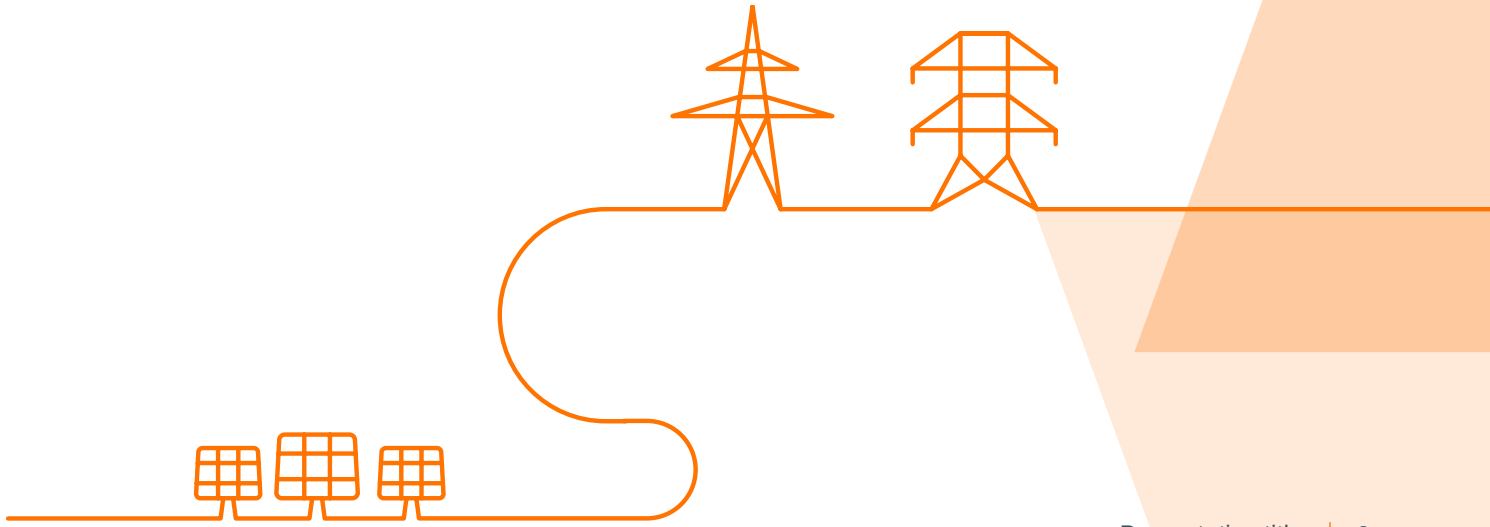
Smart Testing

Consolidated presentation

elia
| Elia Group

Content



Context

1. Smart testing

What is smart testing?

In 2020, Elia proposed a methodology to more specifically determine, by using the available data, when availability tests should be performed and which offers in this context should be triggered. This methodology allows, provided that the BSP passes these tests, to:

What is the goal?

For BSPs:

- Reduce the costs resulting from non-remunerated activations

For Elia:

- Reduce operational burden of test organisation and control
- Reduce impact on grid (each test may create an imbalance)
- Control better, reinforcing grid security

When?

The implementation of this methodology for mFRR had been foreseen to perform in 2024, assuming a go-live of MARI in Q2 2022.

The implementation of Smart testing to mFRR is now an objective for 2024 defined by CREG in the scope of the incentive for the promotion of the system's balance

Context – Smart testing methodology

Smart testing uses **two scoring systems** to select the bids for an availability test:

- A scoring system to **select the CCTU** for an availability test
- A scoring system to **select a bid** within that CCTU for an availability test

The scoring is based on activation control, (past) availability tests and margin control

Additional to the scoring system, **two test regimes** are introduced to limit the impact (in volume) of availability tests:

1. The first test regime **aims to ensure** that a significant part of **the contracted capacities** from a BSP **is compliant**
2. The second test regime aims **to keep in check the compliancy** of a BSP but with a **lower volume of availability tests**

CCTU scoring system determines which CCTU to select for an availability test

The Score per CCTU is based on 3 features:

- **Activation control:** past activations
- **Availability test:** past test
- **Margin Analysis:** ex-post monitoring of contracted capacity

structured data is required (date & time, failure/success, involved bid, DPs and their contribution, off-take metering …).

Features	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 per CCTU		52	39	31	38	48	67

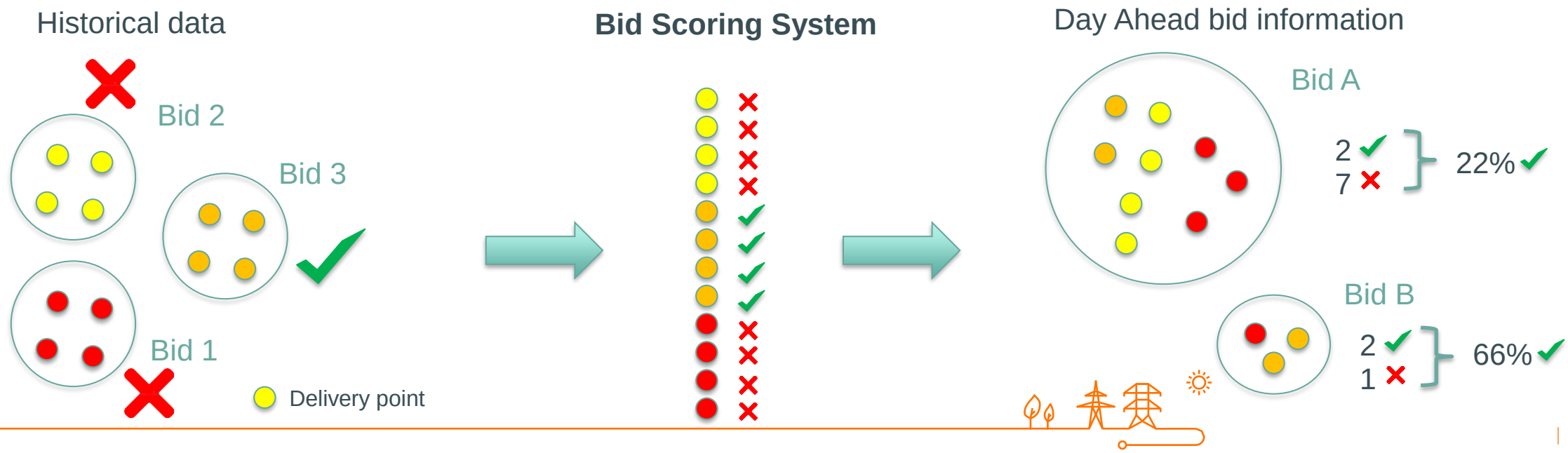
The Score per CCTU ranges from 0 to 100.

- A low value indicates that the CCTU needs to be tested.

Bid scoring system determines which bid to select for an availability test

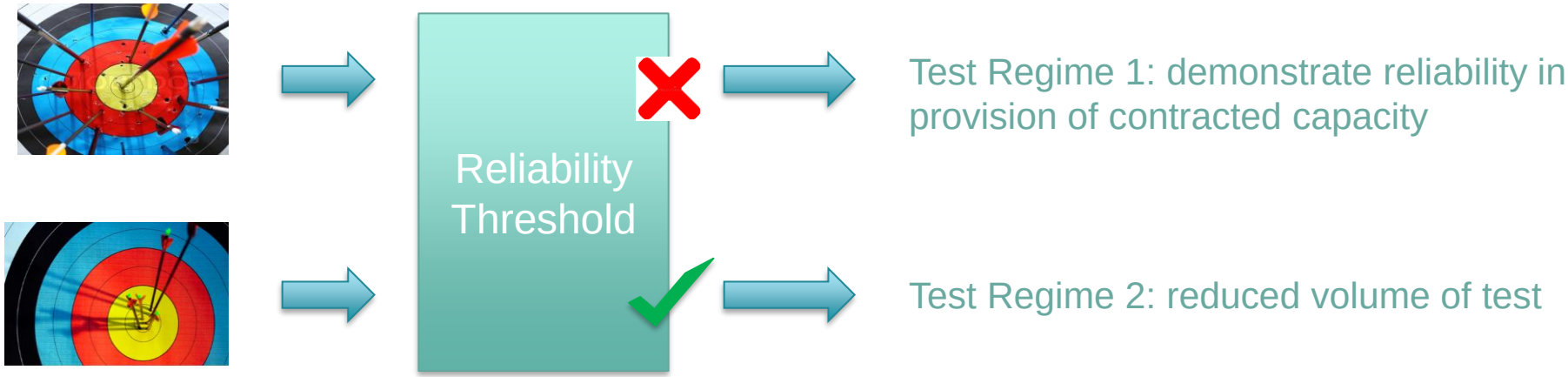
Features	Weight	Bid 1	Bid 2	Bid 3
Volume		60 MW	30 MW	10 MW
Activation Control	33%	39	12	34
Availability test	33%	89	86	50
Margin Analysis	33%	30	18	9
Final Score		52	39	31

- The Score per Bid is based on same 3 features but are adapted to the Bid Scoring System.
- The result of control and test **is disaggregated on a delivery point level**



