

POLYTECHNIC UNIVERSITY OF TURIN

MASTER'S THESIS

**Modelling wind power production using
climate reanalysis**

Author:
Fabio BITETTI

Supervisor:
Prof. Vittorio VERDA



“To understand how something works, figure out how to break it.”

Nassim Nicholas Taleb

POLYTECHNIC UNIVERSITY OF TURIN

Abstract

Energy and Nuclear Engineering
DENERG

Master of science

Modelling wind power production using climate reanalysis

by Fabio BITETTI

The thesis consists in the creation of a wind power production model of mesoscale areas, like states. Inputs of the model are the production history of a geographical area, weather data, total installed capacity and possibly power capacity distribution. In case this last datum is not available, it is still possible to indirectly estimate it and train the model through an inverse mathematical problem. Due to the large amount of calculations, different optimization schemes are compared, in order to achieve a good solution in the most efficient way.

Acknowledgements

I would firstly like to thank Fortum for giving me the opportunity to challenge myself again, in a field I was not experienced enough. In particular, I would like to thank Jussi Mäkelä for the trust he put in me during the whole development of my work and Tommi Laine for the infinite, crucial debates on this model and his help in gathering all the data I needed.

I would also like to thank my supervisor Vittorio Verda for his interest in this project and his continuous support, started with my exchange period more than one year ago.

Last but not least, I thank my family and all my friends, who never stopped believing in me, even when I did.

Contents

| | |
|--|------------|
| Abstract | iii |
| Acknowledgements | v |
| 1 Introduction | 1 |
| 1.1 Historical background | 1 |
| 1.2 Definition of the problem | 1 |
| 1.3 Motivation and objectives | 2 |
| 1.4 Overview of the report | 2 |
| 2 Mathematical formulation and optimization algorithms | 3 |
| 2.1 General physical model | 3 |
| 2.2 ERA-5: reanalysis weather data for wind observations | 3 |
| 2.3 The wind power curve | 4 |
| 2.4 Case of known capacity location | 9 |
| 2.5 Case of unknown capacity location: the mathematical inverse problem of the capacity estimation | 11 |
| 2.5.1 Regularization procedure | 12 |
| 2.5.2 Choice of the regularization parameter: Cross Validation | 13 |
| 2.6 Algorithms in case of unknown capacity location | 14 |
| 2.6.1 Social approach: particle swarm optimization | 16 |
| Modelling of the particle's movements | 17 |
| Stopping criteria and time complexity of the algorithm | 18 |
| 2.6.2 Heuristic approach | 18 |
| 3 Validation of models | 21 |
| 3.1 Validation procedure | 21 |
| 3.2 Power curves validation | 23 |
| 3.3 Complete model validation | 27 |
| 3.3.1 Sweden | 27 |
| 3.3.2 Norway | 32 |
| 3.3.3 Finland | 37 |
| 3.3.4 Denmark | 42 |
| 4 Conclusions | 47 |
| A Basin Hopping algorithm | 49 |
| A.1 Stopping criteria and time complexity of the algorithm | 49 |
| B L-BFGS-B and SLSQP algorithms | 51 |
| B.1 Quasi-Newton methods | 51 |
| B.2 L-BFGS-B approximation | 52 |
| B.3 SLSQP | 54 |
| B.4 Time complexity | 56 |

| | | |
|-------|---------------------|-----------|
| B.4.1 | L-BFGS-B | 56 |
| B.4.2 | SLSPQ | 56 |
| | Bibliography | 57 |

List of Figures

| | | |
|------|--|----|
| 2.1 | ERA-5 nodes in Italy | 5 |
| 2.2 | Wind speeds distribution in Nordic countries in 2017 | 7 |
| 2.3 | Original measures and derived power curve | 8 |
| 2.4 | Power coefficient as a function of pitch angle and tip speed ratio | 9 |
| 2.5 | Graphical explanation of Sigmoid parameters | 10 |
| 2.6 | Graphical explanation of piecewise function parameters | 10 |
| 2.7 | Comparison between Sigmoid power curve and piecewise power curve | 11 |
| 2.8 | General behavior of an overfitting model | 13 |
| 2.9 | Scheme of the Shuffle-Split cross validation | 15 |
| 2.10 | Typical behavior of the average TEE as a function of regularization parameter | 16 |
| 2.11 | Flowchart of the particle swarm optimization scheme | 17 |
| 2.12 | Flowchart of the heuristic optimization scheme | 20 |
| | | |
| 3.1 | Nordic bidding areas | 22 |
| 3.2 | Power output distributions and time series for bidding zone DK2 with known geodata | 25 |
| 3.3 | Comparison of simulated power curves for bidding zone DK2 with known geodata | 26 |
| 3.4 | Time series zoom, DK2, fixed geodata: Sigmoid curve | 26 |
| 3.5 | Comparison of simulated power curves for bidding zone SE1 | 29 |
| 3.6 | Time series zoom, SE1: Heuristic scheme | 29 |
| 3.7 | Power output distributions and time series for bidding zone SE1 | 30 |
| 3.8 | Comparison of wind capacity distributions for bidding zone SE1 | 31 |
| 3.9 | Comparison of simulated power curves for NO3 | 33 |
| 3.10 | Comparison of wind capacity distributions for NO3 | 34 |
| 3.11 | Comparison of power output distributions for NO3 | 35 |
| 3.12 | Time series study for NO3 | 36 |
| 3.13 | Comparison of simulated power curves for Finland | 38 |
| 3.14 | Comparison of wind capacity distributions for Finland | 39 |
| 3.15 | Comparison of power output distributions for Finland | 40 |
| 3.16 | Time series study for Finland | 41 |
| 3.17 | Comparison of simulated power curves for DK2 | 43 |
| 3.18 | Comparison of wind capacity distributions for DK2 | 44 |
| 3.19 | Comparison of power output distributions for DK2 | 45 |
| 3.20 | Time series study for DK2 | 46 |

List of Tables

| | | |
|-----|--|----|
| 2.1 | List of inputs and outputs in wind power production modelling | 4 |
| 2.2 | Wind power production models and featured computational models | 10 |
| 2.3 | List of particle's behaviors and representative parameters | 18 |
| 3.1 | Comparison of results for Piecewise function and Sigmoid, DK2 | 24 |
| 3.2 | Comparison of results for PS and Heuristic method, SE1 | 28 |
| 3.3 | Comparison of results for PS and Heuristic method, NO3 | 33 |
| 3.4 | Comparison of results for PS and Heuristic method, Finland | 38 |
| 3.5 | Comparison of results for PS and Heuristic method, DK2 | 43 |

List of Abbreviations

| | |
|-----------------|--|
| BH | Basin Hopping |
| CD | Capacity Data |
| ECMWF | European Centre (for) Medium-range Weather Forecast |
| NNLS | Non Negative Least Squares |
| LASSO | Least Absolute Shrinkage (and) Selection Operator |
| L-BFGS-B | Limited (memory) - Broyden – Fletcher – Goldfarb – Shanno - Bounded |
| OLS | Ordinary Least Squares |
| OPEC | Oil Producing (and) Exporting Countries |
| PC | Power Curve |
| PD | Production Data |
| PS | Particle Swarm |
| PSO | Particle Swarm Optimization |
| QP | Quadratic Programming |
| RES | Renewable Energy Sources |
| SQP | Sequential Quadratic Programming |
| TEE | TEst Error |
| TRE | TRaining Error |
| WPP | Wind Power Production |

*To my family, my biggest source of inspiration, strength and
determination.*

Chapter 1

Introduction

1.1 Historical background

Power production is going through a genuine revolution, approximately started in 1970. This revolution consists in a lot higher percentage of Renewable Energy Sources (RES) production, such as wind, solar and hydro power. Causes of this movement can be identified in the following factors:

- Two "oil shocks" demonstrated that concentration of the market on a limited spectrum of energy sources makes it fragile [1]. The first one, in 1973, was provoked by the rise of oil prices set by OPEC. The second one rose through the war between Iraq and Iran, both of which aimed at destroying each others refineries. A natural consequence of these events is the decision from the most energy consuming countries to start reducing energy imports, in order to be less influenced by political occurrences [2];
- Scientific community agrees on the impact of greenhouse emissions on global warming. Most of these emissions originate in the energy sector. It has therefore become vital to start and speed up a decarbonization procedure of the power industry [3].
- RES are interesting from an economical point of view. In fact, their Levelized Cost Of Electricity (LCOE) is lower with respect to fossil based power production [4].
- Countries like Germany, Sweden and France have started a reduction of the nuclear energy share. One technical reason is that nuclear plants present a limited ramping potential, which makes them poorly compatible with highly variable energy sources like RES. Furthermore, Fukushima's accident increased social adversity against nuclear energy.

1.2 Definition of the problem

The general problem the thesis proposes to solve consists in a sufficiently detailed description of Wind Power Production (WPP) of a geographical area. In particular, a physical vs. statistical model is built in Python 3, so that results can be easily interpreted and criticized. The most important aspect the model should be able to take into account is the typical variable pattern of wind, directly connected to weather. More details on the inputs and outputs of the model are reported further on.

1.3 Motivation and objectives

Gaining insights into WPP of a geographical area is extremely important today. In fact, increasing decarbonization needs of the most technological advanced countries lead to a growing percentage of RES. However, Solar and WPP are both characterized by a strong weather dependent behaviour and this makes power production highly variable. Thus, in the particular case of Fortum, a model for wind power systems is necessary in order to evaluate their sensitivity to weather volatility. The fundamental idea at the base of this modelling is that, once the model of a wind power system is calibrated, it would be possible to run it with historical weather data and obtain a full spectre of scenarios. Since variety of scenarios is directly linked to the amount of available historical data, weather reanalysis data constitutes a perfect resource for the purpose, due to its abundance and high time and space resolution. More details about reanalysis and its datasets are given in the following parts of the document.

Based on the explanations just reported, the thesis focuses on the following aspects of this complex problem:

- Building a physical statistical model for WPP, in case of known wind capacity location;
- Introduction of a mathematical method for the estimation of wind capacity location in case this piece of information is not available;

1.4 Overview of the report

Chapter 2 starts with a deeper discussion of the WPP modelling from a mathematical and physical point of view. Power Curves (PCs) are modelled through two possible functions and their features are deeply analysed. After that, it is stressed how the model has been implemented and what computational schemes have been tested and adopted in models.

In chapter 3 validation of all proposed models and schemes is described and performed. Its results are consequently analyzed and compared.

Chapter 4 finally summarizes observation developed through the whole work and gives a conclusion about it. Possibilities for improvements in the models are also examined.

Chapter 2

Mathematical formulation and optimization algorithms

The following chapter describes methodology adopted for wind power modelling and discusses how to computationally solve the resulting optimization problem, both in cases of availability or absence of capacity data ¹.

2.1 General physical model

The model built for the description of WPP aims at the minimum amount of input data and a sufficient precision of results. The reason of this choice is due to public availability of information. In fact, in most of the cases, it is not easily possible to have details about installed wind farms in a geographical area ². A natural consequence of this lack is the formulation of a simplified model, which associates one PC to each geographical area with available Production Data (PD). Fortunately, PD can be found in transparency platforms like ENTSO-E TP, whose role is collecting and sharing generation, transportation and consumption data and information for the pan-European market [5]. However, to have a sufficient level of precision, it is still necessary to know (or to estimate) where wind capacity is installed. In this case, unfortunately, data is not always available. Table number 2.1 resumes all the inputs and outputs of the model.

It is important to stress that the model works with normalized PD, as shown in the table. In fact, the advantage of using normalized data is principally linked to computational schemes adopted in the solution of the mathematical problems, which will be described in detail afterwards.

2.2 ERA-5: reanalysis weather data for wind observations

ERA-5 is a reanalysis atmospheric dataset produced by ECWMF. Reanalysis gives numerical descriptions of weather phenomena and consists in an immutable data assimilation scheme, able to process a huge amount of data through consistent meteorological models. Data is acquired from a wide observational network, like radiosondes, satellites, buoys, aircrafts and ships. Estimates are extended to all the areas of the earth, with relatively small time resolutions. There are many powerful reanalysis datasets commonly used today, like MERRA-2 (created by NASA), but ERA-5 has been chosen for four principal reasons:

¹Capacity data consists in the location and value of wind capacity installed in a prescribed region.

²Some providers sell this data. For example, at the moment, a detailed database of European wind farms can be bought for 600 euros from the website thewindpower.com

| Inputs | Outputs |
|--|-------------------------------|
| Normalized hourly production data | Power curve |
| Reanalysis weather data (Normalized capacity geodata) | (Normalized capacity geodata) |

TABLE 2.1: List of inputs and outputs in wind power production modelling. Capacity data is reported between parenthesis because they can be an input or an output of the model according to the case.

- It is easily accessible and processable via Python and its packages ³;
- ERA-5 has the highest time and space resolution among the most used reanalysis datasets: spatial grid has a resolution of 31 kilometers and time resolution of 1 hour [8]. An example of a spatial grid is reported in figure 2.1;
- Recent studies reported the high quality of the dataset in wind modelling in comparison with other important reanalysis datasets [9];
- It is possible to get wind speed observations at 100 meters of altitude. All wind speed measurements processed by the statistical model are referred to this elevation, since it corresponds to the average tower height of modern wind turbines. Furthermore, since wind speeds follow a power law, differences between wind speeds at 100 meters and 90 meters of altitude are a lot less significant than differences between wind speeds at 30 meters and 40 meters. The reason why a power law effectively describes wind speed distributions is that local orography has a relevant impact only on limited altitudes. [10]

2.3 The wind power curve

The wind PC is a function that links wind speed and power output of a wind turbine. An example of PC is reported in figure 2.3a, while figure 2.3b shows where the PC is extrapolated from. As it is possible to notice from the figure, the characteristic parameters describing the curve are three:

- **Cut-in speed:** this parameter describes the minimum wind speed necessary to start producing mechanical (electrical) power. If this value is not reached, there is not enough torque to make turbine blades start rotating;
- **Cut-out speed:** this value corresponds to the maximum wind speed sustainable by a wind turbine before mechanical stress becomes too high. If wind speed is larger than cut-out speed, blades are deviated to reduce wind lift;
- **Rated output speed:** when wind speed is larger than or equal to this parameter, electric generator provides maximum power.

³Reanalysis data is downloaded through ECMWF Web API [6], while the package Iris is used for processing reanalysis data [7]

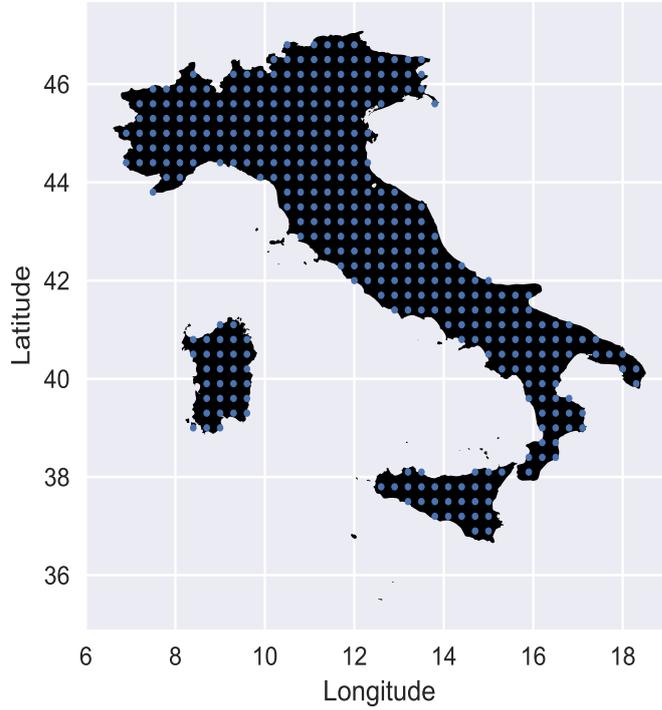


FIGURE 2.1: ERA-5 nodes in Italy.

Another interesting feature to investigate is how the slope of the PC is affected. Ideally, the ascending part of the curve would be proportional to the cube of wind speed passing through the turbine. More precisely:

$$P_{id} = \frac{1}{2} \rho v^3 S \quad (2.1)$$

where S corresponds to the area covered by turbine's blades. In reality, mechanical power delivered by wind turbines is also affected by the pitch angle of blades β and the rotating speed ω_{rt} of the wind rotor:

$$P_{rt} = C_p P_{id} = f(v, \beta, \omega_{rt}) \quad (2.2)$$

A function describing C_p can be found in literature [11]:

$$C_p = C_p(\lambda, \beta) = c_1(c_2/\lambda_i - c_3\beta - c_4)e^{-c_5/\lambda_i} + c_6\lambda \quad (2.3)$$

$$1/\lambda_i = 1/(\lambda + 0.08\beta) - 0.035/(\beta^3 + 1) \quad (2.4)$$

$$\lambda = \frac{\omega_{rt}R}{v} \quad (2.5)$$

where λ is defined as the tip speed ratio and R corresponds to the length of a blade. Parameters c_1, \dots, c_6 are a function of the wind turbine. Plot in figure 2.4 shows the behaviour of C_p for different pitch angles. Generally, in order to keep C_p to its maximum value for a fixed value of β , λ is kept constant by continuously

adjusting the rotor speed ω_{rt} according to wind speed v . Thus, P_{rt} approximately conserves the cubic trend of P_{id} .

Literature shows that this curve can also be adopted to model WPP of a wind farm [12], but in this case it is used to describe every single node of a geographical area. In other words, it is assumed that a geographical area is like a huge wind farm, composed of wind turbines provided with the same normalized PC.

Two models of PC are proposed:

- **Sigmoid function:** a first approximation of a PC can be obtained through a sigmoid function. A sigmoid has the following form:

$$S(x) = \frac{1}{1 + e^{-x}} \quad (2.6)$$

In order to give it a physical sense, the function assumes this particular shape:

$$PC(v) = \frac{\eta}{(1 + e^{(c_1-v)/p_1})(1 + e^{(v-c_{out})/p_2})} \quad (2.7)$$

Where parameters c_1 , c_{out} , p_1 , p_2 correspond to shape factors, η to the efficiency of the curve (asymptotic value) and v to wind speed. A graphical explanation of the role of each parameter is reported in figure 2.5. However, wind speed data show that there is no need for a pole in correspondence of c_{out} . In fact, histograms reported in figures 2.2 show that high wind speeds are rare. This happens because maximum resolution of reanalysis data is in the order of mesoscales, where big deviations from average wind speeds are still very unlikely. Thus, from this moment, the function will be simplified as it follows:

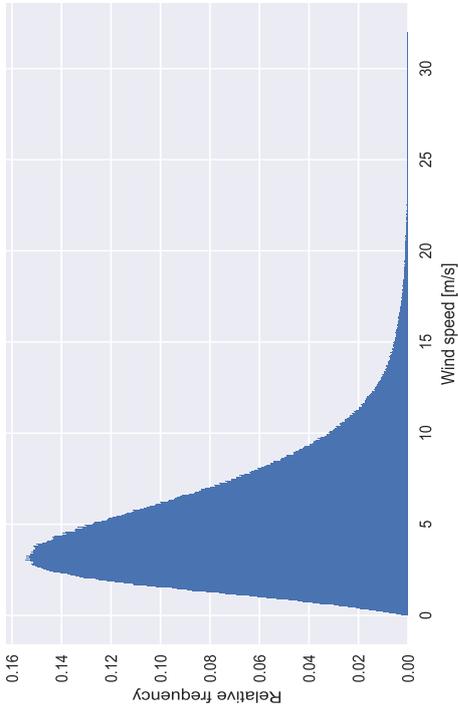
$$PC(v) = \frac{\eta}{(1 + e^{(c-v)/p})} \quad (2.8)$$

Where c_1 is substituted with c and p_1 is substituted with p .

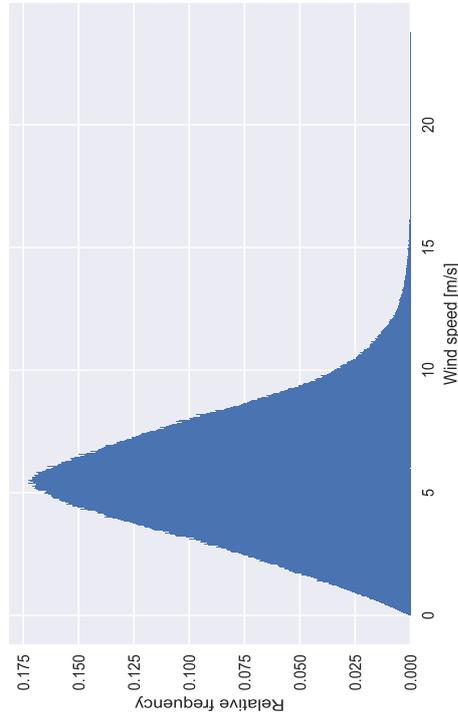
- **Piecewise function:** a more refined approximation of a power curve can be obtained from the proposed piecewise function.

$$\begin{cases} \frac{\eta}{(1+e^{(c-v)/p})} & v > c \\ \eta \left(\frac{v-c}{4p} + 0.5 \right) & z \leq v \leq c \\ \frac{\eta(z-c)}{4p+0.5} e^{\frac{v-z}{z-c+2p}} & 0 \leq v < z \\ 0 & elsewhere \end{cases} \quad (2.9)$$

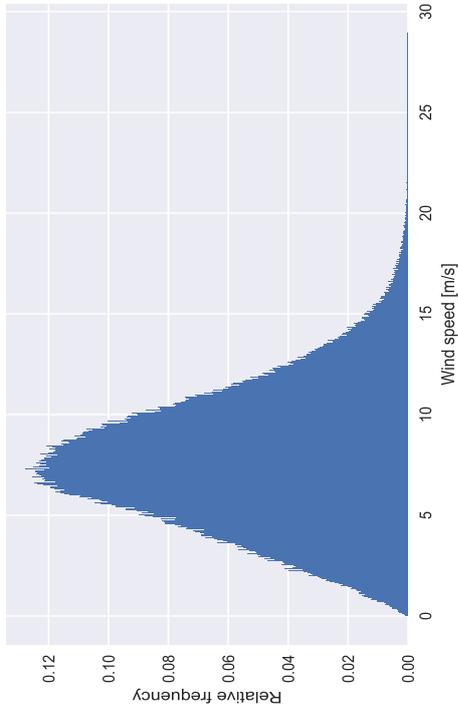
Where c and p correspond to the previous c_1 , p_1 . The first piece of the function is identical to a Sigmoid, the second one is linear and the third one follows an exponential trend. A clearer vision of how each parameter shapes the curve and of its division into different pieces is in figure 2.6, while a comparison of this PC with the previous one is reported in figure 2.7. It is obvious that, thanks to parameter z , it is possible to better model how the curve converges to zero for low wind speeds. Furthermore, it is possible to verify that the derivative of this function is continuous and so the function itself. The increase of complexity of the curve by one parameter is justified by the intention to increase precision in the results.



