The effect of spatial granularity on optimal renewable energy portfolios in an integrated climate-energy assessment model

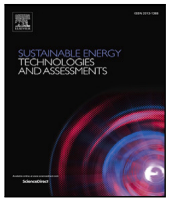
Aina Maimó-Far, Victor Homar, Alexis Tantet, Philippe Drobinski

Sustainable Energy Technologies and Assessments 54 (2022) 102827

Contents lists available at ScienceDirect

Sustainable Energy Technologies and Assessments

journal homepage: www.elsevier.com/locate/seta



The effect of spatial granularity on optimal renewable energy portfolios in an integrated climate-energy assessment model

Aina Maimó-Far [a,*], Victor Homar [a], Alexis Tantet [b], Philippe Drobinski [b]

[a] Meteorology Group, Physics Department, Universitat de les Illes Balears, Cra. de Valldemossa, km 7.5, Palma, 07122, Spain
[b] Laboratoire de Meteorologie Dynamique/Institut Pierre-Simon Laplace, Ecole Polytechnique, IP Paris, Sorbonne Universite, ENS, PSL University, CNRS, Rte de Saclay, Palaiseau, 91120, France

ARTICLE INFO

MSC:
46N99

Keywords:
Renewable energy system planning
High–resolution energy modeling
Pareto front
Climate energy assessment model
Mean–variance analysis

ABSTRACT

Renewable energy planning is key to achieving target levels of renewable-energy penetration and electricity demand in the European Union. Mean–variance analysis can be used to identify the optimal spatial and technological deployment of variable renewable energy (VRE) sources in terms of maximizing VRE penetration and minimizing supply risks. We investigate the extent to which optimizing capacities at the scale of climate-data grid points, instead of administrative regions (a common approach due to data availability and computation costs), helps generate more optimal renewable deployment scenarios. A finer description of climate resources, and thus the VRE capacity factors, results in a better exploitation of complementarities, partly due to the increased degrees of freedom in the optimization. A detailed analysis of the causes behind these improvements shows that better describing local conditions leads to two advantages over less granular counterparts: higher average capacity factors and generation combinations that offer lower covariances. This analysis also reveals that more granular approaches significantly reduce variability in daily and annual climate frequencies in renewable generation under the optimal scenario. These results provide evidence of the need to account for detailed climate information to accurately identify optimal renewable deployment scenarios and support stakeholders and policy makers when it comes to making sustainable commitments.

1. Introduction

The transition towards a society powered by low-carbon energy generation systems has already begun, and it still represents a challenge for the near future. The European Union (EU) has set general objectives and actions to be taken to meet its 2050 decarbonization goals [1], though individual countries have to establish their own specific, realistic measures.

In this study, we focus on the case of Spain. Following the EU directives, the Spanish government has established a National Plan for Adaptation to Climate Change (PNACC) 2021–2030 [2], which sets a path towards the 2030 objectives, a stepping stone for the more ambitious 2050 goals. An interesting aspect of this plan is its multidisciplinary approach to mitigating climate change, setting ambitious objectives in multiple varied fields. When referring to energy, the PNACC highlights the need for generating further knowledge, and particularly the urgent need for research in the field of renewable energy deployment planning. Systemic risks – as well as the need to resolve or, at least, mitigate them – are also pointed out in the plan.

An even more concrete planning approach was set in Spain in the National Integrated Energy and Climate Plan (PNIEC) 2021–2030 [3], which focuses the national energy strategy on the need to make the transition. It promotes the installation of new renewable means of energy generation and underlines the need to further understand the challenges posed by incorporating renewable energy sources into the electricity system.

The transition towards a highly renewable future is led by two main variable renewable energy (VRE) sources, which aim to cover about half of the Spanish energy demand due to their decreasing cost and improving technology: solar photovoltaic (PV) and wind energy [4,5]. In the particular case of Spain, recent years have seen an encouraging increase in the installed capacity (IC) of these sources. From the end of 2018 to the end of 2020, there was an increase in the IC of PV and wind IC of 156% and 17%, respectively [6,7]. For reference, the five previous years saw increases in the IC of PV and wind of only 1% [7,8].

The climate is the sole generator of PV and wind energy, but multiple unrelated factors usually go into the decision-making process when planning for these kinds of renewable installations. Some instances of atlases reflecting the generation potential of wind and PV have been published (for wind, see [9] for Spain and [10,11] for Europe, the

* Corresponding author.
E-mail address: aina.maimo@uib.cat (A. Maimó-Far).

https://doi.org/10.1016/j.seta.2022.102827
Received 31 May 2022; Received in revised form 29 September 2022; Accepted 8 October 2022
Available online 21 October 2022

Nomenclature

2020REF	2020 distribution of solar photovoltaic and wind installations
CF	Capacity Factor
GRID	Full climate grid modeling approach
IC	Installed Capacity
PV	Solar Photovoltaic
REE	Red Eléctrica de España
REG	Aggregated regions modeling approach
SES	Spanish Electricity System
VRE	Variable Renewable Energy

latter including a socio-technical approach; for PV see [12] for a global approach and [13,14] for countries in the EU). Together with capital and operation and management costs, the average capacity factors (CFs) that can be derived from these atlases allow one to compute levelized costs of energy for a given technology and location with some degree of uncertainty (e.g. [15, Ch. 25]). However, these resources lack a key element that should be considered in energy planning: the risk posed by the intermittency of the resource and the complementarity between the different components of the system. Since the main sources of renewable generation are intermittent by nature, taking advantage of the complementarity of the system is fundamental in order to meet the demand and minimize electricity supply risks [16–18].

In this context, identifying scenarios that satisfy certain system requirements becomes fundamental, and optimization methods are commonly applied. There is a wide variety of such methods that can be applied to the renewable energy problem, each with its strengths and weaknesses [19]. Some examples of commonly applied methods can be found in recent applications: [20] for multi-criteria decision-making, [21,22] for the swarm algorithms, [23] for game theory, [24] for the harmony search method, and [25] for the use of neural networks. In addition to the wide variety of methods, objectives and constraints also change significantly from one study to another, depending on their focus: [26] considers the profitability and visual impact of wind farms; [27] explores the smoothest possible output from hydro, wind and PV energy; [28] try to find the balance between economic costs and environmental sustainability; [29] minimizes operational costs and emissions pollution; [22] minimizes the total net annual cost while keeping a constraint on the loss of load supply probability; [30] tries to increase the exergy efficiency while reducing costs and CO_2 emissions; and [31] optimizes cost, grid electricity imports, and building electricity costs considering all relevant factors.

The goal with which the optimization systems are used can also change from one study to another. Most focus on the best way to deploy renewable sources of energy generation (see [32] for a recent example seeking a 100% renewable system and [33] for a recent application to the electrification of buildings in rural areas). However, optimization methods can be applied to investigate different scenarios. Some particular purposes to which optimization has recently been applied are: identifying the potential of adding wind power to cover energy deficiencies in different countries [34], assessing the integration of a new element to the energy system [35,36], typifying different batteries and their usefulness [37], exploring the viability of rooftop PV in terms of economics and grid-worthiness [31], and investigating how high levels of penetration of VRE sources affect the system integration costs and emissions. In turn, any use of optimization in the renewable energy sector may see its results only applicable to the near future only, as the impact of climate change and technological advances may create strong uncertainties in future forecasts [38].

Seeing the vast assortment of elements that can change from one optimization to another, it becomes clear that when planning renewable energy deployment strategies using quantitative models, configuration is key. Multiple factors – such as constraints, time span, optimization method, and region – affect the process and the end result [39]. In this article we apply a common approach to the optimization of VRE source deployment consisting of a simultaneous maximization of the mean and minimization of the variance through modifications to the spatial distribution of IC for each VRE source (see [40] for a wind only configuration and [41] for a combination of wind and concentrated solar power). This simple method allows us to find the lowest variance (as a proxy for supply risk) possible for a given level of penetration, and returns a set of optimal scenarios when multiple penetration levels are considered. More realistic setups, such as considering the cost optimization of the energy system [42], represent different approaches to bi-objective optimizations that can also benefit from the study of the effects of granularity carried out in this article.

When modeling the system, and before considering the use of optimization algorithms, spatial processing and the representation of generation and demand data is decisive. Many examples that make use of regional aggregation of electricity and climate data can be found in recent literature [14,43–48]. More specifically, this approach has been used in combination with portfolio theory to optimize the deployment of renewables between countries or large regions through various metrics. Instances of this include maximizing the return and minimizing the risk of the system [44], and maximizing energy security and minimizing emissions [49]. These optimal results are representative of very large aggregated areas and do not provide information about where the installations should be put within these areas. An intermediate approach consists of having the total regional IC fixed, for instance via the optimization, and then identifying the possible deployment locations in each region, taking into account climate resources and socioeconomic constraints, but without carrying out another full optimization [50].

Inevitably, the aggregation of different renewable energy generation sites and technologies has a smoothing effect on the generation curve, similar to the effect that spatial aggregation has on electric load [51]. A clear reflection of this is highlighted by [44] when mentioning that "the hourly capacity factor of wind power production seems to be much less volatile in larger countries", which implies that the generation series are substantially smoothed by aggregation. However, the optimization problem can be handled in unaggregated models of the system with grids, or even by directly considering existing generation farms as installation points [52,53]. When using higher spatial granularity, the dimension of the problem increases significantly. Nevertheless, the bi-objective optimization required for renewable energy planning can be solved using multiple approaches, e.g., with a Monte Carlo simulation [46,54, Ch. 5]. In the present study we actually solve a direct discretization of the problem using quadratic programming solvers. Despite the range of methodological possibilities, a grid-like approach is often unrealistic because of its high computational cost, and therefore, it is not common [55]. The need for high-resolution data on electricity generation and demand makes this task even more difficult, especially at very high resolutions, at which the topology of the network can become relevant. However, the resolution used is fundamental to the optimization results, especially if the aggregated regions under consideration are very heterogeneous [56,57].

High-resolution climate data more accurately reflects the actual resources that power renewables and thus more realistically models their characteristics [58]. Moreover, this approach provides local data for estimating CFs at an increasing number of locations at which capacity can be optimized. Additionally, a higher level of granularity to model the system allows us to more accurately account for relevant factors in the identification of optimal scenarios. Features such as the feasibility of installations, the distribution cost, and the geographical distribution of demand gain significance when a more detailed and realistic system is considered. This more precise representation of the system, combined with the optimization method, has enormous potential to improve the value and significance of the resulting deployment scenarios.