Test regimes

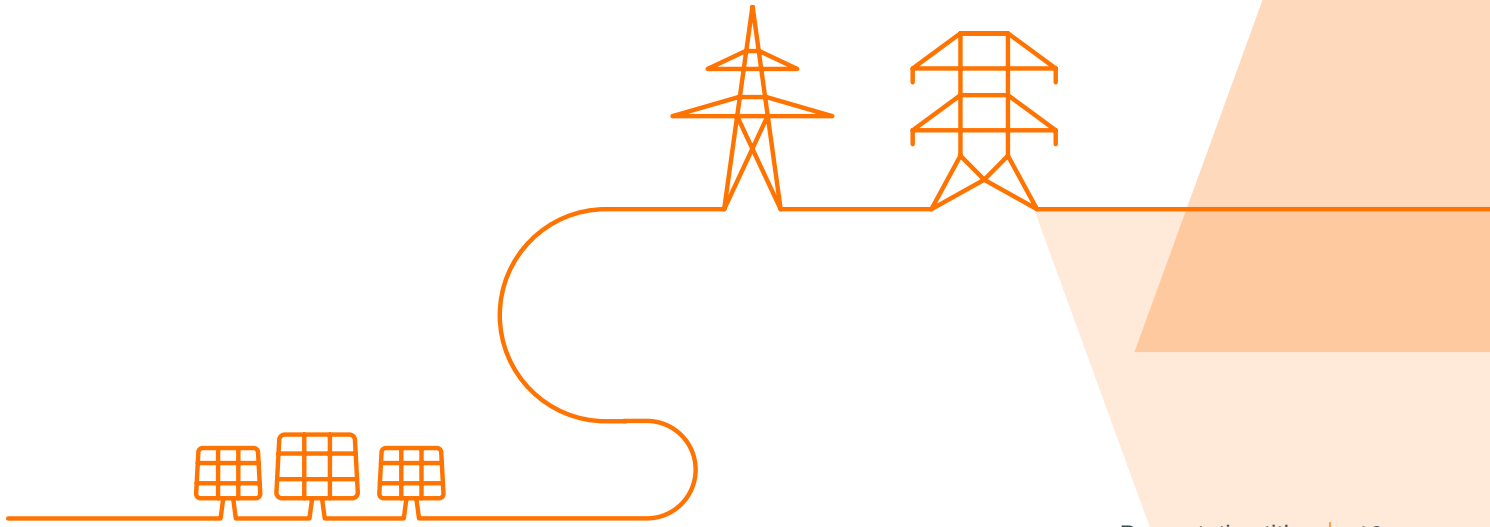
- Additionally, to the scoring system, **two test regimes** are introduced to limit the impact (in volume) of availability tests.
    1. The first test regime **aims to ensure** that a significant part of **the contracted capacities** from a BSP **is compliant.**
    2. The second test regime aims **to keep in check the compliancy** of a BSP but with a **lower volume of availability tests**



- The principles of Smart Testing should be **applicable for all balancing products**.

Final formulas

Content



General factors

General – freshness factor

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.

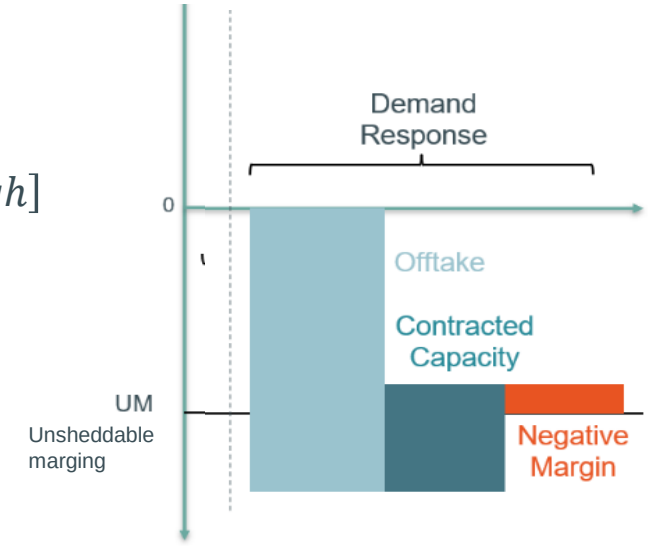
Should the validation process be quicker in the future, the values below would be shifted accordingly.

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

General – margin of a bid (downwards)

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

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



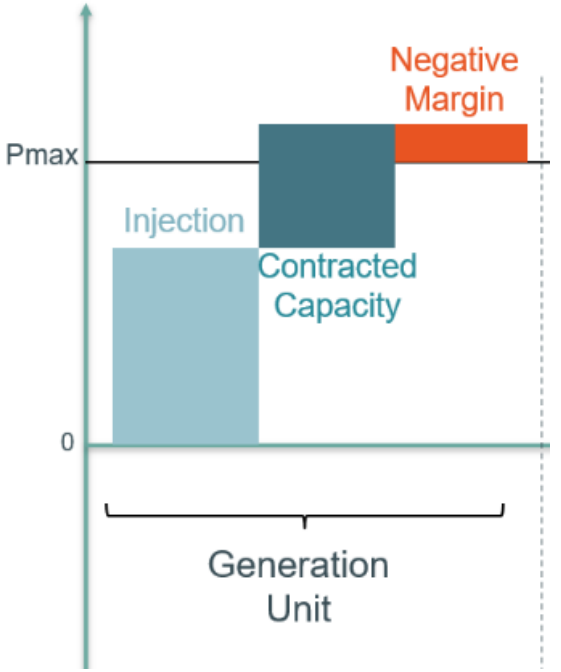
Where

- Offtake [dpi][qh] is the offtake of the Delivery Point dpi at quarter qh.
- Obligation [i][qh] is the allocated capacity for bid i for quarter qh.
- MinOfftake [dpi] is the lowest offtake value reached by Delivery Point DPi for the rolling 12 months.

General – margin of a bid (upwards)

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

$$Margin[i][qh] = \sum_{dp_i \in i} (Pmax[dp_i] - Injection[dp_i][qh]) - obligation[i][qh]$$



Where

- Injection [dpi][qh] is the injection of the Delivery Point dpi at quarter qh.
- Obligation [i][qh] is the allocated capacity for bid i for quarter qh.
- Pmax [dpi] is maximum power of a generation unit by Delivery Point dpi for a given period.

CCTU scoring

CCTU scoring system determines which CCTU to select for an availability test

The Score per CCTU is based on 3 features:

- **Activation control:** past activations
- **Availability test:** past test
- **Margin Analysis:** ex-post monitoring of contracted capacity

structured data is required (date & time, failure/success, involved bid, DPs and their contribution, off-take metering …)

Calibration to be done

Features	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 per CCTU	**100%**	**52**	**39**	**31**	**38**	**48**	**67**

The Score per CCTU ranges from 0 to 100

- A low value indicates that the CCTU needs to be tested

CCTU scoring – activation control

The general formula is as follows:

$$Score_{Activation}(CCTU) = \sum_{M=2}^{13} \left(Score_{refActivation}(CCTU, M) * F_{failure}(CCTU, M) * F_{freshness}(M) \right)$$

The initial score for the activation control component for the CCTU Scoring System is determined as follows for every month M :

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

Weighting factor for the maximum delivered volume wrt the average obligation

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

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

Highest failure

Total number of failures

CCTU scoring – availability test

The general formula is as follows:

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

With the values for the availability test:

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

CCTU scoring – margin control

The general formula is as follows:

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

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.

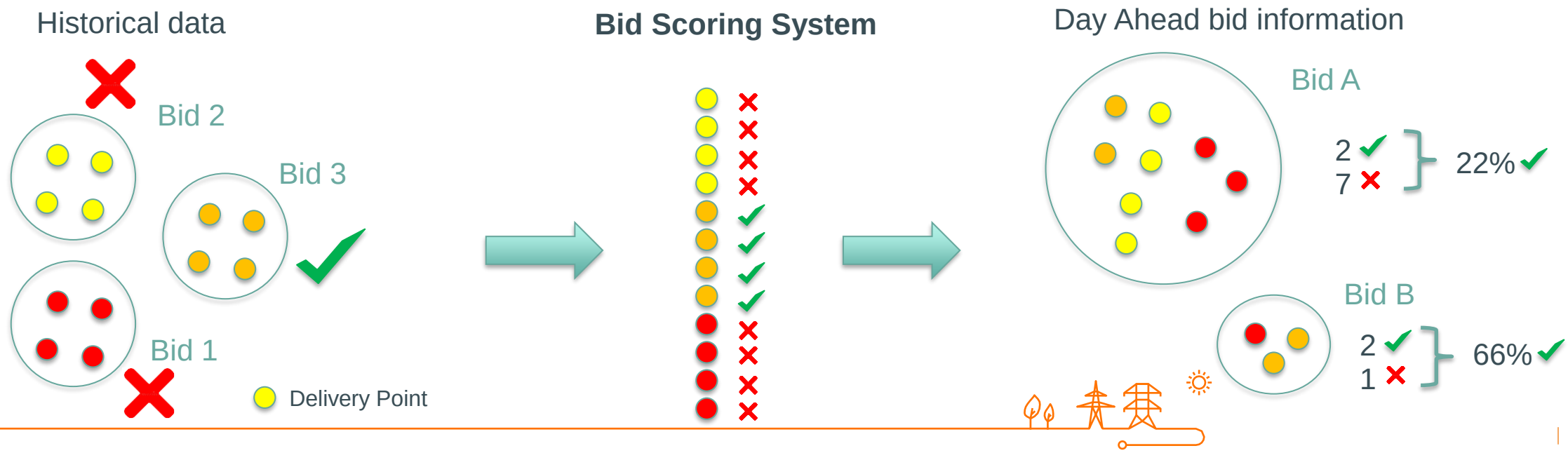
$$Score_{refMargin}(CCTU, D) = \begin{cases} 100\,, if\ for\ all\ bids\ i\ and\ all\ QHs\ included\ in\ the\ CCTU, margin[i][QH] \geq 0 \\ 100 - \frac{\sum_{bid} \sum_{QH \in CCTU} OfferedVolume[bid][QH]}{obligation(CCTU, D)}, else \end{cases}$$

Bid scoring

Bid scoring system determines which bid to select for an availability test

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	**100%**	**52**	**39**	**31**

- The Score per Bid is based on same 3 features but are adapted to the Bid Scoring System
- The results of control and test **are disaggregated on a Delivery Point level**



Bid scoring – activation control

Removal of the bid adjustment factor on the activation control and change to the Fratio:

$$Score_{Activation}(bid) = \sum_{M} F_{freshness}(M) * \left(\sum_{dp \in bid} Score_{refActivation}(dp, M) * Norm(p75, F_{ratio}(dp, M)) * Adjust(dp) \right) * Adjust(bid)$$

$$Score_{refActivation}(dp, M) = \frac{\#of\ QH\ of\ successful\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)}\ , for\ all\ Delivery\ Points\ DP\ which\ are\ "Confirmed\ DPs"$$

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M}$$

Bid scoring – Availability test

Removal of the bid adjustment factor on the Availability Test score:

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

$$Score_{refAvailability}(\mathrm{dp}, M) = \begin{cases} 100\,, if\ successful\ availability\ test \\ 0\,, if\ failed\ availabaility\ test \\ 50\,, if\ no\ availability\ test\ occured \end{cases} \quad , for\ all\ Delivery\ Points\ which\ are\ "confirmed\ DPs"$$

Bid scoring –Margin Analysis

Removal of the bid adjustment factor on the Margin Analysis :

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

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

Changes in comparison to the incentive

Changes in comparison to the incentive for bid scoring

Activation control

Bid scoring – activation control – initial formula from incentive

Changed!!!

