POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

The Duality of Wind: A Comprehensive Study on Lombardy's Renewable Wind Energy Potential and Grid Infrastructure Hazards

Tesi di Laurea Magistrale in
Mathematical Engineering - Ingegneria Matematica

Authors: **Mattia Gentile - Alessandro Sala**

Student ID: 967634 - 968727
Advisor: Prof. Piercesare Secchi
Co-advisor: Ing. Chiara Barbi
Academic Year: 2021-22

Abstract

Wind is a complex natural phenomenon occurring incessantly everywhere on the globe; sometimes it can be harmful, while other times it can be used as a precious resource. In particular, thinking about energy and infrastructures, wind can be disruptive for power lines but, on the other hand, it can be harnessed to produce energy. With this applications in mind, this thesis aims at investigating this duality from a mathematical and statistical point of view, focusing on the area corresponding to the Lombardy region in Italy for an example of practical results. First of all, the general characterization of wind speed distribution is discussed. Then, hazard analysis is carried on with the aim of establishing areas subject to more extreme and dangerous winds. Two clustering strategies from geostatistical literature are evaluated to group together regions with similar winds. And finally, a case study regarding wind energy production is investigated. At all stages, different methods and strategies are compared to determine the most appropriate one in light of the specific application, and interesting insights on the various mathematical methods are highlighted.

Keywords: extreme value analysis, hazard analysis, functional data analysis, geostatistical clustering, wind energy

Abstract in lingua italiana

Il vento è un fenomeno naturale che agisce senza sosta in ogni punto del pianeta; se, a volte, può provocare danni e disagi, altre può essere sfruttato come una preziosa risorsa. In particolare, se si pensa all'aspetto energetico e infrastrutturale, il vento può rappresentare una forza distruttiva per le linee elettriche ma può anche essere sfruttato per produrre energia. Alla luce dei suddetti fattori, questa tesi ha lo scopo di investigare questa dualità da un punto di vista matematico e statistico, concentrando l'attenzione sull'area lombarda per cui verranno forniti dei risultati pratici. Per prima cosa, viene discussa la caratterizzazione generale sulle distribuzioni delle velocità del vento. Segue poi un'analisi sulla pericolosità legata agli eventi ventosi, con l'obiettivo di individuare le aree soggette ai venti più estremi e pericolosi. Due strategie di clustering tratte dalla letteratura geostatistica sono valutate per raggruppare insieme aree con venti simili. In conclusione, vengono investigate le potenzialità di produrre energia eolica. Durante tutti i passaggi dello studio, metodi e strategie diverse sono confrontati per determinare quelli più indicati per l'applicazione specifica che si sta analizzando, e, inoltre, vengono sottolineate le osservazioni più interessanti sui metodi matematici adottati.

Parole chiave: teoria dei valori estremi, analisi della pericolosità, analisi dati funzionale, clustering geostatistico, energia del vento

Contents

Introduction

As the world faces the urgent need of reducing pollution and mitigating the effects of climate change, renewable energy sources are becoming a more and more critical component of our energy systems. Among these sources, wind energy has emerged as a promising solution to help reduce our reliance on fossil fuels and curb greenhouse gas emissions. Wind is indeed an inexhaustible force, present in every place of our planet, that has been largely used by humankind, for example to navigate or for wind mills, throughout its entire history. However, while wind power presents many benefits, it also brings to the table many threats for both human and infrastructure safety.

Figure 1: While the power of wind can be a precious resource to produce energy, the climate change brought to a situation with an always increasing number of extreme events causing infrastructure disruptions.

This thesis aims to investigate this double nature of wind power in the region of Lombardy, Italy. On one hand, wind is a valuable resource that can help generate clean and sustainable energy and, considering that Lombardy alone produces over 22% of the national PIL and about 1/6 of the italian population lives here, reaching a certain degree of autonomous power generation is for sure in the interest of the region. On the other hand, wind can also pose a significant threat, among all, to the electrical grid infrastructure, particularly during extreme weather events. Understanding and addressing this duality is essential for ensuring the reliability, safety and sustainability of our energy system.

To reach this goal, our study analyses wind data measured over 30 years, with the aim of achieving a high level of awareness about both the dangers for the electrical grid and the opportunities to produce clean and renewable energy in Lombardy. The data we are using are open and provided by the RSE S.p.A group in the MERIDA dataset. By examining the potential risks and opportunities associated with wind power, the thesis provides valuable insights for policymakers, energy companies and stakeholders in the region. Ultimately, the findings of this study can help pave the way for more effective strategies to harness the potential of wind energy while mitigating its risks.

Of particular importance is the choice of the most appropriate mathematical procedure to produce proper evaluations. Indeed, in practical applications outdated models or excessive simplifications are often employed and this may cause incorrect results and wrong estimations. For instance, in the literature, the Weibull distribution is often taken to approximate wind speed data (see, for example, Perrin et al. 2006 [34] and Celik 2003 [11]); however, as we will see, not always this model is the best way to proceed and errors on this estimation will cascade down to subsequent steps of any analysis, possibly invalidating the results. One of the aims of this work is a comparison, in each phase, of different procedures and the evaluation of their criticalities and benefits.

The thesis is structured as follows. In Chapter 1 a description of the data used and how they have been collected is given. Chapter 2 illustrates the best fitting distributions for the data of each site of Lombardy, by discussing all the mathematical methods adopted. In Chapter 3 classical extreme value analysis is described as the method to tackle extreme wind events, alongside few alternative applicable techniques. In Chapter 4 hazard assessment is conducted by means of extreme value analysis, with the goal of producing both local results and groupings based on the the risk for electrical grid infrastructure. In Chapter 5 functional data analysis tools used are applied to wind time series in Lombardy, culminating in the definition of a distance that measures the dissimilarities between the wind regimes in two different sites. In Chapter 6 two clustering algorithms to group areas with similar wind behaviours and that exploit the distance defined in the previous chapter are presented and discussed. Chapter 7 tackles the criticalities and opportunities of wind energy production in Lombardy, presenting and analysing in detail the scenario of small wind turbines application in the region. Finally, in Chapter 8 conclusions and final remarks are illustrated. Every computation was performed on `R` (version 4.0.4).

1 | Dataset Overview

The MEteorological Reanalysis Italian DAtaset (MERIDA) has been developed by the "Ricerca Sistema Energetico" (RSE) S.p.A. group to deal with the study of the always more extreme weather conditions which have caused several disruptions to the italian electricity system throughout the years. This dataset has been developed following the indications emerged from the "Working table for resilience" established by the Regulatory Authority for Energy Networks and the Environment (ARERA).

MERIDA has been further improved, focusing on less variables but on a finer grid in order to obtain a High Resolution version of MERIDA (MERIDA HRES). If the main goal of MERIDA is to provide energy stakeholders with meteorological data needed to plan security routines and to design a reliable electrical grid, MERIDA HRES has been developed as a tool to study those variables linked mainly to renewable energy.

Both the datasets are the result of a dinamical downscaling starting from global data and centered on the area of Italy.

1.1. MERIDA

Dataset MERIDA is the first product of RSE group in the field of meteorological analysis. As carefully described in the outline attached to the dataset [8], MERIDA contains many meteorological indices of interest such as the temperature measured at 2 meters from the ground, precipitation data, wind velocity, humidity and so on. MERIDA has been obtained as a downscaling of the ERA5 dataset, the fifth generation of global climate reanalysis produced by the European Centre for Medium-Range Weather Forecast (ECMWF) in 2017. ERA5 consists in a dataset of meteorological related variables that covers all the surface of Earth with a grid of spatial resolution of 31 km (at midlatitudes) and temporal resolution of 1 hour (see Hòlm et al. (2016) [21]). In order to perform a dynamical downscaling from ERA5 data and obtain a finer resolution product, it has been employed the so called Weather Research and Forecasting Model (WRF), and in particular one of its dynamical versions, the Advanced Research WRF (ARW). This algorithm is a numerical weather prediction system designed to serve both atmospheric research and operational forecasting needs (see Skamarock et al. (2008) [43]).

Therefore, applying the WRF-ARW model to ERA5 as described in Bonanno et al. (2019) [7], RSE group managed to obtain the MERIDA dataset, which focuses on the area of Italy. This dataset covers the area between longitudes 5.84 and 18.93, and latitudes 35.37 and 48.25, with a temporal resolution of 1 hour and two types of spatial resolution: the most coarse consisting in a grid of cells of 21 km and the finer one with 7 km resolution. Data are collected over the time period spanning from 1990 to 2020 but are constantly updated and expanded.

1.2. MERIDA HRES

MERIDA HRES is a more refined product produced by RSE group, directly obtained from the previous version of MERIDA. The new dataset counts a smaller amount of variables, focusing on indices strictly linked to renewable energy production, but the grid on which those variables are measured is finer. As highlighted in the dataset description attached [9], these variables are the temperature at 2 meters from the ground, the U and V components of the wind (i.e. the longitudinal, from West to East, and the latitudinal, from South to North, components of the wind vector), measured at both 10 and 100 meters of altitude, the precipitations and the solar radiance. While the temporal resolution remains the same, the spatial one increases, getting a grid with cells of 4 km each; also the area considered and time window of measurements spanned remains the same with respect to MERIDA.

From MERIDA HRES we extracted and used just the data concerning wind; however, the computational load was still huge since our raw data consist in a grid of 323 × 329 cells, with each site containing two time series (one for longitudinal direction, or component U, and one for latitudinal direction, or component V) of 271752 observations, for a total of over 200 GB of data. Due to the computational load of the entire dataset, in the first place, we worked on a single cell at a time, especially when testing new methods, and then enlarged the area considered to an entire region and, in particular, we focused our attention on the territory of Lombardy.

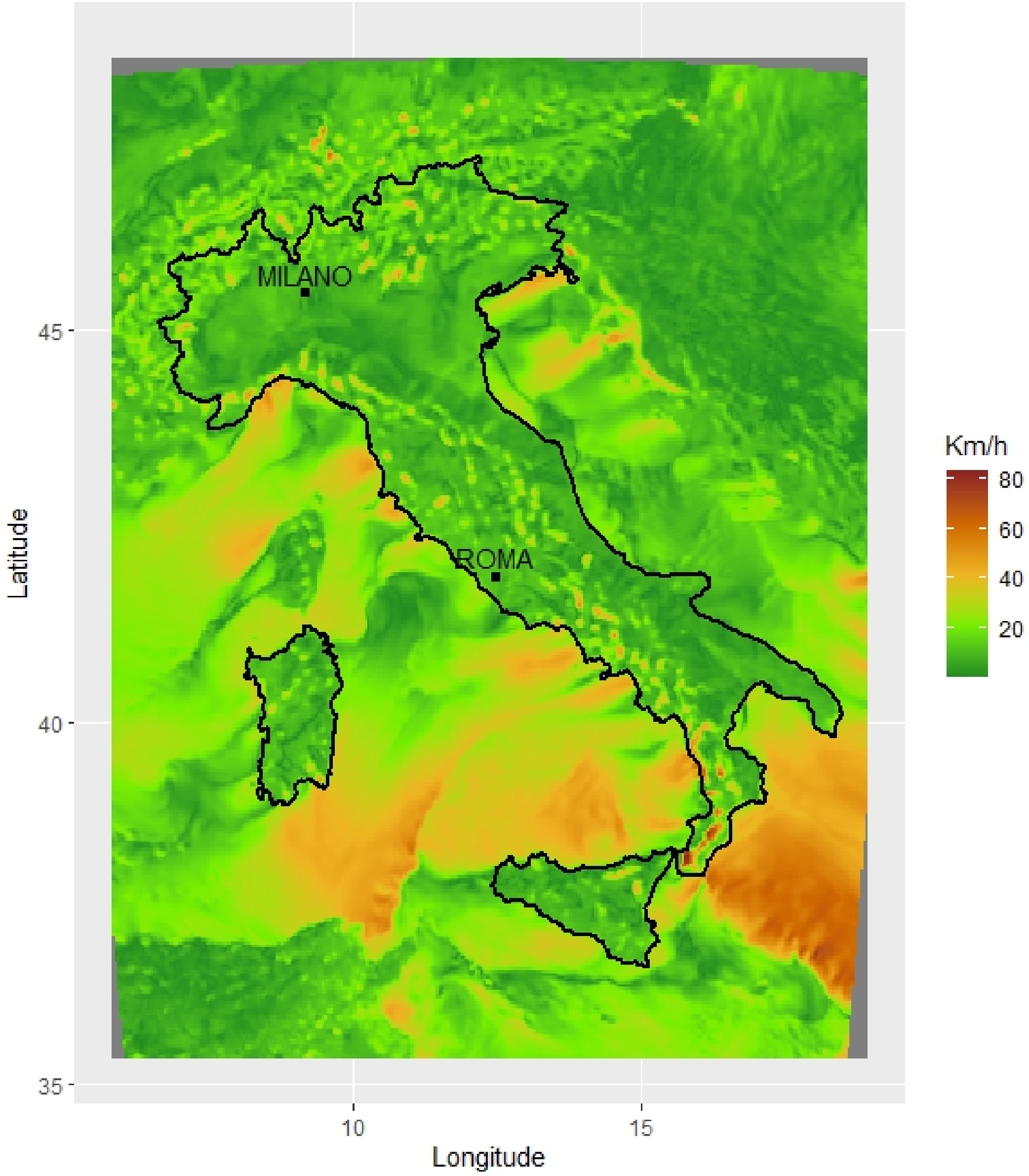


Figure 1.1: An example reporting the wind speeds on the whole area covered by the dataset on the 01/01/1990 at midnight, the first measurement available.

2 | Wind Speed Modeling

A first, fundamental, step in every study concerning wind is the assessment of its characteristics and features. Even though the first part of our work will focus mainly on the extreme values of the phenomenon, a proper introduction on the topic of modelling wind speed as a whole is needed. Two are the main reasons; first, the importance and generality of the discussion is so great that, by itself, it has generated uncountable works in the literature and there is no proper study on the subject of wind, be it a potential energy resource or a risk for infrastructure, that does not start with it. In the same way, the second reason is that even our work cannot be completely blind to a characterization of the total wind series; we will use its features to better understand the underlying phenomenon, it will be used to estimate production of wind energy and, as we will see, it will be the basis of some of the methods analyzed for the extreme winds.

The statistical analysis of the wind speed series follows a pipeline presented for instance in Shi et al. (2021) [41]. The main goal is to determine which probability density function (pdf) better approximates the real distribution of the measured data. Several possible distributions can be considered and, for each of them, the value of its parameters must be determined using an appropriate parameter estimation method. Finally, a goodness of fit criteria must be employed to evaluate the models and choose the most appropriate one.

Previous researches on wind speed distribution have pursued the unification of all existing wind regimes under a single model but the differences in behaviour at the different locations seem to show that such a thing cannot exist and, at present time, it has not yet been proposed a model that can provide a sufficient description at any site. Moreover, different parameter estimation methods and goodness of fit criteria have advantages and disadvantages, bringing difficulties to the assessment of wind speeds.

In this chapter we will investigate the most popular strategies and compare their results on the area of interest.

2.1. Wind Speed Distribution Models

First of all we need to point out the fact that, at this stage and until differently noticed, we refer to "wind speed" meaning only the magnitude of the wind vector, disregarding the direction and considering only the absolute intensity.

Numerous models have been proposed throughout the years but we have chosen to analyze only 4 of the most popular and promising ones: Weibull, Gamma, Lognormal and Generalized Extreme Value distribution (GEV) (Figure 2.1). Notice that the support of the first 3 distributions is (0,$+\infty$) which is a desirable property for the representation of a physical phenomenon like the wind. The GEV instead may also assume negative values but its flexibility has been proven fundamental to achieve surprisingly good results.

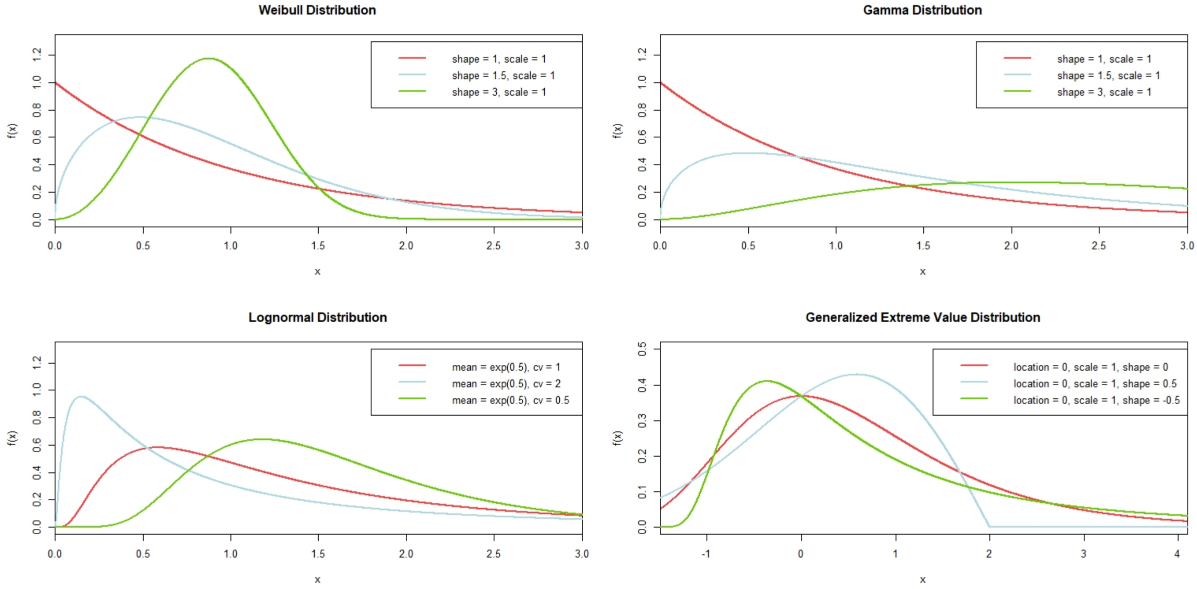


Figure 2.1: Examples of each distribution as the parameters change. As highlighted, the GEV distribution is the only one with three parameters, giving her more flexibility with respect to the others, and a domain that spans all of $\mathbb{R}$.

2.1.1. Weibull Distribution

Probably the most commonly used model in the field, this distribution was promoted in 1951 by the homonym physicist in Weibull et al. (1951) [52] and, since then, it has been employed in a variety of fields including physics, geography, economic, etc. Many variations of it exist but the traditional two-parameter Weibull distribution has been shown to perform particularly well in wind related application, for instance it was the best fit in all sites studied in Alrashidi et al. (2020) [2]. Its pdf is:

$$f(x) = \frac{\kappa}{\vartheta}\Big(\frac{x}{\vartheta}\Big)^{\kappa-1} exp\Big[-\Big(\frac{x}{\vartheta}\Big)^{\kappa}\Big] \tag{2.1}$$

while the Cumulative Density Function (CDF) is:

$$F(x) = 1 - exp\Big(-\Big(\frac{v}{\vartheta}\Big)^{\kappa}\Big) \tag{2.2}$$

where ϑ is the scale parameter and κ is the shape parameter.

2.1.2. Gamma Distribution

Another widely used distribution for wind speed modeling is the Gamma distribution. This model seems to achieve better results in areas characterized by higher wind speeds (Shi et al. 2021 [41]) and it is often proposed as an alternative to the Weibull distribution. The pdf is:

$$f(x) = \frac{1}{\Gamma(\kappa)\vartheta^{\kappa}} x^{\kappa-1} exp\Big(-\frac{x}{\vartheta}\Big) \tag{2.3}$$

The CDF is:

$$F(x) = \frac{\Gamma_{\frac{x}{\vartheta}}(\kappa)}{\Gamma(\kappa)} \tag{2.4}$$

where ϑ is the scale parameter, κ is the shape parameter and $\Gamma_{\frac{x}{\vartheta}}$ is the incomplete Gamma function.

2.1.3. Lognormal Distribution

The third model analyzed is the Lognormal distribution which is almost always proposed as a candidate distribution in similar studies and can sometimes prove to be the best choice, see for instance Tosunoğlu (2018) [48]. The pdf is:

$$f(x) = \frac{1}{x\kappa\sqrt{2\pi}} exp\Big(-\frac{(\ln(x)-\vartheta)^2}{2\kappa^2}\Big) \tag{2.5}$$

The CDF is:

$$F(x) = \Phi\Big(\frac{\ln(x)-\vartheta}{\kappa}\Big) \tag{2.6}$$

where ϑ is the scale parameter, κ is the shape parameter and Φ is the cumulative distribution function of the normal standard distribution.

2.1.4. GEV Distribution

We will dedicate a later chapter (see Subsection 3.1.2) to this distribution in order to discuss its derivation and principal theoretical properties. For the moment it is sufficient to know that it can be employed also to model directly the wind speed distribution and that it can achieve very good results, as shown again in Tosunoğlu (2018) [48], performing better than distributions, such as Weibull and Gamma, with a more consolidated curriculum.

2.2. Parameter Estimation Methods

All the models seen in the previous section are defined by particular parameters and, obviously, determining the best values for such parameters has a significant impact on the fitting of the models. While there exist a plethora of methods for this estimation, just the two most common and easily applicable ones have been considered in this work: the Maximum Likelihood Estimation (MLE) and the Method of Moments (MOM). Indeed, as will be shown in later sections, we have often worked on a big number of sites simultaneously to compare results and, for this reason, other parameter estimation methods, maybe more accurate but more complex and slower, have been discarded for computational reasons.

2.2.1. Maximum Likelihood Estimation

In statistics, the Maximum Likelihood Estimation (MLE) is a method for estimating the parameters of a probability distribution, given some observed data. This is achieved by maximizing the so called likelihood function so that, under the assumed statistical model, the observed data are most probable. In other words, the Maximum Likelihood Estimation method computes that combination of parameters such that the given data have the highest possible probability to be sampled from the assumed distribution with those specific parameters. The point in the parameter space that maximizes the likelihood function is called maximum likelihood estimate.

Although maximum likelihood was largely used by many mathematicians such as Carl Friedrich Gauss and Pierre-Simon Laplace, its widespread use rose between 1912 and 1922 when Ronald Fisher carefully analized, and consequently popularized, the maximum likelihood estimation (see Aldrich (1997) for more informations about Fisher's analysis on MLE [1]). However, MLE method was given a rigourous proof and trascended heuristic justification only in 1938, thanks to Samuel S. Wilks ([53]). The Wilks theorem shows that the error in the logarithm of likelihood values for estimates from multiple independent observations is asymptotically χ^2-distributed, enabling the determination of a confidence region around the estimate of the parameters. Today MLE has become a dominant technique in the context of statistical inference, mainly thanks to its flexibility and intuitive logic.

In order to apply MLE, one has to model data as a random sample from an assumed joint probability distribution with unknown paramaters values. The parameters governing the distribution are written as a vector $\vec{\vartheta} = [\vartheta_1, \vartheta_2, ..., \vartheta_k]^T$ so that the distribution will fall within a parametric family $\{f(\,\cdot\,;\vartheta) \mid \vartheta \in \Theta\}$, where Θ is the parameter space, a finite-dimensional subset of a Euclidean space. Evaluating this joint density at the observed data sample $\vec{y} = (y_1, y_2, ..., y_n)$ gives the Likelihood function

$$\mathcal{L}_n(\vec{\vartheta}) = \mathcal{L}_n(\vec{\vartheta}; \vec{y}) = f_n(\vec{y}; \vec{\vartheta})$$

The goal of MLE is to find the values of the model parameters that maximize the likelihood function over the space parameter, that is

$$\hat{\vartheta} = \arg\max_{\vartheta \in \Theta} \mathcal{L}_n(\vec{\vartheta}; \vec{y})$$

One of the main advantages of MLE method is that, for many distributions, there exists an analytic solution to the previous optimization problem.

2.2.2. Method of Moments

The method of moments (MoM) is an alternative to the method of maximum likelihood to estimate the population parameters of a sample of data. Although the idea of matching empirical moments of a distribution to the sample moments of a population dates back at least to Karl Pearson, the method of moments was formally introduced by Pafnuty Chebyshev in 1887 in the proof of the central limit theorem ([13]).

Just like the MLE method, it requires to assume the distribution of the given data but relies on a much simpler procedure. It starts by expressing the population moments (i.e. the expected values of powers of the random variable under consideration) as function of the parameters of interest and then, those expressions, are set equal to the sample moments computed directly from data. The solutions of this system of equations (note that the number of equations is equal to the number of parameters to be estimated) are estimates of the parameters. The method of moments offers a simple procedure and consistent estimators even if they are often biased; it is quicker than maximum likelihood method and its equations are much easier to solve even without the use of computers.

The idea of a variation of the method of moments was introduced by Greenwood et al. (1979) [16] and later expanded by Hosking et al. (1985) [20] to estimate the parameters of the generalized extreme value distribution (GEVD). This new method, named probability-weighted moments method, was developed specifically to approach the parameter estimation of the GEVD and achieves this goal through an iterative procedure. It starts from the definition of the ijk'th probability-weighted moment:

$$M(i,j,k) = \mathbb{E}[X^i F^j (1-F)^k]$$

where X is a random variable with cdf F.

Following from the fact that Hosking et al. (1985) [20] defined

$$\beta_j = M(i,j,0)$$

and Greenwood et al. (1979) [16] showed that

$$\beta_j = \frac{1}{j+1}\mathbb{E}[\max(X)]$$

it is possible to estimate the parameters of the GEV distribution (see abstract of Hosking et al. (1985) [20] for complete procedure).

2.2.3. Other Methods

Another commonly used method for parameter estimation is Least Squares (LS) where the parameters are estimated by minimizing the sum of squares of the deviation between the empirical CDF and the CDF of the model. However, because of their form, there is no estimator for the Lognormal and Gamma distributions and thus we could not apply this method here.

In addition to this more traditional strategies, some studies like the one conducted by Jiang et al. (2016) [27] have started to consider metaheuristic optimization methods for the search of the optimal parameters. These algorithms are inspired by the behaviour of groups in nature such as ant colonies, predatory behaviours and bat populations and often are named directly after them. However, because of the scarce literature regarding their applications in this field we decided not to study them.

2.3. Goodness of Fit Criteria

After the distribution model and its parameters have been determined, it is necessary to evaluate how well the model fit the real wind speed data to understand its real applicability. The goodness of fit criteria (GOF) give a measure of the distance between reality and model but one always needs to be careful since different criteria can lead to different results.

2.3.1. Root Mean Square Error

The RMSE determines the accuracy of the model through the item-by-item comparison between the observed probability and the estimated one. In particular, given the CDF, the formula for it is:

$$RMSE = \left[\frac{1}{n}\sum_{i=1}^{n}(F_i - \hat{F}_i)^2\right]^{\frac{1}{2}} \tag{2.7}$$

where F_i is the empirical CDF and $\hat{F}_i$ is the estimated CDF.

The closer the RMSE is to 0, the better will be the fitting effect. The main problem of the RMSE is that it is particularly sensible to big errors because of the squaring factor; thus, even a few large distances will increase it by a lot. In this case, however, the comparison is done between two cumulative distribution functions and no big outlier should be present, making this index very suited for the job.

2.3.2. Coefficient of Determination R^2

The R^2 is expressed as the square of the correlation coefficient between the observed and the estimated value: several variants of it exist but the one used here is the one referring to the CDF:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(F_i - \hat{F}_i)^2}{\sum_{i=1}^{n}(F_i - \bar{F}_i)^2} \tag{2.8}$$

where F_i is the empirical CDF, $\bar{F}_i = \frac{1}{n}\sum_{i=1}^{n} F_i$ and $\hat{F}_i$ is the estimated CDF.

The closer the value of R^2 to 1, the better the fit of the model. However, this index gives more weight to values in the middle part of the distribution and thus cannot reflect completely the fitting effect of the theoretical distribution.

2.3.3. Mean Absolute Error

The MAE, similarly to the RMSE, is a measure of error between paired observations but in this case only the absolute value of the differences is considered. The formula to compute it is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|F_i - \hat{F}_i| \tag{2.9}$$

where, again, F_i is the empirical CDF and $\hat{F}_i$ is the estimated CDF.

Also in this case, the closer the MAE is to 0 the better the result.

2.3.4. Wasserstein Distance

The Wasserstein distance, or Kantorovich-Rubinstein metric, is a distance function defined between probability distributions on a given metric space M. Intuitively, if each distribution is viewed as a unit amount of earth (soil) piled on M, the metric is the minimum "cost" of turning one pile into the other, which is assumed to be the amount of earth that needs to be moved times the mean distance it has to be moved. It was first presented in Kantorovich (1939) [28].

In particular, the first order Wasserstein distance between one-dimensional distributions is defined as:

$$W_1 = \int_{\mathbb{R}} |F(x) - \hat{F}(x)|dx \tag{2.10}$$

where F is the empirical CDF and $\hat{F}$ is the estimated CDF.

Clearly, the lower the distance between the empirical distribution and the one obtained from the estimated parameters, the better will be the fitting power of the model.

2.3.5. Other criteria

Numerous other criteria are available, each with advantages and disadvantages but we decided to limit ourselves to previously cited ones. Just to cite some of them, we have Kolmogorov-Smirnov Test, Anderson-Darling test, Chi-Square (χ^2) Test, Akaike Information Criterion and so on.

2.4. Analysis

Following what has been said up until now, we analysed the wind speed distribution of the series of data at our disposal.

We need to specify now that temporal dependence has been completely disregarded at this stage in accordance to the common practice described in the literature; in particular, each hourly sample is considered to be independent and coming from a common distribution not changing throughout the years. This may be considered a bit of a stretch since, clearly, the wind speed at a certain time is dependent on the speed at previous times but this hypothesis is commonly used in all studies of this kind and deciding not to adopt it would make the use of certain methods, such as MLE, simply impossible. Moreover, also the stationarity of data has been taken as an hypothesis for the moment in order to be able to apply the procedure described.

To determine the best distribution at each site, we compared the fit of the four models (Weibull, Gamma, Lognormal, GEV) considering two parameter estimation methods (MLE and MOM) and comparing the results of four GOF criteria (RMSE, R^2, MAE and Wasserstein distance), for a total of 32 indices for each site. In particular, for the estimation of the parameters we exploited `egevd`, `eweibull`, `egamma` and `elnormlAlt` functions from the `R` package `EnvStats` [31].

Moreover, to avoid overfitting and to follow a more proper procedure, we implemented a cross validation process in which we left five years (randomly extracted) of data as validation set; the parameters are estimated on the other 26 years of data and we chose the model with the best goodness-of-fit evaluated on the validation set. This process is repeated 6 times to use (almost) all data as validation and the GOF metrics are averaged for each combination of distribution and estimation model. Finally, the best couple (distribution - estimation model) is trained again on all the data to recover the best parameters estimate.

The results for a couple of individual sites are summarised in Tables [2.1], [2.2], [2.3] and [2.4]. Just from this example, one can find confirmation of what is said in the literature: there is no "best distribution" to model instantaneous wind speed that always achieve the best result at each site, but they need to be analyzed case by case. Moreover, also the technique used for the parameter optimization may affect the results, bringing even more variability to the table. For instance, Tables [2.3] and [2.4] refer to the same location but, changing the optimization method results in a change of the chosen distribution according to 1 of the 4 indices.

	RMSE	R^2	MAE	Wasserstein Distance
GEV	0.02015724	0.9966576	0.01609812	0.3307657
Weibull	0.03668135	0.9929857	0.03164403	0.6789298
Lognormal	0.04258096	0.9855509	0.03658149	0.7595776
Gamma	0.02571846	0.9953128	0.02133410	0.4496034

Table 2.1: Brianza - Maximum Likelihood Estimation

	RMSE	R^2	MAE	Wasserstein Distance
GEV	0.01932669	0.9968728	0.01548286	0.3283987
Weibull	0.03891265	0.9924547	0.03399819	0.7179164
Lognormal	0.02996991	0.9911330	0.02528687	0.4128175
Gamma	0.03017341	0.9941592	0.02573078	0.5025506

Table 2.2: Brianza - Method of Moments

	RMSE	R^2	MAE	Wasserstein Distance
GEV	0.02399874	0.9954963	0.01936980	2.108358
Weibull	0.03554903	0.9873262	0.02970932	1.442257
Lognormal	0.01584138	0.9978477	0.01312012	1.537694
Gamma	0.03080732	0.9913051	0.02477164	1.244782

Table 2.3: Alps - Maximum Likelihood Estimation

	RMSE	R^2	MAE	Wasserstein Distance
GEV	0.02942303	0.9921280	0.02425659	1.596784
Weibull	0.03219626	0.9899551	0.02652980	1.184094
Lognormal	0.04005979	0.9937789	0.03494203	1.815344
Gamma	0.02617657	0.9928818	0.02152196	1.016126

Table 2.4: Alps - Method of Moments

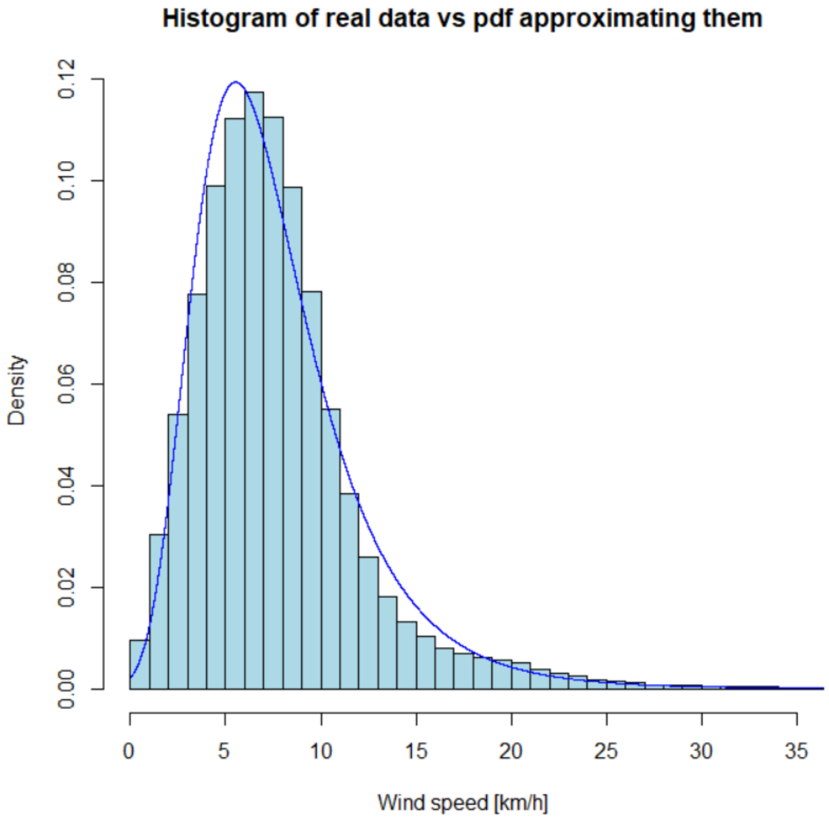


Figure 2.2: Example of comparison between real data histogram and estimated probability density function. While the approximation looks overall good, due to a lack of data in the right tail, this region it's likely to have a non negligible error.

Something to be happy about is that, as highlighted in Figure 2.2, in general, the results are convincing, in the sense that the distributions approximate quite well the true wind speed data. However, keeping an eye on what has to come, this result is biased by the fact that most of the measurements fall in the middle part of the distribution and, while in that region the estimation can be very accurate, the fit may be worse on the tails, and in particular on the right one, where the focus should be in a study regarding extreme winds. More details and comments on this regard will be given in Chapter 3.

2.5. Extension on a Bigger Area

Having looked at the results obtained on single sites, a natural continuation of the process is to study the behaviour of wind speed distributions on a bigger area. In particular, for a number of reasons including the familiarity of the region and the diversified orography, we opted into studying the territory corresponding to the Lombardy region (Figure 2.4a).

A first point of interest is to visualize the results of each model on the whole area (Figure 2.3). For the sake of comparison, we have chosen to employ only the Method of Moments as a parameter estimation method since it is the fastest one and the Wasserstein distance to compare performances since it is the most appropriate method to deal with distributions.