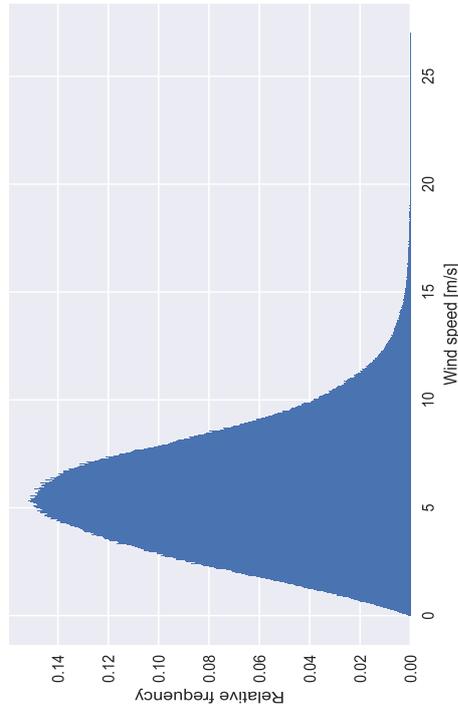
(B) Norwegian wind speeds distribution.



(D) Finnish wind speeds distribution.

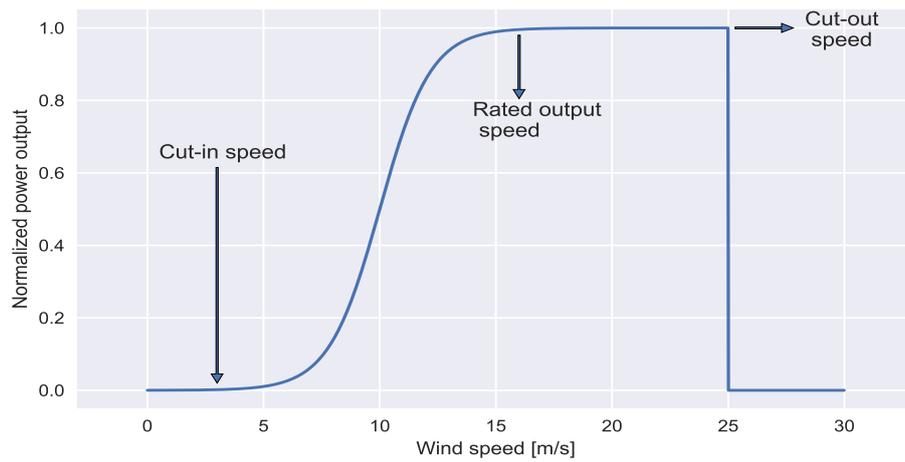


(A) Danish wind speeds distribution.

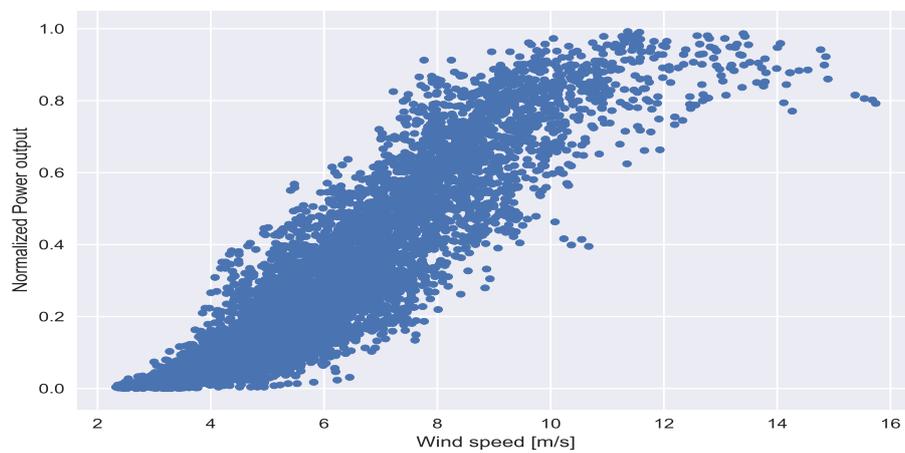


(C) Swedish wind speeds distribution.

FIGURE 2.2: Wind speeds distribution in Nordic countries in 2017. Due to limited space resolution, average high wind speeds are rarely detectable.



(A) Example of wind power curve. [13]



(B) Original measures for wind power curve. [14]

FIGURE 2.3: Original measures and derived power curve.

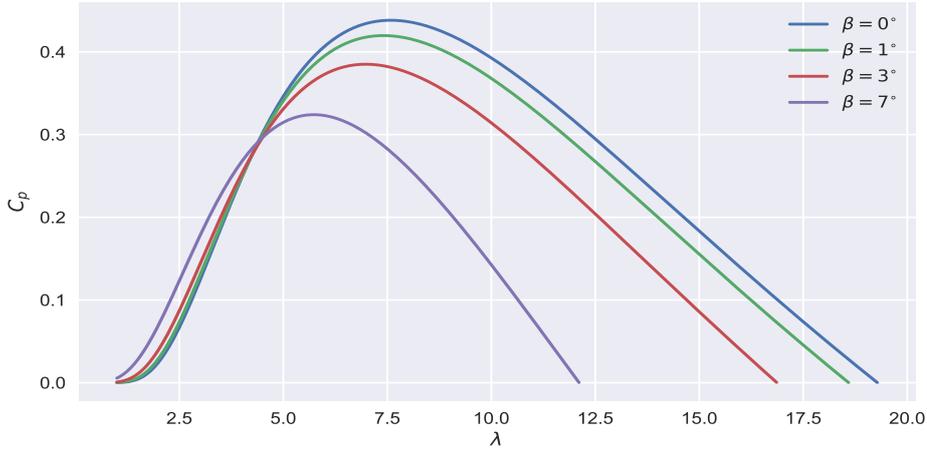


FIGURE 2.4: Power coefficient as a function of pitch angle and tip speed ratio, where $c_1, c_2, c_3, c_4, c_5, c_6$ are 0.22, 120, 0.4, 5, 12.5, 0.

2.4 Case of known capacity location

If the geographical distribution of capacity is known, the only unknown is the PC. If the Sigmoid is used (eq. 2.7), three optimal parameters need to be found. If the piecewise PC is selected (eq. 2.9), four of them must be optimized.

The fitting procedure of the PC can be described as a non linear optimization problem. In fact, the objective is finding the combination of the PC parameters that minimize the distance between the calculated profile and the real one:

$$\min_{PC} \{ \|A \cdot x - b\|_2 \} \quad (2.10)$$

$$A = |PC(v_1), \dots, PC(v_n)|, i \in \{1, \dots, n\}$$

where x corresponds to the vector of the wind capacities, v_i correspond to wind speed history of every single node and b is the total area production data. The selected distance (objective function) in this problem is the L2 norm of the difference of the two profiles. Algorithms chosen for this optimization are reported in table number 2.2.

Both optimizations have bounded solution. In fact, all the parameters must be positive. Furthermore, in order to speed up the convergence of the solution, proper upper bounds are set for each investigated quantity. However, Piecewise PC also needs two constraints:

- z must be smaller than c ;
- $z - c + 2p$ must be larger than zero in order to keep its exponent negative.

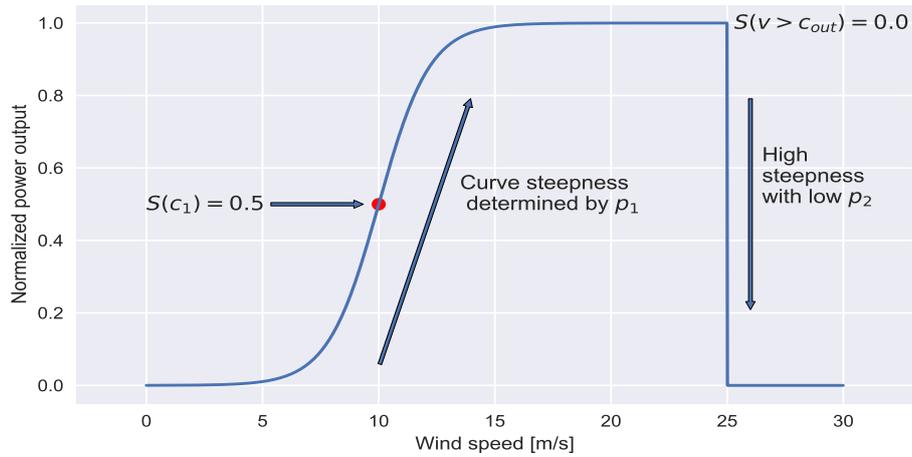


FIGURE 2.5: Graphical explanation of Sigmoid parameters. The asymptotic value of the curve (η) is set to 1 in this case.

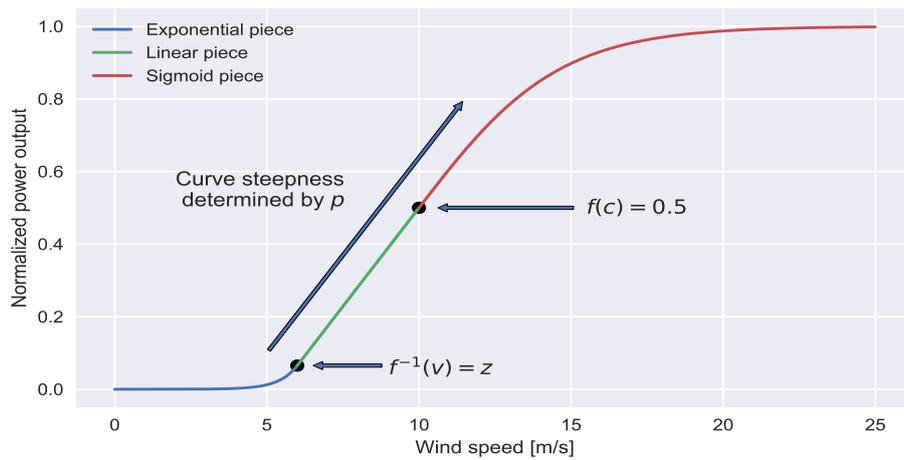


FIGURE 2.6: Graphical explanation of piecewise power curve parameters. The curve is composed of three pieces: the red one corresponds to a Sigmoid, the green one is a straight line that has the same slope of the Sigmoid calculated in c and the blue one follows an exponential trend.

| | Sigmoid PC | Piecewise PC |
|----------|----------------------|------------------------------------|
| Method | L-BFGS-B [15] | SLSQP [16] |
| Typology | Gradient based | Gradient based |
| Feature | Bounded optimization | Bounded & Constrained optimization |

TABLE 2.2: Synthetic description of WPP models and featured computational models.

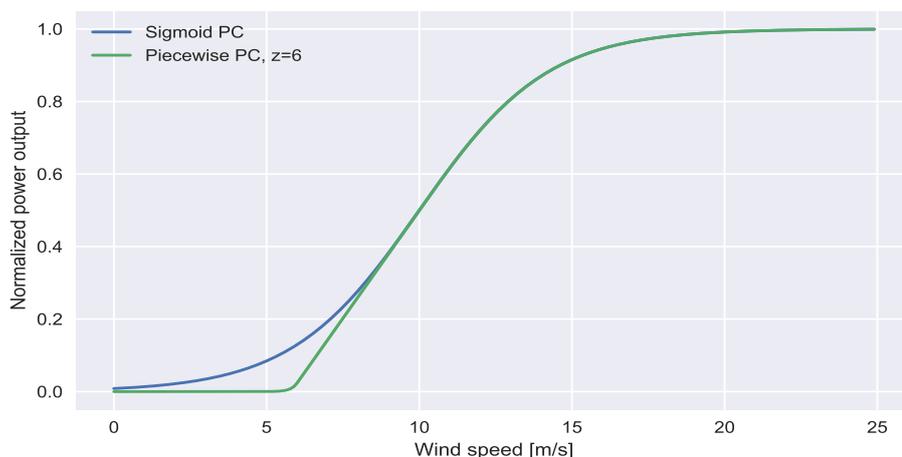


FIGURE 2.7: Comparison between Sigmoid power curve and piecewise power curve. The increase in the complexity of the function permits to model the power curve more freely.

L-BFGS-B is a proper scheme for the Sigmoid case because it is specialized in bounded, unconstrained problems. Instead, SLSQP is suitable for constrained optimizations [17], thus it is used for the second case.

The limit of these two methods is that their result may be heavily influenced by the initial guess the method needs. Thus, in order to have maximum chances of finding the global minimum, both methods are integrated into the Basin-Hopping (BH) algorithm, which combines a global stepping algorithm with local optimization. Further details of the algorithm can be found in the relative appendix A.

2.5 Case of unknown capacity location: the mathematical inverse problem of the capacity estimation

Thanks to transparency platforms (like ENTSO-E TP), accessing to production and installed capacity data has become easier. However, in many countries, it is still not possible to know where wind parks are located with sufficient precision. Thus, estimating the position and the entity of installed wind capacity becomes crucial to obtain sufficiently accurate results. In order to explain the nature of this problem, it is convenient to start from the following simplified reasoning:

Suppose to have a finite number of WPP time series $C = \{p_1(t), p_2(t), \dots, p_n(t)\}$ and a particular WPP profile $b(t)$. Which of their infinite possible linear combinations gives a resulting profile that better fits $b(t)$? In a practical way, the following problem must be solved:

Each grid point of a prescribed geographical area has its own normalized WPP time series, derived from the application of a PC to its wind speed time series. How should the grid points be weighted to recover area production data b ?

In a more formal way, we are trying to minimize the error between two profiles by using proper weights. If the error is defined as the L2 norm of the difference of these two profiles, the problem will consist of the Non Negative Least Squares (NNLS) problem, a particular case of the OLS problem:

$$\begin{aligned}
& \min_{x \in \mathbb{R}^n} \{ \|A \cdot x - b\|_2 \} \\
& A = |p_1, p_2, \dots, p_n| \\
& \text{s.t.} \\
& x_i \geq 0 \forall i \in \{1, \dots, n\}
\end{aligned} \tag{2.11}$$

Furthermore, an important observation should be done about parameters that should be optimized. Since the efficiency of a PC is a linear term:

$$PC(v) = \eta PC(v, \eta = 1) = \eta PC_M(v) \tag{2.12}$$

If equation 2.12 is substituted in definition of A in equation 2.11:

$$A = |p_1, p_2, \dots, p_n| = |PC(v_1), \dots, PC(v_n)| = \eta |PC_M(v_1), \dots, PC_M(v_n)| \tag{2.13}$$

Since η is unknown, it is possible to define a new variable and substitute it to the unknown capacity vector x in problem 2.11:

$$x' = \eta x \tag{2.14}$$

In this way, the number of variables that should be optimized reduced by one.

Now that a general idea has been outlined, it will be easier to understand the introduction of more complexity into this inverse mathematical problem: since the realized production b is subject to estimation and measurements errors and A derives from the application of a simplified model of reality, it is necessary to *regularize* the solution. Otherwise, the resulting model would overfit the profile b , with a consequent loss of predictivity of the model. Plot in figure 2.8 describes this concept more effectively: if the candidate model overfits training data, testing the model on some new data will give worse results than an apparently less precise model.

2.5.1 Regularization procedure

The previously described regularization problem can be solved by the introduction of a regularization procedure. The selected one for this specific problem is LASSO. The reason why it has been preferred to other regularization techniques will become clear soon.

LASSO is formally introduced as the solution of the following optimization problem:

$$\min_{x \in \mathbb{R}^n} \{ \|b - A \cdot x\|_2^2 + \lambda \|x\|_1 \} \tag{2.15}$$

Where λ is a non negative real number. Observing the equation gives the reader lots of interesting observations about the method:

- Since the optimization consists in a minimization, increasing λ will make entries of x smaller, provoking an increase in the difference between $A \cdot x$ and b ;
- The increase of λ increases the sparsity of the solution, a desired phenomenon due to the typically sparse distribution of wind capacity in a country (validation of results shows this pattern);

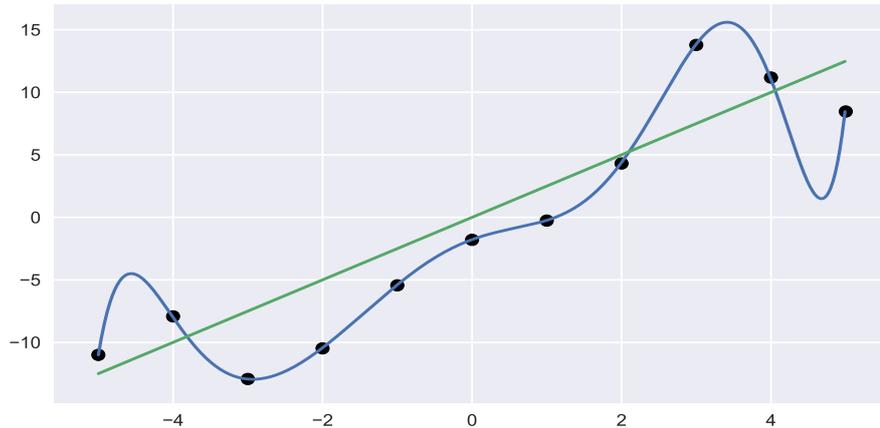


FIGURE 2.8: General behavior of an overfitting model. A polynomial curve explains the measured points perfectly, but it would not be effective in case of an extrapolation. In this case, a linear function would better describe trend of measures, even if the resulting error is larger. [18]

- In case of perfect observations, which means A and b without noise or uncertainty, the solution would not need regularization:

$$\lim_{\lambda \rightarrow 0} \min_{x \in \mathbb{R}^n} \{ \|b - A \cdot x\|_2^2 + \lambda \|x\|_1 \} = \min_{x \in \mathbb{R}^n} \{ \|b - A \cdot x\|_2^2 \} \quad (2.16)$$

which corresponds to the NNLS problem again;

- It is crucial to investigate which value of λ guarantees the best regularization. In fact, picking a too small value will lead to an overfitting solution, while picking a too large value will generate a model that is not even able to explain the fitted data.

2.5.2 Choice of the regularization parameter: Cross Validation

Cross validation is a simple, intuitive and established method for the calibration of λ . Before continuing with the explanation, it is necessary to introduce some definitions.

We define (x_i, y_i) , $i = 1, \dots, n$, the *training data*, which consists in the observation of the phenomenon that the model should describe. Once an estimator f is formulated based on the training data, it is possible to calculate TRaining Error (TRE):

$$TRE(f) = \frac{1}{n} \sum_{i=1}^n (y_i - f(x_i))^2 \quad (2.17)$$

Given a set of *test data* (x'_i, y'_i) , $i = 1, \dots, m$, it is possible to define TEst Error (TEE) in an analogous way:

$$TEE(f) = \frac{1}{n} \sum_{i=1}^n (y'_i - f(x'_i))^2 \quad (2.18)$$

In order to create test data, the model is trained on a subset of the total available data, called *fold*. The test is performed on the excluded part. It follows that the

distinction between training data and test data is labile. For example, if the whole sampled data is divided into K folds, it is possible to train and test the model K times. In fact, it will be sufficient to leave one of the folds out, training the model on the remaining ones and then testing the model on the excluded one. This procedure is called K -fold validation and it is widely used in Machine Learning. However, in this case, the Shuffle-Split cross validation is executed. Indeed, it has the particular feature of assembling the K sets by randomly selecting samples from the total data. The advantage of this choice is that the model will be trained without losing the influence of the seasonal trends in wind power production. The procedure is schematically explained in figure 2.9 and resumed below:

- The total data is divided into K subsets, randomly composed and roughly of the same size. A set of possible values for λ is selected;
- $\forall \lambda \in [\lambda_{min}, \lambda_{max}]$:
 - For every subset, the model is trained on all but one subsets;
 - The TEE of the excluded subset is calculated;
 - The average TEE is calculated from all the subsets.

A typical behavior of λ is reported in figure 2.10. It is obvious that the optimal λ is the one that minimizes prediction errors, i.e. TEE.

2.6 Algorithms in case of unknown capacity location

As already explained, in case of unknown wind capacity location, it becomes necessary to estimate it. However, in this case, gradient based methods are not suitable for the problem under exam. In fact, the amount of variables involved into the problem is the following:

- One variable for each parameter of the modelled PC, excluding the efficiency η ;
- One variable for each grid point of a geographical area ⁴.

In order to apply a gradient based optimization procedure, the gradient of a function should be calculated at each iteration. Computationally, this means that function evaluations in one iteration of the local gradient based optimization would be equal to the number of variables in the problem. If this procedure is integrated into a global optimization procedure like the BH algorithm, it is possible to imagine how slow the optimization would be. As a consequence, a different approach must be used. In the following sections, two approaches are proposed.

⁴As an example, the number of nodes in Stockholm is over 300.

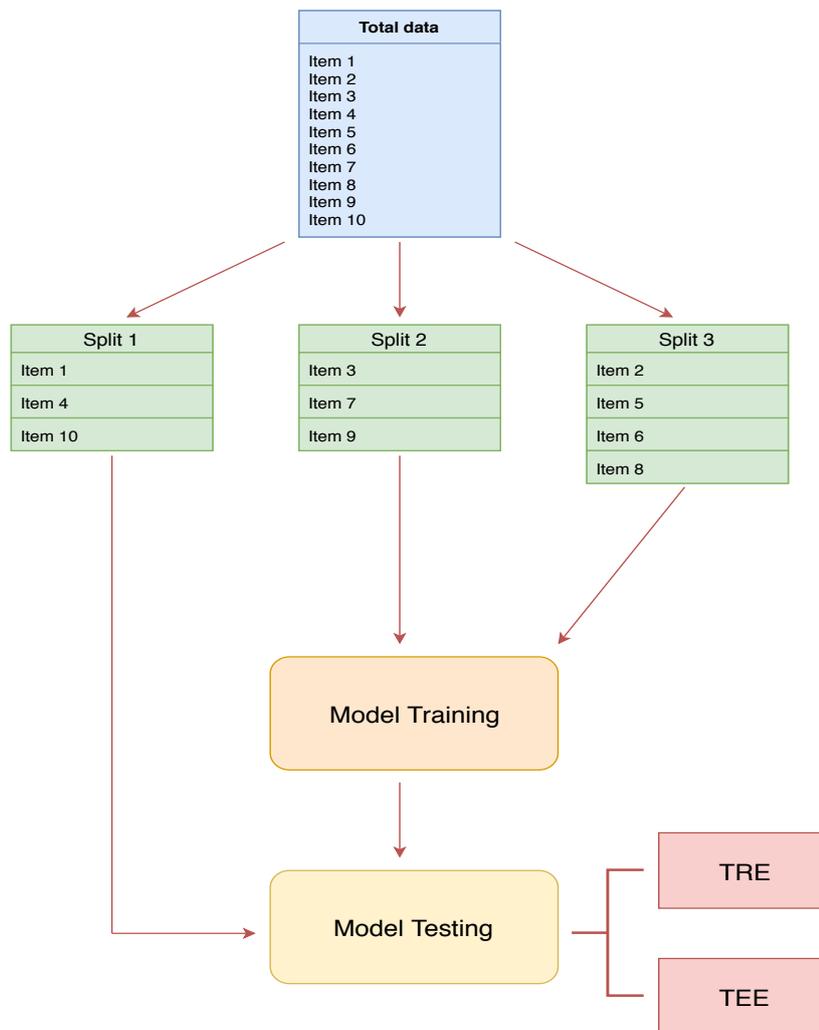


FIGURE 2.9: Scheme of the Shuffle-Split cross validation, for $K=3$ and data composed by ten items. Former data is splitted into K subsets. $K-1$ subsets are used to train the model, while the remaining one is used for its validation.

