

Chapter 2

Renewable Energy Sources—Modeling and Forecasting

2.1 Introduction

Forecasts are essential to the integration of renewable power generation in electricity markets operations, since markets ought to be cleared in advance, while market participants shall then make decisions even before that. This is true for all types of electricity markets, that is, from real-time to futures markets, via the more classical day-ahead (forward) ones. For the reference case of conventional generators, power production forecasts are straightforward since, except for unit failures, one actually controls future electricity generation. In such a case, forecasts directly consist of potential schedules, which then translate to supply offers in the market. When it comes to renewable power generation, one is mostly left with Nature deciding on the future schedule of the power plants: wind power is only there when the wind blows and solar energy when the sun shines. Only hydro power is more dispatchable as the water originating from rainfall and snow melt can be stored in gigantic reservoirs. The nonstorability of other types of renewable energy sources, at least in a technologically and economically efficient manner today, magnifies this need for appropriate forecasts of renewable power generation. Here emphasis will be mainly placed on wind energy, which has so far been the leading form of renewable energy. The ideas and concepts presented could be extended to the case of, e.g., solar and wave energy since, from a mathematical point of view, the modeling and forecasting problems share a high level of similarity. Solar energy is becoming increasingly popular and present in a number of countries like Spain and Germany, among others. Wave energy is finally envisaged to become a natural complement to wind energy in the offshore energy mix, based on a number of demonstration projects today in the UK and Portugal, for instance.

It is sometimes argued that forecasts are there mainly to comfort decision-makers—here, the market and network operators, as well as power producers, and potentially end-consumers—while they are not really used or at least not used in an optimal manner in daily operations. However, employing the appropriate forecasts in a well-defined decision-making problem can tremendously improve the decisions to be made, while allowing controlling the risk brought in by unforeseen events. Indeed, a crucial starting point of this chapter is that forecasts are always wrong to

a certain extent. This should be accounted for in the various operational problems considered.

All aspects of renewable power forecasting cannot be covered within a single chapter of this book, nor can the necessary theoretical background on, e.g., stochastic processes, modeling, and estimation. Forecasting of renewable power generation relies on cross-disciplinary approaches taking roots in mathematics, statistics, meteorology, and power systems engineering. Most importantly, we aim at discussing here the various types of forecasts that exist for renewable power generation, being wind, solar or wave energy, and that are to be used as input to operational problems for electricity markets.

In Sect. 2.2, we introduce some of the necessary notation and definitions while placing ourselves in a stochastic process modeling framework. Necessary concepts related to stochastic processes are further developed in Appendix A. Subsequently, the various types of renewable energy forecasts that may be issued as input to decision-making problems are introduced in Sect. 2.3 based on examples, giving a pragmatic view of their characteristics. Emphasis is then placed in Sect. 2.4 on the quality of these forecasts, by covering their required and necessary properties, as well as some key scores and diagnostic tools for their evaluation. It is of utmost importance to fully appraise the quality of forecasts before to use them as input to decision-making and general operational problems. The way these forecasts may be generated from various sets of input data is then discussed in Sect. 2.5. Further readings are suggested at the end of this chapter.

2.2 Renewable Power Generation as a Stochastic Process

Even though referring to either renewable energy or power modeling and forecasting, focus is always placed on the power variable. This is because it is actually power which is measured at renewable energy generation plants. It is then straightforward to obtain energy values for given periods of time if necessary, by integrating power observations over these time periods.

Owing to the combination of a large number of complex physical processes, also mixed with additional uncertainties in our understanding of these processes, there may always be a part of randomness in our knowledge of energy generation from renewable energy sources. For instance, for a wind farm, even if having a perfect picture of the theoretical power curve of each and every turbine (as provided by the turbine manufacturer), it is close to impossible to know for sure what the power curve of the wind farm composed by all these turbines may be. This uncertainty originates from shadowing effects among the set of turbines, turbulence effects, dust and insects on the blades, etc.