Little research exists on the impact of using different spatial aggregation approaches. Previous studies have focused on each form of aggregation but fail to identify the specific effects of this change in granularity on the optimal solutions [55]. More extensive research exists on general optimization models and specially on the methodologies, optimization functions, and constraints that have been considered, but their analyses still yield no clear answers to the challenges that granularity raises [39,59].

During the current decade, over 240 billion euros are set to be invested in the energy transition in Spain, more than 90 million of which will be spent on the deployment of renewable energy sources [3]. Considering the heterogeneity between, and especially within, the regions of Spain (Fig. 1), we designed our study to identify whether accounting for this heterogeneity could help find better deployment scenarios. Thus, the present study analyzes the impacts of spatial granularity in optimal renewable energy deployment portfolio theory. We focus on the improvements made in terms of higher penetration solutions and more realistic mixes found using a grid-based approach, which benefits from a higher resolution in the allocation of the renewable resources. The novelty of this research lies in our analysis of the differences between optimizations and our quantification of the improvements in the description of the renewable deployment scenarios.

After an overview of the datasets and methods used, the differences between regional and grid-based modeling of the system are presented. Then, Section 3 disentangles the origins of such differences, and finally, the main findings of our study are summarized in the conclusions section.

2. Data and methodology

2.1. Optimization

The challenge that comes with implementing renewable energy sources while guaranteeing supply to the network can be translated into an approximate simplified problem of a maximizing the renewable penetration and minimizing the risks of supply failure. If we consider the risk to be the square root of the variance – and therefore the variance the risk squared – it can be modeled as a bi-objective optimization problem, more specifically, a mean–variance optimization scheme. The result of a mean–variance optimization is a set of optimal solutions defining a so-called Pareto front. For a fixed mean (variance/risk squared), the minimum variance/risk squared (maximum mean) is represented by the Pareto front. In a mean–variance diagram, all possible configurations of the system are either suboptimal or lie on the Pareto front.

When applying these optimization techniques to renewable energy deployment, the vector of the ICs of PV and wind at each location is the decision variable. Each spatial location is associated with a specific CF for renewable energy generation, which quantifies the percentage of the potential maximum technical production that is actually generated over a period of time. The national demand is also considered, as the real objective of the penetration of renewable energy sources is to find the overall generation curve (combining information on CF and IC) that best suits the demand and its fluctuations, not an arbitrarily high level of generation.

In order to carry out the optimization, hourly penetration is defined as the ratio of national hourly VRE generation to national hourly demand, even though the adequacy between generation and demand is not constrained in the model since the reason behind the mean–variance analysis is to prevent having to model non-VRE producers that would contribute to meeting the demand. From this, the total penetration (hereafter simply *penetration*) is defined as the average of the hourly penetration over the study period. *Risk* is defined as the standard deviation of the hourly penetration time series, in which we account for the specificities of the predictable and unpredictable components of PV generation [60]. Two additional constraints are used: positive ICs, which are used in all experiments, and total available IC, which is limited to the actual 2020 Spanish renewable IC for PV and wind technologies (Fig. 1, see Section 2.2 for further detail and Appendix A for the full mathematical expressions that make up the optimization), which is only applied in the experiments labeled *constrained*. We assume no limit on the available capacity to install in the *unconstrained* experiments. Throughout the study, the optimizations are applied under two different modeling approaches: one using aggregated regions (REG) and another using the full climate grid (GRID). The optimizations consist of maximizing penetration and minimizing risk squared.

This setup is used to perform two experiments. By using the full climate grid, the GRID approach increases the granularity of the model with respect to REG, and thus the degrees of freedom of the optimization problem. This allows us to better represent actual local characteristics and more accurately account for the climate and its variability. Note that the only difference in the setup between both experiments is the basic spatial unit considered (region versus grid point), as the regions are assumed to be homogeneous and are therefore represented by the average of the grid points they contain. The degrees of freedom available for the optimization algorithm to identify optimal IC deployment solutions are very different in each experiment (Fig. 1).

The Pareto front naturally is best displayed on a penetration-risk diagram (consider the result in Fig. 2 as an illustrative example), which allows us to represent all possible projections of spatial IC distributions on the penetration-risk plane. Each possible IC distribution is represented by a single point in the diagram. The Pareto front is represented by a curve defined by the IC distributions with the minimum risk attainable at each penetration level. The actual 2020 Spanish IC distribution is also represented by a point. The region above and to the left of the Pareto front is not in the range of possible solutions for that system, so no spatial or technological IC combinations meet such high levels of penetration and low levels of risk for the given available climate resource.

The representation of the optimal solutions in the form of Pareto fronts shows that the optimal penetration-risk ratio (the ratio between the mean penetration and the standard deviation of the hourly penetration series) is linear at low levels of penetration, while the constrained experiment shows curvature at higher levels of penetration, as the limitation on the total available IC to install makes it impossible to reach higher levels of penetration without assigning capacity to locations that carry much higher levels of risk. In the unconstrained case, the penetration-risk ratio remains constant at all penetration levels (see [60–63] for examples of these Pareto-front behaviors). The unconstrained front reveals the maximum mean-risk ratio possible for a given climate resource and associated CFs, and is itself an intrinsic property of the modeled system since it does not depend on the availability of IC. In the unconstrained case, an increase in penetration is achieved by multiplying the ICs by the corresponding constant factor. In this case, and assuming no changes in installed technologies, the climate resource is the only limiting factor. As mentioned, no IC distribution can render lower levels of risk or higher levels of penetration than the distributions that define the unconstrained Pareto front, since these possibilities are limited by the distribution of CFs (in time, space and technology), which are determined by the combination of spatial–temporal climate conditions with specific technologies. The constrained solutions overlap with the unconstrained ones as long as the total IC constraint is not in effect (i.e. the assigned ICs in the optimization solution do not exceed the upper bounds set by 2020 values). Beyond this level of penetration, optimal solutions involve lower penetration-risk ratios, and thus, suboptimal penetration returns for the increased risk assumed.

Given the penetration and risk delivered by the 2020 distribution of PV and wind installations (hereafter *2020REF*), three interesting optimal scenarios emerge: high-penetration, low-risk and full repowering. The high-penetration scenario highlights the IC distribution that

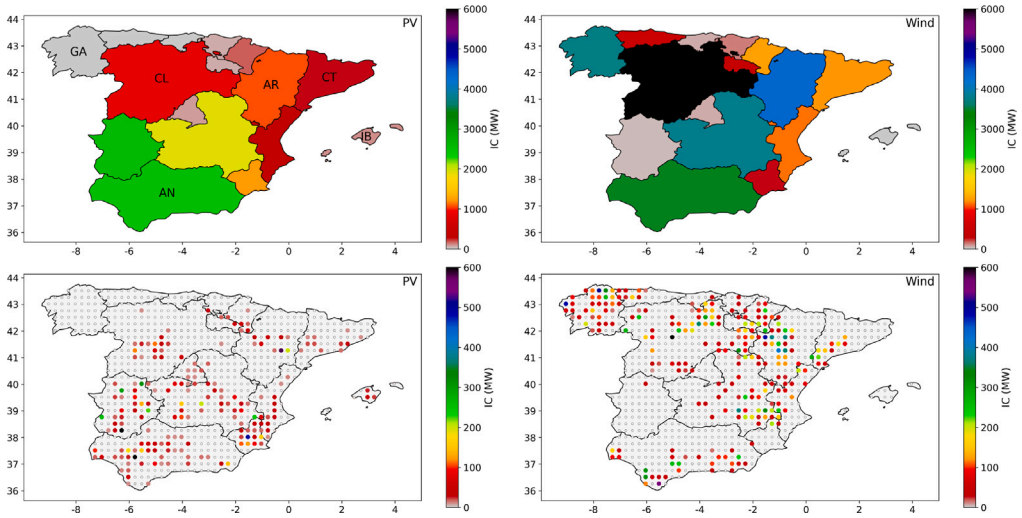


Fig. 1. Observed 2020 IC for PV (left) and wind (right) using aggregated regions (top) and the climate grid (bottom). Grid points with ICs under 10 MW are represented by empty circles and grid points with ICs over 10 MW show their ICs in reference to the color bar. ICs within a grid box are summed at the grid point that represents it. The top left panel shows the regions mentioned in the text: GA, Galicia; CL, Castilla *y* Leon; AR, Aragon; CT, Catalunya; IB. Illes Balears; AN, Andalucia.

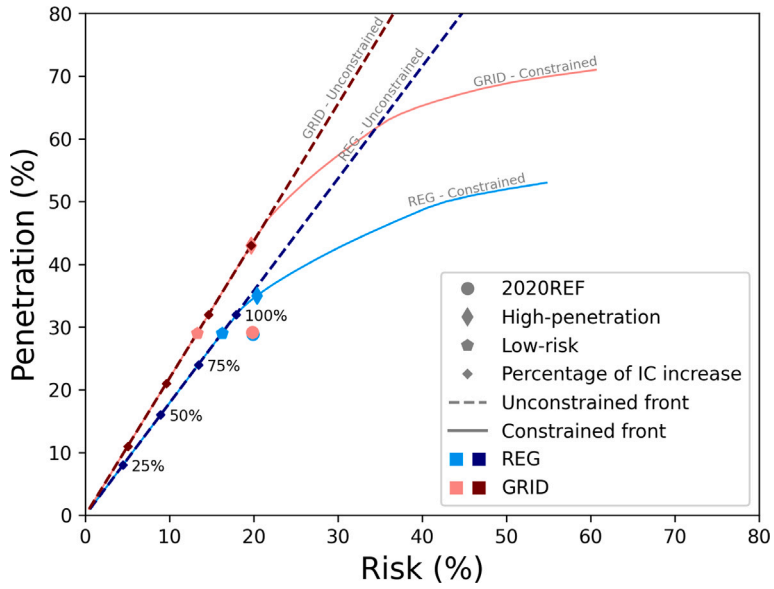


Fig. 2. Penetration-risk diagram, points of interest and optimal Pareto fronts of the unconstrained (dark, dashed, straight) and constrained (light, solid, curved) scenarios for REG (blue) and GRID (red). Circles represent 2020REF, diamonds show the high-penetration scenarios, pentagons indicate the low-risk scenarios, and small dark rotated squares represent every increment of 25% along the unconstrained front in the total IC from 25% to 100% of the total observed 2020 IC. The 100%-point marks the full repowering scenario.

would render maximum penetration at the 2020REF risk level. This scenario is found on the Pareto front, above the 2020REF point in the penetration-risk diagram (Fig. 2). Likewise, the low-risk scenario is defined as the IC distribution that carries the lowest possible risk at the 2020REF penetration level. This point is found on the Pareto front, to the left of the 2020REF in the diagram. On top of that, both the high-penetration and low-risk mixes for GRID and the low-risk mix for REG fall on the linear range of the front, so they are virtually met even without requiring all 2020REF IC. Indeed, the *full repowering* scenario (marked with 100% in Fig. 2) corresponds to the optimal (highest penetration-risk ratio) redistribution in Spain of the total 2020 IC.