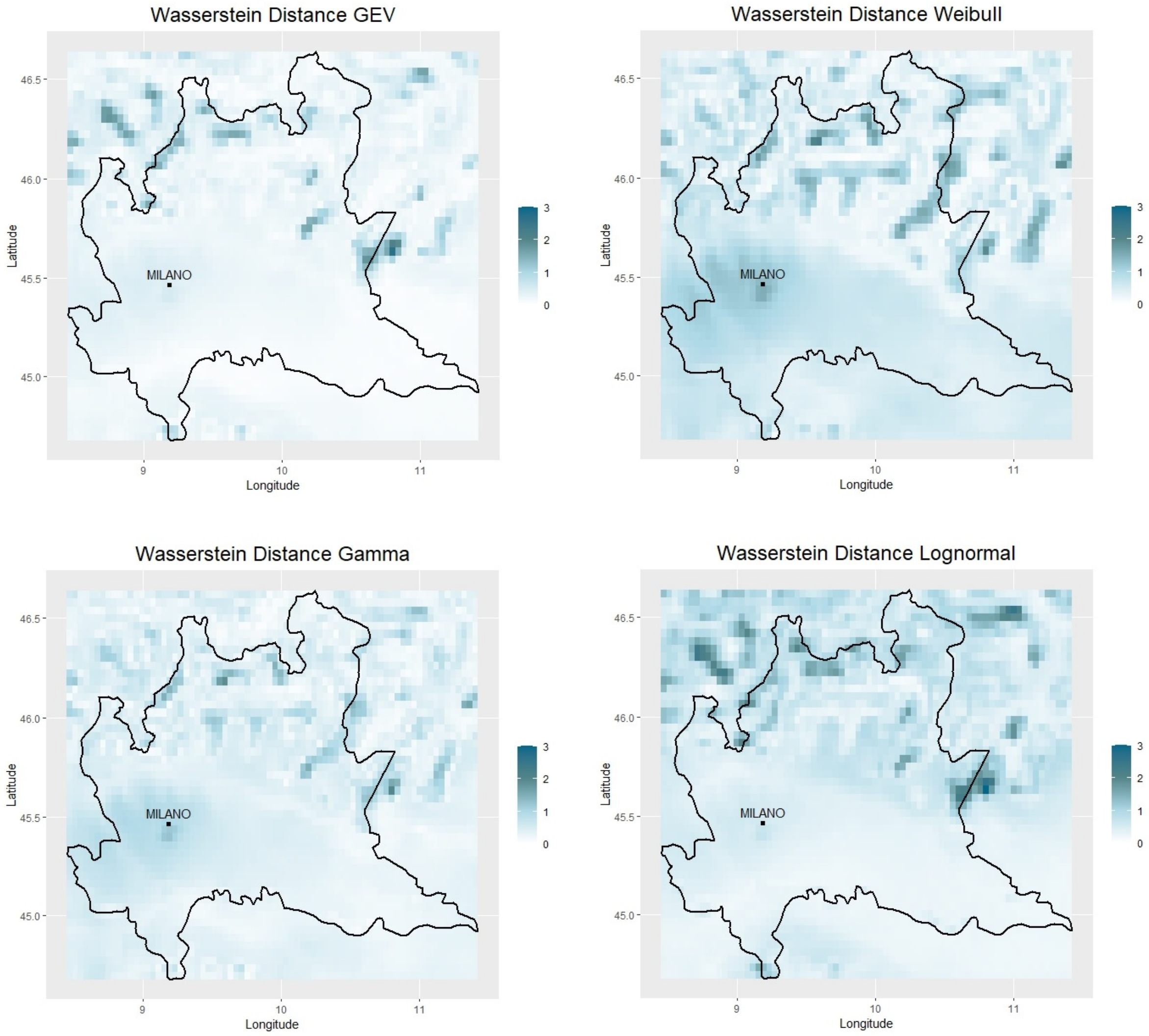
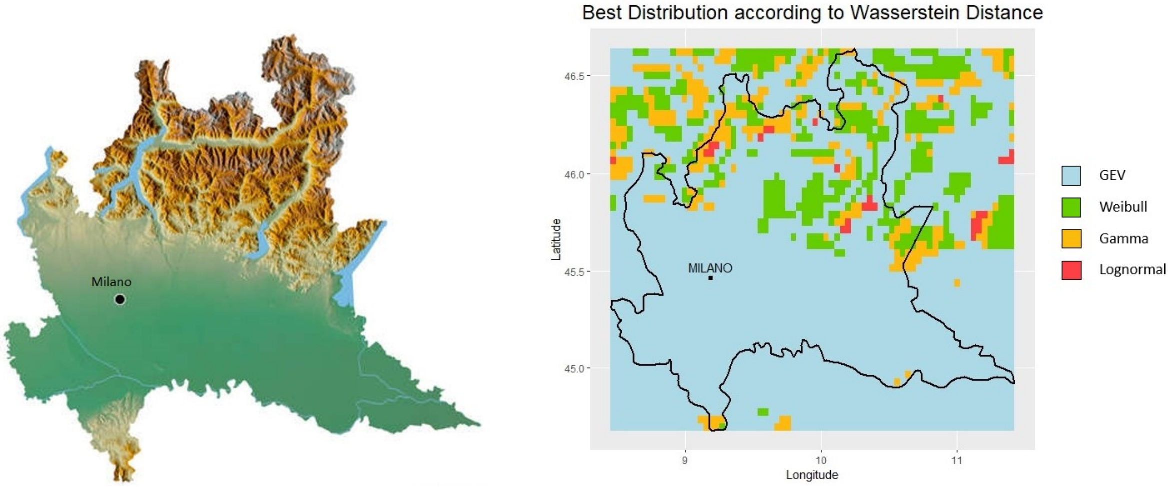

Wasserstein Distance GEV
Wasserstein Distance Weibull
Wasserstein Distance Gamma
Wasserstein Distance Lognormal
Latitude
Longitude
46.5
46.0
45.5
45.0
9
10
11
MILANO
3
2
1
0

Figure 2.3: For each one of the 4 considered cases, we computed the best fitting distribution in each location and measured the Wasserstein distance between that and the empirical distribution of data in that cell.

Although we once again found confirmation of the fact that there is no "best distribution" that takes it all, what seems to be a clear result is that the GEV distribution achieves better results on the majority of the region. Indeed, in Figure 2.4b, a direct comparison is reported, showing the supremacy of this model in particular in the area of the Po Valley. Regarding the other distributions, all the three of them find application over the alpine arc, with the Gamma and Weibull being more present than the Lognormal, which instead is outperformed almost everywhere by at least one distribution and, thus, is nearly never chosen.



(a) Lombardy orography

(b) Best distribution for each location

Figure 2.4: While the predominance of the GEV is cristal clear, especially over the Po Valley, where the land is mountainous and winds are stronger on average, we find all four distributions.

Referring to Figure 2.5 we can find further insights on where each distribution is chosen. Starting with the GEV distribution, it is clearly the best performing distribution when it comes to modeling low wind speed profiles. Furthermore, thanks to the flexibility given by having three parameters, GEV is sometimes chosen also in northern areas, where winds are stronger. Coming to Weibull distribution, it finds application in areas with slightly higher winds than those modeled by GEVD, on hills and mountains, but it is outperformed by Gamma and Lognormal when average winds are consistently higher. Gamma, confirming what stated in the literature (Shi et al. (2021) [41]), is an alternative to Weibull that performs better over high wind speed profiles. Lognormal, on the other hand, finds very little application in our studies, modeling only areas with extreme average wind speeds.

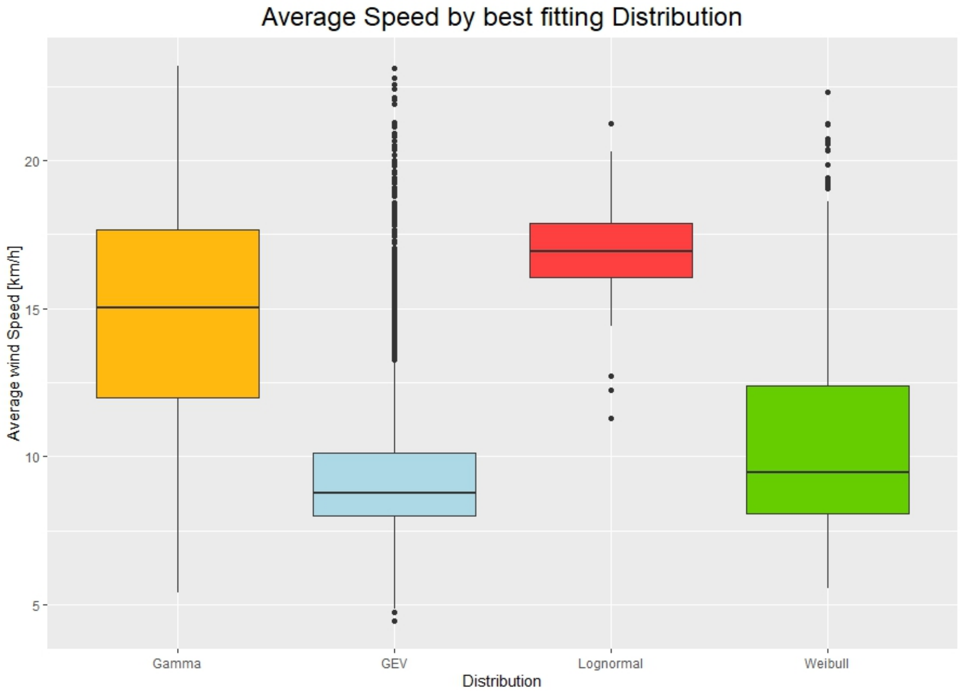


Figure 2.5: Boxplots of average wind speeds by best fitting distribution.

2.6. Final Observations

The main point to bring home from this chapter is the importance of an estimation of wind speed distributions that is the most accurate possible. As we said, this is the cornerstone on which many procedures, regarding both hazard analysis and energy production, are based: an error here will flow down to subsequent steps, altering results. The great number of possible models, estimation methods and goodness of fit criteria presented here and collected in the literature are proof of its importance and the choice of the most appropriate strategy to produce this estimation may be vital in many studies, possibly making the difference with respect to a potential economic loss. In Chapter 7 we will see a practical example of this, where results coming from different estimations will be compared to highlight the importance of this step in relation to wind energy production.

Then, we need to remark that, when we first approached this topic, we were well aware of the fact that the phenomena causing windy perturbations have a continuous nature, in the sense that measurements taken in locations few meters apart will likely be almost the same. However, what we could sense from these preliminary analysis is that simple geographical distance provides too little information to satisfactorily describe winds behaviour. Indeed, what we noticed is that, more often than not, geographically close data in mountainous areas show very different wind distributions (think, for example, of two different sides of the same mountain), while in the Po Valley we were able to discover many examples of locations tens of kilometers apart exhibiting very similar distributions. Following this intuition, we decided to further investigate this path and tried to group the locations of the region under analysis based on the wind characteristics. This topic will be deeply examined in Chapter 4, from the point of view of hazard, and in Chapter 6 for what concerns wind regimes.

3 | Extreme Value Analysis

Having studied the behaviour of the wind speed series as a whole, we now move to one of the cores of this research: the analysis of extreme phenomena.

Extreme wind speeds pose a threat to all infrastructures and in particular to the transmission network of the electric energy, whose pylons are subject to all kind of meteorological events. In this context, the study of high speed winds is of paramount importance to improve the reliability of the network; knowing the areas exposed to the greater risk allows to intervene preventively with appropriate countermeasures, avoiding possible disasters and saving resources that otherwise would be wasted in reparations.

In general, the main aim is to estimate either the return period T of a certain wind speed, i.e. given a threshold value defined by some standard, T is the average number of years before this threshold is excedeed, or, viceversa, the return level over a predetermined period of time T, i.e. the maximum annual wind speed which are exceeded once every T years, with T usually ranging from 10 to 100 years. In practice, what one needs are the $1 - \frac{1}{T}$ quantiles of the distribution of the annual maximum wind speed.

A more complete description would be given by the so-called exceedance probability curves, graphs reporting the probability of exceeding each possible wind speed value (see Figure 3.1). This graphs are easily obtained as 1−CDF (where CDF is the cumulative distribution function of the annual wind speed) but, in return, obtaining this CDF may be a challenging task and lot of studies have been conducted on which are the most appropriate strategies to model extreme values.

This chapter will collect a review of many procedures of extreme value analysis coming from the literature and some proposals from us. The strategies are evaluated on the dataset to understand the criticalities of each one and determine the most appropriate way to proceed.

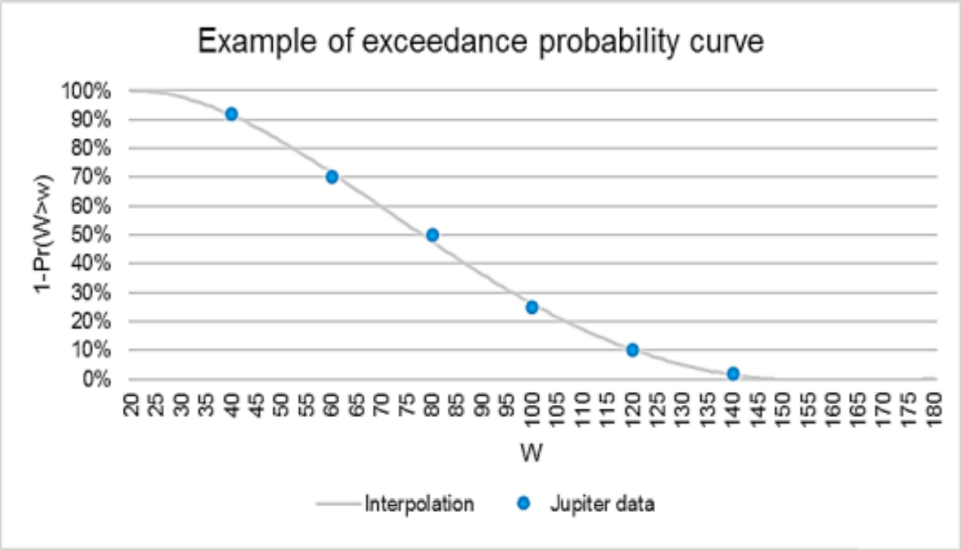


Figure 3.1: The graph shows an example of exceedance probability curve obtained by "Terna - Rete Elettrica Nazionale S.p.A." in a previous study they conducted [45]. The criticality here is that they consider just 6 estimated values, relative to 6 thresholds of risk (the blue dots in figure) and then they perform an interpolation to create a curve. As we will see later on, we aimed at recreating these kind of graphs using just our data and evaluations, obtaining entire curves and not interpolating points.

3.1. Extreme Value Theory

In this section and in the next one, a review of the theory regarding extreme values will be carried out, following in large part the description provided in "An Introduction to Statistical Modeling of Extreme Values" by S. Coles (2003) [14].

Consider $M_n = \max\{X_1, ..., X_n\}$ where $X_1, ..., X_n$ is a sequence of independent random variables having a common distribution function F. In the applications, X_i usually represents the value of a process, such as the wind speed series, measured on a regular time scale, so that M_n represents the maximum of the process over n observation; in particular, n is usually chosen so that M_n corresponds to the annual maximum.

Then, in theory, the distribution of M_n can be derived exactly for all values of n as:

$$\mathbb{P}(M_n \leq x) = \mathbb{P}(X_1 \leq x, ..., X_n \leq x) = \{F(x)\}^n \tag{3.1}$$

However, this is not immediately useful in practice since the distribution F is unknown and one would need to estimate it from data, as described in the previous chapter, and then substitute this estimate into formula 3.1. Unfortunately, as described in Perrin et al. (2006) [34] and confirmed by the results presented later in this chapter, very small discrepancies in the estimate of F can lead to substantial discrepancies for F^n.

The alternative approach is to accept that F is unknown and to look for approximate families of models for F^n to be estimated on the basis of extreme data only.

As commonly done in mathematics, we proceed by looking at the behaviour of F^n for the limit $n \to \infty$. However, one needs to be careful because for any $x < x_+$ where x_+ is the upper end-point of F, i.e. the smallest value of x such that $F(x) = 1$, $F^n(x) \to 0$ as $n \to \infty$, so that the distribution of M_n degenerates to a point mass on x_+. This difficulty is avoided by allowing a linear renormalization of the variable M_n:

$$M_n^* = \frac{M_n - b_n}{a_n}$$

for appropriate sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ and then by looking for limiting distributions for M_n^* instead of M_n.

3.1.1. Extreme Value Theorem

The entire range of possible limit distributions for M_n^* is given by the following Extreme Value Theorem. However, before stating it, it is necessary to define the three distributions that will come into play:

- Gumbel distribution, or Fisher-Tippett Type 1 distribution (FT1):

$$PDF : f(x) = \frac{1}{\sigma} exp\Big\{ - \frac{x-\mu}{\sigma} + exp\Big\{ - \frac{x-\mu}{\sigma}\Big\}\Big\}$$

$$CDF : F(x) = exp\Big\{ - exp\Big\{ - \frac{x-\mu}{\sigma}\Big\}\Big\}$$

  where μ and σ represent respectively the parameters of location and scale.

- Fréchet distribution, or Fisher-Tippett Type 2 distribution (FT2):

$$PDF : f(x) = \begin{cases} \frac{\xi}{\sigma}\Big(\frac{x-\mu}{\sigma}\Big)^{-1-\xi} exp\Big\{ - \Big(\frac{x-\mu}{\sigma}\Big)^{-\xi}\Big\} & \text{if } x > \mu \\ 0 & \text{if } x \leq \mu \end{cases}$$

$$CDF : F(x) = \begin{cases} exp\Big\{-\Big(\dfrac{x-\mu}{\sigma}\Big)^{-\xi}\Big\} & \text{if } x > \mu \\ 0 & \text{if } x \leq \mu \end{cases}$$

where μ, σ and ξ represent respectively the parameters of location, scale and shape.

- Reverse Weibull distribution, or Fisher-Tippett Type 3 distribution (FT2):

$$PDF : f(x) = \begin{cases} \dfrac{\xi}{\sigma}\Big(\dfrac{-x+\mu}{\sigma}\Big)^{-1+\xi} exp\Big\{-\Big(\dfrac{-x+\mu}{\sigma}\Big)^{\xi}\Big\} & \text{if } x < \mu \\ 0 & \text{if } x \geq \mu \end{cases}$$

$$CDF : F(x) = \begin{cases} exp\Big\{-\Big(\dfrac{x-\mu}{\sigma}\Big)^{-\xi}\Big\} & \text{if } x < \mu \\ 0 & \text{if } x \geq \mu \end{cases}$$

where μ, σ and ξ represent respectively the parameters of location, scale and shape.

Theorem 3.1 (Extreme Value Theorem). *If there exists a sequence of constants $\{a_n > 0\}$ and $\{b_n\}$ s.t.*

$$\mathbb{P}\Big(\frac{M_n - b_n}{a_n} \leq x\Big) \rightarrow G(x) \qquad as \quad n \rightarrow \infty, \tag{3.2}$$

then G belongs to one of the three possible families, namely the Gumbel, the Fréchet and the Reverse Weibull distribution.

The remarkable feature of this result is that the three types of extreme value distributions are the only possible limits for the distribution of M_n^*, regardless of the parent distribution F of the population. In this sense, the theorem provides an extreme value analog of the central limit theorem.

The complete proof of the theorem can be found, for instance, in "Extremes and Related Properties of random Sequences and Processes" by M.R. Leadbetter et al. (1983) [29].

3.1.2. Extreme Value Distribution

The three types of limits that arise in theorem 3.1 have distinct shapes, corresponding to the different behaviours of the tail of the distribution F of the X_i. This can be made precise by considering the behaviour of the limit distribution G at the upper point x_+. For the Reverse Weibull, x_+ is finite while for both the Gumbel and the Fréchet distributions $x_+ = \infty$. However the density of G decays exponentially for the former and polynomially for the latter.

In early applications it was common use to adopt a priori one of the three families but this strategy required a technique to choose the most appropriate one and did not account for the uncertainties related to this choice. Later, the three asymptotes have been combined into a single distribution firstly defined by Von Mises (1936) [49] and known today as "Generalized Extreme Value" (GEV) distribution:

$$\mathbb{P}(X < x) = e^{-\Lambda(x)}; \qquad \Lambda(x) = \begin{cases} \Big[1 + \xi \dfrac{(x-\mu)}{\sigma}\Big]^{-\frac{1}{\xi}} & \text{if } \xi \neq 0 \\ exp\Big\{ - \dfrac{(x-\mu)}{\sigma}\Big\} & \text{if } \xi = 0 \end{cases} \tag{3.3}$$

where μ and σ are, respectively, the location and scale parameters, while ξ is the shape factor which determines the asymptotic form adopted by the GEV distribution: $\xi = 0$ corresponds to Gumbel distribution, $\xi > 0$ to Frechet distribution and $\xi < 0$ to reverse Weibull distribution (see Figure 3.2).

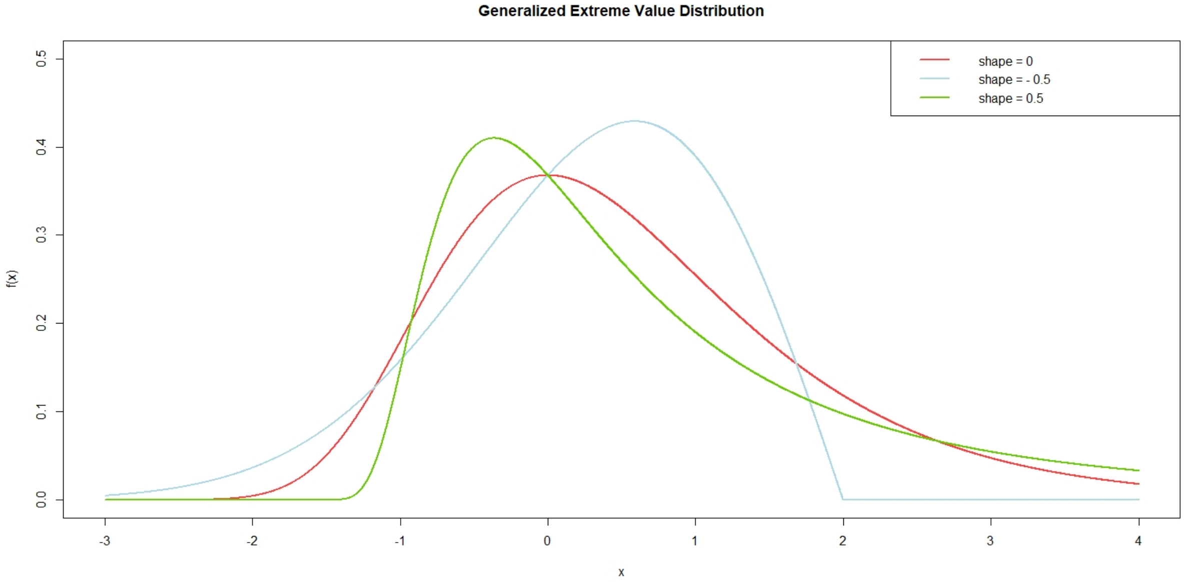


Figure 3.2: The three distributions displayed are GEV with location parameter $\mu = 0$ and scale parameter $\sigma = 1$.

One thing to keep in mind when using this distribution is that no data fitting can yield the condition $\xi = 0$ since it is associated with a singularity of the exponent; this fact makes it necessary to use a test to verify the shape factor estimated and thus an ad-hoc Z-test is employed (see `zTestGevdShape` in `R` package `EnvStats` [31]).

The unification of the original three families of extreme value distribution into a single family greatly simplifies statistical implementation. Through inference on ξ, the data themselves determine the most appropriate type of tail behavior, and there is no necessity to make subjective a priori judgements about which individual extreme value family to adopt. Moreover, uncertainty in the inferred value of ξ measures the lack of certainty on which type among the original three is the most appropriate for a given dataset.

Now we can look back at theorem 3.1; combining formula 3.2 with formula 3.3 one can clearly see the practical meaning of it: the limiting distribution of the maximum of a process over n observation is the Generalized Extreme Value Distribution. From the point of view of wind speed analysis, in particular, this translates to the fact that the wind speed annual maxima can be modeled using the GEV distribution. Thus, one needs only the extreme values to study their behaviour, without the need of the whole time series.

3.2. Classical Extensions

Extreme Value Analysis has an inevitable weakness: extreme values, by definition, are scarce and a small amount of data leads to higher uncertainties in the results (Torrielli et al. (2013) [46]). This is particularly true in a field like wind series analysis, where measurements never exceed 40-60 years and often times are much less. For this reason, there has been a good effort in the literature to find alternative ways to study the annual maximum wind speed distribution and improve the accuracy on the prediction for design values.

This section will be dedicated to a theoretical introduction on two of the most famous of these methods, belonging to the family of thresholding methods: the r largest order statistic model (r-LOS), which selects the r largest observations per epoch, and the peak over threshold (POT) method, which analyzes all values exceeding a predefined threshold.

As observed in Palutikof et al. (1999) [33], a main drawback of techniques like these is that the choice of the censoring greatly affects the estimated parameters of the distribution and this decision has to be taken by the analyst for each temporal series. This flaw made these kind of approaches impracticable for our purposes since we will mainly focus on large areas and the large number of time series involved makes it impossible to tune the censoring site by site. Therefore, we decided not to travel this path for our studies and we will just present these methods in a theoretical way because of their scientific interest.

3.2.1. r-LOS Method

Here we need to extend the result of the previous section by considering $M_n^{(k)} = k$-th largest of $\{X_1..., X_n\}$ and identifying the limiting behaviour of this variable for k fixed and $n \to \infty$. In particular, we usually require the characterization of the whole vector $\mathbf{M_n^{(r)}} = (M_n^{(1)}, ..., M_n^{(r)})$ and its joint distribution is readily given in the following theorem (see again [14] for the proof).

Theorem 3.2. *If there exist sequences of constants $\{a_n > 0\}$ and $\{b_n\}$ s.t.*

$$\mathbb{P}\Big(\frac{M_n - b_n}{a_n} \leq x\Big) \to G(x) \qquad as \quad n \to \infty,$$

for some non-degenerate distribution function G, then, for fixed r, the limiting distribution as $n \to \infty$ of

$$\tilde{\mathbf{M}}_\mathbf{n}^{(\mathbf{r})} = \Big(\frac{M_n^{(1)} - b_n}{a_n}, \dots, \frac{M_n^{(r)} - b_n}{a_n}\Big)$$

falls within the family having joint probability density function:

$$f(x^{(1)}, ..., x^{(r)}) = exp\Big\{ - \Big[1 + \xi\Big(\frac{x^{(r)} - \mu}{\sigma}\Big)\Big]^{-\frac{1}{\xi}}\Big\} \times \prod_{k=1}^{r} \frac{1}{\sigma}\Big[1 + \xi\Big(\frac{x^{(r)} - \mu}{\sigma}\Big)\Big]^{-\frac{1}{\xi}-1} \tag{3.4}$$

where $-\infty < \mu < \infty$, $\sigma > 0$ and $-\infty < \xi < \infty$; $x^{(r)} \leq x^{(r-1)} \leq \dots \leq x^{(1)}$; and $x^{(k)}$ is s.t. $1 + \xi\frac{x^{(k)} - \mu}{\sigma} > 0$ for $k = 1, \dots, r$.

Notice that in the case $r = 1$, formula 3.4 reduces to the GEV family of density functions and the case $\xi = 0$ is to be interpreted as the limiting form $\xi \to 0$ similarly to what has been done for the Gumbell distribution.

Then, this joint distribution provides the basis for the Maximum Likelihood method. Consider R years of data with r-LOS values extracted from each year; the likelihood is simply the product of the R densities:

$$L(\mu, \sigma, \xi) = \prod_{i=1}^{R} f_i(x_i^{(1)}, ..., x_i^{(r)})$$

and the optimal parameters can be recovered by maximizing the log-likelihood as usual. These parameters correspond to those of a GEV distribution of annual maxima but incorporate more of the observed data. Thus, the interpretation is unaltered but precision should be improved due to the inclusion of extra information.

The difficulty of this method is that the selected r-LOS must be independent events.

While this is reasonable to assume for annual maxima, in this case a suitable separation interval should be set between observations of the same year. This issue is strictly related to the choice of r: a practical criterion is to set it so as to minimize the variance associated with the parameters but ultimately it must be decided by the analyst and in previous applications various values between 3 to 10 have been proposed.

3.2.2. POT Method

POT method relies on exctracting the peak values reached in any period of time whose values exceed a certain threshold. This method allows for the use of sub-annual maxima but, on the other hand, the analysis involves fitting two distributions: one for the number of events in a time period and the second for the entity of exceedances.

Consider a variable V having a parent $F_V(v)$ such that the distribution of the largest value in the period T converges to one of the asymptotes combined in equation 3.3. With this assumption, we can exploit the following theorem:

Theorem 3.3 (Pickhands, Balkerna, De Hann - 1975). *Let $(V_1, V_2, ...)$ be a sequence of independent and identically-distributed random variables, and let $(X_1, X_2, ...) = (V_1 - u, V_2 - u, ...)$ be the sequence of the excesses beyond threshold u. Let $F_u(x)$ be the conditional distribution function of the excesses.*

Then, for u large enough, $F_u(x)$ is well approximated by the generalized Pareto distribution, namely:

$$F_u(x) \to G_{\sigma,\xi}(x), \quad as\ u \to \infty$$

where

$$G_{\sigma,\xi}(x) = \begin{cases} 1 - \left[1 - \xi \dfrac{x}{\sigma}\right]^{\frac{1}{\xi}} & if\ \xi \neq 0 \\ 1 - exp\left\{ - \dfrac{x}{\sigma}\right\} & if\ \xi = 0 \end{cases}$$

where σ is the scale parameter and ξ the shape parameter.

Moreover, under the assumption that the threshold u is large enough, we can consider its crossings to be independent, and the number N of values over the threshold in a period T to be Poisson-distributed, with a rate λ_u/year. A possible unbiased estimate of λ_u is n/T where n is the total number of exceedances of u counted directly from the data, and T is the number of years of the record. In such a case, the distribution of the largest value of V in T is given by:

$$\mathbb{P}(\hat{V}_T < v) = \begin{cases} exp\Big\{ -\lambda_u T\Big(\xi\dfrac{v-u}{\sigma}\Big)^{\frac{1}{\xi}}\Big\} & \text{if } \xi \neq 0 \\ exp\Big\{ -\lambda_u T exp\Big\{ -\dfrac{v-u}{\sigma}\Big\}\Big\} & \text{if } \xi = 0 \end{cases} \tag{3.5}$$

From equation 3.5 it is clear that the distribution strictly depends on the value of the threshold u. When choosing the value of this hyperparameter, one has to carefully evaluate the trade-off between number of data and independence of them: on one hand, u has to be set low enough to ensure that a sufficient quantity of data is used to estimate the distribution parameters, while on the other hand the asymptotic requirement $u \to \infty$ must be satisfied.

Once the value of u is chosen and λ_u is obtained from data, the POT method reduces the fitting problem to the estimation of only two parameters instead of the three required by the GEV approach.

3.3. New Maxima from Data Augmentation

Another possible approach we came up with in order to mitigate the lack of data at our disposal is data augmentation. Our main idea was to exploit all the analysis previously conducted to approximate the distribution of instantaneous wind speed data to simulate new data, sampling new observations directly from the distributions found in Chapter 2. In our case, we had 31 annual maxima for each site and this number could be in principle considered sufficient for inference; however, we still tried to apply this method to reduce the uncertainty.

To this end, for each site, we decided to sample 1000 new simulated years, each one consisting of 365×24 values, directly from the best approximating distribution associated to the cell. Then, by extracting the maximum value of each year we created a new pool of 1000 simulated annual maxima and used them to estimate the parameters of the GEV distribution approximating the 31 original maxima. One could argue that by proceeding in this way, the simulated years are not actual time series but only a family of independent samples and we are aware of this criticality but, since our only interest was to collect the maxima, we neglected this matter.

This procedure is somewhat similar, even if more simplified, to what is described in Torrielli et al. (2014) [47] where a more complete characterization of the wind speed series is produced before simulating the new data. In particular, the probability distribution function (pdf) and the power spectral density function (psdf) of the original series are considered and studied in detail and, then, the simulation algorithm tries to match both of them to recreate a process as similar as possible to the actual wind speed series.

This method proposed in Torrielli et al. (2014) [46] seems to achieve good results but, again, it suffers of too much specificity; a precise psdf characterization is needed to produce accurate results and it is not clear if it could be possible to analyse multiple time series automatically and quickly. For this reason and because of its overall complexity, we discarded this method in our comparison, just reporting it here for completeness.

3.4. Results

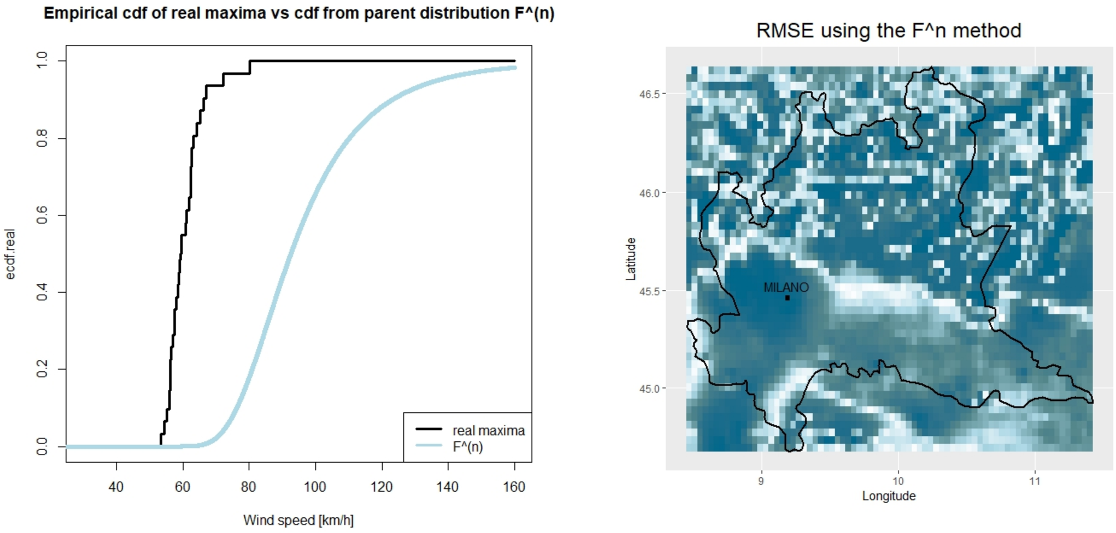
In this section, a comparison between the methods of Extreme Value Analysis described up to now will be carried on with the aim to define the best and most practical way to model extreme values.

In particular, results will be evaluated according to the RMSE obtained from the comparison between the CDF estimated by the method and the empirical CDF coming from the actual 31 annual maxima extracted from the time series. As usual, the comparison will be done on the area corresponding to the Lombardy region to underline possible territorial effects.

The first method is the F^n one, meaning that we are using directly formula 3.1 to model the CDF of the annual maxima. The parent distribution F is obtained as described in Chapter 2, choosing for each cell the one best approximating real data. Results are displayed in Figure 3.3: the average RMSE is around 0.4 and some patters can be noticed in the plain areas.

Confirming what was already to be expected from the theory, Figure 3.4 shows a much better result for the GEV method, i.e. the direct modelling of the 31 annual maxima with the GEV distribution. Here the the average RMSE is 0.04, ten times lower, and no particular pattern can be seen, proving the solidity and generality of this method.

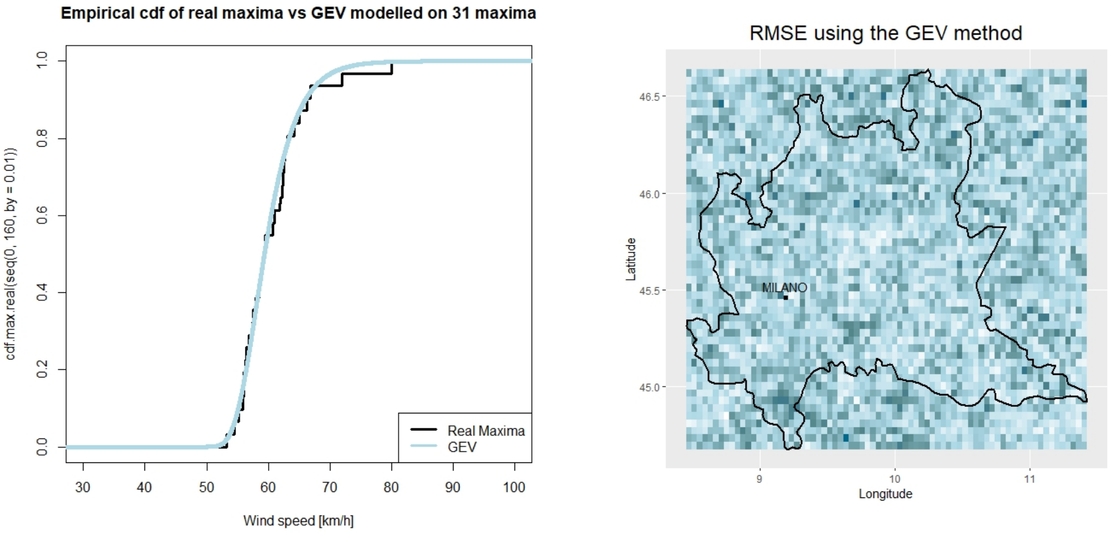
The final technique analysed is the one described in the previous section, consisting in the simulation of new data. Results (Figure 3.5) are similar to the ones of the F^n method since they both inherently suffer from the same flaw: when approximating the parent distribution, most of the effort is devoted where the number of data is greater, i.e. in the middle part, while little accuracy is used in the tails. This then cascades down to



(a) Pavia (Lat: 45°11' N, Lon: 9°09' E) - Parent distribution: $F \sim$ GEV

(b) RMSE computed in each site of Lombardy

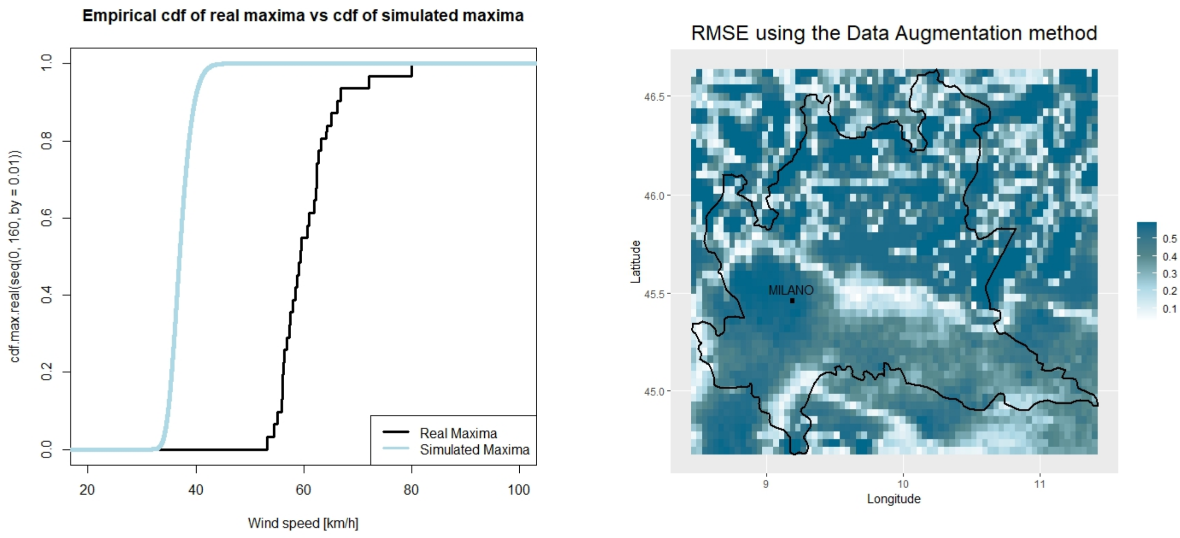
Figure 3.3: Performance of F^n method when it comes to fitting maxima data. RMSE is computed comparing real data empirical cdf and the cdf obtained from F^n, where F is the parent distribution (i.e. the distribution that better fits the instantaneous data).



(a) Pavia (Lat: 45°11' N, Lon: 9°09' E) - Parent distribution: $F \sim$ GEV

(b) RMSE computed in each site of Lombardy

Figure 3.4: Performance of GEV method when it comes to fitting maxima data. RMSE is computed comparing real data empirical cdf and the cdf of the GEV directly fitted from the real maximum data.