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

$$Score_{refActivation}(dp, M) = \frac{\#of\ QH\ of\ successful\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)}\ , for\ all\ Delivery\ Points\ DP\ which\ are\ "Confirmed\ DPs"$$

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M}$$

Bid scoring – activation control – Initial context before changes

The Bid Scoring System looks at the inclusion of a Delivery Point in a bid and, whether the Delivery Point already demonstrated its contribution in satisfying obligations.

$$Score_{Activation}(bid) = \sum_{M} F_{freshness}(M) * \underbrace{Adjust(bid)}_{2} * \left(\sum_{dp \in bid} Score_{refActivation}(dp, M) * \underbrace{F_{ratio}(dp, M)}_{1} * Adjust(dp) \right)$$

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

Ratio of the successful activations versus the total number of activations

$$Score_{refActivation}(dp, M) = \frac{\#of\ QH\ of\ successful\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)}, for\ all\ Delivery\ Points\ DP\ which\ are\ "Confirmed\ DPs"$$

Bid scoring – activation control - Changes versus the incentive

1. **Portfolio bidding → use all DPs in the Business Acknowledgement message**
2. **Fratio**
3. **Adjust(bid)**

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

2

1

Activation control – Portfolio bidding

Bid scoring – activation control

The Bid Scoring System looks at the inclusion of a Delivery Point in a bid and, whether the Delivery Point already demonstrated its contribution in satisfying obligations.

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

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

$$Score_{refActivation}(dp, M) = \frac{\#of\ QH\ of\ successful\ activation\ (dp, M)}{total\ \#\ of\ QH\ of\ activation\ (dp, M)}$$

Ratio of the successful activations versus the total number of activations

$, for\ all\ Delivery\ Points\ DP\ which\ are\ "Confirmed\ DPs"$

These values cannot be determined since all assets in the bids and the supporting group can be used to comply with the request from Elia.

Change versus the incentive:

For every failure, all bids with a non-zero value in the BU ACK (so DPs in the bids and supporting providing group) will receive a “negative score”.

Activation control - Fratio

Bid scoring – activation control

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.

How often is the DP activated compared to the moments that it could have been activated

How often is the DP activated compared to the amount of QHs in the month

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M},$$

$$for\ all\ Delivery\ Points\ which\ are\ part\ of\ an\ activated\ bid$$

"$\#\ of\ QH\ of\ activation\ (dp)$" 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.

Conclusion from test runs

From the results of the initial test runs, it was shown that **these values are too small** and thus **do not allow for a distinction between the quality of the service delivery**. Therefore, a new proposal has been investigated and detailed in the next slides.

Design change for Fratio: Example

Activation control

How often is the DP activated compared to the moments that it could have been activated

How often is the DP activated compared to the amount of QHs in the month

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M},$$

$$for\ all\ Delivery\ Points\ which\ are\ part\ of\ an\ activated\ bid$$

There are 2 different DPs, A and B:

DP A
Correct activation: 100% of the time
Activation control score: **0,004**
Availability test score: 0,5
Margin analysis score: 0,95
Sum: 1,454

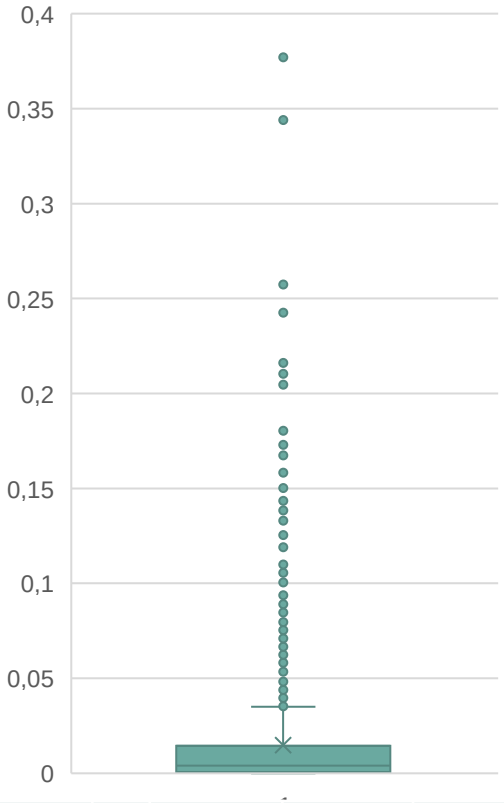
DP is reliable, had no availability test and good margin control

DP B
Correct activation: 2% of the time
Activation control score: **0,00008**
Availability test score: 0,55
Margin analysis score: 0,98
Sum: 1,53008

DP is not reliable, but had an availability test 11 months ago and okay margin control

Even though DP A is much more reliable than DP B (50 times better score), the impact of this score is negligible in comparison to the other scores

Fratio, original formula



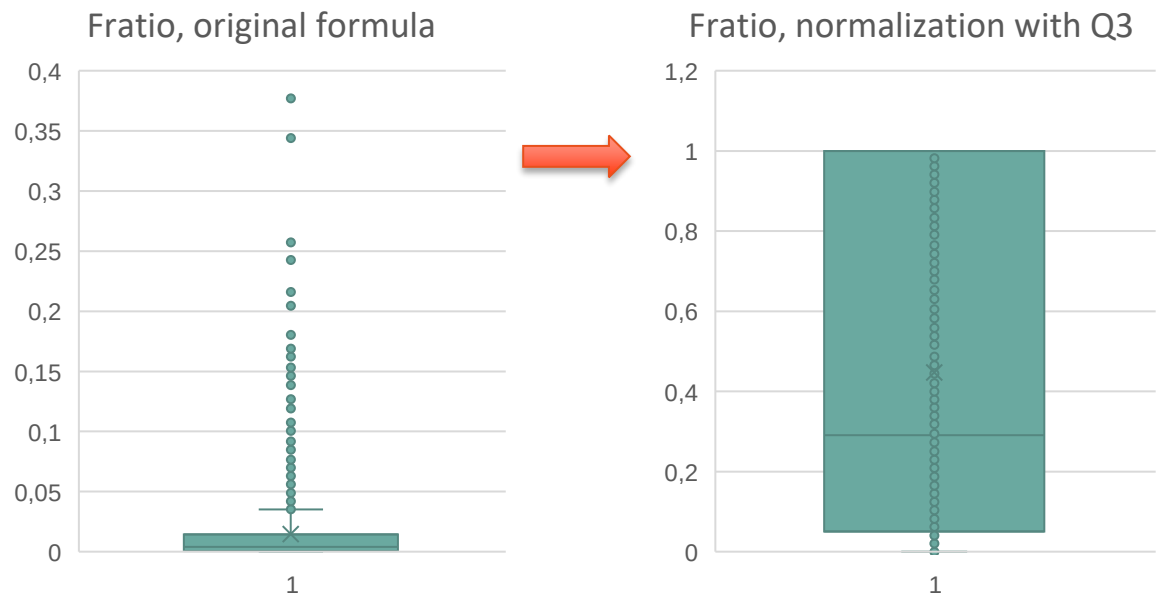
	min	Q1	median	Q3	max
Default	0	0,00061	0,0040	0,014	0,38

In this scenario, DP A has a higher chance to be tested, even though their activation performance is much better (50 times) and there is no significant difference on the other scores. In the end, the impact of the activation control with this implementation is non-existent.

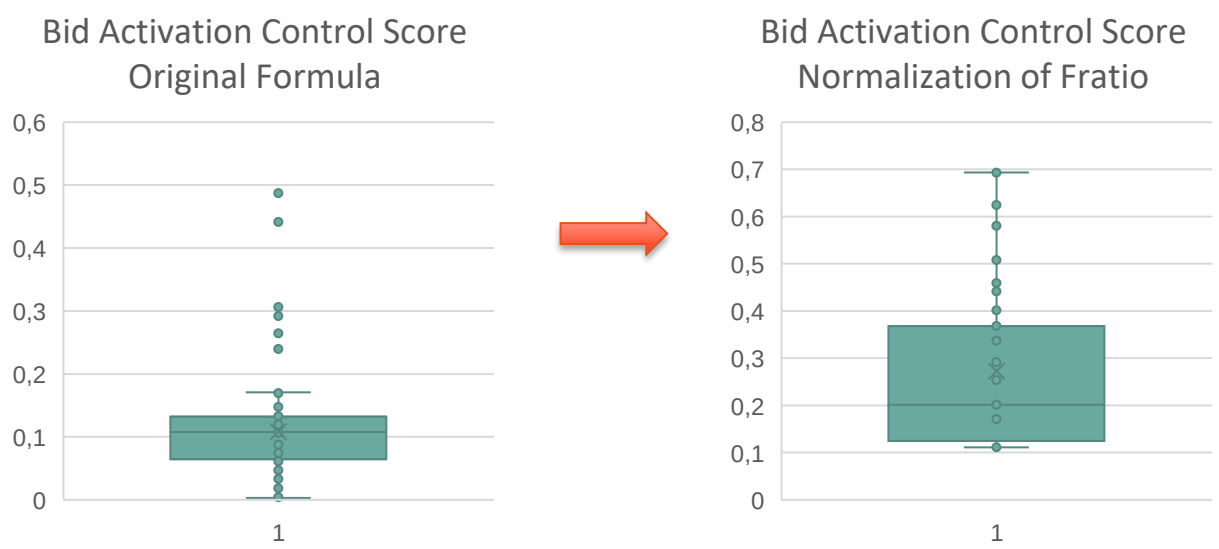
Design change: Normalization of Fratio factor

As shown in the previous example, the **Fratio is too small**, which results in an activation control score that is too small. However, assessing the quality of the information (goal of Fratio) is still important. Therefore, Elia will do **a normalization using the Q3**. More activations do not significantly increase the reliability of the information.