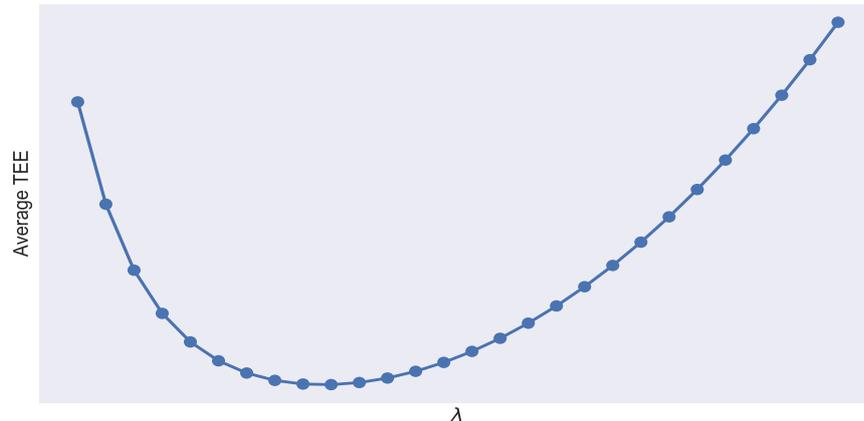


FIGURE 2.10: Typical behavior of the average TEE as a function of regularization parameter. It is possible to maximize the predictivity of the model by selecting the value of λ that minimizes the average TEE.

2.6.1 Social approach: particle swarm optimization

The mathematical nature of the allocation problem has been already described. In particular, it is possible to underline the following points:

- For an assigned combination of matrix A and vector b , there are unique x and λ that solve the minimization problem reported in equation 2.15. In other words, there exist a unique capacity allocation (x) that solves the minimization problem for each collection of wind profiles A ;
- A is the output of the application of a PC to wind speeds of each grid point of a geographical domain at each hour of a prescribed period. In other words, matrix A is a function of the PC;
- The PC's shape is unknown.

As a conclusion, it is possible to state that the only concrete variables in this optimization problem are the parameters of the PC, since all the remaining quantities are an implicit function of them.

Starting from these observations, a Particle Swarm Optimization (PSO) scheme is implemented into a bigger optimization procedure:

- At each iteration, a swarm of PC parameters is found starting from the previous one through the PSO algorithm;
- For each set of PC parameters, the spatial distribution that optimizes the distance between the simulated power profile and the real one is found;
- A new swarm of PC parameters is obtained, until convergence conditions are satisfied or a maximum number of iterations is reached.

Particle swarm optimization is a social stochastic scheme, graphically represented in the flowchart in figure 2.11. Its idea originated from the observation of behavior of large flocks of birds, schools of fishes or swarms of insects. In fact, all of their movements follow two concurrent dynamics:

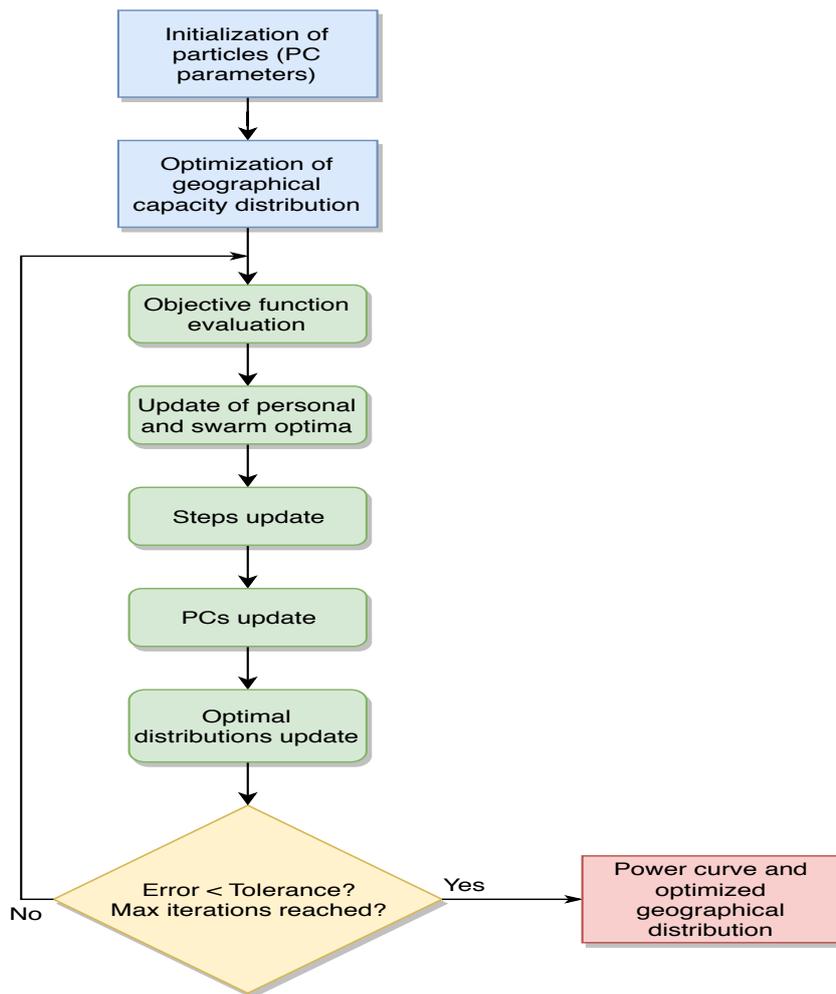


FIGURE 2.11: Flowchart of the particle swarm optimization scheme.

- **Movements of the single individual:** each member of the group collaborates/competes with its neighbors individually;
- **Movement of the whole group:** on a global scale, the set of individuals behaves as a single entity.

Starting from this fascinating movement logic, a computational scheme is modelled in order to emulate it. From this moment, the word *particle* will be used for each individual of the group, while *swarm* will correspond to the group itself.

Modelling of the particle's movements

The movement of each particle in the swarm is modeled through three factors:

- **Cognitive behavior:** each particle is attracted by the optimum found in its movements;
- **Social behavior:** the same particle is also attracted by the global optimum the swarm has found so far;

| Particle's behavior | Representative parameter |
|---------------------|--------------------------|
| Cognitive behavior | ω |
| Social behavior | ϕ |
| Inertia | m |

TABLE 2.3: List of particle's behaviors and representative parameters.

- **Inertia:** a mass (inertia) is assigned to particles. In fact, in the same way a body will oppose resistance to a force due to its mass, a particle will tend to conserve its original cinematic state.

Table 2.3 links each tendency of the particle to its corresponding representative parameter.

In practice, the model is implemented as it follows. At each iteration, the updated position of the particle x_{k+1}^i is:

$$x_{k+1}^i = x_k^i + v_{k+1}^i \quad (2.19)$$

Where v_{k+1}^i corresponds to the particle's velocity at the k+1-th iteration, calculated through the equation below:

$$v_{k+1}^i = mv_k^i + \omega r_1(p_k^i - x_k^i) + \phi r_2(p_k^g - x_k^i) \quad (2.20)$$

Terms r_1 and r_2 are random numbers between 0 and 1. p_k^i and p_k^g represent the optimum found by the single particle and the one found by the swarm at the k-th iteration.

As regards x_0^i and v_0^i , they are randomly initialized.

It is interesting to notice how simple the algorithm and its implementation is. The only values that must be set are behavioral parameters, size of the swarm and convergence parameters (maximum number of iterations and tolerance).

Stopping criteria and time complexity of the algorithm

The algorithm stops iterating in two possible conditions:

- The maximum number of iterations is reached;
- Variations in swarm's optimum are below the prescribed tolerance.

Time complexity of the scheme is $O(\alpha)$ per function evaluation, where α is the size of the swarm. If time complexity of function application is taken into account, time complexity is $O(\alpha n)$, where n is the size of matrix A in equation 2.15.

2.6.2 Heuristic approach

Observations of PSO logs report that there is a poor dispersion of local minima: they all tend to concentrate in a small space of parameters combination and the objective function assumes really similar values in all those points. This leads to two important conclusions: the first one is that all those points can satisfy the required level of precision of the model and the second one is that, if the starting point of a

gradient based method is close enough to one of those local optima, the algorithm will successfully converge. These two important observations constitute the basis for a heuristic approach, which led to very fast and precise enough results.⁵ The algorithm is articulated as it follows:

- A first estimate of the power curve is produced by approximating the whole geographical area with a single node, where wind speed time series corresponds to the average of all the time series of each node. In this way, it is possible to produce a scatter plot that associates wind power production of a geographical area to wind speeds;
- A fitting procedure is applied, in order to extrapolate a PC from the scattered data. The Levenberg–Marquardt algorithm has been adopted for this fitting [19];
- Starting from this first estimation of PC, optimal parameters for the PC are found with a Quasi-Newton scheme, under the assumption of homogeneous capacity distribution;
- A new optimal geographical capacity distribution is found through the solution of its inverse mathematical problem. Lasso Cross-validation is used to regularize the problem;
- The capacity distribution works as an input for the next iteration, where new PC parameters will be determined through the minimization of the L2 norm between the estimated power series profile and the actual one, like in the PSO.

The last three steps of the procedure are repeated until convergence or the maximum number of iterations is achieved. For better clarity, the scheme is also reported in the flowchart in figure 2.12.

⁵In the section dedicated to validation of the models, it will be possible to verify how this method gives results similar to more accurate but slower methods.

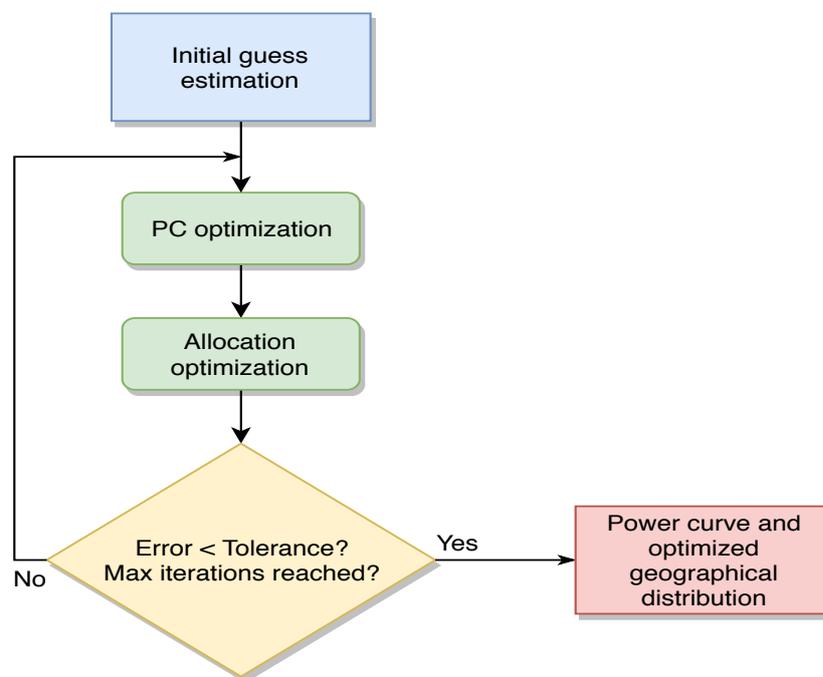


FIGURE 2.12: Flowchart of the heuristic optimization scheme.

Chapter 3

Validation of models

In this chapter, models are tested against production data through five Shuffle-Split cross validations. In a first section, the two proposed wind power curves are validated with a fixed spatial distribution. In the special case of Denmark, thanks to data availability, capacity distribution of every day of 2017 is available. Thus, the model is validated against the actual wind capacity distributions through the whole year of bidding area DK2. Figure 3.1 reports all the bidding zones of nordic countries.

In the second section, the complete allocation algorithm is tested in two ways:

- Simulated time series are validated against actual WPP time series;
- Simulated capacity distribution is compared against the year-ending one [20] [21] [22] [23]. In this case, the former assumes the physical meaning of an average wind capacity distribution. In fact, theoretically, in order to get a daily capacity distribution, the simulated period should be reduced to one day. However, the model would fail, since the available samples for the calibration of the model should be reduced to 24 and they would be too few.

As regards simulated geographical zones in this part, Finland, Sweden, Norway and Denmark are selected. In particular, for each country a bidding zone is studied singularly.

3.1 Validation procedure

The validation procedure is described in detail in this section. As a first step, it is verified that approximating one PC for a geographical region can give good results. In order to assess the quality of the approximation, three indices are taken into account:

- **R² score**: this value, also called *coefficient of determination*, indicates the percentage of variability explained by the model and it's widely used to test the predictivity of regression models. If this value is calculated from training data, its value is limited between 0 and 1. However, since this value is obtained from test data in this case, it only has an upper limit of 1 (the best score). If an estimator f , observations y_i and their average value \bar{y} are introduced, R^2 is mathematically defined in the following way:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (3.1)$$

$$SS_{res} = \sum_i (y_i - f_i)^2 \quad (3.2)$$

$$SS_{tot} = \sum_i (y_i - \bar{y})^2 \quad (3.3)$$

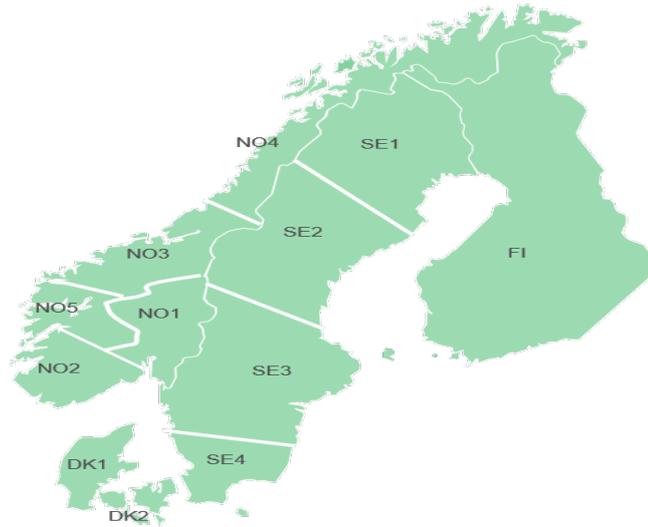


FIGURE 3.1: Nordic bidding areas. [24]

It is observable that if the model is perfect, $f_i = y_i$ for every i , so R^2 converges to 1.

- **MAE:** MAE stands for Mean Absolute Error. As the acronym already explains, it corresponds to the average value of the n absolute values of errors e_i :

$$e_i = y_i - f_i \quad (3.4)$$

$$MAE = \frac{\sum_i |e_i|}{n} \quad (3.5)$$

- **MAD:** MAD stands for Median Absolute Deviation. It corresponds to the median of the absolute values of errors e_i , defined in equation 3.4.

As regards capacity distributions, the allocation error should take into account two important factors:

- Geographical pattern of the estimated capacity with respect to the real one;
- Magnitude of estimated capacity in a certain sub-area with respect to the actual one;

In other words, due to the typically high number of nodes in a geographical area, it is practically impossible that the estimated capacity distribution will ever perfectly coincide with the real one. However, this does not mean that estimations are wrong. For example let us suppose that, in a particular node in Finland, 10% of the country's total capacity is installed. Let us also suppose that the model allocates the same identical amount of capacity in a node right after the actual one. Given that Finland has a surface of $338,424 \text{ km}^2$ and that neighboring nodes have a distance of approximately 30 km (40 km if they are diagonally placed), the error committed by the model can be considered null. In an analogous way, if in one node there is a certain amount of actual wind capacity and the model estimates the same wind capacity, but in many

neighboring nodes, results can still be considered valid ¹. Thus, a clustering of the capacity geodata becomes necessary. In this way, it is possible to represent a particular sub-area of a wider zone with an equivalent node, in which installed capacity corresponds to the total capacity in that sub-area. After that, the allocation error for a single cluster is calculated as the absolute value of the difference between the actual cluster capacity and the simulated one. Finally, the total allocation error is calculated as the sum of all the allocation errors of each cluster.

For these particular applications, Spectral Clustering and Agglomerative Clustering from the Scikit-Learn Python package have been adopted due to their good performances [25].

3.2 Power curves validation

Results of the first validation step are reported in table 3.1 ². For each PC, BH and PSO are adopted as optimization schemes. However, the two numerical schemes always converge to the same solution, so the four optimization procedures results are condensed into two. It is possible to notice that both of the PCs are able to describe WPP effectively. Specifically, it is possible to conclude that both of the modelled PCs assume the same optimal shape. Figure 3.3 shows this result graphically. Plots 3.2-3.4, instead, give a better description of WPP patterns. It is possible to notice that the model tends to underestimate peak load production, as can be observed in figure 3.2a. A further confirmation of this feature can be obtained from figures 3.2b-3.4. In fact, while low load hours are described with relatively small errors, intermediate and high load may present higher offsets. An example of this tendency can be observed in figure 3.4, between the 17th and the 24th of January: the simulated WPP pattern is the same of the realised one, but there is a systematic error that leads to a constant overestimation. An analogous kind of error can be found around the 7th of February, with the difference that there is a systematic underestimation. Nevertheless, in other cases this underestimation (overestimation) is very small, so the simulated time series almost perfectly match realised production, like between the 3rd and the 17th of January. Furthermore, the model is able to describe sudden rampings of the realised production.

Another observable limit of the model consists in the underestimation of hours in which production is equal to zero. This limit can be observed in the general time series in figure 3.2b or more clearly in figure 3.2a, in correspondence of a normalized power production of zero. An explanation of this phenomenon is that the sigmoid adopted in the WPP modelling converges to zero asymptotically. As a consequence, probabilities that the total production of an area is zero are almost null.

Thanks to very similar test errors, it is possible to conclude that both of the PCs are able to describe WPP accurately. Moreover, convergence of both methods is achieved in a comparable number of iterations. However, while the Piecewise function depends on four parameters, the Sigmoid function is built with three. Also, from a numerical point of view, the two curves are optimized through different schemes. In fact, while the Piecewise function is optimized with a SLSQP scheme due to constraints in its correspondent optimization problem, the Sigmoid function is shaped through a L-BFGS-B scheme, suitable for bounded optimization problems. On top of

¹This logic has also sense in the opposite way: if a locally homogeneous distribution is estimated by a concentrated capacity in one node by the model, estimation can still be considered correct.

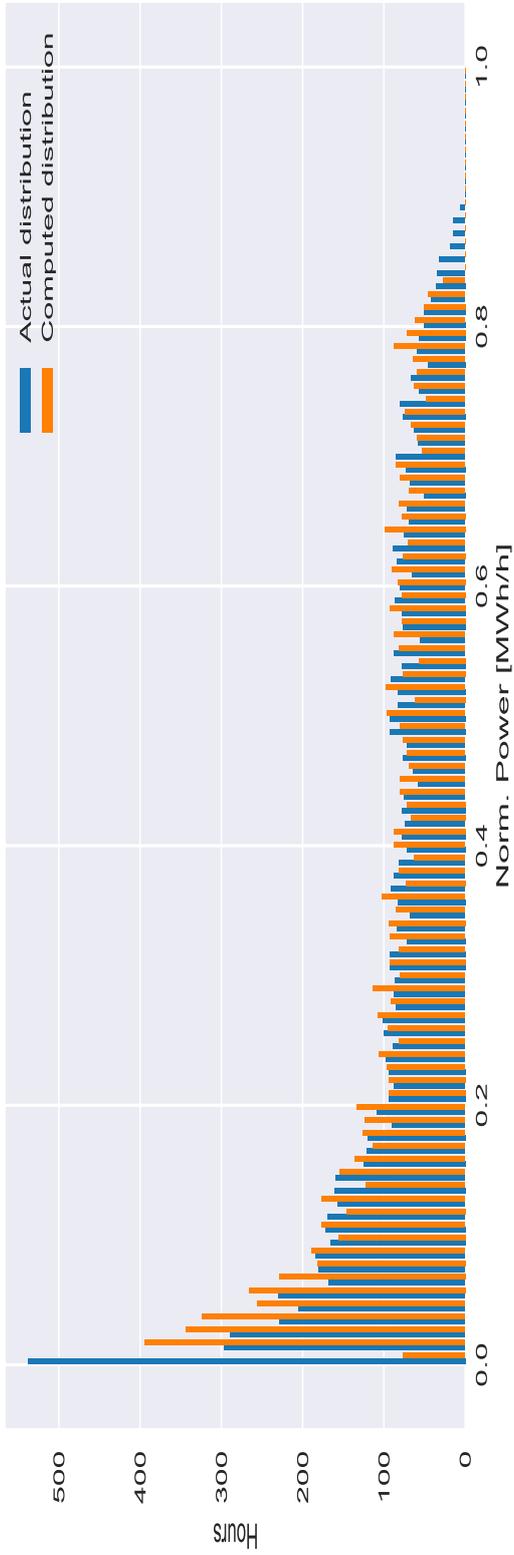
²Results between parenthesis in the table are standard deviations. For example, 3.61(165) means 3.61 ± 1.65 and 5.60(43) means 5.60 ± 0.43 . If the standard deviation is not written, it is equal to zero.

| Test errors | Piecewise function | | Sigmoid | |
|-----------------|-----------------------------|-----------|----------------------------|-------------|
| R^2 score [%] | 94.09(32) | | 94.10(32) | |
| MAE | $4.412(100) \times 10^{-2}$ | | $4.401(96) \times 10^{-2}$ | |
| MAD | $2.991(99) \times 10^{-2}$ | | $3.024(87) \times 10^{-2}$ | |
| Parameters | Piecewise function | | Sigmoid | |
| c | 10.141(6) | | 10.138(9) | |
| p | 1.961(10) | | 1.849(6) | |
| z | 0.848(1) | | 0.839(1) | |
| η | 8.408(41) | | // | |
| Scheme | P. F., PS | P. F., BH | Sigmoid, PS | Sigmoid, BH |
| No. iterations | 27(6) | 40(6) | 24(4) | 61(18) |

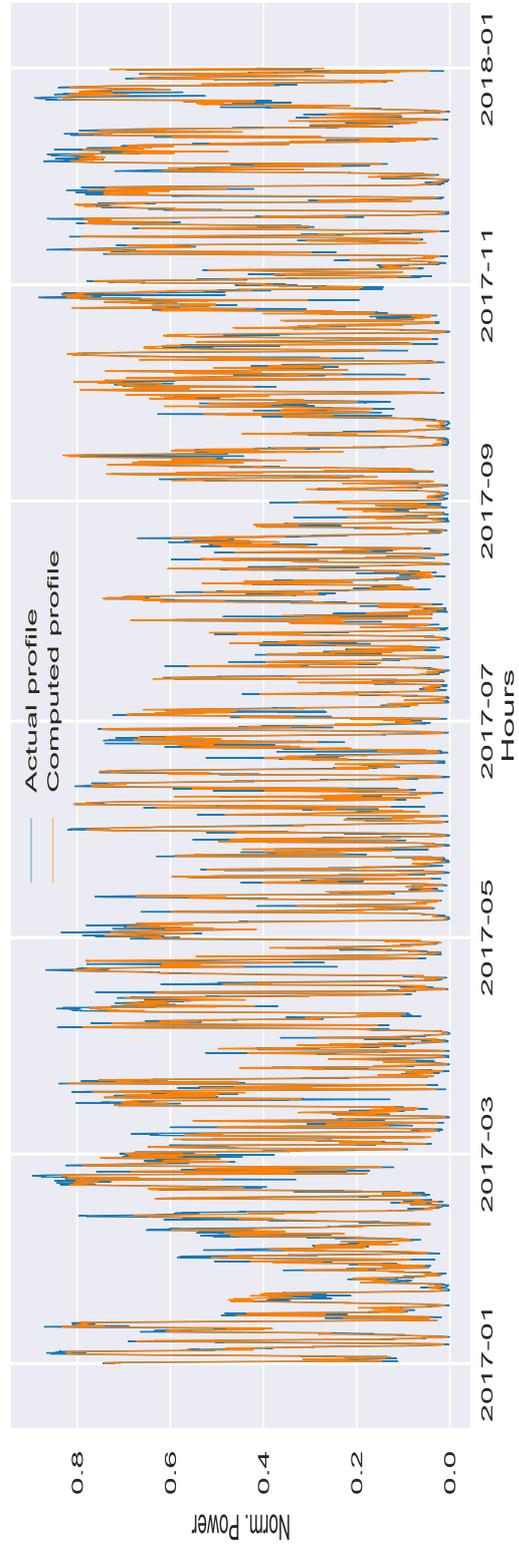
TABLE 3.1: Comparison of results for Piecewise function and Sigmoid, DK2. P. F.: Piecewise Function.

that, L-BFGS-B is computationally cheaper than SLSQP. In fact, L-BFGS-B presents a time complexity of $O(nm)$, where n corresponds to the number of variables of the problem and m to the number of previous optimization steps stored into memory. SLSQP, instead, has a typical time complexity of $O(mn^2)$, where n assumes the same meaning and m is the maximal number of active constraints³. It is possible to get more details of both schemes and their comparison in the relative appendix B. As a conclusion, the Sigmoid curve represents a better option for the description of WPP, due to its greater simplicity and faster optimization procedure. Thus, the following steps of the validation are made with this power curve.

