REPORT – PUBLICATION CONSULTATION

Forecasting of the "Deterministic Frequency Deviation" (DFD) and the ELIA contribution

September 1st, 2023

TABLE OF CONTENTS

TABLE OF FIGURES

1. Introduction

1.1 Context

Deterministic Frequency Deviations (DFD)[1] are phenomena which occur on a regular basis as a result of (quarter-) hourly load and production differences. With respect to the criteria presented in ENTSO-e's report (entso-e, 2019), a frequency deviation is labelled as a DFD if the variation in frequency exceeds 75 mHz for more than 20 successive seconds during a change of Market Time Unit.

As a result of a large amount of DFD events in 2019, a Task Force was approved by ENTSO-e, which concluded that common DFD targets should be established in the Synchronous Area to limit the contribution of each LFC block. The report of ENTSO-e (entso-e, 2019) also mentions that TSO's, which didn't implement any mitigation measures while the number of DFD's remain high according to the established targets, would have to acquire additional reserves as default solution (equivalent to a penalty).

Following these conclusions, Elia made a first analysis of its contribution to DFD's and drafted a report called "Report on Deterministic Frequency Deviations: Lowering the contribution of the Belgian Control Block" (ELIA, 2020). Among a dozen of possible mitigation measures, the retained solutions to be considered for the Belgian control area, were, in the following order:

1. Moving towards 15' cross-border trading products on the intraday market
2. Activate mFRR and/or tune the output of the LFC controller at the change of MTU based on prediction algorithms
3. Discuss with owners of fast acting units to possibly spread the starting and stopping of these units over a longer period.

The first solution has already been implemented in December 2020 on the bidding zone borders between Belgium, Germany, and the Netherlands. In 2023, 15' Operational Time Unit (OTU) has also been released which gives a possibility to exchange 15min cross-border products in intraday.

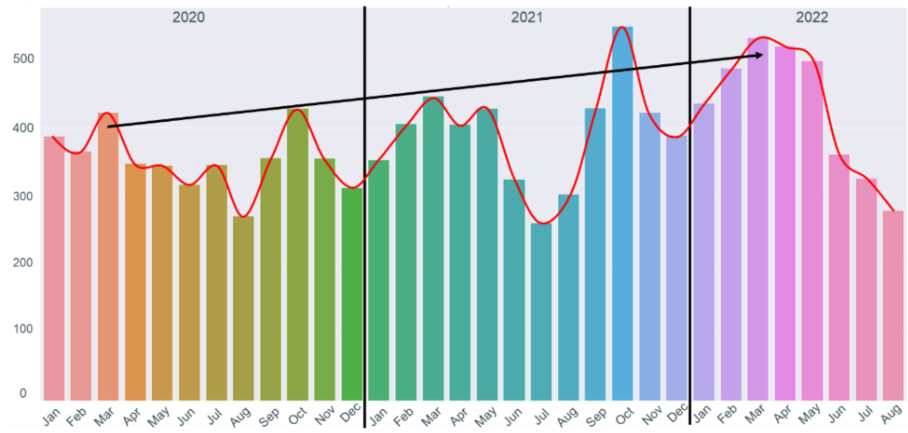


Figure 1 - Evolution of the number of DFD's in the Synchronous Area

In the meantime, the situation regarding DFD's and Elia's contribution is getting worse. The number of DFD's is increasing as from 2020 (Figure 1) and Elia is exceeding its limited contribution established in Article C-9 of the Continental Europe Synchronous Area Framework Agreement (SAFA) (entso-e, s.d.). In fact, the SAFA fixed Area Control Error (ACE) limitations during DFD events to all the TSO's from the Synchronous Area. Additionally, it is mentioned that TSO's shall not violate their own ACE contribution threshold for more than 30% of the DFD events happening in a quarter of a year. Elia

[1] Deterministic Frequency Deviation: See definition in Chapter 2.1

exceeded this threshold for 3 quarters in 2022 and is staying close to it since then (Figure 2).

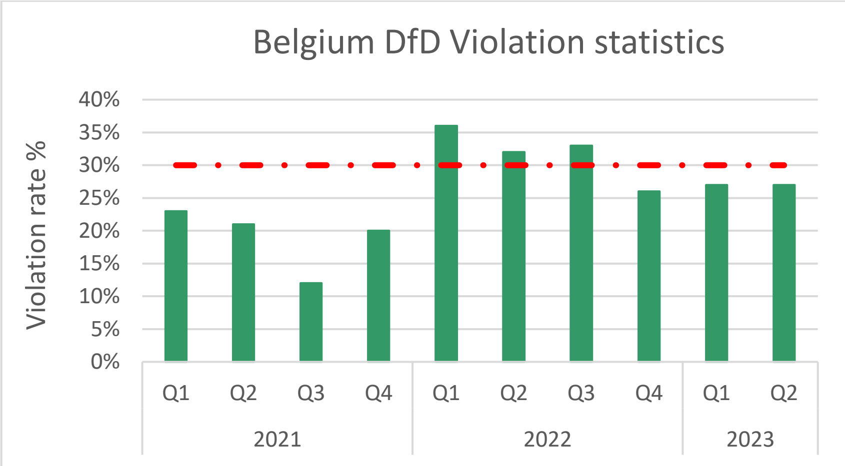


Figure 2 - Evolution of the percentage of DFD cases where Elia does not respect its ACE contribution threshold

Elia is then considering the second solution identified as a mitigation measure to respect ENTSO-e's criteria and avoid having to contract additional reserves (equivalent to a penalty).

Therefore, in agreement with CREG, before effectively applying the second solution, Elia puts it as part of the CREG Balancing Incentives for 2023 (CREG, 2022) which consists in the analysis of the activation of mFRR and/or tuning of the LFC controller based on prediction algorithms.

With this context, it is important to understand the global approach and the role of the incentive. In case Elia is not pro-actively taking action to reduce its contribution in the DFD event, Elia could/will be forced to contract additional reserves as a penalty for trespassing the threshold. Elia will so demonstrate that, under the current hypothesis, contracting the additional reserves is more costly than implementing the second solution as proposed in the incentive and so, it stays preferable to apply the mitigation measures as proposed in this second solution than to face the penalty.

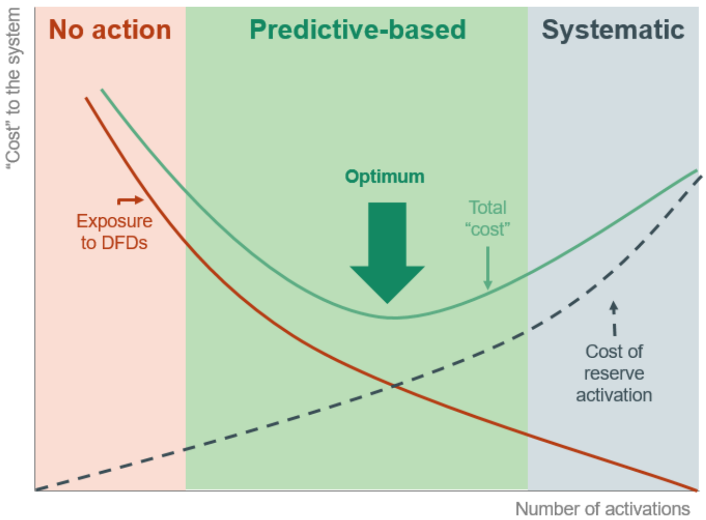


Figure 3 - Predictive-based model for optimal decision-making

Then, Elia will define the quality target to reach in terms of DFD contribution between 0% (being never a contributor among any of the DFD event) and 30% (being 30% of the time a contributor among the DFD event).

Finally, Elia will define the ACE threshold above which an action needs to be taken anyways to avoid endangering the System Security and for all other cases, work to reach the above-mentioned target in a cost-efficient way which means to

- only apply the mitigation measure when there is a DFD and when Elia is contributor thanks to two predictive models,
- define the most efficient product combinations (mFRR activation and/or tuning of the LFC controller) to reduce Elia's contribution.

The third solution consists in discussing with the producers to possibly spread the starting and stopping of fast acting units over a longer period and to all in all restrict their ramping rate. Until now, this solution stays at the level of the discussions between Elia and market parties and no official action has been taken.

1.2 CREG Incentive

As part of the 2023 Balancing CREG Incentives (CREG, 2022), Elia defines with CREG several deliverables regarding Deterministic Frequency Deviations. The content of the incentive (CREG, 2022) is aligned with Section 1.1 as expressed by the extracts hereunder:

« L'objectif de cette étude est d'analyser en détail comment mettre en place un outil de prévision des DFDs et de la contribution d'Elia en lien avec le critère ENTSOe de manière fiable, « cost-efficient » et qui permettrait aux dispatchers d'avoir à leur disposition un indicateur pour activer du mFRR et/ou adapter le réglage de la sortie du LFC dans le cadre de l'équilibrage du système au moment des DFDs. »

"Het doel van deze studie is in detail te analyseren hoe een tool voor de voorspelling van DFD's en de bijdrage van Elia in verband met het ENTSOe-criterium op een betrouwbare, kostenefficiënte manier kan worden opgezet waardoor dispatchers over een indicator zouden kunnen beschikken om mFRR te activeren en/of de LFC-outputregeling aan te passen in het kader van de balancering van het systeem op het moment van de DFD's."

1.3 Timeline

The different milestones are listed here below:

- 3rd February 2023: Selection of most relevant datasets and most performant model
    - Analysis and comparison of the datasets and the models.
- 1st September 2023: Consultation of a draft report
    - Description of the method used to select the dataset and the final model;
    - Results of the comparison based on statistical indicators;
    - Advantages and Disadvantages of the models;
    - Proposal/Relevance of publications related to DFD's;
    - If applicable: recommendations in terms of tool implementation.
- 22nd December 2023: Final report
    - Tests results (minimum 1 month);
    - If applicable: implementation plan.

2. Definition & General Approach

2.1 Definition

Based on the ENTSO-e criteria (entso-e, 2019), **a frequency deviation is defined as a DFD if the variation in frequency exceeds 75 mHz for more than 20 successive seconds during a change of Market Time Unit.**

Using this criteria, ENTSO-e identifies and extracts the DFD events over a period of time with the related amplitudes and timings. ENTSO-e then selects the **instantaneous ACE at the moment the frequency reaches its extremum**, also called the frequency "Nadir", where the ACE value will define the Elia ACE contribution. The event start time corresponds to the observed max rate of change of frequency (ROCOF) computed with moving average of 30s (Johannes Kruse, 2021). On the other hand, the recovery time corresponds to the time needed to return to 50Hz. After having extracted these values, a report is sent every quarter of a year to the control blocks of the Synchronous Area.

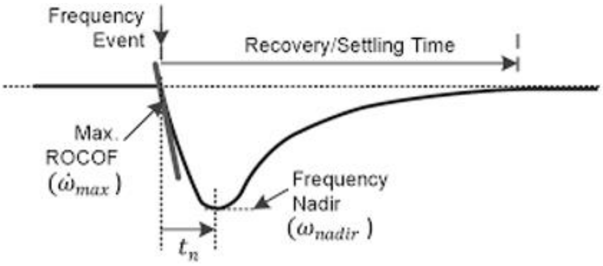


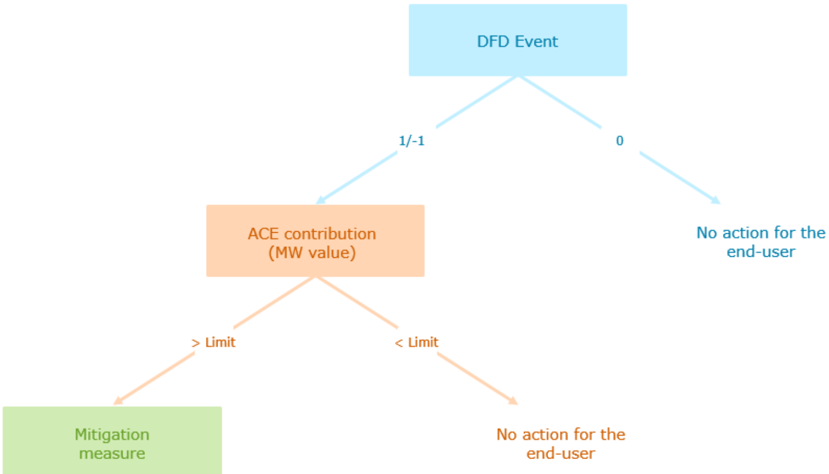
Figure 4 - Scheme of a DFD event with defined metrics

2.2 General Approach

Based on ENTSO-e's monitoring, Elia established conditions to take some actions via the mitigation measures (mFRR activation and/or tuning of the LFC controller), which are

- the occurrence of a **DFD event** (1 for positive DFD/-1 for negative DFD) and
- the **ACE contribution of Elia** that exceeds its own limit[2] (= 217MW in 2022).

This approach will enable Elia to take actions only when there is a DFD while avoiding unnecessary costs by taking mitigations measures in case Elia isn't considered as a DFD contributor (meaning being below the 217MW threshold).



[2] Each TSO has a limit (ACE contribution threshold) that he cannot cross. In case of the TSO is crossing this value, the TSO is considered as a DFD contributor.

Figure 5 - Flowchart of the DFD forecasting process

Three different layers, having each their own specificities and range of application, can be distinguished in the diagram. More precisely**, two different models (DFD event and ACE contribution forecasts** for which details are presented in Chapters 3.3.2 and 3.4.2 respectively) **having their proper dataset** (details in Chapters 3.3.1 and 3.4.1 respectively) are used in the process and based on their combined outputs, some **mitigation measures (mFRR activation and/or tuning of the LFC controller)** have to be applied (details in Chapter 5):

DFD event forecast @Synchronous Area level:

Firstly, the aim of the **DFD event model** is to forecast the likelihood of a DFD using a supervised machine learning model. The DFD will be forecasted as per the criteria defined in Chapter 3.3. The outputs of the model will namely be three categorical variables:

CATEGORY	MEANING
0	No DFD forecasted
-1	Negative DFD (drop of frequency)
1	Positive DFD (rise of frequency)

If a DFD is forecasted, the process continues to the second step which is related to the Elia ACE contribution. Otherwise, the process ends up directly and no action is taken.

ACE contribution forecast @Elia (TSO) level:

When a DFD is forecasted, it is then important to forecast whether or not Elia will exceed its own 217MW contribution threshold as given by the SAFA (entso-e, s.d.). It is why the aim of the **ACE contribution model** is to use a supervised machine learning model to forecast the Elia contribution during a DFD. So, the model will perform a regression in order to provide **a continuous variable representing the instantaneous ACE contribution of Elia at the moment the frequency reaches its maximum/minimum (NADIR) depending on if it is a rise or a drop of frequency.** This output variable will then be compared to the Elia threshold, which corresponds to the maximum acceptable contribution of Elia.

In case the model output, probability of being above this 217MW threshold, is below a certain value, the process ends up and no action is taken. Otherwise, this will warn the System Engineer to take some mitigation measures to reduce Elia's contribution in the DFD.

Mitigation measures @Elia (TSO) level:

Based on Elia-CREG's report on DFD's in 2020 (ELIA, 2020) and on the analysis of Chapter 5, **Elia will define the mitigation measure (adapting the output of the controller and/or activating mFRR) that resolves enough violations while limiting the impact in terms of costs.**

3. Forecasts

This incentive shows a first iteration of the DFD product development on the Elia side. In the future, Elia could potentially consider new datasets, new models and parameters and 'easily' adapt to new dynamics and market behaviors in close collaboration with the markets and CREG in order to improve the results.

For both forecasts (DFD event and ACE contribution), the timing characteristics are:

Start	D-1 10 pm
Update	Each QH
Forecast horizon	96 QH of D+1 or the left-over of the day (D)

For both forecasts (DFD event and ACE contribution), Elia separately performs a dataset selection and a model selection as described in the two following sub-sections.

3.1 Approach for Dataset selection

The initial set of data is defined based on the physical root causes of the variable to predict and on the experience of the Elia experts. The variables are studied over different time horizons past and future (for forecast) when available.

With historical data, no issue (invalid, missing, etc.) was observed. If needed, aberrant outliers are removed, and missing data will be linearly interpolated. The strategy could be revised and adapted depending on real time behavior.

From the defined dataset, a correlation (based on Pearson coefficient) and BorutaSHAP analysis is performed to identify the most correlated features. The variables also undergo Principal Component Analysis (PCA) and Recursive feature elimination (RFE). The results are consolidated in one dataset, (mainly based on BorutaSHAP) that would reduce the variable inter-dependencies and model over-fitting and complexity. Using less variables speeds up training and makes the model less sensitive to outliers and wrong data. Literature about the data analysis and selection methods can be found in Annex 9.1.1.

