POLITECNICO MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE E DELL'INFORMAZIONE

Executive Summary of the Thesis

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

Laurea Magistrale in Mathematical Engineering - Ingegneria Matematica

Authors: Mattia Gentile - Alessandro Sala

Advisor: Prof. Piercesare Secchi

Co-advisor: Ing. Chiara Barbi

Academic year: 2021-2022

---

1. 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. However, while wind power presents many benefits, it also brings many threats for both human and infrastructure safety; this thesis aims to investigate this double nature of wind power, focusing the attention on the region of Lombardy, Italy. In particular, our study analyses wind data measured over 30 years in order to achieve a high level of awareness about both the dangers for the electrical grid and the opportunities to produce clean and renewable energy in Lombardy.

Moreover, we aim to compare, in each phase, different techniques and evaluate their benefits and criticalities. In this kind of practical application, indeed, it is of paramount importance to choose the most appropriate mathematical procedures: employing outdated models or excessive simplifications may lead to serious errors and wrong results.

The data we used are contained in the MEteorological Reanalysis Italian DAtaset (MERIDA), which is open and provided by the "Ricerca Sistema Energetico" (RSE) S.p.A. group. The dataset subdivide the whole territory of Italy in 4×4 km cells , forming a grid over which many climatological features are measured. The measurements are taken hourly and span the time period that goes from 1990 to 2020. For our purposes, we analysed data relative to Lombardy and the series of wind speeds measured at 10 meters from the ground.

2. Wind Speed Modeling

Our first step was the assessment of the main features of wind speed by finding the distribution that better approximates measured data. The importance of this section lies in the fact that it is the starting point of many procedures concerning wind speed analysis, be it extreme values or wind energy production studies, and lack of accuracy here translates into estimation errors in subsequent steps, even with possible economic repercussions.

In this study, we considered four of the most commonly used and promising distributions (Shi et al. (2021) [6]): the Weibull distribution, the

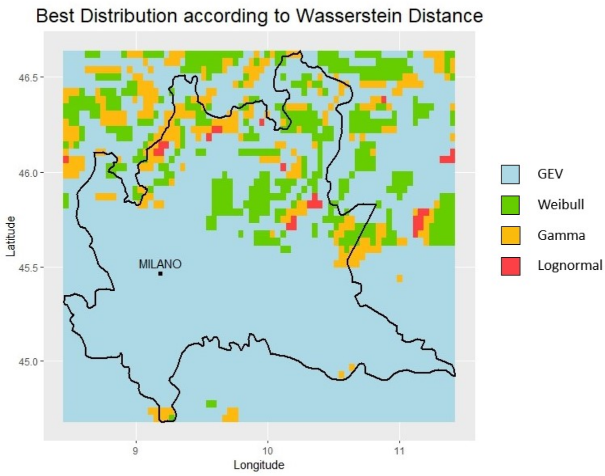

Figure 1: Each cell reports the best approximating distribution obtained with the Method of Moments and according to the Wasserstain distance to the empirical distribution.

Gamma, the Lognormal and the Generalized Extreme Value Distribution (or GEV, which will be described later). The goal here is to find the parameters of these models that produce the best fit and then choose the most appropriate model; to do so we compared two parameter estimation methods, the Maximum Likelihood Estimator (MLE) and the Method Of Moments (MOM), and evaluated their performances with four different Goodness Of Fit criteria, namely the Root Mean Square Error (RMSE), the coefficient of determination R^2, the Mean Absolute Error (MAE) and the Wasserstein Distance. Moreover, for a more proper analysis, we also employed a cross validation procedure in order to avoid overfitting.

Notably, our results confirmed what is the general consensus in the field, i.e. there is no unified model that it is always able to produce the best result independently on the location and wind speed characteristics but each case must be considered individually. What we noticed is that the GEV distribution is the most flexible one and produce the best fit in the largest share of sites, especially the ones with lower speeds. Weibull distribution, instead, finds application in areas with slightly higher average winds, on the hills and mountains, while Gamma and Lognormal represent a better alternative to Weibull in areas with much higher average wind speeds (Figure 1).

While observing wind speed distributions we also noticed that simple geographical distance provides too little information when it comes to satisfactorily describe winds behaviour correlations: more often than not, in the Alps we found very different wind regimes even in adjacent cells, while pretty far away sites in the Po Valley often showed quite similar trends. This insight led us to further investigate this fact later on, when we tried to divide the region in areas with the same wind behaviours.

