

UNIVERSIDAD NACIONAL DE INGENIERÍA
FACULTAD DE INGENIERÍA MECÁNICA



TESIS:

**“OPTIMAL POWER SYSTEM EXPANSION PLANNING
INCLUDING RENEWABLE PLANTS, STORAGE SYSTEMS AND
AC SHORT-TERM OPERATION CONSTRAINTS USING TIME-
SERIES SYNTHESIS AND CLUSTERING: APPLICATION IN
MEDIUM-SIZE PERUVIAN POWER SYSTEMS”**

PARA OBTENER EL GRADO ACADÉMICO DE DOCTOR EN
CIENCIAS CON MENCIÓN EN ENERGÉTICA

ELABORADO POR:

Ruben Said Felix Ruiz

ASESORES:

Dr. Eladio Ocaña Anaya

Dr. Raul Gonzales Palomino

LIMA – PERÚ

2021

Abstract

One of the main challenges the Peruvian Electric System faces is the coordinated expansion of generation capacity and transmission to supply the demand requirements in the medium and long-term. Likewise, this process occurs in a context in which non-conventional renewable technologies such as solar photovoltaic (PV) and wind have experienced considerable cost reductions, which, added to the short development times that they entail, make them the ideal candidates to carry out the expansion of generation capacity. However, it is found that the country currently lacks the tools to carry out medium and long-term expansion planning considering these new technologies. Thus, this thesis presents 1) a methodology to synthesize hourly time series of solar PV and wind generation anywhere in the country, 2) a technique to reduce (cluster) the dimension of planning problems, and 3) an optimization model for the expansion of an electric power system that considers the operation of renewable plants, energy storage systems, and the AC short-term constraints. The results obtained show that these three proposals together make up a complete planning framework that allows evaluating different power systems in a context of scarcity of information, taking care of short-term validity of medium and long-term solutions, and achieving a correct trade-off between the technical and economic benefits of the solutions, versus the operating and investment costs of the decisions.

Acknowledgments

I am deeply indebted to many people for their support, patience, and encouragement. My advisors, Dr. Eladio Ocaña and Dr. Raul Gonzales, have generously offered their time, guidance, and advice. I have learned an incredible amount under their instruction. I am also grateful to Dr. Jaime Luyo for his thoughtful questions and careful stewardship of my doctoral program.

I also want to thank Concytec-Fondecyt, who financed this work and made possible research activities in the field of energy at the School of Mechanical Engineering of the Universidad Nacional de Ingeniería.

Finally, I wish to thank my incredibly supportive family, especially my fiancée, who inspires me and encourages me to keep advancing every time. This thesis is dedicated to her.

Contents

1	Introduction	11
1.1	Renewable Energy Challenges	12
1.1.1	Challenges addressed in Energy Planning.....	13
1.1.2	Challenges addressed in Power System Planning	14
1.2	Motivating Case: Peruvian Power System	15
1.2.1	Overview of Peruvian Power System.....	16
1.2.2	Peruvian regulations regarding renewable energies	18
1.3	Research Statement.....	19
1.4	Thesis Outline.....	20
2	Literature Review	22
2.1	Synthesis methods	22
2.1.1	Solar Energy	22
2.1.2	Wind Energy.....	24
2.2	Time-series clustering techniques.....	26
2.3	Expansion planning optimization models.....	29
2.4	The Gap in the Literature and Research Objectives	31
3	Synthesis of hourly renewable production time-series	33
3.1	Methodology.....	33
3.1.1	Methodology for Solar Energy	33
3.1.2	Methodology for Wind Energy	40
3.2	Validations.....	45
3.2.1	Validation of methodology for Solar Energy	45
3.2.2	Validation of methodology for Wind Energy.....	52
3.3	Chapter conclusions.....	63
4	Time-series clustering to reduce data dimension	64

4.1	Methodology.....	64
4.2	Application	66
4.2.1	Case study.....	66
4.2.2	Parameters tuning	68
4.2.3	Results	70
4.3	Chapter conclusions.....	72
5	Expansion planning using a practical AC model	73
5.1	Methodology.....	73
5.1.1	Conventional Generation Plants	73
5.1.2	Renewable Generation Plants.....	74
5.1.3	Energy Storage Systems	75
5.1.4	Transmission Lines.....	76
5.1.5	Reactive compensators	78
5.1.6	Busbars	78
5.1.7	Losses assessment	79
5.1.8	Objective function	80
5.1.9	Notes on extending the proposed model	81
5.2	Application	82
5.2.1	Case study.....	82
5.2.2	Solution scenarios.....	83
5.3	Chapter conclusions.....	91
6	Case Study: Electrical system of eastern Peru	92
6.1	System description.....	92
6.2	Renewable time-series synthesis	94
6.3	Demand and renewable time-series clustering	96
6.4	Costs of expansion technologies.....	98
6.5	Expansion planning optimization	98
6.5.1	C01: Base scenario	99
6.5.2	C02: Optimized scenario	100
6.6	Chapter conclusions.....	105
7	Conclusions	106
7.1	Findings	106

7.2	Future work.....	107
A	Complementary equations to calculate solar radiation.....	109
B	Complementary equations to calculate wind speed	112
C	Practical AC Optimization Model	113

List of Figures

Figure 1. Historic and projected growth for Total Peruvian System electricity load (based on [47])	17
Figure 2. Methodology for Solar Energy.....	33
Figure 3. Calculation flow to convert Daily Horizontal Radiation (<i>Hsyn</i>) into Hourly Total Radiation (<i>IT</i>)	36
Figure 4. Methodology for Wind Energy	40
Figure 5. Wind diary cycle for a point in Chile [128]......	43
Figure 6. Comparison of GHI (kWh/m ²) from different solar data sources for a point in Chile.	46
Figure 7. Time-series of solar global horizontal radiation (kW/m ²) for a point in Slovakia.	47
Figure 8. Comparison of total monthly GHI (kWh/m ²) for a point in Slovakia.....	48
Figure 9. Location and satellite image of Majes Solar Park in Peru.	49
Figure 10. Monthly average daily total insolation (Wh/m ² /day) over a PV panel for a point in Peru.....	49
Figure 11. Monthly average daily hourly irradiance (W/m ²) over a PV panel for a point in Peru.....	50
Figure 12. Total monthly solar energy output (MWh) for a point in Peru.....	51
Figure 13. Time-series of total solar power output (MW) for a point in Peru.	51
Figure 14. Time-series of total wind power output (MW) for a point in India.	53
Figure 15. Monthly and annual daily (a) synthetic and (b) actual profile of wind speed for a point in India.....	54
Figure 16. Total monthly wind energy output (MWh) for a point in India.....	55
Figure 17. Time-series of total wind power output (MW) for a point in Chile.....	56
Figure 18. Monthly and annual daily (a) synthetic and (b) actual profile of wind speed for a point in Chile.	57
Figure 19. Total monthly wind energy output (MWh) for a point in Chile.	58

Figure 20. Location and site image of Tres Hermanas Wind Farm in Peru.....	59
Figure 21. Monthly and annual daily (a) synthetic and (b) real profile of wind power (MW) for a point in Peru.	60
Figure 22. Time-series of total wind power output (MW) for a point in Peru.	61
Figure 23. Total monthly wind energy output (MWh) for a point in Peru.....	62
Figure 24. Location and satellite image of Talara Wind Farm in Peru.	67
Figure 25. Location and satellite image of Tacna Solar Park in Peru.	67
Figure 26. Power profiles of a load demand (blue), a wind farm (red) and a solar park (green) presented in PRC60 power system.	68
Figure 27. Annual time-series of the PRC60 system rearranged as daily time-series....	69
Figure 28. Clusters obtained for PRC60 system.	71
Figure 29. Real centroids of clusters obtained for PRC60 system.	72
Figure 30. Iterative process to calculate active and reactive losses.....	79
Figure 31. Satellite view of PRC60 system.....	82
Figure 32. One-line diagram of PRC60 without projects in the year 2024.	83
Figure 33. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C01.	84
Figure 34. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C02.	85
Figure 35. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C03.	86
Figure 36. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C04.	88
Figure 37. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C05.	89
Figure 38. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C06.	90
Figure 39. Satellite view of ELOR-C system.....	93
Figure 40. Active power production profiles for three generation alternatives with distinct technologies.....	94
Figure 41. Satellite view of substation Pongo 60 kV and two solar plant projects of ELOR-C system.	95
Figure 42. Active power production profiles for two wind plants located closely together.	95

Figure 43. Active power production profiles for three solar plants located closely together.	96
Figure 44. Clusters obtained for ELOR-C system.....	97
Figure 45. Real centroids of clusters obtained for ELOR-C system.	97
Figure 46. ELOR-C system operation – base case C01.	99
Figure 47. Voltage levels in load bars of ELOR-C system – base case C01.....	100
Figure 48. Satellite view of ELOR-C system with selected renewable projects.	101
Figure 49. ELOR-C system operation – optimized case C02.....	102
Figure 50. Operation of BESS installed in ELOR-C system – optimized case C02. ...	103
Figure 51. Voltage levels in load bars of ELOR-C system – optimized case C02.....	104

List of Tables

Table I. Deviation of solar data sources respect to official values for a point in Chile..	46
Table II. Deviations of total monthly GHI (kWh/m ² /month) values for a point in Slovakia.	48
Table III. Deviation of total monthly solar energy output (MWh) values for a point in Peru.....	50
Table IV. Deviations of monthly mean wind speed (m/s) values for a point in India. ..	52
Table V. Deviation of total monthly wind energy output (MWh) values for a point in India.....	55
Table VI. Deviations of monthly mean wind speed (m/s) values for a point in Chile... 56	
Table VII. Deviation of total monthly wind energy output (MWh) values for a point in Chile.	58
Table VIII. Deviation of total monthly wind energy output (MWh) values for a point in Peru.....	62
Table IX. Silhouette coefficient and processing time for different metrics used in a K-Means Clusterization	69
Table X. Silhouette coefficient and processing time for Soft-DTW metric used in a K-Means Clusterization	70
Table XI. Economic variables for expansion technologies	98

1 Introduction

As non-conventional renewable generation technologies continue decreasing their investment cost and increasing their efficiencies [1]–[4], electrical systems all over the world will get their renewable energy resources (RER) penetration level growing up, as is forecasted by many global organizations like [1] who projects that RER production will account for a 86% of global power generation. At the same time, electricity will have 49% of the share in final consumption by 2050. This fact will bring several benefits to human development and new challenges that energy sector actors shall face to smooth the path of this energy transition.

Renewable resources like wind and solar have some particularities that aggravate the most critical problems that electrical systems usually have in operation and planning. These particularities mainly refer to their intermittent behavior, which is intrinsic to the resources, and their decoupled technology (presented in the majority of existing solar PV and wind applications) that affects the system's reliability, decreasing its capacity to cope with the various natural and human-made phenomena.

Although the trend in prevailing technologies to make new generation capacity additions in the world is clear, developing countries still have a set of barriers that make them vulnerable to the challenges this new energy form brings and prevent them from reaping the benefits of renewable energy. As studied in [5], renewable resources in developing countries represent an opportunity in four significant sectors like socioeconomic development, energy access, energy security, and climate change mitigation. Nevertheless, to take advantage of these benefits, these countries first have to overcome some barriers, which can be classified in market failures like monopolies, externalities, and information asymmetry; and system failures such as legitimizing the process, lack of capabilities and social particularities, as suggested by [6].

The remainder of Chapter 1 provides a brief introduction to the challenges that renewable energies represent for energy and power systems and how they have been addressed

globally, the current situation of Peruvian regulations regarding renewable energies, the kinds of problems this dissertation aims to address, and the thesis outline.

1.1 Renewable Energy Challenges

On the one hand, both operation problems: real-time operation and dispatch scheduling, must deal with the intermittent behavior of solar and wind resources. It is known that while penetration level increases, the system gets more vulnerable to RER generation fluctuation, as studied in [7]–[11]. Cited studies show that short-period fluctuations in solar irradiance or wind speed change drastically load demands as seen by the system operator creating a significant ramp up or down, which make it compulsory to have must-run machines or energy storage systems in the grid. Also, the unpredictable availability of the generation capacity of this kind of plant makes the scheduling process of day-ahead electricity markets difficult. Sometimes, complexity is such that the system operation coordinator assumes a deterministic generation time-series that plant owners prepare and send to them, avoiding the responsibility to accurately forecast this input, as evidenced in [12], [13] for the Peruvian case. This problem affects physical markets and the wholesale electricity market, as studied in [14]. Plenty of approaches have been studied in the literature [15], [16], [25]–[28], [17]–[24] to overcome these problems.

Although uncertainties are present everywhere in the energy sector's value chain, it should be noticed that the farther the time horizon analysis is, the more challenging it becomes to resolve the problem. Thus, on the other hand, system expansion planning problems are at a higher level of complexity [29], [30].

Furthermore, successfully overcoming these problems is conditioned to information on resources measurements and existing plants' historical records. Unfortunately, developing countries that are recently increasing their investment expenditures in RER plants usually lack this information. Although private agents make their field studies, the gathered information is not socialized, and public and academic organisms are obliged to resort to other methods to develop their activities.

Indeed, this is of the points on which recent works on barriers to renewable energy in developing countries strongly coincide. Specifically on the lack of capabilities [6] and information barriers [5], since these countries do not have access to databases on the

potential and production of renewable energies. Neither do they have the capabilities to develop proper mathematical models in order to study its phenomena.

1.1.1 Challenges addressed in Energy Planning

The evaluation of the impacts and the feasibility of the inclusion of renewable energy resources within energy systems have been studied in various works. In [31], a methodology based on the joint and iterative use of two software is proposed. One is EnergyPLAN, which simulates the national energy system on annual, monthly, and hourly time scales; and the other is GenOpt, which solves the problem of optimizing the selection of policies and technical measures using the outputs of the previous software as inputs and also updating its inputs as a result. Consequently, it tries to find the optimal mix of resources for the energy matrix, although it does not consider the variables' uncertainties.

Unlike the previous one that performs a static optimization, a multistage study is presented in [32], where the impacts of implementing renewable energies within the French energy system are studied. In that work, in addition to considering thirteen cut-off years for the analysis until 2050, it is also roughly included the variability of wind energy sources by considering representative scenarios to model the seasons of the year that affect demand and availability of wind resources among others. Due to limitations of the TIMES model that is used, it is not possible to correctly represent the restrictions of the short and medium-term operation, nor to simulate the 8760 hours of a typical year.

More specific cases are also treated in the literature. So in [33], the implications of implementing large-scale wind generation in energy systems with a high presence of nuclear plants are analyzed. This study makes use of the EnergyPLAN and MATLAB software set. Consideration of uncertainties in wind energy production in this research is achieved by using the Monte Carlo method. Several time-series of wind energy production are generated by performing the same number of runs of the energy software and consequently obtaining different answers. However, as the study's objective was not to find an optimal plan, conclusions were limited to describing the implications of each result.

The specific safety implications of planning electric power systems in the presence of variable renewable generation are discussed in [34]. In that research, a similar approach

to the previous cases is used, using two models: TIMES and PLEXOS. The PLEXOS model allows an evaluation of security restrictions evaluating a typical year of operation, while the TIMES model performs the generation of scenarios combining the various expansion policies of the system. The uncertainty of renewable generation is not considered in the study, and on the contrary, the same annual production profile is used for all the alternatives.

Another problem of interest for countries with a lower degree of electrification is the energy planning of isolated systems, which is presented as an alternative to the high investment costs of the traditional expansion of interconnected energy systems. In this field of study, [35] presents a multi-criteria optimization model for selecting the optimal mix of energy sources for the province of Cajamarca in Peru. The study identifies opposing objectives, assesses the importance (weights) of the alternatives for each objective, and finally uses a distance technique to select the optimal combination. The representation of the system, located in an isolated place, lacks detailed modeling of routes and is presented as an energy balance; also, uncertainties are not considered in this work. A complete review of configurations, models, and optimization techniques used in isolated systems planning, is presented in [36].

1.1.2 Challenges addressed in Power System Planning

Regarding the planning of electrical systems, several contributions have been made to consider the integration of variable renewable energy sources. In [37], the penetration of renewable energies is maximized by penalizing the waste of its generation within the transmission expansion planning. However, the problem is addressed in a weekly resolution, ignoring the variations of the wind resource in a shorter time scale. Another approach is presented in [38], where probability density functions are considered to randomly generate wind generation and demand values (using Monte Carlo). With these values, several scenarios are created and used to run the optimal flow sub-problem whose average result will then be used by the master problem that optimizes investments in transmission within an iterative process.

A more comprehensive approach is presented in [39], where generation planning is also carried out evaluating the inclusion of power plants based on renewable energy resources. The variation in demand and renewable sources are taken into account from the analysis

of historical data, which are reduced to twenty-four operating points through a K-means clustering process to make the optimization problem treatable.

Consideration of energy storage systems has also been integrated into the planning of transmission systems, thus in [40], the use of batteries is included as a mechanism to reduce the waste of variable renewable generation and its impact on the investment cost of the transmission system by producing the effect of "cutting peaks and filling valleys" that reduces the required capacity of the installations. The problem case study is based on academic test systems, and its horizon analysis comprises 168 points that make up the 24 hours of the seven days of a week. Another approach is presented in [41], where energy storage systems are optimally located to minimize the construction of transmission lines by considering energy storage as a generation plant. The treatment of this problem considers 48 operating points being 30 min each time step, and application is performed over an academic 6-buses system.

On the other hand, contributions to the operation of electrical systems have shown the security problems caused by the inclusion of intermittent renewable energy sources. In [42], an optimization model is proposed to perform an optimal short-term dispatch, considering within its formulation the restrictions necessary to identify the spinning reserve that can cope with sudden variations in wind resources. Similarly, another study is shown in [43] that determines the machines that must operate to respond to the intermittency of wind power generation in northern Peru.

As seen, when using mathematical models to develop expansion plans considering renewable energies, it is necessary to care about the dimension of the data. A correct data dimension reduction process allows the problem to be tractable and preserve the behavior and impact of renewables in the power system along the planning horizon, focusing on the short-term operation, which gets to be the most challenging part of analyzing due to computational limitations.

1.2 Motivating Case: Peruvian Power System

Peru became a South American region benchmark in 2008 when the Legislative Decree N° 1002 for the Promotion of Electricity Generation Investment with Renewable Energies was approved by the Peruvian Government. However, since that milestone and despite

the critical ratio of new renewable projects development and the clarification of the challenges and opportunities these technologies represent, no progress has been made on regulations.

1.2.1 Overview of Peruvian Power System

According to the report [44] published by the Peruvian ISO (Independent System Operator), in 2019, the maximum electrical load demand reached by the Peruvian System was 7,017.57 MW representing a growth of 1.93% year-on-year compared to 2018, where maximum demand was 6,884.59 MW. As for energy, total production in 2019 was 52,889.14 GWh, 4.08% more than the amount generated in the previous year.

From this amount of energy, 1,646.16 GWh (3.11%) was injected by wind plants whereas 761.73 GWh (1.44%) by solar photovoltaic (PV) plants, totalizing a share of 4.55% for renewable energies.

Additionally, the sector regulator published the yearbook [45], which mentioned that the Peruvian electricity sector represents a US\$ 5,065.26 million market, which is composed of regulated users (residential, commercial, and industrial) who pay a regulated tariff, and free users (industrial) who pay a tariff established by contracts signed directly with distributors or generators. The regulated market represents 60.8% of the total electricity market, while the free market is 39.2%.

Renewable capacity installed in the Peruvian system is 660.5 MW, representing 4.98% of the total installed power of 13,255.3 MW in 2019. Hydropower has 38.95% of total capacity with 5,163.1 MW, while thermoelectric plants achieve a significant generation park with 7,431.7 MW (56.07%). A curious fact reported by [44] is that between 2018 and 2019, no new renewable capacity was added.

Regarding the quality of supply, the system regulator presented in [46] that the System Average Interruption Frequency Index (SAIFI) was 5.12 times a year for Lima, capital of Peru, whereas SAIFI for the rest of the country was 16.5 interruptions by year, that is, more than triple. Moreover, Lima's System Average Interruption Duration Index (SAIDI) value was 14.9 hours a year, while the same index increases up to 36.9 hours/year for the rest of the country. Reported values correspond to the operational year 2018.

1.2.1.1 Projections

The Peruvian ISO published the study [47] on the growth projections for the Peruvian electricity sector until the year 2032. This document analyzes the historic electricity demand from 2008 until 2019, resulting in a Compound Annual Growth Rate (CAGR) of 4.90%. In addition, the ISO estimates a CAGR of 3.41% for the period from 2019 to 2032. This projection means that load demand for the year 2032 will reach 10,995 MW, an increase of about 4,000 MW in the maximum demand from the levels of 2019, as depicted in Figure 1.

In the same document, the ISO indicates that an additional generation capacity of 2,100 MW should be installed to meet system requirements.

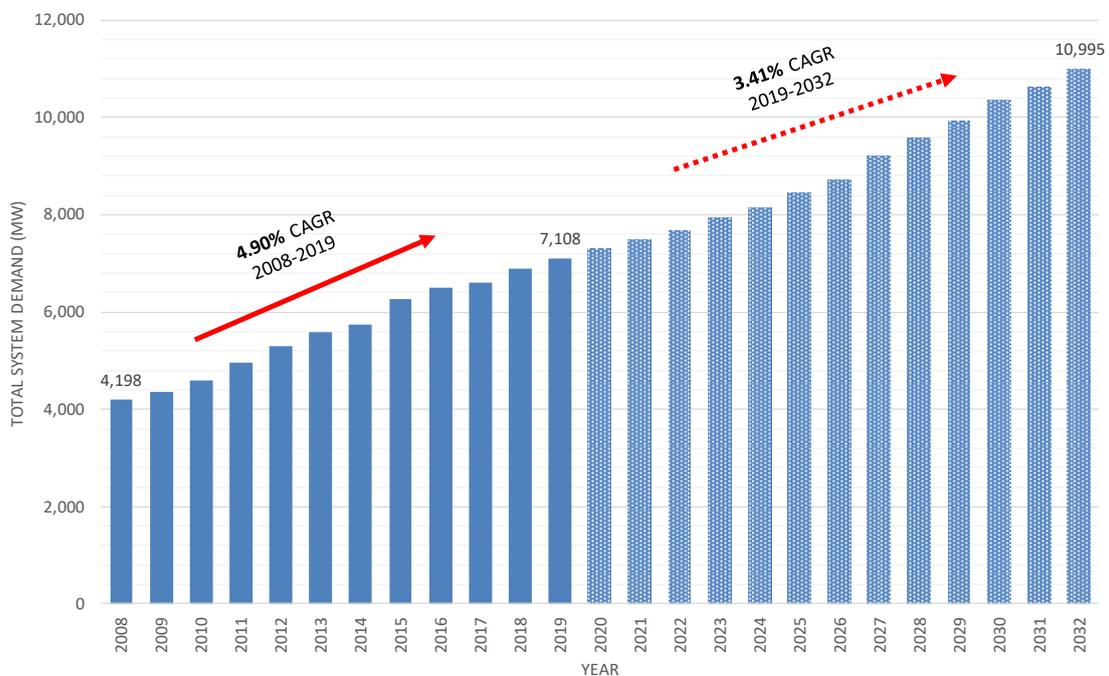


Figure 1. Historic and projected growth for Total Peruvian System electricity load (based on [47])

The actual pipeline of new generation projects presented to the ISO, which in the case of Peru is also the organism in charge of evaluating and approving new installations transmission and generation in the system, accounts for more than 20,000 MW submitted between 2019 and 2021 inclusive, as reported in the web of ISO [48]. From this total capacity, 72.2% belongs to wind and solar PV projects, with 7,615.20 MW and 6,918.70 MW each technology.

1.2.2 Peruvian regulations regarding renewable energies

As evidenced in the previous paragraph, Peru has a significant renewable portfolio, which is even larger than total installed capacity, resulting from private initiatives, given that the government has not announced new tenders since 2015 [49]. Indeed, more than 14 GW of non-conventional and intermittent renewable projects would be installed in the Peruvian system in the following years if they get ISO approval.

However, there is a concern about the impact that renewable projects will cause on the system, as has been stated by the Peruvian ISO in [50] where they mentioned that renewable plants cause a decrease in short-circuit levels, reduction of the inertia of the system, displacement of the dispatch of thermal plants by technical minimums, among others. In conclusion, the ISO proposed the definition of a maximum renewable injection value for each bar, which in practice restricts and delays the implantation of new projects, contrary to national interests.

Legislative Decree N° 1002 announced a national interest to develop and include renewable energies within the electricity generation matrix. It also established that renewable plants must be considered with zero marginal cost and therefore have dispatch priority. Despite this, little or no adjustments have been made to the planning methodologies of the national electricity system.

On the one hand, the criteria and methodology for preparing the Transmission Plan, established by the Ministerial Resolution N° 129-2009-MEM-DM, indicates that the electricity demand to be considered in the study represents the maximum annual values for each year of analysis within a horizon of 10 years. Then, the national planning methodology for the transmission expansion uses hour blocks to represent the behavior of load demand within a period to simulate an economic dispatch. This official methodology also specifies the construction of demand scenarios to perform a sensibility analysis and obtain a robust plan. The Transmission Plan is elaborated every two years by the ISO.

On the other hand, the general criteria for the definition of the Investment Plan of sub-transmission, established by Regulator Resolution N° 217-2013-OS/CD, also contemplates a 10-year analysis horizon using a single demand value for each year, corresponding to the maximum demand. However, it does not consider hour blocks,

economic dispatch, or a sensitivity assessment within the methodology. Each electricity distributor must prepare an Investment Plan every four years for its concession area to be evaluated and approved by the system Regulator.

As evidenced, current Peruvian regulations regarding planning processes are not prepared to consider the implantation of new renewable technologies such as solar PV and wind plants which have an intermittent behavior that cannot be represented by the use of annual maximum values or hour blocks. This fact also limits the evaluation of complementary equipment like energy storage systems. Beyond the modification of time scales both for national and sub-national planning methodologies, planning regulations should consider the use of specialized mathematical models that allows the analysis of new expansion technologies in addition to transmission lines.

1.3 Research Statement

To better understand the opportunities and challenges that renewable energies could have within the power systems of developing countries, this thesis develops an expansion planning framework for the medium and long-term horizon. This dissertation is not designed to be a comprehensive study of every aspect of power system expansion planning in the presence of renewable energies but focused on critical issues identified as the most critical challenges in developing countries. Specifically, this dissertation addresses the following research questions:

- What methodologies should be used to estimate the behavior of solar PV and wind renewable plants in any part of a country in a context of scarcity of information?
- How to reduce the data dimension of a planning problem for the expansion of a power system in the medium or long-term to allow the use of detailed mathematical models of non-conventional renewable energy plants, energy storage systems, and AC short-term operation constraints?
- What is the medium or long-term power system expansion planning optimization model representing renewable plants, energy storage systems, and AC short-time operation constraints?

Any proposals must be consistent with sound scientific and mathematical theory and include two essential aspects. First, it has to be the result of studying the state of art on

solving the problems mentioned above, focusing on the existing barriers in developing countries. Second, any proposals should include an application to an existing case and, if possible, a comparison of the results with historical information or specialized software for power system simulation. In the first element, proposals may not be fully generalizable to countries different from Peru because information accessibility and technical criteria will depend on the studied system.

1.4 Thesis Outline

Chapter 2 provides a literature review about the approaches presented to address the questions mentioned previously. From the analysis of these works, the main gaps in the literature are found. Proposals to fill these gaps are described and discussed in detail in subsequent chapters.

Chapter 3 presents synthesis methods to generate hourly time-series for solar PV and wind plants. The construction of these methods considers the absence of historic resource measurements, as is the case of developing countries. Based on this condition, complete methodologies for gathering information, randomly generating time-series for each resource, and assessing the power produced by solar PV and wind plants located in any place over the country are presented. The chapter concludes with a discussion about the results obtained after applying the presented methodologies in different places worldwide and its comparison with historical values.

Chapter 4 addresses the challenges produced by the dimension of the data sets involved in expansion planning problems, specifically when it is vital to study the impact of increasing the penetration of intermittent renewable plants into the power systems. This chapter examines a cutting-edge method for clustering energy-related time-series that preserve the correlation, simultaneity, randomness, and variability between all the elements present in the power system. The chapter concludes with applying the presented method to an existing system located in the northern part of Peru, which comprises residential and industrial loads and projected solar PV and wind power plants.

Chapter 5 studies the construction of a novel expansion planning optimization model for power systems to analyze the impact of expansion decisions for the medium or long-term within the short-term operation. This chapter proposes the equations required to represent

the principal expansion alternatives of a power system such as transmission lines, capacitors, solar PV and wind plants, and battery energy storage systems. All these equations accomplish to preserve the linearity of the model, at the same time that achieves a correct representation of the AC-related phenomena like the voltage, losses, and reactive power. The chapter concludes with the analysis of the results of applying the proposed optimization model into the system used in the previous chapter, which demonstrates how well the model manages the trade-off between the technical and economic benefits obtained when new equipment is implemented into the system and the investments and operation cost involved in that decisions.

Chapter 6 examines the joint application of the main contributions of this thesis presented in previous chapters into a new existing system located in the eastern part of Peru. The chapter first synthesizes hourly time-series for all identified expansion alternatives of renewable plants. Then, the proposed clustering process is applied to reduce the planning problem's data dimension. The chapter concludes with applying the proposed optimization model, which achieves an optimal plan to resolve the capacity problem of the case study.

Finally, Chapter 7 summarizes the findings and presents the conclusions and recommendations for further work.

2 Literature Review

Expansion of global generation capacity has been dominated by renewable plants for many years [51]. In that sense, plenty of works have treated the problems that intermittency of these non-conventional resources cause to power systems. However, most of the studies started from the existing conditions of developed countries. Since principal renewable investors have recently switched their focus from developed countries to developing countries [51], a significant gap has been evident because of a lack of information, measurement, and technical capacity about renewable resources [5], [6]. This chapter demonstrates the need for further study on synthesis methods adapted to a scarcity of historical information to provide developing countries with the necessary input to start a planning process.

This chapter also addresses the existing literature gap on the mathematical modeling of relevant phenomena present on power systems that integrate renewable resources. The review reveals the need to deepen the study on new methods to tackle the intractability problem produced by the data dimension of expansion planning studies and new equations to represent the short-term operation of new expansion alternatives like renewable plants and storage systems that require small time steps.

The following sections provide a review of state of the art for these gaps, identify areas where existing methods can be improved, and motivate the focus of the research presented in subsequent chapters.

2.1 Synthesis methods

2.1.1 Solar Energy

One of the first proposals to model solar hourly radiation values was presented in [52]. The daily and annual periodicity of hourly insolation values was removed using a Fourier analysis, and time-dependent frequency distribution (TDFD) was employed to synthesize insolation values. Likewise, [53] presented an autoregressive-moving average (ARMA)

technique to generate synthetic values starting from monthly mean values of the clearness index $\overline{K_T}$.

A novel approach was described in [54], where it was demonstrated that the clearness index was the variable that induces the randomness into the series, in contrast with previously published researches that treat solar irradiance as a random variable itself. Then, an ARMA model for clearness index was used to generate synthetic irradiance values.

However, the previously mentioned works coincide in presenting complex statistical analysis and models that became too difficult to understand and implement in practical applications.

On the other side, [55] developed an artificial intelligence (AI) model to synthesize solar hourly radiation time-series. Although formulation does not require a deep statistical analysis of variables' interrelation, it is necessary to have historic time-series values to train the AI model. Indeed, most AI models have the same limitation. As reviewed in [56], AI models that do not need historic solar measurements require historical values for many other parameters.

Another reasonable attempt was presented by [57], where an energy output model was used to generate synthetic hourly radiation values. Although energy output is coherent with historical measurements, results showed that the power generation profile does not reflect actual stochastic solar behavior.

An interesting approach was developed by [58], who, through a simple model based on a beta distribution, generates a 10-min solar irradiance time-series. Validation of model showed that daily and monthly means between actual and synthetic series coincide and that 10-min solar irradiance profile preserves the actual behavior acceptably. However, the proposed procedure needs a historic hourly solar irradiance time-series to start.

[59] developed a simple algorithm to generate hourly solar radiation values implemented using a web page with the same aim. The algorithm needs as arguments distances instead of coordinates, and starting from that, hourly values are generated. Nonetheless, algorithm equations were explicitly defined for the southwest region of Western Australia. Thus applicability to other places is limited and must be carried out carefully since model calibration depends on comprehensive historical measurement data.

Another major group of works looks to generate synthetic time-series at high temporal resolutions. In that line, [60] proposed a model that uses a second-order Markov transition matrix (MTM) to synthesize a 1-min time resolution time-series. While final model only needs as input a daily clearness index (K_T), it is needed data at the same time scale of 1-min to calibrate the model.

Also, [61] presented a simple approach to generate five-minute global horizontal irradiance (GHI) and direct normal irradiance (DNI) interpolating hourly mean values. The method uses 1-min historical observed data (re-scaled to 5-min) to calibrate a model, which then is used to interpolate hourly mean solar irradiance values obtaining good results though validation is performed only in the Australian territory. The authors suggest that the method would be suitable for applying to Typical Meteorological Year (TMY) records.

However, the same authors later recommend using another attempt they proposed that combines Fourier series, autoregressive models, and white noise terms. This new attempt achieves to synthesize daily and hourly time-series since it was demonstrated that generated series include patterns that have not occurred in the recorded data but are equally as likely to occur, thus better suits evaluation and planning requirements [62]. The application of the described method needs to have historical data to force the generated series' distribution.

Recently, a novel approach to downscale DNI time-series from 1-h to 1-min that can be applied in any location without requiring any local adaptation was presented in [63]. After training the model with 14 years of measured 1-min DNI data, it only needs an hourly DNI time-series for any location.

2.1.2 Wind Energy

Similar to the case of the solar resource, research effort has also been made for wind energy. In that way, [64] compared six well-established approaches to generate synthetic wind time-series. According to their test results, authors concluded that methods based on independent and identically distributed values, one-step Markov model, two-step model, and Box-Jenkins model do not generate representative and accurate synthetic time-series. Therefore, recommend using the Shinozuka method or the embedded Markov chain model.

Despite these findings, plenty of works have been published later using supposedly not recommended methods. One of them is [65], where it is used a first-order Markov chain model to synthesize wind speed time-series obtaining good results with more than 90% of the agreement for statistical parameters between historical and synthetic series. However, to improve results, a higher-order Markov model must be used as recommended by the authors.

In [66], first- and second-order Markov chain models are employed to generate synthetic wind speed time-series for two localities in Malaysia. After comparing principal statistics such as mean, standard deviation, autocorrelation, Weibull distribution parameters, and spectral density of real and generated series, the authors concluded that both models have a good performance synthesizing wind speed time-series. Although the cited paper does not analyze similarity in wind speed profile, it must be said that neither the first- or second-order Markov model achieves reproducing actual behavior accurately. However, it is true that a higher-order or larger size transition matrix of Markov models better preserves statistical characteristics of historical data, as studied in [67]. An interesting analysis of the pitfalls of using Markov models to synthetically generate wind speed time-series to be later used in planning processes, specifically for energy storage planning, is presented in [68], where it is demonstrated that generated time-series lacks the persistence of actual data and would predict a radically different storage requirement.

A novel approach was presented by [69], who developed a model inspired by the foundation of Markov transition matrixes, obtaining very realistic wind speed time-series. For the calibration of the model, a set of 7 years' measurements was required.

Similarly, [70] proposed using a model based on Copula theory yielding good results that help perform steady-state analysis of power systems. A comparison between synthetic and natural energy generated by some existing wind farms in Europe was taken as a validation criterion. Regrettably, no graphs of real and generated time-series were provided in the publication, making it impossible to compare the reproduction accuracy of actual behavior. Another attempt using Copula theory was presented by [71], who also addresses the synthesis of wind direction. The statistical validations against one-year measurements data obtained positive results; however, the wind speed profile does not reflect an actual wind speed time-series.

