Università
Ca'Foscari
Venezia

Ca' Foscari
Dorsoduro 3246
30123 Venezia

Master's Degree programme – Second Cycle
(*D.M. 270/2004*)
in Economics and Finance

Final Thesis

Very short-term analysis of wind power generation in a probabilistic forecasting framework

Supervisor
Ch. Prof. CASARIN ROBERTO

Graduate
ZAETTA PAUL
872113

Academic Year
2017 / 2018

Preface

This is a Master thesis in time-series analysis as part of the study program in Economics and Finance at the University of Ca' Foscari in Venice. This Master thesis is achieved within the framework of the double degree in Economics, Econometrics and Finance (MEEF) resulting from a partnership between the University of Venice and Aix-Marseille School of Economics. The thesis was carried out from April to mid June of 2018.

The Master thesis assumes the reader is familiar with notions and common methods within statistics.

Acknowledgements

Abstract

Nowadays, generating very-short term energy power forecasts is a crucial challenge. In particular, wind generation, which exhibits large fluctuations, is not easy to predict. This study is based on a probabilistic forecasting framework and ought to account for the nonlinear and double-bounded nature of that stochastic process. Discrete and continuous mixtures of generalised logit-Normal distributions and probability masses at the bounds serve to provide probabilistic forecasts. Pinson (2012) showed that this framework is superior to classical models for wind power production, which assume that the shape of predictive densities follow (censored) Normal and Beta distributions. Both simple autoregressive and autoregressive moving average models are designed in order to estimate the location and the scale parameters. The first aim of this study is to extend the Pinson (2012) model by introducing a dynamic structure for the location of the wind generation. The second aim is to analyse the predictive ability of the proposed model. The theory approach concerning the different methods is illustrated by assessment and ranking of probabilistic forecasts of wind generation at Galicia in the Spain Northwest (on 10-minute ahead point).

Keyword: bounded times-series; cross-validation; density forecast; dynamic models; predictive distribution; transformation; wind power

Contents

Chapter I

Introduction

One of the most important issues of the 21st century concerns energy production. Indeed, for quite some time we have understood the alarming situation that our energy production system is not stationary and that in face of exponential population growth and to a lesser extent the climate change, the situation will not remain stable for long. Consequently, we have to change our energy production mode in order to have a stationary system. We must simply remember that our energy production system is a current and unavoidable global challenge.

The energy sector is very complex from several reasons. One of its complexities is the highly technical process and also that each industry has his specific features. The energy is a crucial point in economic development of countries. When the electricity collapses in a country, it generates many economic losses. We can measure these losses by the gross domestic product (GDP) of the county multiplied by the time of the collapse. As a result, an efficient electrical network is essential to preserve economic growth.

Like any branch of economics, energy economics is concerned with the basis issue of allocating scarce resources in the economy. Thus, the micro-economic concerns of energy supply and demand and the macro-economic concerns of investment, financing and economic linkages with the rest of the economy form an essential part of a complete energy power analysis. The issues facing the energy industry change, bringing new issues to the fore. In the 1970s, the focus was on understanding the energy industry and especially the oil industry, energy substitution and to some extent on renewable energies.

Nowadays, renewable energies have a complete role in energy production. Its involvement in our energy production system will increase even more in the future. We can quote for instance Gneiting *et al., 2007* saying 10 years ago "*Wind power is the fastest growing source of energy today*". Indeed, global trends are towards renewable energies for many years. Recently this being formalized by the Paris Agreement in April 2016, which is a global agreement signed. This Agreement aims to respond to mitigate global warming through the reduction of high-carbon energies. Therefore, so-called low-carbon energies must be adopted in our energy production system. The renewable energies and nuclear

energy being considered low-carbon, thus these two energy sources solve this previous issue. For some external reasons only the renewable energies are desirable for the future. The chart below allows us to illustrate the situation about the current energy production and future production prospects concerning the European Union (see Figure 1). The Union aims to increase by more than 10% its proportion of renewable energy in its gross final energy consumption over the next ten years. This represents a very important market share for renewable energies.
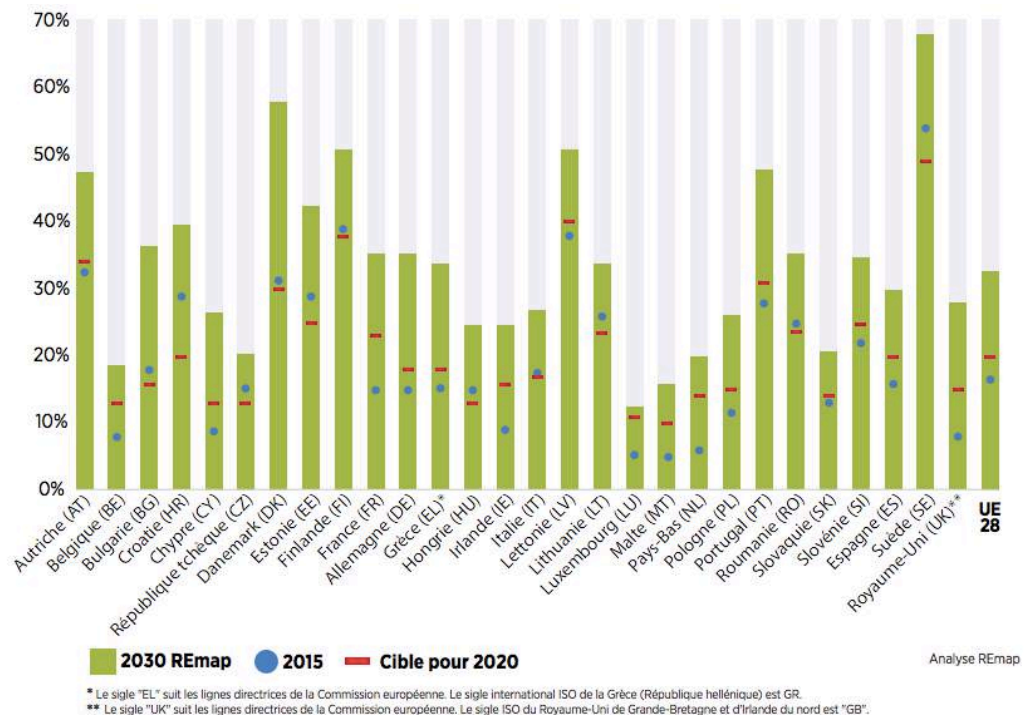
Figure 1. Proportion of renewable energy in gross final European energy consumption - 2015, 2020 target and 2030 potential through accelerated energy adoption renewable, this chart comes from the International Renewable Energy Agency[1].

Another point to illustrative purposes is the Portugal's renewable electricity production, which is a very interesting case due to the central role of the wind production in the Portugal's economy. For instance, in March 2018 its power generation from renewable sources, especially from wind power generation, exceeded monthly consumption according to the nation's transmission operator. Its average renewable generation for the month exceeded 103% of consumption. However, this source of energy (especially from wind and water) is neither deterministic nor constant.

(1) Available at: http://www.irena.org/publications/2018/Feb/Renewable-energy-prospects-for-the-EU

Therefore, at some moment the country need to fossil fuel and also imported electricity to even out supply. Consequently, as mentioned by Gneiting *et al.*, 2007, some arguments against the proliferation of wind energy have been discussed because of the inability to predict wind resources with any degree of precision. But partially, the developments of advanced probabilistic prediction methodologies help to address these concerns.

The electrical energy is not a commodity like any other. Indeed, an important feature is that it cannot be stored and the response to the demand must be instantaneous. Consequently, the market must be managed on a per second basis. There are major economic consequences in the event of a supply disruption. An important characteristic about the renewable energy is that they provide the lowest market prices; it outperforms the nuclear power, which is also energy low-carbon. Nonetheless, the renewable energies from wind power and solar power in particular are extremely volatile compared to other types of energy sources. The electric market depends only on the supply and demand; hence this market is very sensible. For instance, it is possible to have a negative price when the production is significantly higher than the demand. Indeed, the producers may have to pay so that the energy produced can be consumed. It makes this market unique.

As a result, the renewable energy is characterized on the one hand by a low price, low-carbon and stationary energy source, in the sense that this resource is inexhaustible, and on the other hand by the fact that its part in the energy production system is constantly increasing. A direct consequence of this is that tomorrow's energy production will be more volatile than today's. This resulting uncertainty is a real problem. Information is crucial for any decision-making: be it development planning, decisions or business decisions or decisions by individual consumers. Because of all this, the new issues to the fore are related to forecast the future power generation provided by renewable energies. Consequently, the quality of the forecasts concerning the energy produced by these renewable energies is a crucial issue. The quality of the electricity networks, the economic life, prices and climate problems depend on a large part of these forecasts.

This study aims to forecast the wind power generation in the very short-term, more precisely on 10-minute ahead. The main reason to focus on the wind energy is, as mentioned before, that it is one of the most volatile sources of energy. Furthermore, it has been showed that for such lead times statistical approaches are known to outperform the Running Numerical Weather Prediction (NWP) models when forecasts are based on short-term temporal resolutions (Giebel *et al.,* 2011). In addition, the NWP models for these temporal resolutions (very short-term) with frequent updates would be clearly to expansive if not impossible today.

