An Overview of Probabilistic Dimensioning and Scarcity Pricing with a Focus on the Greek Electricity Market

Anthony Papavasiliou

Contents

1. Introduction

This deliverable responds to the request of the Greek Regulatory Authority for Energy (RAE) for describing the principles of a probabilistic dimensioning methodology and a scarcity pricing methodology that could be applied in the Greek electricity market. The following specific tasks are requested in the terms of reference of the regulator:

1. Comparison between the approved methodology for determining reserve requirements and probabilistic dimensioning.

2. Record and evaluate shortage pricing mechanisms that are applied in the EU.

Probabilistic dimensioning and scarcity pricing are electricity market and system operation reforms that respond to the transition of electric power systems towards an increasing reliance on renewable energy sources. The common denominator of both reforms is reserve. Probabilistic dimensioning is a quantity-based reform for adapting the commitment of reserve to the forecast short-term needs of the system based on day-ahead conditions. Scarcity pricing is a price-based reform for revisiting the pricing of electricity in balancing markets which aims at remunerating reserve in a way that better reflects the value that this reserve brings to the system.

Both methods are intrinsically linked to loss of load probability, and are thus aligned with EU legislation which calls for sizing reserves based on probabilistic methods and pricing in a way that reflects scarcity in the system, as quantified by loss of load probability. As both reforms respond to the requirements of recent EU legislation, they are currently being implemented or considered in a number of European markets. Belgium is an interesting case in point, where both measures are under implementation [13, 59]. The report will therefore anchor the discussion around the implementation of these reforms in Belgium, and attempt to draw analogies to the case of the Greek market.

Part I
Probabilistic Dimensioning

2. Overview of Reserve Dimensioning Methodologies

2.1 Literature Review

We provide a literature review on reserve dimensioning which is inspired by a classification of the reserve dimensioning literature as set forth in [13]. We consider three axes along which this literature is classified, and summarize this classification in table 1:

- Sizing methodology: The sizing methodology refers to how the decision-making problem of sizing reserve is quantified. Three predominant approaches in this respect are heuristic methods, probabilistic methods, and bottom-up unit commitment and economic dispatch models.
- Adaptiveness: Adaptiveness refers to whether the sizing methodology is adaptive to the forecast conditions of the system. Two options in this respect are static sizing and dynamic sizing.
- Stochastic models: This dimension refers to the way in which uncertainty is modeled. Two options are considered in this dimension: parametric or non-parametric.

We now explain each of these options in further details.

Heuristic sizing [21, 16, 34, 40, 50, 48, 49]. Heuristic sizing methods refer to sizing methods that determine the amount of reserve that the system should carry on the basis of simple system statistics. These methods have been widely employed in practice in the past, due to their attractive simplicity. This is the current method of choice also in the Greek system. However, these methods are currently under scrutiny on account of not being able to adapt *accurately* to system conditions that can vary significantly as a function of renewable energy supply and other system indicators on which we can perform advanced analytics. An example of a heuristic sizing method based on statistical parameters is the so-called "3+5 rule" of the US National Renewable Energy Laboratory [54], which dictates that the system should carry reserve equal to 3% of forecast load plus 5% of forecast renewable supply. The rationale of such a rule is that higher demand forecasts or higher load forecasts expose the system to greater uncertainty, and should therefore be accompanied by more reserve in the system.

Probabilistic sizing [7, 6, 18, 30, 36, 35, 37, 39, 41, 45, 43, 12]. Probabilistic methods determine the amount of reserve by explicitly aiming to satisfy a certain reliability target. Thus, the reserve decision is essentially the quantile of a distribution of system imbalance. Like the heuristic sizing methods of the previous paragraph, this approach is also attractive in its simplicity. Moreover, this sizing philosophy is aligned with the spirit of EU law, e.g. the Electricity Balancing Guideline [23]. This is the method that is currently being employed in Belgium, and it is also a direction that is currently being considered in the Nordic region.

Unit commitment / economic dispatch models [3, 15, 17, 16, 30, 44, 50, 55, 63, 67, 66]. An alternative to heuristic and probabilistic methods is sizing based on unit commitment and economic dispatch models that endogenously represent uncertainty. Such models attempt to develop a bottom-up description of the system and trade off explicitly the increased cost of running the system more securely (e.g. due to startup and minimum load costs, or the higher fuel costs of reserve) with the increased security that the system enjoys when it carries more reserve. Such models are typically not employed in practice, due to the complexity of the underlying stochastic formulation as well as the ensuing difficulty in solving the resulting model. They are nevertheless widely studied in the academic literature.

Static sizing [15, 21, 18, 30, 39, 12]. Static reserve sizing refers to dimensioning methods that fix the amount of reserve requirements for extended periods of operation, in a fashion that is not adaptive to system conditions. This was, for instance, the case in Belgium until recently, where the sizing of certain types of reserve would be fixed for an entire year or an entire month.

Dynamic sizing [3, 7, 6, 17, 16, 34, 36, 35, 37, 41, 40, 45, 43, 44, 50, 48, 49, 55, 63, 67, 66]. Dynamic reserve sizing refers to methods that adapt the sizing decision to the forecast day of operations. Dynamic sizing methods aim at achieving the same reliability target as static ones by relying on less reserve as well as a constant risk profile throughout every day of operations during the year. A dynamic dimensioning methodology has been adopted in Belgium, and will also be considered in the present report as a point of comparison with a static sizing method which is compatible with the data that was made available by the Greek Regulatory Authority for Energy for the years 2018-2020.

Parametric models [7, 39, 40, 45, 43, 50, 65]. Most reserve sizing methods do not only consist of a sizing decision-making method, but also a statistical model of the underlying uncertainty. Such a model is directly required for probabilistic dimensioning, since it yields the quantiles that are the basis of the sizing decision. It is also required for bottom-up models, since it is used for generating scenarios for unit commitment and economic dispatch models. The modeling of uncertainty is typically centered in the literature on the stochastic model of wind power, solar power and / or load. Parametric models refer to methodologies that employ parametric distributions (such as Levy alpha-stable, gamma, and normal distributions) for fitting each of these processes. The appeal of parametric methods is the fact that only few parameters need to be estimated for the development of a stochastic model of the uncertain input of the problem.

Non-parametric models [4, 5, 30, 36, 35, 37, 38, 47, 67, 65, 66, 61]. Non-parametric methods for fitting uncertain data are also employed in the literature. They often apply kernel density estimation, and their appeal is a better fit to historical data. This advantage needs to be balanced against the risk of over-fitting historical data, which is typically made available in annual records of stochastic processes (imbalances, load, wind, solar, ...) with a resolution of 1 minute to 1 hour.

2.2 Sizing Methodology in the Greek Electricity Market

2.2.1 Target Model Methodology

The existing reserve sizing procedure that is employed by the Greek TSO (ADMIE) was approved by decision 1092/2020 of the Regulatory Authority for Energy. The procedure is laid out by ADMIE in [1]. The methodology described in [1] corresponds to the Target Model of the Greek electricity market. As such, the methodology is effective November of 2020, at which point the new model of the Greek electricity market was launched.

It is interesting to note that ADMIE distinguishes, like other European TSOs, between "normal imbalances" (e.g. forecast errors) which need to be dealt with by aFRR and mFRR, as well as contingencies, which need to be dealt with by FCR, and FRR. Using the classification of section 2.1, ADMIE follows a dynamic sizing procedure based on heuristics related to the statistical parameters of system characteristics.

The existing sizing procedure adopted in the Greek electricity market for upward / downward aFRR is driven by

Table 1: A classification of reserve sizing literature.

	Heuristic	Probabilistic	UC/ED	Static	Dynamic	Parametric pdfs	Non-parametric pdfs
[4]							X
[5]							X
[3]			X		X		
[7]					X	X	
[6]		X			X		
[15]			X	X			
[17]			X		X		
[16]	X		X		X		
[21]	X			X			
[18]		X		X			
[30]		X	X	X			X
[34]	X				X		
[36]					X		X
[35]		X			X		X
[37]		X			X		X
[38]							X
[41]		X			X		
[39]		X		X		X	
[40]	X				X	X	
[45]		X			X	X	
[43]		X			X	X	
[44]			X		X		
[47]							X
[50]	X		X		X	X	
[48]	X				X		
[49]	X				X		
[55]			X		X		
[63]			X		X		
[12]		X		X			
[13]		X			X		X
[67]			X		X		
[65]						X	X
[66]			X		X		X
[61]							X

1. the minimum FRR requirement (which in itself is a function of maximum load in the system),

2. a constant corresponding to the technical minimum of a typical thermal unit (meant to capture the possibility that a unit is asked to turn on but fails to do so),

3. the scheduled interchange, and

4. the scheduled demand.

The way in which one distinguishes the sizing for upward and downward aFRR in these cases is driven by the difference in the coefficients that are used for how each of these factors is assumed to contribute to the total aFRR requirement (c versus d parameters in [1] for upward and downward aFRR respectively).

The existing sizing procedure for upward / downward mFRR is driven by:

1. upward / downward aFRR

2. renewable forecasts

3. demand ramps

4. scheduled interchanges

5. an indicator for extreme conditions (indicatively, unfavorable weather, large renewable forecast deviations, reduced adequacy, contingencies, strikes, reduced fuel reserves for thermal units, low hydro energy levels, or a combination of the above).

The treatment of contingencies is interesting to point out in the new methodology. In particular, it is worth noting that the sizing of aFRR seems to target at least the largest online unit for hours 7 to midnight (see page 23 and appendix B of [1]). Although the SOGL foresees FRR sizing that should be able to cope with at least the largest contingency in the system, the allocation of this requirement to mFRR versus aFRR is typically driven by considerations of how physical operations respond to the occurrence of such an incident.

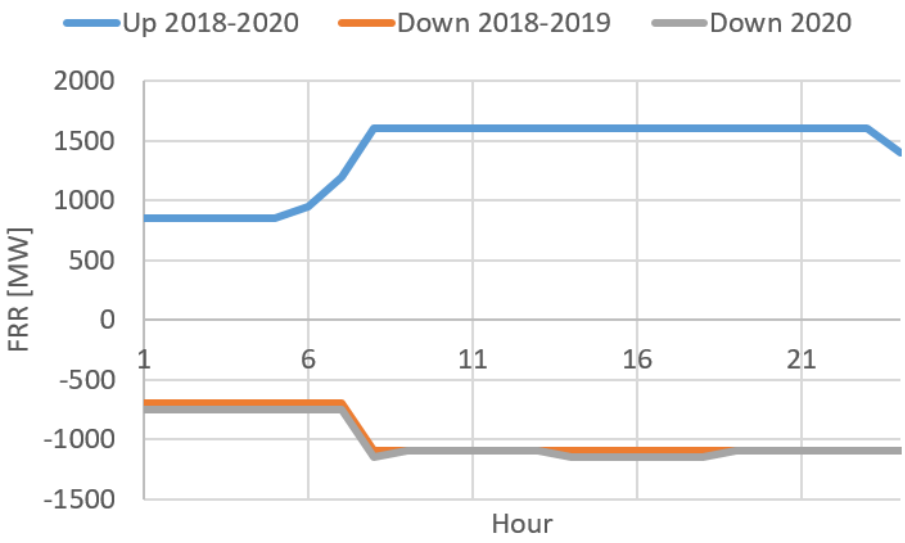

Figure 1: Representative reserve sizing values for upward / downward capacity in the Greek electricity market in 2018 - 2020.

2.2.2 Sizing Before the Target Model

Figure 1 describes representative reserve requirements of the Greek electricity market for January 2018 - October 2020. The data is sourced from the website of the Hellenic Energy Exchange. In addition to reserve requirement data, the website includes the day-ahead schedules of individual units, hourly energy production schedules, as well as commitments of FCR, aFRR and mFRR for individual units in the Greek system. We concentrate on FRR in the table, i.e. the sum of mFRR and aFRR, since the sizing of FCR follows a separate procedure and is out of scope for the present analysis. Note that the mFRR requirements on the aforementioned website are assumed to correspond to both upward as well as downward reserve. In contrast, the .

