FINAL REPORT

Smart Testing

December 18, 2020

Content

Executive Summary

The assessment of the availability of the reserves is a legal obligation and contractual right of Elia. In this regard, Elia is testing throughout the year the procured balancing capacity for FCR, aFRR and mFRR to ensure the availability of the said reserves. The current approach is based on past observations on the reliability of procured services and on the recorded success rate of previous availability tests.

The key objective of the "Smart Testing" is, for a given level of reliability, to reduce the number and volume of availability tests. This should lead a reduction of related costs. Smart Testing aims to use more extensively the available data in order to increase the effectiveness of availability tests.

In order to achieve the objectives of Smart Testing, three guiding principles for the methodology are defined. Firstly, Elia should be able to test any part of the BSP's portfolio. Secondly, the Smart Testing methodology should be unpredictable for the BSPs to truthfully assess the compliance of the BSPs at the moment that the availability test is performed. Finally, the methodology shall not discriminate between BSPs nor technologies.

Applying the principles of Smart Testing, the proposed methodology suggests to implement two scoring systems to provide recommendations for the organization of the tests. The first scoring system determines the Capacity Contracting Time Unit (CCTU) to be tested. The second one determines the bid(s) to be tested.

Additionally to the scoring system, two test regimes are introduced to frame the number and volume of availability tests to be performed. The first test regime aims to ensure that a significant part of the contracted capacities from a BSP is compliant. Once the BSP has demonstrated its reliability, the BSP enters the second test regime in which more limited volumes are tested (on top of a reduced number of tests) with the aim to confirm that the BSP remains compliant.

As for any implementation assessment, an overview of all impacts is necessary to support the implementation decision and associated plan. One aspect to take into account is that Smart Testing methodology is part of the balancing services which undergo a paradigm shift during the years to come with the balancing platform integration projects MARI and PICASSO (amongst others). The assimilation of the new framework continues even after the Go-Live of the platforms. A second aspect is that the application of the Smart Testing methodology would also require the modification of Terms and Conditions for BSPs as well as the development and upgrade of tools at Elia.

Given the significant amount of projects planned in the Roadmap 2021-2022 which are either mandatory or given a higher priority by stakeholders, the proposal of Elia, as discussed in the Working Group Balancing, is to not implement Smart Testing before 2023.

1. Introduction

The assessment of the availability of the reserves is a legal obligation and contractual right of Elia. In this regard, Elia is testing throughout the year the procured balancing capacity for FCR, aFRR and mFRR to ensure the availability of the said reserves. The current approach is based on past observations on the reliability of procured services and on the recorded success rate of previous availability tests.

The availability tests are an integral part of the service offered by the BSP and are therefore not remunerated nor compensated. The Balancing Service Provider (BSP) can take the costs of such tests into account in the bidding price while offering balancing capacity to Elia. Provided this can be found, a more effective and efficient test selection methodology would bring welfare to the society as a whole.

Pursuant to CREG's decision (B)658E/63 of November 21st 2019, Elia has received an incentive on the development of one or more methodologies to ensure that the availability tests of FCR, aFRR and mFRR are performed based on relevant and updated information which will increase the efficiency of the tests and reduce therefore their frequency.

Thus, the proposed "Smart Testing" methodology uses more extensively available data such as activation controls or perform additional analysis on the BSPs' behaviours to get more insights to improve the efficiency of the availability tests.

This report aims to be as transparent as possible on the Smart Testing methodology which will be applied on the market parties. However, the usage of the outputs to determine the testing conditions is subject to confidentiality to keep its unpredictable character for the BSPs, which ensures the effectiveness of the methodology itself.

2. Objectives and Scope of Smart Testing

As the Belgian TSO, Elia is responsible for the operation, maintenance and development of the transmission system in order to ensure the security of supply. Amongst the tasks outlined in the article 8 of the Electricity Law ("loi électricité"), ensuring the availability of ancillary services (and more specifically balancing services) is the focal point of the incentive on Smart Testing.

A priori, a greater number of tests would provide a greater assurance but not a certainty on the availability of contracted reserves. However, a greater number of tests are likely to inflate costs for the society as a whole.

The key objective of the "Smart Testing" is, for given level of reliability, to reduce the number of availability tests and the tested volume. This should lead to a reduction of related costs. Smart Testing aims to use more extensively the available data in order to increase the effectiveness of availability tests.

In order to reach those objectives, the Smart Testing relies on the idea that if tests are successful during identified periods of potential non-compliance or simply because some delivery points were not tested yet, those tests increase the confidence that a BSP will be compliant in general. Successful tests in different moments of a day and of different delivery points indicate a reliability of the BSP in the provision of services and therefore can trigger a reduction of the number of tests and the tested volume for the BSP.

As such, Smart Testing intends to identify "when" a test should be performed and "what" bid(s) should be tested and does not consider "how" a test is carried out. More specifically, the methodology will use available data such as historical activation and availability control data as well as analysis of the behaviour of BSPs in order to determine which moment should be tested. The methodology aims to also indicate which part of a BSP portfolio should be tested.

For the sake of clarity, the conclusions of this report may lead to the modification of the Terms & Conditions for BSPs (T&Cs for BSPs) of each balancing product (FCR, aFRR, mFRR) concerning the maximum number of tests. However, the design of the availability control, activation control, pre-qualification, monitoring and associated penalties of each balancing product is out of scope of this report.

3. Current framework

3.1 Balancing services testing

As of 2020, balancing products (FCR, aFRR and mFRR) experience a major overhaul of the designs. This report focuses on the new designs, in which the rules on the number of availability tests are harmonized across the three balancing products.

It has also to be noted that the future changes in the design of the products (e.g. in the context of the implementation of the MARI and PICASSO projects) may also have an impact on the Smart Testing methodology itself. Those changes may impact (amongst others) activation control and availability test due to new characteristics of Standard Balancing Energy Product for aFRR and mFRR.

3.1.1 General rules applicable for all three products

While evolving the design of the products, Elia has ensured to harmonize as much as possible aspects such as penalties, testing rules, etc. Nevertheless, some elements are still product specific such as activation controls or availability tests. As it will be detailed later on in section 4.2, Smart Testing relies on them.

There exist today a limitation and rules on the number of availability tests that Elia can perform. Under the current rules, Elia has the right to perform at maximum 12 availability tests on the 12 month rolling window:

- In case of two successive successful availability tests, Elia reduces this maximum to 6 availability tests on the rolling window.
- Any failed availability test will automatically set a limitation on the number of availability tests back to 12 for the rolling window.

Although the limit of 12 tests will remain, the detailed rules are to be changed with the Smart Testing methodology.