Meanwhile, [72] presented a simple yet exciting methodology to produce synthetic wind speed time-series. Starting from the premise that wind speed comprises periodic deterministic and stochastic components, a five-step methodology is developed, supported by a genetic algorithm for tuning parameters. The paper proposal's main advantage is that the model only needs aggregate statistical parameters such as yearly mean, monthly mean, and monthly maximum wind speed. Generated time-series properly preserves actual wind behavior, although persistence is still not achieved.

Indeed, persistence is a characteristic that is not discussed in most of the mentioned papers, even though this characteristic can represent the difference between a completely random and a realistic wind speed time-series. An interesting approach to achieve this feature was presented by [73], who uses the optimization model of the Assignment Problem to re-sequence the generated values to accomplish persistence requirements. An hourly measured wind speed data was needed to resolve this optimization problem, similar to [69].

[74] also produced very realistic wind speed time-series using a model that combines Fourier series and ARMA to characterize seasonal trend and autocorrelation in residue, respectively. Nevertheless, again, three years' historical data is employed to train the model.

2.2 Time-series clustering techniques

The traditional way to perform power system planning studies consists of simplifying both model and data to make problems tractable. In that sense, load data usually is represented using demand blocks, and power equations are limited to model just DC flows, ignoring other electric parameters like reactive power, losses, and voltage [75]–[78].

However, recent advances in power generation technologies [1], [3] have brought plenty of new alternatives on capacity additions that go further than typical thermal or hydroelectric plants. In that sense, any current expansion planning study should contemplate as valid alternatives wind farms, solar PV parks, or even energy storage systems, technologies that make the technique of demand blocks obsolete since the new

relevant operation time scales are very short, in the order of seconds, minutes, 15-minutes, 30-minutes or 1-hour, depending on the planning horizon [29], [79].

Therefore, as has been declared in plenty of previous works [80]–[84], these new technologies require planning studies to be conducted in a lower time scale to preserve data variability and randomness. Also, time-series must be congruent between them because simultaneity and correlation among renewable resources and load time-series are crucial to evaluating the expansion plan's quality.

As these new technologies increase their penetration in traditional power systems, it becomes more necessary to downgrade time scale analysis to steps of seconds, minutes, or, at most, 1-hour because, as mentioned in [85]. Although planning is done for the medium and long-term horizon, results have to account for the short-term operational needs.

Nevertheless, the problem of shifting time scale analysis from demand blocks approach to 1-hour simulation step or lower is that optimization models get computationally intractable because of dimensionality.

One method to overcome this problem was presented in [86], where a novel way to use blocks and duration curves is proposed to preserve correlations between load and wind production. Although the proposal reduces the dimensionality of data, it has the disadvantage of forcing all locations to have the same correlation value, which is not realistic since wind resource and demand have different behavior on each bus bar.

Clustering techniques are being used in the literature to overcome this trouble. In [81], a comparison of the differences in investment decisions when using the duration curves modeling and K-means clustering algorithm (to produce representative hours) is performed. The results show that the objective expansion plan change depending on the chosen data modeling technique. Mentioned work reduces a year into 48 representative hours.

Another case is presented in [80], where it is also studied the benefits of using a clustering method (to produce representative days) over traditional load blocks. The mentioned paper demonstrates that clustering is a more suitable way to conduct planning studies since it allows evaluating flexibility performance of final systems, drastically reducing load shedding and curtailed energy—cited paper yields four representative days per year.

As indicated by [29], data aggregation for a short-term evaluation of the system could be done on representative days or representative hours. However, inter-temporal constraints like ramping or storing cannot be considered when selecting representative hours due to the loss of chronology. An essential finding of mentioned work is that the impact of short-term uncertainty modeling is less critical as the number of representative days increases. The application case considers a full-year data set which is then clustered in 10 representative days.

The use of clustering techniques is also helpful when planning large-scale real-size power systems that include renewable plants where dimensionality is an aspect to consider. An interesting case is presented in [87] where the transmission expansion planning of the European Continental South West (CSW) region, conformed by Portugal, Spain, and France, is performed. This investigation makes use of 4 clusters to model load and renewables.

While a clustering process to obtain representative days is the preferred technique for energy planning applications, there is no standard way to perform it. Thus, some authors have expressed the need to pre-group data before applying any clustering or aggregation technique over the complete data set. That is the case of [82], where an aggregation within seasons is performed to evaluate the requirement of dynamic probabilistic reserve in the Mexican power system. The mentioned paper produces three representative days for each season, although it uses a simple averaging process, not a clustering technique. A similar criterion is presented in [30], where a previous month clustering process is performed to define three seasons. Within each one, another clustering process is done to yield three representative days.

Several clustering techniques have been developed for different types of time-series data, as documented in [88]. Depending on the specific application, one or another technique will be preferred to be applied. A revision of the performance of some standard algorithms for load modeling is made in [89].

Another interesting approach is presented in [85], where the main objective is to reduce data dimension to evaluate interday storage devices, complementing typical intraday energy balancing. The authors proposed a modified hierarchical clustering technique similar to a discretization procedure but considering a dynamic time-step chosen

automatically by an algorithm. This proposal produces an equivalent of 28 representative days to perform expansion planning studies.

2.3 Expansion planning optimization models

An important insight from the previous section is that, despite applying clustering methods, most of the mentioned works only employ a simple representation of the transmission system. They apply the transportation model (refer to [90], [91]), which turns their applications into an energy balancing problem. Just a few of them implement a DC model. Indeed, most of the research on Transmission Expansion Planning (TEP) has focused on studying DC models, as evidenced by the comprehensive review presented in [78], where 84% of the more than 70 papers analyzed used this model. However, DC models can not represent all phenomena of power transmission, such as losses, reactive power, or voltage fluctuations. It is essential to mention that most of the literature about Generation Expansion Planning (GEP) does not consider any transmission network at all, as revealed by [75], where it was found that 81% of analyzed papers only consider a single node approach.

Some approaches to electrical system expansion planning in the presence of intermittent technologies like solar PV and wind plants suggest using relevant operational points to represent the production of renewable along the year. That is the case of [92], where four snapshots per year are used, or as in [93], which considers 36 operational points resulting from three blocks for each month. However, when considering the penetration of intermittent renewable energies and fast-operation technologies like energy storage systems, modeling of renewable generation cannot be addressed through hour blocks and must use time-slices of at most 1-hour, as indicated in [29], [79].

Of the models that consider the influence of the hourly behavior of renewables within the planning process, we have on the one hand the approach presented in [84], which performs an analysis of a whole year in hourly steps (intra-year) but in a small system and without a multi-year evaluation. Despite this, obtained results demonstrate how renewable plants could improve the quality of electric service in places where electrical faults and disconnections are widespread.

On the other hand, [82] presents multi-year planning that considers hourly time-series for each season and each typical day (within the season) instead of the whole yearly time-series. However, their analysis does not contemplate intertemporal or dynamic optimization, evaluation of the intra-year operation, and neither model adequately simultaneity and correlation between renewable generation and demand, since a unique renewable time-series is used for all typical days within the same season.

A decomposition method is required to tackle computational power limitations when both characteristics should be implemented together (multi- and intra-year analysis). For example, in [83], a DC model is applied for the South American region using several commercial software, or in [87], where a stochastic DC model is designed to study the expansion planning of the whole European CSW region. However, as mentioned previously, a pure DC flow analysis does not reveal all aspects that a planning study must evaluate when facing renewables and energy storage plants.

Some of the leading power supply security (i.e., generation availability) and flexibility (i.e., ramping requirement) concerns caused by the intermittent behavior of renewable resources have been studied in the literature [7]–[11], [43], [82]. However, other aspects must also be analyzed when facing the presence of renewable plants in the power system, such as reactive power flows, technical losses, and voltage quality. It is necessary to enter into the AC world to address these problems.

Plenty of proposals to represent electrical characteristics of power systems using AC models have been presented in the literature [94]–[100]. Some popular models are based on relaxations of nonlinear constraints and binary decision variables [99]; others use the so-called reformulation-linearization technique (RLT) [98]. Recent AC models are using the special ordered set method (SOS) [96]. However, all these approaches require the utilization of new fictitious variables that increase the dimension of the optimization problem, being not suitable for real case applications or when analysis requires using short time steps, as it is required for renewable and storage technologies, due to dimensionality computational problems.

An interesting approach was presented in [101], where an iterative procedure is used to solve a DC model that considers multistage analysis, network reconfiguration, and technical losses. This approach not just allow the optimization model to remain a mixed-integer linear model (without transforming it into a mixed nonlinear model) but also leave

the calculation of complex functions (used for technical losses) out of the optimization equations constraints avoiding increasing the complexity and dimension of the problem and its solution time.

2.4 The Gap in the Literature and Research Objectives

One of the common challenges that almost all South America and other developing countries face is the scarcity of solar atlas or an official source of information about solar resources within its territory. Hence, obtaining measurement data at scales of 1-min, 5-min, or even 1-hr is almost impossible most of the time, making it very difficult to adopt some of the models proposed in developed countries.

Similarly, public-access databases for wind resources are relatively rare in Latin American countries. As seen, most of the proposed models to synthesize wind speed time-series found in the literature make extensive use of historical data at the scale of 1-h or more frequently, which could be very difficult to find in developing countries.

Two simple but effective methodologies to synthesize hourly solar irradiance time-series and wind speed time-series based on aggregate input data are proposed in Chapter 3 of this dissertation to overcome this gap. These methodologies can be helpful for projects evaluation and system expansion planning purposes.

Another gap is found in defining a clustering technique to be used within an energy planning process. Although many applications have already used clustering techniques, there is no standard procedure for choosing the best clustering technique, metric, and other relevant parameters. In that sense, Chapter 4 explores the principal clustering techniques used in the literature, proposes using an index to evaluate the quality of any clustering process, and finally recommends some parameters tuning based on test results.

Finally, the last identified gap refers to the absence of an optimization model that allows the analysis of power systems, including renewable generators and energy storage systems, with the particularity that considers active and reactive power flows, technical losses, and voltage values. This dissertation addressed this gap in Chapter 5, where the model is constructed, presented, and tested on an existing system.

Therefore, this research aims to expand the literature on power system expansion planning with a particular focus on developing countries by meeting these goals:

- To develop a methodology to generate hourly electricity production time-series for solar PV and wind renewable plants using aggregate data as input.
- To define a clustering technique to reduce the data dimension of an expansion planning problem for power systems to allow the application of a detailed optimization model.
- To design an optimization model for the power system, generation and transmission, expansion planning that considers renewable plants, energy storage systems, and AC short-term operation constraints like voltage fluctuations, reactive power, and technical losses.

3 Synthesis of hourly renewable production time-series

This chapter presents methodologies to synthesize solar and wind production time-series based on aggregate input data using statistical and optimization methods to fill the information gap required for planning purposes in developing countries.

Studied technologies correspond to solar PV and wind since they account for more than 70% of projected new global additions, as estimated by [1].

The content of this chapter has been structured into three sections. The proposed methodology is presented in Section 3.1. Real application cases are performed and discussed in Section 3.2, whereas conclusions are reported in Section 3.3.

3.1 Methodology

3.1.1 Methodology for Solar Energy

The proposed methodology for generating synthetic solar PV power generation data consists of five steps, as depicted in Figure 2.

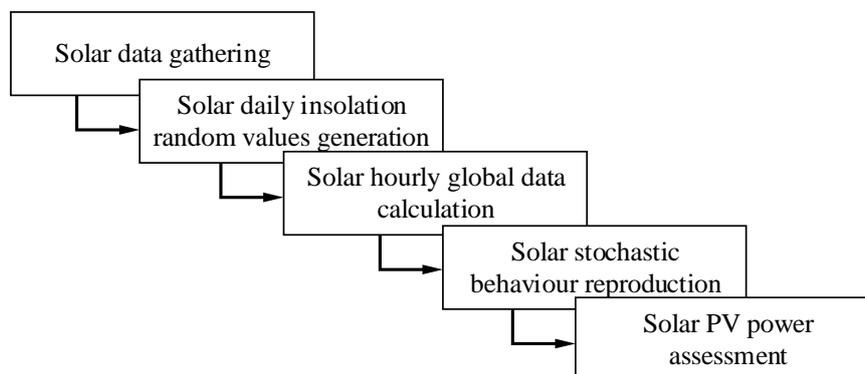


Figure 2. Methodology for Solar Energy

The first step seeks to gather relevant solar data about a specific geographic coordinate. The second step shows the statistical tool and settings used to generate random global

horizontal insolation values. Then, the third step describes the equations needed to calculate the hourly global radiation incident on the PV array starting from the values generated in the previous step. Subsequently, the fourth step presents the process of converting solar radiation into PV power generation values. Finally, step five models the stochastic behavior of clouds that pass over the PV panel and other phenomena.

3.1.1.1 Solar data gathering

Available open-access information about solar irradiance and insolation in Peru is limited to very few sources. On the one hand, the Peruvian Ministry of Energy of Mines released in 2003 a non-interactive solar map [102] that shows average month solar insolation values for the whole country, but the ability to extract precise values for any given coordinate is limited since it was delivered as static images instead of being implemented over a GIS platform.

On the other hand, some international companies and organisms provide this kind of information. While many of them are private commercial services (e.g., Meteonorm or Solargis), there is a couple of open-access services provided by the NASA called POWER [103] and by the European Commission called PVGIS [104]. This chapter employs five parameters provided by POWER service and one by PVGIS to complete this methodology.

For a given pair of latitude (ϕ in $^\circ$) and longitude (λ in $^\circ$), the monthly mean daily global horizontal insolation (\bar{H} measured in $kWh/m^2/day$) is obtained from PVGIS (average of all years available for each month will be used), while the maximum and minimum variations of \bar{H} (H^{max} and H^{min} in %), the maximum and minimum air temperature at 2 m values (T^{max} and T^{min} in $^\circ C$), and the surface albedo (ρ_g) are retrieved from POWER. Parameters mentioned above must be obtained for each month $m = 1..M$ as an average daily value.

3.1.1.2 Solar daily insolation random values generation

A free plug-in for Microsoft Excel called Ntrand [105] is used to generate random values for global horizontal insolation. This plug-in was created by Numerical Technologies as a simple but powerful tool that provides a complete suite of statistical functionalities.

One of the distributions supported by Ntrand is the Truncated Normal Distribution, which is proposed to be employed in this chapter.

A random variable derived from this distribution could be obtained using the following expression:

$$H_{syn} = \Phi^{-1} \left(\Phi(\alpha) + U \cdot (\Phi(\beta) - \Phi(\alpha)) \right) \sigma + \mu \quad (3.1)$$

Where Φ represents the cumulative distribution function of standard normal distribution, Φ^{-1} is its inverse, and U is a uniform random number. Parameters α and β are the normalized values of the lower (a) and upper (b) bounds of truncated normal for a given mean (μ) and standard deviation (σ).

The desired quantity N of random numbers could be obtained using the Nrand matrix function $NtRandTruncnorm(N, a, b, \mu, \sigma, 0)$ in Microsoft Excel, providing required parameters.

In an Excel worksheet, a $(N \times M)$ matrix must be constructed to obtain $N = 8760$ random values for every month ($M = 12$) of a typical year. The content of each matrix column should be the Nrand matrix function considering the monthly mean daily global horizontal insolation, the maximum, and the minimum values for the corresponding month m .

Notice that no value was not recovered for the standard deviation of this parameter. So, it will be assumed equal to the multiplication of a constant k by the mean value ($\sigma = k \cdot \mu$). In this chapter, the value of k is set to 0.5.

Therefore, random generation process results depend on the value of these four parameters. From these, maximums and minimums could be treated as quasi-fixed values since they come from historical measures. On the other side, mean value and constant k value should be considered factors that introduce uncertainty in the process and could be used to generate distinct synthetic time-series to represent multiple scenarios for the stochastic behavior of solar resources.

Two additional terms must be appended when calling the Nrand Excel function to avoid the problem of producing identical random values for places with the same aggregate input parameters. The final formula would be $NtRandTruncnorm(\dots, R_\phi, R_\lambda)$, where R_ϕ and R_λ are the last five digits of ϕ and λ starting from the right, which usually belongs to the decimal part when working with a precision of at least six digits. These constants are used in the function as random seeds.

Although different values for each hour of a typical year are obtained after this process, generated values are still a daily insolation value.

3.1.1.3 Solar hourly global radiation calculation

This third step explains the calculation flow needed to convert the global daily horizontal radiation synthesized in the second step into hourly total radiation over a tilted solar panel. Four primary relations will be used following the order presented in Figure 3 to achieve this objective.

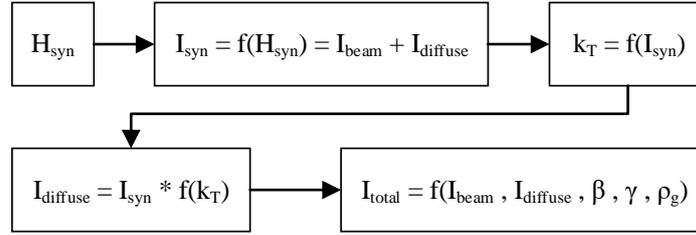


Figure 3. Calculation flow to convert Daily Horizontal Radiation (H_{syn}) into Hourly Total Radiation (I_T)

To start, it should be clear that total radiation incident on the PV panels (H) corresponds to the sum of direct or beam radiation (H_b), diffuse radiation (H_d) and albedo or ground reflected radiation (H_g). However, when working with horizontal radiation values, as synthesized in the previous step, the reflected component is neglected due to this physical disposition.

In that sense, the estimated hourly radiation value (I) obtained through eq. (3.2) from this global daily data will account just for beam and diffuse components ($I = I_b + I_d$).

$$r_t = \frac{I}{H_{syn}} \quad (3.2)$$

The definition of r_t and other needed values are explained in Appendix A, from eq. (7.1) to eq. (7.8).

After this first relation, the hourly global horizontal radiation I is obtained. This value could be related to the extraterrestrial horizontal radiation (I_o) by a ratio called the hourly clearness index (k_T) indicated in eq. (3.3). Special care should be taken to consider only positive values for I , and therefore, for k_T .

$$k_T = \frac{I}{I_o} \quad (3.3)$$

The equations needed for the calculus of I_o are given Appendix A, from eq. (7.9) to eq. (7.10).

The second relation gives us the value of the clearness index (k_T) for each hour. This index can then define the hourly diffuse fraction, which relates diffuse radiation and global horizontal radiation.

$$\frac{I_d}{I} = f(k_T) \quad (3.4)$$

$f(k_T)$ represents a piecewise function whose definition is presented in eq. (7.11) of Appendix A.

Diffuse radiation (I_d) is obtained applying the third relation presented in eq. (3.4). With this value, it is possible to clear the value of the beam radiation since $I_b = I - I_d$.

As suggested by [106], for PV panels placed on the southern hemisphere, the HDKR anisotropic model is suitable. Hence, the HDKR model will be used to calculate the total incident radiation on a tilted surface with slope β (in *rad*) and surface azimuth angle γ (in *rad*) as defined in eq. (3.5). For the sake of simplicity suggested model is used in this chapter, although a selection process could be carried out to choose the best model, similar to [107].

$$I_T = (I_b + I_d A_i) R_b + I_d (1 - A_i) \left(\frac{1 + \cos \beta}{2} \right) \left[1 + f \sin^3 \left(\frac{\beta}{2} \right) \right] + I \rho_g \left(\frac{1 - \cos \beta}{2} \right) \quad (3.5)$$

Complementary equations for this fourth relation are shown in Appendix A, from eq. (7.12) to eq. (7.16).

Having an Excel worksheet with the hourly total radiation values for an entire typical year, it is possible to use the Solver Tool to optimize the slope angle (β) to maximize estimated annual energy. This angle will be close to the latitude ϕ for most places in Peru. The reflectance of the ground (ρ_g) would be the month average recovered in step one.

3.1.1.4 Solar stochastic behavior reproduction

The effect of shadows produced by clouds that pass over the PV panel is represented by a factor f_{CLD} . Although clouds interrupt beam radiation, diffuse radiation continues to be present. Hence, only direct radiation is affected by factor f_{CLD} as indicated in eq. (3.6).

$$I_b^* = f_{CLD}I_b \quad (3.6)$$

Factor f_{CLD} is generated randomly for each hour using the procedure shown in (3.7). Note that global radiation hour-to-hour variation is achieved using Step 2, where random values are obtained, so additional noise is introduced in this step.

```

aux = 0;
for(i in 1..N) {
  if(aux ≥ rand1i) {
    fCLDi = aux;
    aux = 0;
  } else {
    fCLDi = 1;
    aux += rand2i;
  }
}

```

(3.7)

$rand_1$ and $rand_2$ are random numbers between 0 and 1 generated for each hour i by the non-volatile Excel matrix function $NTRAND(N)$.

3.1.1.5 Solar PV power assessment

Air temperature varies within a day; however, available data from the first step only provides the minimum and maximum values for each month. A relationship between global radiation and air temperature has to be used to approximate the value of air temperature for each hour of a day. In that sense, the proposed model presented in [108] will be adapted to this need.

$$\frac{\bar{H}}{\bar{H}_o} = c \cdot \left(\frac{T^{max} - T^{min}}{T^{min}} \right)^{0.5} \quad (3.8)$$

In eq. (3.8), the constant c is a function of the altitude and distance to the sea, and their coefficients are determined using historical data. In this chapter, the value of this constant will be estimated using available data from step one.

To find out the value of constant c , monthly mean daily extraterrestrial radiation \overline{H}_o has to be evaluated using the following expression:

$$\overline{H}_o = \frac{24}{\pi} G_{on} (\cos \varphi \cos \delta \sin \omega_s + \omega_s \sin \varphi \sin \delta) \quad (3.9)$$

Extraterrestrial radiation should be calculated for the recommended average day for months indicated in [106]. Consequently, twelve different values will be obtained for constant c . Thus, the average value must be considered.

$$c = \sum_{m=1}^M \frac{c_m}{M} \quad (3.10)$$

Prieto's proposed model [108] is then extended in this chapter to calculate hourly values. In consequence, the air temperature for a specific hour i is determined by Eq. (3.11). It is assumed that minimum temperature is the same for every day of the corresponding month and that the maximum temperature solved from the equation for each hour is the representative air temperature for that hour.

$$T^i = \left(\frac{I}{I_o} \cdot \frac{\sqrt{T^{min}}}{c} \right)^2 + T^{min} \quad (3.11)$$

Since the temperature has an inertial behavior, the actual temperature hour value and previous hour value are weighted using a factor z . In that sense, the final ambient temperature value is defined as shown in Eq. (3.12).

$$T_a^i = zT^i + (1 - z)T_a^{i-1} \quad (3.12)$$

Now, it is possible to calculate cell temperature (T_c) using eq. (3.13), which is the inferred equation presented in [109].

$$T_c = \frac{T_a + (T_{c,NOCT} - T_{a,NOCT}) \left(\frac{G_T}{G_{T,NOCT}} \right) \left[1 - \frac{\eta_{mp,STC} (1 - \alpha_p T_{c,STC})}{\tau \alpha} \right]}{1 + (T_{c,NOCT} - T_{a,NOCT}) \left(\frac{G_T}{G_{T,NOCT}} \right) \left(\frac{\alpha_p \eta_{mp,STC}}{\tau \alpha} \right)} \quad (3.13)$$

Nominal operating cell temperature ($NOCT$) is defined for solar irradiance $G_{T,NOCT} = 0.8 \text{ kW}/m^2$ and ambient temperature $T_{a,NOCT} = 20^\circ C$. Values for cell temperature under $NOCT$ condition ($T_{c,NOCT}$ in $^\circ C$), efficiency at maximum power point ($\eta_{mp,STC}$ in $\%$), temperature coefficient of power (α_p in $\%/^\circ C$) and cell temperature ($T_{c,STC}$ in $^\circ C$) under standard test conditions (STC) can be found in the solar PV module datasheet.

Although solar irradiance striking the PV panel (G_T) is an instantaneous value, the hourly total incident radiation (I_T) value obtained in step 3 will be used here, assuming that flat radiation occurs for the entire hour bin.

Finally, the following Eq. (3.14) is used to calculate the output power of the PV panel:

$$P_{PV} = P_{STC} f_{PV} \left(\frac{G_T}{G_{T,STC}} \right) [1 + \alpha_p (T_c - T_{c,STC})] \quad (3.14)$$

Rated capacity of PV panel under standard test condition (P_{STC} in W) is obtained from the module datasheet. Solar irradiance for STC ($G_{T,STC}$) is equal to $1 \text{ kW}/m^2$.

In this equation, a derating factor (f_{PV}) is employed to account for losses occasioned by soiling, mismatching, transformation, degradation, etcetera. Do notice that losses by irradiance level, temperature, and clouds are internalized in the proposed procedure, so f_{PV} should only account for other factors. Its value strictly depends on the place conditions where the solar PV panel is mounted, such as the level of dust, rain, among others. Typical values are approximately between 75-95% [110]–[117].

3.1.2 Methodology for Wind Energy

The proposed methodology for wind resources to generate synthetic data consists of four steps, as shown in Figure 4.

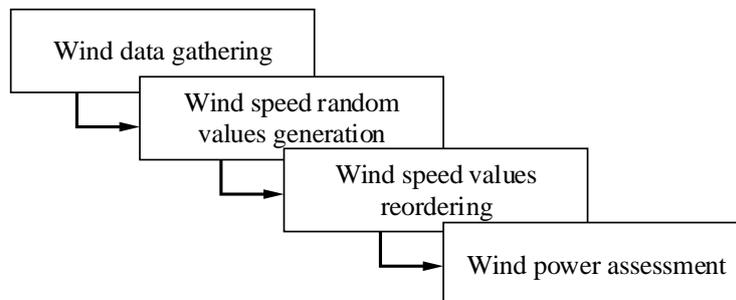


Figure 4. Methodology for Wind Energy

The first step gathers relevant wind data about a specific geographic coordinate. The second step shows the randomly generating process of wind speed values using information retrieved in the previous step. In step three, an optimization model is used to reorder randomly generated values of the second step. Finally, the fourth step presents the process to obtain wind power generation values.

3.1.2.1 Wind data gathering

The Peruvian Government released an interactive tool called Wind Atlas in the year 2016. This tool allows obtaining the yearly scale (β_y) and shape (α_y) parameters for the Weibull distribution approximation at 100m height. Correspondent average wind speed (\bar{v}_y in m/s) and standard deviation (σ_y) can be obtained with the function $NtWeibullStdev(\alpha_y, \beta_y)$ provided by the Excel plug-in Ntrand [105].

Additionally, air density value (ρ in kg/m^3) must be obtained for the total altitude (including wind tower) above sea level (h in m) for the selected location. Needed equations are presented in Appendix B.

Since Peruvian Wind Atlas does not provide monthly average wind speed values, the POWER [103] database will be used to acquire monthly averages ($\bar{v}_{m,p}$ in m/s).

Having these values, a parameter k_m is defined for each month relating the monthly and annual average of these values ($k_m = \bar{v}_{m,p} / (\sum \bar{v}_{m,p} / 12)$).

For other countries, the RE Data Explorer [118], a renewable energy resource geospatial explorer created by The National Renewable Energy Laboratory (NREL), could be used to gather wind speed time-series at different heights.

If another source of data were used, e.g., in-site measurement, Weibull scale and shape parameters could be obtained 1) using one of the methods presented in [119]–[121], 2) using the function $NtWeibullParam(\mu, \sigma)$ being μ the mean and σ the standard deviation of speed values; or 3) employing the $fitdistr(ws, densfun = "weibull", lower = 0)$ function in the software R [122] with package MASS [123], where ws is an array containing wind speed values. Monthly averages, as well, may be obtained from the same data source.

3.1.2.2 Wind speed random values generation

The inverse Weibull distribution will be employed to generate random values through the following expression (3.15), which is implemented in a function available in the Excel plug-in Ntrand [105]. v_{syn}^u represents the set of unordered generated values.

$$v_{syn}^u = \beta \left(\ln \frac{1}{1-U} \right)^{1/\alpha} \quad (3.15)$$

Where U is a uniform random number.

The desired quantity N of random numbers could be obtained using the Ntrand matrix function $NtRandWeibull(N, \alpha, \beta, 0)$ in Microsoft Excel, providing the required parameters.

In an Excel worksheet, a $(N \times M)$ matrix must be constructed to obtain $N = 8760$ random values for every month ($M = 12$) of a typical year. The content of each matrix column should be the Ntrand matrix function considering the scale (β_m) and shape (α_m) parameters values for the corresponding month m , which could be obtained with the matrix function $NtWeibullParam(\bar{v}_m, \sigma_m)$ having $\bar{v}_m = k_m \bar{v}_y$ and $\sigma_m = k_m \sigma_y$. These relationships represent a good approximation of actual parameter monthly variation as inferred from the data shown in [124].

Similar to the solar case, additional terms must be appended when calling the Ntrand Excel function to avoid obtaining identical random values for places with the same aggregate input parameters. The final formula would be $NtRandWeibull(\dots, R_\phi, R_\lambda)$, where R_ϕ and R_λ have the exact definition as indicated in 3.1.1.2.

Wind speed values generally are reported at 100m height. However, the wind turbine hub can be placed at a different height, so values must be scaled using the power-law presented in eq.(3.16) as suggested by [125].

$$\hat{v}_{syn}^u = v_{syn}^u \left(\frac{h}{h_{ref}} \right)^\gamma \quad (3.16)$$

Where \hat{v}_{syn}^u and v_{syn}^u are the wind speed values for heights h and h_{ref} , respectively. Power law exponent γ , also known as Hellman's wind shear or friction exponent, could be approximated using one of the formulas presented in [126], [127], but since both NASA POWER and site measurements most frequently report wind speed values for

more than one height, it is possible to clear the value of γ for each month using the monthly average speed values.

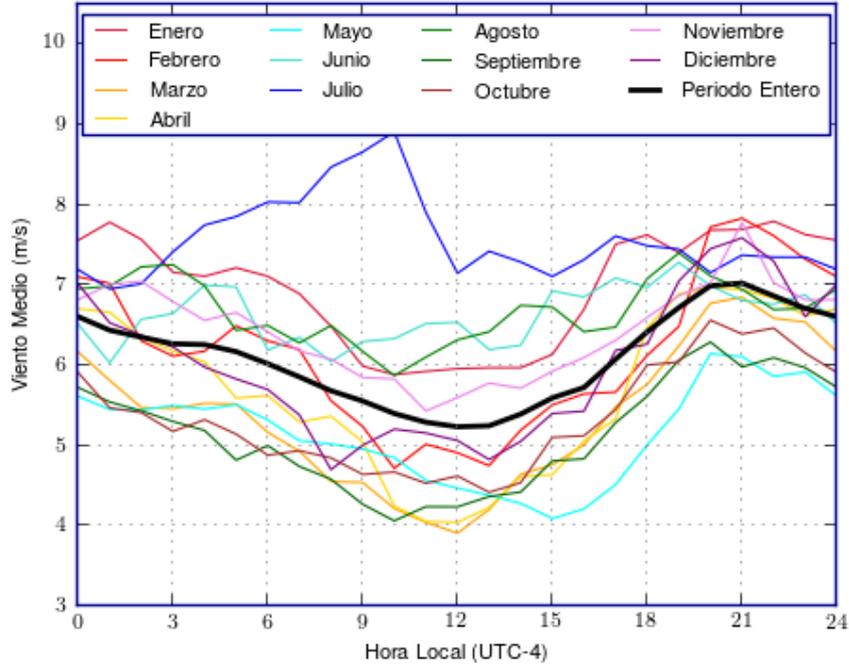


Figure 5. Wind diary cycle for a point in Chile [128].

Since natural wind behavior has a sinusoidal-like form, as evidenced in Figure 5, obtained for latitude -37.09° and longitude -72.67° from [128], diurnal pattern formulation proposed by [109] will be adapted and applied to synthetic values.

$$v_{final} = v_{syn}^o \left(1 + \delta \cos\left[\frac{2\pi}{24}(i - \phi)\right] \right) \quad (3.17)$$

In eq. (3.17), parameter $i \in \{1..24\}$ represents the hours of the day, ϕ is the hour of day at which peak wind speed used to occur, and δ is the diurnal pattern strength whose typical value goes from 0.0 to 0.4. v_{syn}^o is the correspondent synthetic value for the hour of year and month analyzed after being ordered using the algorithm presented next.

3.1.2.3 Wind speed values reordering

Until the previous step, generated values represent a random set of numbers, with almost all wind speed synthesis processes reviewed in section 2.1.2. However, it is possible to reorder the generated time-series to be more realistic, attending to the persistence characteristics that wind speed has.

In that sense, an algorithm that uses the Assignment Problem optimization model is proposed in (3.18), similar to [73].

```

a = 0
b = rand(20: 32);
while(a < N) {
     $\hat{v}_{syn}^{sub} = \hat{v}_{syn}^u[a: b];$ 
    c = rand(12: 24);
    for (j in 1..length( $\hat{v}_{syn}^{sub}$ )) {
        for (i in 1..length( $\hat{v}_{syn}^{sub}$ )) {
             $disordered = \hat{v}_{syn}^{sub}[j]/\max(\hat{v}_{syn}^{sub});$ 
             $pattern = (1 + \cos[\frac{2\pi}{\text{length}(\hat{v}_{syn}^{sub})}(i - c))];$ 
             $distances[j, i] = \text{abs}(pattern^2 - disordered^2);$ 
        }
    }
     $\hat{v}_{syn}^u[a: b] = \text{AssignmentProblem}(\hat{v}_{syn}^{sub}, distances);$ 
    a += b;
    if(a ≥ N) break;
    b = rand(20: 32);
    if(a + b > N) b = N - a;
}

```

(3.18)

This algorithm takes as input the values of \hat{v}_{syn}^u which was obtained in eq. (3.16) and replace its values in chunks. In each iteration, a subset of the generated time-series is created. The size of this subset is chosen randomly between 20 and 32 hours, which tries to represent a natural “diary” cycle length. After that, this subset is compared with a sinusoidal pattern whose hour of peak speed is set by a random function between 12 and 24 since typically wind blows stronger in the afternoon and night. The distance between the pattern and the normalized subset is calculated as suggested in [73].

With these values, the assignment problem is finally solved. As a result, it is obtained a reordered time-series which is used to replace the original correspondent values in \hat{v}_{syn}^u . This newly ordered time-series is represented by v_{syn}^o and is the one that must be used in eq. (3.17).

Parameters of the random functions can be revised to adapt to other locations' diary cycle patterns.

3.1.2.4 Wind power assessment

Although wind turbine manufacturers provide their products' power-wind speed (P-S) curves, no equation exists. In that sense, a piece-wise linearization function should be constructed to represent P-S curves [129].

$$P_W = f(v_{final}) \frac{\rho}{\rho_0} \eta_W \quad (3.19)$$

Eq. (3.19) shows the power calculation formula used to assess the generated power of a wind turbine for a specific wind speed (v_{final}). In this equation, ρ is the air density defined in step one (3.1.2.1), $\rho_0 = 1.225 \text{ kg/m}^3$ is the standard air density at sea level, $f(v_{final})$ is the piece-wise linearization function that models the P-S curve of wind turbine and η_W is a derating factor.

Typical η_W values are in a range of 80-90% [124], [130]–[135] and reflect the losses caused due to wake effect, availability, electrical efficiency, turbine performance, environment, and curtailments.

3.2 Validations

3.2.1 Validation of methodology for Solar Energy

Three tests will be performed. The first test will evaluate the accuracy of proposed data sources. Then, methodology output until the second step will be compared with a historical time-series of irradiance. Finally, an existing PV power plant will be simulated through the proposed methodology, and its results will be contrasted with actual values.

3.2.1.1 Test N° 1 – Comparison of solar data sources for a point in Chile

Both POWER and PVGIS databases employ complex models to approximate solar radiation for a wide range worldwide. Hence, values obtained from these places are not historic measurement values but a mathematical approximation.

