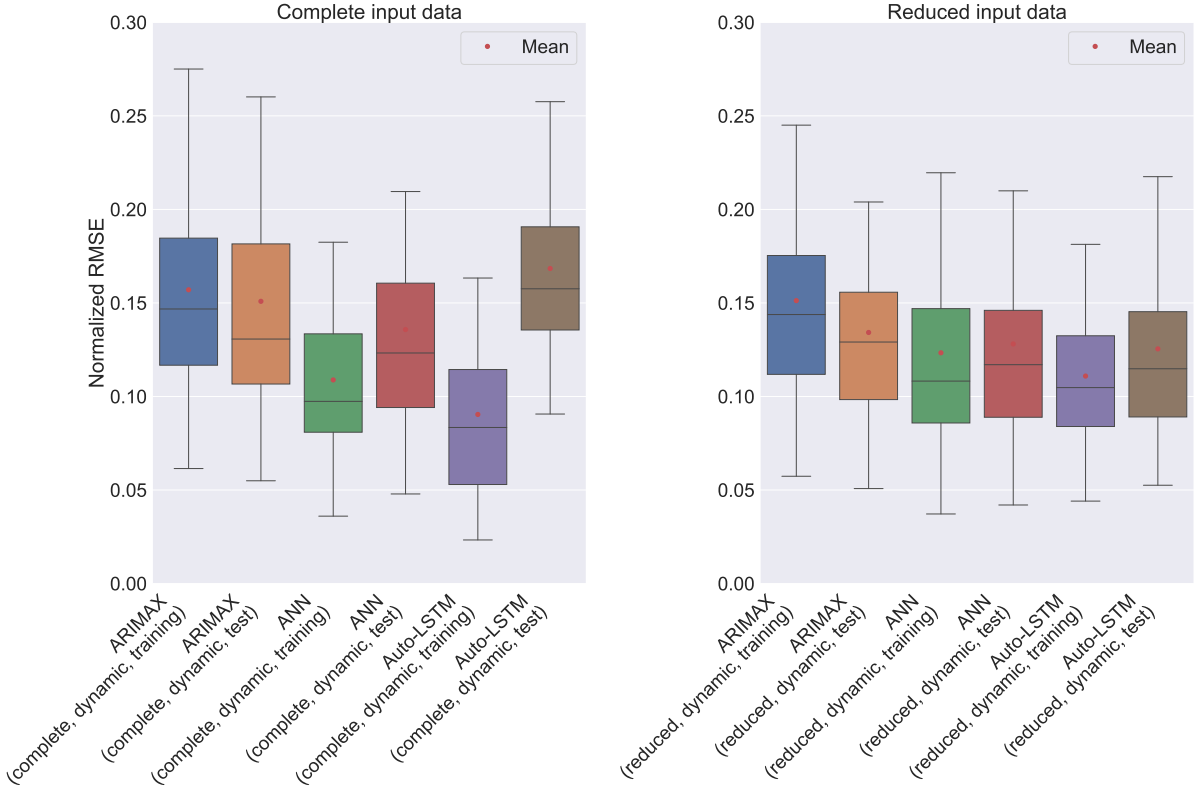


Figure 6: Distribution of training and test errors for the 24-hour dynamic forecast

Overall, it can be seen that the findings of the static forecast horizon can only be partially confirmed: (i) there is no clear trend as to which model approach performs best, especially on the test set; (ii) the finding that the training performance is better than the test performance is especially true for the Auto-LSTM models with complete input data. While the average training error is the lowest of all evaluated dynamic forecast performance measures, the average test error is the highest. The chosen model configurations are obviously not able to generalize the data, a phenomenon called ‘overfitting’. There are several ways to deal with overfitting, such as obtaining more training data, using regularization, or other options discussed, for example, in [Rebala et al., 2019](#). Contrary to this observation, the ARIMAX models perform even better on the test set than on the training set. (iii) similar to the static forecast horizon, no clear difference is observable between models with complete and reduced input data. One exception is the already mentioned poor performance of the Auto-LSTM with complete input data on the test set.

