Identifying Extreme Wind Resource Scenarios for Power System Operation and Planning with High Renewable Energy Penetration

Alex Villamarín-Jácome
Departamento de Ingeniería Eléctrica
Universidad de Chile
Santiago, Chile
avillama@ing.uchile.cl

Danny Espín-Sarzosa
Departamento de Ingeniería Eléctrica
Universidad Técnica Federico Santa María
Valparaíso, Chile
danny.espin@usm.cl

Jorge Vega-Herrera
Departamento de Ingeniería Eléctrica
Universidad de Antofagasta
Antofagasta, Chile
jorge.vega@uantof.cl

Felipe Muñoz
Escuela de Ingeniería Eléctrica
Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile
felipe.munoz.q@mail.pucv.cl

Sebastián Aguirre
Escuela de Ingeniería Eléctrica
Pontificia Universidad Católica de Valparaíso
Valparaíso, Chile
sebastian.aguirre.v@mail.pucv.cl

***Abstract*—This paper presents a methodology designed to identify extreme wind resource scenarios, which could pose critical operating conditions for power systems, especially in areas with high penetration of renewable energies like Chile. Due to its unique geographic characteristics as the longest and narrowest country in the world, Chile serves as an ideal case study. The methodology is applied using historical wind resource data from three distinct zones—Northern, Central, and Southern Chile. The methodology comprises five main stages: Data Collection, data preprocessing, application of multivariate clustering, clustering of extreme scenarios, and selection of representative scenarios. These extreme wind scenarios are then leveraged to assess the performance of wind generators in power production. The analysis reveals daily wind speed profiles, highlighting how these extreme scenarios differ from average conditions and emphasizing the potential challenges in utilizing wind power during extreme weather conditions.**

***Index Terms*—Extreme events, k-means clustering, renewable energy, wind resource scenarios.**

I. Introduction

Power systems are undergoing significant transformations due to the increasing integration of renewable energy sources, such as solar and wind power, as part of global efforts to mitigate climate change. Wind energy, in particular, has experienced substantial growth in recent years. Between 2015 and 2020, the globally installed capacity increased from 426.45 GW to 898.82 GW [1]. This growth is driven by technological advancements, decreasing costs, and the widespread availability of wind resources [2]. However, the inherent variability and uncertainty associated with wind energy pose significant challenges in its integration into power systems. As wind resource variability can fluctuate dramatically, especially under the influence of climate change, power systems are becoming increasingly prone to stability and security issues [3]. To ensure the stability and security of power systems, it is essential to consider these extreme scenarios of wind resource variability in both the operation and planning stages. This requires the development of advanced methods that can properly represent the variability associated with wind energy, particularly extreme wind resource scenarios that may have implications for power system planning and operation [3].

In this context, various clustering techniques have been employed to address the challenges arising from the variability of wind energy. These techniques are widely used in power system applications such as scenario reduction, renewable resource forecasting, fault detection, and many more. For instance, in [4], k-means clustering is applied to determine a set of representative wind and solar scenarios for use in stochastic unit commitment problems, thereby reducing computational burden while maintaining operational accuracy. The self-organization map clustering algorithm (SOM) is also applied to the same problem in [5]. Similarly, in [6] and [7], the k-means are used to classify normal and fault conditions, which are then used for fault detection in distributed networks and photovoltaic systems. Thus, enhancing the reliability of the system.

Clustering techniques have also been applied in generation expansion planning, as discussed in [8]–[10], where scenarios are reduced to capture the variability of renewable generation, ensuring that future generation and transmission capacity can meet the fluctuating demand, and uncertainty associated with renewable energy. Furthermore, [11] and [12] examine the application of clustering for renewable energy forecasting, while [13] and [14] use clustering for customer net-meter energy data from residential consumers for optimal sizing of energy storage optimization and the energy management system of microgrids. These applications demonstrate its versatility in various operational contexts. The reader is referred to [15]