Accepting the fact that there are uncertainties in the process of renewable energy generation, it is hence considered as a stochastic process. Necessary basics related to the definition of stochastic processes are introduced in Appendix A. Consequently, power generation from renewable energy sources, such as wind and solar, will be referred to as *stochastic power generation* in the subsequent chapters.

Definition 2.1 (The Renewable Energy Generation Stochastic Process). In the most general case,

$$\{Y_{r,s,t}, r = r_1, \dots, r_m, s = s_1, \dots, s_n, t = 1, \dots, T\}, \quad (2.1)$$

is a multivariate stochastic process in space and in time, observed at a set of n locations, $s = s_1, s_2, \dots, s_n$, and for successive time points $t = 1, \dots, T$, describing power generation from a number m of different renewable energy sources, $r = r_1, r_2, \dots, r_m$. The corresponding realizations of that stochastic process are denoted by

$$\{y_{r,s,t}, r = r_1, \dots, r_m, s = s_1, \dots, s_n, t = 1, \dots, T\}. \quad (2.2)$$

This stochastic process may be univariate ($m = 1$, in the above definition) if considering one type of renewable energy only, or multivariate ($m > 1$) if jointly considering several forms of renewable energy generation, as for the example of wind and wave energy generation offshore. In the former case, the notation for the stochastic process may be simplified to $\{Y_{(s,t)}\}$. Similarly, while both the time and space dimensions may be jointly considered, it is often the case that (i) focus is on the spatial dimension only, e.g., as input to a power flow calculation, or (ii) focus is on the time dimension only, e.g., if dealing with renewable energy generation for a given location in an optimal storage operation problem. Notation would then simplify even more, by using the relevant subscript only, that is, $\{Y_s\}$ and $\{Y_t\}$ for the space and time cases, respectively.

This stochastic process can be normalized for simplification, hence taking values between 0 and 1 at any time, any location and for all types of renewable energy,

$$Y_{r,s,t} \in [0, 1], \quad \forall i, s, t. \quad (2.3)$$

The above then also necessarily applies to all realizations $y_{r,s,t}$. The normalization is done individually by the nominal capacity of that type of renewable energy at this location and at this point in time. While it is fairly obvious that nominal capacity depends upon the renewable energy plant and therefore its location, one should not forget that nominal capacity can vary in time, e.g., due to maintenance planning and decommissioning/recommissioning of renewable energy assets.

The concepts introduced in the above are illustrated in the following example describing a univariate case, with wind power generation only, though observed at a number of locations, and for a long period of time.

Example 2.1 (Wind Power Generation for 15 Control Zones in Western Denmark)

A dataset with wind power generation over the control area (split into 15 control zones) of Western Denmark, operated by Energinet.dk for a total nominal capacity of 2.515 GW, will be used as a basis for illustration in this chapter. This control area is commonly referred to as DK-1. Wind power generation over these 15 control zones can be considered as a univariate stochastic process $\{Y_{s,t}\}$ in space and in time, in practice observed at 15 locations $s = s_1, s_2, \dots, s_{15}$ only. Figure 2.1 depicts an episode with two days in mid-February 2006 of wind power observations at these 15 control zones, with a hourly temporal resolution. These wind power observations for every zone are normalized by the relevant nominal capacity values. The generation