Fratio Normalization using the Q3 as maximum value



Result on the Activation Control score using the normalized Fratio



This means that after a DP has had a certain number of activations in a month, Elia considers the activation control information as representative for the quality of the service delivery.

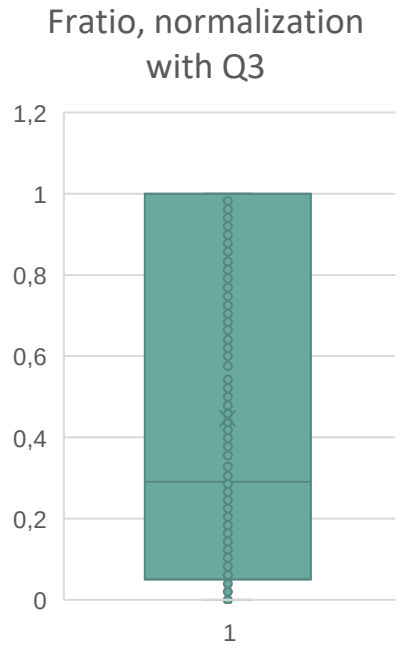
Design change for Fratio (normalization using Q3): Example

Activation control

How often is the DP activated compared to the moments that it could have been activated | **How often is the DP activated compared to the amount of QHs in the month**

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M},$$

$$for\ all\ Delivery\ Points\ which\ are\ part\ of\ an\ activated\ bid$$



	min	Q1	median	Q3	max
Default	0	0,05	0,29	1	1

There are 2 different DPs, A and B:

DP A
Correct activation: 100% of the time
Activation control score: **0,29**
Availability test score: 0,5
Margin analysis score: 0,95
Sum: 1,75

DP is reliable, had no availability test and good margin control

DP B
Correct activation: 2% of the time
Activation control score: **0,0058**
Availability test score: 0,55
Margin analysis score: 0,98
Sum: 1,5358

DP is not reliable, but had an availability test 11 months ago and okay margin control

In this case, DP B is much more likely to be tested than DP A. This better reflects also the quality of service delivery of the DPs

Activation control – Adjust Bid factor

Design change proposal: Remove the Bid adjustment factor

The Bid Adjustment factor is used to weight the offered volume of the bid in the total obligation of the BSP.

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

However, this results in the unwanted effect that smaller bids are more prone to be tested. This would mean that a BSP would be able to game the system easily by providing some smaller, very reliable bids.

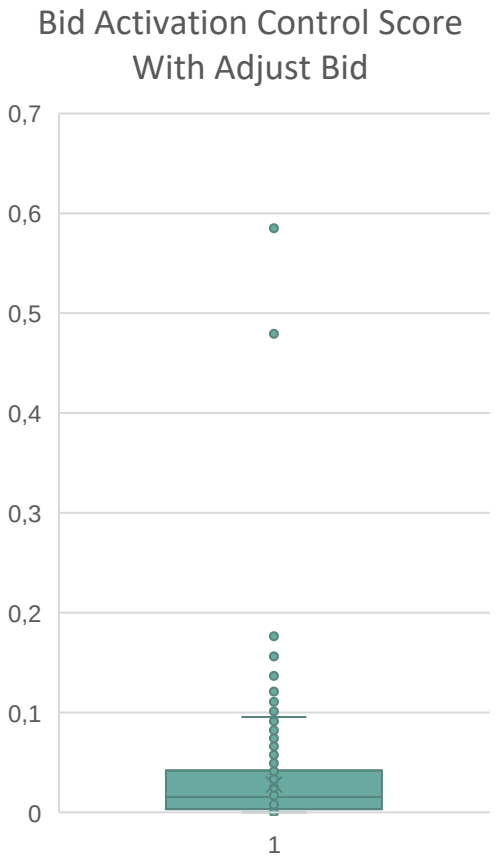
Example:

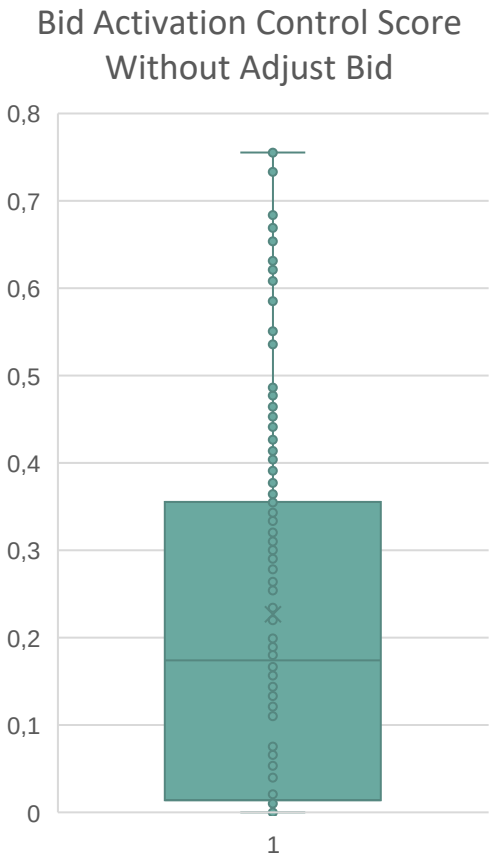
The BSP has an obligation of 100 MW.

Bid A	Bid B	Bid C
1MW Score without adjust(bid) = **0,85** adjust(bid) = 1 MW/ 100 MW = **0,01** Total score = **0,0085**	9MW Score without adjust(bid) = **0,75** adjust(bid) = 1MW / 100MW = **0,09** Total score = **0,0675**	90MW Score without adjust(bid) = **0,2** adjust(bid) = 1MW / 100MW = **0,9** Total score = **0,18**
Bid is reliable and is often activated.	Bid is slightly less reliable but still often activated.	Bid is not reliable and activated infrequently

Bid A is very likely to be tested, even though it is frequently activated and very reliable. **Bid C** on the other hand is not reliable and activated infrequently, but has a comparatively **very low chance to be tested**.

Remove the Bid adjustment factor: Results





- Without the Bid Adjustment factor, the scores behave as expected
- The results are more distributed for the Activation Control

In the formula that will be applied, we will remove the bid adjustment factor given that this factor penalizes small bids (in comparison to the total obligation).

Bid scoring – activation control – final formula

$$Score_{Activation}(bid) = \sum_{M} F_{freshness}(M) * \left(\sum_{dp \in bid} Score_{refActivation}(dp, M) * Norm(p75, F_{ratio}(dp, M)) * Adjust(dp) \right) * Adjust(bid)$$

$$Score_{refActivation}(dp, M) = \frac{\#of\ QH\ of\ successful\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)}\ , for\ all\ Delivery\ Points\ DP\ which\ are\ "Confirmed\ DPs"$$

$$F_{ratio}(dp, M) = \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ of\ activation\ (dp)} * \frac{\#\ of\ QH\ of\ activation\ (dp)}{total\ \#\ of\ QH\ in\ month\ M}$$

Margin analysis / availability testing – Adjust Bid factor

Bid scoring – availability test

Changed!!!

The general formula is as follows:

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

With the values for the availability test:

$$Score_{refAvailability}(\mathrm{dp}, M) = \begin{cases} 100\,, if\ successful\ availability\ test \\ 0\,, if\ failed\ availabaility\ test \\ 50\,, if\ no\ availability\ test\ occured \end{cases} \quad , for\ all\ Delivery\ Points\ which\ are\ "confirmed\ DPs"$$

Bid scoring – margin control

Changed!!!

The Margin Analysis Score of the bid is based on the score of each individual Delivery Point. The score of the Delivery Point excludes periods of activation control and availability tests in order to avoid an overlap of information.

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

Where:

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

Design change proposal: Remove the Bid adjustment factor

Availability testing and margin control

The Bid Adjustment factor is used to weight the offered volume of the bid in the total obligation of the BSP.

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

However, this results in the unwanted effect that smaller bids are more prone to be tested. This would mean that a BSP would be able to game the system easily by providing some smaller, very reliable bids.

Example:

The BSP has an obligation of 100 MW.

Bid A
1MW
Score without adjust(bid) = **0,85**
adjust(bid) = 1 MW/ 100 MW = **0,01**
Total score = **0,0085**

Bid is reliable and is often activated.

Bid B
9MW
Score without adjust(bid) = **0,75**
adjust(bid) = 1MW / 100MW = **0,09**
Total score = **0,0675**

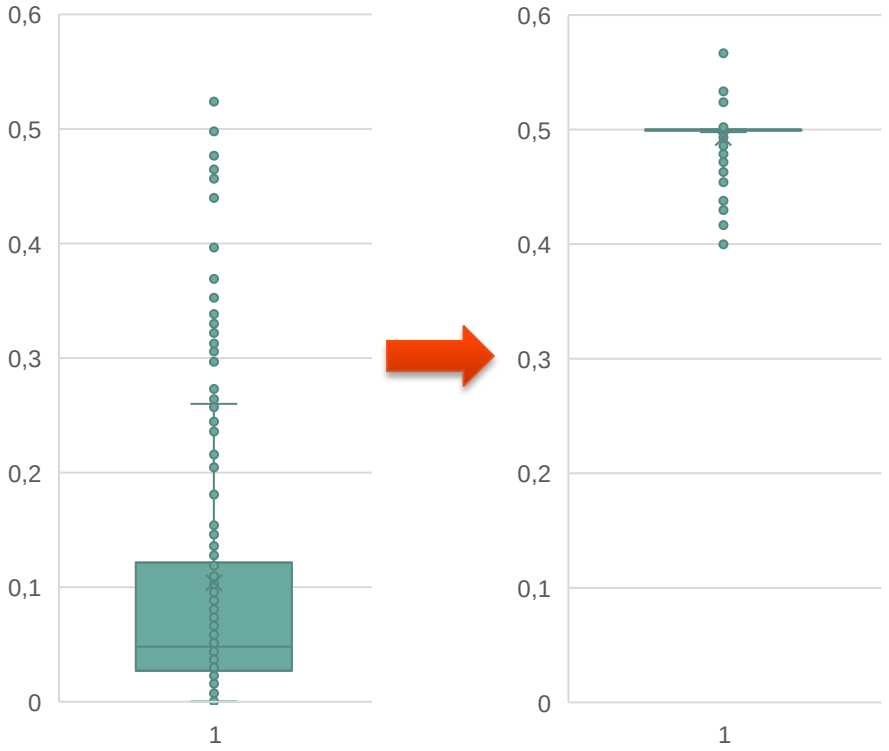
Bid is slightly less reliable but still often activated.