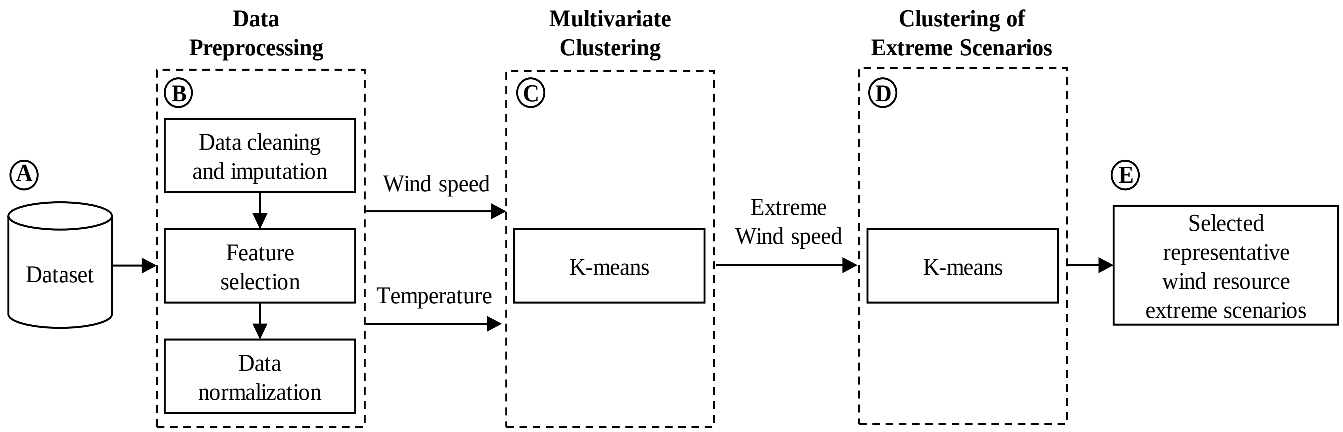


Fig. 1: Methodology for identifying extreme wind resource scenarios.

for a detailed review of the applications of k-means to power system issues. Despite these advancements, existing methodologies often fail to adequately represent extreme scenarios associated with integrating high levels of wind energy into power systems, which is critical to ensuring reliable and secure operation.

In power system operation and planning problems that involve high levels of integration of wind energy, it is common practice to use a set of representative scenarios to capture the variability and uncertainties of wind resources. These scenarios are typically derived from historical records and critical inputs in generation expansion planning. Clustering techniques, such as k-means, are often applied to historical data to identify a set of wind scenarios that appropriately represent the variability and uncertainty inherent in wind resources [5], [9]. An accurate representation of these scenarios is crucial to anticipate unsafe operating conditions and to ensure that power systems are adequately developed to integrate high levels of wind energy [4], [8], [10]. However, traditional clustering techniques in these contexts often overlook extreme events associated with wind resources, significantly affecting the reliability of the system.

In this context, this work proposes a methodology to identify extreme wind resource scenarios, which could represent critical operating conditions for power systems. The methodology is applied to a case study using historical wind resource information from three zones of Chile (Northern, Central, and Southern). Chile is a particular choice due to its singular geographic characteristics (it is the longest and narrowest country in the world), together with the high penetration of renewable energies in its power system. Extreme wind resource scenarios are then used to evaluate the performance of wind generators in power production in zones with high renewable energy penetration.

The remainder of the paper is organized as follows: Section II presents the methodology for identifying the extreme wind resource scenarios. Section III describes the case study considered to apply the methodology and obtain the results of the identification of wind resource scenarios. Section IV provides the analysis of wind generator performance when considering extreme wind resource scenarios. Finally, Section V highlights the main conclusions of the study and future work.

II. Methodology

Fig. 1 displays the proposed methodology for identifying extreme wind resource scenarios that are critical for power system operation and planning, especially in bulk power systems with high renewable energy penetration. As shown in the Fig. 1, the methodology comprises 5 main stages, which are detailed in the following subsections.

