Sustainable Energy, Grids and Networks 44 (2025) 101957

Contents lists available at ScienceDirect

Sustainable Energy, Grids and Networks

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

Future-proofed resource adequacy metrics: A model-based assessment of multi-metric vs. composite-metric reliability standards

Paulo Brito-Pereira [a], Kenneth Bruninx [b], Laurens de Vries [b], Paolo Mastropietro [a,*], Pablo Rodilla [a]

[a] Instituto de Investigación Tecnológica, Universidad Pontificia Comillas, Rey Francisco 4, Madrid 28008, Spain
[b] Faculty of Technology, Policy and Management, TU Delft, Jaffalaan 5, Delft 2628 BX, Netherlands

ARTICLE INFO

Keywords:
Resource adequacy
Reliability standard
Capacity mechanism
De-rating factor
Multi-metric
Composite metrics

ABSTRACT

The rapid decarbonisation of the power sector is challenging the traditional resource adequacy framework. Variable and energy-limited resources are driving the emergence of new correlations that, together with extreme weather events, are rapidly changing the expected scarcity conditions in the electricity system. Traditional resource adequacy metrics are showing their limitations under these new conditions, and many regulators have already started to reform them. This article presents the first model-based comparative analysis of two different approaches that have been proposed to overcome these limitations, i.e., multi-metric standards (imposing a set of different resource adequacy constraints) and composite-metric standards (combining different resource adequacy metrics through weighting factors to build a single reliability standard). These two approaches are quantitatively evaluated in this article through case studies obtained from a simulation model, focusing not only on the impact of the reliability standard on the resource mix, but also on the design of the reliability product to be traded in a capacity mechanism to guide the system towards that mix.

1. Introduction

Security of supply has always been one of the main concerns of regulators in liberalised electricity sectors, where investment (and decommissioning) decisions should be taken by independent market actors based on the efficient signals provided by the electricity market price. Most countries with a liberalised electricity sector rely on so-called resource adequacy assessments, which use simulation models to analyse security of supply and characterise expected scarcity conditions in the system. The risk of shortfall is usually quantified using resource adequacy metrics such as LOLE (Loss of Load Expectation) or EUE (Expected Unserved Energy).[1] These metrics examine a particular type of contingency (e.g., loss of load events or unserved energy) over a number of different scenarios and then condense this information through a statistical measure (e.g., an average, or expectation, or a specific percentile of the probability distribution function). Some regulators impose a reliability standard (or RS) by setting thresholds for these resource adequacy metrics. If the resource adequacy assessment reveals that the expected power system does not meet the reliability standard set by the regulator, a resource adequacy mechanism, usually referred to as a capacity mechanism, may be implemented. The capacity mechanism provides a new economic signal which, together with the energy market price, is intended to steer the system towards an RS-compliant resource mix [1]. Capacity mechanisms have been extensively studied in the academic literature [2,3].

The rapid decarbonisation of power systems, which will need to be accelerated in the coming decades to mitigate climate change, is reshaping the whole resource adequacy debate [4]. As new variable and energy-limited resources enter the power system, new correlations emerge (e.g., between peak demand and the availability of intermittent generation [5–7], between the latter and the availability of hydropower [8,9], or even between the availability of generation and transmission

* Correspondence to: IIT, Universidad Pontificia Comillas, Rey Francisco 4, Madrid, Spain.
E-mail address: Paolo.Mastropietro@comillas.edu (P. Mastropietro).

[1] The terminology used in relation to resource adequacy varies considerably from region to region. The terminology used in this article and some equivalent terms found in the literature are as follows: i) unserved energy is used in this paper to refer to non-served energy and EUE (Expected Unserved Energy) is used to refer to the equivalent concept of EENS (Expected Energy Non-Served); ii) reliability standard, or RS, is used to refer to a reliability or resource adequacy criterion, target or objective; iii) capacity mechanism is used to refer to Capacity Remuneration Mechanism (CRM), e.g., a capacity market; iv) de-rating factor is used to refer to capacity credit.

https://doi.org/10.1016/j.segan.2025.101957
Received 29 January 2025; Received in revised form 31 July 2025; Accepted 31 August 2025
Available online 1 September 2025

resources [10]). In addition, extreme weather events are becoming more frequent and severe than in the past due to climate change, which has a significant impact on the so-called tail risk (risk of events with low probability of occurrence but potentially large consequences or impact) that needs to be considered in modern resource adequacy assessments [11,12].

The traditional resource adequacy metrics commonly used in these assessments are not well suited to characterising the scarcity conditions that decarbonising power systems will face. For example, traditional metrics may be unable to capture the economic damage of a loss-of-load event, if they do not consider the volume of unserved energy; or they may focus on the median of the probability distribution function, without considering tail-risk events. Several experts are calling for a fundamental reform of these metrics [13,14], covering both the type of contingency they assess [15,16] and the statistical measure used to condense information from different scenarios [17]. This dynamic is also affecting regulatory activity. Brazil [18], MISO [19] or Pacific Northwest [20] in the United States, the United Kingdom [21] or Australia [22] are just a few examples of power systems that have reformed or are in the process of reforming their resource adequacy assessments and the metrics on which they are based.

One of the most common arguments found in the academic literature to justify a reform of resource adequacy assessments is that a single resource adequacy metric cannot characterise the new and evolving scarcity conditions [23]. A single metric could represent in the same way two scarcity events which may be significantly different and have very different impacts on electricity consumers. For example, Fig. 1 shows different scarcity conditions, with each blue block representing a 1-MWh shortfall. Cases A and B have the same LOLEv (Loss of Load Events), but very different EUEs or LOLH (Loss of Load Hours), while the opposite is true for cases C and D.

According to several experts, the solution to this problem would be to move to a multi-metric approach [11]. Resource adequacy assessments would use a set of different resource adequacy metrics and evaluate the compliance of the resource mix with multiple reliability standards. A multi-metric approach would make it possible to examine the different facets of security of electricity supply and provide the system operator and the policy-maker with more information on the nature of potential shortfall events.

However, when the resource adequacy assessment informs the design of a capacity mechanism, a multi-metric approach may have a significant drawback. In capacity mechanisms, resources should be remunerated according to their ability to help the system meet the reliability standard by providing their contribution during the expected scarcity conditions. This means that the reliability product to be traded in the capacity mechanism is conceptually related to the reliability standard that the system must comply with. As resources cannot guarantee to be fully available during scarcity events, all resources willing to participate in a capacity mechanism are assigned a de-rating factor that is used to calculate their "firm capacity" [24]. A 100-MW resource with a 30 % de-rating factor would only be allowed to trade 30 MW in the capacity mechanism. In theory, the de-rating factor should reflect each resource's contribution to meeting the reliability standard set by the regulator.[2] If the system has to meet multiple reliability standards, then each resource should be assigned multiple de-rating factors (one for each standard) and the capacity mechanism should be based on the procurement of multiple reliability products. However, there is no obvious solution to coordinate the procurement of these different reliability products and inefficiencies may arise.[3]

One possible solution may come from the so-called composite metrics. These metrics simply combine different primary resource adequacy metrics through weighting factors to build a single metric. For instance, a composite metric based on the unserved energy but assessed by two different statistical measures, i.e., the expectation and the $CVaR_{5\%}$ (Conditional Value at Risk), has been proposed by the Australian regulator [22,26] and is shown in Eq. (1), where w is the weighting factor.

$$w \bullet EUE + (1 - w) \bullet CVaR_{5\%}(UE) \tag{1}$$

Composite metrics allow different facets of the security of supply problem to be captured in a single reliability standard. An advantage of this approach is that the associated de-rating factors for participation in the capacity mechanism can be calculated according to the contribution to this single reliability standard. Therefore, each technology or resource would be assigned a single de-rating, simplifying the design of the capacity mechanism. A disadvantage of a composite-metric standard, also analysed in this article, is that it may not guarantee that all the single standards that compose the composite-metric standard are met.

The objective of this article is to compare multi-metric and composite-metric reliability standards and to quantitatively assess, through some illustrative case studies, the impact of the two approaches on the resulting resource mix and de-rating factors to be used in a capacity mechanism. De-rating factors are analysed as the design element of the reliability product to be traded in the capacity mechanism that is more affected by the selection of the resource adequacy metric. This is achieved through a stochastic expansion planning model,[4] where the optimisation is constrained using different resource adequacy metrics, first by a multi-metric approach and then by setting a composite-metric reliability standard. The modelling exercise complements the theoretical analysis presented in this introduction and allows the effects of the choice of metrics for the reliability standard to be clearly demonstrated. The resource adequacy metrics used in the article to build the multi-metric and composite-metric standards are the same as those used in the Australian proposal, i.e., expected unserved energy, EUE, and $CVaR_{5\%}$ of unserved energy, $CVaR_{5\%}$(UE). However, the main findings of the article are not affected by the choice of the underlying resource adequacy metrics, as explained in the conclusions (Section 4.1).

To the best of the authors' knowledge, multi-metric and composite-metric approaches have been proposed by experts and regulators, but their application has not been independently tested in a model-based comparative analysis. The findings of this article could inform the current debate on the reform of resource adequacy frameworks and the discussions raised in the case studies may be relevant to both system planners and regulators.

The remainder of this article is structured as follows. Section 2 presents the methodology and case studies used to compare multi-metric and composite-metric approaches. Section 3 presents and discusses the results of the simulations. Section 4 summarises the main findings of the article and outlines areas of research that could be explored in future work.

[2] This is achieved, for example, by calculating de-rating factors using the Effective Load Carrying Capacity (ELCC) or similar methodologies, which are the most common approach in modern capacity mechanisms [24,25].

[3] The possibility of introducing a multi-product capacity market in the Chilean power sector was assessed but rejected exactly due to these complexities and potential inefficiencies [27]. A detailed theoretical analysis on a multi-product resource adequacy mechanism and the resulting coordination needs exceeds the scope of this article.

[4] In this article, the central planner model with resource adequacy constraints is used to simulate an electricity market with perfect competition and an incentive to install firm capacity. For further details on this equivalence, the reader may refer to Annex I of [17] or [28].

