Chapter 1

Introduction

These days, life without electrical energy is almost inconceivable. Electricity networks play a key role in supplying customers with electrical energy and are organized hierarchically depending on their main task, as shown in Figure 1.1. Transmission grids are used to transport electricity over long distances from large power plants to distribution grids. They are operated at high voltage level (above 100 kV) to keep power losses low during transport. Distribution networks, located in cities or rural areas, distribute electricity to end customers. They are usually divided into medium-voltage and low-voltage networks. At the medium-voltage level, larger customers such as factories, smaller power plants, or large wind farms are usually connected to the power grid. Households are supplied at the low-voltage level, which in Europe is 400 V (line-to-line voltage). Both the transmission and distribution networks are usually operated by different companies that only loosely communicate with each other. For all voltage levels and grid types, the relationship between the load and generation is key aspect. While generation involves the production of electrical energy, load refers to the consumption of this energy, and generation must cover, but not exceed, the consumption, including network losses. This means that transmission and distribution companies may only feed as much electricity into the grid as can be consumed at any given time.

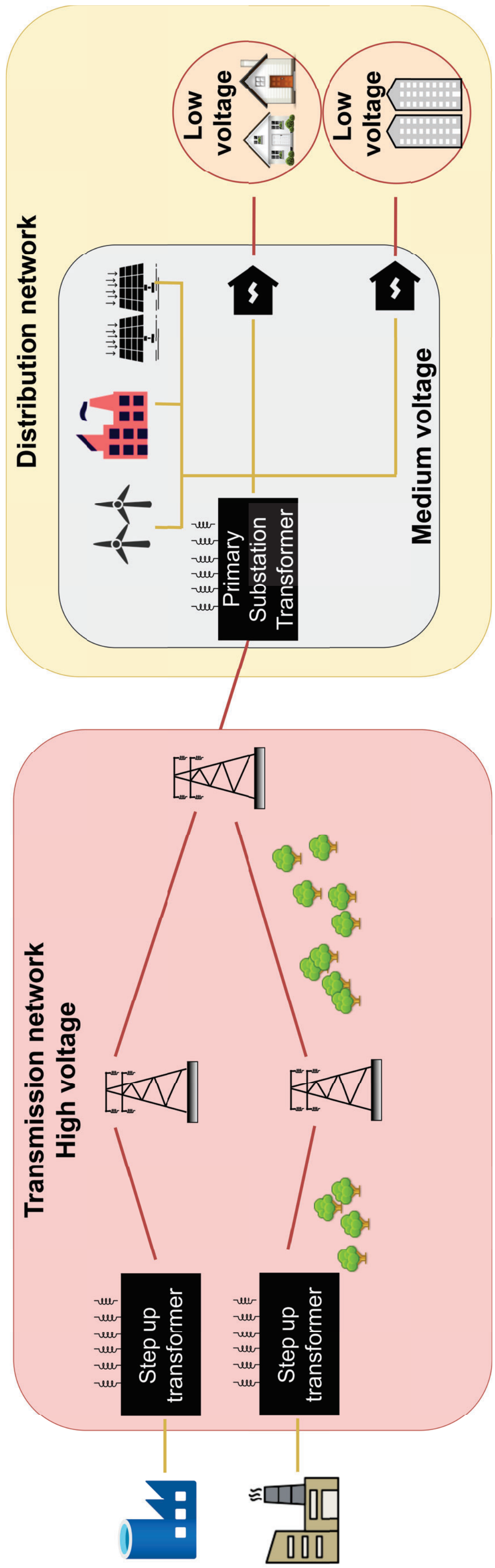

Figure 1.1: Hierarchical structure of power networks.