3. Hazard Analysis

Once we modeled the wind speed distribution, we moved to the analysis of extreme events or, in other words, the annual maximum distribution. The common issue when studying maxima of a given distribution is the scarcity of data at disposal: in our case, having 31 years of measurements translates into 31 annual maxima. Since this may represent an issue, we tested some techniques that exploit the parent distribution of data, i.e. the *pdf* computed in Section 2, to model maxima. For example, if we have a pool of n independent measurements distributed according to a cumulative distribution function F, following basic probability notions, the annual maxima will be distributed as F^n. While very simple, this method can give good results if the parent distribution is accurate.

Another possibility can be found in a data augmentation approach which relies on the simulation of new annual wind series and, thus, of new annual maxima. The idea behind this method relies on sampling new measurements of wind speed from the parent distribution F in order to obtain an annual simulated series; in our specific application, being the measurements hourly taken, we need 24×365 new samples. Then, from these new values we keep only the maximum and, iterating the procedure, we use it to increase the pool of data at our disposal and solve the issue of data scarcity.

Both the methods, however, suffer the same flaw: while the parent distributions obtained following the pipeline described in Section 2 look accurate, the lack of extreme data determines a poor modeling of the right tail and, thus, an inaccurate characterization of the maxima. This error cascades in these two methods, producing estimates that are very far from representing the real measured maxima.

Despite the lack of data, a direct modeling approach of the maxima looks the most promising. Indeed, thanks to extreme value theorem and, subsequently, to the Generalized Extreme Value distribution (GEV) we have a powerful tool to model maxima of a given variable. The extreme value theorem demonstrates, indeed, that the distribution of the maximum of a variable tends (in distribution) to one of three possible asymptotes: Gumbel, Fréchet or Reverse Weibull distributions; the GEV, which is the combination in a unique formula of these three possible asymptotes, is a three parameters distribution. Results obtained by tuning the parameters directly on the maxima are much more consistent and, thus, we eventually decided to use this last approach.

At this point, we were ready to quantify the actual hazard that each area of Lombardy is subject to because of extreme wind events. After having modeled with a GEV the distribution of wind speeds in each cell of the grid, we could compute the so called "Exceedance probability curve" and the "Mean return time". The first one is a function defined as $1 - F$, where F now denotes the cumulative distribution function of maxima, that assigns to each threshold of the wind speed the probability of exceeding that value at least once in a year. The mean return time, instead, is computed as $\frac{1}{1-F(x)}$, where F is again the *cdf* of maxima, and represent the number of years that one has to wait, on average, before the considered threshold x will be exceeded again. Both of them tell us the same story: on one side the Po Valley is an extremely safe area (with the exception of the peculiar and isolated case of Milan, which shows a high level of hazard), while the Alps area is characterized by much higher winds and, thus, a higher hazard level.

In particular, we focused our attention on two main thresholds: 60 km/h and 140 km/h which are denoted, in the literature, as the thresholds of failures caused by, respectively, indirect and direct effects. 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 observed these two thresholds on the area of Lombardy to immediately spot those areas more subject to extreme wind events (Figure 2).

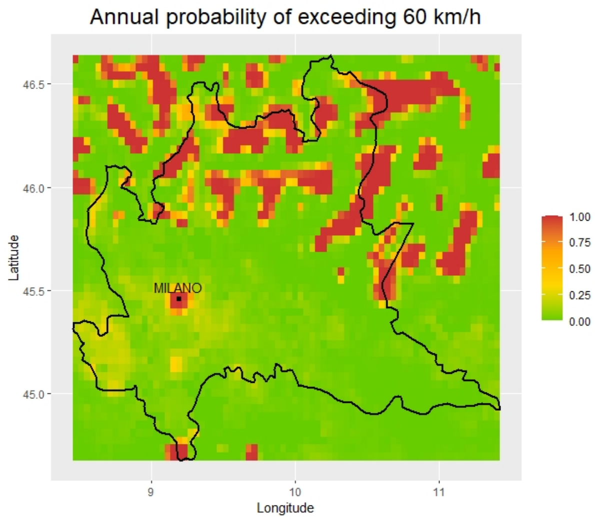

Figure 2: Each cell reports the probability of exceeding the threshold of 60 km/h in a year.

The strenght of the exceedance probability curves as a tool to identify areas of hazard, in particular, is that it is easy to extend the time window considered: instead of computing the probability of exceeding a threshold in a year, we can do it for whichever time window we are interested in, obtaining more forward-looking results.