The dataset for the DFD occurrences is generally imbalanced. As an example, in 2021, Elia observed the following (see Figure 6):

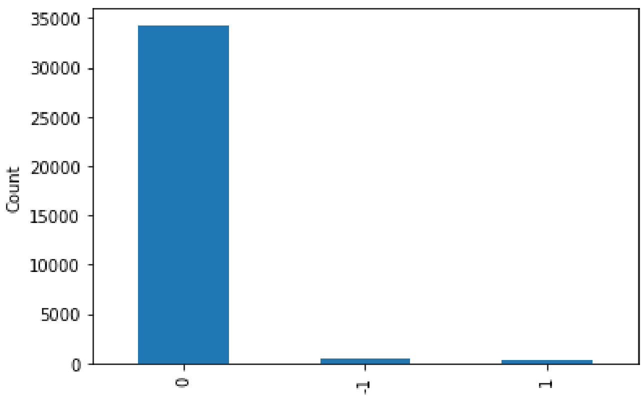


Figure 6 - Value counts of DFDs for 2021

The learning phase and the subsequent prediction of machine learning algorithms can be affected by the problem of imbalanced data set. As a result, in order to handle this imbalance between No DFD

(0), Positive DFD (1) and Negative DFD (-1), an additional imbalance class handling operation is performed on the DFD event dataset via several techniques presented in Annex 9.1.2.

3.2 Approach for Model Selection

As foreseen by the incentive, Elia studies the following 5 families of supervised machine learning models (all models state of the art, advantages and disadvantages are given in Annex 9.1.3.):

- Linear regression and Logistic regression
- Artificial Neural Network (ANN)
- Support Vector Machine (SVM)
- Random Forest
- Gradient boosting

Elia first identifies a baseline for each of the models (DFD event and ACE contribution). This naïve model will serve as a comparison with the developed model.

Then, a model parameters sweep is performed for each of the models to fine-tune their parameters.

After this, the models are ranked based on defined statistic indicators and are evaluated on a full year (from 2021-04-19 to 2022-04-19), what should give a reasonable indication of performance while still being computationally feasible. This evaluation is done in a k-fold[3] timeseries split way.

Finally, for the best model, Elia optimizes the training set length and performs a sensitivity analysis of the model regarding input dataset availability.

3.3 DFD event Forecast

3.3.1 Data selection

By definition, DFDs arise when a large imbalance between load and production is observed in the Synchronous Area in a small amount of time during a change of Market Time Unit (MTU).

A first insight of correlated variables would be the ones related to **the market positions of European control blocks:**

- **Day-ahead Generation data**
- **Day-ahead Load data**
- **Net positions**

The data for these three variables were extracted for Austria (AT), Belgium (BE), France (FR), Germany (DE), Italy (IT), The Netherlands (NL), Spain (ES), Switzerland (CH), Denmark (DK), Czech Republic (CZ), Poland (PL) and Luxembourg (LU). This set accounts for the largest DFD contributors.

Then, as Elia has more detailed information about its own control zone, the correlation of some variables specific to Belgium were also tested:

- **Daily Schedules of DPsu's**
- **HVDC schedules between BE and GB (Nemo Link)**

---

[3] k-Fold timeseries split uses 3 months of train data and 1 day of evaluation. After that, it shifts both the train and test data by 1 day and it predicts again, until it has done the 12 months of data.

- **Regulated volume: aFRR, iGCC and mFRR from previous QH**

This specific analysis could help to detect which of the Belgian units are more involved in DFD events than others.

For all of the variables, first and second derivatives dependencies were also tested. Moreover, the variable dependencies was also studied for different Time Horizons, namely:

- **DFD occurrence** (autocorrelation studied until 96 QH in the past)
- **Time Horizons for all Synchronous Area data** (studied until 8QH in the past)
- **Time Horizons for all variables** at Belgian scale (studied until 8QH in the past) meaning Production programs of Belgian "CIPU" units and HVDC schedules (Nemo Link) between BE and GB.

Finally, time-based features were also tested:

- Day of the month
- Day of the week
- Hour
- Minute

3.3.1.1 Conclusion on Data selection

The retained variables are summed up in Table 1 (here below) with the results of the different methods. The details are available in Annex 9.2.1.

Eight important features were found where Elia could identify time-based dependency (minute and hour) and correlation with load and generation (values, first and second derivative). The Belgian requested aFRR from previous QH, despite with a relatively small correlation, is retained and should highlight the important Belgian contribution to DFD event.

The number of variables retained is relatively limited compared to the identified variables in the Pearson's correlation analysis. This is mainly due to variable collinearity and is aligned with the Principal Component Analysis (PCA) that concludes that most of the information is contained in less than 10 independent components.

Recursive feature elimination (RFE) tends to give flawed results and is only given for information purpose. This is probably due to overfitting of the model, as the latter removes, at each iteration, the least correlated feature.

Table 1 - Overview of dimension reduction techniques

Retained variables	Pearson's correlation coefficient	BorutaSHAP rank position	RFE rank position
Minute	0.045	1	156
IT(Italy)_Gen_first_derivative	0.411	2	2
IT_Gen_second_derivative	0.295	3	53
Hour	-0.098	4	101
NL(Netherlands)_Load_first_derivative	0.098	5	19
PL(Poland)_Load_second_derivative	0.277	6	215

BE(Belgium)_Load_forecast_first_derivative	0.129	7	8
BE_Gen	0.005	8	72
aFRR_previous_qh_second_derivative	-0.012	9	3

3.3.2 Model selection

Regarding the model selection, many statistical indicators can be used to evaluate the performances of the models but most of them are derived from the confusion matrix.

Then, in order to define the metrics, Elia defines the number of True Positives (TP), False Positives (FP), True Negatives (TN) and False Negatives (FN).

Based on it, the following metrics are derived:

- **Precision: TP / (FP + TP)**
- **Recall: TP / (TP + FN)**

For multiclass predictions like here (DFD positive, DFD negative or no DFD), the metrics will be given for each class and for all of them together.

The DFD occurrence of previous hour is used as a baseline. For models to be considered, they need to do at least better than this result. However better can have a lot of meanings:

- Better recall: Less DFD cases missed, but probably more false positives.
- Better precision: Less errors when predicting a DFD, but probably more DFDs missed.

All the other measures mainly derivate from this, and it will always be a tradeoff. To optimize selected models, Elia uses the F1 macro average score:

- F1-score: 2*Precision *Recall / (Precision + Recall)

Confusion matrices and other metrics (as Youden index) can be found in Annex 9.4.1.

3.3.2.1 Conclusion on Model selection

The F1-score results for the models are given in table 2, with their respective optimal parameters and considering naïve oversampling. The latter was found the most performant resampling technique (see Annex 9.3 for details). **The best model is the random forest classifier**, which allows to increase the performance on Negative and Positive DFD detection while not deteriorating too much the performance on no DFD event detection.

As explained in the Annex 9.1.3.5, this decision tree based model is performant when describing non-linearities and can provide indication of the certainty of the model in the form of a probability. However, due to the model complexity, it remains quite opaque to the identification of the DFD root causes.

Then, the huge number of decision trees that make up the forest, presents the disadvantage that they take more time to train. As Elia aims to retrain the model continuously on a rolling window of 3 months, this could be a challenge in terms of implementation. Would the training end up being too slow, several options could still be investigated like further develop out IT capabilities, not retrain the model every quarter hour, but seek for a balance to train it less often or switch to an alternative model like hist gradient boosting which is much faster to train and which presents similar advantages.

Table 2 - F1-score results for DFD model selection

Model	Optimal parameters	F1-score			
		Negative DFD	No DFD	Positive DFD	Macro average
Baseline (previous hour)	-	**0.192**	**0.981**	**0.171**	0.448
Baseline (previous day)	-	**0.294**	**0.983**	**0.244**	0.507
Logistic regression	-	**0.239**	**0.924**	**0.203**	0.455
Neural Network	Layer sizes: (4,4,4,4)	**0.236**	**0.936**	**0.189**	0.454
Support Vector Machine	-	**0.225**	**0.916**	**0.192**	0.444
Random forest classifier	Depth: 16	**0.334**	**0.977**	**0.348**	0.553
Hist gradient boosting classifier	Depth: 6 L2 regularization: 13.78	**0.337**	**0.976**	**0.337**	0.550

Random Forest model is able to catch ~ 50% of the DFDs up and down with high certainty.

There is only 0.3% of prediction in the wrong direction (i.e., predicting a DFD up for a DFD down and the other way around) and the percentage of true negative is 3,29 %[4]; (i.e., predicting a DFD up or down when there should be no DFD).

3.4 ACE Contribution Forecast

3.4.1 Data selection

The Elia ACE contribution forecast corresponds to the imbalance of the TSO control zone during the DFD (at the worst moment of the frequency deviation) without having taken any balancing actions (which means that the Net Regulated Volume is null).

Therefore, a first insight of correlated data would be the data directly associated to the Elia control block.

Based on the physical nature of the ACE, the following variables were selected for the analysis:

- **Daily Schedules of DPsu's**
- **HVDC schedules between BE and GB (Nemo Link)**
- **BE Net position**
- **15' average ACE measurements**
- **Net Regulated volume: aFRR, mFRR, iGCC**

Next to that, Synchronous Area data were also tested similarly as for the DFD event model for the same set of countries:

- **Day-ahead Generation data**
- **Day-ahead Load data**
- **Net positions**

[4] This probability reduces to 0,5% when assuming $|ACE| > 217MW$ (hypothesis of perfect ACE model).

The ACE contribution, due to its nature, is an instantaneous variable, the correlation with past values is then very limited. Despite auto-correlation was not investigated, the average ACE of previous QH should give baseline information of the BE ACE contribution.

For all of the variables, the first and second derivatives dependencies were studied.

Finally, as for DFD event, time-based features were also tested:

- Day of the month
- Day of the week
- Hour
- Minute

3.4.1.1 Conclusion on Data selection

The retained variables are summed up in Table 3 (here below) with the results of the different methods. Some detailed information is available in Annex 9.2.2.

Firstly, the variables of interest for the prediction of DFD events appears naturally as good predictors. This is probably due to the large contribution of Belgium to the DFD event, and thus correlation between those two variables to predict (which may change through time).

Moreover, despite being related to the Belgian zone, there is a correlation with the load and generation of other TSOs. This comes from their interdependencies with BE load and generation.

Regarding the production program, the impact of specific units is correlated, and some other units appear unexpectedly as predictors. The correlation with those unexpected stable non-condonable units comes from their maintenance planning. They are most of the time off during the summer, where reduced number of DFDs appears. This, therefore, indicates the seasonality evolution of DFD.

Finally, the HVDC schedule program change (Nemo Link) and other regulation variables (mFRR and IGCC) are also correlated.

As for DFD event forecast, Recursive Feature Elimination (RFE) ranking is given for information purpose.

Table 3 - ACE forecast data selection summary table

Variable	Pearson's correlation coefficient	BorutaSHAP ranking	RFE ranking
IT_Gen_first_derivative	0.327	1	4
aFRR_previous_qh	-0.061	2	3
BE Load Forecast	-0.118	3	16
CH_Gen_first_derivative	0.278	4	60
Minute	0.072	5	88
BE NET Position_first_derivative	0.093	6	19
AT_Gen_first_derivative	0.272	7	201
Load Forecast_first_derivative	0.085	8	29
Previous_ace_belgium	0.058	9	1
HVDC Nemo_first_derivative	0.113	10	18
HVDC Nemo_second_derivative	0.072	11	117
aFRR_previous_qh_second_derivative	-0.074	12	13

aFRR_previous_qh_first_derivative	-0.005	13	6
Specific Power Unit (Anonymized)	-0.045	14	98
DE_Load	-0.054	15	44
DE_Load_second_derivative	0.105	16	14
BE_Load_first_derivative	0.083	17	29
BE NET Position_second_derivative	0.054	18	8
Specific Power Unit (Anonymized)	0.031	19	38
ES_Gen_second_derivative	0.015	20	116
AT_Gen_second_derivative	0.195	21	25
Hour	-0.033	22	113
NL_Load_first_derivative	0.062	23	7
Specific Power Unit (Anonymized)_first_derivative	0.147	24	22
DK_Load_first_derivative	0.241	25	254
CZ_Load_second_derivative	0.128	26	111
Specific Power Unit (Anonymized)	0.002	27	72
mFRR_previous_qh_first_derivative	0.062	28	80
DK_Gen	-0.009	29	9
CH_Gen_second_derivative	0.198	30	162
Specific Power Unit (Anonymized)	-0.006	31	137
CH_Gen	-0.047	32	57
FR_Load_second_derivative	0.130	33	165
HVDC Alegro	-0.050	34	23
Specific Power Unit (Anonymized)_first_derivative	-0.096	35	78
AT_Load_second_derivative	-0.051	36	37
GCC_previous_qh	-0.019	37	34

Note: References to “Specific Power Unit (Anonymized)” do not necessarily always refer to the same power unit.

3.4.2 Model selection

The most popular performance indicators for continuous dependent variable are the following:

- Root Mean Square Error: $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$
- Mean Absolute Error: $MAE = \frac{1}{N}\sum_{i=1}^{n}|y_i - \hat{y}_i|$
- R² coefficient: $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

where y_i, $\bar{y}$ and $\hat{y}_i$ are the observations, the observations mean and the model output respectively.

The RMSE indicator will be used for the model selection.

We considered as baseline model: Predicting zero result. This baseline performed better than considering ACE from previous QH, hour or day.

3.4.2.1 Conclusion on Model selection

The RMSE indicator for the models is given in table 4 (see here below), with their respective optimal parameter. **The best model is the Gradient boosting model** with the best RMSE (and the best MAE and R^2).

As for DFD event model, decision tree based models performs the best, probably due to the non-linearity of the ACE contribution with the features. The main disadvantage is the physical interpretability.

Table 4 – RMSE (R^2 and MAE) results for ACE model selection

Model	Optimal Parameter	R^2	MAE	RMSE
Predicting zero	-	0	79.67	**112.29**
Linear regression	-	Unstable	Unstable	**Unstable**
Linear regression with PCA	Number of components: 34	0.163	73.15	**102.72**
Neural network	Layer sizes: (4,4,4,4)	0.083	75.32	**107.51**
Random forest	Max depth = 15	0.168	72.93	**102.38**
Gradientboosting	Depth = 15	0.171	72.84	**102.21**

Note: When focusing on DFD up and down events, the RMSE of all models degrades importantly. For our selected GradientBoosting model, it reduces to 171MW. This is due to the instantaneous characteristic of the ACE and its unpredictability (availability of IGCC, position of other TSOs, direction of the aFRR in real-time, human factor, etc.)

3.5 Performance – Conclusion of combined forecasts

Until now, Elia analyzed the models for “DFD events” and “ACE contribution” forecasts separately.

Nevertheless, according to the initial hypothesis, it’s the combination of the outputs of the two forecasts that gives the trigger to apply the mitigation measure or not. So, it’s important to assess how the two forecasts perform with each other sequentially speaking.

The analysis was performed on a period going from 04/2021 to 04/2022.

For DFD event, the number of false positive (DFD up/down detected but no DFD in reality) as well as the number of DFD identified in the opposite direction are relatively negligeable. Therefore, for the following part, Elia will study the performance of the ACE contribution model for (upward and downward) DFD event (meaning Elia will thus disregard the events for which no DFD is forecasted).

Then, when a DFD is detected, the criterion to take action is based on the probability to observe an ACE violation (ACE>217 MW for a DFD up and ACE<-217MW for a DFD down). The probability can be derived under Normal assumption, the ACE point forecast and the model performance (RMSE).

Elia can so adapt the Confidence Level “x” (in the interval $[0; 1]$) to act with a confidence level under a Risk approach. The Confidence Level is defined to not take unnecessary action while reducing the ACE contribution.

> *Eg: There is a negative DFD and if x=0.8, Elia will act only if we are sure at 80% that ACE will be below -217MW. There is, on the other side, a risk of 20% that the ACE is above -217 MW.*

The graph below (Figure 7) shows the combined forecasts performance evolution with the Confidence Level:

- 0% Confidence Level: means that we need to be sure at 0% that the ACE will be above 217MW or below -217MW according to the direction of the DFD and so corresponds to a systematic activation that will only be based on the DFD event forecast
- 100% Confidence Level: means that we want to be 100% sure that the ACE will be above 217MW or below -217MW according to the direction of the DFD and so corresponds to never activate).