We note that the splitting of FRR between mFRR and aFRR is considered as being out of scope for the present analysis, although a number of publications [6, 8, 41, 35, 37, 12, 43] consider this important design question. On the other hand, we consider downward sizing in our analysis, and present representative values in figure 1.

It is worth noting that the FCR, upward aFRR, and mFRR require-

ments have remained fairly constant in Greece in January 2018 - October 2020. In the sample dates presented in the figure, there is a slight change in the requirement of 2020 for downward aFRR, which increases slightly (by 50 MW) in hours 1-8 and 14-18. We also note that the downward aFRR requirements are notably lower than the upward aFRR requirements, which is typical due to the asymmetric exposure of the system to contingencies.

3. Model of Imbalances in the Greek System

In the absence of publicly accessible real-time imbalance data for the Greek market, in this section we propose a simplified model which aims to serve as the basis for our case study of the probabilistic dimensioning methodology. We emphasize that this imbalance model is not meant to be realistic, but rather to convey certain first principles. On the other hand, the probabilistic dimensioning methodology does not depend on this imbalance model, and can be applied directly on historical data. Therefore, the probabilistic dimensioning methodology is robust and directly implementable.

3.1 Modeling Imbalances

We have received the following data from the Regulatory Authority for Energy:

- Load data with hourly resolution from 01/01/2018 until 31/10/2020, thus spanning 2 years and 10 months (namely 1035 days).
- Renewable energy supply data with the same characteristics.
- Import / export data with the same characteristics.

Additionally, we have access to the day-ahead commitment of individual units at the website of the Hellenic Energy Exchange. Notably, however, we do not have access to historical *real-time* imbalance data for the Greek system. We therefore propose a model for imbalances which allows us to conduct our analysis.

In developing our imbalance model, we are interested in a number of features that affect reserve dimensioning:

- Imbalances are driven by both contingencies, as well as "smooth" imbalance drivers such as forecast errors.
- Imbalances can be explained by a number of factors in the system, such as renewable energy forecasts, load forecasts, and scheduled imports. We refer to these factors as *imbalance drivers*. Other imbalance drivers may include the change of the hour (due to market ramps), temperature, and so on. Higher forecasts tend to result in higher imbalances.
- On the other hand, a significant portion of the system imbalance signal may not be possible to explain based on imbalance drivers. Past analyses of the Belgian system [13] have shown that approximately half

of the imbalance signal may not be attributable to imbalance drivers. We assume that this part of the imbalances is representable by white noise.

- Imbalance drivers do not have symmetric distributions in the upward and downward directions. For example, high renewable supply forecasts are more likely to lead to significant negative imbalances (under-supply) and low positive imbalances (over-supply) since the renewable supply can mostly decrease during periods of high output.

We therefore develop a model for imbalances based on the following methodology. The proposed methodology attempts to strike a balance between data availability and a desire to capture empirically relevant effects that drive reserve dimensioning decisions [13]:

- Use representative day types to model contingency risk in the system. The idea is presented in section 3.2.
- Use imbalance drivers (load, RES and imports) to model factors that contribute to the system imbalance based on skewed distributions, whose variance depends on the imbalance drivers. The idea is presented in section 3.3.
- Use "white noise" to model the part of the imbalance signal that cannot be explained by imbalance drivers. The idea is presented in section 3.4.
- Tune the parameters of the model so that the resulting imbalance is consistent with the reliability achieved by the reserve dimensioning that is employed in the Greek market. The idea is presented in section 3.5.
- We then compare the baseline dimensioning methodology to the probabilistic dimensioning methods that we describe in section 4

In the sequel, we refer to "normal imbalances" as the sum of imbalances related to imbalance drivers and idiosyncratic imbalances. These should be contrasted to imbalances resulting from contingencies.

3.2 Contingencies

In order to represent the risk of generator failures, we consider eight representative day types (one for each season, and weekdays versus weekends).

For each of these day types, we fix the generator schedules to historically observed data. Concretely, we consider the following day types in 2018:

- Winter weekday: 15/01/2018
- Winter weekend: 07/01/2018
- Spring weekday: 08/03/2018
- Spring weekend: 11/03/2018
- Summer weekday: 07/06/2018
- Summer weekend: 10/06/2018
- Fall weekday: 06/09/2018
- Fall weekend: 09/09/2018

Alternatively, we could have considered a clustering method for determining different day types, or we could have worked directly with each day of the dataset. The latter option is out of scope for the present assignment due to IT implementation effort given the format in which the data became available by RAE, and could be investigated further in future work.

We consider a failure probability of 1 incident per year. We further assume that each failure corresponds to four imbalance intervals (i.e. the time to clear the fault by repairing the unit or bringing online another unit is assumed to be one hour). This assumption is an intermediate choice between the values that have been assumed in previous analyses of the Belgian and Swedish systems.

We assume that contingencies occur independently of normal imbalances and idiosyncratic imbalances. This allows us to sample contingencies independently from one period to the next. Concretely, since there are no intertemporal constraints in our model, we can assume that the contingencies are sampled for each balancing market time unit, without concerning ourselves about the fact that a failure lasts for four consecutive 15-minute imbalance intervals.

There are 28 thermal units and 18 hydro units in the system, as well as 4 pumping units. Failures between these components are assumed to be independent of each other.

3.3 Imbalance Drivers

The next component of imbalances that we wish to model are normal imbalances. We are specifically interested in capturing two effects: higher forecasts are correlated with higher forecast errors, and the support of the probability density function depends on the imbalance driver, an effect that introduces skewness to the probability density functions of imbalances for reasons that we explain below.

Regarding modeling the first effect, we adopt a simple assumption. We specifically assume that imbalances follow a normal distribution with a mean of 0 MW and a standard deviation of $C \cdot |L|$ (for load forecast errors), $C \cdot |R|$ (for renewable supply forecast errors), and $C \cdot |I|$ (for import forecast errors) respectively, where L is the system load, R is the renewable energy supply, and I are the imports. The idea is to adapt the constant C such that the average sizing value of figure 1 achieves the target unreliability of 3 hours per year. Note that this simple modeling assumption captures the effect whereby higher loads / renewable supply / imports imply larger imbalances.

Regarding skewness, our modeling assumption is driven by the fact that load, renewable supply, and imports are lower and upper bounded. Concretely, based on the data provided by RAE, we can estimate the following values:

- Minimum load over the duration of the dataset is 2840 MW, maximum load is 9529 MW.
- Minimum renewable supply over the duration of the dataset is 103 MW, maximum renewable supply is 4245 MW.
- Minimum imports over the duration of the dataset are -1428 MW (i.e. the maximum amount that is *exported* historically is 1428 MW), maximum imports are 2041 MW (i.e. the maximum amount that is *imported* historically is 2041 MW).

These bounds imply a skewness in the distribution of the imbalances caused by these drivers. We explain this idea concretely in figure 2. We consider, in the left panel of this figure, the probability distribution function of an imbalance driven by renewable supply which is originally symmetric. The vertical line in the left panel corresponds to the installed capacity of renewable generation. Since the total renewable supply, which is the sum of the day-ahead forecast supply and the renewable supply imbalance, cannot exceed the installed renewable capacity, we propose a model that captures

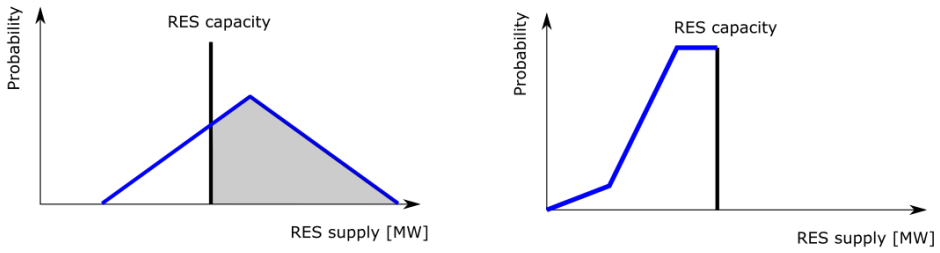

Figure 2: Left panel: hypothetical probability distribution function of renewable supply (RES-driven imbalance plus the underlying imbalance driver) which exceeds RES installed capacity. Right panel: probability distribution function of renewable supply which reflects on the installed capacity boundary.

this physical feature. Concretely, in simulating imbalances driven by renewable supply forecast errors, we assume that the supply "bounces back" / is reflected on the wall of the vertical line of the left panel of figure 2. As a result, we arrive to the probability density function of the right panel of figure 2. Note that, whereas we commence from a probability density function with zero skewness in the left panel, we arrive to a skewed probability density function in the right panel.

We model the effect of figure 2 by drawing the original imbalance from a normal distribution, as indicated in the beginning of this section. We then "reflect" this process against the lower and upper bounds of the load / renewable supply / imports to derive the final simulated imbalance value.

This modeling feature attempts to capture an effect that has been observed to be empirically relevant in the sizing of reserves, e.g. in Belgium [13]: upward and downward imbalances are skewed, and depend on imbalance drivers. As a concrete example, large renewable forecasts pose a significant threat for negative imbalances, and a minor threat for positive imbalances.

To summarize, we generate imbalances that are driven by imbalance drivers as follows:

$$Imb = \begin{cases} 2 \cdot X^- - X - C \cdot |X| \cdot N, & C \cdot |X| \cdot N + X < X^- \\ C \cdot |X| \cdot N, & X^- \leq C \cdot |X| \cdot N + X \leq X^+ \\ 2 \cdot X^+ - X - C \cdot |X| \cdot N, & C \cdot |X| \cdot N + X > X^+ \end{cases}$$

where X denotes the value of the imbalance driver, X^- and X^+ correspond

to its minimum and maximum possible value respectively, C is a tunable parameter that we use in section 3.5 in order to tune the level of uncertainty in the system, and N is a standard normal random variable.

We note that the imbalances that can be attributed to imbalance drivers are assumed to be independent of each other and of the imbalances related to contingencies. Thus, we draw random variables N independently for each of the imbalance drivers.

3.4 Idiosyncratic Noise

Our motivation for introducing an idiosyncratic component to the total imbalance signal is based on the implementation of dynamic dimensioning in Belgium [13]. In that work, it was observed that a significant portion of the imbalance signal could not be explained by imbalance drivers such as renewable supply, load, import, market ramps, and so on. This can be interpreted as a "white noise" component to the imbalance signal which cannot be specifically attributed to observable information in the system.

We concretely model idiosyncratic noise as normal random variables that are drawn independently of the imbalances related to imbalance drivers:

$$Imb = C \cdot D \cdot N$$

where C is the tunable parameter introduced in section 3.3 and D is a parameter that we need to estimate so as to ensure that the idiosyncratic imbalances represent a realistic fraction of the normal imbalance signal.

In order to decide on the variance of the idiosyncratic noise, i.e. on the parameter D, we note that we would like the idiosyncratic noise to represent 40% of the normal imbalance signal. Concretely, we are interested in a variance of idiosyncratic imbalances which corresponds to approximately 40% of the variance of normal imbalances. This can be expressed mathematically as follows:

$$\begin{aligned} T \cdot C^2 \cdot D^2 &= 0.4 \cdot (\sum_{t=1}^{T} C^2 \cdot (|I|_t^2 + |L|_t^2 + |R|_t^2) + T \cdot C^2 \cdot D^2) \\ \Rightarrow D &= \sqrt{\frac{0.4 \cdot \sum_{t=1}^{T} (|I|_t^2 + |L|_t^2 + |R|_t^2)}{0.6 \cdot T}} \end{aligned}$$

3.5 Matching the Model to a Static Sizing Methodology

The next step in our methodology is to tune the parameter C which is introduced in section 3 such that the sizing indicated in figure 1 matches

the reliability target of 3 hours per year. The idea is to scale the normal imbalances according to the parameter C, and to perform a bisection until we find a level of imbalance for which 3 hours of failures occur per year. Note that these 3 hours include failure to cover imbalances in both the upward *and* downward direction.

The chosen value of C is 0.0384. We note that we present a best-case scenario for the dimensioning of figure 1, since we exactly match the observed reliability to that of the target level of 3 hours per year. In this sense, we have no redundant capacity (the observed reliability is not below 3 hours per year), and we are also within the reliability target (the observed reliability is not above 3 hours per year).