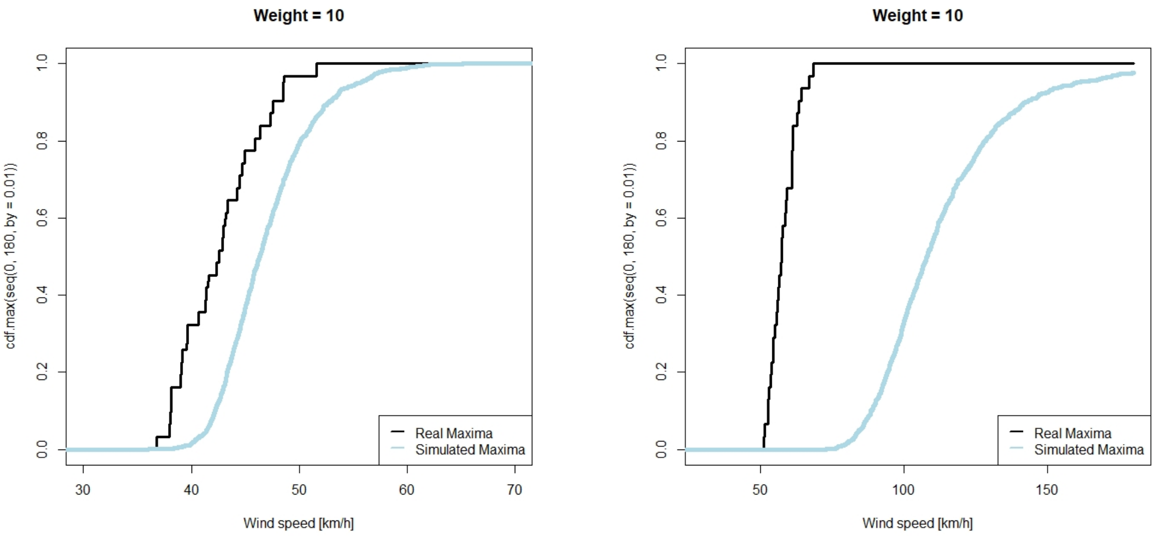
(a) Pavia (Lat: 45°11' N, Lon: 9°09' E) - Parent distribution: $F \sim$ GEV

(b) RMSE computed in each site of Lombardy

Figure 3.5: Performance of data augmentation method when it comes to fitting maxima data. RMSE is computed comparing real data empirical cdf and the cdf computed from the simulated maxima.

the subsequent phases of the procedure and, as a result, maxima estimation is imprecise, being consistently lower or higher (depending on the cases) than actual values.

In order to improve the performance of this last method we also tried a different optimization approach for computing the best distribution from which to simulate data. Since our problem was the scarcity of data in the right tail, we tried to implement a weighted optimization algorithm that gives more relevance to extreme data. In order to do so, we built a grid of possible values for the parameters of the chosen distribution around the values of the parameters obtained from classical optimizazion, and evaluated each point of this grid with a weighted version of RMSE, giving more weight to the errors committed on the tail data. In this way we managed to replicate better the behaviour of real data but we decided not to further explore this procedure for a series of criticalities that affected it. Indeed, first of all the computation load was too heavy and computation time too long, moreover the choice of the weight was to be evaluated case by case as there is no "optimal weight" that always works (see Figure 3.6). These two reasons made this method absolutely inapplicable in our case study, since our goal is to analyse vast areas and we can't afford to apply this procedure on such big zones like Lombardy (which is composed by 3700 sites for example).



(a) Weibull Case

(b) Lognormal Case

Figure 3.6: The two graphs shows the results obtained with same weight in two different sites that are approximate by two different distributions. This highligths how different the degrees of accuracy are just based on the starting distribution and how impractical this method would be on a large scale.

4 | Hazard Analysis

The main result to be taken home from the previous chapter is that modelling the annual maxima by means of the GEV method is the best thing to do in this kind of studies: indeed, the F^n method and data simulation approach both suffer the lack of data in the right tail when trying to fit the best parent distribution. With this newfound security, in this chapter we will present some practical results regarding Extreme Value Analysis.

At first we will have a look at some interesting features revealed by the distribution of maximal values at various scales. Then, the actual hazard analysis will be carried on; the aim is to understand which areas are more likely to be subject to extreme phenomena and, to this end, exceedance probabilities curves at different years and return times will be studied. Notice that there is a difference between "hazard" and "risk" in the sense that the former indicates something with potential to harm people or structures, like extreme winds, while the latter is the likelihood of a hazard causing harm and, thus, needs more data to be investigated, like vulnerability of infrastructures, possibility of people presence and, in general, data on anything that can be damaged. In this work we are only interested in the hazard and we will not delve into technical risk analysis.

To conclude the chapter, some grouping attempts will be shown: the goal is to cluster together those area subject to same risk of extreme events and give a summarizing and practical feedback for hazard quantification and safety design.

4.1. Extreme Wind Speeds

Once the proper GEV distribution has been recovered from the 31 annual maxima at each site, it is interesting to observe the behaviour of some of its parameters. We remember now that we are still working with just the magnitude of the wind vector and that the optimal parameters for the distributions are computed as described in section 2.4 and, in particular, using the probability-weighted moments method as described in subsection 2.2.2.

The first value of interest is the location parameter μ of the GEV. This parameter represents, to some degree, the "middle value" of the distribution in the sense that the mean, the median and the mode are all just small variations of it. Indeed we have:

$$Mean = \begin{cases} \mu + \sigma\dfrac{g_1 - 1}{\xi} & \text{if } \xi \neq 0, \xi < 1 \\ \mu + \sigma\gamma & \text{if } \xi = 0 \\ \infty & \text{if } \xi \geq 1 \end{cases}$$

$$Median = \begin{cases} \mu + \sigma\dfrac{(ln2)^{-\xi} - 1}{\xi} & \text{if } \xi \neq 0 \\ \mu - \sigma ln(ln2) & \text{if } \xi = 0 \end{cases}$$

and

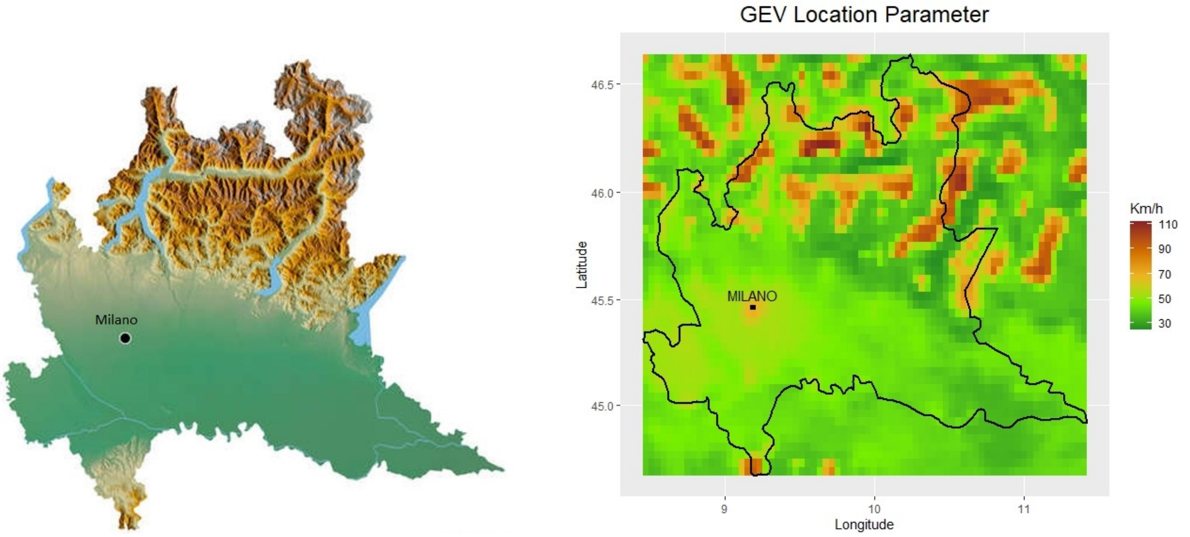
$$Mode = \begin{cases} \mu + \sigma\dfrac{(1+\xi)^{-\xi} - 1}{\xi} & \text{if } \xi \neq 0 \\ \mu & \text{if } \xi = 0 \end{cases}$$

where μ is the location, σ is the scale, ξ is the shape parameter, $g_k = \Gamma(1 - k\xi)$ and γ is the Euler's constant.

Figure 4.1 and 4.2 show a comparison between a map of the elevation and one reporting the location parameter μ of each cell, one at regional level and the other at national level. As intuitively expected, the similarity is striking: mountainous areas are more windy in general and they register higher extreme winds due to the higher differences in temperature (and thus, pressure) in a shorter distance while plains areas have little to no variations and wind speeds tend to be more uniform and lower.

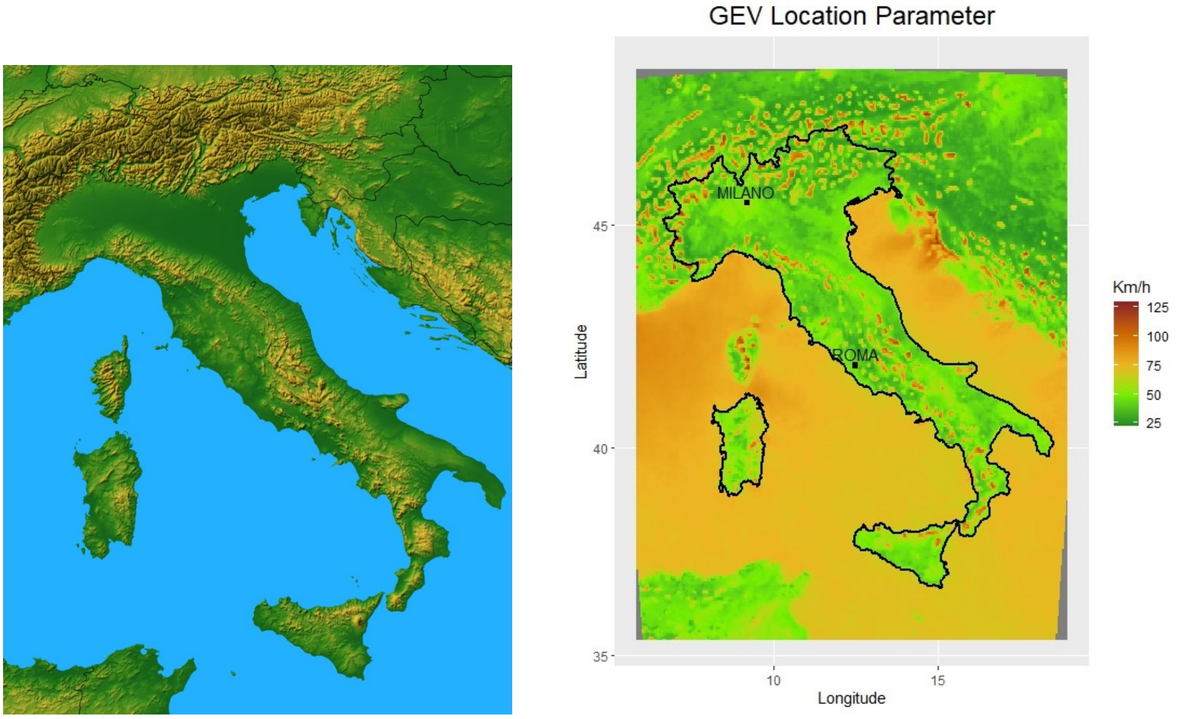
Something odd, however, is found in Figure 4.1, in the area corresponding to the metropolitan city of Milan. The city rises in the middle of the Po Valley but the extreme wind speeds recorded in this location are much higher than those measured in the surroundings. The causes of this peculiar phenomenon are the geographical position of Milan and the temperature in this area: although Milan rises in a completely plain area, it is located south of the Alps, not that many kilometers away from them and it is interested by the Foehn, a typical wind that blows on the Alpine arc, from north to south, descending from the mountains and reaching also the plain, Milan included.

As highlighted by the italian meteorologist and president of the italian meteorological society, Luca Mercalli, this kind of phenomenon, as unexpected as it may seem, is actually ordinary. Indeed, in an interview for the newspaper "Corriere della Sera" [17], few days after an extreme wind event had unroofed the Central train station of Milan,



(a) Lombardy orography (b) Location parameters of annual maxima

Figure 4.1: Comparison between orography and location parameters of the GEV distribution fitted on annual maxima in each site of Lombardy.



(a) Italy orography (b) Location parameters of annual maxima

Figure 4.2: Comparison between orography and location parameters of the GEV distribution fitted on annual maxima in each site of Italy.

Mercalli pointed out that this is quite common, especially during the winter season, and may happen even more than once each year. The reason why the Foehn reaching Milan can have such a disrupting behaviour is the climatic situation that can be found in the metropolitan city. The heavy urbanization of the Lombardy regional capital causes the creation of a microclimate in the city characterized by temperatures that are considerably higher than the surroundings. This fact leads to a sudden change in pressure, analogously to what happens on the mountains, and, consequently, makes the extreme winds stronger.

After the location, also the shape parameter ξ, which determines the type of the distribution taken by the GEV, represents a value of interest. As cited for instance in Holmes et al. (1991) [18], the Type 2 GEV distribution predicts unlimited values as the return time increases; this is obviously unfeasible for a physical phenomenon, such as wind, whose speed must be limited from above and the cases where the fitting results in this type of distribution should be considered outliers. Such anomalous results may come from faults in the original data or possibly if the sample contains mixed populations, for instance two very different storm types causing extreme wind speeds. In any case, in our study, very little sites showed this behaviour, confirming the idea of them being outliers with respect to the expected distributions (see Figure 4.3).

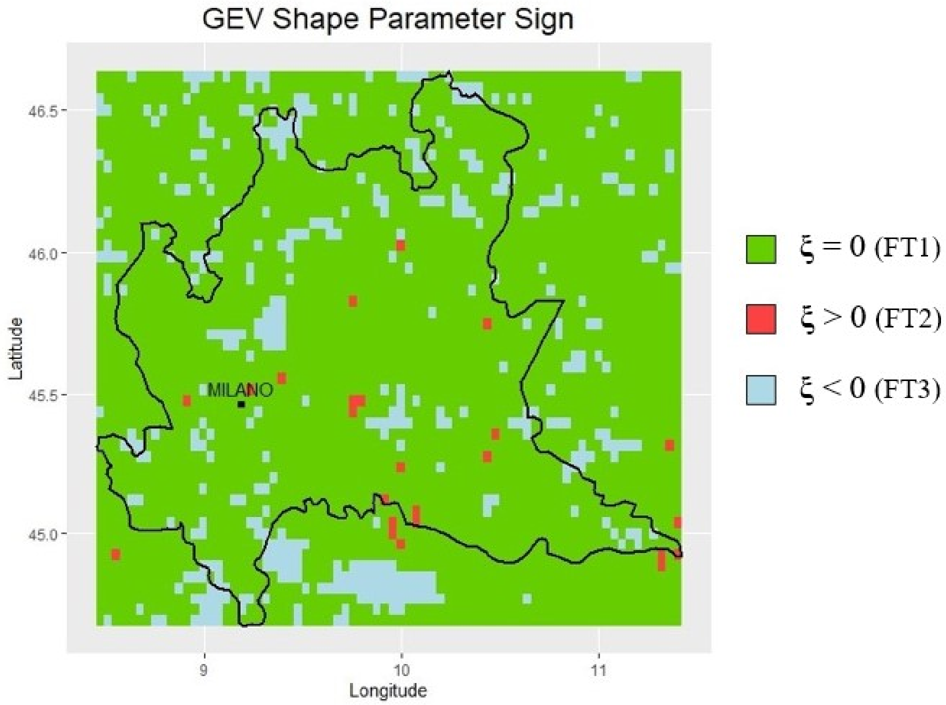


Figure 4.3: As we can see, in the absolute majority of the sites the shape parameter is $\xi = 0$, with some cells having a negative shape and just 24 of them having a positive shape.

4.2. Hazard Assessment

4.2.1. Exceedance Probability Curve

The natural conclusion of this first part of our study starts with the production of the so called Exceedance Probability Curves, which summarise the information related to the hazard. Indeed, these curves describe, for each possible value of the wind speed, the probability of exceeding that threshold in a year. As already mentioned in the introduction of Chapter 3, they are easily obtained once the cumulative distribution function for the annual maxima is available, since they are simply computed as 1−CDF and thus, using the GEV method, there is no difficulty related to this part once the optimal parameters are estimated. In Figure 4.4 we can find an example of exceedance probability curves: once again we wanted to highlight the behaviour of Milan and to do so we compared its curve of annual risk with the one computed in a location just few kilometers away from the centre of the regional capital.

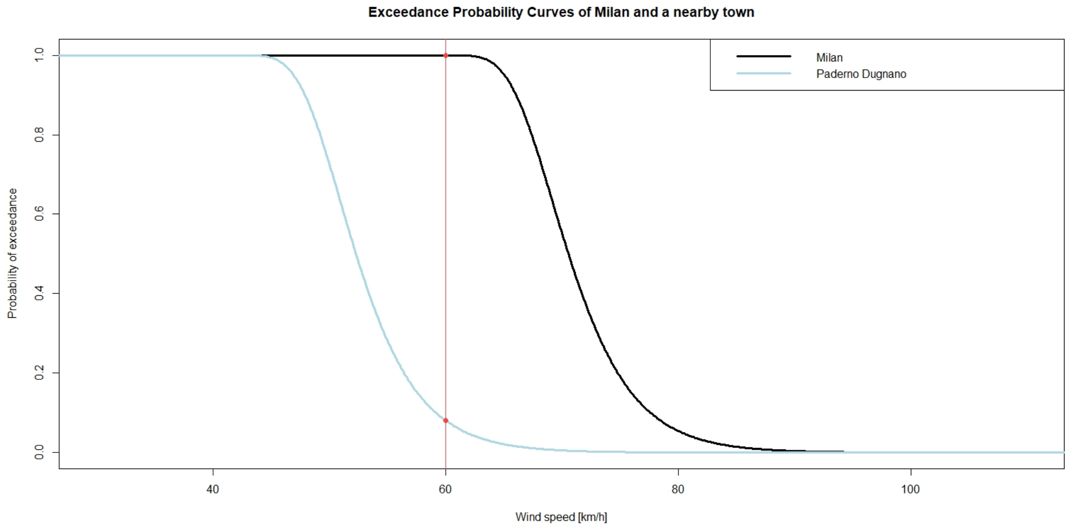


Figure 4.4: Paderno Dugnano is a town about 10 kilometers away from the centre of Milan. As we can see, their exceedance probability curves are very different in terms of hazard; indeed we highlighted the threshold of 60 km/h which is almost surely surpassed in the case of Milan while it is quite improbable for Paderno.

Notice that, in this section of our analysis, among all the others, we will focus our attention on the observation of two particular wind speeds that are 60 km/h and 140 km/h; this because, as pointed out by Terna (2021) [45], those wind speeds are commonly studied when analysing the electrical grid resilience. The reason is that above 60 km/h the electrical grid starts to be affected by those that are called "failures caused by indirect effects", while above 140 km/h also the "failures caused by direct effects" come into play. The first case comprehends all those infrastructure breakdowns due to the action of the wind on other objects, trees in particular, while the second case includes all the failures caused by the action of the wind directly on the electrical power grid components.

With this in mind we proceeded to analyse these two thresholds on a larger area and, as always done so far, we observed what are the areas at risk in Lombardy. In Figure 4.5 we can observe what are the annual probabilities of exceeding the two thresholds of 60 and 140 km/h for each site in Lombardy. While the first graph highlights the areas exposed to the risk of indirect effects failures, the second one shows that the probability of extreme winds over 140 km/h is almost zero everywhere.

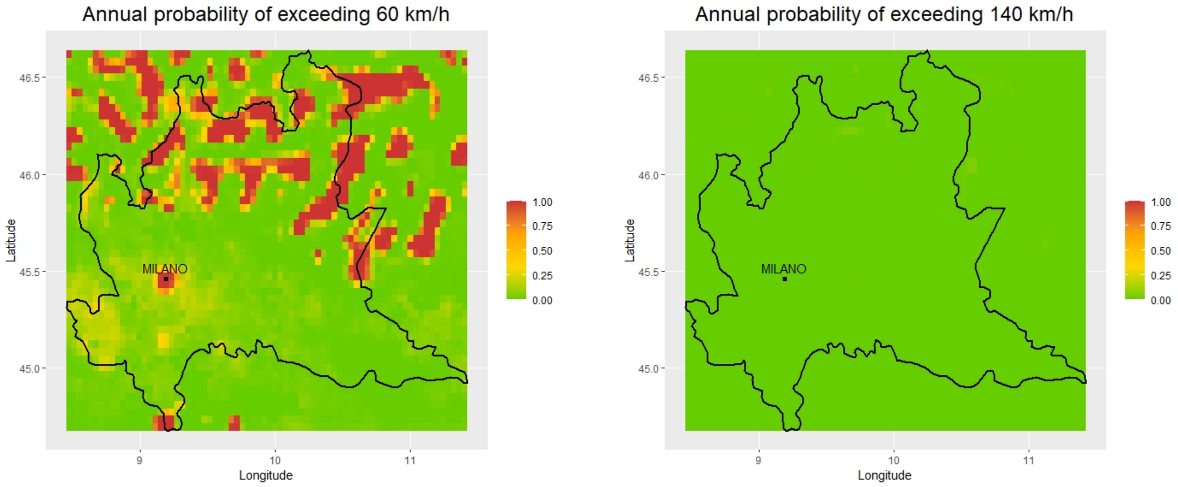

Annual probability of exceeding 60 km/h
Annual probability of exceeding 140 km/h
Latitude
46.5
46.0
45.5
45.0
Longitude
9
10
11
MILANO
1.00
0.75
0.50
0.25
0.00

Figure 4.5: Annual probabilities of exceeding 60 and 140 km/h in Lombardy. The risk of experiencing failures due to indirect effects is very high in the mountainous part of the region while exceeding 140 km/h threshold has nearly null probability everywhere.

4.2.2. Mean Return Time

Another possible approach to quantify the risk of experiencing an extreme wind event consists in estimating how much time should pass (on average) before a specific threshold is exceeded. This notation is known as "Mean Return Time" and is computed as

$$T(v) = \frac{1}{1 - CDF(v)} \tag{4.1}$$

where v is the wind speed relative to which we want to compute the return time.

As highlighted by formula 4.1, the two concepts of Exceedance Probability Curve and Mean Return Time are strictly related, being one the inverse of the other, but they provide two different answers to the same question: the first one estimates how probable it is that, in the time period of one year, there will be a wind event exceeding a certain threshold, while the latter says how much time we expect to have to wait before a certain speed will be sampled. Figure 4.6 shows an example of the mean return time computed on Milan while Table 4.1 shows the values of Exceedance Probability and Mean Return Time computed on some values for Milan.

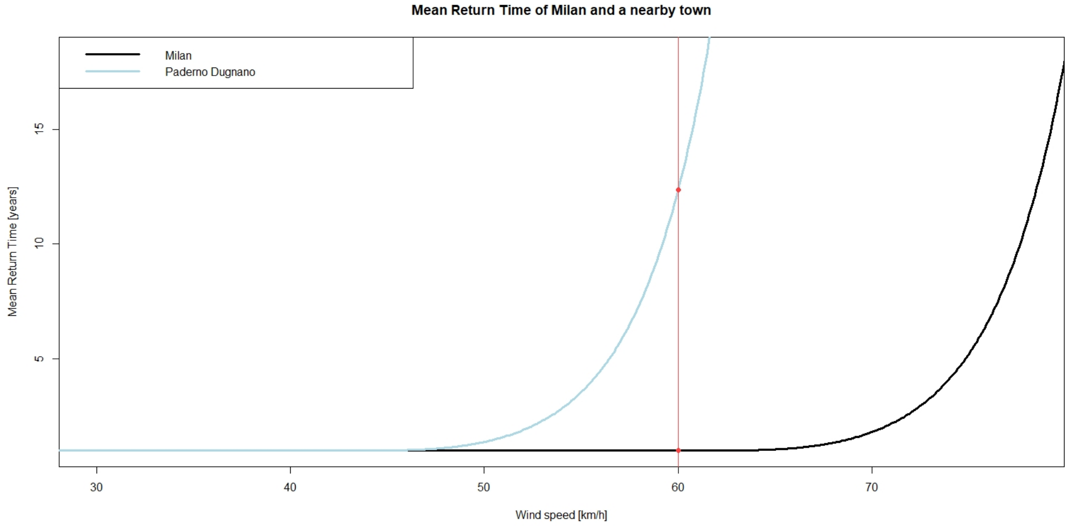
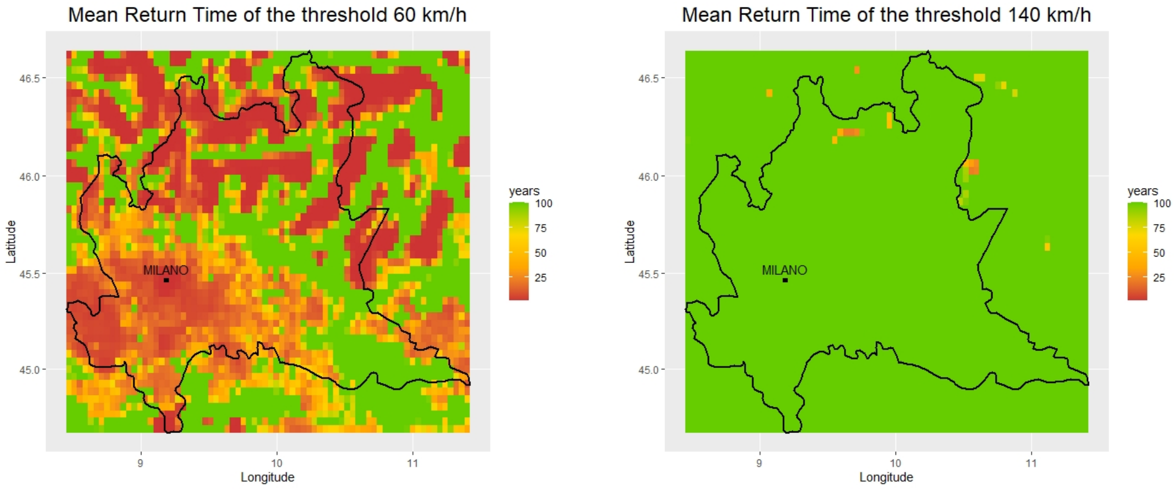


Figure 4.6: Once again we compare Milan and Paderno Dugnano, observing that, for example, if we can expect to have a wind of at least 60 km/h every year in Milan, instead we can measure a wind of the same speed in Paderno once every 12 years.

Speed threshold	60 km/h	70 km/h	80 km/h	90 km/h	100 km/h
Exceedance Probability	1	0.55	0.055	0.004	0.0003
Mean Return Time [years]	1	1.8	18.2	250	3333

Table 4.1: Numerical Examples for the risk in Milan

Analogously to what done before for Exceedance Probabilities, we can picture the quantitative trend of the Mean Return Time on the area of Lombardy (Figure 4.7). Notice that in both the images of Figure 4.7 all sites with a return time higher or equal to 100 years are represented with the same colour. This choice was taken for graphical reasons but also because there is little interest in discriminating cases with return time higher than 100 years. Moreover, as investigated in Ben Alaya et al. (2021) [5], the cases where the return time is particularly high require additional attention.


Mean Return Time of the threshold 60 km/h

Mean Return Time of the threshold 140 km/h

Figure 4.7: Return times of the two thresholds of interest on all the territory of Lombardy. As we can appreciate, many areas have a short return time for what concern 60 km/h while on the side of 140 km/h, according to what already seen, almost everywhere we have very high return time.

4.2.3. Exceedance Probability Curve on Longer Time Window

After having observed these two approaches, we will now discuss exceedance probabilities for a longer time horizon. Indeed, since we are working under the hypothesis that the distribution of annual maxima is stationary (i.e. does not change throughout the years) and that different years are independent, once we have the annual exceedance probability curves we can trivially compute these curves for any time horizon T of interest just by taking $1-\text{CDF}^T$. In this way we can obtain accurate estimates of the probability of exceeding whichever threshold we are interested in, for any time window within the span of a century; in this way we answer the question: "what is the probability that, in the period corresponding to the chosen number of years, we exceed at least once the threshold?".

Once again, let us have a closer look to Milan: in Figure 4.8 we can appreciate how the exceedance probability curve of Milan evolves by increasing the time window considered.

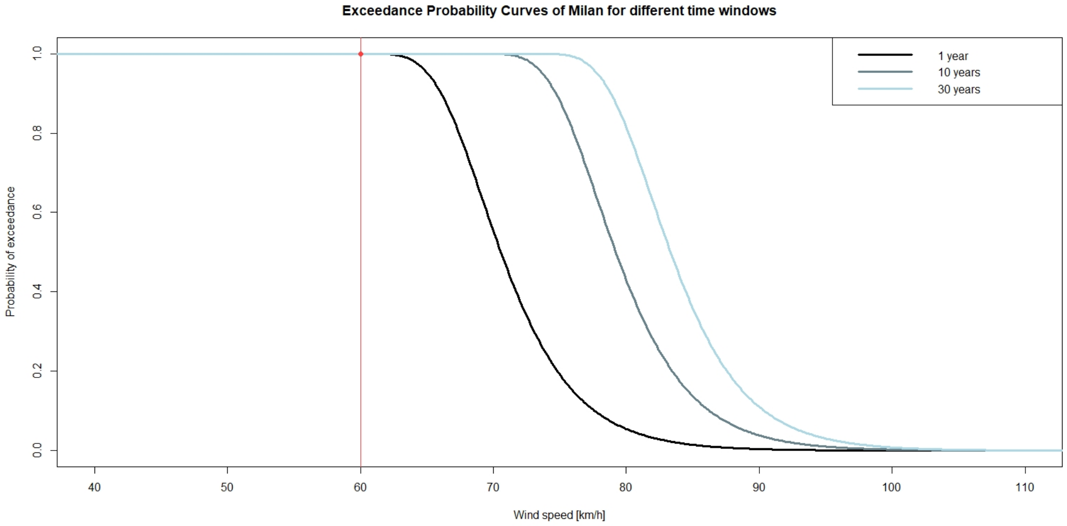


Figure 4.8: We represent the exceedance probability curves computed for Milan for the 1 year, 10 years and 30 years time window cases. As we could expect, by increasing the dimension of the time window considered, the probability of exceeding a given threshold increases.

To conclude, we computed the probability of exceeding the usual 60 and 140 km/h threshold on the entire region of Lombardy when we increase the time window to 10 and 30 years. For completeness we once again report also the case of a 1 year time window (Figure 4.9). The evolution of the probability graph in the case of 60 km/h is evident: the longer the time window, the more numerous are the zones where the risk of surpassing this threshold becomes very high. On the other side, the cases computed for 140 km/h look more stable and conservative. However, although the changes in this case seem less apparent, we can find few areas where the probability of exceeding this crucial threshold of hazard become less and less negligible.

4.3. Grouping by Hazard Class

After having observed where the probability of extreme events is high and where it is moderate through the exploitation of Exceedance Probability Curves and of Mean Return Time, we decided to conclude the hazard analysis by summarizing all these information collected for each site and by grouping together zones with the same level of hazard.

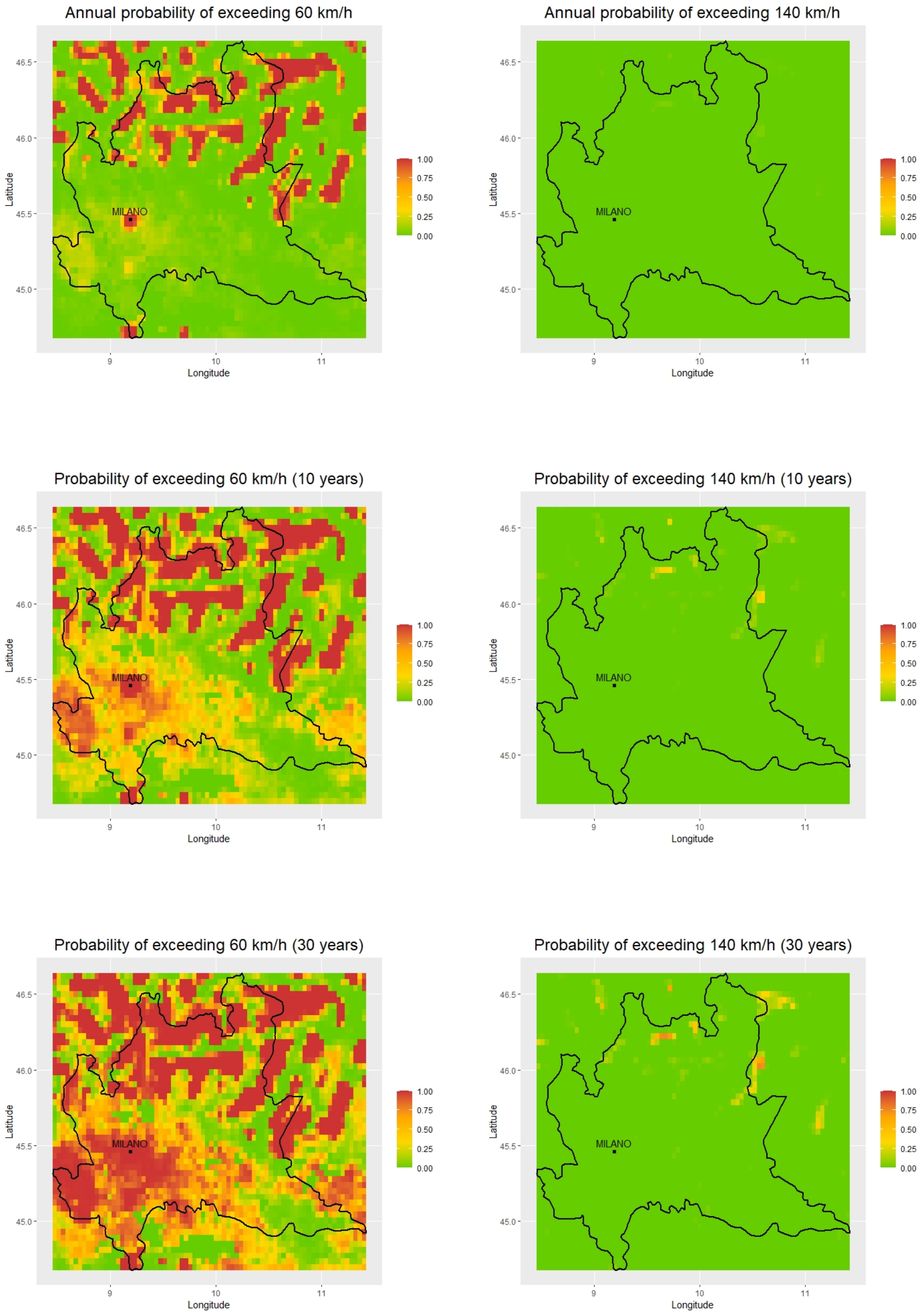
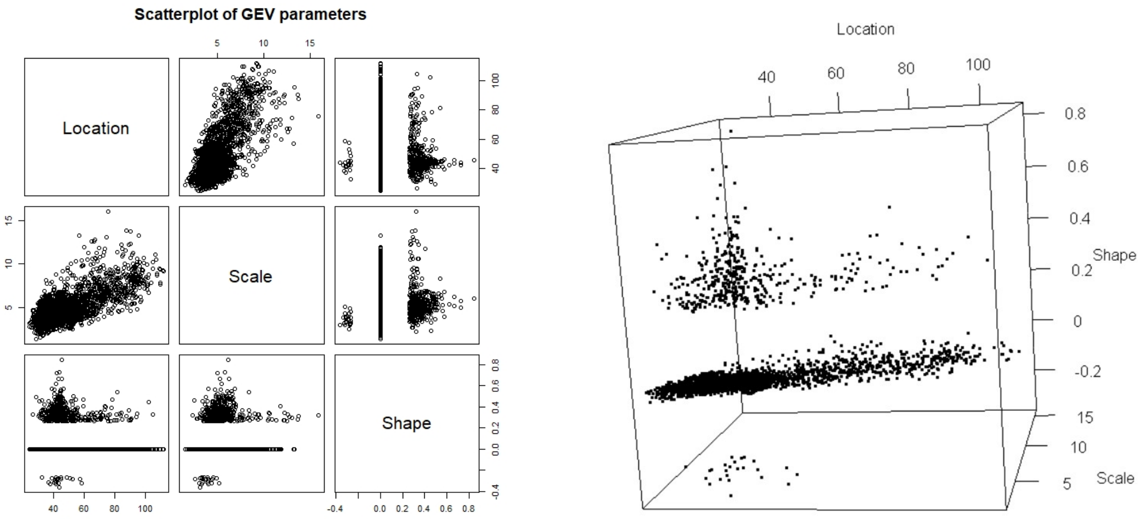

Annual probability of exceeding 60 km/h
Annual probability of exceeding 140 km/h
Probability of exceeding 60 km/h (10 years)
Probability of exceeding 140 km/h (10 years)
Probability of exceeding 60 km/h (30 years)
Probability of exceeding 140 km/h (30 years)
MILANO
Latitude
46.5
46.0
45.5
45.0
Longitude
9
10
11
1.00
0.75
0.50
0.25
0.00

Figure 4.9: On the left side the cases for 60 km/h, on the right side, 140 km/h

4.3.1. Preliminary decisions

The first attempt that we put into practice, following our initial intuition, was to perform some sort of clustering induced by the three parameters of the GEV fitted in each location. However, as highlighted by Figure 4.10, utilizing these 3 indices to cluster may not be the smartest choice.



(a) Scatterplot

(b) 3D plot

Figure 4.10: Parameters of the GEV fitted in each site of Lombardy.

First of all, just like the phenomenon of wind itself, also the parameters show a continuous behaviour (except for the shape parameter that is "conditionally continuous"). Because of that, there is no way that we can find any clear and reasonable division in data that may induct the definition of clusters. Our only opportunity to do such a thing is follow the path traced by the discontinuity in the values of the shape parameter produced by `zTestGevdShape`; however clustering data in terms of their shape parameter would produce results that are not only trivial, but also useless under the light of risk assessment.

For these reasons, we decided to change perspective and so we abandoned the idea of exploiting automatic clustering algorithms (such as hierarchical ones for instance) and moved to a procedure in which we were responsible for the clustering produced, by setting manually the division criterion. One could argue that this way to proceed and the grouping obtained are totally arbitrary and that's actually true, but at this stage we are not looking yet for a "behavioural" clustering of data but we want to group our sites with others that have the same level of hazard: considering that the literature has already identified some thresholds to classify the family of hazard, we deemed that this procedure could actually work properly for our interests.

4.3.2. Results

Diving into practice, we decided to work with the quantiles of our maxima distributions, and to group each site based on which interval of wind speed its quantile falls in. We remind that the quantile of order α is that value of the dominion of the distribution for which the cumulative distribution function evaluated in that point is equal to α, that is:

$$\mathbb{P}(X < q_\alpha) = \alpha$$

where X is a random variable with a certain distribution and q_α is the quantile of order α.