In this setting, we were also interested in grouping different areas by means of their hazard level: to do so we divided areas based on the 99% quantile of the distribution of maxima. After having defined few intervals of wind speed (for example, less than 60 km/h, between 60 and 140 km/h, and more than 140 km/h), we grouped together those cells of the grid whose 99-quantile falls into the same bin. The results showed that the majority of the territory is characterized by no to low hazard, with very few areas subject to the risk of direct failures (Figure 3).

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 most hazardous areas.

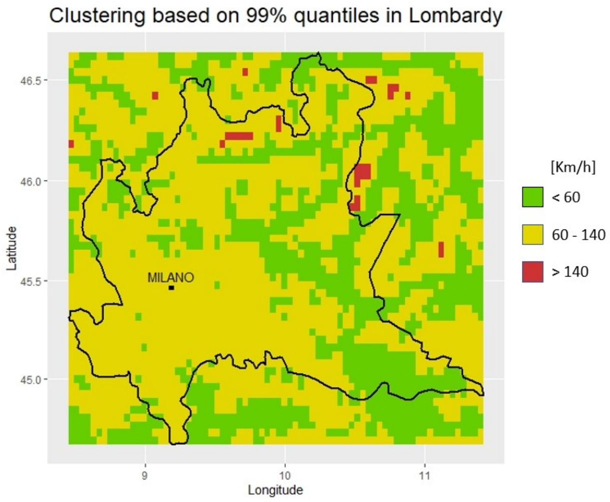

Figure 3: Each cell is colored according to the corresponding cluster in a 3 cluster subdivision based on the 99% quantile.

4. Meteorological Correlations

Before leaving hazard analysis to focus on wind energy production, we conducted a more general study regarding wind speed characterization. In particular, the goal here is to determine whether two sites can be considered to be "subject to the same wind regime", meaning that the same wind blows over them, and, ultimately, to divide the region in geographical areas interested by distinctive winds. This can surely be interesting from a climatological point of view in general, but it can also offer helpful insights on wind characteristics in the area and help with infrastructure design and wind energy production planning.

For this analysis we decided, instead of taking measurements separately as done before, to consider the wind speed series as a whole, in order to grasp the trends of the regime. To do so, we moved to a Functional Data Analysis (FDA) environment, which allowed us to efficiently model the temporal series and recover the most important information. With FDA we were able to produce a smoother representation without losing too much information on the variability and, in particular, we opted to use a number of basis functions that allowed to follow the daily variations in wind speeds, which represented the most influential cycle of the phenomenon.

Moreover, of particular interest is the observation of functional principal components (FPCs), a tool that allows to synthesize the main characteristics of the series in just a few numbers, i.e. the scores relative to the FPCs, which give an indication on some of the features of wind speed trends and can be used to discriminate different behaviours. Following on this, we produced a metric that, measuring distances between scores, is able to measure the difference between wind regimes and, thus, tell how similar or different two locations can be from the point of view of wind regimes.

This newly defined distance can be exploited to cluster together areas with similar wind behaviours and, in particular, we used it as the core for two different clustering algorithms, namely the "Geostatistical Hierarchical Clustering" and the "Bagging Voronoi Classifier". Once again, in this section, we want to compare the results obtained with two different paths, in order to select the better suited method for the job.

The Geostatistical Hierarchical Clustering algorithm (Romary et al. (2015) [3]) revolves around the simple idea of combining a classical hierarchical clustering routine with the concept of "adjacency": basically, at each step, we are allowed to group together only clusters that are adjacent. In our case, the concept of adjacency comes naturally from the grid structure, considering adjacent two sites only if they share at least a vertex in the grid. We report here the general structure of the algorithm to give an idea on how it works:

Algorithm 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

---

While the code is quite straightforward, the results produced are robust and appealing: this method manages to discriminate the big area of the Po Valley from the rest of the region, highlighting also smaller and more collected zones such as the Garda Lake and the Valtellina with part of the Como Lake. The good quality of the results produced can be found also in the trends that characterize each of the aforementioned areas. Every cluster shows an average trend very different from the others, even with a small number of clusters (Figure 4).