We further note that this "training" data has been generated with a fixed seed. We can then generate test data from the same imbalance drivers but different realizations of the imbalances themselves, which is the methodology that will be adopted in section 5.

It is interesting to note that of the 34 incidents that are recorded in the training dataset, 4 are related to shortages in upward balancing capacity, and 30 to shortages in downward balancing capacity. Moreover, we note that none of the incidents are caused by a contingency. Among these, 11 incidents correspond to 2018 (2 upward and 9 downward), 18 incidents correspond to 2019 (2 upward and 16 downward), and 5 incidents correspond to 2020. Thus, the reliability target is upheld over the 2 years and 10 months of the simulation, even if reliability in certain years may be higher than the target and in other years lower than the target. A probabilistic dimensioning methodology is able to achieve a relatively constant risk profile, as we describe in section 5.

4. Probabilistic Dimensioning Methodology

In this section we present a methodology that has been considered for the implementation of probabilistic dimensioning in the Belgian system [13]. This method is then compared to the dimensioning of figure 1 in section 5.

4.1 Overview of k-Means Clustering Applied to Probabilistic Dimensioning

The k-means approach for probabilistic dimensioning is based on [8] and is also one of the methods that is proposed for implementation in the Belgian system [13]. The k-means problem is a clustering problem which aims to cluster a dataset into k groups, such that the sums of the distances of the original data from the means of the nearest clusters are minimized. The problem is computationally hard, since one in principle needs to consider all possible ways in which the original dataset can be clustered, and select the configuration that minimizes the sum of distances of cluster elements from the cluster means. The intuition of the clustering method is that the distance of cluster elements from their mean is a measure of similarity of the data points. Thus, minimizing the sum of distances implies grouping the data such that each group contains as similar data as possible.

In the context of our application, the data that we are clustering are the imbalance drivers, namely day-ahead load forecasts, day-ahead scheduled imports and day-ahead renewable supply forecasts. The idea is that each cluster corresponds to the same day type which, when observed one day in advance of operations, can provide refined information about the distribution of normal imbalances. Thus, if the imbalance drivers indicate a risky day of operations (e.g. due to high load forecasts which imply high load forecast errors) then the reserve sizing can adapt to this information by committing fewer reserves for the following day. Conversely, if a low-risk day is anticipated, then the system can resort to fewer reserves without compromising system reliability, which implies economic savings for the TSO.

4.2 Implementation of Probabilistic Dimensioning based on k-Means

In order to implement the k-means probabilistic dimensioning method with contingencies, we require the following steps:

- Step 1: cluster imbalance drivers in order to determine the day types.

- Step 2: approximate imbalances, e.g. using kernel density estimation or the empirical distribution of the data.
- Step 3: determine the reserve requirement of each day type from the appropriate quantile of the distribution computed in step 2.

Step 1: determine day types. In our analysis, we are clustering along three dimensions, namely forecast load, forecast renewable supply, and forecast imports. We cluster each of these data inputs into two values. This gives us 8 types of days: (High Load, High Wind, High Solar), ..., (Low Load, Low Wind, Low Solar). Regarding the interaction of the method with contingencies, we note that the preliminary analysis of section 3.5 indicates that observed incidents are ones in which the system experiences a large normal imbalance, even if there is no contingency. We therefore opt to work with 8 day types as determined by imbalance drivers, instead of further differentiating day types as a function of how generators are committed in the system in the day ahead.

It is interesting to note that the most commonly used k-means algorithms are inherently non-deterministic. For example, Loyd's algorithm [42] is initialized with a random selection of points which act as centroids. Initializing with kmeans++ [2] also involves a random selection of points at the first step of the initialization procedure. We therefore replicate the clustering ten times and keep the solution with the best performance. We validate that the result is consistent by repeating the sizing three times. The reserve dimensioning decisions at each run are identical, and are presented in table 2.

Step 2: estimation of imbalance distribution. Once we have defined clusters of imbalance drivers, we can observe the imbalance that materialized in the corresponding imbalance period. We can use kernel density estimation (KDE) for the estimation of the distribution within each cluster, or simply the empirical probability density function obtained from the observations within each cluster, assuming that a sufficient number of points within the cluster are observed. For each cluster, we estimate a different reserve target based on the target reliability level. This gives us the results of table 2.

Step 3: probabilistic reserve requirement. In this step we use the appropriate quantile of the distributions obtained in step 2 in order to determine upward and downward reserve requirements.

Table 2: Reserve requirement for each type of day. All quantities are in MW.

Load	Renewables	Imports	Reserve up	Reserve down
6810 (H)	2091 (H)	1331 (H)	1383	1085
6810 (H)	2091 (H)	471 (L)	1282	1043
6810 (H)	782 (L)	1331 (H)	1119	1028
6810 (H)	782 (L)	471 (L)	1187	919
4886 (L)	2091 (H)	1331 (H)	981	855
4886 (L)	2091 (H)	471 (L)	912	845
4886 (L)	782 (L)	1331 (H)	991	802
4886 (L)	782 (L)	471 (L)	970	737

We concretely follow the same procedure, in order to make the results consistent with the sizing of figure 1:

- The upward capacity requirement of figure 1 serves all but 4/99360 incidents, as noted in section 3.5.
- The downward capacity requirement of figure 1 serves all but 30/99360 incidents, as noted in section 3.5.

The results are presented in table 2. We observe a number of effects that are consistent with our intuition: (i) Downward reserve requirements are lower than upward reserve requirements for a given day type, due to the asymmetric risk of contingencies in upward requirements. (ii) Higher load implies higher reserve requirements. (iii) Higher renewable supply implies higher reserve requirements. (iv) Higher imports imply higher reserve requirements. (v) The effect of load on reserve requirements is the strongest, the effect of imports on reserve requirements is the least strong.

5. Case Study of Probabilistic Dimensioning

We proceed to compare the dimensioning of figure 1 to probabilistic dimensioning method based on k-means. The results are presented in table 3. We report, for each sizing policy, the following metrics for both the upward and downward direction:

- Average reserve committed, measured in MW.
- Unreliability: a measure of how many incidents of oversupply or undersupply occur per year, measured in hours per year. This corresponds to the loss of load expectation (LOLE) measure in reliability studies, but is here measured in both the case of upward as well as downward imbalances.
- Shortage or over-supply, measured in MWh/year. This corresponds to expected energy not served in adequacy studies, but is also measured in the downward direction (in the sense of quantity of energy oversupplied).

We note that the probabilistic dimensioning approach achieves a significant improvement in the upward dimensioning requirement, with average upward reserves being reduced by 281 MW, or 20.2 % of the average upward requirement of figure 1. Similarly, the downward dimensioning decreases by 43 MW, or 4.3%, which is less than the savings of the upward dimensioning, but still notable.

In terms of reliability performance, we find that the sizing of figure 1 remains close to the failure target of 3 hours/year. The probabilistic dimensioning results in total failures of 2.3 hours/year, thereby staying below the reliability target of 3 hours/year. The MWh of shortage and oversupply are correspondingly lower in the case of the probabilistic dimensioning method. Thus, the probabilistic dimensioning method is more reliable, while also relying on lower reserve.

The results presented in table 3 are based on an out-of-sample simulation, in the sense that we generate an entirely new sample of 99360 imbalance intervals (2 years and 10 months), based on the imbalance driver data that has been provided by the regulatory authority, and based on the imbalance model that has been developed in section 3.

An alternative way to approach the simulation could be to implement it in a rolling fashion, in the sense of training a sizing model once a year, based on the data of the past year. One then uses the trained clustering

Table 3: Comparative results of the probabilistic dimensioning versus the sizing of figure 1.

	Fig. 1	Probabilistic
Res-Up (MW)	1392	1111
Unrel.-Up (hours/y)	0.3	0.4
Shortage (MWh/y)	40.2	21.5
Res-Down (MW)	993	950
Unrel.-Down (hours/y)	2.8	1.9
Oversupply (MWh/y)	208.4	159.1

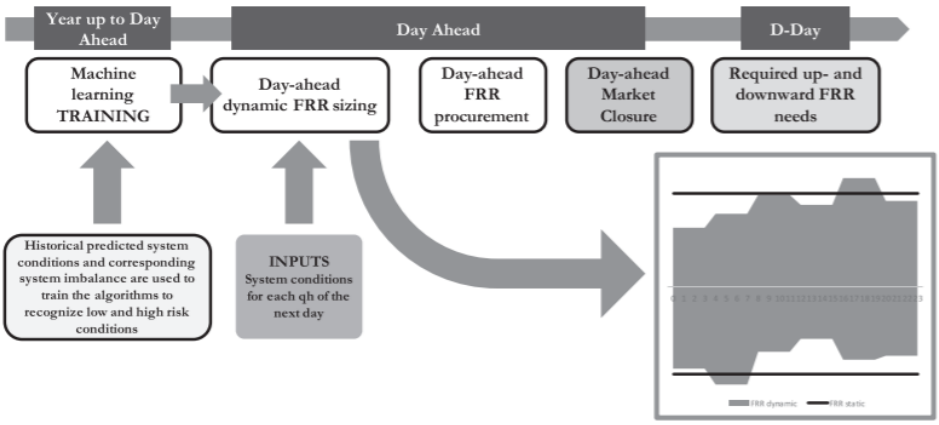

Figure 3: Timeline of a probabilistic dimensioning procedure, where the training of the probabilistic dimensioning algorithm takes place once a year based on the data of the previous year. The figure is sourced from [13].

algorithm one day in advance of operations in order to determine the cluster in which the following day belongs, so as to then decide on the reserve capacity that is procured in the day ahead. This is the approach that has been proposed and considered for implementation in the Belgian system [13]. The timeline of this rolling procedure is depicted in figure 3, which is sourced from [13]. Note that the Greek market operates an Integrated Scheduling Process (ISP) where reserve is committed on a daily basis (with three runs of ISP, in the day ahead as well as two intraday adjustment runs), thus the proposed probabilistic dimensioning procedure is compatible with the timeline of operations of the Greek electricity market.

In table 4 we present the number of intervals that belong to each of the

Table 4: Number of intervals belonging to each cluster of the probabilistic dimensioning method and corresponding number of incidents within each cluster.

Interval type	No. occurrences	Fails Prob. (h/yr)	Fails fig. 1 (h/yr)
LoH-ReH-ImH	13,292 (13.4%)	0.44 (18.5%)	0.44 (14.3%)
LoH-ReH-ImL	9,644 (9.7%)	0.09 (3.7%)	0.09 (2.9%)
LoH-ReL-ImH	12,144 (12.2%)	0.09 (3.7%)	0.26 (8.6%)
LoH-ReL-ImL	10,812 (10.9%)	0.53 (22.2%)	0.09 (2.9%)
LoL-ReH-ImH	7,960 (8.0%)	0.26 (11.1%)	0.18 (5.7%)
LoL-ReH-ImL	6,952 (7.0%)	0.09 (3.7%)	0.09 (2.9%)
LoL-ReL-ImH	26,296 (26.5%)	0.53 (22.2%)	1.59 (51.4%)
LoL-ReL-ImL	12,260 (12.3%)	0.35 (14.8%)	0.35 (11.4%)

clusters of the probabilistic dimensioning method. These are equal in both the training as well as the testing phase, since the same day-ahead data is used for both training and testing. We additionally present the number of failures that occur in each day type in the testing phase, for both the sizing of figure 1 as well as the probabilistic dimensioning. This serves as a measure of the risk assumed by each of the methods.

Concretely, an indication of the extent to which each sizing method is able to maintain a constant level of risk is how well the observed out-of-sample risk of each method is able to track the frequency of each interval type. For example, we observe in table 4 that, although interval type 7 (low load, low renewable supply, high import) corresponds to 26.5% of the intervals in the data sample, the sizing method of figure 1 exhibits a frequency of failures in the seventh interval type which is twice as high as the frequency of this interval type. This indicates that a fixed reserve requirement corresponding to figure 1 tends to under-size for this specific interval type (which corresponds to a quarter of the time).