The implementation of the portfolio theory, as well as the modeling of generation and demand is performed with the e4clim model [63]. This model generates CF time series which give information on the suitability of renewable generation, independent of the eventual IC, and it creates a climate-dependent series of demand to be covered. Hourly climatic time series and monthly electricity data are combined to generate calibrated hourly CFs with realistic values. These calibrated hourly CF series conserve climate variability but have reduced biases with respect to observations, and they are used as fixed parameters in the optimization problem. The demand series is generated following the process described in the updated version of e4clim in its application to Spain [60], and is explained in detail in the data section. Finally, we perform an optimization that maximizes the penetration while minimizing the risk squared. This work represents the first application of e4clim at the grid scale and more importantly, the first explicit analysis of the differences between using two different levels of granularity. Indeed, as a novelty, this setup allows for a direct comparison of the effects of using two spatial configurations for the same optimization problem.

2.2. Data

Climate resources were assessed using ERA5 reanalysis data [64], which has proven reliable for both solar radiation [65] and wind

power [66] estimations, despite some issues related to underestimating winds in mountainous areas [67] and potential improvements that could be made at the local level concerning PV estimates [68]. This climate dataset is particularly suitable for this study because it presents high temporal and spatial resolutions over a long time period. We use hourly climate data at surface levels from 1999 to 2020 at the native ERA5 resolution of 0.25 degrees (the bottom panels of Fig. 1 show the grid points over Spain). The input data for the e4clim model are the temperature at 2 m, wind components at 100 m, surface pressure, mean sea level pressure, mean surface downward short-wave radiation flux, and mean surface downward short-wave radiation flux with clear skies. Electricity information is obtained from two different Spanish national sources. On the one hand, IC data until the end of 2020 at the municipal level is available from the Spanish Ministry for the Ecological Transition and the Demographic Challenge from the time when renewable energy installations were first registered. These municipal ICs are assigned to the nearest model grid point so that each point aggregates IC from the municipalities around it. Then, regional monthly energy generation and demand information, as well as yearly regional IC data, were obtained from the *Red Eléctrica de España* (REE). Since CFs are calculated from a combination of Ministry and REE data, comparisons between monthly regional REE ICs and Ministry ICs are used to validate the outcomes. The resulting IC is therefore an accurate representation of the actual distribution of PV and wind installations over the model domain, either in the regional or the grid point-based approach (Fig. 1 illustrates this accuracy for 2020). As REE data is only available since 2013, the calibration using electricity data in our experiments covers the period from 2013 to 2020.

All of these data are combined in order to obtain the parameters of the optimization: CFs and demand series. In the calibration, the IC is directly assigned from the Ministry's data to the nearest point, and the CF is calculated using two approaches: an hourly estimation at a grid point level directly from the climate data, and a monthly calculation from electricity data, directly comparing generation to potential generation given the IC. The climate-derived CFs have higher temporal and spatial resolutions than the available observed data, but show biases with respect to the observed mean grid box CF. Therefore, a linear calibration of the national monthly CFs is applied to the hourly climate-derived CFs in order to remove persistent biases.

Demand is also modeled in order to permit analysis over periods with no observed demand data. Historical regional monthly demands are disaggregated into hourly values by means of the national hourly measures, which are readily available. This observed demand is then fit to a function of the temperature at two meters and nine parameters representing the part of the demand associated to heating, cooling, and base demand on weekdays, Saturdays and Sundays. For the REG approach, all of these approximations are averaged over Spain's administrative regions, instead of climate grid boxes. This function is used to fill demand series voids when necessary.

No information on the sub-regional scale is used in the REG approach. Thus, each regional CF is calculated as the uniform mean of the corresponding climate grid CFs, and any assigned IC is assumed to be evenly distributed within each region. This is done on purpose, because we want to explore the role that the incorporation of information at the sub-regional scale has on the results of the optimization. Thus, in the GRID approach, all individual grid points are associated with a CF and an IC. In both experiments, the pre-processing before the optimization includes a calibration of the climate-derived CFs via the electricity-derived CFs. Then, these individually calibrated CFs (for every grid point in GRID and for every region in REG) combined with the demand series are plugged into the optimization. The optimization then uses the distribution of IC as the decision variable, which is modified in order to identify the most optimal mixes in terms of the mean–variance analysis using the total 2020 IC as the maximum IC constraint when necessary. Mathematical definitions and expressions are discussed in Appendix A.

3. Results and discussion

Given the series of inputs (national hourly demand and hourly calibrated CF) for the REG and GRID experiments, a Pareto front describing the set of optimal IC scenarios in the Spanish electricity system (SES) (Fig. 2) is found as a solution of the optimization problem. Naturally, the 2020REF scenario is suboptimal, and thus, it is located away from the Pareto front in a penetration-risk diagram (Fig. 2). The simulated penetration and risk by current ICs line up quite well in both experiments, meaning that assuming a uniform distribution of the capacities within each region little affects the mean and the risk of the observed mix. In both approaches, the calculated penetration for 2020 matches that which was reported by REE: slightly under 30% [6].

An additional test of the comparability of both approaches is an analysis of the hourly penetration series. A comparison of the probability density functions (PDFs) of hourly penetration under the 2020REF scenario illustrates how the current mix not only presents similar values of overall penetration and risk (Fig. 2) in both experiments but also shares the overall distribution of hourly penetrations (Fig. 3 left-hand panel). More specifically, the two experiments simultaneously share the same hourly penetration at the vast majority of times (Fig. 3 right-hand panel). This confirms that the representation of the observed mix in both experiments is equivalent, and therefore, in the case of the observed mix, assuming regions are homogeneous does not have a strong effect on the penetration or risk.

When comparing optimal fronts, the first relevant feature that stands out is that the GRID fronts are always above (or to the left of) the REG fronts in the penetration-risk diagram. Indeed, optimal GRID scenarios fall in the impossible region (left and above) of the REG front, revealing the higher levels of achievable penetration (or lower levels of achievable risk) given by more favorable combinations between the selected installations. This effect is more apparent at high levels of penetration, at which locations and technologies with a high mean CF play an important role in meeting the high demand for renewable penetration, regardless of the associated risks (see [61] Section 3 for a mathematical justification). Under such circumstances, accounting for climate resources on a local scale provides higher CFs and more options for beneficial covariances than the aggregated regional averages.

The amount of IC that is installed at low levels of penetration – where the total capacity is not an active constraint – is also representative of the differences between the two approaches. Both optimizations have the same total IC constraint, but the diversity of CFs across the GRID experiment reveals higher penetration scenarios with lower levels of IC. For instance, only 75% of the current IC in the SES would be necessary to reach approximately the 32% penetration level in the GRID scenario, whereas 100% of the current IC would be required under the REG scenario. This confirms the advantage of the GRID method when it comes to identifying more beneficial CF combinations in the optimization. Likewise, the maximum penetration level reachable by the total available IC – at which all capacity is installed at the one location and technology with the largest CF – is 70% in the GRID approach and 50% in the REG approach. This result shows that the same total amount of available IC can yield higher returns in GRID compared to REG, as the smoothing effect of spatial aggregation disappears. It is important to note that this result strongly depends on a proper estimation of the maximum CFs, which are highly sensitive to changes in the sampling resolution.

This difference between the methods is expected, as more degrees of freedom are available in the GRID experiment and their combinations could identify optimal scenarios in the penetration-risk sense between a larger number of possible combinations. When using spatial aggregation, the averaging of the data alters the hourly penetration series, and thus modifies the value of the renewable energy resource. For instance, spatial aggregation changes the intrinsic maximum penetration-risk ratio (previously discussed and visualized as different slopes in the fronts at low levels of penetration). Similarly, the smoothing effect can

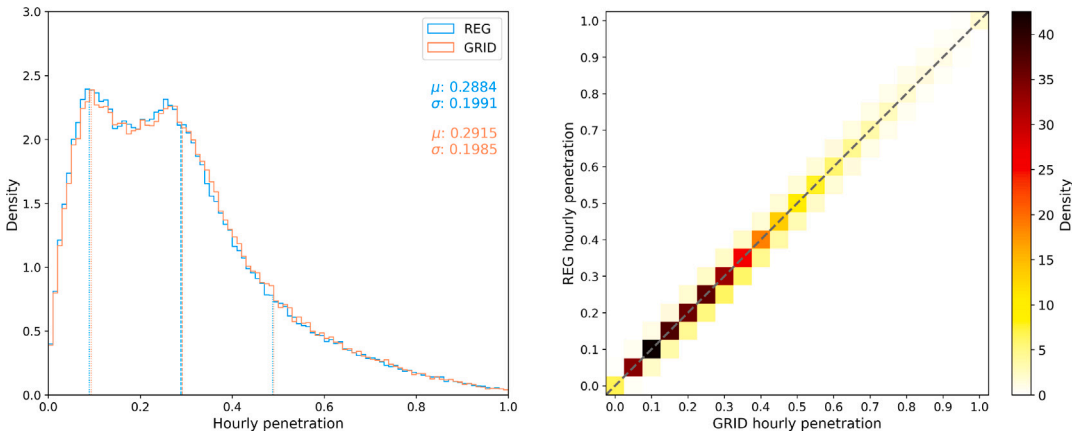


Fig. 3. Representation of CF distribution for the national hourly penetration series of 2020REF for REG and GRID in a histogram estimate of the joint distribution (2D) of hourly penetration (right) and the corresponding histogram estimates of both marginal distributions (1D) (left), both normalized so their integral is one. On the left, vertical dashed lines show the mean of each distribution and vertical dotted lines show one standard deviation from the mean in each direction for REG (blue) and GRID (red). μ and σ indicate the mean and standard deviation for each experiment, respectively. On the right, colors represent the density of total instances of hourly penetration and the gray dashed line represents the diagonal where hourly penetrations between REG and GRID are unaltered.