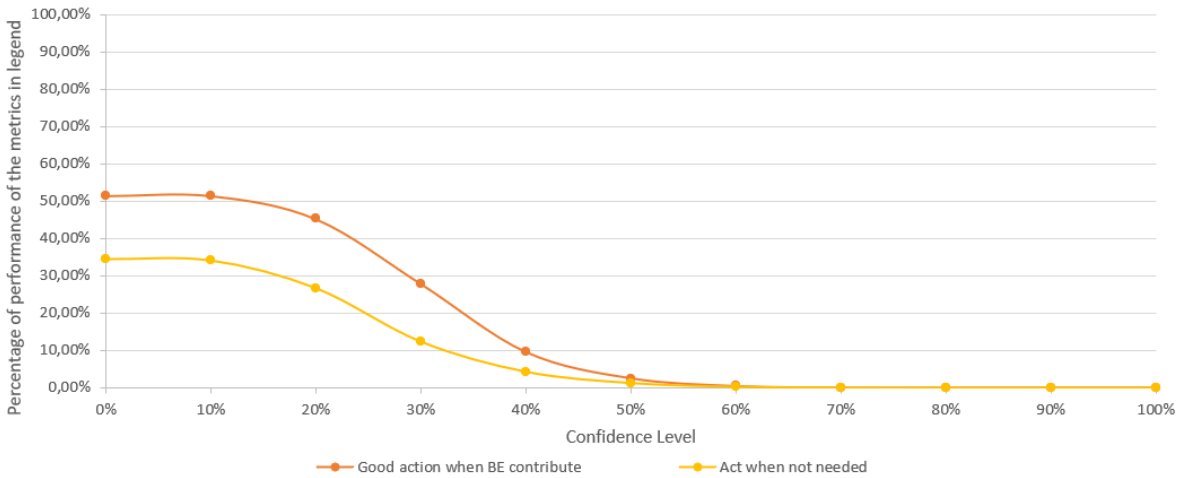


Figure 7 Performance evolution with combined forecasts (DFD event & ACE contribution) depending on the Confidence Level (from activating each time till never activating).

The "orange curve" ("Good action when BE contributes") corresponds to the percentage of the QH where an action is applied whereas a DFD is predicted and Elia is a contributor.

The maximum value is when the Confidence Level is 0% which corresponds to a systematic activation based on DFD event forecast.

Therefore, the performance is limited by the 50% accuracy of the DFD event forecast. On the other hand, increasing the level of confidence will reduce the number of actions until no more actions are taken to reduce the contribution (=actual situation).

The "yellow curve" ("Act when not needed") corresponds to the percentage of unneeded actions whereas a DFD is forecasted meaning there is a DFD which is forecasted but Elia is not contributing to the DFD.

The curve shows that "Act when not needed" reduces rapidly with the Confidence Level. This percentage should be kept at a reasonable level to avoid ACE deterioration (taking action and finally degrading the ACE) and unneeded cost (taking action whereas BE is not a contributor).

E.g.:

Hypothesis: From previous sections we know that the DFD forecast predicts ~50% of the DFD.

For ~30% of the DFDs, we will have some violations and for the other ~70% we will have no violations. We consider a perfect mitigation measure (a measure that solves all events for which it's applied

<u>*Confidence level of 0*</u>*%: For all the DFDs that we forecast, we take action.*

- *Orange curve: We only look at DFD events for which Elia was also contributing. We forecast 50% of all DFD events so we also forecast 50% of those DFD events for which Elia was contributing.*

For all of them we take an action and we solve the violation. So, we have 50% of "Good action when BE contributes".

- *Yellow curve: We look at all the DFD events. For around 70% of events, we had no violation. We forecast half of them, so around 35% and for those we take an action. So for 35% of the time, we "Act when not needed".*

In conclusion, due to the lack of accuracy of the "ACE contribution" forecast (high RMSE), Elia tends to choose a low Confidence Level and acts more upon detected DFD, even if the risk is high that Elia was not contributing to the DFD. Thanks to the negligeable number of false positive as well as the number of DFD identified in the opposite direction, all the actions that would be taken, even if not strictly needed to solve a violation, would not worsen our number of violations.

4. Cost of the penalty

The cost of the penalty is the cost that Elia would face in case Elia does not remain below the 30% ENTSOe threshold imposed by the SAFA (entso-e, n.d.)**. Indeed, in case no effective mitigation measure has been implemented by excessively contributing TSOs, ENTSO-e could theoretically impose that these LFC blocks acquire/procure additional FCR** (entso-e, 2019) **(which will be the equivalent of a penalty).**

Therefore, Elia took the proposed volume of penalty (entso-e, 2019) as hypothesis for this incentive.

> *"Any LFC Block which chooses this as a solution will need to increase its FCR obligation (prescriptions and procurement) by 66%[5] at least during the time period when DFDs occur and will in such a way assist in reducing the DFD with the additional FCR provided.*
>
> *Nevertheless, it is important also to mention that such measure could be requested from non-compliant LFC blocks with respect to the fixed maximum violation ratio."*

This default solution would assist in reducing the DFDs but entails a social cost coming from the increased procurement of reserves. Moreover, this solution doesn't help in reducing the ACE contribution of the LFC block and represents thus well a penalty cost as there is no local efficiency related to this measure. But by adding additional FCR volume, it still assures that enough FCR will be available for its true used and not for DFD management.

For the sake of completeness and as alternative of penalty type, Elia also studied the additional cost in case **additional aFRR** volume was required/procured. Although increasing the procured volume of aFRR has been proven to be inefficient to mitigate ACE contribution during DFD events (entso-e, 2019), except if it could be used in combination with a modification of the aFRR request to increase the cost-efficiency of the measure.

The study has been performed on the same reference period as for the Mitigation Measure Selection in Chapter 5, namely Q3/Q4 2021 and Q1 2022[6].

Some assumptions were considered, such as:

- Market conditions and product design in terms of procurement are kept as they were in Q3/Q4 2021 and Q1 2022
- Bi-directional and symmetrical procurement of FCR and aFRR

4.1 Additional FCR

4.1.1 Hypothesis

- FCR procured for all CCTU's

  For the moment, the FCR volume to be procured is constant over all CCTU's.

  Sensitivity Analysis: 4 CCTU's

  An additional sensitivity has been performed by considering a procurement for the CCTU's that cover the great majority of DFDs. In this context, 4 CCTU's covering 91% of DFDs over the

[5] The proposal to increase the volume by 66% is related to the fact that if this percentage was added to the actual total volume of FCR in the Synchronous Area (3000MW), this would generate 2000MW more FCR. The latter corresponds to 75 mHz (=DFD) if one considers a power/frequency characteristic of 27000 MW/Hz.

[6] Data with a granularity of 4 seconds are needed to run the simulations. The Scada system only allows to go back 2 years in time for 4 seconds data. We could not go back further than Q3 2021. Then we wanted to have 1 year of data for the analysis. Nevertheless, we considered it better to not consider data from Q2 2022 in order to avoid parasite effects from the energy crisis on the results.

studied period have been identified (CCTU's 2,3,5,6). In this case, no additional procurement would thus be considered for CCTU 1 and 4 during which the fewest DFD occurred. The latter could be potentially suggested if it allows to reduce the cost while still covering almost all DFDs.

- 66% of FCR capacity increase

  For Elia, this means an additional FCR volume of 58MW (based on a procured FCR volume of 87MW in 2021). This does not represent exactly the studied period but the volumes of FCR to be procured have been relatively stable over the last years.

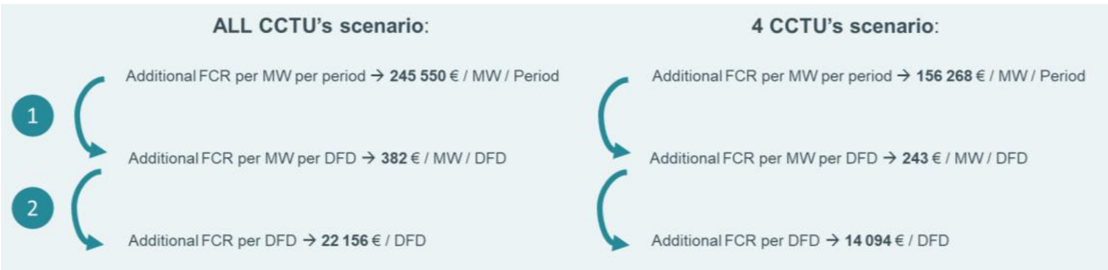
- Marginal pricing is used to estimate the cost of additional FCR

  The estimation of the costs linked to the additional procurement of FCR is done by taking the Local marginal prices observed during FCR auctions over the year 2021. The related data can be retrieved in (Regelleistung, n.d.).

  - The smaller the additional procurement, the more precise the estimation would be regarding the additional costs. As the additional volume to be procured in case of penalty is not negligeable (58 MW), the related penalty costs would probably represent a lower-bound of the actual costs it would really imply, provided that the same type of FCR-supplier can supply most of the additional volume and no other technology has to participate. Moreover, this lower-bound does not take into account any pricing change resulting from different market conditions- or for example from a transfer of volume from one product to another, like aFRR to FCR to fulfill such additional 58MW.
  - FCR volume to be procured remains a hypothesis based on ENTSO-e's report. Actually, due to liquidity issues, it could even be impossible to contract that much additional volume in Belgium.

4.1.2 Cost evaluation

Based on the prices of the 2021 FCR auctions, the total marginal price per MW over the studied period could be computed. As this price will have to be compared to the price of activations per MWh, it has been decided to express every cost per DFD occurrence.



Step 1 aims at converting the FCR cost per MW per period[7] into an FCR cost per MW per DFD. This can be done by taking into account the total number of DFDs over the studied period (643 DFDs/period).

Step 2 aims to assume the hypothesis of a 66% FCR capacity increase. So, the cost per DFD is obtained by multiplying the previous amount by the 58MW additional.

4.2 Additional aFRR

4.2.1 Hypothesis

- aFRR procured for all CCTU's

  For the moment, the aFRR to be procured is stable over the CCTU's.

[7] Period: refers to the studied period covering Q3/Q4 of year 2021 and Q1 of year 2022.

Sensitivity Analysis: 4 CCTU's

An additional sensitivity is realized by considering a procurement for only the 4 CCTU's that cover 91% of all DFDs over the studied period. In that case, no additional procurement is thus considered for CCTU 1 and 4 during which the fewest DFD occurred. The latter could be suggested in the future if it allows to reduce the cost while still covering almost all DFDs.

- Average pricing method is used to estimate the cost of additional aFRR

  As starting point, the total costs over the studied period have been computed by taking the sum of the daily procurement costs of upward and downward aFRR. Then, the estimation is an average taking into account the total cost divided by the total procured volume, which was 145 MW per auction at that time. The related data can be retrieved in (Elia - Auction Results, n.d.).

- Same additional aFRR volume increase is considered as for FCR

  Considering the calculation leading to the 66% increase of FCR volume, in order to compare the two types of penalties, Elia will consider the same aFRR volume increase than for FCR which means +58MW additional.

4.2.2 Cost evaluation

Based on the prices of the aFRR auctions, an average of the marginal price per MW over the studied period could be computed. As this price will have to be compared to the price of activations per MWh, it has been decided to express every cost per DFD occurrence. Our calculations led to the following results:

Step 1 aims at converting the aFRR cost per MW per period into an aFRR cost per MW per DFD. This can be done by taking into account the total number of DFDs over the studied period (643 DFDs/period).

Step 2 aims to assume the hypothesis of a 58 MW capacity increase (as for FCR). So, the cost per DFD is obtained by multiplying the previous amount by the 58MW additional.

4.3 Conclusion

The role of the penalty costs is to define the minimal cost that Elia would face as a penalty if they don't improve the contribution in the DFD thanks to the application of mitigation measure and thus, to have a reference cost to which the costs of the mitigation measures will be compared.

This reference cost will thus serve as an "Economical Parameter" in the Decision Tree presented in Chapter 6.3 (Cost for no action (facing the penalty) versus Cost for action (applying mitigation measure)).

Based on the above simulations, the next chapters will consider the cost of additional FCR as the reference (penalty) cost as it is the lowest one (meaning the most constraining).

The evaluation of the penalty cost is based on several hypothesis that will tend to evolve over time. It means that this cost will have to be reviewed anytime any of the hypothesis changes in a significant way.

5. Efficiency of the mitigation measure (aFRR and/or mFRR)

The aim of this chapter is to determine the most efficient mitigation measure to apply based on a comparative analysis in case of DFD where Elia is a contributor.

The considered mitigations measures are:

i. tuning the output of the LFC controller for the aFRR activation or;

ii. activating mFRR or;

iii. activating a combination of aFRR and mFRR.

5.1 Hypotheses

5.1.1 Common assumptions

- Elia uses historical data (Merit-Order, SI, ACE, ...) from Q3 and Q4 2021 and Q1 2022[8]
- Elia considers current product characteristics (FAT, remuneration, ...) meaning that future product design evolutions is not considered
- A perfect forecast has been considered (for historical data, Elia knows exactly when a DFD happened and what the Elia contribution was)
- The mitigation measures are applied on each DFD event for which Elia's contribution was higher than 217 MW
- An upward (downward) DFD means that downward (upward) activations were applied as mitigation measure.
- The cost is evaluated for a time window of one hour around the DFD event[9].
- The balancing product merit orders are not symmetrical (in volume and price), so Elia studies separately the upward and the downward DFD

5.1.2 aFRR

The aFRR mitigation measure consists of a tuning of the output of the LFC controller, during a certain time window.

	Initial hypothesis	Sensitivity
Window of activation	Start: 5 minutes before the beginning of the quarter hour of the DFD End: 5 minutes after the beginning of the quarter hour of the DFD	All combinations of the following start and end times: Start: 2-5-7 minutes before the beginning of the quarter hour of the DFD End: 2-5-7 minutes after the beginning of the quarter hour of the DFD

[8] The data with a granularity of 4 seconds are needed to run the simulations. The Scada system only allows to go back 2 years in time for 4 seconds data. So, Elia could not go back further than Q3 2021. Elia wanted to have 1 year of representation data for the analysis but decided to not consider any data from Q2 2022 in order to avoid the effects of the energy crisis on the results.

[9] Evaluating a one-hour window allows to catch all the impact of the mitigation measure. For mFRR for example, we activate the mFRR volume already the quarter our before the DFD event, so we needed to make sure that our window would catch it. To be more flexible, we took a one-hour window which is certainly too large but allowed some flexibility in the studied cases. The impact of this hypothesis is cancelled by the fact that we always take the difference between the situation where no mitigation measure is applied and the situation with the measure applied. So, all timesteps for which the measure has no impact will cancel.

Historical aFRR activations	Overwrite of the output of the controller during the tuning window	/
Activated volume	100% of the contribution, limited to the offered aFRR energy bid volume[10]	/

The figure below illustrates a typical application of the aFRR mitigation measure for a specific DFD event considering the initial hypothesis. The purple curve represents the aFRR volume in the situation where the measure is applied and is to be compared to the green curve that is the aFRR volume activated historically without any mitigation measure applied.

We see that the purple curves starts deviating from the green curve 5 minutes before the beginning of the quarter-hour and starts rejoining it 5 minutes after the beginning of the quarter-hour which represents the initial time window for the application of the measure. They result respectively in the ACE curve in red when the mitigation measure is applied instead of the blue historical ACE curve without any measure. At the moment of the DFD we can notice an improvement of the ACE between the blue and the red curve due to the application of the mitigation measure.

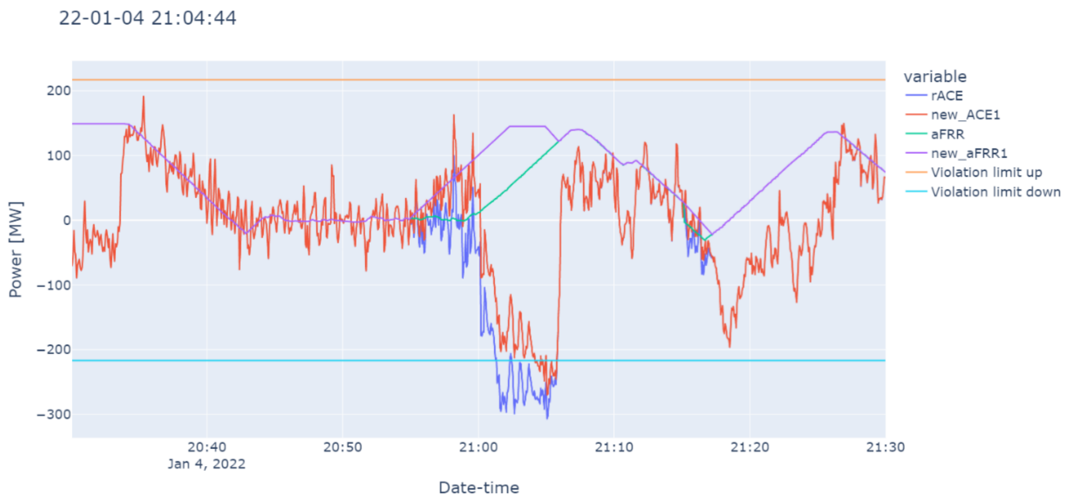


Figure 8 - Illustration of the application of the aFRR mitigation measure

5.1.3 mFRR

The mFRR mitigation measure consists of an additional mFRR activation.

