DYNAMIC DIMENSIONING OF THE FRR NEEDS

31/10/2017

Table of Content

EXECUTIVE SUMMARY

Elia has the legal obligation to determine the FRR needs which are necessary to cover the potential system imbalances. These needs are currently calculated each year for the entire next year via a statistical convolution of the observed system imbalances, the prediction errors of incremental installed renewable generation and the forced outages of generation units. The **FRR needs** are calculated as input of the **FRR means**, i.e. R2, R3 flex, R3 standard, as well as cross-border reserve sharing.

It is expected that the FRR needs will increase in the upcoming years, following the growing share of renewable generation characterized by variable output and limited predictability. In addition, the integration of NEMO-link will affect the forced outage risk. In order to manage reliability and reserve costs in future years, this study investigates the potential of a new method for determining the FRR needs. A '**dynamic sizing approach'** is put forward which determines on day-ahead the FRR needs in function of the potential system imbalance in real-time, based on the perceived risk concerning forecast errors and outages. This could replace the current 'static sizing approach', for which the FRR needs are fixed on yearly basis, regardless of the short term system conditions (e.g. planned maintenances or renewable generation).

In order to assess the potential of such a dynamic approach, a study is conducted in two parts. The first part contains the results of an **Analysis** (including a **Cost and Benefit Analysis**) of six potential methods for the dynamic sizing of FRR needs. The objective of this first part is to give **Recommendations** towards a set of methodologies which need to be further analyzed in a Proof of Concept. The second part contains the final selection of the most promising method(s) for implementation, accompanied by an **Implementation plan.** Besides the Proof of Concept, this part also includes a financial implementation impact assessment.

The FRR needs presented in this study are a best-estimate of the dynamic FRR needs profiles in 2020 and 2027 as assumptions were to be taken to represent the future system conditions (e.g. installed renewable capacity). It cannot be excluded that these results are different to the dynamic FRR needs observed in the future. Furthermore, all results are obtained by means of 'prototypes' which are likely to perform better when optimized upon industrialization and operational experience. Different scenarios and sensitivities have been analysed in order to assure the robustness of the conclusions of this study.

Dynamic sizing refers to dimensioning the FRR needs before the Day-Ahead Market closure, i.e. at least 18 – 36 hours before real-time, in order to guarantee their availability. Hence reaping the full benefits of a dynamic dimensioning approach is conditional to the implementation of short-term (daily) procurement of at least part of the FRR means. This will be one of the deliverables of a study on dynamic mFRR procurement which is planned in 2018. Because an evolution to dynamic mFRR procurement will have implications for product design and procurement process, an involvement of the stakeholders is required. The implementation of dynamic FRR dimensioning depends therefore on the results of the assessment during the dynamic mFRR procurement study and in specific the outcome regarding the implementation of daily procurement.

Dynamics of imbalance drivers

Section 2 of the report identifies the required system conditions that are relevant for the prediction of the imbalance risk in day-ahead. Similar as today, the drivers for system imbalances are categorized as prediction or outage risks, while taking into account the N-1 criterion.

- **Forecast risks**

Scheduling market operations inherently relate to the forecast of variable generation and off-take. Even with intra-day markets and real-time balancing incentives, forecast errors contribute to system imbalances. However, the forecast error risk is tied to the prediction: it is for instance more likely to have a surplus than a large shortage in the system when predicting low wind conditions. A specific category covered in the forecast risk, i.e. the market risk, relates to the ability of market players to balance their portfolio in the absence of quarter-hourly products. Analysis shows that this appears to be an important driver for system imbalances. To capture the dynamic potential of the forecast risk, a statistical model is developed analyzing the correlation between historical time series representing the day-ahead predictions of system conditions, and the observed system imbalances.

- **Outage risks**

When a forced outage on a dispatched power plant occurs, its power output deviates from its schedule, resulting in system imbalances. This means that the probability of an outage is linked to its day-ahead schedule, i.e. a unit which is not scheduled cannot contribute to the outage risk. The scheduling will depend on economic conditions but also planned maintenances. In the future, also the scheduling of NEMO-link will impact the outage risk, even causing unexpected excess generation. As today, the N-1 criterion remains the minimum FRR requirement, which is determined by the largest nuclear unit for upward reserves. The outage risk is determined by means of a Monte Carlo simulation determining a probability density curve of the outages, based on day-ahead schedule of generating units (and NEMO-link).

Potential of dynamic sizing methods

In Section 3, six dynamic sizing methods which incorporate the above-mentioned information are examined. On the one hand, the proposed approaches evolve from intuitive methods (Extreme Cases, Manual Clustering) that relate to user decisions concerning the categorization of system conditions towards system imbalance risks, including an approach which is based on the schedule of NEMO-link and the maintenance planning of power plants (Outage-Only). On the other hand, the approaches also include advanced statistical methods based on machine learning techniques in order to relate predicted system conditions to the system imbalance risk, and the corresponding FRR needs (Continuous Neighboring, Quantitative Clustering, Neural Networks).

Analysis of a dynamic sizing method: approach and cases

In Section 4, the methods presented above have been implemented and tested. Simulations for 2020 are based on the historical data of 2016 including projections towards the renewable capacity towards 2020. The reliability target is always fixed at 99.9% (i.e. the FRR should cover 99.9% of the expected system imbalances), in line with current and generally accepted practice. All six methods are simulated and assessed in Section 5 towards reliability (expressed in percentage), FRR needs (expressed in capacity) and robustness towards future system evolutions. This determines their selection for the Proof of Concept in Section 6.

Overview of the results of the six methods concerning methodologic feasibility, reliability, FRR needs reductions, robustness towards future system conditions, as well as the decision to selection for the Proof of Concept

	Feasibility	Reliability	FRR needs	Robustness	Selection
Outage Only	Yes	=	+	+	**Yes**
Extreme Cases	Yes	No	N.A	N.A.	**No**
Manual Clustering	Yes	No	N.A.	N.A.	**No**
Quantitative Clustering	Yes	+	++	++	**Yes**
Continuous Neighbors	Yes	+	++	++	**Yes**
Neural Networks	No	N.A.	N.A.	N.A.	**No**

A. Reliability (technical assessment)

When assessing the intuitive methods, it is found that the forecast risk is too complex to be explained by means of a straightforward ‘manual’ categorization of the system conditions. The **Extreme Cases** and **Manual Clustering** do not achieve to grasp any benefits without seriously reducing the required minimum reliability. When assessing the machine learning methods, it is explained that an **Artificial Neural Network** method encountered methodologic problems when integrating the outage and prediction risk.

B. FRR needs (assessment of the dynamic potential)

The **Outage-Only** method, focusing only on the outage risk, already brings moderate benefits in terms of FRR needs, while maintaining a similar reliability as the static approach. When only treating NEMO-link in a dynamic way, results for the reference case 2020 show that the average upward FRR needs reductions amount up to 45 MW. In addition, the method facilitates large FRR needs reductions when facing simultaneous planned maintenance of large nuclear units.

The machine learning methods, **Quantitative Clustering and Continuous Neighbors,** do realize higher average FRR needs reductions, both on the up- and downward side. Results for the reference case 2020 demonstrate a substantial reduction in average FRR needs: on average, almost 100 MW is saved up- and downward. Furthermore, the machine learning tools provide a better reliability management by increasing the FRR needs during high risk conditions (typically undersized with static sizing) and lowering the FRR needs during low risk conditions (typically oversized with static sizing). This allows to de facto ensuring 99.9% reliability in all situations.

C. Robustness (sensitivity towards future system evolutions)

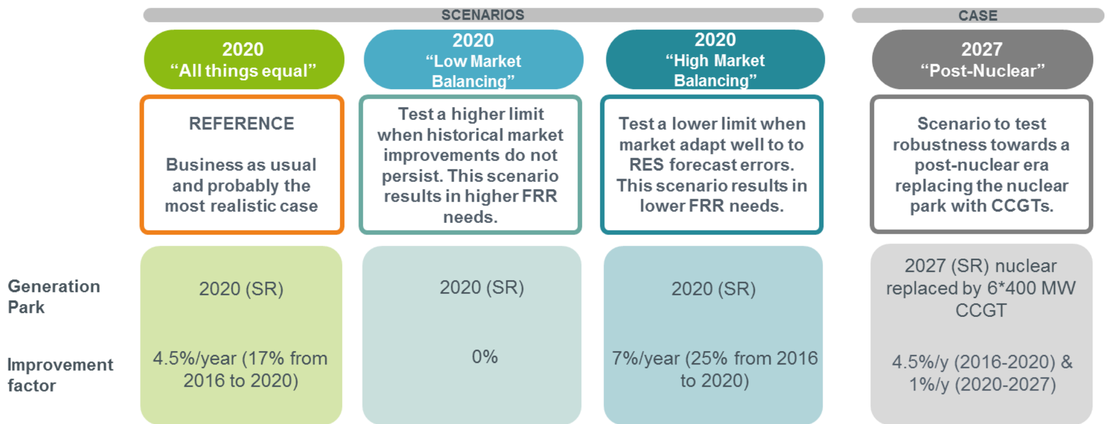
Simulations demonstrate that the increasing share of renewable energy toward 2020 increases the dynamic potential for the machine learning methods, while gradually reducing the dynamic potential of the Outage-Only method. It is also shown that a better performance of balancing markets reduces the magnitude of the FRR needs, and the dynamic potential, although the latter remains positive. This way, Elia's reactive balancing market design does not only reduce activated balancing energy, but also contribute to the reduction of the FRR needs.

The **post-nuclear case** shows a reducing effect on the average upward reserve needs due to the reduction of the outage risk: the dynamic potential increases due to the increasing weight of the forecast risk. Analyses also shows the compatibility of the three methods with an **offshore cut-off risk**, i.e. sudden disconnection of offshore wind turbines following a storm, if this would be decided to be treated as a forced outage incident in 2020.

D. Selection of methods

The costs and benefits are determined based on two criteria: the **technical criteria**, focusing on the improvement of reliability in high risk conditions and the **economic criteria**, expressed in average FRR needs reduction. The **economic criteria** show that three methods have a potential in reducing the average FRR needs while achieving an equal or even better reliability management than the current static methodology. In addition to meeting the minimum technical pre-conditions concerning reliability, the 3 remaining methodologies demonstrate sufficient high benefits, and it is therefore decided to further investigate all of them in detail in the Proof of Concept phase.

Proof of Concept



Representation of the 3 scenarios and 1 case study

The set-up of the Proof of Concept is dealt with in Section 6 explaining that the three selected methods are developed to "prototypes" which are tested in a realistic context for 2020, taking into account the limited predictability of the import-export direction of the NEMO-link, as well as design aspects of the method concerning training frequency (yearly for Outage-Only, monthly for Quantitative Clustering and daily for Continuous Neighbors), resolution (4-hourly) and lead time (Day-Ahead). A sensitivity analysis is conducted on these design aspects to assess their impact. The three methods are simulated in a "virtual" parallel run for 2020 and compared with a static sizing approach. The results are described in Section 7.

Reliability, FRR needs and robustness

Firstly, the Proof of Concept confirms previous findings that a dynamic sizing method results in **a better reliability management by means of ensuring higher FRR needs in higher risk periods**. This is particularly true for the machine learning methods. In contrast, the Outage-Only method maintains the predefined reliability levels, as in the static approach, but its functionality is limited to reducing FRR needs in lower risk periods following the known unavailability of power plants or predicted flow direction of the NEMO-link. Secondly, the results also confirm **average FRR needs saving in all three methods**.

Overview of dynamic potential (average FRR needs reductions compared to the static approach), dynamic spread (difference between the maximum and minimum FRR needs) for the tree methods

MW	Dynamic Potential		Dynamic Spread	
	Up	Down	Up	Down
Outage Only	31	15	54	112
Quantitative Clustering	53	46	346	794
Continuous Neighbors	64	46	408	995

It is important to realize that (1) the values obtained are a best estimate towards 2020, (2) that the results do not represent contracted FRR volumes, (3) improvements and operational experience during industrialization is expected to further increase the performance of the tool and (4) downward FRR needs falling below the 1000 MW limit when the NEMO-link is scheduled in import is an assumption for which the operational and regulatory framework is to be further analyzed.

Thirdly, and not the least, analysis revealed that static methods are extremely sensitive to extreme periods in the data set, which are expected to occur more frequently in future renewable systems. Indeed, a few extreme system imbalances in one unrepresentative week may increase the FRR needs for the entire year. It can be concluded that such **static methods will be become unsuitable in the middle-long term from a technical and economic point of view**. Dynamic methods do not face this disadvantage as such extreme outliers in the data will only result in a high sizing of FRR needs when recognizing similar system conditions.

Transparency

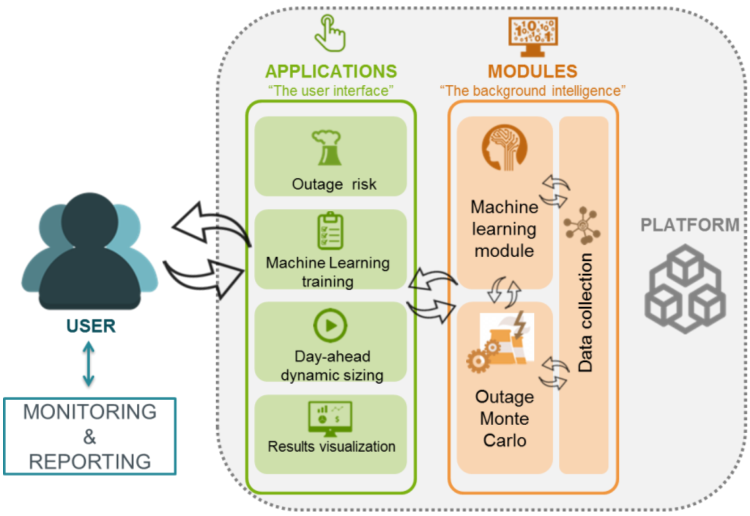
A detailed analysis of the three methods, and in particular the machine learning methods demonstrate that the method is not a black box: a detailed analysis of the correlations allows to understand the relation between system conditions and FRR needs, while a dashboard visualizing the FRR needs and system conditions allow operators to understand and interpret the result as illustrated by the example below.

Cost and Benefit Analysis

After a positive assessment of the business case for the three methods in the different scenarios and a case study for 2027 after the nuclear phase out, including an analysis of the financial gains in the reference scenario, it is concluded to keep the three models when moving towards an industrialization of a dynamic sizing method. Indeed, the Outage-Only method is estimated to bring a yearly financial gain of 1.48 and 1.71 M€, while the machine learning methods are estimated at a yearly gain 2.51 and 2.97 M€. An analysis of the implementation cost of the method show that these are largely exceeded by these financial gains. It has to be kept in mind that these estimations are based on a 'generic' price for mFRR in 2020, subject to uncertainty.

It is concluded that the Outage-Only provide interesting potential for a stepwise implementation, and even later as fallback method, while still providing positive FRR needs gains until 2020. However, this method will become, similar to static sizing unsuitable when renewable capacities increase. In contrast, machine learning methods provide most potential, expected to increase with higher renewable capacity and after a nuclear phase-out. Nevertheless, they will require a longer period of testing before they can be implemented.

Implementation Plan



Schematic representation of the architecture of a dynamic software tool

When putting such a dynamic sizing method in place, the "prototypes" used in the Proof of Concept will be re-developed towards robust calculation modules which will implemented in one integrated software platform. The architecture will be based on calculation modules, a data collection module, while the user will interact with the software platform by means of different applications. It has to be clear that dynamic sizing will not be a one shot implementation, but will require regulator updating of the design of the method, and the algorithms itself.

A first release of the dynamic sizing tool can be used in the market between 9 to 12 months after approval and resources at Elia have been cleared. It is to be stressed that this will contain a first version for the Outage-Only method, while the machine learning methods require further testing in the parallel run and can be used in the market between 19 and 21 months after start of the project.

Including the project development and yearly recurrent cost a dynamic sizing, the implementation cost is estimated at €850,000 to €1,100,000 per year. This includes the first estimations concerning the development of the tools for translating the FRR needs in to FRR volumes to be contracted, as well as the expected operation costs of these tools. These will be further analyzed in a follow-up project on daily procurement.

Recommendations and next steps

This study demonstrates the techno-economic potential for a dynamic sizing approach in Belgium, as well in terms of a better reliability management, as a robust alternative for static sizing methods in a system based more and more on dynamic system conditions. Furthermore, a financial analysis of the potential FRR gains and the implementation costs demonstrate a positive business case as the implementation costs (0.85 – 1.10 M€ per year) are clearly exceeded by the financial gains of the FRR needs (1.48 - 1.71 M€ per year for Outage Only, 2.51 and 2.97 M€ per year for Machine Learning).

It is therefore recommended to further prepare implementation of a dynamic sizing method. As explained before **the effective application of dynamic dimensioning is subject to the implementation of short-term (daily) procurement for at least the mFRR means**. This will be one of the deliverables of a study on dynamic mFRR procurement which is planned to be finished in the course of 2018 in close collaboration with relevant stakeholders.

1 INTRODUCTION

1.1. Operating reserves

Every high-voltage Transmission System Operator (TSO) in ENTSO-e is responsible for maintaining the balance in its control area. However, the market players (BRPs) are, on an individual basis, responsible for maintaining the balance within their portfolio. Elia has established a balancing mechanism in order to on the one hand (1) incentivize the market players as much as possible to maintain a balanced portfolio and help to restore the system imbalance, and on the other hand (2) manage the remaining imbalances in the system by means of contracted and non-contracted power reserves supplied by Balancing Service Providers (BSPs).

These operating reserves are activated according a certain hierarchy (Figure 1). Primary reserves are activated automatically and on a continuous basis, and revised up- or downwards as required to stabilize the frequency of the European grid. Secondary reserves are activated automatically and on a continuous basis and revised up- or downwards to handle sudden imbalances in the control area. Tertiary reserves can be activated manually at Elia's request. It can be used to address a major imbalance in the zone managed by Elia and deal with congestion problems. There are several types of tertiary reserve: contracted reserves (R2, R3 standard and R3 flex) as well as non-contracted reserves. It is generally activated after the primary or secondary reserve and will remain active until the problem is solved.

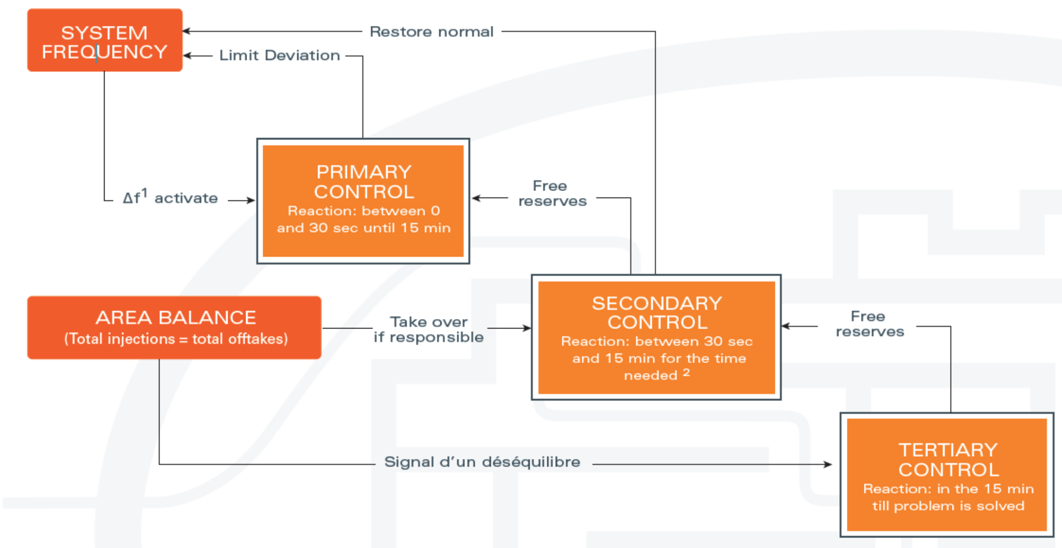


Figure 1: Schematic overview of the activation of operating reserves

The current European regulatory framework, the System Operation Guidelines (SO GL), defines FCR (Frequency Containment Reserve), aFRR (automatic Frequency Restoration Reserve) and mFRR (manual Frequency Restoration Reserve) for respectively primary reserve, secondary reserve and tertiary reserve. It sets the minimum requirements concerning the determination of reserve capacity within the Member States.

1.2. Determining the need for operating reserves

System imbalances result from unexpected demand variations, power plant outages, and unexpected variations of renewable generation. As the primary reserve is deterministically determined on the level of the synchronous zone, the scope of this study is limited to the FRR needs, i.e. the total needs for secondary and tertiary reserves.

FRR is used to restore the balance in the control area. It is sized according to the expected system imbalances. The system imbalance represents the residual imbalances of the market players' following the (1) variability of renewable energy, i.e. wind and photovoltaics, (2) variability of the demand, and (3) unexpected outages of power plants, or relevant transmission assets, and (4) other (new) elements which may impact the current and future system imbalance.

When specifically focusing on sizing power reserve requirements, **the current statistical method is generally accepted as a good method, allowing taking into account variable generation, while trading off complexity and accuracy.** This method is currently implemented by Elia and is compliant with current and new ENTSO-e Guidelines. Furthermore, it is well described in literature and in use by several TSOs (e.g. Germany). Furthermore, it is well understood and accepted by Belgian stakeholders and the regulator. Alternative approaches such as deterministic or system simulations are respectively less suited to cope with increasing decentral generation, or not compatible with the European electricity market design.

With this methodology, statistical indicators, such as the probability density curve of time series of the different causes of system imbalances are convoluted and used as measure for reserve requirements. This is done by sizing the reserve capacity to a volume which achieves a predefined reliability level, i.e. 99.9%. A schematic overview of the method is shown in Figure 2, and further explained in Box 1.

The European System Operation Guidelines instruct such statistical approach for dimensioning the up- and downward FRR, based on historical values of the system imbalance. It is to be combined with a deterministic approach which sets the minimum FRR capacity at the reference incident level (N-1), i.e. the maximum positive or negative power deviation in a control area, e.g. following a power plant outage. The methodology used by Elia is already compliant with these requirements (although no downward volume is currently procured yet).

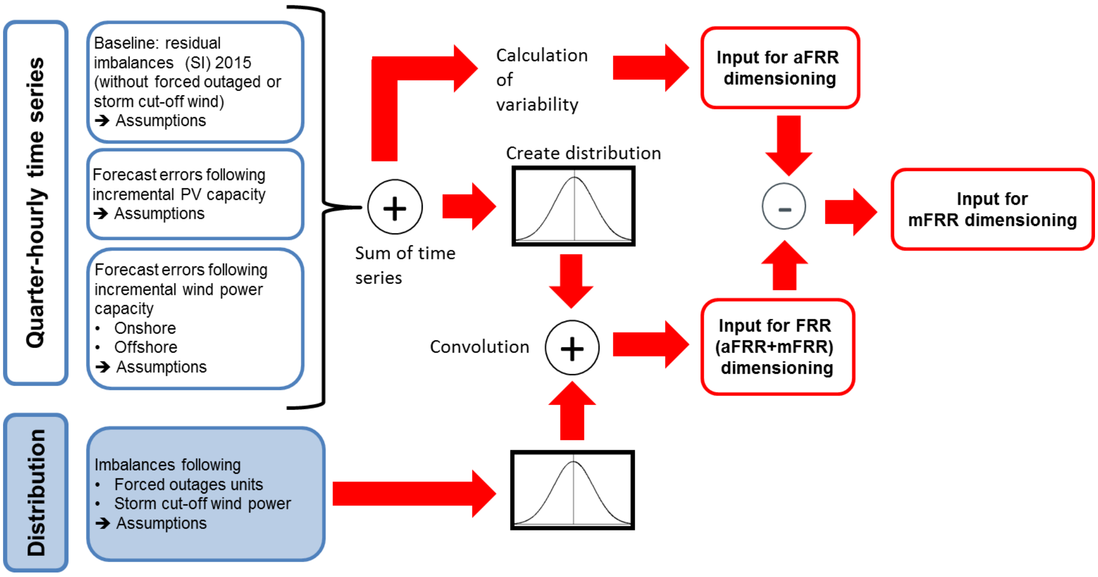


Figure 2: Schematic overview of the design of the FRR volume dimensioning for 2017

Box 1: Elia's approach of the FRR volume sizing since 2015

Sizing total FRR need (aFRR + mFRR) is based on a statistical convolution of the probability distribution curve of the forced power plant outages and of the probability distribution curve of the expected system imbalance. The 99.9% percentile of the probability distribution curve of the expected upward system imbalances determines the upward FRR needs (with N-1 set as lower limit).

The **expected system imbalance** for year Y is derived from the historical time series (with a resolution of 15 min) of the system imbalance for one entire year (Y-2) from which all the forced outages are excluded. In addition, the expected forecast errors of the incremental renewable generation capacity (photovoltaics, onshore and offshore wind) are added, based on assumptions concerning to which extent the market can cover part of these forecast errors. The **forced power plant outages** are derived from a Monte Carlo simulation including the outage probability and characteristics of the expected generation fleet, and the outage probability of the offshore wind parks caused by storm events.

In contrast, aFRR need is sized based on the probability distribution curve of the absolute variations over each time step of 15 min of the expected system imbalance described above. The 79% percentile of this probability distribution curve percentile determines the symmetric aFRR needs. Then, mFRR need is sized based on the difference between the total FRR and aFRR. In a final step, the needs for aFRR and mFRR are translated to product type volumes (R2, R3flex, R3standard), taking into account product availability and availability of cross-border sharing agreements with other TSOs.

More information is provided in the yearly decision of the CREG on the approval of the proposal by Elia on the assessment method for the determination of the primary, secondary and tertiary reserve capacity www.creg.be.

1.3. Increasing RES and static dimensioning

The variability of renewable energy sources (RES), such as wind power and solar photovoltaics, challenges reserve dimensioning approaches through prediction errors. These are an inherent characteristic of such technologies and the impact on the reserve needs is expected to increase with their installed capacity.

Because the current dimensioning methodology is based on a yearly calculation of the FRR needs, it is sensitive to extreme conditions which may set a high value for the entire year, e.g. the forecast risk of the offshore wind park. As this risk is only valid in certain circumstances, the FRR needs might be overestimated for a large part of the year, resulting in inefficiencies.

Although the increasing reserve need following RES will be correctly captured by the current static sizing method, this capacity will be required to be available all year through, even during conditions with lower imbalance risks. It is therefore important to find the appropriate sizing for each moment, avoiding oversizing (and obviously undersizing as well) as much as possible to ensure cost efficiency.

Table 1 shows the historical global FRR needs between 2013 and 2018 as (or is expected to be) approved by the NRA. Although a decreasing trend is observed between 2013 and 2017, this is mainly due to market design improvements, which is not expected to continue due to the further increase of renewable generation towards 2019 (total installed capacity of wind power and solar photovoltaics of 8075 MW) and 2020 (8785 MW). This is already confirmed by the FRR needs of 2018.

Table 1: Evolution of the FRR needs and variable Renewable Energy Sources Installed (vRES)

[MW]	Upward FRR Needs*	Variable RES**
2013	1260	4260
2014	1241	4948
2015	1240	5279
2016	1203	5609
2017	1183	6166
2018	1190	6824

**Volume Assessments (Final Decisions CREG); **Adequacy Study for Belgium (2016)[1]*

[1] Elia, Adequacy Study for Belgium: the need for strategic reserve for winter 2017-2018, November 2016, www.elia.be

The expectation towards an increasing trend is also confirmed in the adequacy study and assessment of the need for flexibility in the Belgian electricity system published by Elia in 2016[2], and put forward a total FRR needs of 1240 MW and 1000 MW for respectively up- and downward FRR towards 2021 (Table 2). These non-binding indicative volumes[3] are based on the 2016 applicable volume determination methodology, excluding any additional measures and volumes that would be required dealing with exceptional situation (e.g. outage of offshore parks due to storm events). While the FRR needs are expected to increase following the integration of renewable energy (including offshore wind power), and NEMO-link, the assumptions are deemed to be still rather progressive concerning the ability of the market to balance the prediction errors from incremental renewable capacity.

Table 2: Estimated evolution of the FRR needs in the Base Case, as published in the adequacy study and assessment of the need for flexibility of 2016

Horizon	FRR+	FRR-	aFRR	mFRR+	mFRR-
2021	1240	1000	175	1065	825
2023	1240	1000	175	1065	825
2027	1240	1000	175	1065	825

1.4. Towards dynamic FRR dimensioning

While static reserve methodologies, e.g. yearly dimensioning, are challenged by the increasing share of variable renewable generation, statistical and mathematical software tools facilitate alternative approaches which can size reserve needs in a more dynamic way. Indeed, advanced statistics allow learning to predict the real-time imbalance risk based on the observed day-ahead forecasted system conditions, and then to adapt the reserve needs accordingly on a regular basis, e.g. each day in function of the estimated risk for the next day.

Consequently, ‘static’ reserve methodologies, which dimension a fixed amount of reserves over a longer period, i.e. typically one year, can be replaced by such a method determining the imbalance risk and reserve needs from day-to-day, for each hour of the next day in its most extreme application. This allows reducing reserve capacity in periods with lower risks, and respectively increasing in periods of higher risks. Such dynamic sizing methodologies are described in scientific literature.

A dynamic dimensioning can be defined as **a dimensioning methodology which sizes a variable reserve capacity over time, e.g. per hour, per block of hours or per day, in function of the expected system conditions, and in particular the**

[2] Elia, Adequacy study and assessment of the need for flexibility in the Belgian electricity system, April 2016, www.elia.be

[3] These volumes have as sole purpose to give an idea of the future trend with respect to volume needs and do by no means substitute for the legally or regulatory determined volume assessment process in place between Elia and the CREG.

associated risk for system imbalances. This in contrast with a static method sizing fixed FRR needs based on a static generation mix, not taking into account if units are scheduled to be dispatched or not. Such elements which allow dynamic dimensioning include for instance:

- Forecast errors of variable renewable generation, which are correlated with day-ahead predictions.
- Forced outages which are correlated with the expected operation schedule of these assets.

These allow to determine the reserve needs in function of the risk, for instance: (1) low renewable generation schedules such as offshore wind power are likely to result in a reduced risk for a shortage following a prediction error; (2) power plants which are not scheduled to inject, due to maintenance or low demand, do not contribute to the risk of an unexpected shortage following an outage; or (3) an HVDC interconnector which is scheduled to import, does not contribute to the risk of an unexpected excess following an outage.

In practice, a major constraint of dynamic dimensioning of reserves is that in the European market design **reserve capacity is generally procured before the day-ahead market outcome**. Obviously, this is to ensure the availability of the required capacity, and the liquidity of the balancing market. This availability would not be guaranteed if procured after day-ahead, e.g. during scarcity, which would leave the TSO with reduced balancing means. This constraint implies the need for estimating the real-time system conditions at least 18 - 36 hours ahead, impacting the potential of dynamic sizing.

The result of this constraint is that an important condition is the short-term (e.g. daily) procurement of reserves, at least for a part of the capacity. This includes certain aspects such as:

- **Tender design:** The full benefits of dynamic reserves will be reaped when facilitating the same lead time of procurement as the dimensioning. Daily sizing without daily procurement does not make much sense (although the inverse might be useful). Furthermore, a dimensioning too far ahead of DAM clearing may facilitate the procurement process while it reduces the ability to accurately predict system conditions. Nevertheless, some parameters such as power plant maintenance or installed capacity can be initialized or updated week-ahead or month-ahead.
- **Product design**: The allocation towards different FRR product types might have important implications, as well as the product duration. The resolution of the reserve sizing has to be aligned with product design (a higher resolution than the product length will not be useful), but it can be expected that lower resolution products (longer activation duration) reduce the potential. A trade-off comparing the variability of the FRR needs and the transaction costs of higher resolution products can probably be found.

As a daily procurement will have strong impact on Elia and the relevant market parties, it has to be thoroughly analyzed and discussed with the stakeholders.

1.5. Objective, scope and structure of the study

A dynamic dimensioning of FRR is expected to reduce on average the reserve needs, and therefore the associated procurement cost, in periods with lower risks for imbalances. Nevertheless, the cost and benefits of such a strategy depend largely on the procurement strategy of reserves.

Nevertheless, the scope of the project is strictly limited to the **FRR needs** (Figure 3) which are defined as the total capacity required covering the expected system imbalances based on uncertainty, i.e. forecast risks and outage risks, within a control zone. These rules are described in the system operation guidelines and are to be translated in the LFC Block Agreement. This determines among other things the obligations of Elia (being the sole TSO in the LFC block) concerning the dimensioning of FRR.

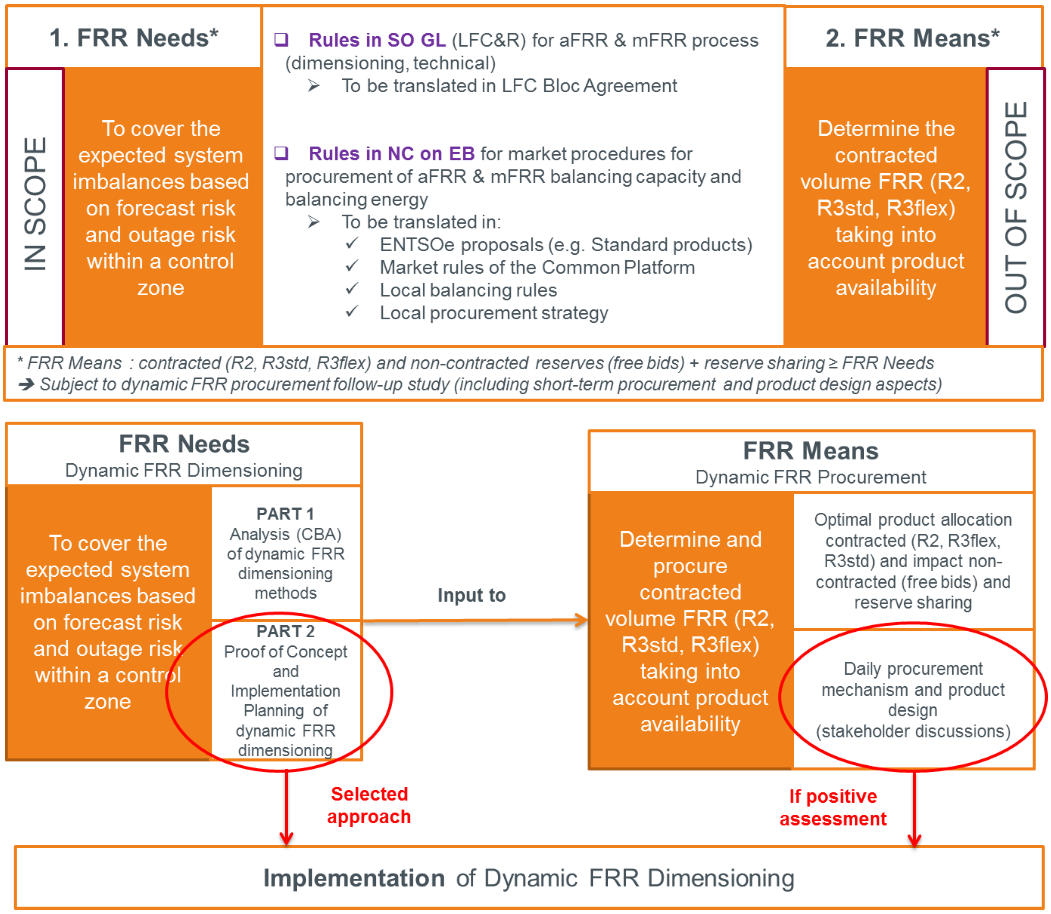


Figure 3: Scope of the study

This is in contrast to **FRR means**, which is defined as the contracted volume FRR (R2, R3 standard, R3 flex), as well as cross-border reserve sharing, and non-contracted reserves (free bids). The determination of the FRR means takes into account product characteristics, such as its availability. These rules are described in the ENTSO-e Network Guidelines on Electricity Balancing which needs to be implemented in proposals on standards products, market rules of procurement platforms and local balancing rules.

As the scope of this study is limited to the total FRR needs, it focusses on the dynamic potential derived from the imbalance risk. Indeed, this crucial step is to be completed before the next step can be taken, i.e. the allocation towards contracted mFRR volumes, while also taking into account non-reserved capacity (free bids) and reserve sharing. This should be studied after the FRR needs analysis is completed, i.e. in a study on dynamic mFRR procurement which also analyses the implications of daily procurement on product design and procurement process.

The first part of the study aims **to conceptualize a methodology for dynamic sizing of the FRR needs in Belgium** which is (1) implementable from a practical point of view (in terms of transparency and complexity); (2) is compliant with European regulation (ENTSO-e Network Guidelines) and can be accepted by the NRA; and (3) is compliant with future power system evolutions (including installed RES capacity and HVDC interconnections), and (4) meeting minimum technical criteria maintaining the pre-defined reliability.

Section 2 of this report investigates the dynamic potential of the system conditions which drive the needs for reserve capacity. Thereafter, Section 3 proposes a selection of dynamic sizing methodologies which will be further investigated by means of a detailed analysis in Section 4. Section 5 describes the Cost and Benefit Analysis (CBA), the selection of the successful methods and practical implementing issues of these methods.

The second part of the study aims to **test the selected methods in a Proof of Concept** representing a realistic setting for 2020, i.e. a 'virtual' parallel run to the static sizing based on a year of simulations. The results allow assessing the business case and presenting a corresponding **implementation plan**. Section 6 describes the set-up of such Proof of Concept, while Section 7 discusses the results. Section 8 presents an implementation impact assessment. It is reminded that this will only concern the dynamic FRR dimensioning, and that the effective implementation will depend on the result of the dynamic FRR procurement study regarding daily procurement of mFRR.

The mathematical modelling and calculations in this study have been conducted together with N-SIDE, which is an external consultant specialized in advanced analytics applied to the energy sector.

