UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

OFFICIAL MASTER'S DEGREE IN THE ELECTRIC POWER INDUSTRY

Master's Thesis

Demand Response integration in Capacity Remuneration Mechanisms: Firm Supply and Cost Allocation

Author: René Inés Villarreal del Ángel
Supervisor: Pablo Rodilla Rodríguez
Co-Supervisor: Paulo Brito Pereira

Madrid, July 2023

Master's Thesis Presentation Authorization

THE STUDENT:

René Inés Villarreal del Ángel

....

THE SUPERVISOR

Pablo Rodilla Rodríguez

Signed: .... Date: ....../ ....../ ......

THE CO-SUPERVISOR

Paulo Brito Pereira

Firmado por BRITO PEREIRA PAULO - ***4898** el día 19/07/2023 con un certificado emitido

Signed: .... Date: ....../ ....../ ......

Authorization of the Master's Thesis Coordinator

Dr. Luis Olmos Camacho

Signed: .... Date: ....../ ....../ ......

UNIVERSIDAD PONTIFICIA COMILLAS

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA (ICAI)

OFFICIAL MASTER'S DEGREE IN THE ELECTRIC POWER INDUSTRY

Master's Thesis

Demand Response integration in Capacity Remuneration Mechanisms: Firm Supply and Cost Allocation

Author: René Inés Villarreal del Ángel
Supervisor: Pablo Rodilla Rodríguez
Co-Supervisor: Paulo Brito Pereira

Madrid, July 2023

Summary

This work seeks to solve how Demand Response may be integrated into Capacity Mechanisms in a way that allows its participation to be efficient and transparent. This means that the design elements of the Capacity Mechanism should strive for technology blindness, so that Demand Response can compete on equal terms with other sources of reliability for the system. Key to achieve this is to define how much Demand Response is contributing (Firm Supply) and how to assign the share of the costs of the program to it (Cost Allocation).

Capacity remuneration mechanisms (CRMs) are now considered a mainstay in liberalized electricity markets, considering they attempt to solve two market failures: consumers are usually highly inelastic in the short term and the regulators tend to limit high prices, even if they appear shortly and serve to signal scarcity. This creates a missing money problem from the point of view of producers, who see their revenues limited. Additionally, there appears to be low liquidity for energy in the long-term from consumers which in turn results in a missing market problem.

Power system reliability is at the front of the regulatory debate, especially in North America & Europe, where their markets are mostly capacity-limited (this wouldn't be the case in South America). The EU has put forward a process for Member States to implement Capacity Markets, albeit with caveats, while investors claim that the extra revenue stream will be more and more essential as time goes by and the system penetration of intermittent sources turns higher.

A couple steps are taken to develop a methodology that allows for an effective implementation of Demand Response resources in Capacity Markets, paving the way for enough transparency so that market agents are prevented from tampering with the program as a form of arbitration. First, a big picture view of programs across North America & Europe reveals that the regulatory designs are ever-changing, with the most experienced, PJM, tweaking its design for over 20 years. Second, a theoretical analysis is carried through describing the different ways that Demand Response resources can participate, for example, should they do so on the demand or supply side of the market? Finally, a theoretical link is proposed between Demand Response's effect on a reliability index and the determination of both Firm Supply and Cost Allocation.

In order to apply these concepts on a practical case example, a convolution model is used to estimate the changes on the reliability of a power system by altering the overall demand. This model takes into account different demand profiles and serves the purpose of determining their corresponding Cost Allocation as well as their Firm Supply according to the way Demand Response is applied. Given the methodology described in the theoretical framework, the link is found between the reliability index and the determination of Cost Allocation and Firm Supply. The conclusion drawn is that with this proposal a market agent would be subject to the same revenues and/or costs regardless of whether they are participating in the Capacity Market as Demand Response or any other source for reliability and therefore the proposal avoids any undesired arbitration.

Foreword

This work would not have been possible without:

- My mother’s loving patience
- My father’s firm ideals
- My sister’s joy

Table of contents

Chapter 1: Introduction

This work will seek to establish why Capacity Mechanisms are at the front of the regulatory debate in the power sector in developed countries and how they strive to increase the reliability of power systems. This is especially in focus as the percentage of intermittent sources becomes ever more present and important for the generation mix in both North America & Europe.

Next the discussion will focus on the design details on which the success of the Capacity Mechanisms lies. Key among them will be for the mechanism to be applicable to a variety of technologies so that any source can participate on equal grounding. Even more so, if it includes not only generation sources but also demand, which can contribute to this by reducing its consumption in key moments. This is usually called Demand Response.

Demand Response fits neatly into overall goals regarding the power sector:

1. Environment: In principle, Demand Response would reduce the need for installing more thermal facilities.
2. Structure: Traditionally a top-down sector, Demand Response would empower consumers to participate actively in the market.
3. Timing: As new installations of generation sources tend to be intermittent; Demand Response would make the system more flexible.

Demand Response would however raise the question of how to determine its net contribution in contrast to a generation source and how to decide the share from the cost it should pay into the mechanism.

Therefore, a review will be presented of how these systems have been implemented across North America & Europe, while also delving deeper into those systems which have had the most experience. It will become apparent that the design of Capacity Mechanisms that also include Demand Response is not purely theoretical. On the contrary, those Capacity Mechanisms that have included Demand Response as a valid source have been changing their design for over 20 years and therefore it might be safe to assume that the design relies strongly on an element of trial and error from both the regulator's and the market agents' points of view.

Nevertheless, a theoretical discussion will go through the defining elements of how the Demand is to participate, by its own choice or by the regulator's. Special focus will be given to two key issues, mentioned briefly before but now more precisely names: Firm Supply (the aforementioned net contribution of Demand Response) and Cost Allocation (previously mentioned as share of the total cost). It is important to mention that the total cost will not be determined as part of this work, Cost Allocation for all purposes will focus only on the share it should be assigned, regardless of what the total cost of the Capacity Mechanism is.

Afterwards, turning towards a practical case example, a convolution model will be implemented in order to determine the effect that different cases can have on a reliability metric. The cases will consist of different demand profiles coming from different kinds of consumers and a couple scenarios of much solar power penetration there will be in the system. Both will be applied in order to have around 8 possibilities of results for Firm Supply and Cost Allocation.

The results will be presented in a way that takes the reader through the step-by-step processing of data as well as the final formulas which will link the reliability metric with the results. The outcome will point out that the marginal contribution to the reliability metric should be the defining aspect to determining the Firm Supply and Cost Allocation.

Finally, conclusions will be drawn as to the overall relevancy of Capacity Markets, the role of Demand Response and how theoretically and practically its particularities of determining Firm Supply and Cost Allocation might be solved.

Chapter 2: Theoretical Framework

2.1 Capacity Mechanisms

2.1.1 Justification

As mercantilism faded after Adam Smith's foundation of modern economics, liberalized markets developed across the Western world during the 1800s (HUMP98). Passing through their Keynesian phase and then back to Monetarists, they found its application on electricity markets with Milton Friedman's Spanish-speaking students at Chicago University, usually called the "Chicago Boys" (KLEI07).

This influenced the market design adopted in the electricity sector across North America & Europe. The price, based on microeconomics principles, was to be set by the point where the marginal costs of producers (supply) and the marginal utility of consumers (demand) met each other:

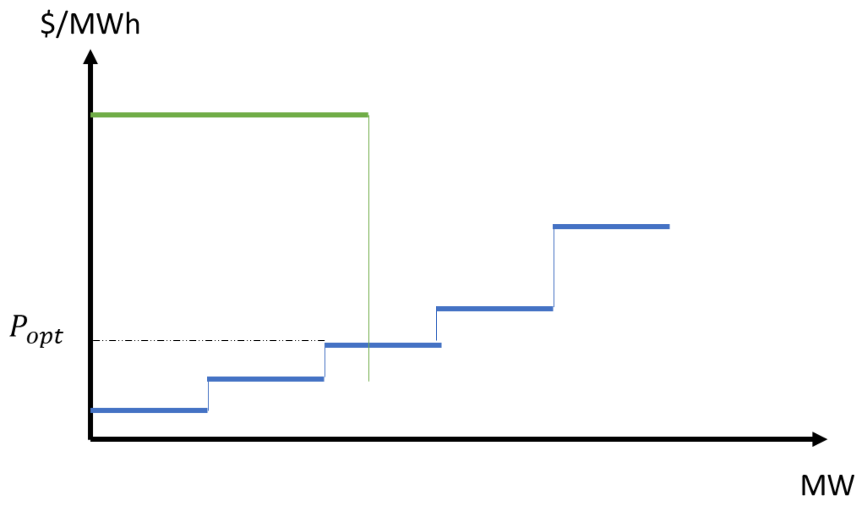


Figure 1, Price setting (own elaboration)

Producers build their selling bids by submitting their marginal costs (represented in Figure 1 in blue). They are then arranged from lower to higher in order to incentivize them to strive for efficiency in their operation. Since most consumers don't react directly to wholesale prices, their retailers usually submit buying bids (represented in Figure 1 in green) based on the monetary damage it would represent for consumers not to have electricity, also called the Value of Lost Load.