³Given a feasible point of an optimization problem, an inequality constraint of the form $c(x) \leq 0$ is active in that point if $c(x) = 0$.



(A) Power output distribution: Sigmoid curve v. validation data.



(B) Time series: Sigmoid curve v. validation data.

FIGURE 3.2: Power output distributions and time series for bidding zone DK2 with known geodata.

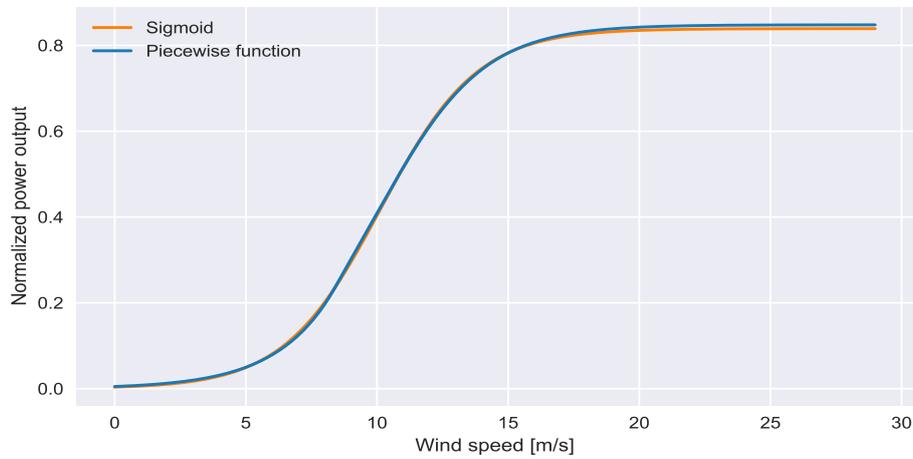


FIGURE 3.3: Comparison of the average simulated power curves for bidding zone DK2 with known geodata. The two power curves are almost identical.

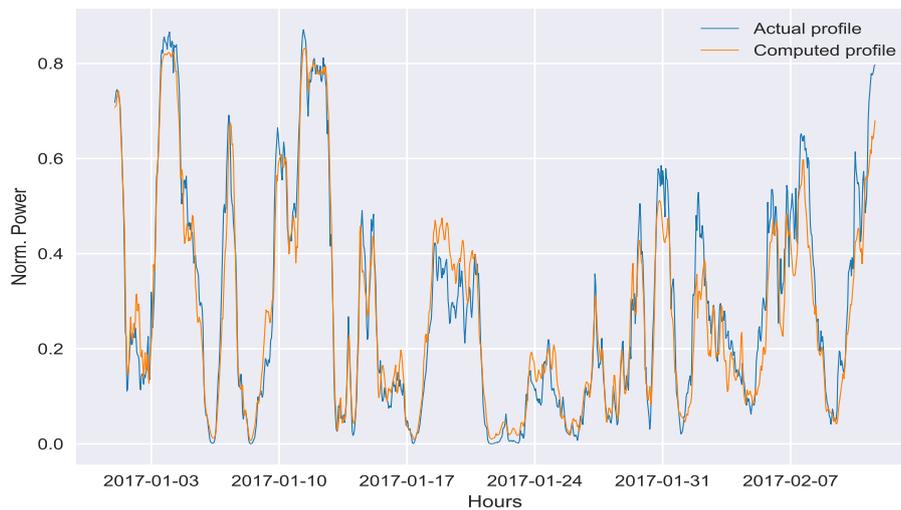


FIGURE 3.4: Time series zoom, DK2, fixed geodata: Sigmoid curve. General pattern of wind power production is successfully described by the model, especially in case of sudden ramping phenomena. However, it is possible to notice the presence of underestimations (overestimation) of peaks in power outputs.

3.3 Complete model validation

3.3.1 Sweden

As regards Sweden, bidding zone SE1 is selected, due to its sparse capacity distribution. In this way, it is possible to test how accurate allocation of wind capacity can be, in both of the possible approaches.

Table 3.2 reports quantitative measures of accuracy obtained for this area and the number of necessary iterations in order to achieve convergence. In order to make understanding of the table easier, supportive plots accompany it. A first interesting observation that can be done is the amount of necessary iterations for the convergence of the solution. While PSO needs an average of 17 iterations, the heuristic method is able to converge in only 3 iterations. Moreover, if the time complexity of each iteration is taken into account, differences in convergence speed become even stronger. In fact, one optimization step of the Heuristic scheme corresponds to one optimization step made by one particle in the swarm. To make an example, if a swarm is composed by fifty particles, PSO will be fifty times slower than the heuristic method.

Standard deviation of test errors is low, which shows that the model has been trained on a sufficient amount of samples. Furthermore, high similarity of test errors and figure 3.5 make the observer suspect that both methods converged to approximately the same solution, since PCs are practically the same.

Quality of capacity allocation can be checked by comparing the real year-ending distribution and the simulated one in figure 3.8. In particular, subplot 3.8a shows how each node of the map is clustered, in order to calculate the allocation error. It is possible to observe that accuracy of allocation is good as well, due to an acceptable total error. On top of that, it is clearly visible that the difference between the two allocation methods is not the pattern of the estimated wind capacity, but the amount of capacity allocated in each cluster, as can be observed from colorbars and in the total estimated capacity in table 3.2. As a consequence, in the allocation error it is also necessary to take into account that the total estimated wind capacity does not necessarily coincide with the real one. In fact, in order to keep the adopted schemes numerically stable, constraints on total capacity have been relaxed. As a conclusion, allocation errors are influenced by both the single cluster allocation error and the total estimated capacity of an area, which may not be equal to one.

A final verification of the two models allocations can be made by comparing their resulting spatial distribution. Using the same clustering logic adopted for validation of capacity distributions, it is possible to evaluate how similar estimated capacities are. Allocation difference in table 3.2 shows that average cluster differences are really low. This means that resulting distributions are very similar and this obviously confirms the small differences in allocation errors.

Observations of curves in figures 3.6-3.7 complete the validation by showing strength and weak points of the model ⁴. At first glance, it is possible to notice that there is a general accurate yearly power distribution in plot 3.7a. However, the model constantly underestimates peak loads, as it is possible to notice at tails of the power output distributions. A further confirmation of this underestimating behavior is confirmed by observing year 2017 time series in figure 3.7b. However, except for this issue, WPP patterns are accurately described by the model. A closer look to a section of time series in figure 3.6 permits to get more details of the model. A sudden

⁴Resulting plots of PSO have not been included because it is impossible to notice differences from results of the Heuristic scheme.

| Parameters | PS scheme | Heuristic scheme |
|-----------------------|-----------------------------|----------------------------|
| c | 6.921(11) | 6.824(12) |
| p | 1.233(7) | 1.190(10) |
| Total capacity | 0.961(3) | 0.946(5) |
| Test errors | PS scheme | Heuristic scheme |
| R^2 score [%] | 92.50(37) | 92.49(39) |
| MAE | 5.073(97) $\times 10^{-2}$ | 5.075(98) $\times 10^{-2}$ |
| MAD | 3.670(113) $\times 10^{-2}$ | 3.659(95) $\times 10^{-2}$ |
| Iterations | 17(2) | 3 |
| Allocation error | PS v. Reality | Heuristics v. Reality |
| Cluster 1 | 2.07(16) $\times 10^{-2}$ | 2.48(17) $\times 10^{-2}$ |
| Cluster 2 | 3.35(68) $\times 10^{-2}$ | 2.57(61) $\times 10^{-2}$ |
| Cluster 3 | 1.04(5) $\times 10^{-1}$ | 1.10(5) $\times 10^{-1}$ |
| Cluster 4 | 4.51(322) $\times 10^{-3}$ | 2.20(174) $\times 10^{-3}$ |
| Cluster 5 | 4.86(38) $\times 10^{-2}$ | 5.45(37) $\times 10^{-2}$ |
| Total A. error | 2.12(9) $\times 10^{-1}$ | 2.17(10) $\times 10^{-1}$ |
| Allocation difference | PS v. Heuristics | |
| Cluster 1 | 4.13(55) $\times 10^{-3}$ | |
| Cluster 2 | 7.78(113) $\times 10^{-3}$ | |
| Cluster 3 | 5.18(52) $\times 10^{-3}$ | |
| Cluster 4 | 4.45(76) $\times 10^{-3}$ | |
| Cluster 5 | 5.95(76) $\times 10^{-3}$ | |
| Total difference | 2.75(24) $\times 10^{-2}$ | |

TABLE 3.2: Comparison of results for PS and Heuristic method, SE1.

drop of WPP from the 25th of September and the 2nd of October is accurately described by the model. It is also possible to notice that majority of errors concentrate in peak load hours, as already noticed in power output distributions plots.

The underestimation of null power production is also present, as already found in the first step of the validation. Unfortunately, this error is inherently part of the model itself, so it can't be removed without changing the PC modelling radically.

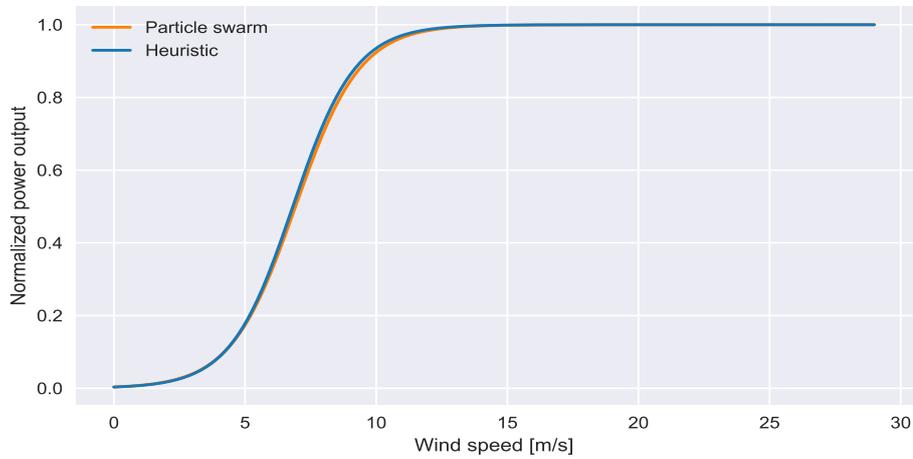


FIGURE 3.5: Comparison of the average simulated power curves for bidding zone SE1. Differences between the two power curves are almost absent.

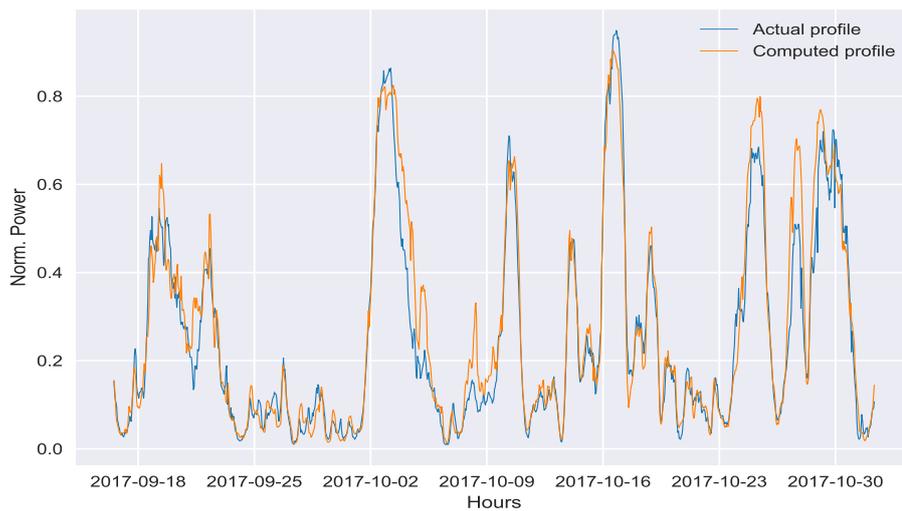
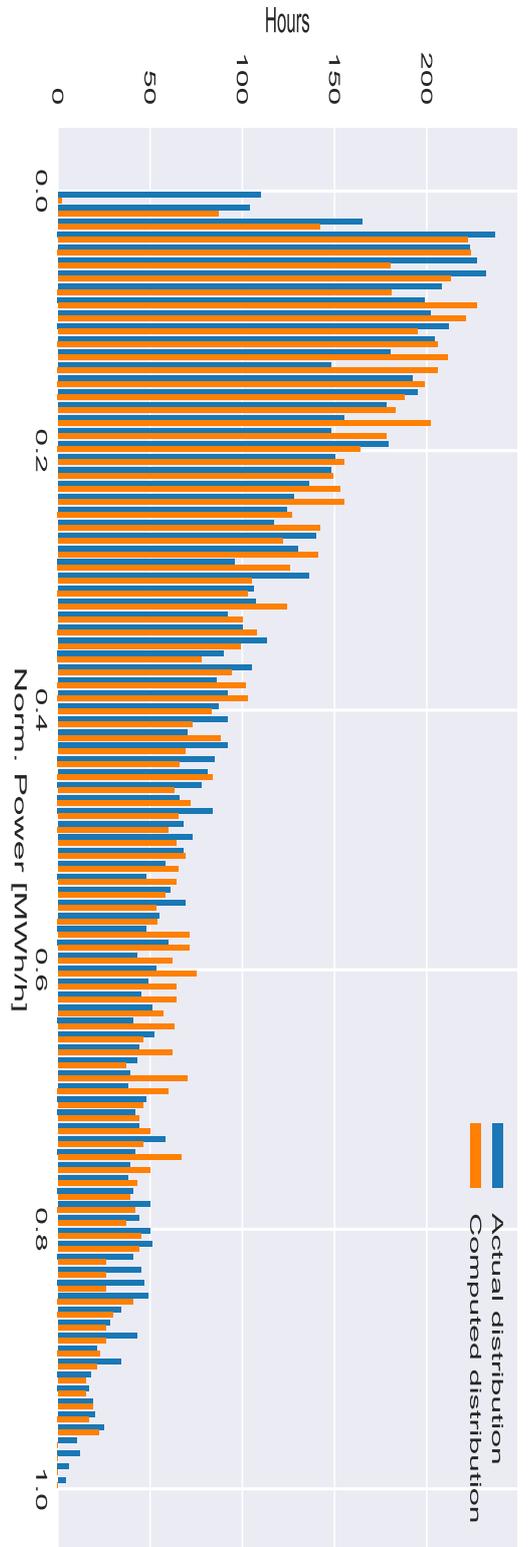
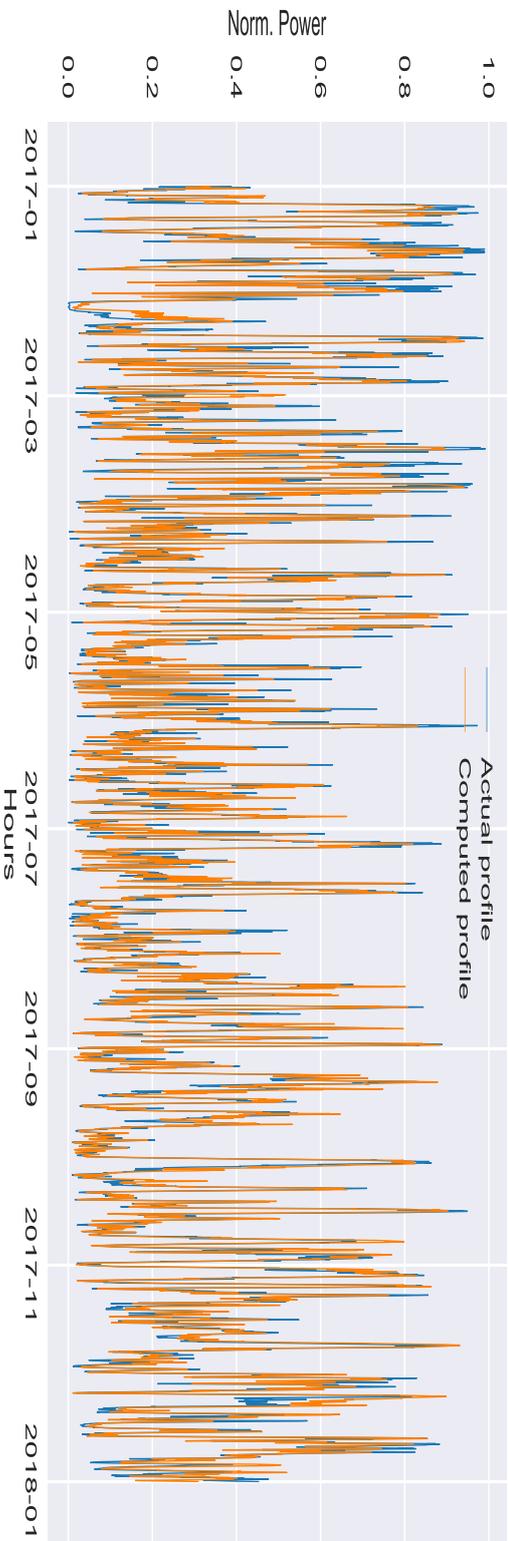


FIGURE 3.6: Time series zoom, SE1: Heuristic scheme. The model is able to describe low loads very accurately, while higher loads are described with a relatively larger error.

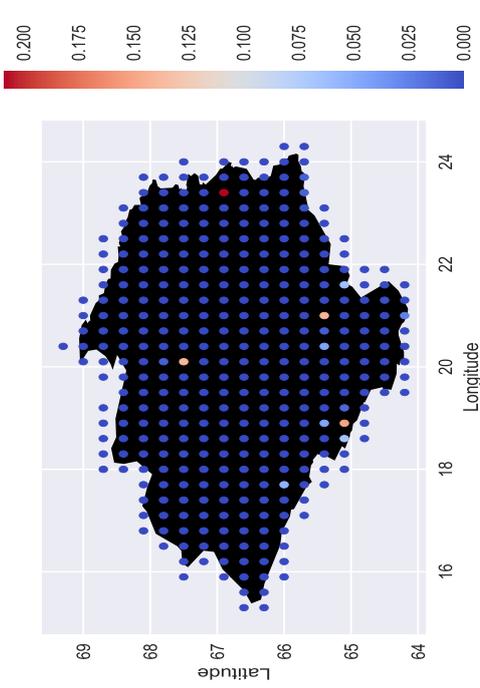


(A) Power output distribution: heuristic scheme v. validation data.

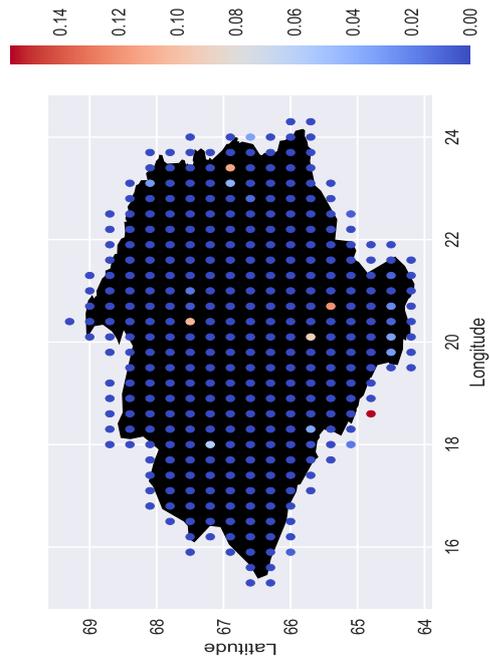


(B) Time series: heuristic scheme v. validation data.