PART 1: POTENTIAL OF DYNAMIC SIZING APPROACHES

2 Potential of dynamic dimensioning

2.1. Imbalance drivers

Figure 4 represents an overview of the different **imbalance drivers** which are investigated. These are categorized in forecast risks and outage risks. **Forecast risks** are determined by unpredicted power variations of variable renewable generation (e.g. wind power and photovoltaics) and demand. Their output variations are an inherent characteristic, together with their forecast errors. A specific imbalance driver which can be identified in this category is the market risk. As market products are based on an hourly resolution, while the system imbalance is, by convention, monitored on quarter-hourly basis, this discrepancy may result in system imbalances.

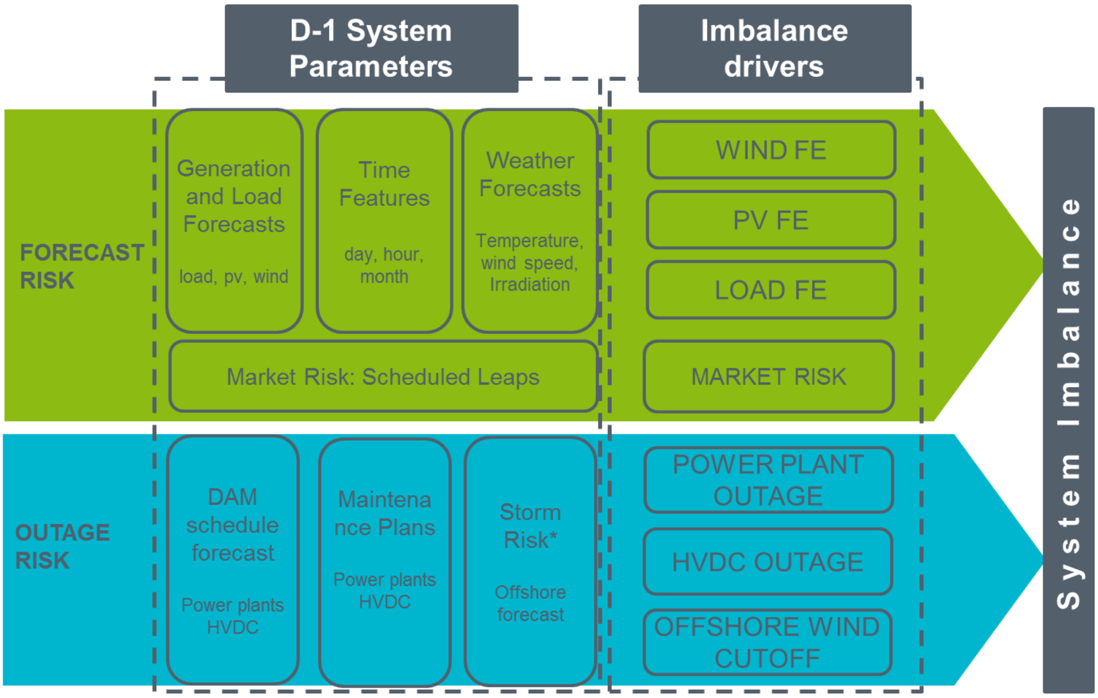


Figure 4: System parameters, imbalance drivers and system imbalance

In contrast, the **outage risk** is driven by the unexpected outages of power plants and the future HVDC interconnectors (NEMO-link). These are characterized by their limited frequency and unpredictable nature, as well as their impact (output falls to zero). A specific imbalance driver is the offshore wind cut off. This refers to a sudden disconnection of offshore wind turbines following a storm. Today, it is treated as a forced outage, but it is uncertain how this risk will be managed in the future. The impact on the system imbalance, the predictability of the event and the possibilities to manage this risk are being investigated by Elia in a separate study.

These imbalance drivers, as well as the system imbalance itself are seen as variables which can be estimated by using system parameters, available in day-ahead, such as predicted weather conditions or expected generation schedules.

2.2. Forecast risk

Market parties face forecast errors of variable renewable generation and demand, which can cause imbalances in their portfolio in real-time. Although part of these may be covered by modifications of positions close-to-real-time (e.g. intra-day), a part of these contribute to the system imbalance. The imbalance drivers studied are the onshore, offshore, photovoltaic and demand forecast error, as well as the market risk.

Although these imbalance drivers contribute to the theoretical explanation of the imbalance, they are not known in day-ahead, so they cannot be directly used in our dynamic sizing methodology. However, behind these imbalance drivers are system parameters that are known in day-ahead (such as wind production forecast, hour of the day, weather forecasts) which impact the forecast error risk.

2.2.1. Statistical characteristics of the forecast risk

A statistical analysis is conducted to analyze the correlation of the system imbalance with the imbalance drivers, and with the system parameters which are known in day-ahead, before the market outcome. The analysis is conducted on time series of quarter-hourly observations and predictions for 2015. The system imbalance data itself is published on the website of Elia. The time series is processed to remove the periods with forced outages, as the forced outage risks are treated in a separate analysis. The system parameters, available in day-ahead, which have shown to impact the system imbalance and ranked in terms of importance according to a statistical indicator called Mutual Information (Box 2).

The load DA forecast is the Day-Ahead Total Load Forecast as published on the website of Elia. The load DA gradient is calculated as the delta between two sequential quarter-hours. The scheduled leaps represents the market risk and is calculated as the difference between the hourly average load DA forecast and the corresponding quarterly-hourly values, and relate to the discrepancy between hourly market products and quarter-hourly imbalance settlement.

The temperature, wind speed and solar irradiation is obtained from internal databases based on predictions of a weather station in Zaventem. The time series represent the day-ahead forecast provided with an hourly resolution. The wind onshore and offshore DA forecast and solar PV DA forecast are the aggregated day-ahead wind power and solar PV power forecasts published on the website of Elia. The wind DA gradient and solar PV DA gradient are calculated as the delta between two sequential quarter-hours.

The results of the mutual information analysis shows how the scheduled leaps appear to be an important factor in the system imbalance, as well as the temperature, wind forecasts and demand forecasts (Figure 5). The solar forecasts, as well as time aspects, seem to have less impact on the system imbalance.

Box 2. Methodology of Mutual Information (MI)

In order to assess the relevance of each system parameter which is used to explain the system imbalance, statistical analyses have been conducted in three steps:

1. ***Mutual information***

The mutual information between each system parameter and the system imbalance has been computed. Mutual information quantifies how much uncertainty of one variable is reduced by knowing another variable. This is expressed mathematically as

$$MI(x,y) = H(y) - H(y|x);$$

where H(•) is the entropy function (i.e. "a measure of uncertainty"). According to the above formula, the mutual information measures how much uncertainty of *y* (e.g. the system imbalance) is reduced when *x* (e.g. the photovoltaic day-ahead forecast) is known. This mutual information, such as the correlation, provides a value that quantifies the link between two variables. But, unlike the correlation, the mutual information can quantify "non-linear" relationships.

2. ***Conditional analysis***

The mutual information provides a first insight of what are the relevant parameters to be considered. However, this is a number to be interpreted cautiously, especially for "cyclic" features such as *hour of the day* or *season*. In order to further analyze the links between system parameters and imbalance, some additional statistical tests have been conducted, to validate or invalidate what has been concluded with the mutual information tests. More specifically, conditional probability distributions of the system imbalance with respect to the system parameters have been analyzed.

3. ***Redundancy check***

The aim of this third step is to identify potential redundancies that could exist between system parameters that have been identified as relevant. Indeed, some of the system parameters may be linked, e.g. wind speed and wind production forecast both provide information useful to explain the system imbalance. However, wind speed and wind production are also linked to each other. So the question arises if the wind speed provides additional information on top of the wind production.

A redundancy check has been conducted in order to find the possible redundancy. This check includes multi-variate mutual information as well as conditional distribution tests.

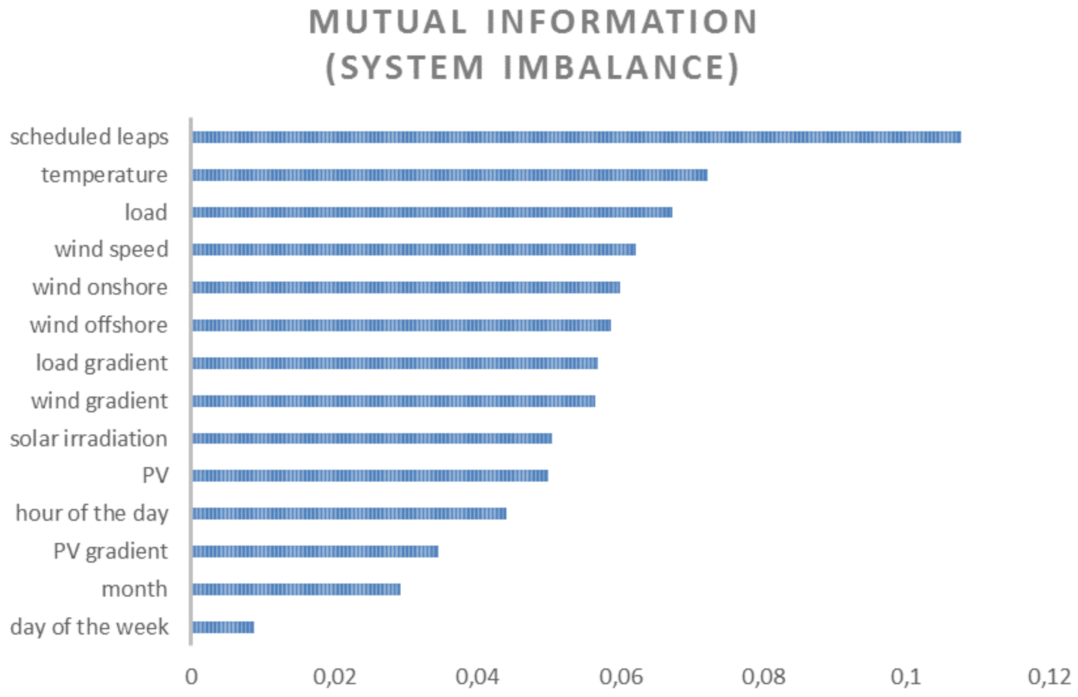


Figure 5. Ranking of the system parameters in order of impact on the system imbalance according to mutual information method (Box 2), Computed using MI(feature, SI), for N=60

An additional analysis has been conducted to realize how much of the system imbalance can be explained, *a-posteriori*, by the imbalance drivers (prediction errors of renewable generation and demand) listed in Figure 4. The objective is to use the system imbalance drivers as an intermediate step towards explaining the system imbalance.

A **linear least square method is applied which estimates** the unknown parameters in a linear regression model, with the objective of minimizing the sum of the squares of the differences between the observed responses (values of the variable being predicted, in this case the system imbalance) in the given dataset and those predicted by a linear function of a set of explanatory variables (in this case the imbalance drivers, i.e. the forecast errors). R-squared is an element between [0;1] representing the proportion of observed response variations which can be explained by the explanatory variables in the model: the higher R-squared, the better the predictability of the model.

Table 3 shows the R-squared value representing how much of the system imbalance can be explained by the imbalance drivers linear model. Three cases are considered: firstly, looking at all the system imbalance values without any discrimination; secondly considering only the high imbalances; thirdly, considering low imbalances.

Table 3: R-Squared value of the linear least square method investigating the relation between the imbalance drivers and system imbalances in 2015. (high R-squared represents a better predictability)

	All System Imbalance	SI < -300 MW ; SI > 300 MW	-70 MW < SI < 60 MW
R-squared	28.2%	58.5%	3.8%

Note that the analysis has shown that a major part of the system imbalance remains unexplained, and that the analysis does not facilitate a model to explain the imbalance drivers, which can be aggregated to the system imbalance. Indeed, on average, only 28.2% of the imbalance can be explained by the identified imbalance drivers listed above. Note that in case of high imbalance, this number improves to 58.5%, while in low imbalance case almost nothing is explained by the drivers. This means that high imbalances are generally caused by forecast errors of renewables and demand while low imbalances are caused by structural issues which are not fully identified yet. This may indicate that on day-ahead level, the system imbalance risk is not entirely predictable, or that the system parameters are not entirely captured yet, and will require further analysis to better understand and predict the system imbalance risk.

2.2.2. Assessment of the dynamic potential

In order to conduct a first assessment of the dynamic potential of the forecast risk, the probability distribution curve of the system imbalance of 2015 is studied, after removal of the forced outages. In order to estimate the future potential, a similar analysis is conducted for the expected system imbalance in 2019. This takes into account the prediction errors of the incremental renewable capacity as explained in Box 3. Table 4 shows the observed and projected renewable capacity installed between 2011 and 2020. It can be seen that the coming years 2017-2020, the renewable capacity is projected to increase more rapidly than in the past years, mainly due to offshore wind power developments.

Box 3: Approach for estimating the future system imbalance

The expected system imbalance is derived from the historical time series, i.e. 2015 (with a resolution of 15 min) of the system imbalance for one year from which all the forced outages are excluded.

In addition, the prediction errors of the incremental renewable generation capacity (photovoltaics, onshore and offshore wind) are added. These are calculated by means of the forecast errors of each individual source, expressed in percentage of the installed capacity. These percentages are then multiplied with the incremental RES capacity over the studied time horizon, i.e. between 2015 and 2019, 2466 MW.

It is noted that the larger the horizon, the more the forecast errors may impact the result, potentially overestimating the expected system imbalance. For this reason, correction factors are applied in the current ‘static sizing’ methodology. Note that in a ‘dynamic sizing’ methodology, these correction factors will become obsolete as the model will inherently take into account the latest system information available.

Table 4: Installed renewable capacities (Adequacy study for Belgium, 2016) expressed in MW

Production capacity [MW] at the end of 20xx										
	11	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**
ONSHORE	891	1005	1014	1360	1.528	1.696	1.857	2.047	2.236	2.431
OFFSHORE	195	380	566	713	713	713	862	1.142	1.996	2.310
PV	1901	2501	2680	2.875	3.038	3.200	3.447	3.635	3.843	4.044
TOTAL	**2987**	**3886**	**4260**	**4948**	**5279**	**5609**	**6166**	**6824**	**8075**	**8785**

**The values for 2020 are based on Elia's estimations*

A. Predicted renewable output

To assess the dynamic potential of the forecast risk, the probability distribution curve of the expected system imbalance for all periods with a low prediction error risk (typically hours with a low predicted renewable output) is studied and compared to all periods with a high forecast risk (typically hours with a high predicted renewable output. Table 5 depicts the capacity required to cover 99.9% of this forecast risk. **It has to be stressed that these values correspond to the observed system imbalance of 2015, without taking into account forced outages, and cannot be compared with the results of analyses conducted in the yearly sizing of the reserves[4].** Without the convolution with the forced outage probability distribution, the presented values here will be much lower as the FRR needs.

In 2015, it is shown that this capacity corresponds to 599 MW (shortage) and 651 MW (surplus) for all periods. However, when categorizing this into scenario's with high and low variable RES predictions, respectively referring to periods of high and low forecast risk, this impact the system imbalance risk[5]. Indeed, in low risks, these figures go down to 453 MW and 445 MW, while increasing up to 848 MW and 944 MW in high risk conditions. These observations are shown in the probability distribution curve presented in Figure 6. The curve for 2015 demonstrates how the tails of the curve show a higher system imbalance in the high risk periods compared to the low periods, or when taking all the periods.

This effect is even higher when studying the results for 2019. In this year, the up- and downward capacity required to cover the expected system imbalance amounts up to 1132 MW and 882 MW, respectively (Table 5). The asymmetry of 2019 results is explained by the asymmetric behavior of the forecast errors of offshore wind in 2015. Indeed, as the 2019 expected system imbalance is calculated by means of extrapolating these based on the elevated incremental capacity for offshore wind between 2015 and 2019, this effect has a large impact on the results. Although this effect might be corrected by the market in future years, this does not impact the main conclusions of this analysis.

---

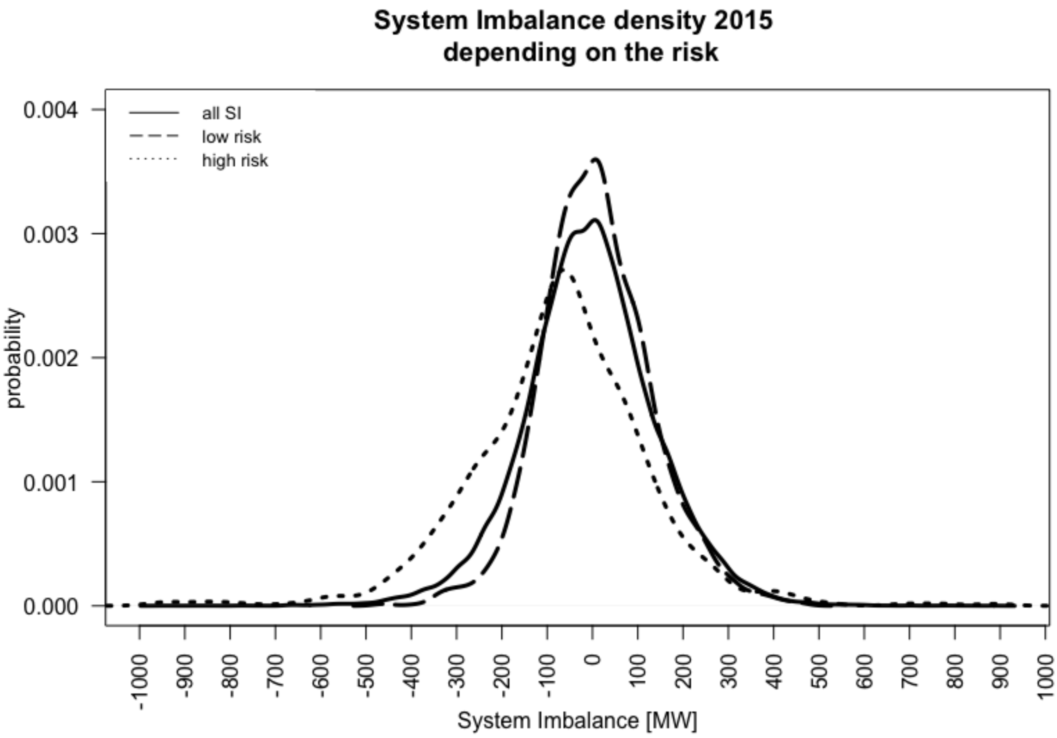
[4] The FRR needs for 2015 are calculated based on assumptions made in 2014, while using system imbalance observations and renewable prediction errors of 2013, for which the probability distribution curve is convoluted with the distribution of the forced outages.

[5] It is explained in Section 4 that extreme conditions such as high wind or low wind do not always straightforwardly correspond to certain risk conditions.

This analysis confirms that a dynamic approach may allow to reduce the amount of FRR needs in low risk conditions, while increasing the amount of reserve needs during high risk conditions. This will also increase the reliability in these high risk periods. In addition, as low risk conditions (low renewable generation) occur more frequently than high risk conditions (high renewable generation), this may result in a reduced average FRR needs. As the analysis for 2019 demonstrates, this potential is expected to increase with growing shares of renewable capacity.

Table 5: Up- and downward capacity to cover 99.9% of the forecast risks following classifications based on the predicted renewable generation.

	RES Forecast	**Upward [MW]**	**Downward [MW]**
2015			
'All'	"All"	599	651
'Low Risk'	Wind < 6% of installed capacity ~ 110 MW PV < 5% of installed capacity ~150 MW	453	445
'High Risk'	Wind > 40% of installed capacity ~ 900 MW PV > 27% of installed capacity ~ 800 MW	848	944
2019			
'All'	"All"	1132	882
'Low Risk'	Wind < 6.5% of installed capacity ~ 200 MW PV < 6% of installed capacity ~200 MW	467	565
'High Risk'	Wind > 66% of installed capacity ~ 2000 MW	1443	854



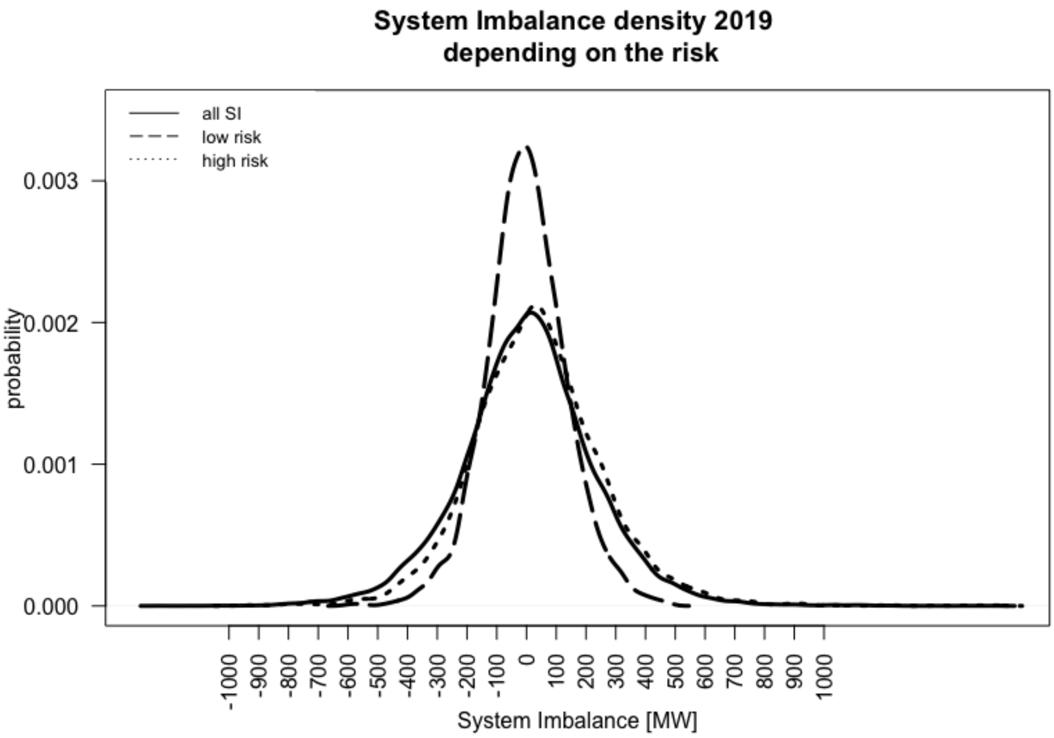


Figure 6: Probability Distribution Curves of the (expected) system imbalance in 2015 (up) and 2019 (down) for different forecast risks. SI>0 represents an upward capacity (positive system imbalance, i.e. shortage)

B. Scheduled Leaps

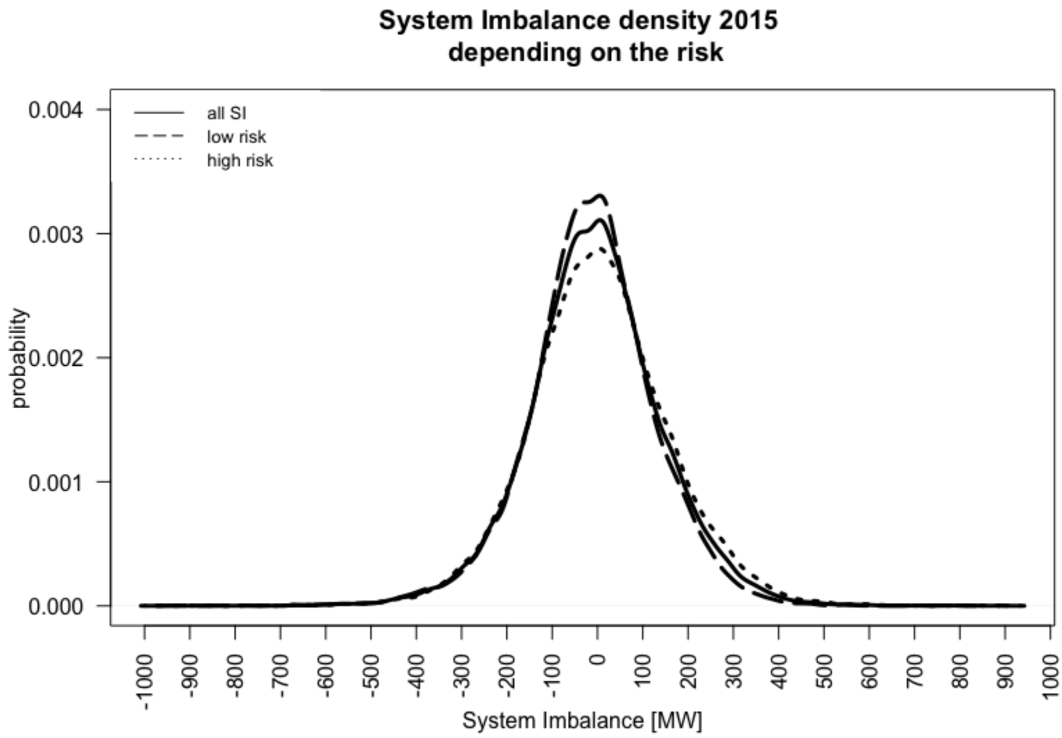
The market risk is studied by means of one parameter, i.e. 'the market schedule deviations' or 'scheduled leaps (SL)'. This is defined as the 15' deviations from the 60' average expected demand. It should represent the potential difficulty of market parties to balance their portfolio in absence of 15' market products.

To assess the dynamic potential of this imbalance driver, the probability distribution curve of the expected system imbalance for all periods with a high market risk (typically periods with a high expected SL) is studied and compared to all periods with a low market risk (typically periods with a low expected SL).

Table 6: Up- and downward capacity to cover 99.9% of the forecast risks following classifications based on specific situations concerning the market risk (absolute value 'Abs' Scheduled Leaps 'SL')

[MW]	Scheduled Leaps SL	Upward [MW]	Downward [MW]
2015			
'All'	"All"	599	651
'Low Risk'	Abs(SL) ≤ 50 MW	548	616
'High Risk'	Abs(SL) > 50 MW	613	659
2019			
'All'	"All"	1132	882
'Low Risk'	Abs(SL) ≤ 50 MW	1125	829
'High Risk'	Abs(SL) > 50 MW	1151	931

Table 6 depicts the capacity required to cover 99.9% of the expected system imbalances (without outages). The results indicate an impact as well in 2015 and 2019. In general, it is shown that periods with low leaps result in reduction of the required up- and downward capacity to cover the expected system imbalances to 548 MW, 616 MW, respectively, and increased up to 613 MW and 659 MW. This effect becomes more asymmetric in 2019 where the potential is higher for the downward capacity required. This effect is also confirmed by the tails of the probability distribution curves.



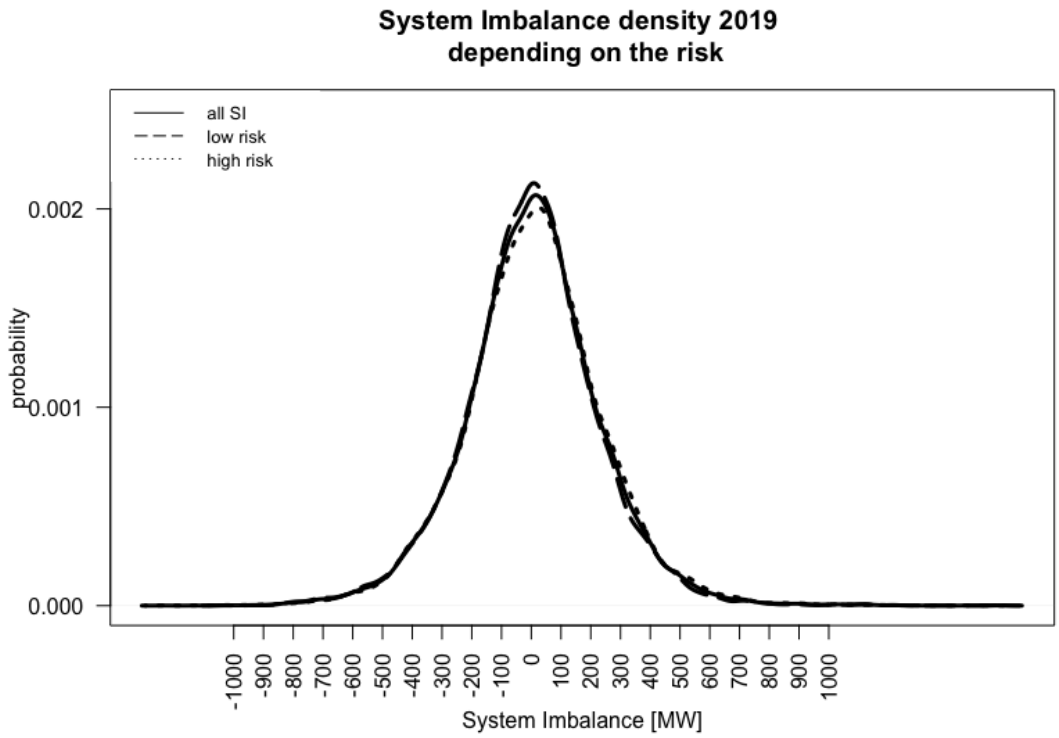


Figure 7: Probability Distribution Curves of the (expected) system imbalance in 2015 (up) and 2019 (down) for different market risks. On these graphs, SI>0 represents an upward capacity (positive system imbalance, shortage)

2.3. Forced outage risk

Market parties cope with unexpected events such as equipment failures, causing instantaneous imbalances in their portfolio. Although they might try to limit their impact on their portfolio by means of re-scheduling their positions, such events are likely to impact the system imbalance.

The current methodology is applied to analyse the outage risk, i.e. assessing the probability distribution curve after conducting a Monte Carlo analysis with the generation fleet for 2015, as well as the expected generation fleet for 2019 (including NEMO-link). The latter is based on information published in the Input Data of the Adequacy Study for the Strategic Reserves (2016), published by Elia. This analysis takes into account, as today, all the generation units larger as 100 MW, together with their outage risk and outage duration. Table 7 depicts the capacity required to cover 99.9% of this outage risk.

Table 7. Up- and downward capacity to cover 99.9% of the outage risk analysis following classifications based on specific situations concerning maintenance and NEMO-link schedule (as from 2019)

[MW]	**Situation**	**Upward**	**Downward**
2015			
No Outages		1040	N.A.
1/2/3/4 Nuclear Out		1040/1010/1010/490	
2019			
'All'		1100	1000
Nemo Import	No Outages	1070	0
Nemo Import	1/2/3/4 Nuclear Out	1040/1010/1010/1000	0
NEMO Export	No Outages	1040	1000
NEMO Export	1/2/3/4 Nuclear Out	1040/1040/1010/640	1000

It is obvious that this risk depends on the scheduled generation portfolio. Although this is unknown before the day-ahead market outcome, some reliable estimation can be made following maintenance plans or dispatch predictions. A first analysis **shows that the impact of power plants schedules is limited**, which is explained by the limited impact of one power plant in the total outage risk while having 4 larger nuclear base load units of 1 GW.

- In the context of 2015, Table 7 shows that a reduction of the required capacity to cover the outage risk from 1040 MW to 490 MW can be achieved when no nuclear unit of 1 GW is scheduled. This is similar for 2019 in which reduction from 1100 MW to 640 MW can be realized only when NEMO is scheduled in export. Although the occurrence of such an event is extremely rare, the simultaneous outage of multiple nuclear units impacts the outage distribution curve.

- In a future context without nuclear, the outage risk is driven by the gas-fired power plants, but again the scheduling of one or more units will have limited impact on the outage risk.

On the other hand, dynamic potential might be provided following the scheduling of NEMO-link. This new asset is integrated in the outage risk as it is expected to impact the system imbalance, similar to a power plant, when facing an outage. However, due to the fact that the risk of losing energy or facing surplus energy depends on the schedule, the impact of NEMO schedule is investigated which show an effect, particularly for the downward capacity.

Table 7 shows that in 2019, the up- and downward capacity required to cover the outage risk is 1100 MW and 1000 MW, respectively. When NEMO is in export, the required downward capacity becomes 0 MW, and when NEMO is in import, the required upward capacity decreases slightly to 1040 MW. It is to be noted that the offshore wind power cut-off following storms is not included in this analysis.

The potential of dynamic dimensioning based on scheduling of assets is constrained following the ability to predict this before the day-ahead market schedule is known. This requires reliable forecast tools based on day-ahead generation unit commitment simulations or probabilistic approaches based on historical observations.

2.4. Wrap up: combining the forecast risk and outage risk

The analysis of the imbalance drivers shows how a certain potential might be attained when treating the forecast error risk (including the market risk) dynamically, as well as the outage risk. One remark is that predicting the system imbalance risk on day-ahead is complex, and will require further investigation to increase its performance. The analysis also shows enough potential for dynamic sizing based on the forecast risk. On the other hand, this also means that improving the ability to understand the system and how the system imbalances are explained, is likely to further increase the potential of dynamic sizing methods.

Nevertheless, the simplistic approach above is only a pre-assessment to identify the dynamic potential of certain imbalance drivers and this potential shall only be fully exploited when succeeding to combine the prediction and outage risk in one model. The challenge is to access to the dynamic potential of both imbalance driver categories while each one depends on a different methodology. Furthermore:

- The forecast risk is highly dependent on installed RES evolutions: the higher the installed RES capacity, the more the total risk will be impacted by the forecast risk. Nevertheless, in system conditions with low forecast risks, the outage risk may still impact the result. Consequently, the dynamic potential will be lower in small control zones with large conventional units.

- The downward outage risk is highly dependent on the scheduling of NEMO-link. This effect will decrease with increasing RES as the forecast risk grows in importance.

- The N-1, set by nuclear units (upward) or NEMO (downward) will impact the final result, being a floor for the FRR needs. The impact of the largest unit is however expected to become less with increasing renewable generation. In addition, the N-1 can even be treated dynamically although that this is mainly relevant for the downward N-1 following the scheduling of NEMO, after the nuclear phase-out this will also be the case for upward capacity.

3 DYNAMIC SIZING METHODOLOGIES

3.1. Principles of the methodology

When deploying a dynamic sizing method, three steps are considered. The **determination of the general methodology** is the first step. This is conducted long before (e.g. one year) before the actual sizing, and can be in line with the current reserve sizing process where the methodology is determined one year ahead and approved. It concerns determining the main principles of the sizing approach as well as the process of the next steps. Note that in principle, a first simulation on historic data allows to already determine or even procure a certain part of the reserve on long-term (e.g. yearly).

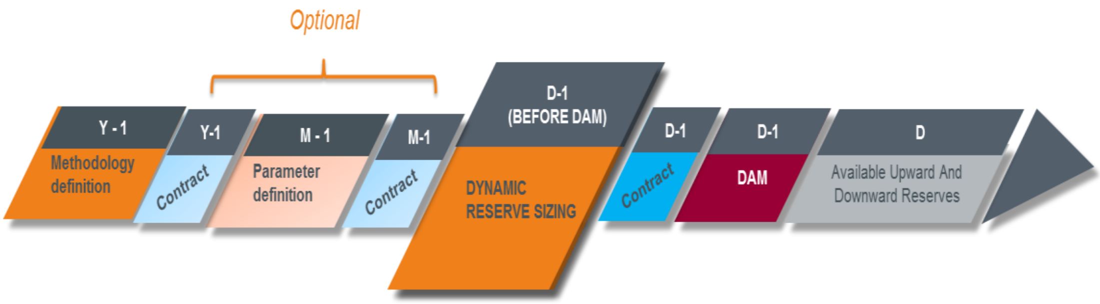


Figure 8: General Dynamic Sizing Process

A second, but optional, step is to **determine the parameters of the dynamic sizing methodology, also referred to as training of the model**. This can be done closer to the actual dimensioning in order to capture the latest information concerning system conditions or installed generation capacity. Some methodologies are based on pre-defined scenarios, which might be adapted over time (e.g. on a seasonal or monthly basis). This phase can be seen as optional as the modification can occur on a yearly basis, and some methodologies do not need to conduct this intermediate step before directly sizing the FRR needs. Such inter-temporal calculations allow an optional procurement phase (e.g. yearly, monthly, weekly).

Finally, the **actual sizing of the reserves** is conducted close to real-time, but at least before day-ahead market closure. Based on the methodology and the scenarios, it determines the up-and downward FRR needs for the 24 hours of the next day (or even further ahead). The resolution is determined in the methodology and can in theory be hourly, or even quarter-hourly (depending on the resolution of the input data). In practice, it might depend on the product length. This capacity, taking into account the procured capacity on the longer term, is contracted day-ahead.

Depending on the resolution of the method, this results in a daily FRR needs profile for the next day, illustrated in Figure 9.

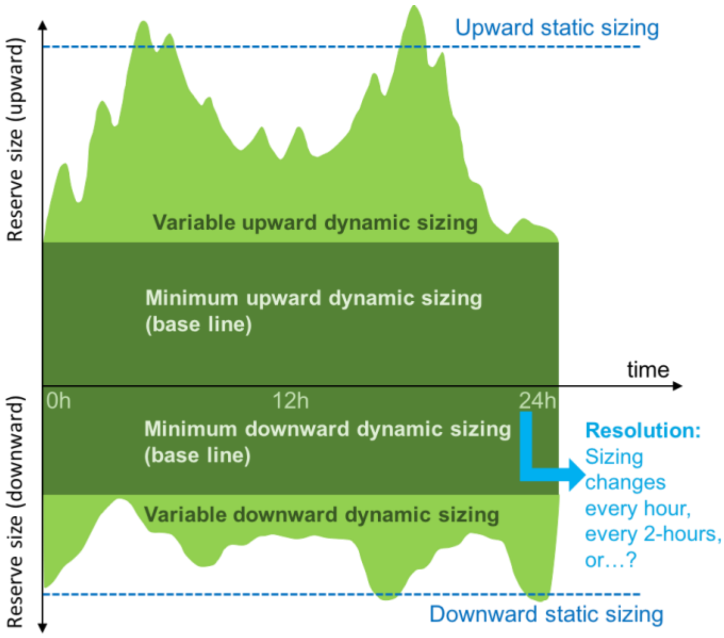


Figure 9: Illustration of a possible daily profile of Dynamic FRR needs

3.2. Imbalance drivers

The methodology is built on the imbalance driver categories, i.e. the forecast risk and outage risk, described in Section 2. This corresponds to the current static sizing methodology, complemented by the market risk, which is integrated in the forecast risk.