Figure 4 represents the percentages of table 3 visually. The closer the curves remain to the blue curve, the more consistent they are in terms of maintaining a constant risk. Deviating too far above the blue curve indicates an exposure to a disproportionately high risk (i.e. undersizing), while deviating too far below the blue curve indicates an exposure to a disproportionately low risk (i.e. oversizing). It is clear that the probabilistic dimensioning method is able to remain closer to the blue curve, thereby indicating an improved risk profile relative to the sizing of figure 1. Similar conclusions emerge in the probabilistic dimensioning method that is employed in Belgium [13].

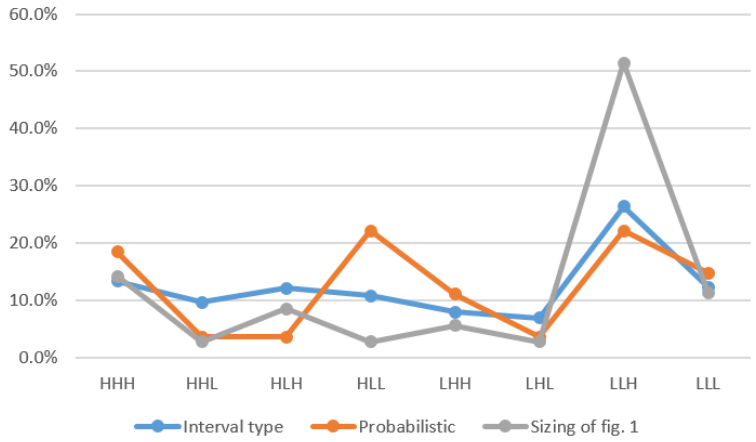

Figure 4: Frequency of each interval type, and frequency of failures for the probabilistic dimensioning method and the sizing of figure 1.

.

Part II
Scarcity Pricing

6. Introduction to Scarcity Pricing

In this section we introduce scarcity pricing. Scarcity pricing is a disciplined approach for pricing energy in periods of system stress at a value above the marginal cost of the marginal unit. The concept relies on quantifying the value of reserve towards enhancing system reliability. This value can be quantified using loss of load probability calculations. Thus, probabilistic computations are the unifying theme with the first part of the report. In part I, loss of load probability is used for quantifying reserve requirements. In part II, loss of load probability is used for computing the value of reserve to the system and adapting energy and reserve prices accordingly.

6.1 Motivation for Scarcity Pricing

The increasing integration of renewable resources in electric power systems is inducing a value shift in electricity markets. The low marginal cost of renewable resources such as wind and solar power is exerting a downward pressure on marginal costs, due to the near-zero marginal cost of these resources, which are not required to burn fuel in order to produce energy. Despite their attractive marginal cost, renewable resources are inherently unreliable. Therefore, the more renewable capacity that we integrate in the system, the more we need reserves in order to operate the system reliably. Our focus in the subsequent discussion, as in part I, will be on aFRR and mFRR.

Reserves are typically provided by resources with high marginal cost, such as CCGT units. Due to their high marginal cost, such resources are the first to be pushed out of the market when the supply function of the market is shifted to the right due to the integration of renewable resources, with near-zero marginal cost. This leads to the following paradoxical effect: although the integration of renewable resources increases the need for high-cost flexible resources such as CCGT units, these are also the units that are facing the greatest economic pressure as a result of the low marginal cost of renewable resources. This effect has been documented by Capros [9] insofar as the Greek electricity market is concerned. The problem is also relevant for the Belgian electricity market [56], where a strategic reserve mechanism has been employed in the past in order to resuscitate CCGT units that are headed for mothballing or retirement.

This apparent paradox is the result of incomplete market design, and in particular the absence of a real-time market for reserve capacity in the European balancing market [52]. Concretely, reserves offer value to the sys-

tem by virtue of increasing system reliability. This reliability value should in principle be reflected (and paid for) in real time by a reserve imbalance mechanism. More renewable resources in the system would then imply a higher value for real-time reserve capacity, and higher payments to those resources that can offer this reserve capacity. This would create a revenue stream for reserve resources that would contribute towards covering their missing money. This would also uplift balancing and imbalance prices (due to the fact that real-time energy prices should be such that the profit margin in the real-time energy market should equal the profit margin in the real-time reserve market), thereby lifting the real-time value of energy prices above the marginal cost of the marginal unit. These two effects, in tandem, work towards resolving the apparent paradox described above, and enable a future-proof market design where flexible resources, even those with high marginal costs, can survive in a competitive electricity market with large amounts of installed renewable supply capacity. This same mechanism further mobilizes demand response, storage, and other sources of flexibility, thereby generating a sustainable economic signal for attracting flexibility in electricity markets with significant levels of installed renewable capacity.

6.2 EU Legal Context

Scarcity pricing is becoming increasingly relevant in European legislature. The predominant examples in this respect are article 44(3) of the European Commission Electricity Balancing Guideline [23] and article 20(3) of the European Parliament Clean Energy Package [25].

Concretely, article 44(3) of [23] reads as follows:

"*Each TSO may develop a proposal for an additional settlement mechanism separate from the imbalance settlement, to settle the procurement costs of balancing capacity pursuant to Chapter 5 of this Title, administrative costs and other costs related to balancing. The additional settlement mechanism shall apply to balance responsible parties. This should be preferably achieved with the introduction of a* ***shortage pricing function****. If TSOs choose another mechanism, they should justify this in the proposal. Such a proposal shall be subject to approval by the relevant regulatory authority.*"

Note that the text suggests that scarcity pricing (referred to in the text as shortage pricing) should be considered as a default. Although scarcity pricing and capacity mechanisms co-exist (as we discuss in section 8.5), precedence matters [52]: the implementation of scarcity pricing is a no-regret measure towards improving electricity market design; if capacity markets are further needed for restoring missing money, this can be accommodated in a

market with scarcity pricing.

In a similar spirit, article 20(3) of the the Clean Energy Package [25] reads as follows:

"*Member States with identified resource adequacy concerns shall develop and publish an implementation plan with a timeline for adopting measures to eliminate any identified regulatory distortions or market failures as a part of the State aid process. When addressing resource adequacy concerns, the Member States shall in particular take into account the principles set out in Article 3 and shall consider:*

...

(c) introducing a ***shortage pricing function*** *for balancing energy as referred to in Article 44(3) of Regulation 2017/2195; ...*

"

The CEP article thus makes a direct reference to article 40(3) of the EBGL [23]. These articles open the door to the introduction of scarcity adders in balancing energy prices (used for settling BSPs) and imbalance prices (used for settling BRPs).

6.3 Scarcity Pricing Based on ORDC

At a high level, scarcity pricing is the pricing of energy at a value that is higher than the marginal cost of the marginal unit, which aims towards contributing to the mitigation (or eradication) of the missing money problem. The specific instantiation of scarcity pricing that rationalizes this design based on the value of reserve capacity has been developed in detail by Hogan [32], with subsequent refinements of the original theory [31]. The adaptation of the theoretical framework to European balancing markets has been developed by Papavasiliou [52, 53].

In order to understand the mechanism, let us first commence from a simple economic dispatch model:

$$\begin{aligned}
(ED): \quad & \max_{p\geq 0, d\geq 0} VOLL \cdot d - \sum_{g\in G} MC_g \cdot p_g \\
(\lambda): \quad & d - \sum_{g\in G} p_g = 0 \\
(\mu_g): \quad & p_g \leq P_g, g \in G \\
(\nu): \quad & d \leq D
\end{aligned}$$

Here, G represents the set of generators in the market. The decision variables are the production p_g of unit g and the demand d. The parameter $VOLL$

is the value of lost load. Parameter MC_g represents the marginal cost of unit g. The maximum production of unit g is represented as P_g, and D corresponds to the inelastic demand in the system. The dual multipliers of the model have economic interpretations: λ is the equilibrium energy price, μ_g is the scarcity rent of generator g, and ν is the scarcity rent of consumers.

The KKT conditions of the problem (we do not repeat equality constraints) can be described as follows:

$$
\begin{aligned}
&0 \leq p_g \perp MC_g - \lambda + \mu_g \geq 0, g \in G \\
&0 \leq d \perp \lambda - VOLL + \nu \geq 0 \\
&0 \leq \mu_g \perp P_g - p_g \geq 0, g \in G \\
&0 \leq \nu \perp D - d \geq 0
\end{aligned}
$$

Let us consider a scarcity situation, in the sense of a situation where demand cannot be served. Then, since $0 < d < D$, the complementarity conditions above imply that $\lambda = VOLL$, so the equilibrium energy price is the value of lost load. For situations where capacity is adequate, the equilibrium energy price can be set to $\lambda = MC_g$, where g is the marginal cost of the marginal unit. This design can, in theory, support an optimal investment plan. In practice, however, scarcity periods (which correspond to periods where the system is out of capacity, hence $\lambda = VOLL$) occur at a frequency that is very challenging for investors to anticipate, and price caps are sometimes enforced during these scarcity periods.

These two effects, risk and price caps, act towards discouraging investment. The latter results in a missing money problem. We will refer to the above design as an *energy-only market with VOLL pricing* [62].

Scarcity pricing based on operating reserve demand curves (ORDC) mitigates the issues described above by allowing the market to better reflect the value of reliability, and hence also the value of energy during periods when the system is tight. We can use the following underlying model, which generalizes the energy-only model, in order to fix concepts:

$$
\begin{aligned}
(EDR): \quad & \max_{p \geq 0, d \geq 0, r \geq 0, dr \geq 0} VOLL \cdot d + \sum_{l \in LR} VR_l \cdot dr_l - \sum_{g \in G} MC_g \cdot p_g \\
(\mu_g): \quad & p_g + r_g \leq P_g, g \in G \\
(\lambda): \quad & d - \sum_{g \in G} p_g = 0 \\
(\lambda R): \quad & \sum_{l \in LR} dr_l - \sum_{g \in G} r_g = 0
\end{aligned}
$$

$$(\nu): \quad d \leq D$$
$$(\nu R_l): \quad dr_l \leq DR_l, l \in LR$$

The additional notation introduced here can be explained as follows. The set LR indicates the TSO balancing capacity bids: bid $(VR_l, DR_l), l \in LR$, is a bid to buy up to DR_l units of balancing capacity at a price of no more than VR_l. The variable dr_l indicates how much of bid $l \in LR$ is accepted in the auction. The dual variable λR represents the equilibrium price of reserve, while μR_l represents the net profit of reserve bid l. Note the linking of energy and balancing capacity bids for resource g: the first constraint of the model is requiring that activated bids for energy and balancing capacity cannot exceed the total capacity P_g of supplier g.

Note that the collection of TSO demands $\{(VR_l, DR_l), l \in LR\}$ constitutes an *operating reserve demand curve* (ORDC). *All markets have ORDCs*, even those with hard reserve requirements (e.g. Greece, Belgium): it is simply that these ORDCs are price-inelastic, i.e. they consist of a single bid for DR_l equal to the reserve requirement and VR_l equal to a very high penalty factor [28]. Other markets have multiple steps in their demand curve (e.g. MISO, ISO-NE, NYISO), and others yet base these ORDCs on loss of load probability (e.g. Texas, and the proposed design in Belgium). What really distinguishes scarcity pricing based on ORDC is not the presence or absence of ORDCs (since all markets have ORDCs), rather whether or not a real-time market for reserve is present. All US markets have a real-time market for reserve. The EU balancing design is missing such a market [52], although Belgium is contemplating the introduction of such a market in the context of its scarcity pricing reforms.

Note that the model can be generalized in a number of ways: (i) ramp rates apply, (ii) multiple reserve products can be introduced, (iii) demand can also offer reserve, and so on. Here we are rather interested in fixing concepts, and therefore develop the discussion with the simplest possible example that brings home the key insights.