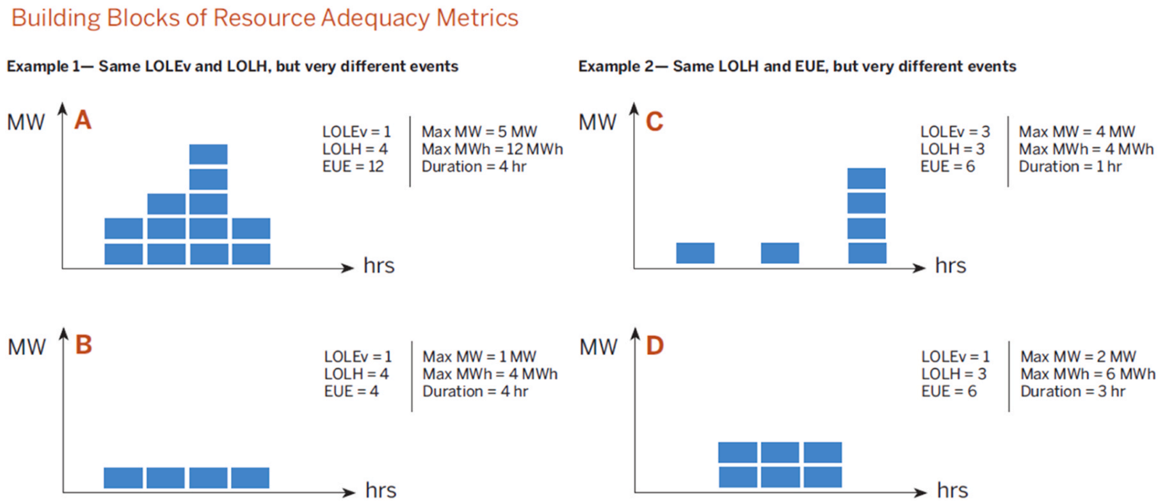


Fig. 1. Illustrative example of the need for a multi-metric approach; chart from [1].

2. Methodology

The simulation model used to compare different resource adequacy metrics is based on stochastic expansion planning with 500 scenarios, which optimises the resource mix from scratch (greenfield) to supply a given load curve at minimum expected cost (see Annex 1 for the detailed mathematical formulation). The load curve is inspired by the real demand of the Spanish power sector and consists of 672 hourly values, corresponding to four weeks, one for each season of the year. The corresponding load duration curve is shown in Fig. 2.

To serve this demand curve, the model can decide to install and dispatch five different technologies: nuclear, combined cycle gas turbines (CCGTs), wind, solar photovoltaic and diesel.[5] The cost of unserved energy (acting in the model as a price cap) is set at 3 000 €/MWh. The stochastic simulation considers 500 scenarios of availability of thermal resources (nuclear, CCGT and diesel), modelled by availability matrices built by a two-state Markov chain using a Monte Carlo approach [30,31]. These availability matrices have been modified to simulate the effects of a polar vortex similar to the one that hit the ERCOT power system in 2021 [32]. In 2.5 % of the scenarios, CCGT availability drops by 80 % for three consecutive days. Similar cold snaps, albeit with less dramatic impacts, have affected other electricity systems in the United States in the past, such as the polar vortex experienced in the Midwest, South Central, and East Coast regions in 2014 [33]. The polar vortex is used in the model to represent an extreme weather event that may affect the tail risk in the resource adequacy assessment, decoupling the different resource adequacy metrics assessed in the simulations to better illustrate the findings of the research. However, these findings can be generalised to any type of scarcity condition (see Section 3.3.3 for a sensitivity analysis of the impact of the polar vortex). The profiles of demand and availability of renewable resources are the same in all scenarios.

Several simplifications (four weeks to represent one year, deterministic profiles for demand and availability of renewable energy sources, absence of energy-limited resources among the potential technologies, etc.) have been introduced in the modelling exercise to meet computational constraints. The aim of the simulation model, however, is not to represent or predict the real operation of a power system, but rather to compare multi-metric and composite-metric reliability standards through illustrative, yet realistic, case studies that allow their performance to be assessed.

2.1. Case studies

Four case studies are generated by the simulation model, as shown in Fig. 3. In the first case study, the stochastic expansion planning is run without any reliability standard, i.e., without any constraint on resource adequacy. In this base case, the model simply seeks an economic equilibrium between increasing the cost of new entrants and reducing the cost of unserved energy valued at the 3 000-€/MWh price cap. The outcome is evaluated quantitatively in terms of EUE and $\text{CVaR}_{5\%}(\text{UE})$, and the resulting resource mix is set as the reference for comparison with the other case studies.

In the second case study, the model is run with single reliability standards. The EUE and $\text{CVaR}_{5\%}(\text{UE})$ constraints are imposed separately in the stochastic expansion planning, creating two sub-cases. The outcome of each sub-case is evaluated by examining the variation in installed capacity of different technologies with respect to the base case and their de-rating factors. De-rating factors, in this and other case studies, are calculated as the contribution of each technology to the reliability standard using an ELCC methodology: i) first, the installed capacity of a given technology is marginally incremented and the resulting variation in the resource adequacy metric is calculated; ii) second, the same marginal increment in perfect generation (i.e., a unit with full availability and no outages) is added to the resource mix and the resulting variation in the resource adequacy metric is calculated; iii) the two variations are compared and the de-rating factor is calculated as their ratio; iv) the process is repeated for each technology.

In the third case study, the model is run with a multi-metric reliability standard, i.e., the EUE and $\text{CVaR}_{5\,\%}(\text{UE})$ constraints are imposed together, with reliability standards that result in the simultaneous activation of both constraints. Again, the outcome is evaluated by examining the variation in installed capacity of different technologies with respect to the base case and their de-rating factors. In this case, since two reliability standards are active at the same time, two de-rating factors are calculated for each technology (see Section 3.3).

In the fourth case study, the model is run with a composite-metric reliability standard. A weighting factor, w, is used to combine EUE and $\text{CVaR}_{5\,\%}(\text{UE})$ into a single metric and a constraint is applied based on this. The reliability standard and the constraint is formulated as in Eq. (2), with EUE_{RS} and $\text{CVaR}_{5\,\%}(\text{UE})_{\text{RS}}$ being the standards used in the second and third case studies. The normalisation of each resource adequacy metric by its standard allows metrics that may have different orders of magnitude, or even different units of measurement, to be combined.

[5] Investment and variable costs for these technologies have been taken from [29].

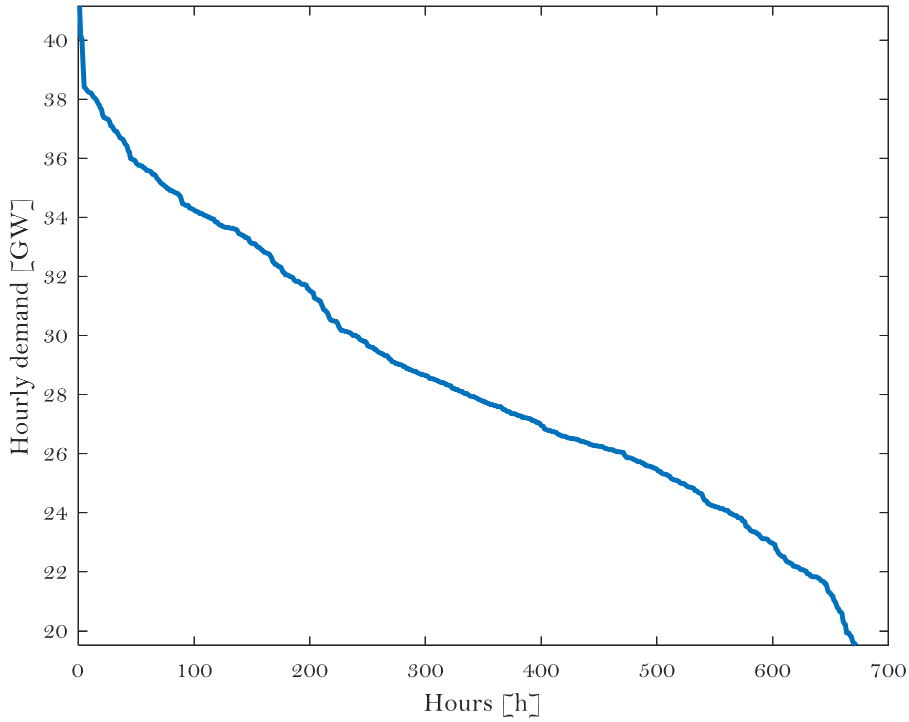


Fig. 2. Load duration curve of the demand used in the stochastic expansion planning.

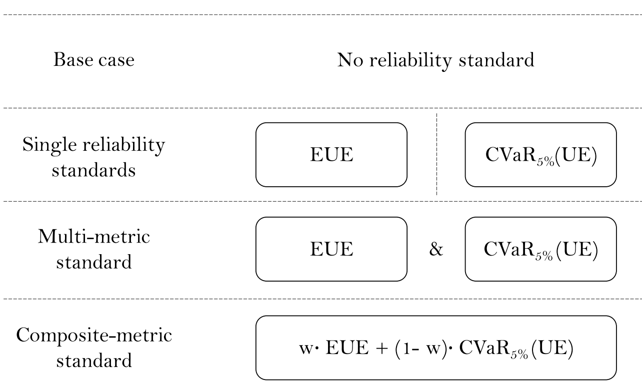


Fig. 3. Schematic representation of the four case studies.

$$w \bullet \left(\frac{EUE}{EUE_{RS}}\right) + (1 - w) \bullet \left(\frac{CVaR_{5\%}(UE)}{CVaR_{5\%}(UE)_{RS}}\right) \leq 1 \tag{2}$$

The model is run for values of the weighting factor ranging from 0, corresponding to a $\text{CVaR}_{5\,\%}$(UE) constraint, to 1, corresponding to an EUE constraint. Again, the outcomes are evaluated by examining the variation in installed capacity of different technologies relative to the base case and their de-rating factors.