The recovery of capital costs is supposed to come from periods where prices rise above the short-run marginal costs. As demand increases, prices adjust to avoid shortages. Two factors create a fundamental challenge: the inelasticity of short-term demand and cap limits on high energy prices. The first stems from the fact the aforementioned reality that most consumers do not change their behavior based on the prices of the wholesale market. There usually are limited signals and/or penalties for consumption on critical moments, even if the cost of the system as a whole increases disproportionally. The second results from the willingness of regulators to protect consumers from high prices they may deem unacceptable for them.

The combination of these two factors contributes to producers not getting high enough prices, if even sporadically, to recover their investment costs. This has been called the missing money problem (BUSH17).

The missing money problem might even result in producers investing less in generation that would be used during critical moments, since it has been shown to them that the market won't allow them to collect high enough prices. The regulator's intervention to cap prices would actually result in less investment and, most probably, higher scarcity situations in the future.

There are two main paths to combat this. A preventive one would be to go back and improve price formation in short-term markets to begin with; the second is a corrective one and would add a new revenue stream for producers in the form of a Capacity Remuneration Mechanism (BOTT20). This, in turn, will create a market that was missing.

All the way back to the Chicago Boys in Chile during the 1980s the ability of short-term marginal prices to provide sufficient incentives for investment in generation was put into question. Most Latin American countries (with the exception of Brazil) and some in northeastern United States (like PJM), introduced in their market designs some sort of complementary market for capacity (BATL22).

Capacity Mechanisms can be either based on quantity or price. Quantity-based mechanisms define a desired quantity of capacity and then let the market decide the price of the transaction. On the other hand, Price-based mechanisms define a price and then expect the market to provide the desired quantity:

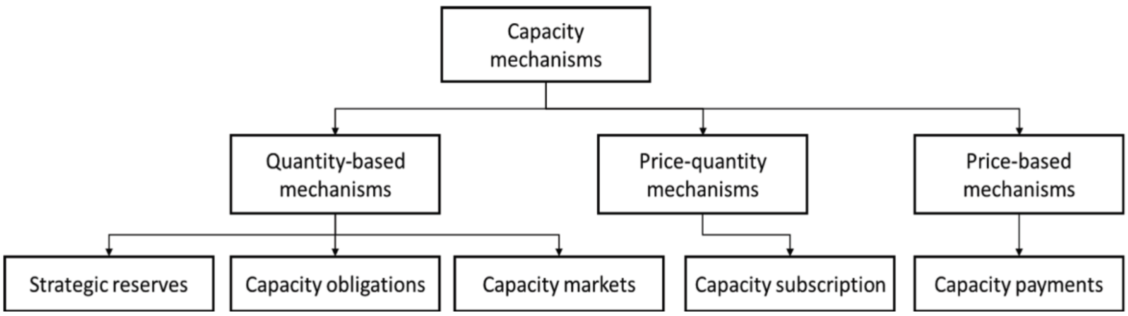


Figure 2, Capacity Remuneration Mechanism types (IAEE)

In order to avoid a step back in the process of liberalization, a market-oriented mechanism with limited intervention is preferred, resulting in Capacity Markets having been one of the most developed among them.

Moreover, Capacity Markets may be designed to envision not only generation facilities, but also alternative resources such Demand Response, or DR, which the Federal Energy Regulatory Commission, or FERC, defines as:

> "changes in electric usage by demand-side resources from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive

payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized" (FERC18).

In integrating Demand Response to Capacity Markets, there are two key aspects to define:

1. How Firm Supply is determined:
   That is, the difference between what was expected of the user to consume (usually called the Baseline) and what was actually consumed at the moment the Demand Response was enforced, which should show an overall reduction.
2. How Cost Allocation should be determined:
   According to regulatory theory, Firm Supply for generation resources will be directly proportional to their generation during scarcity conditions. This regulatory principle could be translated to determine an efficient Cost Allocation methodology for the Capacity Market.

2.1.2 Current regulatory debate

The current regulatory discussion for Capacity Mechanisms) at the European level is set forward in the Internal Market for Electricity Regulation 2019/943 (EURO19), specifically in Chapter IV Resource Adequacy. Its main takeaways are:

- Article 20: Resource adequacy in the internal market for electricity
  - Member States shall monitor resource adequacy concerns.
  - The Member State shall identify any regulatory distortions or market failures.
  - Member States shall develop and publish an implementation plan as a part of the State Aid process.
  - The Member States concerned shall submit their implementation plans to the Commission for review.
- Article 21: General principles for capacity mechanisms
  - To eliminate residual resource adequacy concerns, Member States may, as a last resort, introduce capacity mechanisms.
  - The Member States concerned shall conduct a comprehensive study of the possible effects of such mechanisms on the neighboring Member States.
  - Member States shall assess whether a strategic reserve is capable of addressing the resource adequacy concerns.
  - When designing capacity mechanisms Member States shall include a provision allowing for an efficient administrative phase-out.
  - Capacity mechanisms shall be temporary. They shall be approved by the Commission for no longer than 10 years.
- Article 22: Design principles for capacity mechanisms

- Select capacity providers by means of a transparent, non-discriminatory, and competitive process.
- Provide incentives for capacity providers to be available in times of expected system stress.
- Be open to participation of all resources that are capable of providing the required technical performance, including energy storage and demand side management.
- Apply appropriate penalties to capacity providers that are not available in times of system stress.

There seems to be differences between the points of view of academics and the input from market agents. They have also been identified by a few analysts from the Florence School of Regulation within the European University Institute. Some of their arguments are:

I) Adequacy and Capacity Remuneration Mechanisms (POTO22)

On the one side, investors in the sector claim that Capacity Markets will be more and more essential going forward, to provide investors with the stability in their revenue streams that a more binomial distribution of short-term electricity prices threatens. On the other camp, the European Commission among others, claims that a fully functioning short-term energy market, providing efficient investment signals, is the best means to ensure resource adequacy and security of electricity supply.

A composition of this debate might point to the fact that, while, in theory, markets can always clear (at the value of lost load in case of scarcity), this might happen at prices which are so high to be socially disruptive and politically unacceptable. To avoid or minimize the probability of the market clearing at such prices, Capacity Markets might be implemented, not only as a transitory solution.

II) Recent energy price dynamics and market enhancements for the future energy transition (ROBE22)

Investors may be worried that sudden increases in prices, reaching levels which are politically too high, might attract government attention, with the risk of interventions to introduce caps or take the higher profits earned in these hours away. A hedging requirement for suppliers might promote the development of financial hedging instruments, which could also provide a more stable stream of revenue for operators of resources selling electricity on the market and therefore reduce investment risk

In this respect, EU legislation allows Member States to introduce capacity remuneration mechanisms (CRMs) but only as a last resort, to address residual resource adequacy concerns and as a temporary remedy while any identified regulatory distortions or market failures are addressed.

2.2 Demand Response Integration

2.2.1 Participation

The challenge at hand is that demand is to cover the cost incurred while implementing a Capacity Market design, whose goal is to provide reliable supply during scarcity events, while at the same time demand should also be able to strive to participate to lower the cost of the mechanism.

This participation would entail demand either explicitly or implicitly participating in a capacity mechanism auction. Therefore, a method is to be defined to determine how much Firm Capacity (or, Supply) is being provided and how to compute the distribution of the Cost Allocation (RODI23).

Firm Supply is the expected contribution of a resource while scarcity conditions are present in the system. The focus will be in Capacity since that is the sort of constraint that is present in North America & Europe (this wouldn't be the case in South America).

Cost Allocation is the other design element that deeply affects demand participation. In principle, it is the methodology by which the costs of the Capacity Market are recovered. This should come from the signature of capacity contracts between consumers and reliability providers.

Now, delving into how the demand for Firm Supply is to be determined, current systems tend to set the demand for Firm Supply with no participation from their consumers. The demand curve and its elasticity are defined administratively. In order to set the appropriate signals from the beginning, the consumers should participate setting their demand quantity and elasticity.

Theoretically, a self-declared demand for Firm Supply would be the equivalent to the de-rating process commonly applied to generation sources and would consequently be the best cost driver on which to apply Capacity Market charges. Each consumer would pay the costs of the Capacity Market according to its "negative" contribution to the reliability of the system (considering that buying 1 MW of Firm Supply is equivalent to providing negatively).