Let us now consider scarcity conditions in this model. We can consider two levels of scarcity:

- Energy shortage: $0 < d < D$
- Shortage of reserve: $0 < d_l < DR_l$ for some $l \in LR$

We analyze each of these situations in turn using the KKT conditions, which generalize those of the energy-only market with VOLL pricing:

$$0 \leq p_g \perp MC_g - \lambda + \mu_g \geq 0, g \in G$$

$$
\begin{aligned}
&0 \leq d \perp \lambda - VOLL + \nu \geq 0 \\
&0 \leq \mu_g \perp P_g - p_g - r_g \geq 0, g \in G \\
&0 \leq \nu \perp D - d \geq 0 \\
&0 \leq \nu R_l \perp DR_l - dr_l \geq 0, l \in LR \\
&0 \leq dr_l \perp \lambda R - VR_l + \nu R_l \geq 0, l \in LR \\
&0 \leq r_g \perp \mu_g - \lambda R \geq 0, g \in G
\end{aligned}
$$

In the first case (energy shortage), the same comments as in the energy-only model apply: the energy price becomes $\lambda = VOLL$, as one can conclude from the second and fourth complementarity conditions above.

In the second case (reserve shortage), we have $dr_l < DR_l$ implying $\nu R_l = 0$. And since dr_l, it follows that $\lambda R = VR_l$, hence it is the l-th segment of the ORDC that sets the price. Assuming that there exists a generator g that is splitting its capacity between energy and reserve, we can conclude that, for this generator, we have $p_g > 0$ and $r_g > 0$. Then from the first complementarity condition we have $\mu_g = \lambda - MC_g$, i.e. the profit of the generator is the energy price minus its marginal cost. Then from the last complementarity condition we have $\mu_g = \lambda R$, hence the profit margin of the generator is also equal to the price of reserve. Thus, we arrive to the following key scarcity pricing relationship:

$$\lambda = MC_g + \lambda R \tag{1}$$

In other words, the energy price is no longer simply equal to the marginal cost of the marginal unit, but to the marginal cost of the marginal unit plus an ORDC "adder", or the price of reserve.

Note that this identity does not hold in more complex settings with ramp rates and multiple reserve products, but serves as an approximation for these more complex settings and has been the basis of the scarcity pricing mechanism in Texas [22] as well as the Belgian scarcity pricing proposal [58].

What this identity implies in practice is the following, which has been the basis of the scarcity pricing market design proposal for Belgium [58, 52]:

- BSPs that trade energy in real time are settled not only the marginal cost of the marginal unit, but an additional adder λR
- BRP real-time imbalances are settled against the same price
- BSPs are settled for their real-time reserve against λR

The design is intended to overcome the aforementioned drawbacks of energy-only markets with VOLL: (i) The mechanism is intended to result in more frequent scarcity pricing periods, and therefore more predictable energy streams for flexible resources. (ii) Market power mitigation can be enforced without suppressing energy prices: even if BSP bids are truthfully submitted, energy prices can be uplifted during periods of scarcity because the ORDC sets the price for reserve capacity λR, which in turn uplifts the balancing and imbalance price λ.

7. International Experience with the Implementation of Scarcity Pricing with ORDCs

All markets have ORDCs. Fixed reserve requirements are a specific form of ORDC. The real distinction between markets that implement scarcity pricing with ORDC and those that do not is therefore not whether an ORDC exists or not, but rather whether a real-time market for reserve is put in place or not. US ISOs typically have such a mechanism in place, whereas the European balancing market design is handicapped in this respect [52].

In what follows, we summarize as past and ongoing efforts for implementing scarcity pricing in Europe, summarize the pioneering efforts of ERCOT and PJM for implementing scarcity pricing, and overview the design of a number of other US markets. In section 8 we deep-dive in the description of the Belgian scarcity pricing proposal.

7.1 Belgium [52]

The Belgian regulatory authority began investigating the implementation of scarcity pricing in 2014. Since then, a number of reports have been published related to the implementation of scarcity pricing in Belgium.

The first study that was conducted on behalf of the Belgian regulatory authority [56] focused on the topic of how electricity prices would change if we were to introduce ORDC in the Belgian market. The report found that such a market design change could enable the majority of combined cycle gas turbines, which were at the time operating at a loss in Belgium, to recover their investment costs

The second study [57], which was performed in 2016, investigated how scarcity pricing depends on a number of factors, including the presence of strategic reserve, the value of lost load, the restoration of nuclear capacity in Belgium, and the day-ahead (instead of month-ahead) clearing of reserves.

Following these two studies, the Belgian regulator launched a study for developing a market design proposal which would outline a possible way forward for the implementation of scarcity pricing in Belgium. The underlying question was whether we can adopt a US-inspired design and plug it into the existing European market? The resulting analysis [58] underscored the essential role of a real-time market for reserve capacity for the back-propagation of ORDC adders to forward reserve markets

In 2018, the Belgian TSO ELIA performed an ex-post simulation [19] of how scarcity prices would have transpired in the Belgian market for 2017, given the telemetry data recorded for that year. The finding was

that 2017 was a comfortable year, during which there would be an infrequent occurrence of non-zero scarcity pricing adders. Since October 2019, ELIA publishes adders one day after operation. The adders are published in the following link: (https://www.elia.be/en/electricity-market-and-system/studies/scarcity-pricing-simulation).

A number of recent studies have focused on various practical aspects related to the implementation of scarcity pricing in Belgium. The interaction of scarcity pricing with the MARI balancing platform is investigated in [51], where the interaction of the market design proposal with foreign BSPs is clarified. A crucial aspect related to the implementation of scarcity pricing is to introduce ORDC adders for the following purposes:

- imbalance energy
- balancing energy, and
- the settlement of reserve imbalances.

This is crucial in order to achieve a back-propagation of ORDC adders to forward reserve capacity markets. This design is justified in detail in [52] and also using an analytical model in [53].

In its national implementation plan [26], the Belgian government argued that its current imbalance settlement approach exhibits characteristics of scarcity pricing. The response of the European Commission [24] was that (i) an adder which depends on the remaining available reserve in the system should be contemplated, and (ii) an alignment should be considered between balancing and imbalance prices. These two positions are perfectly aligned with the market proposal set forth in [52], nevertheless it remains crucial to also apply the ORDC adder to reserve imbalance settlement in order for reserve value to back-propagate effectively to day-ahead reserve markets [52].

In October 2020 the Belgian TSO ELIA released a public consultation regarding the scarcity pricing proposal set forth in [58]. The ELIA scarcity pricing proposal is limited to an application of ORDC adders in imbalance prices, but the new element is the so-called "Omega" component, which indeed responds to the market design proposal of [58] and the request of the European Commission [24] to make the scarcity adder dependent on the remaining real-time reserve capacity in the system. The public consultation of ELIA is currently being evaluated by the Belgian regulatory authority.

7.2 UK [58]

The UK system operator balances the system using a mix of balancing market bids (the analog of free bids, in Belgian market terminology) and the so-called Short-Term Operating Reserve (STOR), which is the analog of frequency responsive reserve (aFRR and mFRR) in Belgian market terminology. It appears that the UK market does not involve real-time reserve capacity payments, but only real-time energy payments.

Balancing market bids can adjust their activation cost in real time. By contrast, STOR receives so-called availability payments which can be interpreted as activation costs for energy, but the value of which is not closely linked to the real-time stress of the system but is rather based on an ex-post calculation (see article 3.46 and figure 3.48 of [27]). In this sense, STOR unit owners do not have the freedom of adapting their real-time bids for energy. This creates a challenge in creating a real-time energy signal which accurately reflects scarcity.

For this reason, the UK regulator (OFGEM) recently proposed the introduction of a real-time operating reserve demand function that would set the real-time energy price and more accurately reflect scarcity in the system. The UK ORDCs were introduced in early winter 2015/16 (article 3.51, [27]). The ORDC is constructed by using the product of VOLL with loss of load probability as a function of available reserve capacity. The original estimate for VOLL was equal to 3000 British pounds per MWh, and was planned to be raised to 6000 British pounds per MWh by early winter 2018/2019 (article 3.55, [27]).

The intent of OFGEM is to use a dynamic LOLP for computing ORDC, as is the case in Texas, and as also recommended in [58] for the Belgian market. The LOLP uses both STOR capacity, as well as free balancing bids, as recommended by Papavasiliou [58] for the Belgian market. The LOLP is computed using data of the current balancing interval, and the LOLP information is published shortly thereafter to balancing market participants. Indicative LOLPs are recommended to be published shortly in advance of real time (e.g. 4, 3 or 2 hours ahead of real time) in order to reduce the risk of balancing market participants. Advance information is not required in central dispatch systems such as Texas or PJM, since it is the system operator that dispatches resources in a way that is automatically consistent with real-time prices.

7.3 ERCOT [58]

7.3.1 Performance of ORDC Adder in the Texas Market

In 2021 the ERCOT system experienced an extreme cold weather front. This led to severe power outages related to disruptions in the gas supply chain, increased demand, and a number of other shocks to system operation. In such scarcity conditions, where involuntary demand curtailment takes place, prices anyway soar to value of lost load, and price formation due to ORDCs is overshadowed or displaced by price formation due to energy demand shortages. Pending a market monitoring report for 2021, we repeat here certain statistics sourced from [58].

According to the Potomac 2017 state of the market report [60], the ORDC adder contributed to 0.24 $/MWh to the real-time price, which corresponds to less than 1% of the annual average real-time price in Texas. This is due to the fact that the system was rarely short of reserves in 2017. The most notable impact of ORDC on real-time energy prices was in July and August of 2017.

7.3.2 Loss of Load Probability

ERCOT calibrates ORDCs on the basis of the so-called historical reserve error, which is the difference between the hour-ahead available reserves and the real-time available reserves. The amount of hour-ahead reserves measured in ERCOT is the difference between generator capacities and load forecasts. In this sense, "free bids" (i.e. resources that are available in real time, even if they have not been cleared for reserve capacity) are counted towards real-time available capacity. The real-time reserve is computed as the difference between the measured generator capacities and their set-points determined by security constrained economic dispatch (SCED).

ERCOT assumes a normal distribution for the reserve error, and estimates a mean and standard deviation for characterizing this distribution. The parameters of the normal distribution are changed in 4-hour blocks for every season. The ERCOT ORDC is computed based on a value of lost load of 9000 $/MWh, and implements a minimum operating reserve requirement of 2000 MW. This is referred to as the minimum contingency level.

7.3.3 ORDC adder

ERCOT computes one adder for resources that can respond within 30 minutes (which is called the real-time online reserve price adder), and one adder

for resources that can response within 60 minutes (which is called the real-time offline reserve price adder). Both of these adders are computed every five minutes, i.e. every time that SCED runs. However, because the Texas real-time market settles transactions every 15 minutes (as opposed to every 5 minutes), the real-time online and offline reserve price adders are averaged over three time intervals into the real-time reserve price for online and for offline reserve respectively.

ERCOT currently does not perform a real-time co-optimization of energy and reserves. Nevertheless, the Texas market computes real-time reserve prices by virtue of the ORDC adder. Resources that can respond within 30 minutes are paid the online reserve price, while resources that can respond within 60 minutes are paid the offline reserve price. This idea has inspired the Belgian scarcity pricing proposal [58].

ERCOT uses hourly estimates of the reserve error for the offline reserve adder, and uses an assumption of independent increments for estimating the online reserve price. Concretely, if μ and σ are the hourly mean and standard deviation for the offline adder, the online adder is computed by assuming a mean of $0.5 \cdot \mu$ and a standard deviation of $0.707 \cdot \sigma$. This is also explained in [31].

7.3.4 How Many Adders?

Texas has the following reserve products: regulation up (responds in 3-5 seconds), regulation down (responds in 3-5 seconds), restoration reserve services (responds within 10 minutes), and non-spin (responds within half an hour). Day-ahead prices for these reserves are posted with hourly resolution.

The ORDC adder is broken into two components ([22], slide 29), one component that corresponds to capacity that can be made available immediately (the online adder), and one component that corresponds to capacity that can respond in 30 minutes (the offline adder). We therefore interpret the online adder as being applicable to reserve capacity that can be offered by regulation and restoration reserve, and the offline adder as being applicable to reserve capacity that can be offered by all reserves (including non-spinning reserve). The change in the energy price is driven by the online adder.

7.4 PJM [58]

PJM is currently moving forward with the implementation of ORDC [33] in the day-ahead and real-time market. An important ongoing debate in PJM

is about whether the ORDC should be present in the day-ahead market clearing model. The argument in favor of introducing it to the day-ahead market clearing is that it would better align real-time and day-ahead pricing.