In order to have an efficient electrical network, it is crucial to have some 'robust' forecasts for horizons of a few minutes ahead. Characterizing and modelling the power fluctuation at these time scales is known as a current challenge.

The analysis is focused on a Spanish wind farm, which is the site of Galicia in the northwest region of Spain. In particular, this country is one of the European countries that use wind energy the most. It is the second European country in term of production. This aspect directly motivated the choice to focus on this site.

All forecasts are associated with some sort of uncertainty and probabilistic forecasts are often preferred over deterministic forecast to better express this uncertainty. This is the reason why this study places us in a parametric probabilistic forecasting framework. The probabilistic forecast takes the form of a probability density function, which is denoted by PDF or a cumulative density function, which is denoted by CDF (Gneiting *et al.*, 2007). The way a probabilistic forecast represents uncertainty allows more nuanced decision-making. As a result, probabilistic forecasts have seen increased impact in many applications, such as economics, meteorology and climatology (Gneiting and Ranjan, 2013). The aim of a probabilistic forecast is to maximise the sharpness subject to calibration. Calibration refers to the statistical consistency between the probabilistic forecast and the observations, whereas sharpness refers to the concentration of the predictive distributions (Gneiting *et al.*, 2007). If we declare an interval or event to have probability p, the forecast is calibrated if the event happens a proportion p of the time on average (Raftery *et al.*, 2005). The continuous ranked probability score (CRPS) combines both sharpness and calibration to evaluate the performance of a forecast. The score is minimized when the probabilistic forecast is identical to the distribution of the process we want to forecast, and is thus a proper scoring rule (Gneiting *et al.*, 2007).

The main objective of this study is twofold. On the one hand, it is to extend the Pinson (2012) model by introducing a dynamic structure for the location of the wind generation. On the other hand, it is to analyse the predictive ability of the proposed model within a parametric probabilistic forecasting framework.

The central characteristic of the wind power variable is that it is double-bounded in the sense that it can be between a minimum production, which is zero and a maximum production, which is the nominal capacity. This second value is related either to the wind turbine or wind farm or to the energy portfolio considered, therefore depending to your framework. Here, the nominal capacity represents the maximum production of the wind farm at Galicia. To formalize this feature the wind power measurements and forecasts are normalized by the nominal capacity, which is denoted by P_n, and then the outputs take values in the

unit interval [0, 1]. Moreover, we note that the wind power generation is a nonlinear function of the wind speed, in the form of a sigmoid according to Pinson (2012). As a result, these characteristics show that the predictive densities in this case cannot be Gaussian for any temporal resolution of wind power time-series or whatever the forecast horizon considered. In order to illustrate the purposes, the next figure (see Figure 2) depicts 4 days of wind power measurements at the Galicia wind farm (1152 time-steps, with a temporal resolution of 10 minutes).
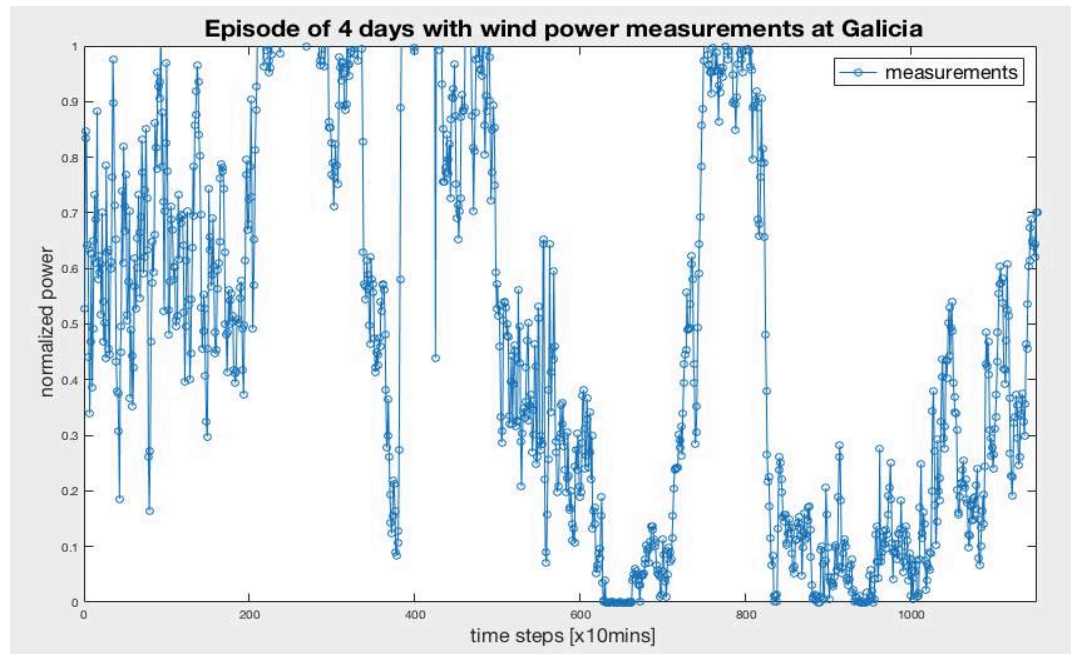
Figure 2. Episode of 4 days (1152 time-steps, with a temporal resolution of 10 minutes) with wind power measurements at the Galicia farm in Spain.

A consistent probabilistic forecasting of wind power time-series has to take into account all the characteristics mentioned here above.

The master's thesis is organized as follows. In a first chapter, a suitable distribution is introduced in order to represent the predictive densities of wind power generation. This distribution is a generalised logit-Normal distribution inspired by Pinson (2012). The second chapter concerns definitions and background about the distributions, which are required for the continuation of the study. The following chapter carries out the forecasting methodology concerning the adaptive predictive density based on the generalised logit-Normal distribution. The chapter 4 presents the different dynamic models for the shape and location parameters. The

chapter 5 puts forward all the methods concerning the evaluation of forecasts i.e. on the one hand the evaluation of point forecasts and the evaluation of density forecasts and on the other hand the cross-validation exercise. Following that, the chapter 6 is assigned to the probabilistic forecasting of wind power generation at lead-time of 10 minutes at the Galicia wind farm in Spain. The results obtained are compared between themselves and also against some benchmarks. To conclude, the last chapter summarizes the results and discusses possible extensions.

Chapter II

Definitions and Background

In this chapter we introduce the key assumptions of this work about wind generation distribution and data transformation.

The normal distribution has played a major role in theoretical and applied statistics since Laplace's time. Nonetheless, it was obvious that the normal distribution could not correctly represent all distributions in statistical domain. At the end of the 19th century, many mathematicians then attempted to construct systems of frequency curves in order to provide a wider variety of distributions than a normal curve. Thus, on the basis of many works done in the past, we could have a better distribution than the normal distribution in order to model the wind power generation.

A good model should account for some well-known feature of wind data. As mentioned in the introduction, an essential feature about wind power variable is that it is double-bounded in the sense that this variable can take value between zero (a null production) and its nominal capacity (the maximum production). We can denote the nominal capacity by P_n. It can depend of the framework, in the sense that if the study is applied on a special wind turbine or a general wind farm or also on a wind portfolio energy. In our case study the nominal capacity is considered on the Galicia wind farm in Spain. Moreover, as mentioned by Pinson (2012), the wind power variable is characterised by a nonlinear function of the wind speed. In addition, the wind power measurements and forecasts are normalised by the nominal capacity (P_n). As a result, variables can take the values between the range [0,1]. Thus, in the rest of this work the wind power measurements will always be considered as standardised. These various characteristics about modelling the wind power generation demonstrate that predictive densities cannot be Gaussian. Furthermore, this crucial consequence is true for any temporal resolution of wind power times-series and for any forecast horizon considered. In particular, Pinson (2012) showed the superiority of the generalised logit-Normal distribution over the classical distributions for wind generation modelling, which are the (censored) Normal and Beta distributions.

At present, an introduction of the background theory has to be demonstrated concerning the generalised logit-Normal distribution. That is necessary in order to develop the forecasting methodology, which is discussed in the next chapter. The starting point of this section is the

concept of "method of translation", which called also "transformation" in the literature. Many authors have covered this issue, in particular Johnson (1949). According to previous works, the purpose of using a transformation of a variable is to obtain a transformed variable, which is approximately normally distributed. As a result, this method is regularly used for variance stabilization and also to work in a Gaussian framework.