The reason why we decided to deal with quantiles for the grouping is that, in a sense, the quantile itself expresses a "confidence margin". Indeed, by considering the α-quantile of a specific site we are already taking the value of wind speed that has a probability equal to alpha not to be exceeded. Consider, for example, the distribution of the maxima fitted for Milan: the quantile of order 95% for Milan is approximately equal to 80 km/h. This means that, each year, there is a probability of 95% that there will not be any wind of speed over 80 km/h in Milan. Recalling the concepts of Section 4.2, we have an annual probability of exceeding 80 km/h equal to 5% and a Mean Return Time equal to 20 years. In our analysis we evaluated both the 95% and the 99% quantiles of maxima distribution. In Figure 4.11 we can see the 95-quantiles and 99-quantiles of each site in Lombardy.

From this point on we will focus our attention only on the 99-quantile, as it provides a larger margin of confidence: indeed, the 99-quantile means looking at that "once in a century" value. In this way we will perform a grouping based on the value associated to the 1% annual risk. Having decided to exploit the 99-quantiles, we simply need to decide how to subdivide the range of wind speeds. In this work we propose two examples of clustering: a finer grouping and a more coarse one. The first clustering defines 6 intervals, less than 60 km/h, between 60 and 80 km/h, between 80 and 100 km/h, 100-120 km/h, 120-140 km/h and more than 140 km/h, while the second one focuses more on the already defined thresholds of hazard, describing 3 clusters, one for values smaller than 60 km/h (and thus subject to nearly no danger at all), the middle group between 60 and 140 km/h (which means those areas subject to indirect effects), and finally the group subject also to direct effects with quantiles over 140 km/h. In Figure 4.12 we show the two versions of the grouping produced.

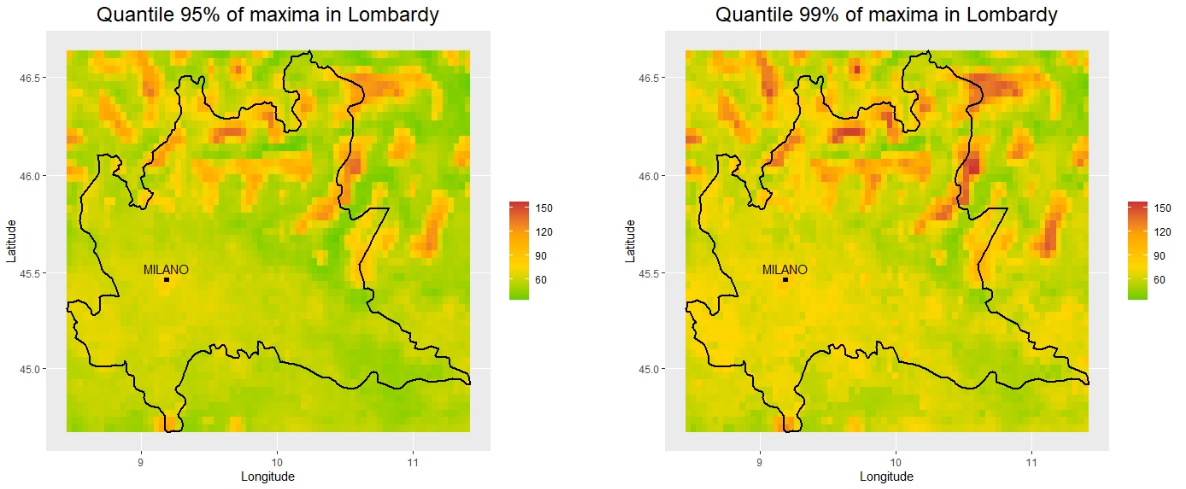
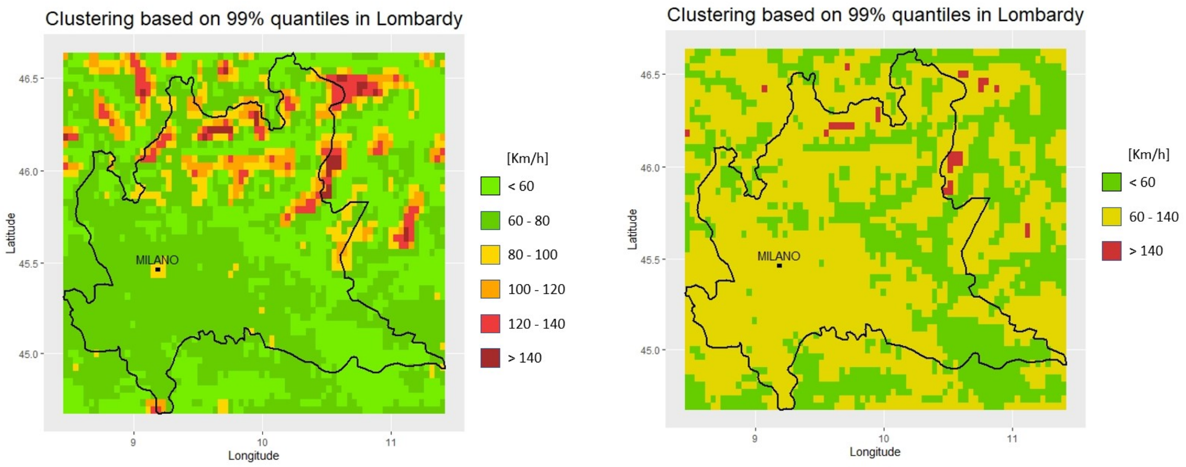

Quantile 95% of maxima in Lombardy
Quantile 99% of maxima in Lombardy
46.5
46.0
45.5
45.0
Latitude
MILANO
9
10
11
Longitude
150
120
90
60

Figure 4.11: The quantiles of maxima of each site over Lombardy repeat the same trends we already found in previous graphs.



(a) 6 clusters

(b) 3 clusters

Figure 4.12: We can notice that in Lombardy the majority of the areas are subject to very low risk or events causing indirect breakdowns.

At this point of the work we managed to produce a useful result in the setting of hazard assessment. Thanks to this grouping, we have a clear visual tool to immediately identify which are the more hazardous areas.

4.3.3. Group distribution

The last question we wanted to answer is if it was possible, once we grouped data, to model each of the clusters with a single GEV distribution. What we can notice from Figure 4.13 is that, if we consider data within each of the 6 clusters previously defined as a whole, they seem to describe very well a single GEV.

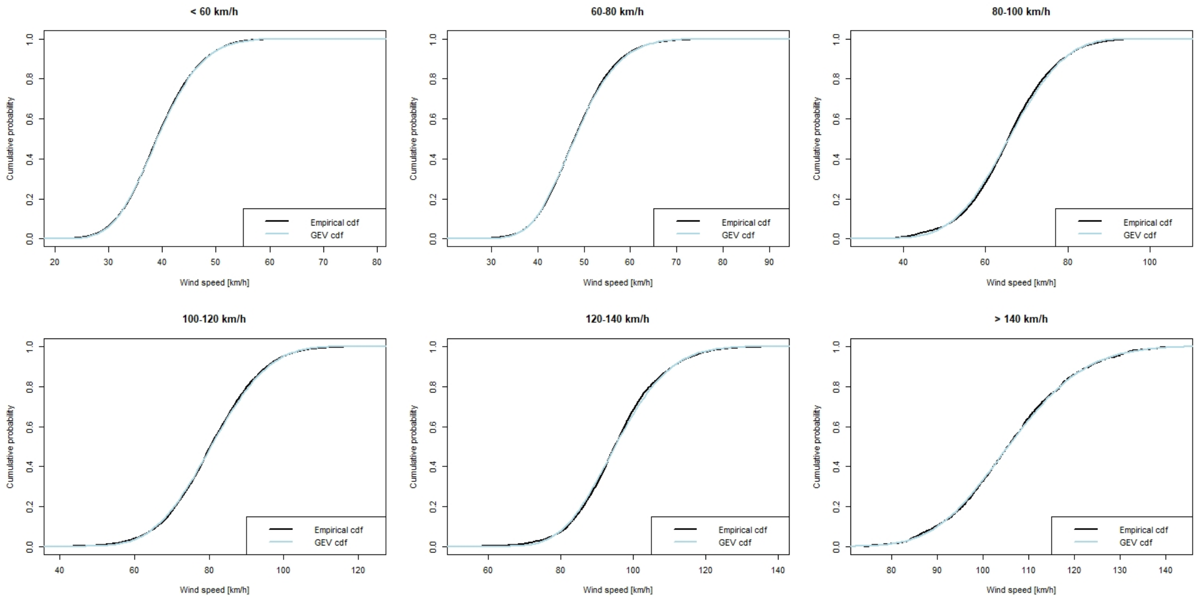


Figure 4.13: Comparison between empirical cdf and GEV cdf fitted on all data of each cluster.

Recalling the criticality of the lack of data highlighted in Section 3.2, this may look as a possible solution: we grouped together data with the same characteristics and that, for this reason, may be considered as coming from the same distribution and thus increase the number of data at our disposal to provide a more accurate estimate of the GEV parameters. However, if we look closely to data within the same cluster we can notice that, in there, are grouped curves characterized by very different behaviours (Figure 4.14).

Thinking of summarizing the characterization of a whole function in a single value is a bit of an optimistic task. For this reason, we gave up the idea of and summing up the behaviour of all sites in a group into a unique pdf, as it would have resulted in substantial underestimates or overestimates of risk in many sites.

To support this decision, we also employed a non-parametric technique, performing a permutation test. The idea at the base of this procedure is that you want to understand how "special" a particular configuration of samples is with respect to all the other possible

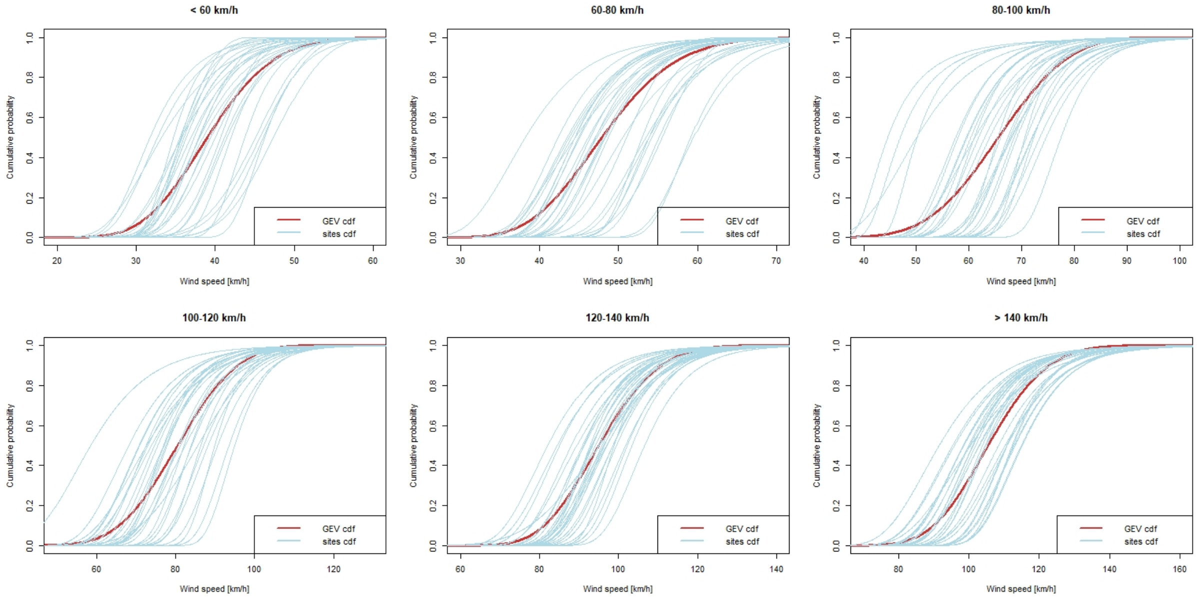


Figure 4.14: Comparison between GEV cdf fitted on all data of each cluster and some of GEV cdf fitted in sites of the cluster.

permutations of the data. The first idea for the Permutation tests dates back to Fisher (1936) [15]; we will report from this article an example made by Fisher that clearly explains the logic of Permutation Tests:

"Let us suppose, for example, that we have measurements of the stature of a hundred Englishmen and a hundred Frenchmen. It may be that the first group are, on the average, an inch taller than the second, although the two sets of heights will overlap widely. [...] The simplest way of understanding quite rigorously, yet without mathematics, what the calculations of the test of significance amount to, is to consider what would happen if our two hundred actual measurements were written on cards, shuffled without regard to nationality, and divided at random into two new groups of a hundred each. This division could be done in an enormous number of ways, but though the number is enormous it is a finite and a calculable number. We may suppose that for each of these ways the difference between the two average statures is calculated. Sometimes it will be less than an inch, sometimes greater. If it is very seldom greater than an inch, in only one hundredth, for example, of the ways in which the sub-division can possibly be made, the statistician will have been right in saying that the samples differed significantly. For if, in fact, the two populations were homogeneous, there would be nothing to distinguish the particular subdivision in which the Frenchmen are separated from the Englishmen from among the aggregate of the other possible separations which might have been made. Actually, the statistician does not carry out this very simple and very tedious process, but his conclusions

have no justification beyond the fact that they agree with those which could have been arrived at by this elementary method."

Coming to our specific case study, we tried to apply this idea to the data within a cluster: we considered each site $X_1, X_2, ..., X_n$ within a group (of numerosity n) as a population composed by 31 elements (the 31 maxima) and computed the average value $m_i = \mathbb{E}[X_i]$ for each site, following the idea described in the example of Fisher. Our aim is to verify whether or not the hypothesis H_0: $X_1 \overset{d}{=} X_2 \overset{d}{=} ... \overset{d}{=} X_n$ is true or not.

However, at this point, we could not proceed as in the example because our task is not to compare just two populations but we have hundreds of them within a cluster. For this reason, instead of using as test statistic T_0 the difference between the mean of the populations, we used the variance of the mean values computed $T_0 = Var(\vec{m})$. The reason behind this choice is that the variance gives a sufficiently good idea of the dispersion of our data, indeed, if the variance results to be small, it means that all the means are quite similar in value, while if the variance starts to raise then it means that the values are more different from each other.

Having our T_0, we are left with the iterative part: it's enough to shuffle all the maxima together and create new groups of 31 elements, compute the new means and, then, the new variance. After doing this procedure for hundreds, or even thousands, of iterations we want to count how many times we obtained a value of variance greater than T_0. This value divided by the total number of iterations gives an estimate of the p-value for the test. Table 4.2 highlights that in all 6 clusters the T_0 is always considerably higher than the variance obtained during the iterative steps and, thus, the configuration is a very particular case and, thus, we have to reject the hypothesis of homogeneity of the distributions.

Cluster [km/h]	**< 60**	**60 - 80**	**80 - 100**	**100 - 120**	**120 - 140**	**> 140**
T_0	19.30	23.51	55.63	62.67	38.05	42.68
mean(T_1)	1.34	1.77	3.40	4.35	4.37	5.22
p-value	0	0	0	0	0	0

Table 4.2: Permutation test results

4.4. Final Considerations

In this chapter and in the previous one we confronted some methods to estimate extreme values of wind speeds, coming to the conclusion that approximating their distribution by means of a GEV produces the best results in general. Different methods can be more accurate in specific cases but, neither here nor in the literature, considerable improvements have been found. Then, once these extreme wind distributions are recovered, it is easy to produce exceedance probability curves for any value of interest and with any (reasonable) time span. In particular, we produced estimates for the most common thresholds at different time horizons and compared results on the region of interest to determine hazard level on the area.

In conclusion, we can sum up by saying that Lombardy is, for sure, a safe region from the point of view of extreme winds causing direct infrastructure failures, but, on the other side, is quite subject to the threat of indirect failures of the electrical grid. If we exclude the areas around Milan, which we found out to be a peculiar exception characterized by a high level of hazard with respect to the surroundings, the locations more interested by the risk of extreme winds are those on the mountains, while plain areas, where most of the infrastructure is found, are less interested by extreme phenomena.

However, if we consider that, on the mountains, the power grid is also less dense and connected, we realize that a single fault in a small portion of the grid can mean cutting off connection in a large geographical area and leaving many people without electricity, even for a long time. For this reasons, it is of paramount importance to carefully analyze the wind regime and the related threats when designing the infrastructure or when planning maintenance on it, so that interventions carry out result adequate and effective to assure safety and reliability of the service.

5 | Meteorological Correlations

This chapter and the next one will work as a bridge between the first part, where wind has been considered a hazard, and the second one, where it will be evaluated as a possible resource. In particular, now we are interested in the individuation of meteorological correlations, i.e. the determination of areas that can be considered to be subject to the same meteo events and, in particular for our case, to the same wind regimes.

This can be interesting from a climatological point of view per se, as it provides a natural subdivision of the territory according to the dominant wind regimes; but it is also important regarding infrastructural safety, for instance for an electrical line, in the sense that it allows to know which sections of the grid can be handled and maintained in the same way.

Here, we are not interested anymore in just the extreme values to determine the hazard related to a particular site but we want to give a more general characterization to the wind speed series to understand if the same winds are blowing over different zones. This means that different hazards can be associated to two sites even if we consider them to be subject to the same winds: indeed, hazard is related just to the absolute intensity of the wind speed, while we can consider two locations to be under the same wind regimes even if this intensity is slightly different. Still, as we said, it can be important to know for instance the direction of the principal winds in an area and, in general, their main characteristics to provide appropriate countermeasures regarding safety design.

To better suit this analysis, from this point on, we will change the way we look at our data. First of all, we now need to consider both the longitudinal and the latitudinal components of the wind (see the dataset description in Chapter 1), since also the direction is clearly a relevant factor here. Then, we can't limit ourselves to a subset of values but we need the whole time series; to lighten the computational weight we only considered the series corresponding to one year (2020, the most recent at our disposal) since we can safely assume that it is a sufficient span of time for this study and encapsulates also possible seasonal variability.

For these reasons, the approach taken in this chapter and in the next one will be a Functional Data Analysis (FDA) driven one. This chapter, in particular, will be dedicated to a throughout study on FDA, where theoretical notions will be alternated with applications to the case in exam; it will cover topics like smoothing and Functional Principal Component Analysis (FPCA) and will culminate in the definition of a distance between wind time series. The next chapter will instead present two clustering algorithms, developed with the aim of incorporating geographical information to account also for spacial dependence.

5.1. Introduction on FDA

The discussion carried on in this chapter is mainly based on two books by Ramsay: "Functional Data Analysis" (2005) [25] and "Functional Data Analysis with R and MATLAB" (2009) [26] with applications referring directly to the case in exam; but, first of all, a small introduction on Functional Data Analysis as a whole is due.

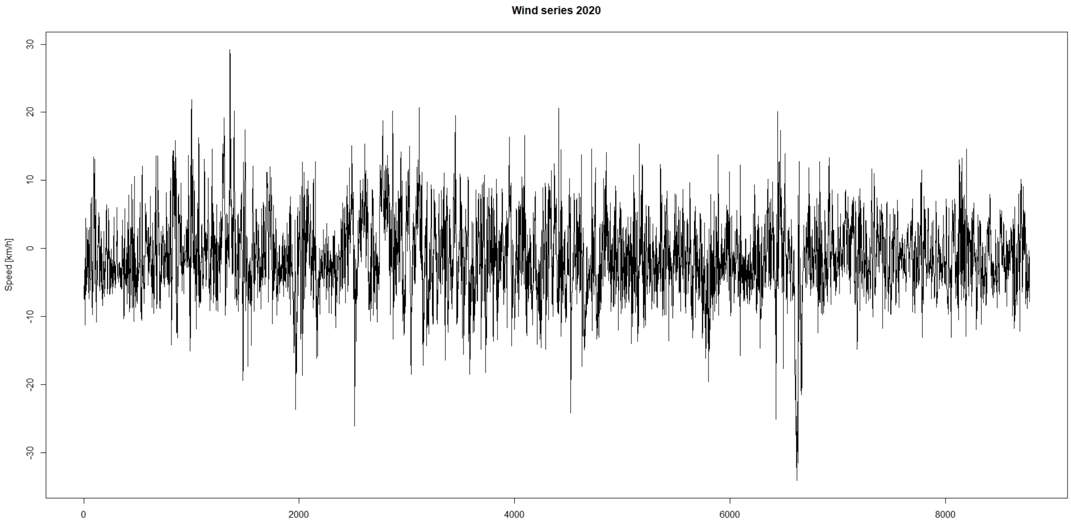


Figure 5.1: An example of temporal series for the U component of wind in 2020.

Let us have a look for instance at Figure 5.1, where the temporal series of longitudinal winds for a particular site in the year 2020 is reported. While observations in our possession are actually discrete, it is clear that they reflect a smooth variation of the phenomenon and that there exists an underlying function describing the evolution in time of the variable of interest. Recovering an approximation of this function allows us to represent data in a simpler way, displaying their fundamental features and emphasizing patterns and variations and this is where the power of FDA comes in handy.

Before starting with the actual analysis we need to understand the context where functional data live. A common choice for the space structure of functional data is the one of separable Hilbert spaces; even if a discussion on such topics is beyond the scopes of the present work (and more info can be found for example in "Functional Analysis, Sobolev Spaces and Partial Differential Equations" by H. Brezis (2010) [10]), we will remind the basic concepts for convenience.

Definition 1: A (real) Hilbert space H is a vector space equipped with an inner product that defines a distance function for which the space is a complete metric space.

Definition 2: Let H be a linear space; an inner product on H is a bilinear, symmetric and positive definite form $\langle \cdot \ , \cdot \rangle \ : H \times H \to \mathbb{R}$.

Definition 3: An Hilbert space (and also a general metric space) is complete if it contains all the limit points of its Cauchy sequences.

Definition 4: An Hilbert space (and also a general metric space) is separable if it contains a dense countable subset.

At this point we are ready to give a proper definition of functional data. Let H be an Hilbert space whose points are functions defined on a closed interval $T = [t_{min}, t_{max}]$; then:

Definition 5: A functional random variable is a random element defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in H: $X : \Omega \to H$.

Definition 6: A functional datum x is a realization of a functional random variable: $x = X(\omega) : T \to \mathbb{R}, \ \forall \ \omega \in \Omega$.

The most used space for FDA is the space L^2 of square-integrable real functions. In particular, the latter will be our structure of choice since it is separable and perfectly adequate for unconstrained data analysis. To complete the overview, two more definitions can be useful: let $X : \Omega \to H$ be a functional random variable in H.

Definition 7: We call Fréchet mean of X the unique element μ of H that solves $arg \inf_{x \in H} \mathbb{E}[\parallel X - x \parallel_H^2]$.

Definition 8: We call covariance operator of X the operator $C : H \to H$ defined as $Cx = \mathbb{E}[\langle X, x \rangle X], \ x \in H$.

In L^2 the Fréchet mean coincides a.e. with the point-wise mean and can be estimated with the sample estimator: $\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i$; moreover, the covariance operator can be defined through a kernel operator $[Cx](t) = \int_T c(s,t)x(s)ds, \ x \in L^2$, where the covariance kernel

$c(s,t)$ is precisely the point-wise covariance $c(s,t) = \mathbb{E}[X(s)X(t)]$ and can be estimated with the sample covariance operator $\hat{c}(s,t) = \frac{1}{N}\sum\limits_{i=1}^{N} X_i(s)X_i(t)$.

5.2. Smoothing

The first, fundamental, step in any analysis involving functional data consists in turning raw discrete data into smooth functions. Indeed, the temporal series of discrete data points often consist in thousands of elements (8784 in our case, 24 hourly measurements for each of the 366 days of 2020) and we need to find a representation that allows both to reduce the computational load and to work with the familiar tools of matrix algebra. Fortunately, we have exactly what we need: basis functions.

A basis function is a set of known functions ϕ_k that are independent of each other and have the very convenient property that we can approximate arbitrarily well any function by taking a linear combination of them. The two most famous basis functions systems are the monomials: $1,\ t,\ t^2, \dots,\ t^k, \dots$ and the Fourier series:

$$1,\ sin(\omega t),\ cos(\omega t),\ sin(2\omega t),\ cos(2\omega t), \dots,\ sin(k\omega t),\ cos(k\omega t), \dots$$

In general, a function $f(t)$ can be represented by means of basis functions as:

$$f(t) = \sum_{k=1}^{K} c_k \phi_k(t)$$

where $\{c_k\}$ are coefficients to be estimated and $\{\phi_k\}$ is the set of basis functions.

Consider a series of n discrete data $\{y_j\}$, $j = 1, 2, \dots, n$; then, an exact interpolation is always reached when $K = n$ since the coefficients can be chosen to yield $f(t_j) = y_j\ \forall j$. Therefore, the degree of smoothing in relation to the value of K determines the quality of the operation: the lower the value of K is, the better the basis functions reflect the characteristics of the data. In other words, if we consider two different basis systems, the one that requires the smaller number of basis K to achieve the desired level of smoothing, is the better choice.

Obviously, many basis systems have been developed, each with its pros and cons and with specific applications; just to cite some of them, we have the Fourier basis, B-splines (and splines in general), wavelets, polynomials and many others.

5.2.1. Choice of the Basis System

In the choice of a basis system for a dataset, one of the most important things to do is to look whether or not data under analysis are periodic. Natural phenomena, such as wind, have a degree of periodicity in the sense that there are cycles (daily, seasonal, annual) of repeating trends and, for this reason, it is common practice to consider them as periodic.

In this case, the undoubtedly best basis system is also one of the most famous one, the Fourier series, which is given by:

$$f(t) = c_0 + c_1 sin(\omega t) + c_2 cos(\omega t) + c_3 sin(2\omega t) + c_4 cos(2\omega t) + \dots \tag{5.1}$$

where the period is determined by the parameter ω as $T = \frac{2\pi}{\omega}$.

This basis, in general, is not recommended for unstable data: functions with strong local features such as spikes or discontinuities (both in the function itself and in its derivatives) are not good elements to be approximated with a Fourier basis. However, this is not the case for wind series which, being a natural phenomenon, are never really discontinuous nor they present extremely sudden changes in intensity and thus, this basis system is perfect for the job.

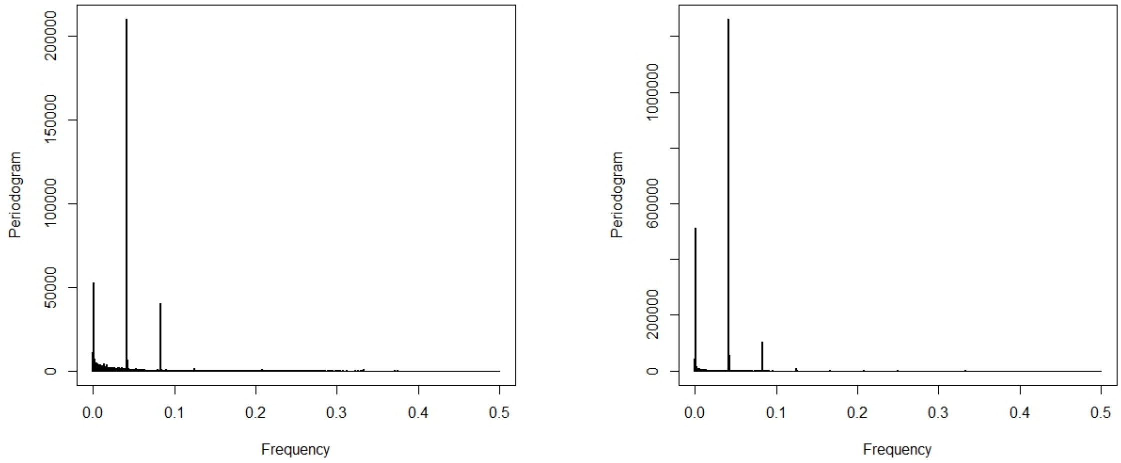
5.2.2. Choice of the Number of Basis

Some interesting comments can be done regarding the choice of the number of basis to be taken to better approximate the temporal series once the Fourier series has been chosen. Rewriting equation 5.1 expliciting the period T, we have:

$$f(t) = c_0 + c_1 sin(\frac{2\pi}{T}t) + c_2 cos(\frac{2\pi}{T}t) + c_3 sin(2\frac{2\pi}{T}t) + c_4 cos(2\frac{2\pi}{T}t) + \dots \tag{5.2}$$

For a function with period T, the frequencies of sines and cosines are $\frac{1}{T}$, $\frac{2}{T}$, $\frac{3}{T}, \dots$, i.e. integer multiples of the fundamental frequency $\frac{1}{T}$. The frequency $\frac{n}{T}$ is called the n-th harmonic and we need to understand up to which value of n we need to encapsulate all the relevant frequencies of the time series (see, for instance, Majoral et al. (2017) [30]).

The right tool for this job is the periodogram (Schuster (1898) [39]). A periodogram is an estimate of the spectral density of a signal used to identify its dominant periods (or frequencies) and to determine dominant cycles in the series. Spikes in the periodogram represent frequencies of interest and are exactly what we are looking for to decide the number of basis to be used.



(a) Periodogram U

(b) Periodogram V

Figure 5.2: Periodograms of the U and V components of wind.

Figure 5.2 represents the periodogram, computed with the function `periodogram` from the `R` package `TSA` [12], for the winds of the same location as in section 5.1, with the exception that all the 31 years available alongside both U and V components have been considered. It is clear that, in both cases, there are three main frequencies: the first peak is at $f = 0.000113$ or $T = \frac{1}{f} \simeq 8850$ which, remembering that we have 24 measurements for each day and accounting for leap years, correspond to a period of 1 year; the second, and most prominent, spike is in correspondence of $f = 0.041664$ and represent the daily cycle; the third one has $f = 0.083332$ and accounts for a period of 12h. These results reflect exactly what we were expecting from the common knowledge on meteorological phenomena and confirm that wind speeds follow a daily and annual cycles. Moreover, basically identical results can be retrieved on the whole area in exam, showing that this is not an isolated exception and the 1 year and 24 hours cycles are the predominant periods of interest when studying wind speeds.

At this point it is clear that we need to account at least for the 24 hour cycle in the choice of the number of basis; thus, not forgetting that 2020 was a leap year, we need to take at least 366 harmonics (remember that, since we are considering just 1 year of measurements in this section, our total number of values is 8784 and we are taking this value to be equal to the period T of equation 5.2). The cycle of 12 hours has a much smaller impact and we don't want to have too many basis to avoid an excessive interpolation, for these reasons the total number of basis will be 733: 366 sines, 366 cosines and the constant term.

5.2.3. Results

Results of the smoothing for our usual site of reference are represented in Figure 5.3 for the u and v components where, for a major visual clarity, only the month of January has been reported. The functional approximation is able to represent the daily behaviour of real data, following the up and down trends of the 24 hour cycles. This is particularly evident for the latitudinal component while the longitudinal one presents a more chaotic trend with many oscillations and no clear daily cycle. This behaviour comes as no surprise and the reason can be found again looking at the predominant wind regimes in the area in that time of the year; indeed, the area south of the Alps is mainly subject to North to South winds while the longitudinal ones are generated by more local and random phenomena, without a clear, regular regime.

All in all, the approximation is quite satisfactory and we can move on to the next steps.

5.3. Functional Principal Component Analysis

Now that we have our dataset in functional form, Functional Principal Component Analysis (FPCA) is the next key step to take. This technique works exactly like its multivariate counterpart, allowing the user to explore the data in a simpler way through data reduction, to see the main features that characterize them and estimating how much of the total complexity is explained by such features. A principal component analysis provides a way of looking at the covariance structure that can be much more informative than a direct examination of the variance-covariance function in the sense that functional principal components can be directly interpreted as variations of the mean trend. For this reason, through the corresponding scores, principal components can be used to characterize the single datum, showing its most distinguishing features, and PCA is often used as an intermediate step to simplify information before moving onto larger investigations such as regression and cluster analysis.

5.3.1. Definition of FPCA

First of all we recall the concept of principal components in the traditional multivariate setting (see for instance "Applied Multivariate Statistical Analysis" by Johnson and Wichern (2007) [35] for a more complete presentation).

Consider a dataset of N multivariate observations in $\mathbb{R}^p$, $\mathbf{X}_1,\ \mathbf{X}_2, \dots,\ \mathbf{X}_N$ with sample mean $\bar{\mathbf{X}}$ and sample covariance $\mathbf{S}$. First, these observations are centered in order to have

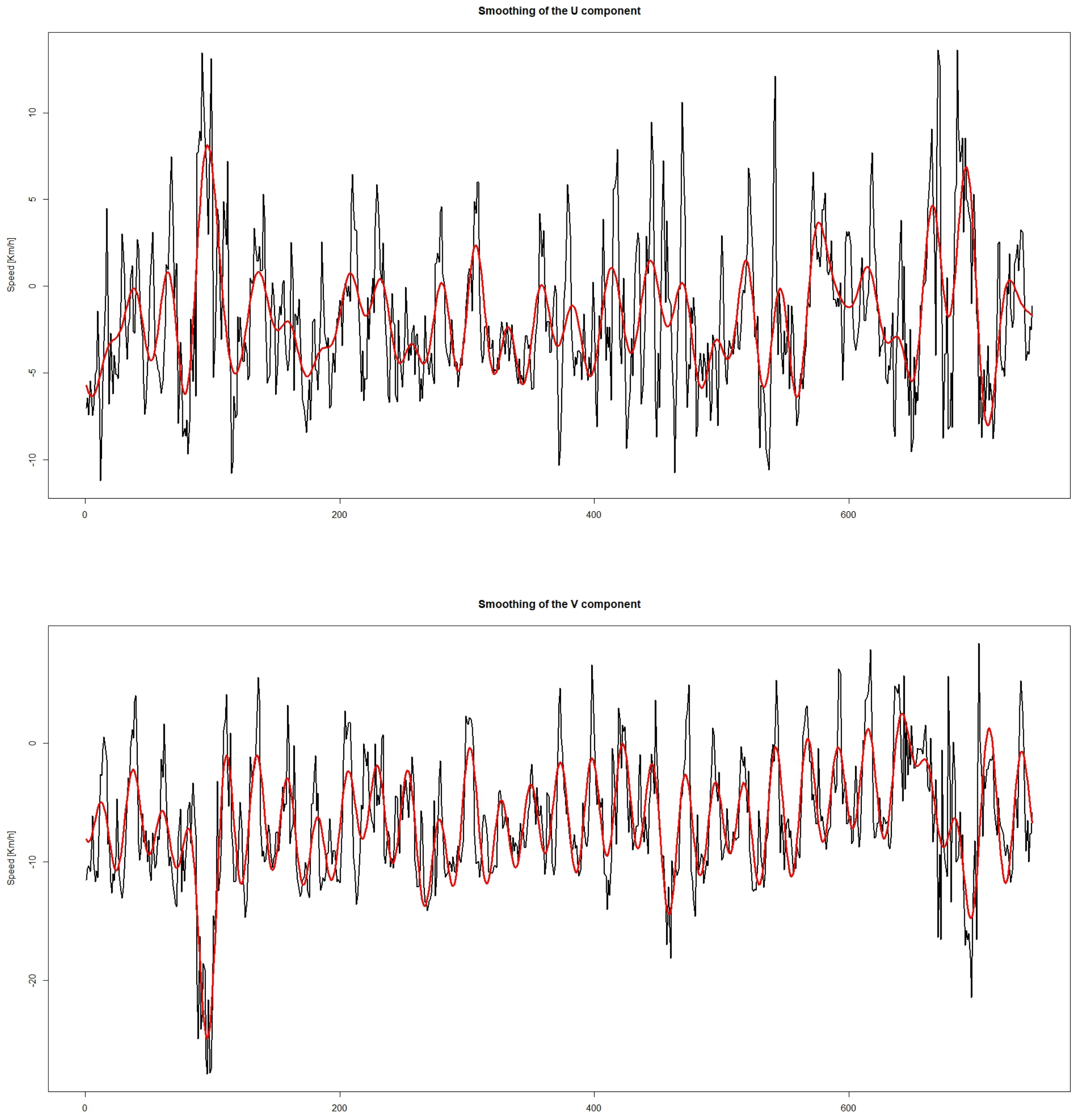


Figure 5.3: In black the original temporal series, in red the result of the smoothing.

zero mean. Then, the principal components are defined as those linear combinations of the variables yielding directions $\mathbf{a}_1,\ \mathbf{a}_2, \dots,\ \mathbf{a}_p$ of maximum sample variance; moreover, we want them to be orthonormal, i.e. $\mathbf{a}_i^{'}\mathbf{a}_j = 0\ \forall i \neq j$ and $\mathbf{a}_i^{'}\mathbf{a}_i = 1$. In practice, the principal components are determined as the eigenvectors of the covariance matrix $\mathbb{S}$, i.e. for $k = 1,\ 2, \dots,\ p$ they are the solution of the eigen-equation:

$$\mathbb{S}\mathbf{a}_k = \lambda_k \mathbf{a}_k$$

The eigenvalue λ_k associated with eigenvector $\mathbf{a}_k$ represents the variability along the direction $\mathbf{a}_k$ and we call score x_{ik} the projection of variable $\mathbf{X}_i$ along the direction $\mathbf{a}_k$: $x_{ik} = \mathbf{X}_i^{'}\mathbf{a}_k$.

In words, principal components are linear combinations of the variables that explain the highest share of the variability of the dataset; in this way, one can take just few of them and discard the others and still be able to have insight on the problem while working in much smaller dimensions.

With some changes to adjust to the structure of Hilbert spaces, we can now give the definition of PCA in the functional setting. Indeed, consider a dataset of N functional observations in an Hilbert space H (particularly, in L^2) $X_1,\ X_2, \dots,\ X_N$. The objective is again to find an orthonormal system of "directions" (in the sense of H, directions are functions) $\xi_1,\ \xi_2, \dots$ that maximize the variability of the projections of the actual observations along them.

Similarly to the multivariate case, the functional principal components, now also called "harmonics", are defined as the eigenfunctions of the covariance operator C as said in section 5.1, i.e. the solution of the eigen-equation:

$$C\xi_k = \lambda_k \xi_k$$

Again, the eigenvalue λ_k associated to ξ_k represents the variability along the "direction" ξ_k and we call functional score x_{ik} the projection of observation X_i along ξ_k, i.e. $x_{ik} = \langle X_i, \xi_k \rangle$.

5.3.2. Number of PCs

An important result in PCA (both multivariate and functional) is that the eigenvalues λ_k are proportional to the percentage of variation explained along the corresponding eigen-direction. In particular, for the multivariate case it holds that $\sum_{i=1}^{p} Var(X_i) = \lambda_1 + \lambda_2 + \dots + \lambda_p$ and an equivalent relation is true also for the functional case. Consequently, the proportion of total variance due to the k-th principal component is $\frac{\lambda_k}{\lambda_1+\lambda_2+\cdots+\lambda_p}$ and, in general, if most of the total variance can be attributed to the first one, two or three components, then these components can replace the original variables without much loss of information.