Nevertheless, in most Capacity Markets, the demand for Firm Supply is estimated in an aggregated way. For example, the demand curve in capacity auctions is usually defined through an estimation of the entire system demand and its expected evolution in the future. This approach constrains the participation of demand resources, giving into the missing market problem.

As for the Cost Allocation, once the aforementioned capacity contracts are signed, for all intents and purposes, their costs should be considered sunk costs. Therefore, these sunk costs should be allocated in a way that distorts the market the least, which is why fixed costs are the preferred solution.

Theoretically, such fixed costs should reflect the contribution by each consumer to the demand for Firm Supply of the system. In the absence of such information, the charge could be proportionally set to the historical consumption of each consumer during scarcity conditions. Even a moving average of a certain number of years could be implemented.

Nevertheless, international practice favors simple volumetric charges applied over a large number of hours. This is equivalent to socializing the sunk costs without providing any efficient signal to consumers and without guaranteeing cost recovery, either.

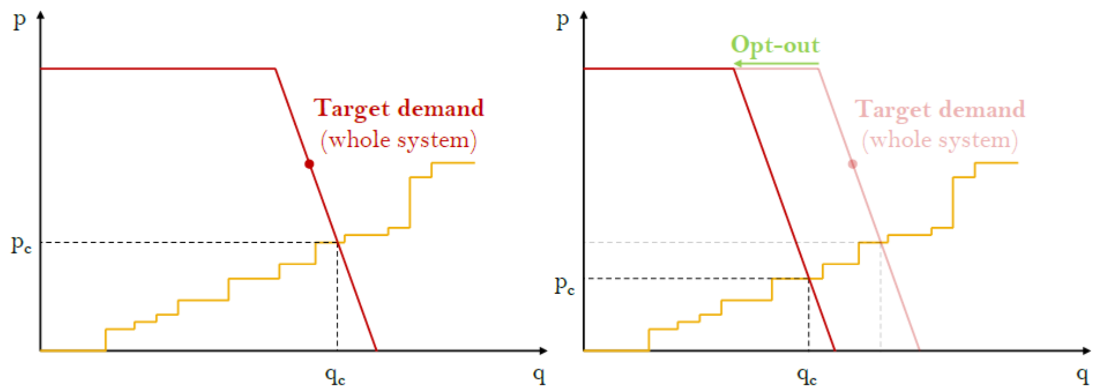
Regarding how demand is to participate in general, there are two broad categories, explicit and implicit participation. Implicit participation depends on the design of the Cost Allocation. Since most systems use volumetric charges instead of fixed ones this participation scheme is almost non-existent. Explicit participation requires consumers to take part in some phase of the capacity market and assume binding commitments. This chapter will focus on such schemes.

The most obvious way to involve consumers in Capacity Market would be for them to participate on the demand side. That is, they would actively define the capacity they expect to demand during scarcity conditions. This would function as a limit to actual consumption during such events as it would be a commitment not to exceed what they previously defined.

This would move the responsibility of defining the demand for Firm Supply from the regulator to the consumer's shoulders. However, it may prove difficult to implement, in both a regulatory and political point of view.

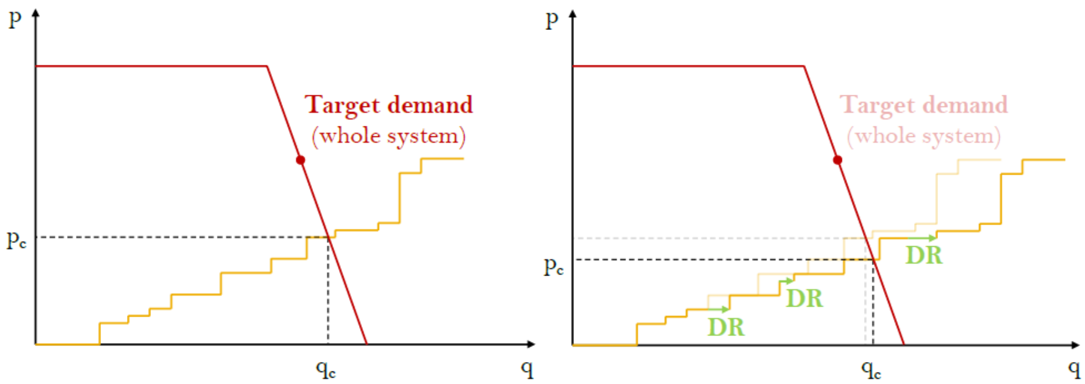
A first alternative would be for the regulator or system operator to estimate an aggregated demand for Firm Supply and then compute a disaggregated estimation for consumer categories (perhaps following tariffs). Then, the consumers would be given the chance to opt-out.

The opt-out would generate a commitment that allows the system operator to limit consumption during scarcity conditions. However, it would also exempt the consumer from paying Capacity Market charges for the opted-out capacity. Furthermore, it would also simplify the Cost Allocation of the Capacity Market, since every consumer would have a certain demand for Firm Supply defined for them.



What is actually more common is for the demand to be defined for the entire system with no opting-out allowed. However, consumers could still participate in the Capacity Market by offering DR. In order for this to be feasible, the regulator would have to design a reliability product such that the resources traded it within the capacity market. In principle, the product should be technology blind. That is, resources should compete on the same playing field with Demand Response being just one among all the reliability options.

Another way for demand to participate in the market is on the supply side. This is done through price-quantity bids that go into the supply curve of the market. However, the consumers involved would be represented twice in the auction; once in the demand curve (as part of the demand of the system as a whole) and another on the supply curve. This is susceptible to arbitrage.



In theory, the regulator should design a reliability product that is the same for all resources competing in the Capacity Market. However, many regulators in both North America & Europe have designed specific products for the characteristics of Demand Response resources.

Delving into the supply-side issue, there's also the potential for double remuneration. Once a consumer participates on both the demand and supply curves it will be remunerated once for reducing its load during scarcity conditions and twice by reducing its contribution to the share of the cost it should pay into the mechanism. Even more important, its net position in the Capacity Market could result in them getting a net revenue from its participation.

However, since the final goal of the agent is to avoid the payment of the Capacity Market charges, without benefitting from the coverage of the mechanism, its net position should be zero at most. The service provided by a Demand Response resource cannot be provided to a third party, in contrast to the service provided by a generator, whose net position is evidently expected to be larger than zero.

Paradoxically, the double-remuneration problem has been usually avoided thanks to the inefficient volumetric charges which were previously mentioned. When these volumetric charges cover a very large number of hours, they simultaneously reduce the benefit that can be achieved by Demand Response activating during scarcity conditions. However, Cost Allocation based on capacity chargers during scarcity conditions could increase the risk of a double remuneration. The most efficient way of dealing with this issue is again by introducing fixed charges, based for example on historical consumption. This way, Demand Response services would pay a fixed amount and could only offset this quantity by the Capacity Market revenues, which if every design element is harmonized, should result in a net position which approaches but never reaches zero.

2.2.2 International experience

United States

Now, let's look at the state of art where this is actually implemented:

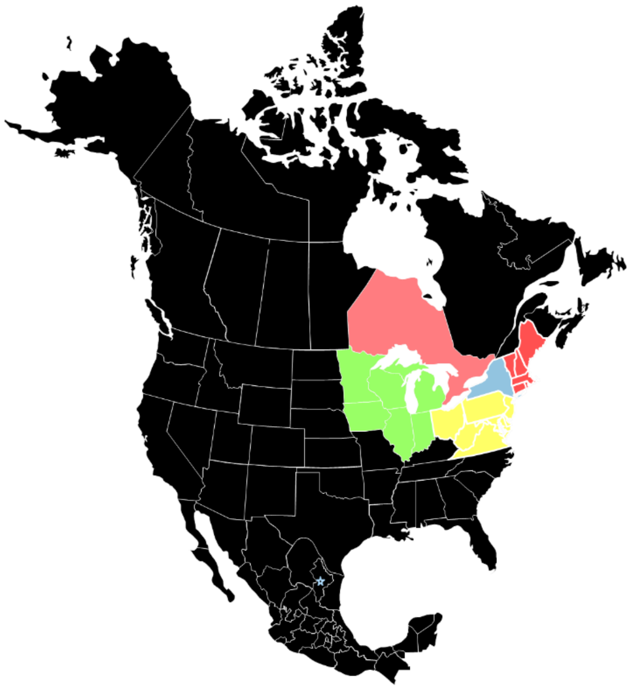


Figure 3 Main markets in North America with Demand Response in their Capacity Markets

PJM

One of the most experienced markets in Capacity Markets is PJM in northeastern USA (marked yellow in Figure 3). They have integrated Demand Response since at least 1999 and so a look will be taken at their design (PJML23).