In case of several failures in controls and tests as defined in the respective T&C of each balancing product, different procedures are in place to remove the contribution of the unreliable delivery points from the prequalified volume.

3.2 Analysis of the key drivers in the performed number of tests in the current framework

As one of the goals of Smart Testing is to reduce the number of tests and tested volumes (provided that this does not affect the effectivity of the tests), an analysis and quantification of the impact of the key drivers in the number of tests being performed is of relevance. Two key drivers have been identified:

- the success rate of the BSP for any given test (or activation);
- the actual limitation on the number of test.

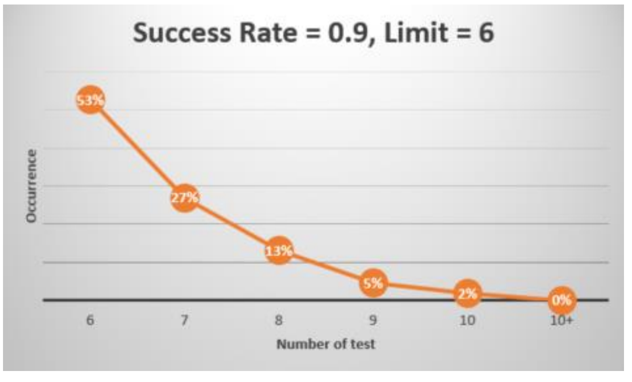
The analysis is supported by simple Monte Carlo simulations based on the assumption that the success rate of the BSP remains stable.

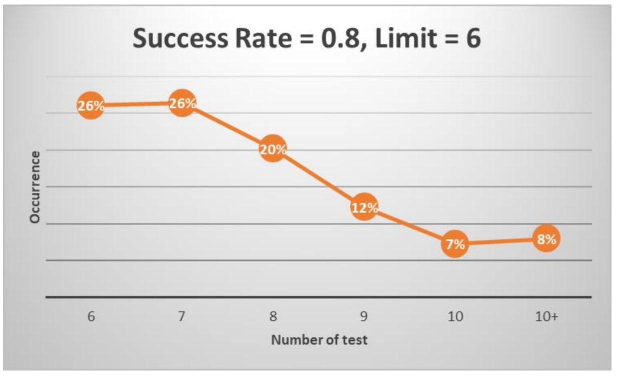
Impact of the success rate of the BSP

The two figures show the result of a Monte Carlo simulation with 10000 iterations of availability tests. Each availability test is represented with a binominal distribution with a success rate of 80% and 90%. In other words, for each of the 10000 iterations, a function will output whether a test is successful or not. The function continues to provide results until the maximum number of tests (6 or 12 in the current framework) is reached.

As an example, if the function provides 6 successful tests in a row, the maximum test (of 6) is reached (cf. 3.1.1) and the iteration is considered complete. If for example, the function provides 5 successful tests in a row and then one failed test, the maximum number of tests is then put back to 12 according to the current rules. The function will continue to output test results until a row of 2 successful tests is reached (which brings back the maximum to 6) or that the number of tests reaches 12. Continuing with the previous example, the best case scenario would yield a number of tests of 8 (7 successful tests and 1 failed test). In the worst case, the total number of tests would be 12

In the figures below, the percentage shows the number of occurrences for a given number of tests. For example, the 53% in the left figure indicates that on the 10000 iterations, 53% of the cases had a number of tests of 6, which means that all tests were performed successfully until the limit.





Any percentage for a number of tests above 6 indicates situations where at least one test failed. In any case, the number of tests is capped at 12.

The simulations show clearly that the number of tests (average of 6.7 for a success rate 90% and 7.8 activations for a success rate of 80%) is directly dependant on the success rate. The higher the success rate is, the lower the number of tests performed by Elia.

Additionally, the shape of the distribution of the number of tests is sensitive to the success rate. The higher the success rate, the higher the probability of having 2 successful tests in a row. This means that a lower success rate reduces such occurrence and therefore increases the chance to have alternating successful and failed tests. This leads to an increase of the number of tests overall.

Impact of the limitation on the number of tests

The other driver is the maximum number of availability tests. The figure shows a simulation where the maximum number of tests drops to 4 instead of 6. The reduction in the maximum of number of tests decreases the average number of tests by 2.2, to reach 4.5. One can also observe that the lower the limitation is, the higher is the occurrence of tests at the limit (65% for a limit of 4 and 53% for a limit of 6). The tail of the distribution is also thinner. Intuitively, a lower number of tests also reduces the chance of failure.

Conclusions

As conclusions, this short analysis shows that the proficiency of the BSPs in delivering their obligation is key in order to minimize the number of tests to be performed. The other element which will reduce the number of tests is the actual test limitation which directly impacts the number of tests to be performed.

Smart Testing is designed to reduce testing for BSPs which have demonstrated reliability. Before this reduction, relevant availability tests, supported by activation controls, are used to check if the BSP's reliability is sufficient to lower the number of tests and tested volumes, which is further explained in section 4.3. A good balance is necessary between less tests and enough confidence from the test results.

4. Smart Testing

The Smart Testing methodology is developed based on the objectives of efficiency of testing and reduction of the number of tests and tested volumes. To reach those objectives, some guiding principles are first defined, based on which the actual methodology will be developed. The methodology is composed of two Scoring Systems to determine when and what to test, and two test regimes to diminish the number of tests and tested volumes.

4.1 Guiding Principles of Smart Testing

In order to achieve the objectives of Smart Testing, three guiding principles for the methodology are defined. Firstly, Elia should be able to test any part of the BSP's portfolio. Secondly, the Smart Testing methodology should be unpredictable for the BSPs to truthfully assess the compliance of the BSPs at the moment that the availability test is performed. Finally, the methodology shall not discriminate between BSPs nor technologies.

All contracted capacities are susceptible to be tested

As Elia needs to ensure that all contracted capacities are readily available at all time, all capacities may be tested. However, testing all delivery points several times is not practical nor effective. The goal of Smart Testing is then to test as many delivery points as possible and put a priority on those which have not been tested yet or failed availability tests or activation controls in the past. The methodology will propose the most relevant bids for an efficient testing based on frequency of activation of the delivery points in the bids, contribution of the delivery points to the portfolio, size of the bids compared to the obligation of the BSPs, etc.

For example,

- A bid which has delivery points which were frequently and/or recently successfully activated is less susceptible to be tested.
- A bid which has delivery points which were frequently activated but failed to deliver is more susceptible to be tested.
- A bid which has delivery points which were unfrequently or not recently activated is more susceptible to be tested.
- A bid which represents a larger part of the obligation is more susceptible to be tested compared to bids with a smaller volume. However, as successful activations and testing increase, the focus may shift to bids with smaller volumes within a BSP's portfolio.