The following demonstration is inspired by the work of Johnson (1949) and Mead (1965). The general form of transformation can be defined by $z = f(x)$ where z is a standard normal variable will be referred to as the transformation of x, which is the observed variable. It has been shown that the introduction of four parameters associated to the possible functions $f(x)$ is convenient. These functions can be considered as being normally distributed. Let us resume the Mead's generalisation of one of the transformations discussed by Johnson. This generalisation as the following form

$$z = \gamma + \delta \, log\left(\frac{x - \xi}{\xi + \lambda - x}\right), \quad (\xi < x < \xi + \lambda) \tag{1}$$

where ξ and $(\xi + \lambda)$ are the lower and higher limits of x, respectively. In a variety of setting the bottom limit of x is zero. Consequently, we can rewrite the previous equation as follows,

$$z = \gamma + \delta \, log\left(\frac{x}{U - x}\right) \tag{2}$$

where x is within the range [0,U]. Following Nelder (1961) the generalised transformation can be obtained by adding a power parameter, θ. The complete transformation is defined as follows,

$$z = \gamma + \delta \, log\left(\frac{x^\theta}{A^\theta - x^\theta}\right) \tag{3}$$

or

$$z = \gamma + \delta \, logit\left(\frac{x}{A}\right)^\theta. \tag{4}$$

The distribution of x related to this previous transformation corresponds to the generalised logit-Normal distribution, which can be denoted by GL-Normal distribution.

With regard to wind power an important feature concerns the different inflection of the power curves for low and high power values. Nonetheless, this previous transformation developed by Mead (1965) seems to be an appropriate transformation that takes this asymmetry into account.

We can apply this method on the normalised power measurements $\{x_t\}$, which can take values between 0 and 1. We assume that $\gamma = 0$ and $\delta = 1$. The generalised logit transform y_t is defined by

$$y_t = T(x_t; \theta) = log\left(\frac{x_t^\theta}{1 - x_t^\theta}\right), \quad \theta > 0, \quad x_t \in (0,1) \tag{5}$$

in parallel, we can note also the inverse transformation as follows,

$$x_t = T^{-1}(y_t; \theta) = \left\{1 + \frac{1}{exp(y_t)}\right\}^{-1/\theta}, \quad \theta > 0, \quad y_t \in \mathbb{R} \tag{6}$$

corresponding to the inverse generalised logit transformation (IGL). Note that the simple logistic transformation is derived from the generalised logistic transformation when the power parameter θ is equal to one. This parameter aims to influence the evolution of variance and skweness of these distributions as a function of their mean, as confirmed by Pinson (2012). As stated previously the purpose of using such a transformation is to be able to work whit a time-series $\{y_t\}$, which is assumed that conditional densities may be Gaussian.

Following the previous reasoning, if $Y \sim \mathcal{N}(\mu, \sigma^2)$ is a Gaussian variable over the interval $[-\infty, \infty]$, the transformed variable $X = T^{-1}(Y; \theta)$ follows a generalised logit-Normal (GL-Normal) distribution, $X \sim L_\theta(\mu, \sigma^2)$ over the interval $[0,1]$. From these assumptions and according to Mead (1965), the probability density function of the GL-Normal variable X is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}\left(\frac{\theta}{x(1 - x^\theta)}\right) exp\left\{-\frac{1}{2}\left[\frac{T(x; \theta) - \mu}{\sigma}\right]^2\right\}, \quad x \in (0,1). \tag{7}$$

Chapter III

Forecasting Methodology

Many authors such as Lange (2005) and Pinson (2006), have demonstrated that the standard deviation of the distribution of forecast errors is directly related to its conditional expectation. This specificity is independent of the period, horizon and forecast model. Consequently, the predictive densities related to the wind power production shouldn't ignore this feature. In particular, Pinson (2012) mentioned this issue and, explains that the use of a suitable GL-transform to the wind energy time-series mitigates the effect by assuming that the variance of the predictive densities of the transformed is independent of its mean. Taking into account these characteristics we assume the predictive density for the wind power generation X_{t+k} at time $t+k$ is

$$X_{t+k} \sim \omega_{t+k}^0 \delta_0 + (1 - \omega_{t+k}^0 - \omega_{t+k}^1) L_\theta(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_1 \qquad (8)$$

where δ_0 and δ_1 are Dirac delta functions at 0 and 1, respectively. Note that $\delta_a(x)$ takes value 1 if $x = a$ and 0 otherwise. They represent the potential concentration of probability mass at the bounds of the unit interval. In addition, to obtain a well-defined density, we assume that

$$\omega_{t+k}^0, \ \omega_{t+k}^1 \in [0,1]$$

and

$$\omega_{t+k}^0 + \omega_{t+k}^1 \in [0,1].$$

The introduction of this concentration of probability mass at the bounds is explained by the fact that the original variable, which is the normalised power measurement to be GL-transform, belongs to the unit interval. Furthermore, for this type of variable the bounds are regularly reached. To illustrate this characteristic; the next figure (see Figure 3) represents the histogram issuing of observed wind power production at the Galicia site.
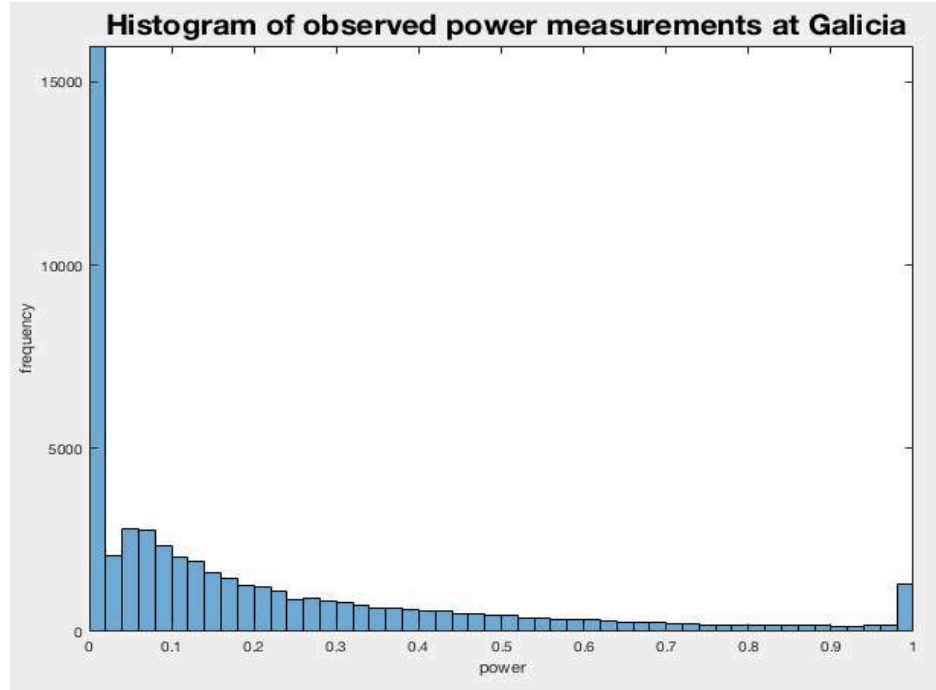
Figure 3. The histogram of observed wind energy production at Galicia in the northwest of Spain.

According to the previous chart the concentration of probability mass at these bounds over the unit interval make full sense. This representation reflects the fact that when the forecast takes value 0 or 1 (the nominal power), there is always some form of uncertainty.

From the predictive density of the process X_{t+k}, we can deduce the predictive density of its GL-transform, which is denoted by Y_{t+k}. Therefore, we can write

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{-\infty} + (1 - \omega_{t+k}^0 - \omega_{t+k}^1)\mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2) + \omega_{t+k}^1 \delta_{+\infty}. \tag{9}$$

In the same way that Pinson (2012) and Lesaffre *et al.*, (2007), for the modelling of outcome scores in [0,1] using classical logit-Normal distribution, a threshold value ϵ is defined, being in the order of measurement precision, such as $\epsilon \leq 10^{-3}$. Thus, wind power values below or equal to this threshold (respectively above or equal to $1 - \epsilon$) are considered to be null (respectively equal to 1). This allows a limitation of the range of potential variations of the GL-transformed variables, i.e. $y_t \in D_y$, where $D_y = [T(\epsilon; \theta), T(1 - \epsilon; \theta)]$. Thus, we can re-write the previous predictive density as follows

$$Y_{t+k} \sim \omega_{t+k}^0 \delta_{T(\epsilon;\theta)} + \mathcal{N}(\mu_{t+k}, \sigma_{t+k}^2)\mathbb{I}_{D_y} + \omega_{t+k}^1 \delta_{T(1-\epsilon;\theta)} \qquad (10)$$

where $\mathbb{I}_{D_y}$ is the indicator function for the set D_y, i.e.

$$\mathbb{I}_{D_y} = \begin{cases} 1, & y \in D_y \\ 0, & otherwise \end{cases}.$$

The weight $(1 - \omega_{t+k}^0 - \omega_{t+k}^1)$ disappears, for the reason that the predictive density for Y_{t+k} takes the form of censored Normal distribution instead. This allows for obtaining the both weights ω_{t+k}^0 and ω_{t+k}^1:

$$\omega_{t+k}^0 = \Phi\left(\frac{T(\epsilon;\theta) - \mu_{t+k}}{\sigma_{t+k}}\right), \quad \omega_{t+k}^1 = 1 - \Phi\left(\frac{T(1-\epsilon;\theta) - \mu_{t+k}}{\sigma_{t+k}}\right),$$

where Φ is the cumulative distribution function (CDF) of a standard Normal variable. In this way, predictive densities are fully represented by their location (μ_{t+k}) and scale (σ_{t+k}) parameters only.