To start, they define two periods per year:

- Capacity Performance: Year-round with different schedules, one for November-April and another one for June-October and the following May.
- Summer Period: Seasonal with hours defined just for June-October and the following May.

There is no limit to the duration of the interruption on either of them. The hours are defined as 10 AM to 10 PM for June-October and May, while they are 6 AM to 9 PM for November-April.

Then, PJM defines two types of measurement:

- Firm Service Level (FSL) – achieved by a customer reducing its load to a pre-determined level upon notification from the market operation's center.
    - The customer must be able to reduce its load below the pre-determined level which must be lower than the amount of capacity reserve for the customer as represented by the peak load contribution ("PLC") in summer and winter peak load ("WPL").

- Guaranteed Load Drop (GLD) – achieved by a customer reducing its load below the PLC in summer and WPL when compared to what the load would have been absent of the PJM emergency or test

Now let's look at how the Actual Load Reduction is determined, taking into account both the period and the type of measurement:

	FSL	GLD
June-October and May	PLC - Metered load * Loss factor	MIN(Comparison load - Metered load * Loss factor ; PLC - Metered load * Loss factor)
November-April	WPL * Weather Adjustment * Loss factor - Metered load * Loss factor	MIN(Comparison load - Metered load * Loss factor ; WPL * Weather Adjustment * Loss factor - Metered load * Loss factor)

Some notes:

- ✓ Summer load reductions are only recognized if:
  Metered load * Loss factor < PLC
- ✓ Non-summer load reductions are only recognized if:
  Metered load * Loss factor < WPL * Weather adjustment * Loss factor

PJM assures that the method to be used should result in the best possible estimate of what load would have occurred in the absence of an emergency or test event.

The minimum amount of data to be required will be 24 hours for one full calendar day.

As for the Comparison Load, there are two main possibilities to define it:

1. Compare Day:
    - The customer's actual hourly loads on one of the prior 10 calendar days before the test or emergency event day selected which best represents what the load level would have been absent the emergency or test event.

    - PJM must approve the use of any alternative day
2. Same Day:
    - The customer's average hourly integrated consumption for two full hours prior to notification of an emergency event or prior to one full hour before a test and for two full hours after skipping first full hour after the event or test.
    - This option is appropriate for high load factor customers with no weather sensitivity.

In order to also get the perspective from a market agent, a report from Enel X provides the following feedback (ENEL23):

- It considers the inclusion of winter months a direct response to an extreme weather event in 2014 called the Polar Vortex.
- It deems the participation registration too complex and suggests installing metering devices to monitor consumption in real time.
- It recommends focusing reductions of consumption on non-essential lighting, manufacturing processes, HVAC equipment, pumps, and industrial freezers.
- It reports that PJM has called around 1-2 dispatches per year in any given zone. If no event is called, PJM will schedule a two-hour audit.
- It concludes that the relatively low incidence is attributable to PJM's healthy reserve margin.

NY ISO

Next on the list is the NY ISO (marked blue in Figure 3) they have designed the scheme called Installed Capacity – Special Case Resource (ICAP-SCR), which among their Demand Response programs is the only one that allows for remuneration for capacity (NYSO20).

The minimum requirement for participation is 100 kW of net load reduction, although it is allowed to aggregate multiple resources if and only if they are located within the same Load Zone and the total load reduction reaches the aforementioned minimum of 100 kW. NY ISO manages two types of baselines (NYSO18):

- Average Coincident Load (ACL)
    - May be provisional
    - Always subject to changes
- Customer Baseline Load (CBL)
    - Declared value

The one that is actually used to calculate the SCR capacity is ACL. It used the average of the highest 20 resource loads that occurred during the Capability Period SCR Load Zone Peak Hours. Any load supported by generation produced from a Local Generator may not be included in the SCR's metered Load Values reported for the ACL.

A provisional value may be used when the resource was not previously enrolled in the ICAP-SCR program and/or when they did not have meter data from the Prior Equivalent Capability Period. The changes that ACL is subject to are justified in either an increase or a decrease during a specific month. The change must be equal or greater than 30% or 5 MW in New York City or 10 MW in any other zone.

The declared value used in CBL is identified upon initial enrollment and represents the amount of capacity the SCR could make available. It cannot be greater than the resource's Net Average Coincident Load.

The settlement is based upon a calculated ACL and closed in an auction environment where the Strike Price is determined by aggregation. The offer may not exceed 500 $/MWh, is set for the entire month and is mandatory when the SCR supplies capacity.

Last on the North American side is ISO NE (marked red in Figure 3) which used to have separate systems for Real-time Demand Response and Real-time Emergency Generation Resource, but is in the process of integrating both into a unique Active Demand Capacity Resource (ADCR) which allows for capacity participation (SONE17).

The minimum requirements are to have an interruptible capacity of at least 10 kW, although multiple facilities may be aggregated to achieve this goal. These may be called Demand Response Assets (or DRAs) unless they reach 5 MW, in which case they turn into Demand Response Resources (or DRRs). New England is split into 20 DRR aggregation zones.

The Unadjusted Baseline (UBL) is calculated for each DRA every day. The current methodology considers:

- Non-holiday weekdays:
    - Ten-day average taken from past 30 non-holiday weekdays
    - Most recent non-performance days used first, then dispatch days, then curtailment days if needed
- Two additional day types, Saturday, and Sunday/Holiday
    - 5-day average, maximum lookback of 42 calendar days
    - Non-performance days prioritized over performance or curtailment days
- Baseline adjustment period shortened to three intervals that ended prior to the dispatch instruction, unless they were part of a prior dispatch

ISO NE

ISO NE establishes that the market participant shall not take any action to create or maintain a Demand Response Baseline that exceeds the typical electricity consumption levels of its end-use metered customers expected in the normal course of business. This seems like something its design should avoid, but it is mentioned, nonetheless.

Europe

Crossing over to the other side of the Atlantic, some systems have Demand Response participation too. Some with a positive experience (France and Belgium) and another one with a negative one (Italy).

France

France has the power system which has probably attracted the most Demand Response at the European level. This is mainly due to a market segmentation; the decentralized mechanism for capacity obligations designed in 2016 was complemented in 2018 by an auction specifically designed for Demand Response. The goal of the auction was to define a variable premium which is then paid to the Demand Response resources on top of the market price for capacity.

As reference, the French market price for capacity in 2021 was 31 €/kW-year, while the auction closed at 55 €/kW-year. In 2022 almost 2500 MW of Demand Response were purchased. The creation of a specific product for Demand Response and its segregation from the rest of the market for capacity has doubtlessly allowed for higher participation. However, this segregation is not the optimal solution from a theoretical point of view because it blocks direct competition among capacity resources.

Belgium

Belgium recently introduced a capacity mechanism based on reliability options, which will begin starting on 2025. These are basically financial options which will clear according to the Day-Ahad Market with a strike price that in 2021 was estimated to be 300 €/MWh.

The Belgian mechanism is open to Demand Response participation and the minimum capacity is 1 MW, with aggregation being allowed as well. In order to facilitate the risk management involved in Demand Response, the regulator decided that Demand Response resources could determine their own strike price, which in turn would be determined by their activation cost. This possibility evidently gives preference to Demand Response resources (287 MW were cleared in the first auction). Again, regulatory theory maintains that all resources should provide the same product. By allowing Demand Response to offer a lower strike price, its service is deemed lower in value although it does compete in equal conditions.

Italy

Italy has a capacity mechanism which is also based on reliability options. Demand Response resources with over 1 MW available may participate, but with very different rules than those that apply for regular resources. Specifically, in order to avoid double remuneration, Demand Response resources are subject to the following economic conditions:

- Not receiving the resulting premium from the auction to supply capacity, but being exempt from the payment of recovery costs of the mechanism
- Not receiving the difference between the market and strike prices, but not having to return it either as Demand Response resources
- Having to be disconnected by the system operator during scarcity conditions

These conditions have some grounding on the theoretical need to avoid Demand Response resources from obtaining a net positive remuneration from their contribution. However, the absence of a direct remuneration has failed to incentivize market agents and no Demand Response resource has been cleared in the first two auctions of the Capacity Market.

2.3 Theoretical considerations

2.3.1 Baselining

The first item is to define a Baseline, which represents the "business as usual" or what a consumer's load would have been if not for the implementation of DR. This directly affects both key issues studied in this work, since Firm Supply is the difference between actual consumption and the Baseline and this in turn has an impact on the Cost Allocation. Since a Baseline is intangible, it has various characteristics to be studied (ENER09).