Unpredictability of the test

The moment of testing should not be easily predictable by the BSPs as it may otherwise give incentives to specifically ensure the availability at those moments to avoid penalties and not provide all contracted capacities during all other periods.

The Smart Testing methodology is transparent on the Scoring Systems so that BSPs are able to fully comprehend the methodology and how it impacts them. The BSP is able to reproduce the methodology based on the information provided in this report, assess the potential points of attention and address them.

The unpredictability is introduced in the selection of the CCTU and bid(s) for the availability tests. Therefore, on top of the uncertainty of the day of the test, a probabilistic approach in the selection of the CCTU and bid(s) renders the availability tests even more unpredictable.

Non discriminatory

The Smart Testing methodology should not discriminate as a principle. More specifically, Smart Testing does not discriminate between BSPs by their size or by the technologies of their portfolio (power plants, storage, demand response, decentral generation based on aggregation).

Big or small, same performance equals same outcome

Currently, it may be argued that a smaller BSP faces more chances that their whole portfolio is tested compared to a big player since only a number of tests is considered. However, the Smart Testing methodology ensures that all BSPs remain on a level playing field by considering the volumes being tested, proportionally to the size of the BSP.

Technology

Smart Testing is also technology neutral. However, based upon objective data, the methodology may naturally yield results which may be technology dependent. In this first version of the methodology, no variable directly or indirectly linked to technology is in use.

As an example, a portfolio, which is heavily weighted by power generation, is likely to be temperature sensitive. If such correlation is identified, the methodology could be amended to introduce the temperature as an additional component which plays in the selection of the period and/or the bids to be activated.

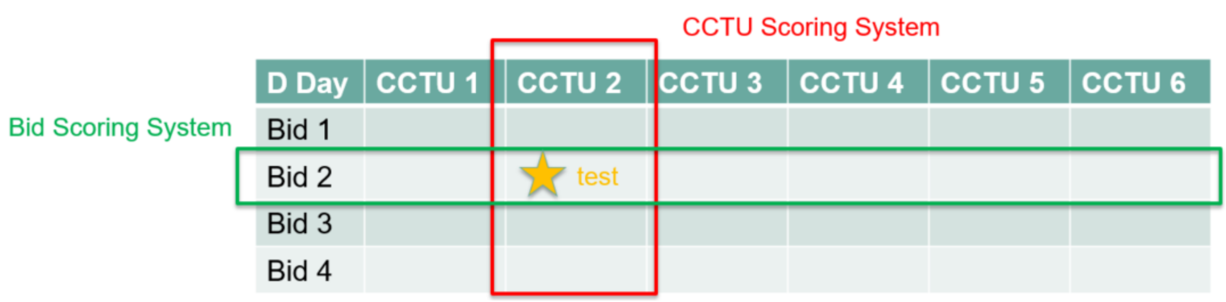
Non-discriminatory but adaptable

Smart Testing will be adapted to each BSP respecting the above-mentioned principles as data such as historical test performance is BSP specific. Moreover the system is designed to adapt naturally to the activation pattern of the BSP. A BSP which does not have any activation control is equally likely to be tested on its whole portfolio. Comparatively, a BSP which had a part of its portfolio successfully activated is less likely to be tested on this very part of its portfolio during an availability test.

4.2 Smart Testing in practice: the Scoring Systems

Applying the principles of Smart Testing, the proposed methodology suggests to implement two Scoring Systems to provide recommendations for the organization of the tests. For the sake of clarity, Smart Testing does not change the principle that the triggering of a test remains at the discretion of Elia as it is today. It <u>does not impact either the success or failure of an activation control</u>. The Scoring Systems only aims to provide insights for Elia to make better testing decisions.

The first Scoring System determines the CCTU to be tested. The second one determines the bid(s) to be tested.



As illustrated with the figure above, the combination of the information from the two Scoring Systems yields the availability test to be performed.

4.2.1 CCTU Scoring System

The CCTU Scoring System aims to indicate which moments are more relevant to be tested. The time resolution for the scoring system is fixed per CCTU. A CCTU resolution allows for capturing fluctuation and change of behaviour within a day and is directly linked to the obligation of the BSP.

The following example illustrates the mechanism of the CCTU Scoring System.

Component	Weight	CCTU 1	CCTU 2	CCTU 3	CCTU 4	CCTU 5	CCTU 6
Activation Control	33%	39	12	34	29	74	73
Availability test	33%	89	86	50	2	12	79
Margin Analysis	33%	30	18	9	82	58	50
Final Score		**52**	**39**	**31**	**38**	**48**	**67**

The scoring system is based on 3 components, namely activation control, availability test and margin analysis. Each will have a score per CCTU. The final score will be the weighted average of the 3 components. The weights are to be fine-tuned and fixed by Elia after a parallel run (cf. 5.33). The meaning of the final score can be interpreted or used in different ways (cf. 4.2.3). The weights will be made available to the stakeholders.

The system is designed to accommodate new data as they become available. A new BSP without any historical data will have the same likelihood to be tested on each CCTU. As time progresses, the score value for each component and each CCTU will be modified by the new data, that provide indications on the CCTU(s) to be tested next.

All scores in both scoring systems follow the same principle: an initial score is determined and several modification or adjustment factors are applied in order to change the final score. The scoring for each component is described in the following sub-sections. As a general remark, the scoring goes from 0 to 100 and factors from 0 to 1. The general principle is that a CCTU or a bid showing a low scoring has an increased probability of being selected for testing.

A detailed example of the scoring for each component of both Scoring Systems is given in Annex 3.

Activation control

This component relies on data from activation controls performed after the activation of the balancing services. A higher score is granted if the successful activation relates to a bigger part of the obligation of the BSP. On the contrary, a failed activation will be detrimental to the final score ($F_{failure}$). Finally, the final score is influenced by the freshness of the activation control ($F_{freshness}$), i.e. recent information is more valuable than old one. This last modification factor is used for all scores.

$$Score_{Activation}(CCTU) = \sum_{M} Score_{refActivation}(CCTU, M) * F_{failure}(CCTU, M) * F_{freshness}(M)$$

The initial score for the activation control component for the CCTU Scoring System is determined as follow:

$$Score_{refActivation}(CCTU, M) = \min(100; 100 * \frac{\max(requested\ volume\ (CCTU, M))}{average\ obligation\ (CCTU, M)})$$

The higher the requested volume by Elia is compared to the average obligation of the BSP for that month (for a given CCTU), the higher the initial score. In case the requested volume is above the average obligation, the score is capped at 100. It has to be noted that the initial score does not take into account any failure or success of activation control.