The predictive density is illustrated with the next figure (see Figure 4), which represents the cumulative distribution function of wind power generation. This CDF is related to a zero wind power measurement. We can observe on the chart the concentration of probability mass on the left bound. Moreover, values approximately below -22 and above 5 are not achievable in real world, which represent the minimum and maximum production, respectively.
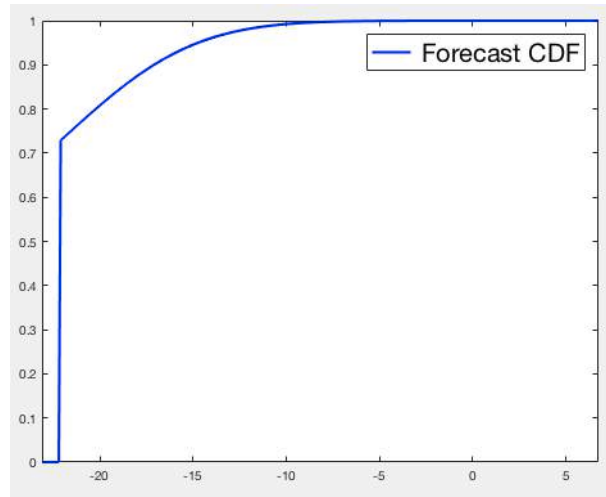
Figure 4. Forecast CDF of the predictive density of the process Y_{t+k}.

Chapter IV

Dynamic and Reference Models

In this chapter, different dynamic models are specified for the location and scale parameters. Both Autoregressive (AR) and Autoregressive Moving Average (ARMA) models are proposed for the location parameter. Adaptive estimation of the scale parameter is updated each period according to the dynamic model used. Several authors such as Gneiting *et al.*, (2006), Pinson and Madsen (2011), Sanchez (2006) or Vincent *et al.*, (2010) have shown the fact that parameters in dynamic models should be assumed slowly changing over time. The reason is that such a process should certainly vary at time scales of weeks, months, seasons and years. Consequently, the dynamic models are adaptively and recursively estimated based on an exponential forgetting of past observations.

4.1. Autoregressive dynamics

Placing us in the case of AR dynamics, the equation linking the location parameter to previous observations is written as follows

$$\mu_{t+k} = \phi_t^T \Theta_t \tag{11}$$

where $\Theta_t = [\theta_{0,t},\ \theta_{1,t}, \dots, \theta_{p,t}]^T$ and $\phi_t = [1, y_t, y_{t-1}, \dots, y_{t-p+1}]^T$. This study is based on one-step ahead forecasts, so the horizon k is assumed equal to one. Note that the equations below are easily transformable for horizons greater than one. The expression of the scale parameter is developed at the end of this section. The number of lag denoted by p can be determined from many methods. For instance after examining the autocorrelation and partial autocorrelation functions (ACF and PACF) of y_t, which is the time-series corresponding to realisations of censored Normal variables. Other possibilities are also by minimisation of either Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) or finally through a cross-validation exercise. The last technique mentioned is applied in this study. It can be showed that parameters AR model Θ_t can be estimated using the Least Squares or Maximum

Likelihood methods, denoted respectively LS and ML. Note that a greater bias may be feared in the case of censored Normal variables using the LS method. Nonetheless, Pinson (2012) admitted that the LS-estimator applied to wind energy production is tolerable comparing to the ML-estimator. This technique is preferable in the sense that it has the advantage of being easier to implement than the other. Furthermore, as Pinson (2012), we place ourselves in a recursive framework, fore more details refer to Ljung and Söderström (1983) and Madsen (2007). The main benefit is that only the last available observations are used at each time step for updating the model parameters. Thus, this method offers the advantage of reducing the computational cost. The minimization problem for the estimate of the vector of AR parameters Θ_t with forgetting factor at time t takes the following form

$$\widehat{\Theta}_t = argmin_{\{\Theta\}} \Psi_t(\Theta) \qquad (12)$$

where

$$\Psi_t(\Theta) = \sum_{i=1}^{t} \lambda^{t-i}(y_i - \phi_{i-1}^T \Theta_t)^2 \qquad (13)$$

with λ is the forgetting factor, which belongs to the unit interval [0,1], allowing the exponential forgetting of past observations, for which the corresponding number of observations is defined as $n_\lambda = (1 - \lambda)^{-1}$. The value of the forgetting factor is typically slightly below 1. In that sense, a smallest value is assigned to the older observations. It assumes that the recent observations have more information for the future. According to the Recursive Least Squares (RLS) framework and from the minimization problem mentioned above, we can obtain the final updating parameters based on the newly information, which are summarized by the following equations

$$\widehat{R}_t^{-1} = \lambda \widehat{R}_{t-1}^{-1} + \phi_{t-1}\phi_{t-1}^T \qquad (14)$$

$$\widehat{\Theta}_t = \widehat{\Theta}_{t-1} + \widehat{R}_t \phi_{t-1} \varepsilon_t \qquad (15)$$

with $\varepsilon_t = y_t - \phi_{t-1}^T \widehat{\Theta}_{t-1}$ and $\widehat{R}_t$ is the inverse variance-covariance matrix at time t. For more details, the demonstration of the RLS method is demonstrated in the appendix (page 30).

The adaptive estimation at time t of the scale parameter can be performed using ARMA dynamics with some restrictions. The first restriction is that parameters of the model are fixed according to the forgetting factor and, secondly, the residual term is not related to the scale parameter but to the location parameter i.e. $\varepsilon_t^2 = \left(y_t - \phi_{t-1}^T \widehat{\Theta}_{t-1}\right)^2$. The proposal concerning the updating of the scale parameter is

$$\hat{\beta}_t = \lambda \hat{\beta}_{t-1} + (1 - \lambda)\varepsilon_t^2. \tag{16}$$

4.2. Autoregressive Moving Average dynamics

The Autoregressive Moving Average model (ARMA) is a very famous model used in the time-series domain. This model is similar to the previous one but integrate also the MA part, which involves modelling the error term as a linear combination of error terms occurring contemporaneously and at various times in the past. It has been shown to be a very tough competitor within renewable energy forecasting (David, Lauret, Trombe, Ramahatana, 2016). The equation linking the location parameter to previous observations inspired by Tsay *et al.*, (2005) is written as follows

$$\mu_{t+k} = \phi_t^T \Theta_t \tag{17}$$

with $k = 1$ is the horizon fixed to one, $\Theta_t = [\theta_{0,t}, \ \theta_{1,t}, \dots, \theta_{p,t}, \beta_{1,t}, \dots, \beta_{q,t}]^T$ the vector of parameters to be estimated and $\phi_t = [1, \ y_t, \ y_{t-1}, \dots, y_{t-p+1}, \varepsilon_t, \varepsilon_{t-1}, \dots, \varepsilon_{t-q+1}]^T$ the vector of inputs, where p is the order of the AR part and q is the order of the MA part. Similarly to the AR dynamic model, the final updating parameters based on the newly information can be written as

$$\hat{R}_t^{-1} = \lambda \hat{R}_{t-1}^{-1} + \phi_{t-1}\phi_{t-1}^T \tag{18}$$

$$\widehat{\Theta}_t = \widehat{\Theta}_{t-1} + \hat{R}_t \phi_{t-1} \varepsilon_t \tag{19}$$

with $\varepsilon_t = y_t - \phi_{t-1}^T \widehat{\Theta}_{t-1}$ and $\hat{R}_t$ is the inverse variance-covariance matrix at time t. In addition, the scale parameter is updated in the same way as in the previous section.

4.3. Reference models

It exists in the literature many benchmarks, which can be applied to this type of data. According to Perez *et al.,* (2010), persistence, smart persistence and climatology could be used. The two reference models in this study, as Pinson (2012), are the Moving-Average and the persistence models for the horizon of one step ahead, which are both very easy to use and also very efficient in the short term, especially the second one.

The Moving-Average benchmark consists to take the mean of the previous observations. In order to be consistent with the rest of the study, the number of lags here is selected the same that the AR dynamics. Thus, this first reference model can be written as follows

$$\mu_{t+k} = \frac{1}{p} \sum_{i=0}^{p-1} y_{t-i} \tag{20}$$

where $k = 1$ in this study case and p is the number of lags.

The second benchmark, which is the persistence model, is known to be very difficult to outperform for short-term forecasts. It is expressed as follows

$$\mu_{t+k} = y_t. \tag{21}$$

This model assumes that the wind power generation for each time depends only on its previous value, which means that the power production remains invariant between time t and time $t + k$.

In addition, for both reference models, the scale parameter is obtained with the same formula expressed in the section 4.1.

Chapter V

Evaluation Methods

The aim of this master thesis is to develop some new dynamic models to estimate the location and shape parameters related to the GL-Normal distribution and to predict wind energy production. We compare our models with some benchmarks discussed by several authors such as Giebel *et al.*, (2011), Costa *et al.*, (2008) and Pinson and Madsen (2011), which are know to be very difficult to outperform in the case of wind generation for short lead time, i.e. less than one hour. Model performance is assessed based on predictive sharpness and calibration, as these are desired predictive properties. As Gneiting *et al.*, (2007) indicate calibration refers to the statistical consistency between the predictive distribution and the observations, whereas sharpness refers to the concentration of predictive distributions. With actual values and predictions in hand, we need to find ways to quantify the model's effectiveness with respect to sharpness and calibration. This topic is discussed in this section.