A. Data Collection

The first stage involves collecting historical wind speed and temperature data to provide a thorough analysis of extreme wind conditions. This is because the ambient temperature also influences the operation of wind generators. The data can be obtained from reliable and publicly available meteorological sources, such as weather stations, satellite data, or other databases [16]–[19], ensuring the dataset encompasses a sufficiently long period to capture the variability and extremes in both wind speed and temperature. This timeframe is selected to represent different seasons and weather patterns, providing a robust basis for subsequent analysis.

B. Data Preprocessing

The collected data undergoes comprehensive preprocessing to ensure its suitability for clustering analysis. This process includes data cleaning and data imputation, where erroneous or missing data points are addressed through interpolation or exclusion [20], and outliers that could skew clustering results are detected and removed. Wind speed and temperature are selected as the primary variables of interest. However, additional features such as air pressure or humidity can be considered depending on the specific analysis objectives. To ensure that each feature contributes equally to the clustering process, the selected data is normalized, avoiding bias introduced by differing magnitudes.

C. Application of Multivariate Clustering

With the preprocessed data, multivariate k-means clustering is applied [15], utilizing both normalized wind speed and temperature as input features. The k-means algorithm splits the dataset into distinct clusters, where each cluster represents a group of days with similar wind speed and temperature profiles. The optimal number of clusters, k, is determined using the Elbow method [21], which identifies the point where adding more clusters does not significantly improve clustering performance.

To identify extreme wind scenarios, specific thresholds are established based on the operational limits of wind turbines. Extreme scenarios are identified by considering a cut-in speed of 3 m/s and a cut-out speed of 25 m/s, which correspond to the operational range of most commercial wind turbines [22]. Temperature extremes are defined using the upper 5% and lower 5% percentiles, based on the assumption that turbines typically operate within a temperature range of -10°C to +40°C. Data that meet these extreme criteria in either wind speed or temperature are extracted for further analysis. After this, the daily wind resource profiles to which the data meeting the established thresholds belong are identified and extracted in a subset.

D. Clustering of Extreme Scenarios

To the subset of daily wind resource profiles, identified as extremes in the previous stage, the k-means clustering algorithm is applied. This step focuses on grouping similar extreme profiles, providing a detailed understanding of the variety within extreme scenarios.

E. Selection of Representative Scenarios

Finally, representative scenarios are selected considering the centroid of each cluster of extreme days. These scenarios are chosen based on their ability to exemplify the characteristics of their respective clusters. This selection provides a manageable set of scenarios for further analysis in power system operation and planning. The selected extreme scenarios are compared against average daily wind speeds. For this purpose, the operation of a wind generator is simulated through typical wind turbine operating curves to evaluate the impact of the wind resource scenarios (extreme and average) on electricity production.

III. Case study and Results

In this section, the methodology presented above is applied to identify and analyze extreme wind resource scenarios that impact wind power plant operations, particularly in zones with high penetration of renewable energy sources.

A. Dataset

The dataset were collected from three distinct geographical locations in Chile: the Northern zone, Central zone, and Southern zone. We focused on the locations of three operational wind farms (WF), one in each zone, as depicted in Fig. 2.