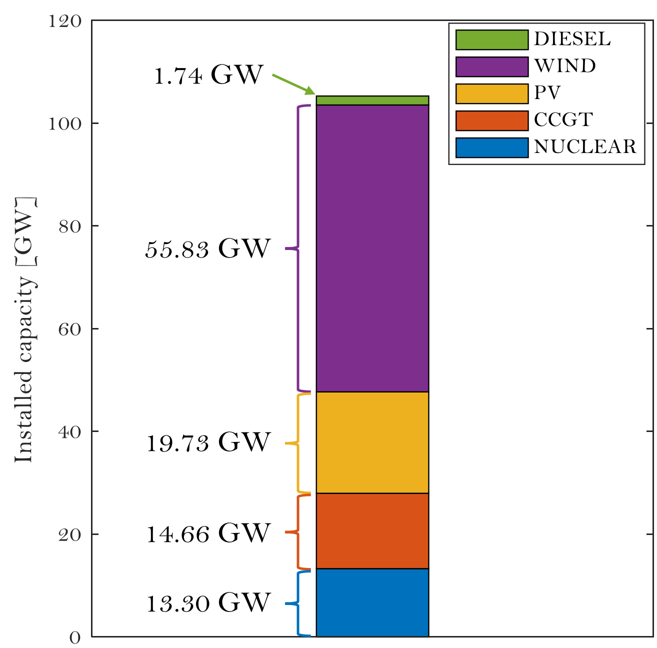


Fig. 4. Resource mix in the base case.

3. Results and discussion

This section presents the main results of the simulation model, divided according to the four case studies identified in Section 2.

3.1. Base case - no reliability standard

As discussed in Section 2, in the absence of a resource adequacy constraint, the stochastic expansion planning seeks an economic equilibrium between increasing the cost of new entrants and reducing the cost of unserved energy. The resulting resource mix is shown in Fig. 4, while Fig. 5 presents the dispatch of a typical day. As can be seen, nuclear is used for baseload, CCGTs are responsible for supporting the variability of the large installed renewable capacity, especially wind, while diesel is used as a peaker, especially in the polar vortex scenarios when CCGTs are not available.

The resource adequacy of the base case is assessed using the EUE and the $\text{CVaR}_{5\,\%}$(UE). In the upper chart of Fig. 6, these parameters are plotted together with the unserved energy for each scenario, ordered by increasing values. It can be observed how the unserved energy grows slowly until the polar vortex scenarios are reached, where it becomes much higher. In the lower chart of Fig. 6, the same information is

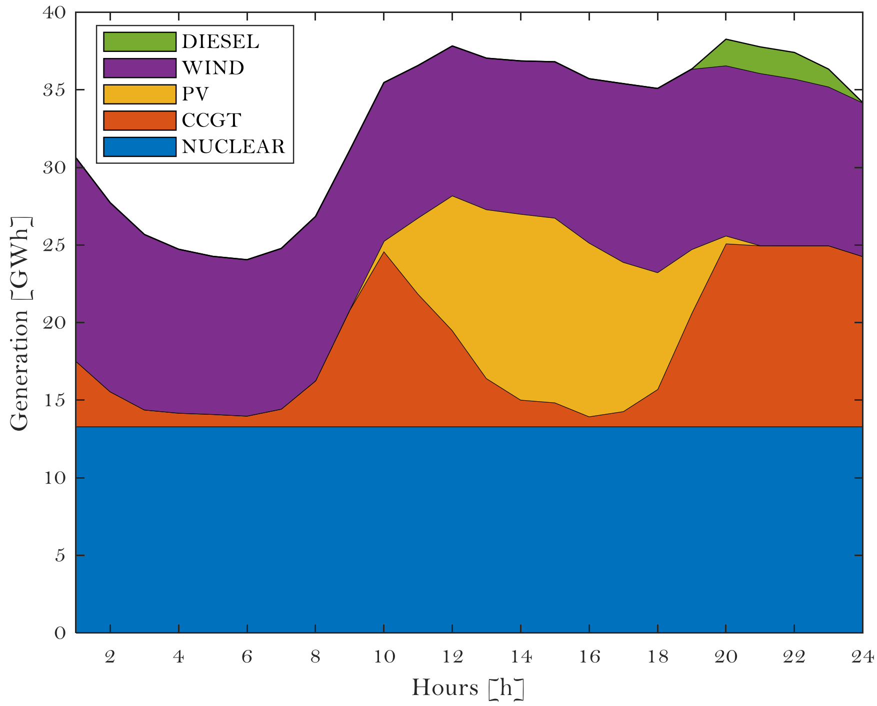


Fig. 5. Dispatch of different technologies in a typical day.

presented as a probability density distribution function, which makes it easier to visualise the so-called tail risk, represented by events with low probability but high magnitude (very high unserved energy).

The $CVaR_{5\,\%}$(UE) is ten times higher than the EUE because it focuses on the 5 % scenarios with the highest unserved energy.

3.2. *Single-metric standards*

In this case study, the stochastic expansion planning is run twice with different resource adequacy constraints, one applied to the EUE and one to the $CVaR_{5\,\%}$(UE). The outcomes are shown in Fig. 7, with blue representing the results of applying the EUE constraint and red representing the results of applying the $CVaR_{5\,\%}$(UE) constraint. Here, as in other charts, attention is focused on the two technologies that are more affected by a change in the resource adequacy metric used to set the reliability standard, i.e., CCGTs, which suffer severe unavailability in polar vortex scenarios, and diesel, which is the peaking technology called upon by the model to make up for the missing CCGT capacity.

When a resource adequacy constraint is imposed, i.e., when the model seeks a more reliable resource mix, the installed capacity of CCGTs decreases, due to their poor performance in the polar vortex scenarios. However, the drop is much larger for the $CVaR_{5\,\%}$(UE) constraint than for the EUE constraint. In fact, when a $CVaR_{5\,\%}$(UE) constraint is imposed, the model is asked to reduce the unserved energy in the worst scenarios, which encompass the polar vortex. Therefore, more CCGTs are replaced by diesel units, as shown in the upper chart in Fig. 7. For the same reason, the de-rating factor assigned to CCGTs drops from over 80 % when the reliability standard is based on EUE to 55 % when the constraint is based on $CVaR_{5\,\%}$(UE). The diesel de-rating is high in both sub-cases, but is slightly higher when the reliability standard is based on $CVaR_{5\,\%}$(UE).

3.3. *Multi-metric standard*

3.3.1. *The multi-activation area*

In this case study, the stochastic expansion planning is required to meet two reliability standards simultaneously. The first step is to find values for the two resource adequacy constraints that allow this simultaneous activation. To find these values, the multi-metric domain was characterised as shown in Fig. 8. Each point of the multi-metric domain represents an EUE-$CVaR_{5\,\%}$(UE) pair that could be imposed in a multi-metric reliability standard. The starting point is the base case, which defines the initial values for the EUE and the $CVaR_{5\,\%}$(UE). From this point, the blue line is drawn by constraining the EUE and assessing the $CVaR_{5\,\%}$(UE) of the resulting mix, repeating this operation for small decrements in EUE. The red line is drawn in a similar way by constraining the $CVaR_{5\,\%}$(UE) and assessing the EUE of the resulting mix.

The two lines allow three areas to be identified in this double-metric domain.[6] In the EUE-dominated area, the imposition of two reliability standards would make the $CVaR_{5\,\%}$(UE) standard redundant as only the EUE constraint would be activated. In the CVaR-dominated area, only the $CVaR_{5\,\%}$(UE) constraint would be activated and the EUE standard would be redundant. However, in the area between the blue and red lines, referred to here as the multi-activation area, both resource adequacy constraints would be activated simultaneously.

3.3.2. *The application of multi-metric standards*

To demonstrate this multiple activation, the stochastic expansion planning was run with two reliability standards corresponding to the point[7] shown in Fig. 9, which is in the multi-activation area.

The outcomes of applying this multi-metric standard are shown in yellow in Fig. 10, where they are compared with the results of applying single-metric standards. From a computational point of view, the activation of both constraints has been demonstrated by their dual

[6] In this article, the multi-metric domain is composed by two resource adequacy metrics and this results in a bi-dimensional multi-activation area. Multi-metric or composite-metric reliability standards can also consider three or more resource adequacy metrics, resulting in multi-activation volumes or hypervolumes.

[7] This point was also used to define the reliability standards in the previous case study, where the resource adequacy constraints were applied separately. The same reliability standards are also the starting point for the following case study, where the constraints are combined. This is done to ensure comparability of results.