Unfortunately, since the storage of a significant amount of electrical energy needs further development, the generation must be balanced with consumption as overproduction of electrical energy can lead to equipment damage or emergency shutdowns. On the other hand, excessive consumption relative to generation results in under-voltage, malfunction of electrical equipment, and even failure of the supply. Therefore, knowing the actual and expected consumption is critical for utilities at all voltage levels to provide an adequate amount of generated power and, in this way, provide reliable service to the end customer. In real-time operation, this information is obtained from on-site measuring devices, and based on this information, utilities decide whether measures such as load shedding or generation curtailment are immediately necessary. For planning, information about expected consumption is obtained from load forecasts.

1.1 Basic terms

1.1.1 Time series definition and properties

Before examining electrical load forecasting in more detail, some terms related to time series should first be clarified. A time series is defined as a set of data collected at ordered points in time. If the time intervals are equal, the time series is called regular, whereas otherwise, it is irregular [Jos22]. The interval length, that is frequently used in the area of short-term load forecasting, is either 15 minutes or one hour and the time series considered in this work are regular.
Another important characteristic is the stationarity of the time series. If the distribution of a time series does not change over time, the time series is called stationary, otherwise, it is non-stationary [Jos22]. The time series discussed in this work are non-stationary.
Common measures such as the mean, median, standard deviation, distribution type etc. are also used to describe the characteristics of time series signals.
According to [HA18], many time series can be divided into three elements: trend and cycle, seasonality, and the reminder component. Trend refers to a long-term change in a time series, such as a steady increase or decrease over

time. Cycle describes repeating patterns in the time series that do not have a fixed frequency. Both trend and cycle are often combined into a trend-cycle component of the time series. Seasonality comes into play when the time series is affected by seasonal factors such as the day of the year, the day of the week, or the hour of the day, whereby the frequency is fixed and known in advance. The reminder component represents the random part of the time series. The time series value y at time instance t is composed from the three components, either additively or multiplicatively, as expressed in (1.1) and (1.2) [HA18] with S_t, T_t and R_t, which represent the seasonality, the trend and the reminder component at time instance t respectively. The additive model is used if the seasonal or trend-cycle components do not vary with the level of the time series [HA18]. In contrast, multiplicative decomposition is advised for time series for which the change in a seasonal or trend-cycle component is proportional to the level of the time series [HA18].

$$y_t = S_t + T_t + R_t \tag{1.1}$$

$$y_t = S_t \times T_t \times R_t \tag{1.2}$$

Pure random time series that can not be broken down into the above-mentioned components, such as stock data, can thus hardly be forecasted. Hence, the ability to decompose a time series into such components is a precondition for being able to provide meaningful forecasts of the time series. According to [HA18], the tools to detect the components within a time series are:

a) Plots of a time series

b) The calculation of the moving average T_t at time t for a trend-cycle component detection using the time series values k steps before and k steps after t.

$$T_t = \frac{1}{m} \sum_{j=-k}^{k} y_{t+j}, m = 2k+1 \tag{1.3}$$

c) Seasonal and trend decomposition based on Loess (STL) with the latter being a method for estimation of nonlinear relationships [HA18]. This approach is used for time series with additive composition. For time series with multiplicative characteristics, an application of the logarithm is recommended before applying STL.

1.1.2 Time series forecasting

Time series forecasting can be defined as the computation of future values of a time series based on the past inputs of the time series and optionally based on values from other time series that are correlated with the original time series. While the process is similar to prediction in machine learning (ML) or classification in pattern recognition, in the case of forecasting, the temporal aspect is important [Jos22]. Usually, forecasting consists of multiple time instances that are forecasted sequentially or simultaneously according to their temporal sequence [Jos22]. The external time series used for forecasting are often referred to as exogenous or independent variables, and the time series being forecasted is referred to as the target variable, dependent variable, or endogenous variable [Jos22]. Depending on the model defined, forecasts for multiple exogenous time series may be used. In addition, previously forecasted values of the target variable or previous values of exogenous variables are often used as input for forecasting the next time instance. This is shown schematically in Figure 1.2. The forecast window includes the forecast of four time points in the future with t_0 as the first forecast time point. To obtain the forecast value for t_0, the values of the exogenous variable from time points $[t_{-2}, t_0]$ and the values of the target variable at time points $[t_{-3}, t_{-1}]$ are used. After the forecast for t_0 is obtained, the forecast for t_1 is executed, whereby the values of the independent variable at time points $[t_{-1}, t_1]$ and the values of the dependent variable at time instances $[t_{-2}, t_0]$ are utilized with the previously obtained forecast value at t_0. The forecast stops when all the time instances within the forecast window have been processed. The various models use different types of exogenous variables, varied numbers of past time points and diverging sizes of the forecast window.