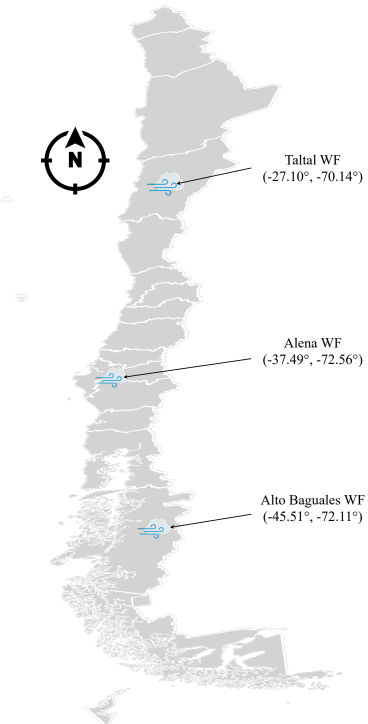


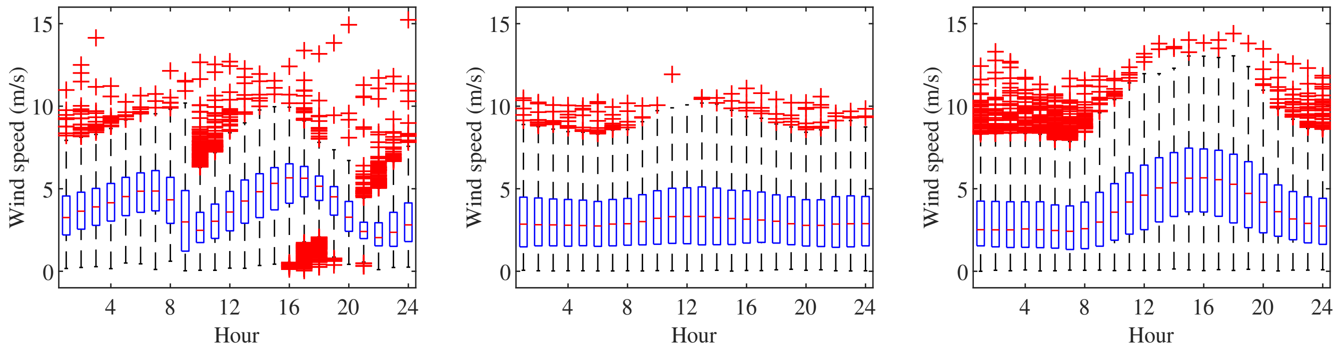
Fig. 2: Different geographic locations in Chile considered for the study.

These zones were specifically chosen due to their diverse climatic conditions, which significantly influence the availability and variability of wind resources throughout Chile. As of 2024, wind power capacity reached 4,831 MW, accounting for 14% of the total national capacity with distributions of 2,097 MW in the North, 1,418 MW in the Central, and 1,316 MW in the South according to data from the National Energy Commission [23]. In addition, projections estimate a significant increase to over 35 GW of wind power by 2050 (38% of total capacity), as part of the country's strategy to achieve carbon neutrality [24].

The dataset consists of hourly measurements of wind speed and temperature from the period of 2004 to 2016 (totaling 113,952 data for each variable). The data were sourced from resource assessment tools provided by the Ministry of Energy [16], with wind speed measurements recorded at a height of 5.5 meters.

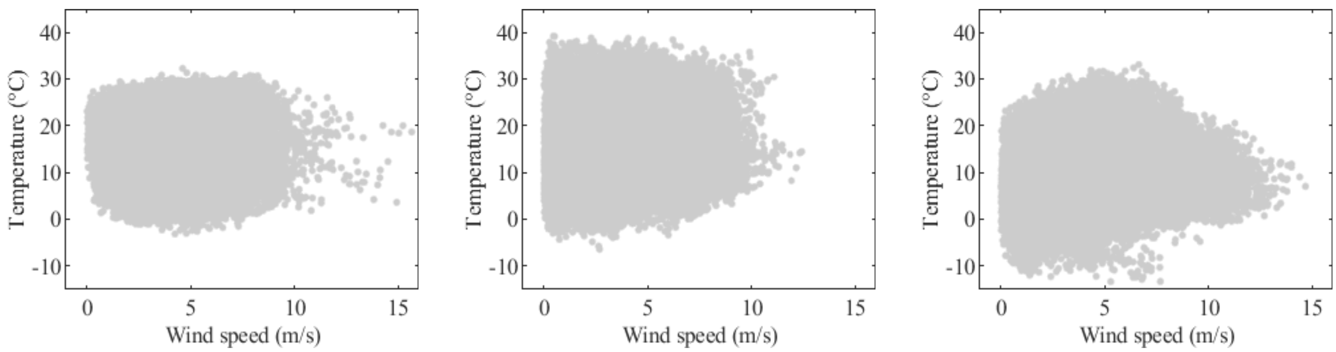
B. Data Analysis and Results

The analysis began by examining the daily wind speeds and temperatures from the datasets corresponding to each zone. Fig. 3 shows box plots of daily wind speeds for each zone, illustrating the distribution and variability of wind speeds. For instance, the Northern zone exhibits a wide range of wind speeds and notable outliers, indicating highly variable wind conditions. In contrast, the Central zone exhibits low variability with sporadic outliers, suggesting more predictable wind



(a) (b) (c)

Fig. 3: Box plot of daily wind speeds: (a) Northern zone, (b) Central zone, and (c) Southern zone.