The **forecast risk** is modelled by means of historically observed system conditions and imbalances as described in Section 2. However, it is explained that future insights or new system evolutions may require adding or removing parameters from the model in future years. In fact, updating the statistical model is an activity which has to be conducted every few years. The statistical model is able to link day-ahead predicted system conditions to a certain forecast risk, represented by a probability density curve of the total prediction error, or other statistical indicator.

Outages are rare events, compared to the forecast error risk, which happen on a continuous basis. Similar as today, the **outage risk** is considered by means of a Monte Carlo simulation creating a specific probability distribution curve of the total capacity in outage. On day-ahead, an estimation can be made of the scheduled generation fleet (economic dispatch, security constraints and planned maintenance), allowing to create a specific day-ahead probability density curve.

This probability distribution curve representing the outage risk is convoluted with the outcome of the statistical model of the forecast risk. While the complexity of the outage risk model is the ability to predict the schedule of units before market closure, the complexity of forecast risk model is the application of predictive statistics to train a model and make accurate estimations on the imbalance risk based on historical observations.

3.3. Methods

The aim of this chapter is to use statistical learning methods to understand the relation between system conditions and the imbalance risk, and use this information to predict the imbalance risk in day-ahead. The presented methods will be based on ‘machine learning’ or ‘artificial intelligence’ methods trained on historical observations of pairs of input objects (predicted system conditions) and output values (system imbalance).

Different types of machine learning methods exist and supervised learning methods are investigated for this study following the nature of the problem where data concerning the output and input variables are both available. The two main methods in supervised learning are classification (clustering) and regression. The first is based on grouping input data which is similar in terms of the corresponding output value, and is often used when working with qualitative data. The latter is based on linear and non-linear regressions to estimate the output value, and is generally used when working with quantitative data. These models described in the literature a complemented with some heuristic methods based on ‘human intuition’ and are less ‘smart’ as the presented machine learning methods.

Six potential methodologies are identified which are potentially eligible for dynamic dimensioning of reserve capacity, and which are put forward for the analysis. The methods are ranked in following an increasing complexity in what concerns statistical methodology (Figure 10). And can be categorized according to type of statistical models in three groups:

1. “Minimal changes” methods applying rudimentary statistics or human intuition to predict dynamically the imbalance risk in some specific situations/days (semi-dynamic methods);
2. “Discrete” methods applying discrete statistical models (i.e. relying on scenarios) based on human decision or machine learning (artificial intelligence) to predict dynamically the imbalance risk in every situation (fully dynamic methods);
3. “Continuous” methods applying continuous statistical based on machine learning (artificial intelligence) to predict dynamically the imbalance risk in every situation (fully dynamic methods).

A distinction is made between the models which are trained by means of an automatic learning process, often referred to as ‘machine learning’ and other methods requiring a ‘human intuition’ when determining parameters manually during the learning phase.

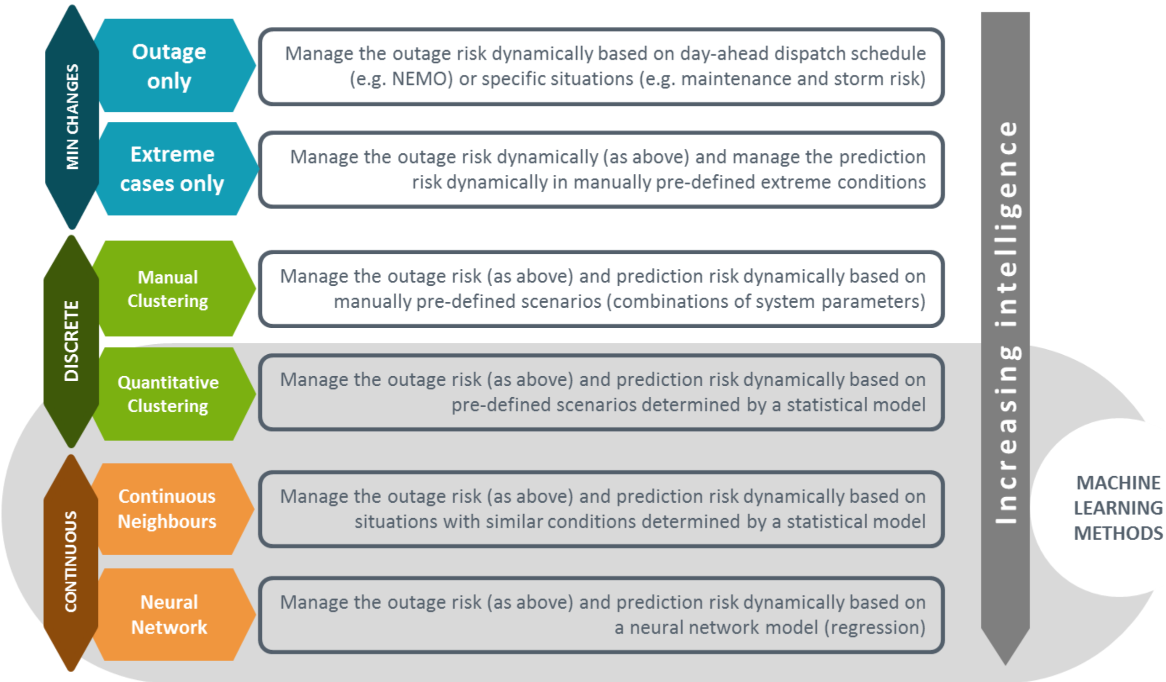


Figure 10: General overview of the 6 presented methods for dynamic sizing

3.3.1. Minimal changes

These methods applying rudimentary statistics or human intuition to predict dynamically the imbalance risk in some specific situations. They are based on the current 'static' approach, sizing fixed FRR needs based on time series of the expected system imbalance (Figure 11).

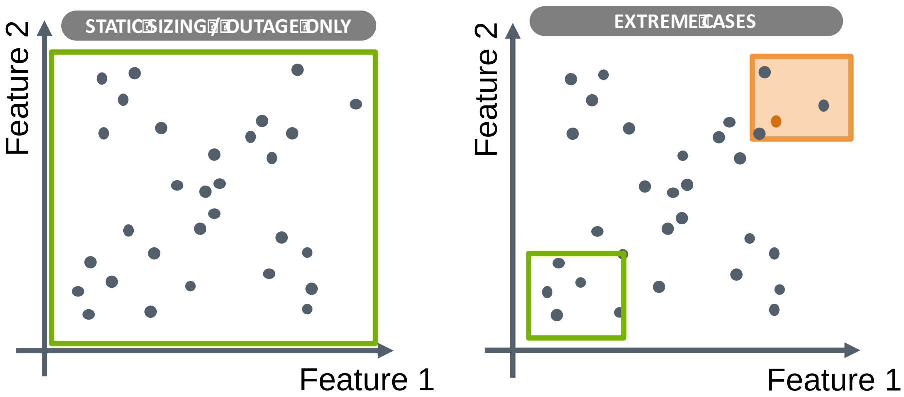


Figure 11: Visual representation of the outage-only (left) and extreme cases method (right)

A. Outage-Only

The 'Outage-Only' method is conceived in a framework where power plants or relevant transmission assets dispatch can be forecasted, allowing a 'dynamic' treatment by means of only integrating the relevant power plants in the forced outage distribution curve. Although a previous analysis in Section 2.3 has shown that the impact of power plant schedules is limited because the impact of one or few power plants on the total outage risk is low, some specific cases such as the scheduling of NEMO-link, the simultaneous outage of large nuclear units and the offshore cut-off risk hold an interesting potential.

The expected system conditions on day-ahead concerning the scheduling of generation or transmission assets (NEMO-link scheduled as import, export, or in between, or the identification of an offshore cut-off risk if relevant) result in a particular forced outage probability density curve which is convoluted with the forecast error risk. The latter is built on the yearly expected imbalance, based on historic data of system imbalances and forecast errors just as the current "static" sizing method.

B. Extreme Cases

The "Extreme Cases" method treats, on top of the outage risk as described above, also the forecast risk dynamically, but only in extreme conditions. This is based on a statistical analysis of the historic data identifying 'extreme' conditions (Box 4).

In order to "determine the parameters of the methodology", during the calibration of the method (e.g. year-ahead, month-ahead), the user manually defined "extreme cases". For instance, in the Figure 11, these are the green and the orange boxes. These are ranges of system parameters values for which the situation is considered as "extreme". For these conditions, a specific probability density curve is applied based on historical observations during these conditions.

In the "actual sizing of the reserves" phase, there are two possibilities: (1) or the forecasted system conditions for the next day falls in one of the extreme cases boxes. In this case, the "specific" sizing is applied; (2) the forecasted system conditions for the next day does not fall in one of the boxes and the usual static sizing is applied (no discrimination depending on the system parameters).

3.3.2. Discrete predictive statistics

The next two approaches are based on 'clustering' where a categorization is made by the user of all different system conditions (Figure 12).

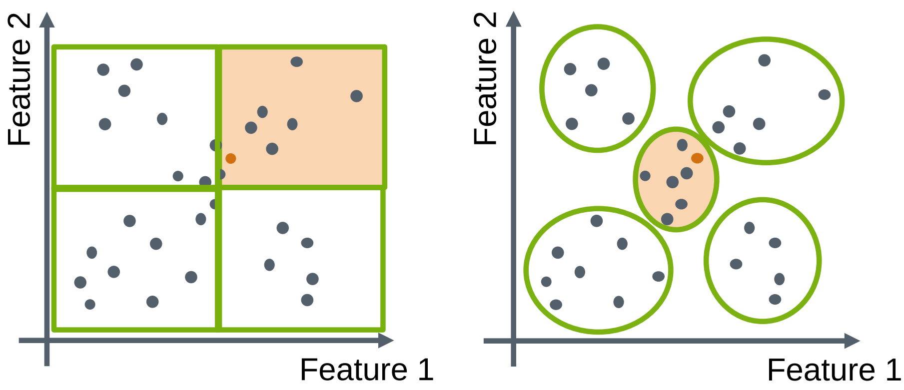


Figure 12: Visual representation of manual clustering (left) and quantitative clustering method (right).

Box 4: Extreme cases in practice

The Figure 11 provided a schematic view of what the method is doing. In practice, the design choices retained were:

- Three features were used to classify the historical imbalances:
  - PV day-ahead (DA) production forecast,
  - Wind (offshore + onshore) DA production forecast
  - Total load DA forecast
- Five "extreme cases" were defined by the user (note that other combinations have been tested but are not retained following lower performance)

Categorization is based on the percentile of the yearly observed production or consumption, for instance low wind in the first row considers the 0% to 25% highest observations for wind production.

1	Low wind < 25% 2015=185 MW 2019=340 MW	No PV	Low load <25% 2015=8850 MW 2019=8900 MW
2	Low wind <30% 2015=235 MW 2019=440 MW	Low PV <65% 2015=250 MW 2019=300 MW	Low load <40% 2015=9530 MW 2019=9580 MW
3	High wind >65% 2015=710 MW 2019=1370 MW	High PV >75% 2015=525 MW 2019=665 MW	Low load <40% 2015=9530 MW 2019=9580 MW
4	High wind >70% 2015=810 MW 2019=1560 MW	High PV >80% 2015=720 MW 2019=910 MW	High load >70% 2015=10810 MW 2019=10870 MW
5	High Wind >70% 2015=810 MW 2019=1560 MW	No PV	High load >70% 2015=10810 MW 2019=10870 MW

A. Manual Clustering

"Manual Clustering" method is a kind of extension of the Extreme Cases method. In order to "determine the parameters of the methodology", the user manually divides the space in boxes. Unlike in the Extreme Cases method, these boxes cover the whole space. The historical imbalance observations are thus classified according to multiple features (e.g. two axes, e.g. wind and photovoltaic production forecast). In Figure 12, four clusters are illustrated, corresponding to “low wind - low pv”, “high wind – low pv”, “low wind – high pv” and “high wind - high pv”).

For each of the clusters, a specific probability density curve is used, based on historical imbalances during the system conditions of the corresponding box. In the “actual sizing of the reserves” phase, the cluster in which the forecasted system condition for tomorrow falls is identified and the corresponding “specific” sizing is applied (Box 5).

Box 5: Manual Clustering in practice

Manual Clustering implies the manual definition of the clusters by the user. This choice can be based arbitrary, based on intuition or based on a statistical pre-assessment.

Several possibilities have been investigated when implementing the method. For the reason explain in the previous remark, it is decided to use only 3 features: PV forecast, wind forecast and load forecast. Categorization is based on the percentile of the yearly observed production or consumption, for instance medium wind considers the 33% to 66% highest observations for wind production.

%	Low	Medium	High
PV	0% 2015 = 0 MW 2019 = 0 MW	]0-75]% 2015 = 0-525MW 2019 = 0-665MW	]75-100]% 2015 =525-2266MW 2019 =665-2870MW
Wind	[0-33]% 2015 =0-270MW 2019 =0-500MW	]33-66]% 2015 =270-740MW 2019 =500-1425MW	]66-100]% 2015 =740-1900MW 2019 =1425-3670MW
Demand	[0-33]% 2015 =6500-9240MW 2019 =6530-9280MW	]33-66]% 2015 =9240-10690MW 2019 =9280-10750MW	]66-100]% 2015 =10690-13620MW 2019 =10750-13700MW

Both Extreme Cases and Manual Clustering methods relate on the user’s “manual” decisions. It should be emphasized that the prediction of the imbalance risk in day-ahead is a tough problem which requires looking for smarter methods. Let’s particularly notice that thousands of possible choices of clusters/extreme cases are possible. Many combinations have been tested. However, it would be a tremendous effort to test every possible choice manually.

B. Qualitative Clustering

More intelligent is the Quantitative Clustering, where the categories are identified without human interference, based on 'machine learning' techniques. In a sense, "Quantitative Clustering" automatizes the process to find the best choice of clusters. In order to "determine the parameters of the methodology", the algorithm computes the best clusters and the associated FRR needs. In the "actual sizing of the reserves" phase, the cluster in which the forecasted system conditions for tomorrow falls is identified and the corresponding "specific" sizing is applied.

Box 6: Qualitative Clustering

Quantitative clustering actually refers to a well-known machine learning clustering methods usually called the "k-means" algorithm.

The objective of k-means clustering is to split n observations into k clusters such as each observation belongs to the cluster with the nearest mean. This leads to a partitioning of the data space that looks like "Voronoi" cells. Mathematically, this means that the sum of square distances from observations to the assigned cluster centers is minimized, i.e.

$$\min \sum_i^k \sum_{x \in S_i} \left|\left| x - \mu_i \right|\right|^2;$$

where $(x_1, x_2, \ldots, x_n)$ is the set of observations and $(S_1, S_2, \ldots, S_k)$ are the set of the clusters of center μ_i. Of course, the type of norm remains the choice of the user. In this project, the norm picked is the 2-norm which is a popular choice.

The dimensions of the space (i.e. the system features) were normalized to [0;1] and weights were introduced on top of that to favor one or another dimension. Let's notice that this is a NP-hard problem which is therefore solved using a variety of heuristics (e.g. "Loyd", "Hartigan and Wong").

3.3.3. Continuous predictive statistics

While discrete methods relate on an "a-priori" portioning (manual or automated) of the space into clusters, continuous predictive statistics relates to continuous notions which do not involve the design of clusters (Figure 13).

A. Continuous Neighbors

The 'Continuous neighbors' method is based on a statistical analysis in which a model is trained to recognize similar system conditions (k closest historical measures). The observed day-ahead system conditions allow associating a probability density curve of the forecast risk, which is then convoluted with the outage risk.

In such methods, there is no need to "determine the parameters of the methodology". While the discrete methods could already define the clusters and the associated FRR needs (e.g. 1 month in advance), such intermediate step is not relevant in this method. However, this step can still be used at most to update the database.

In the “actual sizing of the reserves” phase, the nearest observations to the forecasted system condition for tomorrow are identified (the grey dots in the orange area on the figure) and used to compute the proper sizing.

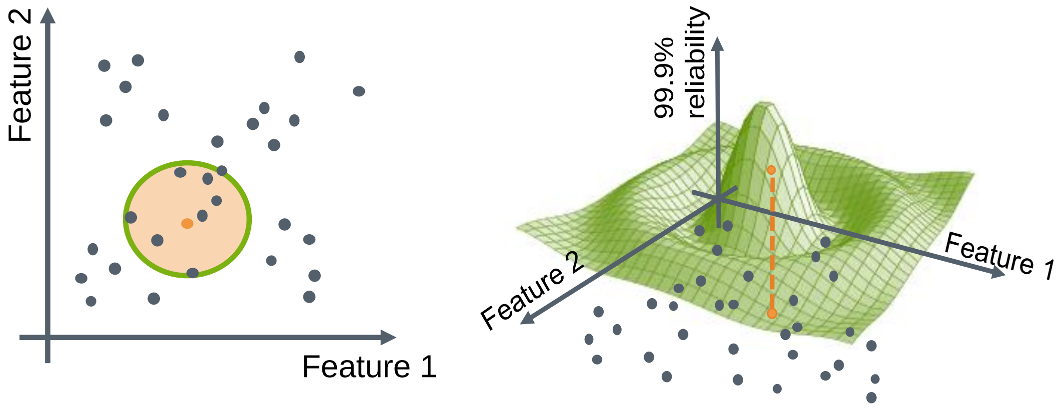


Figure 13: Visual representation of continuous neighboring method (left) and neural network method (right).

Box 7: Continuous Neighbors

“Continuous neighbors” actually refers to a popular machine learning method called “k-nearest neighbors” (KNN) algorithm.

Mathematically, it consists of finding the k observations $S = (x_1^*, x_2^*, \dots, x_k^*)$ among n observations $D = (x_1, x_2, \dots, x_N)$ such that

$$\left|\left|x_i^* - y\right|\right|^2 \leq \left|\left|x_j - y\right|\right|^2 \; \forall x_i^* \in S, x_j \in D \backslash S;$$

where y is the forecast system condition for tomorrow.

The type of the norm remains a user decision. In this project, the 2-norm was picked. Such as the k-means, the features were normalized to [0;1] and weights were added on top of that to favor the most relevant input features.

B. *Neural Networks*

In a continuous regression such as Neural Networks, the statistical analysis is used to determine a function of the system conditions, typically by means of “Neural Network” algorithms. In contrast to the other methods, it is based on regression which generates a continuous function model that estimates the FRR needs (Box 8), without explicitly building a distribution curve.

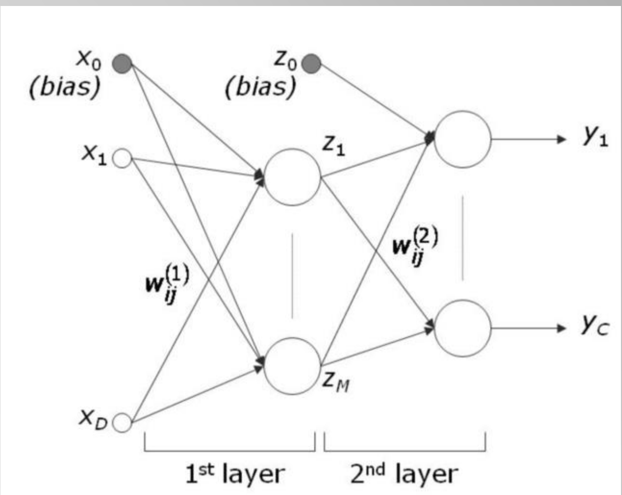
In order to “determine the parameters of the methodology”, the parameters of the function are estimated through a training process. In the “actual sizing of the reserves” phase, the function is evaluated at the coordinates corresponding to the forecasted system of tomorrow.

Box 8: Continuous Regression or Neural Networks

"Continuous Regression" actually refers to a popular regression machine learning method called "Artificial Neural Network" (ANN) algorithm. ANN aims to build a multi-layer non-linear regression model based on historical observations of an output y_k^x, i.e.

$$y_k(x) = h(\sum_i w_{ki}^{(2)} g(\sum_j w_{ij}^{(1)} x_j)) \ ,$$

where h and g are non-linear functions, y_k is the estimated output model, x_j are the inputs (i.e. each historical observation is a vector $(x_0, x_1, \dots, x_D)$ of the system features), y the output (e.g. the imbalance in our situation) and w_{ij} are the weights (i.e. parameter of the model). The previous expression can be schematically expressed which explains the name "neural network". The optimal weights are estimated through an error back-propagation process.



As the purpose in this study is to estimate the FRR needs, corresponding to a 99.9% quantile of the imbalance, the error function is designed as a "quantile regression" function. Whereas a classical error function, corresponding to the least squares, results in estimates that approximate the conditional mean of the response variable, quantile regression aims to estimate either the conditional median or other quantiles of the response variable. This error can be formulated as

$$E = \ q_\tau(y^x - y(x)) \text{ where } q_\tau(x) = \begin{cases} (\tau - 1)x, x < 0 \\ \tau x, x \geq 0 \end{cases}$$

where τ is the target quantile (e.g. 99.9% in our case).

4 ANALYSIS OF THE METHODS

4.1. Methodology

In order to assess the performance of the six methodologies, an assessment methodology is put forward based on a typical training, validation and testing process applied for machine learning methods.

- **Training**

The proposed statistical methods are based on mathematical models which need to be trained with historical data. Indeed, the models involve parameters such as cluster shapes for Quantitative Clustering or regression weights for Neural Network that need to be computed. In the training step, these parameters are determined.

For instance, when aiming to run a Quantitative Clustering method based on 10 clusters, the training step is the step where, based on a historical data set, the shape of these clusters is computed as well as the corresponding FRR needs.

- **Validation**

As highlighted in the previous section, the methods involve design decisions. The validation step is the step where the best design decisions are taken.

In the previous example it is decided to have 10 clusters. However, alternatively, it could also be based on 8, 9 or 11 clusters. The validation determines the best option. For a specific design decision, e.g. 10 clusters, the method is trained and then its performances are evaluated on a validation data set. This process is repeated iteratively and the design choices showing the best performance are retained.

- **Testing**

When the method has been designed optimally through the validation step, the next step is to test it on a new data set. This will contain sets of randomly chosen sequences of days as explained in Box 9. This is necessary to evaluate its performance because the performances should not be biased by any decisions. That is why the train, validate and test data sets should be independent.

In the context of this study, the models have been trained and validated on a data set of 2015 (and 2019), and then trained and tested on a data set of 2016 (respectively 2020). For this, the observed time series of predicted system conditions and system imbalances are used (without forced outage events). For 2019, and 2020, incremental renewable capacity is taken into account amounting up to 2466 MW and 3176 MW from end-2016 to respectively begin 2020 and end 2020 by means of the same approach presented in Section 2. To ensure the most efficient use of the limited database (two years[6]), an n-fold cross validation and testing is used (Box 9).

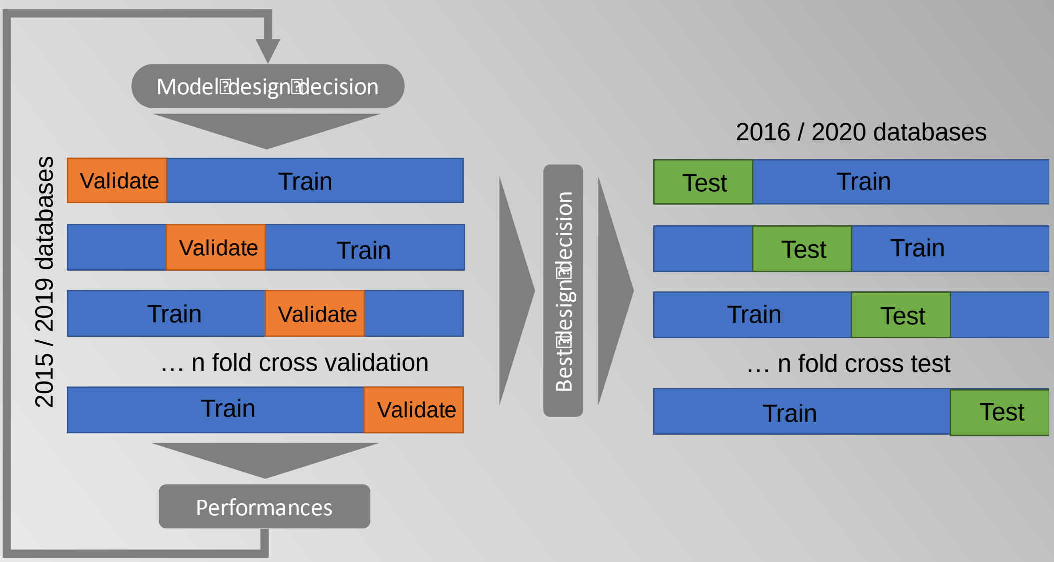
[6] When using historical observation, the time horizon which can be used is limited in order to ensure representativeness of the data (following system and market evolutions).

Box 9 : N-fold cross validation

The amount of available data is generally an issue when designing machine learning methods. As highlighted above, in other not to bias the results, the training, validation and testing sets should be independent.

The trade-off is that it would be preferable to have enough data in the training set such that the model "learns" correctly and at the same time, there should be enough data in the validation and test sets in other to draw performance conclusions on a trustable basis. Therefore, instead of making a unique split between training, validation and test sets, the split is repeated n times, to get the most out of the limited years of representative data.

On top of that, notice that this "split" is made randomly, in order to avoid user interference with the output performances.



For the simulations presented in this study, a 16-fold cross validation has been used, using each time a set of 22 days for testing and 343 days for training. Each day of the set is picked randomly without replacing it (not possible to have two identical days in one set). On top of that, outages have been added to the system imbalance via Monte Carlo simulation time series. As these are random events, 50 simulations with different forced outages have been conducted.

4.2. Scenarios

A reference case is constructed for 2016 and 2020. The 2016 case is not only a crucial step in assessing the performance by means of the observed system imbalance; it also allows assessing the potential of dynamic sizing in systems with lower renewable generation, neglecting the incremental capacity of wind power and photovoltaics. It is a scenario to study the dynamic potential when having limited renewable generation in the system.

Nevertheless, it will be only in the years to come that such a methodology can be implemented. Therefore, 2020 is in fact the real reference case, taking into account an incremental installed capacity of renewable generation (photovoltaics, onshore and offshore wind) between 2016-2020, as well as the observed system conditions of 2016 (weather, demand and renewable generation forecasts).

The statistical model is combined with an outage risk assessment based on the power plant generation mix of the corresponding years. In both years 2016 and 2020, NEMO-link is implemented by means of a perfect forecast model, i.e. the schedule is known in day-ahead[7], while not taking into account the offshore wind power cut-off following storm (being subject to further assessments in a separate study), but will be incorporated as a separate sensitivity check. The predefined reliability level is set at the current target of 99.9%.

The large incremental capacity of renewable capacity between 2016 and 2020 is expected to have strong impact in the expected imbalance and might therefore overestimate the FRR needs. As the reference case does not include any correction to take into account the ability of market players to improve their ability to cover part of the incremental prediction errors of renewable energy, this scenario can be seen as rather conservative. This allows identifying the maximum potential of dynamic sizing.

Nevertheless, a correction factor may be needed to avoid an overestimation of the final FRR needs, particularly when sizing several years ahead towards 2020. Therefore, the reference case is accompanied by the “progressive market case 2020” including a correction factor, based on historical observations. This assumes the ability of the market to cover part of the prediction errors, similar as observed in the past. This correction factor is determined at 4.5% a year, amounting up to a total reduction of 17% of the total system imbalance towards 2020 (Box 10). Note that when implementing the dynamic approach, and thus determining the FRR needs close to real-time, taking into account the latest installed capacities, such correction factors on incremental capacity becomes obsolete.

In order to further test robustness of the methods, a sensitivity is also conducted for a post-nuclear context (replacing 1 GW nuclear units by 6 CCGTs of 400 MW).

[7] In reality, the schedule is the result of the market and therefore not known at the time of sizing. A reliable forecast of the market outcome is therefore required, and will, depending on its accuracy impact the potential of dynamic sizing. This is further assessed in the second part of the study.

Box 10: Correction factor for incremental wind power and photovoltaics

An analysis has been conducted on the system imbalance from 2012 to 2016 in order to derive a consistent figure representing the expected impact of incremental renewable capacity on the system imbalance. The analysis determines the expected system imbalance (without forced outages) for 2016 as if it was conducted in 2012, 2013, 2014, and 2015, taking into account the prediction errors of the incremental renewable capacity. The result of this estimation shows the up- and downward capacity needs corresponding to the 99.9% reliability level, and is compared to the observed system imbalance in 2016.

	historical SI converted to 2016 (with RES capacity)				
	2012	2013	2014	2015	2016
99.9% UPWARD [MW]	801	761	745	611	666
99.9% DOWNWARD [MW]	1039	1027	713	663	713
Upward Correction [%]	-16.79	-12.46	-10.60	8.99	N.A.
Downward Correction [%]	-31.39	-30.59	-0.11	7.50	
Upward Correction [%/y]	4.49	4.34	5.45	-8.99	

RES: Renewable Energy Sources, i.e. wind power and photovoltaics

The results show that only part of the prediction error contributes to additional system imbalances, resulting in overestimations of the impact of incremental renewable generation on the forecast risk. In the current 'static' sizing methodology, working with a time horizon of 1 year, this is resolved with improvement factors representing system imbalance improvements and forecast improvements in intra-day. The analysis here puts forward a yearly correction factor of 4.5% which is rather consistent with the 2012-2014, and results in a total improvement of 17% of the expected system imbalance. Note that the upward side is selected to avoid extrapolating the impact of the market design changes (moving from dual to single imbalance prices). It is also to be noted that market design changes may have let to historic improvements which may not be attained in the future and can therefore be rather optimistic.

4.3. Assessment criteria

In order to assess the dynamic sizing methods, a cost and benefit analysis is conducted based on reliability criteria and FRR needs capacity criteria.

The **reliability criteria** are important to assess to which extent the method is able to attain the pre-defined reliability, e.g. 99.9% in the reference case, and contains three sub-criteria:

- Average reliability (expressed in %) which assesses the overall reliability over all periods over which it is assessed. It represents the amount of hours in which the FRR needs are not able to cover the system imbalance.
- Instantaneous reliability (expressed in %) which assesses the reliability in certain system conditions. It measures the average reliability in periods which are associated with high and low risks, respectively periods with high and low FRR needs. This indicator is further explained in Box 11.
- Reliability indicators such as the maximum uncovered system imbalance (excess or shortage), at least the part which is uncovered (expressed in MW), as well as the average uncovered system imbalance per failure (expressed in MWh).

The reliability criteria, and in particular the average reliability, is also used as a minimum technical criterion, i.e. a pre-condition, for a method to be eligible as method for dimensioning FRR needs. Methods which are found to not provide stable pre-defined reliability criteria will be discarded.

Box 11 : Instantaneous Reliability

A dynamic sizing method is in fact a kind of forecast model that attempts to predict the risk of a system imbalance. Therefore, the output FRR needs computed by such method can be interpreted as a measure of the risk: a high FRR needs means that the method foresees a high risk of imbalance while a low FRR needs means that the method foresees a low risk of imbalance.

The instantaneous reliability is therefore computed in three steps: (1) the method computes FRR needs, after which (2) the hours for which the sizing is "low" is classified as a low risk period while the hours for which the sizing is "high" is classified as a high risk period. Finally (3), the reliability is computed in both the "low risk" period and the "high risk" period set. This is done for both up- and downward FRR needs corresponding to up- and downward risk.

The exact definition of the high and low risk is subject to a trade-off: ideally the "high" should be high enough (and same for the low), i.e. only considering "extreme" situations but at the same time there should be enough hours in the high (and low) group to be representative. In the results, the low sizes are the 25% quantile while the high sizes are the 75% quantile.

While the reliability criteria are used as a minimum condition (if a method is not able to guarantee a stable, pre-defined reliability level the method will be discarded), the **capacity criteria** are important to study the economic potential. It assesses to which extent the method varies the FRR needs over time and also contains three criteria:

- Average capacity which assesses the yearly reduction of increase in FRR needs compared to a static approach (expressed in MW).

- Capacity distribution which assesses the percentage of the time certain FRR needs occur, as well as the spread between the maximum as the minimum FRR (expressed in MW).
- Capacity indicators which assess the dynamic behavior of the FRR needs, i.e. the hourly, daily variations of the FRR needs, as well as the capacity requirements during high and low risk periods.

These assessment criteria are used to assess the reference cases, but also the other cases defined in Section 4.2. This allows assessing the **robustness** of a dynamic method. Indeed, a successful methodology should attain stable results over the different cases representing different system conditions. Indeed, future system evolutions are uncertain and it is important that the selected method is applicable in each scenario.

4.4. Results of the analysis

A detailed overview of the results of the simulations can be found in 4 tables in Annex. Table A represents the average reliability and FRR needs, while Table B shows the spread, i.e. the minimum and maximum FRR needs. Table C focuses on the instantaneous reliability, showing reliability and FRR needs during high and low risk periods. Finally, Table D depicts two additional reliability criteria, i.e. maximum uncovered imbalance, and the average uncovered imbalance per failure. Table 8 provides a general overview which is elaborated further into detail in the next sections.

Table 8: General Overview of the results of the 6 methods in all scenarios

	RELIABILITY [%]					FRR NEEDS [MW]									
	Average Reliability					Average UPWARD FRR needs					Average DOWNWARD FRR needs				
	2016	2020				2016	2020				2016	2020			
		Ref	Wind cutoffs	Low Reserve	Post Nuke		2020	Wind cutoffs	Low Reserve	Post Nuke		2020	Wind cutoffs	Low Reserve	Post Nuke
Static	99.93	99.87	99.91	99.87	99.87	1261	1510	2101	1240	1341	1099	1539	1537	1287	1536
Outage Only	99.90	99.86	99.86	99.87	99.86	1233	1465	1466	1215	1263	969	1530	1510	1280	1533
Extreme cases	99.90	99.84	-	-	-	1230	1478	-	-	-	968	1495	-	-	-
Manual Clustering	99.89	99.76	-	-	-	1237	1401	-	-	-	949	1334	-	-	-
Quantitative Clustering	99.89	99.81	99.83	99.85	99.81	1236	1416	1418	1179	1160	945	1388	1386	1159	1391
Continuous Neighbors	99.89	99.86	99.86	99.88	99.87	1232	1426	1428	1188	1202	945	1448	1448	1209	1450
Neural Networks	A preliminary appraisal is conducted for the prediction risk.														

4.4.1. Reference cases 2016 and 2020

(A) Outage-Only

In the Outage-Only method, only the power plant outages and the HVDC interconnectors are treated dynamically. In Section 2, it was already concluded that:

- The dynamic potential of power plant schedules following economic dispatch is very limited as the outage risk is largely determined by largest unit, being nuclear power plants, which are operated as base load and hence running most of the time. Therefore, only extreme situations which have an effect on the FRR needs are taken into account, reducing the reserve needs when dealing with simultaneous maintenance of three 1 GW nuclear units.
- The dynamic potential of interconnectors following economic dispatch might be relevant, certainly for the downward reserves for which the outage risk is only determined by means of the interconnector. Reaping this dynamic potential remains subject to the ability to predict its schedule before day-ahead. For now, a perfect forecast model is considered assuming the NEMO-link is 60% of the time in export, 30% import, and 10% having no flow (cfr. Footnote 7).

When testing the dynamic potential based on the scheduling of the NEMO-link, Table A and C show a stable reliability compared to the static approaches. This method is indeed only going to increase the FRR needs in periods when scheduling NEMO in import, and providing more downward reserves when scheduling NEMO in export. As these simulations are conducted based on a perfect forecast assumption of the interconnector, no fundamental impact on average reliability level should be observed.

The results of the reference case 2020 show average volume reduction of 45 MW respectively 9 MW for up- and downward reserve needs, compared to a static method for which up- and downward FRR needs are determined respectively at 1510 and 1539 MW (Table 9). The average reduction for upward is explained as the NEMO-link is assumed to be scheduled most of the time in export mode.

The moderate dynamic potential is confirmed by looking at the spread between the minimum and maximum FRR needs of 43 MW and 26 MW for respectively up- and downward reserve needs in 2020, at least in situations without nuclear plant maintenance (Table 9).

Table 9: Average, Minimum and Maximum FRR needs [MW] in the reference case 2016 & 2020 for the Outage-Only method

2016	Static	Avg	Min	Max	2020	Static	Avg	Min	Max
up	1261	1233	1220	1260	up	1510	1465	1460	1503
down	1099	968	730	1100	down	1539	1530	1512	1538

In the reference case 2016, due to the larger role of the outage risk following lower renewable capacity in the system, a dynamic reserve sizing provided larger potential on the downward side with an average FRR needs reduction of 131 MW (compared to a static reserve need of 1099 MW) and a spread of 370 MW. In contrast, the upward potential is limited due to the existence of several 1 GW nuclear power plants, i.e. 28 MW reserve need reduction compared to a static sizing of 1261 MW. The spread is here only 40 MW.

However, in some particular situations the method allows to realize high temporary reductions by incorporating rare but scheduled events such as planned maintenances of power plants, or NEMO-link. An analysis of the outage risk shows that the reserve requirement change when having a few units in simultaneously planned maintenance (Table 10). It can be seen that for 2020, the static upward FRR needs of 1510 MW can be reduced to 1370 MW, 1320 MW and even 1230 MW if NEMO-link is in maintenance during simultaneous maintenance of 2, 3 and 4 nuclear units of 1 GW. This is not fundamentally different if NEMO-link is scheduled in export, and even in import mode a lower but still substantial reduction is observed.

The same method also allows increasing FRR needs in a case that a storm risk is predicted. The compatibility of the methods with storm risks is further discussed in Section 4.4.2.

Table 10: FRR needs [MW] in 2020 in particular situations related to the outage risk

		static	no plant in maintenance	1 nuclear maintenance	2 nuclear maintenance	3 nuclear maintenance	4 nuclear maintenance
NEMO import	Up	1510	1500	1480	1430	1380	1330
	Down	1539	1510	1510	1510	1510	1510
NEMO export	Up	1510	1460	1430	1380	1320	1240
	Down	1539	1530	1540	1530	1540	1540
NEMO out	Up	1510	1470	1430	1370	1320	1230
	Down	1539	1510	1510	1510	1510	1510

In conclusion, the Outage-Only approach is a rudimentary dynamic sizing method. Although the average upward FRR needs reductions are moderate (up to 45 MW), the benefits in some moments can amount to 200 MW and higher when facing simultaneous outages of the larger nuclear units. Following the decreasing importance of the outage risk, this method will not be suitable for a stand-alone implementation on long term, but can be useful in a gradual or hybrid implementation, i.e. combined with the advanced statistical methods.