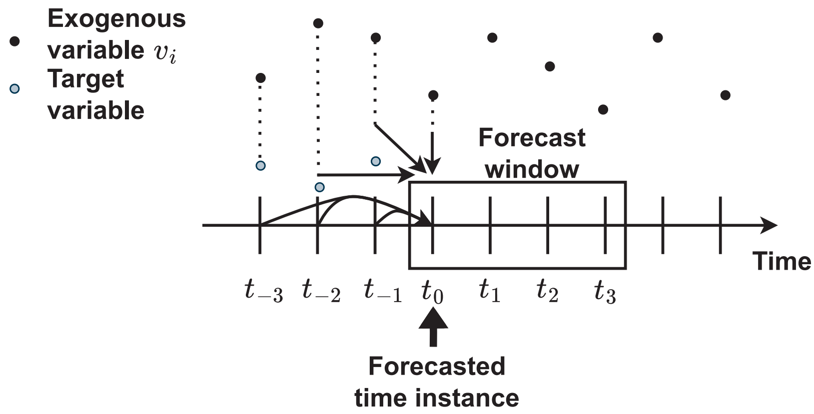

Figure 1.2: Time series forecasting principle.

1.1.3 Feature vector and classifier pipeline

Since a forecast algorithm can also be described in terms of a classifier that takes the exogenous and past (or already forecasted) values of the target variable as inputs and returns the class of target variable for a time instance t_c as output, some terms related to the definition of a feature vector and the pipeline for developing a classification algorithm are introduced. In pattern recognition, the feature vector describes the vector of random variables whose components in combination describe a unique pattern used to distinguish different classes of objects [TK09]. For l different features, the i-th feature vector is formed as described in (1.4), where T stands for transposition.

$$x_i = [x_{1,i}, x_{2,i}, .., x_{l,i}]^T \tag{1.4}$$

Each feature vector can be assigned to an object class via a classification algorithm (classifier). If the classes are known in advance, the feature vector is called a training feature vector [TK09]. Theodoridis and Koutroumbas [TK09] present the pipeline for developing a classification algorithm that can be directly used for designing a forecasting algorithm. The pipeline consists of six stages as shown in Figure 1.3:

- data collection from sensor data,
- extraction of numerical features from non-numerical raw data,

- data preprocessing including treatment of outliers, missing values and different data ranges
- selection of non-redundant features for which the classifier performs best
- design of the classifier,
- evaluation of the implemented classifier.

For the development of a load forecasting algorithm, all of these steps - except feature extraction - are important. Since the available raw data for load forecasting is numerical or can easily be mapped to numerical values, no feature extraction step is included here and the feature vector can directly be created.

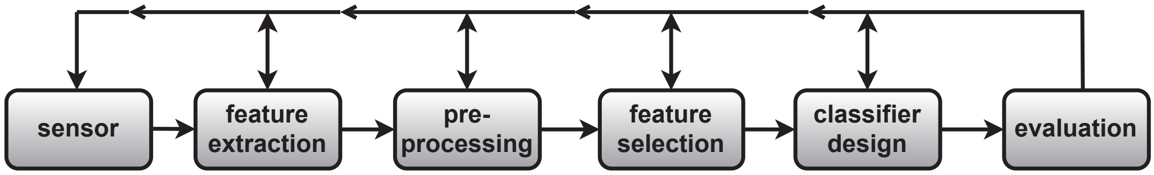

Figure 1.3: A pipeline for the development of a classification algorithm based on [TK09].