5.1. Mean Absolute Error and Root Mean Square Error

The Mean Absolute Error and the Root Mean Square Error allow us to evaluate the point forecasts by measuring differences between observed and predicted values. They are regularly used in model evaluation studies.

The Mean Absolute Error, which is denoted by MAE, measures the average magnitude of errors in a set of predictions, regardless of their direction. It represents the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight. Consequently, the MAE associates the same weight to all errors. We let $\hat{y}_t$ be the prediction of the observed y_t at time t. If $t = 1, \dots, N$ then the MAE can be written as follows

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |y_t - \hat{y}_t|.$$

(22)

We note that if the absolute value is not defined (the signs of the errors are not removed), the average error becomes the Mean Bias Error (denoted by MBE) and usually aims to measure average model bias. MBE can transmit useful information, however, it must be interpreted with caution for the reason that positive and negative errors cancel each other out.

The Root Mean Square Error, which is denoted by RMSE, is a quadratic scoring rule that also measures the average magnitude of the error. It represents the square root of the average of squared differences between prediction and actual observation. As a consequence, the RMSE penalizes variance as it gives errors whit larger absolute values more weight than errors with smaller absolute values. The RMSE can be writing as follows

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (y_t - \hat{y}_t)^2}. \tag{23}$$

We note some similarities between them. Both MAE and RMSE evaluated average model prediction error in units of the variable of interest. Both metrics can vary from zero to infinite and are indifferent to the direction of errors. They are negatively oriented scores, which by definition means lower values are better. Obviously, when both measures are computed the RMSE cannot be smaller than the MAE. There are also some differences; taking the square root of the average squared errors has some interesting implications for RMSE. Since the errors are squared before they are averaged, the RMSE attributes a relatively high weight to large errors. Hence, the RMSE should be more interesting to calculate when large errors are particularly undesirable. An important remark is that if the error distribution is expected to be Gaussian and there are enough samples, the RMSE has an advantage over the MAE to represent the error distribution. Indeed, referring to Chai and Draxler (2014), the RMSE is more appropriate to represent model performance than the MAE when the error distribution is expected to be Gaussian.

Another remark is that, the RMSE does not necessarily increase with the variance of errors. In parallel, the RMSE increases with the variance of the frequency distribution of error magnitudes.

The underlying assumption when presenting the RMSE is that the errors are unbiased and follow a normal distribution. The MAE is more suitable to describe uniformly distributed errors. For the reason that model errors are likely to have a normal distribution rather than a uniform

distribution, the RMSE is a better measure to highlight than the MAE when errors following a Gaussian distribution.

5.2. Continuous Rank Probability Score

The Continuous Rank Probability Score denoted by CRPS is a verification tool related to probabilistic forecast systems, and it is a quantity that highlights to both forecast sharpness and calibration. In the case of a deterministic forecast, the CRPS takes the value of the mean absolute error (MAE) and, therefore, has a clear interpretation. According to the work done by Hersbach (2000), we can define the CRPS as below.

Let us denote with x the parameter of interest. Suppose that the Probability Density Function (PDF) forecast by an ensemble system is given by $f_t(x)$, and the corresponding observation is denoted by x_t, which is the value that actually occurred. Both Cumulative Distribution Functions (CDFs) for f_t and x_t are then respectively represented by

$$F_t(x) = \int_{-\infty}^{x} f_t(y)dy \tag{24}$$

and

$$P_t(x) = H(x - x_t) \tag{25}$$

where

$$H(x) = \begin{cases} 0 & \text{for } x < 0 \\ 1 & \text{for } x \geq 0 \end{cases}$$

this last equation $H(x)$ is the well-known Heaviside function. Note that, $F_t(x)$ is the forecasted probability that x_t will be smaller than x. Obviously, whatever the cumulative distribution $P_t(x) \in [0,1]$, $F(-\infty) = 0$, and $F(+\infty) = 1$. This remains true for parameters that are only defined on a subdomain. In that case $f_t(x) = 0$ and F_t constant outside the domain of definition.

We can show through a simple case the useful of the CRPS, using a simple implementation on Matlab software. We consider that the PDF has a standard normal distribution, and $x_t = 1$. Then, we can observe on the figures below the PDFs and CDFs plots at time t.
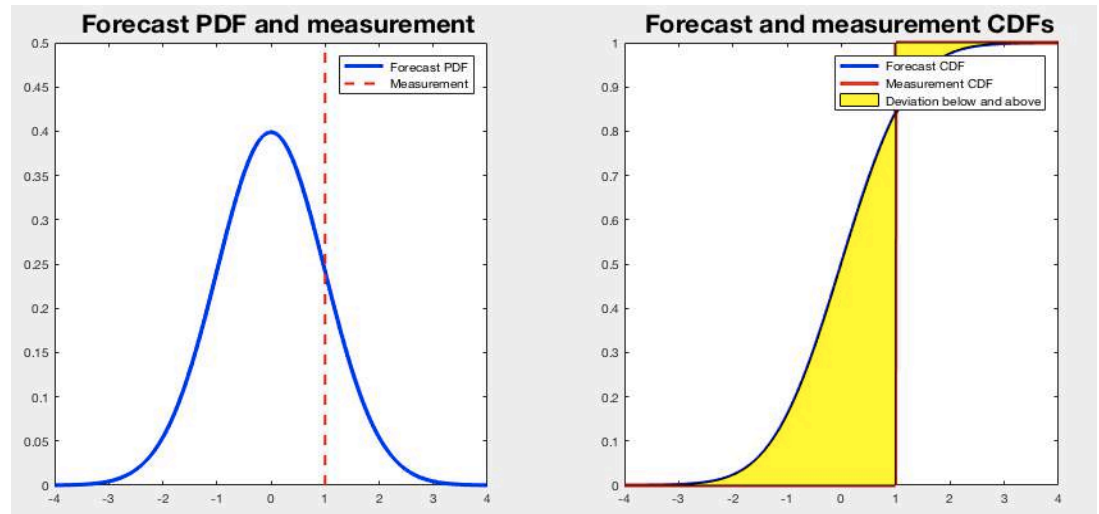
Figure 4. Visual illustration of CRPS when the probabilistic forecast is assumed to be a standard normal distribution.

The CRPS measures the difference between the predicted and occurred cumulative distributions. Its minimal value of zero is only achieved for $F_t(x)$ is equal to $P_t(x)$, that is, in the case of a perfect deterministic forecast. This measurement can be written like this

$$CRPS_t = CRPS(F_t, x_t) = \int_{-\infty}^{\infty} [F_t(x) - P_t(x)]^2 dx. \tag{26}$$

Figure 4 represents a visualization of the CRPS at a given time t. The blue and red lines represent F_t and P_t respectively. The particularity of the CRPS is that it is negatively oriented, in this way, a smaller CRPS indicates a better forecast. The area between F_t and P_t marked yellow (on the right chart above) is straight related to $CRPS_t$, it means that minimizing this area also minimizes $CRPS_t$. The area and thus the $CRPS_t$ are minimized when the observation takes the value of the median of the probabilistic forecast. In an empirical study, the CRPS is averaged over an interval of time, so that we obtain only one value for the CRPS:

$$\overline{CRPS} = \frac{1}{T} \sum_{t=1}^{T} CRPS_t. \tag{27}$$

Note one of the benefits of the CRPS is that it takes the form to absolute error if the forecast is deterministic, that is, when there is no randomness in the forecast. Empirically, this allows us to have a consistently comparison between a deterministic forecast and a probabilistic forecast.

5.3. Cross Validation

The cross validation method is a well-known method for adjusting and calculating predictions to estimate the performance of the prediction model. This technique so called rotation estimation or also out-of-sample analysis considers a dataset denoted by S. And, it divides the data in two sets: the training (so-called learning) set and the testing set. The purpose of cross-validation is to test the predictive model's ability to predict new data that have not been used to estimate it. Moreover, it allows to report problems such as overfitting issues and to provide an overview of how the model will generalize to an independent dataset.

In this section, the description of a famous category of cross validation called k-fold cross validation is presented. A brief outline of this technique is described in three points as follows:

1. Cut the dataset S into k folds (parts) $S_1, ..., S_k$, which are approximately of the same size. The type of data and the model used can influence how the dataset S should be divided.

2. For $i = 1, ..., k$, apply the following:

   2.1. For the i-th fold (testing set), fit the model to the other $k-1$ folds (learning set).

   2.2. Test the model on the i-th fold.

3. Merge results for all test sets.

In this master thesis, the k-fold cross validation method is applied in the fallowing chapter on wind power measurements. For more information about this technique, we can refer to Stone (1974).

Chapter VI

Application and Results

The aim of this part is to analyse the performance of the proposed dynamic models and the associated adaptive estimation method for modelling and forecasting wind energy fluctuations. These assessments are applied to a real case study. The exercise consists of one-step ahead forecasting of wind power generation. In a first stage, the onshore wind farm data are described. Then in a second stage, both models configuration and estimation setup are commented. The way forecasts are evaluated, as well the benchmarks is also presented. Moreover, the results obtained from point and density forecasts are discussed.