In order to quantify how accurate are data provided by these platforms, they will be compared with values obtained from the Solar Explorer published by the Ministry of

Energy of Chile [136], which has a normalized root-mean-square deviation (nRMSD) of 5.7% (observed vs. simulated) for the northern part of Chile (between latitudes -17° and -30°), according to [137].

Values for the comparison are gathered for a place located at latitude -18.37° and longitude -70.14° .

Table I. Deviation of solar data sources respect to official values for a point in Chile.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
EU-PVGIS	-2.1%	-1.8%	-1.4%	0.5%	1.7%	3.8%	6.0%	5.5%	0.3%	2.6%	2.0%	-0.3%
USA-POWER	-0.5%	2.8%	-5.3%	-18.0%	-28.8%	-37.0%	-36.5%	-36.8%	-27.2%	-23.9%	-16.7%	-7.7%

Values offered by PVGIS-NSRDB (2005-2015) are significantly more accurate than those provided by POWER, as shown in Table I. NASA's data display information lower by more than 10% from Chilean's values for 8 out of 12 months.

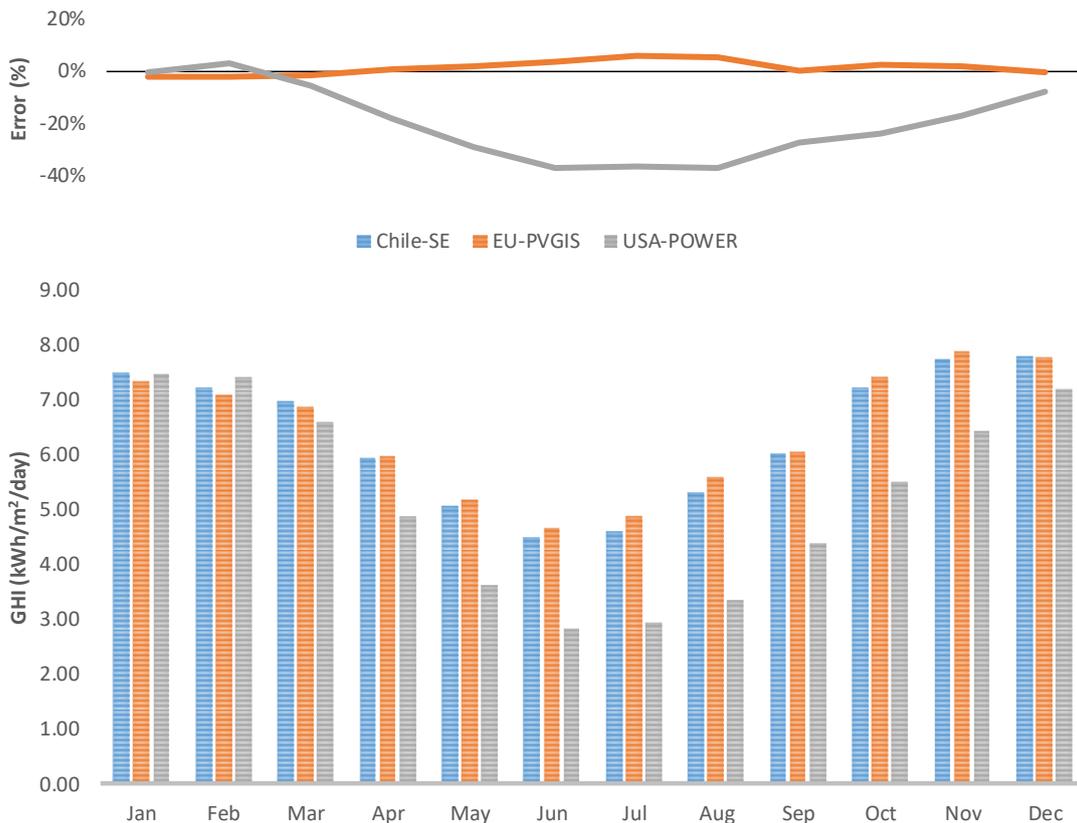


Figure 6. Comparison of GHI (kWh/m²) from different solar data sources for a point in Chile.

In that sense, it is validated that using PVGIS to obtain information about Global Horizontal Insolation (GHI) should be preferred over the POWER database, as proposed in step one (3.1.1.1).

3.2.1.2 Test N° 2 – Comparison of total GHI for a point in Slovakia

The second test validates the quality of synthetic hourly values generated. Hence, historic hourly measurement data is used to compute the total monthly and annual global horizontal irradiance received, then compared with the amounts produced by synthetic values.

Available data of Slovakia for years 2014, 2015, and 2017 were downloaded from [138] to perform this analysis. Geographic coordinates correspond to latitude 49.03° and longitude 20.32° , for which information was gathered from POWER and PVGIS-COSMO (2005-2015).

Figure 7 represents the daily behavior of measured and synthetic values only for the recommended average day for months indicated in [106], although both time-series correspond to a full year.

It should be noticed that daily behavior for the same month is different from year to year. Therefore, synthetically generated values roughly coincide with some days.

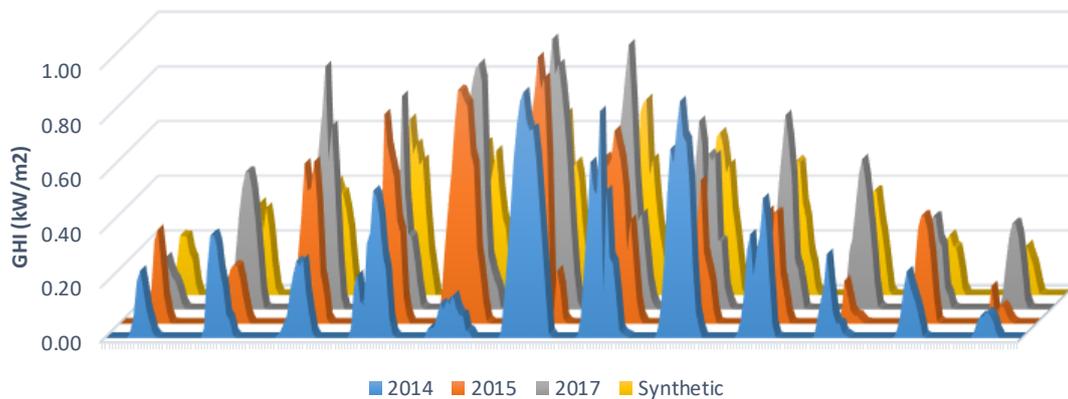


Figure 7. Time-series of solar global horizontal radiation (kW/m^2) for a point in Slovakia.

Figure 8 reflects that the total monthly global horizontal insolation received each month fluctuates year to year. Despite this, synthetic amounts seem similar to actual values for some months (Table II). The mean value of yearly nRMSD values comparing the total monthly amount of energy surpass 13.5%, having 12.6%, 9.8%, and 18.3% for years 2014, 2015, and 2017, respectively.

Nevertheless, if a mean aggregates the three historic yearly time-series, resulting amounts tend to be more likely than synthetic ones. Specifically, deviations get reduced to a mean absolute percentage error (MAPE) of 5% (each year independently reaches 10.9%, 7.1%, and 11.1%), while the nRMSD totals 10.6%.

Table II. Deviations of total monthly GHI (kWh/m²/month) values for a point in Slovakia.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
Avg	35.58	56.59	103.45	128.78	159.78	188.11	172.05	149.96	98.41	67.16	38.37	29.89	1228.13
Syn	35.01	56.05	102.16	139.97	152.69	155.01	168.13	140.56	97.82	63.95	38.38	26.41	1176.14
Var	-1.6%	-0.9%	-1.2%	8.7%	-4.4%	-17.6%	-2.3%	-6.3%	-0.6%	-4.8%	0.0%	-11.6%	-4.2%

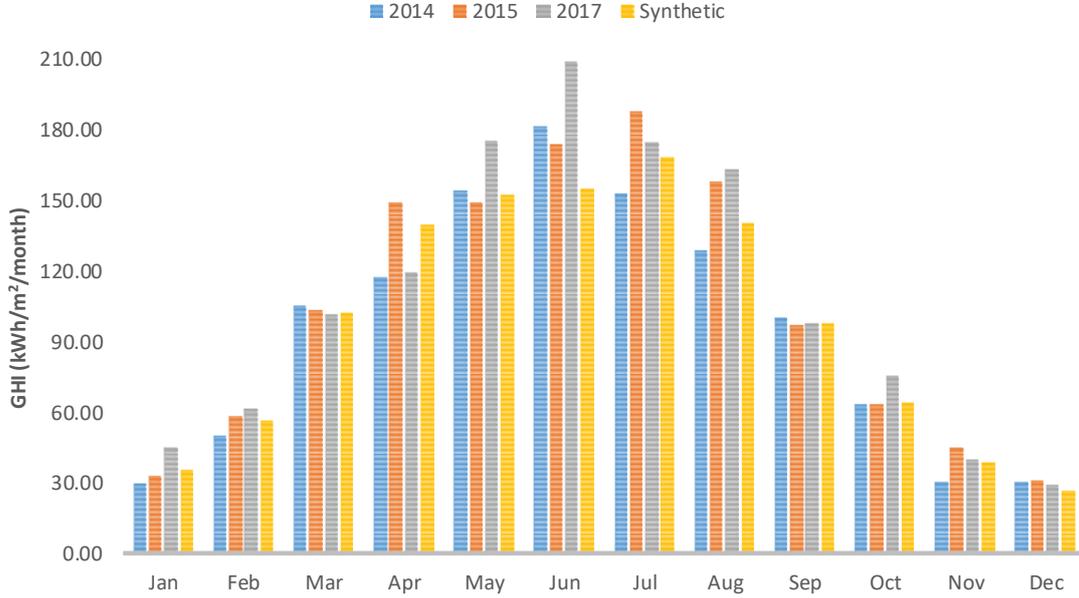


Figure 8. Comparison of total monthly GHI (kWh/m²) for a point in Slovakia.

These results validate the proposed procedure until step four (3.1.1.4) and also suggest that to deal with uncertainty, synthetic series should also be generated for monthly mean daily GHI (\bar{H}) values in a range around $\pm 10.6\%$.

3.2.1.3 Test N^o 3 – Comparison of solar energy with an existing solar park in Peru

The third test corroborates the results of all steps (from 3.1.1.1 until 3.1.1.5). In that sense, the historical power output of an existing PV plant in Peru is used to evaluate the quality of synthetic time-series generated for a similar plant.

The studied plant is Majes Solar Park (Figure 9) which is located at latitude -16.44° and longitude -72.22° , so data is gathered for this point from POWER and PVGIS-NSRDB (2005-2015).

This plant uses fixed PV panels sloped at 15° facing the north [139]. Although the existing plant comprises modules of 350, 370, 390, and 410 W manufactured by TSolar, simulation is made using technical specifications of Hanwha Solar Q.Peak Duo L-G5.2 for a rated power of 380 W [140]. A total of 57 894 panels are considered for the simulation, giving a total installed power of 22 MW, as it is in reality.



Figure 9. Location and satellite image of Majes Solar Park in Peru.

Solar resource for the plant's location is good and almost stable throughout the year, as shown in Figure 10. The yearly average daily total insolation received over a PV panel is 6,365 Wh/m²/day. Furthermore, the expected daily behavior of radiation is almost constant for every month of the year, as represented in Figure 11.

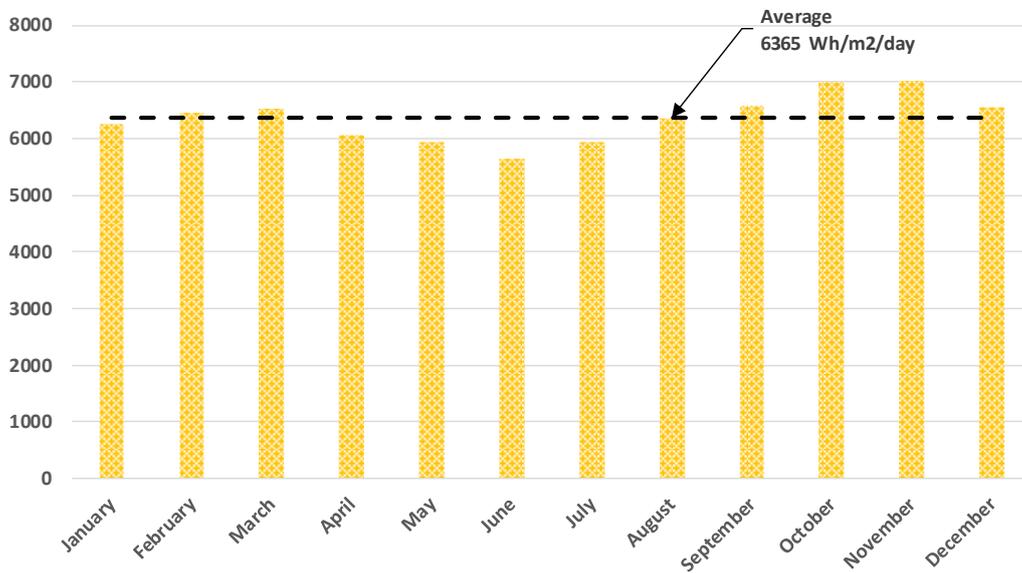


Figure 10. Monthly average daily total insolation (Wh/m²/day) over a PV panel for a point in Peru.

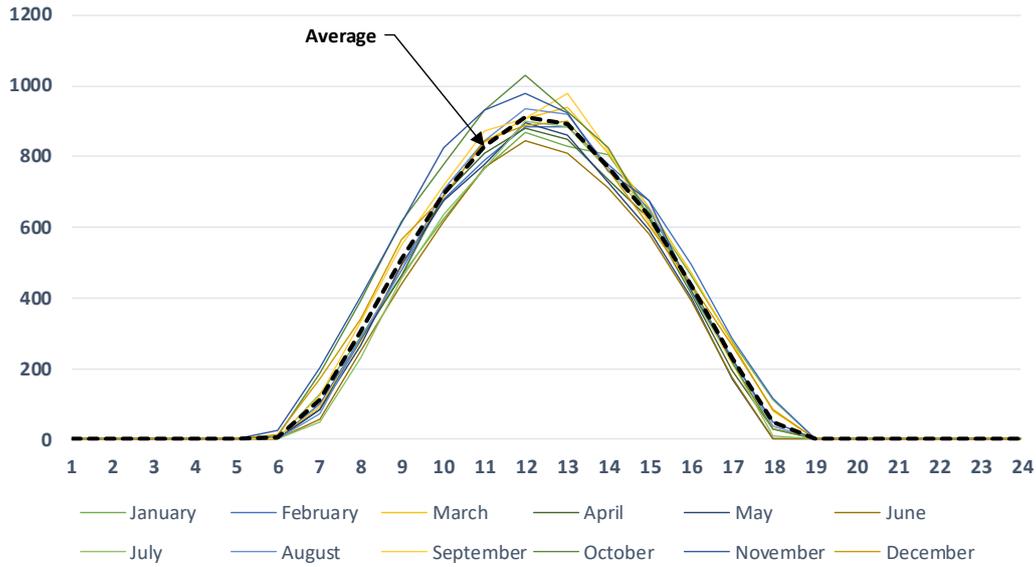


Figure 11. Monthly average daily hourly irradiance (W/m^2) over a PV panel for a point in Peru.

For the simulation, a value of 86% was used for the derating factor f_{PV} . Table III reports the total monthly energy produced (actual and synthetic) by the solar park. It also shows the variation between both values. Correspondent MAPE is 4.9%, while the nRMSD is 5.4%.

Table III. Deviation of total monthly solar energy output (MWh) values for a point in Peru.

	Actual	Synthetic	Variation
January	3630.41	3512.84	-3.2%
February	3345.25	3246.72	-2.9%
March	3474.82	3624.32	4.3%
April	3615.47	3280.31	-9.3%
May	3210.09	3319.55	3.4%
June	3164.76	3074.66	-2.8%
July	3061.85	3338.34	9.0%
August	3806.71	3551.88	-6.7%
September	3818.65	3555.02	-6.9%
October	4070.51	3888.83	-4.5%
November	3841.99	3774.42	-1.8%
December	3811.57	3667.79	-3.8%
Total	42852.08	41834.70	-2.4%
Mean	3571.01	3486.22	

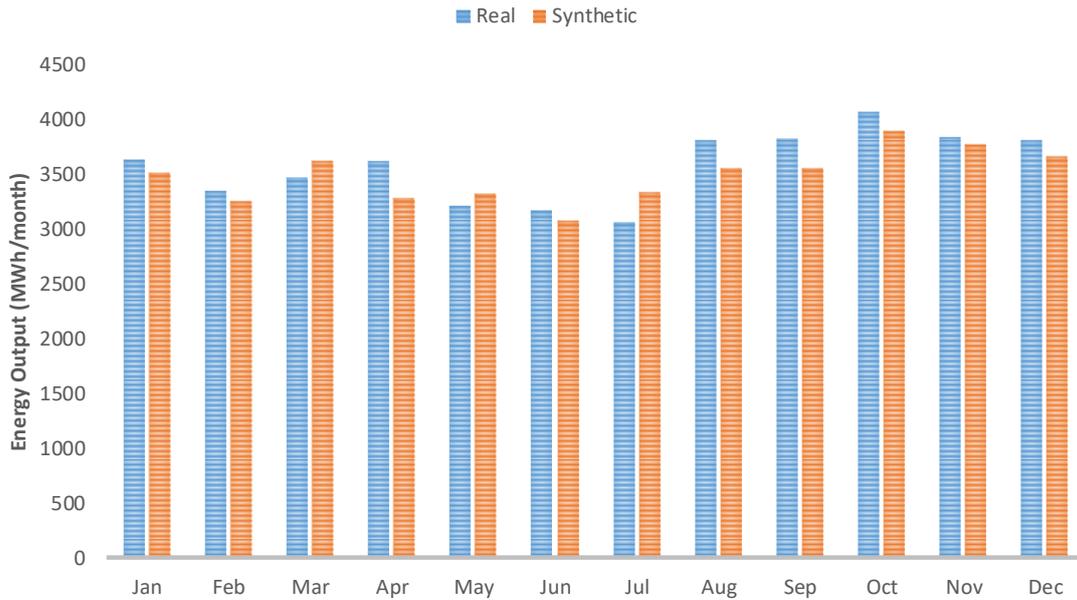


Figure 12. Total monthly solar energy output (MWh) for a point in Peru.

Besides validating total energy output between actual and synthetic values, a comparison between output power profiles is also needed. Figure 13 displays power output time-series for the first and second weeks of the year. It can be noticed that synthetic series achieves reproducing stochastic behavior of solar resources. It should not be forgotten that the proposed methodology does not try to adjust natural solar curves but to generate realistic probable ones.

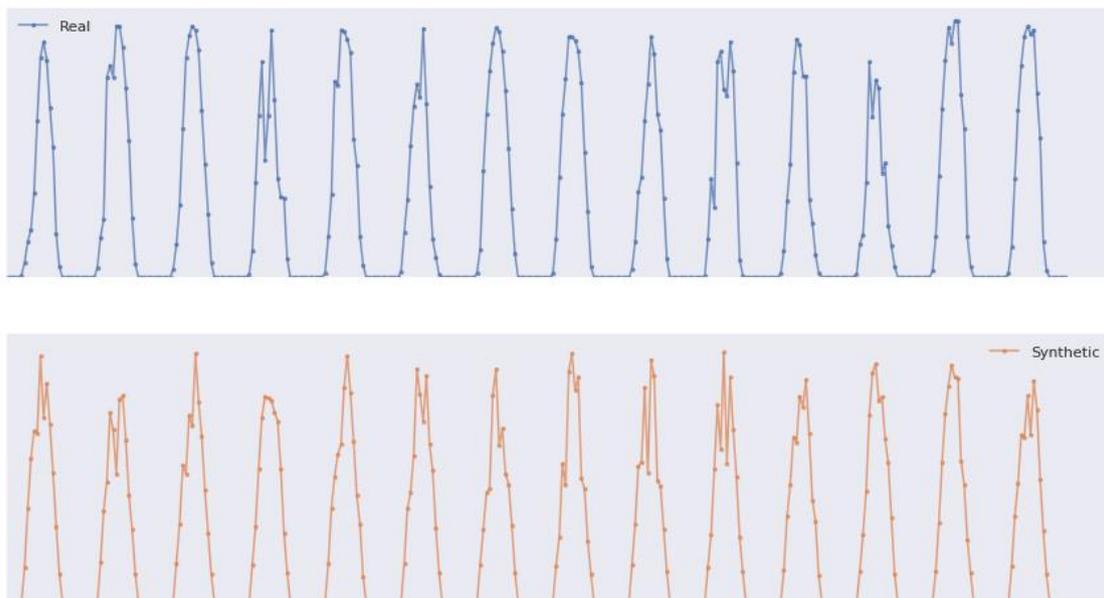


Figure 13. Time-series of total solar power output (MW) for a point in Peru.

3.2.2 Validation of methodology for Wind Energy

Three tests will be performed to validate the proposed procedure. Both tests will compare total energy produced when wind speed is converted using 42 wind turbines of a specific model for all data sources. Used turbine model corresponds to Siemens SWT-2.3-108 [141], a turbine of 2.3 MW nominal power, with a hub height of 80 m and wind cut-in, cut-out, and rated speed values of 3, 25, and 11 m/s, respectively.

3.2.2.1 Test N° 1 – Comparison of wind energy for a point in India

In contrast with solar resource data availability, gathering data for wind resources for free is not an easy task. The National Renewable Energy Laboratory (NREL) has developed RE Data Explorer [118], a renewable energy resource geospatial explorer. Although this explorer covers only a minor part of the world, it is a valuable tool to obtain free data.

For this test, a place in India at 248 masl, latitude 25.44° , and longitude 78.57° is analyzed. The RE Explorer allows obtaining a full-year measurement data of wind speed at different heights, besides other parameters.

Weibull parameters β_y and α_y are calculated using the obtained wind speed time-series at 100 m as the average result of the R function indicated in 3.1.2.1 and methods 1, 2, and 3 of [119].

Then, monthly means are calculated for wind speed at 40 m and 80 m, which is then used to clear the value of power-law exponent γ for each month. Starting from these monthly values for mean wind speed at 40 m and γ , correspondent mean wind speed values at 100 m ($\bar{v}_{m,p}$) is calculated. Now, it is possible to find the values of the parameter k_m described in 3.1.2.1. A value of $\phi = 22$ and $\delta = 0.2$ is used in this test.

Since the process is intended to be applied in Peru, where there are no free wind speed time-series at 100 m, monthly mean wind speed values at 100 m were inferred from measurements at 40 m, although data retrieved from RE Data Explorer give us measurements at this height. Then, a measure of accuracy is done between actual and approximate values.

Table IV. Deviations of monthly mean wind speed (m/s) values for a point in India.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual	4.995	5.255	5.159	5.485	7.093	8.751	6.828	6.984	6.592	4.697	3.774	4.250
Approx	5.133	5.419	5.302	5.645	7.293	8.855	6.864	7.016	6.684	4.866	3.977	4.418
Var	2.8%	3.1%	2.8%	2.9%	2.8%	1.2%	0.5%	0.5%	1.4%	3.6%	5.4%	4.0%

As shown in Table IV, approximate values differ from actual mean values 2.6% on average, having a maximum on November when variation reaches up to 5.4%, which is still acceptable.

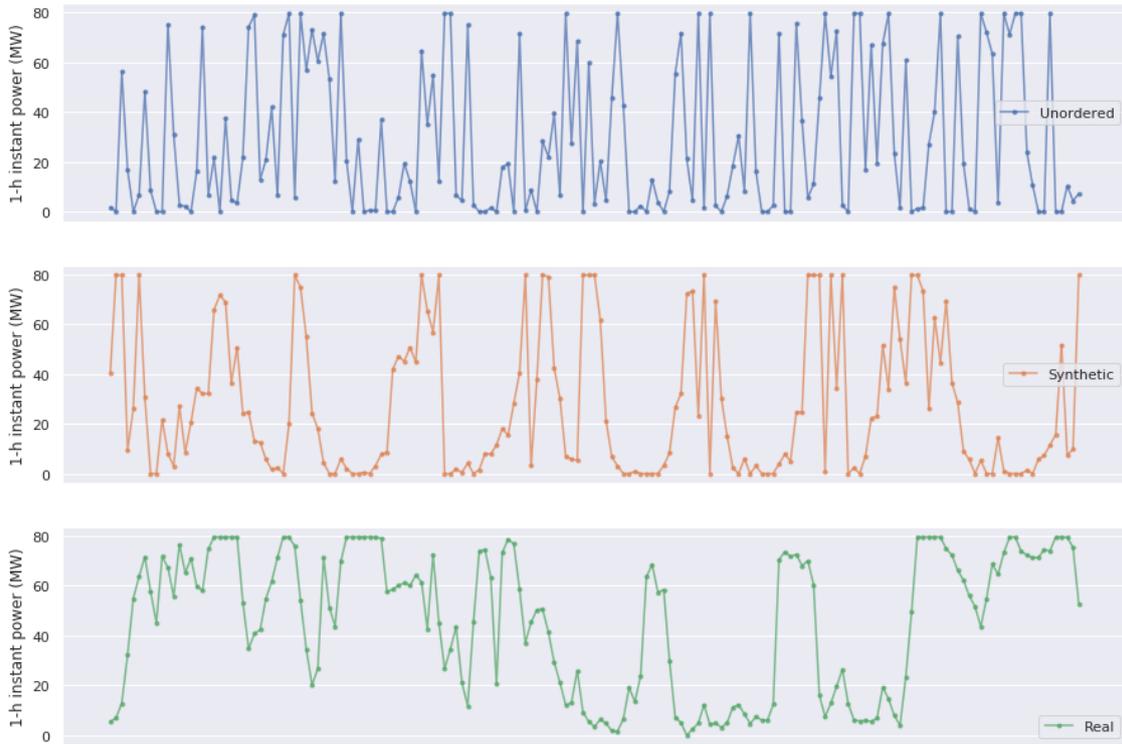
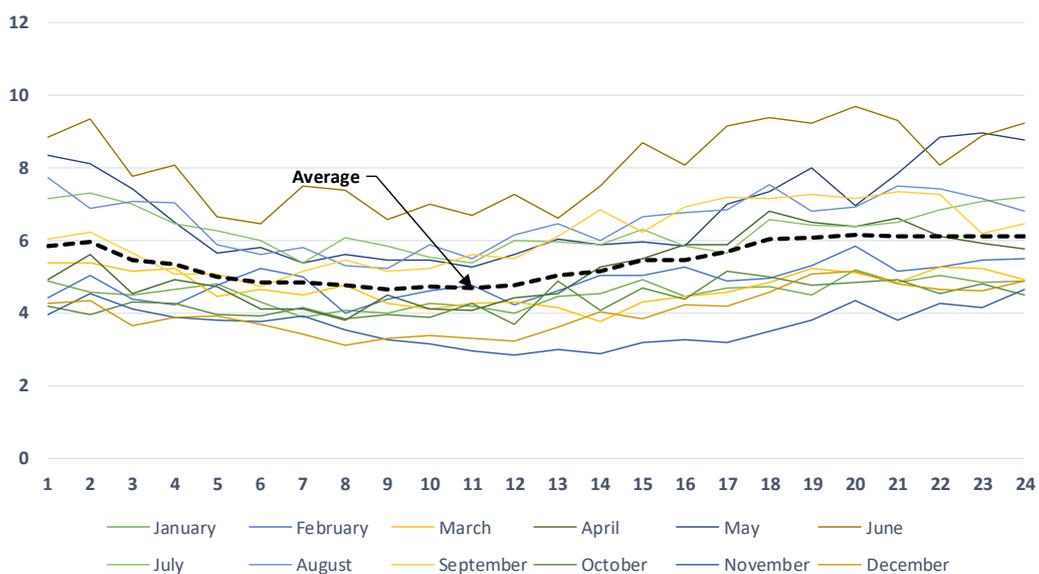
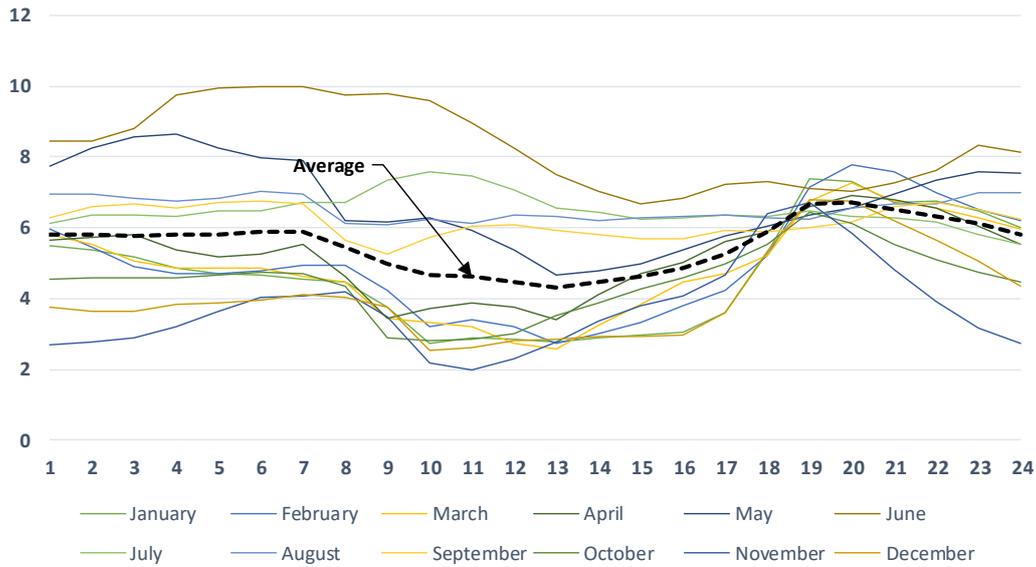


Figure 14. Time-series of total wind power output (MW) for a point in India.

If the power generation profile calculated using the unordered wind speed time-series is revised (Figure 14), it is found that it has a more erratic behavior than the actual one.



(a)



(b)

Figure 15. Monthly and annual daily (a) synthetic and (b) actual profile of wind speed for a point in India.

In this context, it could be said that the unordered time-series represents a more complicated-to-integrate-into-power system wind project since it will require a more robust and flexible system to absorb such variation in power. This, somehow, could be a positive attribute for planning activities but, in contrast, could increase the expansion plan cost.

A more realistic time-series is successfully obtained after applying the reordering algorithm indicated in 3.1.2.3, as shown in Figure 14. It is interesting to notice that peak and valley moments coincide acceptably between both series, confirming that a good value of ϕ was chosen.

Indeed, the monthly and annual daily profiles (Figure 15) reveal a similar pattern between synthetic and natural time-series. However, it can be noticed that historic time-series have a smoother behavior and more dispersed values during the whole day, with a notable exception near hour 19. In contrast, dispersion is minor and almost the same along the year in the synthetic time-series.

On the other side, the proposed methodology produces quite realistic total monthly wind energy values as shown in Table V. Correspondent MAPE is 8.3%, while the nRMSD is 13.7%. The maximum discrepancy occurs in June and September with an underestimation of 17.8% and an overestimation of 6.0% in wind energy production, respectively.

Remember that even the actual production of wind farms is not the same each year. For the simulation, a value of 85% was used for the derating factor η_W .

Table V. Deviation of total monthly wind energy output (MWh) values for a point in India.

	Actual	Synthetic	Variation
January	10445.60	9298.15	-11.0%
February	11372.26	9729.97	-14.4%
March	11382.02	10400.56	-8.6%
April	13069.84	13220.91	1.2%
May	23178.49	22987.15	-0.8%
June	35351.94	29044.86	-17.8%
July	22610.76	20006.27	-11.5%
August	21399.94	22009.06	2.8%
September	17773.97	18839.08	6.0%
October	7906.84	7679.01	-2.9%
November	3798.43	4582.08	20.6%
December	6819.88	6919.43	1.5%
Total	185109.99	174716.53	-5.6%
Mean	15425.83	14559.71	

The actual energy calculation used wind speed measurements for 80 m height because of the hub height of the wind turbine model. Likewise, ordered synthetic values were scaled to that height through power law.

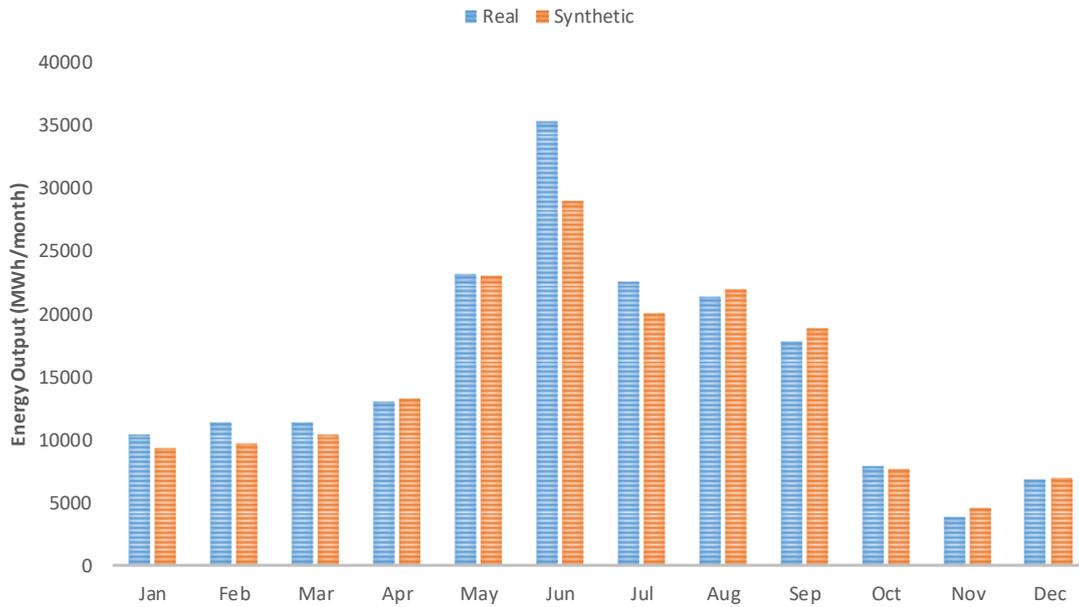


Figure 16. Total monthly wind energy output (MWh) for a point in India.

3.2.2.2 Test N° 2 – Comparison of wind energy for a point in Chile

A place in Chile with latitude -24.65° , longitude -70.24° , and an altitude of 2588 masl is analyzed in this second test. Measurement data is obtained from [142] and corresponds to yearly wind speed measurements at different heights and other parameters.

β_y and α_y are calculated using the time-series of wind speed at 50 m in the same way described in the previous test 3.2.2.1. Then, γ is cleared for each month using the monthly mean wind speeds at 10 and 40 m. Next, using the calculated value of γ , mean wind speed values at 50 m ($\bar{v}_{m,p}$) is calculated. To generate synthetic series, a value of $\phi = 11$ and $\delta = 0.2$ is used in this test.

Table VI. Deviations of monthly mean wind speed (m/s) values for a point in Chile.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Actual	5.553	6.578	6.705	6.420	7.104	8.996	9.777	7.459	5.895	8.997	6.163	6.538
Approx	5.588	6.550	6.642	6.382	7.037	9.006	9.727	7.372	5.736	8.896	5.986	6.383
Var	0.63%	-0.43%	-0.93%	-0.59%	-0.94%	0.11%	-0.52%	-1.17%	-2.71%	-1.12%	-2.87%	-2.38%

In Table VI, it could be seen deviations between actual monthly mean wind speed and approximate value estimated using power law for a hub height of 50 m.