(a) (b) (c)

Fig. 4: Scatter plot of wind speed against temperature: a) Northern zone, (b) Central zone, and (c) Southern zone.

patterns. Meanwhile, the Southern zone maintains a stable range of wind speeds, reflecting uniform wind conditions.

Next, the wind speed and temperature data were cleaned and imputed to address any missing values and remove outliers that could distort the clustering results. The data was normalized to ensure that both features contributed equally to the clustering process. Fig. 4 displays scatter plots of wind speed against temperature for each zone, providing insights into the relationship between these two variables. Notably, the Southern zone exhibits wind speeds below 3 m/s (cut-in speed) frequently coinciding with lower temperatures (see Fig. 4(c)).

With the preprocessed data, we applied the multivariate k-means clustering algorithm to each of the three zones separately. Both wind speed and temperature were used as input features, allowing the algorithm to identify clusters of days with similar environmental conditions. Fig. 5 is an elbow plot derived from the clustering analysis of the Southern zone, used to determine the optimal number of clusters. This plot helps in identifying a cut-off point where the addition of another cluster does not significantly improve the within-cluster sum of squared errors (SSE).

Fig. 6(a) shows a scatter plot of normalized wind speed and temperature in the Southern zone, color-coded by clusters identified through k-means clustering. This figure also highlights wind speeds classified as extreme scenarios, based on cut-in speed of 3 m/s, along with temperatures in the top and bottom 5% percentiles, as detailed in Section II-C.

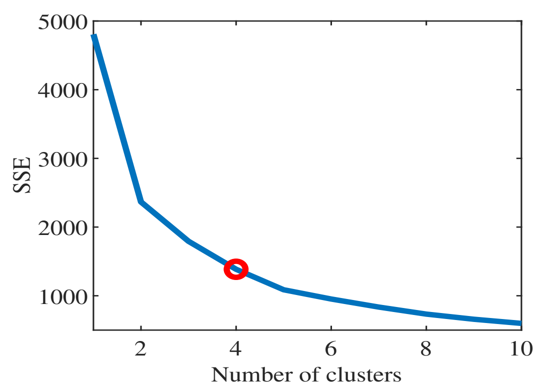


Fig. 5: Elbow plot for Southern zone data.

Fig. 6(b) illustrates a box plot of the extreme wind speeds identified in Fig. 6(a) for the Southern zone. Interestingly, it shows that while the medians decrease, they follow the same trend as the original data depicted in Fig. 4(c).

After identifying the extreme wind speed scenarios, we further refined our analysis by clustering similar days within each extreme scenario subset. This additional round of k-means clustering allowed us to group days that shared similar characteristics of extreme wind speeds. Fig. 6(c) illustrates