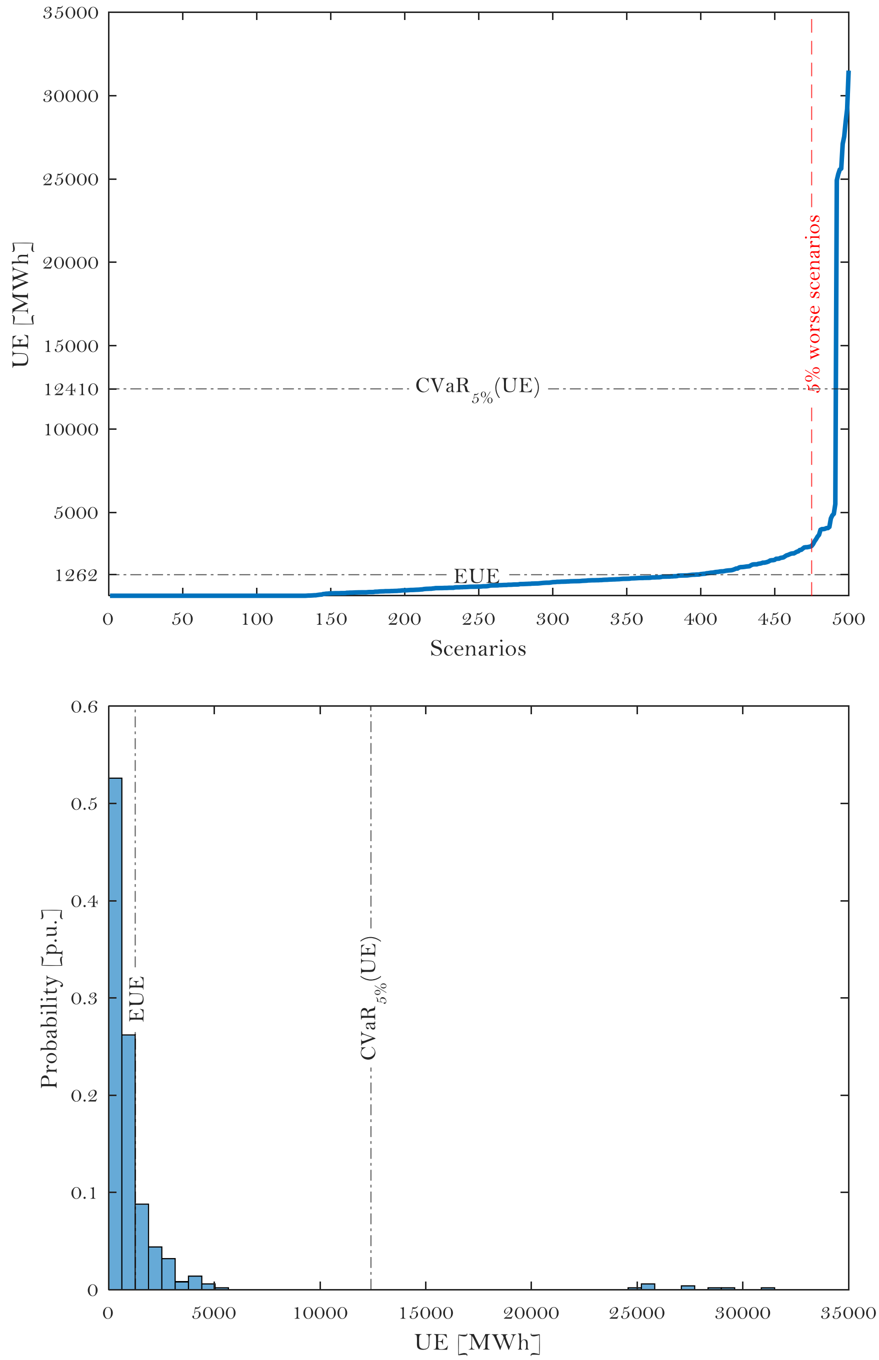


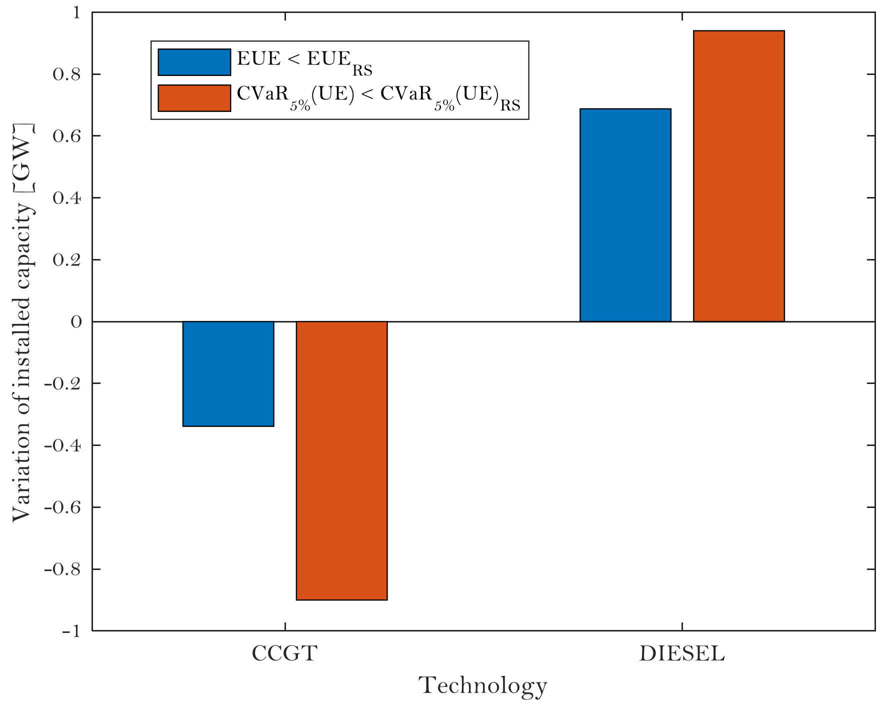
Fig. 6. Resource adequacy assessment for the base case.

variables, both of which assume a non-zero value.

The resource mix resulting from the application of the multi-metric standard shows an intermediate variation in installed capacity compared to the single-metric standards. However, as shown in the lower chart in Fig. 10, the application of a multi-metric standard results in the calculation of two different de-rating factors for a single technology, in this case CCGT, with a higher de-rating factor corresponding to its contribution to the EUE standard and a lower de-rating factor corresponding to its contribution to the $CVaR_{5\,\%}$(UE) standard. In the capacity mechanism, these two de-rating factors should be reflected in the procurement of two reliability products. Procurement of multiple products in the capacity market forces the regulator to define different requirements and may lead to different prices for different reliability products, increasing the complexity of the design of the regulatory instrument [17,34].

3.3.3. *Sensitivity analysis without polar vortex*

It should be remarked that resource adequacy metrics tend to be highly correlated, meaning that a scarcity event is likely to affect different metrics in a consistent way. In the case studies presented here, the polar vortex creates two types of scarcity conditions in this particular power system: shortages simply related to sporadic outages of thermal power plants, and shortages related to an extreme weather event that mostly affects a single technology (CCGT). This approach made it possible to reduce the correlation between the two resource adequacy metrics under study, EUE and $CVaR_{5\,\%}$(UE). This has an impact on the multi-activation area, which becomes larger when metrics in the multi-metric domain are less correlated. To investigate the impact of the



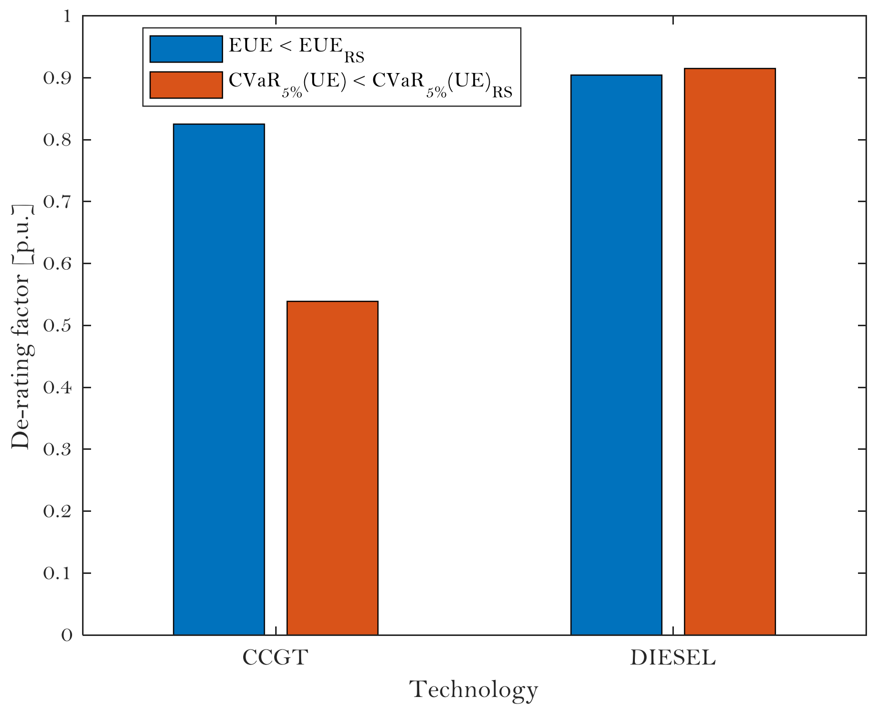


Fig. 7. Outcomes of the imposition of single-metric standards.

correlation between metrics on the multi-activation area, the model was run without polar vortex scenarios. The result is shown in Fig. 11, which has the same axis ranges as Fig. 8 for comparison.

The base case without polar vortex obviously shows lower values for both EUE and $\text{CVaR}_{5\,\%}$(UE). However, it can also be observed that the multi-activation area, although present (as shown in the zoomed-in graph), is much thinner than in the polar vortex case study. This means that, without the polar vortex, the two resource adequacy metrics are much more correlated. The imposition of a multi-metric standard, in this case, is much more likely to result in the activation of only one of the two initial standards.

Of course, the size of the multi-activation area depends on the type of scarcity conditions expected in the system and the type and number of metrics selected to assess resource adequacy in the system. However, any discussion on the use of multi-metric standards should consider the possibility of several redundancies and the eventual activation of a single reliability standard.

3.4. Composite-metric standard

The last case study examines the application of a composite-metric standard. The resource adequacy constraint is imposed by the

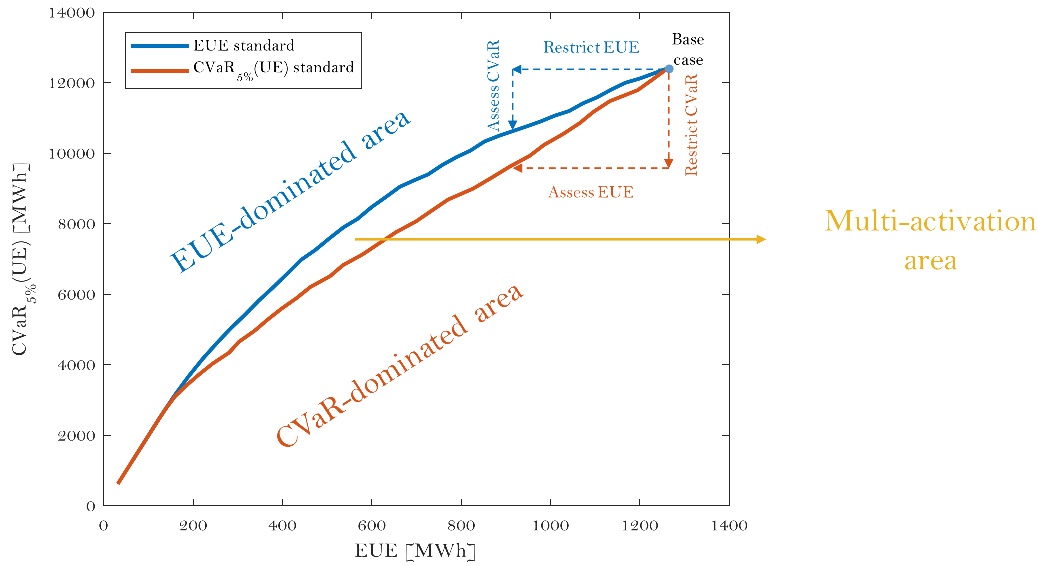


Fig. 8. Characterisation of the multi-metric domain and identification of the multi-activation area.