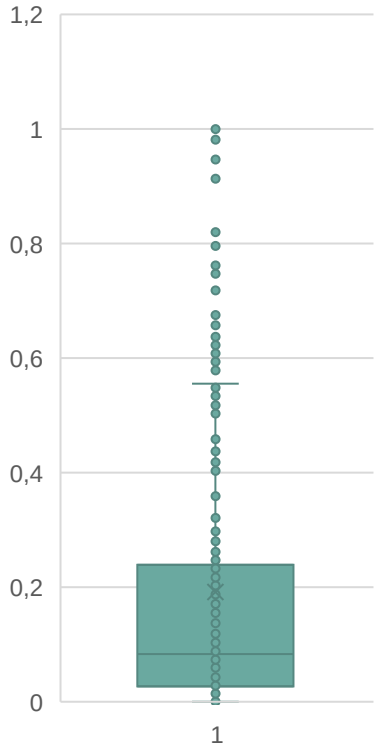
Bid C
90MW
Score without adjust(bid) = **0,2**
adjust(bid) = 1MW / 100MW = **0,9**
Total score = **0,18**

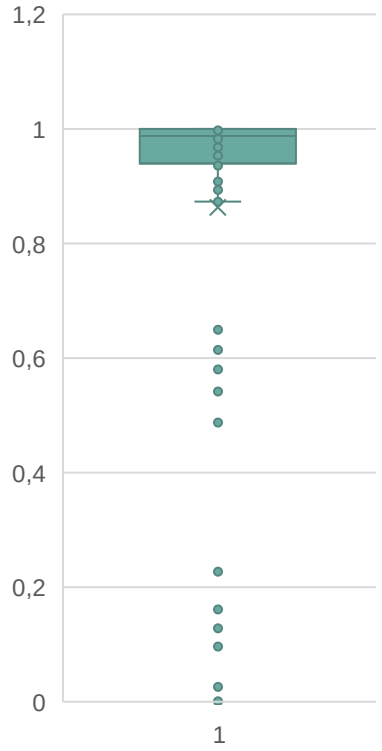
Bid is not reliable and activated infrequently

Bid A is very likely to be tested, even though it is frequently activated and very reliable. **Bid C** on the other hand is not reliable and activated infrequently, but has a comparatively **very low chance to be tested**.

Remove the Bid adjustment factor: Results







- Without the Bid Adjustment factor, the scores behaves as expected
- The results are more distributed for the Activation Control
- The availability test score is concentrated around 0.5 as a lot of units are not tested

Similarly to the activation control score, the values are more in line with the expectation (0,5 for the availability test and 1 for the margin control)

Bid scoring – Availability test and Margin Analysis

Removal of the bid adjustment factor on both the Availability Test score and Margin Analysis :

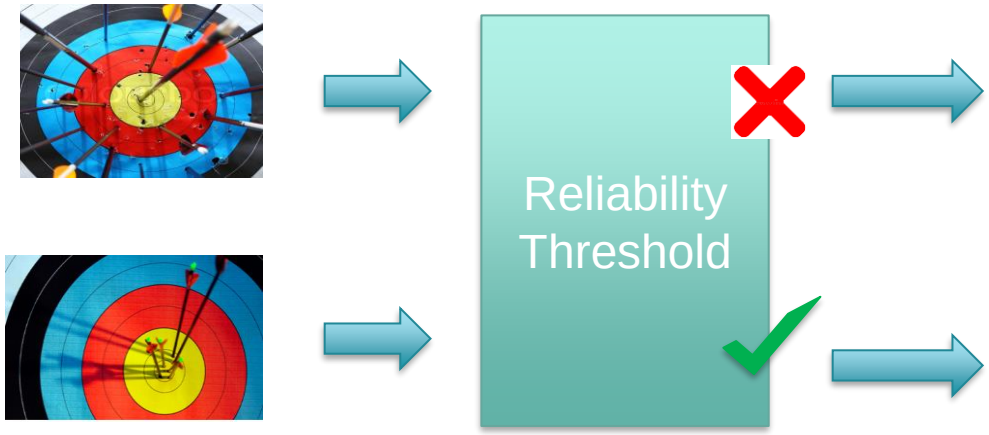
$$Score_{refAvailability}(bid) = \sum_{M} F_{freshness}(M) * Adjust(bid) * \left(\sum_{dp \in bid} Score_{refAvailability}(dp, M) * Adjust(dp) \right)$$

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

Test regimes

Test regimes

- Additionally to the scoring system, **two test regimes** are introduced to limit the impact (in volume) of availability tests.
    1. The first test regime **aims to ensure** that a significant part of **the contracted capacities** from a BSP **is compliant.**
    2. The second test regime aims **to keep in check the compliancy** of a BSP but with a **lower volume of availability tests**



Test Regime 1: demonstrate reliability in provision of contracted capacity

$$\boldsymbol{If}: \ valid\ activated\ volume < threshold$$
$$BSP\ is\ in\ \boldsymbol{test\ regime\ 1}$$

Test Regime 2: reduced volume of test (**maximum 4 tests**)

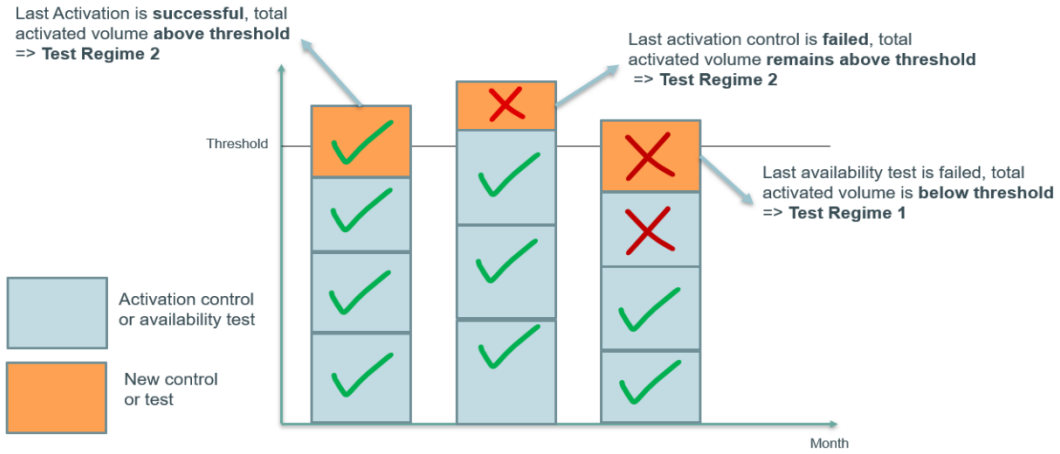
$$\boldsymbol{If}: valid\ activated\ volume \geq threshold$$
$$BSP\ is\ in\ \boldsymbol{test\ regime\ 2}$$

Threshold & valid activated volume

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

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

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.



Test Regimes – updated formula

– **Threshold for each BSP** that will determine the test regimes :

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

Average Obligation of the BSP in the last 12 months

– **Valid Activated Volume**: for each BSP, sum of the valid activated volume of each DP of its portfolio.
At **DP level**, consist of the **maximum activated volume** since the **last failed activation control**.
When control is **failed**, the volume **returns to 0**.

If the current maximum volume was measured more than 12 months ago, a new maximum is defined based on the maximum in the last 12 months

For all mFRR activations:
    For all DP in BuAck:
        if control is successful:
            $ValidVolume(DP) = \max\left(mFRR\ supplied(DP), ValidVolume(DP)\right)$
            if Freshness($ValidVolume$)>12 months:
                $ValidVolume(DP) = \max_{12<M}\left(mFRR\ supplied(DP)\right)$
        if control is failed:
            $ValidVolume(DP) = 0$

Test regimes – number of tests during rolling 12 months

Alternative scoring implemented

Scoring as defined in incentive

The scoring as defined in the incentive is as follows:

> We look at the rolling 12 months (in the past) and during this period we can only test as defined in the test regime (so **12 tests** in test regime 1 and **4 tests** in test regime 2)

However, this means that when we transition from test regime 1 to test regime 2, it is possible that for an extended period of time we cannot perform a test **(see example next slide).**

Alternative scoring

For the alternative scoring, we would give a value for an executed test in test regime 1 and a different (larger) value for a test executed in test regime 2:

> A test executed in **test regime 1 counts for 1 executed test**. A test executed in **test regime 2 counts for 3 executed tests** (12 / 4 = 3, max tests / number of tests in test regime 2).

Like in the other scoring method, we would sum up the values of the rolling 12 months and make sure that this value is always lower than or equal to 12 **(see example next slide).** This resolves the issue following from the scoring as defined in the incentive.

Test regimes – number of tests during rolling 12 months

Alternative scoring implemented

	Month	Jan-25	Feb-25	Mar-25	Apr-25	May-25	Jun-25	Jul-25	Aug-25	Sep-25	Oct-25	Nov-25	Dec-25	Jan-26	Feb-26	Mar-26	Apr-26	May-26	Jun-26	Jul-26	Aug-26	Sep-26	Oct-26	Nov-26	Dec-26
12 per year	Test execution normal (1)	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x	x
4 per year	Test execution normal (2)	x			x			x			x			x			x			x			x		
Rolling 12 months	Test executed normal (1)	1	2	3	4	5	6	7	8	9	10	11	12	12	12	12	12	12	12	12	12	12	12	12	12
	Test executed normal (2)	1	1	1	2	2	2	3	3	3	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
	From 1 --> 2																								
Rolling 12 months	Test executed normal (1) --> (2)	x	x	x	x	x	x									x			x			x			x
		1	2	3	4	5	6	6	6	6	6	6	6	5	4	4	4	4	4	4	4	4	4	4	4
		+1	*+1*	*+1*	*+1*	*+1*	*+1*									*+1*			*+1*			*+1*			*+1*
Rolling 12 months	Test executed alternative scoring (1) --> (2)	x	x	x	x	x	x			x			x			x			x			x			x
		1	2	3	4	5	6			9			12			12			12			12			12
		+1	*+1*	*+1*	*+1*	*+1*	*+1*			*+3*			*+3*			*+3*			*+3*			*+3*			*+3*

When we pass from test regime 1 to test regime 2, the number of tests that we can execute reduces. If we just consider the test executed in the previous 12 months, there is a gap of 8 months where we cannot test the BSP. Therefore, alternative scoring was implemented.