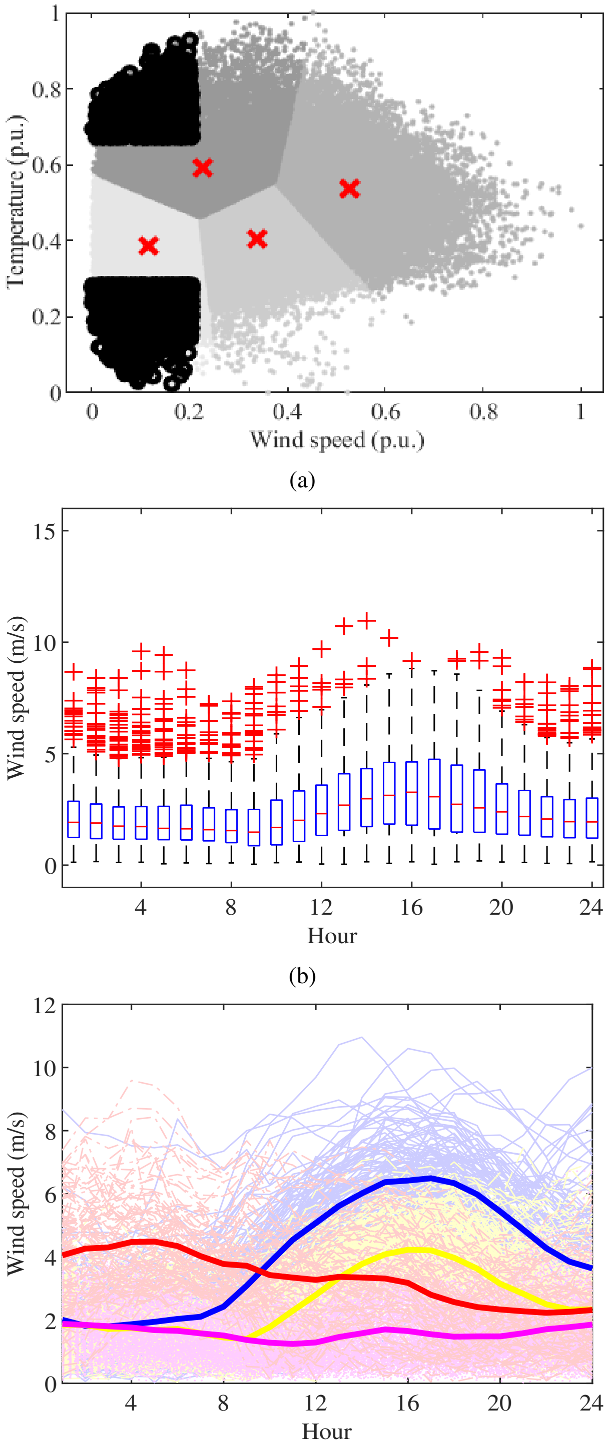


Fig. 6: Extreme scenarios for Southern zone: (a) Scatter plot of wind speed against temperature, (b) Box plot of daily wind speeds, and (c) Clusters of extreme wind speeds.

the daily profiles of these clustered extreme wind speeds, alongside the profiles of each centroid for the Southern zone. This visualization aids in understanding the typical patterns of extreme wind, which can have varied implications for wind power generation.

Finally, we selected the lower centroid of subset of extreme profiles as the representative wind resource extreme scenario. This process was also performed for the North and Central zones to obtain representative scenarios for each zone. Fig. 7 shows the selected lower centroids of extreme wind scenarios alongside the average wind profile of the original dataset for each zone. This comparison highlights how the extreme scenarios deviate from average conditions and underscores the potential challenges in wind power utilization during extreme conditions.

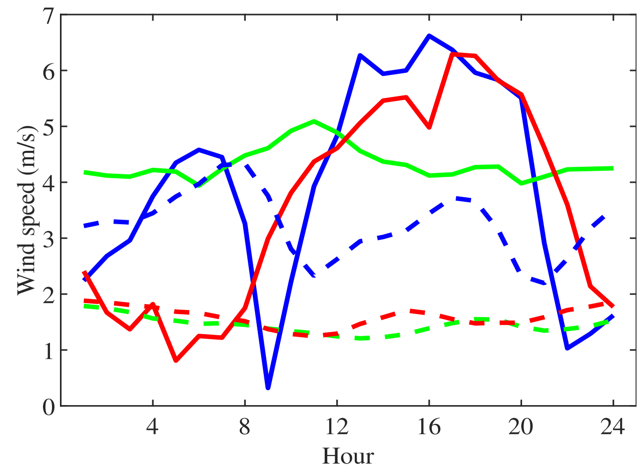


Fig. 7: Representative wind extreme scenarios versus and average resource profiles. Northern zone (blue), Central zone (green), and Southern zone (red).