(B) Extreme cases and Manual Clustering

The 'Extreme Cases' method and "Manual Clustering" follow the same nature, i.e. the user defines high and low risk conditions based on historical observations and statistical analyses to define a table with pre-defined volumes related to certain system conditions (e.g. low wind and low photovoltaics and high demand).

It is found that both methodologies do not attain an acceptable performance on the reliability criteria. Table A shows how in the reference scenario 2020, the average reliability is drastically reduced. Looking at the instantaneous reliability, it seems that both methods are not capable to determine the FRR need which corresponds to the risks.

Table C and Figure 14 show how the method reduces the downward reserves during low risk conditions, but also reduces reliability far under the predefined reliability level of 99.9%, during these low risk conditions. Furthermore, it slightly increases the FRR needs during high risk conditions, while the reliability target of 99.9% was already exceeded during these high risk periods. This demonstrates that the method fails to correctly recognize high and low risk periods. Furthermore, reliability criteria demonstrate that when facing a failure, this results in a larger loss or excess of energy and loss or excess of power compared to the static approach (Table D). Indeed, simulations show a maximum loss of power of 809 MW, instead of 401 MW in the static case.

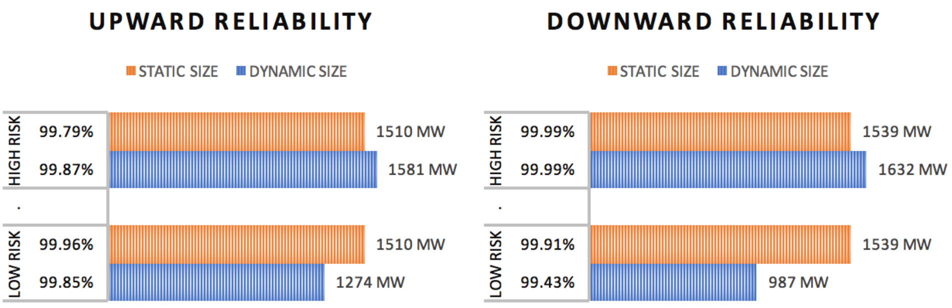


Figure 14: Up- and downward reliability and FRR needs [MW] in low and high risk cases for the reference case 2020

As no solutions are found to improve the behavior of these methods, these are discarded for further analysis following the minimum technical pre-condition on reliability. Indeed, the reliability criteria, which is seen as a minimum technical requirement is not attained. It seems that intuitive approaches to pre-define risk conditions fail to grasp the real complexity of the behavior of the system imbalance.

(C) Quantitative clustering and continuous neighbors

Although the "Quantitative Clustering" and "Continuous Neighbor" method are based on a different mathematical approach (discrete versus continuous), both are machine learning techniques, and it is shown that both obtain similar results. The **reliability**

criteria in the reference case show that both methods attain an acceptable average reliability, similar as the static approach (Table A).

Further analysis of the **instantaneous reliability** shows how the methods are both capable to identify high risk periods and increase the reliability during these periods by foreseeing more reserves compared to the static approach (Table C and Figure 15). On the other hand, they reduce the reserve needs during the low risk periods, lowering the reliability to the pre-defined level but avoid oversizing during these periods as is the case with static reserve. Such method allows achieving a more stable reliability over time, and therefore a better reliability management.

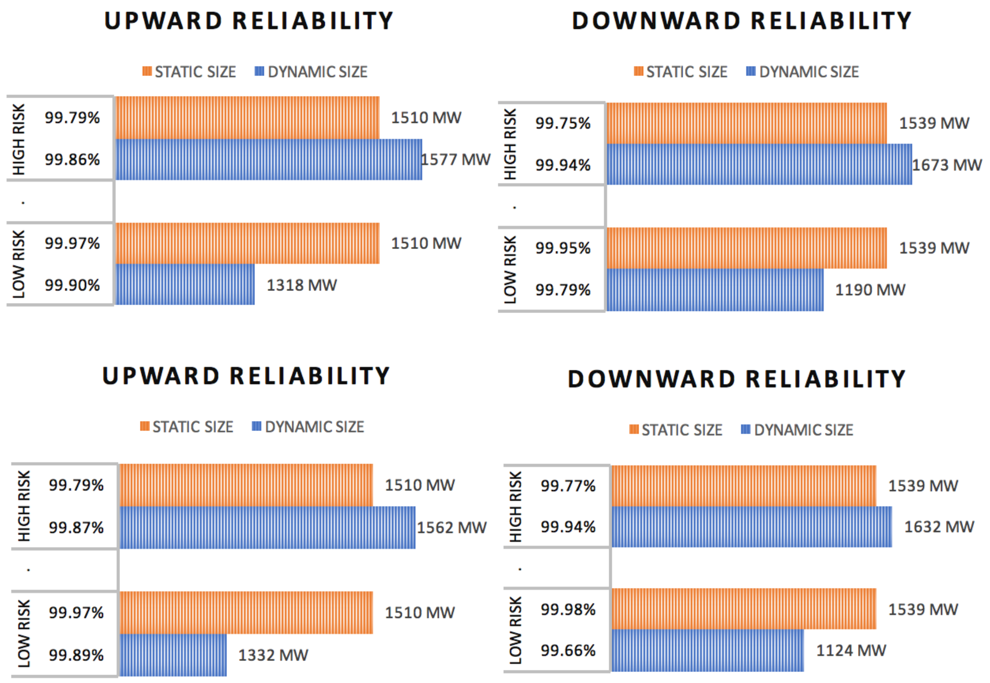


Figure 15: Up- and downward reliability and FRR needs [MW] in low and high risk cases for Continuous Neighbors (up) and Quantitative Clustering (down) for the reference case 2020

When studying the impact on the **FRR needs** (Table 11), it is observed that there is a potential in terms of volume reductions, both for up- as well as downward FRR. This is a difference compared with 2016 where limited upward potential is observed as the FRR needs is mainly driven by the outage risk (a reduction of 25 MW on the average FRR needs compared to static in continuous neighbors method).

Table 11: Average, Minimum and Maximum FRR needs [MW] in the reference case 2016 & 2020 in the Continuous Neighbors (above) and Quantitative Clustering method (below)

2016	Static	Avg	Min	Max	2020	Static	Avg	Min	Max
up	1261	1236	1170	1300	up	1510	1426	1262	1746
down	1099	945	500	1150	down	1539	1448	1038	1745

2016	Static	Avg	Min	Max	2020	Static	Avg	Min	Max
up	1261	1232	1175	1290	up	1510	1416	1322	1779
down	1099	945	500	1140	down	1539	1389	997	1698

Indeed, in 2016 there is not enough renewable capacity in the system yet to have a substantial impact of the forecast risk, limiting the potential of the machine learning methods. On the downward side there is potential in 2016 following the scheduling of NEMO, similar to the Outage-Only case (a reduction of 154 MW on the average FRR needs compared to static in the Continuous Neighbors method). This is not fundamentally different for the Quantitative Clustering method.

In 2020, this potential increases substantially following the increased capacity of renewable energy. A reduction on the average FRR needs of 84 MW and 91 MW for up- and downward FRR needs is realized for the Continuous Neighbors compared to the static approach. Furthermore, the spreads between the minimum and maximum are between 1262 MW and 1746 MW for upward FRR needs and between 1038 MW and 1745 MW for downward FRR needs. This is not fundamentally different for the Quantitative Clustering method.

A criterion to study the time in which the reserve needs are reduced or increased is the **FRR needs duration curve** (Figure 16). When studying the duration curves, these do not only show the gap between the minimum and the maximum, but also show that for instance that the dynamic upward FRR needs are under the static FRR for 82% of the time for the Continuous Neighbors. Of course, there are also moments where the FRR needs are increased following high risk conditions.

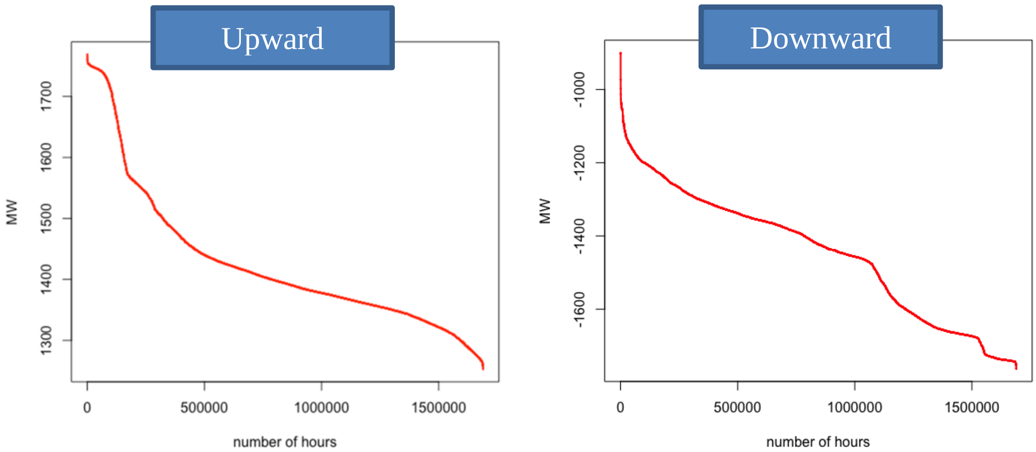


Figure 16. FRR needs duration curve in the Continuous Neighbor's method for the reference case 2020. The left graph is upward FRR needs duration curve while the right graph is downward FRR needs duration curve

An example of the dynamic reserves is given in Figure 17 showing the FRR needs for a representative day. It shows how the method increases the upward FRR needs during high wind production forecasts (perceiving a high risk for a forecast error resulting in a potential shortage), and reducing the upward FRR needs during low wind conditions and night (perceived as less risk). Another trend is shown in the downward profile showing lower FRR needs when facing low wind generation and vice versa. Such analysis of the system parameters allow to explain the FRR needs.

In conclusion, the Quantitative Clustering and Continuous Neighboring are advanced statistical tools which achieve a reduction of the average FRR needs while obtaining a better reliability management. In both directions, average FRR reductions up to 100 MW can be obtained, with a spread between minimum and maximum which can amount up to 500 MW and higher. Furthermore, they show increasing potential with the renewable capacity installed.

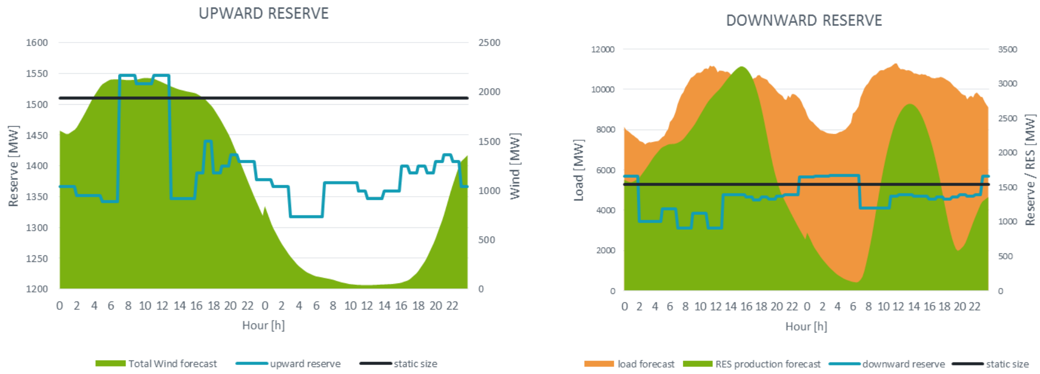


Figure 17: Example of a time series of the FRR needs during a representative two days (Continuous Neighbors) in the reference case 2020

(d) Neural networks

During the implementation of the method, it was found that the Neural Networks method did not facilitate a combination of the forecast risk and the outage risk. As the outage risk is a fundamental driver of the system imbalance and the FRR needs, and resolving the methodological issues was not possible following mathematical problem concerning the convolution of both imbalance driver categories, this method is not further investigated (Box 12).

In addition, a preliminary assessment based on the forecast risk shows that in terms of reliability, the model did not perform any better as the “Continuous Neighbor” and “Quantitative Clustering” method (Figure 18). In addition, no substantial volume savings were observed compared to these two methods.

Box 12: Technical complexity of the Neural Network methods

Machine Learning methods can be categorized in clustering (classification) methods and regression methods. In the scope of this study, the Neural Network method was put forward as a regression method: it directly estimates the desired FRR needs and does not build clusters (KNN) or scenarios (KMEANS).

Part 1 explains that regression methods are not compatible with the methodology of combining a prediction risk distribution with a forced outage risk distribution as a regression, by definition, directly estimates a value and not a distribution (the output of the algorithm is not a distribution but a single numerical value). This is a substantial obstacle in the use of ANN for the dynamic sizing following this pre-defined methodology implying the convolution of probability distributions.

However, simulations were conducted comparing the results of an ANN with the other machine learning methods, without taking into account the forced outage risk. These simulations show that ANN was not superior to the other machine learning methods what concerns this specific problem. Therefore, as ANN did not match with the pre-defined methodology and did not put forward better performances, the method was discarded for the scope of this study.

However, even if the ANN was not selected for further analysis, it remains a possibility to further investigate the method in a later stage. Indeed, several strategies could allow going around the “distribution obstacle”. A first obvious strategy is to abandon the sizing methodology based on two probability distribution curves. Indeed, in theory, it would be possible to simulate e.g. 500 years of expected system imbalance, directly incorporating forced outages based on a Monte Carlo analysis. Another approach can be to compose percentile per percentile a model of the distribution of the prediction risk. Indeed, the ANN can be used to estimates several quantiles (e.g. 0.1%, 0.2%,…, 99.8%, 99.9%). This distribution can be convoluted with the outage risk distribution. Both methods would require intensive computations and should be further investigated before considering implementation.

4.4.2. Robustness

(a) Offshore wind power cut-off

When including the offshore wind power cut-off risk, i.e. sudden disconnection of offshore wind turbines following a storm, in the simulations for a 99.9% reliability level, this obviously impact the static reserves substantially, increasing the capacity over the level which can be lost when wind speeds exceed 25 m/sec (2101 MW). Of course, this is a rough assumptions and it is to be investigated:

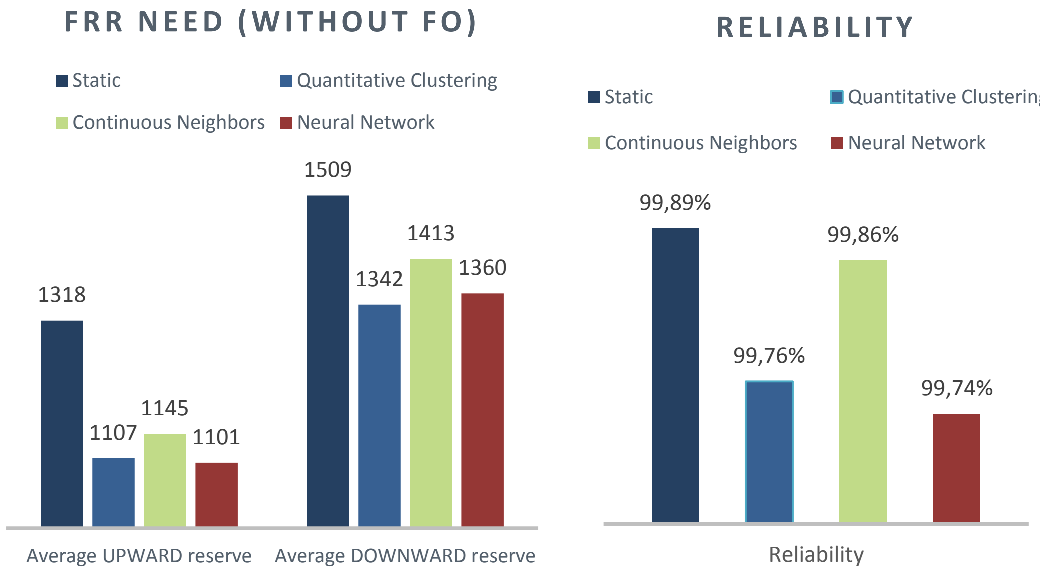


Figure 18: Results of the forecast risk (no outage risk is taken into account) on the FRR need and reliability for the reference case 2020

- What is the impact of a storm cut-off on the system imbalance? Should this be considered in the FRR volume determination and will storm cut-off lead to an increase of the reserve needs? Will a storm cut-off risk be predictable in day-ahead?
- What is the best approach to cope with this risk from a technical, economic and regulatory perspective? Will Elia need to procure additional reserves or will the BRPs of the offshore parks be fully responsible for the dealing with this this issue?

Answering these questions is not in scope of the project and is subject to further assessments. This explains why the storm risk is not incorporated in the reference case, but assessed as a sensitivity, in order to check compatibility if it would be decided to treat it as an outage (as today).

Some rough assumptions are taken to incorporate this event as a dynamic event in the outage risk. For this, it is assumed that a storm risk can be predicted and has an impact on the FRR needs, and is identified when the offshore wind speeds exceed 20 m/sec. In this case, the entire capacity with a cut-off speed of 25 m/sec is incorporated in the Monte Carlo simulations, largely impacting the result.

Results show that the impact on the average upward FRR needs is completely mitigated, when a dynamic sizing method is used due to the rarity of the event (69 hours per year). Nevertheless, this will result in some periods with very elevated reserves, i.e. 2093 MW or higher, and the question remains how to source these means in such moments. Again, it is to be stressed that at the time of writing it is still uncertain to which extent such events would impact the system imbalance and how to treat the balancing responsibility of the market parties in such cases.

Table 12: Average, Minimum and Maximum FRR needs [MW] (Continuous Neighbors) in the offshore cut-off case

	2020 Ref	2020 Offshore Cut-Off			
FRR NEEDS [MW]	**Static**	**Static**	**Average**	**Min**	**Max**
up	1510	2101	1426	1262	2101
down	1539	1599	1448	1038	1745

(b) Progressive market case 2020

It is recognized that the reference case 2020 might be rather conservative in terms of FRR needs. This is due to the fact that the FRR needs are determined incorporating the incremental capacity between 2016 and 2020, without assuming that the market will cover part of the incremental prediction errors as perceived in the past. This extrapolation may potentially oversize the reserves and a correction might be necessary.

Therefore, a sensitivity analysis is conducted with a correction of the expected imbalance in 2020. This correction factor is set at 4.5%/y, when expressed in incremental capacity, as explained in Box 10. This factor is used to correct the expected system imbalance (17% in total), and reduces the reserve needs substantially. In addition, it is also assumed that market parties will be able to better cover their forced outages.

Table 13: Average, Minimum and Maximum FRR needs [MW] (Continuous Neighbors)in the progressive market 2020 case

	2020 Ref	2020 Progressive Market			
FRR NEEDS [MW]	**Static**	**Static**	**Average**	**Min**	**Max**
up	1510	1240	1188	1099	1435
down	1539	1287	1209	865	1458

It is found that such a high performing market would reduce the static up- and downward FRR needs up to 1240 MW and 1287 MW respectively. It is found that the dynamic potential is reduced, but still pertinently available (25 MW, 61 MW and 51 MW average reductions on the upward side for Outage Only, Quantitative Clustering and Continuous Neighbors) and 7 MW, 129 MW and 78 MW for downward FRR needs. As this scenario can be seen as highly optimistic in terms of imbalance risks, the result will serve as a floor for the potential of dynamic sizing.

(c) Post-nuclear case

All results show that the outage risk of the large nuclear units (1 GW) have a substantial impact on the results. It is therefore investigated what the effect would be of a post-nuclear case. Therefore, an analysis is conducted where the largest nuclear units are replaced by 6 CCGT units of 400 MW in 2020. Although this context is theoretical, it teaches us the sensitivity towards the generation mix.

It is found that such a scenario has a reducing effect on the upward FRR needs, for which the static FRR needs are reduced to 1341 MW due to the reduction of the outage risk. But, in contrast to the progressive market case 2020, the dynamic potential increases due to the increasing importance of the forecast risk, showing average upward reserve reduction savings of 78, 182 and 139 MW in the Outage Only, Quantitative Clustering and Continuous Neighbors.

In contrast, the static needs for downward reserves are not impacted and remain at a level of 1536 MW. The potential is thus lower than in the upward side with 6, 145 and 86 MW, however with a substantial difference between minimum and maximum sizing.

Table 14: Average, Minimum and Maximum FRR needs [MW] (Continuous Neighbors) in the post nuclear case

	2020 Ref	2020 Post Nuclear			
FRR NEEDS [MW]	**Static**	**Static**	**Average**	**Min**	**Max**
up	1510	1341	1202	1039	1694
down	1539	1536	1450	1038	1744

5 COST AND BENEFITS ANALYSIS AND METHOD SELECTION

5.1. Cost and Benefit Analysis (CBA) and Method Selection

The previous section explains how the minimum technical requirement, the ability of a method to guarantee a stable pre-defined reliability level, allowed discarding 3 methods: Extreme Cases, Manual Clustering and Neural Networks.

In order to assess, rank and select the remaining three methods, the first element of the assessment, i.e. the reliability which is expressed in percentage, shows how a better reliability management is obtained. It is difficult to express this in financial terms as a better reliability management translates into better Area Control Error quality which is difficult to valorize in monetary gains. Nevertheless, this analysis allowed discarding the 3 methodologies which are not able to deliver the required reliability levels.

The second criterion of the assessment has an economic nature, i.e. the FRR needs reductions which are expressed in capacity. It is used to confirm the dynamic potential of the three successful methods and recommend these for the Proof of Concept. It is also not straightforward to express the FRR needs reductions in 2020 in financial terms. This is due to (1) lack of reference prices following recent introduction of new product types, (2) uncertainty concerning on-going and future market evolutions such as the evolution of the reserve needs, appearance of new service providers and future technical requirements of the European standard products for balancing energy, (3) the lack of experience with daily auctions, and corresponding price elasticity. Moreover as previously explained there is no one-to-one link between the reserve needs and means as a part of reserve needs can be covered via reserve sharing and non-contracted available reserves.

At this stage, a financial analysis is not necessary as all 3 remaining methodologies, which are respecting the minimum technical requirement, are selected for the Proof of Concept. In the second part of the study, the Proof of Concept will provide a time series of the FRR needs for an entire year, while also providing information on the reserve sharing potential and the predicted system conditions (e.g. availability of downward wind power reserve power). These will enable to better estimate the effective amount of contracted reserves which need to be contracted and hence a more accurate estimation of the financial benefits will be possible.

Nevertheless, a rough estimation is conducted in this section to provide the order of magnitude of the potential gains of the average FRR reductions, assuming that these benefits occur from mFRR reductions. Therefore, an overview for the average FRR reductions of the three successful methods is presented in Table 15, showing a clear reduction in the average volume. This demonstrates the value of a dynamic sizing method. This is even confirmed when looking at the minimum potential provided by the high market balancing scenario (Table 16).

Table 15: Summary of average up- and downward FRR needs reductions (benefit) and maximum spread for three successful methods in reference case 2020 (= max potential)

[MW]	Static	Dynamic	Benefit	Min	Max	Spread
	Upward FRR needs [MW]					
Outage-Only	1510	1465	-45 (-3%)	1460	1503	43
Quantitative Clustering		1416	-94 (-6%)	1322	1779	457
Continuous Neighbors		1426	-84 (-6%)	1262	1746	484
[MW]	Downward FRR needs [MW]					
Outage-Only	1539	1530	-9 (-1%)	1512	1538	26
Quantitative Clustering		1388	-151 (-10%)	997	1698	701
Continuous Neighbors		1448	-91 (-6%)	1038	1745	707

Table 16: Summary of average up- and downward FRR needs reductions (benefit) and maximum spread for three successful methods in progressive market case 2020 (=min potential)

[MW]	Static	Dynamic	Benefit	Min	Max	Spread
	Upward FRR needs [MW]					
Outage-Only	1240	1215	-25 (-2%)	1214	1241	27
Quantitative Clustering		1179	-61 (-5%)	1120	1457	337
Continuous Neighbors		1188	-52 (-4%)	1099	1435	336
[MW]	Downward FRR needs [MW]					
Outage-Only	1287	1280	-7 (-1%)	1265	1287	22
Quantitative Clustering		1159	-128 (-10%)	834	1417	583
Continuous Neighbors		1209	-78 (-6%)	865	1458	593

While the relative (%) reduction in average FRR needs (compared to the static approach) could already give a rough indication of the relative (%) procurement cost reductions of the contracted FRR needs, a rough assessment in absolute terms is conducted. Generic FRR price scenarios are used to estimate the benefits in financial terms based on the average FRR needs savings. Therefore, assumptions are made on the average up- and downward reserve cost.

Upward reserve prices (including inflation) are assumed to vary between 2.5 (low price), 4.0 (medium price) and 5.5 (high price) €/MW.hour. This is derived from observed prices for mFRR in 2016 and 2017 as the FRR needs reductions will mainly incur mFRR reserves. It is to be noted that due to new product type introductions (R3 flex), and short-term auctioning, these prices cannot be considered stable (Figure 19). Furthermore, the impact of system evolutions towards 2020 on these prices is uncertain. The results are to be interpreted in a cautious way, and are only to be used as a first estimate of the potential.

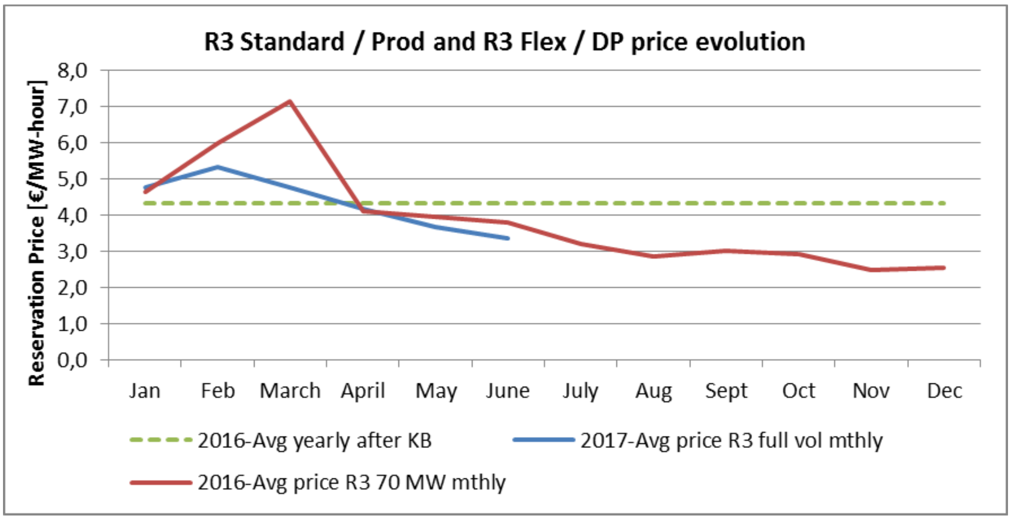


Figure 19: Prices for tertiary reserve products: R3 Standard, R3 flex and R3 Dynamic Profile (DP) (source: Elia)

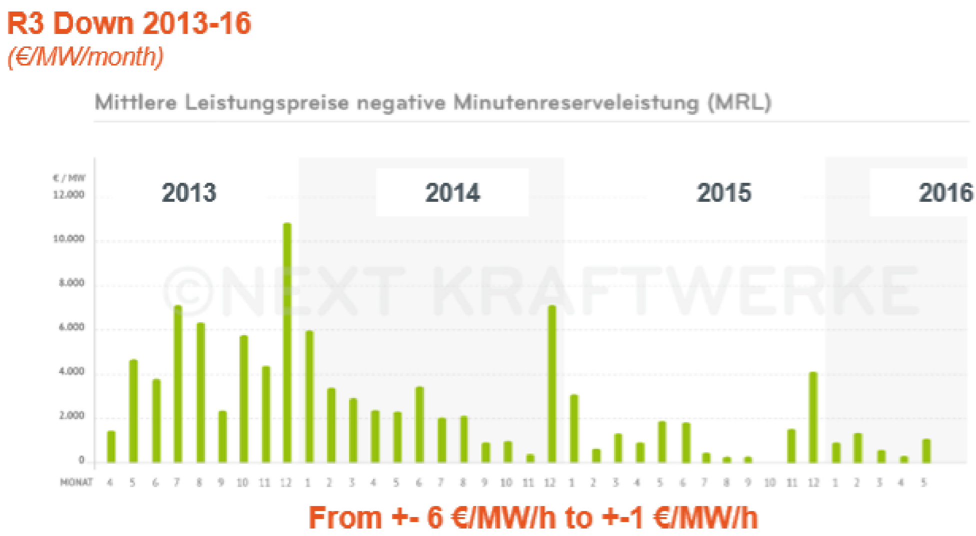


Figure 20: Prices for downward mFRR in Germany (source: Next Kraftwerke)

Downward reserve prices are assumed to vary between 1 (low price), 3 (medium price) and 6 (high price) €/MW.hour. In some cases with high wind power availability, this can potentially even become zero. The minimum is determined by the availability of downward flexibility following large wind farms, and the high price from observations of the maximum price for downward reserves in Germany. It is reminded that there is no experience with downward reserve in Belgium.

Table 17 shows how the benefits in the reference case 2020 of the average FRR need reductions can provide a yearly gain for respectively up- and downward reserves between €985,500 and €2,168,100, and €78,320 and €473,040 when implementing the intuitive Outage-Only method. With the machine learning tools, implementing the Continuous Neighbors or Quantitative Clustering, this would result in a yearly gain of €1,839,800 and €4,528,920 for upward, and €797,160 and €7,936,560 for downward FRR needs. Note that these values may further increase when facing extra-ordinary conditions such as scheduled maintenance of multiple nuclear units of 1 GW.

Table 17: Yearly benefits of average FRR needs saving in 2020 expressed in € in the reference case

	Upward FRR saving [€]			**Downward FRR saving [€]**		
Price [€)	Low 2.5	Medium 4.0	High 5.5	Low 1.0	Medium 4.0	High 6.0
Outage-Only	985,500	1,576,800	2,168,100	78,840	236,520	473,040
Quantitative Clustering	2,058,600	3,293,760	4,528,920	1,322,760	3,968,280	7,936,560
Continuous Neighbors	1,839,600	2,943,360	4,047,120	797,160	2,391,480	4,782,960

Table 18: Yearly benefits of average FRR needs saving in 2020 expressed in € in the progressive market case 2020

	Upward FRR saving [€]			**Downward FRR saving [€]**		
Price [€)	Low 2.5	Medium 4.0	High 5.5	Low 1.0	Medium 4.0	High 6.0
Outage-Only	547,500	876,000	1,204,500	61,320	183,960	367,920
Quantitative Clustering	1,335,900	2,137,440	2,938,980	1,121,280	3,363,840	6,727,680
Continuous Neighbors	1,138,800	1,822,080	2,505,360	683,280	2,049,840	4,099,680

Table 18 shows how the benefits in the progressive market case 2020 of the average FRR needs are reduced and can provide a yearly gain for respectively up- and downward reserves between €547,500 and €1,204,500, and €61,320 and €367,920 when implementing the Outage-Only method. With the machine learning tools, implementing the Continuous Neighbors or Quantitative Clustering, this would result in a yearly gain of €1,138,800 and €2,938,980 for upward, and €683,280 and €6,727,680 for downward FRR needs. Note that these values may further increase when facing extra-ordinary conditions such as scheduled maintenance of multiple nuclear units of 1 GW.

In addition to technical feasibility, the CBA demonstrates a reduced average FRR needs, and consequently a positive financial benefit **for the Outage-Only, Quantitative Clustering and Continuous Neighbors and these methods are therefore recommended to be further investigated in a Proof of Concept**.

Besides the fact that future reserve prices are uncertain, it is expected that these prices depend on the total volume to be procured as this volume varies from low to high values over time. Furthermore, these prices will be driven by the system conditions: it is expected that there will also be a strong interaction with the system schedule conditions: if a lot of conventional units are running (high demand), there is generally high downward flexibility; if conventional units are running at minimum load, there is generally high upward flexibility, and if limited conventional units are running, there is limited up- and downward flexibility. No experience is available at this stage with the procurement of dynamic FRR needs, and it is difficult to estimate the final impact on the benefits. This will be further investigated in the second part of the report based on the results of the Proof of Concept.

5.2. Practical implementation aspects of the methods

In preparation of the second part of the study, a first analysis is conducted analyzing the regulatory, procurement and implementation aspects. The Proof of Concept in the second part of the study will quantify the impact on the FRR needs and reliability by running the dynamic sizing method for entire year in 2020, while taking into account realistic constraints of the method.

5.2.1. Regulatory aspects

Concerning the regulatory aspects, the three proposed methods are compliant with the ENTSO-E SO GL as they are all based on the system imbalance, include a full year of data, apply a probabilistic approach and include a deterministic N-1 criterion.

Nevertheless, the dynamic methods are obviously more complex in terms of validation of the method and the results (although dynamic sizing does not require correction factor assumptions which are inherent to a static approach sizing at least 1 year ahead). Although the increase in complexity is marginal for the Outage-Only, this is not the case for the machine learning methods implementing statistical models calculating the FRR needs.

Indeed, as the Outage-Only only deals with the outage risk, a manageable table with the reserve needs for each schedule of assets can be provided ex ante, month- or year-ahead. The same principle applies for the Quantitative Clustering where this ex ante table with scenarios becomes more complex as it can include various system conditions determining the forecast risk. This might give a multi-dimensional table which may require deeper analyses to explain. In contrast, the Continuous Neighbors method does not allow such an ex ante check, and determines the reserve needs directly on day-ahead. Of course, an ex post check is still possible to explain the reserve needs.

A dynamic FRR dimensioning will change the current framework in which Elia determines its yearly volumes. Although Elia can still propose a yearly methodology, the calculation of the volumes themselves will be only be conducted a (few) days ahead.

5.2.2. Tender and product aspects

A dynamic dimensioning of FRR needs requires de facto a daily procurement to be useful. It is already put forward that this should at least be conducted before the day-ahead market closure. Dynamic procurement contains elements which are relevant to investigate. A first element is which **dynamic reserve share** is to be procured in short-term (i.e. daily), and which part can be procured on longer terms (e.g. monthly). The analysis learns that the reserve needs contain a rather large fixed amount, and only a small part varies following system conditions. The fixed volume does not necessarily need to be procured by means of daily tenders, although it can.

Another aspect is the **product length** of the FRR products, which is likely to reduce the dynamic potential when showing a longer duration. On the other hand, one hour products are not necessarily realistic from an implementation point of view, and are also bound to European standardization evolutions, which seem to evolve towards 4 hour products. Also the **sizing lead time** will be relevant as sizing too long before day-ahead market closure will result in less accurate results (and less potential), a lead time which is too short allows less time for the follow up processes (procurement).

These elements will be taken into account in the Proof of Concept, though the objective is not to determine the optimal tender process, or product design, but only to assess their potential implications. A study on **dynamic FRR procurement** is to be conducted to analyze the full potential of dynamic FRR dimensioning:

- Impact on the FRR means: How to translate the FRR needs in FRR means, i.e. R2, R3 flex and R3 standard while taking into account non-reserved capacity and reserve sharing?
- Impact on the procurement: How to evolve to a daily (e.g. daily) procurement? The latter will have strong implications for the market parties; this is to be discussed with stakeholders.

5.2.3. Business requirements

Although no insurmountable barriers are found, an important effort is needed from Elia to implement the tools for dynamic dimensioning. First of all, databases with available data to be coupled and elaborated (e.g. offshore wind observations) to have enough training data to develop a performant reserve dimensioning tool based on machine learning. This effort is less predominant in case of the Outage-Only method. Secondly, one main challenge is the development of the algorithms to be used in a real-life environment which is a large step beyond the application of these methods in a study. Besides the statistical tools themselves, tools are required which can predict the schedule of the NEMO-link (further investigated in second part of the study). In the second part of the study, a high level estimation is made what the impact and cost will be of an integration, storage, management and protection of the required information sources.

Secondly, the algorithms are to be developed and trained. Even if the development of such a tool would be outsourced to external experts, people are to be trained to operate the tool, understand the results and exploit them. Computing power of the methods might be less of a problem but specific software will need to be developed and replace the current excel tool. Fall-back mechanisms are to be put in place if the tool suffers an unexpected problem. These development and maintenance costs are to be assessed in the second part of the report.

PART 2: PROOF OF CONCEPT AND IMPLEMENTATION PLAN

6 SET-UP OF THE PROOF OF CONCEPT

6.1. Introduction

Now that the potential of dynamic sizing is demonstrated, the next step is to conduct a Proof of Concept in realistic conditions towards 2020. Therefore, the three selected methods, i.e. Outage Only (OO), Quantitative Clustering (KMEANS) and Continuous Neighbors (KNN) are simulated from hour to hour for the entire year 2020. These simulations can be seen as **a 'virtual' parallel run** comparing dynamic sizing approaches with the current static sizing approach, while taking into account real-life constraints which will be faced when operating a dynamic approach from day to day. This includes sizing resolution, sizing lead time and training frequency and the limited predictability of NEMO-link.

Therefore, three realistic scenarios are put forward towards 2020, taking into account the incremental capacity of renewable generation, as well as other expected system evolutions, while taking into account improvement factors following the ability of the market to cope with system imbalances. It is explained in Section 4.1 that these improvement factors are only relevant when assessing future system imbalances and FRR needs one or more years ahead. This correction becomes unnecessary in a real-life implementation of dynamic sizing which calculates FRR needs close to real-time. In addition to these scenarios, a separate analysis will be conducted on a "post nuclear" case, representing 2027, i.e. the Belgian power system after the decommissioning of the nuclear units, including incremental renewable capacity towards that year.

Firstly, the results, being a time series of FRR needs for an entire year are used in Section 7 to **analyze the impact on reliability, FRR gains, and robustness.** It is investigated if the conclusions made in Part 1 of the study are also valid when taking into account real-life constraints. Furthermore, the results are used to conduct a more **detailed analysis of the financial FRR gains**, taking into account the time series of FRR needs over an entire year, while taking into account price elasticity on the mFRR prices.