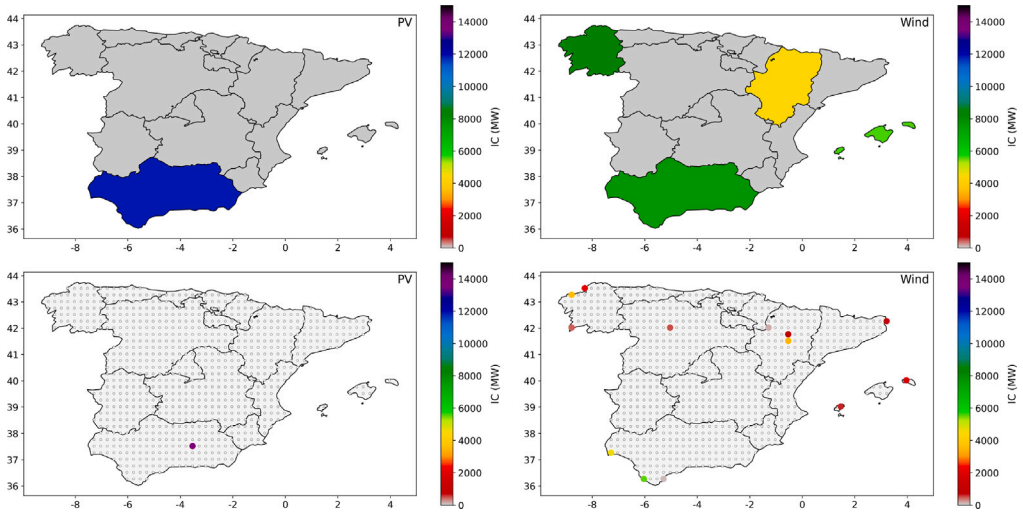


Fig. 4. Full repowering scenario IC for PV (left) and wind (right) using REG (top) and GRID (bottom). Colors represent the IC in each grid point or region. Grid points with ICs under 10 MW are represented by empty circles.

severely affect the complementarity between installations, as regional series can have very different variabilities from those seen at each of the grid points comprising them. The behavior and limits of the Pareto fronts as the granularity of the system changes are explored in Appendix B.

This effect can be seen clearly in the low-risk, high-penetration (GRID only), and full repowering mixes (Fig. 4 for the full repowering mix). For each experiment, these different mixes belong to the linear part of the front, and are thus colinear. Indeed, when the capacity constraint does not come into play (i.e., in the unconstrained case, or the optimal risk for a penetration level is reached without all available IC being assigned), the IC distribution is proportionally rescaled but no reallocation of capacity occurs. For instance, the low-risk mix can be obtained by scaling the full repowering mix by a factor of 0.680 in the GRID experiment and 0.913 in the REG experiment, these factors being calculated as the ratio between the IC in each scenario and the total available IC. This IC pattern produces the minimum penetration-risk ratio, and therefore returns the minimum risk at any positive level of penetration.

Before focusing on the differences between the two mixes, let us interpret the regional mix (Fig. 4 top panels) in terms of climate patterns. First, focusing on the IC of wind, Galicia has a climate that is strongly dominated by persistent winds from Atlantic depressions. This is also the case for western Andalucia. Analogously, the channeling effect produced by the Pyrenees and the Iberian system results in a singularly high frequency of significant persistent winds in the Aragon region, and thus represents a valuable asset for wind generation. Another region with wind IC in the linear optimal results is the Illes Balears, which contain the particularly windy island of Menorca, with multiple dynamic causes for highly persistent wind resources, such as sea breezes and the Tramontana winds that get channeled between the Pyrenees in northeastern Spain and the Massif Central in the middle of southern France. In a similar fashion, the IC of PV is all located in Andalucia, the southernmost region of Spain and the one with a particularly high level of solar radiation.

The areas with wind IC in the full repowering scenario and all colinear optimal scenarios are mostly consistent between the REG and GRID scenarios (Fig. 4). We say that the REG and the GRID mixes are consistent if regions with positive capacities are the same in both experiments and they are strongly consistent when the regional capacities are practically the same. The IC distribution for PV is consistent between REG and GRID scenarios, as there is only PV in Andalucia. However, the GRID approach concentrates all of that IC into one grid point, possibly the most favorable one of the region in terms of penetration and risk.

The IC distribution for wind is consistent in some regions (Andalucia, Illes Balears, and Galicia) and very consistent in Aragon. However, two regions (Castilla *y* Leon and Catalunya) without a presence in the REG mix, have one grid point with IC in the GRID mix. It is likely that the assumption of homogeneity in the REG scenario covered these points, but allowing heterogeneity within the regions in the GRID scenario reveals their contribution to the mix.

We can use some information from the last decade of analyzed data (2011–2020) to visualize the differences between GRID and REG. The total demand over this 10-year period was 2534 TWh. Maintaining the level of risk from 2020REF, the GRID generation reaches 1098 TWh. In comparison to only 875 TWh generated in REG throughout this decade, the increment in the generation from VRE sources shown by GRID is promising and manages to cover almost 10% of the total demand.

We would expect similar results in more realistic or complementary models of optimal investment in renewable capacities such as EOLES or PyPSA [69,70]. In fact, we would expect these results to be applicable in the real world. Having climatic series that accurately represent the potential of specific locations instead of a generic average can substantially change the outcome. Our results reveal the impact of accounting for more precise information, as long as the spatial variability is adequately captured by the data.

No assumption is made on the specifics of the Spanish case. Therefore, this study is applicable to any location conditioned to data availability. Generation and installed capacity information make for more precise estimations, but high frequency (i.e., hourly) climate and demand data are strictly necessary to run the optimization problem. Given a suitable location, we would expect similar results to the ones in this study considering that the decorrelation scale of the meteorological fields for that area is shorter than the scale of the high-resolution grid.

3.1. Attribution of optimal scenario improvements between REG and GRID

The changes in the IC distributions have clearly modified the penetration and risk values. For instance, the high-penetration mixes which have the same level of risk by definition (the standard deviation of the national hourly penetration series) in both scenarios, show higher levels of penetration (mean) under the GRID scenario than the REG scenario (Fig. 2). Contrarily, the low-risk hourly penetration series share their levels of penetration (mean), but the GRID solutions show lower levels of risk (standard deviation) than in the REG scenario.

3.1.1. Variance decomposition

A first approach to understanding these differences comes from an analysis of the different terms that make up the variance (risk squared). To this end, we use the national hourly penetration data in both experiments, and we analyze the variance (risk squared) of the series, as it can help pinpoint the mechanisms that have been exploited in the optimization in order to achieve the lower levels of risk.

We specifically use the variance in order to take advantage of the decomposition into a linear sum of individual variances and covariances. This provides an explicit expression of the different elements that contribute to the overall variance and a comparison between individual contributions in the REG and GRID scenarios. Considering P_k as the hourly penetration (generation divided by total demand) series for the grid point-technology pair k (and considering the alternative index r for each region-technology pair), and $P_{national} = \sum_k P_k = \sum_r \sum_{k\in r} P_k$, then the total variance can be expressed as:

$$\mathbb{V}(P_{national}) = \sum_r \sum_{k\in r} \mathbb{V}(P_k) + \sum_r \sum_{k,l\in r;k\neq l} \mathrm{Cov}(P_k, P_l) + \sum_{r,s;r\neq s} \sum_{k\in r;l\in s} \mathrm{Cov}(P_k, P_l), \quad (1)$$

where the first term is the sum of the individual variances of each grid point-technology pair, the second term is the covariance between grid point-technology pairs that share a region and the third term is the covariance between grid point-technology pairs in different regions. If we consider the division of each grid point-technology pair into its corresponding grid point (i) and technology, then we can further expand each one of these terms. From this point on, the index R will refer strictly to the spatial division represented by regions, and is not to be confused with each region-technology pair (r). The first term of the variance expression in Eq. (1) can be decomposed as

$$\sum_r \sum_{k\in r} \mathbb{V}(P_k) = \sum_R \sum_{i\in R} \mathbb{V}(P_{i,PV}) + \sum_R \sum_{i\in R} \mathbb{V}(P_{i,wind}), \quad (2)$$

which separates the sum of the individual variance of each grid point into the contributions it sees from PV (first term) and wind (second term). The second term in Eq. (1) refers to covariances between points in the same region and can be decomposed as

$$\begin{aligned}\sum_r \sum_{k,l\in r;k\neq l} \mathrm{Cov}(P_k, P_l) &= \sum_R \sum_{i,j\in R;i\neq j} \mathrm{Cov}(P_{i,PV}, P_{j,PV}) + \\ &+ \sum_R \sum_{i,j\in R;i\neq j} \mathrm{Cov}(P_{i,wind}, P_{j,wind}) + \sum_R \sum_{i,j\in R;i\neq j} \mathrm{Cov}(P_{i,PV}, P_{j,wind}) + \\ &+ \sum_R \sum_{i\in R} \mathrm{Cov}(P_{i,PV}, P_{i,wind}),\end{aligned} \quad (3)$$

where the first term represents the covariance between grid points with PV technology in the same region, the second term represents the covariance between grid points in the same region with wind technology, the third term represents the covariance between the PV technology at a grid point and the wind technology at a different grid point in the same region, and the fourth term represents the covariance between wind and PV technologies at the same grid point. Therefore, this decomposition differentiates whether the technology is the same (PV and wind are separate terms) or different and, in the case of different technologies, whether the grid point is the same or not. Finally, the last term in Eq. (1), which describes the covariances between grid points in different regions, can be further broken down as

$$\begin{aligned}\sum_{r,s;r\neq s} \sum_{k\in r;l\in s} \mathrm{Cov}(P_k, P_l) &= \sum_{R,S;R\neq S} \sum_{i\in R;j\in S} \mathrm{Cov}(P_{i,PV}, P_{j,PV}) + \\ &+ \sum_{R,S;R\neq S} \sum_{i\in R;j\in S} \mathrm{Cov}(P_{i,wind}, P_{j,wind}) + \sum_{R,S;R\neq S} \sum_{i\in R;j\in S} \mathrm{Cov}(P_{i,PV}, P_{j,wind}),\end{aligned} \quad (4)$$

where the first term represents the covariance between grid points in different regions with PV technology, the second term represents the covariance between grid points in different regions with wind technology and the third term represents the covariance between the PV technology at a grid point in one region and the wind technology at a grid point in a different region. This decomposition allows us to discern which technology is associated with these terms (PV or wind) or if it is a combination of the two.