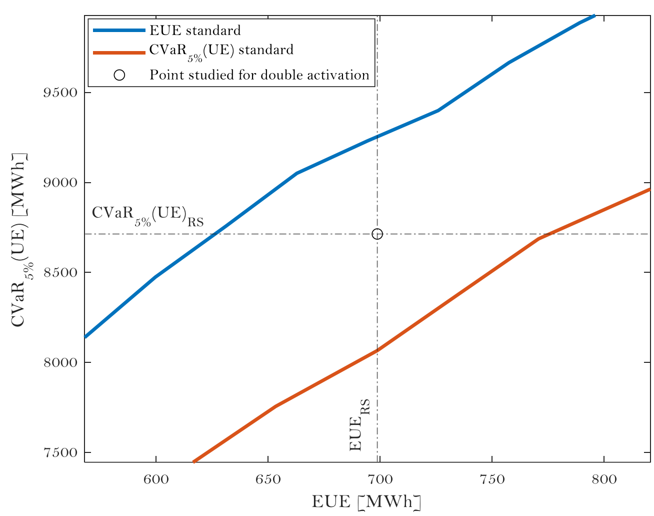


Fig. 9. Multi-activation area and reliability standards under study.

normalised formula presented in Eq. 2. The model is run for different values of the weighting factor and the results in terms of the adequacy of the resulting resource mix are shown in the multi-metric domain in Fig. 12.

When the weighting factor is equal to 0, the composite-metric standard coincides with the $CVaR_{5\,\%}(UE)$ constraint and the resulting point in the multi-metric domain lies on the red curve. When the weighting factor is equal to 1, the composite-metric standard coincides with the EUE constraint and the resulting point lies on the blue curve. For values of the weighting factor between 0 and 1, the optimisation model is allowed to seek an equilibrium between the two initial reliability standards in the pursuit of cost minimisation and compliance with the composite-metric standard. For low values of the weighting factor (between 0 and 0.4 in this case study), the $CVaR_{5\,\%}(UE)$ constraint is still dominant and the resulting resource mix exceeds the initial $CVaR_{5\,\%}(UE)$ standard (being below that reference) but is above the initial EUE standard. For high values of the weighting factor (between 0.5 and 1 in this case study), the EUE constraint becomes dominant and the resulting resource mix exceeds the initial EUE standard but is above the initial $CVaR_{5\,\%}(UE)$ standard.

The outcomes in terms of installed capacity and de-rating factors with the application of a composite-metric standard are shown in Fig. 13, focusing on combined cycles for the sake of clarity. Both the installed capacity of CCGTs and their de-rating factor (calculated as their contribution to the composite-metric reliability standard) gradually decrease as the weighting factor goes from 1, corresponding to the EUE constraint, to 0, corresponding to the $CVaR_{5\,\%}(UE)$ constraint. This behaviour is consistent with the effects shown in the previous case studies. Annex 2 presents an assessment of the energy/capacity revenue shares for this case study, based on the dual variables associated with the constraints in the model.

The main difference between the application of a composite-metric and a multi-metric reliability standard is that the composite-metric standard allows a single de-rating factor to be calculated for each technology. If a capacity mechanism is implemented, this feature allows the procurement of a single reliability product, avoiding the complexities of a multi-product capacity market.

4. Conclusions and policy implications

The traditional resource adequacy framework is being challenged by the decarbonisation of the power sector. New correlations in the resource mix and extreme weather events are rapidly changing the expected scarcity conditions in the electricity system. Conventional resource adequacy metrics are revealing their limitations in this new environment. Several experts are calling for a reform of resource adequacy metrics and reliability standards, and many policy-makers have already started to update their regulations. Some experts argue that a single resource adequacy metric cannot characterise the new and evolving scarcity conditions and call for the introduction of multi-metric reliability standards. Other experts have proposed the introduction of composite-metric reliability standards that combine different resource adequacy metrics through weighting factors.

This article presents the first model-based exercise to compare these two approaches using case studies obtained from a simulation model. The latter is based on stochastic expansion planning, which is run i) without any reliability standard, ii) with single reliability standards, EUE or $CVaR_{5\,\%}(UE)$, separately, iii) with a multi-metric standard that imposes both constraints simultaneously, and iv) with a composite-metric standard that combines the two resource adequacy metrics in a single standard. The main findings are summarised below.

The outcomes of the case studies show the impact of the choice of resource adequacy metric(s) on the resource mix, as well as on the associated de-rating factors to be used in a capacity mechanism to guide

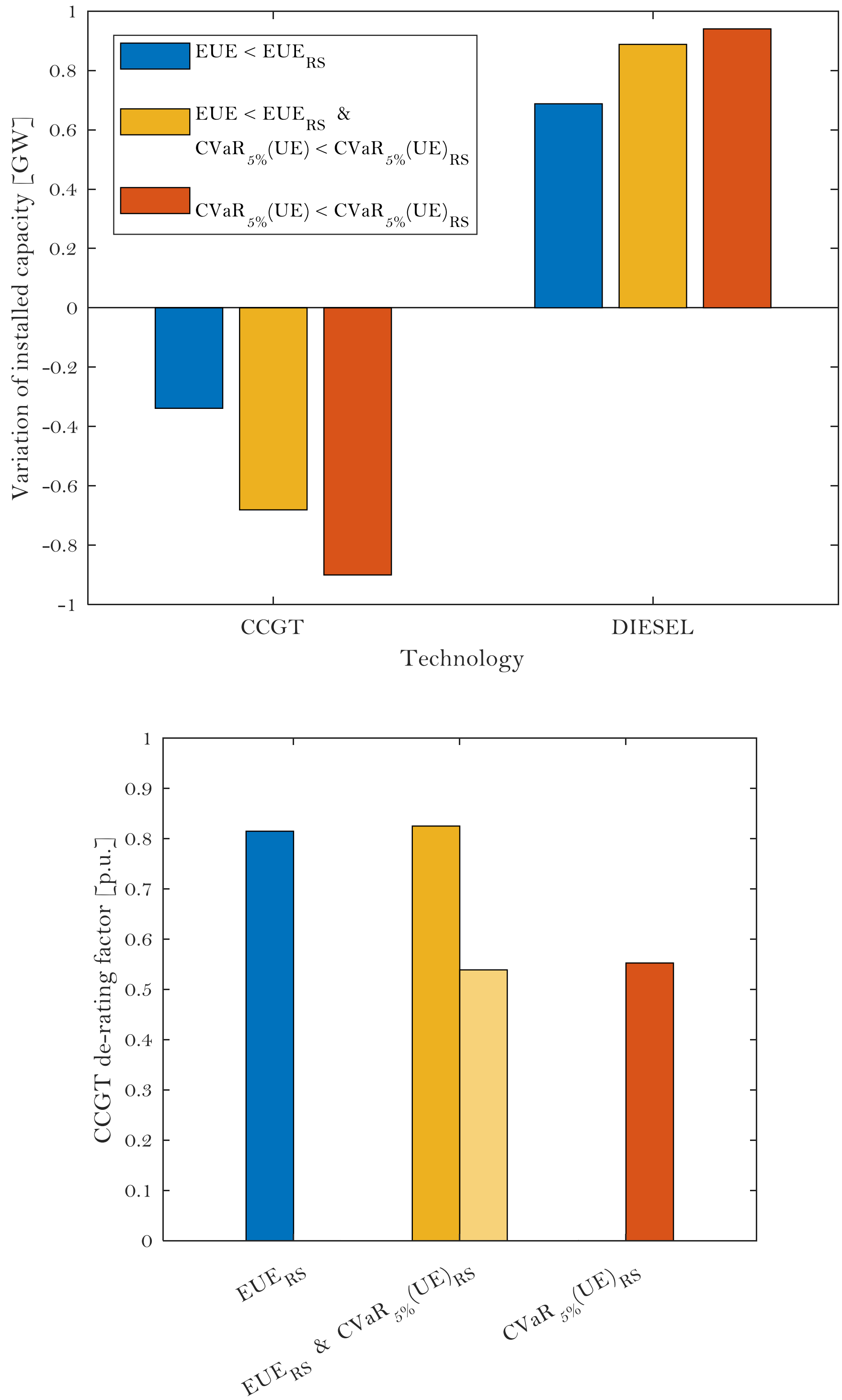


Fig. 10. Outcomes of the imposition of a multi-metric standard.

the system towards that mix. In the case studies, the selection of a metric that focuses on the scenarios with the highest unserved energy, such as $CVaR_{5\,\%}$(UE), leads to a reduction in the installed capacity and de-rating factor of those technologies that perform poorly in these scenarios (CCGTs in this model).

The case studies also show that resource adequacy metrics can be highly correlated. Strong correlations result in a very small multi-activation area. This can significantly reduce the scope of multi-metric reliability standards, as they may often result in the activation of a single standard, with the others being redundant. However, where a multi-activation area exists and is large enough for a set of reliability standards to be active at the same time, multiple de-rating factors can be calculated for each technology, reflecting the contribution of that technology to the different reliability standards imposed simultaneously. In the case studies, this was demonstrated by the calculation of two different de-rating factors corresponding to CCGTs. In the capacity mechanism, this approach should be reflected in the procurement of multiple reliability products (one for each metric included in the multi-metric standard), but this may increase the complexity of the design of the regulatory instrument (multiple capacity requirements, multiple

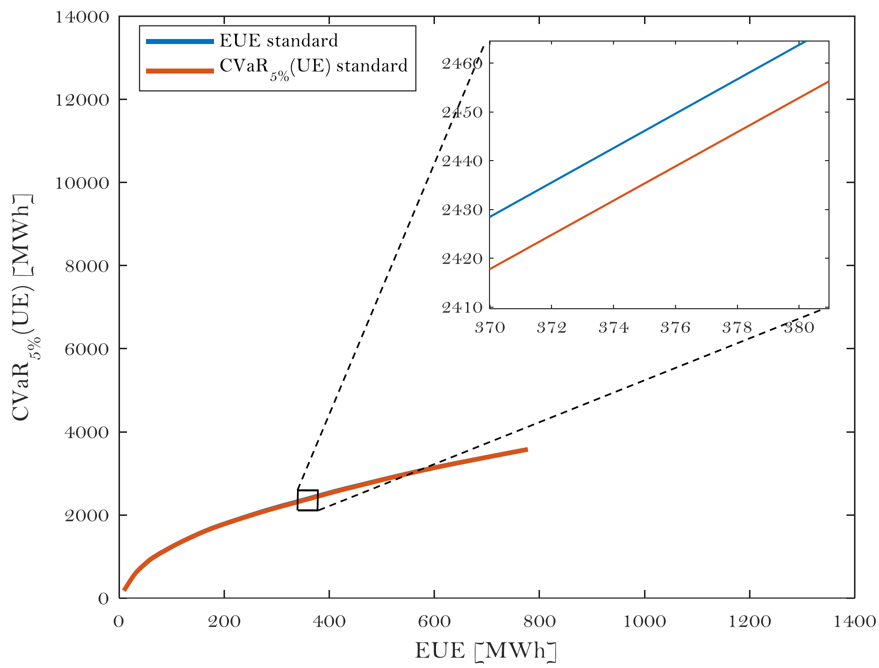


Fig. 11. Characterisation of the multi-metric domain for a case study without polar vortex.