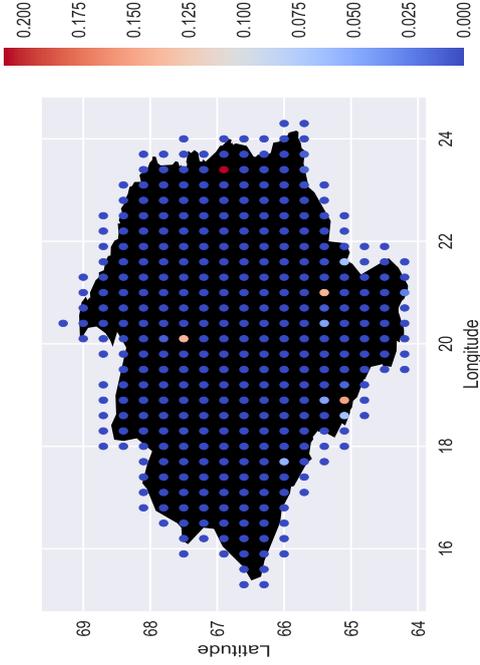
FIGURE 3.7: Power output distributions and time series for bidding zone SE1.



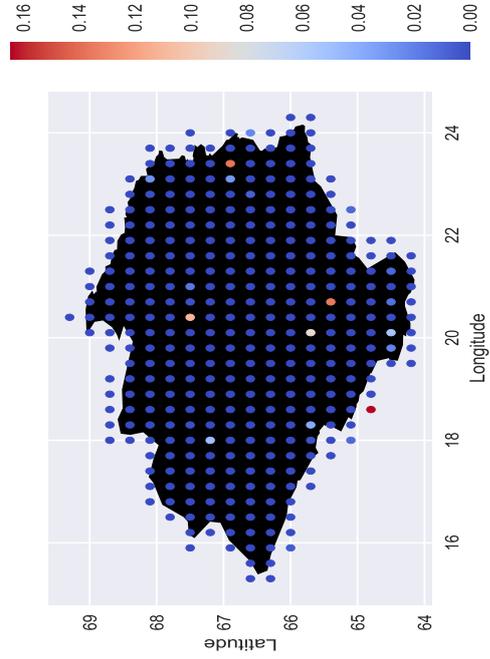
(A) Wind capacity distribution: clustering of the area



(C) Wind capacity distribution: heuristic scheme.



(B) Wind capacity distribution: actual distribution.



(D) Wind capacity distribution: PS scheme.

FIGURE 3.8: Comparison of wind capacity distributions for bidding zone SE1. It is possible to notice a very similar pattern in capacity allocation among the simulations and reality. Differences are in the entity of these allocations.

3.3.2 Norway

In the case of Norway, NO₃ is tested. In fact, the diffused presence of mountains and the concentration of wind capacity close to the Norwegian Sea constitutes an interesting setting for testing the model.

Results of the validation can be found in table 3.3. First of all, it is possible to notice how PSO and the Heuristic schemes converged to different solutions but with comparable test errors, as can be verified in figure 3.9. Diversity in results was expected for this area, due to abundance of mountains. As a matter of fact, this leads to bigger gradients in velocity distributions and more difficulties to achieve a global minimum. Consequently, capacity distributions are allocated in different ways, since they are optimized according to the shape of the PC. In fact, as reported in figures 3.10c-3.10d, the Heuristic scheme tends to return a sparser solution with respect to the PSO. This higher sparsity makes the estimated distribution more similar to the actual one, reported in figure 3.10d. However, allocation errors reported in table 3.3 are similar in both approaches and allocation differences are small. This means that, even if the PSO scheme returns a solution with a less pronounced sparsity, on a smaller scale, the estimation of the total cluster capacity is still accurate enough. Furthermore, even if the Heuristic scheme is able to better estimate the sparse pattern of the actual capacity distribution, the order of magnitude of estimated wind capacities is quite different from the real one. Indeed, while the color bar of the realized normalized capacity distribution has a maximum of 0.4, in the Heuristic scheme the maximum is slightly below 0.3. Probably, difference of the two models in the total estimated capacity (roughly five percent) balances out their sparsity differences with respect to the real capacity distribution, making the two distributions equally accurate.

Estimated power output distributions and time series are reported in figures 3.11-3.12. It is quite clear from figure 3.11 that the Heuristic scheme presents a recurrent underestimation of peak power hours, while PSO seems to be more effective. Moreover, like in the previous cases, both of the schemes show a large underestimation of hours with no production. However, time series of the two schemes are hardly distinguishable, so complete time series are reported only for the Heuristic scheme to give a general idea of the model's behaviour (figure 3.12). As a confirmation, the underestimating behaviour of the model for peak load and null production hours is visible.

A more detailed comparison of the two schemes is made in figures 3.12b-3.12c. It is clear that, despite the phenomenon of the underestimation, large rampings events are explained by the model effectively. Also low loads hours are correctly described, with relatively smaller errors.

Number of iterations performed by the two methods are similar. However, as already explained in the validation of SE1, every iteration of the Heuristic scheme is sensibly faster, since function evaluations are not repeated per each particle of the swarm, like in PSO. It is possible to have a clearer idea of how the schemes differ by reading appendix ??.

| Parameters | PS scheme | Heuristic scheme |
|-----------------------|----------------------------|-----------------------------|
| c | 9.822(75) | 8.442(71) |
| p | 1.957(20) | 1.510(18) |
| Total capacity | 0.882(5) | 0.833(5) |
| Test errors | PS scheme | Heuristic scheme |
| R^2 score [%] | 92.30(34) | 91.93(44) |
| MAE | $4.818(97) \times 10^{-2}$ | $4.914(132) \times 10^{-2}$ |
| MAD | $3.270(60) \times 10^{-2}$ | $3.381(133) \times 10^{-2}$ |
| Iterations | 18(4) | 15(1) |
| Allocation error | PS v. Reality | Heuristics v. Reality |
| Cluster 1 | $7.62(41) \times 10^{-2}$ | $8.17(38) \times 10^{-2}$ |
| Cluster 2 | $4.33(34) \times 10^{-2}$ | $3.78(29) \times 10^{-2}$ |
| Cluster 3 | $1.12(3) \times 10^{-1}$ | $1.23(6) \times 10^{-1}$ |
| Cluster 4 | $2.78(49) \times 10^{-2}$ | 0.00 |
| Total A. error | $2.60(9) \times 10^{-1}$ | $2.42(6) \times 10^{-1}$ |
| Allocation difference | PS v. Heuristics | |
| Cluster 1 | $7.26(230) \times 10^{-3}$ | |
| Cluster 2 | $5.83(312) \times 10^{-3}$ | |
| Cluster 3 | $1.02(52) \times 10^{-2}$ | |
| Cluster 4 | $2.78(49) \times 10^{-2}$ | |
| Total difference | $5.11(65) \times 10^{-2}$ | |

TABLE 3.3: Comparison of results for PS and Heuristic method, NO3.

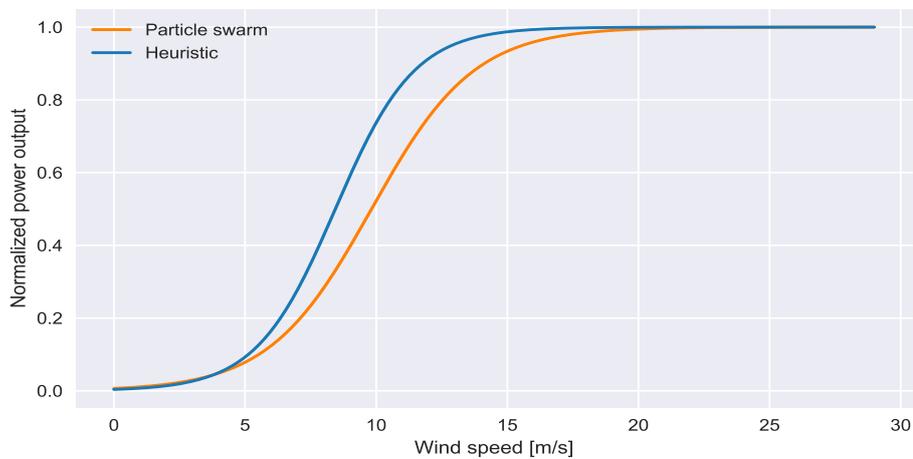
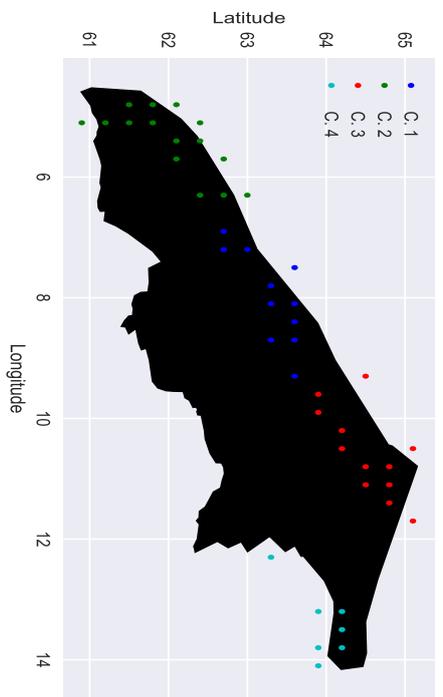
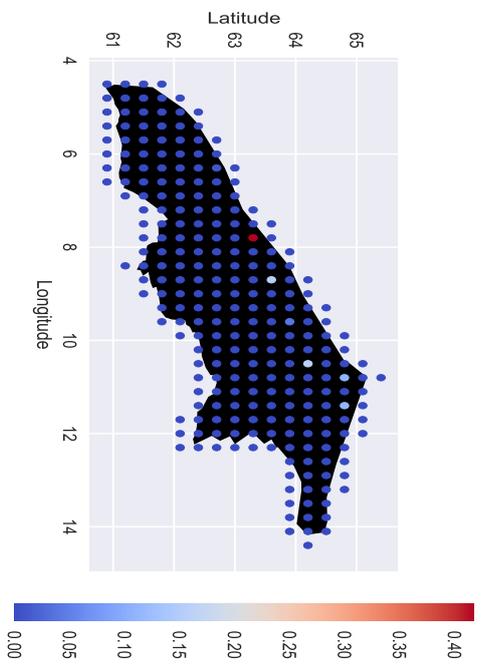


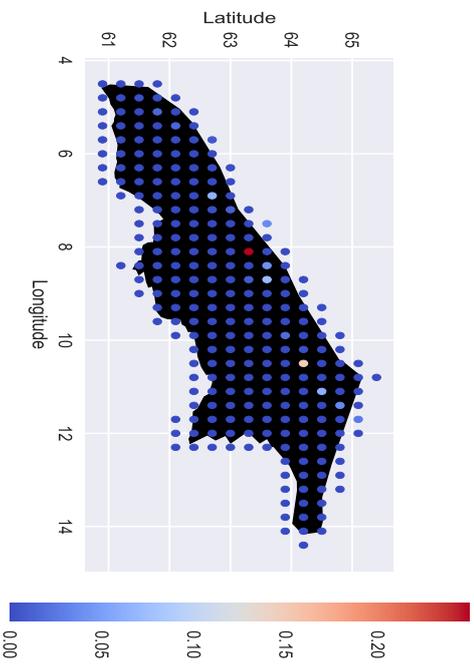
FIGURE 3.9: Comparison of the average simulated power curves for NO3. The Heuristic scheme returned a steeper power curve.



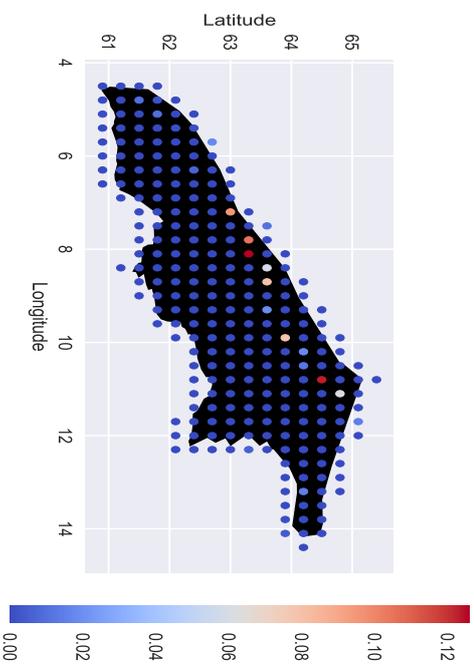
(A) Wind capacity distribution: clustering of the area



(B) Wind capacity distribution: actual distribution.

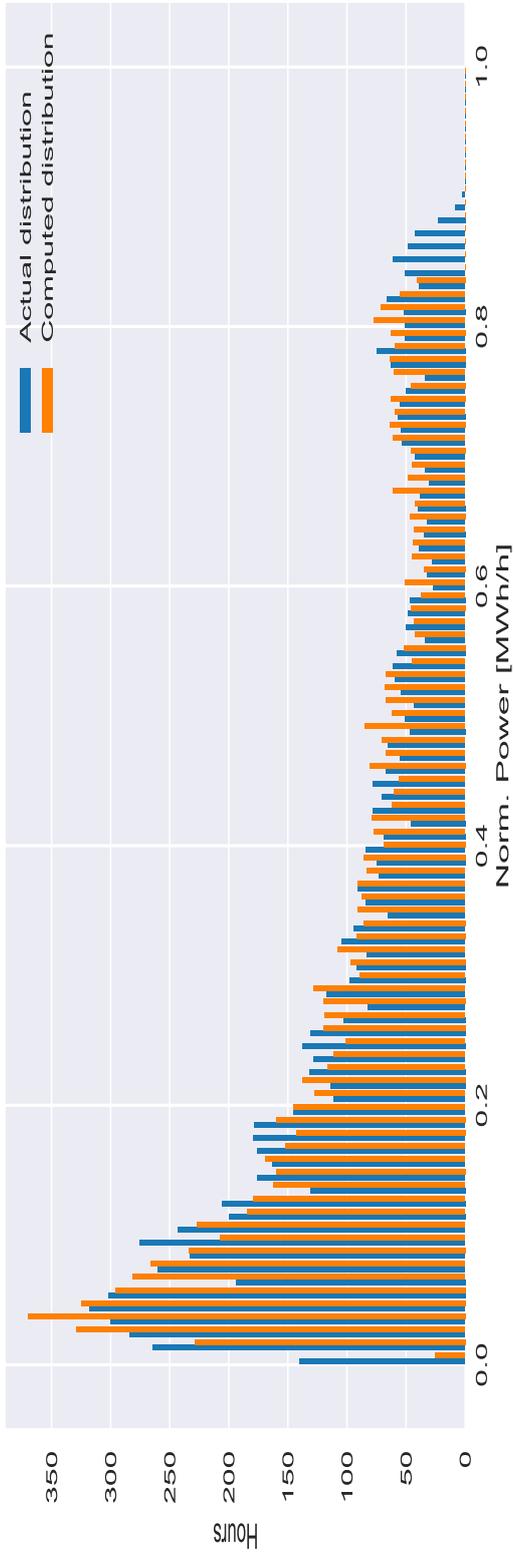


(C) Wind capacity distribution: heuristic scheme.

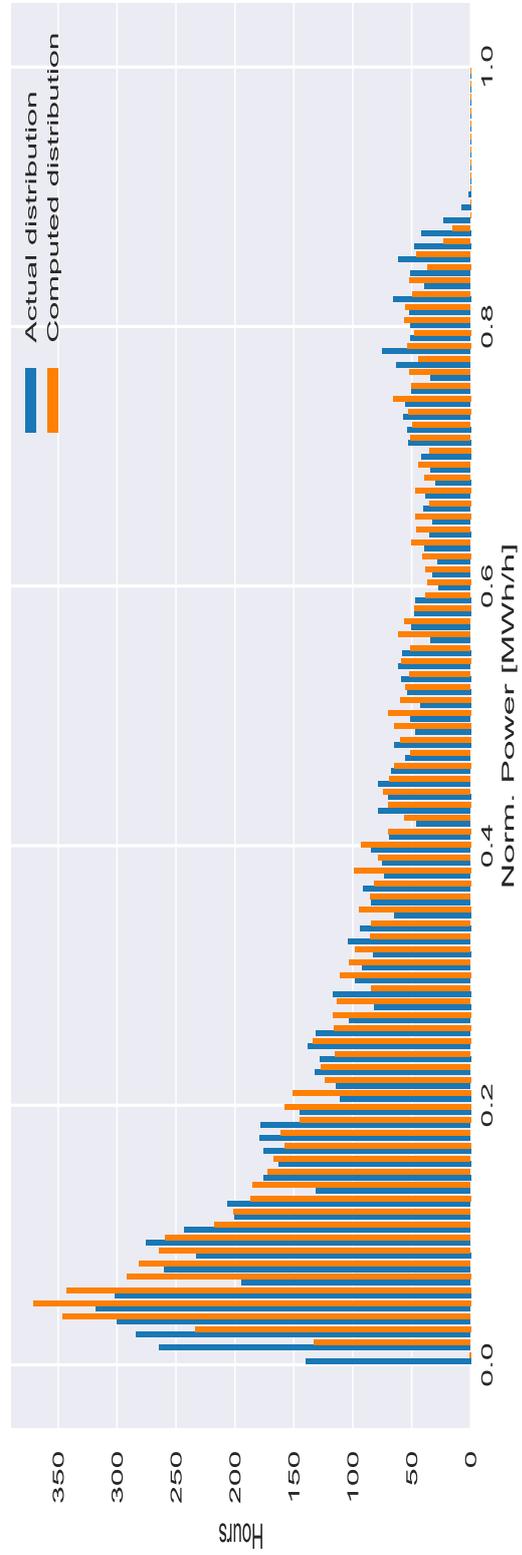


(D) Wind capacity distribution: PS scheme.

FIGURE 3.10: Comparison of wind capacity distributions for NO₃. In this case, the two schemes converged to distinguishable solutions.

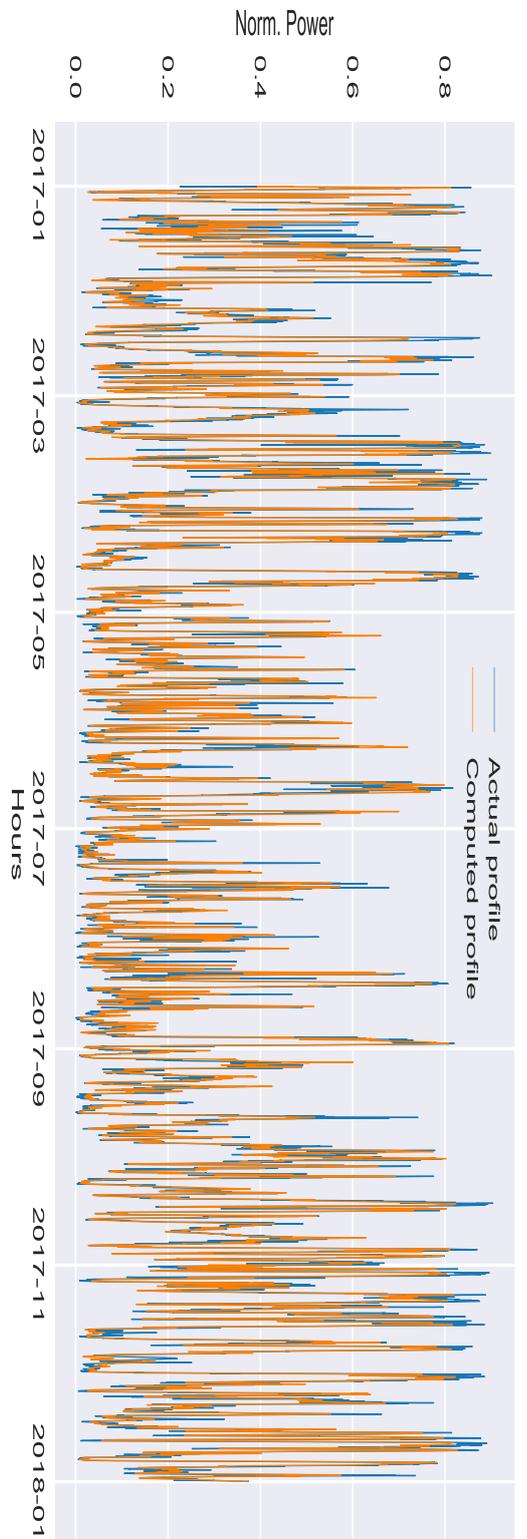


(A) Power output distribution: heuristic scheme v. validation data.

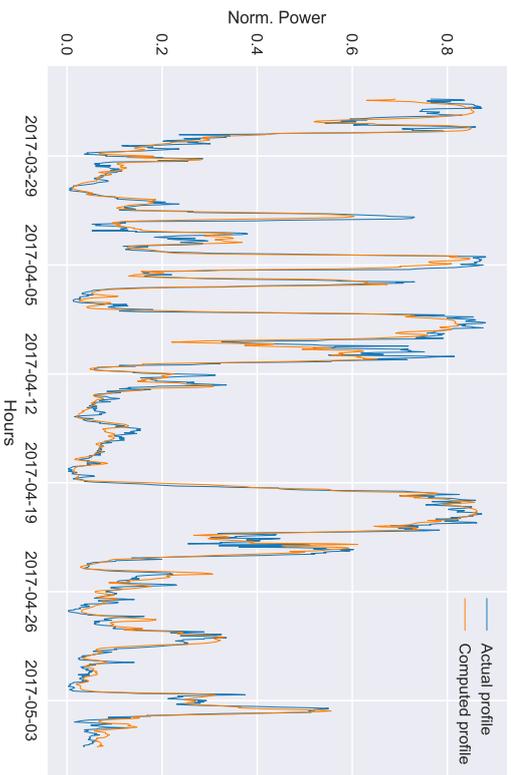


(B) Power output distribution: PS scheme v. validation data.

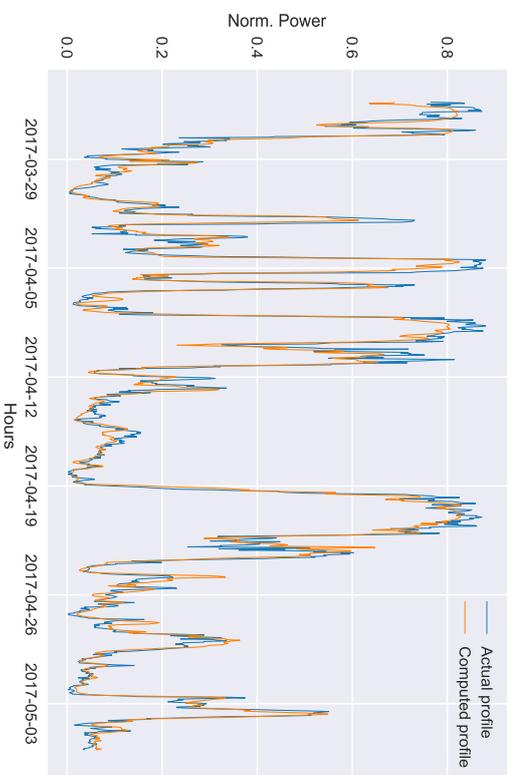
FIGURE 3.11: Comparison of power output distributions for NO₃.