	Initial hypothesis	Sensitivity	
Window of activation	Start: scheduled activation in the quarter hour preceding the quarter hour of the DFD End: end of the QH of the DFD (ramping down takes place in the next QH but not remunerated)	Start: direct activation 3 minutes after the beginning of the quarter hour preceding the quarter hour of the DFD End: end of the QH of the DFD (ramping down still takes place in the next QH but not remunerated)	Start: scheduled activation at the beginning of the quarter hour of the DFD End: end of the QH of the DFD (ramping down still takes place in the next QH but is not remunerated)
Historical	The mitigation measure is		

[10] During the observed period, the contracted volume was of 145MW. The offered energy bid volume was often higher than this volume, especially in the downward direction due to non-contracted energy bids. No cap has been applied, the contracted and non-contracted energy bid volume offered were considered in the simulations.

mFRR activations	applied additionally to the historical mFRR activations		
Activated volume	80% of the contribution limited to the offered mFRR energy bid volume		160% of the contribution (large overshoot to cover the contribution despite ramping) limited to offered mFRR energy bid volume

The figure below illustrates a typical application of the mFRR mitigation measure for a specific DFD event considering the initial hypothesis. The purple curve represents the mFRR volume in the situation where the mitigation measure is applied (contrarily to aFRR, the mFRR volume that is activated in the framework of the mitigation measure does not replace the historical mFRR activation of the green curve but is added to it).

We see that the purple curves starts ramping up 15 minutes before the beginning of the quarter-hour of the DFD and starts ramping down at the end of that quarter-hour which represents the initial time window for the application of the measure. We can notice a larger improvement of the ACE between the blue historical curve and the red curve due to the application of the mitigation measure, especially at the exact moment of the DFD.

For the remaining part of the quarter-hour of the DFD, the measure worsens the ACE.

Figure 9 - Illustration of the application of the mFRR mitigation measure

5.1.4 Combination of aFRR and mFRR

The combined aFRR and mFRR mitigation measure consists in

- the tuning of the output of the controller for part of the ACE contribution and
- the activation of mFRR bids for the remaining part of this contribution.

All the initial hypotheses of aFRR and mFRR are used.

	Initial hypothesis	Sensitivity	
Product share	aFRR: 30% mFRR: 70%	aFRR: 50% mFRR: 50%	aFRR: 70% mFRR: 30%

The figure below illustrates a typical application of the mFRR mitigation measure for a specific DFD event considering the initial hypothesis. In light green and pink, we observe the aFRR and mFRR volume activated in the framework of the mitigation measure. As expected, the result is like the combination two previous mitigation measure (aFRR and mFRR).

We notice the same improvement of the ACE red curve with regards to the ACE blue curve without any mitigation measure applied especially at the exact moment of the DFD.

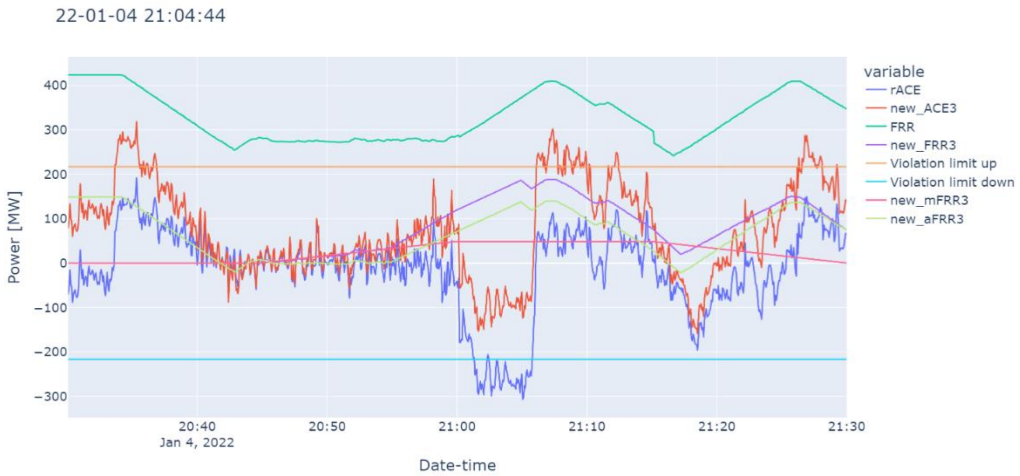


Figure 10 - Illustration of the application of the aFRR and mFRR mitigation measure

5.2 Selection of the type of mitigation measures

In order to select the mitigation measure, Elia defines **two KPIs** which are **the cost per avoided violation** and **additional cost per action taken by the TSO** and compares the mitigation measure between them.

After selecting the best mitigation measure, Elia will define its exact characteristics thanks to the sensitivity analysis.

5.2.1 KPIs definition

5.2.1.1 KPI: Cost per avoided violation

The KPI is the ratio between the delta of NRV cost and the number of avoided violations:

$$Cost\ per\ avoided\ violation\ = \frac{\Delta\ NRV\ cost}{\Delta\ Violation\ \ \#}$$

i) The delta NRV cost is the difference between the initial situation without mitigation measure and the situation in which a mitigation measure is applied. As it takes the difference between the two situations, all timestamps without any mitigation measure will cancel each other so that it only provides the additional cost of the mitigation measure itself.

ii) The number of avoided violations is the difference between the number of violations without applying any measure and the number of violations observed after applying the mitigation measure.

The KPI should be seen as an additional cost, which considers the effectiveness of the measure. The smaller the ‘cost per avoided violation’ the better, as it either reflects small implied costs or a measure that solves more violations.

5.2.1.2 KPI: Additional cost per action taken

The KPI is the ratio between the delta NRV cost and the total amount of violations.

$$Additional\ cost\ per\ action\ taken = \frac{\Delta NRV\ cost}{Total\ Violation\ \#}$$

i) The delta NRV cost – See Chapter 5.2.1.1 above

ii) The total amount of violation is the number of cases for which the measure has been applied

This indicator only gives an insight on the cost of the measure and not on the number of violations solved meaning it can be a low cost without resolving any violation or on the contrary a high cost resolving a lot of violations.

5.2.2 Comparison between the type of mitigations measures

	Upward DFD				Downward DFD			
	Status Quo	aFRR	mFRR	aFRR_mFRR	Status Quo	aFRR	mFRR	aFRR_mFRR
# violations	63	35	16	34	91	54	6	34
% Resolution		44%	75%	54%		41%	93%	63%
aFRR_costs	€ 584.558	€ 467.213	€ 584.558	€ 501.275	€ 1.419.815	€ 1.669.095	€ 1.419.815	€ 1.546.968
mFRR_costs	€ 246.806	€ 246.806	€ 1.139	€ 41.533	€ 844.061	€ 844.061	€ 5.897.944	€ 2.559.846
Total costs	€ 831.364	€ 714.019	€ 585.697	€ 542.808	€ 2.263.877	€ 2.513.157	€ 7.317.759	€ 4.106.814
Costs per avoided violation		-€ 4.191	-€ 5.227	-€ 9.950		€ 6.737	€ 59.457	€ 32.332
Additional cost per action taken		-€ 1.863	-€ 3.899	-€ 4.580		€ 2.739	€ 55.537	€ 20.252

Note: a positive number means additional costs that Elia would have to pay and negative numbers represent money Elia would receive or “save”.

The outcome of this first comparison respects a certain logic that **upward activations would imply additional costs while downward activations would lead to additional savings**. The latter is inherent to the observed prices in the merit-order lists of the concerned period. Indeed, upward DFD will lead to the deactivation of upward bids and/or the activation of downwards bids which often have positive prices (cost to receive from the BSP) at the beginning of the merit-order.

For upward DFDs, the best solution seems to be the combination of aFRR and mFRR and for downward DFDs, the best solution seems to be the aFRR mitigation measure.

Explanation:

- It is quite logical that mFRR doesn’t offer a good mitigation measure. Indeed, the activation of mFRR will probably cover the DFD contribution quite correctly but constitutes a large overshoot for the remaining part of the quarter-hour and for a significant part of the adjacent quarter-hours where the rampings (up & down) are performed.
- aFRR is probably not the cheapest reserve to activate but allows to cover a very definite time window that allows to mitigate the DFD contribution while avoiding overshoots.
- Fewer aFRR volumes than mFRR volumes are at disposal so the percentage of violations that are solved with the only aFRR mitigation measure is smaller.
- The combination of aFRR and mFRR partially offers both the advantages and the disadvantages mentioned for each individual measure. It is a bit higher in price due to the use of mFRR but also solves more violations as mFRR compensates the volume of aFRR that is sometimes not sufficient to avoid the violation.

5.2.3 Conclusion

Firstly, the aFRR mitigation measure clearly seems to be a good option for downwards DFDs. Its cost-effectiveness is inherent to its tunability, which enables to target the short timeframe of a DFD event and to activate a non-negligeable volume that helps reducing the amount of time Elia behaves as a principal contributor.

Secondly, using mFRR only as a mitigation measure is not an option that is worth investigating. So, Elia didn't perform any of the identified sensitivities on mFRR as none of them has a chance of changing the conclusions because:

- Using the direct activation could allow to delay the activation by a few minutes but the delay will prevent to reach the full volume at the moment of the DFD. So, Elia will not be able to delay this activation by more than 2 to 3 minutes which will not drastically reduce the cost.
- Using the direct activation and delay the start of the activation by more than a few minutes will have a positive impact on the price but will not cover the DFD contribution correctly and will still imply an overshoot.
- Using a scheduled activation starting at the beginning of the quarter hour of the DFD will require a large overshoot to make sure that a sufficient volume is activated at the beginning of the DFD event (i.e. a few minutes after the beginning of the activation). There is no reason to think that this might have a positive impact on the price and the overshoot would be even stronger.

Thirdly, the aFRR and mFRR combination is interesting for upward DFDs. Indeed, seeing the difference of results between the aFRR only and mFRR only mitigation measures, changing the proportion of each of the products could completely change the conclusions on this solution.

5.3 Sensitivity Analysis

This section aims to provide different implementations of the selected type of mitigations that show the balance between the cost for action and the percentage of resolution.

In this section, the sensitivities on each measure are presented for the category of DFD for which they were identified as most efficient in the previous section (e.g. aFRR sensitivity analysis is performed for Downwards DFDs) but the complete tables are provided in Annex 9.5

5.3.1 aFRR

As initial hypothesis, Elia considered tuning the output of the LFC controller during a time window starting 5 minutes before the beginning of the quarter hour of the DFD event and ending 5 minutes after the beginning of the quarter-hour of the DFD event.

This initial hypothesis is justified by the fact that the DFD event usually occur 2-3 minutes after the beginning of the quarter-hour and starting 5 minutes before the quarter-hour allows the full aFRR activation by the beginning of the DFD event. Then, ending 5 minutes after the beginning of the quarter hour assures that the whole event is covered before deactivating the aFRR.

For aFRR, the moment to Start and the length of the interval (difference between start time and end time) impact the cost and the percentage of resolution.

	Downwards DFD									
	SQ	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR
Start		-7	-7	-7	-5	-5	-5	-2	-2	-2
End		7	5	2	7	5	2	7	5	2
# violations	91	48	48	55	54	54	61	69	69	76
% of resolution		47%	47%	40%	41%	41%	33%	24%	24%	16%
aFRR_costs	€ 1.419.815	€ 1.840.176	€ 1.794.549	€ 1.704.670	€ 1.712.070	€ 1.669.095	€ 1.584.245	€ 1.514.140	€ 1.482.904	€ 1.439.492
mFRR_costs	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061
Total costs	€ 2.263.877	€ 2.684.237	€ 2.638.610	€ 2.548.731	€ 2.556.132	€ 2.513.157	€ 2.428.306	€ 2.358.201	€ 2.326.966	€ 2.283.553
Costs per avoided violation		€ 9.776	€ 8.715	€ 7.913	€ 7.899	€ 6.737	€ 5.481	€ 4.287	€ 2.868	€ 1.312
Additional cost per action taken		€ 4.619	€ 4.118	€ 3.130	€ 3.212	€ 2.739	€ 1.807	€ 1.037	€ 693	€ 216

As expected, on the one hand, the longer the interval the more violations will be solved to a certain extend but also the higher the cost. On the other hand, the shorter the interval the less violation solved and the smaller the cost.

Similarly, the later Elia starts applying the mitigation measure, the largest number of violations is missed or not entirely solved.

The combined model performances and the percentage of resolution of the measures need to be analyzed together to draw conclusions on the exact mitigation measure to apply. This study is performed in Chapter 5.3.4 Conclusion.

5.3.2 mFRR

As mentioned here above, it is not relevant to perform any sensitivity analysis for this mitigation measure.

5.3.3 Combination of aFRR and mFRR

For the combination of aFRR and mFRR, considering the large difference between the results observed for aFRR and mFRR, changing the proportion of each product could also change the result of this mitigation measure with regards to the two KPIs.

For the combination aFRR and mFRR, it is the split of the combination 30%(aFRR)/70%(mFRR), 50%/50% and 70%/30% that impacts the cost for action and the % of resolution.

	Upwards DFD			
	SQ	aFRR_mFRR		
aFRR		70%	50%	30%
mFRR		30%	50%	70%
# violations	63	30	31	34
% of resolution		52%	51%	46%
aFRR_costs	€ 584.558	€ 582.202	€ 533.417	€ 501.275
mFRR_costs	€ 246.806	-€ 1.168	-€ 11.825	€ 41.533
Total costs	€ 831.364	€ 581.034	€ 521.592	€ 542.808
Costs per avoided violation		-€ 7.586	-€ 9.680	-€ 9.950
Additional cost per action taken		-€ 3.974	-€ 4.917	-€ 4.580

Similarly as in Chapter 5.3.1, the combined model performances and the percentage of resolution of the measures need to be analyzed together in order to draw conclusions on the exact mitigation measure to apply. This study is performed in Chapter 5.3.4 Conclusion.

5.3.4 Conclusion

As mentioned here above, none of the mitigation measure solves all the violations even when applied based on a perfect forecast. This is because all events are all slightly different with a standard measure. Sometimes the DFD will start or stop at different timings or reach the NADIR at a slightly different moment and the measure will so not be set up correctly.

Then most of the time, a mitigation measure that solves more violation is also more expensive (longer window of application, larger activated volume, …). Moreover, some mitigation measures solve a very limited number of violations which might be too limited to reach to solve sufficient violations to remain below the 30% ENTSOe threshold.

Elia thus needs to estimate a minimum percentage of DFD resolution to ensure that the selected mitigation measure would not be over-efficient and so over-expensive as well.

As working assumption, Elia took Q1 2022 as reference. At that time, Elia reached 36% of violations on a total of 286 DFDs.

Considering the aFRR mitigation measure for Downward DFD and the combination aFRR/mFRR mitigation measure for Upward DFD and using the percentage of resolution and the additional cost per action taken from the tables presented in 5.3 Sensitivity Analysis.

We crossed those results with the results from Chapter 3.5 on the combined model performances (Figure 7).

For Q1 2022, Elia needs to reduce its violation rates by 6% minimum to not exceed the 30% ENTSOe threshold. Dividing those 6% of improvement by the percentage of resolution gives us the number of "Good Actions to be taken when BE contributes".

In Figure 7 from chapter 3.5, we can derive from this value the corresponding percentage of actions that are taken when it's not needed.

E.g.: For the mitigation measure aFRR -7/7, we see on the figure 7 that a percentage of "Good action when BE contributes" of 13% corresponds to a "Confidence level" of 37% and for this "Confidence level", the curve "Act when not needed" indicated 7%.

The total number of actions taken is obtained by summing up the percentage of "Act when not needed" and the percentage of "Good action when BE contributes" multiplied by the amount of contributions of Elia (as a reminder, the "Act when not needed considers all DFDs while the "Good action when BE contributes" only consider the DFDs for which BE contributes) and leads us to the total cost for all actions.