The difference in the contribution to the total variance made by each of these terms in the REG or GRID scenario (Fig. 5) can help us better understand whether the improvements seen in the GRID scenario come from the identification of better individual CFs or more beneficial combinations of VRE sources. It is clear that the main improvement in the GRID experiment is the reduction of covariances, although different regions contribute less to the reduction of the total variance than points within the same region, which is clearly a consequence of the assumption made in REG of homogeneous regions, whereby all grid points contained in a selected region have some allocated IC. Additionally, although low levels of penetration show a strong reduction of PV covariances, the effect becomes small in comparison to the wind covariances as levels of penetration get higher. It is also worth mentioning that although covariances between different grid points in the same region dominate the change in variance, this is not the case for covariances between different technologies at the same grid point (see the narrow orange section at the top of Fig. 5). The change in this effect is orders of magnitude lower than the rest, which possibly

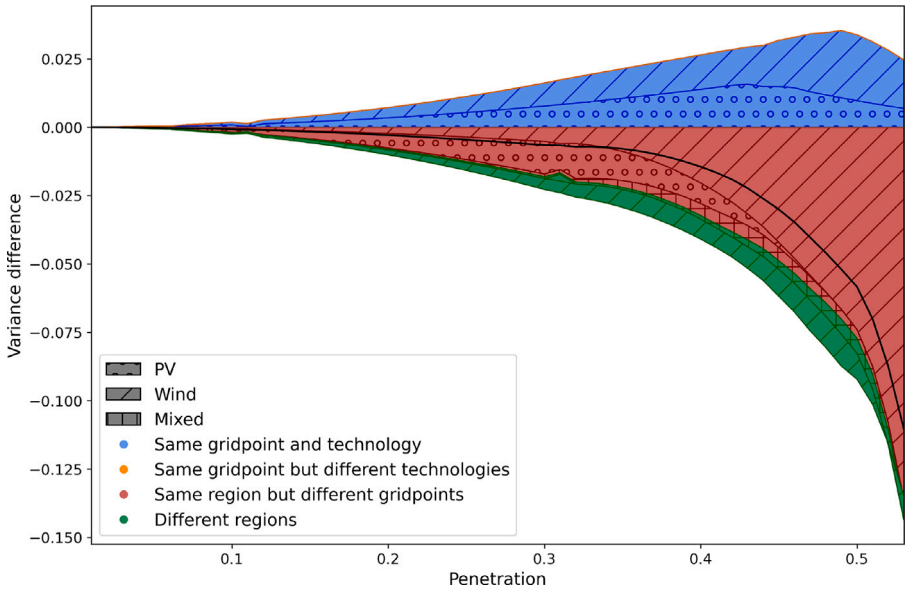


Fig. 5. Difference between the components of the variance for REG and GRID (GRID-REG) at each penetration level. Colors represent the relationship between two grid points: blue for the variance of the same grid point and technology, orange for the covariance in the same grid point and with different technologies, red for covariances between different grid points in the same region, and green for covariances between grid points in different regions. The patterns represent the relationship between a given technology at two points: circles represent PV, diagonal lines represent wind and a squared mesh represents a combination of both. The thick black line represents the difference in total variance.

stems from the lower number of terms contributing to it. All in all, the national variances for different penetrations in the REG scenario are higher despite a decrease in grid point-technology variances, clearly highlighting the fundamental role of covariances as fundamental in the lower levels of risk associated with the GRID experiment.

3.1.2. Histogram analysis

In an additional step towards understanding the more optimal results obtained in the GRID experiment with respect to the REG experiment, we compare the distributions of hourly penetrations for the low-risk and high-penetration scenarios (Fig. 6). Analyzing the general characteristics of these hourly time series beyond their means and variances can provide additional insights into the underlying processes that affect the penetration and risk values between experiments (other than the direct increase in granularity). For the high-penetration scenario, the marginal distributions (Fig. 6 top right) show that a general shift of the GRID solution towards higher hourly penetrations with respect to REG allows GRID to reach a higher overall penetration while maintaining the level of risk.

The GRID experiment results in less variable penetrations than REG by means of increasing the hourly penetrations that are below (or decreasing the hourly penetrations that are above) the mean value of nearly 30% (Fig. 6, top left). Indeed, a large portion of hourly instances below the 30% level in the REG experiment show increased levels of penetration in the GRID experiment (densities below the diagonal in Fig. 6 bottom left). Similarly, instances in the REG experiment that are above 30% are seen to be lower in the GRID experiment, and thus, overall hourly penetration variability is reduced.

The overall shift in the distribution for the high-penetration mixes (Fig. 6 top right) is attributed to a general increase in the hourly penetrations achieved by the GRID optimization (i.e., the highest densities below the diagonal in Fig. 6 bottom right). Therefore the higher penetration achieved by GRID with respect to REG is not attributable to changes in specific ranges of hourly penetration but to a mostly homogeneous increase in penetration. This reveals the improved IC distributions rendered by the GRID method, which achieve higher levels of penetration without the cost of any additional risk.

3.1.3. Decomposition into cycles and spectral decomposition

A powerful tool for dissecting the variability of a time series and analyzing its behavior is decomposition, and we take two different approaches to breaking down the series into its components. The first is a decomposition based on cycles that are not harmonic and the second is a power spectral density estimate using the Welch estimator [71] with a square one-year window and no overlapping. We perform these analyses on the national hourly penetration series in both the REG and GRID experiments. More specifically, we analyze and compare the high-penetration mixes as they have fixed variance by definition and therefore, the total power is constant between the experiments.

Our REG and GRID scenarios show similar distributions between the cycles (Fig. 7 left), with a daily and a seasonal cycle that are mainly attributable to the climate resource, a less important weekly cycle coming from the demand, and very small inter-annual variability: the most important contribution comes from all the other variability at sub-annual scales.

An analysis of the cycles (Fig. 7 left) reveals that the main difference in the GRID solution, with respect to REG, is the reduction of variability in the two main climate-driven cycles. The variance of the seasonal and diurnal cycles is reduced and moved to intermediate scales (e.g., the rest of sub-annual variability).

To get further detail on how the variance is distributed between cycles, we turn to the Welch estimate (Fig. 7 right). Despite the smoothing effect at the one-year frequency (associated with the seasonal cycle), it is clear that the most energetic frequencies belong to the seasonal and diurnal cycles. Note that the frequencies associated with the rest of the sub-annual scales are increased overall in the GRID scenario with respect to REG, but none of them exceed the most energetic climate-associated peaks. Therefore, the optimal solution achieved by GRID can reach the same level of risk as its REG equivalent, while reduced contributions of the most important climatic frequencies. This lower variability in the frequencies with the strongest climatic forcing allows the GRID method to identify more ambitious scenarios in terms of penetration, while maintaining the same level of risk as the REG method.

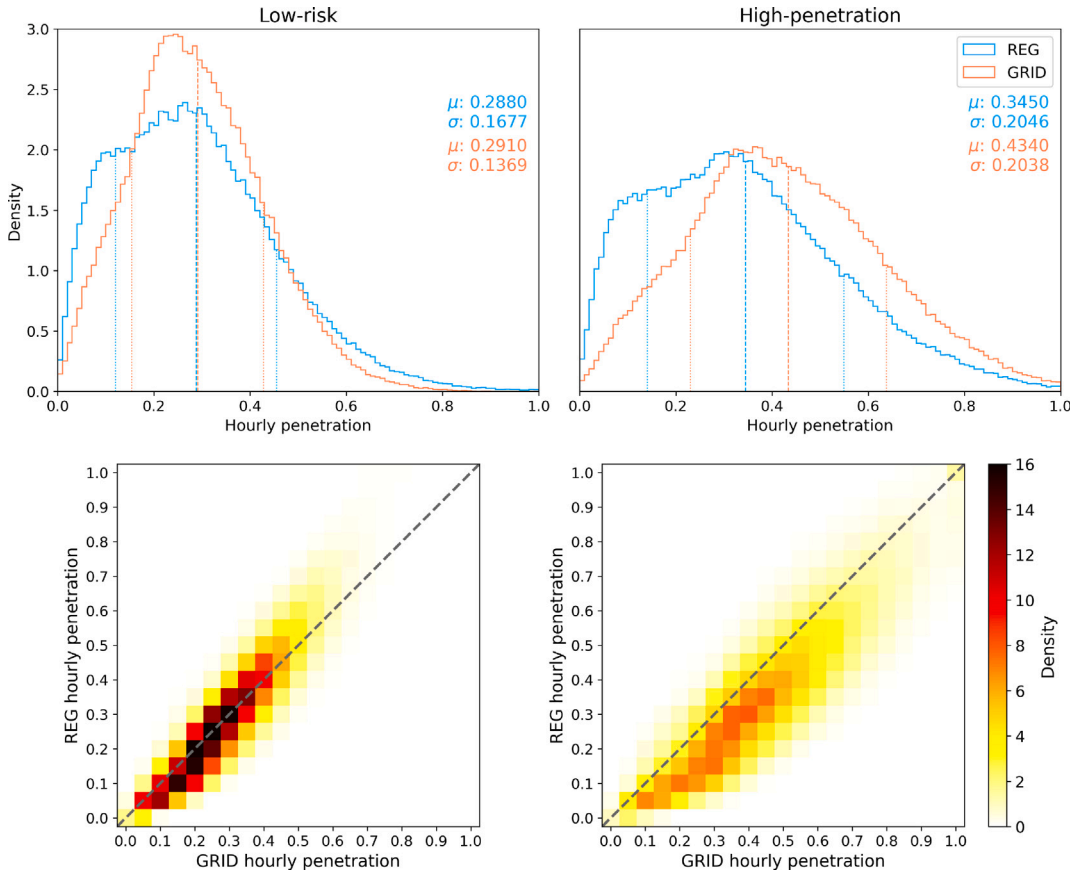


Fig. 6. Same as Fig. 3 for the low-risk (left) and high-penetration (right) mixes. Representation of CF distribution for the national hourly penetration series of each mix for the REG and GRID experiments in a histogram estimate of the joint distribution (2D) of hourly penetration (top) and the corresponding histogram estimates of both marginal distributions (1D) (bottom), both normalized so their integral is one.