Secondly, the results are used to **study the behavior of the FRR needs over time**, and the **correlation with the predicted system conditions**. This is an important element to assess the feasibility of the method, as well as its transparency. Indeed, it is to be shown that the method can provide credible, non-volatile, explainable FRR needs. Furthermore, a sensitivity analysis is conducted on the scenarios and case study, as well on different design aspects including sizing resolution, training frequency and lead time. The results allow making the final selection of method and potential algorithms for a dynamic sizing approach to be used for an implementation phase.

Thirdly, the **implementation plan and corresponding cost** (including the development and operation of the necessary tools to implement a dynamic sizing method) are composed in Section 8. This completes the cost and benefit analysis by comparing the FRR needs reductions (benefit) with the implementation (cost). This provides all the information to assess the business case, and decide upon implementation

6.2. Dynamic sizing algorithms

The three selected dynamic sizing methods implemented in the Proof of Concept are Outage Only (OO), Quantitative Clustering (KMEANS) and Continuous Neighbors (KNN). In Part 1 of the study, these methods have proven to meet the minimum technical pre-conditions concerning reliability, while showing robust potential in terms of reliability management and FRR needs reductions.

As elaborately explained in Section 3.3, the **Outage-Only (OO)** method is an intuitive method which focuses on the outage risk by means of capturing planned maintenances of power plants and NEMO-link, as well as the scheduled import and export direction of NEMO-link. In contrast to the outage risk, the prediction risk is still treated statically. In contrast, the **Quantitative Clustering (KMEANS) and Continuous Neighbours (KNN)**, which are based on 'machine learning' techniques, provide the largest dynamic potential.

Figure 21 depicts the overall methodology implemented in the Proof of Concept which is based on the convolution of two probability density curves. The Outage-Only relies only on a dynamic behaviour of the outage risk, assuming different probability distribution curves over time depending on the status of the generation fleet and relevant HVDC-interconnectors, while assuming the prediction risk as one fixed probability curve. In contrast the machine learning methods also assume a dynamic behaviour of the prediction risk, assuming different probability distribution curves over time depending on the predicted system conditions. Box 11 explains the specific assumptions made for the algorithms in the Proof of Concept.

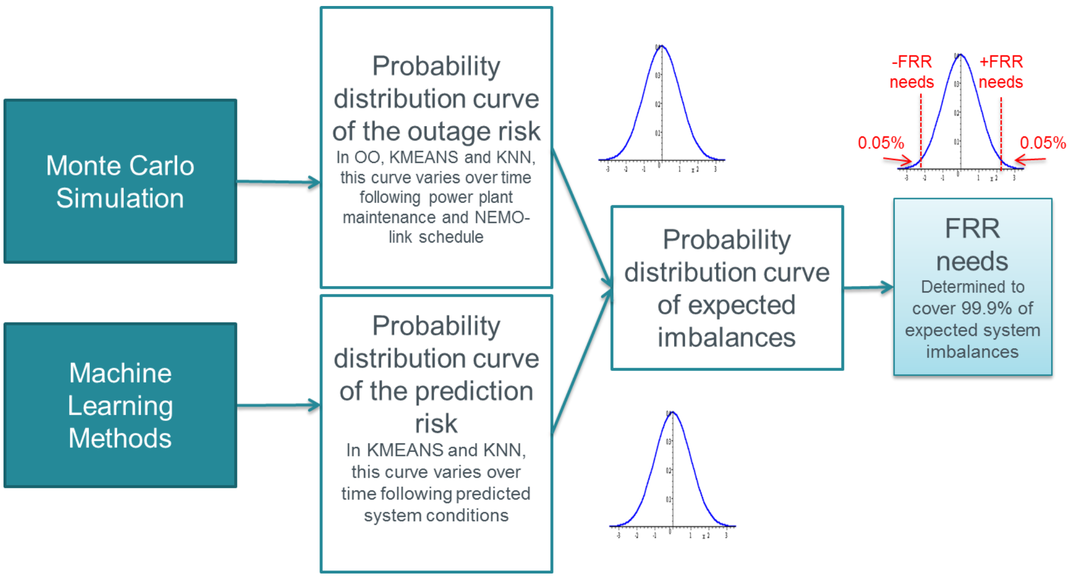


Figure 21: Schematic representation of the dynamic sizing methods

Box 11: Implementation of the methods in the Proof of Concept

Each method consists of several design decisions which allows optimizing the performance of the model. For the simulations presented in the report, the “design” decisions are identical for all simulations (features and parameters are selected after conducting sensitivities and analyses in Part 1 of the study):

- **Forced outage risk**: assumes an identical representation for the three methods, based on three forced outage distributions covering NEMO-link in export, import and both in import and export (when prediction is indecisive).
- **Prediction risk** (machine learning methods) is captured based on a training period of 1.5 year of input data based on following :
    - Inputs features:
    - Load forecast: normalized by its maximum
    - PV forecast: normalized by its maximum
    - Wind onshore forecast: normalized by its maximum
    - Wind offshore: normalized by its maximum
    - Hour of the day: sinus and cosinus (to represent cyclic behavior)
    - Load gradient: normalized by its maximum
    - PV gradient: normalized by its maximum
    - Absolute Scheduled leaps: normalized by its maximum
    - Temperature forecast: normalized by its maximum

    - Parameters of Quantitative Clustering (KMEANS)
    - Number of clusters: 10
    - Weights on the inputs: all 1
    - Norm: 2-norm
    - Probability distribution based on a kernel model with parameters:
        - sigma = (1/0.6745)*median(abs(cluster-median(cluster)))
        - bandwith = sigma*((4/(3*length(cluster)))^(0.2))

    - Parameters of Continuous Neighbors (KNN)
    - Number of neighbors: 3500
    - Weights on the inputs: all 1
    - Norm: 2-norm
    - Weights observation (in distribution): 1/sqrt(distance to observation), in order to favor the closest historical observations
    - Probability distribution based on a kernel model with parameters:
        - sigma = (1/0.6745)*median(abs(cluster-median(cluster)))
        - bandwith = sigma*((4/(3*length(cluster)))^(0.2))

6.3. Scenario's and assumptions

6.3.1. Scenarios for 2020

Figure 22 represents an overview of the scenarios and case study investigated in the Proof of Concept. In order to conduct this Proof of Concept, the dynamic sizing algorithms have to be trained, tested and implemented on historic data of system imbalance and system conditions. For this, a database is constructed for 2015, 2016 and 2017.

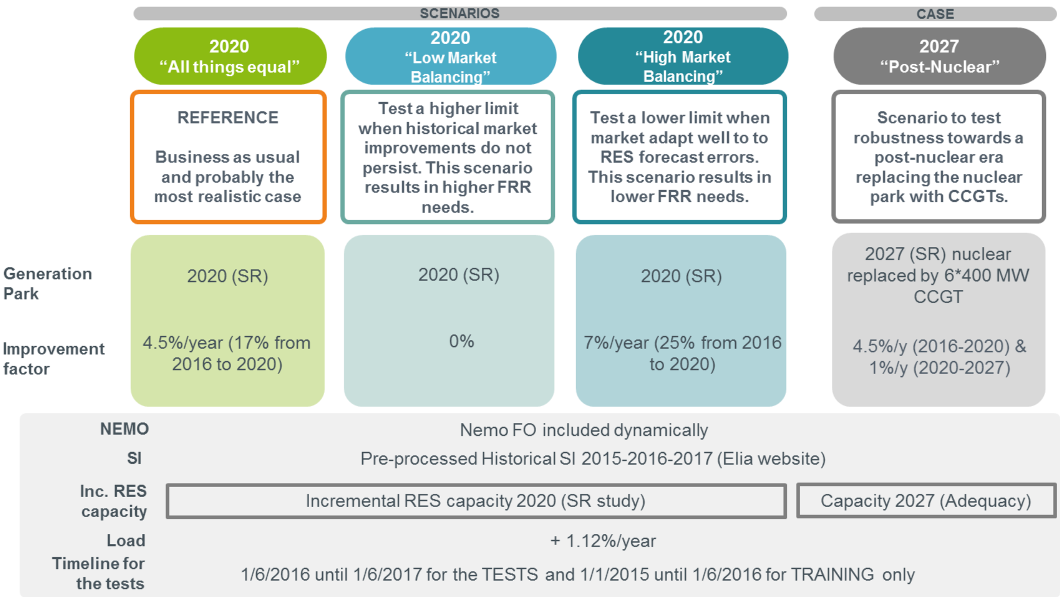


Figure 22: Overview of scenarios and case study investigated in the Proof of Concept

However, the Proof of Concept aims to investigate the dynamic sizing methods in 2020, including a dynamic behavior following the outage risk of the NEMO-link, and facing higher shares of renewable energy sources, with corresponding forecast errors which have a substantial impact on the FRR needs. As explained in Box 3, this is taken into account by means of extrapolating the forecast errors on the incremental capacity of onshore, offshore and photovoltaics between these years, and 2020.

It is concluded that this extrapolation is likely overestimating the FRR needs due to improving forecasting tools, and market mechanisms which allow market parties to absorb part of this need. As explained in Box 10, an analysis of historic data shows how the expected system imbalance can be assumed to improve with 4.5% per year the sizing is conducted closer to real-time. Therefore, a correction factor is taken into account in the reference scenario of the Proof of Concept: unlike the reference case of Part 1 "Reference 2020", the reference case of the Proof of Concept ("all things equal") assumes the 4.5%/y as it is the more likely improvement foreseen.

As implementing such a correction factor is likely to be more realistic than applying no correction factor, a reference scenario **"All Things Equal"** assumes an identical correction factor as observed in the past (2012-2016). It represents a business as usual scenario, assuming that the historic ability of the market to cope with imbalances will be reproduced in the next years (2016-2020). Although this is probably the most realistic scenario, two additional scenarios are considered:

- A **"Low Performant Balancing Market"** in which the correction factor is assumed to be 0% per year. This corresponds to a scenario where the historic ability of the market to cope with imbalances was rather exceptional, e.g. following market design improvements, and is not likely to be reproduced in the coming years. This scenario results in higher FRR needs, providing an upper limit of the dynamic sizing potential.
- A **"High Performant Balancing Market"** in which the correction factor is assumed to be 7% per year (the highest annual improvement factor observed over the period 2012-2016). This corresponds to a scenario where the historic ability of the market to cope with imbalances is related to the incremental renewable capacity. As this capacity has grown between 2012 and 2016 with 1723 MW, and is expected to grow with 3176 MW between 2016 and 2020 (Table 19). This scenario results in lower FRR needs, providing a lower limit of the dynamic sizing potential.

Table 19. Historic and projected installed capacity [MW] of renewable energy sources in Belgium. (a. Adequacy Study 2016; b. Elia Estimations; c. Adequacy and Flexibility Study 2016).

[MW]	2012[a]	2016[a]	2020[b]	2027[c]
Onshore wind	1005	1696	2431	3542
Offshore wind	380	713	2310	2312
Photovoltaics	2501	3200	4044	4988

It is explained in Part 1 that when implementing a dynamic sizing method, such corrections following assumptions on the impact on incremental capacity becomes obsolete. Besides these correction factors, these scenarios are based on the same data and assumptions taken in the "Reference Case 2020" of Part 1 (Section 4.2). **It is to be stressed that the cut-off risk of offshore wind power is not treated as an outage risk, for which the impact is already investigated as a specific sensitivity in Part 1.**

6.3.2. Case study 2027: post-nuclear

Similar to the first part of the study, one additional case study is dedicated to a nuclear phase out. This case analyzes the behavior of a dynamic sizing method after the nuclear units are phased out, and partially replaced by six 400 MW CCGT units. Objective is to verify the effect of replacing large 1 GW units by smaller 400 MW units, and not to make an appraisal on the required capacities or technologies to replace the nuclear units.

As explained in the previous part of the study, this has a large impact on the outage risk, reducing the upward FRR needs while increasing the dynamic sizing potential. In addition, the incremental renewable capacity towards 2027 is included in the simulation. Therefore, a correction factor of 4.5%/y is considered too optimistic, also taking into account that the incremental capacity (projections in the adequacy and flexibility study published by Elia in 2016) foresee no new additional offshore wind power capacity after 2020). It is therefore assumed that this correction factor is reduced to 1%/year between 2020 and 2027.

6.4. Real-life constraints: temporal aspects

The key feature of the Proof of Concept is to include a full year simulation, taking into account temporal aspects. While in Part 1 of the study, training, testing and validation are conducted on a random selection of time periods in 2016, the Proof of Concept implements a realistic order of the different steps using a rolling training, testing over 2015, 2016 and 2017. This allows taking into account more practical considerations, including (1) the training frequency which sets the frequency with which the algorithms are re-trained with updated historical data and re-compute the optimal set of parameters; (2) the sizing resolution of the dynamic needs sizing which has relevance in the framework of the FRR products length and (3) the lead time determining how much time in advance the sizing is conducted. An overview of the assumptions made for each method is given in Table 20.

Table 20: Overview of assumptions taken considering the constraints in the Proof of Concept

	Resolution	Training Frequency	Lead Time
OO	4-hourly	Yearly	Day-Ahead
KMEANS		Monthly	
KNN		Daily	

6.4.1. Training frequency

In general, machine learning methods are used by means of three steps: (1) design the model and parametrize, (2) train the model and (3) apply the model to make the actual predictions.

The design phase is generally not done regularly as it mainly consists of choosing the overall approach and the suited algorithm. However, the parametrization of the model can be done as often as the training. This training frequency is therefore a key parameter: this is the frequency when the historical data used to feed the model will be updated as well as the parameters of the model re-optimized. Training and applying the model to make actual prediction (i.e. testing) are therefore done following a "rolling horizon" procedure, as illustrated at Figure 23.

In a monthly training (which is the reference used for the KMEANS), before conducting the sizing for every day in the month M, the model is trained using the data from M-17 until M-1, and taking into account the required extrapolations from 2016 to 2020 (Box 13). E.g. the model is trained with March 2015 until July 2016, after which this trained model is used to size the reserves every day-ahead in August 2016.

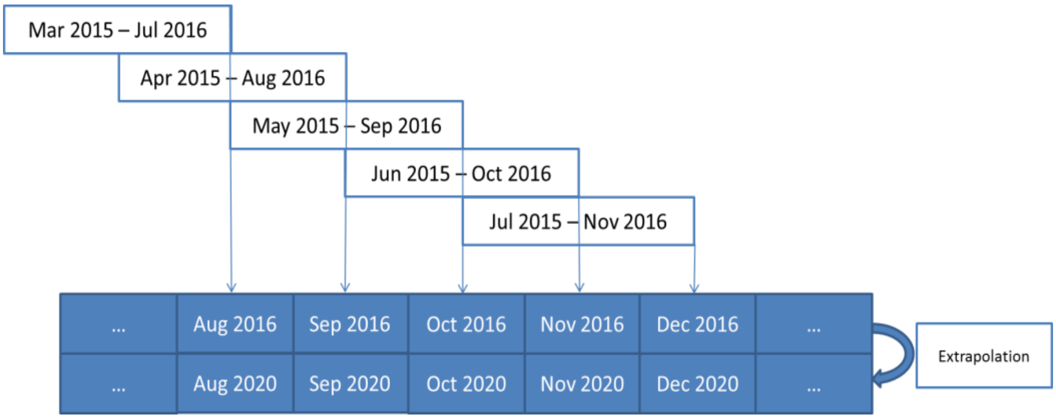


Figure 23: Visual representation of the "rolling horizon" training process of the machine learning methods

Box 13: Extrapolation of the data from 2016 and 2020

The month, e.g. June 2016, for which FRR needs are sized is extrapolated towards 2020 by adapting the forecast errors of 2016 with the incremental capacity (taking into account the correction factor) towards that month (linear increase of incremental capacity between begin 2016/2020 and end 2016/2020). Furthermore, also the demand and generation fleet are updated to represent that year 2020. The training data is also adapted to represent the installed renewable capacity of the same month June 2020.

It is stressed again that these aspects of extrapolations are only needed in this study because of the 4-year horizon to assess the potential of dynamic sizing in 2020. This would become irrelevant for the operation of a dynamic sizing of FRR needs on daily basis. Analysis of the results observed that 2015 is a year with a bias towards particular high positive forecast errors (representing a shortage), compared to downward forecast errors, which has a strong impact on the 'static' FRR needs when extrapolating towards 2020. In contrast, 2016 is a year with a bias towards particular high negative forecast errors (representing an excess), compared to upward forecast errors, impacting the 'static' sizing results for 2020. As the static sizing has to be used as a benchmark, providing a point of comparison for dynamic sizing, the high sensitivity of the training year on the static sizing presents the risk to distort the comparison.

The impact is mitigated by implementing a symmetrisation under assumption that these biases are an anomaly and not repeated next year (which is also confirmed). Therefore, in the scope of this report, a basic model has been established, in order to provide a representative benchmark for dynamic sizing, in which the forecast errors, which are used to extrapolate the past system imbalance towards 2020, are assumed to follow a symmetric distribution. Indeed, errors are mathematically often represented as symmetrical variable.

Mathematically, when using past data to evaluate a certain quantity, it is a good practise to define a mathematical model which is then calibrated based on historical data rather than using raw historical data. E.g. instead of using a histogram (i.e. distribution of the raw data), the targeted variable is firstly defined as a Gaussian variable, the parameters of the Gaussian model being then estimated based on the data. This is why, for instance, in the dynamic sizing approach, the imbalance distributions are based on a 'kernel' model which is estimated with historical data and not on histograms.

The same principles are followed in a daily training process (KNN) for which the training is conducted from day to day, including the day-ahead (D-1) until M-17. In other words, results for August 2, 2020 are based on training for March 2, 2015 until August 1, 2016, and extrapolated accordingly. Note that the length of the training data, together with the training frequency is part of the design decisions of dynamic sizing. Ideally, a model is trained on as much data possible to improve the predictive capacity of the model. However, as data is also impacted by certain events and evolutions, not necessarily captured by the model (e.g. such as market design changes or new forecast tools) it is chosen to limit the training period on 1.5 year.

A monthly training frequency is selected for the quantitative clustering method as a reasonable trade-off between accuracy and required efforts. A daily training is selected for the continuous neighbors as more in line with the more continuous philosophy of the method. However, sensitivities with daily, monthly and yearly training are conducted on each method. It is to be noted that for the OO-method, this training frequency choice is less relevant, as the prediction risk is still determined by means of static method. Nevertheless, while the status of the NEMO-link is adapted on a daily basis, as well as announced planned maintenances, the generation fleet can be adapted when information becomes available, e.g. monthly basis.

Furthermore, it is stressed that although the methods are developed to take into account planned maintenances, such events are not simulated in the yearly run of the Proof of Concept. This is to avoid distortions in the results as these events are too rare to include them in the yearly analysis, and draw conclusions on the aggregated FRR needs savings. However, separate analyses with simulated events of multiple nuclear units and NEMO-link outages are conducted.

6.4.2. Sizing resolution

The data available are generally based on a 15-minute resolution. This means that the sizing can be as granular as on quarter-hourly basis. However, the objective is to assume a sizing resolution which is realistic and aligned with what would be a plausible product length. Therefore, the sizing resolution is set at 4 hours, in line with European evolutions towards product standardization, i.e. 6 timeframes of 4 hours. Sensitivities are conducted to assess what the impact is of having lower or higher resolution: for 2h, 4h, 8h and 24h.

6.4.3. Lead time

The lead time represents the time between the period for which the FRR needs are sized and the period the FRR needs the day-ahead market closes. Section 3 explains that a sizing before this day-ahead market closure is a pre-requisite to ensure availability of the FRR means to cover the needs. However, the longer the lead time, the more time available for system operator and the market to conduct the required procurement processes of the contracted FRR means.

It is required to find a trade-off between accuracy (prediction accuracy of system conditions tend to become less accurate with a longer lead time) and the required time to organize the tendering process. For this study, the FRR needs are assumed to be sized day-ahead (D-1), although a sensitivity is conducted for a D-3 lead time.

6.5. Prediction of the NEMO-link flow direction

In Part 1 of the study, the exact schedule of NEMO-link (import, export or no flow) was assumed to be known in D-1 at the time where the dynamic sizing computation is conducted. However, this is not a realistic assumption when sizing before day-ahead market results are known. Therefore, this part of the study addresses the predictability of the NEMO-link schedule and is based on a simple methodology for predicting in day-ahead the hourly flow direction of the NEMO-link. Furthermore, to assess the impact of this forecast of NEMO-link schedule on the dynamic sizing reliability, it is also important to simulate the exact schedule of NEMO-link in 2020. This is achieved through 3 steps:

1. Assess the estimated import/export ratio of NEMO-link in 2020
2. Determine an hourly real-time direction of flows of NEMO-link in 2020 that takes into account correlations with other system conditions
3. Determine the corresponding day-ahead (before market matching) forecast of the hourly direction of flows of NEMO-link in 2020

A. Estimated import / export ratio in 2020

The yearly import and export ratio of NEMO-link has a direct impact on the potential of dynamic sizing. It is expected that this ratio will be related with the electricity price difference between Belgium and the United Kingdom, being the two countries interconnected by NEMO-link. This “intuitiveness” is even enforced in the day-ahead market operation assuming that electricity always flows from a country with a lower price to a country with a higher price.

An analysis of the historical import/export ratio of BritNed (the HVDC link between the Netherlands and the UK) and IFA (the HVDC link between France and the UK) confirms that day-ahead price differences are an accurate indicator providing a correct indication on the flow direction in a line. In cases where this deviates, this is due to a reversal of the flow following market evolutions during intra-day. Analysis has shown that this only occurs when price difference is small.

Based on the historical day-ahead price differences between both countries (2015-2017), an import/export ratio of NEMO of 15/85% is put forward for 2020. Such a ratio representing a dominant export status is in line with the current observations of BritNed and IFA. Furthermore, the conventional generation fleets for UK-BE-FR-NL are not expected to change fundamentally, while the shares of RES are expected to increase everywhere in similar proportion. Therefore, this ratio is assumed to remain stable until 2020. Towards 2027, more substantial changes could occur and modify the situation such as the closure of the Belgian nuclear fleet and the high share of RES. For this reason, an increasing ratio import/export of 30/70% is assumed.

B. Hourly real-time flow directions in 2020

To determine the hourly time series defining the hourly scheduled import/export directions in 2020, the hourly time series of the price differences between the two countries is used as a proxy.

Analysis shows that the historical dispatch direction of IFA and BritNed are 95% consistent with the time series of day-ahead price differences between these countries. Furthermore, the remaining 5% corresponds to hours when the price differences between the countries are low and when the intra-day or real-time modifications are observed to impact the dispatch compared to the one determined in the day-ahead market.

The Proof of Concept incorporates the prices of the period 2015-2017 to determine the expected real-time NEMO-link flow directions:

(1) if the day-ahead market prices are lower than UK prices, NEMO-link is assumed to export;
(2) if Belgian day-ahead market prices exceed UK prices, NEMO-link is assumed to import.

In addition, 5% of the hours in which the price difference is small are randomly selected and assumed to show an inverse schedule compared to the price difference.

C. Hourly day-ahead prediction of flow directions in 2020

As the price difference between Belgium and UK has been proven to be a good indicator of the scheduled direction of the NEMO-link, a day-ahead forecast of the schedule can be based on day-ahead electricity price forecasts for these two countries (before market closure). Such commercial forecasts of day-ahead prices are provided by different solutions providers, based on statistical tools and machine learning tools[8]. Based on historical day-ahead price forecast, the price difference between both countries is determined:

(1) if the UK price forecast is higher than the Belgian price forecast by at least €7, then NEMO-link is forecasted to be in export.
(2) if the Belgian forecasted price is higher than the UK forecasted by at least €7, then NEMO-link is forecasted to be in import.

Analysis have shown that if the difference between these two forecasts is smaller than €7, the risk of a forecast error cannot be neglected, and it is decided to relate these situations with an "indecisive" NEMO-link forecast for which both import and export is covered. This threshold of €7 is calculated a trade-off between forecast reliability and FRR needs reduction, represented by the ratio inaccurate forecasts versus informative forecasts (Figure 24). A higher threshold results in a decrease of the risk

[8] Note that day-ahead forecasts have been considered for the thereafter analysis as day-ahead is the main lead time considered in the Dynamic Reserve Study. However, the methodology proposed in this document is completely compatible with other lead times (D-3, D-7…) as forecasts of Belgian and UK prices are available for D-1 to D-7. Naturally, going to longer lead times is expected to reduce the accuracy of the predictions.

of forecast errors (i.e. forecast NEMO-link in import or exports while it is measured in the other state), but the marginal effect is gradually reduced with an increasing threshold. In contrast, a higher threshold reduces also the time an informative forecast occurs, i.e. a forecast in import or export, able to reduce the FRR needs. The ratio of these two curves allow to find an optimum on 7€/MWh, allowing to reduce FRR needs while keeping the impact on reliability low (maintained at 99.9% reliability).

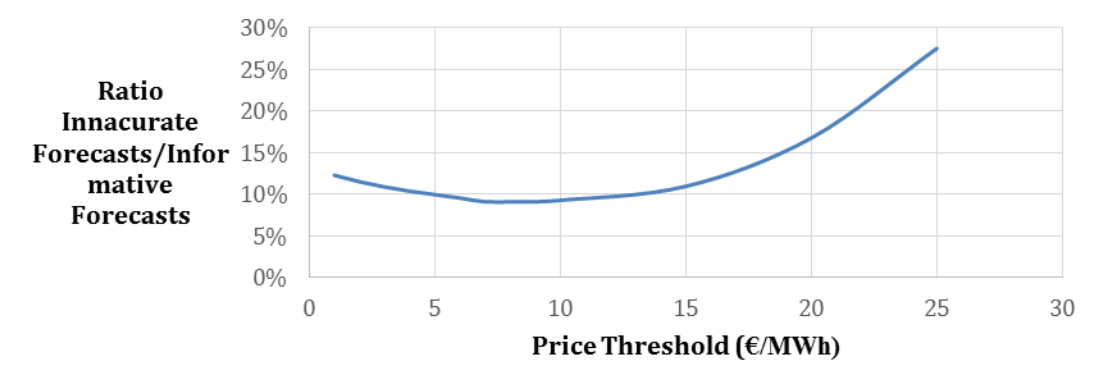


Figure 24: Trade-off between inaccurate forecast (e.g. predicted in import while it is in export) versus informative forecasts (i.e. predicted with state import or export)

Table 21 represents a matrix based on a prediction tool of NEMO-link for the training and test period, following the prediction tool and the 7 €/MWh forecast. The results represent the time NEMO-link is observed and predicted in a certain state (based on observations over 2015 to 2017). NEMO-link is expected to be predicted 64% of the time in export, which means the method will cover the export in these periods. On the other hand, the method will only predict NEMO-link to be in import 8% of the time. Finally, 29% it is seen as undefined as no reliable prediction can be made and both import and export are covered. The percentages are shifted towards import for 2027.

Table 21: Matrix representing the occurrence of prediction and observed schedule of NEMO-link over 2015-2017.

		2020			2027		
	Forecast / Observed	**Export**	**Import**	**Total**	**Export**	**Import**	**Total**
FULL DATA (2015-2017)	**Export**	61%	3%	64%	44%	4%	48%
	Import	2%	6%	8%	2%	9%	11%
	Undefined	22%	7%	29%	25%	16%	40%
	Total	85%	15%	100%	71%	29%	100%
TEST DATA (June 2016 – June 2017)	**Export**	52%	2%	54%	34%	4%	38%
	Import	3%	11%	14%	3%	15%	18%
	Undefined	24%	8%	32%	25%	19%	44%
	Total	80%	20%	100%	62%	38%	100%

Table 21 also allows deducting the errors of the forecast, i.e. the moments when NEMO-link is forecasted in import while it is observed in export (2%) and inversely being forecasted in export while it is in import (3%), resulting in a total forecast error of 5% (which is relatively small). While this basic forecast tool provides reasonable performance for the Proof of Concept, this prediction tool can be further investigated towards implementation.

7 RESULTS OF THE PROOF OF CONCEPT

The results of the Proof of Concept confirm the results obtained in Part 1, i.e. that a dynamic sizing method can facilitate 'stable' reliability by means of adapting the FRR needs to the risk of the system and obtain a constant reliability of 99.9%, as well during high as low risk periods. In contrast to the analyses of Part 1, the objective of this chapter is to focus more on the behavior of the FRR needs in a realistic context for 2020. A full profile for the reference scenario is given in Figure A and Figure B in Appendix.

Table 22 provides a general overview of the average, minimum and maximum FRR needs for the simulations for an entire year 2020 and 2027. These KPIs allow determining the dynamic potential (average FRR needs reduction compared to the static sizing). A few disclaimers are important: firstly, the simulations provide a best-estimate of the dynamic FRR needs and profiles in 2020 and 2027, and will therefore be different from the real dynamic FRR needs observed. Projections towards the incremental renewable capacity and its impact on the system imbalance are subject to uncertainty which increases with the time horizon. Secondly, the results do not represent the FRR volumes which will be contracted as part of the FRR needs (part of the FRR needs may be covered with non-contracted FRR needs, and reserve sharing agreements) while considering product availabilities. Thirdly, the results are obtained by means of prototypes which are likely to perform better when optimized upon industrialization and the operational experience).

Table 22: General results of the Proof of Concept simulations (FRR needs, expressed in MW) and dynamic potential (FRR needs reduction compared to a static sizing)

	[MW]		Low Market Balancing	Reference Case	High Market Balancing	Post Nuclear
UP	**STATIC**		1564	1417	1364	1284
	OO	**Average**	1546	1387	1355	1253
		Max	1565	1418	1382	1284
		Min	1532	1364	1335	1205
	KMEANS	**Average**	1477	1364	1325	1186
		Max	1968	1616	1472	1534
		Min	1327	1270	1243	905
	KNN	**Average**	1436	1353	1318	1162
		Max	1977	1616	1491	1532
		Min	1245	1208	1190	793
DOWN	**STATIC**		1426	1251	1180	1340
	OO	**Average**	1420	1237	1160	1327
		Max	1427	1252	1181	1340
		Min	1377	1140	1027	1272
	KMEANS	**Average**	1357	1205	1142	1286
		Max	1745	1589	1443	1700
		Min	960	795	715	866
	KNN	**Average**	1362	1205	1145	1286
		Max	2031	1693	1527	1778
		Min	844	698	628	841
UP potential	**OO**		19	31	26	31
	KMEANS		87	53	39	98
	KNN		128	64	46	122
DOWN potential	**OO**		7	15	21	13
	KMEANS		69	46	38	54
	KNN		64	46	35	54

Furthermore, extra-ordinary system or network conditions (e.g. during a solar eclipse) may not be covered by the dynamic dimensioning and may require temporal additional FRR needs in practice. Although the main trends in the previous chapter are confirmed, the results are now impacted by constraints which are implemented in the Proof of Concept, i.e. training frequency, resolution and lead time, as well as an imperfect prediction of NEMO-link.

7.1. Main results provided by the Reference Scenario

7.1.1. Improved reliability management

The three methods confirm that the average reliability remains at 99.9% (which is set as objective for the algorithm). Furthermore, the machine learning methods confirm the ability to increase the reliability in high risk periods, while avoiding oversized FRR needs in low risk periods. Figure 25 shows how KNN and KMEANS adapt the FRR needs to the risk, facilitating a better risk management.

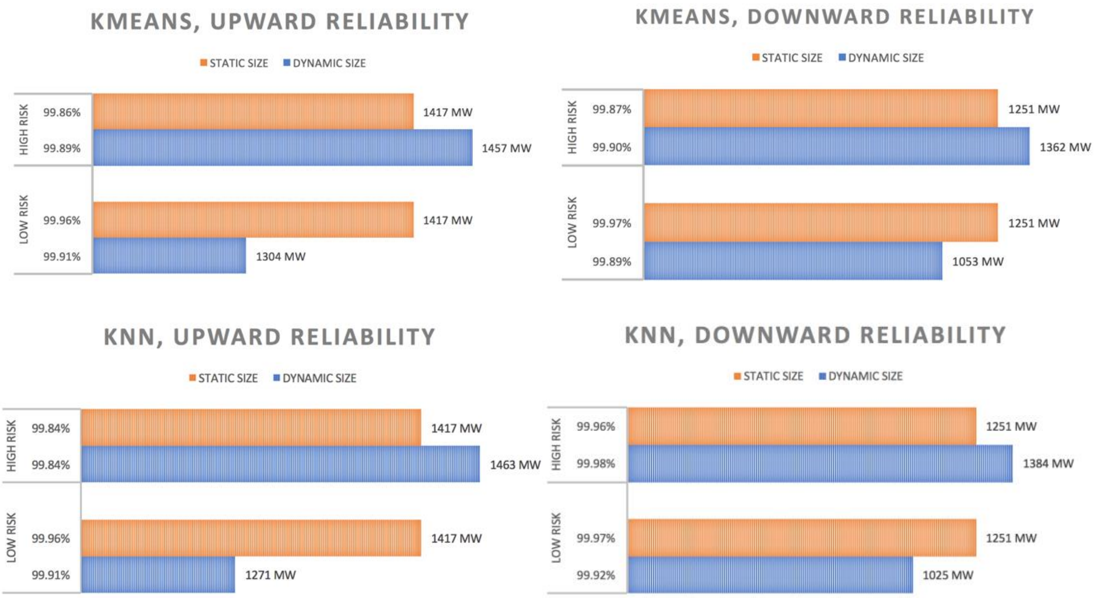


Figure 25: FRR needs [MW] and reliability [%] during high and low risk periods

This improved reliability management is not a characteristic of the OO. This method is an extension of the static sizing method, and cannot be used to detect high risk periods. In contrast, the OO method reduces up- or downward FRR needs during periods in which NEMO-link is predicted to be in export or import, respectively, or reduce the FRR needs when the 1 GW nuclear power plants are in planned maintenance. This avoids overestimating the outage risk, and therefore overestimating the FRR needs. Therefore, the OO does not increase reliability during high risk periods, but only maintains the same reliability levels as in the static sizing method.

The impact on reliability can be expressed in key performance indicators concerning the average loss and excess of energy during hours where the system imbalance is uncovered by the FRR needs (Table 23). It is shown that these values are equal or lower for the machine learning methods (particularly for the KMEANS), indicating a better reliability management, while the values of the OO are more in line with the static method. It is observed that for all methods, the effect on this average loss or excess energy is not large.

Table 23: Reliability indicators expressing the average loss and excess of energy during hours in which the system imbalance is not covered by the FRR needs.

	Loss of Energy per hour of failure [MWh/h]	Excess of Energy per hour of failure [MWh/h]
STATIC	142	112
OO	154	107
KMEANS	**132**	**90**
KNN	137	113

Figure 25 provides further insight by showing the percentiles of the loss and excess of energy. These figures show the 0%, 25%, 50%, 75%, 85% and 95% percentile when ranking the loss and excess energy observations from low to high. It shows that dynamic sizing has a reducing effect in the more extreme cases (e.g. periods representing the 75%, 85% and 95% percentile of the loss or excess of energy). Further analysis shows that while 46% of the static excess of energy is above 100 MW, this percentage is only 39% and 38% for respectively KMEANS and KNN.

Despite that these figures indicate a better reliability management, these improvements remain limited following the fact that dynamic sizing is not trained to reduce loss or excess of energy, but only to maintain a stable 99.9% reliability. This effect can therefore be seen as secondary to obtaining a 'stable' reliability at 99.9%.

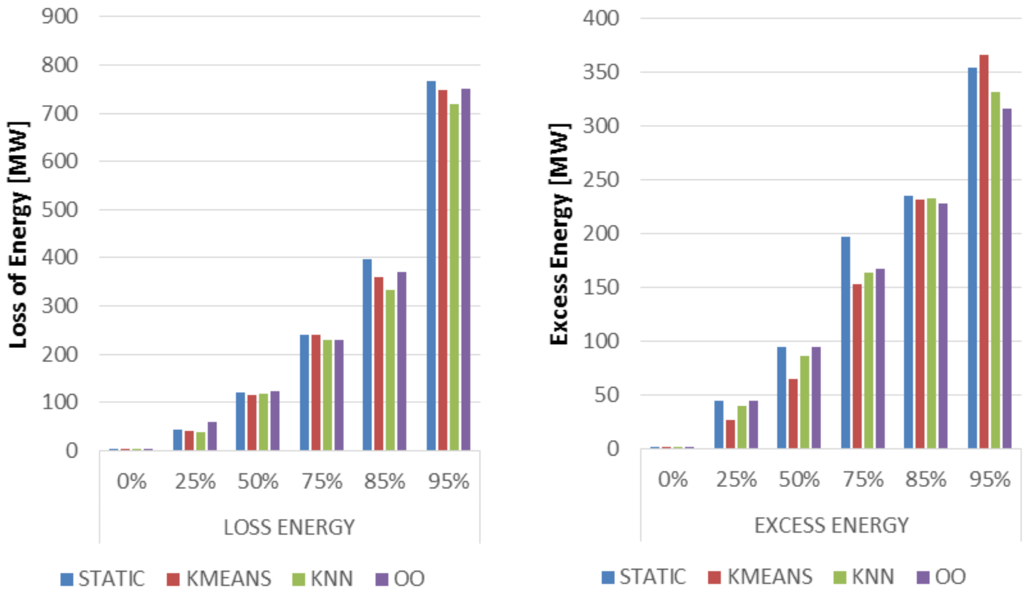


Figure 26: Percentiles of the Loss of energy [MW] and Excess of energy [MW].

7.1.2. Average FRR needs reductions

Table 24 shows the average, minimum, maximum, the dynamic potential (difference between static and average dynamic FRR needs) and dynamic spread (difference between the minimum and maximum FRR needs).