In a second step, the Failure Factor modifies the initial score:

$$F_{failure}(CCTU, M) = [1 - \min(1; \frac{\max failed\ bid\ volume\ \ (CCTU, M)}{average\ obligation\ (CCTU, M)})] * [1 - \frac{\#\ QH\ of\ failed\ activation\ control\ (CCTU, M)}{\#\ QH\ of\ activation\ (CCTU, M)}]$$

Failed activation control will always result in a lowering of the scoring of the concerned CCTU. The Failure Factor has 2 underlying factors. The first one represents the failed requested volume. An activation control is considered failed as defined in the Terms and Conditions of the relevant product[1]. If a control is failed, the total volume of the bid is considered as failed (*failed bid volume*). If several bids were simultaneously requested for activation, the total volume is the sum of each bid. The rationale is that this factor should go down to 0 regardless of the fact the obligation is offered in one bid or split into several bids. This does not have any impact in the determination of Valid Activated Volume which is used for the reliability threshold (cf. 4.3).

The second factor represents the "occurrence" (combination of frequency and length) of the failed activation control. Simply, all quarter hours for which a failed activation control occurred will be accounted for and divided by the total number of quarter hours per month for which an activation control occurred.

Finally, the freshness of the data is weighted per period of (rolling) three months. The most recent available data which is currently foreseen to be used is 2 months old due to the validation process of the data. Should the validation process be quicker in the future, the values will be shifted accordingly.

$$F_{freshness}(M) = \begin{cases} 4/30, if\ X = 2, 3, 4 \\ 3/30, if\ X = \ 5, 6, 7 \\ 2/30, if\ X = 8, 9, 10 \\ 1/30, if\ X = 11, 12, 13 \\ 0, if\ X = else \end{cases} , where\ X\ is\ the\ \#\ of\ past\ months\ compared\ to\ month, M$$

Availability test

Similar to activation control, availability tests information will also enrich the system in order to determine the best CCTU for performing a test. Availability tests only have a score modifier to reflect the "freshness" of the availability test ($F_{freshness}$). A successful test yields a high score of a 100 while a failed test yields 0 as an initial score. Given the limited amount of availability tests, a score of 50 is used in order to not adversely handicap the final scoring via this component. A score of 50 has been chosen[2] to differentiate the situation where there are no test performed and failed tests. A failed test will impact more negatively the score than no test.

[1] The concept of success or failure is not defined *per se* in the T&C BSP aFRR. Based on the current design and available inputs, Elia believes however that the activation control scoring may be computed, in line with the proposed methodology. The implementation details will be sorted out during the implementation phase of the aFRR product.

[2] *The weights of scoring systems are used to calibrate and achieve a balanced effect of each component on the final score.*

$$Score_{Availability}(CCTU) = \sum_{M} Score_{refAvailability}(CCTU, M) * F_{freshness}(M)$$

$$Score_{refAvailability}(CCTU, M) = \begin{cases} 100, if\ succesful\ availability\ test \\ 0, if\ failed\ availability\ test \\ 50, if\ no\ availability\ test\ occured \end{cases}, for\ the\ CCTU\ and\ month\ M$$

If more than one availability test occurs for a given CCTU and for a given month, $Score_{refAvailability}(CCTU, M)$ will be calculated as the average of each score weighted by the volume of the availability test.

Margin analysis

The margin analysis (cf. Annex 2) is an ex-post monitoring of the capability of the BSP to deliver its obligations. A margin analysis which indicates a potential inability of the BSP to provide its obligation will see the initial score to be lower than 100.

$$Score_{margin}(CCTU) = \sum_{M} \sum_{D \in M} \frac{Score_{refMargin}(CCTU, D)}{\#\ days} * F_{freshness}(M)$$

Concretely, the margin analysis is performed on a bid level. If all the bids of a CCTU have a positive margin then the initial score is at maximum. Should the margin be negative, the offered volume of the bid compared to the obligation of the BSP for the CCTU is removed from the maximum score of 100. Therefore, the bigger the bid with a negative margin is, the lower score is.

$$Score_{refMargin}(CCTU, D) = \begin{cases} 100,\ if\ margin(for\ all\ bids) \geq 0 \\ 100 - \frac{\sum Offered\ Volume\ (bid)}{Obligation\ (CCTU, D)}, else \end{cases}$$

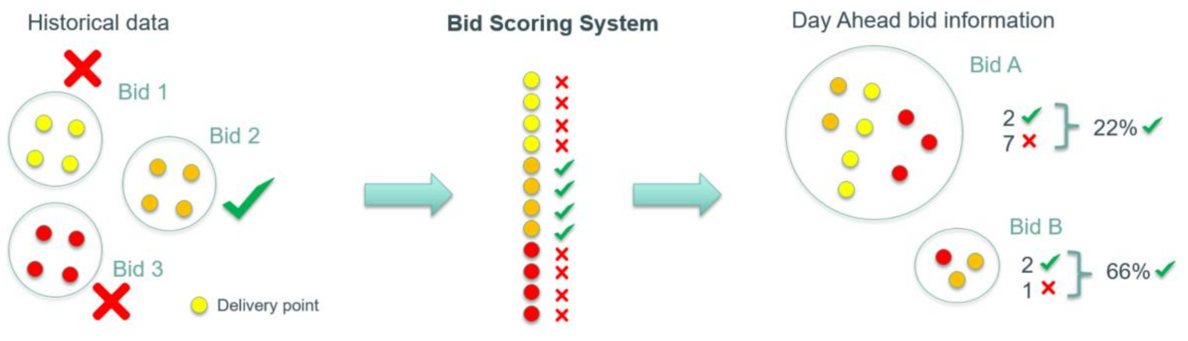
For the sake of clarity, if during at least one quarter hour of a CCTU, a negative margin is identified, the $Score_{margin}$ of the CCTU is negatively impacted proportionally to the total volume of the bids which have a negative margin, as contracted capacity should be available at any time.

4.2.2 Bid Scoring System

The Bid Scoring System aims to indicate which bids are more relevant to be tested. As the delivery points of a bid may change over each quarter hour, it is necessary to disaggregate the information from a bid level to a delivery point level, in order to have stable and comparable information.

The information gathered in the past on a delivery point level is then re-aggregated in accordance with the new submitted bid(s) (which include(s) the underlying delivery point(s)). As an illustration, the figure below shows 3 bids with 2 failures and 1 success, either for the availability tests or activation controls.