IV. Evaluation of the Impact of Extreme Scenarios on Wind Generator Operation

To evaluate the impact of wind resource scenarios (extreme and medium) on electricity production, typical operating curves are used to simulate the operation of the wind farms shown in Fig. 2. In the case of Taltal, this wind farm consists of 33 Vestas V112 wind turbines of 3.3 MW each [25], totaling a gross installed capacity of 99 MW. Then, the wind resource profiles (extreme and average) are used and applied to the power curve of the Vestas V112 wind turbine. Note that the wind data are measured at the height of 5.5 meters, so it is necessary to correct them by the power law profile equation. This requires the hub height of the wind generator, which for Taltal is 140 meters. After this, 24 hours of wind farm operation is simulated and the daily energy produced is quantified using each wind resource profile.

Analogously it is performed for Alena WF considering that it comprises 18 Nordex N119 wind turbines of 4.8 MW each and the hub height is 164 meters [26], and Alto Baguales WF consists of 3 wind turbines of 0.66 MW each and the hub height is 160 meters.

Table I shows the daily energy produced with the two wind resource profiles for each WF considered in the case study.

TABLE I: Daily energy production for each wind farm by zone.

Zone	Wind resources scenarios Average (GWh)	Extreme (GWh)	Difference (%)
Taltal WF	686.72	1.44	99.79
Alena WF	518.07	0.00	100
Alto Baguales WF	5.21	0.00	100

As you can see in the Table, there is a considerable reduction in the wind energy produced as expected.

V. Conclusions

The application of the proposed methodology to the Chilean case study successfully identified and validated extreme wind resource scenarios across the Northern, Central, and Southern zones. The use of multivariate k-means clustering, incorporating both wind speed and temperature data, proved to be a robust approach for capturing the complex environmental conditions that characterize extreme wind events. The wind energy produced considering extreme scenarios is significantly lower than when considering typical average profiles. The resulting scenarios provide critical insights for power system operation and planning, particularly in the context of increasing renewable energy penetration. Overall, the proposed methodology provides a systematic framework for identifying and analyzing extreme wind resource scenarios, contributing to more informed decision-making in power system operation and planning.

Acknowledgment

This work was supported by the SERC Chile FONDAP/CONICYT grant number 1523A0006, the ANID BECAS/DOCTORADO NACIONAL 21191290, and the Scholarship SENESCYT/ARSEQ-BEC-006295-2018.

References

[1] Our World in Data. [Online], Available: https://ourworldindata.org/. Accessed: 2024-08-22.

[2] S. Zhang, J. Wei, X. Chen, and Y. Zhao, "China in global wind power development: Role, status and impact," *Renewable and Sustainable Energy Reviews*, vol. 127, p. 109881, 2020.

[3] D. Zhang, Z. Xu, C. Li, R. Yang, M. Shahidehpour, Q. Wu, and M. Yan, "Economic and sustainability promises of wind energy considering the impacts of climate change and vulnerabilities to extreme conditions," *The Electricity Journal*, vol. 32, no. 6, pp. 7–12, 2019.

[4] L. Qian, S. Lin, B. Zhou, W. Wang, X. Bian, F. Li, and D. Li, "Stochastic unit commitment based on energy-intensive loads participating in wind and solar power consumption," *IET Renewable Power Generation*, vol. 18, no. 4, pp. 589–603, 2024.

[5] D. Ortiz-Villalba, J. Vega-Herrera, J. Llanos-Proaño, and C. Mayol-Cotapos, "Stochastic unit commitment with transmission constraint using self-organized maps (som) for scenarios reduction," in *2018 IEEE International Autumn Meeting on Power, Electronics and Computing (ROPEC)*, pp. 1–6, IEEE, 2018.

[6] S. D. Roy and S. Debbarma, "A novel oc-svm based ensemble learning framework for attack detection in agc loop of power systems," *Electric Power Systems Research*, vol. 202, p. 107625, 2022.

[7] A. Et-taleby, Y. Chaibi, M. Boussetta, A. Allouhi, and M. Benslimane, "A novel fault detection technique for pv systems based on the k-means algorithm, coded wireless orthogonal frequency division multiplexing and thermal image processing techniques," *Solar Energy*, vol. 237, pp. 365–376, 2022.