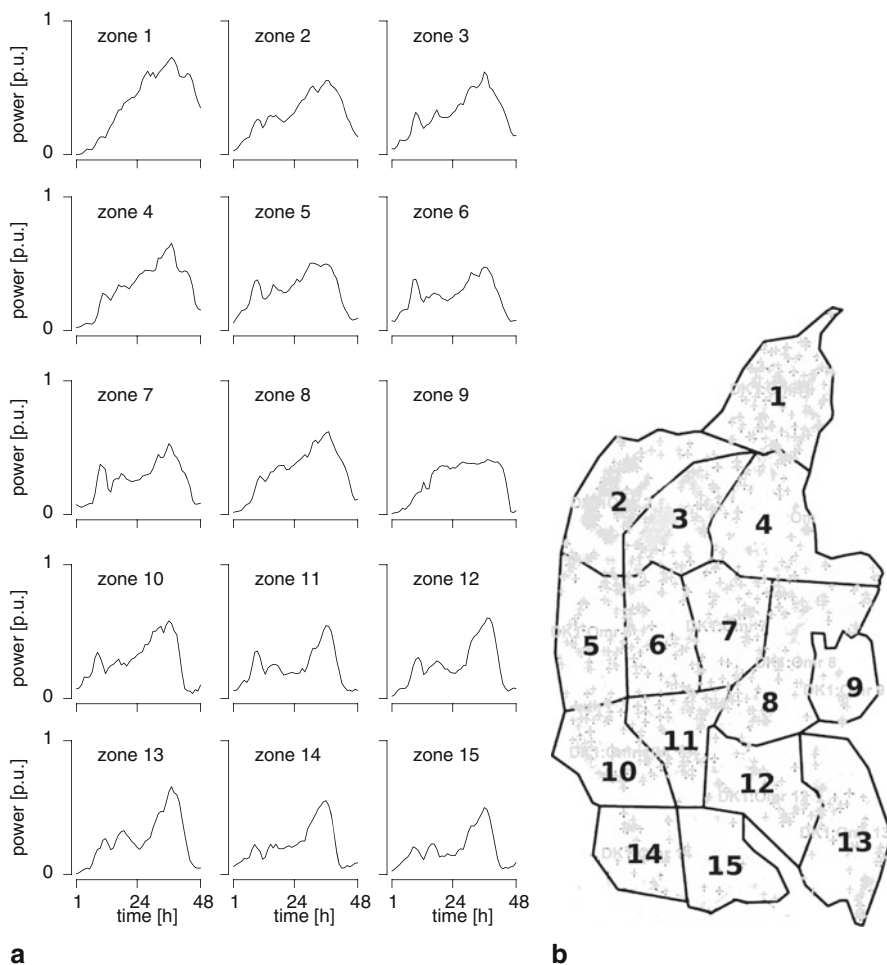


Fig. 2.1 Episode from mid-February 2006 with two days of wind power observations at the 15 control zones forming the control area DK-1 of Energinet.dk. These measurements have an hourly temporal resolution and are normalized by the respective nominal capacities at every control zone. **a** Normalized power observations. **b** 15 control zones

patterns for neighboring zones have similar characteristics, while there is also a clear temporal dependence. These are important aspects when it comes to the modeling and forecasting of such a stochastic process.

2.3 The Various Types of Renewable Power Forecasts

Predictions of renewable energy generation can be obtained and presented in a number of different manners. The choice for the type of forecasts and their presentation somewhat depends upon the process characteristics of interest to the decision-maker,

and also upon the type of operational problem. For instance, a wind farm operator aiming to plan maintenance over the coming week may only be interested in simple deterministic-type of forecasts for wind and power generation at the level of this wind farm, and not in detailed space–time scenarios over the whole country.

The various types of renewable energy forecasts and their presentation are introduced below, starting from the most common point forecasts and building up towards the more advanced products that are probabilistic forecasts and scenarios. We finally mention some of the more exotic forecasts that are currently being issued with focus on predefined events.

2.3.1 Common Features of Renewable Power Forecasts

Forecasting is about foreseeing the future state of the process of interest, in this case, renewable energy generation, at a given location s or for a set of n locations $s = s_1, s_2, \dots, s_n$, potentially with different forms of renewable energy sources at every location. Even though several locations and renewable energy forms may be considered, it is the temporal dimension that is of importance here. In contrast to spatial forecasts, we do not aim in this chapter at predicting the dynamics of the stochastic process at new locations. We do not attempt at issuing forecasts for new types of renewable energy sources either. The set of locations s and the energy mix are both fixed. Let us then place ourselves at time t and look at a future point in time $t + k$. For ease of notation, we only use time indices in the following when referring to values for the stochastic process. One should not forget that these may also be for several locations and types of renewable energy sources.