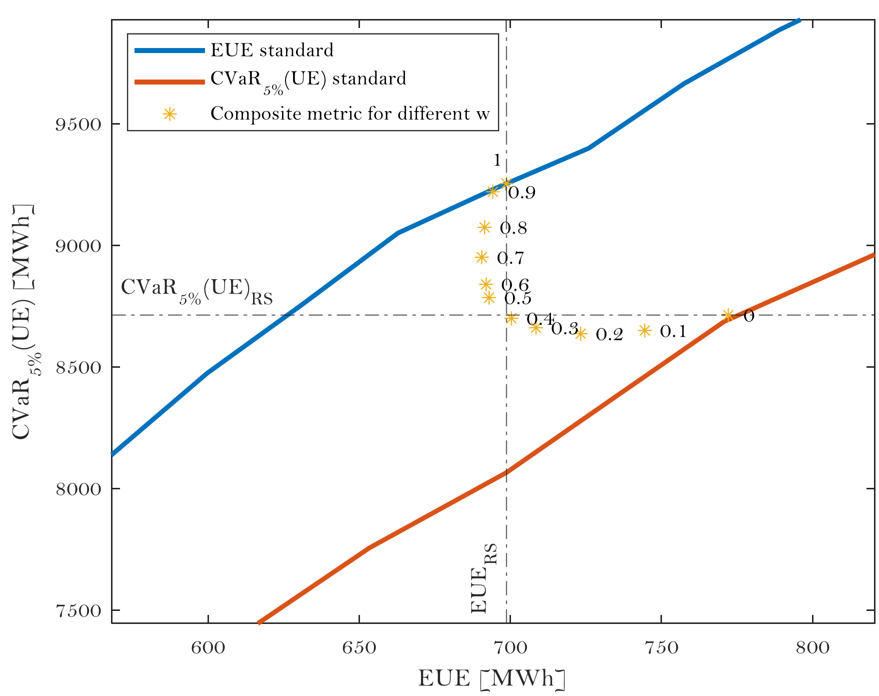


Fig. 12. Outcomes of the imposition of a composite-metric standard represented in the multi-activation area for different values of the weighting factor.

clearing prices, etc.).

Composite-metric reliability standards allow different facets of the security of supply problem to be captured in a single reliability standard by means of weighting factors. This article proposes a normalised formulation for these composite-metric standards, which allows resource adequacy metrics of different orders of magnitude, or even different units of measurement, to be combined. In the case study, this formulation was effective in combining the EUE and $\text{CVaR}_{5\,\%}$(UE) constraints, with the solution, both in terms of installed capacity and de-rating factors, moving between these two extremes for different values of the weighting factor. The advantage of a composite-metric approach is that it allows a single de-rating factor to be calculated for each technology, which could simplify the design of the capacity mechanism. The main findings of this study can therefore be summarised as follows.

- The selection of the resource adequacy metric used to define the reliability standard influences the de-rating factors calculated for each technology.
- Resource adequacy metrics can be highly correlated, which may reduce the scope of multi-metric reliability standards and result in one or more standards becoming redundant.

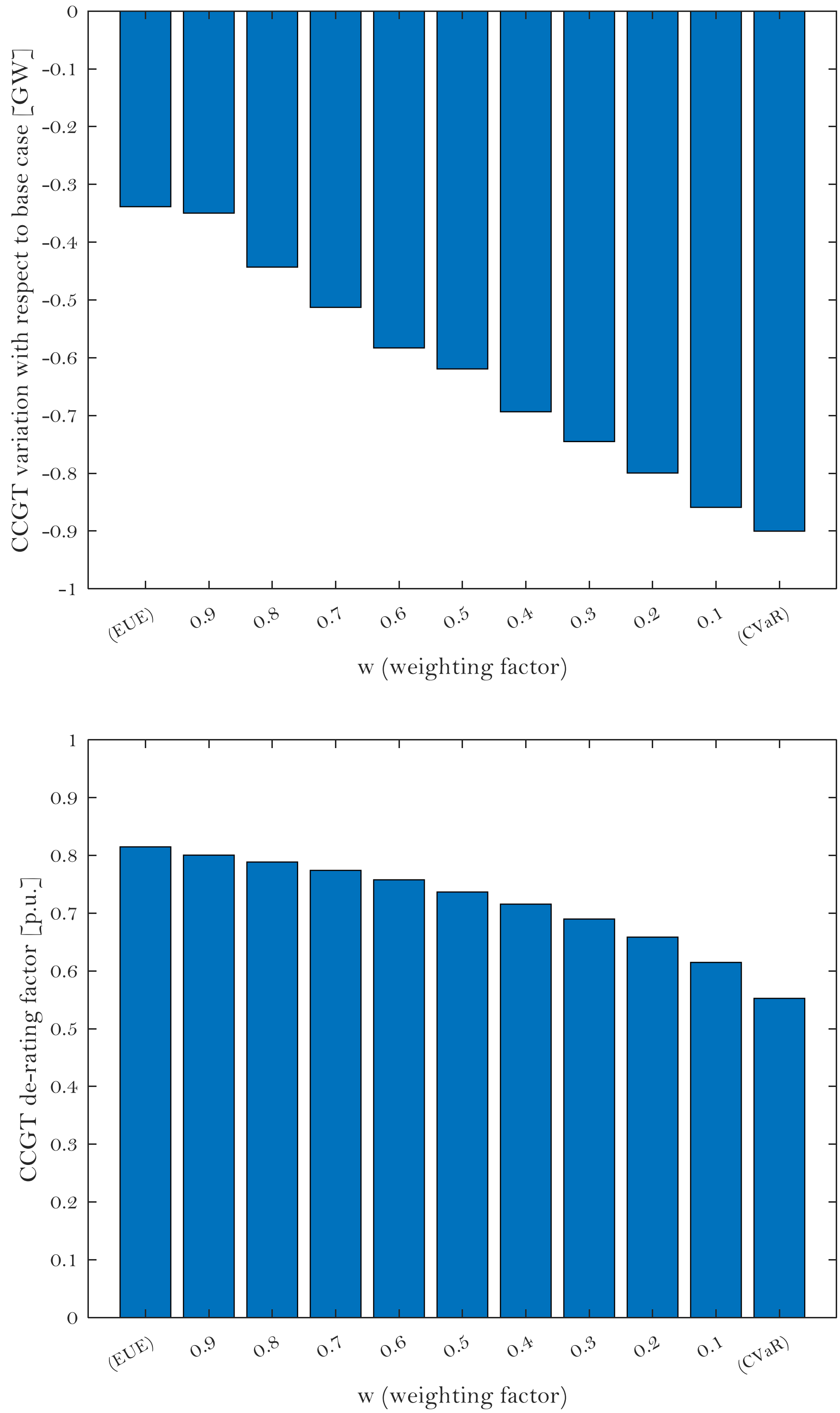


Fig. 13. Outcomes of the imposition of a composite-metric standard on the CCGT technology for different values of the weighting factor.

- When different reliability standards are active at the same time, multiple de-rating factors can be calculated for each technology.
- Composite-metric reliability standards enable different resource adequacy metrics to be merged into a single reliability standard, allowing a single de-rating factor to be calculated for each technology.

4.1. Future work

In this article, the comparison between multi-metric and composite-metric reliability standards has been carried out focusing on two specific resource adequacy metrics, namely EUE and $CVaR_{5\,\%}(UE)$. This selection does not affect the main findings of the article. The existence of a multi-activation area, the need to calculate different de-rating factors when applying multi-metric reliability standards within this area, the possibility of combining the same metric in a composite-metric reliability standard that results in a single de-rating factor are all conclusions that do not depend on the choice the underlying resource adequacy metrics.

However, the same exercise could be repeated with other metrics that may be more or less correlated than those assessed in these case studies. As mentioned in Section 3.3.3, this may affect the shape of the multi-activation area. The latter should also be investigated when more than two resource adequacy metrics are used, adding an additional dimension to the problem and resulting in a multi-activation volume.

Another area for further research is the formulation of the composite-metric reliability standard. The case study presented in this article is based on a normalised formulation, but other solutions are possible. These alternative formulations could be explored and compared in future work.

Future work could also analyse some aspects that could not be modelled in this study due to computational constraints, but which could have a relevant impact on the effectiveness of different approaches to setting reliability standards. These aspects include i) the stochasticity in the load profile or the availability of intermittent renewable resources; ii) other technologies on the supply side, such as hydropower, energy storage, demand response, or prosumers; iii) other reliability stressors or extreme weather events, such as heatwaves or droughts that affect the generation potential of hydropower plants; iv) a larger time horizon for the simulation and time-evolving parameters; v) transmission constraints that can segment scarcity conditions in the system.

CRediT authorship contribution statement

Paulo Brito Pereira: Writing – original draft, Visualization, Validation, Software, Formal analysis, Conceptualization. **Kenneth Bruninx:** Writing – review & editing, Validation, Supervision, Software, Methodology. **Laurens J. de Vries :** Writing – review & editing, Validation, Supervision, Methodology. **Paolo Mastropietro:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Investigation, Conceptualization. **Pablo Rodilla:** Writing – review & editing, Validation, Supervision, Methodology, Conceptualization.

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.

Annex 1

This annex presents the mathematical formulation of stochastic expansion planning model used for the case studies detailed in Section 2.1. The model was solved as an RMIP (Relaxed Mixed Integer Program) problem.

Indexes and sets

$g \in G$Generating technologies
$t \in T$Hourly periods
$sc \in SC$Scenarios

Parameters

C_g^VVariable cost of a unit of technology g [€/MWh]
C_g^{NL}No-load cost of a unit of technology g [€/h·p.u.]
C^{NSE}Non-served energy price (in this case 3000 €/MWh) [€/MWh]
$\overline{P}_g$Maximum power output of a unit of technology g [MW]
$\underline{P}_g$Minimum power output of a unit of technology g [MW]
$AV_{sc,g,t}$Availability of technology g available in period t in scenario sc [p.u.]
$EFOR_g$ Equivalent forced outage rate [p.u.]
AIC_gAnnualised investment cost of units of technology g [M€/p.u.]
D_tDemand in period t [MWh]
LimEUEEUE limit in the system [MWh]
LimCVaR CVaR(UE) limit for the system [MWh]
αPercentile threshold for the CvaR [p.u.]
wWeighting factor for the composite standard [p.u.]