[8] C. A. Moraes, L. W. de Oliveira, E. J. de Oliveira, D. F. Botelho, A. N. de Paula, and M. F. Pinto, "A probabilistic approach to assess the impact of wind power generation in transmission network expansion planning," *Electrical Engineering*, pp. 1–12, 2022.

[9] C. Lin, C. Fang, Y. Chen, S. Liu, and Z. Bie, "Scenario generation and reduction methods for power flow examination of transmission expansion planning," in *2017 IEEE 7th International Conference on Power and Energy Systems (ICPES)*, pp. 90–95, IEEE, 2017.

[10] C. Arêdes Moraes, E. J. de Oliveira, D. Fioresi Botelho, L. Willer de Oliveira, and M. Faria Pinto, "Wind generation impact in transmission expansion planning," *Journal of Control, Automation and Electrical Systems*, vol. 31, pp. 247–256, 2020.

[11] B. Zhou, X. Ma, Y. Luo, and D. Yang, "Wind power prediction based on lstm networks and nonparametric kernel density estimation," *Ieee Access*, vol. 7, pp. 165279–165292, 2019.

[12] F. Najibi, D. Apostolopoulou, and E. Alonso, "Enhanced performance gaussian process regression for probabilistic short-term solar output forecast," *International Journal of Electrical Power & Energy Systems*, vol. 130, p. 106916, 2021.

[13] R. Tang, B. Yildiz, P. H. Leong, A. Vassallo, and J. Dore, "Residential battery sizing model using net meter energy data clustering," *Applied Energy*, vol. 251, p. 113324, 2019.

[14] K. Li, F. Yang, L. Wang, Y. Yan, H. Wang, and C. Zhang, "A scenario-based two-stage stochastic optimization approach for multi-energy microgrids," *Applied Energy*, vol. 322, p. 119388, 2022.

[15] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "K-means and alternative clustering methods in modern power systems," *IEEE Access*, 2023.

[16] Explorador Solar. [Online], Available: https://solar.minenergia.cl/exploracion.

[17] S. Pfenninger and I. Staffell, "Long-term patterns of european pv output using 30 years of validated hourly reanalysis and satellite data," *Energy*, vol. 114, pp. 1251–1265, 2016.

[18] I. Staffell and S. Pfenninger, "Using bias-corrected reanalysis to simulate current and future wind power output," *Energy*, vol. 114, pp. 1224–1239, 2016.

[19] Explorador Eólico. [Online], Available: https://eolico.minenergia.cl/inicio.

[20] E. de Jonge and M. van der Loo, "An Introduction to Data Cleaning with R," 2013. Accessed: 2024-08-22.

[21] T. Kodinariya and P. Makwana, "Review on determining of cluster in k-means clustering," *International Journal of Advance Research in Computer Science and Management Studies*, vol. 1, pp. 90–95, 01 2013.

[22] S. Oyedepo, M. Adaramola, and S. Paul, "Analysis of wind speed data and wind energy potential in three selected locations in south-east nigeria," *International Journal of Energy and Environmental Engineering*, vol. 3, 2012.

[23] Energía Abierta. [Online], Available: http://energiaabierta.cl/visualizaciones/capacidad-instalada/.

[24] Ministerio de Energía, "Planificación energética de largo plazo 2023-2027," August 2021. [Online]. Available: https://energia.gob.cl/pelp/repositorio.

[25] Enel Green Power, "Informe técnico determinación de parámetros para los procesos de partida y detención parque eólico taltal," April 2019. [Online]. Available: https://www.coordinador.cl/wp-content/uploads/2019/06/Informe-T

[26] Estudios Energéticos Consultores, "Informe de determinación de mínimo técnico parque eólico alena," April 2022. [Online]. Available: https://www.coordinador.cl/wp-content/uploads/2022/06/Informe-Final-Minimo-Tecnico-PE-Alena.pdf.