However, in many applications this is not the case and a more accurate analysis is needed to choose the number of principal components to retain. In these situations, there are two main graphs that we can look at to decide this number but, unfortunately, there is no mathematical procedure that tells us the optimal result, and much of the work is left to the interpretation of the statistician, who will need to take a subjective decision based on the task and the current objective.

The first one, on the left in Figure 5.4, reports the cumulative variance explained by subsequent principal components; the second, on the right, represents the eigenvalue associated with each of them and it is called scree plot. In both cases what we look for is an "elbow" in the sequence. This elbow shows a point at which the remaining eigenvalues are small and account for little of the remaining variability.

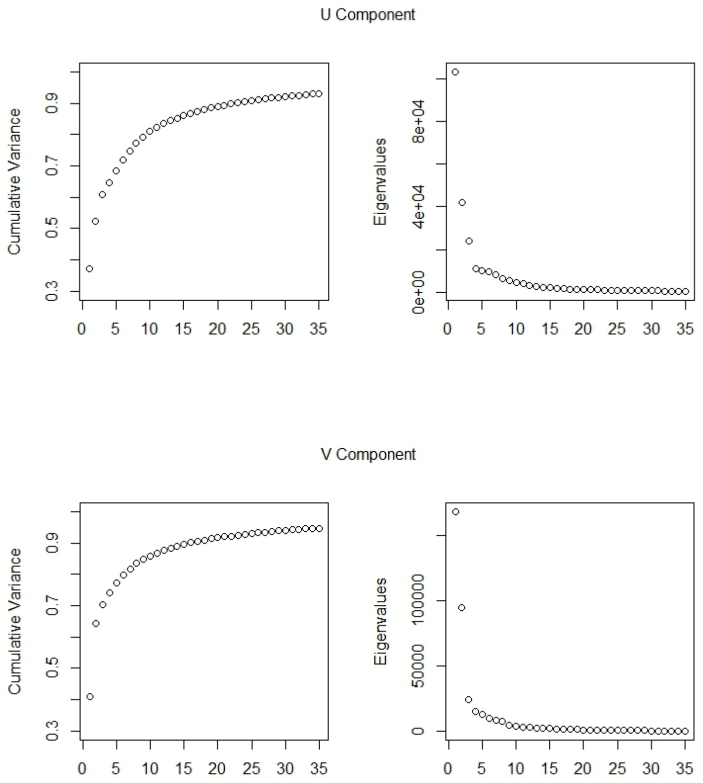


Figure 5.4: Cumulative variance and eigenvalues associated to each functional principal component.

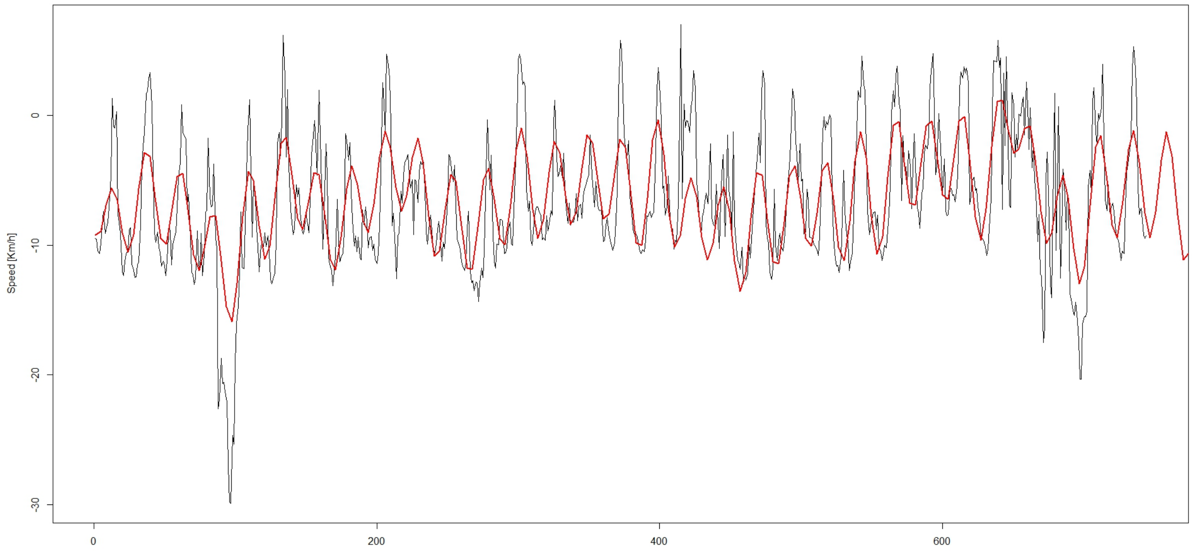


Figure 5.5: In black, the original temporal series; in red, the projection on the first four principal components.

In both cases of longitudinal and latitudinal winds, no clear elbow can be seen in the cumulative variance graph, while a more evident reduction is visible in correspondence of the fourth eigenvalue. The cumulative variance explained by the first four PC would be around 65% for the U component and around 75% for the V component. Another little jump can be seen after the 7th or 8th eigenvalues but, taking 8 principal components, would lead in an increase of explained variability of just 10%, while reducing by a lot the interpretability; thus, we deemed sufficient, for our interests, the use of four principal components.

A confirmation of the goodness of this choice comes from figure 5.5, where the projection on just the four PCs is compared with the original series. In formula this means:

$$X_i \simeq \sum_{k=1}^{4} x_{ik}\xi_k = \sum_{k=1}^{4} \langle X_i, \xi_k \rangle \xi_k$$

and we can see that this number of principal components is already sufficient to recover a good approximation of the trends in the series, showing that four scores should be able to characterize decently the main features of wind in each site.

5.3.3. Visualization

At this point, it is of general interest in any study involving PCA to try and give an interpretation to the principal components, to understand which characteristics of the phenomenon are represented by them and how they are reflected by the data.

In a functional setting such as the one we are working with, it is helpful to plot the principal component functions as perturbations of the mean. This is done by multiplying the harmonic by a fixed value and then summing and subtracting it to the mean. The fixed value is usually chosen to be the square root of the eigenvalue corresponding to the harmonic to be represented, but can be chosen subjectively by the user for major clarity.

Figure 5.6 shows the results of this procedure for the first principal component: the mean is the curve in black, to the green one has been added the first PC, while to the red one it has been subtracted. Figure 5.7 instead, shows the relative scores in each location, giving an idea of the behaviour in that site; more positive scores indicate that the trend will be more similar to the green curve, while negative ones indicate similarities with the red curve.

Given the quite high number of bases and the consequent difficulties in interpretability, we will provide comments regarding just the first principal component for U and for V.

For longitudinal winds, the first PC seems to characterize variability mainly during winter months (remember that the series goes from the first of January to the 31 of December), with positive scores associated to predominant winds with positive sign (West to East) while negative scores associated with negative winds (East to West). A score around zero, as most areas have in Figure 5.7 on the left, indicates that no particular trend can be seen during winter months.

The first PC for latitudinal winds, instead, shows that positive scores are associated to generally lower North-South winds throughout the year, while negative scores indicate the presence of predominant negative (North to South) winds. This is confirmed on the map where, for instance, we can see a long stripe of positive scores around the 46th parallel; there, indeed, we find the Valtellina, a long valley extending West to East where we expect winds to mainly follow this direction and, by contrast, we expect latitudinal winds to be less evident due to the conformation of the territory.

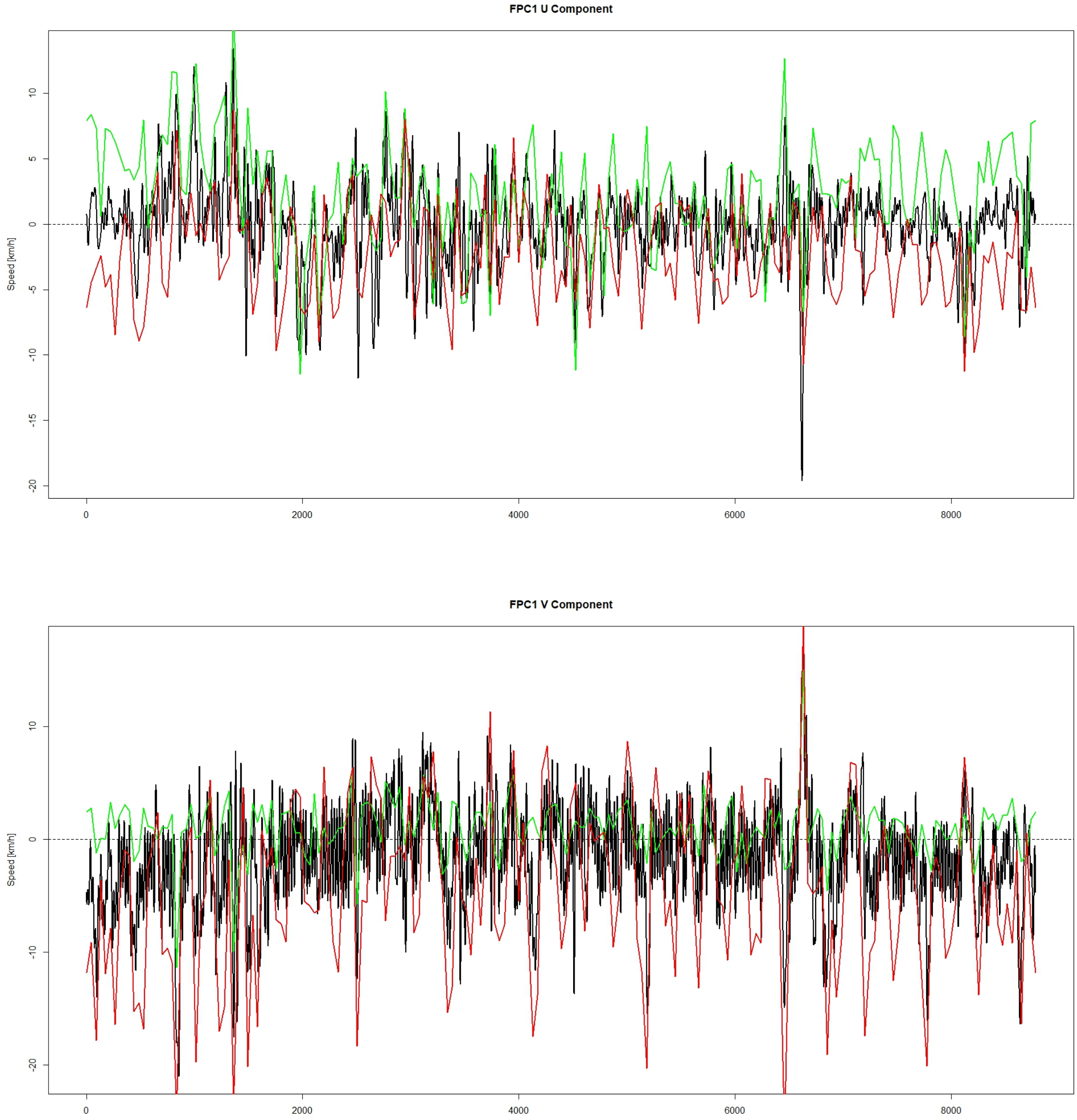


Figure 5.6: Behaviour of the first principal component as perturbation of the mean wind series.

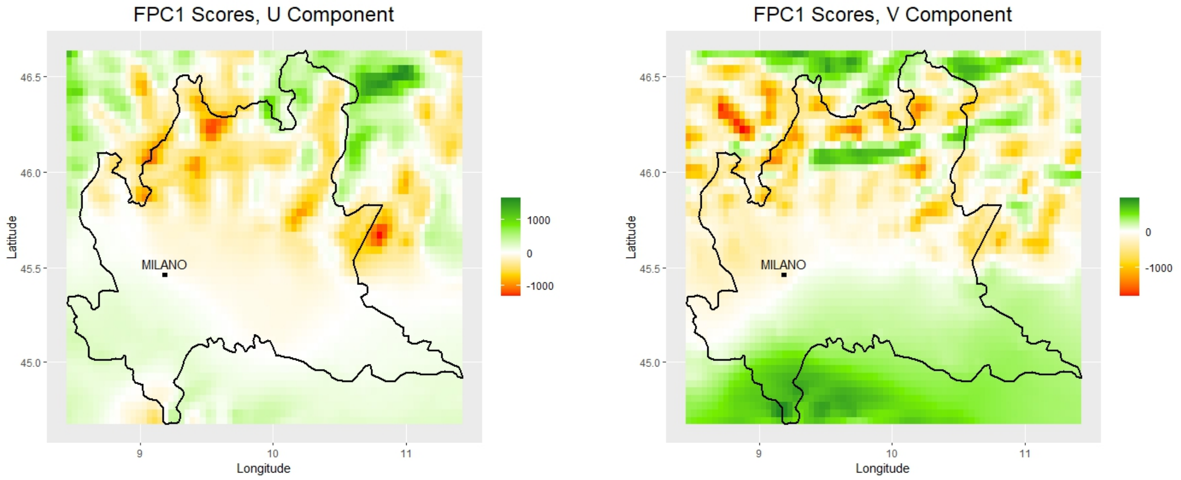


Figure 5.7: Behaviour of the first principal component as perturbation of the mean wind series visualized on the map of Lombardy.

5.4. A Distance between Wind Regimes

This section will be devoted to the definition of a metric allowing us to measure when two wind time series are close and, thus, if the sites interested by them can be considered to be subject to the same wind regime.

The reasoning is intuitive: as one can guess, temporal series coming from locations that are close are extremely similar, while if two sites are distant, their temporal series will look quite different. However, one needs to be careful not to think that geographical distance alone determines similarity in wind behaviour since, for instance in mountains areas, wind can change by a lot also in short distances. For this reasons, we want a metric that accounts just for the temporal series and is able to highlight similarities in the global trends. If this distance will be good enough, spacial dependence should be recovered "automatically", in the sense that close locations will have close wind regimes and thus will be close also according to it.

Considering the procedure followed in the chapter, the most natural idea is to look at the temporal series as functional data, find their principal components and then compute the distance between the corresponding scores. In this way, hopefully, we should be able to recover a measure of similarity between the trends of the series that involves only four values but that is able to capture the variability along the whole temporal interval.

In particular, we used the euclidean distance between scores in the four-dimensional space generated by them and, then, we considered the sum of the distances between longitudinal components and between latitudinal components to define a distance between two sites; in this way, differences in the angle of blowing are automatically accounted for. In formula:

$$D(i,j) = \sqrt{\sum_{k=0}^{K}(scores_{u,k}(i) - scores_{u,k}(j))^2} + \sqrt{\sum_{k=0}^{K}(scores_{v,k}(i) - scores_{v,k}(j))^2} \quad (5.3)$$

where i and j are the indices of the two sites we are considering, K is the number of scores we take into account, and $scores_{u,k}$ and $scores_{v,k}$ represent, respectively, the k-th score value of the component U and of the component V of the wind.

Figures 5.8 and 5.9 report two examples of application of this metric on the dataset. In both cases, in the left panel are shown the distances of all cells from a predetermined one, represented with a black square while, on the right, the temporal series of U component for the reference site (in black) is confronted with the ones of two other sites with similar geographic distance but with very different behaviours: one (in lightblue) similar to the first one and the other (in red) very different. For major visual clarity, only the series relative to the month of January is reported.

The first figure shows that, in a plain area, wind series can be quite close even at long geographical distances since there is no obstacle for the wind, which can blow unaltered through the whole plain. The second figure chooses as a point of reference a mountain and it clearly shows that wind patterns can change very rapidly in such areas. Interestingly, some mountain sites quite far away seem to have similar behaviour; this is not to be considered an oddity since it is very possible that two different locations present similar patterns in the time series. The quality of this distance in measuring the dissimilarity between wind regimes of two sites is portrayed also in the right panels of Figures 5.8 and 5.9: areas that are close with respect to this distance have also similar trends in the wind time series, while areas that are distant show also very different wind behaviours.

Overall, it seems that this distance is quite suited for the job and we will use it in the next chapter as a foundation for a clustering algorithm able to identify areas with the highest similarity and produce a grouping based on wind correlation.

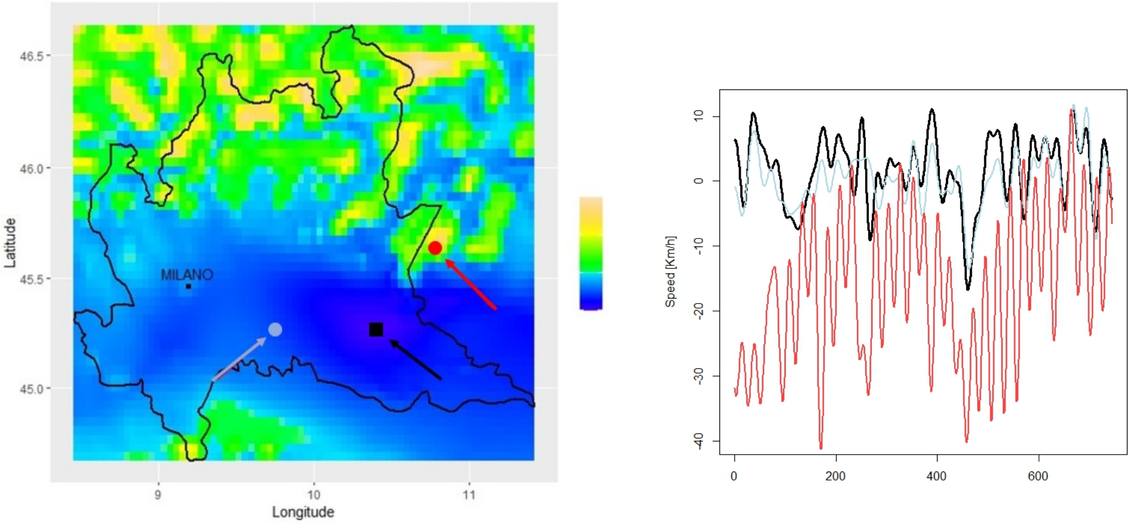


Figure 5.8: Example of distance between wind regimes measured taking a location in the Po Valley (black square) as reference point.

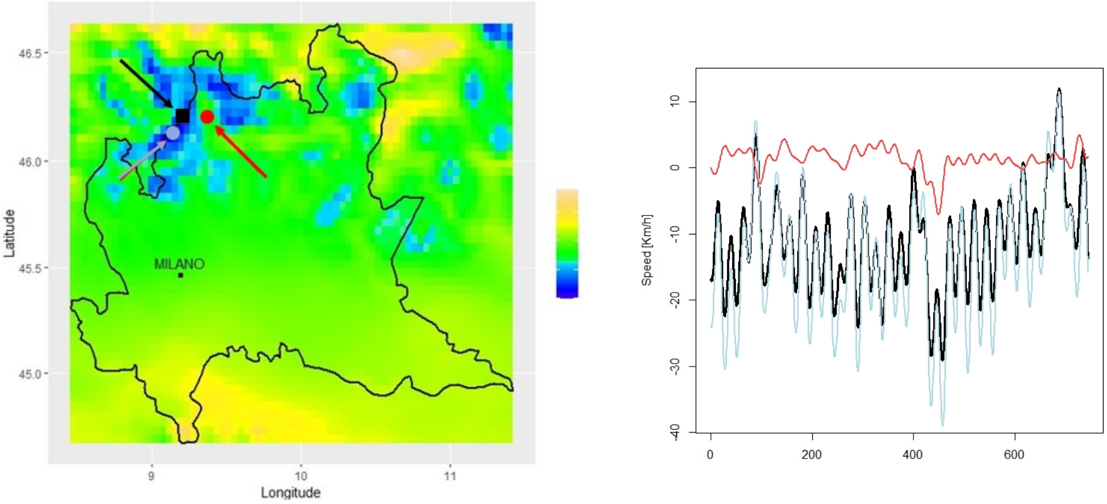


Figure 5.9: Example of distance between wind regimes measured taking a location in the Alps (black square) as reference point.

6 | Clustering Algorithms

Having defined a distance to measure how similar or different the wind events recorded in two sites are, we now aim at providing a new kind of clustering that, differently from the one presented in chapter 4, which focused on the risk of infrastructure failures, will group our sites based on the whole characteristics of wind. As previously mentioned, we will look at our data as functions and try to exploit the features captured using the techniques described in chapter 5.

To this end, we employed two different clustering algorithms, characterized by two different approaches to accomplish the task: the first and more immediate one, the Geostatistical Hierarchical Clustering algorithm, keeps into account the spatial adjacency of every site with each other, while the second one, the so called Bagging-Voronoi Classifier, is based on a bagging strategy followed by a phase of aggregation of all the information into a single result. Both these algorithms will be described into details in the following sections and for each one of them we will present the results obtained.

6.1. Geostatistical Hierarchical Clustering

The Geostatistical Hierarchical Clustering has been presented in the Romary et al. (2015) [36] as a possible solution to take into account the spatial structure of data when it comes to cluster them. When analysing data which it is safe to assume that are spatially correlated or even dependent like in our specific case, it is crucial to define the notion of neighborhood: data sampled in a geospatial setting define a geometrical set, namely a set of points in the geographical space. We can represent this structure as an undirected graph in which each node represents an observation and each edge represents the existence of an adjacency relation between the nodes that it connects. In this way we can say that the set of observations linked to a given datum represents its neighborhood.

Coming to our case study, a geometrical structure is already present since we are working on a grid, and the concept of adjacency follows trivially. The choice we can make (see

Figure 6.1 for a graphical explanation) is to consider adjacent two sites only if they share an entire edge in the grid (4 neighbors) or to consider adjacent also those sites that share only a vertex with each other (8 neighbors). We opted for the 8 neighbors structure to give a bit more flexibility to the algorithm and because of the nature of the phenomenon we are studying (since it is obvious that the wind does not blow only in four cardinal directions).

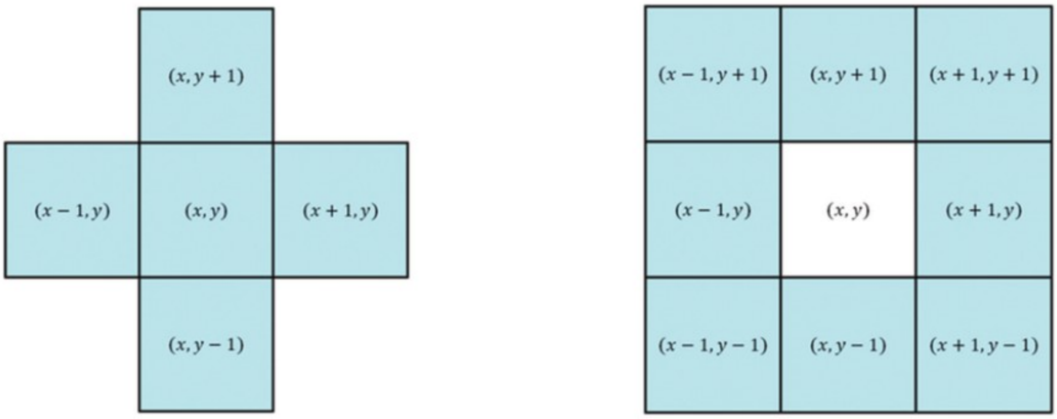


Figure 6.1: Example of 4 and 8 neighbors structure.

The concept of neighborhood comes into play because, at each iteration, the algorithm is allowed to cluster together only those sites that are adjacent. In this way we will be able to divide the territory under analysis in the chosen number of clusters and those will be continuous in the space. Except for this additional constraint, this method works as a common hierarchical clustering algorithm and thus requires the choice of a distance $d(x, y)$ (Euclidean for instance). In our application we will exploit the distance we have defined in Section 5.4 since it captures well the differences between wind functions.

Having defined both the distance and the adjacency structure to employ, the algorithm starts with the computations of the matrix of distances between sites and of the binary matrix attesting which sites are adjacent to each others, and then proceed iteratively clustering together the sites that are both adjacent and the closest in terms of the distance chosen.

Notice that we are going to change the notation used to address a site in this section: indeed so far, since we are working on a grid, to denote a site we used two indices (one for longitudinal and the other for latitudinal component). Now we are going to move to a single index notation, counting by row top to bottom and by column left to right: for example, in the case of Lombardy with 50 rows and 74 columns, site in row 5 and column

54 will be indexed with $74 \times (5-1) + 54 = 350$ (see Figure 6.2). The reason why we need to perform this shift in notation is that, in the Geostatistical Hierarchical Clustering, the two matrices of distances and adjacencies need to compare each single site to each other and thus these two matrices will have dimension 3700×3700 (where $3700 = 74 \times 50$) and will be symmetric.

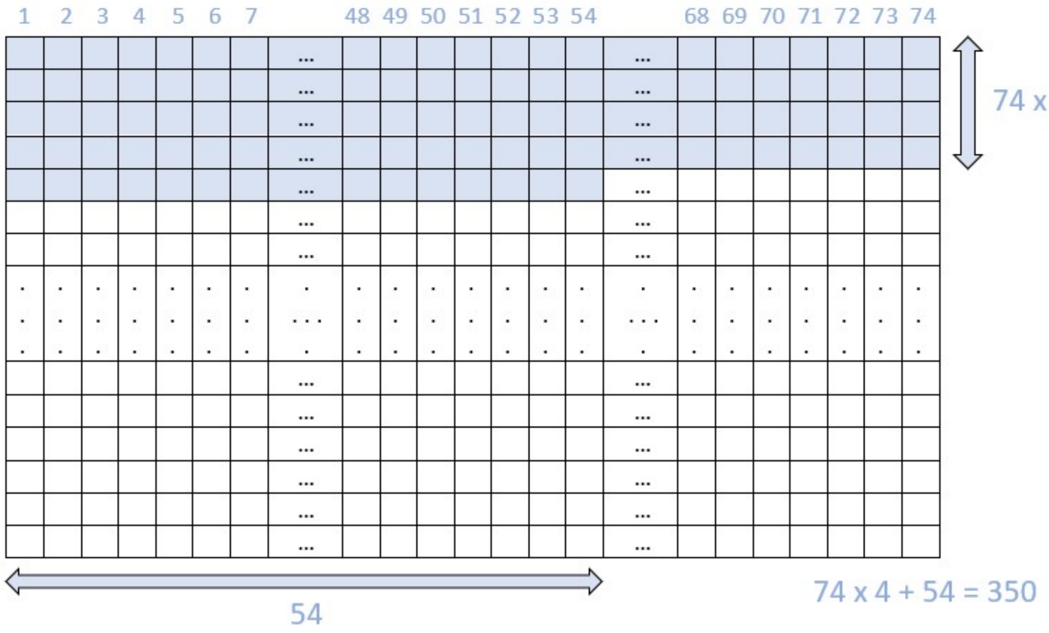


Figure 6.2: Example of the new indexing system.

Having clarified the new indexing system, we will proceed by reporting the pseudocode of this algorithm (Algorithm 6.1) and, in the next section, describe in detail each step of the procedure.

Algorithm 6.1 Geostatistical Hierarchical Clustering

```
1: Initialize the matrix of distances $D$ between each site using the chosen distance $d(x, y)$.
2: Initialize the binary matrix of adjacency $A$: $A(i, j) = 1$ if sites with indices $i$ and $j$
   are adjacent following the chosen definition of adjacency, 0 otherwise.
3: repeat
4:     Identify which adjacent sites (or clusters) are the closest by means of matrix $D$ and
       group them together
5:     Update both matrices $A$ and $D$ to take into account which sites or clusters have
       been grouped together.
6: until There is only one cluster with all the sites in it
```

6.1.1. Details of the Algorithm

1: Initialization of the matrix of distances D

The distance matrix is initialized as a square and symmetric matrix of dimension $N \times N$ where N is the total number of sites in the area we area considering. In our specific case, when we analyse Lombardy territory, N is equal to $74 \times 50 = 3700$. All we have to do is to simply compute the distance between each site. The distance we employed is the one defined in Section 5.4 in formula 5.3.

Notice that the main diagonal of this matrix is composed of zeros only (since on the diagonal we measure the distance of a site from itself) and that the matrix is symmetric, thus we can lower the computational cost of this part by simply calculating only a triangular matrix, halving the number of operations.

2: Initialization of the matrix of adjacency A

Once again, the matrix under consideration is square, symmetric and has only zeros on the main diagonal. It is, moreover, binary. Differently from the computation of the distance matrix, however, the adjacency $A(i, j)$ is a bit more tricky to compile, since there exist different cases based on the position of the sites into the grid. Indeed, there are three of them, as highlighted in Figure 6.3:

1. the site i is placed in one of the four corners of the grid: in this case its neighborhood is composed by only 3 other sites.
2. the site i is placed on one of the edges of the grid; in this case its neighborhood is composed by 5 other sites.
3. the site i is in the interior of the grid: in this case its neighborhood is composed by all the 8 sites around it.

3: Identification of closest adjacent clusters and grouping

Having initialized all the principal quantities we need, we enter in the iterative part: steps 3 and 4 regard all commands that are computed inside a "while" loop. The first step of each iteration is to compute the auxiliary matrix C. Each element of C is obtained as the product of the corresponding elements in the matrices of adjacency A and of distances D:

$$C(i, j) = A(i, j) \times D(i, j) \quad \forall \ \ i, j \in \{1, 2, ..., M\}$$

where M is the size of the matrices A and D.

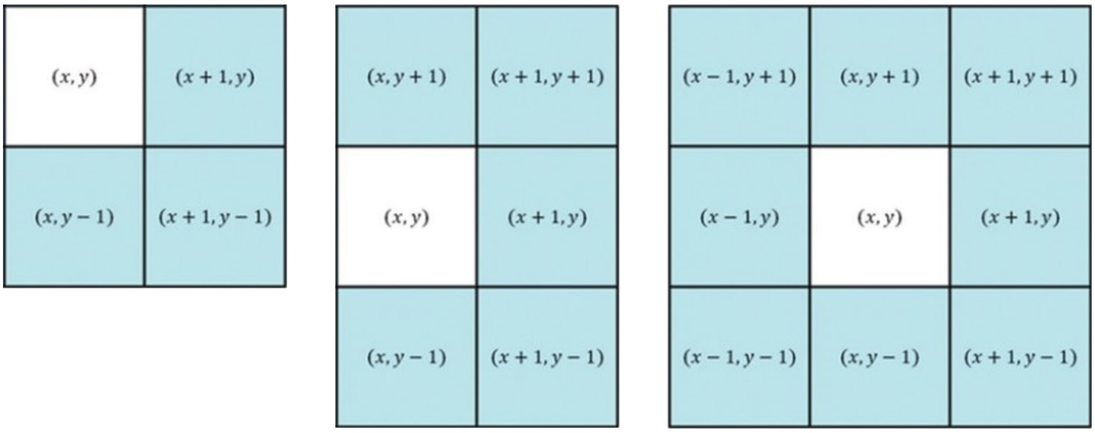


Figure 6.3: From left to right, examples of the 3, 5 and 8 neighbors cases.

The auxiliary matrix C obtained in this way is, then, a square and symmetric matrix with, at each iteration, the same dimension of A and D. This matrix resume the information brought by both A and D since each elements $C(i, j)$ is equal to zero if the two sites (or clusters) are not adjacent while is positive if i and j are actually neighbors and, thus, the value $C(i, j)$ is equal to the distance between the two of them. For faster computations, we then reduced the matrix C to a upper triangular matrix (since it is symmetric).

To find which are the two clusters that are going to be grouped together at each iteration it's enough to look for the smallest value greater than zero in C and get the indices of the corresponding sites or clusters (notice that if we don't consider only the triangular matrix we would get two pairs of the same indices). At this point we just need to assign each element of cluster j to cluster i, creating a bigger new cluster.

4: Update of the matrices A and D

Once we grouped together two clusters in a new one, we have to update both matrices A and D to reflect the changes in the clustering structure.

First of all, the adjacency matrix changes because every time two clusters merge together, we have to compute the neighborhood of the new group and, to this end, it's sufficient to do the union of the neighborhoods of the two old clusters.

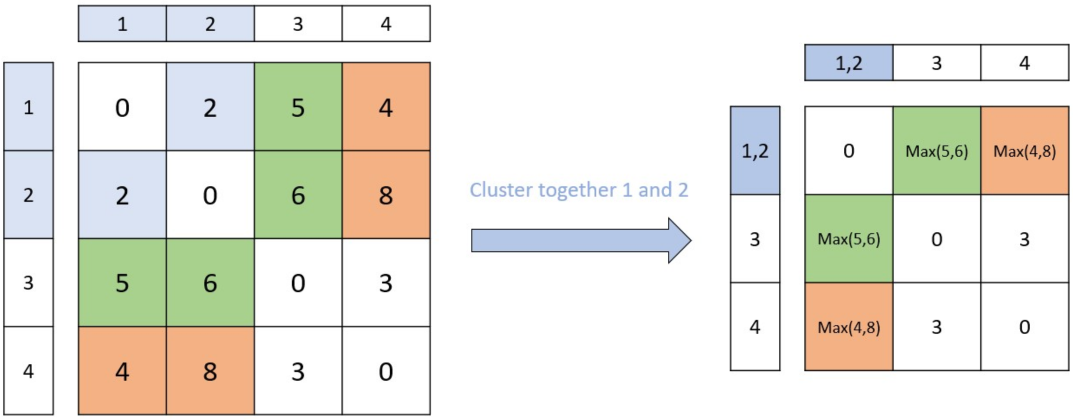
For what concerns D, instead, the situation is a bit more convoluted: as long as we are evaluating two simple sites, computing the distance between them is straightforward since we proceed as done in the initialization of matrix D, but when it comes to evaluate how far two clusters composed by more than one site each are, we have to decide a new standard. There are several techniques adopted in hierarchical clustering applications but the most commonly used are:

- **Single Linkage**: measures the distance between two clusters as the minimum of the distances between each element of the first cluster and each element of the second one:
$$d(c_1, c_2) = \min d(x, y) \quad \textit{with} \;\; x \in c_1, \;\; y \in c_2$$
Single linkage tends to cause what is called a "chain effect" for which the clusters produced look like long chains of elements, grouping together elements that are actually very different or far away from each other.
- **Complete Linkage**: measures the distance between two clusters as the maximum of the distances between each element of the first cluster and each element of the second one:
$$d(c_1, c_2) = \max d(x, y) \quad \textit{with} \;\; x \in c_1, \;\; y \in c_2$$
Differently from the single linkage, complete linkage produces ellipsoidal clusters.
- **Average Linkage**: measures the distance between two clusters as the mean of the distances between each element of the first cluster and each element of the second one:
$$d(c_1, c_2) = \frac{1}{|c_1||c_2|} \sum_{x \in c_1, y \in c_2} d(x, y)$$
Also average linkage tends to produce ellipsoidal clusters.
- **Centroid Linkage**: measures the distance between two clusters as the distance between the centroids of the two clusters:
$$d(c_1, c_2) = d(\bar{x}, \bar{y})$$
where $\bar{x}$ and $\bar{y}$ are respectively the centroids of c_1 and c_2. With the term centroid we can address many objects, like the barycenter of the set for instance. With respect to the previously mentioned techniques, this one is more rarely used.

To best meet our needs, we decided to employ the complete linkage: first of all we avoided the flaw of the chain effect connected to single linkage, so that we discarded this method, the computational load is lower than that of average linkage and centroid linkage, and, moreover, we wanted the data in our clusters to be as close as possible (in terms of distance $D(i, j)$, not geographical distance) to each other and we thought that complete linkage was the best one to pursue this goal. Following this decision, to update the matrix D, at each iteration we have to substitute the two rows and two columns corresponding to clusters i and j with a row and a column compiled with the maximum distance between clusters i and j, and every other cluster in the matrix:

$$D(c_1, k) = \max[D(i,k), D(j,k)] \quad \forall k$$

where c_1 is the new cluster formed by the union of clusters i and j. Figure 6.4 shows an example of this operation.



	1	2	3	4
1	0	2	5	4
2	2	0	6	8
3	5	6	0	3
4	4	8	3	0

	1,2	3	4
1,2	0	Max(5,6)	Max(4,8)
3	Max(5,6)	0	3
4	Max(4,8)	3	0

Figure 6.4: In this example, clusters 1 and 2 are the closest and, thus, are grouped: the distance of the new cluster {1,2} from the clusters 3 and 4 are respectively $\max[5, 6]$ and $\max[4, 8]$.

Notice that, following this procedure, the size of A and D, and thus the size of the auxiliary matrix C, shrinks by 1 at each iteration.

Additional details

To conclude the detailed description of the algorithm, we want to highlight the fact that this method have been built in order to allow the user to travel backwards into the clustering steps and observe the evolution of the grid. Indeed, the output is structured in two elements:

- **Clusters Matrix**: this item is a $N \times N$ square matrix where N is the total number of sites composing the grid considered. Each row of this matrix is filled with indices denoting the cluster to which the site has been assigned at that step, with the last row representing the starting situation (when all sites are independent from each other and form a single site cluster) and the first row being the case where all the sites are clustered together. In this way, users have a nice and easy way to observe the clustering situation that they are interested into: if they are looking for the grouping with n clusters, it's enough to look at the n-th row of this matrix. Notice that the choice of always proceeding with clustering until we get a single group is not demanding in terms of computational cost since the time needed for each iteration is proportional to M^2 where M is the size of the matrices A and D at that specific iteration. So, the time needed to complete an iteration decreases at each step since the size of those matrices shrinks by 1 every time. For this reason, unless we are interested in a huge number of final clusters (which is not the case in our applications), the gain in comfort is largely greater than the computational drawbacks.

- **Merging Distances Vector**: with "merging distance" we refer to the distance measure between the two clusters that, at each iteration, are grouped in the same cluster. In other words, it's the minimum distance discovered in matrix C (and thus the minimum distance measured between adjacent elements) at each iteration. Notice that another quality of complete linkage is that it guaranties that the merging distance is a strictly non-decreasing quantity. This very desirable feature of the merging distance comes into play when we need a visual tool to decide how many clusters to consider. A reasonable common practice in this kind of applications, is to look for the largest jump in the sequence of merging distances because that represents the moment where two clusters quite different from each other have been grouped together. An example of merging distances curve is shown in Figure 6.5. Once again, since our algorithm proceeds until it clusters together all the sites, having at disposal the merging distances sequence may help in identifying which are the clustering steps to look at.

Notice that, after having chosen how many clusters we want, it may be reasonable or necessary (based on the specific needs or on the code following this run) to relabel the clusters in order to assign them indices that go from 1 to K, where K is the total number of clusters chosen. This because the algorithm assigns to each cluster a unique index but not necessarily in the former range.

Moreover, also a procedure to return to a double index notation is suggested: we recall that each clustering is denoted by a vector of dimension N, where N is the total number of sites in the area. By reverting the conversion done at the beginning, it is possible to return to a matrix in which each cell contains the label associated to the cluster that site belongs to. This configuration is much better for plotting purposes and representing results.