Algorithm 2 Bagging Voronoi Classifier

1: Initialize the hyperparameters of the algorithm: B, n, p, K and choose the distance $d(\cdot, \cdot)$ used in the Voronoi Tessellation.
2: **for** $b := 1$ to B **do**
3: 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.
4: Compute the representative g_i^b for each element i of the tessellation.
5: 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.
6: Cluster the scores in K groups according to a chosen unsupervised method.
7: **end for**
8: Perform cluster matching, i.e. match the labels across the B bootstrap replicates of the clusters, to ensure identifiability.
9: **for all** $x \in S_0$ **do**
10: 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.
11: Compute spatial entropy for the site x
12: **end for**

The Bagging Voronoi Classifier (Secchi et al. (2012) [5]), instead, is composed of two parts: a Bootstrap phase, in which the grid is first divided according to geographical proximity (Voronoi Tessellation) and then clusters are produced, and an aggregation phase, in which all the information brought by the many clustering is resumed in a single, final result. This algorithm is much more convoluted than the previous one and consists of many more technicalities. The structure of the Bagging Voronoi Classifier is reported in Algorithm 2.

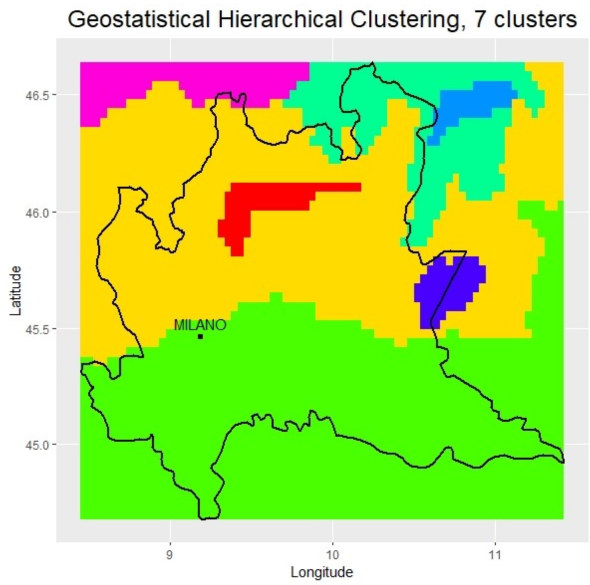

Figure 4: Grouping produced by the Geostatistical Hierarchical Clustering Algorithm with 7 clusters.

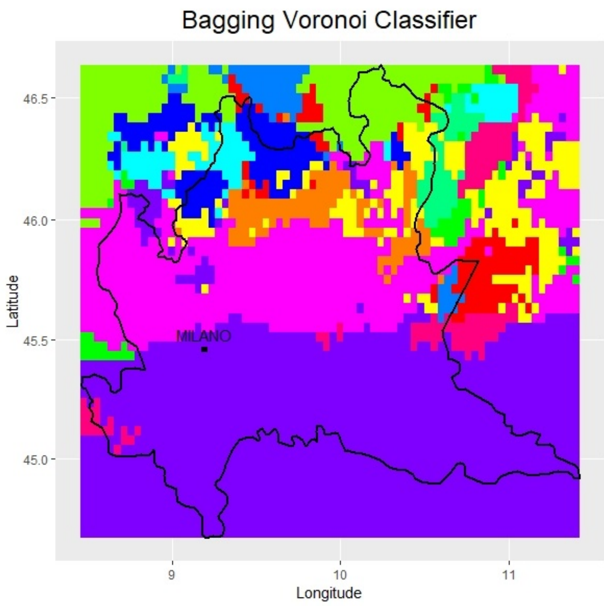

Figure 5: Grouping produced by the Bagging Voronoi Classifier Algorithm with $n = 100$ elements in the Voronoi Tessellation and 12 clusters.

While the code is much more complex, the results are underwhelming if compared with its counterpart. The main issues revolves around the fact that here we don't have a notion of adjacency and, thus, the clusters produced are more scattered; moreover, the step of `cluster matching` may lead to assign small areas or single sites to the wrong group (Figure 5).

In conclusion, for the purpose of defining wind correlations, we determined that the Geostatistical Hierarchical Clustering would be better suited for the job, as the Bagging Voronoi seems to struggle on very heterogeneous areas.

5. Wind Energy

Lombardy is notoriously a region inadequate for wind energy production and many studies have proven, beyond any reasonable doubt, the inapplicability of traditional wind farms. However, the growing need of energy production from renewable sources and energetic independence brought us to consider a possible alternative in the form of the so called small wind turbines.

The major flaws of Lombardy when it comes to wind energy are two: generally low wind speeds and a lack of feasible locations where to install turbines; indeed traditional turbines require large uninhabited plain areas with no obstructions, which are difficult to find in the Po Valley due to dense urbanization. Small wind turbines potentially overcome both these problems since they require lower wind speeds to work and can be installed even in an urban context and, thus, are the perfect candidate for a possible source of clean energy.