6.1. Wind Farm Data

The observations applied to this study come from the website of the Sotavento Galicia Foundation, which manages the Galicia wind farm site located in the northwest Spain. This wind farm has a nominal capacity of 17.56 megawatts (MW). An important remark is that Galicia is not a purely commercial wind farm, but above all an experimental site. During the analysis of data, we have noted that the nominal capacity was never reached. Galicia being a training area then, this is one of the possible reasons why it does not maximize its power generation on its wind farm. For this reason, the nominal capacity has been deliberately reduced to 13.56 MW. This considers the highest 20% measures as the nominal capacity. This assumption brings us closer to the industrial reality, which seeks to maximize the power generation i.e. the nominal capacity is reached on regular basis. Obviously, this assumption doesn't impact the aim of this study.

The data processing was carried out with a personal implementation in Python. The measurements have been gathered for the period ranging from 1th January 2016 to 31th December 2016, consisting of 52703 data points. Some few data were dropped due to measurements errors. After the cleaning procedure, the data set for wind power generation in Galicia consists of 52123 valid data points.

6.2. Model configuration, estimation setup and benchmarking

We have to consider some different periods, that are used for identification (learning sample) of the statistical models, and periods that are used for evaluating the performance of these models concerning out-of-sample analysis (testing sample). The first 17419 data points compound the learning set, which represents approximately the first four months the year studied. While the rest of the data, are employed for out-of-sample evaluation of one-step ahead forecast performance, which represents exactly 34704 data points.

Over the learning period, a part of data is used for one-fold cross validation exercise, represented by the last 10000 data points. This permits to evaluate the optimal values such as the optimal lags for the dynamic models, more precisely the AR order and also the MA order, the optimal forgetting factor and the optimal shape parameter of the GL transformation. Note that only the variance covariance matrix (equation 14) is updated before data point 10000 to get enough information before updating the vector of parameters (equation 15).

We remind that in a similar way to several studies such as Pinson (2012) or also Madsen and Pinson (2012) a suitable transformation is applied to the forgetting factor, it is preferred to use the corresponding effective number of observations denoted by n_λ, which can be written as follows

$$n_\lambda = \frac{1}{1 - \lambda}.$$

(28)

Indeed, this value is more comparable with the 'sliding window' in the adaptive estimation of the dynamic model parameters, as viewed by Gneiting *et al.*, (2006) and Hering and Genton (2009).

The vector of parameters for each model has been initialised as a white noise i.e. with a vector of zeros. The scale parameter has been initialised with a small value and finally, the inverse variance covariance matrix as an identity matrix times a small constant. Drawing on Pinson's (2012) study with equivalent data, the censoring parameter ε is fixed to 0.001, representing the resolution of the power measurements. Moreover, both the effective number of observations and the shape parameter of the GL transformation are selected from the results of Pinson in order to evaluate the first computations. These concern the estimation of the optimal lags for dynamic models using a one-fold cross validation exercise. More precisely, this procedure gives the optimal AR order and

the ARMA order. The metric to be minimised over the cross-validation set is the CRPS of 1-step ahead density forecasts. This method is an appropriate skill score to evaluate the performance of density forecasts (Gneiting *et al.*, 2007).

The location parameters for both models are computed with the RLS method as presented in the chapter IV. The results from the cross validation exercise show that the optimal lag to be considered is $p = 3$ for capturing the AR dynamics in wind power fluctuations with a 10-minute resolution, at Galicia wind farm. Regarding the ARMA dynamics the optimal lag to be considered is $p = 3$ for the AR part and $q = 2$ for the MA part. Using the optimal lags estimated for both models, another cross validation exercise is performed in order to determine the optimal n_λ and v applied to the case study. Pinson (2012) mentioned that for this type of data these parameters could be estimated by a train-and-error manner, by evaluating the results obtained from different setups. Moreover, it has been shown that the results are similar using a simultaneously or iteratively ways. Hence, for simplicity the parameters are reached by the iterative method i.e. optimising one after the other. After testing different setups, the parameters selected in the paper of Pinson (2012), which are applied to a different wind farm, are reasonable for our study case. As a result, the forgetting factor and the GL transform shape are $\lambda = 0.9996$ (with $n_\lambda = 2500$) and $v = 3.2$, respectively. Note that the last parameter is used only with AR and ARMA models, where a generalised logit-normal distribution is assumed. While, the benchmarks are assumed to fallow a (censored) normal distribution.

6.3. Point forecasting results

In this subchapter, we evaluate the point forecasts of different models proposed. The evaluation is performed with the Root Mean Square Error (RMSE). Note that, we dropped the Mean Absolute Error (MAE) criterion for two reasons. The first reason is that several authors used only the RMSE criterion to evaluate point forecasts for wind generation (Giebel *et al.*, 2011, Costa *et al.*, 2008, Madsen *et al.*, 2005), which gives the information sought. Indeed, this criterion permits to evaluate more precisely the large fluctuations, which are totally undesirable for energy production. The second reason is referring to Chai and Draxler (2014); the RMSE is more appropriate to represent the model performance than the MAE when the error distribution is expected to be Gaussian. The GL transform allows to process with a Gaussian framework. In addition, the

fact that we are in a probabilistic forecasting framework, the median of predictive densities is deliberately selected for the RMSE exercise.

Month	Moving-Average	Persistence	GL-Normal AR	GL-Normal ARMA
May	3.1349	2.3931	**2.1585**	2.1595
Jun.	3.5802	2.6940	2.1880	**2.1824**
Jul.	3.5477	2.7603	2.4162	**2.4029**
Aug.	3.7101	2.7635	2.3057	**2.2898**
Sep.	4.3401	3.0979	2.3350	**2.3178**
Oct.	3.3241	2.3297	**1.8338**	1.8341
Nov.	2.8369	2.0812	1.9029	**1.8993**
Dec.	2.5351	1.8646	**1.6286**	1.6295
All	3.4157	2.5263	2.1118	**2.1045**

Table 1. Monthly and overall assessment of point forecasts obtained using a RMSE criterion for the median with the test sample.

The RMSE is negatively oriented scores, then, the best scores in table 1 are identified in bold. The evaluation is carried out on a monthly basis and on overall the test sample, which more precisely represents the data from 1th May to 31th December 2016. We observe a high improvement when the predictive density assumed moves from the (censored) Normal distribution to the GL-Normal distribution. Indeed, we note a decrease in the RMSE criterion of approximately 16.5% when going from the Normal predictive density to GL-Normal predictive density. Nonetheless, the persistence model stays a competitive reference model comparing to the Moving-Average benchmark, which has much more high values whatever the period considered. We note also a slightly improvement concerning the GL-Normal predictive density when we use the ARMA dynamics instead of the AR dynamics. This table confirms the results obtained by Pinson (2012), who demonstrated the superiority of the GL-Normal distribution over the (censored) Normal distribution for point forecasting.

6.4. Density forecasting results

From the setup proposed in the second section of this chapter, the CRPS is computed monthly and overall for each model in order to evaluate the quality of predictive densities. The evaluation is applied on the test sample, which is the same than previously. This method evaluates the score with the GL transform data as for the point forecasts. The table 2

gathers the results obtained from the Galicia site. Note that, the best scores are indicated in bold.

Month	Moving-Average	Persistence	GL-Normal AR	GL-Normal ARMA
May	4.6900e+05	1.4694e+05	1.2293e+05	**1.2252e+05**
Jun.	5.3126e+05	1.6067e+05	1.1936e+05	**1.1901e+05**
Jul.	4.4896e+05	1.8517e+05	1.5701e+05	**1.5699e+05**
Aug.	4.5236e+05	1.8394e+05	1.5344e+05	**1.5293e+05**
Sep.	5.2377e+05	1.9037e+05	1.3973e+05	**1.3893e+05**
Oct.	4.5043e+05	1.6066e+05	1.3103e+05	**1.3052e+05**
Nov.	3.7970e+05	1.3904e+05	1.2137e+05	**1.2067e+05**
Dec.	4.7944e+05	1.0923e+05	8.6053e+04	**8.6030e+04**
All	3.7340e+06	1.2757e+06	1.0307e+06	**1.0274e+06**

Table 2. Monthly and overall assessment of density forecasts using a CRPS criterion.

The Moving-Average benchmark is extremely outperformed whatever the period considered. While, the persistence model appears to be a competitive benchmark. Nonetheless, the improvements in terms of the CRPS criterion are significant when going from the Normal predictive density to GL-Normal predictive density. For both AR and ARMA models with a GL-Normal predictive density, have a significantly higher probabilistic forecast skill if compared with the persistence model assuming a Normal predictive density, with a decrease in the CRPS criterion of 19.21% and 19.46%, respectively. Note that the ARMA model proposed in our study faintly outperforms the AR model. The results obtained from this density forecasting evaluation confirm the conclusion of Pinson (2012), who demonstrated the superiority of the GL-Normal distribution over the (censored) Normal distribution for density forecasting.