The construction of an ORDC in PJM would involve two steps. The first step is the computation of a minimum reserve requirement (Texas has a corresponding quantity in the real-time ORDC, which amounts to 2000 MW). The idea is to compute this quantity by using forced outage rate data, load forecast error data, and adapting this quantity by season and also by the occurrence of extreme weather events. The second step is the construction of the part of the curve that would relate to loss of load probability, and would rely on the same data.

PJM currently operates two tiers of reserve products. Tier 1 has a 10-minute response time, an obligation to respond, is subject to a non-compliance penalty, and is paid for response to an event. It is available generation capacity that is synchronized and can be loaded within 10 minutes. Tier 2 has a 10-minute response time, an obligation to respond, is subject to a non-compliance penalty, and is paid the market clearing price regardless of deployment. It consists of resources that are committed or dispatched out of merit in order to provide reserves. In reforming its reserve market, PJM is interested in consolidating tier 1 and tier 2 reserves.

PJM plans to propose ORDC for all reserve products (tier 1 and tier 2). PJM uplift is currently 300 thousand dollars per day on average. The introduction of ORDC is expected to reduce this uplift. The distribution of uncertainty that will be used for the computation of the ORDC will be based on real-time load forecast errors, renewable (solar and wind) forecast errors, and conventional generator failures. This effort is driven to a significant extent by the projected increase of wind nameplate capacity installation to 36,159 MW by 2029 (it is currently at approximately 10 GW).

PJM currently performs real-time dispatch with 5-minute frequency. The penalty factor for falling below a capacity of 1500 MW is currently set at 850 $/MWh, and to 300 $/MWh for falling below a capacity of 1.7 GW. The goal is to introduce a downward sloping ORDC curve in the real-time dispatch. The curve jumps to 2000 $/MWh at the 1.5 GW level.

In addition to consolidating its tier 1 and tier 2 products, PJM is moving to three types of reserve requirements. These are synchronized, primary, and 30-minute reserve. In implementing ORDC, each reserve product will be associated with a penalty factor, a minimum requirement, and an associated probability distribution. The idea is that: (i) spinning reserve will contribute to all requirements; (ii) non-spinning reserve will contribute to primary reserve and 30-minute reserve, and secondary reserve will contribute

to the 30-minute requirement. PJM proposes to base its reforms on a 30-minute look ahead for uncertainty for the synchronized and primary reserve requirements and a 60-minute look ahead for the 30-minute reserve requirement. For the first half of the 60-minute period, the 30-minute uncertainty will apply only to the valuation of two types of 10-minute reserves. For the second half of the period, the full one-hour uncertainty will apply to the combined levels of 10-minute and 30-minute reserves.

7.5 New York ISO

The New York ISO operates a day-ahead and real-time market for locational reserves of different types [46]. There are four types of reserve in NYISO:

- Regulation, which follows a 6-second set-point (and is thus akin to European aFRR products).
- Spinning reserve with a 10-minute response time.
- 10-minute total reserve, which includes spinning reserve and 10-minute non-synchronized reserves.
- 30-minute reserve, which includes synchronized and non-synchronized reserve with a response time of 30 minutes.

Operating reserves refer to the second, third and fourth reserve product category above. Operating reserve demand curves are used for these three products. ISO-NE also uses a day-ahead and a real-time demand curve for regulation. Interestingly, regulation shortages were the most frequent to occur in the interval from July 2016 until July 2019, when they occurred 9% of the time.

The pricing of ISO-NE accounts for the substitutability of different reserve products, and thus the prices of higher quality reserves are guaranteed to be no less than the prices of lower quality reserves.

The New York ISO uses step curves for constructing its operating reserve demand curves, with no clear connection to VOLL or loss of load probability. The NYISO market monitor has recently recommended that NYISO consider VOLL and loss of load probability as factors for designing its operating reserve demand curves [46].

7.6 ISO New England

ISO New England defines the following reserve products [46]:

- Local 30-minute operating reserve.
- System 30-minute operating reserve.
- System 10-minute non-synchronized reserve.
- System 10-minute spinning reserve.

ISO-NE currently employs stepped operating reserve demand curves. As an alternative, ISO-NE is considering an estimation of loss of load probability on the basis of the risk of forced outages, see figure 9 of [46].

In addition to scarcity pricing, ISO-NE currently applies pay-for-performance incentives related to the capacity market that is put in place in the ISO-NE market. Nevertheless, as noted by [46]:

"*Pay-for-performance programs do not utilize real-time energy market prices to incent resources. Market-based price signals and ancillary services shortage pricing enables the NYISO to incent resource performance.*"

7.7 Other US ISOs

MISO distinguishes two types of reserve in its market, spinning and non-spinning reserve. The ISO relies on a stepped operating reserve demand curve, and accounts for reserve substitutability by employing additive reserve prices. It has recently been proposed [46] that MISO move to an operating reserve demand curve based on loss load probability and value of lost load. The proposed VOLL amounts to 12000 $/MWh, and the proposed procedure for estimating loss of load probability is to be based on forced generator outages, load forecast error, and scheduled interchange error.

The California ISO defines two types of operating reserves, 10-minute spinning reserves and 10-minute non-spinning reserves. CAISO employs a stepped operating reserve demand curve.

SPP defines two reserve products, 10-minute spinning reserves and 10-minute non-synchronous reserves [46]. SPP utilizes a stepped operating reserve demand curve.

8. Detailed Description of the Belgian Mechanism

In this section we describe the Belgian scarcity pricing mechanism in further detail. The implementation of scarcity pricing in Belgium is ongoing, and the author has supported the Belgian regulatory authority and system operator in this implementation since 2014, when the implementation of the mechanism in Belgium was first investigated. The present chapter highlights some of the main features of the proposed mechanism. Section 8.1 describes the calibration of the ORDC curve. Section 8.2 describes the adaptation of the scarcity pricing theory based on ORDC to the European balancing market design. Section 8.3 discusses the application of the theory to the pricing of multiple products. Section 8.4 discusses the unilateral implementation of the mechanism and its integration with EU balancing platforms, whereby a single Member State proceeds with implementation even if neighboring Member States do not follow suit. Section 8.5 discusses the relation between scarcity pricing and capacity markets.

8.1 Estimation of the ORDC

The proposal for the estimation of an ORDC for the Belgian market [58, 19] is based on the computation of loss of load probability based on historical imbalance data. The connection between loss of load probability and a demand curve for operating reserve is established by Hogan [31] in the appendix of his article. The argument is also developed in section 2.1 of [56]. The idea is that, if one ignores ramp constraints in a model with a single reserve product, a demand curve that reproduces the equilibrium energy price of an energy and reserve co-optimization is given by the following expression:

$$V(R) = (VOLL - \hat{MC}) \cdot LOLP(R) \tag{2}$$

Here, $V(R)$ is the ORDC, $VOLL$ is the value of lost load, $\hat{MC}$ is a proxy of the marginal cost of the marginal unit, $LOLP(\cdot)$ is a function that maps the level of reserve in the system to loss of load probability, and R is the amount of available reserve capacity in the system. The $LOLP$ function can be estimated on the basis of historical system imbalance data, as follows:

$$\begin{aligned} LOLP(R) &= \mathbb{P}[Imb \geq R] \\ &= 1 - \mathbb{P}[Imb \leq R] \\ &= 1 - \Phi(\frac{R-\mu}{\sigma}) \end{aligned}$$

where μ is the mean and σ is the standard deviation of the system imbalance and Φ is the cumulative distribution function of the standard normal distribution.

The Belgian system operator records imbalance data with one-minute resolution (see footnote 13 in [57]). One can then use this data in order to estimate the mean and the standard deviation of imbalances for different parts of the year. We propose 24 ORDCs for the implementation of scarcity pricing in Belgium [58], following the example of ERCOT: one for each season, and six for each four-hour block of the day. What is required for the estimation of the demand curves is the mean and standard deviation of the system imbalance.

There are important ongoing discussions about the details of how equation (2) is applied [64]. Three specific attributes of the formula are being currently investigated in the Belgian market design proposal [58]:

- Correlated versus independent imbalances: When estimating the standard deviation of the system imbalance, do we assume that imbalance increments within an imbalance interval are independent or correlated? We revisit this question in section 8.3, since it relates to the scarcity price formulas for aFRR and mFRR capacity. An assumption of correlated imbalances implies that imbalance increments within a balancing interval are assumed to be perfectly correlated, whereas an assumption of independent increments implies that these increments are independent of each other. The former would imply that the imbalance increment in the first part of an imbalance interval exhibits lower variance [31], and would therefore imply a narrower ORDC for aFRR capacity. Empirical evidence based on statistical analyses of Belgian imbalance data [57] suggests that the assumption of correlated increments is better aligned with physical observations.

- Before- or after-activation variants: The scarcity pricing formula computes the loss of load probability as a function of R. In this design variation we examine whether R should be considered as the remaining available reserve capacity before or after clearing the imbalance in the current imbalance interval. Measuring the remaining reserve capacity after the clearing of an imbalance has the potential to increase the scarcity prices during tight periods of operation when the system is stressed significantly.

- Value of VOLL: Our analysis considers two possible values of VOLL, namely 13,500 €/MWh and 8,300 €/MWh, with the latter being the

estimate of the Belgian Federal Planning Bureau [57, 14]. A higher value for VOLL clearly implies higher scarcity prices. Similar design dilemmas have been posed in the implementation of scarcity pricing in a number of US markets [46].

Ongoing work on behalf of the Belgian regulatory authority is testing these design choices by explicitly considering the tradeoff between operational security and cost. Concretely, our team has developed a bottom-up simulation model of the day-ahead, intraday and real-time operations of the Belgian system. The goal is to quantify how different choices of ORDCs balance the fixed cost of carrying reserve against the increased security that the system enjoys when equipped with more reserve capacity. This analysis supports the proposal of the Belgian regulator for the ORDC formulas that will be proposed for the implementation of scarcity pricing in Belgium.

In order to compute the scarcity prices of equation 2, the parameters μ and σ of the LOLP function can be estimated once a year, based on the data of the past year. Then, in real time, the remaining available reserve capacity R can be measured based on telemetry data. The tighter the system, the lower the remaining reserve capacity in the system, and the higher the adder computed in equation 2. This leads to an increasing remuneration of flexible assets that can respond to system needs when the system is tightest. The mechanism is therefore by construction paying for performance.

8.2 BRP and BSP Settlements Implied By Scarcity Pricing

The implementation of scarcity pricing based on operating reserve demand curves is a straightforward process in markets that co-optimize energy and reserves. The fundamental energy and reserves co-optimization model is described as model (EDR) in section 6.3. One implements scarcity pricing by using the dual multiplier λ of this model in order to price real-time energy and the dual multiplier λ^R in order to settle reserve. The ORDC is the term $\sum_{l \in LR} VR_l \cdot dr_l$ that appears in the objective function of this model. The presence of this demand function in the model introduces price elasticity for reserve, which implies that reserve prices increase when the system is tight. Because of equation (1), energy prices also increase by the scarcity adder λ^R.

Interestingly, the European balancing market is of the type (ED), described in section 6.3. More specifically, there is no real-time market for reserve capacity [52]. This is a remarkable oversight of EU market design, which drastically undermines European flexible technologies (as also corroborated empirically in [9] for the case of the Greek market, and in Belgium

with the presence of strategic reserve). Equally remarkable is the fact that we operate forward (e.g. day-ahead) markets for reserve in Europe, even though we have no real-time markets for those same reserves. This implies that any price formation related to reserve in day-ahead markets cannot be driven by the back-propagation of the value of reserve in real time (as determined by an ORDC), since there is no real-time market for reserve. Non-zero day-ahead reserve prices may be driven by other factors, e.g. (i) the fixed commitment costs for bringing reserves online, or (ii) the ability of market agents to affect reserve prices through their bidding behavior in tight system conditions (see for example the increase in mFRR prices in Belgium in November 2018, section 6.1.7 of [20]). No matter what the factors that drive day-ahead reserve price formation are, they are certainly not related to the back-propagation of the real-time value of reserve capacity.