Emphasis is placed in the following on model-based approaches to forecasting. There exists a number of other approaches, e.g., based on expert judgments. For the example case of forecasting the electric demand (commonly referred to as load), it is often said that such expert judgments are very difficult to outperform by any model-based approach. For renewable energy forecasting, however, model-based approaches are to be preferred, since it would be much more difficult for experts to sharply foresee weather developments and their impact on corresponding renewable energy generation. Note that a difference should be made between a model, which comprises a mathematical representation of the processes considered, and a forecasting method, which is, instead, the process of issuing a prediction, based or not on a model.

Definition 2.2. A (model-based) forecast $\hat{\cdot}_{t+k|t}$ of renewable power generation is an estimate of some of the characteristics of the stochastic process Y_{t+k} (where Y is for all locations and types of renewable energy sources) given a chosen model g , an estimated set of parameters $\hat{\Theta}_t$ and the information set Ω_t gathering all data and knowledge about the processes of interest up to time t . That information set is commonly employed to identify a model g and the set of parameters Θ_t .

In the above definition, k is the *lead time*, though sometimes also referred to as *forecast horizon*. The ‘hat’ symbol expresses that $\hat{\cdot}_{t+k|t}$ is an estimate only: it reflects

the presence of uncertainty both in our knowledge of the process and inherent to the process itself. The notation ' $t + k|t$ ' is based on the conditional symbol ' $|$ ' in probability theory. The forecast for time $t + k$ is conditional on our knowledge of the stochastic process up to time t , including the data used as input to the forecasting process, as well as the models identified and parameters estimated.

Whatever the type of forecast, forecasting is to be seen as a form of extrapolation. A model is built and fitted to a set of data, then applied for prediction purposes on totally new data. This conditionality of forecasts makes that they should implicitly be formulated as: “*given the information set and assuming that the identified dynamics continue in the future, we can predict that . . .*”. A forecaster somewhat makes the crucial assumption that the future will be like the past.

Forecasts are issued as series of consecutive values $\hat{y}_{t+k|t}$, $k = 1, 2, \dots, K$, that is, for regularly spaced lead times up to the *forecast length* K . That regular spacing is called the *temporal resolution* of the forecasts. This will be illustrated when introducing the various types of renewable energy forecasts below. For instance, when one talks of 48-hour ahead forecasts with hourly resolution, this means that forecasts actually consist in forecast series gathering predicted power values for each of the following 48 h. Similarly, if predictions were to be issued on a regular spatial grid, one would talk of their spatial resolution. Here forecasts are for specific locations, not uniformly distributed on a grid, and therefore, the concept of spatial resolution does not make much sense.

2.3.2 Point Forecasts

When the renewable energy forecast issued at time t for $t + k$ is single-valued, it is referred to as a point prediction and denoted by $\hat{y}_{t+k|t}$. The fact this forecast is single-valued makes that point forecasts issued in a deterministic or stochastic process framework look similar. However, they are not in essence. In a deterministic framework, the forecaster is somewhat sure that the prediction ought to realize—there is no uncertainty involved. In a stochastic process framework, instead, $\hat{y}_{t+k|t}$ is an estimate only, hence acknowledging the presence of uncertainty.

Definition 2.3. A point forecast $\hat{y}_{t+k|t}$ corresponds to the conditional expectation of Y_{t+k} given g , $\hat{\Theta}$, and the information set Ω_t ,

$$\hat{y}_{t+k|t} = \mathbb{E}[Y_{t+k}|g, \Omega_t, \hat{\Theta}]. \quad (2.4)$$

In everyday words, the conditional expectation is the mean of all that may happen given our state of knowledge up to time t . Providing decision-makers with a forecast in the form of a conditional expectation translates to acknowledging the presence of uncertainty, even though it is not quantified and communicated.