A successful activation of a bid translates into a successful activation of all underlying deliver points listed in the confirmation message. This information can then in a later stage be re-used when scoring the bids to be tested. Each bid will receive a score based on the underlying delivery points (DP_{pg} or DP_{su}) composing the bid.



The following table illustrates the mechanism of the Bid Scoring System:

Features	**Weight**	**Bid 1**	**Bid 2**	**Bid 3**
Volume		60 MW	30 MW	10 MW
Activation Control	33%	39	12	34
Availability test	33%	89	86	50
Margin Analysis	33%	30	18	9
Final Score		**52**	**39**	**31**

The weights used in the Bid Scoring System may be different from the ones used in the CCTU Scoring System and are also subject to further fine-tuning in the parallel run. The weights will be made available to the stakeholders.

Activation control

As in the CCTU Scoring System, activation controls also serve as input into the Bid Scoring System. The modification factors are applied on a delivery point level to be further aggregated onto a bid level.

In the CCTU Scoring System, the scores are mainly based on volume aspects. The same approach cannot be taken for the Bid Scoring System as the actual contribution of a delivery point cannot be easily determined. The Bid Scoring System rather looks at the inclusion of a delivery point in a bid and whether the delivery point already demonstrated its contribution in satisfying obligations.

$$Score_{Activation}\,(bid) = \sum_{M} F_{freshness}(M) * Adjust\,(bid) * \Big(\sum_{dp \in bid} Score_{refActivation}\,(dp, M) * F_{ratio}(dp, M) * Adjust\,(dp) \Big)$$

The higher the contribution of a delivery point (in volume) is in an activated bid, the higher is its initial score. Only delivery points which are listed in the confirmation message are taken into account as those are the ones effectively activated.

$$Score_{refActivation}\,(dp, M) = \frac{\#\text{ of QH of succesful activation (dp)}}{total\ \#\ of\ QH\ of\ activation\ (dp)}\ , for\ all\ delivery\ points\ which\ are\ Confirmed\ DPs$$

The Activation Ratio (F_{ratio}) aims to get a better grasp of the quality of the information in the initial score. For example, the information about a delivery point which is always activated but fails from time to time is more reliable than the

information about a delivery point which has only a limited number of activations even if these would all be successful. The first part of the factor of the Activation Ratio is a relative measure and gives an indication of the selection rate of delivery point (by the BSP) when the bid, it is part of, has been requested for activation. For the sake of clarity, "# of QH of activation (dp)" represents the number of QH where a certain delivery point is actually used by the BSP while " total # of QH of activation (dp)" represents the number of QH where a certain delivery point was in an activated bid and could have been used by the BSP.
The second part of the factor is an absolute measure and takes a look at the selection rate of a delivery point in a month. .

$$F_{ratio}(dp, M) = \frac{\text{\# of QH of activation (dp)}}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\text{\# of QH of activation (dp)}}{total\ \#\ of\ QH\ in\ month\ M},$$
$$for\ all\ delivery\ points\ which\ were\ part\ of\ an\ activated\ bid$$

As the contribution of each delivery point to a bid is likely to be different, a delivery point adjustment factor, Adjust(dp), is used in order to weight the contribution of each delivery point in the bid. Only delivery points in the confirmation message are considered.

$$Adjust\ (dp) = \frac{DP_{contribution}(dp)}{\sum_{dp \in bid} DP_{contribution}(dp)}, for\ all\ delivery\ points\ which\ are\ Confirmed\ DPs$$

Similar to the delivery point adjustment factor, a bid adjustment factor, Adjust(bid) is also used to weight the offered volume of the bid in the total obligation of the BSP.

$$Adjust\ (bid) = \frac{Offered\ Volume\ (bid)}{Obligation\ (CCTU)}, for\ all\ submitted\ bids\ for\ the\ CCTU$$

Availability test

Similar to activation control, availability tests information will also enrich the system in order to determine the best bid to be tested. Availability tests only have a score modifier to reflect the freshness of the availability test ($F_{freshness}$).

$$Score_{Availability}\ (bid) = \sum_{M} F_{freshness}(M) * Adjust(bid)\ * (\sum_{dp \in bid} Score_{refAvailability}(dp, M)\ * Adjust(dp))$$

The initial score for the availability test for each delivery point follows the same principle as for the CCTU Scoring System.

$$Score_{refAvailability}(dp, M) = \begin{cases} 100, if\ succesful\ availability\ test \\ 0, if\ failed\ availability\ test \\ 50, if\ no\ availability\ test\ occured \end{cases}, for\ all\ delivery\ points\ which\ are\ Confirmed\ DPs$$

If several availability tests including the same delivery points are performed the same month, the score will be 100 if all tests are successful and 0, otherwise.

Margin

The Margin Analysis Score of the bid is based on the score of each individual delivery points. The score of the delivery point excludes periods of activation control and availability tests in order to avoid an overlap of information. When a delivery point is part of a bid, and such bid has a positive margin, it receives a score of 100. A score of 0 will be attributed in case of negative margin.

$$Score_{margin}(bid) = \sum_{M} F_{freshness}(M) * Adjust(bid) * (\sum_{dp \in bid} \sum_{qh \in M} \frac{Score_{refmargin}(dp, qh)}{\# qh} * Adjust(dp))$$

$$Score_{refmargin}(dp, qh) = \begin{cases} 100, & if\ bid\ margin \geq 0 \\ 0, & else \end{cases}, \text{only when a DP is part of a non} - \text{activated bid}$$

4.2.3 Smart Testing recommendations based on scoring

The actual selection of the CCTU and bid(s) remains confidential in order to keep the unpredictable nature of the availability testing. Elia will keep a balance between an optimal selection of the CCTU and/or bid(s) and the randomness of the test. Moreover, further considerations including the impact on operational processes need to be taken into account during an implementation phase.

The test recommendation for CCTU and bid(s) can have several forms, from straightforward to more complex. For example, the recommendation for the selection of a CCTU (same reasoning applies for bids) can be to simply select the lowest available score. It has the disadvantage however to be predictable and to not account for the relative scoring value of the said CCTU compared to other CCTUs. A score of 10, 15 or 20 can be considered equally bad in the sense that they all require testing.

Another form of recommendation may be based on the conversion of (all or some of) the scores into probabilities of selection. Then, the selection is automatically performed based on those probabilities. This method provides a fair and transparent approach based on data but it also has a potential downside on the effectiveness of Smart Testing. A CCTU or bid which is often activated successfully may still be tested if the probability is not zero.

Smart Testing recommendations require hands-on experience and depends on the calibration of the Scoring Systems. As a start, the recommendation may follow simple rules which can be further elaborated as experience is built up.