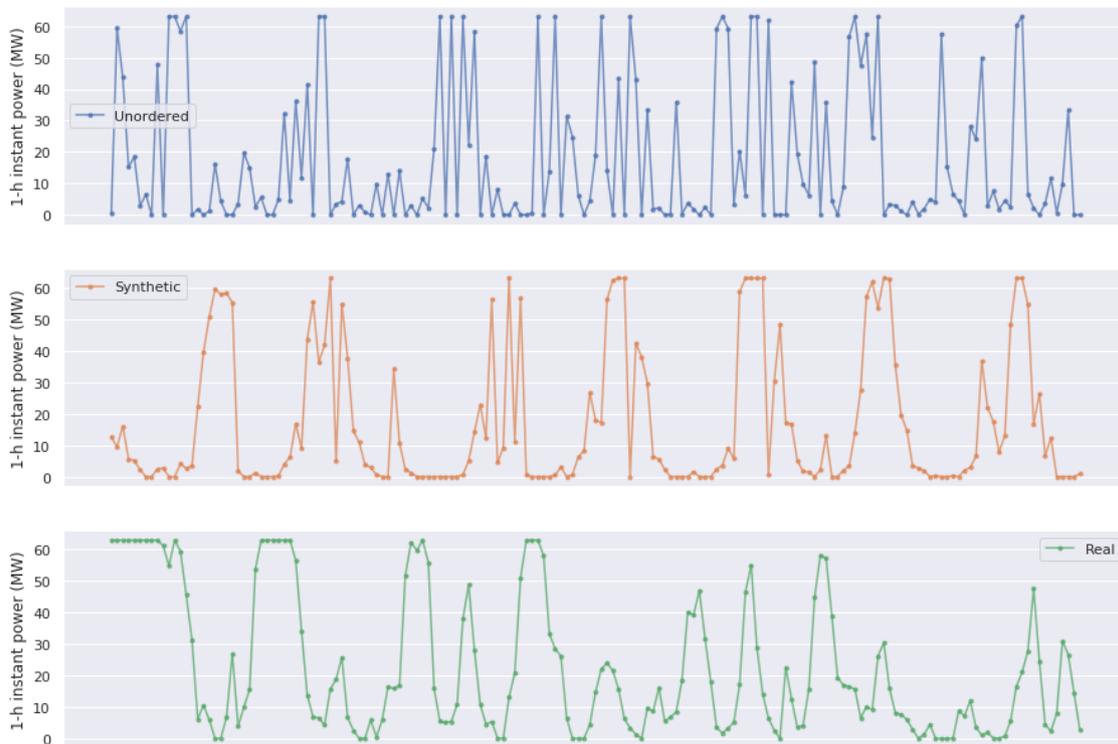
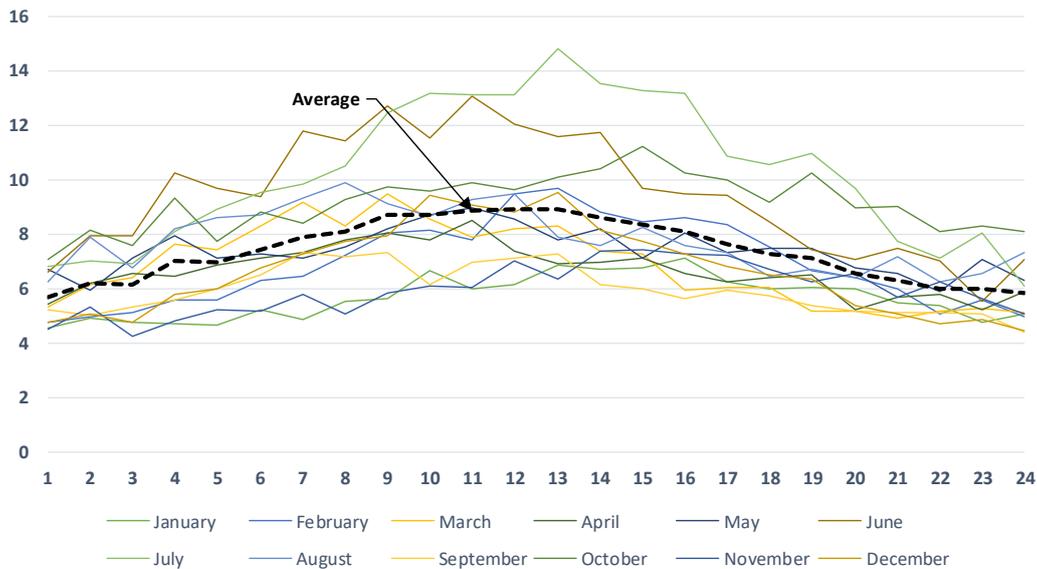


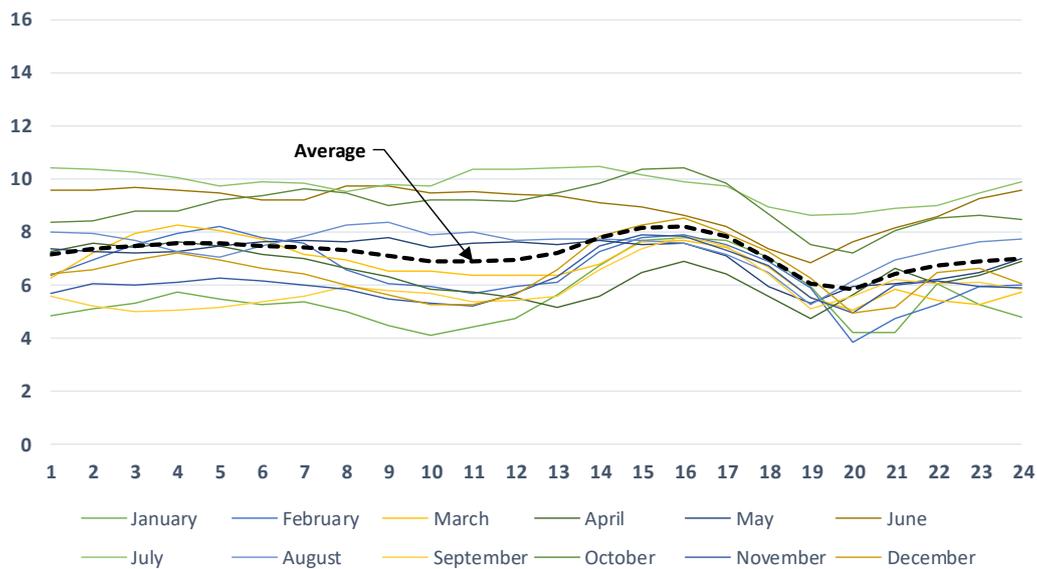
Figure 17. Time-series of total wind power output (MW) for a point in Chile.

Similar to the previous test 3.2.2.1, it is also evident that unordered time-series present a more erratic behavior than the actual one. This behavior is improved after applying the proposed reordering algorithm, as depicted in Figure 17.

This Chilean point's monthly and annual daily wind speed profile do not have a standard performance (Figure 15-b). Hardly this profile will fit a sinusoidal-like form, but a good result is achieved.



(a)



(b)

Figure 18. Monthly and annual daily (a) synthetic and (b) actual profile of wind speed for a point in Chile.

A MAPE of 4.2% and an nRMSD of 6.2% are obtained when comparing total monthly energy. Maximum discrepancies occur in January and July when deviations reach 5.9% and -11.4%, respectively. For this simulation, a value of 85% was used for the derating factor η_W .

Table VII. Deviation of total monthly wind energy output (MWh) values for a point in Chile.

	Actual	Synthetic	Variation
January	12547.62	13287.62	5.9%
February	16525.08	16823.45	1.8%
March	19579.51	18759.00	-4.2%
April	16451.77	17333.11	5.4%
May	21838.07	21177.52	-3.0%
June	26015.00	25655.07	-1.4%
July	32140.78	28477.46	-11.4%
August	23289.61	22459.02	-3.6%
September	14595.80	14283.17	-2.1%
October	26930.13	27670.87	2.8%
November	15543.86	14243.40	-8.4%
December	18047.32	18158.87	0.6%
Total	243504.55	238328.55	-2.1%
Mean	20292.05	19860.71	

Both actual and synthetic wind speed values had to be scaled to wind turbine model hub height of 80 m since they were initially referenced for a height of 50 m.

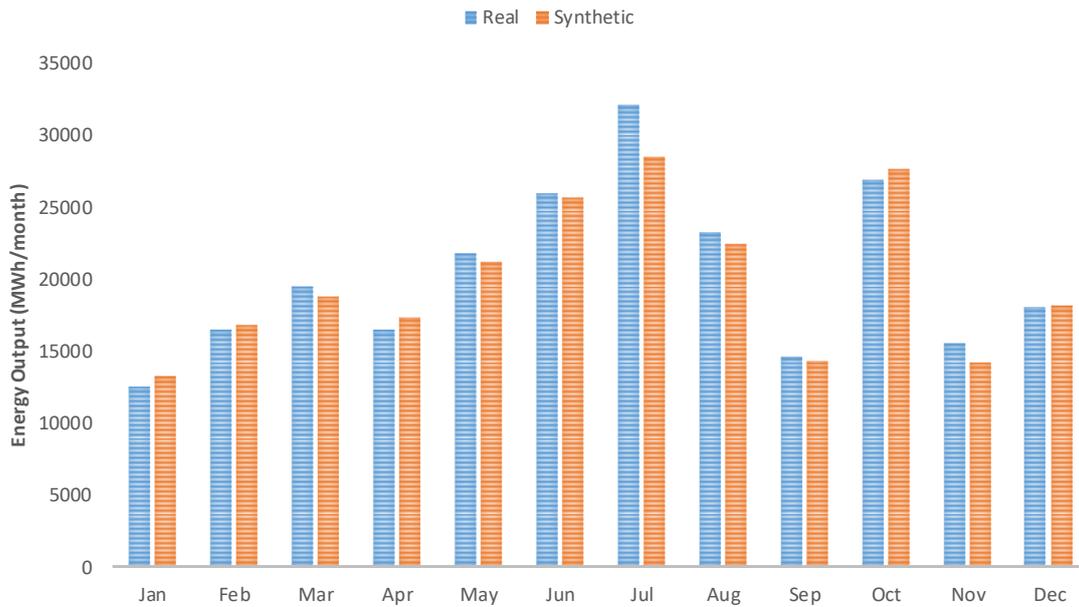


Figure 19. Total monthly wind energy output (MWh) for a point in Chile.

3.2.2.3 Test N° 3 – Comparison of wind energy with a real wind farm in Peru

The third test simulates an existing wind farm in Peru and compares the synthesis results and its historical power output time-series.

The plant studied is Tres Hermanas Wind Farm, located at latitude -16.39° and longitude -75.08° and has an altitude of 217 masl, so data is gathered from POWER and [143].

This plant uses two models of wind turbines: 8 turbines Siemens SWT-2.3-108 and 25 turbines SWT-3.15-142 [139]. However, simulation employs just one type, as mentioned in 3.2.2. Forty-two turbines are considered for the simulation, giving a total installed power of 96.6 MW, being 97.15 MW in reality.

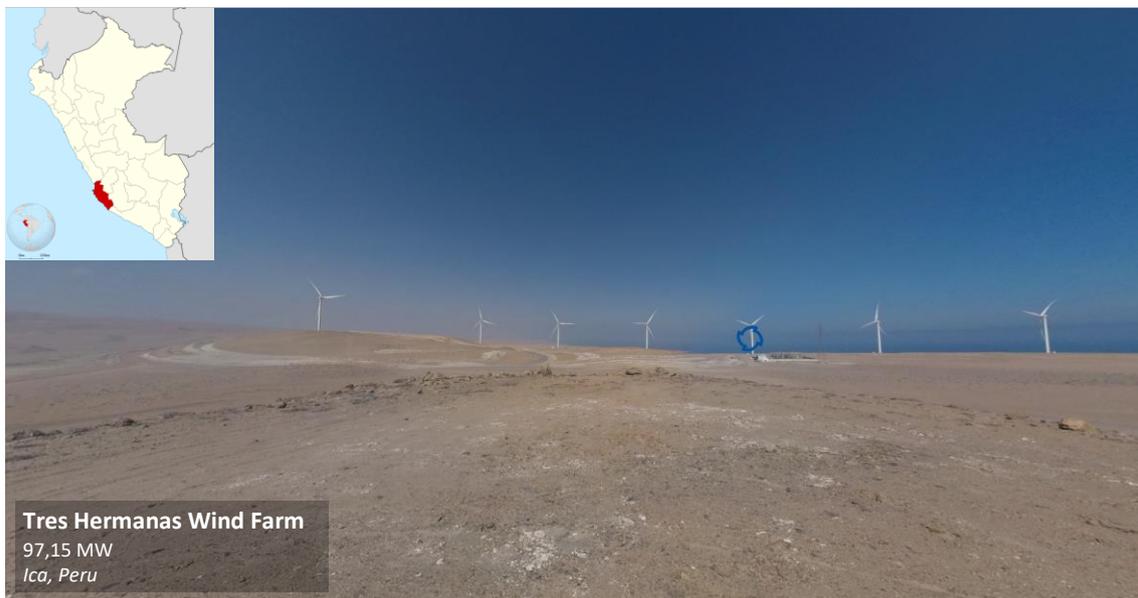


Figure 20. Location and site image of Tres Hermanas Wind Farm in Peru.

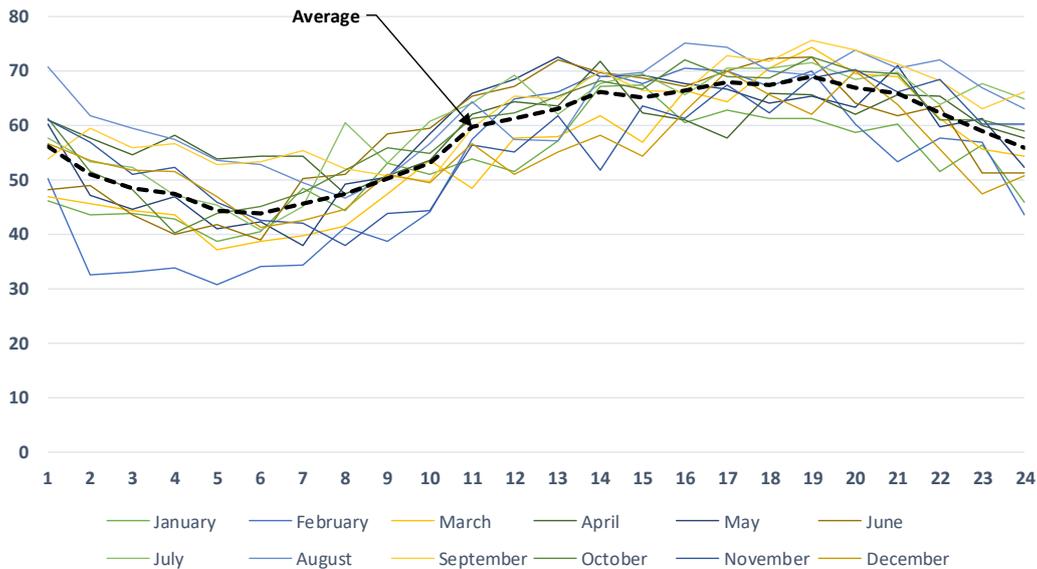
From [143] it is retrieved Weibull parameters α_y and β_y at 100 m height and from NASA POWER, monthly mean wind speed values at 10 and 50 m. Having these values it is possible to cleared power law exponent α and then calculate monthly mean wind speed values at 100 m height.

A value of $\phi = 17$ and $\delta = 0.2$ is used in this test to generate synthetic series. Also, a value of 85% was used for the derating factor η_W .

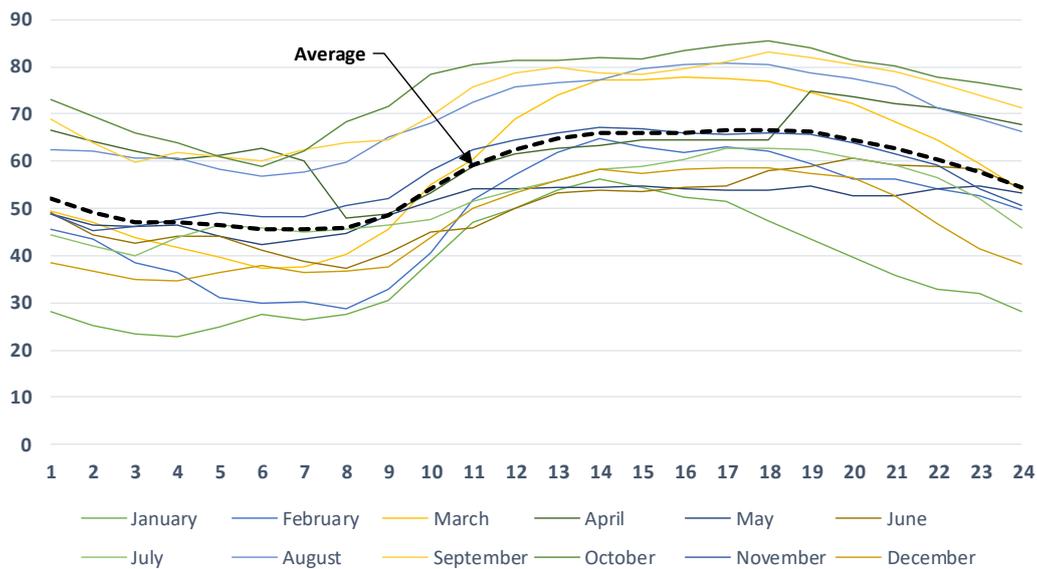
In this case, there is no availability of a time-series of actual wind speed but a series of power plant output. Hence, Figure 21 shows the monthly and annual daily profile of power generation for both synthetic and actual time-series. It is evidenced that synthetic

time-series have a noisier behavior when comparing month profiles, although the annual mean is quite similar.

A common finding in both tests is the noisier behavior of monthly daily profiles since a random generation process is employed. It is impossible to affine the monthly daily profiles because available data allows modeling only some parameters at a precision scale of months when it would be necessary to model hourly parameters.



(a)



(b)

Figure 21. Monthly and annual daily (a) synthetic and (b) real profile of wind power (MW) for a point in Peru.

This fact is not entirely a bad aspect of the proposed methodology whenever it produces a more exigent-to-integrate wind farm time-series, as indicated in a previous test. Indeed, developing planning studies considering this kind of synthetic time-series could guarantee that when actual implementation is done, the power system will not get into a more stressed situation than it was considered when simulated, which in essence represents the worst case.

In terms of energy, the synthesis methodology yields an annual amount greater in 5.6% than the average of measured values with an nRMSD of 9.4%, as shown in Table VIII. Month-to-month variations have a mean error of 7.7% (MAPE).

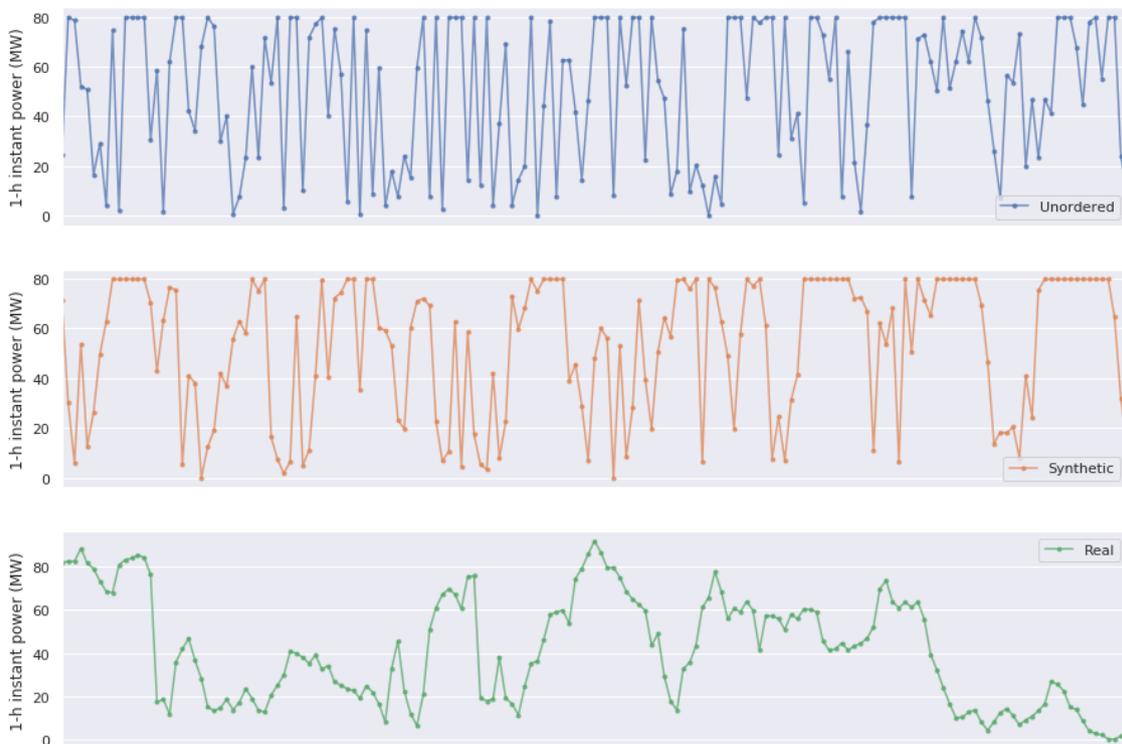


Figure 22. Time-series of total wind power output (MW) for a point in Peru.

However, the nature of wind resources makes energy production change year to year. As shown in Figure 23, total monthly energy production is very different every year, with some exceptions in July and November, where production is quite the same.

If synthetic results are compared to each year independently, it can be found nRMSD values of 12.5%, 15.8%, and 14.9% for years 2016, 2017, and 2018, respectively. However, when averaged, this value is reduced until 9.4%, as indicated previously.

Table VIII. Deviation of total monthly wind energy output (MWh) values for a point in Peru.

	Average	Synthetic	Variation
January	32302.01	39340.20	21.8%
February	29939.20	34978.23	16.8%
March	41116.80	40548.28	-1.4%
April	41852.42	42975.77	2.7%
May	42732.15	43293.23	1.3%
June	36318.82	41967.99	15.6%
July	38270.57	45320.03	18.4%
August	45750.20	46922.11	2.6%
September	44846.52	44895.43	0.1%
October	46864.77	44286.17	-5.5%
November	39551.08	40588.14	2.6%
December	39490.29	40716.75	3.1%
Total	479034.82	505832.32	5.6%
Mean	39919.57	42152.69	

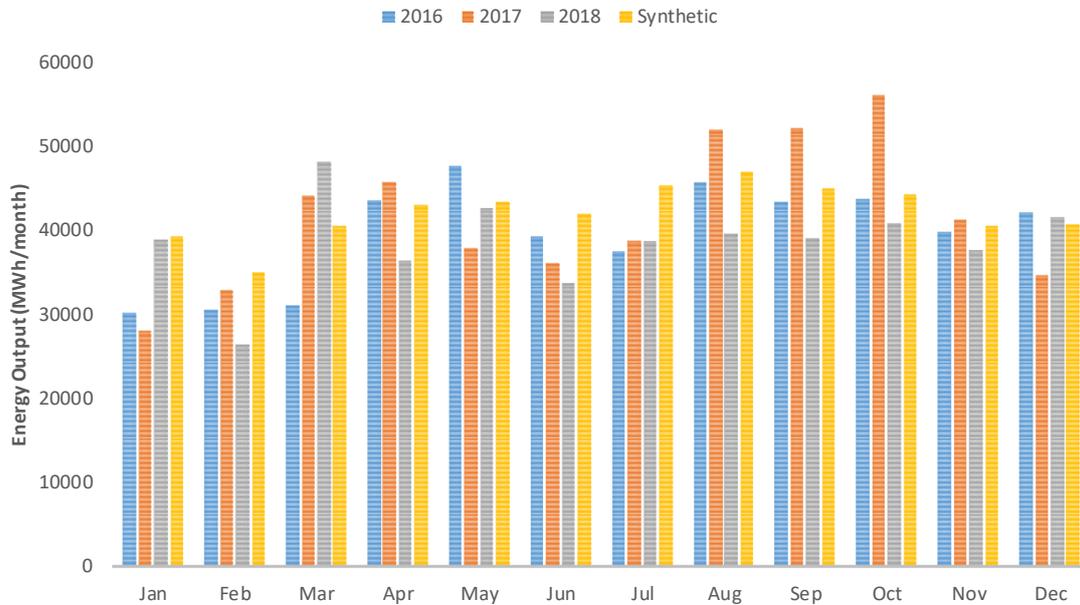


Figure 23. Total monthly wind energy output (MWh) for a point in Peru.

According to the results, it could be said that the proposed methodology tends to overestimate the total wind energy production. This aspect should be taken into account when using these synthetic time-series to perform medium and long-term planning. It would be recommendable to generate multiple time-series varying parameter values to internalize methodology uncertainties into the planning process.

3.3 Chapter conclusions

The proposed methodology presents a simple but effective approach for producing synthetic hourly production values for solar and wind renewable plants.

According to the results presented, the methodology for solar resources successfully synthesizes time-series of output power, which correctly replicates typical behavior. On the other hand, the calculus of produced solar energy is slightly underestimated with the proposed methodology, having a discrepancy of 2.4% in the yearly totals.

For wind resources, the randomness of the proposed methodology was encountered as a problem since it produces time-series in which variability is not realistic. This unrealistic time-series could be used as a worst-case scenario because such variability makes it hard to integrate its corresponding wind plant into a non-flexible power system.

This behavior is corrected by solving the assignment optimization problem, which gives a more realistic profile for the output power of wind farms. On the other hand, the produced energy calculation over or under-estimates actual production within a variation range of -5.6% to 5.6%.

Comparison with existing solar and wind plants could help estimate the expected error for future uses of synthetic time-series.

The presented methodology is novel because it uses aggregate parameters as input and does not require historical time-series, suitable for developing countries lacking renewable information. Besides, this flexible and parametric methodology can generate multiple time-series scenarios modifying aggregate input parameters to achieve enough range of cases to incorporate uncertainties that may be used in future research work.

4 Time-series clustering to reduce data dimension

This chapter evaluates the most used clustering technique presented in the literature and proposes applying a quality index to identify the best parameters to use within the clustering process.

Additionally, parameter tuning is suggested for future applications based on the results obtained after applying the technique to an existing data set.

4.1 Methodology

As suggested by [80], [81], [85], [86], it is essential that the reduced study scenarios maintain the correlation, simultaneity, randomness, and variability within all the time-series presented in the whole power system.

Some approaches have been presented in previous works analyzed in Section 2.2. However, none of them have applied their methodology into an existing power system, where each load and renewable plant have a particular time-series, in contrast to academic problems where it is generally considered that all bus bars have the same load or resource profile.

Existing literature on clustering for energy planning suggests that a pre-aggregating step must be done to combine all available data into a unique value for each time-step when looking for representative hours or to get a unique time-series when looking for representative days to start the process. However, presented cases only consider historical information to perform the pre-grouping step. An average between load and renewable generation is proposed in [81], while the net load (load minus renewable generation) is recommended in [80].

In this chapter, the aim is not only to cluster time-series of the existing infrastructure of loads and renewable plants but also to allow the evaluation of new time-series of

candidate plants that must be evaluated into the planning process through an optimization model. Therefore, the proposed method for pre-aggregating data is an averaging of normalized time-series in order to:

- Accomplish to maintain the correlation, simultaneity, randomness, and variability between all the time-series without the influence of big loads or plants.
- Allow the analysis of other elements that also depends on time-series to pre-set their operation like a run-of-the-river hydroelectric plant.

Concerning the technique used to perform the clustering process, most works have chosen the K-means method [30], [81], [87], [144] over the hierarchical method [80], [85]. In addition to choosing the clustering method, another critical selection is the metric used to calculate the distance between elements and clusters centroids. In that sense, this chapter explores and evaluates different metrics for the K-means clustering technique is developed.

The proposed methodology is presented below.

Having a set of raw time-series, where each series can represent a different element such as a generator or a load, the proposed procedure is described below:

1. Normalize each time-series in order to have all values between 0 and 1.
2. Add up all time-series and then normalize the new aggregated time-series. At this point, we obtain a vector of dimension Y , where $Y=H*N$, H is the number of time steps per day, and N is the number of days of the time window analyzed (typically 365).
3. Rearrange the vector into days to have a matrix of $H \times N$.
4. Apply the Time-series K-Mean function, which is provided by the Python library `TSLearn` [145]. This function will group days into clusters.
5. Identify the most representative day for each cluster.
6. Expand each representative day into its original raw time-series (generators, loads, etcetera).

The following code is executed using the Python 3 environment provided for free by Google Colab to perform step 4.

```
pip install tslearn  
  
from tslearn.clustering import Time-seriesKMeans
```

```
km = Time-seriesKMeans(n_clusters=C, metric=M, n_init=I,  
random_state=0)  
  
df_cc = km.fit_predict(df_X)
```

In the variable `df_cc` we obtain the cluster label to which each time-series (day) provided in `df_X` belongs. Dataframe `df_X` corresponds to the result of steps 1 to 3. Values of `C` and `I` allow choosing, respectively, the specific number of clusters we are looking to have and the number of times the model will run in order to find the best fit. The metric used to calculate the K-means algorithm is set to Euclidean by default, but it can be changed to DTW (Dynamic Time Warping) [146] or Soft-DTW [147], specifying it in the parameter 'metric' (`M`).

Once clusters are defined, each cluster's most representative time-series (day) must be identified, as indicated in step 5. This task is performed by calculating the distance (the same as the chosen metric for step 4) between each member (day) and cluster centroid (provided in `km.cluster_centers_`). The chosen representative time-series for each cluster represent a day of the year. Consequently, having this day, it is possible to obtain the individual series (generators, loads, etcetera) that composed the aggregated chosen time-series.

4.2 Application

As shown in the methodology, the clustering process can be tuned by changing the number of clusters, the number of initial runs, and the selected metric. In that sense, in this section, data of an existing power system will be used to choose the best set of parameters.

4.2.1 Case study

The proposed methodology is applied in the same system called PRC60 that will be presented, detailed, and analyzed in section 5.2.1. This power system, located in the northern part of Peru, has no presence of renewable plants. Two existing plants near the country's areas with the best solar and wind resources have been chosen and added to complement the system's data.



Figure 24. Location and satellite image of Talara Wind Farm in Peru.

On the one hand, a 30 MW plant near the application power system is used for wind resources. This plant started its commercial operation in 2014. Its generator's hub height is 80 m and has an estimated annual production of 119,673.0 MWh.



Figure 25. Location and satellite image of Tacna Solar Park in Peru.

On the other hand, a 20 MW plant that started its commercial operation in 2012 is selected for solar resources. This plant has an estimated annual production of 47,196 MWh and was installed in the southern part of Peru.

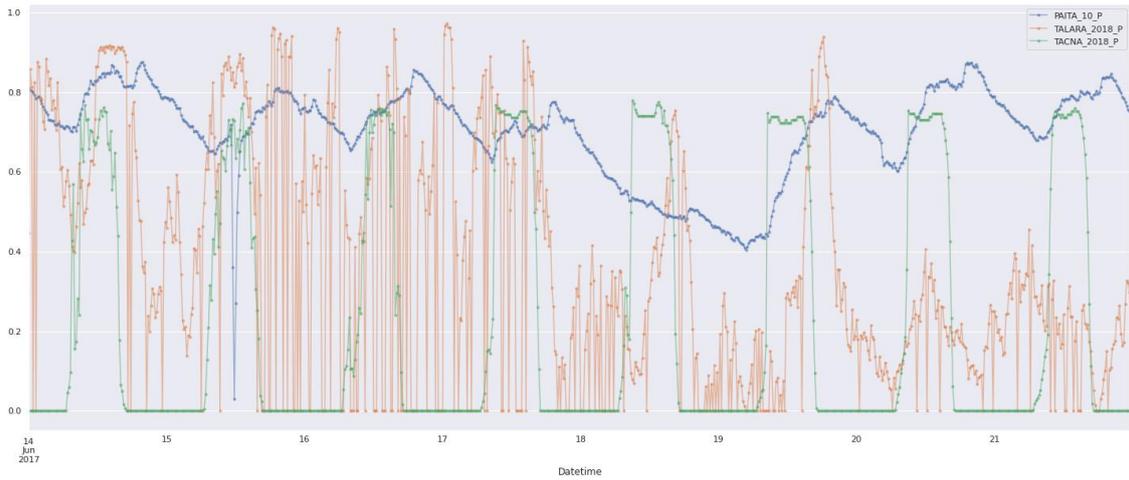


Figure 26. Power profiles of a load demand (blue), a wind farm (red) and a solar park (green) presented in PRC60 power system.

An annual generation time-series of each plant is appended to the 20 demand registers the system initially has. These demands correspond to loads in 10 kV, 14 kV, 22.9 kV, and 60 kV.

Each demand register is composed of two time-series: one for active power and one for reactive power. However, just active power time-series are considered for the clustering process, although both series could be included if the optimization model can use them as input.

As a result, the complete dataset comprises 22 time-series, all of them corresponding to actual measurements.

4.2.2 Parameters tuning

The goal of the proposed procedure is to reduce the number of series for analysis without losing the essential system's insights as a whole. As reviewed in section 2.2, there is no consensus on the number of representative days to use. In this case, we look to find seven representative days which can emulate each typical weekday, as explained later. For this reason, we set parameter $C = 7$.

Trying to perform an expansion planning optimization without this clustering process will have to consider the $365 \text{ days} \times 96 \text{ steps/day} = 35,040$ points a typical yearly time-series has, which is an intractable number of points for an existing system expansion planning.

It must be noticed that the procedure converts each time-series into per-unit values before aggregating, making the actual magnitude of each series irrelevant and prioritizing simultaneity and correlation among renewable resources and load time-series.

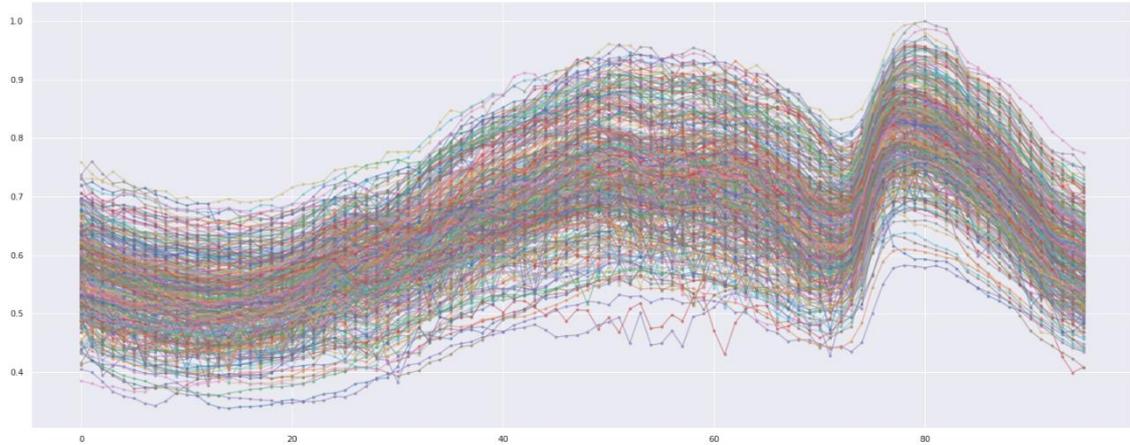


Figure 27. Annual time-series of the PRC60 system rearranged as daily time-series.

Although [80] recommends using DTW [146] as the metric for the clusterization process, the three different K-Means Clustering metrics implemented by the library TSLearn [145] are evaluated in order to find the optimal clustering metric. Comparison is made using the Silhouette coefficient.

As indicated in [148], the Silhouette coefficient is an internal measure widely used for cluster validation since it expresses the cluster separation and cohesion quality. Well-separated clusters with lower within-cluster variations must tend to have Silhouette coefficient (S) values near to 1.

Table IX. Silhouette coefficient and processing time for different metrics used in a K-Means Clusterization

	Euclidian	DTW	Soft-DTW
10 initial runs	S = 0.236515 (53.768 s)	S = 0.254996 (382.588 s)	S = 0.341826 (1322.590 s)
30 initial runs	S = 0.233946 (1403.559 s)	S = 0.275713 (2360.074 s)	S = 0.398355 (5108.184 s)

According to the results presented in Table IX, the best metric for clustering the presented power system time-series is Soft-DTW [147] since its Silhouette coefficients are greater than those achieved by other metrics, although its processing time is by far more considerable. Furthermore, it was found that increasing the number of initial runs (run with different centroid seeds) improved the S coefficient. In that sense, more tests are done only with this metric to find the best number of initial runs.