1.2 Problem statement

1.2.1 Load forecasting in power systems

From a transmission system perspective, the distribution system can be aggregated to a load or a generator (positive or negative injection), and the network structure downstream of the connection point can be disregarded. Therefore, forecasting the demand of the underlying distribution network is usually reduced to providing forecasts at the connection points to the distribution network.
At the transmission level, all consumption and generation points are measured in terms of active and reactive power, current magnitude, and bus voltage and transmitted to the grid control centers so that the required time series data is available for the forecast.
In the distribution grid, the substation, the lines branching off the substation called feeder heads, larger loads, or generators are measured while most parts

of the distribution network are not observed. Here, forecasts are also made at some level of aggregation, for example, at the supplying transformer [HWS20], at the feeder heads [Sun+16], or for larger wind farms, where mostly active and reactive power measurements or current magnitudes are available.

At both the transmission and distribution levels, load forecasts are usually calculated for active power values, and the most important input for the forecast is the historical load data obtained from measurements. These values have a high correlation between loads in subsequent hours. Frequently, the days that lie 24 hours or a week before the forecast day form the set of input variables [RB12]. In addition to the corresponding historical data from field measurements, calendar data relating to the type of day (holiday or working day, day of the week, season) is also included as an independent variable in the forecast models because electricity consumption changes with weekly or seasonally recurring patterns of end-user behavior. Furthermore, weather forecasts such as ambient temperature, humidity, wind speed, or radiation are also often used as independent variables [PHK92; ZYC17], whereby the relationship between load and weather factors can often be non-linear in the case of high or low temperatures [FC06].

1.2.2 Forecasting horizons

Depending on the task the forecast is prepared for, a distinction is made between short, medium, and long-term load forecasts. Typical horizons for short-term forecasts range from a few hours to a week in advance, whereby time intervals of 24 or 48 hours are very common. The forecast frequency ranges from one to four times per hour.

Short-term forecasts are mainly used for planning generation schedules (unit commitment) and economic dispatch (adjustment of the generation to meet the demand in case network congestions are expected). If the demand cannot be met, the utility can purchase or sell the difference on the energy market. For the European Transmission System Operators, for example, there is the European intraday and day-ahead market called EPEX SPOT. While energy markets

are well established on the transmission level, they still comprise an area of research in distribution networks.
The earlier and more accurately the utility knows its demand, the more efficiently it can reduce the cost of its supply. According to [Hon15], a utility with a yearly peak load of 1 GW can save between 300 k EUR and 600 k EUR per year if the forecast error is reduced only by 1 percent.
A medium-term forecast usually refers to time intervals from one week to one year in advance, whereas long-term forecasts are conducted for periods between one and several years ahead [SA10]. For the medium- and long-term forecasts, the forecast frequency decreases from once a day to once a week or month. Medium- and long-term forecasts help utilities plan future networks, estimate the amount of fuel required and design the power supply contracts [Don19; WS21].

1.2.3 Challenges in load forecasting in modern power grids

One of the most significant challenges for modern power grids is the increasing number of Renewable Energy Sources (RES) or Distributed Energy Sources (DER). In its report [NYI17a], the New York Independent System Operator (NYISO) outlines the expected impact of DERs on its power grid as shown in Figures 1.4 and 1.5. The installation of a large number of DERs drives the transformation of the hierarchical network structure (Figure 1.4) into a decentralized structure that is similar to the Internet (Figure 1.5). Although the transformation process takes time, in the near future, there will be no unidirectional flow of generated power through transmission network towards the load placed in the distribution network. Electric power will flow in different directions across the power network with the result that there will be more dynamics at different voltage levels. This reversed power flow can already be observed in distribution networks containing high percentage of RESs such as the network of the German utility EWE-Netz [HWS20].