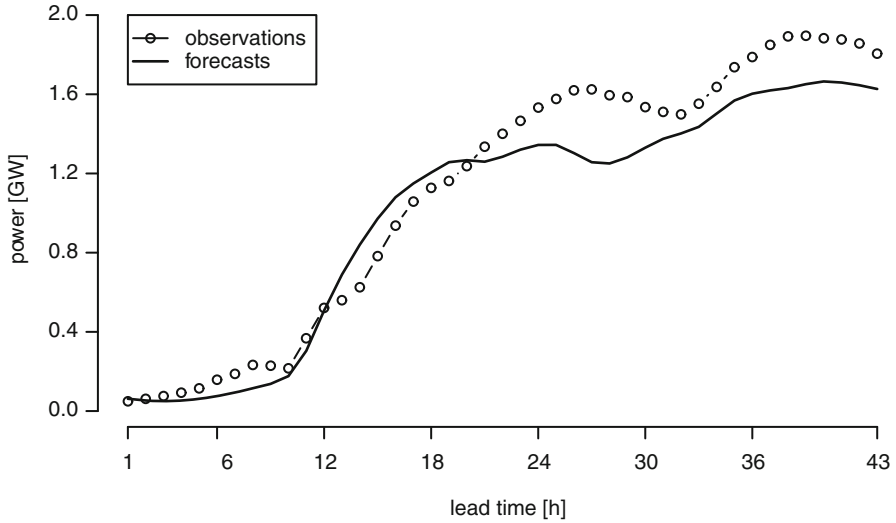


Fig. 2.2 Point forecasts of wind power generation issued on the 4th April 2007 at 00:00 UTC for the whole onshore capacity of Western Denmark (for a nominal capacity of 2.515 GW on that day)

Example 2.2 (Point Forecasts of Wind Power Generation) Let us consider the example of point forecasts issued on 4th April 2007 at 00:00 UTC¹ for the whole onshore capacity of Western Denmark (2.515 GW at the time, see Example 2.1, depicted in Fig. 2.2, along with the corresponding observations obtained a posteriori. This forecast series has a hourly temporal resolution up to 43 h ahead.

It informs that the expected power generation on 5th April 2007 at 00:00 UTC should be 1.32 GW. There what the forecaster really says is that the predicted mean of all potential power production values is 1.32 GW. He or she is not telling about what could really happen, however. The actual power generated 24 h after the forecast is issued could range anywhere between 0 and 2.5 GW, and that would make a big difference! This will all depend upon the forecaster's skill and the inherent forecast uncertainty. In this case, the forecast error made a posteriori appears fairly small, since the observed power generation at that time was of 1.466 GW (still a 146 MW difference).

2.3.3 Probabilistic Forecasts

This shortcoming of point predictions not giving the full picture about what *could* happen is of crucial importance when it comes to operational problems, where the costs potentially induced by the whole potential range of realizations that are likely

¹ UTC actually stands for *Coordinated Universal Time*, which is a time standard by which we regulate time and clocks.

to occur is to be accounted for. This has therefore motivated the substantial research effort invested in the development of probabilistic forecasting methodologies for energy applications, with a strong emphasis on their optimal integration in operations research problems.

In contrast to point predictions, probabilistic forecasts aim at providing decision-makers with the full information about potential future outcomes. Let us use the same notation as before while dropping out the subscripts for location and type of renewable energy source. Recall that y_t is the power production measured at time t and corresponds to a realization of the random variable Y_t . Then let f_t and F_t be the *probability density function* (abbreviated pdf) and related *cumulative distribution function* (abbreviated cdf) of Y_t , respectively.

Definition 2.4. A probabilistic forecast issued at time t for time $t + k$ consists in a prediction of the pdf (or equivalently, the cdf) of Y_{t+k} , or of some summary features.

Deterministic forecasts may be reinterpreted in a probabilistic framework as probability masses of 1 placed on these values predicted for the future state of the process—there is no uncertainty. The various types of probabilistic forecasts are detailed below, from quantile to density forecasts, and through prediction intervals.

2.3.3.1 Quantile Forecasts

Let us now introduce the concept of quantile forecast based on the definition of the quantile of a cumulative distribution function as given in Def. A.6 of Appendix A.