Variables

n_gNumber of units installed of technology g
$nse_{sc,t}$Non-served energy in period t and scenario sc [MWh]
$p_{sc,g,t}$Power output above minimum output of all technology g units in period t and scenario sc [MW]
$u_{sc,g,t}$Number of units of technology g committed in period t and scenario sc
Ω_{sc}Auxiliary variable used to characterize the CvaR (amount of non-served energy in scenario sc above the VaR) [MWh]
VaRValue at risk (if and only if the CVaR constraint is active, else it will not represent the VaR) [MWh]
$nse_{sc,t}, p_{sc,g,t}, VaR, \Omega_{sc} \in \mathbb{R}_{\geq 0}$
$n_g, u_{sc,g,t} \in \mathbb{Z}_{\geq 0}$

Input data

Generation technology technical and economic parameters are presented in Table A-i.

Table A-i
Techno-economic parameters for generation technologies

	Nuclear	CCGT	PV	Wind	Diesel
$\overline{P}_g$ [MW]	1000	1000	1000	1000	1000
$\underline{P}_g$ [MW]	1000	400	0	0	200

(continued on next page)

Table A-i (*continued*)

	Nuclear	CCGT	PV	Wind	Diesel
C_g^V [€/MWh]	6.52	155.6	0	0	710
C_g^{NL} [€/h·p.u.]	0	7200	0	0	4000
$EFOR_g$ [p.u.]	0.05	0.1	0.00	0.00	0.12
AIC_g [M€/p.u.]	626.25	78.57	90.10	128.61	50.77

The four-week demand curve used in the simulation, with each week representing a different season, is presented in Figure A-1.

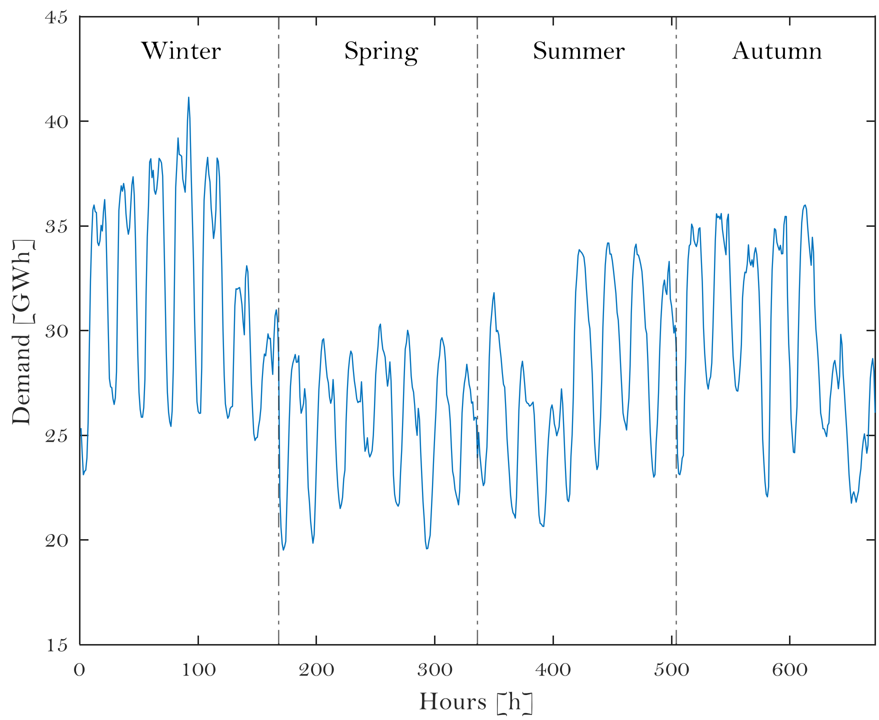


Figure A-1. Demand curve used in the simulation

Formulation

$$\min \sum_{sc} \left[\frac{1}{|SC|} \sum_{t \in T} \left[\sum_{g \in G} \left[C_g^{NL} u_{sc,g,t} + C_g^V \left(\underline{P}_g u_{sc,g,t} + p_{sc,g,t} \right) \right] + C^{NSE} nse_{sc,t} \right] \right] + \sum_{g \in G} n_g AIC_g \tag{1'}$$

s.t.

$$\sum_{g \in G} \left[\underline{P}_g u_{sc,g,t} + p_{sc,g,t} \right] = D_t - nse_{sc,t} \quad \forall sc \in SC,\ \forall t \in T \quad (\lambda_{sc,t}) \tag{2'}$$

$$p_{sc,g,t} \leq \left(\overline{P}_g - \underline{P}_g \right) u_{sc,g,t} \quad \forall sc \in SC,\ \forall g \in G,\ \forall t \in T\ (\overline{\rho}_{sc,g,t}) \tag{3}$$

$$u_{sc,g,t} \leq AV_{sc,g,t} \quad n_g \quad \forall sc \in SC,\ \forall g \in G,\ \forall t \in T\ (\overline{\mu}_{sc,g,t}) \tag{4}$$

$$\sum_{sc \in SC} \left[\frac{1}{|SC|} \sum_{t \in T} nse_{sc,t} \right] \leq \quad EUE_{RS} \quad (\beta_{EUE}) \tag{5}$$

$$VaR + \frac{\sum_{sc \in SC} \left[\frac{\Omega_{sc}}{|SC|} \right]}{\alpha} \leq CVaR_{RS} \quad (\beta_{CVaR(UE)}) \tag{6}$$

$$\Omega_{sc} \geq \sum_{t \in T} [nse_{sc,t}] - VaR \quad \forall sc \in SC\ (\tau_{sc}) \tag{7}$$

$$w \frac{\sum_{sc \in SC} \left[\frac{1}{|SC|} \sum_{t \in T} nse_{sc,t} \right]}{EUE_{RS}} + (1-w) \frac{VaR + \frac{\sum_{sc \in SC} \left[\frac{\Omega_{sc}}{|SC|} \right]}{\alpha}}{CVaR_{RS}} \leq 1 \quad (\beta_{comp}) \tag{8}$$

Formulation explanation

Eq. (1) represents the objective function, which includes both the operational and investment costs of the different resources, as well as the costs of non-served energy, all of which are to be minimised by the model. Eq. (2) represents the demand-generation balance equation for each scenario and time period. The model will attempt to meet demand at each point in time; otherwise, it will incur a penalty represented by the non-served energy (and its associated cost). Eq. (3) limits the power output of a technology above its minimum, depending on the number of units of that technology that have been committed. Eq. (4) limits the number of units that can be committed for a given technology according to its hourly availability for each scenario. For PV and wind technologies, availability is represented according to their ex-ante defined profile. For conventional generation, availability is determined as detailed at the beginning of Section 2. Eq. (5) represents the restriction based on the EUE reliability standard, which limits the amount of non-served energy throughout all scenarios. Eqs. (6) and (7) represent the restriction based on the CVaR(UE) reliability standard, following the formulation presented in [35]. Eq. (8) represents the restriction based on the composite reliability standard, using the normalisation process detailed in Section 2.1.

Models used

Without reliability standards: Eqs. (1)-(4)
With an EUE standard: Eqs. (1)-(5)
With a CVaR(UE) standard: Eqs. (1)-(4), (6) and (7)
With both an EUE and a CVaR(UE) standard: Eqs. (1)-(7)
With a composite EUE-CVaR(UE) standard: Eqs. (1)-(4), (7) and (8)

Annex 2

In this annex, the dual variables from the stochastic centralised expansion problem are used to investigate the revenue shares from the energy and firm capacity markets. The central planner model with resource adequacy constraints is used in this article to simulate an electricity energy market with perfect competition and an additional incentive to install firm capacity. Following the equivalence between these two problems as presented in the academic literature (e.g., Brito-Pereira et al., [17]; Pérez-Arriaga and Meseguer, [28]):

- The energy market price can be derived from the dual variable of the demand balance constraint in the central planner problem ($\lambda_{sc,t}$).
- The resource adequacy remuneration (or capacity market payment, CMP) received by a resource with an installed capacity K_i is equal to its marginal contribution to the reliability standard, $\frac{\partial RS}{\partial K_i}$, multiplied by the value of the dual variable associated with the reliability standard in the central planner problem, β:

$$CMP_{Ki}\ K_i = \frac{\partial RS}{\partial K_i} K_i \bullet \beta$$

By calculating the dual variable(s) of the resource adequacy constraint(s) in the central planner model, it is therefore possible to simulate the hypothetical capacity revenues of different resources and assess their share in total revenues.

This exercise is carried out in this annex for the composite-metric standard, in order also to assess how this share between the energy and firm capacity market revenues varies for different values of the weighting factor (for values of the weighting factor equal to 0 and 1, the results show the revenue shares also for single-metric standards). Figure A-2 shows the revenue shares for CCGT and diesel with a composite-metric standard. Note that the optimal resource mix defined by the model varies for different values of the weighting factor (Fig. 13) and this also affects the marginal energy prices that determine the energy revenues of generating resources.