Table X. Silhouette coefficient and processing time for Soft-DTW metric used in a K-Means Clusterization

	Soft-DTW
10 initial runs	S = 0.341826 (1322.590 s)
20 initial runs	S = 0.341826 (17824.393 s)
30 initial runs	S = 0.398355 (5108.184 s)
50 initial runs	S = 0.398355 (9630.030 s)
70 initial runs	S = 0.398355 (15924.537 s)

S coefficient showed in Table X demonstrate that the best number of initial runs is 30. Before this quantity, the Silhouette coefficient is lower, and after that value, the coefficient does not improve.

In conclusion, subsequent analysis will consider the clustering process using the Soft-DTW metric and 30 initial runs. Nevertheless, it must be noticed that this parameters tuning process can be replicated for each new set of data If extra precision is required.

4.2.3 Results

The expected result of the application of the proposed methodology is composed of two things:

- The actual centroid of each cluster, i.e., representative days
- Weight or number of time-series of each cluster

The first element helps us define our base values or scenarios for simulation, while the second tells us how much these simulation results must be considered in the aggregate evaluation.

Figure 28 shows the seven clusters created and all of its time-series. Additionally, it shows the calculated centroid plotted in red and above their representative weight (ρ_{day}^t) which means the fraction of the total number of daily time-series that belongs to each cluster, or in this case, the number of days of a year that belongs to each cluster.

In the previous sub-section 4.2.2, it was mentioned that the clustering technique would look to have 7 clusters representing each weekday. Obtained results capture this objective producing, for example, clusters 4 and 7 that capture Saturdays and Sundays, cluster 6 that groups holidays, clusters 1, 3, and 5 that represent the most productive days of a week and, finally, cluster 2 that summarizes the least productive day of a typical workweek.

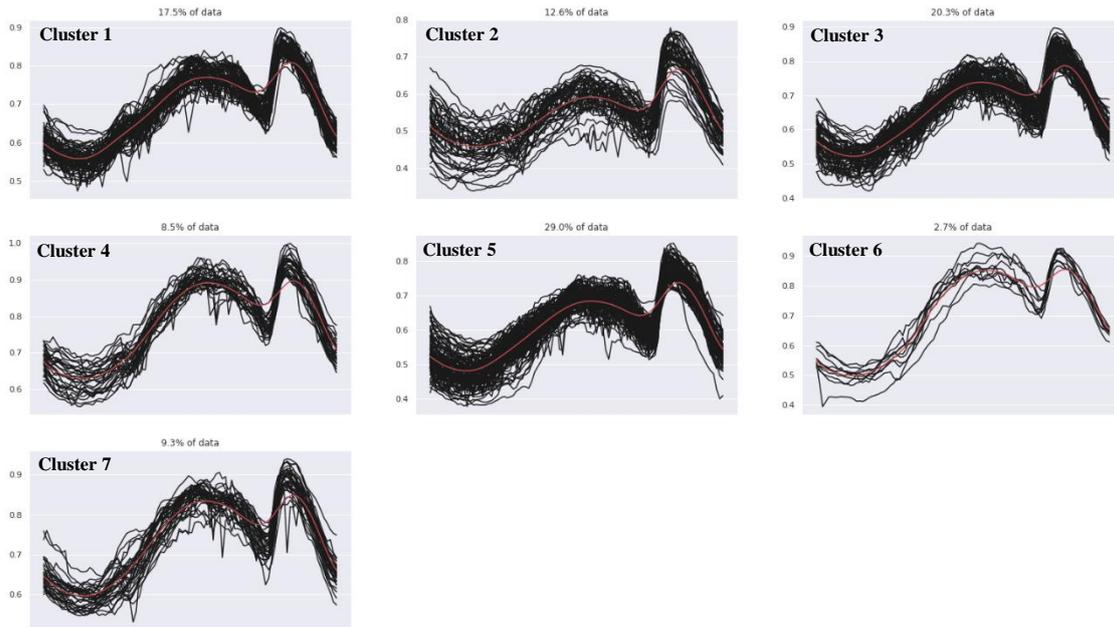


Figure 28. Clusters obtained for PRC60 system.

It is essential to mention that this centroid is fictitious and does not correspond to any actual time-series. In order to have an actual centroid, an iterative process have to be done to find the time-series with the closest distance to this fictitious centroid. This process can be performed using the cross-similarity matrix for the correspondent metric as follows:

```
import numpy as np

dist = cdist_soft_dtw(df_X.reshape(df_X.shape), df_cc.cluster_centers_
.reshape((C, df_X.shape[1])))
df_C = np.zeros(X)

for cluster_n in range(C):
    d_init = 100
    for index in range(df_X.shape[0]):
        if df_y[index] == cluster_n:
            if dist[index][cluster_n] < d_init:
                df_C[cluster_n] = index
                d_init = dist[index][cluster_n]
```

Once identified every actual centroid, the correspondent date will be known, making it possible to identify the raw time-series to use in a later simulation process.

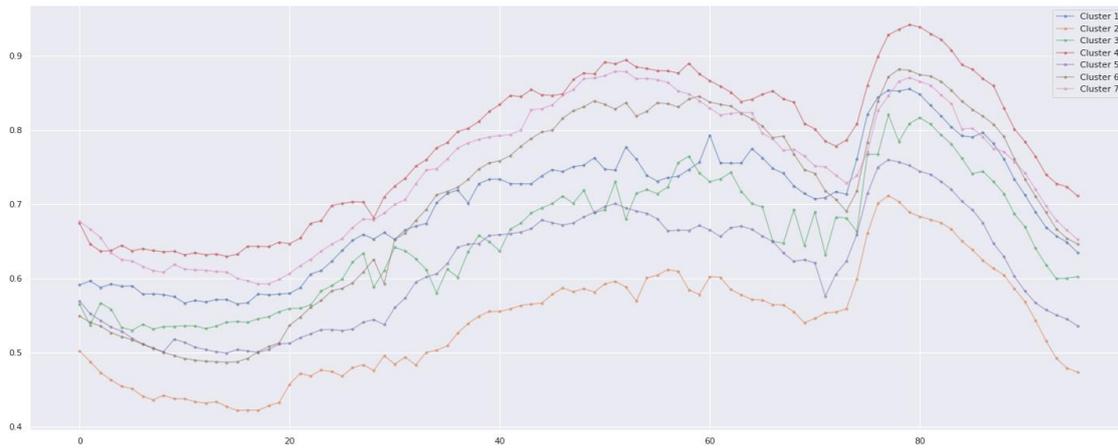


Figure 29. Real centroids of clusters obtained for PRC60 system.

Last Figure 29 plots the actual centroid of all created clusters. These centroids correspond to the aggregated time-series formed following the proposed methodology. Therefore, they do not represent a specific load demand or renewable plant itself.

Resulted clusters will be used in Chapter 5 to plan PRC60 system expansion.

4.3 Chapter conclusions

The proposed methodology reduced the 35,040 points of the existing yearly time-series into 7 days x 96 steps/day = 672 points maintaining the correlation, simultaneity, randomness, and variability within all the time-series.

In order to define the best parameters to use in a time-series clustering process for energy planning studies, the Silhouette coefficient was used to measure the quality of clustering results. This coefficient can help, also, to evaluate new clustering techniques or metrics.

5 Expansion planning using a practical AC model

In this chapter, a practical optimization model that considers an AC optimal power flow is presented to address the gap found in existing literature, which generally only performs a DC power flow.

The presented model includes equations representing several elements like conventional generation plants, renewable plants like solar PV, wind, and hydro whose operation is pre-set by time-series, battery energy storage systems, transmission lines, and capacitors for reactive compensation. All equations accomplish to be written considering active and reactive power flow, losses assessment, and voltage fluctuation.

The proposed model performs an intra-year study in a deterministic way, avoiding the analysis of year-to-year variation of renewable resources. Finally, the implementation of the model is done using the AMPL language [149] and the Gurobi solver [150].

5.1 Methodology

The proposed model is based on the ones presented in [101], [151]. Over this model, many additions and modifications have been made to represent conventional and renewable generators, energy storage systems, reactive compensation equipment, and AC-related parameters.

5.1.1 Conventional Generation Plants

In order to take into account the reactive power flows that occur on existing systems, conventional power generation plants are represented using equations (5.1) to (5.4). The first equation (5.1) limits the maximum and minimum value of the active power (ga_i^t) of each generator. Then equations (5.2) and (5.3) restrict its capacitive and inductive reactive power (gr_i^t), respectively, using the current value of the active power and a fixed power

factor for each case. Finally, active and reactive power are related by equation (5.4) which guarantee apparent power would not exceed its maximum generation capacity (g_i^{max}).

$$g_i^{min} \leq ga_i^t \leq g_i^{max} \quad , \forall i \in G, t \in T \quad (5.1)$$

$$gr_i^t \leq ga_i^t \tan(\text{acos } 0.95) \quad , \forall i \in G, t \in T \quad (5.2)$$

$$gr_i^t \geq -ga_i^t \tan(\text{acos } 0.99) \quad , \forall i \in G, t \in T \quad (5.3)$$

$$(ga_i^t)^2 + (gr_i^t)^2 \leq (g_i^{max})^2 \quad , \forall i \in G, t \in T \quad (5.4)$$

Capacitive and inductive reactive power are limited differently according to Peruvian technical criteria [152]. Consequently, it must be modified depending on the application case.

5.1.2 Renewable Generation Plants

In contrast with conventional plants where dispatched power is decided during the optimization process; in the case of renewable plants, available renewable power (ζ_p^t) is an independent (predefined) value for each time step.

$$za_p^t = \zeta_p^t \cos(\text{acos } 0.95) \quad , \forall p \in Z, t \in T \quad (5.5)$$

$$zr_p^t \leq za_p^t \tan(\text{acos } 0.95) \quad , \forall p \in Z, t \in T \quad (5.6)$$

$$zr_p^t \geq -za_p^t \tan(\text{acos } 0.99) \quad , \forall p \in Z, t \in T \quad (5.7)$$

Given that all renewable plants must participate by supplying reactive power (zr_i^t) to the system, equations (5.5) to (5.7) conduct a simple calculation to determine the value that each power component must have considering recommended power factors [152], similar to the approach presented earlier in 5.1.1.

$$\zeta_p^t = ts_p^t u_p \omega_p \quad , \forall p \in Z, t \in T \quad (5.8)$$

Available power for each time t is determined by multiplying the number of units installed ($\omega_p \geq 0$: integer) within a plant, the power capacity of each unit (u_p) and the per-unit power value (ts_p^t) for time t obtained from clustered time-series (see Chapter 4). This way of modeling non-dispatchable plants can also be extended to cover the case of hydroelectric plants.

An approach of considering modular expansion capacity values of supply-side alternatives (renewable generators, energy storage systems, and reactive compensators)

is implemented in the proposed model. As explained in [153], this approach avoids the need to discretize the continuous capacity values obtained in a post-processing phase, losing the optimality properties.

5.1.3 Energy Storage Systems

Equations proposed in [154] are used as a basis to model the operation of battery energy storage units, which consider a fixed capacity of units. However, as the number of units to install must be determined within the optimization problem, equations were reformulated to maintain their linearity.

As a convention, charge power ($epc_b^t \leq 0$) is considered negative, whereas discharge power ($epd_b^t, epda_b^t$ and $epdr_b^t \geq 0$) is considered positive, denoting that device behaves as a load or a generator, respectively. Both variables are limited by the product of battery nominal power (ep_b^{nom}) and an integer variable ($\omega_b \geq 0$) that represent the number of units installed (if any), as seen in equations (5.9) and (5.10).

$$epc_b^t \geq -ep_b^{nom} \omega_b \quad , \forall b \in B, t \in T \quad (5.9)$$

$$epd_b^t \leq ep_b^{nom} \omega_b \quad , \forall b \in B, t \in T \quad (5.10)$$

$$epda_b^t = epd_b^t \cos(\text{acos } 0.95) \quad , \forall b \in B, t \in T \quad (5.11)$$

$$epdr_b^t \leq epda_b^t \tan(\text{acos } 0.95) \quad , \forall b \in B, t \in T \quad (5.12)$$

$$epdr_b^t \geq -epda_b^t \tan(\text{acos } 0.99) \quad , \forall b \in B, t \in T \quad (5.13)$$

$$epa_b^t = epda_b^t + epc_b^t \quad , \forall b \in B, t \in T \quad (5.14)$$

$$epr_b^t = epdr_b^t \quad , \forall b \in B, t \in T \quad (5.15)$$

$$ec_b^t = ec_b^{t-1} - \left(\frac{epd_b^t}{\eta_b} + epc_b^t \right) \Gamma \quad , \forall b \in B, t \in T / t \geq 2 \quad (5.16)$$

$$ec_b^t \leq ec_b^{nom} \omega_b \quad , \forall b \in B, t \in T \quad (5.17)$$

$$ec_b^t \geq (1 - D_b) ec_b^{nom} \omega_b \quad , \forall b \in B, t \in T \quad (5.18)$$

$$\sum_{s=t-\tau_b+1}^t epd_b^s \leq ep_b^{nom} \omega_b \tau_b \quad , \forall b \in B, t \in T / t \geq \tau_b \quad (5.19)$$

$$epd_b^1 = 0 \quad , \forall b \in B \quad (5.20)$$

$$ec_b^1 = 0 \quad , \forall b \in B \quad (5.21)$$

Similarly to conventional and renewable generation, battery discharge power (epd_b^t) is divided into active ($epda_b^t$) and reactive ($epdr_b^t$) components as written in equations from (5.11) to (5.13). It is understood that electronic devices such as inverters would do the job of splitting discharge power. Net active and reactive power outcomes are calculated in equations (5.14) and (5.15), respectively.

Equation (5.16) estimates actual energy accumulated on the battery (ec_b^t). Having as a base the amount of energy of the previous time step (ec_b^{t-1}), it is added the energy charged (through epc_b^t) and reduced the energy discharged (through epd_b^t) which is affected by the efficiency of this process (η_b). The energy added or reduced is inferred from its respective power multiplying by Γ , which in the application case of this chapter will be $\frac{1}{4}$ that indicated a time step of 15 min (a quarter of an hour). Maximum energy stored on the battery is limited by equation (5.17), while minimum charge level is defined in equation (5.18) using the parameter depth of discharge (D_b).

Batteries can not discharge at maximum power for long periods. For that reason, equation (5.19) limits, linearly, the number of continuous cycles (τ_b expressed in number of time steps) the battery can discharge under this condition.

The last two equations set the initial conditions of this device.

It is essential to mention that, unlike the case of conventional generators, it is not implementing a quadratic equation like (5.4) for renewable plants nor energy storage systems in order to avoid the increment of computational complexity since, in both cases, the value of maximum capacity installed depends on an integer variable which would create a Quadratic Constraints Problem (QCP). This fact does not affect the quality of the solutions produced by the model, as demonstrated in section 5.2.

5.1.4 Transmission Lines

Power flows are modeled using transmission lines' susceptance (bl_l) and conductance (gl_l). Additionally, for each busbar, a voltage phase angle (θ_i^t) and deviation voltage (δV_i^t) is considered. Deviation voltage is limited to be between $\pm 0.05 pu$. Subindex ij denotes the difference between the value of bar i and bar j , in that order. Proposed model is based on the ones presented in [99], [101], [155].

$$|fa_l^t - (gl_l \Delta \delta V_{ij}^t - bl_l \Delta \theta_{ij}^t)| \leq (1 - \omega_l)M \quad , \forall l \in L, t \in T \quad (5.22)$$

$$|fr_l^t + (bl_l \Delta \delta V_{ij}^t + gl_l \Delta \theta_{ij}^t)| \leq (1 - \omega_l)M \quad , \forall l \in L, t \in T \quad (5.23)$$

$$|fa_l^t| \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.24)$$

$$|fr_l^t| \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.25)$$

$$(fa_l^t)^2 + (fr_l^t)^2 \leq (f_l^{max})^2 \omega_l \quad , \forall l \in L, t \in T \quad (5.26)$$

$$fa_l^t + \frac{1}{2} qa_l^t \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.27)$$

$$-fa_l^t + \frac{1}{2} qa_l^t \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.28)$$

$$fr_l^t + \frac{1}{2} qr_l^t \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.29)$$

$$-fr_l^t + \frac{1}{2} qr_l^t \leq f_l^{max} \omega_l \quad , \forall l \in L, t \in T \quad (5.30)$$

$$\left(fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(fr_l^t + \frac{1}{2} qr_l^t \right)^2 \leq (f_l^{max})^2 \omega_l \quad , \forall l \in L, t \in T \quad (5.31)$$

$$\left(fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(-fr_l^t + \frac{1}{2} qr_l^t \right)^2 \leq (f_l^{max})^2 \omega_l \quad , \forall l \in L, t \in T \quad (5.32)$$

$$\left(-fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(fr_l^t + \frac{1}{2} qr_l^t \right)^2 \leq (f_l^{max})^2 \omega_l \quad , \forall l \in L, t \in T \quad (5.33)$$

$$\left(-fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(-fr_l^t + \frac{1}{2} qr_l^t \right)^2 \leq (f_l^{max})^2 \omega_l \quad , \forall l \in L, t \in T \quad (5.34)$$

Active and reactive power flow through line l at hour t are represented by fa_l^t and fr_l^t , respectively. The values of these variables are calculated in equations (5.22) and (5.23), while total power flow is bounded to be within the line's maximum capacity (f_l^{max}) by equations from (5.24) to (5.34) which aggregate flow components in distinct ways. Total power flow also includes active (qa_l^t) and reactive (qr_l^t) losses, which are calculated externally as explained in 5.1.7.

It is essential to mention that most commercial linear solvers such as CPLEX and Gurobi can handle the quadratic constraints considered within the formulation since equations are still convex.

Additionally, a binary decision variable ω_l is introduced to decide whether a candidate line should exist or not. Existing elements have a fixed value of $\omega_l = 1$.

Big M constant is set as suggested in [156], while voltage phase angle is restricted by $|\theta| \leq \pi$ for each bar.

5.1.5 Reactive compensators

The proposed model also considers a block of constraints to solve the Reactive Power Planning (RPP) problem, the exclusion of which would produce expansion plans with cost overruns since all reactive power needs must be attended by generators and transmission lines, as explained in [157].

Consequently, within the model, static compensators are considered whose production is injected to the corresponding bus, following equation (5.35):

$$cr_c = u_c \omega_c, \forall c \in C \quad (5.35)$$

Reactive power is determined by multiplying the number of units installed or activated ($\omega_c \geq 0$: integer) and the capacity of each unit (u_c).

For existing capacitor banks, ω_c represents the number of steps which is considered a fixed value during the whole analysis period. If the application case considers capacitor banks that can be controlled remotely in real-time, this integer variable must be indexed by time and include additional equations to set limits on up or down steps.

For new banks, ω_c means the number of units that should be installed in a busbar in order to accomplish objective function and power system constraints.

5.1.6 Busbars

Active and reactive power balance must be performed at each bar j every time t in order to attend active (da^t) and reactive (dr^t) power requirements. In this model dr^t is calculated from da^t using a fixed power factor. Resource variables have been added to help convergence capturing renewable production wastage ($zloss_k^t$), demand curtailment (ra_k^t) and capacitive (rrc_k^t) and inductive (rri_k^t) reactive compensation additional needs.

$$\sum ga_i^t + \sum za_p^t + \sum epa_b^t + \sum \left(fa_i^t - \frac{1}{2} qa_i^t \right)_{in} - \sum \left(fa_i^t + \frac{1}{2} qa_i^t \right)_{out} = \sum da^t - ra_k^t + zloss_k^t, \forall k \in N, t \in T \quad (5.36)$$

$$\sum gr_i^t + \sum zr_p^t + \sum epr_b^t + \sum \left(fr_i^t - \frac{1}{2} qr_i^t \right)_{in} - \sum \left(fr_i^t + \frac{1}{2} qr_i^t \right)_{out} + \sum cr_c = \sum dr^t + rri_k^t - rrc_k^t, \forall k \in N, t \in T \quad (5.37)$$

Active resource variables are always positive and limited by the following equations:

$$0 \leq ra_k^t \leq \sum da^t, \forall k \in N, t \in T \quad (5.38)$$

$$0 \leq zloss_k^t \leq \sum za_p^t, \forall k \in N, t \in T \quad (5.39)$$

Reactive resource variables are also positive but have no upper limits.

Additionally, default values for voltage phase angle (θ_{slack}^t) and voltage deviation (δV_{slack}^t) should be set for slack busbar. These values could be used to represent the criticality of the present situation. For example, if the current system has voltage problems, then voltage deviation could be set at -0.025 , which means the voltage at slack is $0.975 pu$.

5.1.7 Losses assessment

As demonstrated in [101], an iterative process to calculate losses is a correct way to perform these complex calculations without affecting the linearity and dimension of the optimization model.

In that sense, the algorithm presented in Figure 30 iteratively resolves the optimization model and calculates active and reactive losses of transmission lines while the deviation of current and previous objective function value is more than 0.5%.

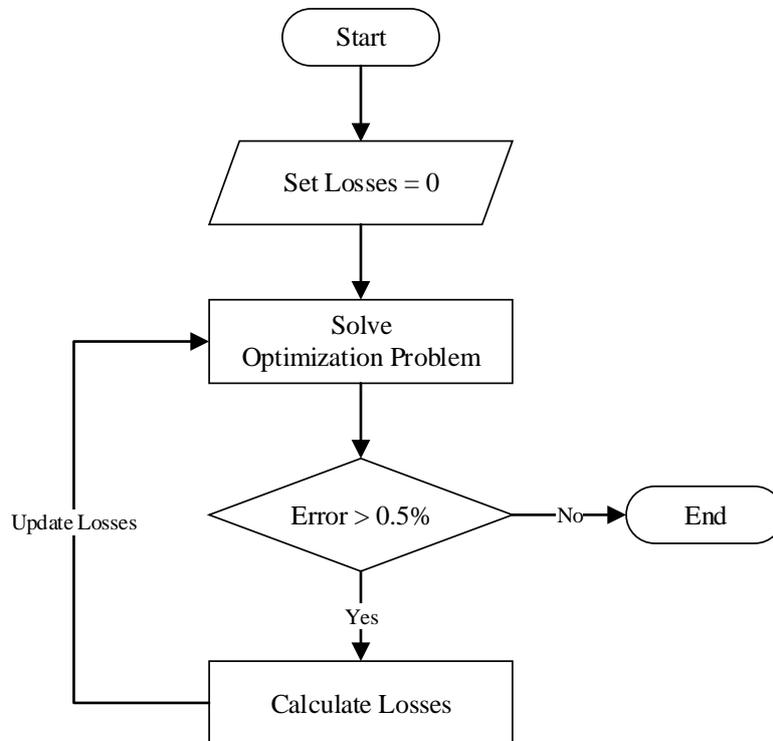


Figure 30. Iterative process to calculate active and reactive losses.

The equations needed to calculate losses are expressed in (5.40) and (5.41):

$$qa_i^t = \omega_l g_l \left((1 + \delta V_i^t)^2 + (1 + \delta V_j^t)^2 - 2(1 + \delta V_i^t)(1 + \delta V_j^t) \cos \Delta \theta_{ij}^t \right) \quad (5.40)$$

$$qr_i^t = -\omega_l b_l \left((1 + \delta V_i^t)^2 + (1 + \delta V_j^t)^2 - 2(1 + \delta V_i^t)(1 + \delta V_j^t) \cos \Delta \theta_{ij}^t \right) \quad (5.41)$$

These must be applied for each transmission line in each time step ($\forall l \in L, t \in T$).

5.1.8 Objective function

The model's objective is to minimize the cost of generation, losses, curtailment, wastage, and investment. It is understood that the investment cost for existing elements is zero, as well as for the new elements, the binary or integer variables will have non-zero values.

$$\begin{aligned} \min v = & 8760\Gamma \sum_{t \in T} \rho_{hour}^t \left(f_c \sum_{i \in G} C_g^i g a_i^t + C_{Mg} f_p \sum_{l \in L} (q a_l^t + q r_l^t) \right. \\ & \left. + C_{ENS} f_c \sum_{k \in B} (r a_k^t + r r i_k^t + r r c_k^t + z l o s s_k^t) \right) \\ & + \sum_{p \in Z} @I_p u_p \omega_p + \sum_{b \in B} @I_b e c_b^{nom} \omega_b + \sum_{c \in C} @I_c u_c \omega_c \\ & + \sum_{l \in L} @I_l \omega_l \end{aligned} \quad (5.42)$$

Since model data is a reduced version of complete registers, each scenario must be multiplied by a weight factor ($\rho_{hour}^t = \frac{1}{24} \rho_{day}^t$) which transmit the importance of each time step in the evaluation.

Power values are converted into energy values using the load factor (f_c) for the conventional generators and resource variables that indeed represent fictitious generators and loads; while the loss factor ($f_p = 0.7f_c^2 + 0.3f_c$ [158]) is used for active and reactive transmission line losses.

The energy generated by conventional plants are valued at C_g^i (in $\$/MWh$) which depends on the combustible used by each power plant. Losses of energy are valued at the market marginal cost C_{Mg} (in $\$/MWh$). Finally, the energy of fictitious loads and generators are valued using a curtailment cost value, which is considered $C_{ENS} = 6,000 \$/MWh$ for the Peruvian electrical system. Energy injected into the system by the renewable generators and batteries is valued at zero cost. The model will try to install the optimal capacity since

renewable energy injected above system requirements (i.e., renewable wastage $zloss_k^t$) is penalized in the objective function.

The investment cost is calculated as the product of the installed power of renewable plant, energy storage system or capacitor bank, multiplied by the annuity cost ($@I$), which is the sum of their annualized CAPEX cost, calculated considering an estimated lifetime (in years) and an annual discount rate (in %), and their OPEX cost. For the case of transmission lines, parameter $@I$ represents the total annualized cost of the element (CAPEX and OPEX), not a unitary cost, so it is not necessary to be multiplied by power or length.

All terms are expressed in the same monetary unit, which could be thousands or millions of US\$ in this case, in order to avoid big order numbers that can lead to numerical instabilities during the optimization process.

5.1.9 Notes on extending the proposed model

For the sake of reading, the complete model is presented in Appendix C.

If a particular application case needs to represent grid dynamics (e.g., rotations of power transformers) and grid reconfiguration (e.g., divisions or junctions of transmission lines), the equations (7) to (28) presented in [101] could be appended to this model.

Alternative energy storage systems like pumped storage hydropower plants can also be included in the proposed model without losing linearity, employing the formulation suggested in [159].

Also, the presented model, which performs an intra-year analysis only, can be re-written, indexing all relevant parameters and variables to a year ($\gamma \in Y$) if a multi-year analysis is needed.

Environmental constraints could also be included following an approach of weights that promote or penalize different technologies based on the convenience for each country or region. Such weights can be obtained from additional studies similar to [160].

However, additions of new equations and constraints will, inevitably, increase problem complexity making necessary the use of big-scale methods to resolve the optimization problems. Some of the most used methods are presented and compared in [161]. These methods should also be considered when applied to real-sized cases.

Since the main objective of this thesis is to present a novel planning framework to perform medium or long-term analysis considering short-term implications, implementation of such methods has been left out of the scope of this work.

5.2 Application

An existing system existing in the northern part of Peru is used as the case study. We will refer to this system as the Paita Ring Circuit 60 kV or PRC60.

5.2.1 Case study

The PRC60 system is located near the Pacific Ocean, surrounded by sand and water, as shown in Figure 31. This place has the Paita seaport, the second most important in Peru, after Callao in Lima.

Load behavior in this area is predominantly domestic, although some industries mainly linked to the extraction and processing of lobsters are also present.

As mentioned in Chapter 4, this system has available registers for only 20 loads, to which two additional time-series: one for a wind plant and one for a solar plant, have been appended.



Figure 31. Satellite view of PRC60 system.

The whole system is a ring circuit, and their central node (slack node) is the substation Piura Oeste 220/60 kV, which is interconnected to the national system (SEIN).

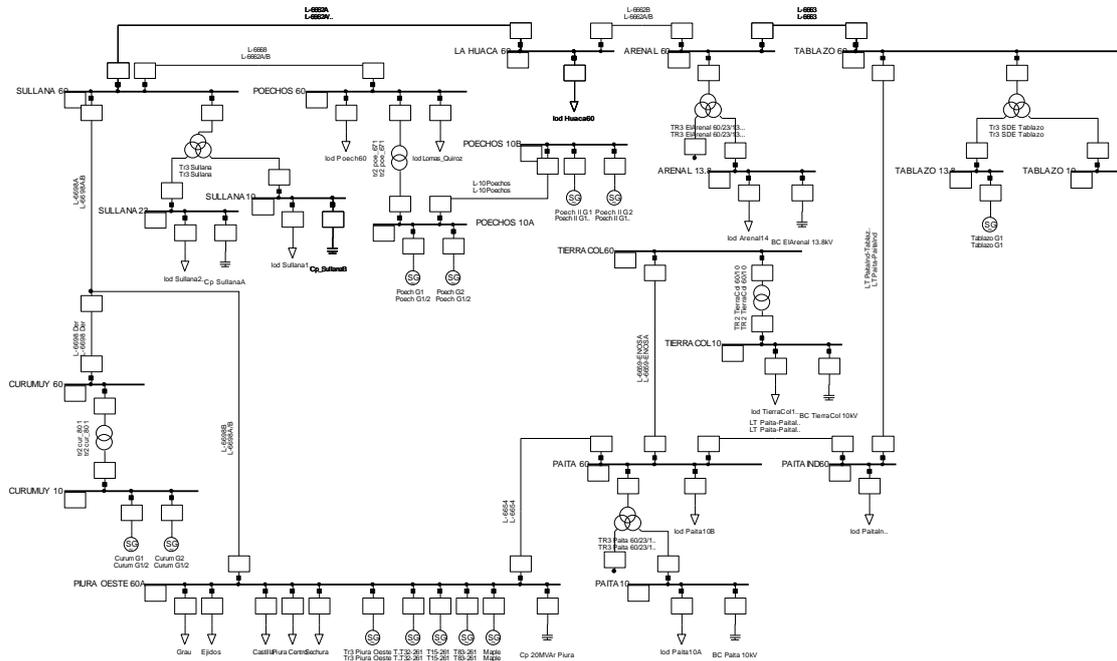


Figure 32. One-line diagram of PRC60 without projects in the year 2024.

All simulation and optimization processes consider that slack node voltage is $0.975 pu$ since the main problem in this system is referred to low voltage profiles that start upstream.

Validation of the results of the optimization model is performed using the specialized electrical software DigSilent PowerFactory. All PowerFactory simulation models use the same parameters as those used in the optimization model. Moreover, the active power of all generators (conventional and renewable) and energy storage systems are configured as indicated by the optimization model results, leaving the software calculates the reactive power flow, considering the same limits used in the proposed model. The number of active units of capacitor banks is also set as specified by the optimization process.

On the other hand, the losses of all components and the power flow values of transmission lines and transformers are calculated by PowerFactory simulation software.

5.2.2 Solution scenarios

Six different scenarios are studied to validate the correct operation of the proposed optimization model. The business-as-usual situation for the year 2024 is C01. Following

scenarios from C02 to C05, implement one or two technologies that would solve existing voltage problems on some bus bars of PRC60. Finally, scenario C06 makes all the technologies supported by the optimization model compete.

5.2.2.1 C01: Base scenario

Figure 33 shows the results of the electrical simulation of the PRC60 system at the moment of maximum coincident demand when working under scenario C01. This moment would be used in subsequent simulations.

This case represents the BAU scenario. Simulation evidences voltage problems on almost all the bars of the right side of the ring circuit (indicated in blue). Voltage values drop to 0.89 pu, which is out of a normal operational scenario where values should always be between 0.95 pu and 1.05 pu.

This system already has five capacitor banks distributed on medium-voltage bars and one in the slack node, a high-voltage bus bar. Nevertheless, even though they are activated at maximum capacity, voltage problems persist in this scenario.

Notice that some equipment is colored red because they are overloaded, although the optimization model constrained them to avoid exceeding their maximum capacity. The reason lies in the simplified mathematical representation used to keep model linearity. Fortunately, overloads are 1.6% on average, which is a small and acceptable value.

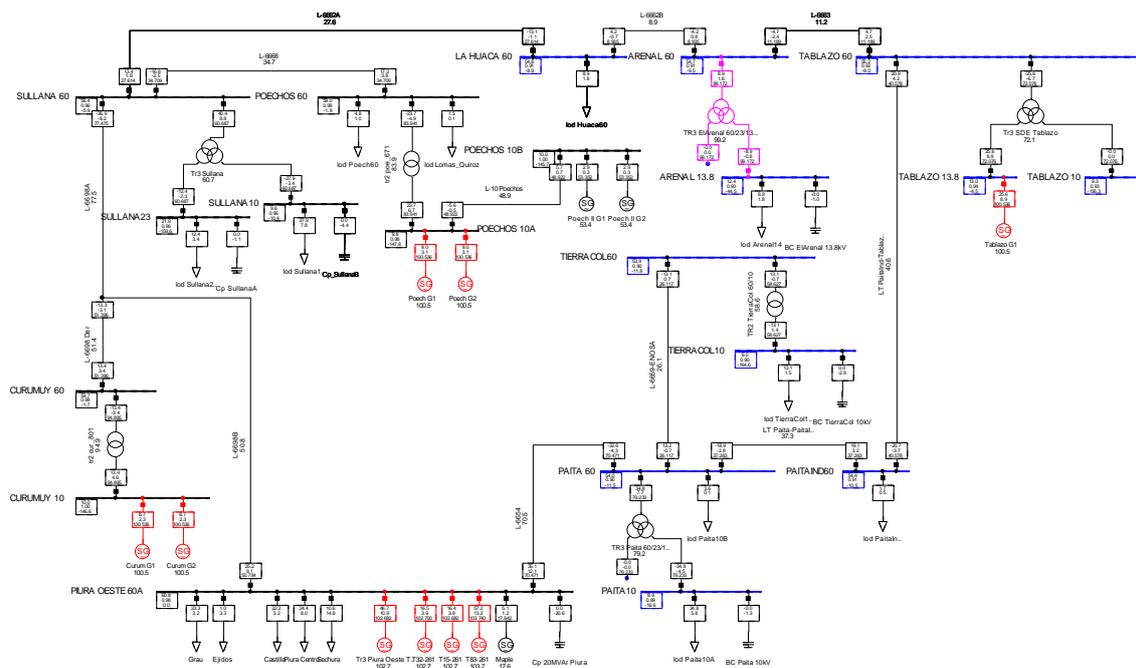


Figure 33. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C01.

The annual operation cost valorizes the production of conventional generators. The cost of fictitious generators is mainly originated by capacitive power that must be injected (through resource variable) to accomplish bus bars voltage limits.

Elapsed time: 11m 22.812s

Total scenario C01 cost is \$ 375'932,219

- Operation cost: \$ 42'153,810
- Fictitious cost: \$ 333'778,409

5.2.2.2 C02: Capacitor banks

Voltage problems of the PRC60 system can be economically solved in scenario C02 by considering only new capacitor banks. These new banks, and all system additions, are supposed to take place on bars Arenal 60 kV, La Huaca 60 kV, or Paita 60 kV.

As shown in Figure 34, the solution proposed by the optimization model manages to correct voltage problems of the right-side bus bars of the system by adding two new capacitor banks on bars La Huaca 60 kV (C_2) and Paita 60 kV (C_3). However, it must be noticed that simulation evaluates a specific moment where maximum coincident demand takes place.