## 6.5. Illustration of quantile forecasts

To conclude this chapter, an illustration of quantile forecasts (with nominal of 5% and 95%) is presented with the figure 5. These forecasts are extracted from GL-Normal predictive densities with our best model, which is the ARMA dynamic structure. The corresponding normalized measurements are also represented. We can easily observe that the predictive densities are naturally bounded between 0 and 1. Moreover, we can observe that the variance of GL-Normal densities is larger when the conditional expectation of the process moves away from the bounds. In

particular, the following chart shows this feature when the conditional expectation deviates by 0. This remark confirms Pinson's (2012) conclusion on the variance of GL-Normal densities when applied to wind power generation.
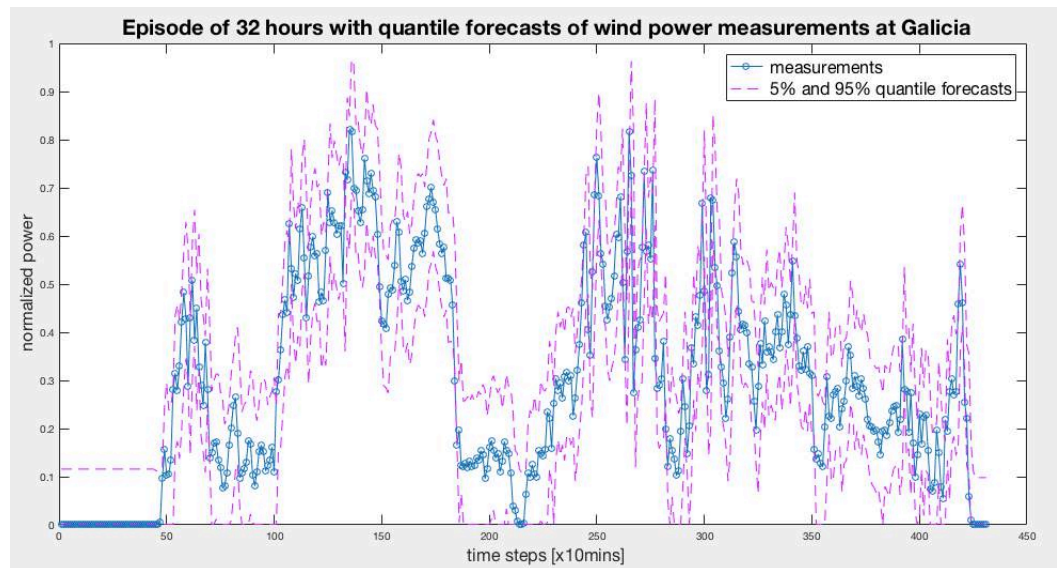
Figure 5. Sequence of 32 hours (with a 10-minute time step) of wind power measurements (solid line) with quantile forecasts of nominal proportions 5% and 95% (hashed lines). Quantile forecasts come from predictive densities with ARMA dynamics.

Chapter VII

Conclusion and Discussion

The first aim of this study was to extend the Pinson (2012) model by introducing a dynamic structure for the location of the wind generation. The second aim was to analyse the predictive ability of the proposed model. A GL-Normal distribution was proposed to better represent the wind power process in short-term probabilistic forecasting, whereas classical assumptions assumed (censored) Normal and Beta distributions. The ARMA dynamic for the location parameter is proposed to extend the existing dynamic models in the literature related to wind power forecasts. The forecasting methodology is applied to one-step-ahead prediction with a lead-time of 10 minutes at Galicia wind farm in Spain.

The proposed methodology has been largely inspired by Pinson (2012). Wind energy predictive densities are represented by discrete-continuous mixtures of GL-Normal distributions with potential concentration of probability mass at the bounds of the unit interval [0,1]. Probability masses are essential since the wind power measurements reach the bounds on regular basis. The location parameters of different setups are updated with the Recursive Least Square method with a forgetting factor. The scale parameters are updated at each step as a function of its previous values and the previous errors concerning the location parameters. These two parameters can fully represented the predictive density considered here. To evaluate point forecasts and density forecasts of different setups, the Root Mean Square Error and the Continuous Rank Probability Score methods have been employed, respectively. The models proposed were compared to different reference models, which are the moving-average and persistence benchmarks.

The proposal forecasting methodology applied to the Galicia wind farm in Spain allows many conclusions to be drawn. Firstly, results confirm the Pinson's belief, which is that the GL-Normal distribution is more appropriate to describe the wind power process in short term probabilistic forecasting and could be considered as the reference distribution. Indeed, evaluations according to the point forecasts and density forecasts confirmed the superiority of the GL-Normal distribution over the classical distribution, which is the (censored) Normal. As a result, this forecasting methodology seems to be robust for the process considered in short term analysis. Secondly, we note also that slightly improvements can be

obtained when more complex dynamics structures are applied to model the location parameter, which in our case ARMA dynamics is used instead of AR dynamics. Note that, even faintly improvements for forecasts are not negligible when applied to the overall electrical network.

Pinson and Madsen (2011) explained that regime switching might be present in the time-series of wind power generation, this feature was also confirmed by Vincent *et al.* (2011). They showed that the switches relate to different meteorological effects, which are very important for forecasts but extremely difficult to measure for short lead times. Pinson and Madsen (2011) demonstrated the superiority of the Markov-Switching Autoregressive (MS-AR) model over the Autoregressive model for very short-term wind power generation forecasts. However, they mentioned that many improvements could still be made. Consequently, furthers works on MS-AR models could be realised in order to extend this master thesis.

Appendix

The mathematical developments necessary for the predictive models' estimation are gathered in this section. More precisely, it gathers the recursive estimation procedure applied to the Autoregressive structure defined by the equations 14 and 15. In addition, to be coherent with our study case, demonstrations are applied for a horizon of one-step ahead. Nonetheless, this demonstration can easily be adapted for horizons greater than one. The recursive estimation procedure applied to the Autoregressive Moving Average model is not demonstrated due to look very similar with the AR model.

Recursive Least Squares of the location parameter with an AR structure

The expression relating the location parameter at time $t+1$ to previous observations can be written as follows

$$\mu_{t+k} = \phi_t^T \Theta_t \tag{29}$$

where $k = 1$, $\Theta_t = [\theta_{0,t} \; \theta_{1,t} \; \dots \; \theta_{p,t}]^T$ and $\phi_t = [1 \; y_t \; y_{t-1} \; \dots \; y_{t-p+1}]^T$. Hence the estimate of the parameters Θ_t at time t is defined as

$$\widehat{\Theta_t} = argmin_{\{\Theta_t\}} \Psi_t(\Theta_t) \tag{30}$$

with

$$\Psi_t(\Theta_t) = \sum_{i=1}^{t} \lambda^{t-i}(y_i - \phi_{i-1}^T \Theta_t)^2 \tag{31}$$

where λ is the forgetting factor, which belongs to the unit interval [0,1], allowing the exponential forgetting of past observations, for which the corresponding number of observations is defined as $n_\lambda = (1 - \lambda)^{-1}$. The value of the forgetting factor is typically slightly below 1. In that sense, a smallest value is assigned to the older observations. It assumes that the

recent observations have more information for the future. We can rewrite the previous objective function into matrix form as follows

$$\Psi(\Theta_t) = [Y_t - \Phi_{t-1}\Theta_t]^T \Lambda_t [Y_t - \Phi_{t-1}\Theta_t] \tag{32}$$

where $Y_t = [y_1 \ y_2 \ \dots \ y_t]^T$, $\Phi_t = [\phi_0 \ \phi_1 \ \dots \ \phi_{t-1}]^T$ and Λ_t the diagonal matrix, which associate the forgetting factor to each observation.
In order to find optimal parameters we have to apply the first order condition to our minimization problem, such that

$$\frac{\partial\Psi(\Theta_t)}{\partial\Theta_t} = 0 \tag{33}$$

$$\Leftrightarrow -2\Phi_{t-1}^T\Lambda_t(Y_t - \Phi_{t-1}\Theta_t) = 0$$

$$\Leftrightarrow \Phi_{t-1}^T\Lambda_t Y_t = \Phi_{t-1}^T\Lambda_t\Phi_{t-1}\Theta_t.$$

As a result, the optimal solution takes the following form

$$\widehat{\Theta}_t = [\Phi_{t-1}^T\Lambda_t\Phi_{t-1}]^{-1}\Phi_{t-1}^T\Lambda_t Y_t \tag{34}$$

with

$$\Lambda_t = \begin{bmatrix} \lambda^{t-1} & 0 & \cdots & 0 & 0 \\ 0 & \lambda^{t-2} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & \cdots & \lambda^1 & 0 \\ 0 & 0 & & 0 & \lambda^0 \end{bmatrix}. \tag{35}$$