(a) Time series: heuristic scheme v. validation data.



(b) Detail of time series: PS scheme v. validation data.



(c) Detail of time series: Heuristic scheme v. validation data.

FIGURE 3.12: Time series study for NO₃. Complete time series in both of the schemes are very similar, but it is possible to appreciate differences by focusing on a smaller time interval.

3.3.3 Finland

As already mentioned, Finland is the only studied case in which the whole area is not divided into sub areas, due to data availability. For this reason, a large area is simulated and this allows to test models in a more extensive way.

Results of the optimization are reported in table 3.4. The first clear result that can be observed is the presence of distinct, yet similar, solutions for the two schemes. In fact, figure 3.13 shows that resulting curves are distinguishable. As a consequence, test errors and capacity distributions are different, as it can be verified in the table and in figure 3.14. However, the general pattern of the original distribution is still effectively emulated by simulations in both approaches. Moreover, the two capacities distributions are similar on a cluster scale, since differences in capacity allocation per each cluster are relatively low. What changes significantly in the allocation procedure is the total estimated capacity, since PSO returns 0.807 and the Heuristic scheme gives 0.872. This can be identified as the principal cause of the higher total allocation error of PSO with respect to the Heuristic scheme.

In this case, PSO shows a greater tendency in underestimating WPP, as can be observed in figure 3.15. However, both of the approaches are not capable to describe hours with no production. Furthermore, from a practical point of view, differences in time series are hardly distinguishable. Thus, time series resulting from the Heuristic scheme are the only ones reported, in figure 3.16a. In order to get more details of the model's behaviour, the period between the 9th of August and 20th of October is studied in figures 3.16b-3.16c. Despite the concentration on a smaller range of hours, the two models are hardly distinguishable. However, it is possible to notice a slightly better accuracy in the description of low load hours in PSO, for example around the 29th of September. A confirmation of this fact is in figure 3.15: bins of the histogram corresponding to ranges 0.01-0.02 and 0.02-0.03 of the normalized production are more accurate in PSO than in the Heuristic scheme. From a general perspective, both of the schemes are capable to describe sudden rampings in WPP, both in case of ramp ups or ramp downs.

The convergence speed of the Heuristic scheme is particularly higher than previously. To give a concrete idea, four iterations of the Heuristic scheme are performed in less time than a single iteration in PSO. An explanation of this phenomenon is that the number of effectively active nodes in the grid are a very small portion of the total grid points contained in Finland and they all concentrate on the western coast. As a consequence, individuating these nodes becomes easier for the algorithm and a lower number of iterations is necessary.

| Parameters | PS scheme | Heuristic scheme |
|-----------------------|-----------------------------|----------------------------|
| c | 7.389(10) | 8.079(25) |
| p | 1.297(7) | 1.532(9) |
| Total capacity | 0.807(5) | 0.872(8) |
| Test errors | PS scheme | Heuristic scheme |
| R^2 score [%] | 94.76(29) | 94.46(34) |
| MAE | $3.308(97) \times 10^{-2}$ | $3.460(67) \times 10^{-2}$ |
| MAD | $2.378(113) \times 10^{-2}$ | $2.552(50) \times 10^{-2}$ |
| Iterations | 18(5) | 4(1) |
| Allocation error | PS v. Reality | Heuristics v. Reality |
| Cluster 1 | $2.49(80) \times 10^{-2}$ | $1.23(68) \times 10^{-2}$ |
| Cluster 2 | $4.20(68) \times 10^{-2}$ | $1.33(76) \times 10^{-2}$ |
| Cluster 3 | $1.26(3) \times 10^{-1}$ | $1.05(4) \times 10^{-1}$ |
| Total A. error | $1.93(4) \times 10^{-1}$ | $1.30(7) \times 10^{-1}$ |
| Allocation difference | PS v. Heuristics | |
| Cluster 1 | $1.45(26) \times 10^{-2}$ | |
| Cluster 2 | $2.87(18) \times 10^{-2}$ | |
| Cluster 3 | $2.15(16) \times 10^{-2}$ | |
| Total difference | $6.46(35) \times 10^{-2}$ | |

TABLE 3.4: Comparison of results for PS and Heuristic method, Finland.

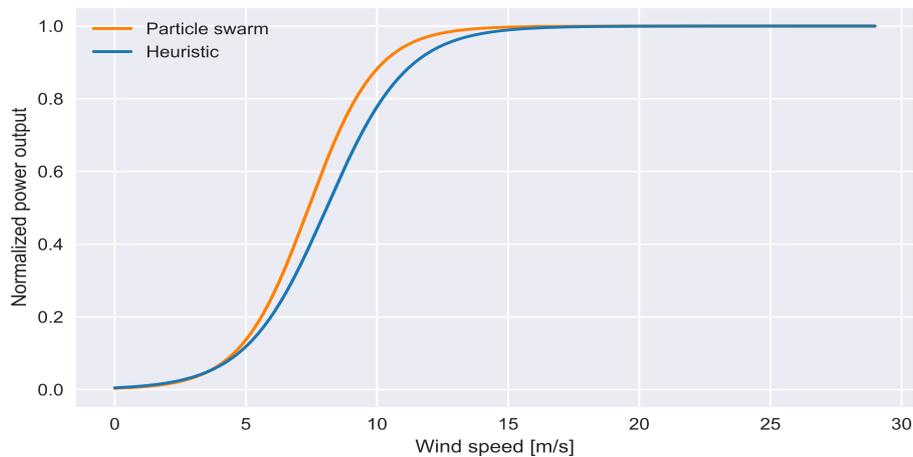
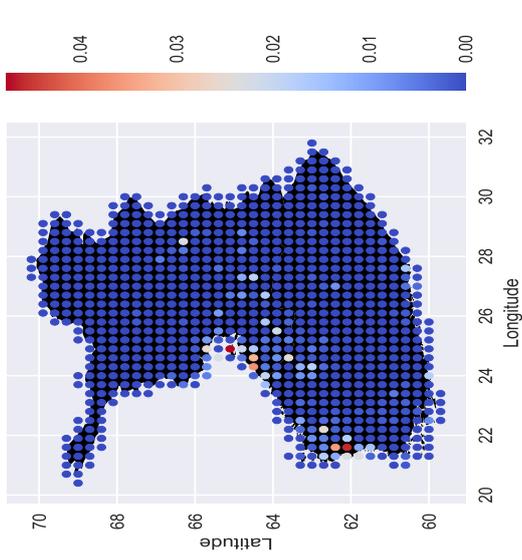
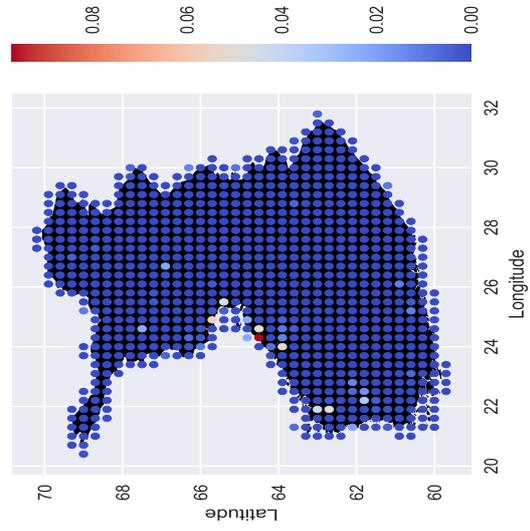


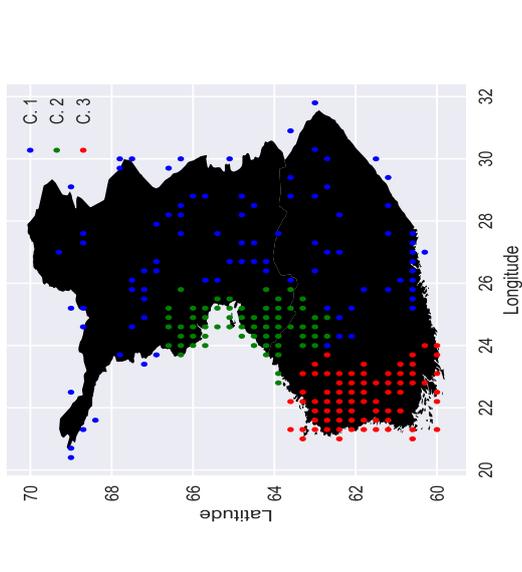
FIGURE 3.13: Comparison of the average simulated power curves for Finland. In this case, curves are distinguishable.



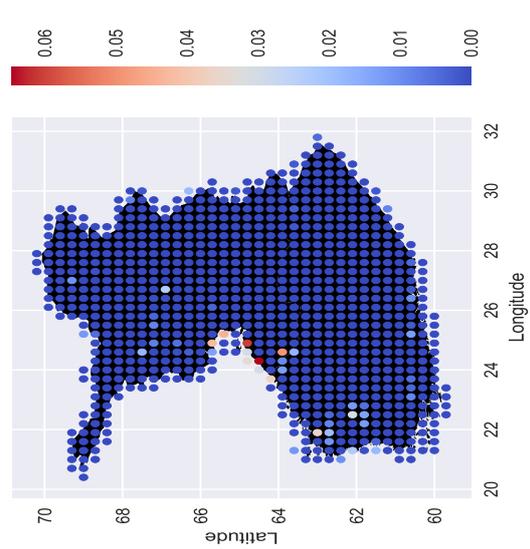
(B) Wind capacity distribution: actual distribution.



(D) Wind capacity distribution: PS scheme.



(A) Wind capacity distribution: clustering of the area



(C) Wind capacity distribution: heuristic scheme.

FIGURE 3.14: Comparison of wind capacity distributions for Finland. In this case, schemes converged to similar solutions.

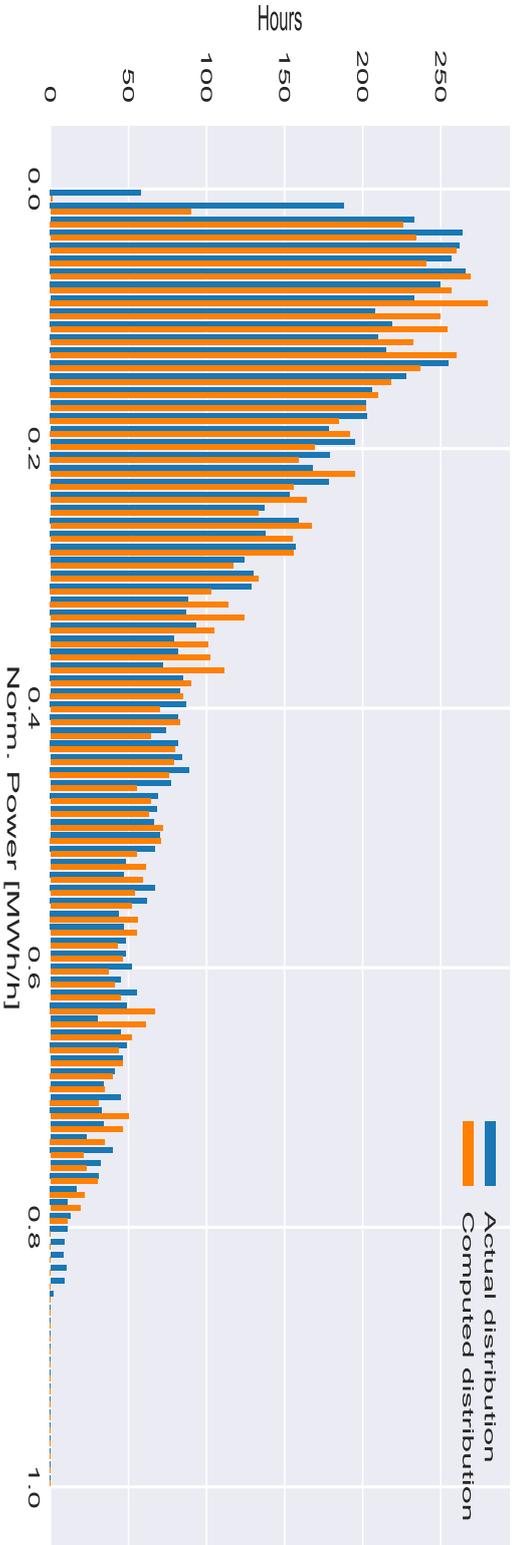
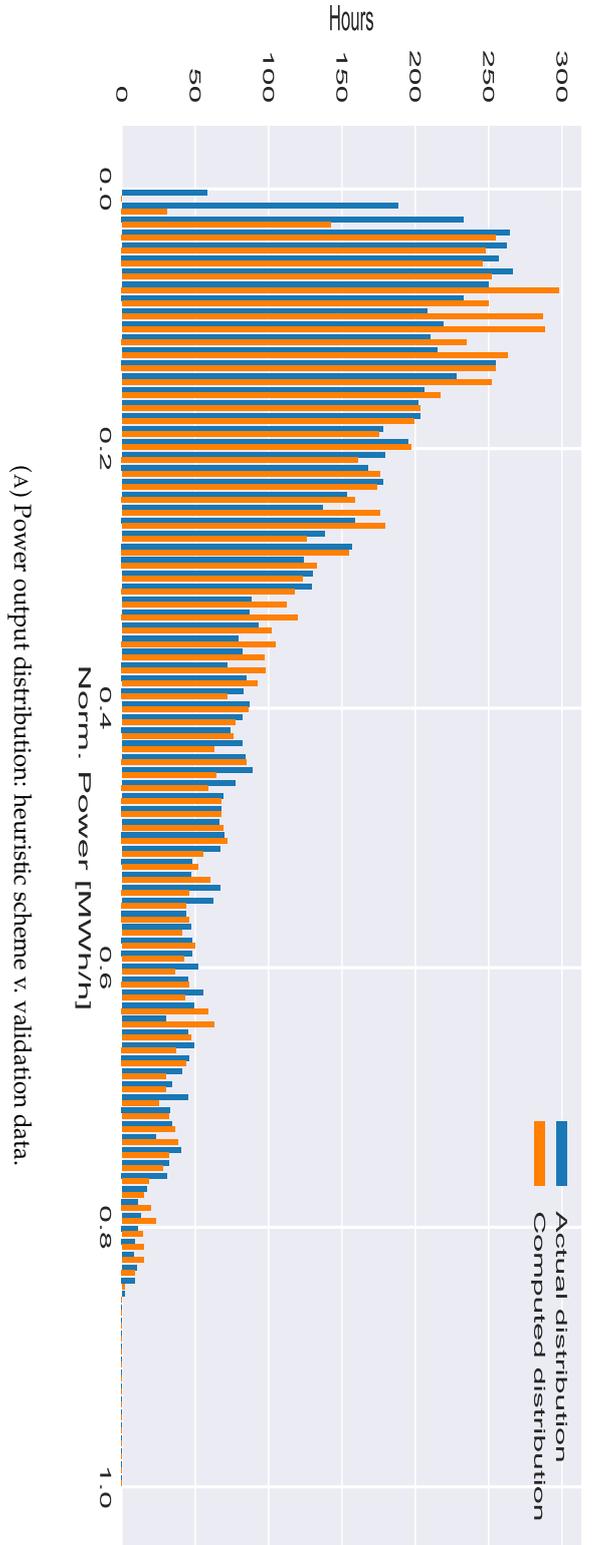
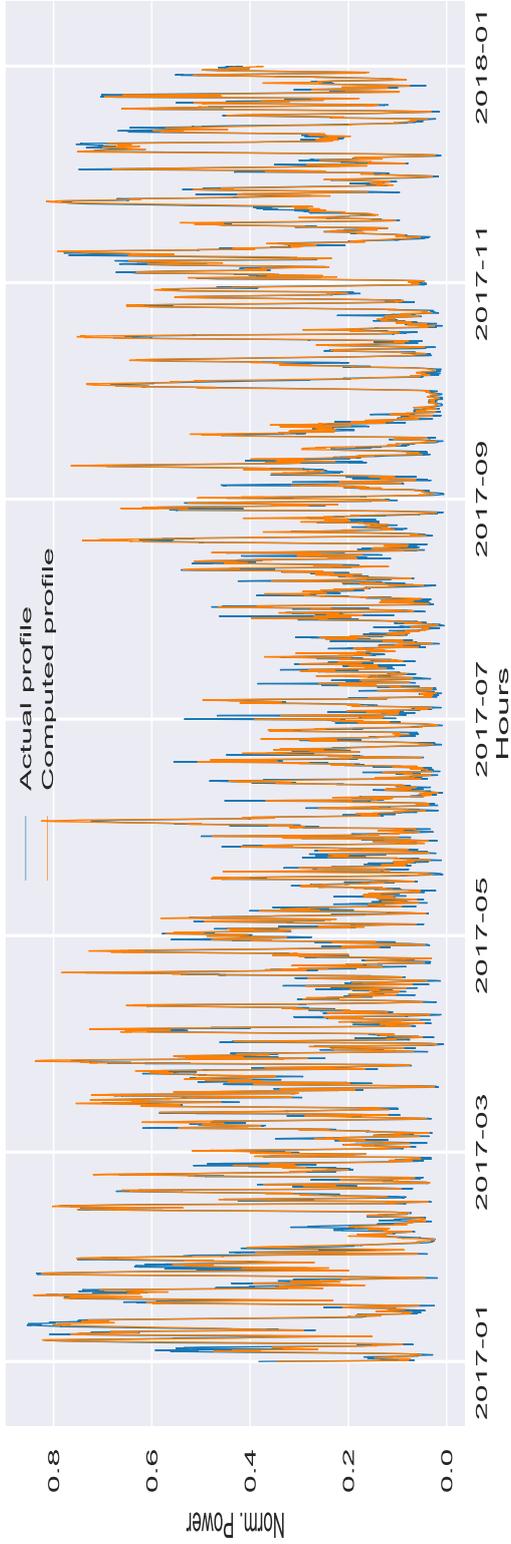
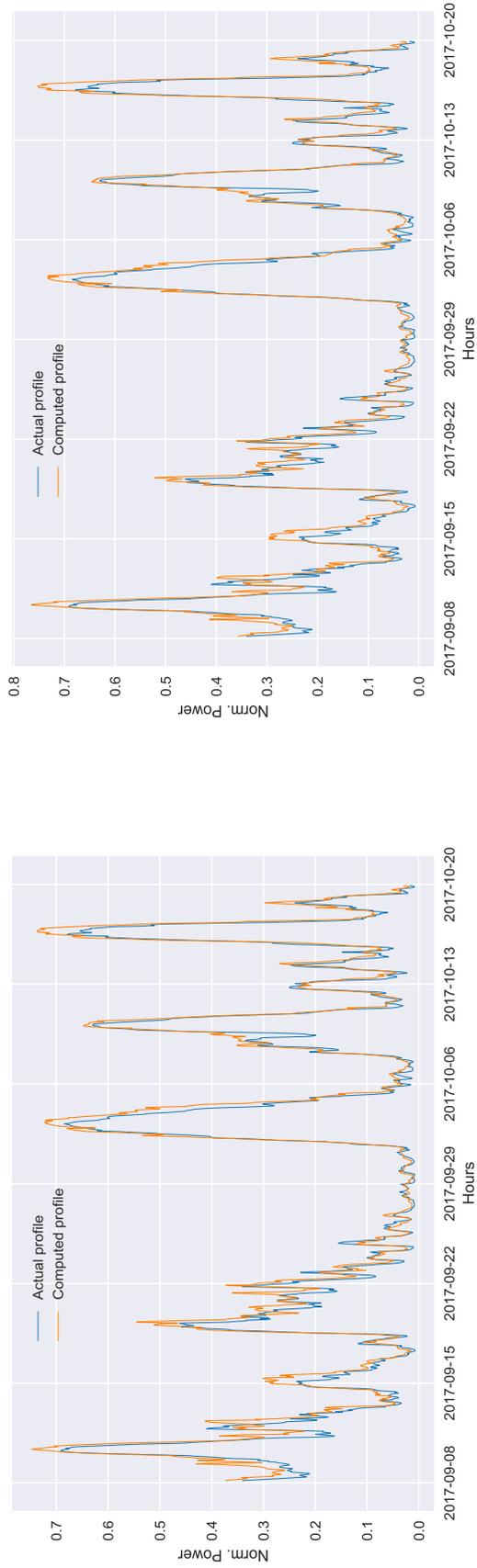


FIGURE 3.15: Comparison of power output distributions for Finland.



(A) Time series: heuristic scheme v. validation data.



(B) Detail of time series: PS scheme v. validation data.

(C) Detail of time series: Heuristic scheme v. validation data.

FIGURE 3.16: Time series study for Finland.

3.3.4 Denmark

DK2 represents an interesting challenge for the model, since this area presents a really small extension, a sparse capacity distribution and it is composed by two separated areas. Despite the severity of the optimization, results in table 3.5 report good test errors, in both the schemes. Furthermore, even if the two schemes return distinguished PCs, they are very similar, as the reader can observe in figure 3.17. However, resulting spatial distributions are quite different, as figure 3.18 and allocation differences report. From allocation errors in table 3.5 it is also possible to conclude that the Heuristic scheme is more capable of allocating wind capacity correctly, since its total allocation error is half of the total allocation error in PSO. In particular, it is possible to see that error committed in estimating capacity installed in cluster 2 is particularly high. However, it is also important to underline that allocation errors are calculated for a very small area compared to previous cases. In fact, the whole area is mapped with less than sixty nodes. As a consequence, each cluster contains a smaller amount of grid points and the allocation error becomes more sensible to small variations of the distribution.

Like in all the previous cases, figure 3.19 shows the typical phenomenon of underestimation of the zero production hours in both schemes. As regards peak load hours, the Heuristic scheme seems to be more accurate than PSO. Analogously to previous areas, complete time series of the Heuristic scheme reported in figure 3.20a confirm this general tendency. Figures 3.20b-3.20c focus on a smaller production period, in order to observe more details. It is possible to notice that the two schemes report almost undistinguishable time series. Both of them are able to explain ramping events. In this particular segment, both of the results recursively overestimates spikes in WPP.