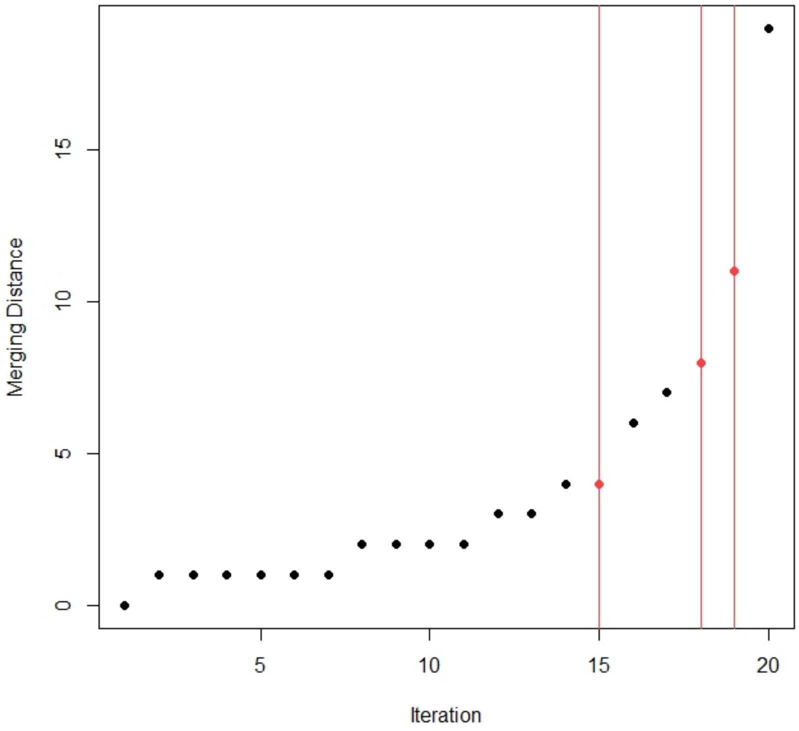


Figure 6.5: This toy example suggests that there are 3 main candidates for the number of clusters: 2, 3 and 6. It is choice of the user which number to adopt based on the specific necessities.

6.2. Bagging Voronoi Classifier

The Bagging Voronoi Classifier is an algorithm proposed in Secchi et al. (2012) [40] as a method to perform clustering of spatially dependent data. The algorithm is composed of two main parts: the first one, the Bootstrap phase, in which the grid is divided in many sets just on the base of geographical information, a representative is computed for each set and then this sets are clustered together on the basis of the representatives similarities. The second phase is devoted to performing cluster matching and, thus, produce the final grouping that will represent the output of the algorithm.

The main feature of this method is its bootstrap nature and what the algorithm does in the iterative part. This phase starts by producing what is called a Voronoi tessellation of the grid, which is nothing different from a partition of the plane into regions. Given a set of points, named centres or nuclei, to produce a Voronoi tessellation we have to assign each point of the plane to the region corresponding to the closest centre (see Figure 6.6 for an example of this structure). An informal use of Voronoi tessellation can be traced back to Descartes [32] but these diagrams take name from Georgy Feodosievych Voronoy, who defined and studied the general n-dimensional case in 1908 [50] [51].

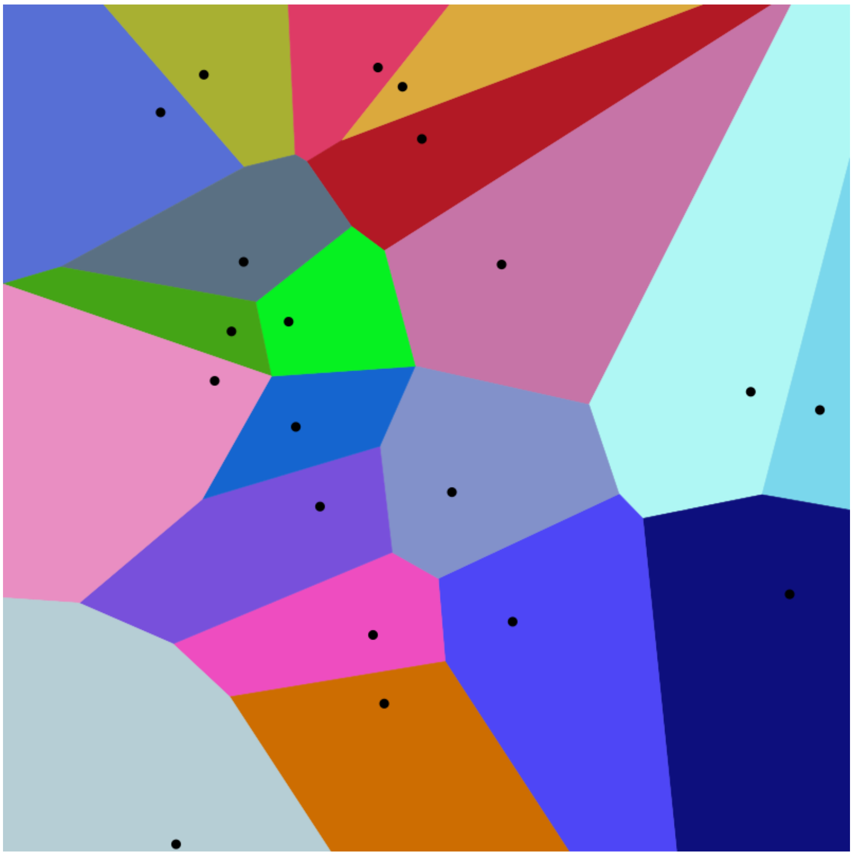


Figure 6.6: Example of Voronoi Tessellation with 20 nuclei.

After having produced the Voronoi Tessellation, a representative quantity is computed for each one of the elements and, based on these representatives, the clustering is obtained, measuring the differences between them and grouping the most similar ones. The main difference between the Geostatistical Hierarchical Clustering and the Bagging Voronoi stands right in this part: while the first method defined a concept of adjacency to cluster together only areas that are actually close both in space and in "behaviour" of the phenomenon studied, the Bagging Voronoi does not impose any constraint of neighborhood and let the data free to act for themselves: the idea is that those regions that are similar in behaviour are also close in space and, thus, will form compact clusters and not scattered ones.

The main reason why we decided to test Bagging Voronoi on our dataset is that this algorithm is thought for functional data. Indeed, once the Voronoi tessellation is formed, each representative is computed as the "weighted mean" of all the curves included in the subset (we will see later what we intend by means of weighted mean). This, with respect to Geostatistical Hierarchical Clustering, represent for sure an improvement for our purposes since it allows to capture more information in intermediate steps of the procedure.

As done before we now proceed to report the pseudocode of the algorithm (Algorithm 6.2) and, in the next section, describe in detail every step of the method.

Algorithm 6.2 Bagging Voronoi Classifier

```
Initialize the hyperparameters of the algorithm: $B$, $n$, $p$, $K$ and choose the distance $d(\cdot,\cdot)$ used in the Voronoi Tessellation.
for $b := 1$ to $B$ do
    Choose a set of $n$ sites $\Phi_n^b = \{Z_1^b, ..., Z_n^b\}$ of the starting grid $S_0$ to play the role of centres and compute the Voronoi Tessellation with these nuclei.
    Compute the representative $g_i^b$ for each element $i$ of the tessellation.
    Perform dimensional reduction of the representatives by projecting them on the space spanned by a proper $p$-dimensional orthogonal basis and, thus, obtaining the $p$-dimensional scores.
    Cluster the scores in $K$ groups according to a chosen unsupervised method.
end for
Perform cluster matching, i.e. match the labels across the $B$ bootstrap replicates of the clusters, to ensure identifiability.
for all $x \in S_0$ do
    Calculate the frequencies of assignment of the site $x$ to each one of the $K$ clusters and assign the site under consideration to the most frequent group.
    Compute spatial entropy for the site $x$
end for
```

6.2.1. Details of the Algorithm

1: Initialization of the Hyperparameters

- B: the number of bootstrap replicates to produce. This number should be set high enough to obtain reliable results but consider that the computational cost increases with the number of iterations.

- n: the number of elements that constitute the Voronoi tessellation. The choice of this hyperparameter is crucial and when choosing it, it is necessary to evaluate a trade-off: on one side, if we set n too low we will get elements of tessellation that are too big and, thus, group together sites that are fairly different. On the other side, if n is, instead, too large, we risk to obtain a final result with extremely scattered clusters since we are not imposing adjacency in this algorithm. Moreover the computational cost in this case would increase drastically. When setting n we have to keep an eye on the structure of data we are working with: in our case, applying this method to Lombardy, we have a big uniform area corresponding to Po Valley, that will likely require few clusters, and the other half with mountains and hills, characterized by wind regimes changing every few kilometers.

- p: the dimension of the orthogonal space where we project the representatives. This choice depends strictly on the method used. In our case, we used the most classical projection method for functional data, i.e. the functional Principal Component Analysis, and thus we set p by observing the curves of variance explained by the method. Referring to Figure 5.4 in Section 5.3, we opted for the first 4 principal components and, thus, set $p = 4$.

- K: the number of clusters in the final result. Differently from the Geostatistical Hierarchical Clustering, we have to choose this number a priori and not after running the algorithm and, because of this, we have to carefully set it exploiting preliminary information at our disposal.

- $d(\cdot, \cdot)$: distance utilized to perform the Voronoi Tessellation. Since we are working on a grid of data described on the surface of Earth, the most obvious choice is to compute the distance between each cell of the grid as the geodesic distance between their centres. Another possibility, which simplifies the computation, is to exploit the Pythagoras theorem and use the differences of the rows indices and of the columns indices. In other words:

  $$d(X_1, X_2) = \sqrt{(r_1 - r_2)^2 + (c_1 - c_2)^2}$$

  where r_1 and r_2 are the row indices for sites X_1 and X_2 and c_1 and c_2 are the column indices for sites X_1 and X_2.

2: Voronoi Tessellation

First of all, we have to select the nuclei for this step and notice that they must not be same at each iteration; indeed the meaning of the bootstrapping procedure is to change at each iteration the starting point of the algorithm but, hopefully, get to similar conclusions every time at the end. So, a possibility for the selection of centres is to sample them from a Uniform distribution defined on the bidimensional grid S_0:

$$\Phi_n^b = \{Z_1^b, ..., Z_n^b\} \quad with \;\; Z_i^b \sim \mathcal{U}(S_0) \;\; \forall i \in \{1, ..., n\}$$

The uniform distribution is the most general and simple way to sample the centres but, if we have any reason to think that it is present any sort of structure in the grid we are considering, it may be appropriate to employ a different distribution, placing more nuclei in some areas. For example, in our case we have seen that Lombardy is basically split in two parts: Alps and Po Valley. About the plain, based on what we have observed so far, we can say that it is a very steady area, with a uniform wind behaviour and thus we may need to place less nuclei in this area with respect to the mountainous area where, instead, wind changes quickly and substantially in just few kilometers of distance. For this reason, we set the probability of placing centres on the Alps higher to the one assigned to plain, to have more flexibility where it is needed without increasing the computational cost. To do so, in particular, we exploited a feature offered by the `R` function `sample`, which allow to set the probability of extracting each element of the pool. In practice, we set the probability of extracting the first 25 rows (corresponding to the mountainous area) four times higher than that of sampling the other rows (corresponding to the plain). In this way, in the end, about 80% of the nuclei are placed in the northern part of the region while only the remaining 20% in the south.

Once the nuclei are sampled, the Voronoi tessellation can be computed by simply assigning each site $x \in S_0$ to the nearest nucleus Z_i^b according to the specified distance $d(\cdot, \cdot)$.

3: Compute Representatives

Now, for each element i of the tessellation it is required to compute the representative g_i^b. We remind that, in this part of the analysis we are considering separately the two components (longitudinal and latitudinal) of the wind vector because we want to maintain a certain degree of information also under the light of the direction in which the wind blows. So that, at this step, we are actually computing two representatives for each element of the tessellation: one for component U and one for component V. The representatives g_i^b have the role of capturing the "average" behaviour of data contained in each element of the tessellation and can be computed in any way that satisfies the needs of the specific problem.

In our case study, we employed a weighted mean with weights that are functions of the distance from the nucleus. In other words, the more a site is far away from the centre it has been assigned to, the less its "importance" is in the computation of the representative, while cells that are close to the nucleus will have greater relevance. In particular the weights we adopted are computed as:

$$w_i(x) = \frac{1}{1 + d(x, Z_i^b)}$$

where $w_i(x)$ denotes the distance of site x assigned to i-th nucleus. In this way, the nucleus itself will have $w_i(Z_i^b) = 1$, the weight will be $\frac{1}{2}$ for sites with distance equal to 1 and so on.

In practice, the representatives we computed for each element of the tessellation are the two weighted averages (one for component U and one for component V) of the temporal series of data in the subset. At this point, we managed to sum up the information of the N sites composing the grid S_0 into $2n$ representatives.

4: Projection on the p-dimensional space of the representatives

At this point, the algorithm performs a step of dimensional reduction: the goal is to concentrate the information contained in the functional data that are the representatives into multivariate data. In our specific case, we simply followed the same procedure of smoothing and Functional Principal Component Analysis described in Section 5.1. First of all we performed a smoothing with Fourier basis on both the representatives of each element of the tessellation and then projected them on the 4 dimensional space of the principal components. This because we aim at applying the distance defined in formula 5.3 to perform the clustering between the elements of tessellation.

5: Clustering into K groups

To end the bagging part of the algorithm, we need to cluster the n elements of the Voronoi tessellation into K clusters. To do so, any unsupervised clustering method can work: the authors of the algorithm proposed, for example, to apply K-mean clustering but, for our purposes, we applied a hierarchical clustering procedure using (5.3) to compute the distance between each element of the tessellation. In particular, first of all, we produced the $n \times n$ distance matrix D, symmetric and with each element on the main diagonal equal to zero.

Once the matrix of distances D is computed we proceed like in a common hierarchical clustering procedure: at each iteration, find the smallest value (greater than zero) in D and identify the corresponding elements or clusters. Assign them to the same cluster and update the distance matrix D to keep into account the fact that two clusters have been merged together. In an analogous way to what was done in the Geostatistical Hierarchical Clustering, we once again adopt the max to measure the distance between clusters composed of more than an element. Indeed, when it is needed to compare the similarity between clusters, we have to choose a linkage approach, as already described at point 4 of Subsection 6.1.1, and also in this case we deemed better for our needs to use the max:

$$D(c_1, k) = \max[D(i, k), D(j, k)] \quad \forall k$$

where c_1 is the new cluster formed by the union of clusters i and j.

Notice that, after each iteration, the dimension of D decreases to reflect the fact that two elements have been merged together. These steps are repeated until the dimension of D is reduced to K and, in other words, until we are left with the chosen number of clusters.

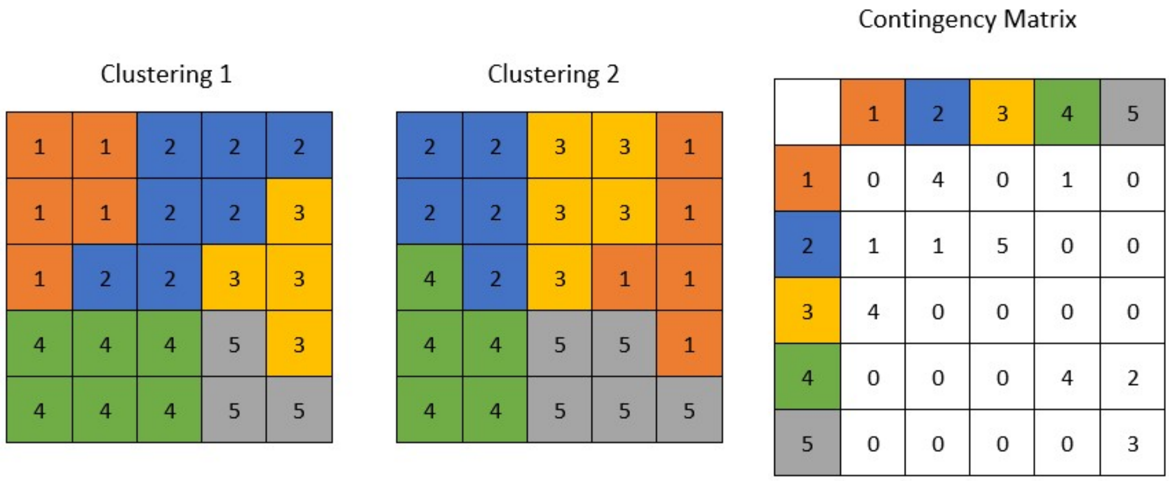
6: Cluster matching

The points from 1 to 5 represent the bagging part of the algorithm and, thus, have to be repeated B times in order to obtain, at the end, B different clustering. The fact that the centres are sampled randomly at each iteration, leads to the fact that these B resulting clustering are likely different from each other, both in the grouping produced and in the labeling. For this reason, the following steps are needed to exploit all the information extracted in the bagging part of the algorithm and produce, in the end, a single clustering.

First of all, we need to perform what is called a "cluster matching" step, to match the cluster labels across bootstrap replicates and ensure identifiability. In other words, we need that all the clustering produced in each bootstrap iteration are coherent and, thus, "agree" on which one is cluster 1, which one is cluster 2 and so on to cluster K. To do so, we first have to choose how to perform the comparisons to make the labeling coherent: one possibility is to compare subsequent clustering replicates (c_1 with c_2, then c_2 with c_3 and so on). Another possible comparison routine is to choose one replicate as reference (say the first one) and then compare each other clustering with that. Once the strategy is chosen, to compare two clustering one has to first compute a contingency matrix H. This $K \times K$ matrix measures how coherent the labeling of the two clustering considered are by counting how many times the same label is assigned to each site in both clustering and how many times instead two different labels are related to the same site. In particular:

$$H(i, j) = c(i, j)$$

where $c(i, j)$ is equal to the number of times that a site with label i in the first clustering has label j in the second clustering under consideration. This means that the diagonal elements of this matrix tell us how many times the labeling is coherent across the two replicates (see Figure 6.7 for a graphical example).



Clustering 1

1	1	2	2	2
1	1	2	2	3
1	2	2	3	3
4	4	4	5	3
4	4	4	5	5

Clustering 2

2	2	3	3	1
2	2	3	3	1
4	2	3	1	1
4	4	5	5	1
4	4	5	5	5

Contingency Matrix

	1	2	3	4	5
1	0	4	0	1	0
2	1	1	5	0	0
3	4	0	0	0	0
4	0	0	0	4	2
5	0	0	0	0	3

Figure 6.7: In this example of contingency matrix we can see that the clusters that have formed are quite similar but in some cases the labels assigned are different.

Once the contingency matrix has been computed we have to perform the relabeling in order to ensure identifiability. As shown in the example of Figure 6.7, it may happen that the cluster produced are very similar but the labeling is different in some cases. The idea behind cluster matching is to maximize the total number of times the same label is assigned to a site in both the clustering. In terms of contingency matrix, then, the goal is to minimize the sum of the elements out of the main diagonal or, in other words, maximize the sum of the elements on the diagonal. To do the job, we employed the function `solve_LSAP` in the `R` package `clue` [19] which computes the permutation of labels maximizing the sum of the elements on the diagonal of the contingency matrix.

7: Resulting clustering computation

The only thing left to do is the computation of the final clustering. To do so, we look at each cell of the grid one by one, and count how many times each label was assigned to it (after cluster matching of course) across all the replicates. In this way we want to compute the frequencies of assignment of each of the K clusters to the site along iterations

$\pi_x^k = \#\{b \in \{1, ..., B\} | x \in C_k^b\}/B \quad \forall k = 1, ..., K$, where C_k^b is the set of sites whose label is equal to k at replicate b. The most frequent label will be the one assigned to the considered site in the final clustering.

In this last phase we compute also another quantity for each site of the grid, called spatial entropy, which tells us how much "stable" or not our labeling is. This is a fact that we want to study to understand if, in any site, we have a much more frequent label with respect to others, which is the best case scenario, or all the labels are more or less equally probable, which instead is bad. The spatial entropy is computed as:

$$\eta_x^k = -\sum_{k=1}^{K} \pi_x^k \log(\pi_x^k) \tag{6.1}$$

Notice that the minimum of expression 6.1 is equal to zero and is achieved when the frequency of a single label is equal to 1 while equal to zero for all other labels; the maximum, instead, is obtained when all the labels have frequency equal to $\frac{1}{K}$ and, thus, the maximum value for the spatial entropy is equal $-\log(\frac{1}{K})$. This means that when we have little uncertainty on the label to be assigned to a site, the spatial entropy is low while when frequencies are more uniformly spread the spatial entropy reaches higher values. Since the spatial entropy is a quantity with range that varies based on the number K of clusters we choose, it is not easily interpretable, while it may be a good idea to normalize the values of spatial entropy on the interval [0,1] to get a more understandable index.

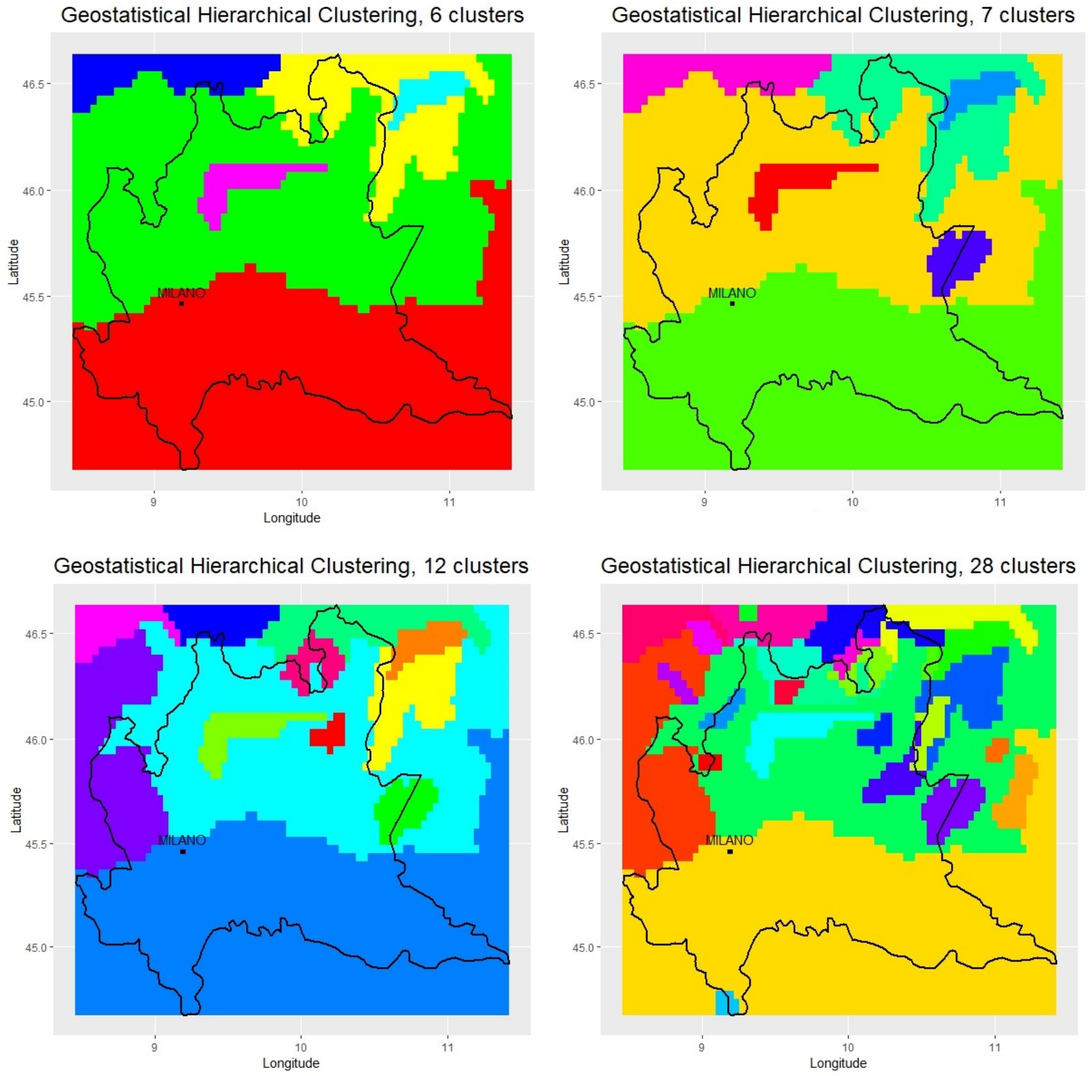
6.3. Results

In this section we will present the results obtained employing the two algorithms in the context of grouping areas with the same "meteorological regime". First, we will show the results separately and, then, compare their performance.

6.3.1. Geostatistical Hierarchical Clustering

The geostatistical hierarchical clustering relies on a simple but effective idea and this fact makes the algorithm easy to understand and quite fast in the computation. Its main merit, however, is for sure the fact that you need to run it only once and then can retrieve every result possible by just setting the desired number of clusters.

Also under the light of the results produced, the algorithm looks robust: in Figure 6.8 we show the results related to 6, 7, 12 and 28 clusters, while in Figure 6.9 we show the



Geostatistical Hierarchical Clustering, 6 clusters

Geostatistical Hierarchical Clustering, 7 clusters

Geostatistical Hierarchical Clustering, 12 clusters

Geostatistical Hierarchical Clustering, 28 clusters

Figure 6.8: Results of the geostatistical hierarchical clustering with 6, 7, 12 and 28 clusters. The higher the number of clusters, the more areas of interest are depicted in the graph.

related curve of merging distances.

There are many things to highlight concerning the results shown in Figure 6.8, the first of them being the stability of the cluster covering the entirety of the Po Valley: across all the examples shown it is needed just one cluster to group basically the southern half of sites we are considering. This is clearly in accordance with what we expected since the plain area is much more uniform than the Alps, thus the wind regimes across all the Po Valley are very similar to each other and can be grouped all together. On the other side, mountainous part of Lombardy is very heterogeneous, both in geographical characteristics, with frequent alternation of peaks and valleys, and in wind behaviour. Because of this, we see that in the northern part of Lombardy more and more groups are

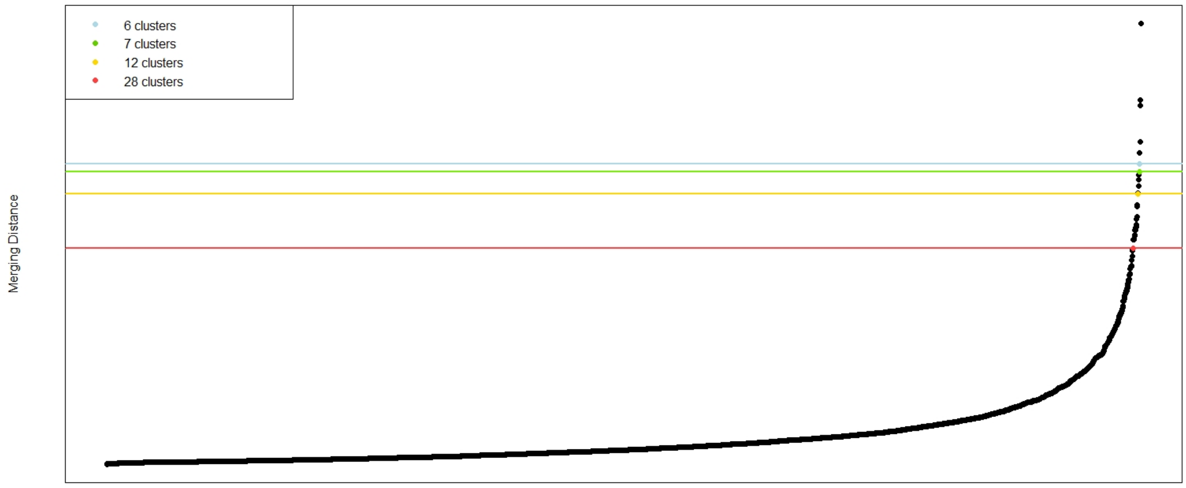


Figure 6.9: The merging distances curve relative to the application of the geostatistical hierarchical clustering to the area of Lombardy. The cases shown in Figure 6.8 are highlighted here.

represented as we increase the number of clusters. The groups that we can see in the various examples of clustering can have three different kind of nature.

The first case is the one characterized by big clusters like the one covering all the Po Valley or the big one that we can always see represented on the Alps. While this first case shows those more uniform areas, the more interesting ones are the other two, both highlighting some details of the territory.

Indeed the second case groups together areas characterized by very peculiar wind regimes; if we look at the example with 7 clusters of Figure 6.8, we can appreciate the clusters covering the shape of the Valtellina and part of the Como Lake in red and of the Garda Lake in dark blue, one with very low winds blowing mostly from East to West, and the other showing mid to high winds with very sudden daily changes in direction, a phenomenon typical of areas with large stretches of water. Another example of this case is the small lightblue cluster under the Po Valley, appearing for the first time in the case with 28 clusters and representing a small tract of Appennines.

The third case shows, again, areas with strong local characteristics but with generally little spatial extension like all the small clusters appearing in the case with 28 groups.

To prove the quality of the grouping produced, Figure 6.10 reports the (smoothed) trends of the average winds in three of the most recognisable clusters; colors are taken to match the ones relative to the 7 clusters case in Figure 6.8. For the Po valley, we can see that both U and V components are low throughout the year, with almost no seasonal difference; regarding Valtellina, we can notice once again its low latitudinal winds while, for the Garda Lake, we have more intense winds, especially during winter months.

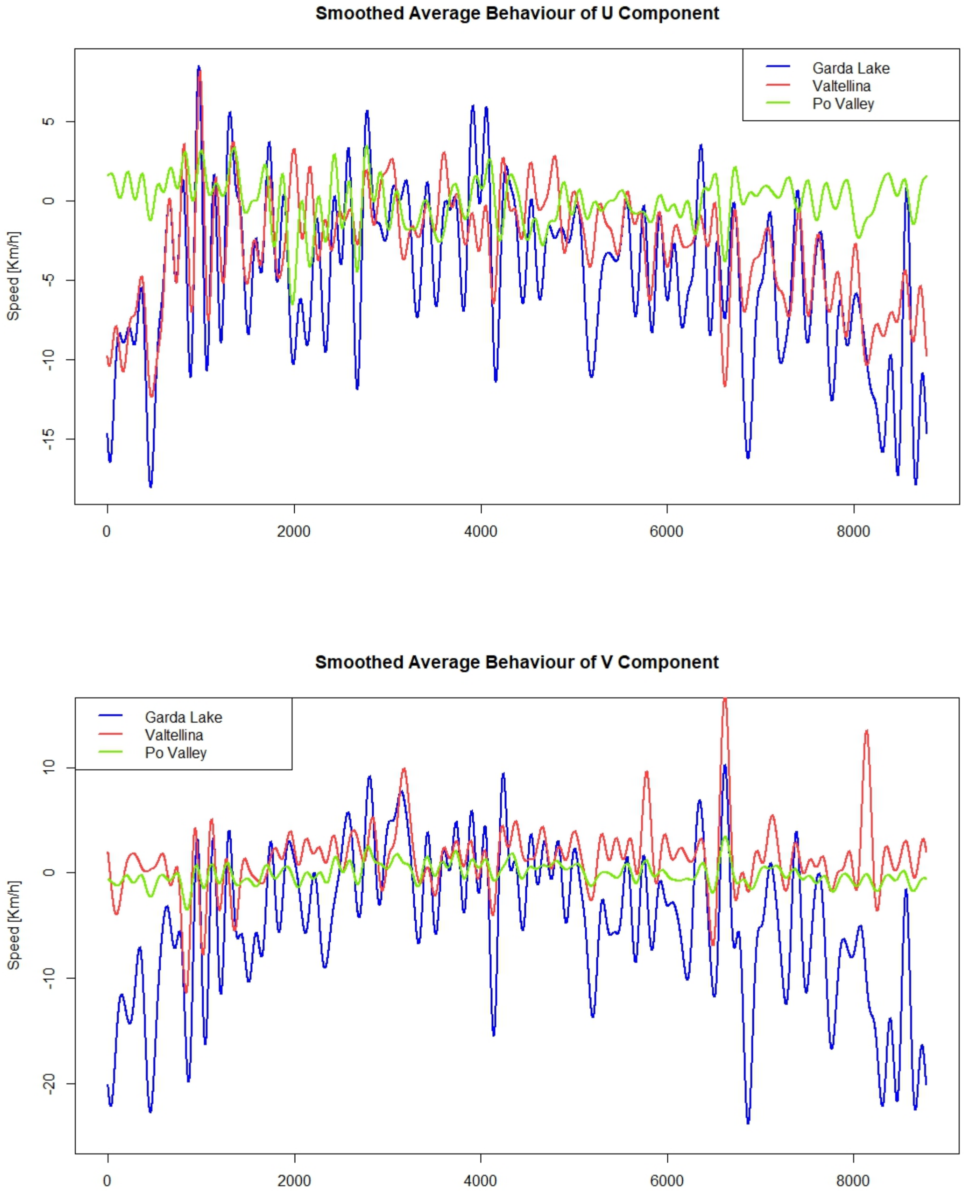


Figure 6.10: The two images show the smoothed average trends of the U and V components of wind for the clusters of Po Valley, Garda Lake and Valtellina.

This behaviour also confirms the results obtained in Chapter 5 regarding principal component analysis; indeed, here we can appreciate some examples of the interpretations of the principal components (see Figure 5.7). For instance, in correspondence of the Garda Lake, we could see strongly negative scores for the first PC of the longitudinal winds; like we said, this translates into very negative winds, especially during winter months, as we can see from the average trend for the cluster reported here.

To conclude, in Figure 6.11, we report a comparison between the results with the Geostatistical Hierarchical Clustering and the grouping produced in Section 4.3 where we analyzed the hazard linked to high winds for each cell of the grid.

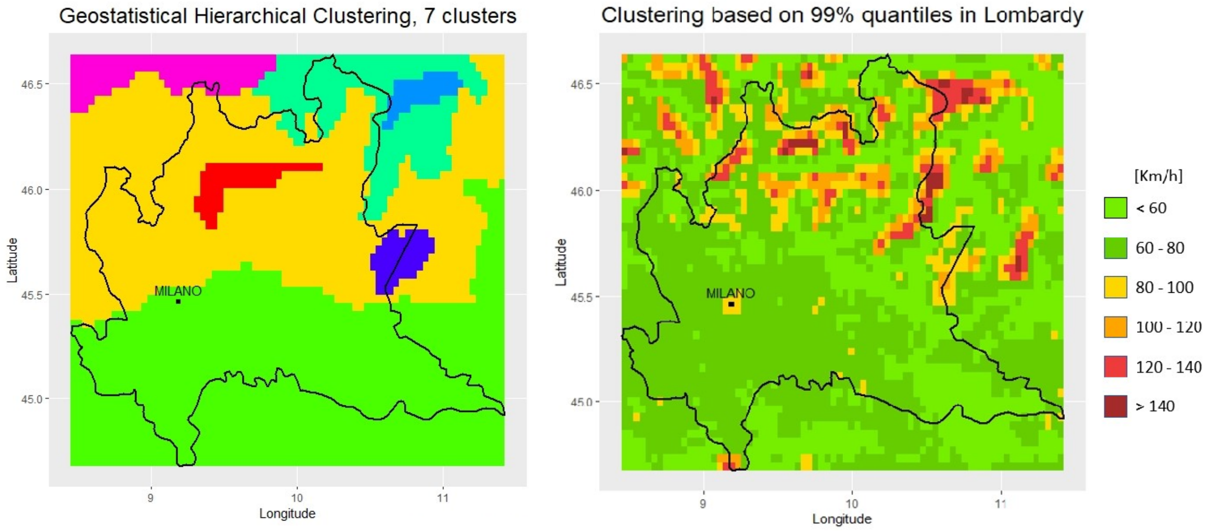

Geostatistical Hierarchical Clustering, 7 clusters
Clustering based on 99% quantiles in Lombardy
46.5
46.0
45.5
45.0
Latitude
MILANO
9
10
11
Longitude
[Km/h]
< 60
60 - 80
80 - 100
100 - 120
120 - 140
> 140

Figure 6.11: While they are very different, some areas are recognisable in both the graphs.

Of course the two graphs show very different characteristics, since they capture different kind of information (one the hazard, the other groups together areas with similar wind behaviours), however we are able to recognize, in both images, few areas of interest like Valtellina, the Garda Lake or the very high risk areas in the north-east part of the region.