Summary and example

Smart Testing methodology

Summary and example

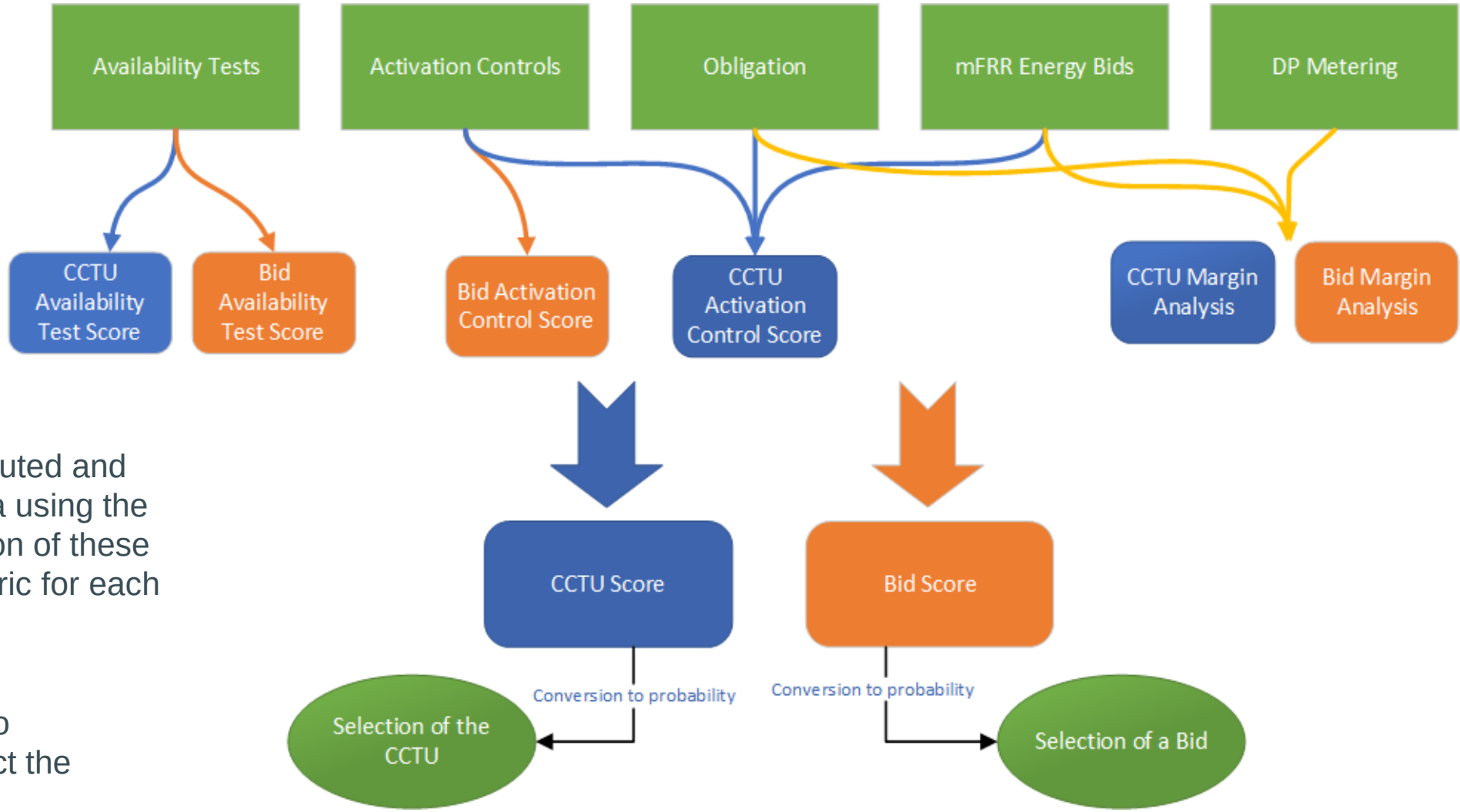
Bids & CCTU Scores : computation

The scores are computed on the data from **M-13** to **M-2**

For each month, **6 sub-scores** are calculated: 3 for the evaluation of the **CCTU** for each BSP and 3 for the evaluation of the **bids** based on their delivery points.

These monthly scores are computed and weighted to prioritize recent data using the freshness factor. The combination of these sub-scores then results in a metric for each Bid and CCTU to test.

The final scores are converted to probability to unpredictably select the CCTU and the Bid



Calculation of CCTU Score

Activation Control Score:

Weighting factor for the maximum delivered volume wrt the average obligation

Highest failure And Total number of failures

$$Score_{Activation}(CCTU) = \sum_{M=2}^{13} \left(Score_{refActivation}(CCTU, M) * F_{failure}(CCTU, M) * F_{freshness}(M) \right)$$

Measure the performance of past Activations

Availability Control Score:

Availability Test Status

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

Target Failed / Untested CCTU for Availability Tests

Margin Analysis Score:

Volumes of the bids with negative margin compared to the obligation

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

Monitor capacity available for contracted bids outside of activations periods

Illustration of CCTU Score calculation

The calculation of the different corposants of the CCTU score can be illustrated as bellow **for a given month**:

- Activation Control Score:
  - $Score_{refActivation}(CCTU, M) * F_{failure}(CCTU, M) * F_{freshness}(M) = \frac{30}{173} * \left(1 - \frac{60}{173}\right) * \left(1 - \frac{50}{512}\right) * \frac{4}{30} = 0{,}101 * 0{,}133$

$Freshness$

$\frac{Max\ requested\ volume}{average\ Obligation}$ $\frac{Max\ failed\ bid\ volume}{average\ Obligation}$ $\frac{\#QH\ of\ failed\ Activations}{\#QH\ of\ Activations}$

- Availability Test Score:
  - $Score_{refAvailability}(CCTU, M) * F_{freshness}(M) = 1 * 0{,}133$

$Successful\ Availability\ Test$

- Margin Analysis Score:
  - $\sum_{D \in M} \frac{Score_{refMargin}(CCTU, M)}{\#days} * F_{freshness}(M)$ = 0,8 * 0,133

For one day :

$$\frac{volume\ of\ bids\ with\ negative\ margin}{CCTU\ Obligation} = \frac{40}{130} = 0{,}69$$

Calculation of Bid Score

Activation Control Score:

Ratio of the successful activations versus the total number of activations

DeliveryPoint Adjustment Factor

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

Activation Frequency of the asset

Measure the performance of past Activations

Availability Control Score:

Availability Test Status

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

Target Failed / Untested CCTU for Availability Tests

Availability Control Score:

Margin of the bids in which the asset participates

$$Score_{margin}(bid) = \sum_{M} \left(F_{freshness}(M) * \left(\sum_{dp \in bid} \sum_{qh \in M} \frac{\{1 \; if \; margin \geq 0; 0 \; otherwise\}}{\#qh} * Adjsust(dp) \right) \right)$$

Monitor capacity available for contracted bids outside of activations periods

Illustration of Bid Score calculation

The calculation of the different corposants of the Bid score can be illustrated as bellow **for a given month**:

- Delivery Point Adjustment factor:

$$Adjust(dp) = \frac{DP_{contribution}(dp)}{\sum_{dp \in bid} DP_{contribution}(dp)} = \frac{2}{12} = 0{,}17$$

$$Fratio = \frac{QH\ Activations\ Delivered}{total\ QH\ activations} * \frac{QH\ Activations\ Delivered}{total\ QH\ in\ month} = \frac{114}{119} * \frac{114}{2976} = 0{,}036$$

Normalization of Fratio

- Activation Control Score *(detail for one delivery point)*:
  - $Score_{refActivation}(dp, M) * F_{ratio}(dp, M) * Adjust(dp) = \frac{98}{119} * 0{,}93 * 0{,}17$

$$Score_{refActivation} = \frac{QH\ successful\ activations}{total\ QH\ activations}$$

Results of availability tests per DP

- Availability Test Score:
  - $\sum_{dp \in bid} Score_{refAvailability}(dp, M) * Adjust(dp) * F_{freshness}(M) = \left[\left(\frac{2}{12} * 0\right) + \left(\frac{6}{12} * 0{,}5\right) + \left(\frac{4}{12} * 1\right)\right] * 0{,}133 = 0{,}58 * 0{,}133$

- Margin Analysis Score *(detail for one delivery point)*:
  - $\sum_{qh \in M} \frac{\{1\ if\ margin \geq 0; 0\ otherwise\}}{\#qh} = \frac{2678}{2976}$

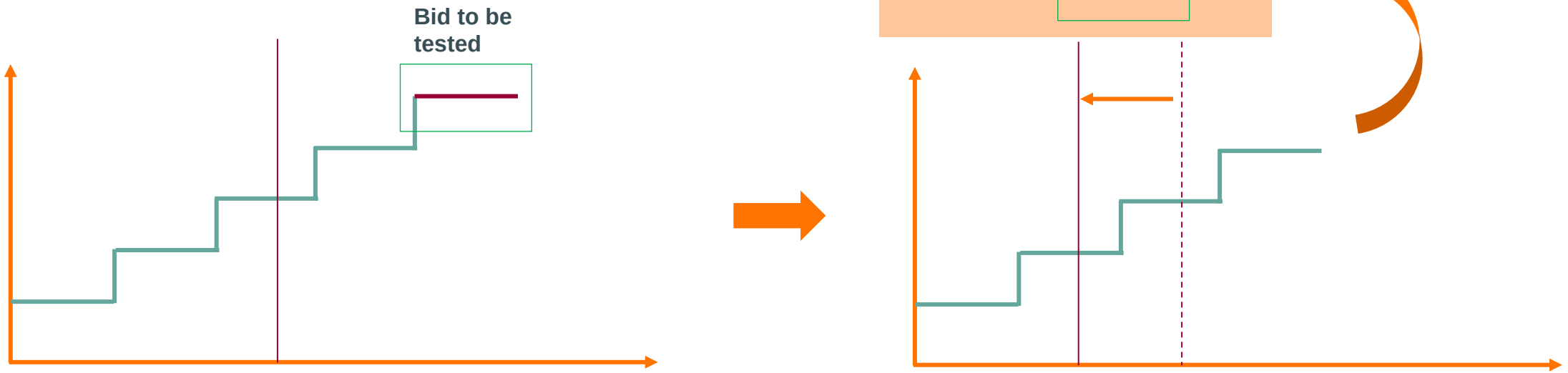
$$\frac{QH\ with\ sufficient\ margin}{total\ QH\ in\ month}$$