Figure 7 shows a heat map of the static and dynamic test errors for all selected best-fitting models for the 43 electricity load times series. The vertical axis show the 43 electricity load



Figure 7: Static and dynamic test errors for the 43 load times series

time series on which the models were trained and tested. The horizontal axis shows each of the model approaches (*ANN*, *ARIMAX*, and *Auto-LSTM*) combined with the two input data variants (*complete* and *reduced*), and the two forecast horizons (*static* and *dynamic*). The lighter the color for a dataset-model combination, the larger the error and thus the worse the forecast performance.

Overall, the heat map illustrates two aspects: (i) a pattern between the static and dynamic forecasts can be observed that underlines the expected characteristic of forecast models that static forecast errors are in general lower than dynamic forecast errors; (ii) the second salience is the comparably poor performance of the Auto-LSTM models with complete input, which has already been discussed above.

Summarizing the results of our empirical application, the following insights arise:

- Each of the three model approaches can provide forecasts for short-term electricity consumption. There is no general tendency towards a best-suited model approach.
- The static error is in general smaller than the dynamic error, which is in line with the expectation that the error term accumulates with every additional forecast time step.
- There is almost no difference in model performance between models based on the full input dataset and models based on the reduced dataset.
- ANN and Auto-LSTM models show a lag between training and test errors. Hence, the models describe the training data more effectively than the test data. This effect cannot be observed for the ARIMAX models. This is in line with the expectation that more complex models are more likely to overfit the data.

## 6. Conclusion and Policy Implication

The present paper brings together a standard topic in energy economics, namely energy forecasts, with the approach of machine learning. Machine learning is a fundamental concept in data-driven science with a broad spectrum of applications, including the field of energy consumption forecasts. Documenting these machine learning approaches forms a large part of the technically driven literature. Although economic studies also use machine learning approaches, there is still a substantial unused potential of sophisticated techniques.

To give a flavor of the idea of machine learning, we introduce the systematic machine learning model selection process and brought it into the context of three different model approaches, namely the classical economic time series analysis approach, ARIMAX, one of the most used non-linear machine learning approaches, ANN, and a further development of ANNs, Auto-LSTM. We apply the introduced machine learning model selection process to each of the model approaches and implement forecast models for local hourly electricity demand data. In addition, we use two different input datasets with, respectively, 35 and five different variables as exogenous explanatory data.

The results show that each of the approaches used is, in principle, able to provide short-term forecast models. How the approaches perform varies depending on the different datasets. The evaluation of the model performance suggests that ANNs perform slightly better on a static forecast than do ARIMAX models, while this difference cannot be evidenced for dynamic forecasts. The highly sophisticated approach of Auto-LSTM, which only generates dynamic forecasts, tends to overfit the training data and therefore does not generally outperform on unknown test data. Nevertheless, in some cases, Auto-LSTM provided the best model performance in the evaluated examples.

Based on the exemplary application and evaluation, three main implications can be derived:

- (i) Our results suggest that there is no general relationship between high model complexity and good model performance. On the contrary, the more complex Auto-LSTM models tend to overfit the underlying data in some cases. Therefore, if a classical model approach leads to a model performance of the desired quality, it is inadvisable to choose a more complex approach than necessary.
- (ii) We did not find an improved performance with more complex input data. The reduced input dataset (with only five variables) provided forecast results similar to the complete input dataset (with 35 variables). This aspect can be interpreted in two ways, depending on which data are available. If data collection is costly, it may be advisable to only use basic data as the explanatory basis. If access to a larger data base exists, it can be reasonable to use the complete dataset without a manual selection of variables.
- (iii) All applied approaches are in principle able to provide forecasts with comparable quality. However, it may be worthwhile to compare models of different types and with different inputs in order to select the one that is best suited for the particular objective. Therefore, the introduced machine learning process should be used as a framework for a structured model selection. This can be applied not only to forecast models, but to any model approach that aims at generalizing from training data to unknown new (test) data.

Our application covers only a small part of the vast methodological diversity of machine learning and it uses a relatively small dataset, from the perspective of other machine learning applications. However, the model selection process introduced here has become standard for many applications, especially in the field of forecasting. This diversity is expected to grow still more in the coming years, with the approaches becoming more intuitive and



straightforward to use and also more easy to interpret. Economic studies relating to energy use should in particular include an appropriate mix of these innovative approaches with those classical approaches based on economic theory.