The last point worth mentioning is that the Smart Testing provides only recommendations. In any case, tests are to be authorized and performed by the operational personnel of Elia depending on the state of the network at the time of the tests.

4.3 Test regimes: reduction of the volume of tests

The current limitation of the number of tests (described in 3.2) only looks at the number of tests performed and not actually at the total amount of MW tested. This last metric is more relevant to ensure availability of the contracted capacity. The criterion on the number of tests is therefore completed by a total volume to be tested on a rolling window period, which allows to also take into account activation controls as requested by the stakeholders in meetings and public consultations.

The Smart Testing methodology proposes the implementation of two Test Regimes, with a threshold to pass from one regime to the other. The threshold aims to quantify the reliability of a BSP in the provision of its obligation. The rationale behind the test regimes is the following:

- Test Regime 1: ensure that contracted capacity is generally available - perform availability tests in order for the BSP to demonstrate that its portfolio is available (if necessary) and pass the threshold. Activation controls are taken into account in order to reduce the necessity of availability tests and may even remove the necessity of performing an availability test to pass to the Test Regime 2.
- Test Regime 2: maintain the confidence level on the already controlled and tested volumes through a more limited number of tests and tested volumes.

Threshold

The threshold is the average of the obligations from the last 12 months, adjusted by the freshness of the data.

$$Threshold = \sum_{M} F_{freshness}(M) * Average_{M}[Max_{D}\left(Obligation(CCTU, D)\right)]$$

Once the threshold is reached, the BSP is considered as reliable and tests should only be used to confirm punctually the compliance of the BSP. In Test Regime 2, Elia has the right to perform up to 4 tests on a rolling 12 months basis. The activated volume of each availability test is limited to 50% of the maximum obligation of the day of the test. It has to be noted that availability tests performed in the Test Regime 1 do not count in the maximum number of tests in the Test Regime 2.

Based on the public consultation, Elia will keep the limitation of 12 tests per rolling window of 12 months for the Test Regime 1 to allow BSPs to better manage their risks.

Valid Activated Volume

The Valid Activated Volume is the activated volume (from a successful activation control or a successful availability test) which is considered as valid in the calculation to reach the threshold. The figure below illustrates the concept of Valid Activated Volume.

As soon as the total Valid Activated Volume is above the threshold, the BSP passes to the Test Regime 2. As long as the total Valid Activated Volume is above the threshold (even in case of some failed activation), the BSP remains in the Test Regime 2. Once there are too many failed activation controls or availability tests, the total Valid Activated Volume drops below the threshold and the BSP returns to the Test Regime 1.

Each Valid Activated Volume is constructed based on a delivery point. A generation unit which has succesfully activated 40 MW has a Valid Activated Volume of 40 MW linked to its delivery point. For a portfolio bidding, an activation of 30 MW will also yield a Valid Activated Volume of 30 MW, which will be split proportionally to the contribution of each delivery point in the confirmation message. For example, if the confirmation message has 3 delivery points of the same size, the Valid Activated Volume of each delivery point is 10 MW each.

The Valid Activated Volume for a delivery point is always the maximum successfully activated volume. For example, a delivery point which had previously 10 MW of Valid Activated Volume and is successfully activated for 15 MW, will see its Valid Activated Volume increase to 15 MW.

In case of a failed availability test or activation control, the Valid Activated Volume associated to the participating delivery points is set to 0 MW. For the next successful test, the Valid Activated Volume will be the volume successfully activated.

5. Implementation of Smart Testing

As for any implementation assessment, an overview of all impacts is necessary to support the implementation decision and associated plan. One aspect to take into account is that Smart Testing methodology is part of the balancing services which undergo a paradigm shift during the years to come with the balancing platforms integration projects MARI and PICASSO (amongst others). The assimilation of the new framework continues even after the Go-Live of the platforms. A second aspect is that the application of the Smart Testing methodology would also require the modification of Terms and Conditions for BSPs as well as the development and upgrades of tools at Elia.

Given the significant amount of projects planned in the Roadmap 2021-2022 which are either mandatory or given a higher priority by stakeholders, the proposal of Elia, as discussed in the Working Group Balancing, is to not implement Smart Testing before 2023.

5.1 Modification of terms and conditions for Smart Testing

The implementation of Smart Testing requires a modification of the Terms and Conditions (T&C) of the balancing products in order to change the limitation on the number of tests and introduce new clauses regarding the Test Regimes. The Go-Live of Smart Testing can be independent between each product from a contractual point of view.

A common Go-Live would ensure harmonization between the balancing products but require also that all other aspects such as IT implementation of Smart Testing into tools are ready. In such a case, the implementation time for Smart Testing would be increased.

Elia estimates that the changes in the T&C can be performed in parallel to the implementation and should roughly take 4 months (including a one-month public consultation and regulatory approval), if limited to the changes for Smart Testing. Elia is willing to limit the number of public consultations and approval processes to not overburden market parties, CREG and Elia with successive (minor) changes. The changes to the T&Cs that are needed for the implementation of Smart Testing would be grouped with other modifications. Dependencies in the planning would need to be considered globally with the market parties and the CREG when considering an implementation.

5.2 Other major changes in balancing services and others

Major changes in product designs are expected for the upcoming European balancing platform of aFRR and mFRR. The updates are a change of paradigm instead of an incremental evolution and require significant changes in the tooling and the way of working of Elia but also of the BSPs. Currently, the expected connections of Elia to the platforms are scheduled in Q1 (PICASSO, aFRR) and Q2 2022 (MARI, mFRR).

Others projects such as iCAROS also have strong interactions with balancing services which add further challenges and dependencies. Those changes should be taken in considerations while assessing the eventual implementation.

5.3 Smart Testing: step wise implementation

The implementation of the Smart Testing methodology was discussed with the stakeholders during the Working Group Balancing meetings. As a result, the proposal for a step-wise implementation where the first step includes the mFRR product only is confirmed. The rationale is the following:

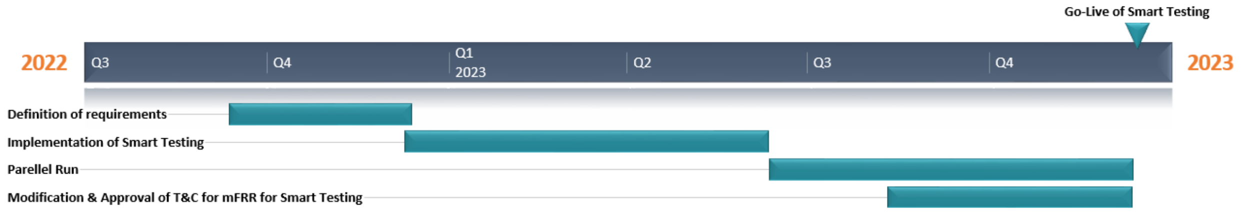
- The implementation of all three products at the same time will take more time than for a single product. The requests from the market for Smart Testing are mainly related to the mFRR product.