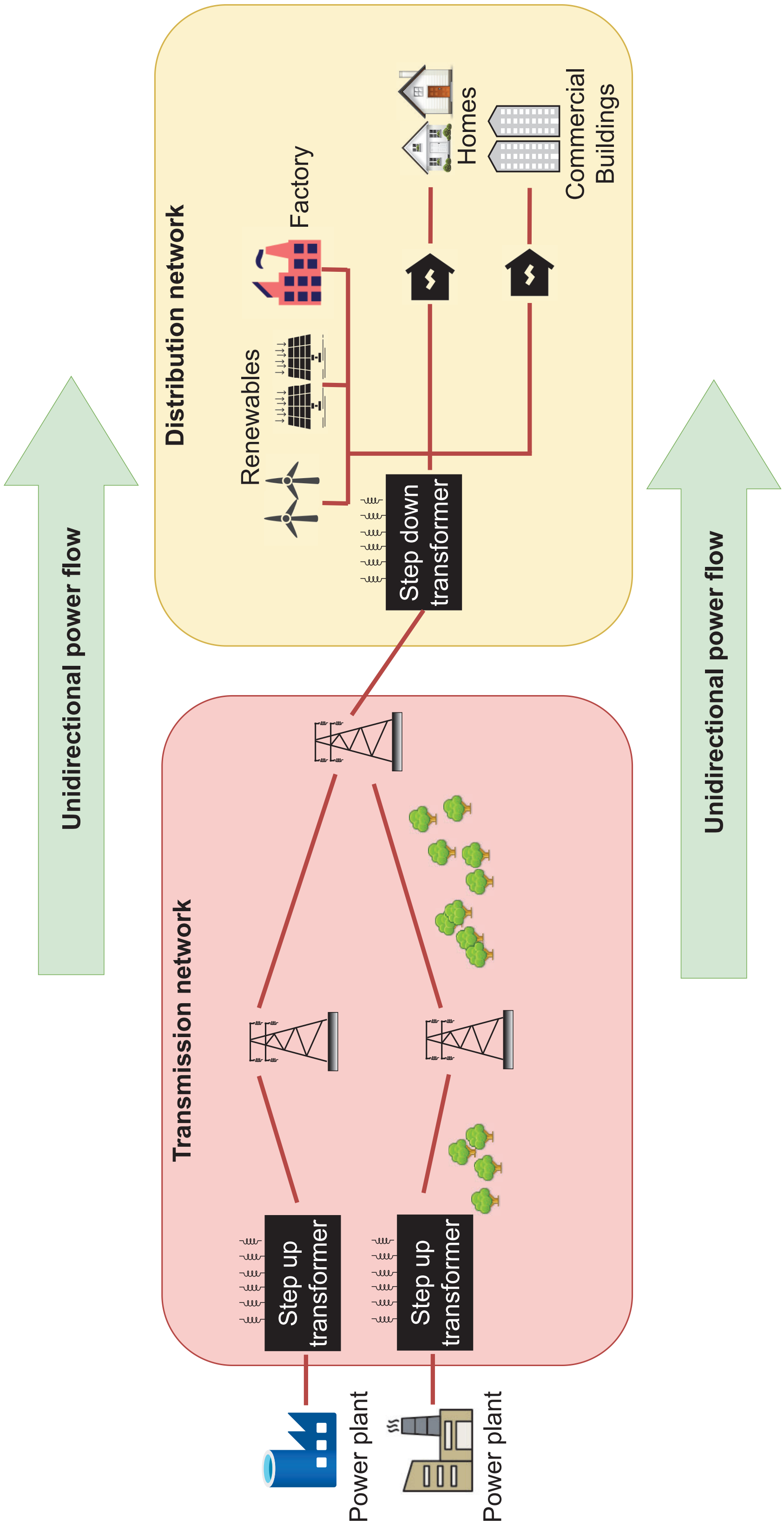

Figure 1.4: Current grid structure based on [NYI17a].