Table 5 - Total cost for action per mitigation measure taking the model performances into account

	Mitigation	% of Resolution	Additional cost per action taken	Good Action when BE contributes	Act when not needed	Confidence level	% Total action taken	Total costs for all actions
Down	aFRR -7/7	47%	4.619,00 €	13%	7%	37%	12%	91.719,65 €
	aFRR -7/5	47%	4.118,00 €	13%	7%	34%	12%	81.771,28 €
	aFRR -7/2	40%	3.130,00 €	15%	7,50%	26%	13%	69.143,26 €
	aFRR -5/7	41%	3.212,00 €	15%	7,50%	36%	13%	70.230,24 €
	aFRR -5/5	41%	2.739,00 €	15%	7,50%	36%	13%	59.888,12 €
	aFRR -5/2	33%	1.807,00 €	18%	8%	35%	15%	45.009,19 €
	aFRR -2/7	24%	1.037,00 €	25%	10%	32%	19%	33.740,24 €
	aFRR -2/5	24%	693,00 €	25%	10%	32%	19%	22.547,72 €
	aFRR -2/2	**16%**	**216,00 €**	**38%**	**20%**	**25%**	**34%**	**12.391,23 €**
Up	aFRR/mFRR 70/30	52%	- 3.974,00 €	12%	7%	37%	11%	- 50.865,96 €
	aFRR/mFRR 50/50	**51%**	**- 4.917,00 €**	**12%**	**7%**	**37%**	**11%**	**- 63.395,64 €**
	aFRR/mFRR 30/70	46%	- 4.580,00 €	13%	7%	37%	12%	- 61.470,21 €

In conclusion (from the Table 5 here above),

- **<u>for downward DFDs</u>, the aFRR mitigation measure starting 2 minutes before the quarter hour and ending 2 minutes after the beginning of the quarter hour suits, our need to solve a sufficient number of violations**.

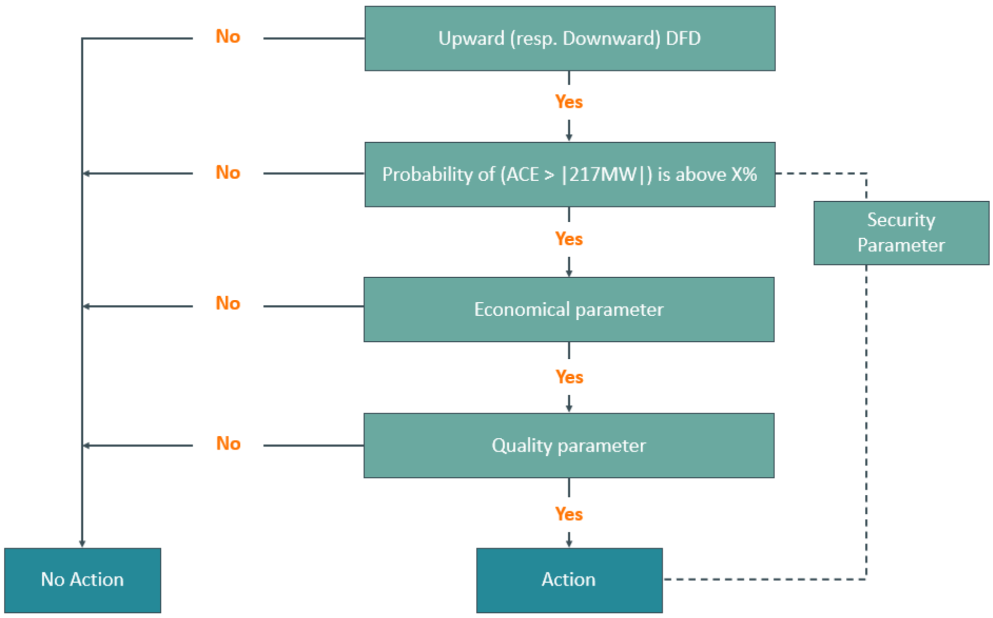
  The mitigation measure is so cheap (compared to the other ones) that it remains cheaper than other options even though Elia would apply it more often and so would increase the "Action not needed".

  The window of application of the measure might seem short but in fact, applying the measure on this small window allows to start deactivating the bids already activated in the wrong direction and/or activating the bids in the right direction. This gives the necessary impulse that is sufficient to solve some violations.

- **<u>for upwards DFD</u>, the combined aFRR and mFRR mitigation measure with the split 50/50 is the retained option.** A higher confidence level is even allowing which limits the number of "Action not needed".

Would the forecast performance increase in the future, the current conclusions would remain valid and the percentage of "Actions not needed" would simply decrease, what would further reduce the Total cost for action.

6. Decision tree



6.1 Upward (resp. Downward) DFD or not: DFD event forecast (model)

Elia uses the DFD event forecast for the last quarter hour before the quarter hour where DFD occurs to know whether or not a DFD (Upward/Downward) is forecasted for the next quarter hour.

6.2 Probability of (ACE > |217MW|) is above X% : ACE contribution forecast (model)

Using the output of the ACE contribution forecast model as such would lead to cover a very limited number of violations. If it follows a normal distribution, a forecast output below 217MW still implies a certain probability of the real ACE to be above the 217MW. That's this probability (P>X%) that Elia needs to consider in order to tackle a sufficient number of violations.

From Chapter 5.3.4, Elia knows that according to the mitigation measure selected, the X is 37% for upward DFDs and 25% for downward DFD. This implies, with the current model performances, to apply the mitigation measure for an ACE prediction above 157MW for upwards DFD and below -105MW for downwards DFD. Those MW values being linked to the RMSE of the model and based on a normal distribution, they'll be subject to change in case of model improvement.

6.2.1 Security Parameter = Probability of (ACE > 2*|217MW|) is above X%

In some cases of DFD prediction, whatever the other aspects, a very large ACE contribution might severely impact the network security. We fix this threshold at the value necessary to have a probability of an ACE > 2*217MW above X% (again, X is 27% for upward DFDs and 12% for downward DFD).

In those cases, Elia will take actions and apply the mitigation measure in any case.

6.3 Economical parameter = Cost for no action < cost for action

In normal circumstances Elia will compare the Cost for no action (meaning the cost for supporting the penalty from ENTSO-e) and the Cost for action (meaning the cost for applying the mitigation measure).

If the cost for getting the penalty is cheaper than the cost for applying the mitigation measure, then Elia will not take any mitigation measure and support the penalty (out of the financial impact, it would have a reputational impact,…).

Current evaluation of the Cost for no action versus Cost for action

Elia observes a total of 258 upwards DFD and 385 downwards DFDs so 643 DFDs over the studied period[11]. Over the same period, Elia had 63 violations for upwards DFD and 91 for downwards DFDs so 154 violations in total.

- Cost for no action: In Chapter 4.1, the cost per additional FCR MW of FCR over the period of 3 quarter is 245.550,00€. Under the hypothesis that the penalty would consist in contracting 58MW additional, it reaches a total of 14.241.900,00€ over the 3 quarter.

  It gives a Cost for no action of 92.479,87€/violation.

- Cost for action: By reusing the "Additional cost per action taken" calculated in Chapter 5.3, for the selected mitigation measure, it reaches the following:

	Downwards DFD		Upwards DFD	
	SQ	aFRR	SQ	aFRR_mFRR
		Start -2		aFRR 50%
		End 2		mFRR 50%
# violations	91	76	63	31
Total costs	€ 2.263.877	€ 2.283.553	€ 831.364	€ 521.592
Additional cost per action taken		€ 216		-€ 4.917

  Hypothesis: the cost is computed in activating the exact volume forecasted by the model.

So, even if it is slightly underestimate, the cost of the mitigation measure is clearly below the cost for no action (= the penalty).

In other words, it really deserves from an economical point of view to apply mitigation measures and so to work to fall under the 30% ENTSOe threshold instead of supporting the penalty cost.

In this decision tree, as long as the current hypothesis are valid, Elia will always move through the part of the decision tree for "Action".

6.4 Quality Parameter = (Current) percentage of violation (=Y) ≥ 15%

As a starting point, Elia fixes the quality target parameter at 15% in terms of percentage of violation in order to make sure Elia avoids the penalty and guarantees a certain quality of regulation at the European level (Y = 30% is the extreme value of the quality parameter as it corresponds to the 30% ENTSOe threshold).

This initial starting point is adaptable depending on our ambition in terms of quality regulation.

11

	Q3 2021	Q4 2021	Q1 2022	**Total**
f>50Hz	47	101	110	**258**
f<50Hz	92	117	176	**385**

The methodology to monitor our quality level though the time will evolve with the different implementation step as the complexity to monitor it in real-time might increase exponentially as detailed in Annex 9.6.

6.5 Action = applying the mitigation measure

If all the conditions of the decision tree are met, then the mitigation measure would be applied according to the conclusions of Chapter 5.3.4.

Considering the current ACE forecast performances, we analyzed which amount/volume (MW) of mitigation measure to apply.

As shown here below, it is more efficient to activate the volume forecasted by the "ACE contribution" forecast and not something else as a fixed volume.

In the following graph (Figure 11), we draw two different indicators:

- The percentage of ACE violation (full line) which is based on the hypothesis of the application of a perfect mitigation measure (meaning following exactly the ACE).

  The smaller the percentage of violations the better.

- The ACE deterioration (dashed line) where even though, according to the DFD forecast, we will almost always activate the mitigation measure in the right direction. But we might end up with an ACE that is evaluating after the application of the mitigation measure further from 0 than what the initial ACE was. It would so mean an ACE deterioration.

  E.g. If we forecast an ACE contribution of 100MW and that the real ACE was of 30MW. Then we will activate -100MW, so instead of facing an ACE of 30MW, we will have an ACE of -70MW after mitigation measure and so we deteriorated the ACE.

  The smaller the deterioration the better it is.

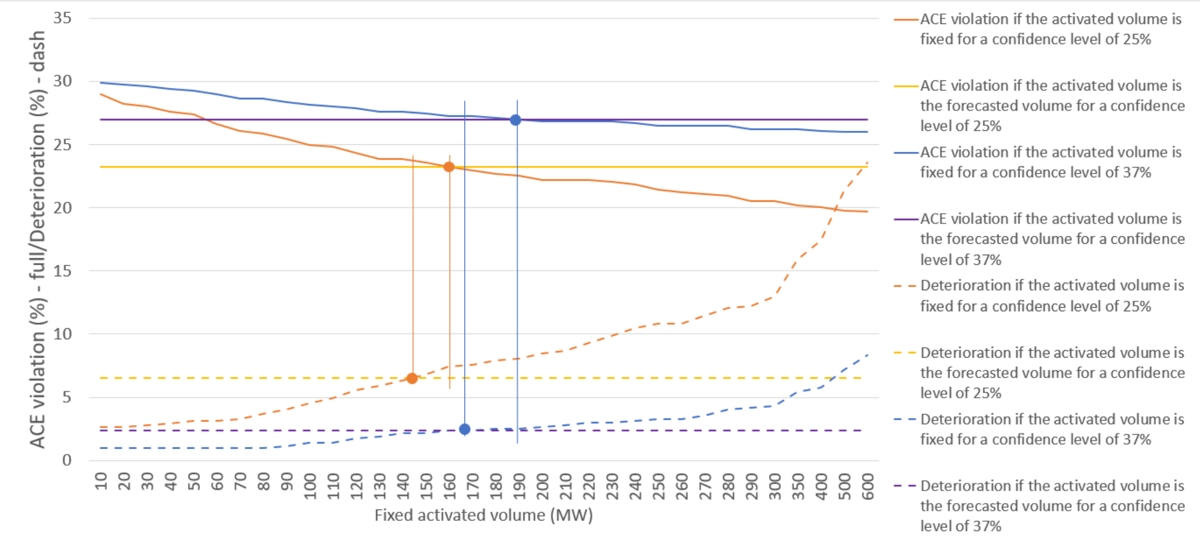


Figure 11 Percentage of ACE violation according to the volume activated for the mitigation measure

We compare:

- the blue (activation of a fixed volume) and purple (activation of the forecasted volume) curves (full and dashed) for the confidence level of 37%
- the orange (activation of a fixed volume) and yellow (activation of the forecasted volume)

curves (full and dashed) for the confidence level of 25%

So we are looking for if any point of the blue (resp. orange) curve has at the same time

- a smaller deterioration AND
- a smaller ACE violation

than the purple (resp. yellow) curve.

This would indicate that there is a fixed volume to activate that would ensure better performances than activating a volume equivalent to the ACE forecasted.

As a result, the graph shows that there is no fixed volume which provide better result than the ACE forecasted volume.

As a conclusion, our forecast is certainly not optimal yet; but it's still better to activate the volume based on the ACE contribution forecast than a fixed volume.

7. Proposal/Relevance of publications related to DFD's

According to statistical analysis, whenever a DFD is observed, the average of the ACE and System Imbalance remain in the direction of the DFD and so in the same direction for most of the quarter hour. This means that it would thus be in the interest of the Market Parties to help the system and to reduce the DFD contribution.

Under those considerations and taking into account the current performances of the DFD forecast that avoids false positive or false negative results, **Elia suggests publishing at least the DFD event forecast for the next 2 hours as indicator. This proposal has to be integrated to the implementation plan.**

The publication will contain, **thanks to a quarter hour refresh, the categorical DFD prediction per quarter hour as given by the forecast for the next 2 hours, knowingly "1", "0" or "-1", its degree of certainty per category and an indication on the data quality of the forecast**. Last but not least, the data will be made available via the Elia open data platform in the form of an API.

According to the current decision tree, any action taken by the market party will not prevent Elia to react in real time as the DFD forecast is one of the main elements of decision for Elia to take or not a mitigation measure. Nevertheless, all actions from the Market Parties will hold us away from the 15% threshold under which Elia takes no action.

Elia currently identifies no counter indications to the publication of the DFD forecast. As far as we know, Market Parties currently do not take any specific actions to prevent DFD contribution. We thus do not risk that Market Parties limit their reaction to cases where the forecast foresees a DFD event. Any action that could be taken by the Market Parties based on our publication would so only be beneficial.

The DFD forecast is not considered as an inside information and has no REMIT impact. A publication would nevertheless be triggered in case of failure of the publication. Moreover, the publication will only be informative and will come with the necessary disclaimer to prevent any implicit transfer of responsibility from any Market Party towards Elia linked to the publication.

The implementation of such publication will be discussed in Chapter 8 with the general recommendations for implementation.

The publication of the ACE contribution forecast is not recommended. Indeed, its current quality (RMSE,..) is not considered as sufficient to be published.

8. Recommendation

Elia recommends a **three-step approach** towards the application of the mitigation measures:

- **Step 1 Improvement:** Considering the current results of the DFD event and ACE contribution forecasts, Elia advises to continue the investigation during a period of at least one year, in order to improve those models predictions.

  The following elements could be investigated further in order to improve the forecasts:

  i. New input dataset, e.g. detailed production and load from other TSOs
  ii. Feature engineering
  iii. New model families
  iv. Change the level confidence – risk- approach

  The target forecasts will aim to reduce the non-needed actions (and so the sunk costs), to improve the percentage of resolutions and the mitigation measure types (activated volume, start time, end time, …) under some pre-defined quality targets.

  The conclusions about the selected type of the mitigation and their technical characteristics, will be subject to revision after this improvement period.

  After this additional year of investigation, a gate (GO/NO-GO) must be fixed with CREG and Market Parties to decide on the industrialization.

- **Step 2 Industrialization:** If positive (GO), Elia will plan **an industrialization of the forecasting models**, foresee an adaptation of the regulated document related to the implementation and foresee the necessary operational processes to apply the mitigation measure.

  The timeline for industrialization will be quantified at that time depending on the final product (dataset, model selection, complexity, IT performance, complexity, operational process,…). A comprehensive timeline will be provided after further alignments with other projects/initiatives that are ongoing or are planned at that time.

  This industrialization will include **a publication at least for the DFD event forecast** because, as already mentioned,

  i. the DFD forecast predicts very limited false positive and false negative
  ii. any action from market Parties could have a positive impact on our DFD contribution (as long as it stays in the same direction as the DFD)

  If negative (NO-GO), within 6 months, Elia will propose a new forward to agree on with CREG and Market Parties at that time.

- **Step 3 Evaluation Report:** After 1-year in production, Elia will provide an Evaluation report to closely monitor the effectiveness of the mitigation measure.

  i. An operational feedback on the application of the measure
  ii. An efficiency feedback on the performances of the predictions, on the performances of the mitigations measure, on the status with regards to the ENTSOe target and on an approximation of the costs incurred.
  iii. Recommendations on the follow up of the mitigation measure application and on the continuous developments.

For the rest, the present consultation will be used as an opportunity to receive feedback from stakeholder and the recommendation which would help to finalize the study.

9. Annex

9.1 State of the art

9.1.1 Data analysis and selection

9.1.1.1 Correlation analysis

Correlation analysis deals with association between two or more variables. Correlation analysis can be performed between both continuous and categorical feature variables however the approach is different.

Pearson Correlation coefficient measures the strength and direction of linear relationship between two continuous feature variables. The value ranges between –1 and +1. A Pearson correlation coefficient:

- greater than 0 is positive correlation: when one feature changes, the other one changes in the same direction
- smaller than 0 is negative correlation: when one feature changes, the other one changes in the opposite direction.
- equal to 0 represents no correlation.