- As the Smart Testing methodology is a first-of-a-kind within Elia in terms of data management, the implementation of all balancing product altogether would likely be highly inefficient. The development of one product and its return on experience would smoothen the implementation of the methodology for the other products. It is difficult to envisage an eventual implementation for the three balancing products without putting a significant risk on the fulfilment of other obligations.

Based on its latest analysis, Elia expects that an implementation of Smart Testing should roughly take one year, which includes development, testing and parallel run.
A significant number of IT tools are impacted by the implementation of Smart Testing, some of which are undergoing significant refactoring or development from scratch in order to cope with the new requirements imposed particularly by the MARI project. The preliminary impact analysis confirm that the main challenge in Smart Testing is data collection and formatting rather than the algorithm itself. The parallel run should provide sufficient data in order to fine-tune the methodology's parameters.

At the same time, the T&C for mFRR would be amended to include the new rules on the limitation of the number of tests and tested volumes.

5.4 Smart Testing: implementation plan



The proposed planning takes into account the Go-Live of MARI in Q2 2022 as well as iCAROS Phase 2 expected in 2023.
The definition of requirements is planned to start soon after the delivery of the MARI project, including some potential aftercare. The implementation is currently foreseen for approximately 6 months with a parallel run of 6 months for fine-tuning.
The amendment of the Terms and Conditions for mFRR is planned for 4 months and can be done in parallel with the parallel run.

6. Conclusions

In the present report, Elia has defined the objectives of Smart Testing, a methodology which aims to perform availability tests more efficiently in order to reduce the number of tests and tested volumes. This reduction should be beneficial to the social welfare as the tests are priced by the BSPs into their offers.

Three guiding principles frame the bid and CCTU scoring systems in order to ensure that the BSP's portfolio is tested, in an unpredictable and non-discriminatory manner. Once a BSP has demonstrated its reliability in fulfilling its obligations, the BSP is granted a reduction of tested volumes and number of tests.

The implementation of the Smart Testing methodology must be looked at together with all upcoming projects. Given the significant amount of projects planned in the Roadmap 2021-2022 which are either mandatory or given a higher priority by stakeholders, the proposal of Elia, as discussed in the Working Group Balancing, is to not implement Smart Testing before 2023.

A first implementation effort assessment is provided. A step-wise implementation of the Smart Testing methodology applied to mFRR presents advantages in terms of development time and reduces potential inefficiencies from a first-of-a-kind project in data management for monitoring purposes by taking advantage of a return of experience on one product.

Annex 1: definition

Confirmed DP: the DP_{SU} or the list of DP_{PG} in the confirmation 10.C for mFRR product or the delivery points effectively supplying aFRR or FCR

Obligation: the FCR Obligation, aFRR Obligation or mFRR Obligation as defined in the relevant Terms and Conditions for BSPs.

Requested Volume: the FCR Requested, aFRR Requested or mFRR Requested as defined in the relevant Terms and Conditions for BSPs.

Annex 2: Margin analysis

The margin analysis aims at quantifying, at any time, the capacity margin that a BSP has, considering the delivery points associated. The analysis intends to assess whether the margin is high enough to fulfil its obligation.

Definition of a bid margin

The BSP is allowed to provide one or several bids to fulfil its obligation. Delivery points dp_i are associated to each bid *i*. The margin shall be calculated for each quarter hour *qh* and for each bid *i*.

The margin is computed as follows for demand response for upward product:

$$Margin\,[i][qh] = \sum_{dpi \in i} (Offtake\,[dp_i][qh] - MinOfftake\,[dp_i]) \; - Obligation\,[i][qh]$$

Where

- Offtake [dp_i][*qh*] is the offtake of delivery point dp_i at quarter *qh.*
- Obligation [*i*][*qh*] is the allocated capacity for bid *i* for quarter *qh*.
- MinOfftake [dp_i] is the lowest offtake value reached by delivery point dp_i for a given period.

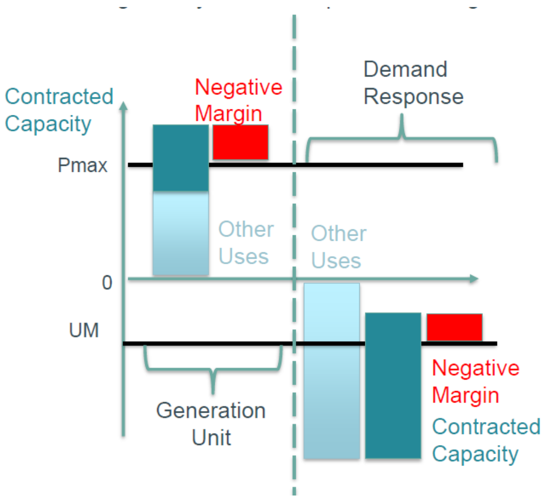
The *MinOfftake* [dp_i] (Unsheddable Margin, UM) is based on the lowest offtake (consumption) value for the considered 12 months rolling window. Elia is aware of the underlying shortcoming of the hypothesis regarding maintenance of assets behind the delivery points, which may drop the UM to zero consumption. The calculation of the UM may be improved with later phases of iCAROS project.

The margin is computed as follows for generation units for upward product:

$$Margin\,[i][qh] = \sum_{dpi \in i} (Pmax\,[dp_i] - \; Injection\,[dp_i][qh]) \; - Obligation\,[i][qh]$$

Where

- Injection [dp_i][*qh*] is the injection of the delivery point dp_i at quarter *qh.*
- Obligation [*i*][*qh*] is the allocated capacity for bid *i* for quarter *qh*.
- Pmax [dp_i] is maximum power of a generation unit by delivery point dp_i for a given period.



For downward products, the formulae have to be amended accordingly. The reference to be used for a generation unit will be the P_{min} instead of P_{max}. For DSM, the maximum measured off-take can be taken as a proxy to calculate the margin.

The interpretation of the margin is simple

- If Margin > 0, the BSP is expected to have enough margin to provide its obligation.
- If Margin = 0, all the nominated delivery points having to reach their (assumed) minimum offtake value (in case of demand response) if all the contracted capacities were to be activated. For a generation unit, the power output is at its maximum, if all the contracted capacities were to be activated. The BSP is not expected to violate its obligation in case of a full activation, but this could indicate a tight situation for the BSP.
- If Margin < 0, it is expected that some delivery points must exceed their (assumed) minimum offtake value to meet the obligation of the BSP, which may not be impossible but would be a strong indicator of potential non-compliance.