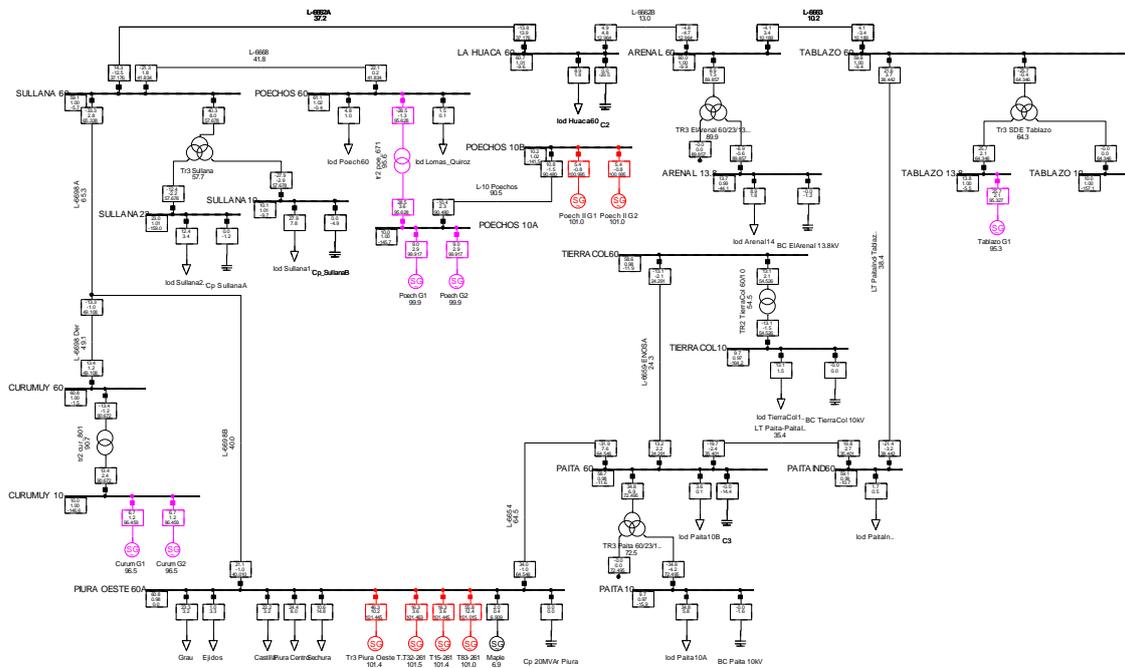


Figure 34. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C02.

Installed power is 20 MVAR for C_2 and 15 MVAR for C_3 , both banks are static and remain at the same step (maximum capacity) throughout the period. Furthermore, this operation scenario does not require existing capacitor banks installed on bars Tierra Colorado 10 kV and Piura Oeste 60 kV.

Elapsed time: 44m 33.700s

Total scenario C02 cost is \$ 65'996,362

- Operation cost: \$ 34'107,197
- Fictitious cost: \$ 31'816,115
- Capacitors cost: \$ 73,050

Fictitious cost comprises inductive power (43.7%) and capacitive power (56.3%), denoting that voltage levels are under or over specified limits on some periods. These problems occur since capacitor banks have a static behavior.

The total cost of this scenario is -82.4% lower than scenario C01.

5.2.2.3 C03: Energy storage systems

For scenario C03, the optimization model only considers the installations of energy storage systems to solve voltage problems of PRC60.

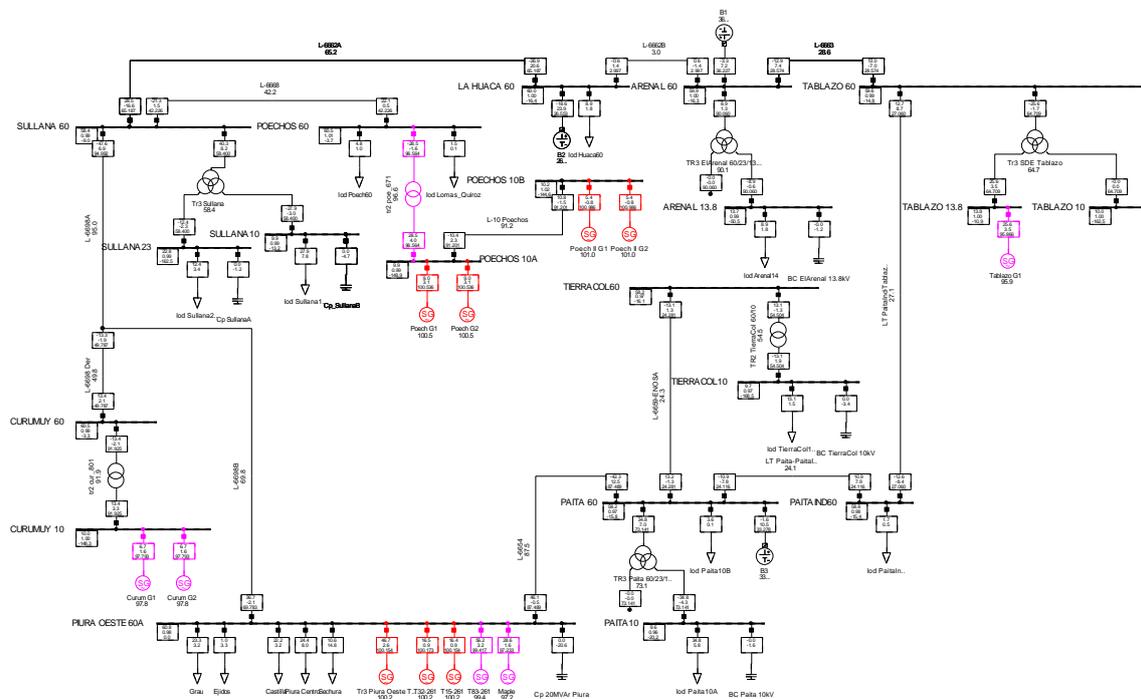


Figure 35. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C03.

As a result, three BESS are added to the PRC60 system and, as shown in Figure 35, manage to solve voltage problems at the moment of maximum demand. A BESS of 22 MVA is installed at Arenal 60 kV (B_1), 114 MVA at La Huaca 60 kV (B_2) and 32 MVA at Paita 60 kV (B_3). All BESS have 4 hours of storage.

Although batteries are faster enough to respond to load fluctuations, additional capacitive power is needed when operational limits of BESS prevent them from injecting more reactive power. The model could have installed a higher capacity of BESS to gain more reactive power into the system, but economic benefits did not justify additional investment. In that sense, all existing capacitor banks are used in this scenario.

Elapsed time: 2h 13m 31.670s

Total scenario C03 cost is \$ 97'524,802

- Operation cost: \$ 37'432,446
- Fictitious cost: \$ 16'280,087
- Batteries cost: \$ 43'812,269

The total cost of this scenario is -74.1% lower than scenario C01.

All subsequent scenarios consider that a maximum of 100 MW can be installed per BESS. Nevertheless, in this scenario, this value was increased up to 150 MW since previous solutions set B_2 to 100 MW, denoting that this parameter was limiting the optimal solution.

5.2.2.4 C04: Solar plants and energy storage systems

In this fourth scenario, the optimization model is limited to using energy storage systems and solar plants as solution alternatives, and, again, results achieved to correct voltage profiles on right-side bus bars of the PRC60 system.

Additions in this scenario consist of one solar plant and three BESS. As in the previous scenario, storage systems are installed in all three available bars but with distinct capacities, which are 31 MVA for B_1 , 53 MVA for B_2 and 19 MVA for B_3 . For the case of renewable plants, just one plant was decided to be built in La Huaca 60 kV with a nominal power of 30 MVA solar (S_2).

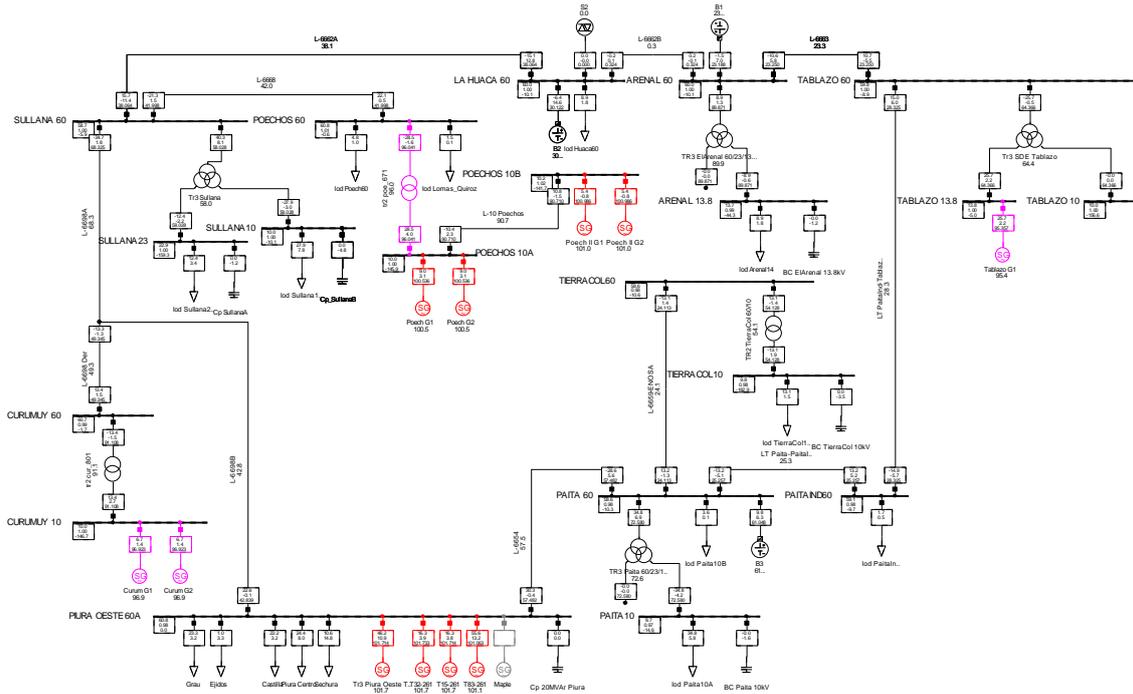


Figure 36. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C04.

To complement the operation scenario, the optimization model decides to deactivate the existing capacitor bank of slack bar Piura Oeste 60 kV and, also, to shut down conventional generator Maple, located in the same bus bar.

Elapsed time: 3h 24m 58.500s

Total scenario C04 cost is \$ 76'750,884

- Operation cost: \$ 35'006,033
- Fictitious cost: \$ 8'847,293
- Batteries cost: \$ 26'861,093
- Solar plants cost: \$ 6'036,465

The total cost of this scenario is -79.6% lower than scenario C01.

Fictitious cost in this scenario represents the need for additional capacitive power in an amount that is almost half that of the previous scenario C03.

5.2.2.5 C05: Wind pants and energy storage systems

As an alternative scenario, in this fifth case, the optimization model solves voltage problems by considering energy storage systems combined with wind plants.

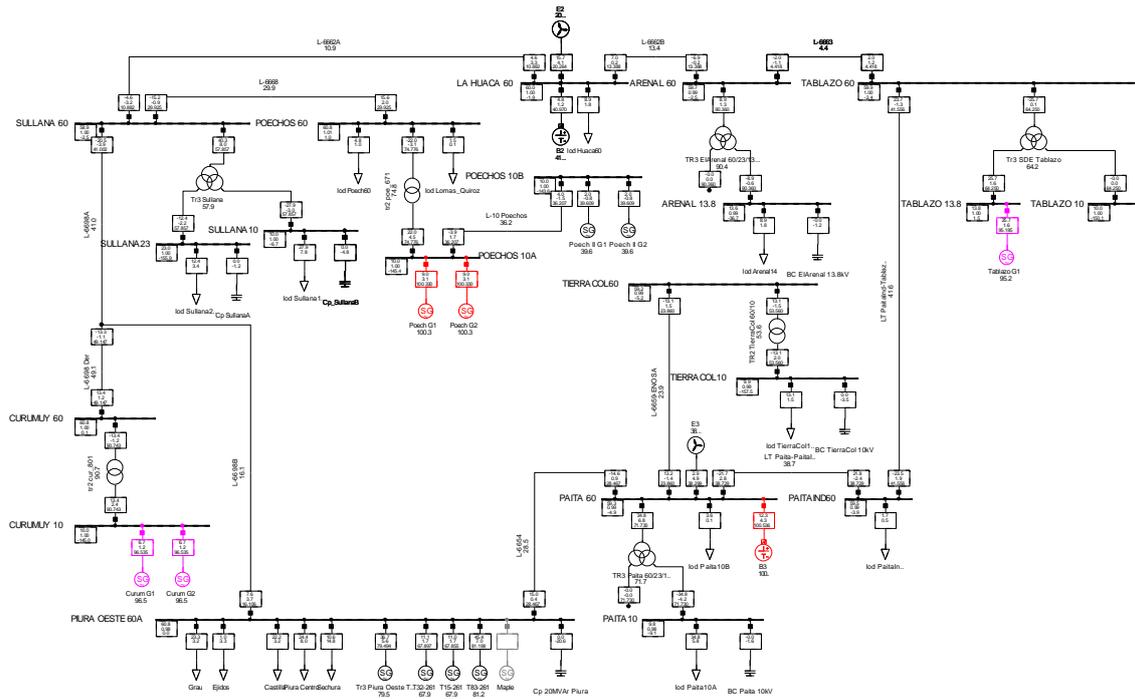


Figure 37. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C05.

Specifically, results indicate that two renewable plants of 80 MVA (E_2) and 15 MVA (E_3) must be installed in La Huaca 60 kV and Paita 60 kV, respectively. Additionally, and unlike the previous scenario, two BESS are considered in this solution: B_2 with a capacity of 12 MVA, and B_3 with 13 MVA.

All existing capacitors are activated in this scenario, and just the Maple generator is off.

Elapsed time: 1h 42m 16.840s

Total scenario C05 cost is \$ 59'646,091

- Operation cost: \$ 28'072,355
- Fictitious cost: \$ 5'938,582
- Batteries cost: \$ 6'519,683
- Wind plants cost: \$ 19'115,471

The total cost of this scenario is -84.1% lower than scenario C01.

The fictitious cost in this case also represents the need for additional capacitive power.

5.2.2.6 C06: Solar and wind plants, energy storage systems, and capacitor banks

Finally, scenario C06 makes all available technologies compete. In this way, the final solution can take advantage of the main benefits of each technology, optimizing expansion plan cost.

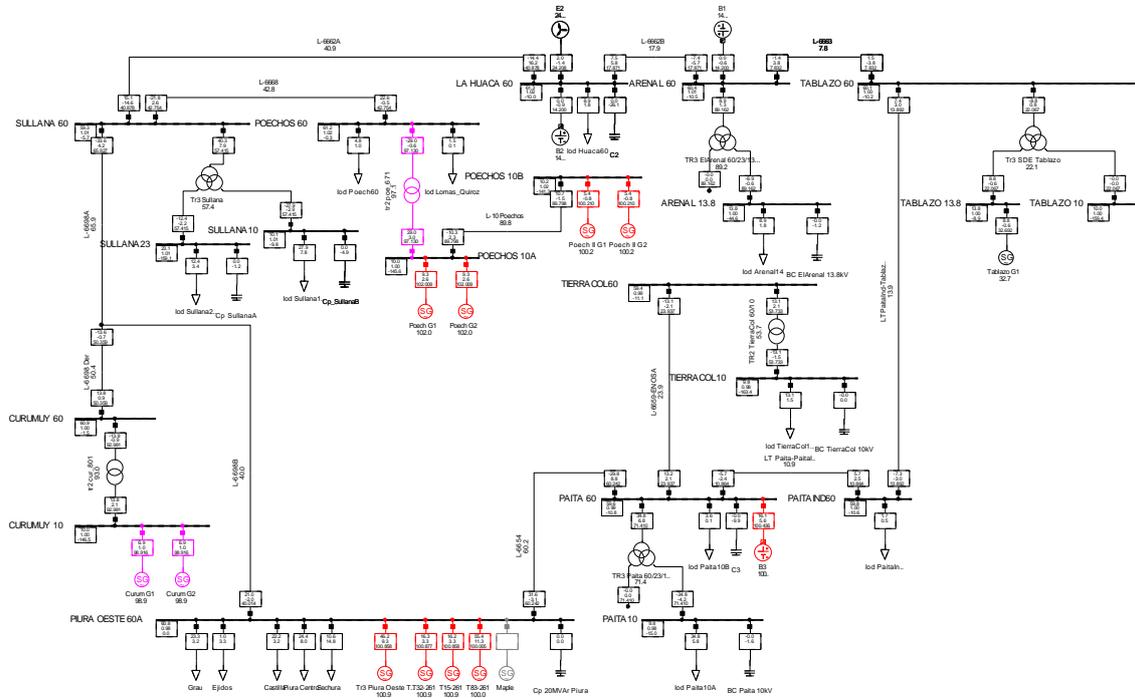


Figure 38. Electrical simulation of PRC60 system at maximum coincident demand - Scenario C06.

The scenario solution includes the installation of two capacitor banks: C_2 of 25 MVA and C_3 of 10 MVA; three energy storage systems: B_1 of 4 MVA, B_2 of 6 MVA and B_3 of 17 MVA; and finally just one renewable plant: E_2 of 10 MVA wind.

Existing capacitor banks in bars Tierra Colorada 10 kV and Piura Oeste 60 kV are disconnected in this operation scenario, and also Maple generator is turned off.

The reason why this conventional generator is off in some scenarios is the fact that it is one of the most expensive plants in the PRC60 system, with a variable cost of 130 \$/MWh, the same that Tablazo G1; while other conventional plants have a value of 30 \$/MWh which is also considered as the spot price of the market (C_{Mg}).

Elapsed time: 5h 22m 53.900s

Total scenario C06 cost is \$ 46'322,795

- Operation cost: \$ 30'395,787

- Fictitious cost: \$ 6'800,545
- Capacitors cost: \$ 73,050
- Batteries cost: \$ 7'041,258
- Solar plants cost: Not installed
- Wind plants cost: \$ 2'012,155

The total cost of this scenario is -87.7% lower than scenario C01.

Again, the fictitious cost is mainly additional capacitive power. Although it represents a significant amount of money, it has to be mentioned that this reactive energy is being valued at a considerable value (C_{ENS}). If this fictitious energy were valued at C_{Mg} , the fictitious cost would represent only \$ 34,000.

5.3 Chapter conclusions

Cost Case	Operation	Fictitious	Capacitors	Batteries	Solar PV	Wind	Total
C01	42'153,810	333'778,409	–	–	–	–	375'932,219
C02	34'107,197	31'816,115	73,050	–	–	–	65'996,362
C03	37'432,446	16'280,087	–	43'812,269	–	–	97'524,802
C04	35'006,033	8'847,293	–	26'861,093	6'036,465	–	76'750,884
C05	28'072,355	5'938,582	–	6'519,683	–	19'115,471	59'646,091
C06	30'395,787	6'800,545	73,050	7'041,258	–	2'012,155	46'322,795

Presented optimization model achieve to evaluate electrical systems taking into account AC operation restrictions like technical loses, reactive power, and voltage values; without losing linearity and yielding consistent and verifiable results.

Also, the proposed model accomplish to solve electrical system problems at the minimum cost. Furthermore, demonstrate that, when allowed, it can choose the optimal mix of technologies to provide the best expansion plan in term of cost.

6 Case Study: Electrical system of eastern Peru

In this final chapter, an application of the three main contributions of this thesis will be performed on an electrical system elaborated, taking as a base an existing system of eastern Peru.

As indicated in 5.1.9, the application case for the proposed planning framework must be performed on a medium-size system due to the detailed representation of system elements. In that sense, an existing system belonging to the distribution company Electro Oriente, called the ELOR system, was taken as the base case. Over this system, some considerations have been established to build the ELOR-C system, a critical version of the ELOR base system.

6.1 System description

The system is located in Loreto, which belongs to the northeast jungle region of Peru. Demand is primarily domestic. Industrial load represents less than 10% of total installed power.

The original ELOR system has two points of interconnection to the National Electrical System (SEIN). One is found in San Martín, at the southernmost substation denominated Bellavista 138 kV, from which a transmission line starts towards substation Juanjuí 138 kV. Since this line is usually open thus it is not considered into the ELOR-C system. The other point is in Moyobamba, at the northernmost substation called Belaunde Terry 220/138 kV, towards which a 220 kV transmission line arrives. This 220 kV line crosses the Andes mountain range. It is subjected to several climatological phenomena that cause an average of 8 hours of disconnection each year (considering historic failures from 2017 to 2020 as reported by Peruvian ISO: COES), being one of the reasons why it was considered a critical electrical system by the regulator Osinergmin [162]. This line's

capacity was limited when constructing the ELOR-C system, forcing the optimization model to obtain an expansion plan that turns this system into a resilient one.

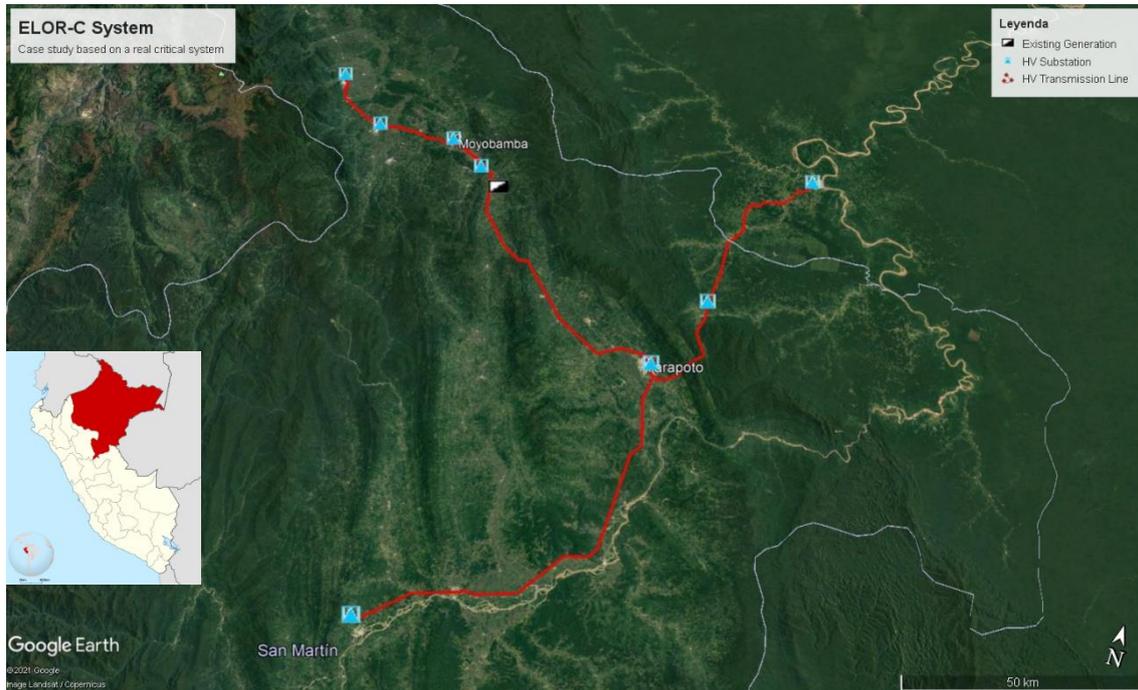


Figure 39. Satellite view of ELOR-C system.

Load power values used in the ELOR-C system correspond to projected values for 2026, approved by the regulator Osinergmin [163]. Moreover, system topology and capacities also correspond to 2026, except the 220 kV transmission line capacity that arrives at Belaunde Terry 220 kV, which is limited to its 2018 maximum active power flow value. A generator represents this injection within the ELOR-C system, and its capacity corresponds to the maximum power injected during the system maximum coincident demand moment.

The total installed load capacity of the ELOR-C system equals 127.79 MW, while the total installed generation capacity is just 75.80 MW. Each load has an associated demand profile (12 available yearly registers). Also, it is considered that the slack bar (Belaunde Terry 220 kV) has a voltage of 0.975 pu as an initial condition, affecting the voltage quality of all bus bars downstream.

In that sense, the optimization problem involves finding the best mix of technologies that provide the necessary generation capacity to supply system demand requirements with an adequate voltage quality.

ELOR-C system consists of 44 bars, 3 existing generators, 14 loads, 3 existing capacitors, 42 existing transmission lines, 28 renewable projects, 6 BESS projects, and 6 capacitor projects. For each project, the model has to decide the optimal capacity to install.

6.2 Renewable time-series synthesis

Generation technologies considered as expansion alternatives for the ELOR-C system are solar photovoltaic (PV), wind and hydro plants, in addition to battery energy storage systems (BESS) and capacitive reactive power compensators. All alternatives are supposed to be installed in one of the following bars: Belaunde Terry 138 kV, Gera 60 kV, Pongo 60 kV, Rioja 60 kV, Tarapoto 138 kV, and Yurimaguas 33 kV.

While the injections of BESS and capacitors are calculated by the optimization model, as formulated in Chapter 5, renewable plants need a time-series to use as an input to deduce their production, having each technology a particular profile as depicted in Figure 40.

In the case of hydro plants, three alternatives presented in [164] are considered for this study. For each hydro project, a time-series obtained from three existing hydro plants is associated. The selected plants are Platanal G2, Chaglla G2, and Cerro del Águila G3, whose power production records were obtained from COES for 2018.

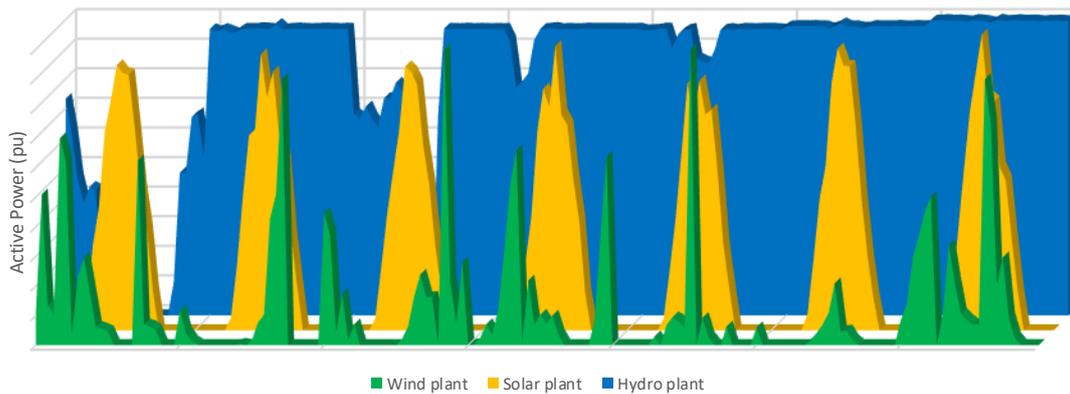


Figure 40. Active power production profiles for three generation alternatives with distinct technologies.

On the other hand, a synthesis process had to be performed to obtain the expected production time-series for each solar PV and wind plant formulated as expansion options. Twenty-five time-series were synthesized using the proposed methodologies presented in Chapter 3: 13 for wind plants and 12 for solar PV plants.

An average of 2 projects for each technology, solar PV and wind, were formulated for each of the six available bus bars. This situation caused the projects to be very close to each other., as shown in Figure 41, making gathered climatological data very similar.

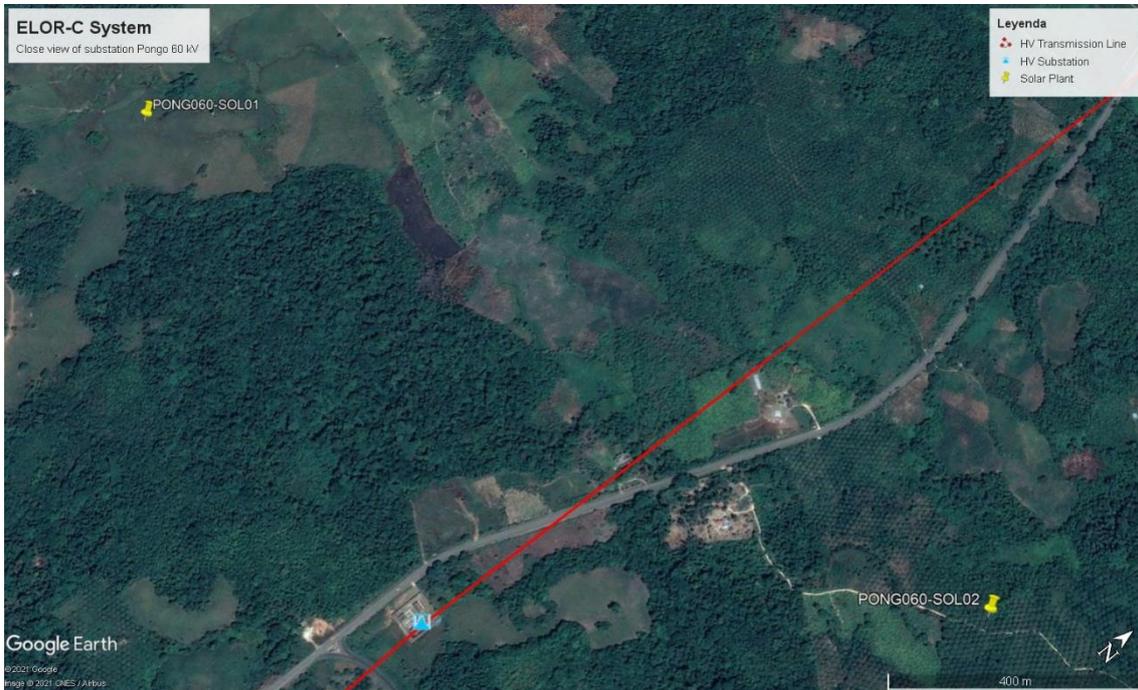


Figure 41. Satellite view of substation Pongo 60 kV and two solar plant projects of ELOR-C system.

Different seeds were used during the random generation processes To overcome that situation. The last five digits of the decimal part of its coordinates served as seeds for each plant, as described in Chapter 3.

As a result, the proposed methodology generates unique time-series for each project no matter how close they are.

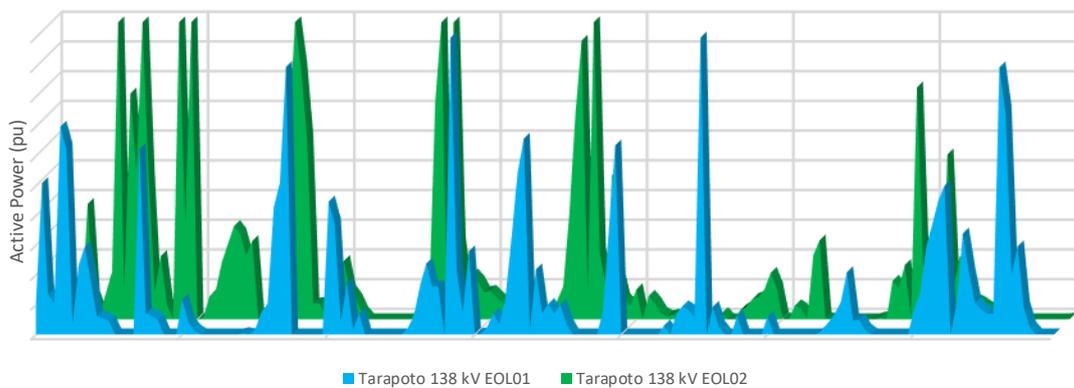


Figure 42. Active power production profiles for two wind plants located closely together.

Figure 42 shows the power profile of two wind plants formulated to be connected in substation Tarapoto 138 kV and demonstrates that synthetic time-series are sufficiently different despite this.

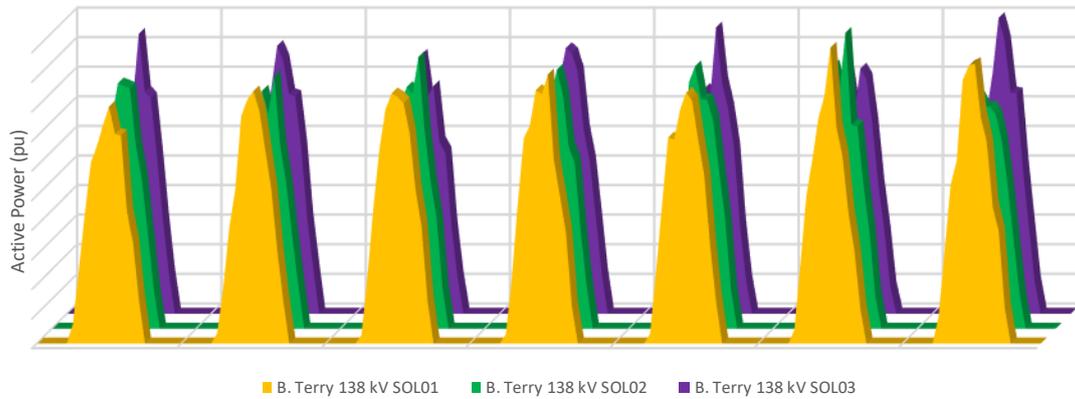


Figure 43. Active power production profiles for three solar plants located closely together.

Similarly, Figure 43 serves to verify that power profiles for the three solar projects formulated for substation Belaunde Terry 138 kV are not equal despite the proximity between them.

Although the quality of wind resources in the geographic area of the ELOR-C system is not as good as solar resources, both technologies were treated in the same way when formulating the generation alternatives since implantation decision should be evaluated by the optimization model instead of a priori evaluation or judgment.

6.3 Demand and renewable time-series clustering

Forty time-series are available for the ELOR-C system. 12 time-series correspond to demand records obtained from feeders at 10 kV, 22.9 kV, and 33 kV. Another 3 time-series belong to the three existing hydro plants selected to be taken as dispatch model. Finally, 25 time-series come from the synthesis process carried out in 6.2.

The time scale for all time-series is 1-hour, having each one 8760 values. In that sense, this section aims is to reduce their size to only 168 values, i.e., 7 typical days.

As proposed in Chapter 4, a K-Means Clusterization process is used to accomplish this goal. The metric used to cluster time-series is Soft-DTW with 30 initial runs.

The clustering process was performed on the free cloud service provided by Google called Colab and was executed in 157.827 seconds. The obtained Silhouette coefficient was 0.339409.

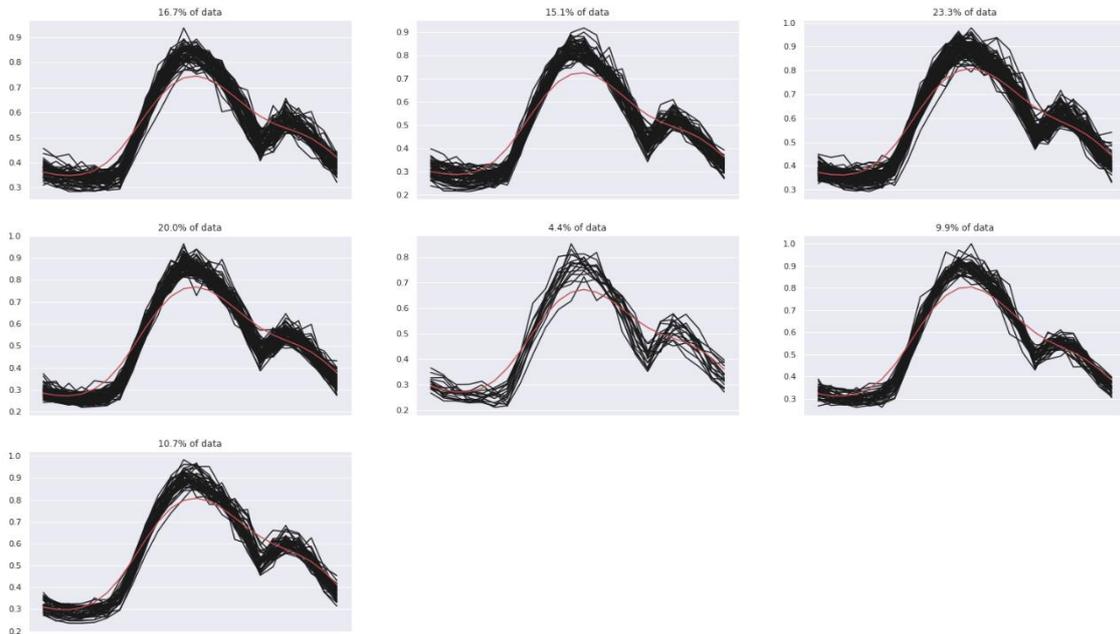


Figure 44. Clusters obtained for ELOR-C system.