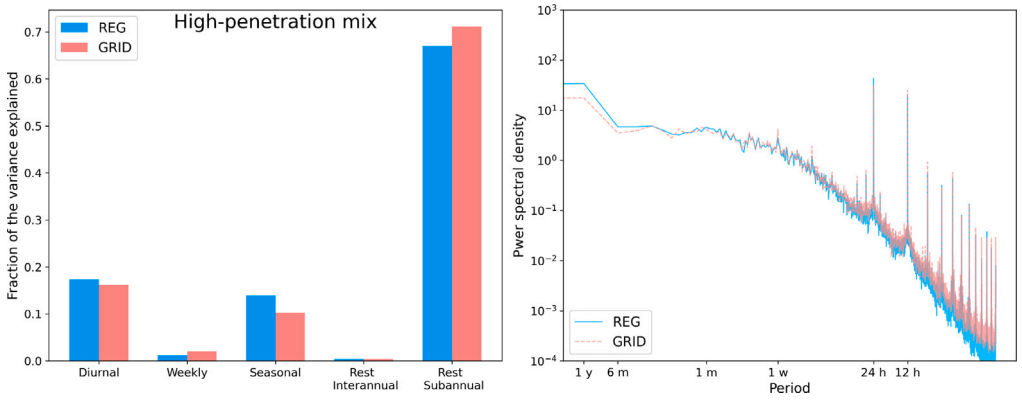


Fig. 7. Representation of the variance distribution in frequencies. On the left, the fraction of the variance of the high-penetration scenario explained by the daily, seasonal and weekly cycles and for all other frequencies for REG (blue) and GRID (red). On the right, the Welch estimate of the power spectral density for REG (blue) and GRID (red).

4. Summary and conclusions

The need for an energy transition towards a more renewable future is imperative, and thus, we must investigate the most favorable pathways for such a shift. Despite any political and social impacts on their development, climate variability in space and time is a key factor in deciding where and how much VRE capacity should be installed. Thus, adequately accounting for climate conditions is crucial in planning the deployment of this transition. Small modifications to the distribution of capacities and the methods for finding them can severely affect the success of any given scenario.

The combination of different technologies is key to reducing the variability of the VRE production in the context of energy system planning. Therefore, taking the highest possible resolution of information provides a more concrete and realistic vision of the possible combinations. The present study has shown how this consideration affects optimal energy deployment scenarios in the mean–variance sense. Using higher granularity and more detailed descriptions of actual CFs leads to optimal higher penetration and lower risk mixes, this is mainly attributable to the effect of complementarity in addition to the obvious benefit of providing more precise locations of installations in the optimal IC scenarios. Using the same total IC, scenarios with over 10% more penetration (e.g., from 32% to 44% in the case of the full

repowering scenarios) are reachable only by more precisely describing the system characteristics. A more refined representation of the climate resources in order to optimize VRE capacities at finer scales allows us to better account for and adapt to the challenges of variability posed by the different scales of climatic variability.

In a direct comparison of modeling experiments, the high-granularity grid-based approach results in optimal renewable deployment scenarios with higher levels of penetration and precision in the location of installations than more highly aggregated approaches due to two main factors. First, considering the complementarity between resources identifies more instances for favorable generation, whether the goal is to achieve a certain penetration level or to not exceed a given level of risk. Secondly, this approach favors the allocation of capacity to sites not contributing to the main energetic climatic frequencies of energy production, thus reducing overall variability.

The effects of increasing granularity in the description of optimal renewable energy deployment scenarios in our representation of the Spanish system have particular bounds. The climate resource sets an intrinsic limit to the attainable penetration-risk ratio, which can be represented by an asymptotic Pareto front (finding an intrinsic penetration-risk ratio of 2.22; intrinsic meaning independent of the granularity used for the optimization, not necessarily independent of the climate data and its resolution). This fundamental front provides an estimate of what could be an upper limit in terms of renewable penetration-risk pay off given the available resource. We identify the frontier reached when using the full climate grid optimization as being very close to the asymptotic behavior of increasing granularity, with differences in penetration and/or risk under 1% at any given point.

These new findings show that higher granularity will provide more informed and robust descriptions of optimal renewable deployment scenarios to best support stakeholders and policy makers so that they may safely and fully reach their sustainable commitments.

It is important to note that there is a limit to the level of granularity that can be used in e4clim, as the model is designed under certain assumptions. The feature that most limits the granularity is the power distribution grid, since it is not considered at the scales used in this study, but could become a key factor at higher resolutions. Therefore, besides the intrinsic limit to the reachable levels of penetration established by the asymptotic front, there is a limitation to the usable granularity set by the very modeling assumptions themselves.

Additionally, sensitivity to the climate data changes along with the changes in configuration. General climate data biases have the same effect whether administrative regions are considered or the whole climate grid is considered, as they represent a shift of the entire set of climate data. However, spatially located errors in the climate data become more important when the whole climate grid is considered. This occurs because when administrative regions are considered, the averaging processes can act as error compensation mechanisms in the climate data. Therefore, when aiming for a high level of granularity, rigorous quality control of the data is crucial.

Another consideration involves limitations to the maximum IC, as the scenarios with very high levels of penetration (which have all capacity allocated to one point and one technology) are unrealistic in real-world applications. As an additional layer of realism, further developments of the model will address limitations of maximum ICs for reasons related to demographic, geographical, or political factors.

CRediT authorship contribution statement

Aina Maimó-Far: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Victor Homar:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – review & editing, Funding acquisition, Project administration, Resources, Supervision. **Alexis Tantet:** Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. **Philippe Drobinski:** Conceptualization, Formal analysis, Investigation, Writing – review & editing, Funding acquisition, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The gridded data of the optimal scenarios, as well as the simulated generation time series are available upon request.

Acknowledgements

This research is framed within the TRAMPAS (PID2020-113036RB-I00/AEI/10.13039/501100011033) project. Author A. Maimó-Far received funding from the Spanish Ministerio de Ciencia Innovación y Universidades (FPU18/00520).This research made use of e4clim, an open source software, developed in the frame of the Energy4Climate Interdisciplinary Center (E4C) of *Institut Polytechnique de Paris* and *Ecole des Ponts ParisTech* and supported by the *3rd Programme d'Investissements d'Avenir* [ANR-18-EUR-0006- 02] and by the Foundation of *Ecole Polytechnique* (*Chaire Défis Technologiques pour une Énergie Responsable*).

Appendix A. The math of the optimization problem

Our approach to the optimization problem, shown here, follows the process presented in [63]. Given the hourly capacity factor $CF_{i,t,j}$, the hourly generation at a given location i and time t by renewable technology j is:

$$g_{i,t,j} = IC_{i,t,j} \cdot CF_{i,t,j} \tag{A.1}$$

where $IC_{i,t,j}$ is the installed capacity for that technology, location, and time.

Given the hourly demand $d_{i,t}$, the hourly penetration at a given location i and time t by technology j is defined as the percentage of the total demand covered by that method of generation:

$$p_{i,t,j} = \frac{g_{i,t,j}}{\sum_i d_{i,t}} \tag{A.2}$$

The average penetration (or penetration) at a given location i by a technology j is defined as the average over time of the hourly penetration:

$$P_{i,j} = \frac{1}{T} \sum_t p_{i,t,j} \tag{A.3}$$

where T is the number of hours in the hourly series.

Similar to this, the overall hourly penetration of the system can be defined as the sum of all of the penetrations of its components:

$$P_t = \sum_{i,j} p_{i,t,j} \tag{A.4}$$

Combining the two previous definitions, the overall penetration of the system can be defined as:

$$P = \frac{1}{T} \sum_{i,t,j} p_{i,t,j} \tag{A.5}$$

Since the demand $d_{i,t}$ and the $CF_{i,t,j}$ are independent of the installed capacity, and they reflect the climatic and demographic behaviors as well as the characteristics of the VRE technologies, we aim to find the $IC_{i,j}$ that maximizes the coverage of the demand given the $CF_{i,t,j}$ produced by the climatic conditions and given the technologies. Notice how the $IC_{i,j}$ is no longer time dependent, as the aim is to find the best combination given a specific climatic and demographic state. Therefore, the first objective of the optimization is to find the following maximum:

$$\max_{IC_{i,j}} P \equiv \max_{IC_{i,j}} \frac{1}{T} \sum_{i,t,j} p_{i,t,j} \equiv \max_{IC_{i,j}} \frac{1}{T} \sum_{i,t,j} \frac{IC_{i,t,j} \cdot CF_{i,t,j}}{\sum_i d_{i,t}} \tag{A.6}$$

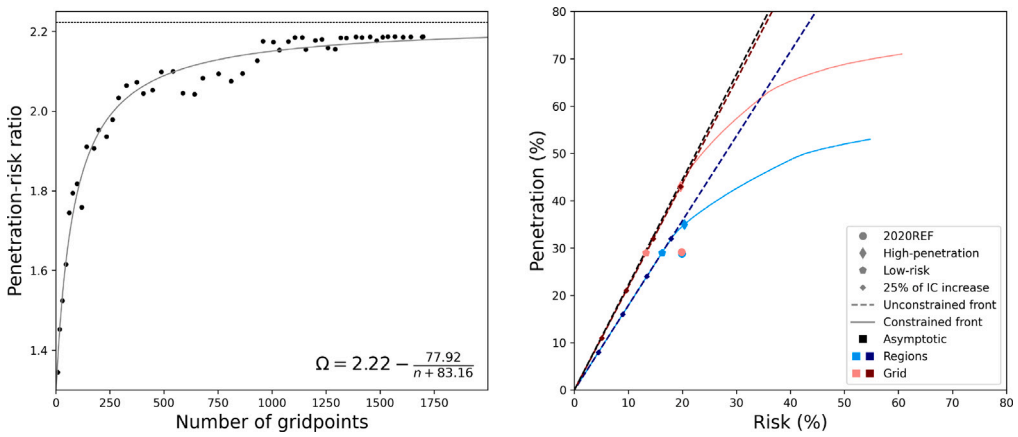


Fig. B.8. Asymptotic Pareto front estimation (left) and corresponding finite granularity fronts. in Fig. 2 (right). In the left-hand panel, black points represent the penetration-risk ratio (slope of the unconstrained Pareto front) for each number of grid points, the solid gray line represents the fitted function of all these points and the thin dashed black line represents the asymptotic value of said function. The right-hand panel presents the same elements as Fig. 2 with the addition of a black dashed line representing the asymptotic unconstrained Pareto front.