Let $Y_{t+1} = [Y_t, \ y_{t+1}]^T$, $\Phi_t = [\Phi_{t-1}; \ \phi_t^T]$ and $R_t = [\Phi_{t-1}^T\Lambda_t\Phi_{t-1}]^{-1}$, where R_t is the inverse variance-covariance matrix at time t. We can write the variance-covariance matrix at time t and $t+1$ as the following:

$$R_t^{-1} = \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 & 1 \\ y_0 & y_1 & \cdots & \cdots & y_{t-2} & y_{t-1} \\ y_{-1} & y_0 & \cdots & \cdots & y_{t-3} & y_{t-2} \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ y_{1-p} & y_{2-p} & \cdots & \cdots & y_{t-1-p} & y_{t-p} \end{bmatrix} \begin{bmatrix} \lambda^{t-1} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda^{t-2} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda^{t-3} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda^0 \end{bmatrix} \begin{bmatrix} 1 & y_0 & y_{-1} & \cdots & \cdots & y_{1-p} \\ 1 & y_1 & y_0 & \cdots & \cdots & y_{2-p} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & y_{t-2} & y_{t-3} & \cdots & \cdots & y_{t-1-p} \\ 1 & y_{t-1} & y_{t-2} & \cdots & \cdots & y_{t-p} \end{bmatrix} \tag{36}$$

$$R_{t+1}^{-1} = \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 & 1 \\ y_0 & y_1 & \cdots & \cdots & y_{t-1} & y_t \\ y_{-1} & y_0 & \cdots & \cdots & y_{t-2} & y_{t-1} \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ \vdots & \vdots & \cdots & \cdots & \vdots & \vdots \\ y_{1-p} & y_{2-p} & \cdots & \cdots & y_{t-p} & y_{t+1-p} \end{bmatrix} \begin{bmatrix} \lambda^t & 0 & 0 & \cdots & 0 & 0 \\ 0 & \lambda^{t-1} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \lambda^{t-2} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda^1 & 0 \\ 0 & 0 & 0 & \cdots & 0 & \lambda^0 \end{bmatrix} \begin{bmatrix} 1 & y_0 & y_{-1} & \cdots & y_{1-p} \\ 1 & y_1 & y_0 & \cdots & y_{2-p} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & y_{t-1} & y_{t-2} & \cdots & y_{t-p} \\ 1 & y_t & y_{t-1} & \cdots & y_{t+1-p} \end{bmatrix}. \tag{37}$$

We note that the covariance-variance matrix at time $t+1$ is linear dependent of the previous period, such that

$$R_{t+1}^{-1} = \lambda R_t^{-1} + \phi_t \lambda^0 \phi_t^T \tag{38}$$

hence, we have

$$R_{t+1}^{-1} = \lambda R_t^{-1} + \phi_t \phi_t^T. \tag{39}$$

In this first demonstration we showed the linear dependence of variance-covariance matrices related to two successive periods. At present, we have to demonstrate how to obtain the LS-estimate of the parameters in our recursive framework. To show that we have to start from the optimal solution the equation 34:

$$\widehat{\Theta}_t = [\Phi_{t-1}^T \Lambda_t \Phi_{t-1}]^{-1} \Phi_{t-1}^T \Lambda_t Y_t$$

at time $t+1$ we have

$$\widehat{\Theta}_{t+1} = [\Phi_t^T \Lambda_{t+1} \Phi_t]^{-1} \Phi_t^T \Lambda_{t+1} Y_{t+1} \tag{40}$$

$$\widehat{\Theta}_{t+1} = R_{t+1}[\Phi_{t-1}^T \Lambda_t Y_t \lambda + \phi_t \lambda^0 y_{t+1}] \tag{41}$$

$$\widehat{\Theta}_{t+1} = R_{t+1}\big[\lambda R_t^{-1}\widehat{\Theta}_t + \phi_t y_{t+1}\big] \tag{42}$$

$$\widehat{\Theta}_{t+1} = R_{t+1}[R_{t+1}^{-1} - \phi_t \phi_t^T]\widehat{\Theta}_t + R_{t+1}\phi_t y_{t+1} \tag{43}$$

hence, we have

$$\widehat{\Theta}_{t+1} = \widehat{\Theta}_t + R_{t+1}\phi_t[y_{t+1} - \phi_t^T\widehat{\Theta}_t] \qquad (44)$$

it is equivalent to write

$$\widehat{\Theta}_{t+1} = \widehat{\Theta}_t + R_{t+1}\phi_t\varepsilon_{t+1} \qquad (45)$$

where $\varepsilon_{t+1} = y_{t+1} - \phi_t^T\widehat{\Theta}_t$. This quick demonstration inspired by Madsen (2007) gives us the recursive system in order to obtain the updating parameter based on the newly available information at time t+1. The updates parameters can be summarised by the following equations

$$\widehat{R}_{t+1}^{-1} = \lambda\widehat{R}_t^{-1} + \phi_t\phi_t^T \qquad (46)$$

and

$$\widehat{\Theta}_{t+1} = \widehat{\Theta}_t + \widehat{R}_{t+1}\phi_t\varepsilon_{t+1}\,. \qquad (47)$$

References

Aitchison, J. and Shen, S. (1980) Logistic-normal distributions: some properties and uses. *Biometrika*, **67**, 261-272.

Chai, T. and Draxler, R.R. (2014) Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geosci. Model Dev.*, **7**, 1247-1250.

Costa, A., Crespo, A., Navarro, J., Lizcano, G., Madsen, H. and Feitose, E. (2008) A review on the youg history of the wind power short-term prediction. *Renew. Sust. Energ. Rev.*, **12**, 1725-1744.

David, M., Lauret, P., Trombe, P.J. and Ramahatana, F. (2016) Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models. *Solar Energy, Elsevier*, **133**, 55-72.

Frederic, P. and Lad, F. (2008) Two moments of the logit-normal distribution. *Comm. Stat. B-Simul.*, **37**, 1263-1269.

Giebel, G., Brownsword, R., Kariniotakis, G., Denhard, M. and Draxl, C. (2011) The state of the art in short-term prediction of wind power – A literature overview, 2nd edition. Technical report, EU project Anemos.plus. [Available online: http://orbit.dtu.dk]

Gneiting, T. (2011) Quantiles as optimal point predictors. *Int. J. Forecasting*, **27**, 197-207.

Gneiting, T., Balabdaoui, F. and Raftery A.E. (2007) Probabilistic forecasts, calibration and sharpness. *J. Royal Stat. Soc.* B, **69**, 243-268.

Gneiting, T., Larson, K., Westrick, K., Genton, M.G. and Aldrich, E. (2006) Calibrated probabilistic forecasting at the stateline wind energy center – The regime-switching space-time method. *J. Am. Stat. Assoc.*, **101**, 968-979.

Gneiting, T. and Ranjan, R. (2013) Combining predictive distributions. *Electron. J. Stat.*, **7**, 1747-1782.

Hering, A.S. and Genton, M.G. (2009) Powering up with space-time wind forecasting. *J. Am. Stat. Assoc.*, **105**, 96-104.

Hersbach, H. (2000) Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**, 559-570.

Johnson, N. (1949) Systems of frequency curves generated by methods of translation. *Biometrika*, **36**, 146-176.

Lange, M. (2005) On the uncertainty of wind power predictions – Analysis of the forecast accuracy and statistical distribution of errors. *J. Solar Energ.-T. ASME*, **127**, 177-184.

Lesaffre, E., Rizopoulos, D. and Tsonoka, R. (2007) The logistic-transform for bounded outcome scores. *Biostatistics*, **8**, 72-95.

Ljung, L. and Söderström, T. (1983) *Theory and Practice of Recursive Estimation*. MIT Press: Boston.

Madsen, H. (2007) *Time Series Analysis*. Chapman & Hall/CRC: London.

Madsen, H., Pinson, P., T.S., Nielsen, H.Aa. and Kariniotakis, G. (2005) Standardizing the performance evaluation of short-term wind power prediction models. *Wind Eng.*, **29**, 475-489.

Mead, R. (1965) A generalised logit-normal distribution. *Biometrics*, **21**, 721-732.

Nelder, J.A. (1961) The fitting of a generalization of the logistic curve. *Biometrics*, **17**, 89-110.

Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D. and Hoff, T.E. (2010) Validation of short and medium term operational solar radiation forecasts. *US. Sol. Energy*, **84**, 2161-2172.

Pinson, P. (2006) Estimation of the uncertainty in wind power forecasting. Ph.D.thesis, Ecole des Mines de Paris, Paris, France. [Available online: http://pastel.paristech.org]

Pinson, P. and Madsen, H. (2011) Adaptive modelling and forecasting of wind power fluctuations with Markov-switching autoregressive models. *J. Forecasting*, available online.

Pinson, P. (2012) Very short-term probabilistic forecasting of wind power with generalised logit-normal distributions. *J. Royal Stat. Soc.* C, **61**, 555-576.

Raftery, A.E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005) Using bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, **133**, 1155-1174.

Sanchez, I. (2006) Short-term prediction of wind energy production. *Int. J. Forecasting*, **22**, 43-56.

Stone, M. (1974) Cross-validation and assessment of statistical predictions (with discussion). *J. Royal Stat. Soc*. B, **36**, 111-147.

Tsay, R.S. (2005) Analysis of financial time series. *Wiley, Hoboken, N.J.*

Vincent, C.L., Giebel, G., Pinson, P. and Madsen, H. (2010) Resolving non-stationary spectral signals in wind speed time-series using the Hilbert-Huang transform. *J. App. Meteoral. Clim*., **49**, 253-267.