Table 24: Average, Minimum, Maximum FRR needs, Dynamic Potential (Δ) and Dynamic Spread (expressed in MW)

[MW]	Upward					Downward				
	Avg	**Max**	**Min**	**Δ**	**Spread**	**Avg**	**Max**	**Min**	**Δ**	**Spread**
STAT		1417		-	-		1251		-	-
OO	1387	1418	1364	31	54	1237	1252	1140	15	112
KMEANS	1364	1616	1270	53	346	1205	1589	795	46	794
KNN	1353	1616	1208	64	408	1205	1693	698	46	995

The static up- and downward FRR needs is respectively 1417 MW and 1251 MW. This is slightly lower as the reference case in Part 1 due to the correction factor (4.5%) which is reducing the expected system imbalance. The asymmetry between up- and downward FRR needs is explained by the forced outage risk which has an increasing impact on the upward FRR needs.

The dynamic potential is confirmed, with average dynamic FRR needs of 1353-1364 MW for the upward FRR needs, depending on the machine learning method, and 1205 MW on the downward side. The upward dynamic potential is found to be 54 - 64 MW for the machine learning methods, and 31 MW for the OO, while the downward potential is 46 MW for the machine learning methods and 15 MW for the OO. This reduction compared to Part 1 is explained by the realistic constraints following the temporality set in the Proof of Concept, and particularly by the limited predictability of the NEMO-schedule. The assumption that NEMO-link is predominantly scheduled in export has an important impact on the downward dynamic potential.

Moreover, the up- and downward dynamic spread remains stable compared to Part 1, i.e. up to 408 MW (upward) and 995 MW (downward) for the machine learning methods, for which the KNN is observed to show a larger spread as the KMEANS. This spread is lower in the OO, i.e. 54 MW and 112 MW for downward. Note that the higher spreads on the downward side result from the absence of an N-1 limit on 1000 MW when NEMO-link is predicted in import, and can result in large instantaneous reductions of the FRR needs. A higher downward range is explained as NEMO-link is sometimes forecasted in import which gives the freedom to the dynamic sizing to have volumes below 1000MW. This is in contrast to the upward side in which there is always a limit defined by the nuclear plants. It is assumed that the operational and regulatory framework allows FRR needs falling below the downward N-1 (1000 MW) if this unit is predicted to be in import with a predefined certainty level. This assumption is however subject to further clarifications during the implementation of the network guidelines on system operation requiring further investigating the effect of a change of flow direction between day-ahead and real-time.

The duration curves in Figure 27 represent the time a certain FRR needs is identified. The figure shows first of all that the two machine learning methods are very similar, and that most of the time the volumes remain between [1300; 1400] MW and [1100; 1300] MW for respectively up- and downward FRR needs. However, differences are observed in the extreme high or low FRR needs, in which the KMEANS results in higher FRR needs. This is the opposite in the downward FRR needs in which the KMEANS seems to result in lower FRR needs, at least at the upward side.

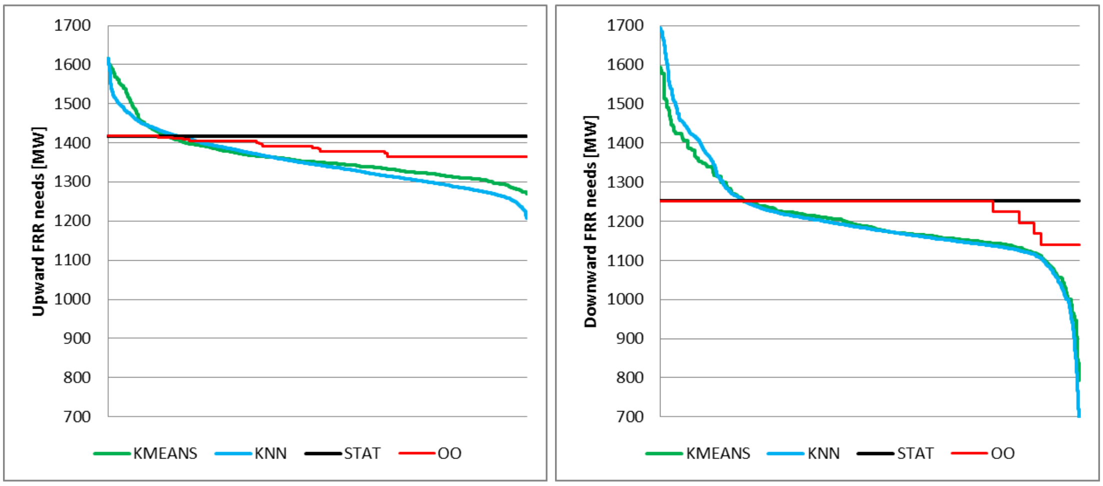


Figure 27: Duration curve of upward (left) and downward (right) FRR needs

The duration curve of the OO is obviously different, resulting in regular (82% of the time) but low (max 54 MW) reductions on the upward FRR needs, and high (max 154 MW) but rare (21% time) reductions in the downward FRR needs. This follows the prediction scheme of NEMO-link, subject to a predominantly export assumption towards 2020. It can also be seen that reductions below 1000 MW happen only a very limited time, i.e. 2.6% (KNN) and 2.3% (KMEANS) (only for downward FRR needs).

7.1.3. Robust FRR needs

While determining the training and test period lengths, it is confirmed that the static sizing is sensitive to the historical input data. As the static sizing computes the FRR needs based on the full year data, without taking into account the expected system conditions and corresponding imbalance risk, a few extreme system observations in the historic system imbalances or forecast errors of wind power or photovoltaics impact the FRR needs for the entire year. This is a general problem of static sizing methods which is observed by other TSOs as well.

Figure 28 (left) illustrates the effect of using several training and testing schemes on the original data: (A) train the sizing for 2020 based on a past year 2015 and then test on 2020 (based on 2016); (B) train the sizing of 2020 based on a past year 2016 and then test on 2020 (based on 2015); (C) train the sizing of 2020 based on a past year

2015 + first part of 2016 and then test on 2020 (based on second part of 2016 + first part of 2017). As explained in Section 6.3, it is observed that 2015 is a year where some extremely high positive forecast errors occurred, resulting in a high impact on the expected system imbalance when projected to 2020, while 2016 is a year where some extremely high negative forecast errors occurred. Therefore, the results of the static sizing (up- and downward), as shown in Figure 28 (left), vary depending on which training set is used (100 MW difference on the upward side and more than 150 MW difference on the downward side).

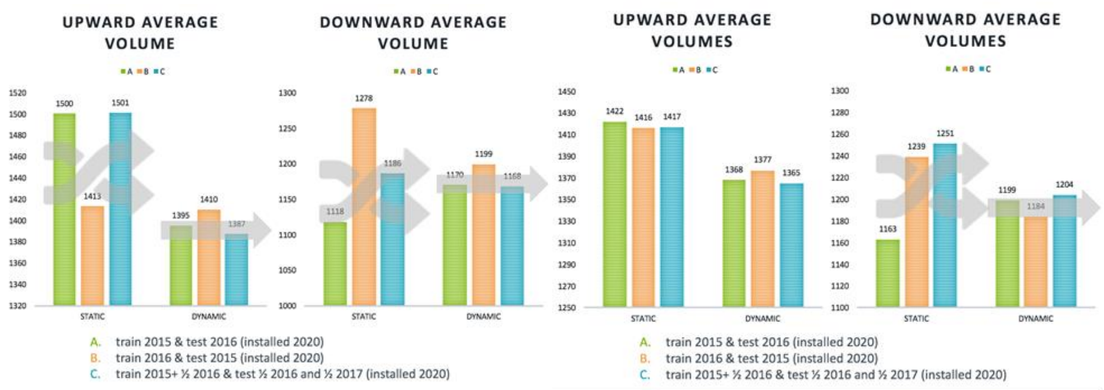


Figure 28: illustration of the average FRR needs over the year depending on the data set which has been used to train the method before (left) and after (right) pre-processing the data

As dynamic sizing computes the sizing based on the risk evaluated for the next day, it is less sensitive to the training set. Extreme measures impact the sizing only during similar system conditions and low system imbalances in the training set still result in higher FRR needs when recognizing a high risk. Indeed, in the example given, the very high system imbalance due to forecast errors will correspond to a "high risk" situation. This explains why the average FRR needs in Figure 28 for dynamic sizing remain stable, even after the data is treated to remove the asymmetry (Box 13). However, following to the processing of the input data, the static sizing instability on the upward side seems to be resorbed, while the downward instability remains. Figure 28 (right) shows that depending on the testing period, results of the downward static sizing vary between 1163 MW and 1251 MW.

It can be concluded that static sizing will become less suitable on the middle and long term. Indeed, it is expected that following the integration of renewable generation, rare but extreme events will occur which set the reserve requirements for an entire year in the static methodology, resulting in expensive oversizing.

7.2. Behavior of the FRR needs

The result of the Proof of Concept, i.e. a time series of hourly FRR needs, allows assessing the behavior over time for each method. Analysis of the monthly, daily and hourly variations allows investigating the volatility of the FRR needs, and is therefore an indicator for the robustness of the method (a high volatility FRR needs may result in difficulties towards explaining these variations and undermine confidence in the method. Furthermore, an analysis towards the relation between the FRR needs and the corresponding system conditions is conducted to understand the results, which is a prerequisite towards transparency.

7.2.1. Variations

Hourly variations

Figure 29 and Figure 30 depict the box plots representing the distribution of the 4-hourly variations (i.e. the absolute difference between two successive periods of 4 hours, as the FRR needs are kept constant within each period of 4 hours) of the FRR needs, as well as the maximum variation within a single day. These box plots clearly show that (1) the variations in the machine learning methods are substantially higher as in the OO; (2) the variations in KNN are slightly higher as in KMEANS, (3) and downward variations are higher as upward variations.

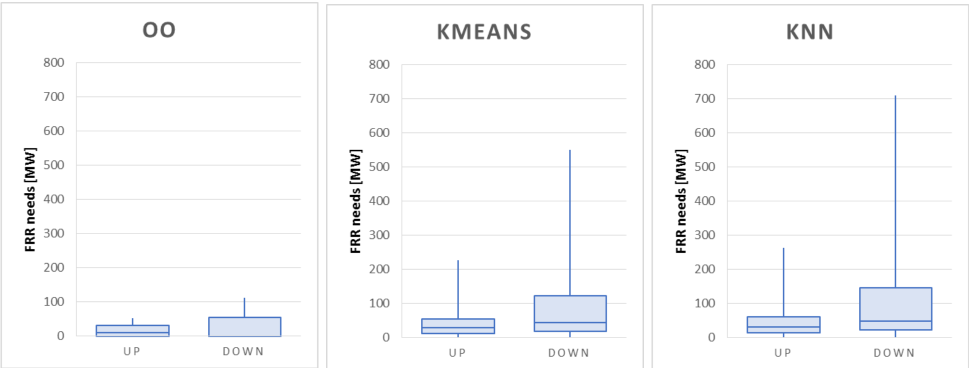


Figure 29: Box plot of the 4-hour FRR needs variations representing min, max and p25, p50 and p75 percentile

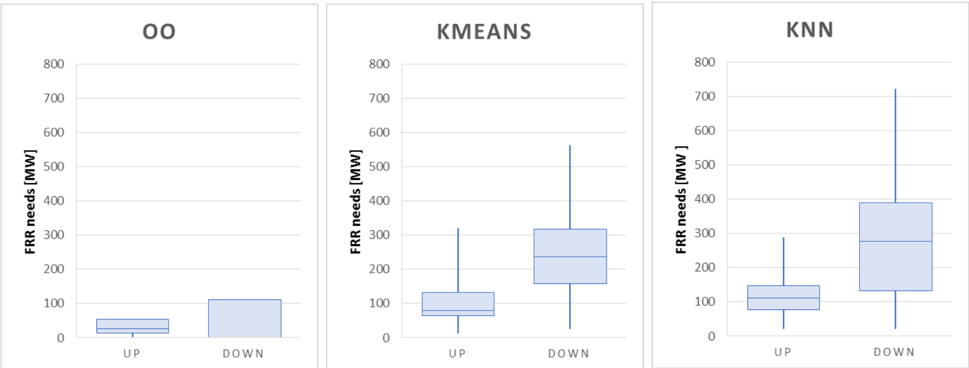


Figure 30: Box plot of the maximal FRR needs variations withing a day representing min, max and p25, p50 and p75 percentile

An analysis of the average 4-hourly and intra-day variations, and p90 percentile (Table 25) confirms that variations in the OO remain low, i.e. respectively 12 and 11 MW for average up- and downward FRR needs variations and 30 and 26 MW in the p90 percentile. However, the machine learning methods results in higher variations, i.e. 42 MW and 103 MW for respectively up- and downward FRR needs in the KNN method. Nevertheless, these variations can reach values above 118 MW and 288 MW, as shown by the p90 percentile. It appears that these values are slightly lower in the KMEANS method. The higher values for the downward FRR needs, compared to the upward FRR needs, are explained by the larger dynamic spread

(Section 7.1.2). It can be concluded that while the dynamic spread (maximum and minimum) is substantial, the FRR needs variations within a day remain lower.

Table 25: Average and p90 percentile of the 4-hourly variations and maximum variations within a day (intra-day)

[MW]	4-hourly variations				Intra-day variations			
	AVG		P90		AVG		P90	
	Up	Down	Up	Down	Up	Down	Up	Down
OO	12	11	30	26	30	26	53	112
KMEANS	39	87	108	238	108	238	219	404
KNN	42	103	118	288	118	288	190	550

Average daily profiles

Figure 31 shows the average FRR needs for each hour of the day (assuming a 4-hourly resolution) for the three methods. The daily profiles for the machine learning methods show that higher upward FRR needs occur during day-time and evening while showing lower upward FRR needs during night-time. This is explained by the demand which is lower during the night (giving more risk for a negative imbalance, i.e. excess), while the opposite is true for the day and evening (higher demand giving more risk for a positive imbalance, i.e. shortage). In addition, there is no uncertainty of photovoltaic generation during the night. When studying the downward FRR needs, the opposite behaviour is confirmed.

For the OO, this effect is less prominent, although downward FRR needs are notably reduced during periods in which NEMO-link is observed to have a higher probability to be in export, i.e. during the night. These observations are further discussed in Section 7.2.3 when assessing the relation of the FRR needs with the predicted system conditions.

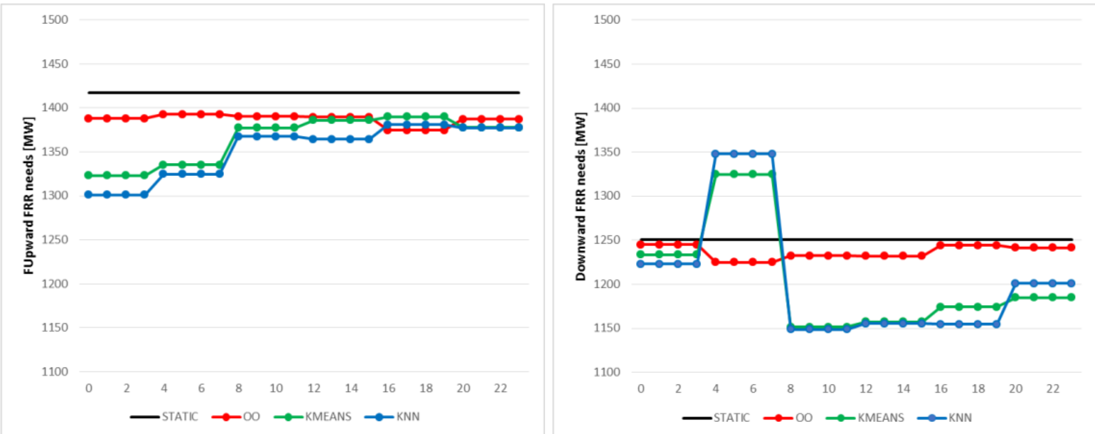


Figure 31: Average FRR needs per hour of the day

Average weekly profile

Figure 32 shows the average FRR needs for each day of the week are similar. Only, a small reduction of the upward FRR needs is observed during the weekend, due to lower demand. In contrast, a small increase of the downward needs is observed during the weekend. This is also the case for OO as the NEMO-link has a higher probability to be in export during low demand.

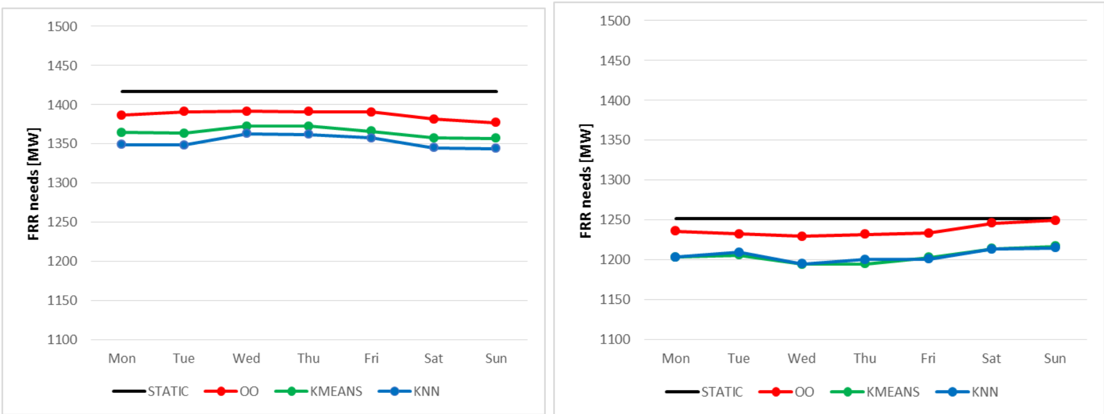


Figure 32: Average FRR needs per day of the week

Average monthly profile

Figure 33 shows the average FRR needs for each month for each method[9]. For the machine learning methods, the upward FRR needs faces two effects: (1) the increasing need along the year which mainly comes from the increasing renewable capacity installed along the year; (2) a higher reserve during the winter which is explained by the higher wind generation during the winter months, especially between January and March (see also Section 7.3.2). For the OO, only trained once, the FRR needs reductions during spring and summer is due to the higher probability that NEMO-link is scheduled in export.

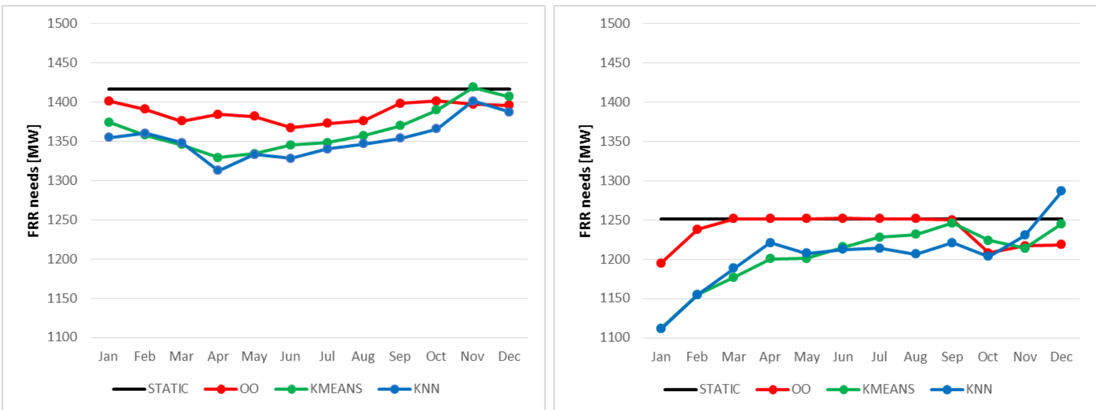


Figure 33: Average FRR needs per month

[9] The monthly trend needs to be interpreted with caution as it is based on a single year of analysis.

Regarding the downward FRR, the needs determined with the machine learning methods are linearly increasing along the year which is again explained by the increasing installed renewable capacity. The low values between January and March are also consistent with the higher wind generation during these months. The OO faces lower FRR needs reductions in the summer period following the higher probability that NEMO-link is scheduled in export.

7.2.2. Illustration of the FRR needs profiles

A. Illustration of Upward FRR needs

The dynamic FRR needs over a day or a week can be plotted in a dashboard together with the expected system conditions. Such dashboards are used to explain why the FRR needs during a certain period are low or high. Figure 34 provides an illustration for the week of September 24 until 30 (note that the timestamp is based on the corresponding week in 2016). In this example, most important drivers are depicted including the forecasted load, NEMO-link flow direction forecast and the total forecasted wind and photovoltaics.

Figure 34 shows a low FRR needs on September 24, Saturday between 5 and 9 AM, down to almost 1300 MW, close to the minimum FRR needs. This is a period with low renewable generation (limited generation of photovoltaics during the morning while facing low wind power generation). Furthermore, demand is low during the weekend and the import of NEMO-link is not to be covered as the interconnector is predicted to be in export (or at least part of that period).

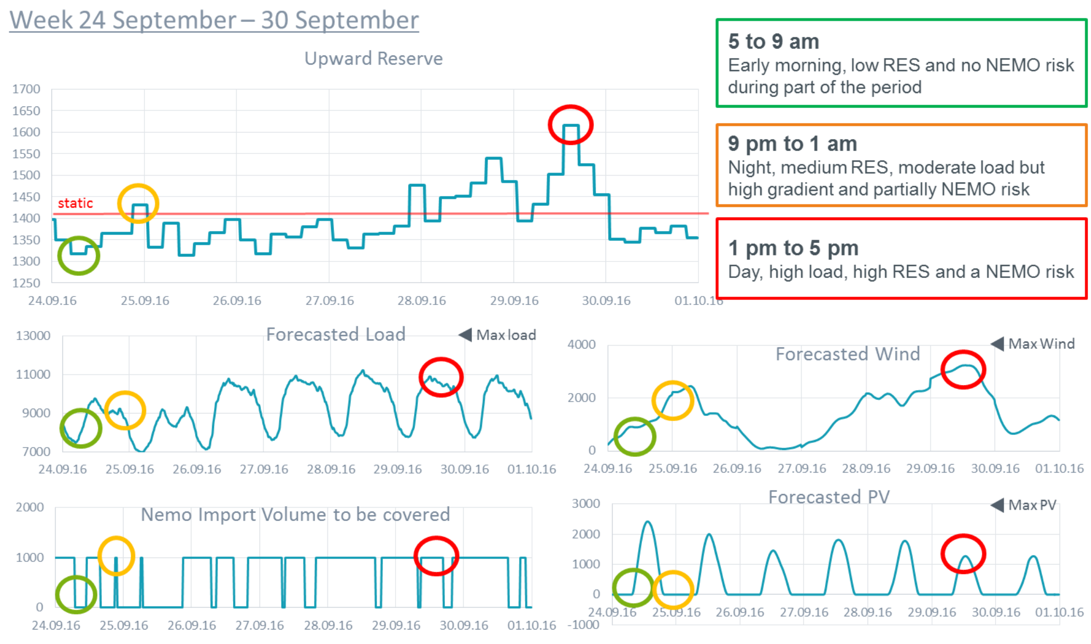


Figure 34: Illustration of upward FRR needs and corresponding predicted system conditions (KMEANS)

A higher FRR needs occurs during the same day between 9 PM and 1 AM around the static FRR needs, i.e. 1417 MW. This follows a moderate renewable generation prediction (relatively high wind generation and gradient while no photovoltaic generation. Simultaneously, the demand is rather moderate (day-time during the weekend) and the import of NEMO-link is to be covered as it is not predicted in export. However, an extreme high FRR needs occurs during September 29, Thursday between 1 to 5 PM, exceeding 1600 MW, close to the maximum FRR needs. This follows a very high renewable generation (wind power and photovoltaics), during the demand peak of the day, and the import of NEMO-link has to be covered as it is not predicted in export.

B. Illustration of downward FRR needs

Figure 36 illustrates the downward FRR needs during the week of October 15 – 21, determined with the KMEANS method. The same dashboard shows how a high downward FRR needs occurs during October 15, Saturday between 5 and 9 AM, up to almost 1500 MW and close to the maximum FRR needs. This is follows a very low prediction of renewable generation, together with the high load gradient during the morning peak. In addition, the export of NEMO-link has to be covered as it is not predicted in import.

The opposite occurs on October 19, Wednesday between 1 and 5 AM, where a downward FRR needs below 800 MW is observed, which is the minimum FRR needs. This follows a period with very high predicted wind power (nevertheless low photovoltaic power), together with a minimum demand level at night. Furthermore, the export of NEMO-link is not to be covered as it is predicted to be in import.

These illustrations show that FRR needs can vary from minimum to maximum in the course of a week but also that this can be explained by means of and analysis of the system conditions. These correlations are further investigated in Section 7.2.3.

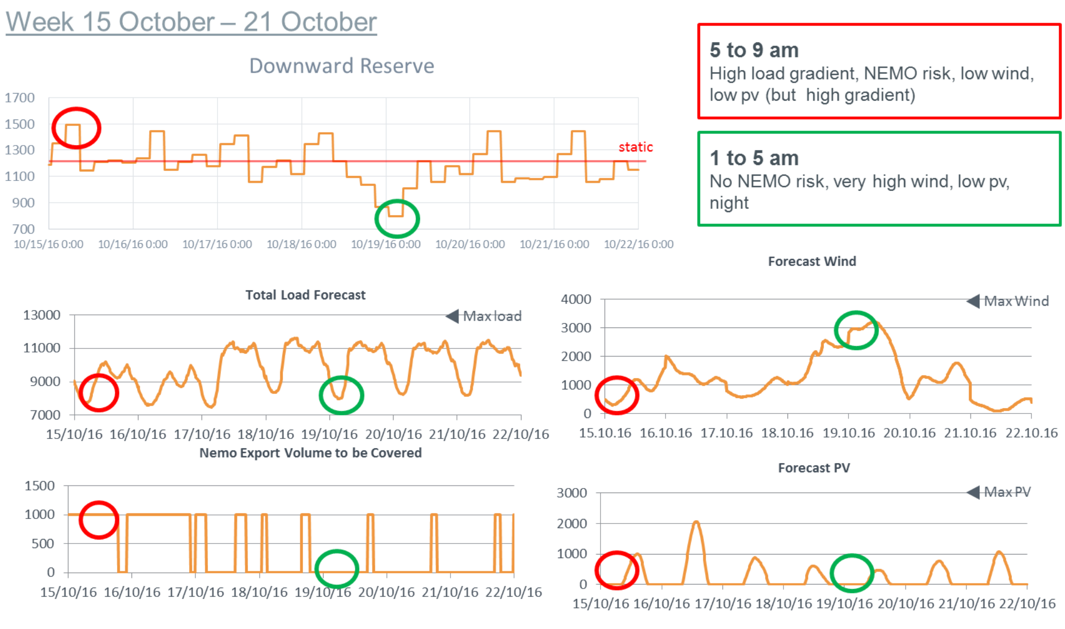


Figure 35: Illustration of downward FRR needs and corresponding predicted system conditions (KMEANS)

C. Illustration of FRR needs during planned maintenances

It is explained that the three methods allow taking into account planned maintenances by excluding these units when determining the outage risk. In the Proof of Concept, such events are not included as such rare events would distort the results if included on a one year analysis. To show the potential effect of such maintenance events, Figure 34 illustrates a week with extreme events: a planned outage of NEMO-link in the first two days of the week, while facing the planned outage of two nuclear units during the last two days.

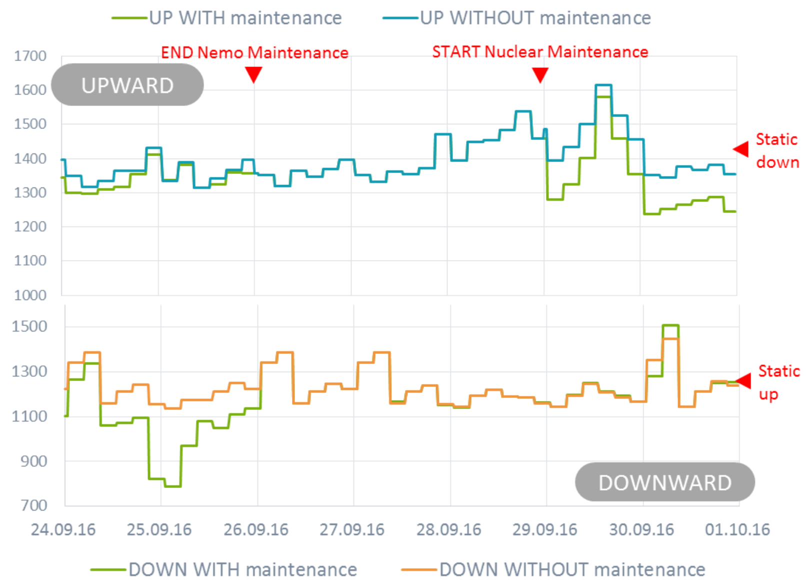


Figure 36: Illustration of upward and downward FRR needs during simulated planned outages (KMEANS)

It is shown that the FRR needs in the KMEANS are adapted accordingly, particularly reducing the downward FRR needs in case of the NEMO-link maintenance, and reducing the upward FRR needs in case of the nuclear power plant outages. Although that these results are not included in the Proof of Concept, it is clear that capturing such events which will results in additional benefits.

D. Illustration of FRR needs with different methods

Figure 37 represents the up- and downward FRR needs for the three methods during the week of September, 24 -30. First of all, it is confirmed that the OO shows limited variation compared to the static sizing. The upward FRR needs reductions, up to 54 MW results from periods where NEMO-link is scheduled in export. In contrast, the example shows rare and little reduction of the downward FRR needs, due to the fact that NEMO-link is only reduced part of the 4-hour period.

The FRR needs variations are larger in the machine learning methods and a comparison shows that deviations between the two machine learning methods can occur. This example shows how differences of 50 to 100 MW occur. This follows the differences in method, and their parametrization. Nevertheless, other illustrations show that both methods can also depict similar results (illustrated in Figure 38), even if they do not match exactly at each hour of the year.

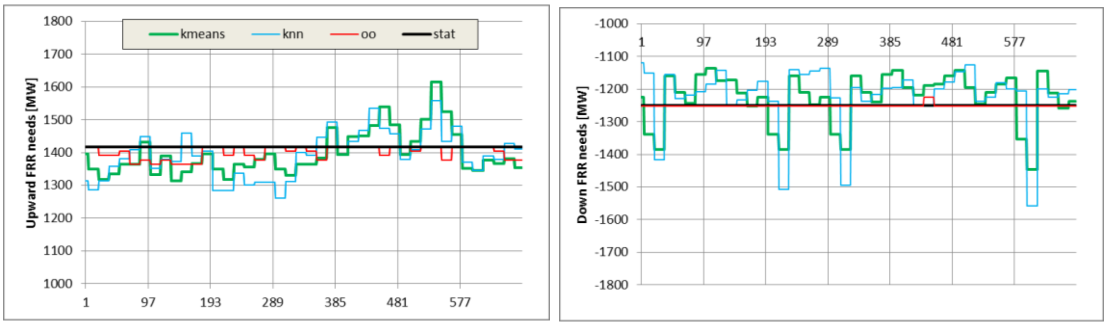


Figure 37: Illustration of the FRR needs in the three methods in the week of Saturday September 24 until Friday 30

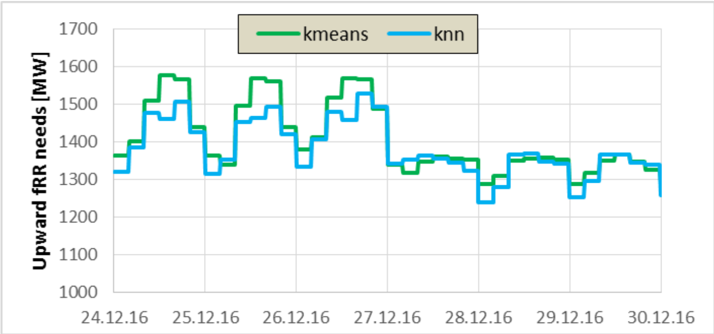


Figure 38: Comparison of the behavior of the machine learning approaches.

Analysis shows that most of the time, the FRR needs in the machine learning methods follow the same trends. Indeed, it is found that high FRR needs generally match even if not with the same magnitude. Further analysis by means of a scatter plot (Figure 39) highlights the correlation between both approaches: 0.79 on upward and 0.83 on downward which proofs that most of the time the sizing are consistent with each other. Furthermore, Figure 40 shows an histogram of the differences between KMEANS and KNN: more than 75% (60%) the difference between KNN and KMEANS upward (downward) FRR needs differs less than 50MW.

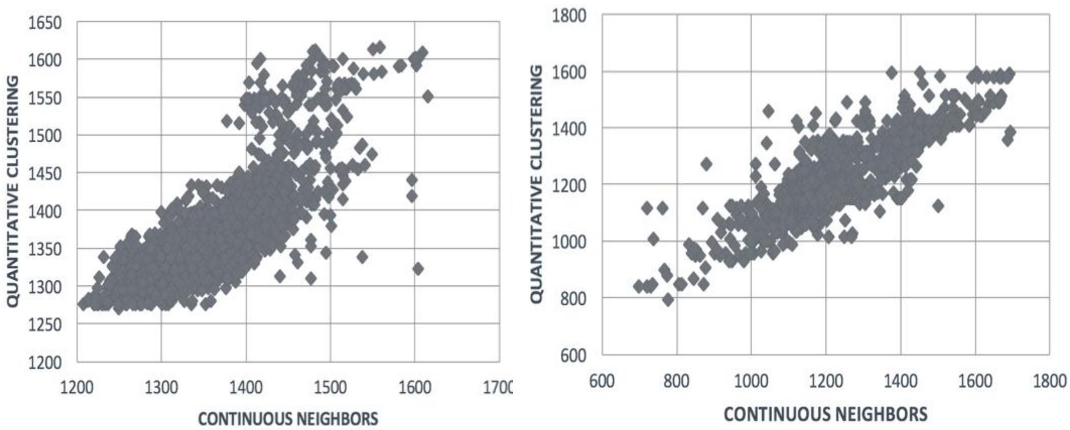


Figure 39: Up- (left) and downward (right) scatter plot of the FRR needs [MW] over the test period for KMEANS and KNN

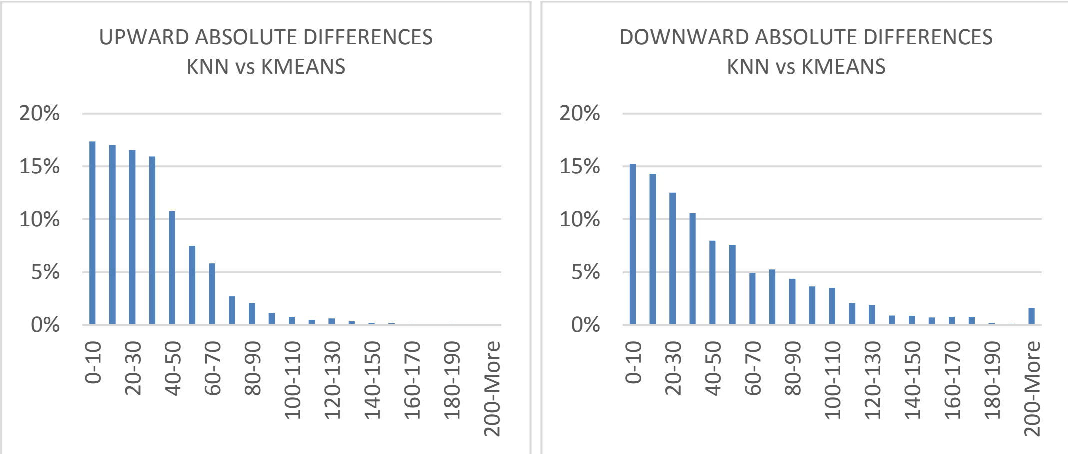


Figure 40: Histogram of the difference between FRR needs (expressed in MW) between KMEANS and KNN

It is observed that the KMEANS results in slightly lower downward FRR needs, while KNN results in slightly lower upward FRR needs. Furthermore, KNN seems more volatile FRR needs while the KMEANS results in more extreme FRR needs (higher maxima) following the use of discrete scenarios (clusters).

7.2.3. System conditions correlation

In the framework of transparency, it is important to be able to explain all results of each calculation of the FRR needs. Indeed, a black box model has to be avoided, and operators have to be able to interpret and explain the FRR needs at all time. Therefore, this section studies the relation with the system conditions.

Transparency can be ensured by means of studying the historic correlation between the historic FRR needs and system conditions. Figure 41 provides an overview of these correlations for the up- and downward FRR needs during the testing period for the KMEANS method (correlations are similar to KNN method).

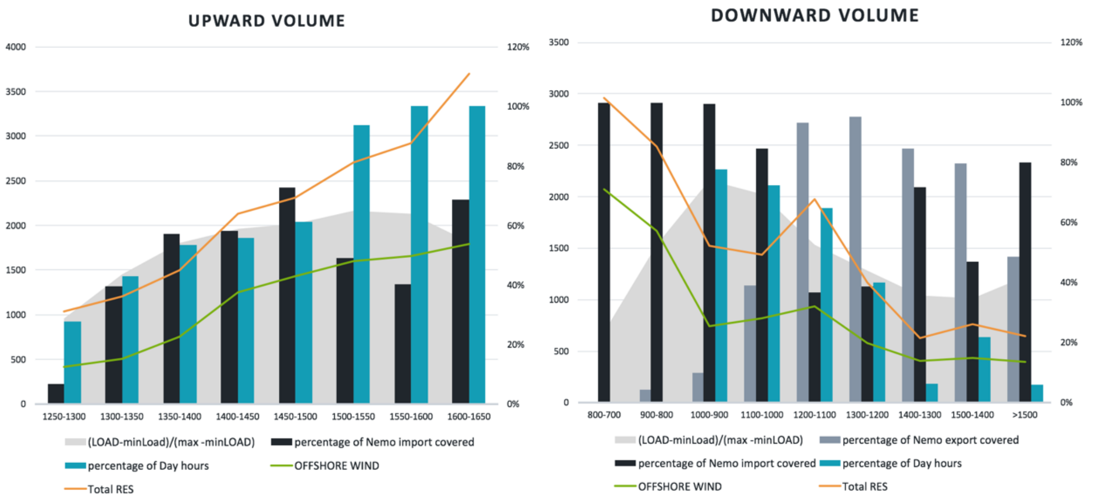


Figure 41: Graphical representation of the FRR needs (KMEANS). The x-axis represents the different categories of sizing, split in ranges of 50 MW for upward and 100 MW on downward, in increasing order. The status of NEMO-link, the percentage of day hours as well as the normalized demand are expressed in %, rated on the right y-axis. The renewable generation, as well as the offshore generation is rated in MW on the left y-axis.