Another important distortion of the EU balancing market design is the misalignment between BSP settlements and BRP settlements, with the former being settled at balancing prices and the latter being settled at imbalance prices. The operational distinction between BRPs and BSPs is significant from a system operation perspective (since BSPs effectively act as reserve resources, whereas BRPs are the entities causing the needs for such balancing actions). The economic distinction between the two is that the former correspond to price-inelastic producers / consumers of real-time energy, whereas the matter correspond to price-elastic producers / consumers of real-time energy. There seems to be no clear economic justification of why the two would face different prices for the trade of real-time energy, and even if they do, they are in a position to arbitrage it away by virtue of the fact that a BSP can present itself as a BRP and therefore respond to the price that is most favorable (be it the balancing or the imbalance price).

These two distortions of EU balancing market design may at first glance obscure the application of the first principles of scarcity pricing to the European market. A number of creative combinations have been proposed by industry stakeholders for "implementing" scarcity pricing, which fail to correctly apply the first principles of scarcity pricing. The most persistent misunderstanding is that applying the adders of formula (2) as a top-up to imbalance prices is sufficient for implementing scarcity pricing [20, 29]. What one can demonstrate through an analytical model of perfect competition [52, 53] is that this induces low-cost BSPs to preemptively assume long positions by inducing active imbalances in their portfolios, in order to exploit the asymmetry between balancing and imbalance prices when the scarcity adders are wrongfully applied to imbalance prices alone. This behavior leads to two adverse side-effects: (i) BSPs keep flexibility out of the

balancing market, and thereby restrict the options of the TSO for balancing the system in real time. (ii) In assuming positive imbalances within their portfolio, the BSPs depress balancing prices and thereby cancel out the effect that scarcity pricing aims to achieve by uplifting energy prices.

A correct application of the first principles of scarcity pricing relies on two economic principles [59]:

- **Economic principle 1**: the law of one price [10]. Real-time energy is a unique product, therefore the buyer and seller should exchange it at the same price.
- **Economic principle 2**: back-propagation. If we put in place a real-time market for reserve capacity, then agents will only sell reserve capacity in forward markets at the value that they would need to buy it back in real time. This second principle is especially crucial, since it allows the value of reserve capacity to back-propagate into forward reserve auctions, and send the signal to investors that the market can support investments in reserve capacity.

These principles rationalize the following market design proposal for implementing scarcity pricing [58, 59]:

- **Market design proposal 1**: the introduction of a scarcity adder to the imbalance price.
- **Market design proposal 2**: the application of the same adder to the balancing energy price.
- **Market design proposal 3**: the implementation of an EU real-time market for reserve capacity (equivalently, a market for "reserve imbalance", in the same way that we operate a market for energy imbalances), which is a missing market in the existing EU balancing design.

The application of the mechanism as proposed above achieves two desirable effects simultaneously, which alternative proposals [20, 29] fail to achieve: (i) BSPs are induced to offer their entire flexible capacity to the balancing market, and (ii) real-time scarcity prices back-propagate to forward (e.g. day-ahead) reserve markets.

The above settlement rule is best illustrated with the following example. Consider a generator that can provide secondary reserve. Suppose that the day-ahead energy price is $\lambda^{DA} = 20$ €/MWh, and that the day-ahead price

Table 5: Example of BSP settlement without scarcity pricing.

Settlement type	Formula	Price [€/MWh]	Quantity [MW]	Cash flow [€/h]
DA energy	$\lambda^{DA} \cdot p^{DA}$	$\lambda^{DA} = 20$	$p^{DA} = 0$	0
DA reserve	$\tilde{\lambda}^{R,DA} \cdot r^{DA}$	$\tilde{\lambda}^{R,DA} = 65$	$r^{DA} = 25$	1,625
RT energy	$\lambda^{RT} \cdot (p^{RT} - p^{DA})$	$\lambda^{RT} = 300$	$p^{RT} - p^{DA} = 125$	37,500
Total				39,125

Table 6: Example of BSP settlement with scarcity pricing.

Settlement type	Formula	Price [$/MWh]	Quantity [MW]	Cash flow [$/h]
DA energy	$\lambda^{DA} \cdot p^{DA}$	$\lambda^{DA} = 20$	$p^{DA} = 0$	0
DA reserve	$\tilde{\lambda}^{R,DA} \cdot r^{DA}$	$\tilde{\lambda}^{R,DA} = 65$	$r^{DA} = 25$	1,625
RT energy	$\lambda^{RT} \cdot (p^{RT} - p^{DA})$	$\lambda^{RT} = 1,529.2$	$p^{RT} - p^{DA} = 125$	191,150
RT reserve	$\tilde{\lambda}^{R,RT} \cdot (r^{RT} - r^{DA})$	$\tilde{\lambda}^{R,RT} = 1,229.2$	$r^{RT} - r^{DA} =$ -25	-30,730
Total				162,045

for secondary reserve is $\lambda^{R,DA} = 65$ €/MWh. Suppose furthermore that the real-time price for secondary reserve price is $\tilde{\lambda}^{R,RT} = 1229.2$ €/MWh, and that the real-time energy price of energy is $\lambda^{R,RT} = 1539.2$ €/MWh. Finally, suppose that the maximum production capacity of the BSP is $P^{+} = 125$ MW.

The settlement of a BSP in a market without scarcity pricing is presented in table 5. This should be contrasted to table 6 which implements scarcity pricing as discussed in the present section. Note the increased remuneration that the BSP receives in the case of table 6 as a result of responding with upward balancing energy at a moment when the system is tight.

8.3 Multiple Products

As in the case of most international markets, also the European balancing market employs multiple reserve products, which can be distinguished on the basis of their full activation times. Concretely, in Europe there are two predominant types of frequency restoration reserves (which are the reserve categories for which we consider the application of scarcity pricing):

- Automatic frequency restoration reserve (aFRR): this reserve adjusts its setpoint every four seconds in response to an automatic controller which is driven by system frequency.

- Manual frequency restoration reserve (mFRR): this reserve obeys a full activation time of 15 minutes.

The theory developed by Hogan [31] is generalizable to the case of multiple types of reserves. Special attention in the development of this theory is attributed to the fact that these reserves are substitutable, in the sense that aFRR capacity is of "higher quality" than mFRR. Said otherwise, a unit that can provide aFRR is also fast enough to provide mFRR.

A disciplined approach towards pricing different reserve products requires the application of a co-optimization model, where reserve prices / adders and the energy price are derived as dual multipliers of this co-optimization model. In the absence of real-time co-optimization (as is the case in Europe, and was also the case in the original ERCOT design [22]), the next best option is to *approximate* the reserve prices of a co-optimization model from first principles [31]. What is interesting to note is that the resulting adder formulas depend on the interpretation of multi-dimensional ORDCs, since the system operator may approach the valuation of reserves in different ways. Consider the following concrete illustrations.

Approach 1: Separate ORDC for aFRR and mFRR. The fact that the EU balancing market operates two separate platforms for balancing energy in real time (MARI and PICASSO) may lead one to conclude that the valuation of mFRR and aFRR capacity are independent[1]. One would then define two functions, $V^{mFRR}(d^{mFRR})$ and $V^{aFRR}(d^{aFRR})$.

Approach 2: Separate ORDC for FRR and aFRR. Wertain EU system operators size reserves following EBGL using a probabilistic methodology which first defines an FRR target, and subsequently split this FRR capacity between mFRR and aFRR. The split has to obey certain rules related to the minimum capacity that the system should carry in aFRR capacity. Thus, we may instead interpret the system operator as employing two functions, $V^{FRR}(d^{aFRR} + d^{mFRR})$ and $V^{aFRR}(d^{aFRR})$.

The formulas that have been proposed for the Belgian market [58] are based on the ERCOT market design proposal [22] and can be expressed as

[1]MARI and PICASSO are markets for balancing energy, but we use their separate operation in order to motivate / rationalize a separate valuation of real-time balancing capacity.

follows:

$$\begin{aligned} V^{R,F} &= 0.5 \cdot (VOLL - \hat{MC}) \cdot LOLP_{7.5}(r^F) \\ V^{R,S} &= 0.5 \cdot (VOLL - \hat{MC}) \cdot LOLP_{15}(r^S) \end{aligned}$$

where $V^{R,F}$ and $V^{R,S}$ corresponds to the adders for fast and slow capacity respectively, r^F and r^S are the fast and slow reserve capacity available in real time, and $LOLP_{7.5}$ and $LOLP_{15}$ correspond to the loss of load probability with respect to 7.5-minute and 15-minute imbalances. The interaction of these adder formulas with BSP incentives is the focus of ongoing research.

8.4 Cross-Border Interactions

Since Belgium is considering the unilateral implementation of scarcity pricing, there have been questions related to how exactly foreign BSPs should be treated and what the interplay of the mechanism could be with neighboring zones. These issues are discussed in turn.

The unilateral implementation of scarcity pricing in a cross-border setting relies on the fact that the future EU balancing platforms (MARI and PICASSO) produce zonal dispatch signals that are consistent with the BSP activations that the BSPs within a zone receive. Concretely, a French BSP is *not* expected to receive a Belgian scarcity price if the Belgian system is tight [51]. The MARI or PICASSO platform are not matching BSP offers bilaterally to TSO demands. The network capacity is rather implicitly allocated, and the MARI or PICASSO price produced by the platform should be sufficient for inducing the balancing action that is requested of a non-Belgian BSP.

Since MARI and PICASSO will not be co-optimizing energy and reserves, it is important to point out that certain business rules that are familiar in an energy-only paradigm (but do not hold true in the more general setting of energy and reserve co-optimization) can no longer necessarily be satisfied. Concretely, one needs to choose between

1. enforcing the network equilibrium conditions of an energy-only design (i.e. no congestion between areas implies equal energy prices in the areas),

2. enforcing the balancing platform dispatch decisions, and

3. applying the valuation of reserves in real time.

One may be tempted to enforce choices 1 and 2 in the EU balancing platforms. However, in future power systems dominated by renewable resources where it will be important to remunerate flexibility adequately, it may be more crucial to enforce choices 2 and 3.

8.5 Scarcity Pricing and Capacity Markets

It is misleading to claim that one must choose between scarcity pricing and capacity markets [59]. Scarcity pricing co-exists with capacity markets in a number of US designs, including ISO-NE and PJM.

ERCOT does not implement a capacity market, and one may incorrectly conclude that the recent forced outages in Texas were a result of scarcity pricing. The Texas blackouts were driven by disruptions in the gas network and a number of other extreme weather-related factors. Thus, the problem was one of resilience, not adequacy. For example, even if Texas had more gas capacity available, a failing gas network that would not be able to provide fuel to these generators would render this additional capacity useless. If anything, scarcity pricing can help in providing incentives for market participants to invest in tecnologies that reduce system load under such conditions [11].

It is unclear whether Texan ratepayers would have been willing to insure against such extreme conditions by procuring the required redundant standby capacity in order to ride through such extreme weather events. Even if this is the case, it does not invalidate the valuable role of scarcity pricing in attracting flexible resources in systems with increasing amounts of renewable resources.

It is also important to note that, in conditions of forced outages, an energy-only design anyway results in price spikes, and the role of scarcity adders fades away. An important discussion about "circuit breakers" is emerging following the Texas post mortem [11].

To conclude: systems without capacity markets benefit from scarcity pricing, and systems with capacity markets benefit from scarcity pricing. In both cases, scarcity pricing adds value to a market design. The fact that scarcity pricing diminishes the missing money problems and therefore the scope of capacity markets should not be confused as implying that scarcity pricing cannot coexist with a capacity market.

References

[1] ADMIE. Methodology for the determination of zonal / system needs for balancing power. Technical report, 2020.

[2] David Arthur and Sergei Vassilvitskii. k-means++: The advantages of careful seeding. Technical report, Stanford, 2006.

[3] D. Bertsimas, E. Litvinov, X. A. Sun, J. Zhao, and T. Zheng. Adaptive robust optimization for the security constrained unit commitment problem. *IEEE Transactions on Power Systems*, 28(1):52–63, February 2013.