In this area, convergence speed of the two schemes is in line with previous cases (except Finland): as usual, the Heuristic scheme is capable to achieve converge relatively faster than PSO.

| Parameters | PS scheme | Heuristic scheme |
|-----------------------|-----------------------------|----------------------------|
| c | 10.067(25) | 10.526(56) |
| p | 1.825(8) | 1.995(25) |
| Total capacity | 0.841(1) | 0.875(3) |
| Test errors | PS scheme | Heuristic scheme |
| R^2 score [%] | 94.50(30) | 94.38(24) |
| MAE | $4.245(101) \times 10^{-2}$ | $4.289(72) \times 10^{-2}$ |
| MAD | $2.937(108) \times 10^{-2}$ | $2.965(51) \times 10^{-2}$ |
| Iterations | 18(1) | 16 |
| Allocation error | PS v. Reality | Heuristics v. Reality |
| Cluster 1 | $7.23(32) \times 10^{-2}$ | $5.22(47) \times 10^{-2}$ |
| Cluster 2 | $1.24(3) \times 10^{-1}$ | $4.93(59) \times 10^{-2}$ |
| Cluster 3 | $6.88(28) \times 10^{-2}$ | $7.38(589) \times 10^{-3}$ |
| Cluster 4 | $3.11(8) \times 10^{-2}$ | $3.07(17) \times 10^{-2}$ |
| Total A. error | $2.97(6) \times 10^{-1}$ | $1.40(14) \times 10^{-1}$ |
| Allocation difference | PS v. Heuristics | |
| Cluster 1 | $2.01(30) \times 10^{-2}$ | |
| Cluster 2 | $7.50(76) \times 10^{-2}$ | |
| Cluster 3 | $6.16(54) \times 10^{-2}$ | |
| Cluster 4 | $4.25(849) \times 10^{-4}$ | |
| Total difference | $1.57(13) \times 10^{-1}$ | |

TABLE 3.5: Comparison of results for PS and Heuristic method, DK2.

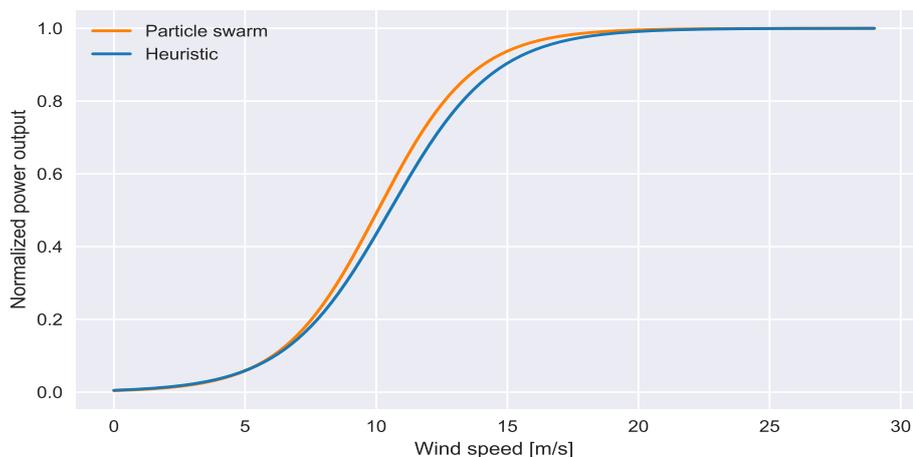
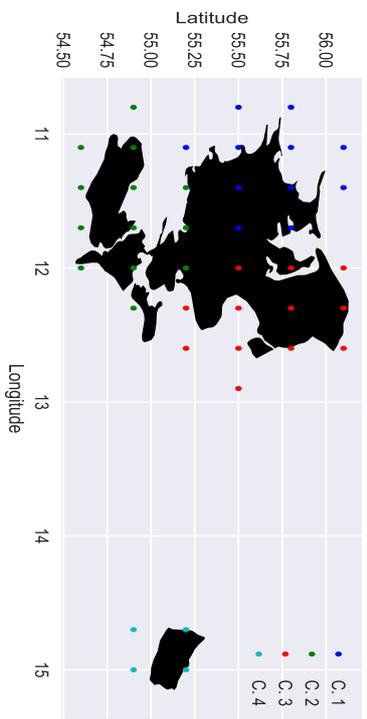
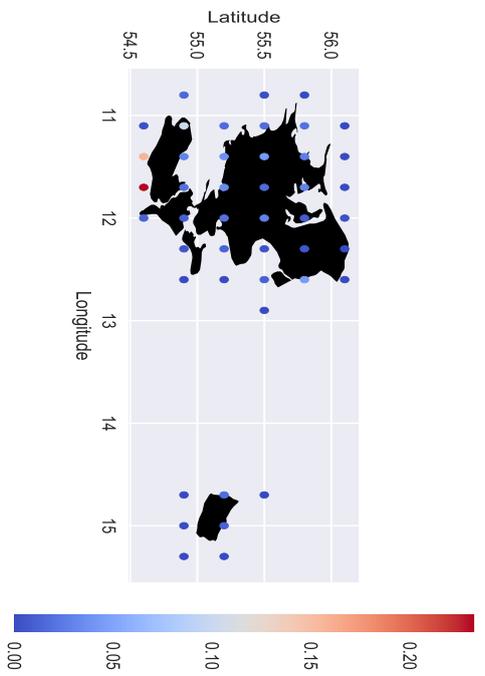


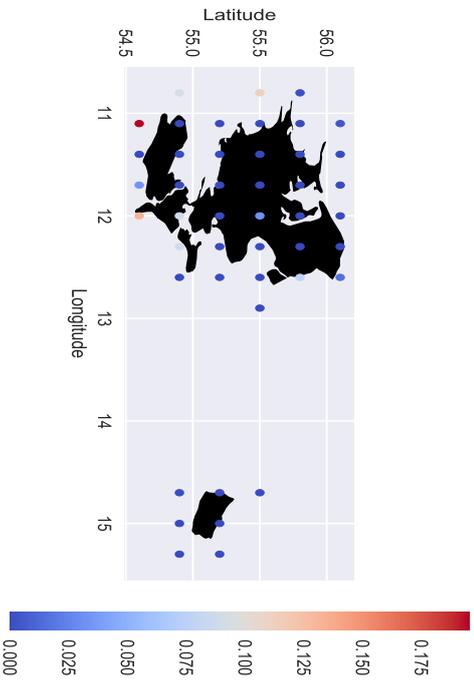
FIGURE 3.17: Comparison of the average simulated power curves for DK2. In this case, power curves returned by the two methods are similar.



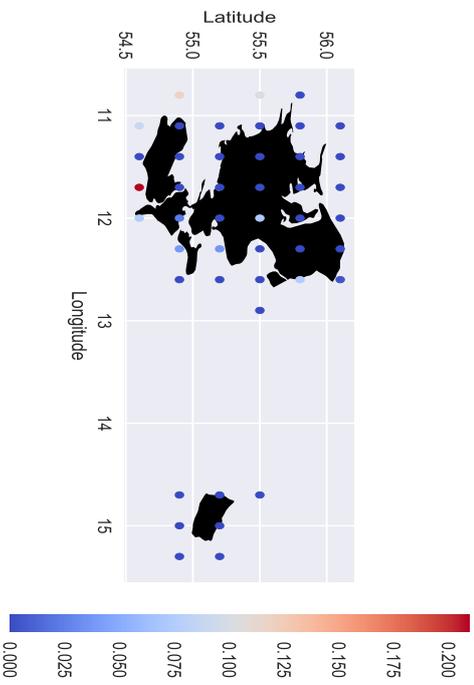
(A) Wind capacity distribution: clustering of the area



(B) Wind capacity distribution: actual distribution.

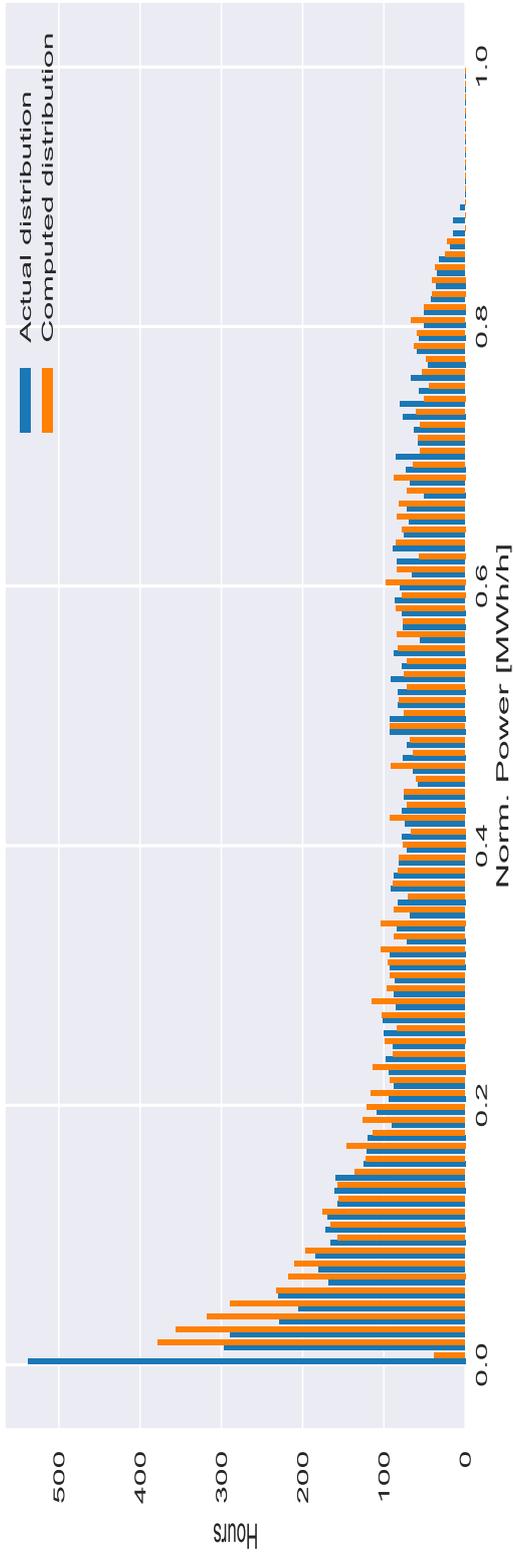


(C) Wind capacity distribution: heuristic scheme.

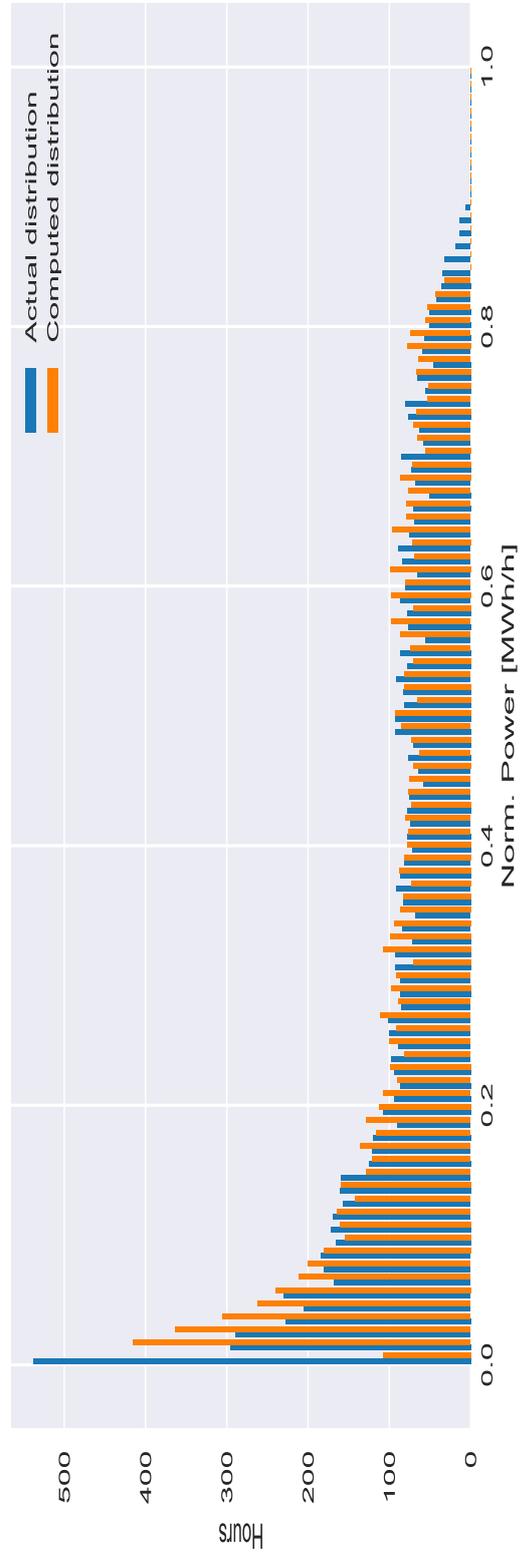


(D) Wind capacity distribution: PS scheme.

FIGURE 3.18: Comparison of wind capacity distributions for DK2.

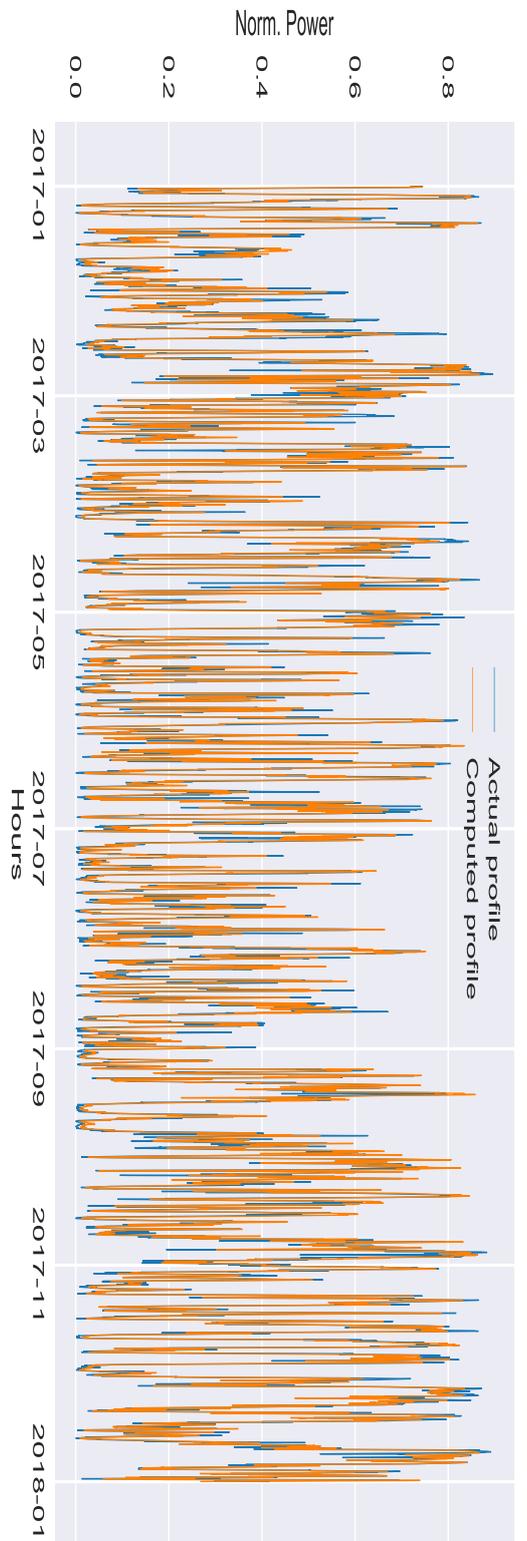


(A) Power output distribution: heuristic scheme v. validation data.

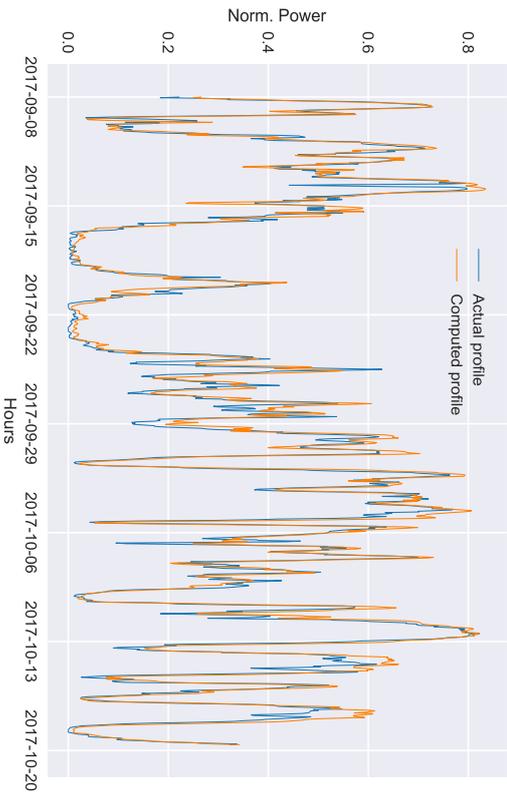


(B) Power output distribution: PS scheme v. validation data.

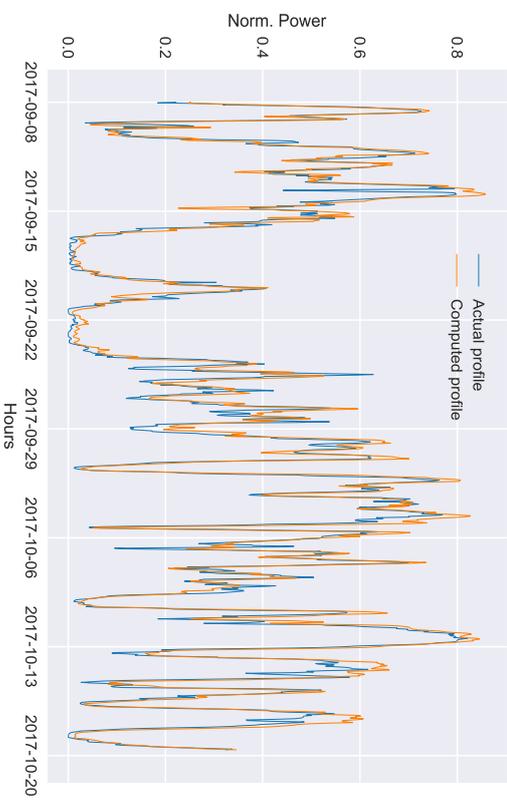
FIGURE 3.19: Comparison of power output distributions for DK2.



(A) Time series: heuristic scheme v. validation data.



(B) Detail of time series: PS scheme v. validation data.



(C) Detail of time series: Heuristic scheme v. validation data.

FIGURE 3.20: Time series study for DK2.

Chapter 4

Conclusions

In order to estimate wind power production of wide areas, a statistical model has been developed. The model associates a power curve to the area and individuates its optimal shape, by minimizing the error between the simulated wind power production series and the realized ones. If wind capacity distribution is unknown, it is estimated by splitting the original optimization problem into two sequential sub-problems (Heuristic scheme) or by directly linking a generic power curve to its optimized capacity distribution (PSO).

As a first step, two possible power curves have been proposed and tested, in the case of known capacity geodata. In this step it has been possible to conclude that a Sigmoid describes wind power production of an area as accurately as a more complex, piecewise curve. However, its correspondent optimization procedure resulted to be faster. In fact, shaping the Sigmoid requires the solution of a bounded optimization problem, while shaping the piecewise function requires solving a constrained optimization problem, which is clearly more complex and computationally expensive. As a conclusion, a Sigmoid represents a better choice for the purposes of the model.

In the following step, the complete model has been tested against four zones (SE1, NO3, DK2 and Finland), by using two different schemes. The first scheme is PSO, while the second one is the Heuristic scheme. Both approaches showed good approximation of reality, with similar spatial patterns in most of the studied cases. In particular, PSO presented slightly higher scores in validation tests, while the Heuristic scheme seemed to be more accurate in the capacity allocation. However, the Heuristic method is significantly faster than PSO in achieving convergence. For this reason, the Heuristic method is a better choice, if both of the quality of the results and time complexity are taken into account in the choice of an algorithm.

Despite promising results from associating a power curve to an extensive area, the complete validation of the model reported some limitations in the quality of the results. In fact, the tendency to underestimate hours without power production was present in all the test results. This is due to the shape of the Sigmoid, which converges to zero asymptotically. Thus, as long as a sufficient number of nodes presents wind speeds different from zero, wind production would always be larger than ten percent of the total installed capacity. A similar problem found in the validation of the problem is the tendency to underestimate power outputs in peak load hours. A plausible explanation of this systematic error is the association of a single power curve to an area, instead that one power curve per wind park.

A possible improvement of the model would consist in building a model based on more available data. For example, if production data for each wind park in a prescribed area is available, it would be possible to associate a single power curve to each wind park and optimize its shape according to production history. An even more detailed model would be developable if the power curve and the height of

each turbine in a prescribed area are known. In this case, there would not be any unknown quantities and the whole optimization problem would collapse in a direct calculation of the estimated production based on given information.

A final critical aspect must be taken into account in the discussion of the results and in possible applications of the model. The model has been trained and validated on realized data. However, once trained, the model would be also capable of taking wind forecasts as inputs. This means that, based on the current level of accuracy of weather forecasts, it would be possible to forecast wind power production of an area.

As a conclusion, based on validation data, it is possible to state that the model developed in this work constitutes a good approximation of reality.

Appendix A

Basin Hopping algorithm

BH algorithm (or Monte Carlo optimization) is a stochastic global optimization coupled with local optimization methods (such as gradient based methods). Its working principle can be shortly resumed in the following procedure:

- At each iteration, coordinates of the previous optimum are casually perturbed through a customized stepping procedure and the local optimization is run again, until a new optimum is found. The chosen stepping procedure for this application is reported below. At each BH iteration:

$$x_{0,k+1}^i = \text{rand}((x_{opt,k}^i - x_{lb}^i) \cdot s, (x_{ub}^i - x_{opt,k}^i) \cdot s) \quad (\text{A.1})$$

Where $x_{0,k+1}^i$ is the starting point of the i-th variable at the k+1-th BH iteration, $\text{rand}(x,y)$ is the function that extract one random value from the homogeneous probability distribution in the interval $[x,y]$. $x_{opt,k}^i$ is the local optimum found at the k-th BH iteration, x_{lb}^i and x_{ub}^i are bounds of the optimized i-th variable and s is the step taken by the algorithm. The step s is periodically updated by the algorithm, in order to maximize the exploration of the domain in which variables are bounded/constrained.

- If the new solution is better than the previous one, it becomes the new global optimal candidate. However, if the local optimum is worse than the previous one, an acceptance test is performed. The probability of the new solution to pass the test is the following:

$$p = e^{-\frac{f(x_{new}) - f(x_{old})}{T}} \quad (\text{A.2})$$

Where T is a parameter calibrated on what is the typical difference in value between two local minima.

The first iteration's minimum will depend on the initial guess if BH is coupled with a gradient based optimization, but a sufficiently high number of perturbations should make the algorithm converge to the global optimum.

A.1 Stopping criteria and time complexity of the algorithm

Two stopping criteria regulate the convergence of the algorithm:

- Iterations stop if the best candidate is not replaced in a certain number of BH iterations;

- The algorithm stops if a maximum number of BH iterations is reached.