Definition 2.5. A quantile forecast $\hat{q}_{t+k|t}^{(\alpha)}$ with *nominal level* α is an estimate, issued at time t for lead time $t + k$, of the quantile $q_{t+k}^{(\alpha)}$ for the random variable Y_{t+k} , given a model g , its estimated parameters $\hat{\Theta}_t$ and the information set Ω_t , i.e.,

$$P[Y_{t+k} \leq \hat{q}_{t+k|t}^{(\alpha)} \mid g, \Omega_t, \hat{\Theta}] = \alpha. \quad (2.5)$$

By issuing a quantile forecast $\hat{q}_{t+k|t}^{(\alpha)}$, the forecaster tells at time t that there is a probability α that renewable energy generation will be less than $\hat{q}_{t+k|t}^{(\alpha)}$ at time $t + k$.

Quantile forecasts are of interest for a number of operational problems, since for a variety of loss functions (quantifying the cost of making a suboptimal decision, to be further introduced and discussed in Sect. 2.5.3), optimal decisions always relate to quantile forecasts with given nominal levels [2]. This is, for instance, the case for the design of optimal offering strategies by wind power producers, where optimal bids are quantile forecasts whose nominal level is a simple function of day-ahead and balancing market prices (see Chap. 7). Furthermore, quantile forecasts also define prediction intervals and, more generally, nonparametric probabilistic forecasts, as will be explained more extensively in the following. The concept of quantile forecasts is further illustrated below by building on the previous examples with wind power generation in Western Denmark.

Example 2.3 (Quantile Forecasts of Wind Power Generation) Fig. 2.3 depicts an example episode with quantile forecasts with a nominal level $\alpha = 0.5$ (i.e., the

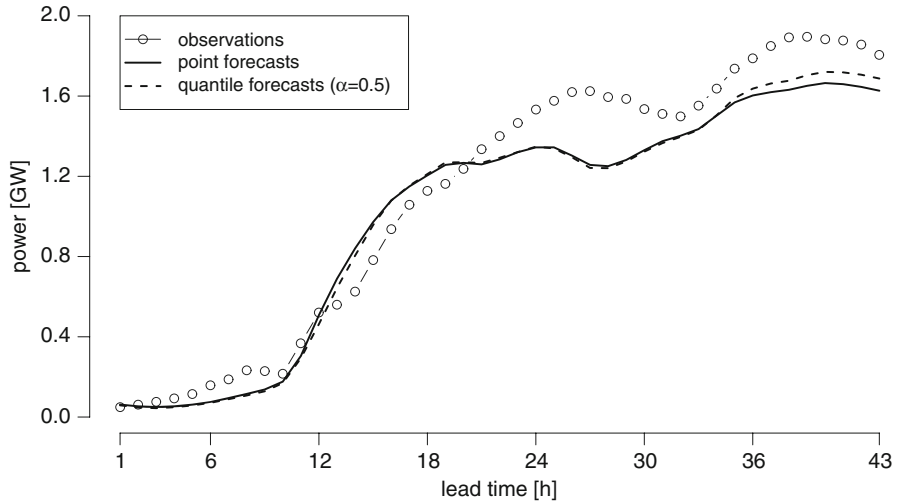


Fig. 2.3 Quantile forecasts of wind power generation with a nominal level of 0.5 (i.e., the median) issued on 4th April 2007 at 00:00 UTC for the whole onshore capacity of Western Denmark (for a nominal capacity of 2.515 GW on that day). Point forecasts and the corresponding observations are also shown

median), for the same period than in Fig. 2.2 and the same set-up as introduced in Example 2.1. For each lead time, these forecasts tell that wind power generation has a 50 % probability of being below (and, therefore, also above) the value they indicate. Their interpretation is hence quite different from that of the point forecasts considered before, since point forecasts, as conditional expectations, are not associated to any form of probability level. Note that, if forecast uncertainty is perfectly symmetric around point predictions, then $\hat{q}_{t+k|t}^{(0.5)} = \hat{y}_{t+k|t}$.