At the same time, the squared risk associated with a given selection of $IC_{i,j}$ is defined using variance as a proxy, while considering the predictable and unpredictable variability of the solar radiation [60]. Let us define a new index for each location-technology pair: $k = (i, j)$. Therefore, the overall systematic variance is:

$$\sigma^2 = \mathbb{V}\left[\sum_k IC_k \frac{CF_{k,t}}{\sum_i d_{i,t}}\right] \tag{A.7}$$

The variance is the risk squared used for wind generation, but not for PV. Therefore:

$$r^2_{wind} = \mathbb{V}_t\left[\sum_i IC_{i,j=wind} \frac{CF_{i,j=wind,t}}{\sum_i d_{i,t}}\right] \tag{A.8}$$

For the PV part of the risk, if we consider x to be PV generation, then it can be split into its predictable (derived from clear-sky radiation) and unpredictable components:

$$x_t = x_{P_t} - x_{NP_t} \tag{A.9}$$

In order to avoid naturally variable differences from having any influence on the risk, we take the spatial (i) mean (*) over the whole system (i.e., all regions or all grid points, depending on the experiment being considered) and apply the same predictable component for each grid point in the definition of risk:

$$x^*_t = x^*_{P_t} - x_{NP_t} \tag{A.10}$$

where $x^*_{P_t}$ is the predictable national average component of PV generation and x^*_t is the estimated PV generation using the national average of the predictable component. Therefore, we can take the variance of this variable, and separate its components:

$$\sigma^2 = \mathbb{V}(x^*_t) = \sigma^2_{x^*_{P_t}} + \sigma^2_{x_{NP_t}} - 2\mathrm{cov}(x^*_{P_t}, x_{NP_t}) \tag{A.11}$$

Finally, considering that the covariance term also contains the predictable variability, it can be averaged over all spatial elements i (i.e. regions or grid points) in order to avoid any influence of differences in predictable spatial variability on the definition of PV risk. This allows us to consider only the contribution of unpredictable variability on the risk, while maintaining comparability to the wind risk. Therefore, the definition of risk associated with PV is:

$$r^2_{PV} = \sigma^2_{x^*_{P_t}} + \sigma^2_{x_{NP_t}} - 2\langle \mathrm{cov}(x^*_{P_t}, x_{NP_t})\rangle_i \tag{A.12}$$

It is therefore the sum of both definitions of risk (which are of comparable magnitudes, as just mentioned) that is minimized. It should be noted that this correction is only applied to the variance of each PV location and does not affect the covariances between different location-technology pairs, which are calculated using the original series for PV as well.

In addition to the risk and penetration definitions, given that installed capacity in the real world cannot not be negative, the installed capacity is bounded:

$$IC_k \geq 0 \quad \forall k \tag{A.13}$$

Finally, in some experiments we impose a bound on the total installed capacity so that it does not exceed the total of the reference year 2020. Therefore:

$$\sum_k IC_k \leq IC_{2020} \tag{A.14}$$

Appendix B. Asymptotic Pareto front

In order to better understand the impacts of effective spatial granularity on the identification of optimal scenarios for deploying renewable energy systems, we designed a set of academic experiments. In these experiments we gradually modify granularity of the model and analyze the effect on the results of the optimization. To this end, we change the granularity of the data by averaging the grid climate values over increasingly large grid boxes, starting at the full climatic grid resolution and ending with only 4 total grid boxes over the domain. For each level of granularity in the model, an optimization is performed and an unconstrained Pareto front is computed. As the modeling granularity gets finer, the intrinsic penetration-risk ratio of the system tends to increase (i.e., we see steeper Pareto front slopes), with our GRID scenario setting the highest ratio. Since the system is driven by climate resources, the observed increase in the penetration-risk ratio is bounded by the level of correspondence between these resources and the demand. This increasing and bounded behavior suggests the existence of an asymptotic Pareto front, which is a reflection of an intrinsic property of the underlying driving climate. By intrinsic, we mean that it does not depend on the granularity of the grid used for the optimization (although it could depend on the grid used for the climate-model simulation).

We define this asymptotic front by extrapolating the fronts obtained from the set of experiments with varying granularities. In particular, we identify the penetration-risk ratio (slope of the Pareto front) for the unconstrained case for each level of granularity. These points are then fitted to a parameterized curve:

$$\Omega(n) = a + \frac{b}{n - c} \tag{B.1}$$

where Ω is the penetration-risk ratio; n is the number of grid points in the domain; and a, b, and c are fitting parameters. The result of fitting Eq. (B.1) using non-linear least squares can be seen in the left-hand panel of Fig. B.8. Considering this fitted function, we can compute the asymptotic penetration-risk ratio of 2.22. This allows us to represent an asymptotic Pareto front on the penetration-risk diagram as a straight line with the asymptotic penetration-risk ratio slope.

The resulting Pareto front is, by definition, located in the penetration-risk diagram above all fronts computed at any other granularity, and it represents an intrinsic characteristic of the system (Fig. B.8, right panel). This front is basically determined by the climate resource, though an important role is played by spatial and temporal covariances in the climate region. Certainly, the calculation of this asymptotic front is affected by the characteristics (e.g., resolution or spatial and temporal correlations) of the climate and demand data considered. However, the existence of a systemic Pareto front intrinsic to a renewable energy system is appealing as it suggests the existence of a limit to how much penetration can be secured without assuming further risk given specific climate data at a certain granularity. In the specific case of the Spanish system considered in this study, it is clear, by the location of the asymptotic front (Fig. B.8, right-hand panel), that the increase in penetration achievable with respect to the GRID scenario is negligible. This indicates that the level of granularity seen in the GRID experiment approaches the limit in terms of the highest achievable penetrations (or lowest achievable risks) in the SES.

References

[1] European Commission. A clean planet for all. A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy. 2018, COM/2018/773 final.

[2] Ministerio para la transición ecológica y el reto demográfico. Plan nacional de adaptación al cambio climático 2021–2030. 2020.

[3] Ministerio para la transición ecológica y el reto demográfico. Plan nacional integrado de energía y clima (PNIEC) 2021–2030. 2020.

[4] IEA. World energy outlook 2019. France: International Energy Agency; 2019.

[5] Victoria M, Zhu K, Brown T, Andresen GB, Greiner M. Early decarbonisation of the European energy system pays off. Nature Commun 2020;11(1):1–9.

[6] Red Eléctrica de España. The spanish electricity system 2020, report. 2021.

[7] Red Eléctrica de España. The spanish electricity system 2018 report. 2019.

[8] Red Eléctrica de España. The spanish electricity system 2013 report. 2014.

[9] Fueyo N, Sanz Y, Rodrigues M, Montañés C, Dopazo C. High resolution modelling of the on-shore technical wind energy potential in Spain. Wind Energy 2010;13(8):717–26.

[10] Eerens H, de Visser E. Wind-energy potential in Europe 2020–2030. Technical paper, The European Topic Centre on Air and Climate Change (ETC/ACC); 2008.

[11] Enevoldsen P, Permien F-H, Bakhtaoui I, von Krauland A-K, Jacobson MZ, Xydis G, Sovacool BK, Valentine SV, Luecht D, Oxley G. How much wind power potential does europe have? Examining European wind power potential with an enhanced socio-technical atlas. Energy Policy 2019;132:1092–100.

[12] Prăvălie R, Patriche C, Bandoc G. Spatial assessment of solar energy potential at global scale. A geographical approach. J Cleaner Prod 2019;209:692–721.

[13] Šúri M, Huld TA, Dunlop ED, Ossenbrink HA. Potential of solar electricity generation in the European Union member states and candidate countries. Sol Energy 2007;81(10):1295–305.

[14] Castillo CP, e Silva FB, Lavalle C. An assessment of the regional potential for solar power generation in EU-28. Energy Policy 2016;88:86–99.

[15] Cretì A, Fontini F. Economics of electricity: markets, competition and rules. Cambridge University Press; 2019.

[16] Sun W, Harrison GP. Wind-solar complementarity and effective use of distribution network capacity. Appl Energy 2019;247:89–101.

[17] Weschenfelder F, Leite GdNP, da Costa ACA, de Castro Vilela O, Ribeiro CM, Ochoa AAV, Araujo AM. A review on the complementarity between grid-connected solar and wind power systems. J Cleaner Prod 2020;257:120617.

[18] Santos-Alamillos FJ, Brayshaw DJ, Methven J, Thomaidis NS, Ruiz-Arias JA, Pozo-Vázquez D. Exploring the meteorological potential for planning a high performance European electricity super-grid: optimal power capacity distribution among countries. Environ Res Lett 2017;12(11):114030.

[19] Minai AF, Malik H. Metaheuristics paradigms for renewable energy systems: advances in optimization algorithms. In: Metaheuristic and evolutionary computation: algorithms and applications. Springer; 2021, p. 35–61.

[20] Asakereh A, Soleymani M, Ardebili SMS. Multi-criteria evaluation of renewable energy technologies for electricity generation: A case study in Khuzestan province, Iran. Sustain Energy Technol Assess 2022;52:102220.

[21] Nguyen TT, Phan TM, Nguyen TT, et al. Maximize the penetration level of photovoltaic systems and shunt capacitors in distribution systems for reducing active power loss and eliminating conventional power source. Sustain Energy Technol Assess 2022;52:102253.

[22] Maleki A. Optimization based on modified swarm intelligence techniques for a stand-alone hybrid photovoltaic/diesel/battery system. Sustain Energy Technol Assess 2022;51:101856.

[23] Alshamrani AM, Alrasheedi AF, Alnowibet KA. A game-theoretic model for wind farm planning problem: A bi-level stochastic optimization approach. Sustain Energy Technol Assess 2022;53:102539.

[24] Güven AF, Yörükeren N, Samy MM. Design optimization of a stand-alone green energy system of university campus based on Jaya-Harmony Search and Ant Colony Optimization algorithms approaches. Energy 2022;253:124089.