## Data Availability

Datasets related to this article can be requested at <https://www.uni-kassel.de/eecs/en/sections/intelligent-embedded-systems/downloads>, as 'HessianLoad data set'. The related Python codes can be found at <http://dx.doi.10.17632/7xxbfsybxz.1>, an open-source online data repository hosted at Mendeley Data (Heilmann, 2021).

## Acknowledgments

This work is financed by the German Federal Ministry for Economic Affairs through the smart energy showcases (SINTEG) program (funding code 03SIN119).

## References

- Ahmed, A., Khalid, M., 2019. A review on the selected applications of forecasting models in renewable power systems. *Renewable and Sustainable Energy Reviews* 100, 9–21. doi:[10.1016/j.rser.2018.09.046](https://doi.org/10.1016/j.rser.2018.09.046).
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de Pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. *Solar Energy* 136, 78–111. doi:[10.1016/j.solener.2016.06.069](https://doi.org/10.1016/j.solener.2016.06.069).
- Athey, S., 2017. Beyond prediction: Using big data for policy problems. *Science* (New York, N.Y.) 355, 483–485. doi:[10.1126/science.aal4321](https://doi.org/10.1126/science.aal4321).
- Athey, S., 2019. The Impact of Machine Learning on Economics, in: Agrawal, A., Gans, J., Goldfarb, A. (Eds.), *The economics of artificial intelligence*. The University of Chicago Press, Chicago. National Bureau of Economic Research conference report.
- Azadeh, A., Ghaderi, S.F., Sohrabkhani, S., 2008. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management* 49, 2272–2278. doi:[10.1016/j.enconman.2008.01.035](https://doi.org/10.1016/j.enconman.2008.01.035).
- Barthelmie, R.J., Murray, F., Pryor, S.C., 2008. The economic benefit of short-term forecasting for wind energy in the UK electricity market. *Energy Policy* 36, 1687–1696. doi:[10.1016/j.enpol.2008.01.027](https://doi.org/10.1016/j.enpol.2008.01.027).
- Bhattacharyya, S.C., Timilsina, G.R., 2010. Modelling energy demand of developing countries: Are the specific features adequately captured? *Energy Policy* 38, 1979–1990. doi:[10.1016/j.enpol.2009.11.079](https://doi.org/10.1016/j.enpol.2009.11.079).
- Bishop, C.M., 2009. *Pattern recognition and machine learning*. Information science and statistics. corrected at 8th printing 2009 ed., Springer, New York, NY.
- Carvalho, J.P., Larsen, P.H., Sanstad, A.H., Goldman, C.A., 2018. Long term load forecasting accuracy in electric utility integrated resource planning. *Energy Policy* 119, 410–422. doi:[10.1016/j.enpol.2018.04.060](https://doi.org/10.1016/j.enpol.2018.04.060).