6.3.2. Bagging Voronoi

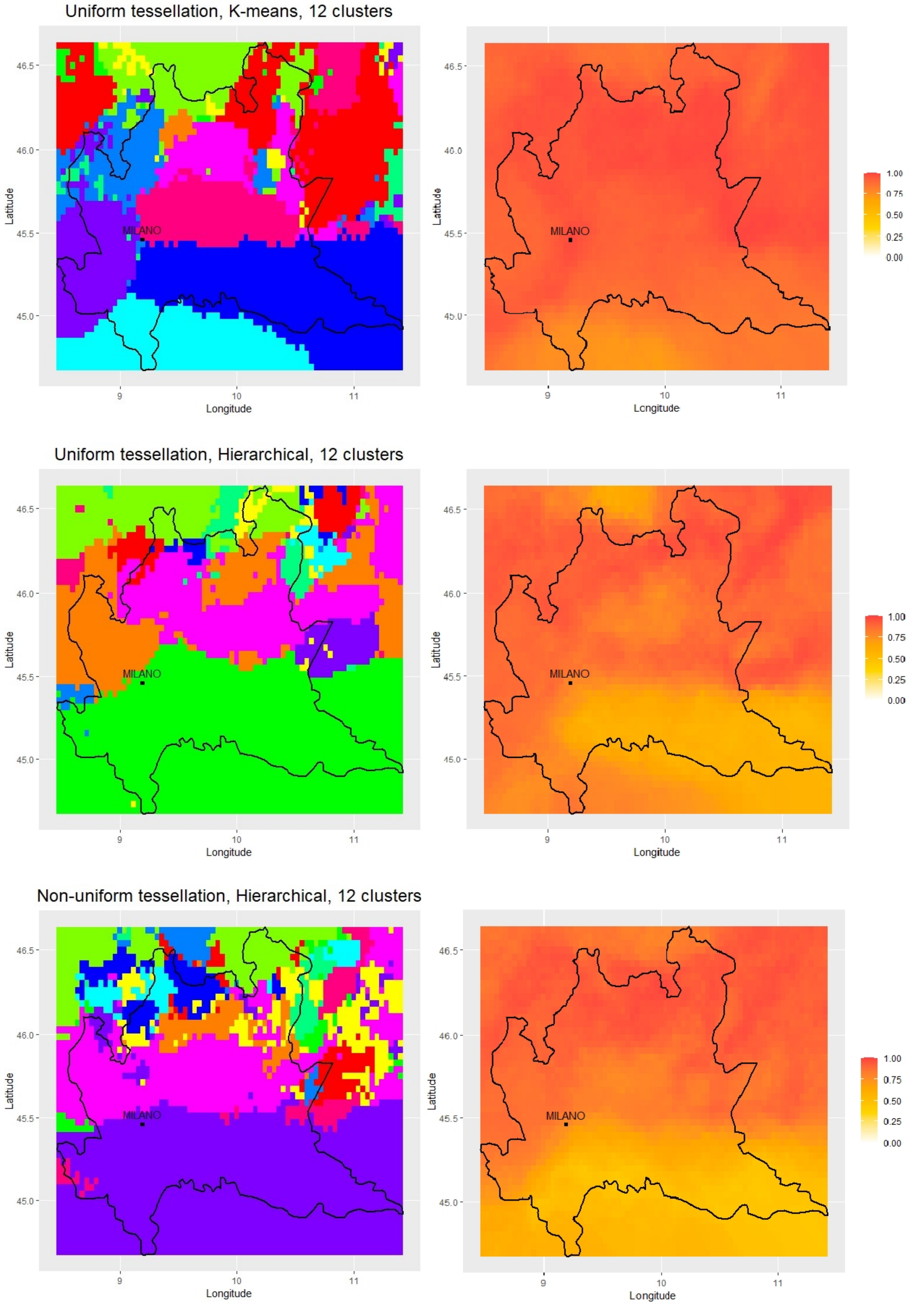
Coming to Bagging Voronoi, this algorithm has for sure many desireable features for our analysis and a quite convoluted procedure that brought us think that it would have been a promising and more sophisticated alternative to Geostatistical Hierarchical Clustering. Moreover, consisting of many steps, the Bagging Voronoi is quite open to modification of the original structure and allows the user to apply the techniques he deems more appropriate for the situation. In our specific case, we tried many ways of personalization of the algorithm, like non-uniformly distributed nuclei for the Voronoi tessellation or an unsupervised procedure that drifts from the original K-means suggested. Here we will report an example of the results obtained with each variation together with the corresponding spatial entropy graph, and then dive deeper into details of the best case. In Figure 6.12 are shown the original case, with uniform tessellation and k-means, the case with uniform tessellation and an unsupervised clustering procedure based on (5.3), and finally the case with both non-uniform tessellation and clustering with (5.3), all three with their respective graphs of normalized spatial entropy.

We can see that in all the three cases the results are not so appealing at a first look. The main issue is that the clusters are quite scattered all around the grid: in the first place this is caused by the fact that the Bagging Voronoi algorithm does not introduce anywhere the concept of adjacency and, thus, the clusters that form are allowed not to be connected. On the other side, we can see many isolated pixels in the middle of clustering to which they do not belong and that is because of the cluster matching part.

Another thing to notice is the graph of spatial entropy; as already said, the spatial entropy gives an idea of the uncertainty linked to the assignment of a label to a given cell of the grid. The ideal behaviour would be to have low spatial entropy in the middle of the clusters and high on the borders of between clusters, but in our cases we see that the spatial entropy is actually pretty high everywhere. We can appreciate, however, an improvement in the performance when we drift from the original model towards the non-uniform tessellation with hierarchical clustering case, in which at least the Po Valley shows a less uncertain profile.

Based on this last insight due to spatial entropy, confirmed by the first panel of Figure 6.13 where we confront the mean spatial entropy of the three cases, we decided to further analyze the third case shown in Figure 6.12 and hopefully retrieve better results. For this reason we performed some attempts with different values of the hyperparameter n, the number of nuclei for the Voronoi tessellation; in particular, we computed clustering for n = 100, 500, 1000, applying the non-uniform tessellation described in subsection 6.2.1, 12 clusters and hierarchical clustering based on (5.3). The results are portrayed in Figure 6.14 with their respective spatial entropy.

It is quite apparent that, by increasing the number of elements in the Voronoi tessellation, both the clustering and the spatial entropy improve. About the clustering, in the cases with higher n we lose the sharp division between mountainous and plain areas that we see with $n = 100$ but we can appreciate the presence of those features highlighted by the Geostatistical Hierarchical Clustering, like Valtellina and Garda Lake. On the side of the spatial entropy, while the case with $n = 500$ is slightly worse than $n = 100$, the case with $n = 1000$ shows a huge improvement, as highlighted in the right panel of Figure 6.13.



Uniform tessellation, K-means, 12 clusters

Uniform tessellation, Hierarchical, 12 clusters

Non-uniform tessellation, Hierarchical, 12 clusters

Figure 6.12: Examples with number of clusters K=12 for the three cases of Bagging Voronoi considered with respective normalized spatial entropy.

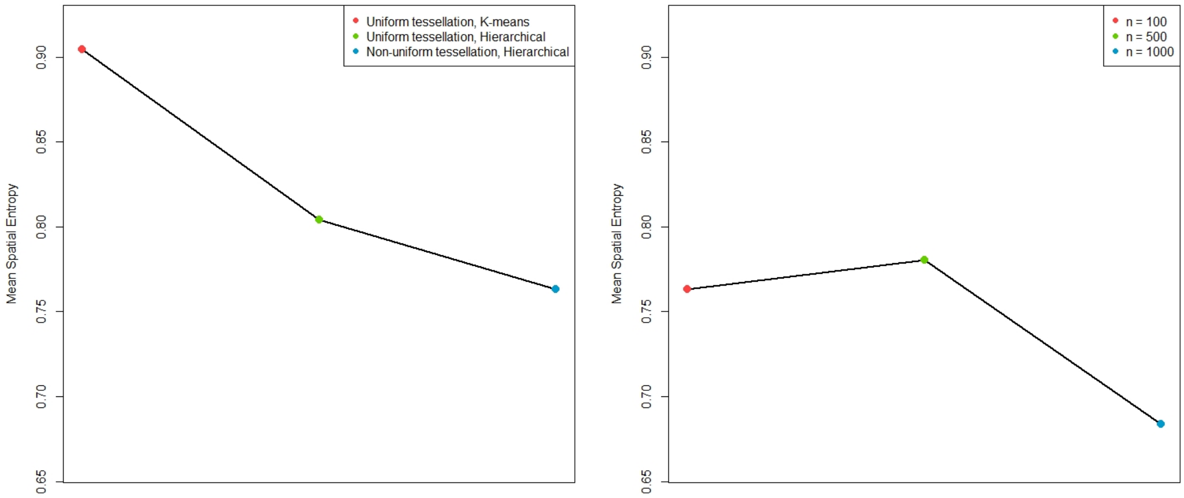


Figure 6.13: In the first graph we show the values of mean spatial entropy for the three possible cases, with the same value of n. In the second one, the case of non-uniform tessellation with hierarchical clustering for different values of n.

While this last case with a lot of elements in the tessellation shows encouraging results, we have to point out the fact that, by increasing significantly the number of starting nuclei we get closer and closer to what Geostatistical Hierarchical Clustering does, basically: we lack the adjacency concept and have the boostrap steps but, with a lot of subsets composing the tessellation, the variations obtainable in the phase of tessellation are limited and the representatives computed are not so different from the single site's time series (since each one of these subsets will be small).

6.3.3. Comparing the two algorithms

To conclude this part of analysis we want to compare the performances of the two algorithms described in this chapter. In our opinion, the Geostatistical Hierarchical Clustering is much better suited than Bagging Voronoi for this case study. The results are more convincing, more coherent and, considered the high spatial entropy shown by the Bagging Voronoi, does not depend on the initialization like the counterpart. The appealing results of the Geostatistical Hierarchical Clustering are for sure due to the constrain of adjacency, which makes totally sense since we are looking for areas with the same wind regime. The Bagging Voronoi, which instead lack this concept but relies on the bootstrap strategy, shows scattered clusters and does not seem to highlight areas of interest like the Geostatistical Clustering, unless we set a very high value for n. Moreover, the Bagging Voronoi

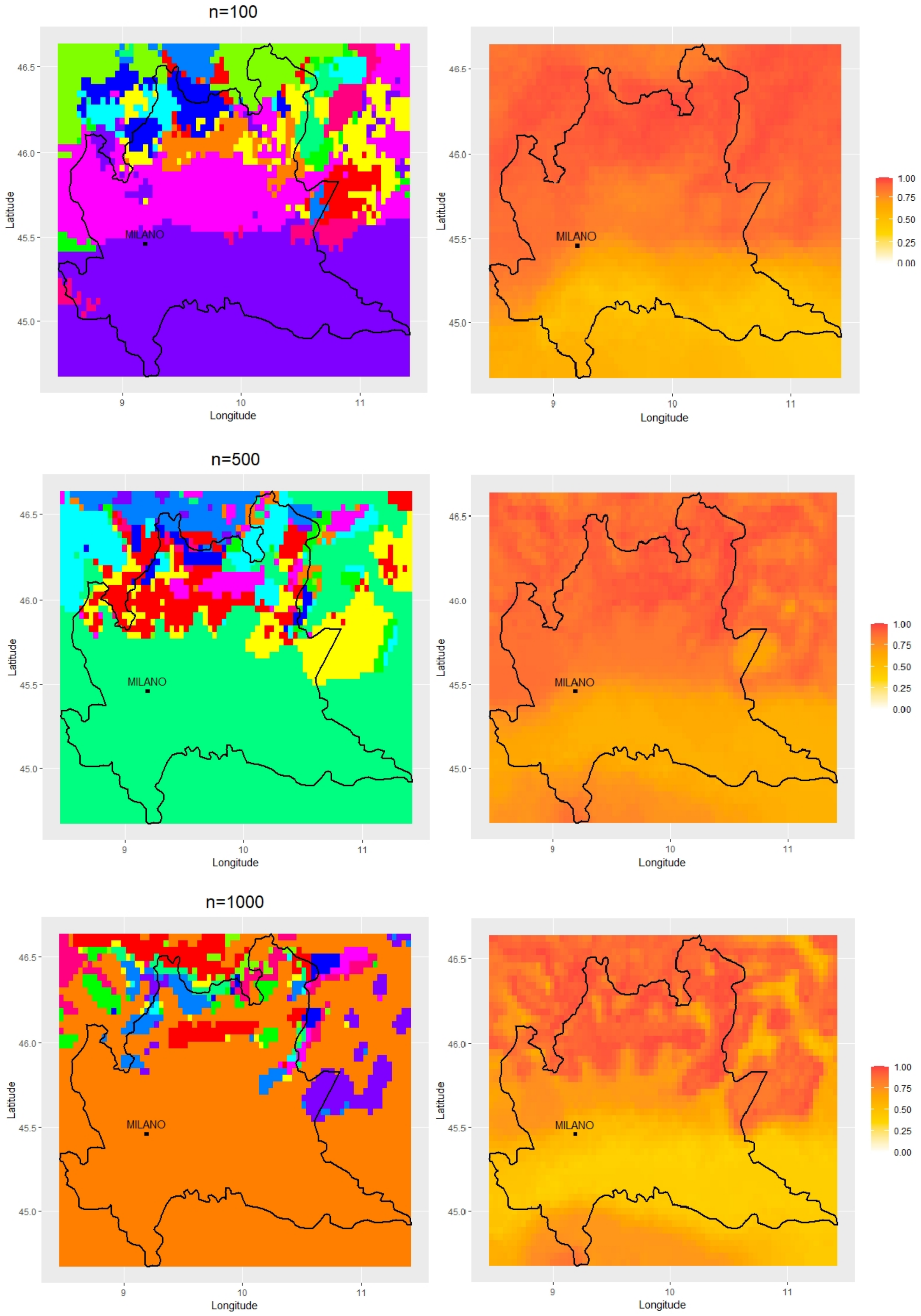


Figure 6.14: Bagging Voronoi results relative to different values of n.

is much heavier than the other one also from the point of view of code and computational costs. We estimated that the Bagging Voronoi requires, with $B = 100$ iterations in the bootstrap phase, on average, 10 to 15 times more time to finish with respect to the time needed by the Geostatistical Hierarchical Clustering. One can argue that the Bagging Voronoi computational time can be significantly decreased exploiting parallelization, and this is for sure true, but the great number of hyperparameters makes this algorithm really demanding in terms of tuning. Each time we want to change, for example, the number of clusters to be produced, we have to run the entire routine from the beginning. Conversely, the Geostatistical Hierarchical Clustering requires just a single run from which we can retrieve all the possible results we may want, thanks to its hierarchical nature.

In conclusion, for this specific purpose, we felt that the Geostatistical Hierarchical Clustering was better suited than its counterpart; the Bagging Voronoi, we deem, shows to have difficulties in performing on such heterogeneous areas and it's probably thought for application on small uniform areas or on low resolution, very large areas, while it struggle in this mid scale applications.

7 | Energy from Wind

In this final chapter we will look at the wind from a different perspective. Indeed, if up until now wind has been considered as a threat and the effort has been focused on the understanding of the risks related to it, from this point on we will consider it as a possible resource, investigating its capability of producing renewable energy.

After an introduction on wind energy in the world and in Italy, we will take a look specifically at the Lombardy region, where the possibility of producing energy from wind is notoriously almost non-existent. The major drawbacks will be presented also in light of what has emerged in past chapters and we will try to investigate the applicability of small scale wind turbines which, necessitating of lower wind speeds, may offer a solution to produce sustainable energy in this region.

Throughout the chapter, we will use notions taken mainly from "L'energia elettrica dal vento" (2017) [37], a publication, in Italian, by RSE (Ricerca sul Sistema Energetico), the same Italian research society that produced the dataset we are using.

Notice that the work done in past chapters will show to be related to this topic and many mathematical procedures can be useful also from this point of view. Take for instance the assessment of wind characteristics and modeling done in Chapter 2, knowing precisely wind features of a site is of paramount importance in determining its value from an energetic point of view. Indeed, efficiency of wind power generation is a key factor in the development of this technology; wind energy resource assessment is an essential part of the feasibility of wind farm projects and whether this assessment is reasonable or not directly impacts the cost/benefits analysis (see Shi et al. 2021 [41]). Therefore, to reduce the uncertainty of wind power estimation, it is necessary to accurately understand the distribution characteristics of wind speed and the work done in Chapter 2 will come in handy to properly estimate distributions.

Similarly, clustering algorithms presented in chapter 6 will be reconsidered to form clusters with similar wind energy productivity, highlighting their flexibility and wide range of applicability once the proper precautions have been taken.

7.1. Wind Energy in the World

Since ancient times, humankind has used wind as a resource, exploiting eolic energy for its benefits in applications like sail navigation and windmills. The importance of this resource has become even more evident in recent years, when energy demand has become a more and more pressing problem. With the continuous growth of the population and the rapid development of the global economy, energy demand is also increasing. Traditional fossil fuels (such as coal, oil, natural gas, etc.) have been widely used in almost all areas of daily life; however, fossil fuel reserves are limited and their excessive usage is leading to climate change, with the well known possibility of disastrous consequences.

For these reasons, a great effort is being made to develop sustainable alternatives aimed at environmental protection, emission reduction and social development; finding energy sources more broadly available and with less negative impact on the environment. Wind energy is one of the potential renewable energy sources that can be used for commercial purposes and many countries have started to move towards this direction to improve their energy production.

As reported by the International Energy Agency (IEA) in their last report [23] (relative to the year 2021), wind remains the leading non-hydro renewable technology, generating 1870 TWh in 2021, almost as much as all the others combined, with a growth of 17% with respect to the previous year. However, they underline the necessity to accelerate even more this growth in order to reach the objective of the Net Zero Emission by 2050 scenario, which has established the need to reach 7900 TWh of wind electricity generation by 2030 (Figure 7.1).

The leading country in the world for wind energy production is China, with over one third of the whole wind capacity installed, while Italy figures at the 11th place (data from the IRENA report of 2021 [24]). Moreover, China is also the country with the biggest growth in capacity in the last years, showing a positive leading attitude in the development of renewable energy and a commitment towards sustainability.

Wind keeps being exploited mainly by means of wind farms of big dimensions, both on shore and off shore but also the diffusion of wind turbines of small dimensions is slowly increasing.

7.1.1. Wind Energy in Italy

Italy is on a good track regarding renewable energy: objectives for production have been reached some years earlier than expected, reaching the pre-established goal of 17% of the

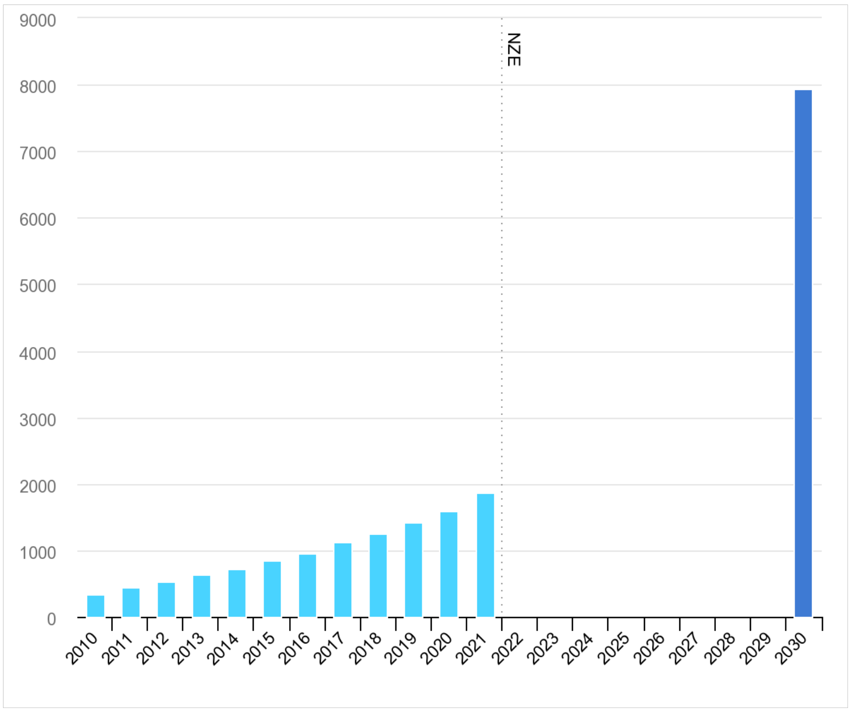


Figure 7.1: Wind power generation in the Net Zero Scenario, 2010-2030 (IEA report 2022 [23]).

total consumption in 2017 instead of 2020. In particular, wind energy was responsible for 15-20% of the total production from clean sources in the last years (see ANEV report of 2022 [3]). Now, the "Piano Nazionale Integrato Energia e Clima" (PNIEC), asks for an even bigger effort for 2030, with 30% of energy from renewable sources and, in particular, 41.5 TWh coming from wind.

To aid this effort, the Wind Atlas of Italy (ATLAEOLICO [38]) has been produced with the help of various research institutes. It offers an interactive online tool where average speeds at different heights is reported, producibility is analysed both on shore and off shore and already existing wind parks are reported.

Unsurprisingly, most of the productivity is found in the southern regions of the peninsula, where winds are, on average, stronger due to the proximity with the sea. Consequently, most of the wind turbines are found there while northern region produce little to no energy. Figure 7.2 shows data of 2022 collected by ANEV (Associazione Nazionale Energia del Vento) and relative to the production of wind energy in each region of Italy [3].

In particular, Lombardy, which has been the main subject of this work, produce basically zero wind energy. This is due, on one side, to the low average wind speeds in the region but, on the other side, also to its conformation: plain areas are densely inhabited and wind turbines cannot be placed too close to cities, while mountain areas (where also wind speeds are more promising) lack sufficient open spaces where there is enough room to build wind farms and are, again, inadequate for turbines.

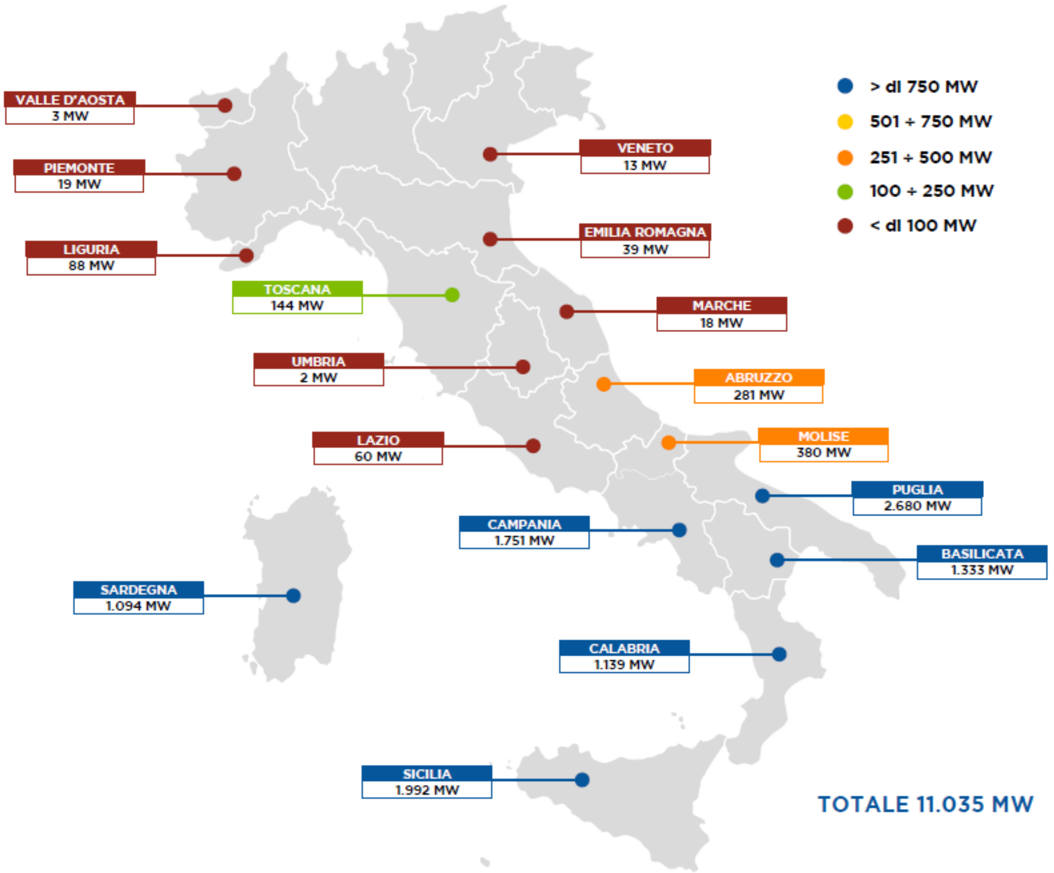
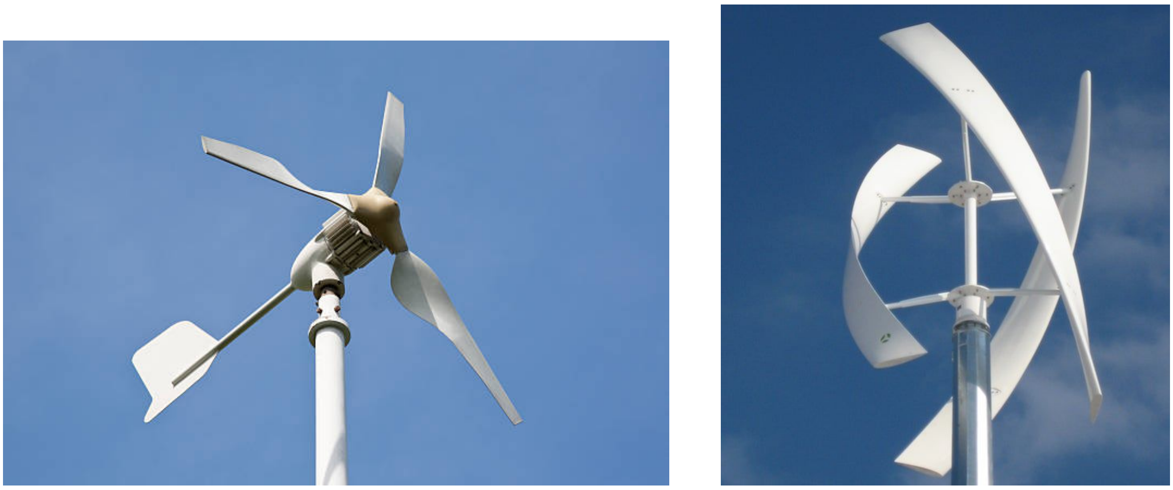


Figure 7.2: Energy power produced in each region of Italy (ANEV report 2022 [3]).

For these reasons, we wanted to try to investigate a different way of producing energy from wind in the form of small wind turbines (known as "mini eolico" in Italy). This kind of turbines require lower wind speeds to work and don't need big open spaces but can be installed potentially everywhere, making them interesting for applications in Lombardy.

7.2. Small Wind Turbines

Internationally, are classified as small wind turbines the ones up to 100 kW of power, even if, formally, according to the normative IEC 61400-2 (Design requirements for small wind turbines), are part of this category systems with swept area less than or equal to 200m^2, corresponding to a diameter of less than 16m and a power of no more than 50-60 kW.



(a) Horizontal axis (b) Vertical axis

Figure 7.3: Two examples of small wind turbines.

In the last years, technological innovation has had a key role in the commercial evolution of the sector, which has seen a real impact in the generation of energy. Indeed, recent machines have inherited components and technologies that have determined the success of bigger turbines, with important results from the point of view of reliability and performance. Importantly, in its Wind Implementing Agreement [42], IEA has set the target to develop an international standard on quality of small wind turbines in order to further help the development of the sector.

However, absolutely central in market orientation, is the role of an incentive system which can pull producer towards one kind of turbines more than another. In particular, in Italy, after some year of advantageous incentives for small wind turbines, the DM 4/7/2019 has changed the situation, reducing economical help and pivoting the operators towards the installation of bigger scale models, in general between 500 and 1000 kW.

Small wind turbines can be divided into two groups: horizontal axis (Figure 7.3a) and vertical axis (Figure 7.3b). The most commonly used turbine in today's market is the horizontal-axis wind turbine which, typically, have two or three blades that are usually made of a composite material such as fiberglass. However, also vertical axis one are starting to see some diffusion and we will take one of this kind as example to evaluate their applicability in Lombardy. Contrary to horizontal-axis turbines that can produce energy only thanks to winds blowing according to their optimal direction, the vertical-axis ones are actually independent on the direction of the wind, making them suited for many more cases.

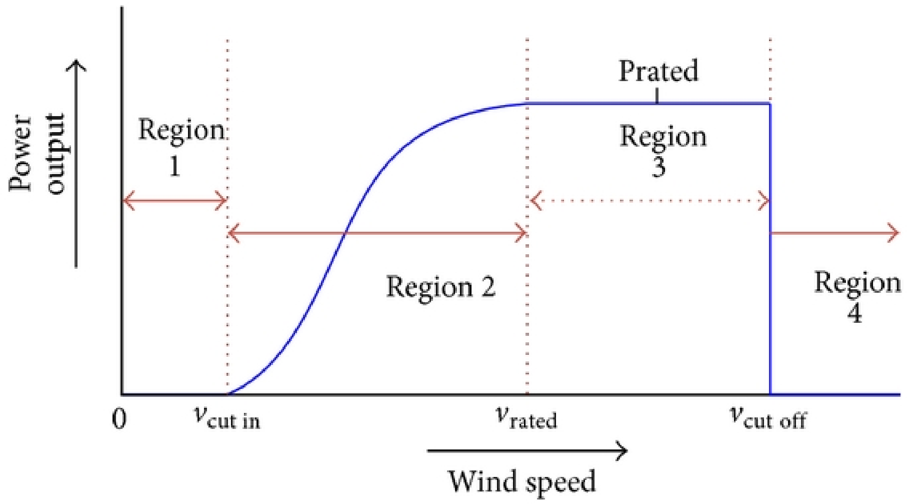


Figure 7.4: An example of power curve of a wind turbine (Sohoni et al 2016 [44]).

The great advantage of small scale turbines is their reduced requirements from the point of view of finding a suitable site for installation with respect to traditional, bigger models. While large turbines require careful planning, investigating many more factors than just wind speeds such as conformation of the territory on a large area, visual impact, ecological impact, noise evaluation, etc., small models have much less limitations and can be installed even in domestic contexts or in more remote areas, such as mountain villages, where connection to the electrical grid may be more difficult, and thus offer energetic independence. The next section will be devoted to a case study to evaluate their potential in the region.

7.3. Case Study

To evaluate the potential energy produced by a turbine we mainly need two things: the distribution of wind speeds, which we have analysed in chapter 2, and the power curve specific to the turbine (Figure 7.4).

The power curve must be made public by the producer as it summarize the fundamental features of the turbine. It shows, for each speed of the wind at rotor height, the electrical power (in kW) produced by the machine. One can notice that, to start production, it is necessary for wind speed to be higher than a threshold called cut-in, which is usually around 5 m/s for a large turbine and less for a small one. For growing speeds, power increases until reaching nominal power; at this point, in most machines, the control system regulate power emission, producing a flat stretch in the power curve. For wind speeds above the maximum (cut-out speeds), the system shut off the turbine for safety reasons and no energy is produced until wind does not go back down to lower speeds.

Once we have both the wind speed distribution $f(v)$ and the power curve $P(v)$ we can compute the theoretical producibility of the turbine. We remark theoretical, since computations are always done assuming that there are no major turbulence effects changing the power curve and that the machine is always available for production. Under these assumptions, the total energy producible by a single turbine in one year, E, can be computed as:

$$E = 8760 \int_0^\infty P(v)f(v)dv \tag{7.1}$$

where 8760 are the hours in one year (indeed, energy is measured in kWh).

7.3.1. Ecolibrì 10 kW Generator in Lombardy

For a practical example of application, we have chosen to study the energy produced by the 10kW turbine developed by the Italian manufacturer Ecolibrì and to compare results across Lombardy.

The turbine has vertical axis and an height of 10m. It occupies an area of just 100 m^2 making it ideal for domestic use or installation in industrial areas and can be easily combined in small grid systems made of multiple turbines, solar panel and batteries. The power curve of this machine is reported in Figure 7.5. The cut-in speed is 3.5 m/s but production does not start until wind speed is at least 5 m/s; cut-of speed is 15 m/s.

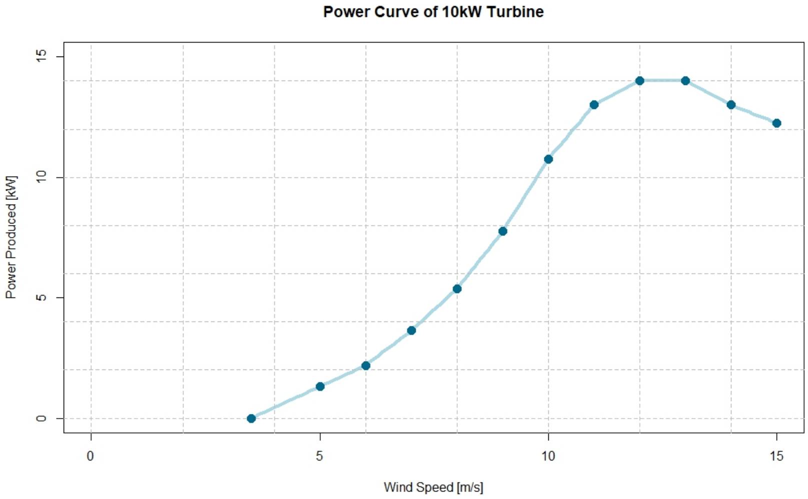


Figure 7.5: Power curve of the Ecolibrì 10 kW generator.

Annual power producible has then been computed starting from the temporal series of measured values of the wind; all 31 years available have been taken in consideration and the final result is the average of the theoretical production computed in these years. Computations have been done using a numerical approximation of equation 7.1 with linear interpolation.

Results are presented in Figure 7.6. Unsurprisingly, this graph is very similar to the ones produced in chapter 4, and it shows again that mountain areas, with higher winds, also have higher capability of producing wind energy.

To evaluate performance, the index of interest is the capacity factor, a parameter representing the fraction of energy generated with respect to the one producible if the turbine would have worked at nominal power for all the 8760 hours of the year. For the Ecolibrì generator, this value is 87.6 MWh since, indeed, its nominal power is 10 kW.

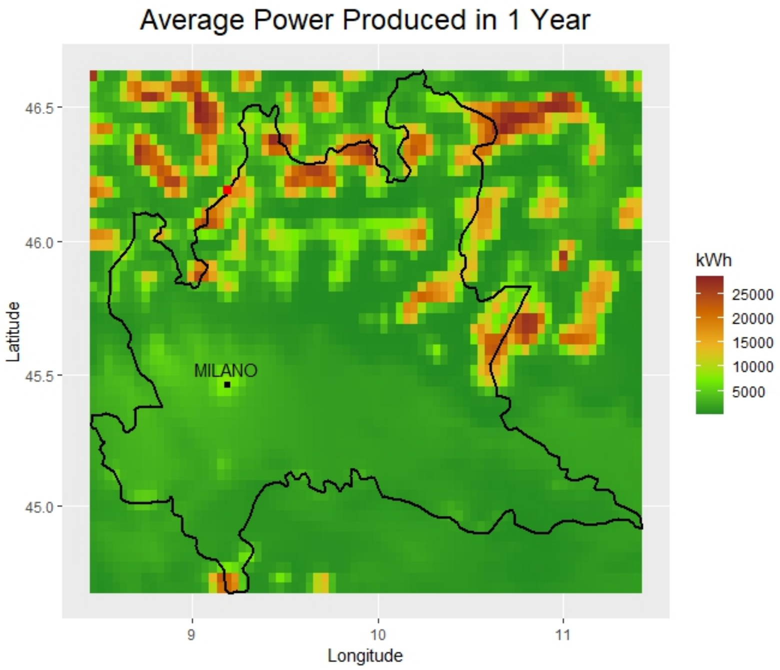


Figure 7.6: Average power produced in Lombardy.

Results are not particularly excellent: the highest capacity factor in the whole area of interest is 33.23%, which is in line with values for on shore wind farm according to IEA measures [22]. However, only 14% of locations achieve a capacity factor of at least 10% (i.e. they can produce at least 8.76 MWh) and most of them are located in mountain.

7.3.2. Comparison of Results Obtained with Different Estimation Methods

An interesting observation can be made on this topic regarding the choice of the best approximation method.

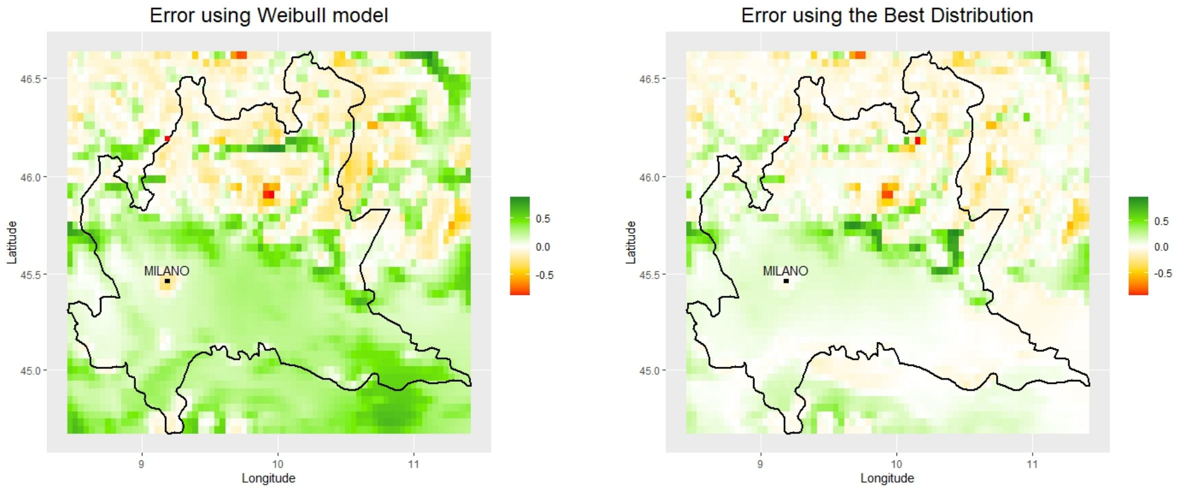
Indeed, for practical reasons, in most productivity analysis for wind energy, estimation of wind speeds distribution is done using directly a Weibull model (see Celik 2003 [11], Bilang et al. 2021 [6], Zagubien et al. 2022 [54]) or, in some cases, even a Rayleigh one, which corresponds to a Weibull with shape $k = 2$ and, thus, has even less flexibility. This is done mainly because a sufficiently long series of in-loco observations is not available and some kind of approximation must be done. As it turns out from our analysis, however, this procedure may be imprecise and return results that overestimate or underestimate production possibilities of a site.

To compare precision in the estimations, we established, as a point of reference, the results obtained in the previous subsection, i.e. using measurements directly from the whole temporal series and then averaging by year. Then, this was confronted with values of produced energy obtained starting from an approximation of the wind speed distribution made using a Weibull model and parameters estimated with the method of moments (see Chapter 2).

Results are in Figure 7.7a, where the percentage differences between the two methods have been reported. The mean error across the whole region, in absolute value, is around 20.34% but this number is greatly inflated by those locations where production is particularly scarce (sometimes as low as few hundreds of kWh) and thus percentage variation is greater. Still, if we consider only sites with at least 10% capacity factor, the mean error of the Weibull approximation method is 10.35%.

However, as the careful reader would remember, always in chapter 2, we determined the best approximating distribution for each site between GEV, Weibull, Gamma and Lognormal distributions and the Weibull model had this title only on a minority of them (see figure 2.4b). For this reason, here, we tried to approximate wind speeds using the probability function of the best fitting distribution and then we repeated computations as before to obtain producibility.

Figure 7.7b reports the results of this new procedure and shows that choosing a more appropriate distribution to model data translates also in a more accurate estimation of the energy producible. Indeed, now, the mean error on the whole region is 9.76% and the one computed only on the most potentially productive sites is down to just 3.79%.



(a) Weibull model

(b) Best fitting distribution

Figure 7.7: Percentage error with respect to the value computed from the temporal series.

These results underline once again two important points. On one side, they highlight the need of data; as in every other field, more data means better models and approximations, while having few of them leads to great uncertainties in the results. On the other hand, they work as a real and practical example for the importance of proper mathematical analysis and estimations in the process of developing new power production plants and, in general, in the installation of wind turbines. A too large approximation in any part of the cost-benefit estimation procedure may lead to wrong decisions with potential economic repercussions and it is fundamental that, at every stage, the most accurate strategy (in relation to the available data) is followed.

In particular, of great importance is the correct estimation of wind speed distributions and model parameters, a field already flourishing with researches which should be considered more in practical applications. Indeed, our work has studied and compared some estimation methods which have shown to achieve better results and many more are available in the literature but, as it has always been, translating research results into common practice for applications require time.