It is important to notice that, due to the purely stochastic nature of the algorithm, a high number of perturbations is necessary. Furthermore, time complexity of a BH step exactly corresponds to the local optimization's one. In fact, the function of global optimization is essentially perturbing coordinates of known local optima, which can be considered instantaneous with respect to the actual local optimization. Thus, this scheme can become extremely expensive in case a local optimum requires a significant amount of time to be found.

Appendix B

L-BFGS-B and SLSQP algorithms

L-BFGS-B and SLSQP are two quasi-Newton schemes adopted for nonlinear optimization problems. While the former solves unconstrained, bounded problems, the second one solves constrained, bounded problems. Formally, SLSQP solves problems of this form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \{f(x)\} \\ \text{s.t.} \\ Ax = b \\ c(x) = u \\ l < x < u \end{aligned}$$

It is important to specify that the presence of equalities in the linear and nonlinear constraints does not make the problem lose generality, since it is always possible to convert an inequality in an equality through the introduction of slack variables. L-BFGS-B, instead, solves simpler problems in this form:

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \{f(x)\} \\ \text{s.t.} \\ l < x < u \end{aligned}$$

B.1 Quasi-Newton methods

Quasi-Newton methods are iterative computational schemes in which Jacobian and/or Hessian of an objective function is not available or too expensive to be calculated at each iteration. In Newton's method, the scalar objective function $f(x)$ is parabolically approximated through Taylor's series expansion:

$$f(x_k + \Delta x) \approx f(x_k) + \nabla f(x_k)^T \Delta x + \frac{1}{2} \Delta x^T H \Delta x \quad (\text{B.1})$$

Where H is the Hessian matrix of function f . In order to search for the direction of maximum variation of the function, the gradient of equation B.1 is applied with respect to Δx :

$$\nabla f(x_k + \Delta x) \approx \nabla f(x_k) + H \Delta x \quad (\text{B.2})$$

This equation is also called *secant equation*. Setting $\nabla f(x_k + \Delta x)$ to zero (which means searching for stationary points of the function), it is possible to finally calculate the optimal step:

$$\Delta x = -H^{-1}\nabla f(x_k) \quad (\text{B.3})$$

If H is positive definite, it will always be invertible and step Δx will lead to a local minimum of $\nabla f(x_k + \Delta x)$. Furthermore, if the problem is convex, the iterative procedure will converge to a global optimum.

At this point, the particular feature of Quasi-Newton methods is introduced: $\nabla f(x_k)$ and H may be unavailable. In particular, H may be computationally expensive to be computed. These two quantities are estimated as follows:

- Since f is a scalar function, the gradient $\nabla f(x_k)$ is numerically approximated. A possible numerical approximation is the FE method:

$$\nabla f(x)_i \approx \frac{f(x_i + \Delta x) - f(x_i)}{\Delta x} \quad (\text{B.4})$$

Where $\nabla f(x)_i$ is the i-th component of the gradient $\nabla f(x)$ and Δx is an increment, small enough. It is possible to demonstrate that the precision of this scheme is of the first order, which means that its error is proportional to Δx [26]. A faster convergence of the error can be achieved with more sophisticated methods, like Central Differencing [27] or Crank-Nicolson schemes [28].

- The Hessian matrix H is substituted with an approximated version of it, B. The way B is estimated depends on the scheme adopted and the kind of problem in question. However, in general, two conditions must be satisfied:
 - B must satisfy the secant condition, which corresponds to equation B.2.
 - B must be definitive positive and symmetric, so that it can be inverted and used in the equation B.3;

B.2 L-BFGS-B approximation

In order to maintain symmetry and positive definitiveness, matrix B^{-1} at the k-th iteration is updated in the following iterations through the sum of two rank-one matrices U_k and V_k [29]:

$$B_{k+1}^{-1} = B_k^{-1} + U_k + V_k \quad (\text{B.5})$$

At the first iteration, B_0^{-1} can be set equal to the identity matrix. A rank-one matrix A_1 can also be identified with the following notation:

$$A_1 = wv^T \quad (\text{B.6})$$

In words, a rank-one matrix is the collection of a vector w and its multiples, contained in vector v. U_k and V_k are defined to have the following form:

$$U_k = \alpha uu^T \quad (\text{B.7})$$

$$V_k = \beta vv^T \quad (\text{B.8})$$

Notation can become more readable by setting $\nabla f(x_{k+1}) - \nabla f(x_k) = y_k$ and $x_{k+1} - x_k = \Delta x = s_k$. By imposing the secant condition of equation B.2, it is possible to write:

$$B_k^{-1}y_k + \alpha(u^T y_k)u + \beta(v^T y_k)v = s_k \quad (\text{B.9})$$

By arbitrarily imposing $u = s_k$ and $v = B_k^{-1}y_k$:

$$\alpha(s_k^T y_k)s_k + \beta(y_k^T B_k^{-1}y_k)B_k^{-1}y_k = s_k - B_k^{-1}y_k \quad (\text{B.10})$$

It is sufficient to observe equation B.10 to notice that it is solved for:

$$\alpha = \frac{1}{s_k^T y_k} \quad (\text{B.11})$$

$$\beta = -\frac{1}{y_k^T B_k^{-1}y_k} \quad (\text{B.12})$$

Substituting the result in equation B.9, the Davidon Fletcher and Powell (DFP) rank 2 update formula is found:

$$B_{k+1}^{-1} = B_k^{-1} + \frac{s_k s_k^T}{s_k^T y_k} - \frac{B_k^{-1}y_k y_k^T B_k^{-1}}{y_k^T B_k^{-1}y_k} \quad (\text{B.13})$$

If the same update logic of equation B.9 is applied to matrix B, it is possible to obtain the following equation:

$$B_{k+1} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k} \quad (\text{B.14})$$

Which introduces the concept of *Duality*, independently discovered by the four authors Broyden, Fletcher, Goldfarb and Shanno. In order to ease computation of matrix B_{k+1}^{-1} , the Sherman-Morrison-Woodbury formula is adopted:

$$(A + UV^T)^{-1} = A^{-1} - A^{-1}UC^{-1}V^T A^{-1} \quad (\text{B.15})$$

$$C = I + V^T A^{-1}U \quad (\text{B.16})$$

Thus, the new expression of B^{-1} becomes:

$$B_{k+1}^{-1} = \left(I - \frac{s_k y_k^T}{s_k^T y_k}\right) B_k^{-1} \left(I - \frac{y_k s_k^T}{s_k^T y_k}\right) + \frac{s_k s_k^T}{s_k^T y_k} \quad (\text{B.17})$$

Since the initial value B_0^{-1} can be set to the identity matrix I, it is possible to use the secant equation B.2 to obtain B_{k+1}^{-1} without never explicitly calculating the intermediate inverse Hessian matrices $B_1^{-1}, \dots, B_k^{-1}$. However, the recursive nature of equation B.17 forces the storage of the history of coordinates x_k and $\nabla f(x)_k$. In order to limit the amount of memory dedicated to this data, the last m iteration are the only ones kept in account while calculating the Hessian. Bounds in variables are taken into account by increasing the complexity of the algorithm, but the core idea of the Hessian approximation remains essentially the same.

B.3 SLSQP

In the case of a constrained problem, at each iteration k , the original problem is splitted into two subproblems:

- The first subproblem is a QP problem, in which the best optimization direction d_k is found. $f(x)$ is approximated through a quadratic approximation and constraints are linearized. If we assume for simplicity that all the variables are non-negative:

$$\begin{aligned} \min_{d_k \in \mathbb{R}^n} \{ & f(x_k) + g(x_k)^T d_k + \frac{1}{2} d_k^T H(x_k, \pi_k) d_k \} \\ & \text{s.t.} \\ & c(x_k) + J(x_k) d = 0 \\ & d_k > 0 \end{aligned}$$

Where $J(x_k)$ is the Jacobian matrix of constraints $c(x)$, $g(x)$ is the gradient of $f(x)$ and $H(x, \pi)$ is the Hessian matrix of the Lagrangian function:

$$H(x, \pi) = \nabla_{xx}^2 L(x, \pi, x) = \nabla^2 f(x) - \sum_i^m \pi_i \nabla^2 c_i(x) \quad (\text{B.18})$$

$$L(x, \pi, x) = f(x) - \pi^T c(x) - z^T x \quad (\text{B.19})$$

Quantities π and z are technically defines as *dual variables* in the context of mathematical optimization. In a more traditional way, they correspond to Lagrange multipliers. The reason why the Hessian of Lagrange function is chosen instead of the Hessian of $f(x)$ can be explained through the following explanation. It is possible to demonstrate that the QP constrained subproblem is mathematically equivalent to the unconstrained optimization of the *modified Lagrangian*:

$$L(d_k, x_k, \pi_k) = f(x_k + d_k) - \pi_k^T (c(x_k + d_k) - \hat{c}_k(x)) \quad (\text{B.20})$$

$$\hat{c}_k(x) = c(x_k) + J(x_k) d_k \quad (\text{B.21})$$

where x_k and π_k are fixed. The quantity $c(x_k + d_k) - \hat{c}_k(x)$ is also called *departure from linearity* and this explains the choice of H : the objective function optimized in the QP subproblem constitutes a local quadratic model of f that incorporates the curvature of the constraints $c(x) = 0$ [16]

- The second subproblem consists in estimating how big the step in the d_k direction should be. If α is defined so that:

$$x - x_k = \alpha_k d_k, \quad \alpha \in [0, 1] \quad (\text{B.22})$$

It is possible to estimate the optimal step magnitude through the minimization of the scalar penalty function Φ :

$$\Phi(\alpha_k, \rho) = f(x_k + \alpha_k d_k) + \rho \sum_i^m |c_i(x_k + \alpha_k d_k)| \quad (\text{B.23})$$

Where ρ corresponds to the penalty of the function. If this value is large enough, it guarantees global convergence. In a more practical way, the minimization of Φ relaxes the resolution of the global optimization problem, since constraints do not have to be perfectly satisfied. ρ works as a parameter, corrected at each iteration if convergence is not reached.

The way these two subproblems are solved is not trivial, especially for the first one. In fact, while the second subproblem essentially corresponds to the optimization of a scalar function of one variable, solvable in multiple fast ways, the first one can be really computational expensive. The generality of the approaches for the QP problem come from the resolution of the KKT system. The KKT system is the following:

$$\begin{bmatrix} H(x_k) & J(x_k) \\ J(x_k) & 0 \end{bmatrix} \begin{bmatrix} d_k \\ -\pi_{k+1} \end{bmatrix} = - \begin{bmatrix} g_k \\ c_k \end{bmatrix} \quad (\text{B.24})$$

and it derives from the first order optimality conditions for a nonlinear optimization problem. If the problem is in the form:

$$\min_{x \in \mathbb{R}^n} \{f(x)\} \text{ s.t. } c(x) = 0, x \geq 0 \quad (\text{B.25})$$

first order optimality conditions are the following¹:

$$c(x) = 0, \quad (\text{B.26})$$

$$g(x) - J(x)^T \pi - z = 0, \quad (\text{B.27})$$

$$x \cdot z = 0 \quad (\text{B.28})$$

Some of the computational schemes addresses system [B.24](#) directly, while some others are based on the so called *nullspace factorization of the KKT system*. With Q_k a nonsingular matrix defined so that $J_k Q_k = (0 \ U_k)$, with U_k sized $m \times m$, the KKT system is algebraically manipulated and turned into an equivalent system:

$$\begin{bmatrix} Q_k^T H(x_k) Q_k & (J(x_k) Q_k)^T \\ J(x_k) Q_k & 0 \end{bmatrix} \begin{bmatrix} d_Q \\ -\pi_{k+1} \end{bmatrix} = - \begin{bmatrix} Q_k^T g_k \\ c_k \end{bmatrix} \quad (\text{B.29})$$

Where $d_k = Q_k d_Q$. Since J_k has shape $n \times m$ and rank m , U_k is nonsingular. By partitioning Q_k into two submatrices Z_k and Y_k so that:

$$J_k Q_k = J_k (Z_k \ Y_k) = (0 \ U_k) \quad (\text{B.30})$$

it is possible to state that vectors of Z_k constitute the basis for the null space of J_k and. If equation [B.30](#) is substituted into [B.29](#), the final system is obtained:

¹First order optimality conditions are necessary conditions for the optimality of a point, but not sufficient.

$$\begin{bmatrix} U_k & 0 & 0 \\ Z_k^T H_k Y_k & Z_k^T H_k Z_k & 0 \\ Y_k^T H_k Y_k & Y_k^T H_k Z_k & U_k^T \end{bmatrix} \begin{bmatrix} d_Y \\ d_Z \\ -\pi_{k+1} \end{bmatrix} = - \begin{bmatrix} c_k \\ Z_k^T g_k \\ Y_k^T g_k \end{bmatrix} \quad (\text{B.31})$$

Where d_Y and d_Z are subvectors of d_Q . The advantage of this approach is in the possibility to determine whether the KKT system is singular or not. In fact, from equation B.31 it is possible to asses that the system is not singular if $Z_k^T H_k Z_k$ is not singular and J_k has maximum rank m . In order to reduce time complexity of the algorithm, H_k is substituted with an approximation, analogue to the BFGS approximation reported in section B.2. In most of the algorithms, the Jacobian is exactly calculated [30].

B.4 Time complexity

Time complexity is a critical factor to take into account. In some cases, computational resources may be limited or obtaining a good enough, fast solution may be more important than an analytical, computational expensive solution.

B.4.1 L-BFGS-B

Time complexity of L-BFGS-B algorithm is $O(nm)$, where n corresponds to the number of variables of the problem and m to the number of previous steps stored into memory. Coherently, the original BFGS has a time complexity of $O(n^2)$.

B.4.2 SLSPQ

Generally, time complexity of SLSPQ is polynomial. In particular, it can vary according to how each subproblem is solved and how the Jacobian of the constraints and the Hessian of the Lagrangian functions in equation B.24 are calculated or approximated. The shape of the Jacobian is also important. For example, in the nullspace factorization approach, matrix Z_k of equation B.31 can be directly calculated through a QR factorization of J_k if the matrix is dense [31]. In case J_k is sparse, the idea is to transfer this property to Z_k . Through some manipulations, it is possible to obtain the desired results at the cost of a LU factorization [32], which has the same time complexity of QR factorization². Some advanced algorithms rely on the approximation of the Jacobian as well, provoking a significant reduction in time complexity. For example, the *ZED is DEAD* approach permits a reduction of the whole SQP system complexity from $O(mn^2)$ to $O(n \cdot \max(m, l))$, where n corresponds to the number of variables, m corresponds to the maximal number of active constraints³ and l the number of stored updates [33].

²Both of them have complexity $O(n^3)$, where n is the size of the matrix to be factorized.

³Given a feasible point of an optimization problem, an inequality constraint of the form $c(x) \leq 0$ is active in that point if $c(x) = 0$.

Bibliography

- [1] D. J. Arent, A. Wise, and R. Gelman, “The Status and Prospects of Renewable Energy for Combating Global Warming”, *Energy Economics*, Special Issue on The Economics of Technologies to Combat Global Warming, vol. 33, no. 4, pp. 584–593, Jul. 1, 2011, ISSN: 0140-9883. DOI: [10.1016/j.eneco.2010.11.003](https://doi.org/10.1016/j.eneco.2010.11.003). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0140988310001908> (visited on 07/30/2018).
- [2] “EUR-Lex - 52015DC0080 - EN - EUR-Lex”, [Online]. Available: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52015DC0080> (visited on 07/30/2018).
- [3] T. Abbasi, M. Premalatha, and S. A. Abbasi, “The Return to Renewables: Will It Help in Global Warming Control?”, *Renewable and Sustainable Energy Reviews*, vol. 15, no. 1, pp. 891–894, Jan. 1, 2011, ISSN: 1364-0321. DOI: [10.1016/j.rser.2010.09.048](https://doi.org/10.1016/j.rser.2010.09.048). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1364032110003357> (visited on 07/30/2018).
- [4] M. Ram, M. Child, A. Aghahosseini, D. Bogdanov, A. Lohrmann, and C. Breyer, “A Comparative Analysis of Electricity Generation Costs from Renewable, Fossil Fuel and Nuclear Sources in G20 Countries for the Period 2015-2030”, *Journal of Cleaner Production*, vol. 199, pp. 687–704, Oct. 20, 2018, ISSN: 0959-6526. DOI: [10.1016/j.jclepro.2018.07.159](https://doi.org/10.1016/j.jclepro.2018.07.159). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0959652618321486> (visited on 10/16/2018).
- [5] “ENTSO-E Transparency Platform”, [Online]. Available: <https://transparency.entsoe.eu/> (visited on 07/24/2018).
- [6] “How to Download ERA5 Data via the ECMWF Web API - Copernicus Knowledge Base - ECMWF Confluence Wiki”, [Online]. Available: <https://confluence.ecmwf.int/display/CKB/How+to+download+ERA5+data+via+the+ECMWF+Web+API> (visited on 07/31/2018).
- [7] “Iris Reference Documentation — Iris 2.0.0 Documentation”, [Online]. Available: <https://scitools.org.uk/iris/docs/v2.0/iris/iris.html> (visited on 07/31/2018).
- [8] “ERA5 Data Documentation - Copernicus Knowledge Base - ECMWF Confluence Wiki”, [Online]. Available: <https://confluence.ecmwf.int//display/CKB/ERA5+data+documentation> (visited on 07/31/2018).
- [9] J. Olauson, “ERA5: The New Champion of Wind Power Modelling?”, *Renewable Energy*, vol. 126, pp. 322–331, Oct. 1, 2018, ISSN: 0960-1481. DOI: [10.1016/j.renene.2018.03.056](https://doi.org/10.1016/j.renene.2018.03.056). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960148118303677> (visited on 07/20/2018).
- [10] *Wind profile power law*, in *Wikipedia*, Page Version ID: 815325859, Dec. 14, 2017. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Wind_profile_power_law&oldid=815325859 (visited on 10/30/2018).

- [11] J. Dai, D. Liu, L. Wen, and X. Long, "Research on power coefficient of wind turbines based on SCADA data", *Renewable Energy*, vol. 86, pp. 206–215, Feb. 1, 2016, ISSN: 0960-1481. DOI: 10.1016/j.renene.2015.08.023. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0960148115302238> (visited on 11/15/2018).
- [12] Y. H. Wan, E. Ela, and K. Orwig, "Development of an Equivalent Wind Plant Power-Curve", p. 23,
- [13] (). Wind Statistics and the Weibull distribution, [Online]. Available: http://www.wind-power-program.com/wind_statistics.htm (visited on 10/30/2018).
- [14] L. Quéval, C. Joulain, and C. Casillas, "Measuring the Power Curve of a Small-Scale Wind Turbine: A Practical Example", MDPI, Mar. 14, 2014, p. c011. DOI: 10.3390/ece-1-c011. [Online]. Available: <http://www.sciforum.net/conference/ece-1/paper/2338> (visited on 07/23/2018).
- [15] *Limited-memory BFGS*, in *Wikipedia*, Page Version ID: 863036695, Oct. 8, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Limited-memory_BFGS&oldid=863036695 (visited on 10/30/2018).
- [16] *Sequential quadratic programming*, in *Wikipedia*, Page Version ID: 855083286, Aug. 15, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Sequential_quadratic_programming&oldid=855083286 (visited on 10/30/2018).
- [17] "Optimization and Root Finding (Scipy.Optimize) — SciPy v1.1.0 Reference Guide", [Online]. Available: <https://docs.scipy.org/doc/scipy/reference/optimize.html> (visited on 07/24/2018).
- [18] *Overfitting*, in *Wikipedia*, Page Version ID: 864108319, Oct. 15, 2018. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Overfitting&oldid=864108319> (visited on 10/30/2018).
- [19] *Levenberg–Marquardt algorithm*, in *Wikipedia*, Page Version ID: 860162799, Sep. 18, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Levenberg%E2%80%93Marquardt_algorithm&oldid=860162799 (visited on 10/30/2018).
- [20] (). Etha Wind, [Online]. Available: <https://ethawind.com/> (visited on 10/29/2018).
- [21] (). Om Vindlov, [Online]. Available: <http://www.vindlov.se/sv/om-vindlov/> (visited on 10/29/2018).
- [22] (). NVE data nedlast, [Online]. Available: <https://nedlasting.nve.no/gis/> (visited on 10/29/2018).
- [23] (Aug. 25, 2016). Overview of the energy sector, [Online]. Available: <https://ens.dk/en/our-services/statistics-data-key-figures-and-energy-maps/overview-energy-sector> (visited on 10/29/2018).
- [24] "Markets Divided into Bidding Areas.", [Online]. Available: <http://www.nordpoolspot.com/the-power-market/Bidding-areas/> (visited on 10/15/2018).
- [25] "2.3. Clustering — Scikit-Learn 0.20.0 Documentation", [Online]. Available: <http://scikit-learn.org/stable/modules/clustering.html#clustering> (visited on 10/16/2018).
- [26] "Forward and Backward Euler Methods", [Online]. Available: http://web.mit.edu/10.001/Web/Course_Notes/Differential_Equations_Notes/node3.html (visited on 08/06/2018).

- [27] *Central differencing scheme*, in *Wikipedia*, Page Version ID: 822475259, Jan. 26, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Central_differencing_scheme&oldid=822475259 (visited on 10/30/2018).
- [28] *Crank–Nicolson method*, in *Wikipedia*, Page Version ID: 829403297, Mar. 8, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Crank%E2%80%93Nicolson_method&oldid=829403297 (visited on 10/30/2018).
- [29] E. Bertolazzi, “Quasi-Newton Methods for Minimization - Lectures for PHD Course on Numerical Optimization”, p. 63,
- [30] M. Diehl, A. Walther, H. G. Bock, and E. Kostina, “An Adjoint-Based SQP Algorithm with Quasi-Newton Jacobian Updates for Inequality Constrained Optimization”, *Optimization Methods and Software*, vol. 25, no. 4, pp. 531–552, Aug. 1, 2010, ISSN: 1055-6788. DOI: 10.1080/10556780903027500. [Online]. Available: <https://doi.org/10.1080/10556780903027500> (visited on 10/04/2018).
- [31] *QR decomposition*, in *Wikipedia*, Page Version ID: 866419792, Oct. 30, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=QR_decomposition&oldid=866419792 (visited on 10/30/2018).
- [32] *LU decomposition*, in *Wikipedia*, Page Version ID: 865402665, Oct. 23, 2018. [Online]. Available: https://en.wikipedia.org/w/index.php?title=LU_decomposition&oldid=865402665 (visited on 10/30/2018).
- [33] T. Bosse, V. Schloßhauer, A. Griewank, T. Bosse, V. Schloßhauer, and A. Griewank, *On Hessian- and Jacobian-Free SQP Methods- A*.