A. Renewable generation

When looking to the FRR needs, it can be seen that the results are strongly correlated with the expected renewable generation. Higher predicted renewable generation results in higher upward FRR needs, while downward FRR needs decrease. This trend is intuitive following the higher risk for an unexpected shortage, when predicting high renewable generation and vice versa for low renewable generation, which provides a higher risk for unexpected surplus. Further analysis demonstrates that this trend is mainly driven by offshore wind generation and in less extent to the solar generation.

B. Time of the day and demand

It is shown that the highest upward FRR needs occur during day time (expressed by the percentage of day-hours). This is related to the expected demand, for which a correlation is found (Section 7.2.1): a high demand generally results in a higher upward FRR needs and vice versa. This trend is in general similar for the downward FRR needs: a low demand results in higher downward FRR needs and vice versa. Furthermore, higher FRR needs are observed during night. Nevertheless, Figure 41 shows that also the very low FRR needs may concur with low demand periods. Furthermore, an analysis demonstrates that for downward FRR needs, the ramping rate of the demand increases the FRR needs while this is not the case for the upward FRR needs.

C. NEMO-link

It is observed that the minimum downward FRR needs only occur when NEMO-link is predicted to be in import (where the N-1 does not need to be covered and so the needs can decrease below 1000 MW). This is similar to the upward FRR needs showing minimum needs occur only when NEMO-link is predicted in export. However, for moderate to high up- and downward FRR needs, the impact of NEMO-link is less clear, although in general, high up / downward FRR needs occur when the import / export state of NEMO-link is to be covered.

7.3. Sensitivities on design aspects

7.3.1. Training frequency

In theory, both machine learning models allow a daily training at the same time of the daily sizing of the methods (or at least just before). While this would theoretically result in the highest accuracy, explained by taking into account latest data, installed capacities and market trends, this might also be seen as resourceful, as well in terms of computation (ensuring the accessibility of all reliable data very close to the actual sizing), as well as human resources (validations, monitoring and reporting), even when automating the training and sizing process as much as possible. Furthermore, such a daily sizing may provide limited benefits as installed generation capacity and market trends do not change from day-to-day. Finally, this may result in higher day-to-day variations of the FRR needs, while this reduced stability makes it more complicated to explain the results, certainly on such short notice.

In order to test the impact in the paragraph above, sensitivity is conducted where the monthly training in the KMEANS is compared with a yearly training, and the daily training in KNN is compared with a monthly and yearly sizing (Figure 42). Of course, this is mainly relevant for the machine learning methods as the OO only requires ‘ad hoc’ updates of the generation fleet (entry or exit conventional power plants). Note that the objective of the analysis is not to derive the optimum training frequency (which is an exercise to be done towards industrialization), but to identify what would be the implication of one or another training choice.

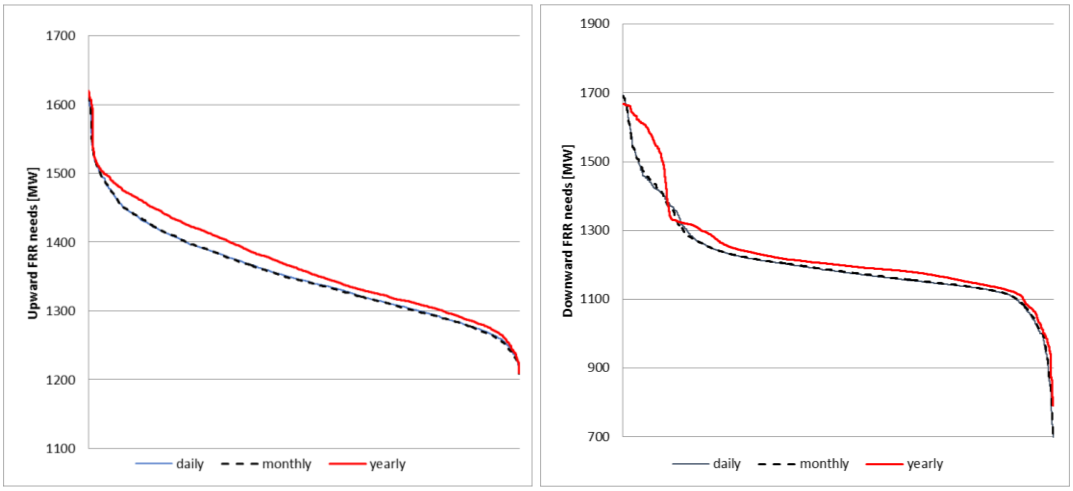


Figure 42: Duration curves of the FRR needs for different training frequencies

The load duration curves in the graph above show that increasing the training frequency from monthly to daily has only limited effect on the FRR needs while reducing the training frequency from monthy to yearly have a substantial increasing effect. It is explained by the fact that the installed capacity of renewable generation has a large effect on the result (Section 7.2.1) and this incremental capacity changes on a monthly basis. This effect is therefore not properly captured in a yearly sizing. Indeed, the yearly training is designed such that it uses the average installed RES capacity of the year when trained[10]. This will respectively result in over- or underestimations, or both which consequently also impacts reliability. Furthermore, a yearly sizing obviously does not allow taking the latest observations available (of the last days and months). Similar conclusions are obtained with the KMEANS where the dynamic potential is also reduced substantially when moving from monthly to yearly sizing. It is concluded that a monthly training frequency provides a good trade-off between accuracy and effort.

[10] Another choice could be to compute 12 sets of sizing, one for each month, each considering the installed capacity of the corresponding month.

7.3.2. Sizing resolution

All analyses in Part 2 are conducted with a resolution of 4 hours, based on trends concerning future product length (as it is expected that products will evolve to 6 blocks of 4 hours). Nevertheless, the more granular resolution that could be considered is determined by the resolution of the input data, which is on quarter-hourly basis.

It is expected that reducing the resolution would reduce the dynamic spread and reliability[11]. However, if the dynamic behavior was only seasonal, it would not matter to use 24-hourly sizing resolution rather than a 4-hourly. In contrast, as shown in Section 7.2.1, there are non-negligible variations and trends inside a day which are wished to be captured by a dynamic sizing (Figure 43 and Figure 44). Such intra-day trends are well captured by 4-hourly resolution (and also higher resolution) which allows them to capture the right risk all along the day, while an 8-hourly resolution roughly allows capturing the trend night/day for upward FRR needs. However, as the downward FRR needs typically show low reserve between 0 and 4 AM and high reserve between 4 and 8 AM. This trend can therefore not be captured with an 8-hourly sizing resolution.

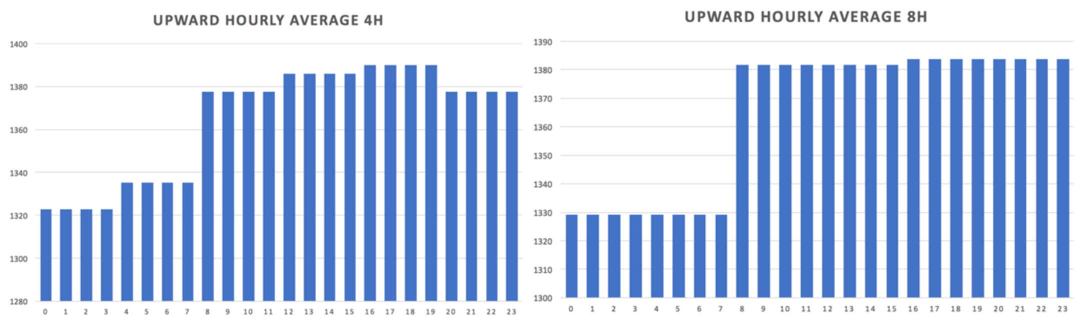


Figure 43:Hhourly average for upward reserve needs over the full period of tests, depending on the sizing resolution

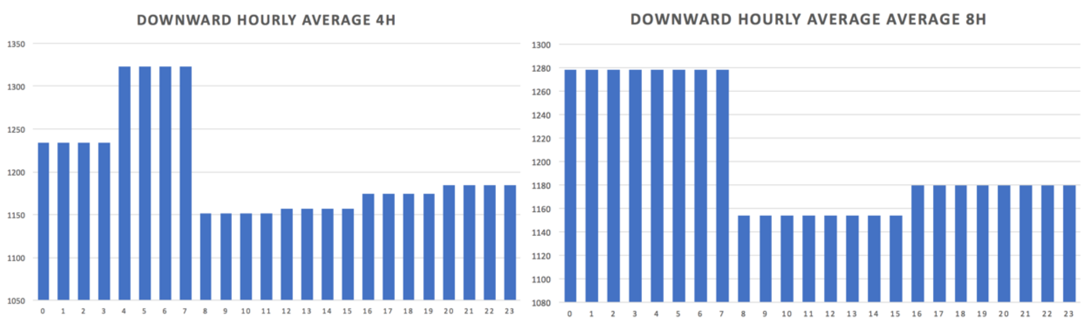


Figure 44: Hourly average for downward reserve needs of the full period of tests, depending on the sizing resolution

[11] The latter can be compensated by reducing FRR needs savings (setting the FRR needs in every period to its maximum FRR needs level)

It is concluded that resolutions up to 4 hours maintain a satisfying performance. However, as from 8 hours, dynamic spread becomes reduced, as well as the ability of the model as well as the FRR needs during high risk periods. This is also demonstrated in Figure 45 showing large reductions of the spread when approaching a 24-hourly resolution.

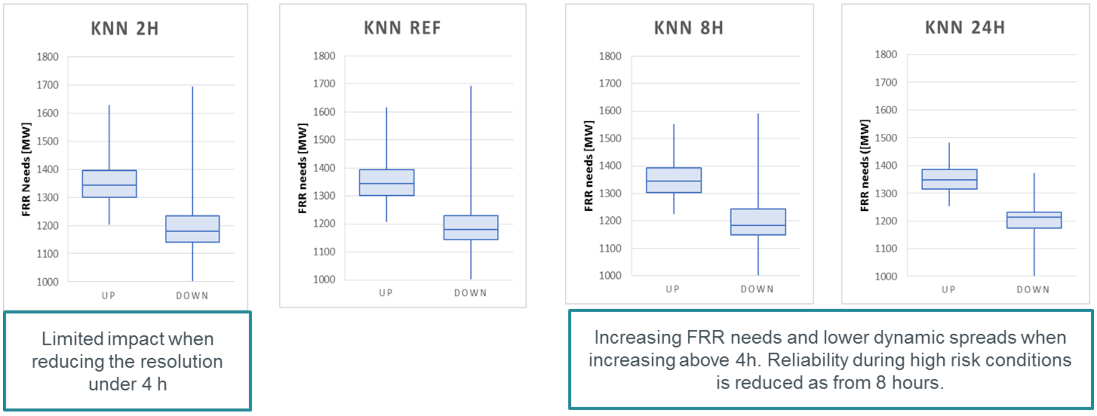


Figure 45: Box plot representing different resolution times

7.3.3. Lead time

All analyses are conducted with the prediction data representing a day-ahead forecast, as published on the website of Elia. However, a dynamic procurement may require a longer lead time between procurement (before Day-Ahead market closure) and sizing of the FRR needs. This is needed to facilitate the procurement process.

To assess the impact of longer lead times, expected to reduce the quality of the forecasts, a worst case analysis is conducted with the D-3 forecasts. It is to be noted that not all D-3 data was available for a period long enough to make conclusive analysis. Furthermore, demand data is not available D-3, and a W-1 forecast has been taken. It is clear that alternative scenarios between D-1 and D-3 are possible, but are not investigated at this stage.

An analysis of the forecast quality (Figure 46), expressed by the Mean Absolute Error, shows a reduction of the forecast accuracy, and in particular for offshore wind and solar when going from D-1 to D-3 forecasts. It is to be noticed that the forecast accuracy of the demand, load, and NEMO-schedule direction remains relatively stable. It is not excluded that future prediction tool improvements can increase this accuracy, although this is uncertain.

When implementing the D-3 forecasts in the dynamic sizing, results show that the dynamic potential is deteriorated, but not entirely decimated. However, further assessment is required as the length of the time series were not adequate to conduct proper analysis on the reliability. However, it is expected that with the D-3 approach reliability problems may arise as can be seen in the illustration in Figure 47. Although the FRR needs are well correlated, it is seen that a prediction errors of wind results in a large FRR needs deviations. It is therefore concluded that moving towards a D-3 horizon would not be desirable, unless substantially improving forecast accuracy, and a D-1 is advised, probably using an earlier forecast on that day.

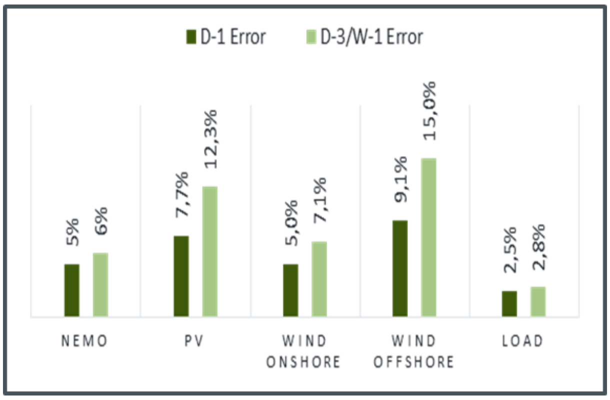


Figure 46: Analysis of the Mean Absolute Error (MAE, expressed as the average error in terms of the installed capacity) of the predictions of system condition

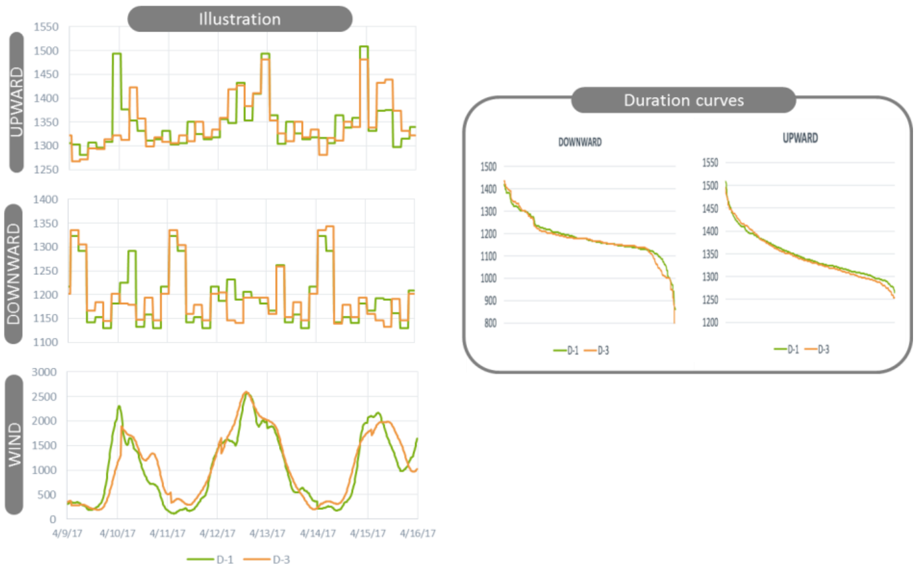


Figure 47: Duration curve (right) and Illustration of FRR needs during 2nd week of April 2020 for a D-1 and D-3 lead time

7.4. Sensitivities on scenario's and case study

7.4.1. Scenario with a low and high market balancing

A sensitivity analysis is conducted to account for a higher and lower ability of the market to deal with future system imbalances. The general conclusions are however not expected to alter from the previous analysis: a low reactive market is expected to result in higher FRR needs, and consequently a higher dynamic spread and potential. In contrast, a high reactive market is likely to result in lower FRR needs, and lower dynamic spreads and potential.

A. Low market balancing

Results in Table 26 confirm that FRR needs increase with less reactive markets. Indeed, the static FRR needs increase up to 1564 MW and 1426 MW for up- and downward FRR needs respectively. For the machine learning methods, the average dynamic FRR needs is also increasing but to lower extent: 1436 - 1477 MW for upward, and 1357 - 1362 MW for downward FRR needs (depending on the machine learning method). This results in an increasing dynamic potential of 87 - 128 MW for upward, and 64 - 69 MW for downward FRR needs. Obviously, also the spread increases, up to 641 – 732 MW for upward and 785 – 1187 MW for downward FRR needs. This is also observed in the duration curves of the three scenarios in Figure 48.

Table 26: Average, Minimum, Maximum FRR needs, Dynamic Potential (Δ) and Dynamic Spread (expressed in MW) for the low market balancing scenario

[MW]	Upward					Downward				
	Avg	**Max**	**Min**	**Δ**	**Spread**	**Avg**	**Max**	**Min**	**Δ**	**Spread**
STAT		1564		-	-		1426		-	-
OO	1546	1565	1532	19	33	1420	1472	1377	7	50
KMEANS	1477	1968	1327	87	641	1357	1745	960	69	785
KNN	1436	1977	1245	128	732	1362	2031	844	64	1187

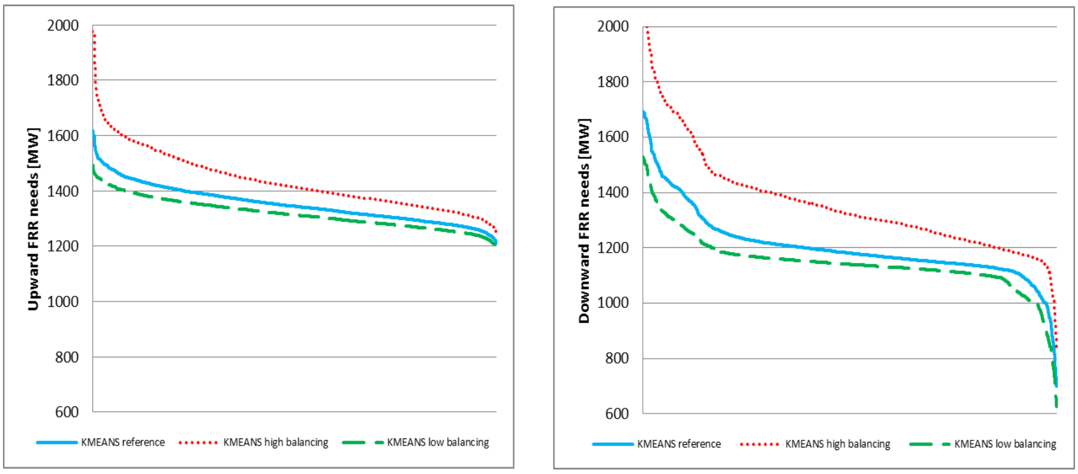


Figure 48: FRR needs duration curves for upward (left) and downward (right) FRR needs for the low market balancing scenario

In contrast to the machine learning methods, the dynamic potential is reduced in the OO compared to the reference scenario (Figure 49). This is due to the decreasing importance of the outage risk compared to the prediction risk following the correction factor.

B. *High market balancing*

Studying the FRR needs in a high reactive market allows investigating if a dynamic sizing method is still worth its merits in a case where the market successfully covers large part of the imbalances. Table 27 confirms that this is still the case with positive dynamic potential for upward of 39 - 46 MW and 35 - 38 MW for downward FRR needs. The reduction in potential for the machine learning methods is due to the reduction in average static FRR needs to 1364 MW and 1180 MW for respectively up- and downward, as well as dynamic FRR needs to 1318 - 1325 MW for upward and 1142 – 1145 MW for the downward FRR needs. Nevertheless, it is observed that the spread remains substantial 229-301 MW for upward and 728 – 899 MW for downward. This can also be observed in Figure 48.

In contrast to the machine learning methods, the dynamic potential of the OO increases compared to the reference scenario due to the increasing importance of the outage risk compared to the prediction risk following the correction factor (Figure 49).

Table 27: Average, Minimum, Maximum FRR needs, Dynamic Potential (Δ) and Dynamic Spread (expressed in MW) for the high market balancing scenario

[MW]	Upward					Downward				
	Avg	**Max**	**Min**	**Δ**	**Spread**	**Avg**	**Max**	**Min**	**Δ**	**Spread**
STAT		1364		-	-		1180		-	-
OO	1355	1382	1335	26	47	1160	1181	1027	21	154
KMEANS	1325	1472	1243	39	229	1142	1443	715	38	728
KNN	1318	1491	1190	46	301	1145	1527	628	35	899

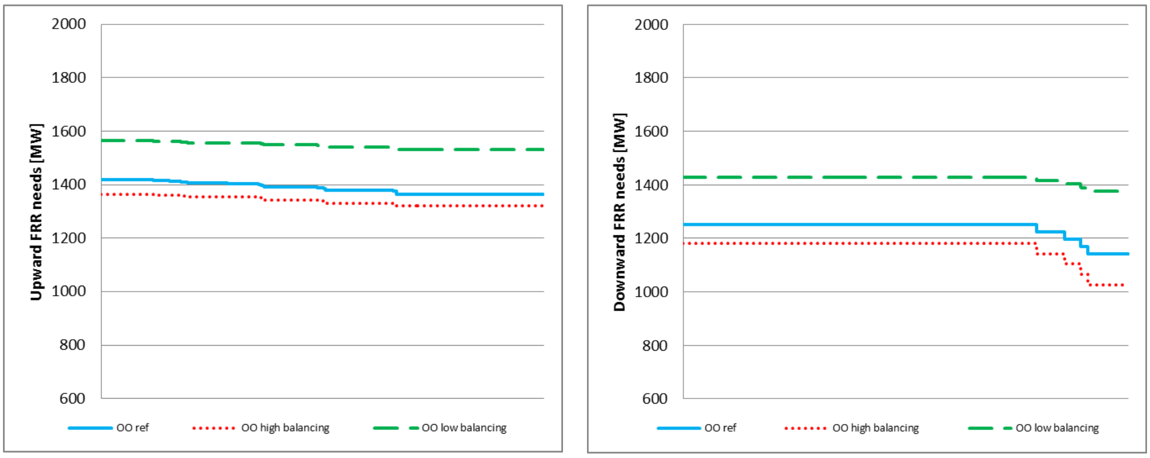


Figure 49: FRR needs duration curves for upward (left) and downward (right) FRR needs for the high market balancing scenario

C. General observations following the scenarios

For the machine learning methods, it can be concluded that the better the market players will balance prediction errors in their portfolio, the lower the FRR needs to be foreseen by the TSO. Despite that this reduces the dynamic potential, it remains substantial and one can confirm the potential of dynamic sizing in either scenario. This is an important conclusion as the goal remains to maximize the balancing actions in the market, and minimize FRR needs.

A particular conclusion is drawn on the OO method in which the dynamic potential increases with higher performance of the balancing market. Although this may seem counterintuitive, this is explained by the increasing importance of the outage risks which gains weight in the convolution of the outage risk and prediction risk the FRR needs. Nevertheless, this effect is expected to wear off with increasing renewable generation after 2020, increasing the importance of the prediction risk.

Furthermore, it is shown that the dynamic behavior of NEMO-link becomes more pronounced in the high market balancing scenario. This is confirmed in Figure 49. Indeed, further analysis shows an average difference in downward FRR needs of 100 MW between import and export periods, compared to 42 MW (reference) and 0 MW (low market balancing). This is due to the fact the downward N-1 is more often binding, providing in a positive effect when removed in import periods.

Regarding the correlations of the obtained dynamic sizing volumes with the predicted system conditions, they remain similar as in the reference, although more or less prominent in the low or high balancing market scenario respectively.

7.4.2. Post-nuclear 2027 case study

The analysis for a post-nuclear era confirms an increasing dynamic potential, even when taking into account further growth of the renewable energy sources for electricity towards 2027.

As already discussed in the previous part of the study, removing the larger 1GW nuclear units from the outage risk reduces the average upward FRR needs, as well with the static method, i.e. 1284 MW, as with the dynamic methods based on machine learning, i.e. 1162 -1186 MW (Table 28). Indeed, this reduction in capacity, even compensated by the additional prediction risk of additional renewable capacity towards 2027, follows the removal of the nuclear units as contingency, and therefore the N-1 criteria. This means the upward FRR needs would be able to drop below 1000 MW, at least when NEMO is predicted in export. It is observed in Figure 50 that this only happens during a very limited amount of the time, max 7% of the time. This will have an increasing effect on the dynamic spread increasing from 346- 408 MW to 629-739 MW, and also the dynamic potential of the upward FRR needs reductions of 53 – 64 MW to 98 – 112 MW.

Table 28: Average, Minimum, Maximum FRR needs, Dynamic Potential (Δ) and Dynamic Spread (expressed in MW) for 2027

[MW]	Upward					Downward				
	Avg	**Max**	**Min**	**Δ**	**Spread**	**Avg**	**Max**	**Min**	**Δ**	**Spread**
STAT		1284		-	-		1340		-	-
OO	**1253**	**1284**	**1205**	**31**	**79**	**1327**	**1340**	**1272**	**13**	**68**
KMEANS	1186	1534	905	98	629	1286	1700	866	54	834
KNN	1162	1532	793	122	739	1286	1778	841	54	937

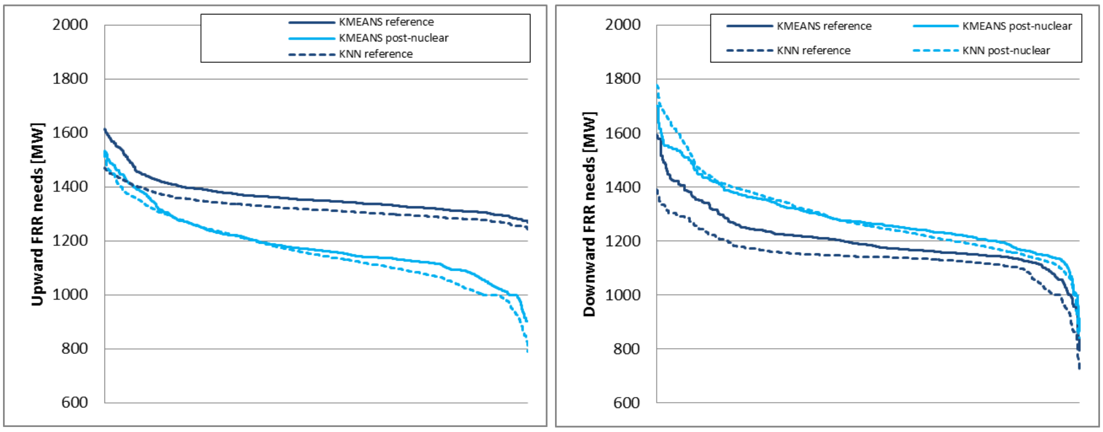


Figure 50: FRR needs duration curves for upward (left) and downward (right) FRR needs for 2027

This is not the case for the downward FRR needs where the outage risk is not impacted, and thus not compensating the prediction risk of the additional renewable generation. Therefore, the FRR needs increase towards 1340 MW in a static method, and 1286 MW in the dynamic method. It can be seen that although the dynamic potential and spread increases, this effect is fairly limited. Similar to the upward FRR needs, the downward FRR needs are rarely lower than 1000 MW (max 1% of time).

In conclusion, the dynamic sizing tool remains functional for a post-nuclear context, although the algorithms may require some further improvements by that time. The upward FRR needs potential is significantly increased, especially on the upward side. In contrast, results show that OO still provides some potential, although this is further reduced.

7.5. Analysis of the financial gains on FRR

In Part 1 of the study, a financial analysis of the FRR needs reduction is already conducted, finding a yearly benefit around 1.1 M€ and 2.7 M€ for the OO, and 2.6 and 7.9 M€ for the machine learning methods. These values are calculated by valorizing the average FRR gains observed in the simulations, i.e. the dynamic potential, for the reference and progressive scenario at a “generic” price for mFRR. It is explained how these prices are subject to high uncertainty, and therefore translated in three fixed price scenarios, resulting in the large margins in the financial gains.

The results the Proof of Concept are used to further refine these results. Firstly, the hourly FRR needs are now determined for an entire year in a more realistic context. Secondly, these FRR needs are now valorized taking into account a price elasticity. It is assumed that when FRR needs increase, the laws of demand and supply will determine a higher price per MW-hour.

The supply curve for two scenarios is shown in Figure 51. These FRR needs can be considered low, moderate or high and are based on the minimum, average and maximum FRR needs for the KMEANS method, respectively attached with the minimum, moderate and high price for the upward and downward FRR needs as defined in Section 5. Note that these figures remain subject to the disclaimers mentioned in the first part of the study. These figures allow to construct a piece-wise linear function representing the first scenario.

Nevertheless, as this scenario may underestimate the price increases with high FRR needs, a second more conservative scenario assumes a price which increases to 11.5 and 12 €/MW-hour for up and downward FRR needs respectively. It has to be seen as a worst case scenario, reducing the financial benefits of dynamic sizing.

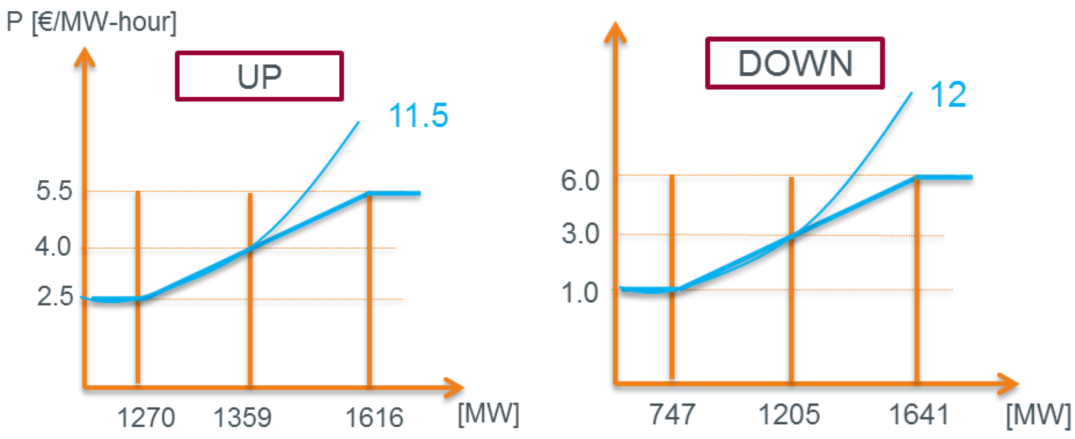


Figure 51: Representation of the generic mFRR price in function of the FRR needs [MW] for two scenarios in 2020

The financial gains are confirmed for the reference scenario. Note that the analysis is only conducted for the most realistic scenario, i.e. "All Things Equal". Considering the large uncertainty concerning the prices, it not deemed useful to conduct the analysis for each scenario. Results in Table 29 show that even in a price inelastic scenario the results remain largely positive, but with lower spread between scenarios compared to the analysis in Part 1. Results show that a yearly saving can be obtained between M€2.51 and M€2.97 for machine learning methods, while this is limited between M€1.48 and M€1.71 for the outage only method.

It is observed that the largest potential is found on the upward FRR needs, which is in line with the previous analysis. It is to be noted again that the results depend on the static sizing which is used as a reference under assumption that this method would be used as alternatively. It is explained however that this method depends strongly on the period covered, and outliers may strongly influence the results and is not suited as sizing methodology on the middle and long term.

Table 29: Financial gains of dynamic sizing methods compared to the static approach

[€]	Scenario 1: Price Elastic			Scenario 2: Price Inelastic		
	UP	DOWN	**TOTAL**	UP	DOWN	**TOTAL**
OO	1.104.388	379.735	1.484.123	1.308.069	400.458	1.708.527
KNN	2.122.126	845.976	2.968.102	2.219.783	343.086	2.562.869
KMEANS	1.765.411	1.003.736	2.769.146	1.785.307	725.788	2.511.096

7.6. Selection of algorithms

The general conclusion is that static sizing is not an enduring solution for the middle and long term from a technical point of view. Indeed, future systems are expected to be more prone to extreme system conditions, which will unnecessarily set high reserve needs for the entire upcoming period, or even harm reliability by not timely responding to new system evolutions.

A dynamic sizing method is therefore an evident solution as it resolves this problem, and additionally results in a better reliability management and obtains average FRR needs reductions. Two categories of dynamic sizing approaches are put forward (Figure 52) : on the one hand, the machine learning methods (Quantitative Clustering and Continuous Neighbors) that are fully dynamic methods, on the other hand, the outage only method that can be seen as an adaptation of the static method that only manages the outage risk dynamically. The machine learning methods provide the ability to determine the FRR needs matching the perceived imbalance risk with a 99.9% reliability level, while reducing the average FRR needs. Improved performance of the algorithms, as well as potential improvement of the market balancing renewable prediction errors, are expected to further reduce average FRR needs in the future.

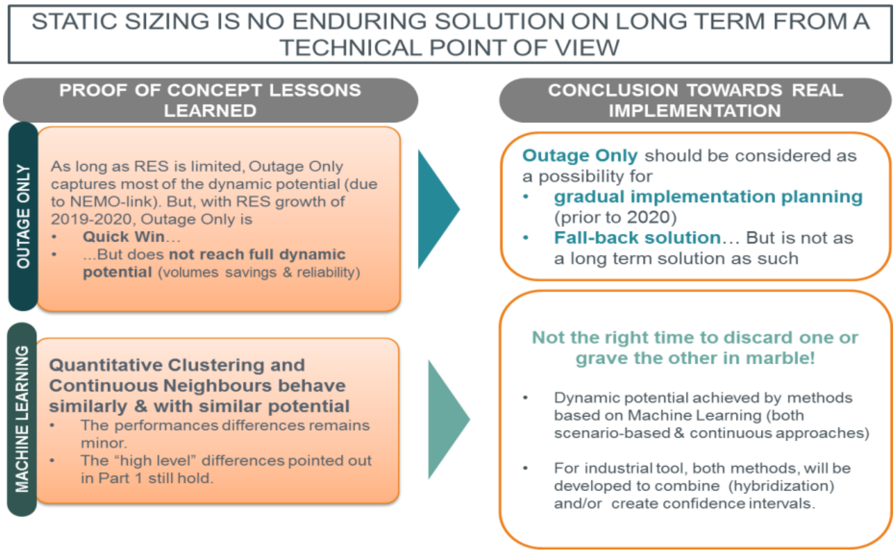


Figure 52; Representation of final selection of dynamic sizing categories

At this stage, two machine learning algorithms are investigated, which show similar general characteristics in terms of reliability and FRR needs reductions, but may also provide different FRR needs at certain moments. It is found difficult to assess which algorithm works best as their performance depends on the system conditions. In the implementation phase, algorithms are further improved on the specific problem, the methods will be re-developed towards a combination of both, or add an additional layer of artificial intelligence will be added to lead to a fully robust operational tool. Furthermore, it is not excluded in the future that other new types of algorithms may outperform the current ones, and call for replacements of the algorithms. It is to be stressed that this will only impact the mathematical aspects, and not the entire design of the dynamic sizing approach or tool.

The other category of Outage-Only can be seen as a heuristic method. Based on a static treatment of the prediction risk, it is clear that this method is unsuitable for a future with high renewable generation as the current static approach. Nevertheless, this method allows providing average FRR needs reductions in situation with lower outage risk while the same level of reliability than the static method. This volume reduction potential of the Outage-Only method will gradually disappear with increasing renewable generation and is therefore only considered useful before 2020. This as a quick win due to its low complexity and easy-implementation and can serve as a step in the implementation of the machine learning methods. Furthermore, the Outage-Only can keep its merits even after as a fallback method in the unlikely situations where machine learning calculation would not provide feasible results.

8 IMPLEMENTATION ASSESSMENT AND PLAN

Figure 53 depicts the different stages towards the industrialization of a dynamic sizing tool. Part 1 of the study is a **feasibility study** in which a general assessment is conducted on the potential of dynamic sizing. Six methods were developed, simulated and assessed towards technical criteria (reliability), economic criteria (reductions of the FRR needs), as well as the sensitivity towards different system evolutions (robustness). However, before moving towards implementation, the three methods are tested in a **Proof of Concept**. The methods are developed to "prototypes" and tested in a realistic environment for 2020. This includes a "virtual" parallel run in which the methods are run from day-to-day for an entire year, taking into account the practical implications of the methods concerning training frequency, sizing resolution and lead time towards day-ahead market closure.

Upon implementation, the selected algorithms will be subject to **industrialization** from "prototype" to a reliable and efficient tool integrated in a software platform ensuring automatic access to the required data, and the user interfaces to operate the method and validate the results. Implementation will be integrated in a wider mFRR Roadmap that will consider and prioritize in time all identified elements influencing the dimensioning methodology.

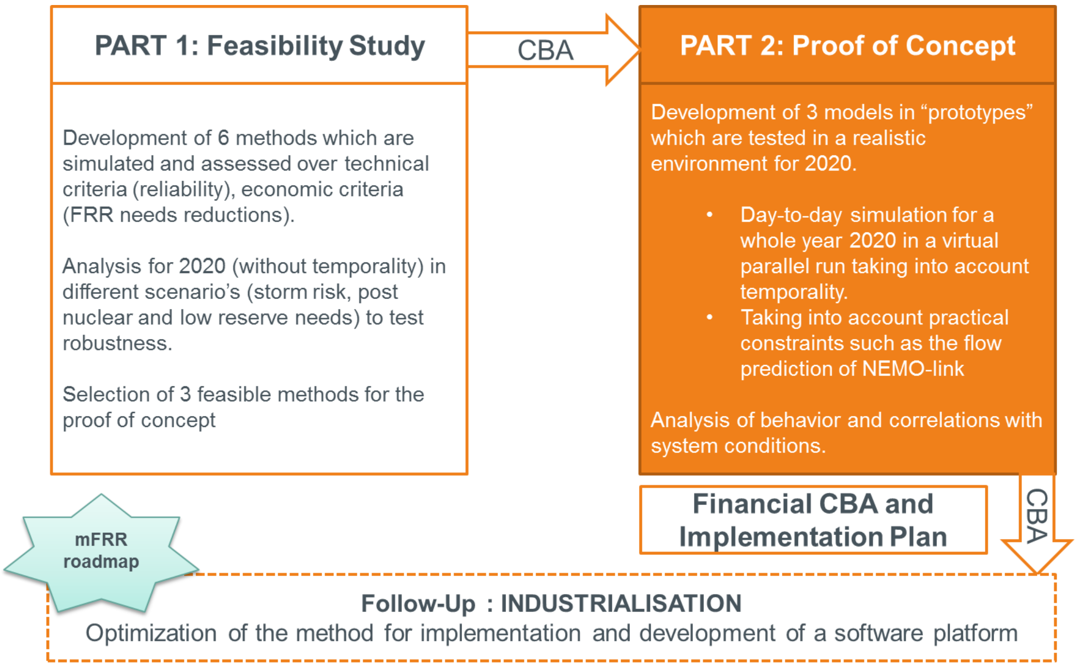


Figure 53: Schematic representation of the implementation trajectory of a dynamic sizing approach

The objective of the presented study is to demonstrate the feasibility of a dynamic sizing approach in the Proof of Concept. In order to achieve this objective, three methods with corresponding algorithms are developed as a "prototype". Nevertheless, implementing an operational tool for dynamic sizing requires further efforts, outside the scope of this study.