[4] Ricardo J Bessa, Joana Mendes, Vladimiro Miranda, Audun Botterud, Jianui Wang, and Zhi Zhou. Quantile-copula density forecast for wind power uncertainty modeling. In *2011 IEEE Trondheim PowerTech*, 2011.

[5] Ricardo J Bessa, V Miranda, A Botterud, Z Zhou, and J Wang. Time-adaptive quantile-copula for wind power probabilistic forecasting. *Renewable Energy*, 40(1):29–39, 2012.

[6] Christopher Breuer, Christian Engelhardt, and Albert Moser. Expectation-based reserve capacity dimensioning in power systems with an increasing intermittent feed-in. In *2013 10th International Conference on the European Energy Market (EEM)*, 2013.

[7] Kenneth Bruninx and Erik Delarue. A statistical description of the error on wind power forecasts for probabilistic reserve sizing. *IEEE transactions on sustainable energy*, 5(3):995–1002, 2014.

[8] Michael Bucksteeg, Lenja Niesen, and Christoph Weber. Impacts of dynamic probabilistic reserve sizing techniques on reserve requirements and system costs. *IEEE Transactions on Sustainable Energy*, 7(4), 2016.

[9] Pantelis Capros. Reform of the capacity remuneration mechanism in Greece. Technical report, E3MLab, 2014.

[10] Peter Cramton. Electricity market design. *Oxford Review of Economic Policy*, 33(4):589–612, 2017.

[11] Peter Cramton. Lessons for peru from the 2021 texas electricity crisis. Technical report, 2021.

[12] Kristof De Vos, Joris Morbee, Johan Driesen, and Ronnie Belmans. Impact of wind power on sizing and allocation of reserve requirements. *IET Renewable Power Generation*, 7(1):1–9, 2013.

[13] Kristof De-Vos, Nicolas Stevens, Olivier Devolder, Anthony Papavasiliou, Bob Hebb, and James Matthys-Donnadieu. Dynamic dimensioning approach for operating reserves: Proof of concept in Belgium. *Energy Policy*, 124:272–285, 2019.

[14] Daniel Devogelaer. Belgian blackouts calculation: A quantitative evaluation of power failures in Belgium. Technical Report WP3-14, Belgian Federal Planning Bureau, March 2014.

[15] Kristin Dietrich, J Latorre, Luis Olmos, Andres Ramos, and I Perez-Arriaga. Stochastic unit commitment considering uncertain wind production in an isolated system. In *4th Conference on Energy Economics and Technology*, 2009.

[16] Yury Dvorkin, Miguel A Ortega-Vazuqez, and Daniel S Kirschen. Wind generation as a reserve provider. *IET Generation, Transmission & Distribution*, 9(8):779–787, 2015.

[17] Yury Dvorkin, Hrvoje Pandžić, Miguel A Ortega-Vazquez, and Daniel S Kirschen. A hybrid stochastic/interval approach to transmission-constrained unit commitment. *IEEE Transactions on Power Systems*, 30(2):621–631, 2014.

[18] Elia. Evolution of ancillary services needs to balance the belgian control area towards 2018. *Brussels, Belgium, Tech. Rep*, 2013.

[19] ELIA. Study report on scarcity pricing in the context of the 2018 discretionary incentives, 2018.

[20] ELIA. Preliminary report on ELIA's findings regarding the design of a scarcity pricing mechanism for implementation in belgium. Technical report, Belgian transmission system operator, 2020.

[21] ENTSO-E. Continental europe operation handbook, load-frequency control and performance. Technical report, 2009.

[22] ERCOT. ERCOT market training: Purpose of ORDC, methodology for implementing ORDC, settlement impacts for ORDC, 2015.

[23] European Commission. Commission regulation (EU) 2017/2195 of 23 november 2017 establishing a guideline on electricity balancing. Technical report, 2017.

[24] European Commission. Commission opinion of 30/04/2020 pursuant to article 20(5) of regulation (EC) no 2019/943 on the implementation plan of belgium. Technical report, 2020.

[25] European Union. Regulation (EU) 2019/943 of 5 june 2019 of the European parliament and council on the internal market for electricity. Technical report, 2019.

[26] Federal Public Service Economy. Belgian electricity market: implementation plan. Technical report, 2019.

[27] A. Flamm and D. Scott. Electricity balancing significant code review - final policy decision. Technical report, Office of Gas and Electricity Markets, 2014.

[28] Patricio Rocha Garrido, Lisa Morelli, and Laura Walter. Price formation education 4: Shortage pricing and operating reserve demand curve. Technical report, PJM interconnection, 2018.

[29] Paul Giesbertz. The power market design column - the scarcity of scarcity pricing, 2019.

[30] HB Gooi, DP Mendes, KRW Bell, and DS Kirschen. Optimal scheduling of spinning reserve. *IEEE Transactions on Power Systems*, 14(4), 1999.

[31] W. Hogan. Electricity scarcity pricing through operating reserves. *Economics of Energy and Environmental Policy*, 2(2):65–86, 2013.

[32] William Hogan. On an 'energy only' electricity market design for resource adequacy. Technical report, Center for Business and Government, JFK School of Government, Harvard University, September 2005.

[33] William W. Hogan and Susan L. Pope. PJM reserve markets: Operating reserve demand curve enhancements. Technical report, Harvard University, 2019.

[34] Hannele Holttinen, Michael Milligan, Brendan Kirby, Tom Acker, Viktoria Neimane, and Tom Molinski. Using standard deviation as a measure of increased operational reserve requirement for wind power. *Wind Engineering*, 32(4):355–377, 2008.

[35] D Jost, M Speckmann, F Sandau, and R Schwinn. A new method for day-ahead sizing of control reserve in germany under a 100% renewable energy sources scenario. *Electric Power Systems Research*, 119:485–491, 2015.

[36] Dominik Jost, Axel Braun, and Rafael Fritz. Dynamic dimensioning of frequency restoration reserve capacity based on quantile regression. 2015.

[37] Dominik Jost, Axel Braun, Rafael Fritz, and Scott Otterson. Dynamic sizing of automatic and manual frequency restoration reserves for different product lengths. In *2016 13th International Conference on the European Energy Market (EEM)*, 2016.

[38] Jeremie Juban, Nils Siebert, and George N Kariniotakis. Probabilistic short-term wind power forecasting for the optimal management of wind generation. In *2007 IEEE Lausanne Power Tech*, pages 683–688, 2007.

[39] Jan Kays, Johannes Schwippe, and Christian Rehtanz. Dimensioning of reserve capacity by means of a multidimensional method considering uncertainties. In *IEEE PSCC Stockholm Conference*, 2011.

[40] Sun Kyo Kim, Joon-Hyung Park, and Yong Tae Yoon. Determination of secondary reserve requirement through interaction-dependent clearance between ex-ante and ex-post. *Journal of Electrical Engineering and Technology*, 9(1):71–79, 2014.

[41] Stefan Kippelt, Thorsten Schlüter, and C Rehtanz. Flexible dimensioning of control reserve for future energy scenarios. In *2013 IEEE Grenoble Conference*, 2013.

[42] Stuart Lloyd. Least squares quantization in pcm. *IEEE transactions on information theory*, 28(2):129–137, 1982.

[43] Christoph Maurer, Simon Krahl, and Holger Weber. Dimensioning of secondary and tertiary control reserve by probabilistic methods. *European Transactions on Electrical Power*, 19(4):544–552, 2009.

[44] Peter Meibom, Rüdiger Barth, Bernhard Hasche, Heike Brand, Christoph Weber, and Mark O'Malley. Stochastic optimization model to study the operational impacts of high wind penetrations in ireland. *IEEE Transactions on Power Systems*, 26(3):1367–1379, 2010.

[45] Nickie Menemenlis, Maurice Huneault, and Andre Robitaille. Computation of dynamic operating balancing reserve for wind power integration for the time-horizon 1–48 hours. *IEEE Transactions on Sustainable Energy*, 3(4):692–702, 2012.

[46] New York ISO. Ancillary services shortage pricing. Technical report, 2019.

[47] Henrik Aalborg Nielsen, Henrik Madsen, and Torben Skov Nielsen. Using quantile regression to extend an existing wind power forecasting system with probabilistic forecasts. *Wind Energy*, 9(1–2):95–108, 2006.

[48] Anja Ohsenbruegge, Thole Klingenberg, and Sebastian Lehnhoff. Dynamic data driven dimensioning of balancing power with k-nearest neighbors. In *Power and Energy Student Summit (PESS) 2015, January 13th-14th, Dortmund Germany*, 2015.

[49] Anja Ohsenbrugge and Sebastian Lehnhoff. Dynamic dimensioning of balancing power with flexible feature selection. In *23rd international conference on electricity distribution*, 2015.

[50] Miguel A Ortega-Vazquez and Daniel S Kirschen. Estimating the spinning reserve requirements in systems with significant wind power generation penetration. *IEEE Transactions on Power Systems*, 24(1):114–124, 2008.

[51] Anthony Papavasiliou. *Modeling Cross-Border Interactions of EU Balancing Markets: a Focus on Scarcity Pricing*. Elsevier, 2020.

[52] Anthony Papavasiliou. Scarcity pricing and the missing European market for real-time reserve capacity. *The Electricity Journal*, 2020.

[53] Anthony Papavasiliou and Gilles Bertrand. Market design options for scarcity pricing in European balancing markets. *IEEE Transactions on Power Systems*, 2021.

[54] Anthony Papavasiliou and Shmuel S. Oren. Multi-area stochastic unit commitment for high wind penetration in a transmission constrained network. *Operations Research*, 61(3):578–592, May / June 2013.

[55] Anthony Papavasiliou, Shmuel S Oren, and Richard P O'Neill. Reserve requirements for wind power integration: A scenario-based stochastic programming framework. *IEEE Transactions on Power Systems*, 26(4):2197–2206, 2011.

[56] Anthony Papavasiliou and Yves Smeers. Remuneration of flexibility using operating reserve demand curves: A case study of Belgium. *The Energy Journal*, pages 105–135, 2017.

[57] Anthony Papavasiliou, Yves Smeers, and Gilles Bertrand. An extended analysis on the remuneration of capacity under scarcity conditions. *Economics of Energy and Environmental Policy*, 7(2), 2018.

[58] Anthony Papavasiliou, Yves Smeers, and Gauthier de Maere d'Aertrycke. Study on the general design of a mechanism for the remuneration of reserves in scarcity situations. Technical report, UCLouvain, 2019.

[59] Anthony Papavasiliou, Yves Smeers, and Gauthier de Maere-d'Aertrycke. Market design considerations for scarcity pricing: A stochastic equilibrium framework. *The Energy Journal*, 42(5):195–220, 2021.

[60] Potomac Economics. 2017 state of the market report for the ERCOT electricity markets, 2018.

[61] Z. Ren, W. Yan, X. Zhao, W. Li, and J. Yu. Chronological probability model of photovoltaic generation. *IEEE Transactions on Power Systems*, 29(3):1077–1088, May 2014.

[62] Steven Stoft. *Power System Economics*. IEEE Press and Wiley Interscience, 2002.

[63] Aidan Tuohy, Peter Meibom, Eleanor Denny, and Mark O'Malley. Unit commitment for systems with high wind penetration. *IEEE Transactions on Power Systems*, 24(2):592–601, May 2009.

[64] J Zarnikau, S Zhu, Chi Keung Woo, and CH Tsai. Texas's operating reserve demand curve's generation investment incentive. *Energy Policy*, 137:111–143, 2020.

[65] Yao Zhang, Jianxue Wang, and Xifan Wang. Review on probabilistic forecasting of wind power generation. *Renewable and Sustainable Energy Reviews*, 32:255–270, 2014.

[66] Z Zhou, A Botterud, J Wang, Ricardo Jorge Bessa, H Keko, Jean Sumaili, and Vladimiro Miranda. Application of probabilistic wind power forecasting in electricity markets. *Wind Energy*, 16(3):321–338, 2013.

[67] Zhi Zhou and Audun Botterud. Dynamic scheduling of operating reserves in co-optimized electricity markets with wind power. *IEEE Transactions on Power Systems*, 29(1):160–171, January 2014.