To study the potential applicability of this technology, we conducted a case study using the Ecolibrì 10kW Generator as a model of reference and computing its theoretical annual energy producible. Results show that the situation in Lombardy is not as bad as one could expect since around 30% of the region could potentially produce enough energy to cover the consumption of the average 3 people family. In addition, while most of the plain area has almost negligible productivity, mountain areas, with higher winds, are more indicated for possible applications and installation of small wind turbines could represent a way to autonomously produce energy in sites where could be difficult to bring it.

Moreover, some interesting observations can be done here. Regarding the computation of the annual energy producible, we compared results obtained approximating the wind speed distribution using a Weibull model, as usually done in the literature, with the ones obtained with the best approximating distribution as retrieved in Section 2. The comparison shows that, in most cases, the Weibull model produces a greater approximation error than the other procedure and highlights the importance of using the most appropriate mathematical modeling, to avoid repercussions on the following steps.

Then, the clustering algorithms of Section 4 have been proposed again here to produce a division of the territory in areas with similar production capability but independently on the regime itself. In this case the most appropriate one was the Bagging Voronoi Classifier which was able to produce a subdivision in four clusters with different capabilities.

	1	**2**	**3**	**4**
Sites	165	2596	317	622
Hours	175	464	1750	2566
kWh	440	1505	7409	12624

Table 1: Results for energy clustering. Hours refers to the number of hours of activity in a year and kWh to the total energy produced. Their values are computed as the average over all the sites of the same cluster.

6. Conclusions

Throughout this work we analyzed many aspects of wind. In the first place, we highlighted the importance of choosing case by case the better suited distribution: both from the point of view of hazard analysis and energy production, having even a slight error in modeling can be disrupting for the procedure. In second place, under the light of the trade-off risks/benefits, Lombardy confirms to be a safe area without a great wind energy potential. However, emerging technologies, less demanding in terms of space and wind speeds, can find profitable application in some areas of the region.

We remark that, this work suffers from some limitations due to the characteristics of data at our disposal and to preliminary assumptions necessary to perform the analysis. Indeed, the dataset covers the whole Italy but we decided to focus only on Lombardy; however, the methods applied can be exploited in studying the nature of wind in any location and can be used in broader applications. Moreover, we remember that the grid of the data is composed of 4 by 4 kilometers cells; possible extensions of this study can focus on higher resolution datasets obtained through downscaling or local measurements and obtain results with increased spatial resolution. Finally, throughout the study, we always worked under the assumption of stationarity, meaning that we considered that the phenomenon does not change with time. This was done in order to be able to apply most of the procedures described but we are aware that climate changes may influence wind regimes and it can be appropriate to keep in mind this fact for future researches building up on this one.

7. Bibliography

- Shi et al. (2021) [6], "Wind Speed Distributions Used in Wind Energy Assessment: A Review": An article about the most commonly used distributions used in wind speed modeling.
- Coles (2001) [1], "An Introduction to Statistical Modeling of Extreme Values": A book on Extreme Value Analysis.
- Ramsay et al. (2005) [2], "Functional Data Analysis": A book on Functional Data Analysis.
- Romary et al. (2015) [3], "Unsupervised classification of multivariate geostatistical data: Two algorithms": An article in which it is presented the original version of the Geostatistical Hierarchical Clustering.
- Secchi et al. (2012) [5], "Bagging Voronoi classifiers for clustering spatial functional data": An article in which it is presented the original version of the Bagging Voronoi Classifier.
- RSE (2017) [4], "L'energia elettrica dal vento": A publication focusing on wind energy production.

References

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

[2] B. W. Silverman J. O. Ramsay. *Functional Data Analysis.* Springer New York, NY, 2005.

[3] Thomas Romary, Fabien Ors, Jacques Rivoirard, and Jacques Deraisme. Unsupervised classification of multivariate geostatistical data: Two algorithms. *Computers & Geosciences*, 85:96–103, 05 2015.

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

[5] Piercesare Secchi, Simone Vantini, and Valeria Vitelli. Bagging Voronoi classifiers for clustering spatial functional data. *International Journal of Applied Earth Observation and Geoinformation*, pages 53–64, 04 2012.

[6] Huanyu Shi, Zhibao Dong, Nan Xiao, and Qinni Huang. Wind speed distributions used in wind energy assessment: A review. *Frontiers in Energy Research*, 9, 11 2021.