Availability testing in the market

Summary

Price setting of the bid to be tested (option 0)

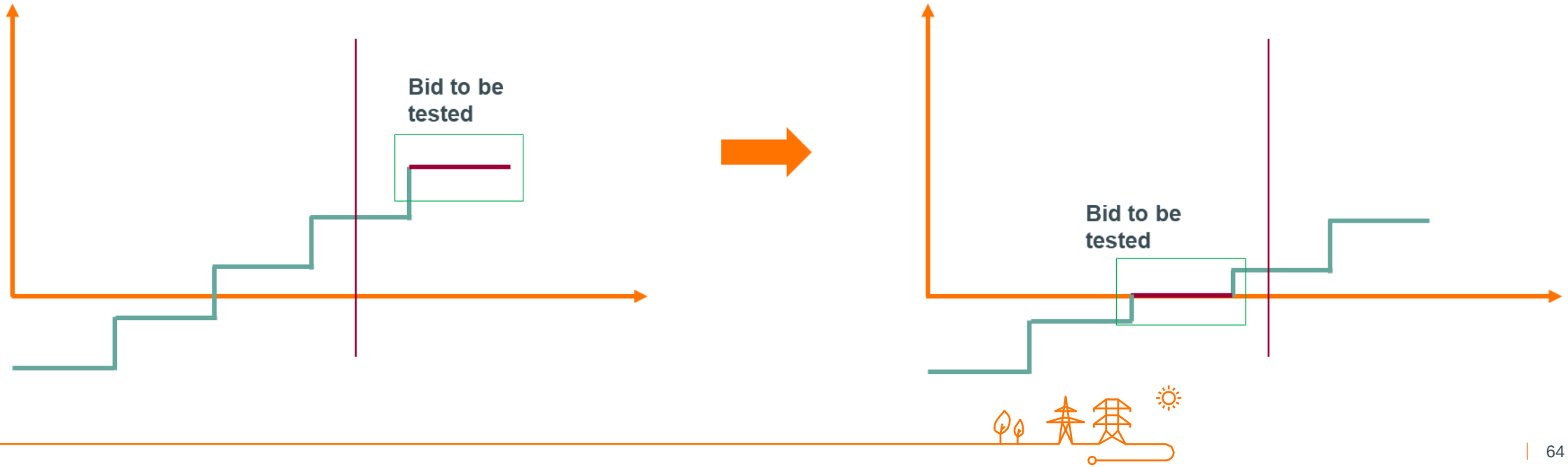
Example



The bid to be tested is taken out of the merit order and activated out of market. The total imbalance is changed based on this activation (in this example reducing the imbalance).

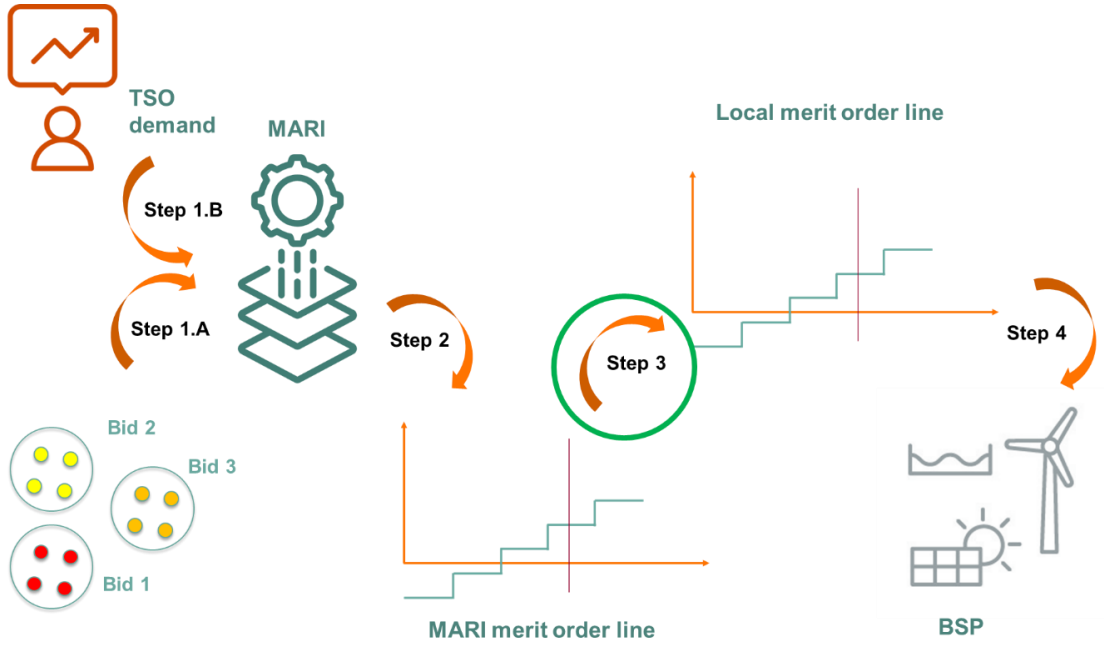
Availability testing in the market : Integration in the MARI merit order (option 1)

Instead of activating a bid for an availability test and then compensating the volume with an aFRR/mFRR activation (if needed), we **move the position of the to be tested bid towards the beginning of the merit order** (and thus modifying the price of the bid). This way the bid remains in the market, but at a different price. In case the bid is activated for an availability test, it is remunerated at the CBMP.



Availability testing in the market : Integration in the local merit order (option 2)

The way to achieve it slightly differs in comparison with the first option. Instead of changing the position of the bid before sending the information towards the Mari platform, Elia would **integrate the bid in the local merit order**. This avoids the issue with the current legal framework to allow for Elia to modify the price of a bid, but introduces new legal challenges. In addition, additional operational actions need to be performed in a very short timeframe which leads to a complex implementation.



Availability testing in the market : Modifying TSO demand (option 3)

In this option Elia would perform a netting between the TSO demand and the bid on which Elia would like to perform an availability test. If this netting is not possible, the test will be cancelled.

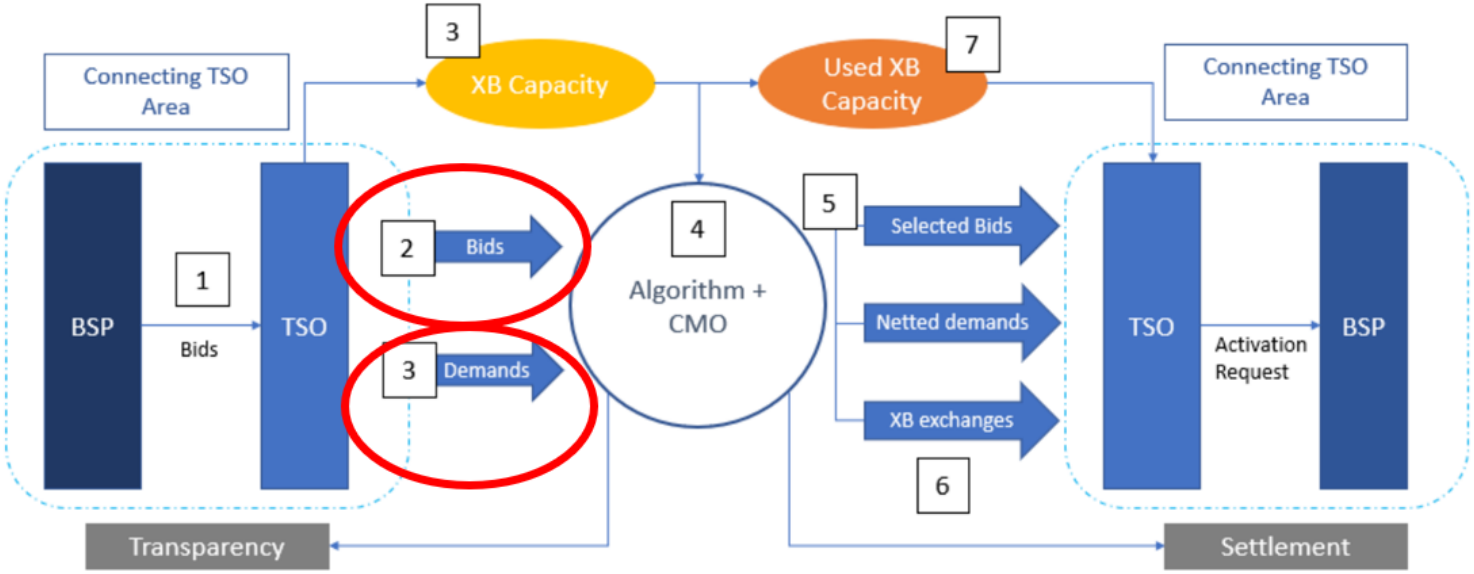


Figure 9:Area of modification to the MARI process in option 3

Conclusion

Conclusion

1. Doing availability tests in the market has some advantages in comparison to doing availability testing out of the market
    1. BSPs can be remunerated at CBMP
    2. The imbalance can only decrease when executing an availability test
2. There is the downside that the execution of an availability test is not always feasible.
3. **<u>However,</u>** depending on the implementation there are additional downside(s):
    1. Operational implementation is difficult
    2. Legal framework needs to be modified

	Ease of implementation	Current legal framework	Brings desired improvements	Price setting of the bid
Option 0	++	++	--	/
Option 1	++	--	++	+
Option 2	--	+	++	++
Option 3	++	++	-	+

Feedback market parties

Availability testing in the market : Feedback Febeliec

1. Febeliec mentions that it is hesitant about introducing a remuneration for availability testing in order to avoid increased costs for the Grid Users. However, they are in favor of reducing the impact on the system imbalance.