For correlation analysis between a continuous and categorical feature, a simple box or violin plot can be used to determine the relationship between them. To quantify the relationship between such features, statistical tests such as Mann-Whitney U test can be used.

Note that Pearson correlation can be a flawed metric to determine if a variable should be included as a predictor or not, especially when working with categorical predictions such as the DFD ones as shown at Figure 12 and Figure 13:

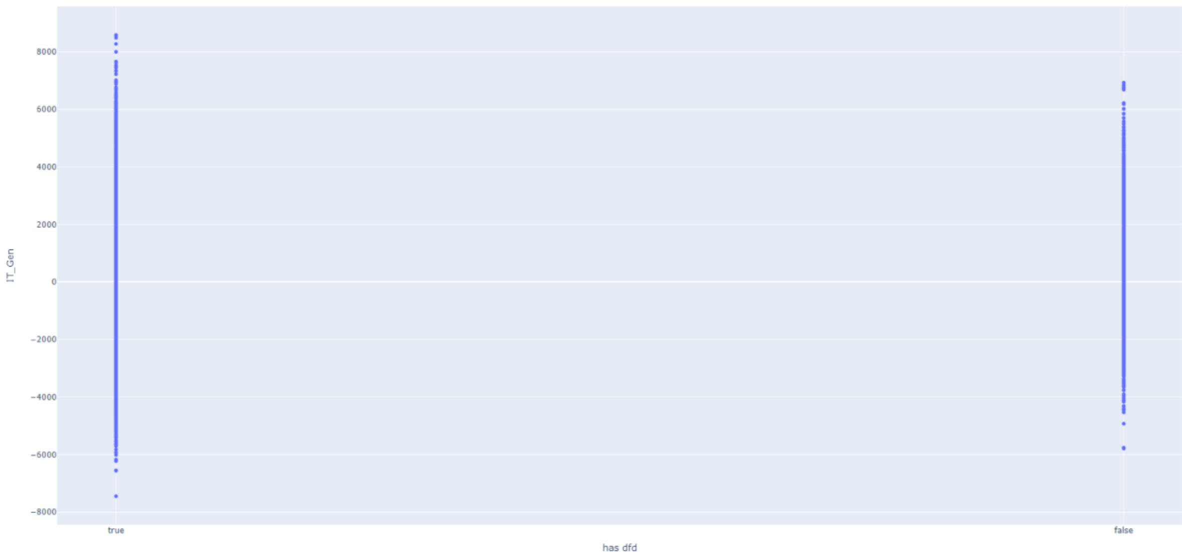


Figure 12 - DFD events vs the generation in Italy
(x-axis: Has a DFD? True/False, y-axis: Italian Delta Generation)

From Figure 12, it is possible to derive a correlation, but even though it mathematically can, it might not tell the same story as the violin plot in Figure 13. There we can really see the distribution difference from the Italian generation when there is a DFD and when there is no DFD. Since the distributions are quite different, a model can probably use this variable to increase its accuracy.

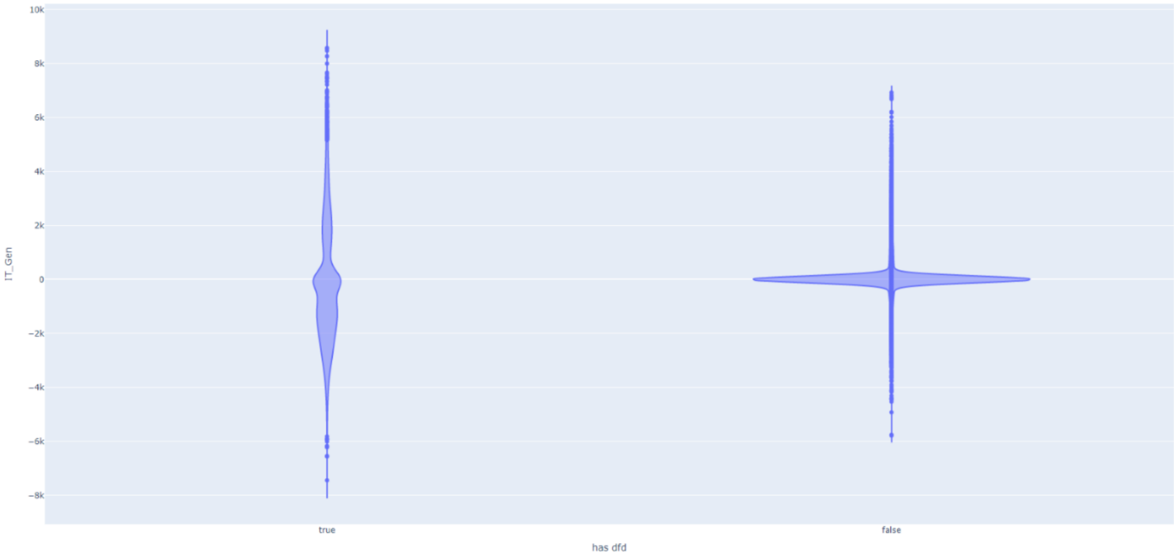


Figure 13 - DFD events vs the generation in Italy violin plot
(x-axis: Has a DFD? True/False, y-axis: Italian Delta Generation)

Correlation can be indicative but should always be cross checked with other measures. Figure 14 is a nice example why correlation doesn't always tell the whole story. All graphs have the same correlation, but totally different behavior. Elia will take care to investigate each variable in detail and will not derive the usefulness of a variable on only 1 indicator.

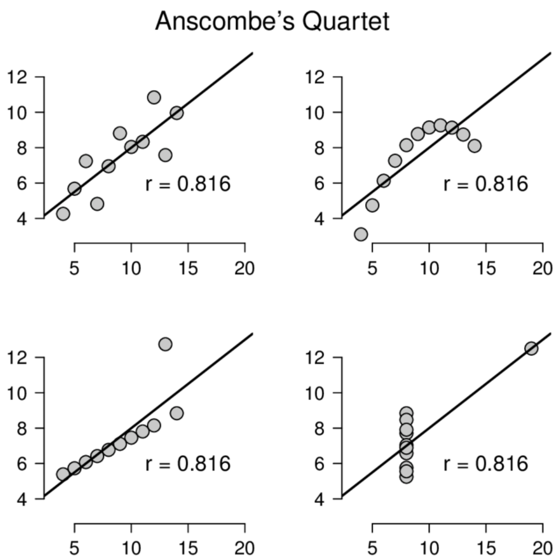


Figure 14 - Anscombe's quartet

9.1.1.2 BorutaSHAP analysis

Boruta analysis consists of the following steps:

- Copy all features and name them "shadow" + "feature_name".
- Shuffle these newly added features to remove their correlation with the target variable.
- Run a classifier on the extended data with the shadow features included.
- Now let the classifier rank the features using an importance measure. Specifically, for

BorutaSHAP, this will be the Z-score of the SHAP[12] values.

- Create a threshold as the maximum value of the shadow features.
- If a feature exceeds this threshold, it gets a hit.

Finally, we need to know how many hits are needed to be relevant. After all, a hit could just be a coincidence. This is where the number of trials comes in. Suppose we don't know whether a feature is useful or not, so we estimate the probability it is at 50 %, or equal to a coin toss.

In our case, we did 100 trials which can be compared to tossing a coin 100 times (i.e. a binomial distribution). The probability of having more than 58 hits is only 5%, so we use that as a threshold. If a feature has less hits, it is probably not useful. Other thresholds can also be set to define when we are hesitant or when we are sure a feature is not useful. They are all based on the binomial distribution.

9.1.1.3 Principal Component Analysis

Principal Component Analysis (PCA) is a statistical technique allowing to reduce the dimensionality of the dataset. It's accomplished by linearly transforming the data into a new orthogonal set of variables containing most of the variance of the initial dataset.

The principal components are obtained by an eigenvalue decomposition of the data covariance matrix of the predictors. Each successive component tends to maximize the remaining variance of the dataset while staying orthogonal to the previous one.

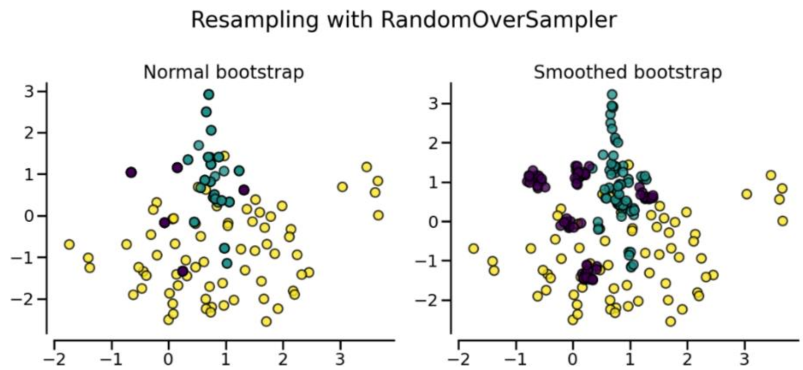
9.1.1.4 Recursive Feature Extraction

Recursive Feature Elimination (RFE) is a backward feature selection method. Starting from a model built on the entire set of predictors, the algorithm removes at each iteration the least important variable according to a defined feature importance criterion. The algorithm stops after reaching the desired number of features.

9.1.2 Imbalance Class handling

9.1.2.1 Naïve random oversampling

One way to fight this issue is to generate new samples in the classes which are under-represented. The most naive strategy is to generate new samples by randomly sampling with replacement the current available samples. Another slightly more advanced technique is to slightly disperse the newly generated samples. This dispersion can be controlled by a para meter and can be described visually as follows:



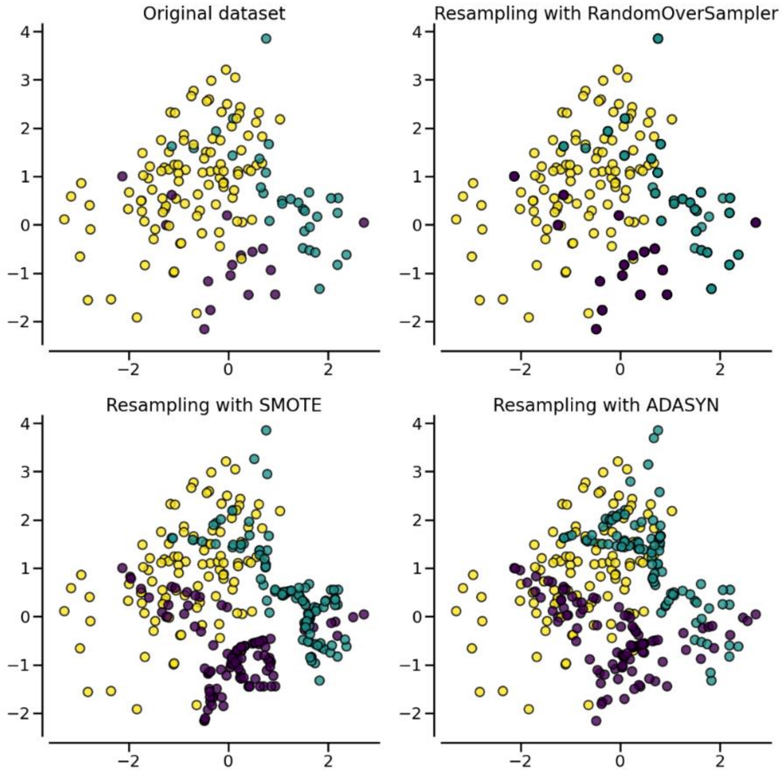
In the smoothened bootstrap you can see that the new samples differ slightly from the old ones.

---

[12] SHAP(SHapely Additive exPlanations) calculates the average marginal contribution of each feature across all permutations at a local level

9.1.2.2 Smote(nc) & adasyn

Synthetic Minority Oversampling Technique or SMOTE and Adaptive synthetic sampling are two popular techniques that are often used to counter imbalances. A visual description of the techniques can be seen here:



While the RandomOverSampler is over-sampling by duplicating some of the original samples of the minority class, SMOTE and ADASYN generate new samples by interpolation. However, the samples used to interpolate/generate new synthetic samples differ. In fact, ADASYN focuses on generating samples next to the original samples which are wrongly classified using a k-Nearest Neighbors classifier while the basic implementation of SMOTE will not make any distinction between easy and hard samples to be classified using the nearest neighbors rule. Therefore, the decision function found during training will be different among the algorithms.

Another important note here is that SMOTE and ADASYN don't work on categorical data. In (Anon., n.d.)our case we have categorical data, mainly time components (month, day, hour, minute). Since they are numeric, we will still try the techniques on the categorical data and consider it numerical.

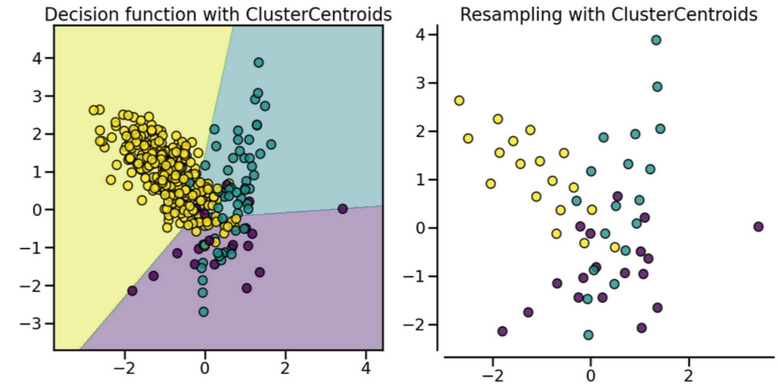
For SMOTE there is a possibility to handle categorical data which is called SMOTENC. Here categorical data are treated differently. When a new sample is generated, each categorical feature value corresponds to the most common category seen in the neighbors belonging to the same class.

All credits of these algorithms and explanations go to https://imbalanced-learn.org/stable/index.html.

9.1.2.3 Undersampling with cluster centroids prototype generation

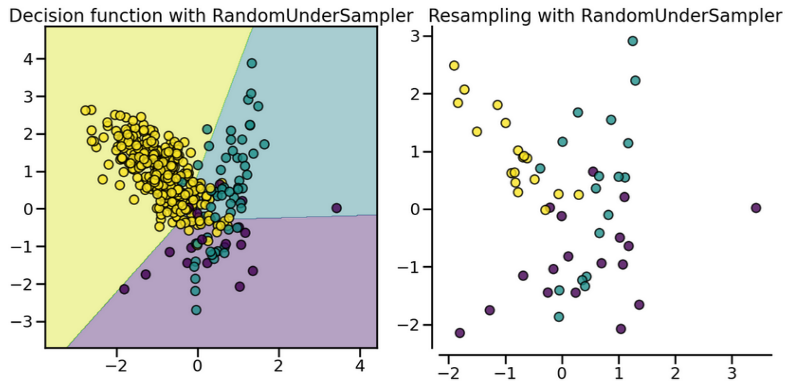
Given an original dataset S, prototype generation algorithm will create a new set S' which is not a subset of S and is smaller than S. In other words, prototype generation technique will reduce the number of samples in the targeted classes but the remaining samples are generated, and not selected, from the original set.

Cluster centroids makes use of K-means to reduce the amount of samples. Therefore, each class will be synthesized with the centroids of the K-means method instead of the original samples. The figure below illustrates such under-sampling:



9.1.2.4 Undersampling with random undersampling

An easy and fast technique where samples are randomly drawn to form a subset of the overrepresented class. The samples can be either drawn with or without replacement.



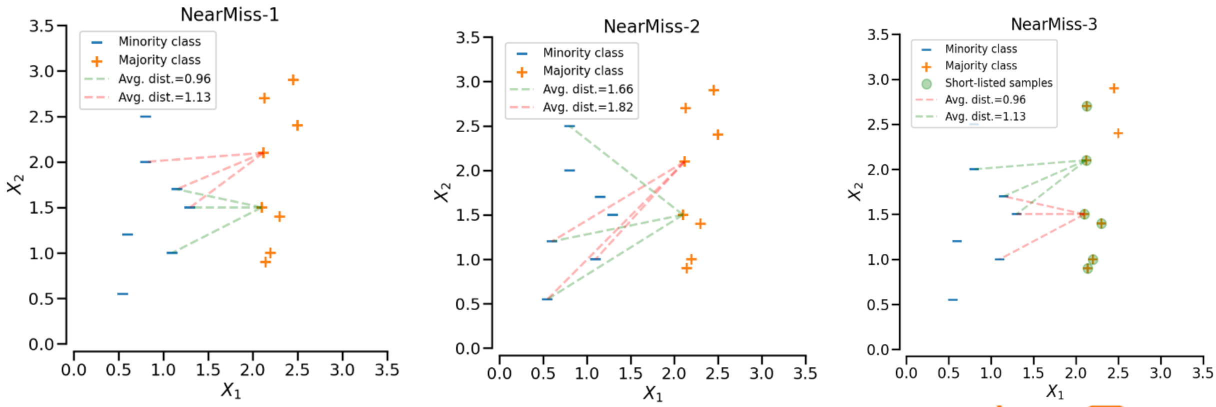
9.1.2.5 Undersampling with nearmiss

Let "Positive samples" be the samples belonging to the targeted class to be under-sampled. "Negative samples" refer to the samples from the minority class (i.e., the most under-represented class).