- Chang, W.Y., 2014. A Literature Review of Wind Forecasting Methods. *Journal of Power and Energy Engineering* 02, 161–168. doi:[10.4236/jpee.2014.24023](https://doi.org/10.4236/jpee.2014.24023).
- Darbellay, G.A., Slama, M., 2000. Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting* 16, 71–83. doi:[10.1016/S0169-2070\(99\)00045-X](https://doi.org/10.1016/S0169-2070(99)00045-X).
- El Moursi, M.S., Al Hinai, A., Ssekulima, E.B., Anwar, M.B., 2016. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review. *IET Renewable Power Generation* 10, 885–989. doi:[10.1049/iet-rpg.2015.0477](https://doi.org/10.1049/iet-rpg.2015.0477).
- Foley, A.M., Leahy, P.G., Marvuglia, A., McKeogh, E.J., 2012. Current methods and advances in forecasting of wind power generation. *Renewable Energy* 37, 1–8. doi:[10.1016/j.renene.2011.05.033](https://doi.org/10.1016/j.renene.2011.05.033).
- Fraunholz, C., Kraft, E., Keles, D., Fichtner, W., 2021. Advanced price forecasting in agent-based electricity market simulation. *Applied Energy* 290, 116688. doi:[10.1016/j.apenergy.2021.116688](https://doi.org/10.1016/j.apenergy.2021.116688).
- García-Ascanio, C., Maté, C., 2010. Electric power demand forecasting using interval time series: A comparison between VAR and iMLP. *Energy Policy* 38, 715–725. doi:[10.1016/j.enpol.2009.10.007](https://doi.org/10.1016/j.enpol.2009.10.007).
- Geem, Z.W., 2011. Transport energy demand modeling of South Korea using artificial neural network. *Energy Policy* 39, 4644–4650. doi:[10.1016/j.enpol.2011.05.008](https://doi.org/10.1016/j.enpol.2011.05.008).
- Gensler, A., Henze, J., Sick, B., Raabe, N., 2016. Deep Learning for solar power forecasting — An approach using AutoEncoder and LSTM Neural Networks, in: 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), pp. 002858–002865. doi:[10.1109/SMC.2016.7844673](https://doi.org/10.1109/SMC.2016.7844673).
- Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M., Draxl, C., 2011. The State of the Art in Short-Term Prediction of Wind Power A Literature Overview, 2nd Edition. doi:[10.13140/RG.2.1.2581.4485](https://doi.org/10.13140/RG.2.1.2581.4485).
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep learning. MIT Press, Cambridge, Massachusetts and London, England. URL: <http://www.deeplearningbook.org/>.
- Günay, M.E., 2016. Forecasting annual gross electricity demand by artificial neural networks using predicted values of socio-economic indicators and climatic conditions: Case of Turkey. *Energy Policy* 90, 92–101. doi:[10.1016/j.enpol.2015.12.01](https://doi.org/10.1016/j.enpol.2015.12.01).
- Hahn, H., Meyer-Nieberg, S., Pickl, S., 2009. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research* 199, 902–907. doi:[10.1016/j.ejor.2009.01.062](https://doi.org/10.1016/j.ejor.2009.01.062).
- Hamilton, J.D., 1994. Time series analysis. Princeton Univ. Press, Princeton, NJ.
- Heilmann, E., 2021. Machine learning in energy forecasts: An application to high frequency electricity consumption data”, mendeley data, v1. doi:[10.17632/7xxbfxybxz.1](https://doi.org/10.17632/7xxbfxybxz.1).
- Henze, J., Schreiber, J., Sick, B., 2020a. Representation Learning in Power Time Series Forecasting, in: Deep Learning: Algorithms and Applications. Springer, pp. 67–101. doi:[10.1007/978-3-030-31760-7\\_3](https://doi.org/10.1007/978-3-030-31760-7_3).
- Henze, J., Siefert, M., Bremicker-Trübelhorn, S., Asanalieva, N., Sick, B., 2020b. Probabilistic upscaling and aggregation of wind power forecasts. *Energy, Sustainability and Society* 10, 15. doi:[10.1186/s13705-020-00247-4](https://doi.org/10.1186/s13705-020-00247-4).
- Hernandez, L., Baladron, C., Aguiar, J.M., Carro, B., Sanchez-Esguevillas, A.J., Lloret, J., Massana, J., 2014. A Survey on Electric Power Demand Forecasting: Future Trends in Smart Grids, Microgrids and Smart Buildings. *IEEE Communications Surveys & Tutorials* 16, 1460–1495. doi:[10.1109/SURV.2014.032014.00094](https://doi.org/10.1109/SURV.2014.032014.00094).

- Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. *Neural computation* 9, 1735–80. doi:[10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- Hong, W.C., 2010. Application of chaotic ant swarm optimization in electric load forecasting. *Energy Policy* 38, 5830–5839. doi:[10.1016/j.enpol.2010.05.033](https://doi.org/10.1016/j.enpol.2010.05.033).
- Hussain, A., Rahman, M., Memon, J.A., 2016. Forecasting electricity consumption in Pakistan: The way forward. *Energy Policy* 90, 73–80. doi:[10.1016/j.enpol.2015.11.028](https://doi.org/10.1016/j.enpol.2015.11.028).
- Inman, R.H., Pedro, H.T., Coimbra, C.F., 2013. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science* 39, 535–576. doi:[10.1016/j.pecs.2013.06.002](https://doi.org/10.1016/j.pecs.2013.06.002).
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2017. *An introduction to statistical learning: With applications in R*. Springer texts in statistics. corrected at 8th printing ed., Springer, New York and Heidelberg and Dordrecht and London.
- Jung, J., Broadwater, R.P., 2014. Current status and future advances for wind speed and power forecasting. *Renewable and Sustainable Energy Reviews* 31, 762–777. doi:[10.1016/j.rser.2013.12.054](https://doi.org/10.1016/j.rser.2013.12.054).
- Karabiber, O., Xydis, G., 2019. Electricity Price Forecasting in the Danish Day-Ahead Market Using the TBATS, ANN and ARIMA Methods. *Energies* 12, 928. doi:[10.3390/en12050928](https://doi.org/10.3390/en12050928).
- Kaur, A., Nonnenmacher, L., Coimbra, C.F., 2016. Net load forecasting for high renewable energy penetration grids. *Energy* 114, 1073–1084. doi:[10.1016/j.energy.2016.08.067](https://doi.org/10.1016/j.energy.2016.08.067).
- Keles, D., Genoese, M., Möst, D., Ortlieb, S., Fichtner, W., 2013. A combined modeling approach for wind power feed-in and electricity spot prices. *Energy Policy* 59, 213–225. doi:[10.1016/j.enpol.2013.03.028](https://doi.org/10.1016/j.enpol.2013.03.028).
- Keles, D., Scelle, J., Paraschiv, F., Fichtner, W., 2016. Extended forecast methods for day-ahead electricity spot prices applying artificial neural networks. *Applied Energy* 162, 218–230. doi:[10.1016/j.apenergy.2015.09.087](https://doi.org/10.1016/j.apenergy.2015.09.087).
- Kirchgässner, G., Wolters, J., Hassler, U., 2013. *Introduction to modern time series analysis*. Springer Texts in Business and Economics. 2nd ed. 2013 ed., Springer, Berlin and Heidelberg. doi:[10.1007/978-3-642-33436-8](https://doi.org/10.1007/978-3-642-33436-8).
- Kleinberg, J., Ludwig, J., Mullainathan, S., Obermeyer, Z., 2015. Prediction Policy Problems. *American Economic Review* 105, 491–495. doi:[10.1257/aer.p20151023](https://doi.org/10.1257/aer.p20151023).
- Kraft, E., Keles, D., Fichtner, W., 2020a. Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting* doi:[10.1002/for.2693](https://doi.org/10.1002/for.2693).
- Kraft, E., Keles, D., Fichtner, W., 2020b. Modeling of frequency containment reserve prices with econometrics and artificial intelligence. *Journal of Forecasting* 39, 1179–1197. doi:[10.1002/for.2693](https://doi.org/10.1002/for.2693).
- Lago, J., de Ridder, F., de Schutter, B., 2018. Forecasting spot electricity prices: Deep learning approaches and empirical comparison of traditional algorithms. *Applied Energy* 221, 386–405. doi:[10.1016/j.apenergy.2018.02.069](https://doi.org/10.1016/j.apenergy.2018.02.069).
- Law, E.W., Prasad, A.A., Kay, M., Taylor, R.A., 2014. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – A review. *Solar Energy* 108, 287–307. doi:[10.1016/j.solener.2014.07.008](https://doi.org/10.1016/j.solener.2014.07.008).
- Lei, M., Shiyao, L., Chuanwen, J., Hongling, L., Yan, Z., 2009. A review on the forecasting of wind speed and generated power. *Renewable and Sustainable Energy Reviews* 13, 915–920. doi:[doi.org/10.1016/j.rser.2008.02.002](https://doi.org/10.1016/j.rser.2008.02.002).