Figure 44 shows the seven clusters obtained for the ELOR-C system. In contrast with the case study of Chapter 4, these clusters seem to be very similar. That is because the 1-hour scale smooths the variations that occurred intra-hour. Differences between clusters would be more visible if the time scale were 15-min or shorter.

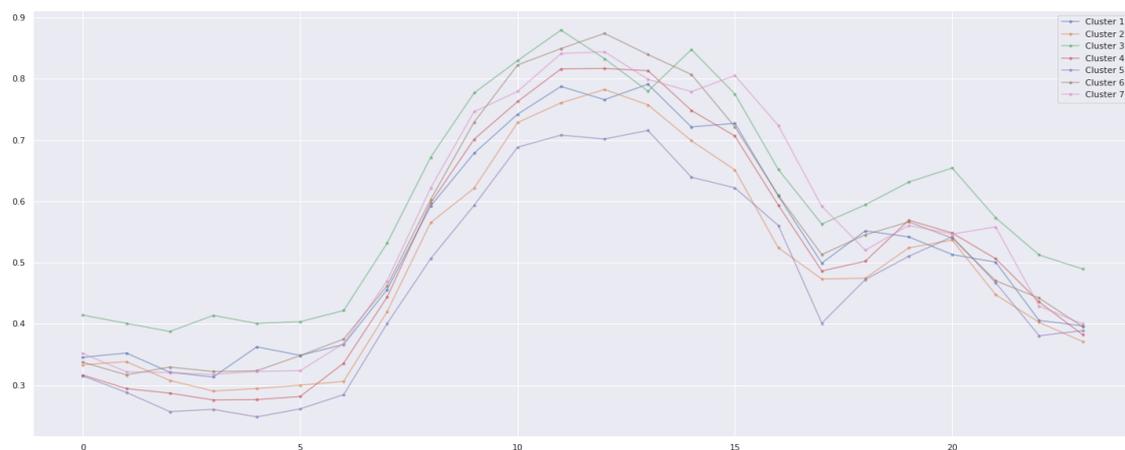


Figure 45. Real centroids of clusters obtained for ELOR-C system.

As indicated, that similarity is just a visual perception since when analyzing real centroids altogether, it is notorious that each cluster has a different magnitude representing the 7

typical days that methodology looked to obtain. Each typical day will be used in the optimization process considering their weight.

6.4 Costs of expansion technologies

Economic variables shown in Table XI have been elaborated using the values presented in [93], where historical investments in generation capacity in Peru are studied. Future cost projections for solar PV and wind technologies presented by IRENA and IEA are also recuperated. Likewise, the cost for battery storage systems (BESS) took as input the projections presented by NREL. It is important to mention that an expert judge finally decided all variables gathered and those missing.

Table XI. Economic variables for expansion technologies

	Investment Cost	OPEX Cost (%)	Useful life (year)
Hydro	2,700 \$/kW	4.0%	30
Solar PV	1,098.5 \$/kW	1.8%	25
Wind	1,516.5 \$/kW	2.0%	25
BESS	500 \$/kWh	2.5%	10
Capacitor	15 \$/kVAR	1.5%	25

Additional considerations for BESS technology are that storage duration is 4 hours, round-trip efficiency is 89%, deep-of-discharge is 100%, and it can operate at maximum power for a maximum of 7 hours.

Energy injected by any of these technologies is valued at 0 \$/MWh. On the other side, generation cost for existing generators is assumed 25 \$/MWh while load curtailment is penalized at 6,000 \$/MWh.

6.5 Expansion planning optimization

As indicated in 6.1, the optimization problem consists in finding the optimal mix of technologies that provide the necessary generation capacity to supply system demand requirements keeping an adequate voltage quality. Results will represent the optimal expansion planning for the year 2026 under the premises established in the previous sections.

In other words, the optimization problem will attempt to resolve a medium-term planning problem considering the short-term operation concerns.

6.5.1 C01: Base scenario

As expected, the base case configuration cannot provide the necessary energy the system needs to operate correctly. Such a situation is depicted in Figure 46, where it is notorious that generation capacity reaches a maximum value of around 75 MW.

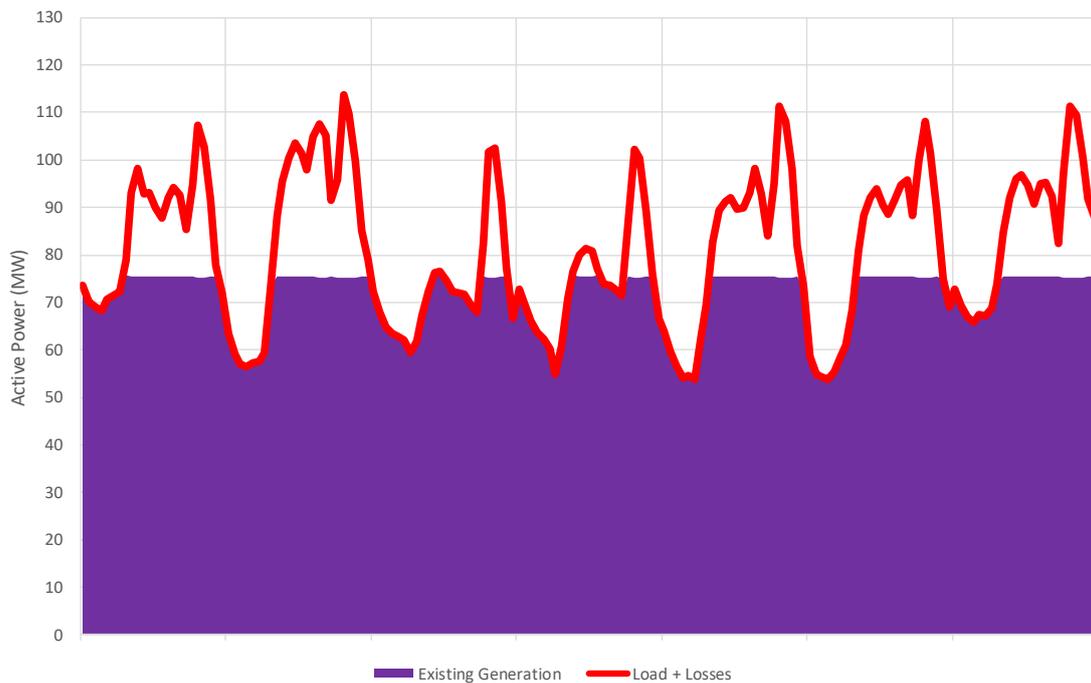


Figure 46. ELOR-C system operation – base case C01.

For this scenario, the model set existing capacitors located in Tarapoto 10 kV and Tarapoto 138 kV at maximum capacity, 10 MVAR, and 5 MVAR, respectively. On the other side, the existing capacitor in Belaunde Terry 220 kV was not activated in this scenario.

Elapsed time: 3m 57.389s

Total scenario C01 cost is \$ 933'518,335

- Operation cost: \$ 10'907,492
- Losses cost: \$ 1'612,949
- Fictitious cost: \$ 920'997,894

35% of the fictitious cost is due to load curtailment, while the rest corresponds to additional capacitive power needed to maintain voltage levels between 0.95 pu and 1.05 pu.

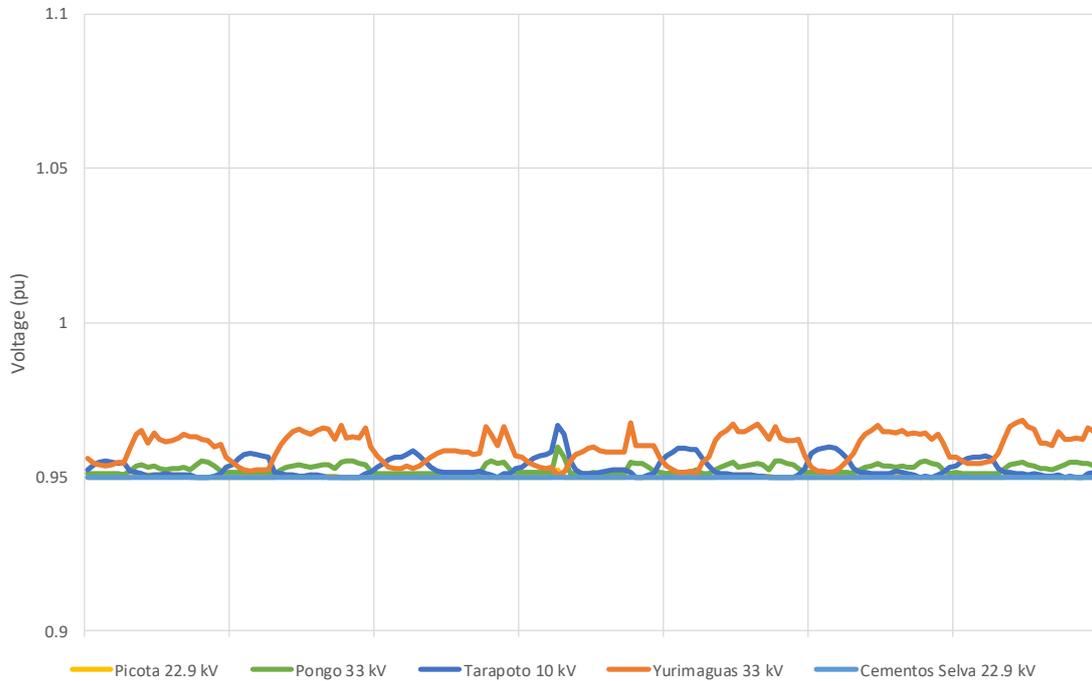


Figure 47. Voltage levels in load bars of ELOR-C system – base case C01.

However, despite the fictitious reactive capacity injected into the system, voltage levels drop to minimum values throughout the analyzed period, as shown in Figure 47.

6.5.2 C02: Optimized scenario

The optimization model obtained an optimal expansion plan that minimizes operation and investment costs while guaranteeing short-term variables like voltage stay in an acceptable range.

The expansion plan consists in installing 80 MVA of solar plants (6 plants) and 15 MVA of hydro plants (3 plants) distributed among the buses Tarapoto 138 kV, Pongo 60 kV, and Yurimaguas 60 kV as shown in Figure 48. Additionally, 60 MVA and 240 MWh of BESS (6 plants) were selected to be installed among all available buses. Finally, the model decided to install 3 capacitors with a total capacity of 25 MVAR between buses Tarapoto 138 kV, Pongo 60 kV, and Rioja 60 kV. No wind plants were chosen since the wind resource quality in this area is not as good as solar.

Existing capacitors in buses Tarapoto 10 kV and Tarapoto 138 kV were activated with 7.5 MVAR and 5 MVAR, respectively. Also, the usage factor of the existing generation got decreased due to the presence of new capacity. The operation cost of the existing generation in this scenario C02 is 23% lower than that obtained in scenario C01.

Another important finding is that transmission losses are reduced by 19% in the optimized scenario C02, where new generation plants are distributed within the ELOR-C system.



Figure 48. Satellite view of ELOR-C system with selected renewable projects.

Elapsed time: 7h 45m 7.100s

Total scenario C02 cost is \$ 81'083,477

- Operation cost: \$ 8'389,313
- Losses cost: \$ 1'302,564
- Fictitious cost: \$ 29'915,707
- Capacitors cost: \$ 53,438
- Batteries cost: \$ 21'988,099
- Solar plants cost: \$ 12'768,538
- Wind plants cost: Not installed
- Hydro plants cost: \$ 6'647,818

The total cost of this scenario is -91.3% lower than scenario C01.

Fictitious cost is 97% lower than the obtained in the base case. 83% of this fictitious cost is produced by additional capacitive power, while load curtailment represents 16%. The presence of load curtailment means that the expansion plan can not meet all demand requirements. A total of 791.75 MWh of non-served energy occurs in this scenario, equivalent to an effective power of 0.09 MW during the whole year. In scenario C01, the non-served energy is 53,182.46 MWh.

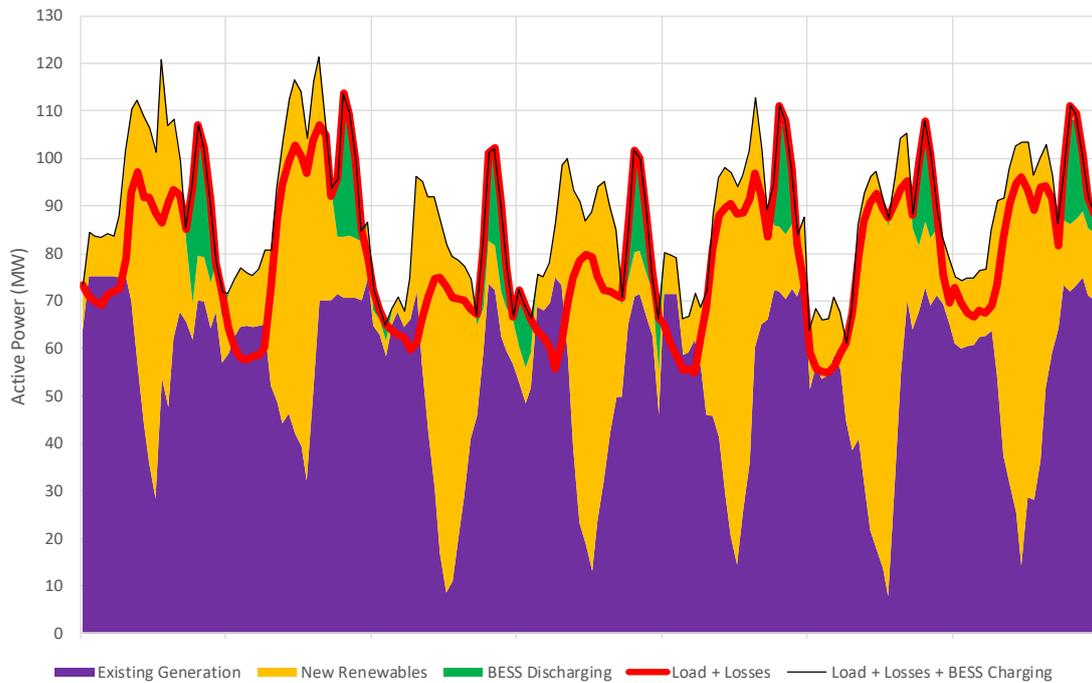


Figure 49. ELOR-C system operation – optimized case C02.

The total energy provided by renewable plants (solar and hydro) is 181.28 GWh, representing a penetration of almost 38% into the ELOR-C system, whose annual demand energy is 477.17 GWh. This fact demonstrates that the optimization model maximizes renewable penetration without causing renewable wastage or voltage problems.

An essential part of the expansion plan is the battery energy storage systems that help save the energy produced in excess due to the considerable amount of solar energy installed and return that energy to the system when demand reaches its peak hours. In Figure 49, this operation dynamic can be seen when generation (purple and yellow areas) is greater than load demand (red line) and when the stored energy is then re-injected to the system (green area).

Stored energy in BESS throughout the analysis period is presented in Figure 50 in color green. Also, power consumed when charging (yellow) and power injected when discharging (red) is presented in the same graphic.

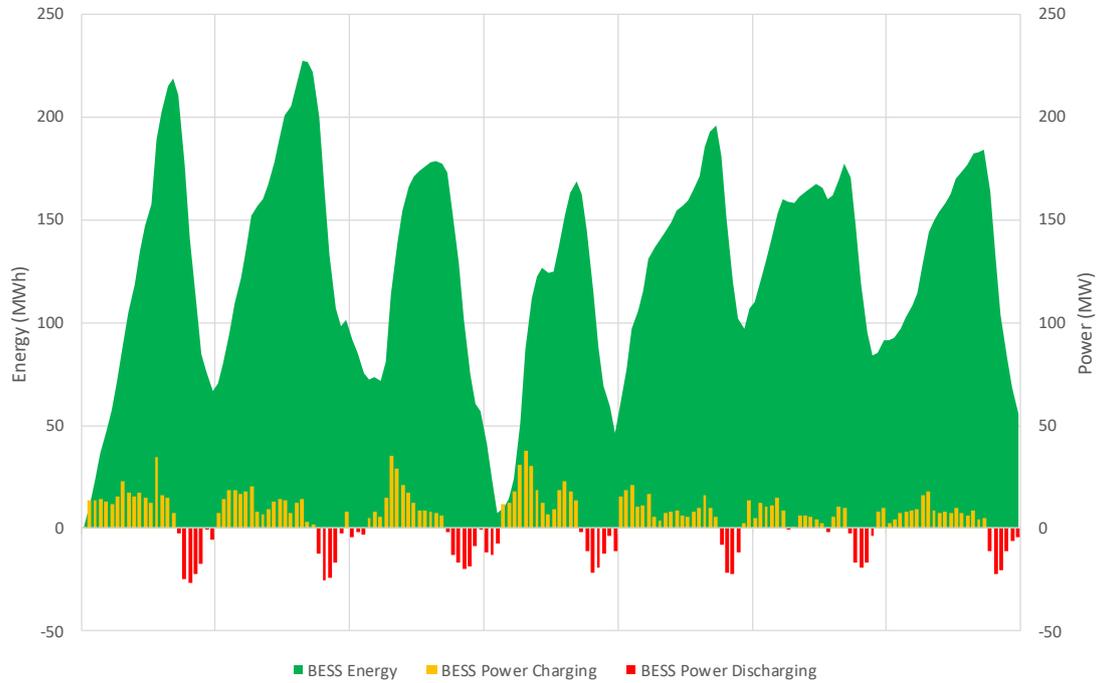


Figure 50. Operation of BESS installed in ELOR-C system – optimized case C02.

Thanks to the fact that the model considers each technology's generation profile, the voltage levels, and the wastage energy, the optimal plan does not include any wind plant. If the analysis were just a cost issue, the model would have decided to install wind plants over BESS plants, but it was not. Installing 60 MW of BESS requires an annual cost of \$ 21'988,099, but if an equivalent wind capacity had been installed, the cost would have been only \$ 13'421,022, 39% less. However, the benefits of having BESS in the short-term operation of the system make the model choose it over other available technologies.

The optimal expansion plan also improves the voltage levels of the bars presented in Figure 47. The same buses are shown in Figure 51, where it can be seen that now the voltage levels fluctuate around 1 pu. It is imperative to mention that voltage levels practically follow the production behavior of renewable energies, having peaks values when injections are at high levels and getting low values when renewable generation decreases. A relevant insight is that although solar variability affects system voltage levels, the optimization model obtained a plan that has better voltage levels than the base case C01 all the time, without the need to have pre-limited the amount of renewables the

model can evaluate. This fact represents a game-changer for the current planning process where the maximum renewable capacity allowed to be installed in a system is calculated ex-ante to precisely avoid voltage problems. With the model presented in this thesis, such a procedure would not be necessary anymore.

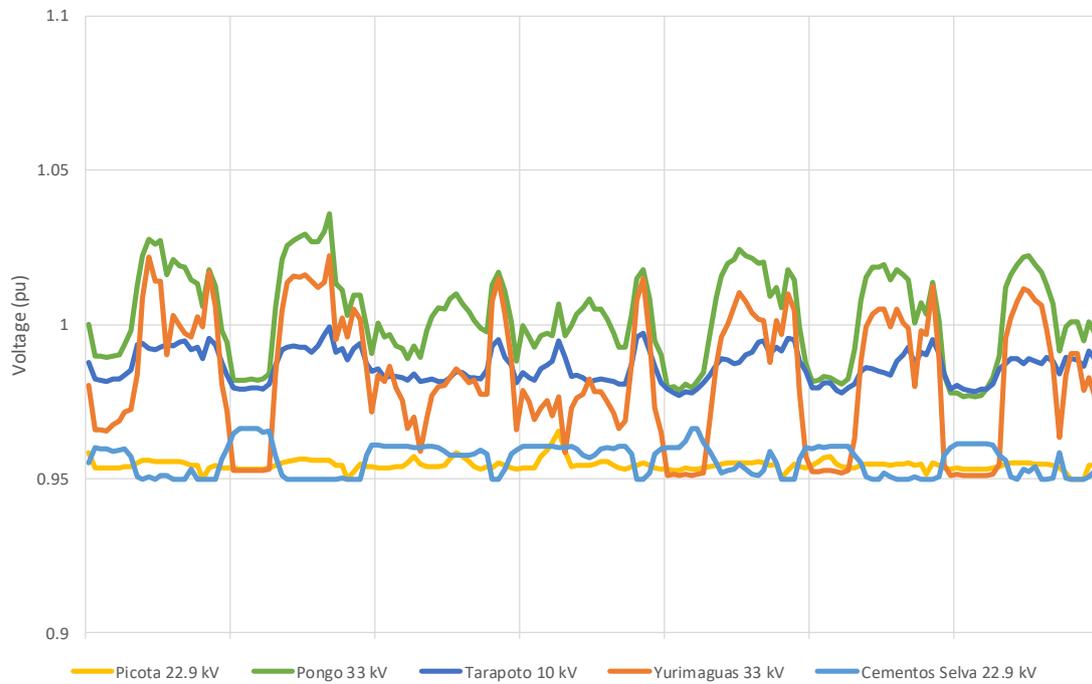


Figure 51. Voltage levels in load bars of ELOR-C system – optimized case C02.

If the ELOR-C system problem were solved using an optimization model that does not consider the voltage and reactive power, the obtained expansion plan would be cheaper by 60.7%. DC expansion plan would contemplate 20 MW of solar plants, 15 MW of hydro plants, and 25 MVA and 100 MWh of BESS, representing a total annual investment of \$ 19'006,161. The new generation capacity was added to bridge the gap between coincident maximum load demand and existing installed generation, which is just one of the existing problems of the ELOR-C system.

It is essential to mention that the proposed model allows us to evaluate the quality of expansion plans comprehensively, considering active and reactive power, losses, and voltage levels. In effect, a DC model would have neglected the influence of intermittent renewables over the voltage values since a DC flow establishes a value of 1 pu to all buses permanently.

6.6 Chapter conclusions

This case study demonstrates that the planning methodology proposed in this thesis is helpful to perform a complete optimal expansion planning study.

Proposed synthesis methodologies allow to generate the required hourly time-series for any identified project without having in-site measurements, which helped speed up the whole planning process.

Likewise, the proposed clustering technique managed to reduce the problem dimensions, allowing the application of a realistic optimization model that considers important electric variables such as voltage, reactive power, and losses.

7 Conclusions

7.1 Findings

- In this Thesis, a comprehensive planning framework was presented to perform optimal expansion planning studies of the coordinated transmission and generation of actual power systems, even when historic measurements on solar and wind resources are not available.
- The presented planning framework could be a relevant input to formulate updates on existing planning regulations or to propose new energy politics.
- In Chapter 3, it is presented two synthesis methodologies to generate hourly time-series of solar PV and wind plants, that are simple but effective in generating synthetic values starting from aggregate parameters without having historical time-series.
- Both synthesis methodologies successfully replicate the typical behavior of solar PV and wind plants, creating unique time-series even when project locations are very close.
- Synthesis methodologies are flexible and parametric, making it possible to generate multiple time-series scenarios modifying input parameters to achieve enough range of cases to incorporate uncertainties in a robust planning process.
- In Chapter 4, it is presented a clustering technique to reduce the dimension of the data from real application cases. This technique produces energy planning-abled time-series since it maintains the correlation, simultaneity, randomness, and variability within all the time-series considered.
- A coefficient was proposed to measure the quality of clustering results, which will be helpful when evaluating new clustering techniques or metrics that may appear in the future.
- In Chapter 5, it is presented an optimization model that achieves to evaluate electrical systems taking into account AC operation restrictions like technical losses, reactive power, and voltage values; without losing linearity and yielding consistent and verifiable results.

- The proposed optimization model manages to solve the coordinated generation and transmission medium-term planning problem while considering realistic short-term operation concerns. It also finds the optimal operation program for battery energy storage systems to maximize the penetration of renewable energy into the system, providing the flexibility required to absorb its intermittencies.
- The proposed optimization model manages the trade-off between the technical and economic benefits gained when implementing new equipment into the system and the investments and operation costs involved in that decision.
- In Chapter 6, the complete planning framework was successfully applied to a Peruvian medium-size power system to obtain a medium-term optimal expansion plan.
- They obtained results show that the proposed model improves voltage quality by deciding the best amount of renewable energies to install in the system without pre-limiting the amount of renewables for each bus bar.
- The presented model allows the analysis of voltage levels as a dimension of the expansion plan's quality, representing a clear improvement over traditional DC models that consider a constant value of 1 pu for all bus bars, hiding the impacts of intermittent generation.

7.2 Future work

- The inclusion of environmental restrictions should be addressed in future research on energy planning to adapt the optimization model to the national objectives and commitments of the energy matrix.
- In order to address, in a better way, the voltage regulation problem within the optimization problem, it could be considered the addition of variable reactive compensators as an expansion option.
- The proposed model could be expanded to include the representation of pumped storage hydropower as an alternative to battery energy storage systems.
- It would be essential to study the best way to cluster energy-related time-series and find a standard or minimum acceptable value for the Silhouette coefficient that evaluates the quality of the clusterization process.
- When it is necessary to analyze a large-sized electrical system, evaluate a model of greater complexity, or study the inter-year expansion (dynamic programming), a big

scale technique must be implemented to allow the optimization model to converge towards an optimal response.

- Further studies must be done on the iterative approach presented in this dissertation to know the limitations of its application in energy planning problems.
- It is recommended to establish a Peruvian System Standard Case on which new planning proposals can be applied and compared to help future research efforts on the Peruvian reality.
- A national software could be implemented to improve access to information to accelerate the adoption of renewable technologies involved in the distributed generation using the methodologies presented in Chapter 3 about renewable time-series synthesis.

A Complementary equations to calculate solar radiation

Relation between Horizontal Daily Radiation and Horizontal Hourly Radiation

This sub-section corresponds to the needed equation to resolve eq. (3.2).

Variable r_t is defined as indicated in [106].

$$r_t = \frac{\pi}{24} (a + b \cos \omega) \frac{\cos \omega - \cos \omega_s}{\sin \omega_s - \omega_s \cos \omega_s} \quad (7.1)$$

Where coefficients a and b are given by the following equations:

$$a = 0.4090 + 0.5016 \sin(\omega_s - \frac{\pi}{3}) \quad (7.2)$$

$$b = 0.6609 - 0.4767 \sin(\omega_s - \frac{\pi}{3}) \quad (7.3)$$

To calculate the hour angle ω , the solar time (t_s) has first to be determined. Solar time does not coincide with civil time (t_c), so some corrections have to be applied as indicated in eq. (7.4). Civil time should be written for the midpoint of the hour. For example, for the hour 10 of a day, civil time must be $t_c = 9.5 \text{ hr}$ because its hour bin comprises from 9 to 10 hr .

$$t_s = t_c + \frac{\lambda}{15^\circ} - \text{GMT} + E \quad (7.4)$$

In this equation, all terms are expressed in hours, GMT is the time zone, and E is defined by [106] as:

$$E = 3.82(0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B) \quad (7.5)$$

Where $B = 2\pi(n - 1)/365$ is calculated for the n th day of the year. Subsequently, the hour angle is then given by:

$$\omega = \frac{\pi}{180^\circ} (t_s - 12) \cdot 15^\circ \quad (7.6)$$

Two additional angles have to be defined also. These are the declination angle (δ) and the sunset hour angle (ω_s) which is a function of latitude and declination [106].

$$\delta = 0.006918 - 0.399912 \cos B + 0.070257 \sin B - 0.006758 \cos 2B + 0.000907 \sin 2B - 0.002697 \cos 3B + 0.00148 \sin 3B \quad (7.7)$$

$$\cos \omega_s = -\tan \phi \tan \delta \quad (7.8)$$

Calculation of Extraterrestrial Horizontal Radiation

This sub-section corresponds to the needed equation to calculate eq. (3.3).

The equation for calculating I_o is given by the following expression [106]:

$$I_o = \frac{12}{\pi} G_{on} (\cos \phi \cos \delta (\sin \omega_2 - \sin \omega_1) + (\omega_2 - \omega_1) \sin \phi \sin \delta) \quad (7.9)$$

Where ϕ is the latitude (ϕ) expressed in radians, ω_1 and ω_2 are the hour angles that define an hour ($\omega_1 < \omega_2$) and G_{on} is the extraterrestrial radiation incident on a plane always normal to the radiation. G_{on} is calculated as shown in eq. (7.10).

$$G_{on} = G_{sc} (1.00011 + 0.034221 \cos B + 0.00128 \sin B + 0.000719 \cos 2B + 0.000077 \sin 2B) \quad (7.10)$$

Solar constant G_{sc} is set equals to $1.367 \text{ kW}/\text{m}^2$.

Relation between Horizontal Hourly Radiation and Horizontal Hourly Diffuse Radiation

This sub-section presents the definition of the piecewise function mentioned in eq. (3.4) [165].

$$\frac{I_d}{I} = \begin{cases} 1.0 - 0.09 \cdot k_T & \text{for } k_T \leq 0.22 \\ 0.9511 - 0.1604 \cdot k_T + 4.388 \cdot k_T^2 - 16.638 \cdot k_T^3 + 12.336 \cdot k_T^4 & \text{for } 0.22 < k_T \leq 0.80 \\ 0.165 & \text{for } k_T > 0.80 \end{cases} \quad (7.11)$$

Complementary equations for HDKR anisotropic model

This sub-section corresponds to the needed equation to calculate eq. (3.5).

Three additional values are required to solve mentioned equation: 1) the ratio between beam radiation on a tilted surface to a horizontal surface (R_b), 2) the anisotropy index

(A_i) and 3) the horizon brightening factor (f). Required definitions are presented below [106]:

$$R_b \approx R_{b,ave} = \frac{a_{R_b,ave}}{b_{R_b,ave}} \quad (7.12)$$

$$\begin{aligned} a_{R_b,ave} = & (\sin \delta \sin \varphi \cos \beta - \sin \delta \cos \varphi \sin \beta \cos \gamma)(\omega_2 - \omega_1) \\ & + (\cos \delta \cos \varphi \cos \beta + \cos \delta \sin \varphi \sin \beta \cos \gamma)(\sin \omega_2 \\ & - \sin \omega_1) - \cos \delta \sin \beta \sin \gamma (\cos \omega_2 - \cos \omega_1) \end{aligned} \quad (7.13)$$

$$b_{R_b,ave} = \cos \varphi \cos \delta (\sin \omega_2 - \sin \omega_1) + \sin \varphi \sin \delta (\omega_2 - \omega_1) \quad (7.14)$$

$$A_i = \frac{I_b}{I_o} \quad (7.15)$$

$$f = \sqrt{\frac{I_b}{I}} \quad (7.16)$$

It has to make sure factor f have a value distinct of zero only when beam radiation (I_b) is positive.

B Complementary equations to calculate wind speed

Calculation of air density

This sub-section presents the equations required to calculate air density used in sub-section 3.1.2.1.

For this purpose, first temperature (T) and pressure (p), both absolutes, at a specific height must be calculated using eq. (7.17) and (7.18):

$$T = T_0 - Lh \quad (7.17)$$

$$p = p_0 \left(1 - \frac{Lh}{T_0}\right)^{gM/RL} \quad (7.18)$$

Where $p_0 = 101\,325\text{ Pa}$ and $T_0 = 288.15\text{ K}$ are the sea-level standard atmospheric pressure and temperature, respectively, $g = 9.80665\text{ m/s}^2$ is the earth-surface gravitational acceleration, $L = 0.0065\text{ K/m}$ is the temperature lapse rate, $R = 8.31447\text{ J/mol} - \text{K}$ is the ideal universal gas constant and $M = 0.028644\text{ kg/mol}$ is the molar mass of dry air.