NearMiss-1 selects the positive samples for which the average distance to the N closest samples of the negative class is the smallest.

NearMiss-2 selects the positive samples for which the average distance to the N farthest samples of the negative class is the smallest.

NearMiss-3 is a 2-steps algorithm. First, for each negative sample, their nearest-neighbors will be kept. Then, the positive samples selected are the one for which the average distance to the nearest-neighbors is the largest.



9.1.3 Models

9.1.3.1 Linear regression

In a linear regression, the dependent variable y, i.e. the variable to predict, is described linearly by a weighted sum of the predictors $(x_1, x_2, \dots, x_n)$:

$$\hat{y} = a_0 + a_1 x_1 + a_2 x_2 + \cdots a_n x_n$$

Where $\hat{y}$ is the predictors and $a_0,\ a_1, \dots,\ a_n$are the model parameters.

The coefficient are generally retrieved by minimizing the residual sum of squares:

$$RSS = \sum_{i=1}^{n} (y - \hat{y})^2$$

Limitations:

- Linear regression assumes that all the predictors are independent. Features selection is then essential.
- Linear regression, by nature, only looks at linear relationship between the variable to predict and the predictors.
- Linear regression is sensitive to outliers and collinearity.

9.1.3.2 Logistic Regression

The logistic regression is a supervised machine learning algorithm used to estimate the probability that a certain event belongs to a specific class. The logistic regression uses a Softmax function to map the linear regression predictions and the probabilities of the event to belong to each one of the classes.

More precisely, for each of the classes y_i, a respective linear regression model $\widehat{y_i}$is trained. The outputs of those linear regressions give scores for each category. The Softmax function then takes the exponential of every score and normalizes it by the sum of all exponentials to retrieve the probabilities $\widehat{p_i}$:

$$\widehat{p_i} = \frac{\exp(\widehat{y_i})}{\sum_{k=1}^{n} \exp(\widehat{y_k})}$$

Limitations: As for the linear regression, this model assumes the independence of the predictors. It is sensitive to outliers and ill-fitted for non-linear problems by construction. Moreover, over-fitting of the training set may arise when considering high dimensional dataset.

9.1.3.3 Artificial Neural Network (ANN)

For the ANN, a relatively simple architecture was selected. The input layer of the ANN has one neuron per variable with a REctified Linear Unit activation function (RELU). The hidden layer has half of the neurons in the first layer with RELU activation. Finally, there is one output neuron with RELU activation. The presented architecture in Figure 15 provides reasonable results while maintaining computation cost of model fitting limited.

Since we are not sure yet that this is the optimal setup, Elia holds the freedom to change the architecture of the ANN if it increases performance of predictions and if it is computationally feasible. i.e :

- Adding more layers to the architecture so the model can catch more complex patterns. Care will be taken that performance on train and test set are similar, to make sure we don't overfit.
- Adding more/less neurons to layers.
- Considering different activation functions

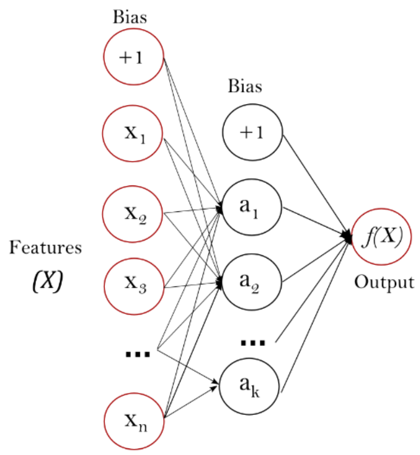


Figure 15 - Artificial neural network architecture used

Theoretically an ANN is a universal function approximator and should be able to generalize on the data to give reasonable predictions. The more neurons, the more accurate the approximation in general will be. But it also increases risk of overfitting and computational complexity.

9.1.3.4 Support Vector Machine

Support Vector Machine is a supervised machine learning model that analyses data for classification and regression. The algorithm builds hyper-plane(s) in a high or infinite dimensional space that separate with the largest distance, the nearest training point of any class. For binary classification, this is done by solving the following primal problem:

$$\min_{w,b,\zeta} \frac{1}{2} w^T w + C \sum_{i=1}^{n} \zeta_i$$

Subject to

$$y_i(w^T \phi(x_i) + b) \leq 1 - \zeta_i$$

$$\zeta_i \geq 0, i = 1, \ldots, n$$

Where

- $\frac{2}{w^T w}$ is the margin between the class
- $C \sum_{i=1}^{n} \zeta_i$ is the penalty term for samples away of a distance ζ_i of their margin boundary
- $\phi(x_i)$ is the kernel function that map the predictors x_i into a higher dimensional space

Multiclass problem is handled by breaking it into multiple binary classification problems (one-to-one or one-to-rest approach).

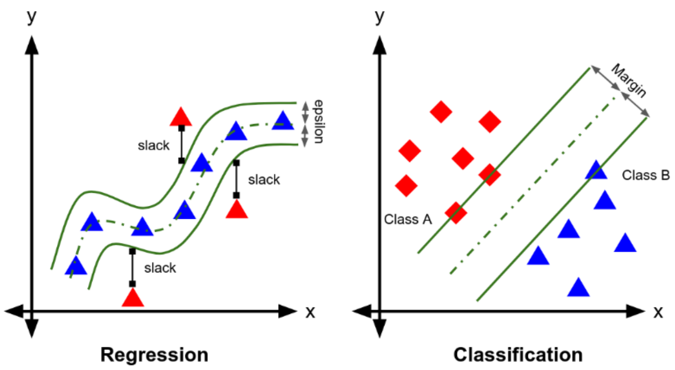


Figure 16 - Illustration of Support Vector Machine regression and classification

For regression, the mathematical formulation is relatively close, except we aim to find the model that minimizes the error of regression considering an error tube ϵ.

SVM is not suited for large data set. Its performance gets also poorer when data is more noise and classes are then overlapping

9.1.3.5 Random Forest

Using a single decision tree when predicting DFD occurrences or ACE contributions may lead to overfitting. Decision trees are good to explain the train data but are not very good at generalizing and predicting unknown data. To handle this shortcoming and to make the predictions more resilient, random forest were invented.

In a random forest a whole bunch of decision trees are trained on bootstrap samples of the data. Each tree only sees a limited subset of features and datapoints. One of those trees is not a strong predictor, but a whole bunch of them (and thus the name forest) offer reduced bias and variance when predicting. In the picture below you can see this more visually:

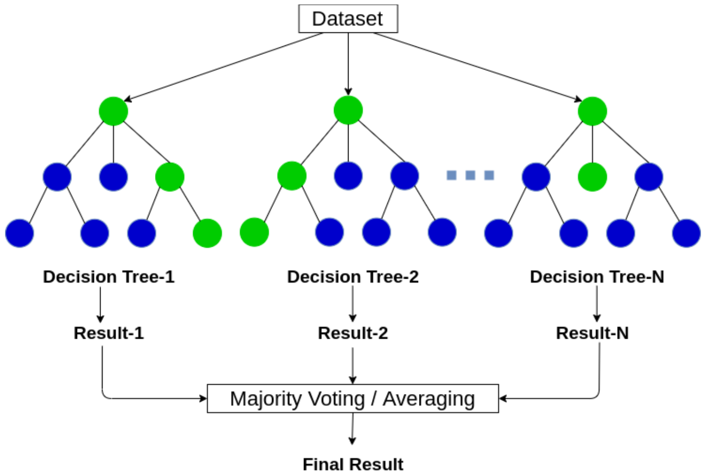


Figure 17 - Illustration of Random forest structure

In our DFD occurrence prediction, the class that receives the most votes will get chosen. For the ACE contribution, the average of a branch will be taken when a selected number of samples is left in the branch.

An advantage of the random forest is that it provides some indication about the certainty of the model. If 90 % of the trees are predicting a certain outcome, it is probably more likely than when only 55 % of the trees would predict.

However, it is not all glitter and glory. Random forests have also some disadvantages. In theory you can exactly follow how they made a decision, in practice they are quite a gray box. It is not really feasible to determine easily how they came to a specific prediction.

Another disadvantage of the random forest model is that it takes considerably more time to train than a normal decision tree. This is to be expected since it consists in our case of 100 decision trees.

A lot of parameters of the random forest can be set; how deep the trees should be, what the minimum samples in a leaf should be, how many trees to use, etc. As new insights are discovered daily, those parameters might need to change. In the results below we give the performance based on a fixed set of parameters. They might be outdated soon and replaced by better ones. Elia will hold the freedom to change them if it makes the model more accurate.

9.1.3.6 Gradient boosting

Another type of ensemble models are the gradient boosting type models. The principle here is the following:

- Start with an initial prediction. For regression in general this is the mean.
- Calculate the residuals, i.e. the errors for each datapoint.
- Use a decision tree to predict the residuals.
- Combine the initial prediction and the prediction of the first decision tree to get a new prediction.
- Get the residuals of the latest prediction and predict them with a new decision tree.

In this way in general around 100 decision trees are stacked.

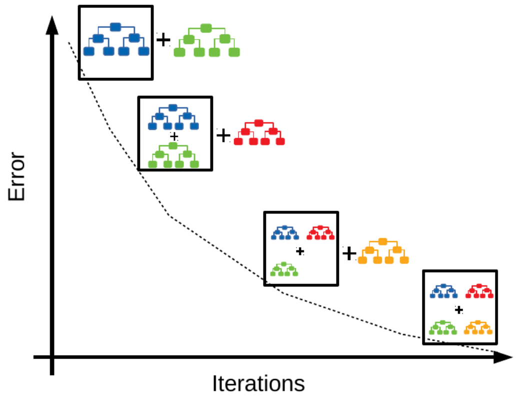


Figure 18 Gradient boosting principle

To train the stacked decision trees, data need to be sorted. This process is quite computationally expensive. To solve this issue, the continuous data is distributed in equally sized bins. In general, we consider 255 bins for data and 1 to indicate NaN values (Not a Number or NaN are unidentified of values with issues). This doesn't hurt performance and speeds up the training process enormously.

Gradient boosting can catch non linearities quite easily and can work with data that consists of different data types and NaN values. When combined with the very decent training speed, they often are a good contender as a best model.

9.2 Details on data selection

9.2.1 DFD event forecast

9.2.1.1 Correlation analysis

When looking at the partial autocorrelation of DFD events from Figure 19, Elia observes that:

- DFD events are reasonably auto-correlated.
- The highest partial autocorrelation (only looks at the direct influence of a variable, contrary to normal correlation) is with DFD events from 4 quarters ago, so the previous hour.
- There is also a relatively strong peak in partial correlation from 96 quarters, or 1 day ago.

So the fact that there was a DFD one hour or one day ago is included as a predictor for a current DFD.

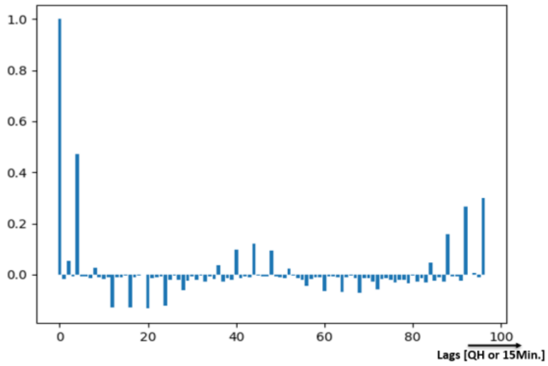


Figure 19 - Partial autocorrelation of DFD occurrences

Figure 20 shows the linear correlation between DFD occurrence and the gradient of load, generation and net position up to 8 quarters lookback horizon.

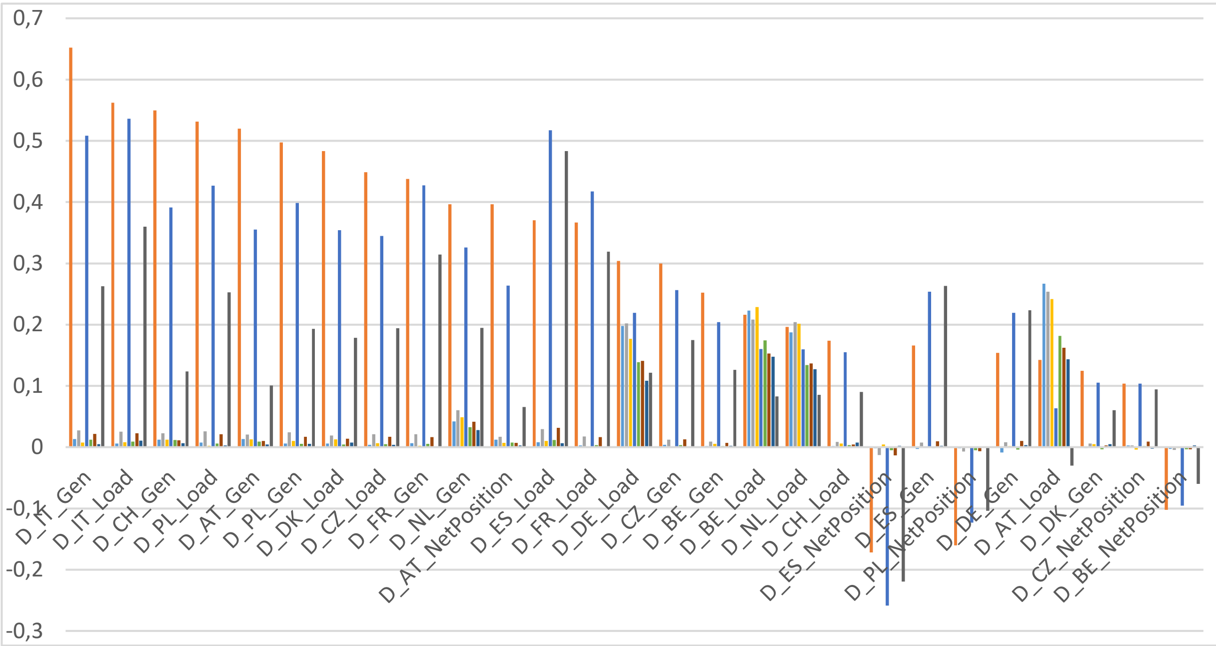


Figure 20 - SA data linear correlation with DFD occurrence up to 8 QH in the past

For each feature, spikes are observed at QH-4 and QH-8 due to the auto-correlation of the dataset and the DFD occurrence. Correlations with previous quarters are observed for the NL, BE, AT and DE load because of the 15 min. granularity of those predictors.

Regarding Elia data and so the correlation with the generation of specific Belgian production units (graph not shown for anonymization), we observe a correlation with some specific units and forecasted load. This is expected from the significant contribution of Elia to the DFDs. Again, spikes were observed at QH-4 and QH-8 due to the auto-correlated characteristic of the DFD occurrence.