For symmetrical products, 2 margins are calculated, one for each direction. Based on the current designs and available inputs, Elia believes that a margin may be computed for aFRR and FCR, in line with the proposed methodology, to be used in the Scoring Systems. The implementation details will be sorted out during the implementation phase of the relevant product.

Annex 3 Example of Scoring

CCTU Scoring System

Activation control

Activation control Score	(A)	(B)	(C)	(A)*100	(1-(B)) * (1-(C))		
Month	Maximum Requested Volume (%) compared to average obligation of the month	Maximum Volume of Failed Activation (%) compared to obligation of the month	'Percentage of time of failure compared to total time of activation request	**Initial Score**	**Failure Factor**	**F freshness**	Final Score
M-2	46%	9%	10%	46	82%	4/30	5
M-3	80%	2%	5%	80	93%	4/30	10
M-4	49%	10%	10%	49	81%	4/30	5
M-5	86%	2%	5%	86	93%	3/30	8
M-6	6%	4%	4%	6	92%	3/30	1
M-7	44%	0%	0%	44	100%	3/30	4
M-8	38%	2%	9%	38	89%	2/30	2
M-9	10%	1%	7%	10	92%	2/30	1
M-10	22%	1%	8%	22	91%	2/30	1
M-11	54%	0%	0%	54	100%	1/30	2
M-12	34%	3%	3%	34	94%	1/30	1
M-13	70%	0%	0%	70	100%	1/30	2
					Total CCTU Activation Score		**43**

Availability test

Availability					Fail = 0, No test = 50; Success = 100		
	Percentage of successful availability tests for a given month	Number of test	Number of succesful test		Initial Score	F freshness	Final Score
M-2	0%	1	0		0	4/30	0
M-3	No Test Performed	0	0		50	4/30	7
M-4	100%	1	1		100	4/30	13
M-5	No Test Performed	0	0		50	3/30	5
M-6	No Test Performed		0		50	3/30	5
M-7	No Test Performed	0	0		50	3/30	5
M-8	No Test Performed	0	0		50	2/30	3
M-9	No Test Performed	0	0		50	2/30	3
M-10	100%	1	1		100	2/30	7
M-11	100%	1	1		100	1/30	3
M-12	100%	1	1		100	1/30	3
M-13	No Test Performed	0	0		50	1/30	2
					Total CCTU Availability Score		**57**

Margin

	CCTU 1	CCTU 2	CCTU 3	CCTU 4	CCTU 5	CCTU 6
Day 1	100	100	100	70	100	90
Day 2	100	60	100	100	50	100
Day 3	100	100	40	100	100	100
Day 4	...	...	...	...	...	...

Month	Average volume (%) compared to average awarded capacity of the month	Initial Score	F freshness	Final Score
M-2	84%	84	4/30	11
M-3	88%	88	4/30	12
M-4	81%	81	4/30	11
M-5	82%	82	3/30	8
M-6	99%	99	3/30	10
M-7	99%	99	3/30	10
M-8	100%	100	2/30	7
M-9	95%	95	2/30	6
M-10	90%	90	2/30	6
M-11	100%	100	1/30	3
M-12	81%	81	1/30	3
M-13	97%	97	1/30	3
		Total CCTU Margin Score		**90**

Bid Scoring System

Activation control

DP level analysis

Month	Activation Control	(A) Percentage of successful activation control	(B) Activation Ratio (%) compared to total activation time	(C) Activation Ratio (%) compared to the month	100* (A) Initial Score
M-2	DP1	85%	59%	15%	85
M-2	DP2	80%	58%	69%	80
M-2	DP3	92%	84%	67%	92
M-2	DP4	93%	72%	86%	93
M-2	DP5	96%	67%	1%	96
M-2	DP6	88%	93%	66%	88
M-2	DP7	100%	100%	58%	100
M-2	DP8	89%	96%	5%	89
M-2	DP9	84%	83%	88%	84
M-2	DP10	91%	60%	70%	91
M-3					

Bid level analysis

	DP Size Adjustment	(A)*100 *(B)*(C) Final Score per DP	(D) Intial Score per bid	(E) Bid Size Adjustment	(1-E) * (D) Final Score per bid
Bid 1	85,0%	6			
Bid 1	5,0%	2			
Bid 1	5,0%	3			
Bid 1	5,0%	3	13	60%	5
Bid 2	25,0%	0			
Bid 2	25,0%	14			
Bid 2	25,0%	15			
Bid 2	25,0%	1	29	30%	20
Bid 3	54,5%	33			
Bid 3	45,5%	17	51	10%	45

Availability test

DP level analysis

Month	Availability test	Percentage of succesful availability test	Initial Score
M-2	DP1	100%	100
M-2	DP2	100%	100
M-2	DP3	100%	100
M-2	DP4	100%	100
M-2	DP5	50%	50
M-2	DP6	50%	50
M-2	DP7	100%	100
M-2	DP8	50%	50
M-2	DP9	50%	50
M-2	DP10	50%	50
M-3			

Bid level analysis

	DP Size Adjustment	Final Score per DP	Initial Score per bid	Bid Size Adjustment	Final Score per bid
Bid 1	85,0%	85		60%	40
Bid 1	5,0%	5			
Bid 1	5,0%	5			
Bid 1	5,0%	5	100		
Bid 2	25,0%	13		30%	44
Bid 2	25,0%	13			
Bid 2	25,0%	25			
Bid 2	25,0%	13	63		
Bid 3	54,5%	27		10%	45
Bid 3	45,5%	23	50		

Margin Analysis

DP level analysis

	Margin	Average time (%) of a DP being part of a bid with positive margin	Initial Score
M-2	DP1	74%	74
M-2	DP2	10%	10
M-2	DP3	54%	54
M-2	DP4	38%	38
M-2	DP5	42%	42
M-2	DP6	96%	96
M-2	DP7	38%	38
M-2	DP8	3%	3
M-2	DP9	58%	58
M-2	DP10	50%	50
M-3			

Bid level analysis

	DP Size Adjustment	Final Score per DP	Intial Score per bid	Bid Size Adjustment	Final Score per bid
Bid 1	85,0%	63			
Bid 1	5,0%	1	44		
Bid 1	5,0%	3			
Bid 1	5,0%	2		60%	17
Bid 2	25,0%	11			
Bid 2	25,0%	24	45		
Bid 2	25,0%	10			
Bid 2	25,0%	1		30%	31
Bid 3	54,5%	32	27		
Bid 3	45,5%	23		10%	24