Finally, air density can be calculated as shown in eq.(7.19).

$$\rho = \frac{pM}{RT} \quad (7.19)$$

C Practical AC Optimization Model

In this appendix, the complete optimization model constructed in Chapter 5 is presented.

$$\begin{aligned} \min v = & 8760\Gamma \sum_{t \in T} \rho_{hour}^t \left(f_c \sum_{i \in G} C_g^i g a_i^t + C_{Mg} f_p \sum_{l \in L} (q a_l^t + q r_l^t) \right. \\ & \left. + C_{ENS} f_c \sum_{k \in B} (r a_k^t + r r i_k^t + r r c_k^t + z l o s s_k^t) \right) + \sum_{p \in Z} @I_p u_p \omega_p \\ & + \sum_{b \in B} @I_b e c_b^{nom} \omega_b + \sum_{c \in C} @I_c u_c \omega_c + \sum_{l \in L} @I_l \omega_l \end{aligned}$$

Subject to

$$\begin{aligned} \sum g a_i^t + \sum z a_p^t + \sum e p a_b^t + \sum \left(f a_i^t - \frac{1}{2} q a_l^t \right)_{in} - \sum \left(f a_i^t + \frac{1}{2} q a_l^t \right)_{out} &= \sum d a^t - \\ r a_k^t + z l o s s_k^t &, \forall k \in N, t \in T \\ \sum g r_i^t + \sum z r_p^t + \sum e p r_b^t + \sum \left(f r_i^t - \frac{1}{2} q r_l^t \right)_{in} - \sum \left(f r_i^t + \frac{1}{2} q r_l^t \right)_{out} + \sum c r_c &= \\ \sum d r^t + r r i_k^t - r r c_k^t &, \forall k \in N, t \in T \\ 0 \leq r a_k^t \leq \sum d a^t &, \forall k \in N, t \in T \\ 0 \leq z l o s s_k^t \leq \sum z a_p^t &, \forall k \in N, t \in T \\ g_i^{min} \leq g a_i^t \leq g_i^{max} &, \forall i \in G, t \in T \\ g r_i^t \leq g a_i^t \tan(\cos 0.95) &, \forall i \in G, t \in T \\ g r_i^t \geq -g a_i^t \tan(\cos 0.99) &, \forall i \in G, t \in T \\ (g a_i^t)^2 + (g r_i^t)^2 \leq (g_i^{max})^2 &, \forall i \in G, t \in T \\ z a_p^t = \zeta_p^t \cos(\cos 0.95) &, \forall p \in Z, t \in T \\ z r_p^t \leq z a_p^t \tan(\cos 0.95) &, \forall p \in Z, t \in T \\ z r_p^t \geq -z a_p^t \tan(\cos 0.99) &, \forall p \in Z, t \in T \\ \zeta_p^t = t s_p^t u_p \omega_p &, \forall p \in Z, t \in T \\ e p c_b^t \geq -e p b^{nom} \omega_b &, \forall b \in B, t \in T \end{aligned}$$

$$\begin{aligned}
epd_b^t &\leq ep_b^{nom} \omega_b && , \forall b \in B, t \in T \\
epda_b^t &= epd_b^t \cos(\text{acos } 0.95) && , \forall b \in B, t \in T \\
epdr_b^t &\leq epda_b^t \tan(\text{acos } 0.95) && , \forall b \in B, t \in T \\
epdr_b^t &\geq -epda_b^t \tan(\text{acos } 0.99) && , \forall b \in B, t \in T \\
epa_b^t &= epda_b^t + epc_b^t && , \forall b \in B, t \in T \\
epr_b^t &= epdr_b^t && , \forall b \in B, t \in T \\
ec_b^t &= ec_b^{t-1} - \left(\frac{epd_b^t}{\eta_b} + epc_b^t \right) \Gamma && , \forall b \in B, t \in T / t \geq 2 \\
ec_b^t &\leq ec_b^{nom} \omega_b && , \forall b \in B, t \in T \\
ec_b^t &\geq (1 - D_b) ec_b^{nom} \omega_b && , \forall b \in B, t \in T \\
\sum_{s=t-\tau_b+1}^t epd_b^s &\leq ep_b^{nom} \omega_b \tau_b && , \forall b \in B, t \in T / t \geq \tau_b \\
epd_b^1 &= 0 && , \forall b \in B \\
ec_b^1 &= 0 && , \forall b \in B \\
|fa_l^t - (gl_l \Delta \delta V_{ij}^t - bl_l \Delta \theta_{ij}^t)| &\leq (1 - \omega_l) M && , \forall l \in L, t \in T \\
|fr_l^t + (bl_l \Delta \delta V_{ij}^t + gl_l \Delta \theta_{ij}^t)| &\leq (1 - \omega_l) M && , \forall l \in L, t \in T \\
|fa_l^t| &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
|fr_l^t| &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
(fa_l^t)^2 + (fr_l^t)^2 &\leq (f_l^{max})^2 \omega_l && , \forall l \in L, t \in T \\
fa_l^t + \frac{1}{2} qa_l^t &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
-fa_l^t + \frac{1}{2} qa_l^t &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
fr_l^t + \frac{1}{2} qr_l^t &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
-fr_l^t + \frac{1}{2} qr_l^t &\leq f_l^{max} \omega_l && , \forall l \in L, t \in T \\
\left(fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(fr_l^t + \frac{1}{2} qr_l^t \right)^2 &\leq (f_l^{max})^2 \omega_l && , \forall l \in L, t \in T \\
\left(fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(-fr_l^t + \frac{1}{2} qr_l^t \right)^2 &\leq (f_l^{max})^2 \omega_l && , \forall l \in L, t \in T \\
\left(-fa_l^t + \frac{1}{2} qa_l^t \right)^2 + \left(fr_l^t + \frac{1}{2} qr_l^t \right)^2 &\leq (f_l^{max})^2 \omega_l && , \forall l \in L, t \in T
\end{aligned}$$

$$\left(-fa_l^t + \frac{1}{2}qa_l^t\right)^2 + \left(-fr_l^t + \frac{1}{2}qr_l^t\right)^2 \leq (f_l^{max})^2\omega_l, \forall l \in L, t \in T$$

$$cr_c = u_c\omega_c, \forall c \in C$$

$$|\delta V_k^t| \leq 0.5, \forall k \in N, t \in T$$

$$|\theta_k^t| \leq \pi, \forall k \in N, t \in T$$

$$epc_b^t \leq 0$$

$$epd_b^t, epda_b^t, epdr_b^t, rri_k^t, rrc_k^t \geq 0$$

$$\omega_p, \omega_b, \omega_l, \omega_c \geq 0: \text{integer}$$

References

- [1] IRENA, “Global energy transformation: A roadmap to 2050 (2019 edition),” 2019.
- [2] IRENA, “Renewable Power Generation Costs in 2018,” 2019.
- [3] IEA, “World Energy Investment 2019,” 2019.
- [4] IRENA, “The Power to Change: Solar and Wind Cost Reduction Potential to 2025,” 2016.
- [5] S. Sen and S. Ganguly, “Opportunities, barriers and issues with renewable energy development – A discussion,” *Renew. Sustain. Energy Rev.*, vol. 69, pp. 1170–1181, Mar. 2017.
- [6] S. Tabrizian, “Technological innovation to achieve sustainable development—Renewable energy technologies diffusion in developing countries,” *Sustain. Dev.*, vol. 27, no. 3, pp. 537–544, May 2019.
- [7] M. R. R. Tabar *et al.*, “Kolmogorov spectrum of renewable wind and solar power fluctuations,” *Eur. Phys. J. Spec. Top.*, vol. 223, no. 12, pp. 2637–2644, Oct. 2014.
- [8] J. Mossoba *et al.*, “Analysis of solar irradiance intermittency mitigation using constant DC voltage PV and EV battery storage,” in *2012 IEEE Transportation Electrification Conference and Expo (ITEC)*, 2012, pp. 1–6.
- [9] M. Anvari *et al.*, “Short term fluctuations of wind and solar power systems,” *New J. Phys.*, vol. 18, no. 6, p. 063027, Jun. 2016.
- [10] J. Traube *et al.*, “Mitigation of Solar Irradiance Intermittency in Photovoltaic Power Systems With Integrated Electric-Vehicle Charging Functionality,” *IEEE Trans. Power Electron.*, vol. 28, no. 6, pp. 3058–3067, Jun. 2013.
- [11] D. K. Critz, S. Busche, and S. Connors, “Power systems balancing with high penetration renewables: The potential of demand response in Hawaii,” *Energy Convers. Manag.*, vol. 76, pp. 609–619, Dec. 2013.
- [12] COES, “Procedimiento Técnico COES PR-01 ‘Programación de la Operación de Corto Plazo.’” Lima, 2014.
- [13] COES, “Procedimiento Técnico COES PR-37 ‘Programación de mediano plazo de la operación del SEIN.’” Lima, 2016.
- [14] F. Wolak, “Long-Term Resource Adequacy in Wholesale Electricity Markets with Significant Intermittent Renewables,” Cambridge, MA, Jul. 2021.
- [15] J. Dyson, H. Mackenzie, N. A. Engerer, and J. Luffman Solcast Canberra, “Utility scale solar short term generation forecasting for improved dispatch and system

security,” in *16th Wind Integration Forum*, 2017.

- [16] J. Dong *et al.*, “Novel stochastic methods to predict short-term solar radiation and photovoltaic power,” *Renew. Energy*, vol. 145, pp. 333–346, Jan. 2020.
- [17] N. Kumar, A. Singh, N. Rai, and N. Chauhan, “Investigation on Short-Term Wind Power Forecasting Using ANN and ANN-PSO,” 2019, pp. 1103–1116.
- [18] A. Heydari, D. Astiaso Garcia, F. Keynia, F. Bisegna, and L. De Santoli, “A novel composite neural network based method for wind and solar power forecasting in microgrids,” *Appl. Energy*, vol. 251, p. 113353, Oct. 2019.
- [19] M. Paulescu and E. Paulescu, “Short-term forecasting of solar irradiance,” *Renew. Energy*, vol. 143, pp. 985–994, Dec. 2019.
- [20] Ü. B. Filik and T. Filik, “Wind Speed Prediction Using Artificial Neural Networks Based on Multiple Local Measurements in Eskisehir,” *Energy Procedia*, vol. 107, pp. 264–269, Feb. 2017.
- [21] M. Caldas and R. Alonso-Suárez, “Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements,” *Renew. Energy*, vol. 143, pp. 1643–1658, Dec. 2019.
- [22] C. Brancucci Martinez-Anido *et al.*, “The value of day-ahead solar power forecasting improvement,” *Sol. Energy*, vol. 129, pp. 192–203, May 2016.
- [23] M. K. Behera, I. Majumder, and N. Nayak, “Solar photovoltaic power forecasting using optimized modified extreme learning machine technique,” *Eng. Sci. Technol. an Int. J.*, vol. 21, no. 3, pp. 428–438, Jun. 2018.
- [24] C. Fu, G.-Q. Li, K.-P. Lin, and H.-J. Zhang, “Short-Term Wind Power Prediction Based on Improved Chicken Algorithm Optimization Support Vector Machine,” *Sustainability*, vol. 11, no. 2, p. 512, Jan. 2019.
- [25] T. Ouyang, X. Zha, L. Qin, Y. He, and Z. Tang, “Prediction of wind power ramp events based on residual correction,” *Renew. Energy*, vol. 136, pp. 781–792, Jun. 2019.
- [26] A. Fentis, L. Bahatti, M. Tabaa, and M. Mestari, “Short-term nonlinear autoregressive photovoltaic power forecasting using statistical learning approaches and in-situ observations,” *Int. J. Energy Environ. Eng.*, vol. 10, no. 2, pp. 189–206, Jun. 2019.
- [27] V. Reddy, S. M. Verma, K. Verma, and R. Kumar, “Hybrid Approach for Short Term Wind Power Forecasting,” in *2018 9th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2018, pp. 1–5.
- [28] S. Tsuzuki and Y. Nishiyama, “Short-Term Wind-Speed Forecasting Using Kernel Spectral Hidden Markov Models,” Nov. 2018.
- [29] H. C. Bylling, S. Pineda, and T. K. Boomsma, “The impact of short-term variability and uncertainty on long-term power planning,” *Ann. Oper. Res.*, vol. 284, no. 1, pp. 199–223, 2020.
- [30] Y. Feng, “Scenario generation and reduction for long-term and short-term power system generation planning under uncertainties,” Iowa State University, Ames,

2014.

- [31] I. Batas Bjelić and N. Rajaković, “Simulation-based optimization of sustainable national energy systems,” *Energy*, vol. 91, pp. 1087–1098, 2015.
- [32] V. Krakowski, E. Assoumou, V. Mazauric, and N. Maïzi, “Feasible path toward 40-100% renewable energy shares for power supply in France by 2050: A prospective analysis,” *Appl. Energy*, vol. 171, pp. 501–522, 2016.
- [33] B. Zakeri, S. Rinne, and S. Syri, “Wind integration into energy systems with a high share of nuclear power - what are the compromises?,” *Energies*, vol. 8, no. 4, pp. 2493–2527, 2015.
- [34] J. P. Deane, F. Gracceva, A. Chiodi, M. Gargiulo, and B. P. Ó. Gallachóir, “Assessing power system security. A framework and a multi model approach,” *Int. J. Electr. Power Energy Syst.*, vol. 73, pp. 283–297, 2015.
- [35] E. J. Falcón-Roque, F. Marcos Martín, C. Pascual Castaño, L. C. Domínguez-Dafauce, and F. J. Bastante Flores, “Energy planning model with renewable energy using optimization multicriteria techniques for isolated rural communities: Cajamarca province, Peru,” *J. Renew. Sustain. Energy*, vol. 9, no. 6, 2017.
- [36] R. Siddaiah and R. P. Saini, “A review on planning, configurations, modeling and optimization techniques of hybrid renewable energy systems for off grid applications,” *Renew. Sustain. Energy Rev.*, vol. 58, pp. 376–396, 2016.
- [37] F. Ugranli and E. Karatepe, “Transmission expansion planning considering maximizing penetration level of renewable sources,” *2013 IEEE Int. Symp. Innov. Intell. Syst. Appl. IEEE INISTA 2013*, pp. 4–8, 2013.
- [38] L. Tang, F. Wen, M. A. Salam, and S. P. Ang, “Transmission system planning considering integration of renewable energy resources,” *IEEE PES Asia-Pacific Power Energy Eng. Conf.*, vol. 3, pp. 5–9, 2015.
- [39] J. B. Nunes, N. Mahmoudi, T. K. Saha, and D. Chattopadhyay, “Stochastic generation and transmission planning considering future renewable and gas opportunities in Queensland, Australia,” *IEEE Power Energy Soc. Gen. Meet.*, vol. 2016-Novem, 2016.
- [40] J. Zheng, F. Wen, M. Zhou, L. Hu, Q. Xu, and Z. Lan, “Transmission planning with renewable generation and energy storage,” *IET Semin. Dig.*, vol. 2015, 2015.
- [41] C. A. G. MacRae, M. Ozlen, and A. T. Ernst, “Transmission expansion planning considering energy storage,” pp. 239–244, Dec. 2014.
- [42] R. A. de Moraes, T. S. P. Fernandes, A. G. B. Arantes, and C. Unsihuay-Vila, “Short-term scheduling of integrated power and spinning reserve of a wind-hydrothermal generation system with AC network security constraints,” *J. Control. Autom. Electr. Syst.*, 2017.
- [43] M. Colorado and M. F. Bedrinana, “Improved planning must-run units considering system reliability and new renewable energy sources,” in *2017 IEEE Manchester PowerTech*, 2017, pp. 1–6.
- [44] COES, “Informe de la Operación Anual del SEIN 2019.” Lima, 2020.
- [45] OSINERGMIN, “Anuario Estadístico 2019.” Lima, 2020.

- [46] OSINERGMIN, “Informe Técnico N° DSE-33-2019 ‘Monitoreo de Sistemas Eléctricos de Transmisión en Alerta (SETA) 2018.’” Lima, 2019.
- [47] COES, “Actualización Plan de Transmisión 2023-2032 ‘Informe de Diagnóstico 2023-2032.’” Lima, 2021.
- [48] COES, “Consultar Información Estudios de Pre Operatividad.” [Online]. Available: <https://www.coes.org.pe/Portal/Planificacion/NuevosProyectos/Consultawebepo>. [Accessed: 29-Oct-2021].
- [49] OSINERGMIN, “Subastas de Energías Renovables en Perú.” [Online]. Available: <https://www.osinergmin.gob.pe/empresas/energias-renovables/subastas>. [Accessed: 30-Oct-2021].
- [50] COES, “Implicancias de Integrar Generación con Recursos Energéticos Renovables en el SEIN,” in *VI Simposio Internacional de Energía*, 2011.
- [51] IRENA, “Global Landscape of Renewable Energy Finance 2020,” Abu Dhabi, 2020.
- [52] A. Balouktsis and P. Tsalides, “Stochastic simulation model of hourly total solar radiation,” *Sol. Energy*, vol. 37, no. 2, pp. 119–126, 1986.
- [53] V. A. Graham, K. G. T. Hollands, and T. E. Unny, “A time series model for Kt with application to global synthetic weather generation,” *Sol. Energy*, 1988.
- [54] V. A. Graham and K. G. T. Hollands, “A method to generate synthetic hourly solar radiation globally,” *Sol. Energy*, 1990.
- [55] L. Hontoria, J. Aguilera, and P. Zufiria, “Generation of hourly irradiation synthetic series using the neural network multilayer perceptron,” *Sol. Energy*, 2002.
- [56] A. Mellit, “Artificial Intelligence technique for modelling and forecasting of solar radiation data: a review,” *Int. J. Artif. Intell. Soft Comput.*, 2008.
- [57] A. N. Celik, “Long-term energy output estimation for photovoltaic energy systems using synthetic solar irradiation data,” *Energy*, 2003.
- [58] J. Polo, L. F. Zarzalejo, R. Marchante, and A. A. Navarro, “A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution,” *Sol. Energy*, 2011.
- [59] D. Laslett, C. Creagh, and P. Jennings, “A method for generating synthetic hourly solar radiation data for any location in the south west of Western Australia, in a world wide web page,” *Renew. Energy*, 2014.
- [60] B. O. Ngoko, H. Sugihara, and T. Funaki, “Synthetic generation of high temporal resolution solar radiation data using Markov models,” *Sol. Energy*, vol. 103, pp. 160–170, 2014.
- [61] A. P. Grantham, P. J. Pudney, L. A. Ward, M. Belusko, and J. W. Boland, “Generating synthetic five-minute solar irradiance values from hourly observations,” *Sol. Energy*, vol. 147, pp. 209–221, 2017.
- [62] A. P. Grantham, P. J. Pudney, and J. W. Boland, “Generating synthetic sequences of global horizontal irradiation,” *Sol. Energy*, vol. 162, no. November 2017, pp.

500–509, 2018.

- [63] M. Larrañeta, C. Fernandez-Peruchena, M. A. Silva-Pérez, and I. Lillo-Bravo, “Methodology to synthetically downscale DNI time series from 1-h to 1-min temporal resolution with geographic flexibility,” *Sol. Energy*, vol. 162, no. October 2017, pp. 573–584, 2018.
- [64] F. C. Kaminsky, R. H. Kirchhoff, C. Y. Syu, and J. F. Manwell, “A Comparison of Alternative Approaches for the Synthetic Generation of a Wind Speed Time Series,” *J. Sol. Energy Eng.*, vol. 113, no. 4, p. 280, 1991.
- [65] A. D. Sahin and Z. Sen, “First-order Markov chain approach to wind speed modelling,” *J. Wind Eng. Ind. Aerodyn.*, 2001.
- [66] A. Shamshad, M. A. Bawadi, W. M. A. Wan Hussin, T. A. Majid, and S. A. M. Sanusi, “First and second order Markov chain models for synthetic generation of wind speed time series,” *Energy*, 2005.
- [67] F. O. Hocaoglu, Ö. N. Gerek, and M. Kurban, “The Effect of Markov Chain State Size for Synthetic Wind Speed Generation,” *Probabilistic Methods Appl. to Power Syst. 2008. PMAPS'08. Proc. 10th Int. Conf.*, 2008.
- [68] K. Brokish and J. Kirtley, “Pitfalls of modeling wind power using Markov chains,” in *2009 IEEE/PES Power Systems Conference and Exposition, PSCE 2009*, 2009.
- [69] N. B. Negra, O. Holmstrøm, B. Bak-Jensen, and P. Sørensen, “Model of a synthetic wind speed time series generator,” *Wind Energy*, 2008.
- [70] S. Hagspiel, A. Papaemmannouil, M. Schmid, and G. Andersson, “Copula-based modeling of stochastic wind power in Europe and implications for the Swiss power grid,” *Appl. Energy*, 2012.
- [71] C. Sarmiento, C. Valencia, and R. Akhavan-Tabatabaei, “Copula autoregressive methodology for the simulation of wind speed and direction time series,” *J. Wind Eng. Ind. Aerodyn.*, 2018.
- [72] R. Carapellucci and L. Giordano, “A methodology for the synthetic generation of hourly wind speed time series based on some known aggregate input data,” *Appl. Energy*, 2013.
- [73] A. Naimo, “A Novel Approach to Generate Synthetic Wind Data,” *Procedia - Soc. Behav. Sci.*, 2014.
- [74] J. Chen and C. Rabiti, “Synthetic wind speed scenarios generation for probabilistic analysis of hybrid energy systems,” *Energy*, 2017.
- [75] O. M. Babatunde, J. L. Munda, and Y. Hamam, “A comprehensive state-of-the-art survey on power generation expansion planning with intermittent renewable energy source and energy storage,” *Int. J. Energy Res.*, vol. 43, no. 12, pp. 6078–6107, Oct. 2019.
- [76] N. E. Koltsaklis and A. S. Dagoumas, “State-of-the-art generation expansion planning: A review,” *Appl. Energy*, vol. 230, pp. 563–589, Nov. 2018.
- [77] N. Gideon Ude, H. Yskandar, and R. Coneth Graham, “A Comprehensive State-of-the-Art Survey on the Transmission Network Expansion Planning Optimization Algorithms,” *IEEE Access*, vol. 7, pp. 123158–123181, 2019.

- [78] P. V. Gomes and J. T. Saraiva, "State-of-the-art of transmission expansion planning: A survey from restructuring to renewable and distributed electricity markets," *Int. J. Electr. Power Energy Syst.*, vol. 111, pp. 411–424, Oct. 2019.
- [79] C. S. Lai, G. Locatelli, A. Pimm, X. Wu, and L. L. Lai, "A review on long-term electrical power system modeling with energy storage," *J. Clean. Prod.*, vol. 280, p. 124298, Jan. 2021.
- [80] B. Maluenda, M. Negrete-Pincetic, D. E. Olivares, and Á. Lorca, "Expansion planning under uncertainty for hydrothermal systems with variable resources," *Int. J. Electr. Power Energy Syst.*, vol. 103, pp. 644–651, Dec. 2018.
- [81] L. Baringo and A. J. Conejo, "Correlated wind-power production and electric load scenarios for investment decisions," *Appl. Energy*, vol. 101, pp. 475–482, Jan. 2013.
- [82] W. S. Morais, R. C. Perez, and A. Soares, "Aplicación de un Modelo de Expansión de la Generación considerando el Requerimiento de Reserva Probabilística Dinámica en el Sistema Eléctrico Mexicano," in *XVIII ERIAC*, 2019, pp. 1–9.
- [83] R. C. Perez, D. Bayma, S. Binato, I. Carvalho, L. Okamura, and W. S. Morais, "Determinación del Plan Óptimo de Expansión de Cada País y Evaluación de Interconexión Energética Regional contemplando Alta Penetración de Fuentes Renovables," in *XVIII ERIAC*, 2019.
- [84] R. Felix and C. Unsihuay-Vila, "A Model to optimize Mix Power Generation Selection of Distributed Renewable Plants for Expansion Planning with Reliability Criteria: An Application in Puno, Peru," in *2018 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D-LA)*, 2018, pp. 1–5.
- [85] S. Pineda and J. M. Morales, "Chronological Time-Period Clustering for Optimal Capacity Expansion Planning With Storage," *IEEE Trans. Power Syst.*, vol. 33, no. 6, pp. 7162–7170, Nov. 2018.
- [86] L. Baringo and A. J. Conejo, "Wind power investment within a market environment," *Appl. Energy*, vol. 88, no. 9, pp. 3239–3247, Sep. 2011.
- [87] S. Lumbreras, A. Ramos, and F. Banez-Chicharro, "Optimal transmission network expansion planning in real-sized power systems with high renewable penetration," *Electr. Power Syst. Res.*, vol. 149, pp. 76–88, Aug. 2017.
- [88] T. Warren Liao, "Clustering of time series data—a survey," *Pattern Recognit.*, vol. 38, no. 11, pp. 1857–1874, Nov. 2005.
- [89] M. S. ElNozahy, M. M. A. Salama, and R. Seethapathy, "A probabilistic load modelling approach using clustering algorithms," in *2013 IEEE Power & Energy Society General Meeting*, 2013, pp. 1–5.
- [90] G. Latorre, R. D. Cruz, J. M. Areiza, and A. Villegas, "Classification of publications and models on transmission expansion planning," *IEEE Trans. Power Syst.*, vol. 18, no. 2, pp. 938–946, 2003.
- [91] R. Romero, A. J. Monticelli, A. Garcia, and S. Haffner, "Test systems and mathematical models for transmission network expansion planning," *IEE Proc. - Gener. Transm. Distrib.*, vol. 149, no. 1, p. 27, 2002.

- [92] D. Min, J. Ryu, and D. G. Choi, "A long-term capacity expansion planning model for an electric power system integrating large-size renewable energy technologies," *Comput. Oper. Res.*, vol. 96, pp. 244–255, Aug. 2018.
- [93] D. G. Cervan Prado, "Modelo multietapa para planificar la expansión óptima de la generación y transmisión eléctrica a largo plazo en el Perú del 2017 al 2040," Universidad Nacional de Ingeniería, 2018.
- [94] T. Akbari and M. Tavakoli Bina, "A linearized formulation of AC multi-year transmission expansion planning: A mixed-integer linear programming approach," *Electr. Power Syst. Res.*, 2014.
- [95] T. Akbari and M. T. Bina, "Approximated MILP model for AC transmission expansion planning: Global solutions versus local solutions," *IET Gener. Transm. Distrib.*, 2016.
- [96] A. Arabpour, M. R. Besmi, and P. Maghouli, "Transmission Expansion Planning with Linearized AC Load Flow by Special Ordered Set Method," *J. Energy Eng.*, 2018.
- [97] A. Arabpour, M. R. Besmi, and P. Maghouli, "Transmission Expansion and Reactive Power Planning Considering Wind Energy Investment Using A Linearized AC Model," *J. Electr. Eng. Technol.*, vol. 14, no. 3, pp. 1035–1043, May 2019.
- [98] H. Zhang, G. T. Heydt, V. Vittal, and H. D. Mittelmann, "Transmission expansion planning using an AC model: Formulations and possible relaxations," *IEEE Power Energy Soc. Gen. Meet.*, pp. 1–8, 2012.
- [99] H. Zhang, G. T. Heydt, V. Vittal, and J. Quintero, "An improved network model for transmission expansion planning considering reactive power and network losses," *Power Syst. IEEE Trans.*, vol. 28, no. 3, pp. 3471–3479, 2013.
- [100] M. Rahmani and M. Rashidinejad, "Integrated AC Transmission Network Expansion and reactive Power Planning," *Iran. J. Sci. Technol. Trans. Electr. Eng.*, vol. 35, pp. 127–140, 2011.
- [101] R. Felix and J. Luyo, "Multistage Transmission Expansion Planning considering Redesign, Repowering and Technical Losses: An Application in Lima, Peru," in *2018 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D-LA)*, 2018, pp. 1–5.
- [102] MEM Peru, "Peruvian Solar Energy Atlas," 2003. [Online]. Available: <http://dger.minem.gob.pe/atlassolar/>.
- [103] NASA, "NASA POWER - Prediction Of Worldwide Energy Resources," 2018. [Online]. Available: <https://power.larc.nasa.gov/>.
- [104] JRC, "JRC Photovoltaic Geographical Information System (PVGIS) - European Commission," 2018. [Online]. Available: http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html. [Accessed: 09-Oct-2018].
- [105] N. Technologies, "Ntrand - Excel Random Number Generator based on Mersenne Twister," 2018. [Online]. Available: <http://www.ntrand.com/>.
- [106] J. A. Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes: Fourth*

Edition. 2013.

- [107] K. N. Shukla, S. Rangnekar, and K. Sudhakar, “Comparative study of isotropic and anisotropic sky models to estimate solar radiation incident on tilted surface: A case study for Bhopal, India,” *Energy Reports*, 2015.
- [108] J. I. Prieto, J. C. Martínez-García, and D. García, “Correlation between global solar irradiation and air temperature in Asturias, Spain,” *Sol. Energy*, vol. 83, no. 7, pp. 1076–1085, 2009.
- [109] HOMER, “HOMER Pro 3.12 User Manual,” 2018. [Online]. Available: <https://www.homerenergy.com/products/pro/docs/3.12/index.html>. [Accessed: 12-Oct-2018].
- [110] B. Shiva Kumar and K. Sudhakar, “Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India,” *Energy Reports*, vol. 1, pp. 184–192, Nov. 2015.
- [111] M. Gostein, J. R. Caron, and B. Littmann, “Measuring soiling losses at utility-scale PV power plants,” in *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*, 2014, pp. 0885–0890.
- [112] J. Zorrilla-Casanova *et al.*, “Analysis of dust losses in photovoltaic modules,” 2011.
- [113] L. Micheli, M. G. Deceglie, and M. Muller, “Mapping Photovoltaic Soiling Using Spatial Interpolation Techniques,” *IEEE J. Photovoltaics*, pp. 1–6, 2018.
- [114] M. G. Deceglie, L. Micheli, and M. Muller, “Quantifying Soiling Loss Directly From PV Yield,” *IEEE J. Photovoltaics*, vol. 8, no. 2, pp. 547–551, Mar. 2018.
- [115] J. R. Caron and B. Littmann, “Direct monitoring of energy lost due to soiling on first solar modules in California,” in *2012 IEEE 38th Photovoltaic Specialists Conference (PVSC) PART 2*, 2012, pp. 1–5.
- [116] IFC, “Utility-Scale Solar Photovoltaic Power Plants: A project Developer’s Guide,” Washington, D.C., 2015.
- [117] A. P. Dobos, “PVWatts Version 5 Manual,” 2014.
- [118] NREL, “RE Explorer - Mapping our energy future,” 2019. [Online]. Available: <https://www.re-explorer.org/>.
- [119] C. G. Justus, W. R. Hargraves, A. Mikhail, and D. Graber, “Methods for Estimating Wind Speed Frequency Distributions,” *J. Appl. Meteorol.*, vol. 17, no. 3, pp. 350–353, Mar. 1978.
- [120] K. S. P. Kumar and S. Gaddada, “Statistical scrutiny of Weibull parameters for wind energy potential appraisal in the area of northern Ethiopia,” *Renewables Wind. Water, Sol.*, vol. 2, no. 1, p. 14, Dec. 2015.
- [121] N. Aghbalou, A. Charki, S. R. Elazzouzi, and K. Reklouei, “A probabilistic assessment approach for wind turbine-site matching,” *Int. J. Electr. Power Energy Syst.*, vol. 103, pp. 497–510, Dec. 2018.
- [122] R Core Team, “R: A Language and Environment for Statistical Computing.” R Foundation for Statistical Computing, Vienna, Austria, 2017.

- [123] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S*, Fourth. New York: Springer, 2002.
- [124] A. Colmenar-Santos, S. Campíez-Romero, L. Enríquez-García, and C. Pérez-Molina, “Simplified Analysis of the Electric Power Losses for On-Shore Wind Farms Considering Weibull Distribution Parameters,” *Energies*, vol. 7, no. 11, pp. 6856–6885, Oct. 2014.
- [125] IEC, “IEC 61400-12-1:2017 Wind energy generation systems – Power performance measurements of electricity producing wind turbines,” 2017.
- [126] J. F. Manwell, J. G. McGowan, and A. L. Rogers, *Wind Energy Explained: Theory, Design and Application, Second Edition*. 2010.
- [127] D. A. Spera and T. R. Richards, “Modified Power Law Equations for Vertical Wind Profiles,” *Wind Charact. Wind Energy Siting Conf.*, 1979.
- [128] ME Chile, “Chilean Wind Energy Explorer,” 2012. [Online]. Available: <http://walker.dgf.uchile.cl/Explorador/Eolico2/>.
- [129] M. Lydia, S. S. Kumar, A. I. Selvakumar, and G. E. Prem Kumar, “A comprehensive review on wind turbine power curve modeling techniques,” *Renew. Sustain. Energy Rev.*, vol. 30, pp. 452–460, Feb. 2014.
- [130] E. Diaz-Dorado, C. Carrillo, J. Cidras, and E. Albo, “Estimation of energy losses in a Wind Park,” in *2007 9th International Conference on Electrical Power Quality and Utilisation*, 2007, pp. 1–6.
- [131] G. S. Böhme, E. A. Fadigas, A. L. V. Gimenes, and C. E. M. Tassinari, “Wake effect measurement in complex terrain - A case study in Brazilian wind farms,” *Energy*, vol. 161, pp. 277–283, Oct. 2018.
- [132] N. Inaba, R. Takahashi, J. Tamura, M. Kimura, A. Komura, and K. Takeda, “A consideration on loss characteristics and annual capacity factor of offshore wind farm,” in *2012 XXth International Conference on Electrical Machines*, 2012, pp. 2022–2027.
- [133] P. K. Steimer and O. Apeldoorn, “Medium voltage power conversion technology for efficient windpark power collection grids,” in *The 2nd International Symposium on Power Electronics for Distributed Generation Systems*, 2010, pp. 12–18.
- [134] A. Madariaga, C. J. Martínez de Ilarduya, S. Ceballos, I. Martínez de Alegria, and J. L. Martín, “Electrical losses in multi-MW wind energy conversion systems,” *Renew. Energy Power Qual. J.*, pp. 322–327, Apr. 2012.
- [135] EWEA, “Wind Farm Energy Loss Factors,” Brussels, Belgium, 2019.
- [136] ME Chile, “Chilean Solar Explorer,” 2017. [Online]. Available: <http://ernc.dgf.uchile.cl:48080/exploracion>. [Accessed: 24-Oct-2018].
- [137] ME Chile, “Modelo de Radiación Solar.” p. 19, 2017.
- [138] WMO, “World Radiation Data Centre,” 2018. [Online]. Available: http://wrdc.mgo.rssi.ru/wrdc_en.htm.
- [139] COES, “Technical Specification of Power Plants,” 2018. [Online]. Available:

- <http://sicoes.coes.org.pe/AppPublico/FichaTecnica/FichaTecnica/>. [Accessed: 26-Nov-2018].
- [140] CivicSolar, “Datasheet of Solar Panels,” 2018. [Online]. Available: <https://www.civicsolar.com/>. [Accessed: 26-Nov-2018].
- [141] Siemens, “Siemens Wind Turbine SWT-2.3-108,” Erlangen, Germany, 2011.
- [142] ME Chile, “Wind and Solar resource measurement campaign,” 2019. [Online]. Available: <http://walker.dgf.uchile.cl/Mediciones/>.
- [143] MEM Peru, “Peruvian Wind Atlas,” 2016. [Online]. Available: http://mapas.minem.gob.pe/map_eolico/.
- [144] D. Z. Fitiwi, F. de Cuadra, L. Olmos, and M. Rivier, “A new approach of clustering operational states for power network expansion planning problems dealing with RES (renewable energy source) generation operational variability and uncertainty,” *Energy*, vol. 90, pp. 1360–1376, Oct. 2015.
- [145] J. Tavenard, V. Romain, and G. Faouzi, “tslearn: A machine learning toolkit dedicated to time-series data.” 2017.
- [146] H. Sakoe and S. Chiba, “Dynamic Programming Algorithm Optimization for Spoken Word Recognition,” *IEEE Trans. Acoust.*, 1978.
- [147] M. Cuturi and M. Blondel, “Soft-DTW: A differentiable loss function for time-series,” in *34th International Conference on Machine Learning, ICML 2017*, 2017.
- [148] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning*. 2017.
- [149] R. Fourer, D. M. Gay, and B. W. Kernighan, *AMPL: A modeling language for mathematical programming*, Second Edi. Duxbury Thomson, 2003.
- [150] L. Gurobi Optimization, “Gurobi Optimizer Reference Manual.” 2020.
- [151] R. Felix, “Modelo multietapa para el planeamiento óptimo de la expansión de redes de subtransmisión considerando reconfiguración y pérdidas técnicas: Caso Lima sur,” Universidad Nacional de Ingeniería, 2018.
- [152] COES, “Procedimiento Técnico COES PR-15 ‘Valorización de Transferencias de Energía Reactiva.’” Lima, 2017.
- [153] C. H. Antunes, A. G. Martins, and I. S. Brito, “A multiple objective mixed integer linear programming model for power generation expansion planning,” *Energy*, vol. 29, no. 4, pp. 613–627, Mar. 2004.
- [154] I. Alsaidan, A. Khodaei, and W. Gao, “Determination of battery energy storage technology and size for standalone microgrids,” in *IEEE Power and Energy Society General Meeting*, 2016.
- [155] N. Alguacil, A. L. Motto, and A. J. Conejo, “Transmission expansion planning: A mixed-integer LP approach,” *IEEE Trans. Power Syst.*, vol. 18, no. 3, pp. 1070–1077, 2003.
- [156] G. Vinasco, M. J. Rider Flores, and R. Romero, “A strategy to solve the multistage transmission expansion planning problem,” *IEEE Trans. Power Syst.*, vol. 26, no.

- 4, pp. 2574–2576, 2011.
- [157] L. H. Macedo, C. V. Montes, J. F. Franco, M. J. Rider, and R. Romero, “MILP branch flow model for concurrent AC multistage transmission expansion and reactive power planning with security constraints,” *IET Gener. Transm. Distrib.*, vol. 10, no. 12, pp. 3023–3032, Sep. 2016.
- [158] OSINERGMIN, “Resolución OSINERGMIN N° 217-2013-OS/CD ‘Norma tarifas y compensaciones para sistemas secundarios de transmisión y sistemas complementarios de transmisión.’” Lima, 2013.
- [159] G. E. Alvarez, “Operation of pumped storage hydropower plants through optimization for power systems,” *Energy*, vol. 202, p. 117797, Jul. 2020.
- [160] M. Sacasqui, “Comparison of Methodologies: Analytical Hierarchy Process and Grey Clustering with Entropy Weight for the Multicriteria Assessment of the Energy Sources of Perú,” in *2018 IEEE PES Transmission & Distribution Conference and Exhibition - Latin America (T&D-LA)*, 2018, pp. 1–5.
- [161] R. E. C. Gonçalves, E. C. Finardi, E. L. da Silva, and M. L. L. dos Santos, “Comparing stochastic optimization methods to solve the medium-term operation planning problem,” *Comput. Appl. Math.*, vol. 30, no. 2, pp. 289–313, 2011.
- [162] OSINERGMIN, “Informe Técnico N° DSE-CT-14-2017.” Lima, 2017.
- [163] OSINERGMIN, “Informe OSINERGMIN N° 086-2016-GART ‘Revisión y aprobación del Plan de Inversiones en Transmisión para el Área de Demanda 6. Regulación para el período 2017-2021 (Prepublicación).’” Lima, 2016.
- [164] MEM Peru, “Peruvian Hydro Projects Atlas,” 2016. [Online]. Available: http://mapas.minem.gob.pe/MAP_HIDROELECTRICO/.
- [165] D. G. Erbs, S. A. Klein, and J. A. Duffie, “Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation,” *Sol. Energy*, vol. 28, no. 4, pp. 293–302, 1982.