Before looking at the elements that define a baseline methodology, there are four characteristics that serve to evaluate the effectiveness of a baseline:

1. Accuracy
    a. Customers should receive credit for the curtailment they actually provide.
2. Integrity
    a. A program should not encourage irregular consumption to "game the system".
3. Simplicity
    a. All stakeholders should be able to make their own calculations.
4. Alignment
    a. The design should consider the goal of the program.

To take two characteristics as an example, a Baseline with maximum integrity (very resistant to manipulation) could be so complex as to be unworkable by stakeholders. On the other side of the spectrum, the most understandable approach could allow participants to exploit the Baseline in their favor. Therefore, it is critical for a design to find a balance among these characteristics.

The first decision is whether to pick a Profile or a Static Baseline. A Profile Baseline is usually referred to as a Customer Baseline, or CBL, and a Static Baseline as a Firm Service Level, or FSL. CBL is determined using granular time interval data whose intent is to simulate the dynamic shape of a customer's demand. An FSL, in turn, usually averages the peak monthly demand for the facility over the previous delivery season. Needless to say, the two approaches can and do result in widely different Firm Capacities when applied to the same curtailment event.

The second issue is the Measurement Granularity. An effective baseline methodology should have an appropriate timing interval for both data collection and calculation. Since most Demand Response programs are intended for peak load conditions, quantity and value should be measured in short time intervals. Therefore, the limit would be set by the metering equipment available, but 5-minute intervals should suffice.

Next is to define the Baseline Window. It is tempting to use the most recent data; however, with longer Demand Response events and advanced notification, using the most recent information might prove to be too short of a window and exhibit problems of accuracy and be susceptible to manipulation. Given enough warning time, a consumer may intentionally increase consumption prior to an event in order to maximize its recognized Firm Supply. It is generally accepted that 10 (non-weather-affected) business days reasonably represent consumption during normal operations.

Now, as hinted in the preceding sentence, there should be some Exclusion Rules. The general rule is that data should be evaluated across similar time periods; weekdays are typically like other weekdays, for example. A counter to this argument would be a consumer who is highly dependent on weather conditions and that by excluding extreme weather days, which coincide with scarcity events, its Baseline will be consistently underestimated. To combat this, a recommendation would be to look only at a set number of highest load days among the non-event data set. A high 5 of 10 approach addresses this issue, ie, among the 10 prior non-event days, the 5 days with the highest peak load are averaged.

Approaching the end of the list are Baseline Adjustments. Like in the previous example, customers' demands tend to be heaviest on event days; therefore, an adjustment based on the load immediately preceding the event is preferable. Such an adjustment can be either calculated with a scalar or additive technique. The scalar technique is based on a percentage comparison, whereas the additive approach calculates the actual demand difference in kW. Furthermore, adjustments may reflect demand conditions symmetrically (ups and downs) or symmetrically (only adjusted up). Downward adjustments may punish consumers for anticipating a Demand Response event and taking action, where they should actually be rewarded.

Second to last is the issue of whether to set an Individual or Portfolio Baseline. The portfolio approach represents essentially a random choice of exclusion days which results in the consumer's incentives not as the firm result of curtailment efforts, but as an arbitrary result within a range of calculation possibilities. Moreover, the individual approach allows the consumer to measure their own performance close to real-time.

Finally, there's the Calculation Type, either Average or Regression. The latter often uses sophisticated statistical tools which promote the highest level of accuracy possible. Unfortunately, its complexity also makes it less hospitable to stakeholders. Conversely, the Average's most important asset is its simplicity.

In sum, the collection of recommendations would be to have a profile baseline with a Ten-day window, high 5 of 10 exclusion rules among similar days (excluding event days), asymmetric day-of adjustments, and a purely individual and average calculation method. The following table illustrates how they comply with the four characteristics to evaluate a baseline's effectiveness:

Table 1, Baseline recommendations of EnerNOC

	Accuracy	Integrity	Simplicity	Alignment
Ten-day window	✕	✕		
High 5 of 10 exclusion	✕			✕
Asymmetric adjustment	✕			✕

Individual approach			×	×
Average calculation			×	

2.3.2 Firm supply and cost allocation calculation

A mathematical demonstration can be used to link reliability metrics to the methodology to measure Firm Supply (BRIT22). Also, it will be helpful to determine how their contribution should be remunerated, which connects to Cost Allocation.

Since there are two ways to solve the dispatch problem; by an ideal central planner and by decentralized individual agent, both cases will be used to compare and contrast. A reliability metric will be set for the centralized case and for the decentralized one the focus will be on the behavior in both the energy and capacity markets.

Finally, by comparing the optimal result for both cases, conclusions will be drawn on how to determine Cost Allocation and Firm Supply.

Centralized case

The ideal central planner looks to maximize the benefit of both generators and consumers. Therefore:

$$MAX_{QK}(NSB) = U\left(\sum_{n} Q\right) - C\left(\sum_{n} Q\right) - I\left(\sum_{n} K\right)$$

(1)

Where:

Q represents the production of every generator

U is the consumer's utility function

C is the generator's cost function

I is the investment cost function

K is the generation capacity installed

As previously stated, there will be a reliability metric involved and so there need to be constraints to that end:

$$Q \leq K \times \alpha$$

(2)

$$RM\ (K) \geq RT \times \beta$$

(3)

Where:

α is the mean value of the tail of the probability function

RM is the reliability metric

RT is the reliability target

β is the dual variable of the constraint

The first constraint (2) represents the maximum power each generator can produce, which corresponds to the capacity of the generator in question, K.

The second constraint (3) forces the reliability metric to reach the pre-defined reliability target.

Now a Lagrangian transformation is used in order to obtain the necessary conditions, also the partial derivates are computed with respect to the variable which is to change:

$$L(Q,K,\alpha,\beta) = U(Q) - C(Q) - I(K) + \sum (Q-K)\Delta\alpha + [RF(K) - RO]\Delta\beta$$

(4)

Then an additional partial derivative is obtained from equation (4) in terms of K:

$$\frac{\partial L}{\partial K} = -\frac{\partial I(K)}{\partial K} - \alpha + \frac{\partial RF(K)}{\partial K}\beta = 0$$

(5)

The equilibrium between producing savings in the short term and limiting costs in the long term can be observed in the first part two terms of equation (5). The last term should only be present if the constraint of equation (3) is applied, as this would change the equilibrium.

Decentralized case

The objective now does not take into account the social benefit, but seeks instead for every generator to maximize its profit assuming there is both an energy market and a capacity market:

$$MAX_{QK}(P) = EMP_Q\Delta Q + CMP_K\Delta K - C(Q) - I(K)$$

(6)

The new variables are:

EMP is the energy market price

CMP is the capacity market price

With no reliability metric, the only constraint this time is the maximum power of each generator:

$$Q \leq K \times \alpha$$

(7)

The Lagrangian transformation is used once again to obtain the necessary conditions, as well as the partial derivates with respect to the variable which is to change:

$$L(Q,K,\alpha,\beta) = SMP_Q\Delta Q + CMP_K\Delta K - C(Q) - I(K) + (Q-K)\Delta\alpha$$

(8)

Similarly, an additional partial derivative is obtained from equation (4) in terms of K:

$$\frac{\partial L}{\partial K} = CMP - \frac{\partial I(K)}{\partial K} - \alpha = 0$$

(9)

The result of equation (9) is very similar to the one obtained for equation (5), even though the constraint described by equation (3) is not present in the decentralized case.

Unification of both cases

By comparing the aforementioned similar equations (9) and (5), the following is obtained:

$$CMP_K = \frac{\partial RM(K)}{\partial K} \Delta\beta$$

(10)

Which could be interpreted as a remuneration result of the capacity market as follows:

$$CMP_K \Delta K = \frac{\partial RM(K)}{\partial K} \Delta K \Delta\beta$$

(11)

And ultimately leads to the conclusion that, given that the CMP represents the price of 1 MW, the marginal contribution to the reliability metric should determine the Firm Supply and its price should be β.

Chapter 3: Case example

3.1 Model description

A convolution model will be used to analyze the effect of demand on how critical moments of the system reflect on a reliability index, in this case Expected Energy Not Served (EENS). Then it will be employed to determine Cost Allocation and Firm Supply of the program.

A convolution model operates by defining the probability of the demand to be greater or equal to a certain value and expressing that as a function. The thermal units of a hypothetical generation fleet are then dispatched in single loading order according to their probability of failure. It's important to note that minimum load requirements or start-up processes are not taken into account. The failure of each plant is regarded as being independent from the operation of any other generation unit.