First of all, additional efforts are required in machine learning, developing the algorithms from “prototypes” to reliable and efficient algorithms. This implies further calibration of parametrization, but also finding an approach to use and combine the two proposed methods to provide a result which can be used for the procurement of the FRR needs. Secondly an integrated software platform is to be developed allowing accessing the required data, computing the FRR needs with minimum human interaction and depicting the results in a user friendly representation. Furthermore, this tool requires features towards testing (e.g. testing environments) and operation (e.g. fallback methods and potential capacity ‘adders’ to account for extra-ordinary system and network conditions such as for instance a solar eclipse), which are not yet present in the prototypes presented in the Proof of Concept.

The scope of this section is to describe the necessary step to evolve towards a tool having five key quality characteristics:

- **Performance:** the tool needs to ensure accurate results in terms of FRR needs and reliability management. The machine learning algorithms have to be tailored for an optimized performance on the specific problem.
- **Computational speed:** the algorithms and software platform as a whole has to facilitate fast computational times.
- **User friendliness:** the tool needs to be easy to use through automation where possible and flexible user interfaces that allows entering the desired inputs and to visualize the output in an optimal way.
- **Transparency:** the tool needs to allow detailed ex post analyses of the results, and even re-produce the results months after the calculation. It also has to track the performance by means of dedicated KPIs.
- **Robustness:** the tool has to be operationally reliable and secure while specific care is required to the development of routines and code assuring the robustness of the code. A fallback procedure has to be foreseen in case of major incident.

It is to be stressed that the development does not stop with the first implementation. System and market evolutions will require regular design modifications (e.g. new imbalance drivers), as well as updating the tool with evolutions in statistical methods and machine learning. Obtaining a performant tool will not be a “one shot” implementation.

8.1. Architecture

Figure 54 presents the software platform which will be composed of four user applications and three mathematical modules. Applications are the user-interfaces that present well-identified functionalities to the user which can be used to monitor and report on the result and performance of the tool. Modules are mathematical algorithms or data management routines that are using mathematical calculations to support business decisions.

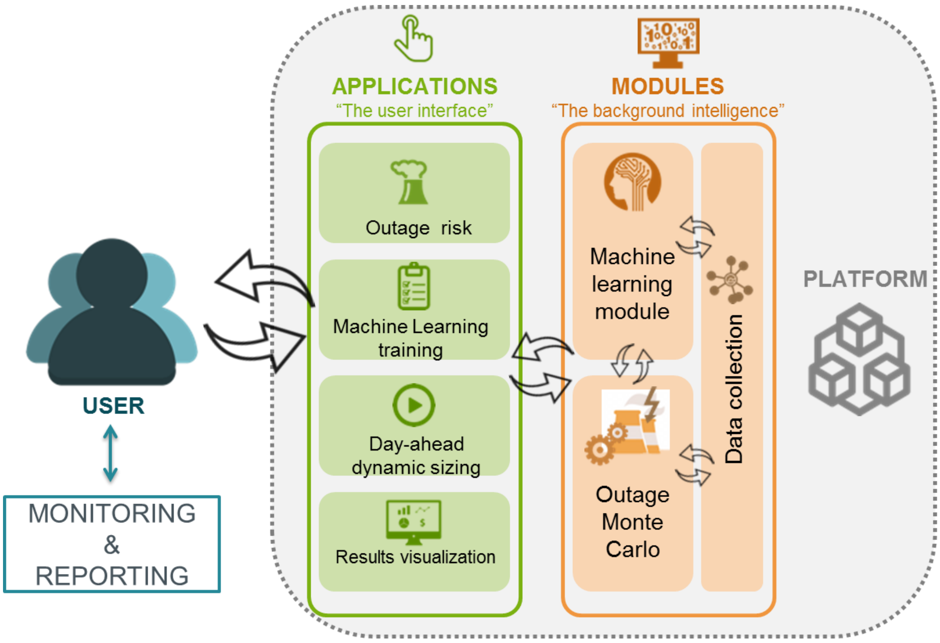


Figure 54: Visual representation of the architecture software platform for dynamic sizing of the FRR needs

A. The mathematical modules

Data collection module

This module automatically provides the data that is needed for the different applications and modules of the solution. The data requirements are different for the different modules.

The machine learning training requires a large amount of data that does not have to include the near past or the near future (day-ahead predicted system conditions and system imbalances). This data can be delivered via a "slow track", using a general database. The day ahead dynamic sizing and outage module on the other hand require a more limited but up-to-date information concerning the day-ahead predicted system conditions. This needs to be available via a "fast track", using a specific service hub. Attention has to be paid to the fact that several versions of data are available and the choice of version can impact the outcome of the result. In particular, the consistency of data sources between the "slow track" and the fast track" is a point of attention.

In order to ensure maximal transparency, it is important that the user can anytime reconstruct the outcome of the dynamic sizing using the correct source data.

Outage risk module

This module computes the distributions of outage risk in terms of MW for the next day (depending on the forecast NEMO-link dispatch and power plant maintenance schedules). This module will implement Monte-Carlo Simulation techniques and run on several hundreds of simulations in order to faithfully estimate the probabilities of outage for the next 24 hours.

Machine learning module

This module computes the dynamic distribution of the past system imbalance based on machine learning technics which allows to find patterns in the data and exploit them. The module includes the implementation of advanced machine learning algorithms. Building on the prototypes developed during the study, the algorithms have to be refined to meet the standards expected by an operational tool as to ensure a high of operational reliability.

B. The user interface applications

Outage risk application

The outage risk application enables the user to manage and operate the outage risk module. This application allows the user to:

- manage the generation fleet (e.g. power plants entering or exiting the market);
- visualize the planned maintenances of power plants (in first instance only the nuclear units) as well as NEMO-link, both obtained in an automated way;
- launch the outage risk module;
- visualize the output of the outage risk module.

This application includes the possibility to take into account planned maintenances of power plants (in first instance only the nuclear units), in an automated way, as well as the predicted schedule direction of the NEMO-link. The latter is set following the output of a NEMO-link forecast tool to be developed upon implementation. The application will launch the Monte-Carlo Simulation Module. This will provide the forecast of likelihood of global outage in MW across the entire fleet of power plants and relevant HVDC-links, and visualize the results concerning the outage risk.

Machine learning training application

The machine learning training application enables the user to manage and operate the machine learning module for training purposes. This application allows the user to:

- enter parameters that will influence the outcome of the training such as the time-window of data to use for the training, the target precision, the method to be used;
- launch the training;
- store multiple training result (based on a different set of parameters) in different ‘scenario’s’.

This application is only relevant for the machine learning methods for which each training session will generate a mathematical model that can be used to forecast the day-ahead dynamic volume needs. This application will allow entering the training parameters that will influence the training algorithm, such as the time-window of data to use for the training, the target precision, the algorithm to be used, the number of clusters or neighbors depending on the machine learning method. The application enables to store and load multiple pre-defined choices.

Day-ahead dynamic sizing application

The day-ahead dynamic sizing application enables the user to manage and operate the machine learning module for daily, operational purposes. This application allows the user to:

- identify which 'scenario' (both outage and machine learning) will be used for the day-ahead dynamic sizing;
- automatically launch the day-ahead dynamic sizing on a daily basis;
- store the output (i.e. a time series for the selected resolution) of each sizing, allowing to ex-post explanations of the FRR needs;
- to compare and assess the results of different outage and prediction risk distributions in a test environment, following different design choices and parametrizations.

This application will forecast the day-ahead dynamic FRR needs. The application computes this value based on the outage risk and prediction risk distributions of the scenarios and parameters selected in the two previous applications. The results are defined as a time series for the selected resolution. Note that the application may need to contain a fallback value for the prediction risk (e.g. a static value calculated over the representative period) in case the machine learning method is not able to generate a result.

This application will store the output of each sizing, allowing to ex-post explanations of the FRR needs. This will application will also include a testing environment where results of different outage and prediction risk distributions, following different design choices and parametrizations, can be compared and assessed.

Results visualization application

The result visualization application is a key factor to ensure the transparency of and confidence in the daily calculated volumes. The goal of this application is to visualize the results so that they can be easily understood and enable the user to reproduce the result. This application allows the user to:

- visualize the results and the selected outage scenario, training scenario;
- present graphs, KPI's and secondary results in order to understand the associated dynamic sizing (i.e. kind of dashboard that can be used to generate standard reports justifying the results);
- monitor the performance allow for monitoring of the performance of the model by means of KPI's (indicating the need for modifications of the method).

As the quality of these visualizations will highly depend on user's experiences, it is likely that the user interface will evolve based on user's feedback.

8.2. Implementation planning

A first release of the dynamic sizing tool can be used in the market between 9 to 12 months after approval and resources at Elia have been cleared. It is to be stressed that this will contain a first version for the outage-only (OO) method, while the machine learning (ML) methods require further testing in the parallel run and can be used in the market between 19 and 21 months after start of the project.

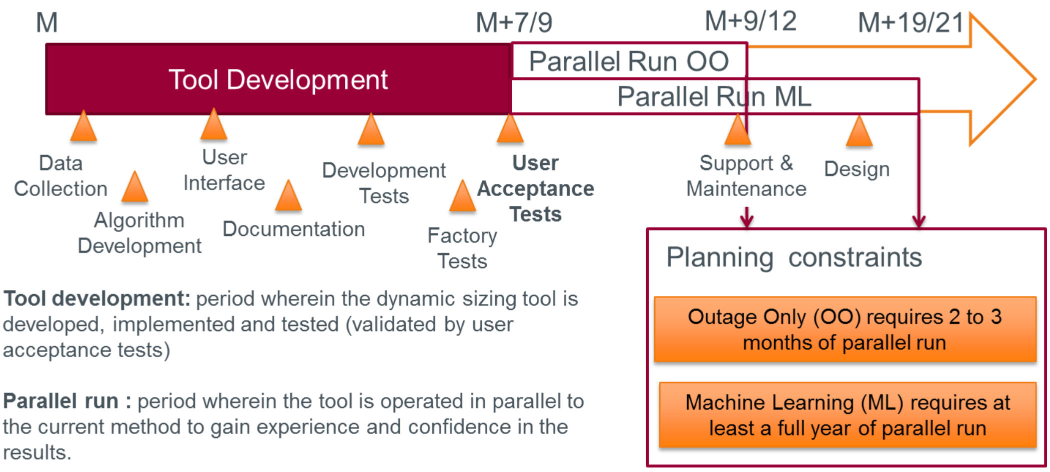


Figure 55: Visual representation of the implementation planning

The project will firstly contain a track for the development of the tool in which the dynamic sizing tool is developed, implemented and tested (validated by means of user acceptance tests). This development is expected to take 7 to 9 months. Thereafter, a parallel run wherein the tool is operated in parallel to the current method to gain experience and confidence in the results taking 2 to 3 months for Outage-Only and at least 12 months for Machine Learning. This difference is explained by the complexity of the machine learning methods compared to the Outage-Only, requiring at least tests for every month of the year. One can notice the typical elements in a development and testing project:

- **Data collection**

An architecture will be put in place to provide the necessary data in order to conduct the (1) regular training and the (2) recurrent sizing (e.g. daily). The training requires a regularly updated database with all historic data concerning the predicted system conditions and the corresponding system imbalances. The dynamic sizing itself requires the availability of a dataset with the predicted system conditions, including the power plant availabilities and the predicted NEMO-link schedule for the next day which is to be accessed each day to conduct the sizing of the FRR needs.

Starting with an in depth requirement analysis an requirement specification, the correct data sources and versions will be identified, the forecast horizons to use for renewable prediction tools will have to be re-assessed and an efficient architecture that allows a timely and performant access to the data will be put in place. Furthermore, a tool will be needed to predict the schedule of the NEMO-link. The latter requires the development of a specific tool.

- **Algorithm development**

This phase concerns the development of an efficient and robust algorithm using machine learning in order to compute the FRR needs for the next day. This work starts with the prototypes developed in the framework of this study. Nevertheless, it is made clear that further developments are required to transform the algorithms in an operational tool. This includes the Monte Carlo module allowing to take into account the imbalance risk of the forced outage risk, and in particular the “Machine Learning” module. The performance of the latter is expected to be increased by means of further calibration of the parameters.

This requires a decision on the training period, training frequency, resolution, lead time as discussed in the Proof of Concept. Furthermore, the tool also needs to incorporate KPIs to monitor the performance of the method. Also, additional tools have to be developed allocating the FRR needs to the different FRR product types.

- **User Interface**

The objective is to develop a User Interface (UI) which is intuitive and easy-to-use for the operator of the tool. The UI has four main components: (1) an outage risk application, (2) a machine learning training application, (3) a dynamic sizing application and (4) a result visualization application, as described in Section 4.1. It is important to assure that the results of the training are easily accessible and re-producible by the user, in order to justify historic FRR needs, but also to serve as a testing environment comparing the FRR needs of calculations with different assumptions (e.g. different parameters).

As is the case for the data collection, the development of the UI starts with a requirement analysis and requirement specification phase to ensure it covers all needs. It is likely to be improved following operational experience.

- **Documentation**

As this tool will implement complex algorithms and is supposed to be used intensively, it is essential to have a clear and structured documentation (decision logs and user manuals).

- **Development Tests**

The development tests are conducted during the development of the tool. They verify the correctness of new functionalities and avoid that new features lead to regression of other functionality. They will include corner cases and rare events as well as normal market situations, and realistic future scenarios.

Development tests include unit tests (functioning of each element of code), component tests (ensure that each module behaves as specified), functional systems tests (ensure compliance of the software platform based on pre-defined inputs) and non-functional system tests (assess the performance of the software platform for pre-defined case studies).

- **Factory Tests or User Acceptance Tests**

After a first operational version of the tool is available, the purpose of this step is to plug the solution in an operational environment and conduct a second round of tests. This enables the full validation of the system, including "non-functional" requirements.

Tests are conducted on a large scale, ideally in production-like environments, including integration of the solution with other platforms. The objective is to simulate operation and make sure everything behaves as expected. Trainings sessions are foreseen in order to provide trainings to the users of the tool.

Tests include historical data, realistic future scenarios for the coming years and stress situations are used to track and fix bugs. Changes can be integrated in the application through new releases.

- **Support and Maintenance**

Support and maintenance will start as soon as the factory tests are initiated. Indeed, as the tool's performance and robustness are critical, a continuous maintenance and support is foreseen in order to mitigate the risk of failure. Furthermore, as the market is constantly evolving a specific care should be dedicated to the adaptation and improvement of the tool in order to assure its performance. A "help desk" should be put in place to track bugs: a platform which allows sharing the bugs and questions..

- **Design improvements and updates**

As already indicated earlier, the development of the tool is not a one shot implementation and will require design improvements along the way. In view of the implementation time for the machine learning methods, design adaptations and algorithm refinements are expected.

8.3. Implementation budget and business case

Table 30 represents an overview of the different cost components of the implementation of the dynamic sizing. First or all, a development project for the development of the software platform, including the development of the algorithms from the prototypes presented in this study is estimated at €700,000 ("Implementation Dynamic Sizing Tool" in Table 30).

In addition, but outside the scope of this study, additional tools to optimize the FRR volumes are to be procured, and tools to support the actual procurement are roughly estimated at €300,000 ("Development of Procurement Tools" in Table 30). It is to be stressed that the latter estimation will be refined when investigating the daily procurement.

Table 30: Estimation of the different cost components for the development and operation of the dynamic sizing tools

		Cost [€]
Project	Implementation Dynamic Sizing Tool	700,000
	Development of Procurement Tools*	300,000
Recurrent	Support and Maintenance of the Tools	200,000
	Design Improvements and algorithm refinement	450,000 to 700,000
	Reporting and monitoring	
	Daily volume determination operations	
	Daily procurement operations*	

**Will be further elaborated in a follow-up project on daily procurement*

An estimate of the recurrent costs is based on five components. First of all, the support and maintenance of the tools itself is estimated at 200,000 € per year. This concerns implementation of required software updates and maintenance of the databases. Additionally, one has to foresee a budget of €450,000 to €700,000 per year to develop recurrent design improvements and algorithm refinements, operations concerning the daily volume determinations (e.g. daily validation of the results), daily procurement operations (assuming a highly automated process) and monitoring and reporting. Again it is to be stressed that the cost estimations of the processes linked to the daily procurement are further investigated in the follow-up project. Together with the support and maintenance of the tools, a total yearly recurrent cost is estimated between €650,000 and €900,000 per year.

When taking into account the annual project development costs (based on a linear depreciation of five years, common for such applications, a yearly budget of €200,000 per year is to be foreseen as total non-recurrent costs for the project), this results in a total cost between 850,000 and 1,100,000 €. This is found to be far below the yearly benefits of reducing FRR needs that are calculated in the CBA. Indeed, the benefits, being 1.48 to 1.71 M€ for the outage-only and 2.51 to 2.97 M€ for the machine learning following the gains in FRR needs do largely exceed the implementation cost. Again, the uncertainty of these benefits is to be stressed following the assumptions that had to be taken on future FRR prices.

An overview of the CBA is given in Table 31. On top of the positive net gains, i.e. 0.38 – 0.86 M€ for Outage Only and 1.41 – 2.21 M€ for machine learning methods, it is to be stressed that the dynamic sizing allows better reliability management (which is difficult to quantify in monetary terms) and the fact that static sizing is in any case not an enduring solution on the middle to long term and needs to be replaced at a certain point in time.

Table 31: Overview of the Cost and Benefit Analysis

	Benefits [M€]	Cost [M€]	Net Benefit [M€]
Outage Only	1.48 - 1.71 M€	0.85 – 1.10 M€	0.38 – 0.86
Machine Learning	2.51 - 2.97 M€		1.41 – 2.21
+Additional Benefit 1	Better Reliability Management		
+Additional Benefit 2	Robust towards dynamic system conditions		

GENERAL CONCLUSIONS

This study investigates the feasibility of a dynamic sizing method to determine the FRR needs, and this as an alternative for the current static sizing method in which the system operator determines fixed FRR needs for an entire year. In contrast, a dynamic sizing method can be used every day, before day-ahead market closure, to calculate the FRR needs for every hour of the next day. This allows adapting the FRR needs to the system imbalance risk, determined by the day-ahead predicted system conditions.

The six proposed methods are based on the current framework in which the probability density curves of forecast risks (i.e. the expected system imbalance without forced outages) is convoluted with the probability density curve of forced outage risks. This first curve is constructed by means of a statistical tool which generates, based on historical data analysis, a probability distribution curve representing the **forecast risk** which corresponds to the day-ahead predicted system conditions. The second curve is constructed by means of a Monte Carlo tool simulating the **outage risk** of power plants, or HVDC-interconnector, based on the scheduling of these assets.

After convolution in one probability distribution curve, this allows making a day-ahead estimation of the imbalance risk which allows determining the corresponding FRR needs to cover the potential system imbalances with 99.9% probability.

Selection of three dynamic sizing algorithms

Analysis of six dynamic sizing methods demonstrated that for middle-long term, only three methods can be considered for implementation in a dynamic sizing method: Outage Only, Quantitative Clustering and Continuous Neighbors. These methods demonstrate a reduction of the average FRR needs while obtaining a better reliability management compared to the current static approach. This is obtained by reducing the reserve needs during low risk conditions, avoiding oversizing the FRR needs, and increasing reserve needs during high risk conditions (at least for Quantitative Clustering and Continuous Neighbors), improving the reliability in these periods.

The **Outage Only (OO)** approach is an intuitive dynamic sizing method based on the outage risk. Based on the scheduling of NEMO-link, as well as foreseen events such as planned maintenances, the outage risk distribution curve is adapted, resulting in a reduction of the FRR needs upon lower outage risk. Analysis has shown that this method is particularly useful as an intermediary, or fallback solutions given its limited complexity. The **Quantitative Clustering (KMEANS)** and **Continuous Neighboring (KNN)** are advanced statistical tools, referred to as machine learning, which are trained on historical observations, and recognize the system imbalance risk based on day-ahead predicted system conditions. Mainly based on the forecast risk, they show increasing potential with the renewable capacity installed.

Two other methods investigated based on 'human intuition' concerning categorizing the system conditions towards system imbalance risk obtained a negative assessment on the minimum technical requirement (as they are found to be unable to guarantee the pre-defined minimum requirements concerning reliability). Another method based on an artificial neural networks, a popular category of machine learning algorithms was found to be incompatible with the current design of the method based on two probability distribution curves, although further investigation in the long-term may overcome this problem.

For the three methods, simulations for a Belgian context reveal substantial potential as alternative for static sizing, showing better reliability management and average FRR needs reductions, and this in several future system conditions represented by sensitivities on the balancing market behavior, offshore wind park evolutions and a nuclear phase-out. For these reasons, the three methods are selected for a Proof of Concept.

Selecting three methodologies is useful as OO provides potential for a transitional implementation, following its lower complexity, while KMEANS and KNN allow further integration in one single 'hybrid' method upon implementation.

Proof of Concept

The Proof of Concept simulates the selected methods for an entire year 2020 (and 2027 where applicable) taking into account real-life constraints following sizing resolution (4-hourly), training frequency (monthly or daily), lead time (day-ahead), as well as imperfect forecast of NEMO-link flow direction. The proof of concept can therefore be seen as a "**virtual" parallel run** of the dynamic sizing methods, together with the current sizing approach.

A **disclaimer** to be made is that all FRR needs presented in this study are a best-estimate of the dynamic FRR needs profiles in 2020 and 2027, as assumptions were to be taken to represent the future system conditions (e.g. installed renewable capacity). It cannot be excluded that these results are different to the dynamic FRR needs observed in the future. Furthermore, all results are obtained by means of 'prototypes' which are likely to perform better when optimized upon industrialization and operational experience.

(a) Reliability, FRR Needs and Robustness

First of all, results confirm that a dynamic sizing method results **in a better reliability management**, as well confirm the **average FRR needs saving in all three methods**. Each method has obtained average FRR reductions, up to 64 MW and 46 MW for respectively up- and downward FRR needs, with a spread between minimum and maximum FRR needs of 408 MW and 995 MW for respectively up- and downward FRR needs. Obviously, these FRR needs reductions are lower in the OO, i.e. 31 MW and 15 MW for respectively up- and downward, with a spread of 54 MW and 112 MW.

Further, the Proof of Concept demonstrated **a higher robustness of dynamic** compared to static sizing. The latter is found to be sensitive to extreme events in the historic data set, which are expected to occur more frequently in future systems with more dynamic conditions (e.g. variable renewable energy). It is concluded that static methods results in volatile FRR needs and become unsuitable in the middle or long term. Dynamic methods do not face this disadvantage as such extreme outliers only results in elevated FRR needs when facing similar system conditions.

Finally, an analysis of the three methods in different scenarios and even a case study for 2027 after the nuclear phase out confirms the robustness of each method. Even a scenario with high market reactivity where market players are able to balance large part of the future system imbalances shows positive potential for dynamic sizing, while the potential is even expected to increase in a post-nuclear era.

(b) Transparency

A detailed analysis of the hourly FRR needs and the corresponding system conditions allows concluding that the **machine learning methods do not behave as a black box**. The analysis of the correlations, visualized in a comprehensive dashboard, allows understanding the relation between system conditions and corresponding FRR needs, while a dashboard visualizing the FRR needs and system conditions allows the operators of the tool, but also the regulator and market parties, to understand and interpret the results. Obviously, this transparency is less of an issue in the Outage Only method based on a limited possible set of outage risk probability distribution curves.

(a) Design aspects of dynamic sizing

A sensitivity analysis on three design aspects, i.e. training frequency, sizing resolution and lead times allows formulating preliminary recommendations toward implementation. Firstly, a sizing resolution of 4 hours (compared to 1, 2, 8 and 24 hours) is put forward as an acceptable resolution in terms of compatibility with the standard product to which Europe is evolving (i.e. 6 blocks of 4 hours) and potential of a dynamic method. Secondly, a monthly training (compared to yearly and daily) frequency is put forward as an acceptable frequency, allowing taking into account incremental capacity and market evolutions, while avoiding unnecessary efforts of a daily training. Thirdly, a day-ahead lead time (compared to 3 days ahead) is put forward as accuracy of the forecasts is found to be reduced substantially when moving towards a larger lead time between sizing and market closure.

(b) Cost and Benefit Analysis

The cost and benefit analysis of the financial gains in terms of FRR savings compared the expected implementation cost show a business case which is largely positive, even without taking into account the additional gains in reliability and robustness compared to the current static approach. The Outage Only method is estimated to bring a yearly financial gain between 1.48 and 1.71 M€, while the machine learning methods are estimated at a yearly gain between 2.51 and 2.97 M€. An analysis of the annual implementation cost, estimated between 0.85 and 1.10 M€ (including estimations of the development of procurement tools based on a 5-year depreciation, and the recurrent costs of operation) of the method show that these are largely exceeded by these financial gains. This results in an annual net gain between 0.38 and 0.86 M€ in case of Outage Only, and between 1.41 and 2.12 M€ in case of the machine learning methods.

Based on this assessment, implementation of a dynamic sizing method is put forward as a no regret decision (subject to the outcome of the daily procurement study) and it is recommended to consider the three models when moving towards an industrialization of a dynamic sizing method.

Implementation plan

An implementation plan is composed putting forward a tool which can be used in the procurement processes after 9 to 12 months resources are cleared. This includes a training period of 2 to 3 months. However, machine learning methods, considering the higher complexity, require a testing period of at least one year and can be implemented 19 to 21 months after start of the project. It is important that dynamic sizing methods will be subject to further improvements, in design (e.g. new imbalance drivers) and algorithms (e.g. evolution of statistical methods), and are capable to account for extra-ordinary system and network conditions (e.g. solar eclipse). Indeed, analysis of the imbalance drivers has shown that predicting the imbalance risk is not straightforward and large of the system imbalance remains unexplained by day-ahead system conditions. The question is to which extent this can be further improved, and which will require continuous efforts to update the statistic models and improve the performance.

A preliminary appraisal does not put forward insurmountable problems towards implementation, e.g. procurement or regulation, but important efforts will be required. Indeed, a necessary condition towards implementation of the proposed method is daily procurement of mFRR. This will be the subject of a follow-up study ‘‘dynamic procurement’ in the course of 2018 analyzing the impact on the mFRR means, i.e. the allocation towards contracted reserves, non-contracted reserves and reserve sharing, while studying the implications of daily procurement and product design in close collaboration with all stakeholders. The implementation of dynamic FRR dimensioning shall depend on the results of this dynamic mFRR procurement study and in specific the implications of daily procurement of the mFRR.

APPENDIX

Table A: Performance of the Methods in terms of Average Reliability (expressed in %) and FRR needs (expressed in MW)

		2020		2016		Cut-Off		Low reserve needs		Post-Nuclear	
		Reliability	**Need**	**Reliability**	**Need**	**Reliability**	**Need**	**Reliability**	**Need**	**Reliability**	**Need**
Static	+	99.871%	1510	99.927%	1261	99.914%	2101	99.875%	1240	99.874%	1341
	-		1539		1099		1537		1287		1536
Outage Only	+	99.858%	1465	99.904%	1233	99.858%	1466	99.866%	1215	99.865%	1263
	-		1530		969		1530		1280		1530
Extreme Cases	+	99.855%	1463								
	-		1482								
Manual Clustering	+	99.756%	1401								
	-		1334								
Quantitative Clustering	+	99.810%	1416	99.891%	1236	99.827%	1418	99.847%	1179	99.814%	1160
	-		1388		945		1386		1159		1391
Continuous Neighbors	+	99.856%	1426	99.894%	1232	99.856%	1428	99.869%	1188	99.876%	1202
	-		1448		945		1448		1209		1450

+: Upward FRR needs, -: Downward FRR needs

Table B: Maximum and minimum observed FRR needs (expressed in MW)

		2020		2016		Cut-Off		Low reserve needs		Post-Nuclear	
		Max FRR	Min FRR	Max FRR	Min FRR	Max FRR	Min FRR	Max FRR	Min FRR	Max FRR	Min FRR
Static	+	1510	1510	1261	1261	2101	2101	1240	1240	1341	1341
	-	1539	1539	1099	1099	1537	1537	1287	1287	1536	1536
Outage Only	+	1503	1460	1260	1220	1762	1460	1241	1214	1347	1260
	-	1538	1512	1100	730	1538	1512	1287	1265	1539	1511
Extreme Cases	+										
	-										
Manual Clustering	+										
	-										
Quantitative Clustering	+	1779	1322	1300	1170	1923	1320	1457	1120	1739	1039
	-	1698	997	1150	500	1694	1001	1417	834	1700	991
Continuous Neighbors	+	1746	1262	1290	1175	1904	1262	1435	1099	1694	1039
	-	1745	1038	1140	500	1745	1038	1458	865	1744	1038

+: Upward FRR needs, -: Downward FRR needs

Table C: Instantaneous Reliability (%) and corresponding FRR needs (MW) in High Risk Periods (HR) and Low Risk Periods (LR)

			2020		2016		Cut-Off		Low reserve needs		Post-Nuclear	
			HR	**LR**	**HR**	**LR**	**HR**	**LR**	**HR**	**LR**	**HR**	**LR**
Outage Only	**Outage Only**	+	99.94%, 1484	99.76%, 1447			99.94%, 1629	99.75%, 1447	99.94%, 1227	99.78%, 1204	99.99%, 1291	99.65%, 1244
		-	99.98%, 1551	99.48%, 1476			99.98%, 1551	99.48%, 1476	100%, 1298	99.51%, 1234	99.99%, 1551	99.49%, 1475
	Static	+	99.95%, 1510	99.78%, 1510			100%, 2100	100%, 2100	99.95%, 1240	99.79%, 1240	100%, 1341	99.73%, 1341
		-	99.98%, 1539	99.51%, 1539			99.98%, 1537	99.48%, 1537	100%, 1287	99.53%, 1287	99.99%, 1535	99.46%, 1535
Extreme Cases	**Extreme Cases**	+	99.87%, 1573	99.86%, 1387								
		-	99.99%, 1551	99.44%, 1232								
	Static	+	99.83%, 1510	99.91%, 1510								
		-	99.99%, 1539	99.53%, 1539								
Manual Clustering	**Manual Clustering**	+	99.87%, 1581	99.85%, 1274								
		-	99.99%, 1632	99.43%, 987								
	Static	+	99.79%, 1510	99.96%, 1510								
		-	99.99%, 1539	99.91%, 1539								
Quantitative Clustering	**Quantitative Clustering**	+	99.87%, 1562	99.89%, 1332			99.87%, 1580	99.85%, 1273	99.86%, 1270	99.87%, 1125	99.86%, 1424	99.84%, 1040
		-	99.94%, 1632	99.66%, 1124			99.94%, 1630	99.74%, 1124	99.94%, 1364	99.85%, 939	99.93%, 1632	99.69%, 1119
	Static	+	99.79%, 1510	99.97, 1510			100%, 2100	100%, 2100	99.81%, 1240	99.92%, 1240	99.74%, 1341	99.99%, 1341
		-	99.77%, 1539	99.98%, 1539			99.77%, 1537	100%, 1537	99.70%, 1287	100%, 1287	99.74%, 1535	99.99%, 1535
Continuous Neighbors	**Continuous Neighbors**	+	99.86%, 1577	99.90%, 1318			99.87%, 1658	99.9%, 1318	99.86%, 1300	99.89%, 1122	99.84%, 1462	99.92%, 1041
		-	99.94%, 1673	99.79%, 1190			99.94%, 1673	99.79%, 1190	99.96%, 1396	99.82%, 992	99.93%, 1673	99.82%, 1192
	Static	+	99.79%, 1510	99.97%, 1510			100%, 2100	100%, 2100	99.80%, 1240	99.94%, 1240	99.74%, 1341	99.99%, 1341
		-	99.75%, 1539	99.95%, 1539			99.75%, 1537	99.93%, 1537	99.77%, 1287	99.96%, 1287	99.74%, 1535	99.93%, 1535

+ : Upward FRR needs, - : Downward FRR needs

Table D: Reliability indicators : Average Reliability, AV (expressed in %), Maximum Uncovered Imbalance, MUI, (expressed in MW) and average Uncovered System Imbalance, USI (when facing an uncovered imbalance) (expressed in MWh)

	2020			2016			Cut-Off			Low reserve needs			Post-Nuclear		
	AV [%]	**MUI [MW]**	**USI [MWh]**	**AV [%]**	**MUI [MW]**	**USI [MWh]**	**AV [%]**	**MUI [MW]**	**USI [MWh]**	**AV [%]**	**MUI [MW]**	**USI [MWh]**	**AV [%]**	**MUI [MW]**	**USI [MWh]**
Static	99.871	400.79	8.97	99.927	96.40	9.06	99.914	394.91	8.06	99.875	336.19	8.10	99.874	569.81	10.04
Outage Only	99.858	443.61	9.59	99.904	145.00	10.08	99.858	443.61	9.59	99.866	370.46	8.44	99.865	636.84	10.93
Extreme Cases	99.855	399.08	8.41												
Manual Clustering	99.756	809.23	9.92												
Quantitative Clustering	99.810	385.80	8.63	99.891	132.60	8.88	99.827	385.84	8.81	99.847	329.52	7.68	99.814	455.97	9.46
Continuous Neighbors	99.856	301.96	8.64	99.894	176.61	9.29	99.856	301.96	8.64	99.869	255.62	7.86	99.876	368.31	8.62

Figure A: Upward FRR needs (expressed in MW) in the reference scenario

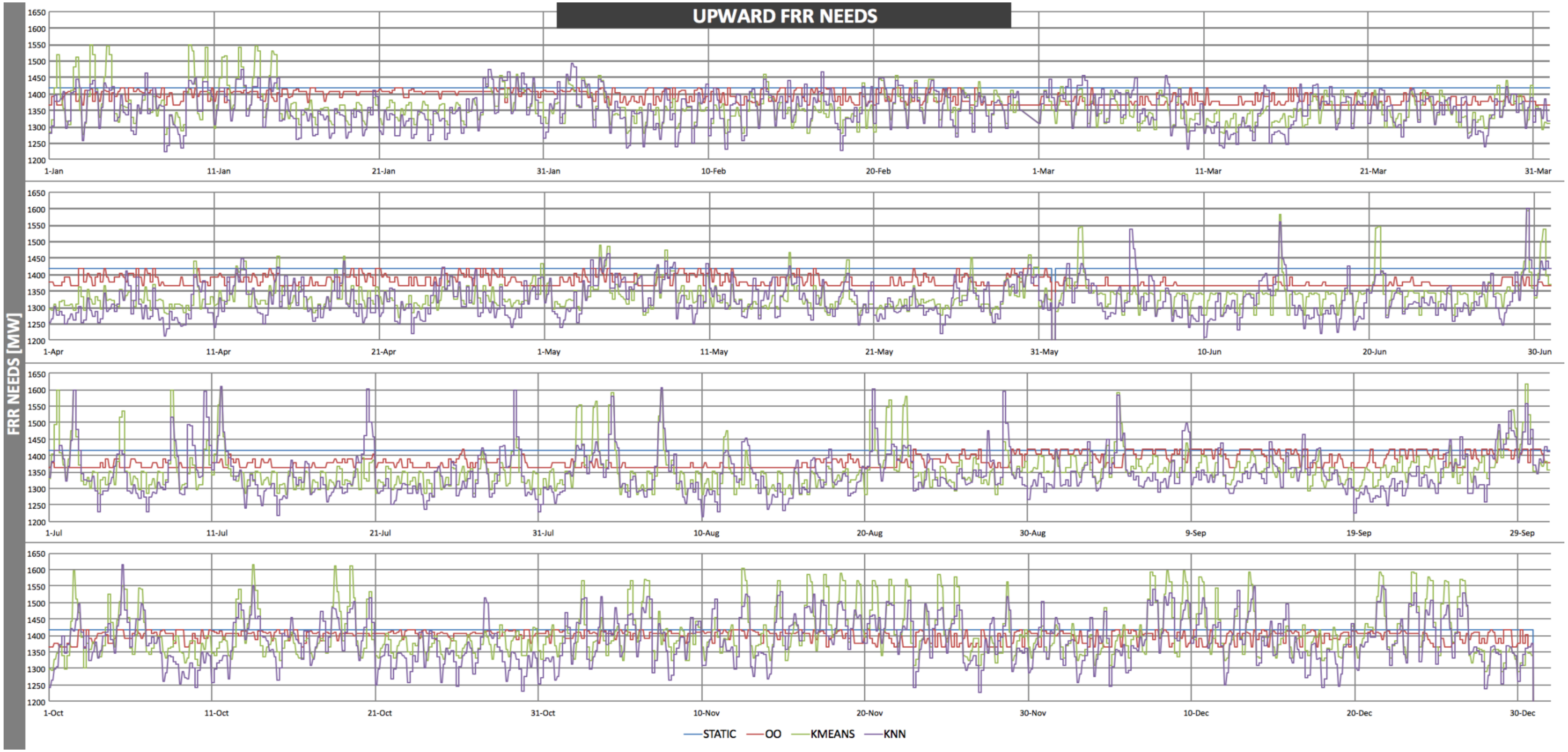


Figure B: Downward FRR needs (expressed in MW) in the reference scenario