   RESPONSE ELIA: The costs would not increase in all of the options because the availability test replaces an mFRR activation. So, there is a shift in remuneration but no increased costs.

2. The proposals from Elia seem optimized towards the BSPs, but not towards the reduction of the system costs. This additional element should be analysed before implementation.

   RESPONSE ELIA: See previous comment.

Availability testing in the market : Feedback FEBEG

1. FEBEG appreciates the efforts from Elia, since the introduction of these costs into the capacity bids is complex.

   RESPONSE ELIA: Elia thanks FEBEG for its feedback.

2. FEBEG believes that Elia can anticipate the evolution of the system imbalance and thus can trigger an availability test when this does not aggravate the system imbalance.

   RESPONSE ELIA: Elia has indeed some capabilities to anticipate the evolution of the system imbalance. However, there is a process behind the selection of the availability tests that takes more time and Elia has no view on the system imbalance at that time. In addition, this could reduce the unpredictability of availability tests

3. Problems with Option 1:

   1. There is still a difference between the CBMP and the bid price. This still needs to be taken into account in the capacity bids.

      RESPONSE ELIA: This gap is indeed still present. However, remunerating the full bid price is not the goal, since this will lead to unwanted incentives.

   2. FEBEG has reservations on modifying bid prices .

      RESPONSE ELIA: In case this option would be chosen, a clear framework would be introduced. This would only be allowable for the execution of availability tests.

Availability testing in the market : Feedback FEBEG

4. Problems with Option 2:
    1. Creation of paradoxically rejected bids

       RESPONSE ELIA: Elia agrees with this point.

    2. The activation of bids not present in the MARI merit order would lead to additional unclarities for the BSP.

       RESPONSE ELIA: Elia indeed considers that this could be detrimental for the transparency in the market which creates additional questions for the BSP.
5. FEBEG proposes option 0 with slight modifications:
    1. Availability testing outside of MARI but with a remuneration equal to the energy bid price

       RESPONSE ELIA: There are 2 issues with this proposition:

        1. Remuneration at bid price creates an incentive to put high bid prices to receive this remuneration when an availability test is executed (unwanted incentive).
        2. A netting with the TSO demand would need to be performed to be able to activate the bid. This leads to inefficiencies when the overall TSO demand is in the other direction.

Regulatory framework

Analysis of changes to the regulatory framework

Topics

The goal of today is to present the elements that can have an impact on the methodology of smart testing that was defined in the incentive from 2020. Before going into the analysis, a short context is given regarding the smart testing methodology:

1. Context smart testing
2. Feedback on the public consultation
3. Potential impacts linked to the mFRR design evolutions
4. Potential impacts linked to the aFRR design evolutions
5. Potential impacts linked to the incentive on PQ & Penalties

Consultation Report
Smart Testing Methodology

Stakeholders feedbacks

Scoring System - Margin Analysis

Flexcity

For sites which use **'high X of Y' baselining the margin score might not be very suitable**. A negative margin in one QH for a site does not mean that, if the site would have been activated in that quarter hour, the site would not have been able to meet the requirements as put forth in the terms and conditions for mFRR.

Flexcity understands the relevance of the different scores (Activation Control, Availability Test & Margin Analyses). However, due to the **complexity** of the formulas, the absence of the weights and the unclarity on the relationship between low scores and the triggering of a test it is very difficult for Flexcity to assess what would be the consequences of this smart testing logic and **whether the derived scores would be a good representation of the reliability of the service and/or a good indication of the need to test** a CCTU or bid.

Therefore we would like to request ELIA to Remain transparent throughout the further process meaning, amongst other things, to give insight in the determination of the weights.

Elia response:

Elia agrees with the stakeholder on the possible impact of the baselining on the $Score_{refMargin}$*. For the sake of simplicity, Elia proposes to not consider such detail for which the added value is questionable. Elia reminds that all scores are designed to provide an indication to Elia on whether to test certain bid(s) or CCTU. It does not impact the success or failure of an activation control. In this case, the indication may be slightly less accurate than if the choice of baselining was taken into account.*

Elia may consider amendments after a return of experience or based on further clarification from the stakeholder on their concerns.

Elia response:

The weights for the scoring systems are subject to ***fine-tuning*** *in the implementation phase and will be made available.*

With regards to the triggering of a test, ***this remains at the discretion of Elia*** *as it is today. Elia does not intend to disclose to the BSP when a test will be performed, nor to let the BSP determine with certainty when it will take place (nor on which bid(s)). Smart Testing does not change this principle and it does not affect the BSP in its obligations.*

Smart Testing only provides additional information to Elia on the selection of the CCTU and the bid(s) to be tested, to give Elia a sufficient comfort on the availability of the bids while reducing the number of tests.

Scoring System - Availability test

As regards the availability test, why a **score of 50** is attributed to the Score ref Availability (CCTU, M) **if no availability test occurred**? What could be the impact on the final score especially for the **CCTU's which are rarely requested for tests** (20:00-00:00h; 00:00-4:00; 4:00-8:00)?

FEBEG

Elia response:

*Regarding the scoring system for availability test, a score of 50 has been chosen 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. The **weights are then used to calibrate** and achieve a balanced effect of each component on the final score.*

The impact of this number will also be seen during the calibration phase and possible amendments based on the return of experience will be proposed during a presentation and integrated.

Scoring System - Margin Analysis

From the supplied materials it does not seem clear how ELIA is planning to identify the Unsheddable Margin (UM). Which period of time will be used to determine UM? Will it be based on the lowest quarter hour consumption or lowest average consumption over a certain time ?

Flexcity

Elia response:

*The Unsheddable Margin (UM) is based on the lowest offtake (consumption) value (lowest quarter hour consumption in case of mFRR and lower granularity for aFRR and FCR) for the considered 12 months rolling window. Elia is aware the **underlying hypothesis regarding maintenance**, which drops the UM to zero consumption. The calculation of the UM may be improved with later phases of iCAROS project with the data on outage planning.*

Scoring System - Margin Analysis

The margin analysis, as described in the note, seems only applicable for mFRR, but not for FCR nor aFRR (symmetrical or down). How is the score computed when a **DP is part of bid that is continuously activated** ?

FEBEG

Elia response:

For downward product, the reference to be used for a generation unit will be the Pmin instead of Pmax. ***For DSM, the maximum measured off-take can be taken as a proxy to calculate the margin****.*

Based on the current designs and available inputs, Elia believes that the margin analysis scoring may be computed for aFRR and FCR, in line with the proposed methodology. The implementation details will be sorted out during the implementation phase of the relevant product.
The final report will contain these additional clarifications.

With Margin Analysis it is very difficult to be **technology neutral** between **Demand Side Management** technology and 'traditional' suppliers of flexibility. There will never be a Negative Margin for the mFRR flexibility delivered by stand-by thermal plants (OCGT operated gas fired power plants, Turbojets, large diesel generators). However it is well known that these plants do have an important 'Forced Outage Rate' and corresponding statistical failure risk at start-up. In this set-up a 95% reliable standby plant will have better scores then a 95% reliable DSM profile.

Elia response:

Smart Testing is technology neutral. However, based upon objective data, the ***methodology may naturally yield score results which may be technology dependent.***
Please note that this should not impact the maximum number and volume of tests that will be performed.

Looking at this from the perspective of an availability test, an asset that does not have sufficient margin available would also have failed the availability test.

Scoring System - Margin Analysis

From the supplied materials it is not clear to Flexcity how the margin score for a CCTU would be determined based on the Margin QH's of Annex 2. Is one quarter hour with a negative margin in a bid enough to consider the CCTU has a negative margin?

Flexcity

Elia response:

Elia confirms the understanding of the stakeholder. If during one quarter hour a negative margin is identified, the Scoremargin of the CCTU is negatively impacted. Contracted capacity should be available at any time.
Elia will clarify this point in the final report.

Scoring System - Activation Control

Concerning the 2 scoring systems, FEBEG agrees with the general principles but expresses its reservation on their concrete application as the note is not fully clear on the calculation methods:

For the **Failure Factor**, "an activation control is considered failed as defined in the T&C of the relevant product" : **this concept is not defined for aFRR.**

Elia response:

On applicability of the Scoring System for aFRR, Elia agrees that success or failure in aFRR activation control 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.

mFRR Contract Evolution & impacts

mFRR Contract November 2023
Changes for connection to MARI platform

mFRR Contract Evolution & impacts

1.	Energy Bidding	→ No impact on Smart testing methodology
2.	Bids selection	→ No impact on Smart testing methodology
3.	Activation	→ No impact on Smart testing methodology
4.	Remuneration	→ No impact on Smart testing methodology
5.	Activation control & penalties	→ **Indirect impact on Smart testing methodology**
6.	CRI Impacts	→ No impact on Smart testing methodology
7.	Penalty for Contracted Bids	→ **Impact on Smart testing methodology**
8.	Update of Bids after BE GCT & Baselines after RDGC	→ **Indirect impact on Smart testing methodology**

Amendments to the T&C aFRR

Terms and Conditions for balancing service providers for aFRR

Only aFRR

aFRR Contract Evolution & impacts

1. Real-time baseline → **Indirect impact on Smart testing methodology**
2. 5’ FAT (Full Activation Time) → **Indirect impact on Smart testing methodology**
3. Move aFRR capacity auction to D-1 → No impact on Smart testing methodology
4. Incentive 2022: activation method → **Impact on Smart testing methodology**
5. CCMD: ind. correction model, opening LV → No impact on Smart testing methodology
6. Connection to aFRR-Platform including the mitigation measures: → No impact on Smart testing methodology
   a) Maintain bid price cap for contracted aFRR Energy Bids
   b) Elastic aFRR demand
   c) Alternative calculation aFRR CBMP based on the global control target

Incentive on Prequalification and penalties

Incentive on Prequalification, Control, and Penalties for the aFRR and mFRR Services

Only aFRR

Incentive PQ & penalties Contract Evolution & impacts

1. Onboarding & prequalification → No impact on Smart testing methodology
2. Penalty MW made available → No impact on Smart testing methodology
3. Activation control aFRR → No impact on Smart testing methodology
4. Baseline aFRR → **Indirect impact on Smart testing methodology**