[25] Samy M, Almamlook RE, Elkhouly HI, Barakat S. Decision-making and optimal design of green energy system based on statistical methods and artificial neural network approaches. Sustainable Cities Soc 2022;84:104015.

[26] Gonzalez-Rodriguez AG, Serrano-Gonzalez J, Burgos-Payan M, Riquelme-Santos J. Multi-objective optimization of a uniformly distributed offshore wind farm considering both economic factors and visual impact. Sustain Energy Technol Assess 2022;52:102148.

[27] Ren Y, Yao X, Liu D, Qiao R, Zhang L, Zhang K, Jin K, Li H, Ran Y, Li F. Optimal design of hydro-wind-PV multi-energy complementary systems considering smooth power output. Sustain Energy Technol Assess 2022;50:101832.

[28] Omar AI, Khattab NM, Aleem SHA. Optimal strategy for transition into net-zero energy in educational buildings: A case study in El-Shorouk City, Egypt. Sustain Energy Technol Assess 2022;49:101701.

[29] Ullah K, Hafeez G, Khan I, Jan S, Javaid N. A multi-objective energy optimization in smart grid with high penetration of renewable energy sources. Appl Energy 2021;299:117104.

[30] Anvari S, Mahian O, Solomin E, Wongwises S, Desideri U. Multi-objective optimization of a proposed multi-generation cycle based on Pareto diagrams: Performance improvement, cost reduction, and CO2 emissions. Sustain Energy Technol Assess 2021;45:101197.

[31] Christiaanse T, Loonen RC, Evins R. Techno-economic optimization for grid-friendly rooftop PV systems–A case study of commercial buildings in British Columbia. Sustain Energy Technol Assess 2021;47:101320.

[32] Doepfert M, Castro R. Techno-economic optimization of a 100% renewable energy system in 2050 for countries with high shares of hydropower: The case of Portugal. Renew Energy 2021;165:491–503.

[33] Mokhtara C, Negrou B, Settou N, Settou B, Samy MM. Design optimization of off-grid hybrid renewable energy systems considering the effects of building energy performance and climate change: Case study of Algeria. Energy 2021;219:119605.

[34] Wimalaratna YP, Afrouzi HN, Mehranzamir K, Siddique MBM, Liew SC, Ahmed J. Analysing wind power penetration in hybrid energy systems based on techno-economic assessments. Sustain Energy Technol Assess 2022;53:102538.

[35] Samy M, Emam A, Tag-Eldin E, Barakat S. Exploring energy storage methods for grid-connected clean power plants in case of repetitive outages. J Energy Storage 2022;54:105307.

[36] Hai T, Dhahad HA, ATTIA E-A, Zakaria Z, Rashidi S, Singh PK, Shamseldin MA, Almojil SF, Almohana AI, Alali AF, et al. Design, modeling and multi-objective techno-economic optimization of an integrated supercritical Brayton cycle with solar power tower for efficient hydrogen production. Sustain Energy Technol Assess 2022;53:102599.

[37] Barakat S, Emam A, Samy M. Investigating grid-connected green power systems' energy storage solutions in the event of frequent blackouts. Energy Rep 2022;8:5177–91.

[38] Russo M, Carvalho D, Martins N, Monteiro A. Forecasting the inevitable: A review on the impacts of climate change on renewable energy resources. Sustain Energy Technol Assess 2022;52:102283.

[39] Deng X, Lv T. Power system planning with increasing variable renewable energy: A review of optimization models. J Cleaner Prod 2020;246:118962.

[40] Santos-Alamillos F, Thomaidis N, Quesada-Ruiz S, Ruiz-Arias J, Pozo-Vázquez D. Do current wind farms in Spain take maximum advantage of spatiotemporal balancing of the wind resource? Renew Energy 2016;96:574–82.

[41] Thomaidis NS, Santos-Alamillos FJ, Pozo-Vázquez D, Usaola-García J. Optimal management of wind and solar energy resources. Comput Oper Res 2016;66:284–91.

[42] Tantet A, Drobinski P. A minimal system cost minimization model for variable renewable energy integration: Application to France and comparison to mean-variance analysis. Energies 2021;14(16):5143.

[43] Holttinen H, Meibom P, Orths A, Lange B, O'Malley M, Tande JO, Estanqueiro A, Gomez E, Söder L, Strbac G, et al. Impacts of large amounts of wind power on design and operation of power systems, results of IEA collaboration. Wind Energy 2011;14(2):179–92.

[44] Roques F, Hiroux C, Saguan M. Optimal wind power deployment in Europe–A portfolio approach. Energy Policy 2010;38(7):3245–56.

[45] Yang Q, Huang T, Wang S, Li J, Dai S, Wright S, Wang Y, Peng H. A GIS-based high spatial resolution assessment of large-scale PV generation potential in China. Appl Energy 2019;247:254–69.

[46] Monforti F, Huld T, Bódis K, Vitali L, D'isidoro M, Lacal-Arántegui R. Assessing complementarity of wind and solar resources for energy production in Italy. A Monte Carlo approach. Renew Energy 2014;63:576–86.
[47] Pfenninger S, Staffell I. Long-term patterns of European PV output using 30 years of validated hourly reanalysis and satellite data. Energy 2016;114:1251–65.
[48] Carpio LG. Efficient spatial allocation of solar photovoltaic electric energy generation in different regions of Brazil: a portfolio approach. Energy Sources B 2021;16(6):542–57.
[49] Forouli A, Doukas H, Nikas A, Sampedro J, Van de Ven D-J. Identifying optimal technological portfolios for European power generation towards climate change mitigation: a robust portfolio analysis approach. Util Policy 2019;57:33–42.
[50] Jerez S, Thais F, Tobin I, Wild M, Colette A, Yiou P, Vautard R. The CLIMIX model: A tool to create and evaluate spatially-resolved scenarios of photovoltaic and wind power development. Renew Sustain Energy Rev 2015;42:1–15.
[51] Sevlian RA, Rajagopal R. A model for the effect of aggregation on short term load forecasting. In: 2014 IEEE PES general meeting| conference & exposition. IEEE; 2014, p. 1–5.
[52] Santos-Alamillos F, Thomaidis N, Usaola-García J, Ruiz-Arias J, Pozo-Vázquez D. Exploring the mean-variance portfolio optimization approach for planning wind repowering actions in Spain. Renew Energy 2017;106(C):335–42. http://dx.doi.org/10.1016/j.renene.2017.01.
[53] Xu D, Bai Z, Jin X, Yang X, Chen S, Zhou M. A mean-variance portfolio optimization approach for high-renewable energy hub. Appl Energy 2022;325:119888.
[54] Shapiro A, Dentcheva D, Ruszczynski A. Lectures on stochastic programming: modeling and theory. SIAM; 2021.
[55] Martínez-Gordón R, Morales-España G, Sijm J, Faaij A. A review of the role of spatial resolution in energy systems modelling: Lessons learned and applicability to the North Sea region. Renew Sustain Energy Rev 2021;141:110857.
[56] Raventós O, Dengiz T, Medjroubi W, Unaichi C, Bruckmeier A, Finck R. Comparison of different methods of spatial disaggregation of electricity generation and consumption time series. Renew Sustain Energy Rev 2022;163:112186.
[57] Aryanpur V, O'Gallachoir B, Dai H, Chen W, Glynn J. A review of spatial resolution and regionalisation in national-scale energy systems optimisation models. Energy Strateg Rev 2021;37:100702.
[58] Zappa W, van den Broek M. Analysing the potential of integrating wind and solar power in Europe using spatial optimisation under various scenarios. Renew Sustain Energy Rev 2018;94:1192–216.
[59] Banos R, Manzano-Agugliaro F, Montoya F, Gil C, Alcayde A, Gómez J. Optimization methods applied to renewable and sustainable energy: A review. Renew Sustain Energy Rev 2011;15(4):1753–66.
[60] Maimó-Far A, Tantet A, Homar V, Drobinski P. Predictable and unpredictable climate variability impacts on optimal renewable energy mixes: The example of Spain. Energies 2020;13(19):5132.
[61] Bouramdane A-A, Tantet A, Drobinski P. Adequacy of renewable energy mixes with concentrated solar power and photovoltaic in Morocco: Impact of thermal storage and cost. Energies 2020;13(19):5087.
[62] Thomaidis NS, Santos-Alamillos FJ, Pozo-Vázquez D, Usaola-García J. Optimal management of wind and solar energy resources. Comput Oper Res 2016;66:284–91. http://dx.doi.org/10.1016/j.cor.2015.02.016.
[63] Tantet A, Stéfanon M, Drobinski P, Badosa J, Concettini S, Cretì A, D'Ambrosio C, Thomopulos D, Tankov P. E4CLIM 1.0: The energy for a climate integrated model: Description and application to Italy. Energies 2019;12(22):4299. http://dx.doi.org/10.3390/en12224299.
[64] Hersbach H, Bell B, Berrisford P, Hirahara S, Horányi A, Muñoz-Sabater J, Nicolas J, Peubey C, Radu R, Schepers D, et al. The ERA5 global reanalysis. Q J R Meteorol Soc 2020;146(730):1999–2049.
[65] Urraca R, Huld T, Gracia-Amillo A, Martinez-de Pison FJ, Kaspar F, Sanz-Garcia A. Evaluation of global horizontal irradiance estimates from ERA5 and COSMO-REA6 reanalyses using ground and satellite-based data. Sol Energy 2018;164:339–54.
[66] Olauson J. ERA5: The new champion of wind power modelling? Renew Energy 2018;126:322–31.
[67] Jourdier B. Evaluation of ERA5, MERRA-2, COSMO-REA6, NEWA and AROME to simulate wind power production over France. Adv Sci Res 2020;17:63–77.
[68] López-Castrillón W, Sepúlveda HH, Mattar C. Too many solar panels? Oversizing or undersizing of hybrid renewable energy systems based on different sources of information. Sustain Energy Technol Assess 2022;52:102264.
[69] Shirizadeh B, Perrier Q, Quirion P. EOLES-elecRES model description. 2020.
[70] Brown T, Hörsch J, Schlachtberger D. PyPSA: Python for power system analysis. 2017, arXiv preprint arXiv:1707.09913.
[71] Welch P. The use of fast Fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms. IEEE Trans Audio Electroacoust 1967;15(2):70–3.