7.3.3. Clustering based on Potential Productivity

When we think about how much wind energy a specific site would be able to produce the first thing to observe is for sure how strong the wind blows there and this translates in observing the wind annual distribution. Although the perspectives of producing energy thanks to the wind in Lombardy does not look very promising for the time being, the scientific development in the field of renewable energy production is proceeding at very high rate, and we don't have to exclude that, in the future, there will be technologies better suited also for Lombardy.

For this reason we want to produce a new clustering, based on the information brought by the pdf and that aims at grouping together those areas with the same potential at producing wind energy. To this end, we thought that, conversely from what seen previously in Chapter 6, the idea behind Bagging Voronoi algorithm could work fairly well. So we decided not to use the algorithm as it was presented, but to take inspiration from it, maintaining the bootstrap approach and simplifying the clustering phase; to do so, in particular, we did not computed functional representatives to later project them on a p-dimensional space, but we decided to directly build the representative of each element of the tessellation by means of the densities computed with the function `hist` on `R`. In practice, we set $M = \max(\text{winds})$ (i.e. the maximum value recorded over the entire grid), divided the interval $[0, M]$ into bins 0.1 km/h wide and, then, for each element of the tessellation, saved the term `density` of function `hist` which measures the fraction of data falling into each bin. Basically, to measure the distance between two elements of the tessellation we measure the l^1 distance between the vectors containing the computed densities. An example of this comparison between histograms can be found in Figure 7.8.

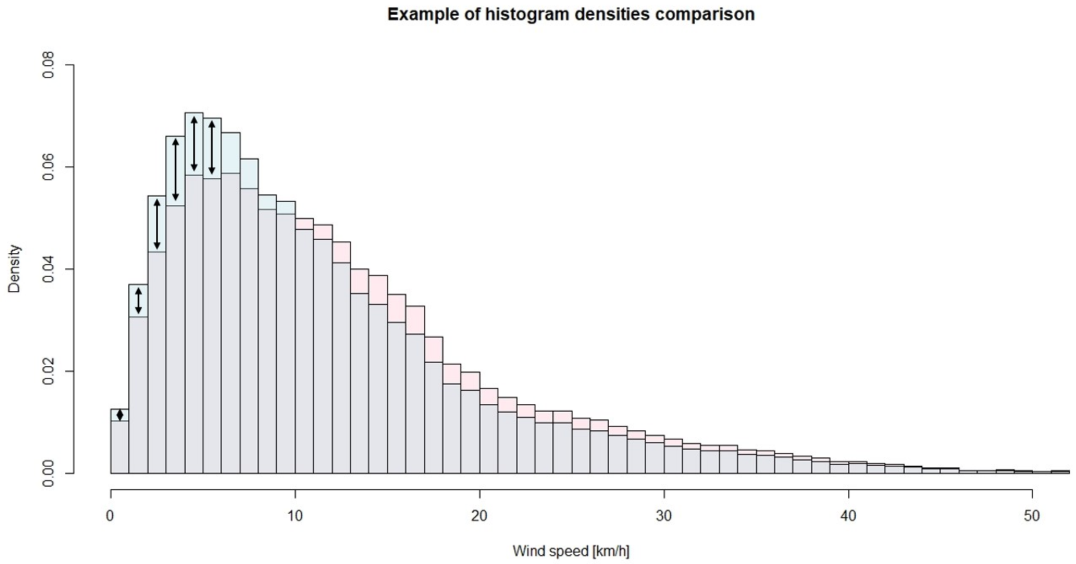


Figure 7.8: In this example we compare the histogram densities of two elements of the tessellation. Here the bins are 1 km/h wide for the sake of interpretability. The distance between two elements of the tessellation is the sum of all the differences between columns of the histograms.

Thanks to this slight simplification in code, we managed to reduce dramatically the computational load and speed up by at least 10 times the algorithm and, consequently, we could afford computing many tries to better tune our hyperparameters.

Starting from the number of clusters K, we didn't want to explore those cases with high number of groups because we are not interested in discriminating too many cases. In particular, we compared the cases with 3, 4, 5 and 6 clusters, for which we report the results in Figure 7.9.

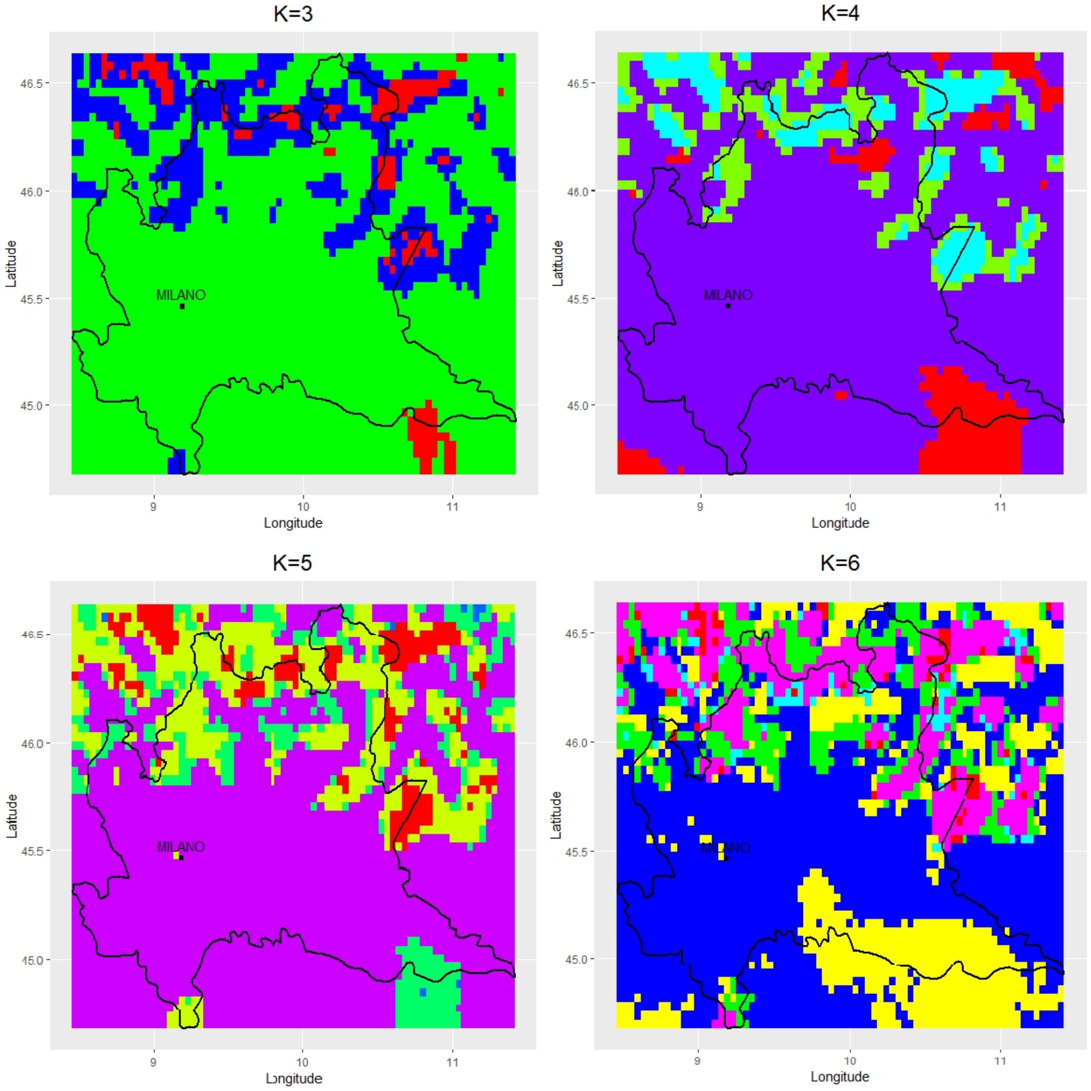


Figure 7.9: 4 examples of clustering, using K = 3, 4, 5 and 6.

By looking at this graphs, we notice that in the cases of $K = 5$ and $K = 6$, respectively 1 and 2 clusters are extremely marginal and actually hard to find on the grid. This means that their role is negligible and, thus, more than 4 clusters are redundant. On the other side, while both 3 and 4 clusters cases look good, if we look at the data that are clustered together in the case of $K = 3$, we can see that the red group contains in itself sites with very different characteristics (see Figure 7.10).

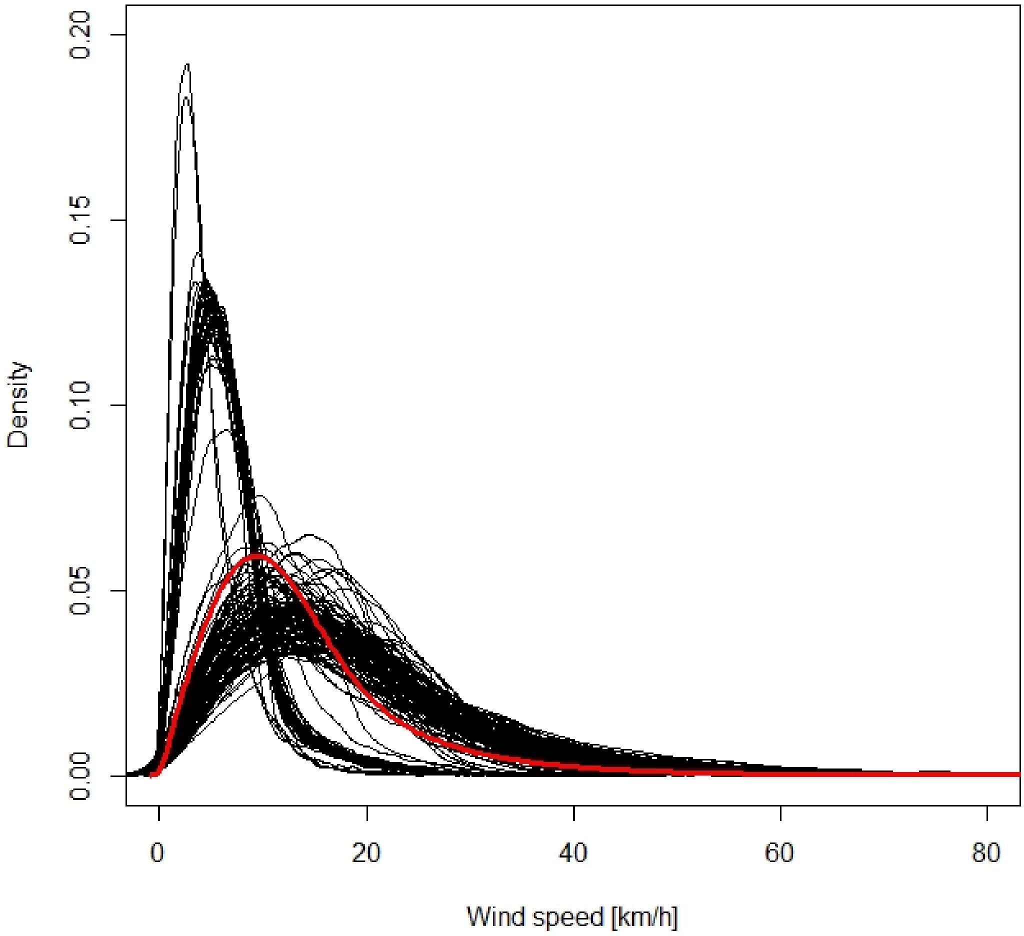


Figure 7.10: In black the pdf of data in the cluster, in red the average curve. It is clear that the data within this group show at least two very different behaviour.

For this reason we opted for $K = 4$ and analyzed deeper this case by tuning the optimal value for the number of elements of the Voronoi Tessellation n.

To do so, we tested the performance of the simplified version of Bagging Voronoi described before with many different values of n through the observation of the spatial entropy; in particular, for each n we computed the average spatial entropy computed over the entire grid. The values we considered in the first place for n are 100, 200, 400, 700, 1000 and 1500:

remember that the total number of sites in our grid is 3700 and that we are employing the same non-uniform tessellation described in subsection 6.2.1, which translates in the fact that about 80% of nuclei are placed on the mountainous area and only 20% on the Po Valley. Since we noticed that both $n = 400$ and $n = 1000$ gave similarly good results, we explored also the cases around this two values: 350 and 450, and 900 and 1100. The average spatial entropy is displayed in Figure 7.11 for each of the said values of n.

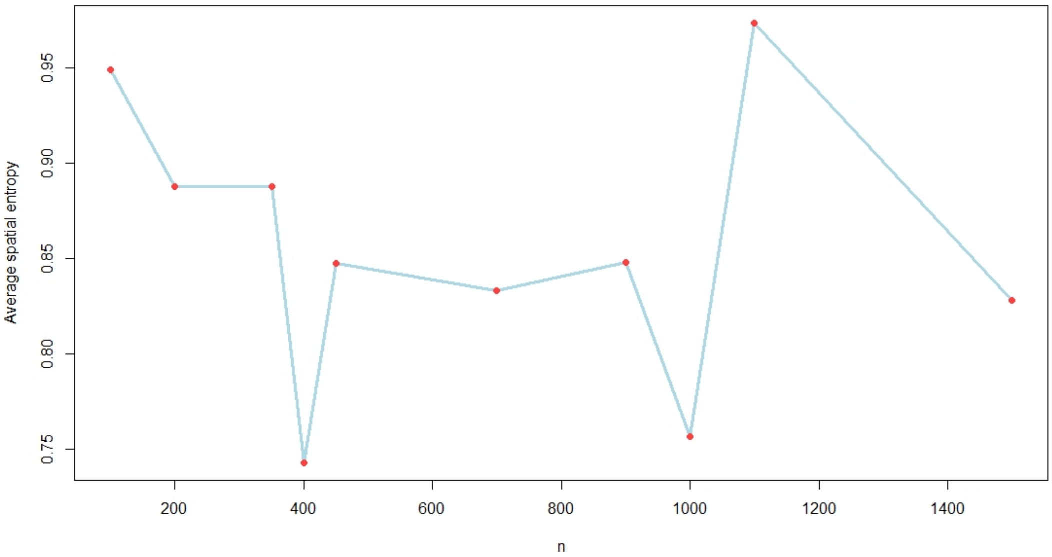


Figure 7.11: The graph shows the trend of the average spatial entropy as the value of n varies. As we can see, $n = 400$ and $n = 1000$ show the best performances among all.

Since $n = 400$ and $n = 1000$ confirm to be the two best performing cases, we now report their respective clustering results and spatial entropy on the grid. As we can see in Figure 7.12, both the cases achieve satisfying results both from the point of view of clustering and of spatial entropy. Starting from the clustering we confirm what already noticed: the major part of Lombardy is not so windy and, thus, at the moment, not really suitable for the production of energy thanks to the wind. But on the mountainous areas or on the Garda lake, as we said, the situation is much more promising and the clustering manages to capture this behaviour. In particular, the case with $n = 1000$, thanks to the higher "spatial resolution" given by the larger number of elements in the starting Voronoi tessellation, manages to portray also those small isolated areas of higher winds like the Appennines in the south and the surrounding of Milan.

Coming to the analysis of spatial entropy, again, both the graphs show good looking behaviours, as we can find that desirable pattern for which the spatial entropy is actually pretty high just where we find the boundary between two clusters. This phenomenon can be seen especially in the areas corresponding to Alps, where clusters alternates rapidly and, once again, it is more pronounced in the case with $n = 1000$. So, while the case with higher number of nuclei has a slightly larger average spatial entropy, the results it achieves look actually better.

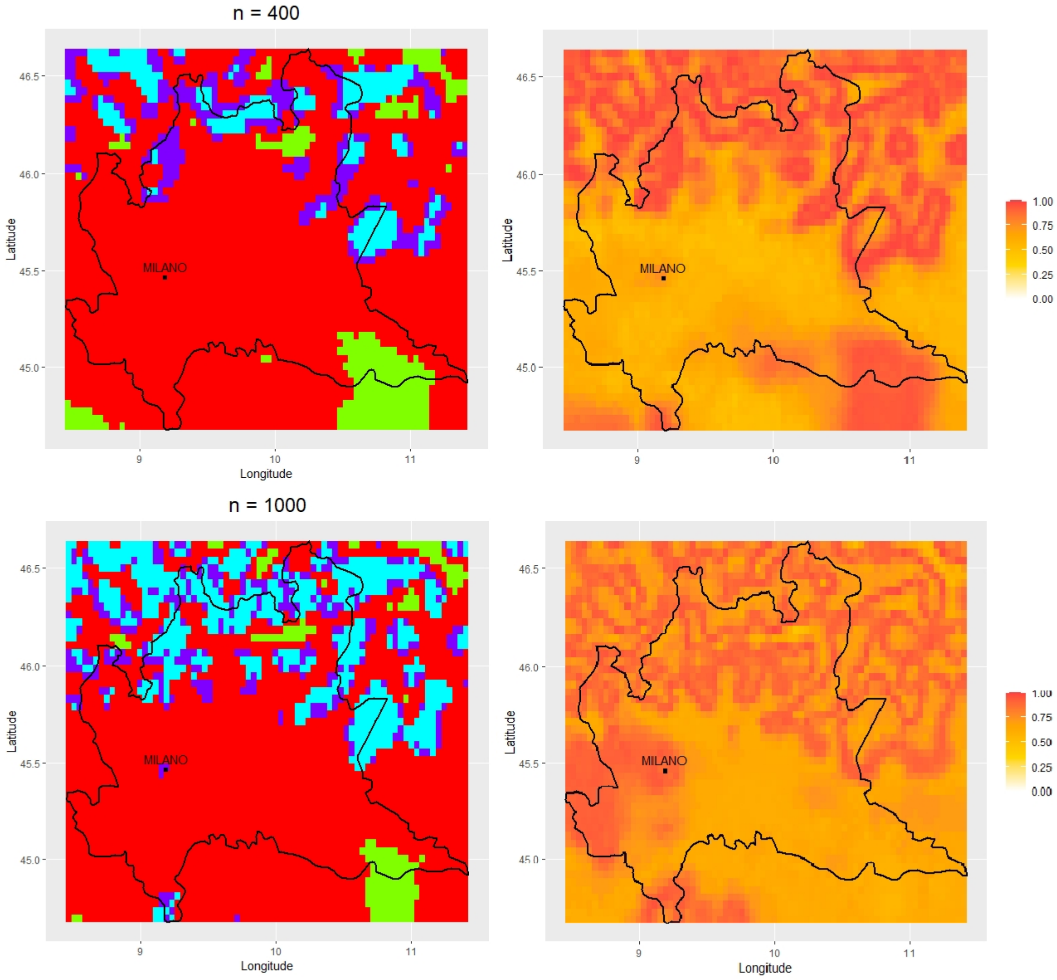


Figure 7.12: Results with the related spatial entropy for the cases with $n = 400$ and $n = 1000$.

To sum up, in Table 7.1 and Table 7.2 we report, for the two cases $n = 400$ and $n = 1000$ and for each cluster, the average number of hours during which the wind blew with speeds in the interval of production of Ecolibrì 10kW and the average energy that could have been produced in a year by a single turbine, alongside the numerosity of each one of the clusters.

	Red	Green	Purple	Cyan
Number of sites	2830	290	327	253
Time Useful [hours]	612	257	2224	3363
Energy Producible [kWh]	2140	734	10264	18006

Table 7.1: Results for energy clustering, $n = 400$. The values of Time Useful and Energy Producible are computed as the average computed over all the sites of the same cluster.

	Red	Green	Purple	Cyan
Number of sites	2596	165	317	622
Time Useful [hours]	464	175	1750	2566
Energy Producible [kWh]	1505	440	7409	12624

Table 7.2: Results for energy clustering, $n = 1000$. The values of Time Useful and Energy Producible are computed as the average computed over all the sites of the same cluster.

As we can notice in these two tables and in the previous graphs, the case with $n = 1000$ has a much bigger cluster (the cyan one) of high productivity and because of this the average energy producible is lower, since it contains also sites with lower winds. In general, the result with $n = 400$ highlights smaller area with higher average productivity while the $n = 1000$ clustering shows that the number of sites with a good prospective in wind energy production is actually larger.

7.4. Final Remarks

All things considered, the situation in Lombardy is not as bad as we were expecting. According to ARERA (Autorità di Regolazione per Energia Reti e Ambiente) [4], on average, a family composed of 3 people consumes 2700 kWh per year, with a total expense of almost 1000 euros per year only to buy electricity. If we look closer to our results, we can see that slightly more than 30% of the sites of the grid could actually produce this amount of energy thanks to the employment of an Ecolibrì wind turbine. This is a very encouraging result: 1/3 of the territory is already eligible, thanks to this technologies, to produce domestically the electric energy needed to satisfy the household needs, and an even larger percentage will when new and more efficient turbines will be developed.

However, considering that most of the population lives in urban areas where wind speeds are low and installation of a turbine is difficult for reasons of space, the possibility of wind energy as a primary source of power seems still a distant future. Probably, at the moment, these application are a reasonable investment only for more isolated locations in high wind sites, for instance mountain farms or small villages, which are ideal in term of requirements and where the installation of autonomous source of power can help achieve energetic independence.

Finally, we remark that before installing a wind turbine, one needs to conduct an accurate analysis of the site at small scale, to confirm that there are no local impediments and obstructions to the wind flow.

8 | Conclusions

With this work we were able to perform a throughout analysis of winds in Lombardy, underlying the dual nature of this phenomenon and studying the mathematical and statistical procedures to better understand and analyse it. During the whole process we looked at wind from many points of view, asking ourselves different questions and trying to find an answer to each of them.

In the first place, regarding the statistical modeling of wind speeds, we confirmed that always using the Weibull distribution regardless of the case in exam, as commonly done in the state of the art, may lead to substantial approximation errors and, in many cases, one can find distributions that better fit the real measurements. Indeed, in locations with generally low wind speeds, we have found out that the Generalized Extreme Value distribution achieves better results and, even with higher speeds, the Gamma and Lognormal models may outperform the Weibull. A practical example of this fact has been displayed under the light of energy production, where we showed that estimates obtained using the best approximating distribution were more accurate, achieving an average error of less than half the one of the Weibull.

Then, we quantified the hazard level in the whole region, focusing in particular on the threat that wind represents for the electrical infrastructure. As it turns out, the majority of the region is subject to low risk of extreme events; in particular, the plain area, where most of the population lives and most of the infrastructures are situated, has consistently low winds, associated with a low hazard level. Higher hazard can be, instead, found in mountain area where, on the other hand, there are also less people and infrastructures. However, here, the very same high wind speeds that can cause disruption to the power grid, can be used to produce renewable energy: on one side we have higher risk for the electrical infrastructure, while, on the other these, high winds translate into an opportunity to generate power and a higher value for the inhabitants of these areas to produce energy in an autonomous way.

In light of a characterization of wind regimes, we managed to produce clusters that group areas of Lombardy that are subject to similar wind phenomena. In particular, we compared two different clustering methods and analysed their respective strengths and weaknesses, finding out, in the end, that the Geostatistical Hierarchical Clustering algorithm was the better suited one. The good quality of the results produced is confirmed by two aspects: on one side, we observed directly the characteristics of the average wind regime of each cluster, finding out that they were clearly different one from another. On the other hand, the clusters often matched well defined geographical areas, and the trends shown were coherent with what we could anticipate knowing the general features of the area.

Noticeably, the area of Milan, which had peculiar characteristics from the point of view of hazard, under the light of wind regime, did not constitute a separate cluster but was grouped with surrounding areas. Indeed, while the average wind in Milan is higher, the trend is pretty similar to the rest of plain and, thus, needs to be placed in the same cluster.

Finally, regarding wind energy, previous studies have largely demonstrated that Lombardy is not suitable for the installation of "traditional" wind farms of big dimensions; for this reason we explored the applicability of the rising technology of small wind turbines. We were able to perform a large scale study covering the whole region, following the standard procedure to estimate energy production but exploiting methods and results obtained in previous chapters. Also in this case, Lombardy shows to be divided in two parts: on one side, plain areas are not particularly suitable for wind energy production, both because of lower wind speeds and also because dense urbanization imposes more constraints on turbine installation, although this technology can still be applied. On the other hand, mountainous areas are much more promising and the application of this rising technology can help remote sites solving the problems of energy supply. In general, we can say that the application of small wind turbines, although the favorable cases are still limited, can be very helpful in some areas.

In conclusion, throughout this thesis, we worked on Lombardy to have a practical case study but we remark that all the methods applied and all the analysis conducted can be extended to whichever area one can be interested in. In particular, our research demonstrated the validity of the techniques exploited in studying the dual nature of wind, allowing for broader applications beyond Lombardy and paving the way for other analysis on the topic.

Bibliography

[1] J. Aldrich. R. A. Fisher and the making of maximum likelihood 1912-1922. *Statistical Science*, 12:162–176, 08 1997. doi: https://doi.org/10.1214/ss/1030037906.

[2] M. Alrashidi, S. Rahman, and M. Pipattanasomporn. Metaheuristic optimization algorithms to estimate statistical distribution parameters for characterizing wind speeds. *Renewable Energy*, 149:664–681, 04 2020. doi: 10.1016/j.renene.2019.12.048.

[3] ANEV. Anev brochure 2022, 2022.

[4] ARERA. Stima spesa annua - elettricità, 2022. URL `https://www.arera.it/it/elettricita/stimaspesa_ele.htm#`.

[5] M. Ben Alaya, F. Zwiers, and X. Zhang. On estimating long period wind speed return levels from annual maxima. *Weather and Climate Extremes*, 34:100388, 2021. ISSN 2212-0947. doi: https://doi.org/10.1016/j.wace.2021.100388. URL `https://www.sciencedirect.com/science/article/pii/S2212094721000785`.

[6] R. G. J. Bilang, J. Olalo, C. Dela Cruz, R. Bonifacio, and E. Paringit. Determination of a potential for the installation of small-scale wind turbine in barangay bagasbas, daet, camarines norte, philippines. *ASEAN Engineering Journal*, 12:17–26, 08 2021. doi: 10.11113/aej.v12.16503.

[7] R. Bonanno, M. Lacavalla, and S. Sperati. A new high-resolution meteorological reanalysis italian dataset: Merida. *Quarterly Journal of the Royal Meteorological Society*, 145:1756–1779, 03 2019. doi: 10.1002/qj.3530.

[8] R. Bonanno, M. Lacavalla, and S. Sperati. Merida: dataset description and main variables, 5 2019.

[9] R. Bonanno, M. Lacavalla, and S. Sperati. Merida hres: dataset description and main variables, 12 2020.

[10] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer New York, NY, 2010. ISBN 978-0-387-70913-0. doi: https://doi.org/10.1007/978-0-387-70914-7.

[11] A. Celik. Energy output estimation for small-scale wind power generators using Weibull-representative wind data. *Journal of Wind Engineering and Industrial Aerodynamics*, 91:693–707, 04 2003. doi: 10.1016/S0167-6105(02)00471-3.

[12] K.-S. Chan and B. Ripley. *TSA: Time Series Analysis*, 2022. URL `https://CRAN.R-project.org/package=TSA`. R package version 1.3.1.

[13] P. Chebyshev. Sur deux théorèmes relatifs aux probabilités. *Acta Math.*, 14:305–315, 1890. doi: https://doi.org/10.1007/BF02413327.

[14] S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer, 2001.

[15] R. Fisher. "the coefficient of racial likeness" and the future of craniometry. *Journal of the Anthropological Institute of Great Britain and Ireland*, 66:57–63, 1936.

[16] J. Greenwood, J. Landwehr, N. Matalas, and J. Wallis. Probability weighted moments: Definition and relation to parameters of several distributions expressable in inverse form. *Water Resources Research*, pages 1049–1054, 05 1979. doi: 10.1029/WR015i005p01049.

[17] F. Guglielmini. Vento forte a Milano, è normale? Luca Mercalli: "Ecco perchè è successo". *Corriere della sera*, 2022. URL `https://milano.corriere.it/notizie/cronaca/22_febbraio_08/vento-forte-milano-normale-ecco-perche-successo`.

[18] J. Holmes and W. Moriarty. Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, 83(1):1–10, 1999. ISSN 0167-6105. doi: https://doi.org/10.1016/S0167-6105(99)00056-2. URL `https://www.sciencedirect.com/science/article/pii/S0167610599000562`.

[19] K. Hornik. *clue: Cluster Ensembles*, 2023. URL `https://CRAN.R-project.org/package=clue`. R package version 0.3-64.

[20] J. Hosking, J. Wallis, and E. Wood. Estimation of the generalized extreme value distribution by the method of probability weighted moments. *Technometrics*, 27: 251–261, 08 1985.

[21] E. Hòlm, R. Forbes, S. Lang, L. Magnusson, and S. Malardel. New model cycle brings higher resolution. *ECMWF Newsletter*, 147:14–19, 04 2016.

[22] IEA. Offshore wind outlook 2019, 2019. URL `https://www.iea.org/reports/offshore-wind-outlook-2019`.

[23] IEA. Wind electricity, 2022. URL `https://www.iea.org/reports/wind-electricity`.

[24] IRENA. Renewable energy statistics 2022, 2022.

[25] B. W. S. J. O. Ramsay. *Functional Data Analysis*. Springer New York, NY, 2005. ISBN 978-0-387-40080-8. doi: https://doi.org/10.1007/b98888.

[26] S. G. James Ramsay, Giles Hooker. *Functional Data Analysis with R and MATLAB*. Springer New York, NY, 2009. ISBN 978-0-387-98184-0. doi: https://doi.org/10.1007/978-0-387-98185-7.

[27] H. Jiang, J. Wang, J. Wu, and W. Geng. Comparison of numerical methods and metaheuristic optimization algorithms for estimating parameters for wind energy potential assessment in low wind regions. *Renewable and Sustainable Energy Reviews*, 69:1199–1217, 12 2016. doi: 10.1016/j.rser.2016.11.241.

[28] L. Kantorovich. Mathematical methods of organizing and planning production. *Management Science*, pages 366–422, 1939.

[29] M. R. Leadbetter, G. Lindgren, and H. Rootzén. *Asymptotic Distributions of Extremes*, pages 3–30. Springer New York, New York, NY, 1983. ISBN 978-1-4612-5449-2. doi: 10.1007/978-1-4612-5449-2_1. URL `https://doi.org/10.1007/978-1-4612-5449-2_1`.

[30] E. H. Mayoral, M. A. H. López, E. R. Hernández, H. J. C. Marrero, J. R. D. Portela, and V. I. M. Oliva. Fourier analysis for harmonic signals in electrical power systems. In G. S. Nikolic, M. D. Cakic, and D. J. Cvetkovic, editors, *Fourier Transforms*, chapter 3. IntechOpen, Rijeka, 2017. doi: 10.5772/66733. URL `https://doi.org/10.5772/66733`.

[31] S. P. Millard. *EnvStats: An R Package for Environmental Statistics*. Springer, New York, 2013. ISBN 978-1-4614-8455-4. URL `https://www.springer.com`.

[32] A. Okabe, B. Boots, K. Sugihara, and S. Chiu. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, volume 43. 01 2000. ISBN 9780471986355. doi: 10.2307/2687299.

[33] J. P. Palutikof, B. B. Brabson, D. H. Lister, and S. T. Adcock. A review of methods to calculate extreme wind speeds. *Meteorological Applications*, 6(2):119–132, 1999. doi: https://doi.org/10.1017/S1350482799001103. URL `https://rmets.onlinelibrary.wiley.com/doi/abs/10.1017/S1350482799001103`.

[34] O. Perrin, H. Rootzén, and R. Taesler. A discussion of statistical methods used to estimate extreme wind speeds. *Theoretical and Applied Climatology*, 85:203–215, 07 2006. doi: 10.1007/s00704-005-0187-3.

[35] D. W. Richard Johnson. *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall, 6 edition, 2007. ISBN 978-0-13-187715-3.

[36] T. Romary, F. Ors, J. Rivoirard, and J. Deraisme. Unsupervised classification of multivariate geostatistical data: Two algorithms. *Computers & Geosciences*, 85: 96–103, 05 2015. doi: 10.1016/j.cageo.2015.05.019.

[37] RSE. L'energia elettrica dal vento, 2017.

[38] RSE. Atlante eolico, 2023. URL `https://atlanteeolico.rse-web.it/start.phtml`.

[39] A. Schuster. On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena. *Terrestrial Magnetism*, 3 (1):13–41, 1898. doi: https://doi.org/10.1029/TM003i001p00013. URL `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/TM003i001p00013`.

[40] P. Secchi, S. Vantini, and V. Vitelli. Bagging Voronoi classifiers for clustering spatial functional data. *International Journal of Applied Earth Observation and Geoinformation*, pages 53–64, 04 2012. doi: 10.1016/j.jag.2012.03.006.

[41] H. Shi, Z. Dong, N. Xiao, and Q. Huang. Wind speed distributions used in wind energy assessment: A review. *Frontiers in Energy Research*, 9, 11 2021. doi: 10.3389/fenrg.2021.769920.

[42] K. C. Sinclair. International Energy Agency (IEA): Implementing Agreement for Cooperation in the Research and Development of Wind Turbine Systems (IEA Wind). 11 2017. URL `https://www.osti.gov/biblio/1409734`.

[43] W. Skamarock, J. Klemp, J. Dudhia, D. Gill, D. Barker, W. Wang, and J. Powers. A Description of the Advanced Research WRF Version 3. 27:3–27, 01 2008.

[44] V. Sohoni, S. Gupta, and R. Nema. A critical review on wind turbine power curve modelling techniques and their applications in wind based energy systems. *Journal of Energy*, 2016:1–18, 01 2016. doi: 10.1155/2016/8519785.

[45] Terna. Metodologia per il calcolo del beneficio per l'incremento della resilienza della rete di trasmissione nazionale, 03 2021.

[46] A. Torrielli, M. P. Repetto, and G. Solari. Extreme wind speeds from long-term

synthetic records. *Journal of Wind Engineering and Industrial Aerodynamics*, 115: 22–38, 2013. ISSN 0167-6105. doi: https://doi.org/10.1016/j.jweia.2012.12.008. URL `https://www.sciencedirect.com/science/article/pii/S0167610512002899`.

[47] A. Torrielli, M. P. Repetto, and G. Solari. A refined analysis and simulation of the wind speed macro-meteorological components. *Journal of Wind Engineering and Industrial Aerodynamics*, 132:54–65, 2014. ISSN 0167-6105. doi: https://doi.org/10.1016/j.jweia.2014.05.006. URL `https://www.sciencedirect.com/science/article/pii/S0167610514001068`.

[48] F. Tosunoğlu. Accurate estimation of T year extreme wind speeds by considering different model selection criterions and different parameter estimation methods. *Energy*, 162:813–824, 2018. ISSN 0360-5442. doi: https://doi.org/10.1016/j.energy.2018.08.074. URL `https://www.sciencedirect.com/science/article/pii/S0360544218316062`.

[49] R. Von Mises. La distribution de la plus grande de n valeurs. *Rev. Math. Union Interbalcanique*, 1:141–160, 1936.

[50] G. Voronoï. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. premier mémoire. sur quelques propriétés des formes quadratiques positives parfaites. *Journal für die Reine und Angewandte Mathematik*, 133:97–178, 1908. doi: 10.1515/crll.1908.133.97.

[51] G. Voronoï. Nouvelles applications des paramètres continus à la théorie des formes quadratiques. deuxième mémoire. recherches sur les parallélloèdres primitifs. *Journal für die Reine und Angewandte Mathematik*, 134:198–287, 1908. doi: 10.1515/crll.1908.134.198.

[52] W. Weibull. A statistical distribution function of wide applicability. *ASME Journal of Applied Mechanics*, pages 293–297, 09 1951.

[53] S. S. Wilks. The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Annals of Mathematical Statistics*, 9:60–62, 1938. doi: 10.1214/aoms/1177732360.

[54] A. Zagubień and K. Wolniewicz. Energy efficiency of small wind turbines in an urbanized area—case studies. *Energies*, 15:52–87, 07 2022. doi: 10.3390/en15145287.

List of Figures

List of Tables

Ringraziamenti Tia

Ringrazio il Professor Piercesare Secchi per la sua disponibilità e per aver rappresentato un punto di riferimento in ogni fase della tesi. Ringrazio anche Chiara Barbi, correlatrice di tesi, per il suo supporto, soprattutto nelle prime fasi del lavoro, e per i suoi preziosi consigli.

Ringrazio il mio più prezioso amico Ale perchè da più di vent'anni è il compagno su cui so di poter sempre contare. Grazie per essere stato un sostegno essenziale durante questi mesi di lavoro, così come lo sei sempre stato nella vita di tutti i giorni.

Ringrazio i miei genitori, Pino e Monica, per avermi lasciato libero di scegliere il mio percorso scolastico e accademico, e avermi sempre dato fiducia in queste scelte.

Ringrazio i miei cari amici del gruppo "Nerding in Progress", Manzo, Poppo e Longo, per essere stati miei compagni nei momenti di svago e di relax, e perchè con voi si ride forte sempre. Grazie Andre, Diba e Nino, con cui ho stretto un rapporto fantastico proprio durante questi ultimi anni. Grazie anche a Rugge e Remo, conosciuti tra i banchi dell'Università, che sono stati prima compagni e ora sono cari amici.

Grazie, infine, ad Ale, la mia dolce metà, perchè sei la persona dal cuore più grande che io abbia mai conosciuto, perchè mi supporti nelle mie scelte e mi sopporti nelle mie stranezze, e perchè da più di 7 anni rappresenti la luce più nitida che rischiara il mio presente e che illumina il mio futuro.

Ringraziamenti Ale

Vorrei ringraziare innanzitutto il professor Secchi per l'opportunità di svolgere questa tesi con lui e la sua guida e Chiara Barbi per il supporto fondamentale che ci ha fornito con la sua disponibilità e prontezza nel seguirci e aiutarci.

Il mio ringraziamento più grande va chiaramente a Tia, compagno in questo viaggio come in tantissimi altri; grazie per aver portato avanti il lavoro quando non ce la facevo e per aver sempre creduto nella buona riuscita di questo progetto, non avrei potuto davvero chiedere una persona migliore da avere accanto.

Grazie ai miei genitori, perché mi hanno sempre supportato senza mai mettermi pressione, lasciandomi libero di seguire la mia strada.

Grazie ai miei amici (in rigoroso ordine alfabetico di soprannome): Ale, Andre, Longo, Manzo, Poppo, Remo e Rugge, per la sincerità e la bellezza del nostro rapporto, perché non importa cosa facciamo, se siamo insieme siamo sicuri di divertirci.

Infine grazie ai ragazzi dell'Oratorio di Lesmo, con i quali, negli ultimi anni, ho passato momenti molto belli e intensi.