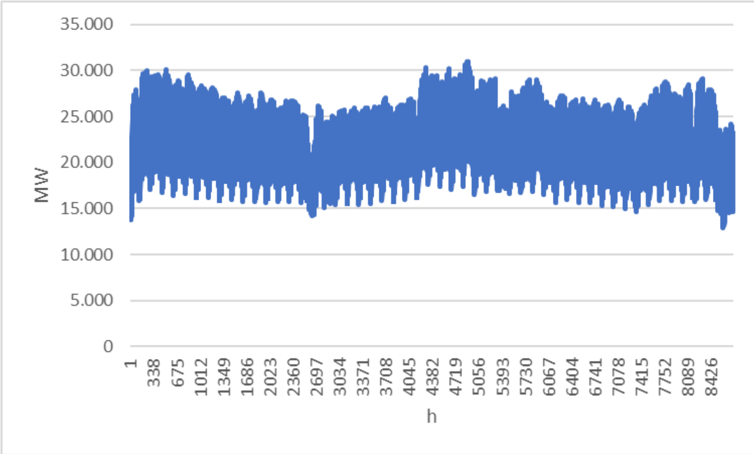
Once the probability of failure is described as a function which changes with total demand, the expected energy output can be calculated. The convolution name comes from piling up the functions of several generation units according to their failure characteristics. The resulting function should show the behavior of the overall power system and its failure probability as a whole.

Very briefly, the definition of the reliability index to be used will be reviewed (BILL96): EENS, measured in MWh, it is the expected energy not to be supplied, either because of generation unavailability or high demand.

3.2 Data used

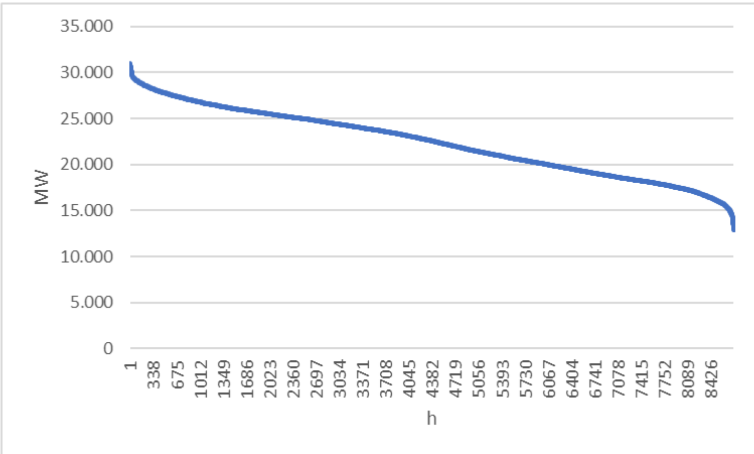
Although the typical graph of demand is usually chronological, the raw data will have to be adjusted to fit a convolution model. The vertical axis will represent hours where a certain demand is recorded, with the x-axis representing how demand grows. This might be understood as a Load Duration Curve with its axis switching places. Let's present every step:

I. Chronological graph: Raw consumption data
   a. Power vs time



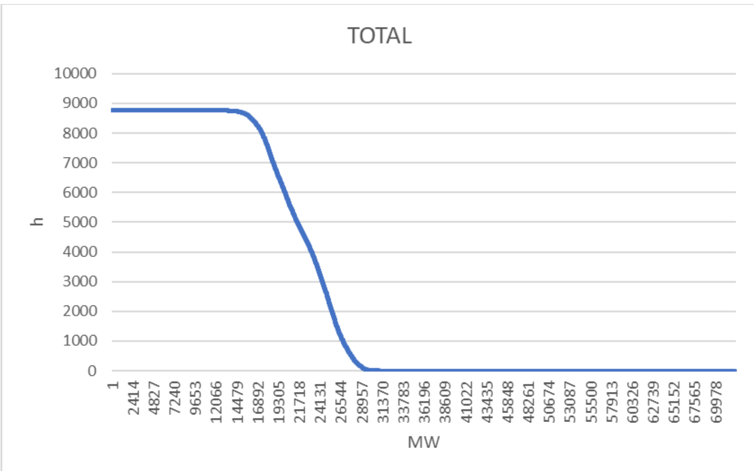
II. Load Duration curve: Data sorted from maximum to minimum

a. Power vs time



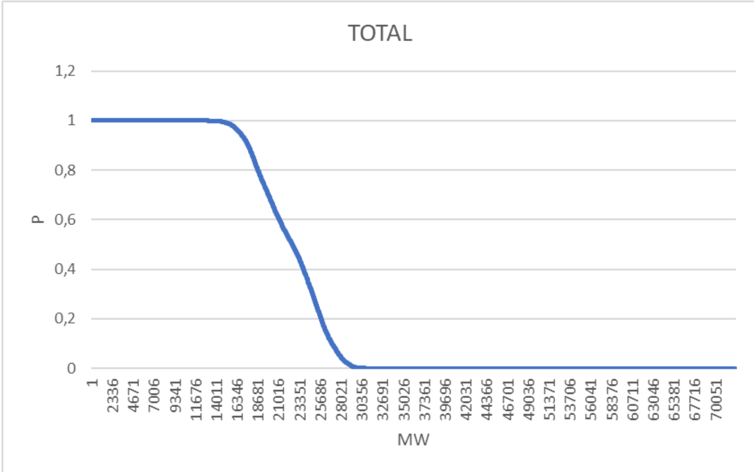
III. Curve for the model: Data rearranged by counting demand by hours

a. Hours vs power



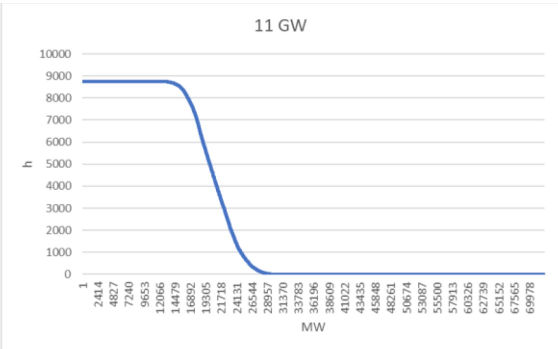
IV. End result: Expressed in probabilistic terms

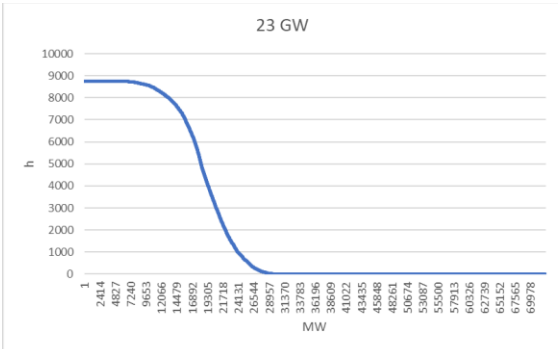
a. Probability of failure vs power



The raw data comes from the consumption of Spaniards during 2019. It was later adjusted and it can be further reworked to filter out specific groups of consumers. The resulting curve provides a useful representation of data for the next calculations.

Now, in order to provide the model with the intermittency of solar power as it stands now and as it will in the future, two scenarios are presented:

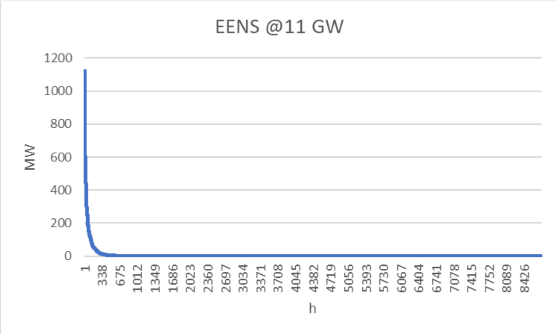


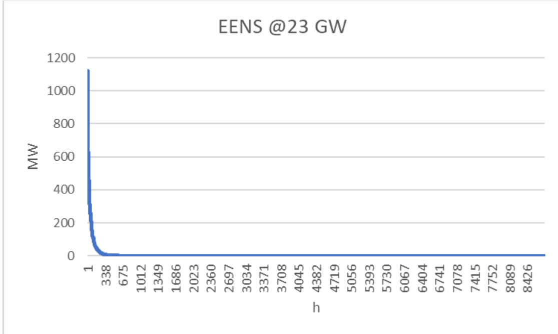


The first one includes, or rather subtracts, what the solar power capacity was at the time (around 11 GW) by the Spanish Transmission System Operator (REES19). The second one follows suit with a solar power participation of 15% (around 23 GW) which is a rough estimation of what solar power should provide in order to comply with the National Energy and Climate Plan for Spain in 2030 (NECP20). Later both curves have been reworked in order to fit the mathematical characteristics of the convolution model.