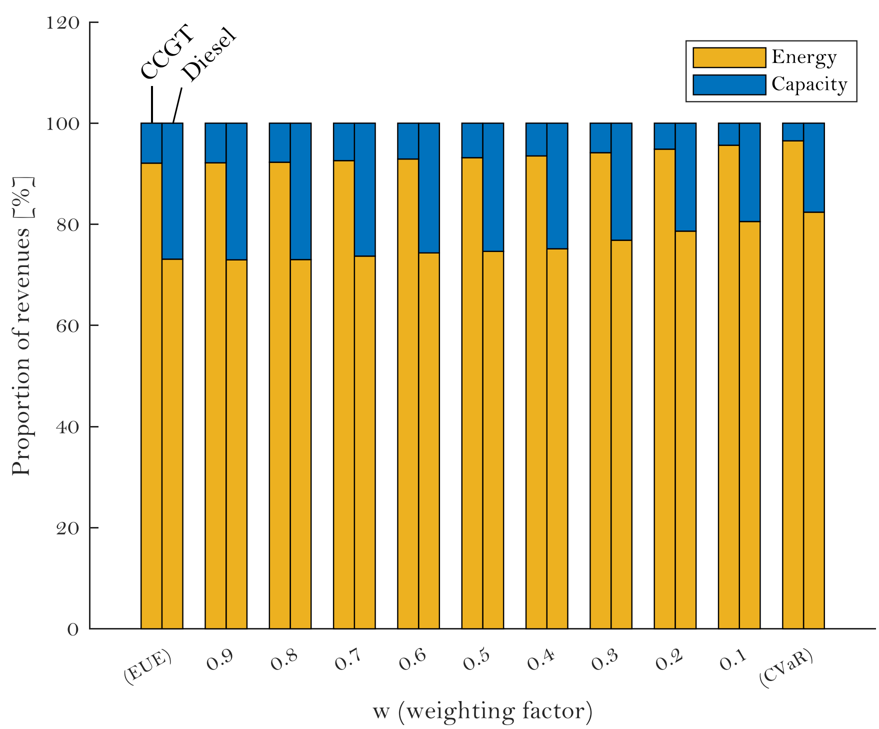


Figure A-2. Revenue shares for CCGT and diesel for different values of the weighting factor

As a general remark, regardless of the value of the weighting factor, diesel resources would receive a higher share of their revenues from the capacity mechanism, consistent with their role as a peaker technology (low fixed costs and high variable costs, with a resulting low load factor). At the same time, the capacity revenues of *both* technologies decrease when the weighting factor decreases, i.e., when the composite-metric standard moves from an EUE to a $CVaR_{5\,\%}$(UE) standard. This is related to the change in the optimal mix mentioned above; at lower values of the weighting factor, more diesel capacity enters the system, increasing the average marginal cost and thus the energy revenues of all resources.

Although there is only a single resource adequacy constraint in the composite metric standard, and therefore a single capacity revenue, it is possible to split this revenue between the contribution of each resource to the EUE and to the $CVaR_{5\,\%}$(UE) metrics. Table A-ii shows the capacity revenues for the whole CCGT and diesel fleets split between the contribution to the EUE and the $CVaR_{5\,\%}$(UE) and how they vary for different values of the weighting factor.

Table A-ii
Capacity revenues for the contribution to each sub-metric within the composite-metric standard

[M€]		w (weighting factor)										
		EUE	0,9	0,8	0,7	0,6	0,5	0,4	0,3	0,2	0,1	CVaR
CCGT	EUE	21,4	20,6	19,4	17,7	16,0	14,3	12,3	9,7	6,9	3,7	0,0
	CVaR(UE)	0,0	0,6	1,3	2,1	2,9	3,8	4,9	5,7	6,6	7,7	9,1
Diesel	EUE	4,0	3,8	3,7	3,5	3,2	2,9	2,5	2,0	1,4	0,8	0,0
	CVaR(UE)	0,0	0,2	0,4	0,6	0,9	1,2	1,5	1,8	2,2	2,5	3,0

Also in this Table the capacity revenues of both technologies are higher when the reliability standard is based on EUE ($w = 1$) than when it is based on $CVaR_{5\,\%}$(UE) ($w = 0$), due to the abovementioned impact of the change in resource mix on average marginal costs and energy revenues. For intermediate values of the weighting factor, the share of capacity revenues relative to the contribution to each metric follows the weight that metric has in the composite-metric reliability standard.

Data availability

Data will be made available on request.

References

[1] L. Hancher, others. Capacity Mechanisms in the EU Energy Markets: Law, Policy, and Economics, second ed., Oxford Academic, 2022 https://doi.org/10.1093/oso/9780192849809.001.0001.
[2] L. De Vries, P. Heijnen, The impact of electricity market design upon investment under uncertainty: the effectiveness of capacity mechanisms, Uti. Policy 16 (iss. 3) (2008) 215–227.
[3] M. Petitet, D. Finon, T. Janssen, Capacity adequacy in power markets facing energy transition: a comparison of scarcity pricing and capacity mechanism, Energy Policy 103 (2017) 30–46.
[4] ESIG, Energy Systems Integration Group, 2021. Redefining Resource Adequacy for Modern Power Systems. Report of the Redefining Resource Adequacy Task Force.
[5] Y. Zhou, P. Mancarella, J. Mutale, Modelling and assessment of the contribution of demand response and electrical energy storage to adequacy of supply, Sustain. Energy Grids Netw. 3 (2015) 12–23.
[6] W.L. Schram, H. Aghaie, I. Lampropoulos, W. van Sark, Insights on the capacity value of photovoltaics, community batteries and electric vehicles, Sustain. Energy Grids Netw. 26 (2021) 100421.
[7] S. Gonzato, K. Bruninx, E. Delarue, The effect of short-term storage operation on resource adequacy, Sustain. Energy Grids Netw. 34 (2023) 101005.
[8] P. De Jong, A.S. Sánchez, K. Esquerre, R.A. Kalid, E.A. Torres, Solar and wind energy production in relation to the electricity load curve and hydroelectricity in the northeast region of Brazil, Renew. Sustain. Energy Rev. 23 (2013) 526–535.
[9] J. Dujardin, A. Kahl, B. Kruyt, S. Bartlett, M. Lehning, Interplay between photovoltaic, wind energy and storage hydropower in a fully renewable Switzerland, Energy 135 (2017) 513–525.
[10] T. Borbáth, D. Van Hertem, Appropriate transmission grid representation for european resource adequacy assessments, Appl. Energy 355 (2024) 122378.
[11] EPRI, Electric Power Research Institute, 2024. Metrics and Criteria - Insights from Case Studies and Recommendations and Considerations for Future Practice. Technical Report.
[12] S. Murphy, L. Lavin, J. Apt, Resource adequacy implications of temperature-dependent electric generator availability, Appl. Energy 262 (2020) 114424.
[13] ESIG, Energy Systems Integration Group, 2024. New Resource Adequacy Criteria for the Energy Transition - Modernizing Reliability Requirements. Report by the ESIG Resource Adequacy Task Force.
[14] EPRI, Electric Power Research Institute, 2022. Resource Adequacy for a Decarbonized Future - A Summary of Existing and Proposed Resource Adequacy Metrics. Technical Report.
[15] L. De Vries, I. Sánchez Jiménez, Market signals as adequacy indicators for future flexible power systems, Oxf. Open Energy 1 (2022) 1–5.
[16] G. Stephen, S.H. Tindemans, J. Fazio, C. Dent, A. Figueroa Acevedo, B. Bagen, A. Crawford, A. Klaube, D. Logan, D. Burke, Clarifying the interpretation and use of the LOLE resource adequacy metric. 2022 17th International Conference on Probabilistic Methods Applied to Power Systems (PMAPS), 2021.
[17] P. Brito-Pereira, P. Mastropietro, P. Rodilla, L.A. Barroso, C. Batlle, Adjusting the aim of capacity mechanisms: Future-Proof reliability metrics and firm supply calculations, Energy Policy 164 (2022) 112891.
[18] EPE, Empresa de Pesquisa Energética, 2020. Critérios de Garantia de Suprimento. Relatório Final do Grupo Temático, Portaria No. 187/2019.
[19] MISO, Midcontinent Independent System Operator, 2024. Exploration of Alternative Resource Adequacy Criteria. Resource Adequacy Subcommittee (RASC-2024-4).
[20] NPCC, Northwest Power and Conservation Council, 2019. Pacific Northwest Power Supply Adequacy Assessment for 2024.
[21] DESNZ, Department for Energy Security & Net Zero, 2023. Exploring Reliability Standard Metrics in a Net Zero Transition. DESNZ research paper No. 2023/050.
[22] AEMC, Australian Energy Market Commission, Reliability Panel, 2022. 2022 Review of the Reliability Standard and Settings. Report..
[23] EPRI, Electric Power Research Institute, Metrics and criteria insights from case studies and recommendations and considerations for future practice, Tech. Rep. (2024) 3002030638.
[24] ESIG, Energy Systems Integration Group, 2023. Ensuring Efficient Reliability - New Design Principles for Capacity Accreditation. Report of the Redefining Resource Adequacy Task Force.
[25] T. Aagaard, A.N. Kleit, Marginal vs. Average effective load carrying capability: how should capacity markets deal with alternative generation forms? Uti. Policy 84 (2023) 101654.
[26] Mancarella, P., 2022. Considerations in Support of the 2022 Reliability Standards and Settings Review. Briefing Note for the Australian Energy Market Commission.
[27] P. Rodilla, P. Mastropietro, T. Gómez, R. Agurto, C. Skerk, C. Batlle, Developing the design of the Chilean power market to allow for efficient deployment of renewables, Oxf. Energy Forum iss. 114 (2018) 19–21.
[28] I.J. Pérez-Arriaga, C. Meseguer, Wholesale marginal prices in competitive generation markets, IEEE Trans. Power Syst. 12 (iss. 2) (1997) 710–717.
[29] Lazard, 2023. Lazard's Levelized Cost of Energy Analysis - Version 16.0.
[30] P. Mastropietro, I. Herrero, P. Rodilla, C. Batlle, A model-based analysis on the impact of explicit penalty schemes in capacity mechanisms, Appl. Energy 168 (2016) 406–417.
[31] A. Lisnianski, D. Elmakias, D. Laredo, H. Ben Haim, A Multi-state Markov model for a Short-term reliability analysis of a power generating unit, Reliab. Eng. Syst. Saf. 98 (iss. 1) (2012) 1–6.
[32] J.B. Busby, K. Baker, M.D. Bazilian, A.Q. Gilbert, E. Grubert, V. Rai, J.D. Rhodes, S. Shidore, C.A. Smith, M.E. Webber, Cascading Risks Understanding 2021 Winter Blackout Texas, Energy Res. Soc. Sci. 77 (2021) 102106.
[33] NERC, North American Electric Reliability Corporation, 2024. Polar Vortex Review. Report.
[34] ERAWA, Economic Regulation Authority Western Australia, 2024. Procedure Change Proposal: Benchmark Reserve Capacity Price. Working document EEPC_2024_01.
[35] S. Uryasev, Conditional value-at-risk: optimization algorithms and applications. Proceedings of the IEEE/IAFE/INFORMS 2000 Conference on Computational Intelligence for Financial Engineering (CIFEr) (Cat. No.00TH8520), New York, NY, USA, 2000, pp. 49–57.