In the present case, if looking more closely at the 42-hour ahead lead time, while the previously discussed point forecasts tell that the expected power generation is 1.646 GW, the quantile forecast informs there is a 50 % probability that power generation will be below (or above) 1.706 GW.

2.3.3.2 Prediction Intervals

Quantile forecasts give a probabilistic information about future renewable power generation, in the form of a threshold level associated with a probability. Even though they may be of direct use for a number of operational problems, they cannot provide forecast users with a feeling about the level of forecast uncertainty for the coming period. For that purpose, *prediction intervals* certainly are the most relevant type of forecasts. Furthermore, prediction intervals are frequently used to make decisions under uncertainty using robust optimization (see, e.g., Chaps. 8 and 9).

Definition 2.6. A prediction interval $\hat{I}_{t+k|t}^{(\beta)}$, issued at time t for time $t + k$, defines a range of potential values for Y_{t+k} , for a certain level of probability $(1 - \beta)$, $\beta \in [0, 1]$, its nominal coverage rate,

$$P[Y_{t+k} \in \hat{I}_{t+k|t}^{(\beta)} \mid g, \Omega_t, \hat{\Theta}] = 1 - \beta. \quad (2.6)$$

It is equivalently referred to as an *interval forecast*.

Such an interval $\hat{I}_{t+k|t}^{(\beta)}$ must be defined by its lower and upper bounds,

$$\hat{I}_{t+k|t}^{(\beta)} = [\hat{q}_{t+k|t}^{(\alpha)}, \hat{q}_{t+k|t}^{(\bar{\alpha})}], \quad (2.7)$$

where these bounds are quantile forecasts whose nominal levels $\underline{\alpha}$ and $\bar{\alpha}$ verify that

$$\bar{\alpha} - \underline{\alpha} = 1 - \beta. \quad (2.8)$$

This general definition makes that a prediction interval is not uniquely defined by its nominal coverage rate. It is thus also necessary to decide on the way it should be centered on the probability density function. Commonly, it is chosen to center it (in probability) on the median, so that there is the same probability that an uncovered realization y_{t+k} lies below or above that interval. This translates to

$$\underline{\alpha} = 1 - \bar{\alpha} = \beta/2. \quad (2.9)$$

With this type of centering, the resulting intervals are called *central prediction intervals*. For example, central prediction intervals with a nominal coverage rate of 90 % (i.e., $(1 - \beta) = 0.9$) are defined by quantile forecasts with nominal levels of 5 and 95 %. Other types of intervals exist, e.g., shortest-length intervals and highest-density regions among others [5], depending upon the way they are chosen to summarize information from the full probabilistic distribution. An illustration is given below, for the case of wind power generation in Western Denmark.

Example 2.4 (Central Prediction Intervals of Wind Power Generation) Central prediction intervals of wind power generation with a nominal coverage rate of 90 % (i.e., $(1 - \beta) = 0.9$), issued for the whole onshore capacity of Western Denmark and for the same period than in Figs. 2.2 and 2.3, are depicted in Fig. 2.4. They give a range of possibilities of power generation for every lead time, for a certain probability level, and therefore tell about how confident one may be about the point forecasts originally provided—the tighter they are, the higher the confidence is. The advantage is that they give a very visual information on the expected range of future events.

In the present case, these intervals, for instance, inform that there is a 90 % probability that, 24 h in the future, wind power generation will be between 0.897 GW and 1.65 GW. There is only a 5 % probability that wind power generation will actually be less than 0.897 GW, and similarly, only a 5 % probability of being more than 1.65 GW.

Integrating Renewables in Electricity Markets
Operational Problems

Morales, J.M.; Conejo, A.J.; Madsen, H.; Pinson, P.;
Zugno, M.

2014, XVI, 429 p. 92 illus., 12 illus. in color., Hardcover
ISBN: 978-1-4614-9410-2