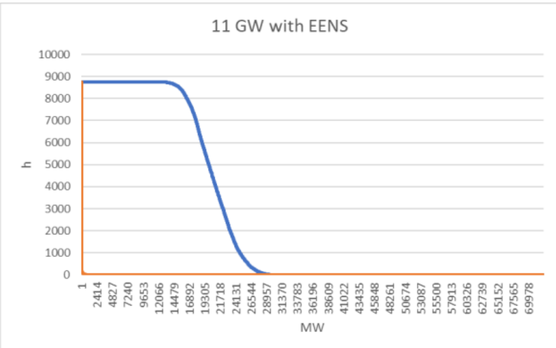
3.3 Results

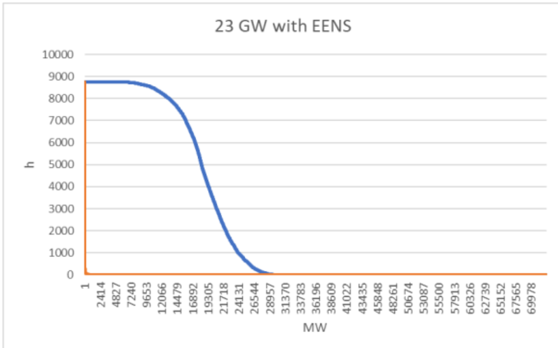
The main result from the convolution model is the measurement of the EENS:





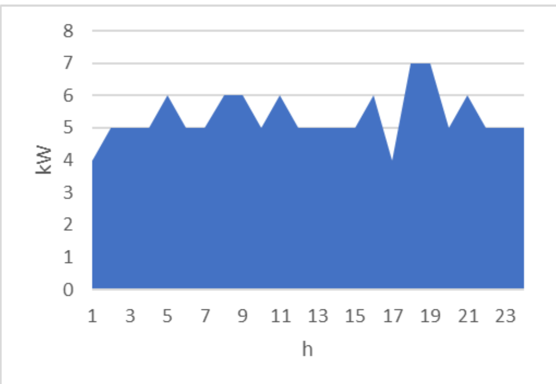
Now, for the sake of conceptualization, the EENS will be rearranged by hours and integrated with the curve for demand:

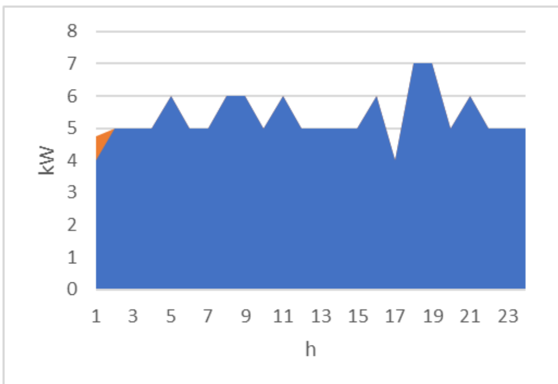




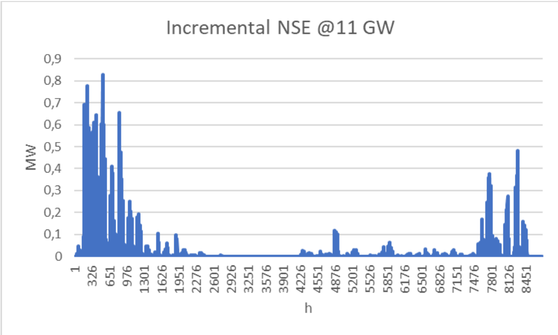
These last graphs do not really add much in terms of visibility, but they do provide a factor of visual confirmation since most hours should show almost no EENS.

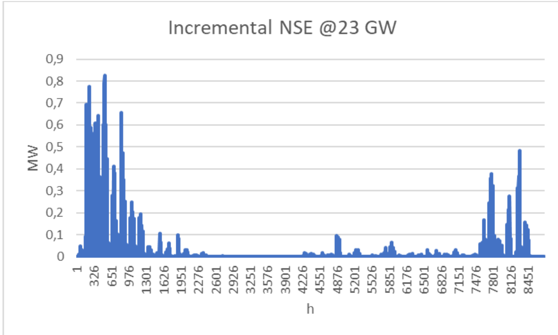
A more mathematically useful output from the convolution model is the incremental EENS. That is, how much the EENS measurement changes as a response to a small uptick on the demand curve in an specific hour each time:



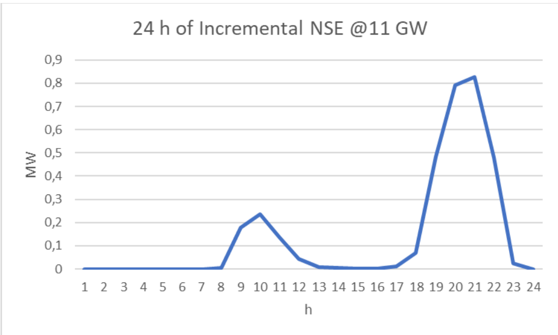


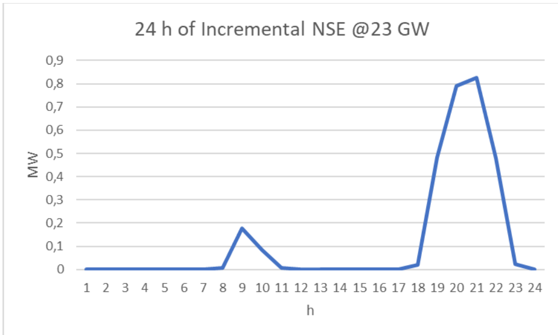
This process is then repeated for every single hour. The result of incremental EENS for the whole year would be:





For enhanced visual sharpness a 24-hour look at the day with the maximum incremental EENS is shown:





Now onto determining the Cost Allocation and Firm Supply that should apply to a Demand Response program. Conceptually, Cost Allocation should be assigned based on how much a consumer contributes to the cost of the program and Firm Supply could be determined by comparing said contribution to a pre-defined minimum one.

Four demand profiles (G) will be taken into account. First an entertainment establishment which works at select times (I), second a business unit which is open at office hours (II), third a middle-sized factory (III) and fourth a big industrial complex (IV):

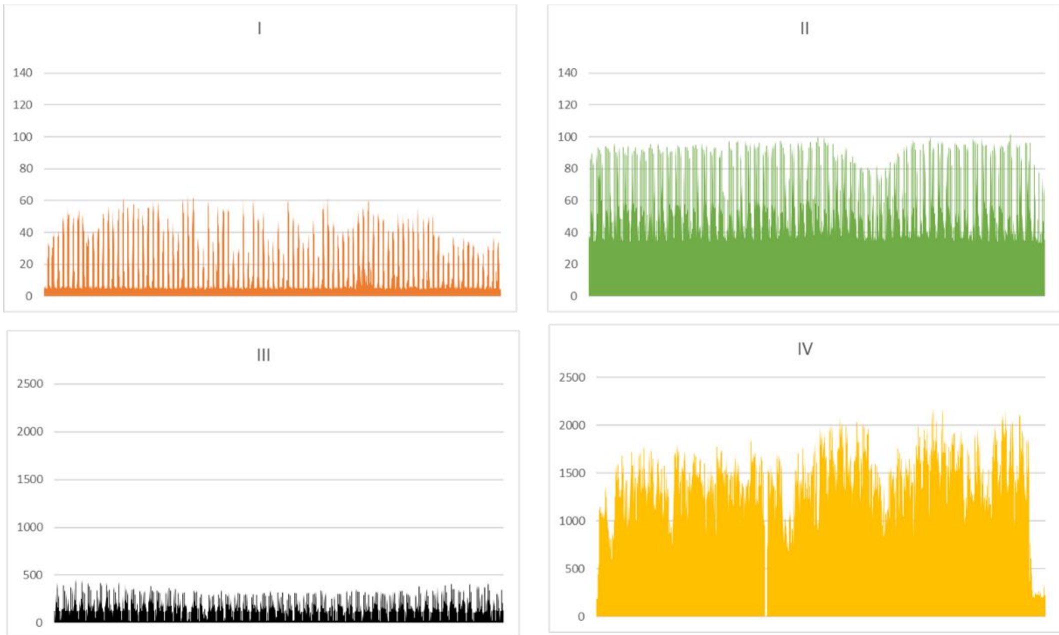


Figure 4 Demand profiles during the year (different scale for I & II and for III & IV)

Each of these profiles will be normalized in order to compare the results to a hypothetical 1 MW constant curve:

$$Normalized\ demand_h^G = Demand_h^G \times \frac{8760}{Annual\ consumption^G}$$