- Limanond, T., Jomnonkwao, S., Srikaew, A., 2011. Projection of future transport energy demand of Thailand. *Energy Policy* 39, 2754–2763. doi:[10.1016/j.enpol.2011.02.045](https://doi.org/10.1016/j.enpol.2011.02.045).
- Mengelkamp, E., Diesing, J., Weinhardt, C., 2019. Tracing Local Energy Markets: A Literature Review. doi:[10.13140/RG.2.2.17644.21128](https://doi.org/10.13140/RG.2.2.17644.21128).
- Mitchell, T.M., 1997. *Machine learning*. McGraw-Hill series in computer science. international ed. ed., McGraw-Hill, New York, NY.
- Mullainathan, S., Spiess, J., 2017. Machine Learning: An Applied Econometric Approach. *Journal of Economic Perspectives* 31, 87–106. doi:[10.1257/jep.31.2.87](https://doi.org/10.1257/jep.31.2.87).
- Murat, Y.S., Ceylan, H., 2006. Use of artificial neural networks for transport energy demand modeling. *Energy Policy* 34, 3165–3172. doi:[10.1016/j.enpol.2005.02.010](https://doi.org/10.1016/j.enpol.2005.02.010).
- Murphy, K.P., 2012. *Machine learning: A probabilistic perspective*. Adaptive computation and machine learning series, MIT Press, Cambridge, Mass. URL: <https://ebookcentral.proquest.com/lib/subhh/detail.action?docID=3339490>.
- Neto, J.C.d.L., da Costa Junior, C.T., Bitar, S.D., Junior, W.B., 2011. Forecasting of energy and diesel consumption and the cost of energy production in isolated electrical systems in the Amazon using a fuzzification process in time series models. *Energy Policy* 39, 4947–4955. doi:[10.1016/j.enpol.2011.06.020](https://doi.org/10.1016/j.enpol.2011.06.020).
- Paoella, M.S., 2019. *Linear models and time-series analysis: Regression, ANOVA, ARMA and GARCH*. Wiley series in probability and statistics, John Wiley & Sons Inc, Hoboken, NJ. doi:[10.1002/9781119432036](https://doi.org/10.1002/9781119432036).
- Papadimitriou, T., Gogas, P., Stathakis, E., 2014. Forecasting energy markets using support vector machines. *Energy Economics* 44, 135–142. doi:[10.1016/j.eneco.2014.03.017](https://doi.org/10.1016/j.eneco.2014.03.017).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* 12, 2825–2830.
- Rebala, G., Ravi, A., Churiwala, S., 2019. *An Introduction to Machine Learning*. Springer International Publishing, Cham. doi:[10.1007/978-3-030-15729-6](https://doi.org/10.1007/978-3-030-15729-6).
- Ren, Y., Suganthan, P.N., Srikanth, N., 2015. Ensemble methods for wind and solar power forecasting—A state-of-the-art review. *Renewable and Sustainable Energy Reviews* 50, 82–91. doi:[10.1016/j.rser.2015.04.081](https://doi.org/10.1016/j.rser.2015.04.081).
- Seabold, S., Perktold, J., 2010. statsmodels: Econometric and statistical modeling with python, in: 9th Python in Science Conference.
- Suganthi, L., Samuel, A.A., 2012. Energy models for demand forecasting—A review. *Renewable and Sustainable Energy Reviews* 16, 1223–1240. doi:[10.1016/j.rser.2011.08.014](https://doi.org/10.1016/j.rser.2011.08.014).
- Tascikaraoglu, A., Uzunoglu, M., 2014. A review of combined approaches for prediction of short-term wind speed and power. *Renewable and Sustainable Energy Reviews* 34, 243–254. doi:[10.1016/j.rser.2014.03.033](https://doi.org/10.1016/j.rser.2014.03.033).
- Varian, H.R., 2014. Big Data: New Tricks for Econometrics. *Journal of Economic Perspectives* 28, 3–28. doi:[10.1257/jep.28.2.3](https://doi.org/10.1257/jep.28.2.3).
- Voyant, C., Notton, G., Kalogirou, S., Nivet, M.L., Paoli, C., Motte, F., Fouilloy, A., 2017. Machine learning methods for solar radiation forecasting: A review. *Renewable Energy* 105, 569–582. doi:[10.1016/j.renene.2017.05.033](https://doi.org/10.1016/j.renene.2017.05.033).

[1016/j.renene.2016.12.095](https://doi.org/10.1016/j.renene.2016.12.095).

Ward, K.R., Green, R., Staffell, I., 2019. Getting prices right in structural electricity market models. *Energy Policy* 129, 1190–1206. doi:[10.1016/j.enpol.2019.01.077](https://doi.org/10.1016/j.enpol.2019.01.077).

Weron, R., 2014. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting* 30, 1030–1081. doi:[10.1016/j.ijforecast.2014.08.008](https://doi.org/10.1016/j.ijforecast.2014.08.008).

Ziel, F., Steinert, R., Husmann, S., 2015. Efficient modeling and forecasting of electricity spot prices. *Energy Economics* 47, 98–111. doi:[10.1016/j.eneco.2014.10.012](https://doi.org/10.1016/j.eneco.2014.10.012).