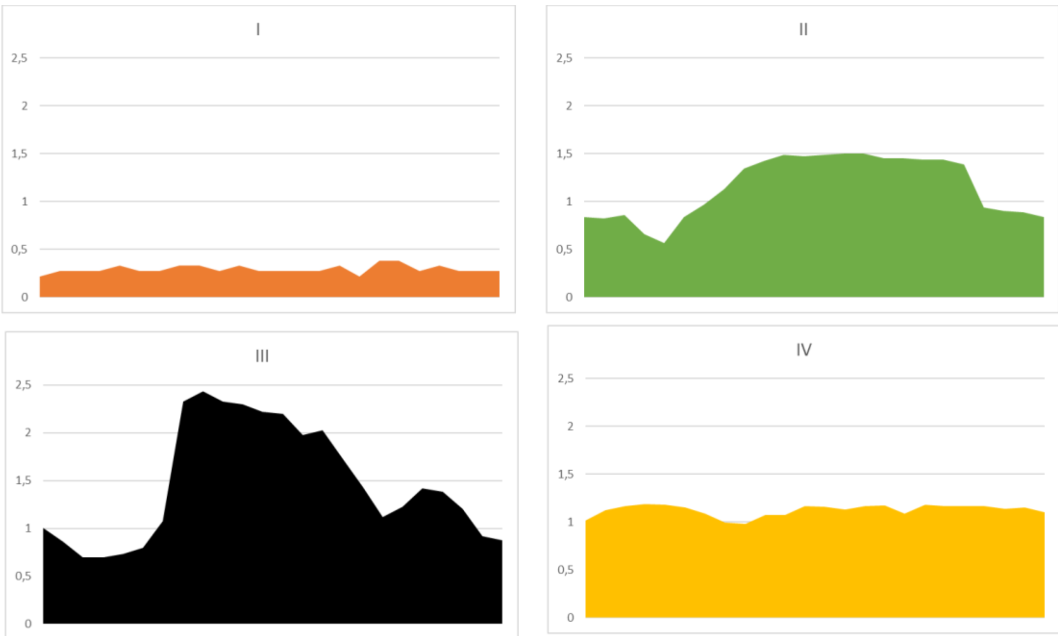


Figure 5 Normalized demand profiles on the day of maximum Incremental NSE (common scale)

As previously stated, Cost Allocation should come from the incremental contribution that a demand profile has on the overall reliability index in question, in terms of percentage:

$$100\% \times \left(\sum_{h}^{8760} Incremental\ NSE_h \times Normalized\ demand_h^G \right)$$

The overall cost of the program is not considered for the purposes of this work, only how it would be allocated among the consumers in relation to the hypothetical 1 MW constant curve. The results are:

Solar scenario	Demand profile	Cost Allocation result
11 GW	I	104,3%
	II	120,3%
	III	140,6%
	IV	102,5%
23 GW	I	106,7%
	II	117,8%
	III	134,5%
	IV	103,0%

Now, Firm Supply will be determined and verified. If a Demand Response of 20% of the original consumption is applied to every demand group, then an 80% demand will remain. Next, by multiplying the 20% Demand Response by the incremental contribution to the reliability index, the Firm Supply of that reduction is determined.

Then, the 80% remaining demand will have an adjusted Cost Allocation, determined through the incremental contribution to the reliability index, which we may compare to the previous result minus the Firm Supply of the 20% Demand Response. This arithmetic operation would result in a Cost Allocation which should be equivalent to the adjusted Cost Allocation of the remaining 80% of demand.

Solar scenario	Demand profile	Previous Cost Allocation	Firm Supply DR 20%	Cost Allocation by arithmetic	Cost Allocation Remaining 80%
11 GW	I	104,3%	-20.86%	**83,43%**	**83,43%**
	II	120,3%	-24,06%	**96,24%**	**96,24%**
	III	140,6%	-28,11%	**112,44%**	**112,44%**
	IV	102,5%	-20,50%	**82,00%**	**82,00%**
23 GW	I	106,7%	-21,34%	**85,35%**	**85,35%**
	II	117,8%	-23,56%	**94,23%**	**94,23%**
	III	134,5%	-26,89%	**107,57%**	**107,57%**
	IV	103,0%	-20,60%	**82,40%**	**82,40%**

As expected, the last two columns show equal values.

Chapter 4: Conclusions

Capacity markets are a useful tool to guarantee the reliability of electricity systems by correcting the market failures described earlier, mainly the inelasticity of consumers in the short-term and limits on prices introduced by regulators. Power system reliability lies at the front of the regulatory debate in recent years, especially in North America & Europe, with various design options being discussed.

A policy program that seeks to make the system more reliable regardless of the technology used, should rely on a market design that allows all resources to participate, including Demand Response. This would fit with the overall goals to be more environmentally friendly, empower consumers and make the system more flexible. A technology-blind capacity mechanism is the proposed design in this work as it would address the aforementioned missing money and the missing market problems.

Some difficulties arise from the challenge to include demand in the regulatory design, though. A review of the theoretical framework reveals that the design rules must meet a compromise among design elements. For example, neither a very simple design that allows agents to "game the system," nor a complicated one that is hard to explain, are adequate.

International experience reveals that the markets with the most experience integrating Demand Response in their Capacity Mechanisms are located in North America and PJM in particular has a design that has evolved for over 20 years with specific rules for both the summer and year-round periods. European designs that have included a unique remuneration for Demand Response tend to attract more participation but fail to treat resources equally.

Demand Response should pay for its participation in the capacity mechanism and, at the same time, it could reduce its participation by reducing its consumption during critical moments. How much reliable Firm Supply it provides to the system and how to determine its share of the Cost Allocation are the main questions of this work.

Through a convolution model that starts from Spanish consumption in 2019 and that singles out 4 kinds of consumers, their different contributions to a reliability index were measured. Finally, conclusions have been drawn towards Firm Supply and Capacity Allocation being similarly determined through the incremental contribution they have on a reliability index. Specifically, the Firm Supply of a particular demand reduction should equal the difference between the Cost Allocation of the original and remaining demand.

References

Anon., 2020. *Comisión Nacional de los Mercados y la Competencia.* [Online]
Available at: https://www.cnmc.es/prensa/circular-3-2020-peajes-electricidad-transporte-distribucion-20200124

Anon., 2020. *NECP.* [Online]
Available at: https://energy.ec.europa.eu/system/files/2020-06/es_final_necp_main_en_0.pdf

Batlle, C., Mastropietro, P., Rodilla, P. & Pérez Arriaga, I. J., 2022. *Resource Adequacy in Decarbonizing Power Systems,* s.l.: s.n.

Billinton, R. & Allan, R., 1996. *Reliability Evaluation of Power Systems,,* s.l.: Springer.

Botterud, A. & Auer, H., 2020. Resource Adequacy with Increasing Shares of Wind and Solar Power: A Comparison of European and U.S. Electricity Market Designs. *International Association for Energy Economics.*

Brito-Pereira, P. et al., 2022. *Adjusting the aim of capacity mechanisms: Future-proof reliability metrics,* s.l.: Elsevier.

Bushnell, J., Flagg, M. & Mansur, E., 2017. Capacity Markets at a Crossroads. *Energy Institute at Haas.*

Enel X, 2023. *Get Paid to Reduce Energy in PJM,* s.l.: s.n.

ENERNOC, Inc., 2009. The Demand Response Baseline.

European Parliament, 2019. *Internal Market for Electricity Regulation 2019/943,* s.l.: s.n.

FERC, 2018. *Federal Energy Regulatory Commission.* [Online]
Available at: https://www.ferc.gov/industries/electric/indus-act/demand-response/dem-res-adv-metering.asp

Humphrey, T., 1998. *Federal Reserve Bank of Richmond.* [Online]
Available at: https://www.richmondfed.org/

ISO NE, 2017. *Fully Integrated Price Responsive Demand,* s.l.: s.n.

Klein, N., 2007. *The Shock Doctrine.* s.l.:s.n.

NY ISO, 2018. *Special Case Resources (SCR),* s.l.: s.n.

NY ISO, 2020. *NYISO Demand Response Programs,* s.l.: s.n.

PJM State & Member Training Dept, 2023. *PJM Load Management Demand Response,* s.l.: s.n.

Pototsching, A., 2022. *Adequacy and Capacity Remuneration Mechanisms.* [Online]
Available at: https://fsr.eui.eu/adequacy-and-capacity-remuneration-mechanisms/

Red Eléctrica de España, 2019. *Series Estadísticas Nacionales.* [Online]
Available at: https://www.ree.es/es/estadisticas-del-sistema-electrico-espanol/series-estadisticas/series-estadisticas-nacionales

Robert Schuman Center, 2022. *Recent energy price dynamics and market enhancements for the future energy transition.* [Online]
Available at:
https://cadmus.eui.eu/bitstream/handle/1814/73597/PB_2022_05_FSR.pdf?sequence=1&isAllowed=y

Rodilla, P., Mastropietro, P. & Brito-Pereira, P., 2023. The Challenge of Integrating Demand Response in Capacity Remuneration Mechanisms.