9.2.1.2 BorutaSHAP analysis

Using BorutaSHAP (theory in Annex 9.1.1.2) with a histogram based gradient boosting classifier on the current data for 100 trials (for computational feasibility) resulted in 9 important features with their importance as below:

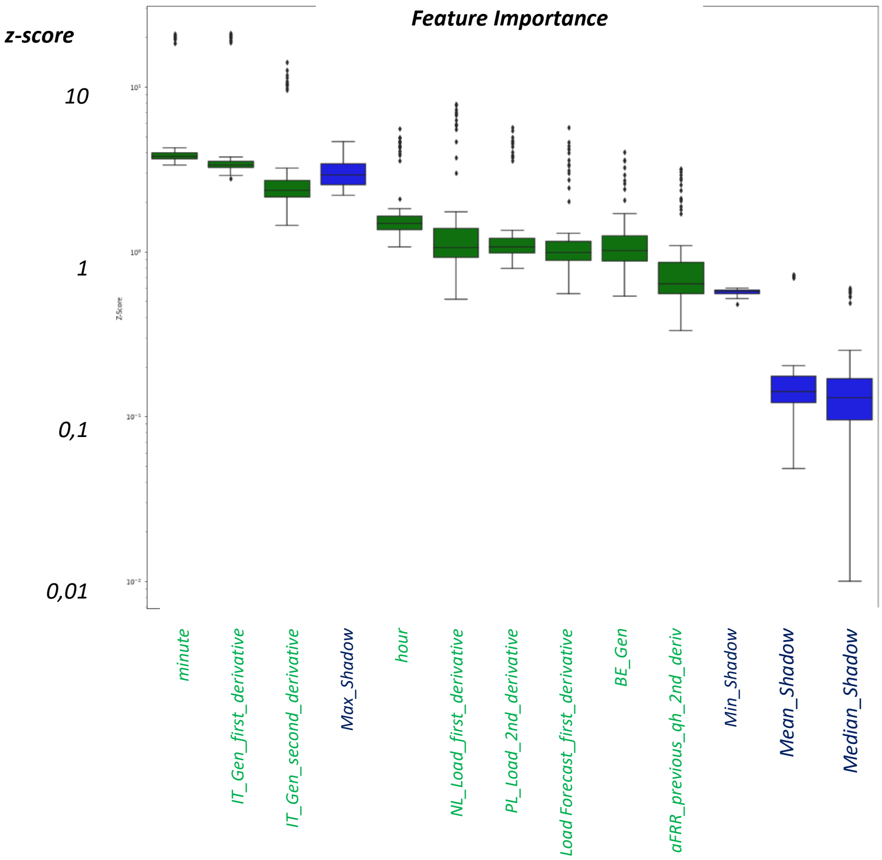


Figure 21 Borutashap accepted features

Important to note is that BorutaSHAP is model dependent. In this case we used a histogram based gradient boosting classifier as it trains fast and can catch non linearities. It differs from a normal gradient boosting algorithm in the sense that it categorizes continuous data in 256 bins (i.e. 1 byte) to enhance calculation speed while preserving accuracy. Calculation time is an order of magnitude (≈100 times) faster than normal gradient boosting.

9.2.1.3 PCA

We applied PCA in the following fashion:

- Categorical data stayed untouched (i.e., datetime components in this case)
- Other variables got reduced dimensionality by applying principal component analysis

The explained variance in percentage by number of components can be seen in Figure 22:

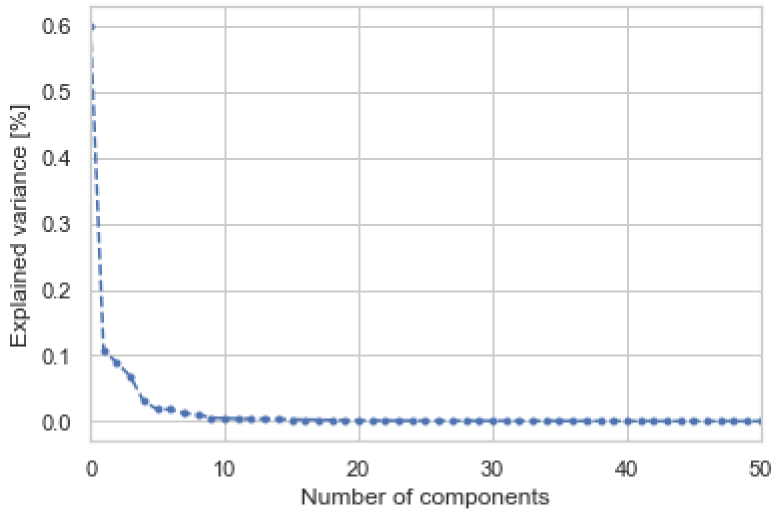


Figure 22 PCA analysis of all features

From Figure 22 we can observe that around 10 to 15 PCA components is enough to explain almost all the variance in the data. One possibility is to transform the ~500 input features with PCA to only 15 abstract features to train the model on. This would be quite beneficial to the model as it will speed up train time and increase stability.

A notable disadvantage of PCA however is its explainability. It is notably difficult to interpret the components into something with physical meaning. We also saw that BorutaSHAP only selected 9 features from the feature set we gave it. If we use only those 9 features, PCA will not have much use. For this reason, we won't apply PCA right now.

9.2.2 ACE forecast

9.2.2.1 Correlation analysis

There should not be auto-correlation as we look for an instantaneous event. We nevertheless still test the correlation of the AVG ACE of previous QH.

The most linearly correlated features are summed up in the bar chart below:

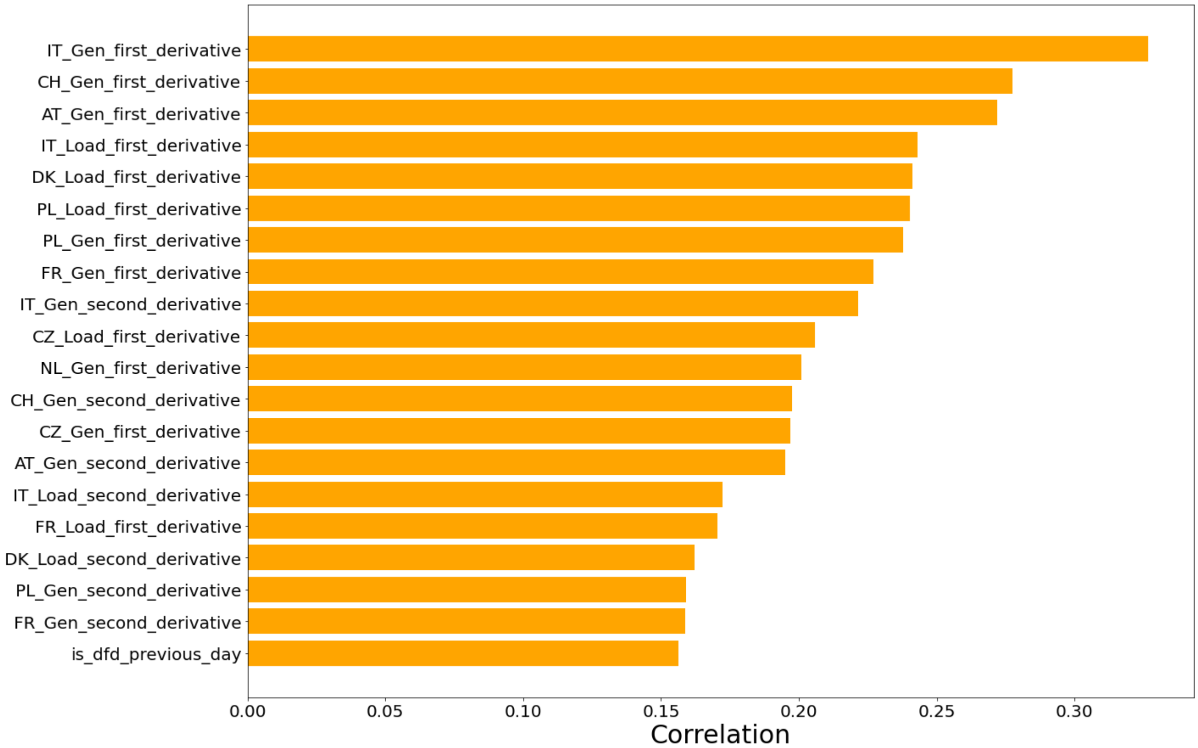


Figure 23 - Features linear correlation with Elia ACE contribution

We see that the first derivative of the total generation in Italy has the highest correlation with Belgian ace from all the variables we tested. A correlation of around 0.35 is not very high, but it is the highest out of all our variables. Note that high or low correlation doesn't say anything about variable importance. For example, adding 2 variables with both a very high correlation might introduce collinearity and deteriorate model performance. Vice versa a model with a very low correlation might still hold valuable information, just not in a linear fashion.

9.2.2.2 BorutaSHAP analysis

Like in the DFD analysis, Borutashap has been performed for 100 trials using a histogram based gradient boosting regressor. The 100 trials give as again a very reasonable amount of certainty about the variables selected. The histogram based gradient boosting is chosen because it is very fast to calculate and can catch non linearities.

The accepted variables (in green) are shown by order of importance on Figure 24. The z-scores of the shadow features, the permuted columns that serve as a reference are also indicated in blue on the figure.

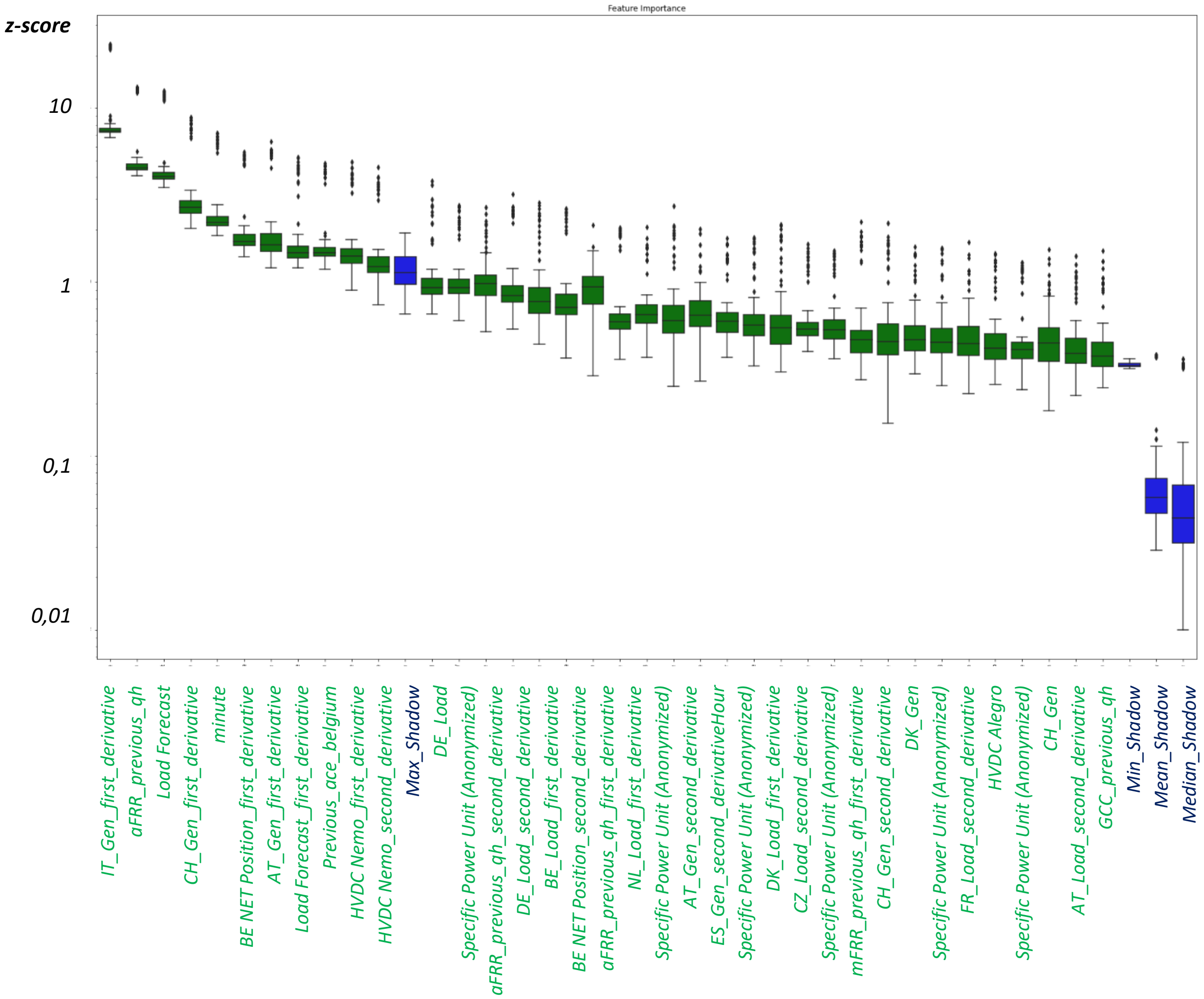


Figure 24 Borutashap ace accepted features

9.2.2.3 PCA

PCA analysis is independent of the model and the variable to predict. It was therefore performed only once on the whole input dataset and is presented in Chapter 9.2.1.3.

9.3 Details on Imbalance Class handling for DFD event forecast

To compare imbalance class handling technique, we consider a fast-training model; the histogram gradient boosting classifier. The classifier is a computationally optimized version of gradient boosting that uses binning to categorize continuous variables. The variables get divided in 256 bins, which speeds up calculations tremendously. Using 256 bins allows the algorithm to use 1 byte per continuous variable to store its information. Categorizing a continuous variable in 256 bins doesn't seem to hurt accuracy but speeds up calculation by an order of magnitude (in our case ≈ 100 times faster). This in turn enables us to do a lot of quick testing with good accuracy. Note that in reality only 255 bins are used to categorize continuous variables, and 1 bin to store if the variable is NaN or not.

We assume that the resampling method will yield similar results for other models. So basically, we tested the imbalance technique on one model. Theoretically, the methods should be tested on every model but it will be highly computational. We selected one fast model (gradient boosting) and we assume the conclusion will be similar for other models.

After doing a hyperparameter sweep (see Chapter 9.4.1.2), the optimal parameters for the hist gradient boosting seemed to be a depth of 6 and a L2 regularization of 13.78. The results for the different resampling techniques are summarized in Table 6. All testing was done on the variables deemed important by BorutaSHAP. The data ranged from 2021-04-19 to 2022-04-19.

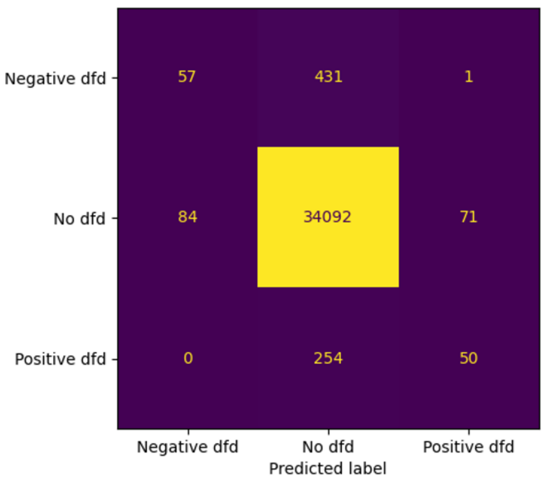
Table 6 - Resampling methods results

Resampling method	F1-score			
	Negative DFD	**No DFD**	**Positive DFD**	**Weighted average**
No resampling	0.181	0.988	0.235	0.468
Naïve random oversampling	0.343	0.969	0.338	0.550
Naïve random oversampling with shrinkage	0.234	0.988	0.273	0.498
Smotenc	0.328	0.963	0.282	0.524
Adasyn	0.313	0.958	0.272	0.514
Undersampling clustercentroids	0.138	0.690	0.050	0.293
Random undersampling	0.225	0.905	0.158	0.430
Undersampling nearmiss v1	0.058	0.468	0.056	0.194
Undersampling nearmiss v2	0.041	0.069	0.038	0.049
Undersampling nearmiss v3	0.037	0.425	0.051	0.171

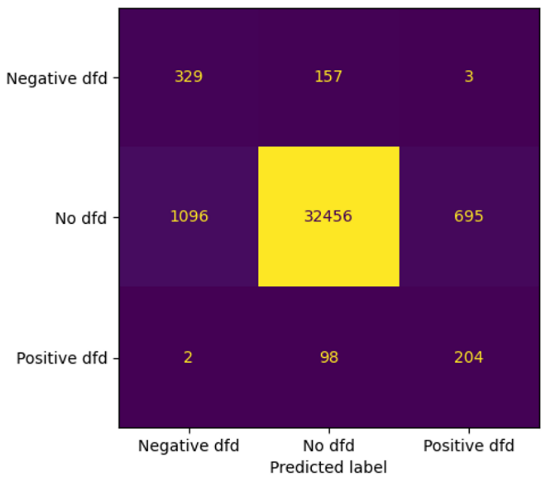
9.3.1.1 Confusion matrix and complete results

In the images below you can see the confusion matrixes of the different resampling techniques that were tried in combination with a gradient boosting model. The X-axis shows the predicted labels from the model, the Y-axis the actual label. A perfect predictor would have only numbers on the diagonal.

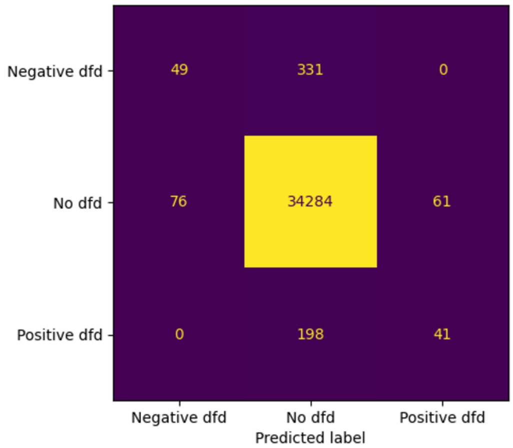
No handling



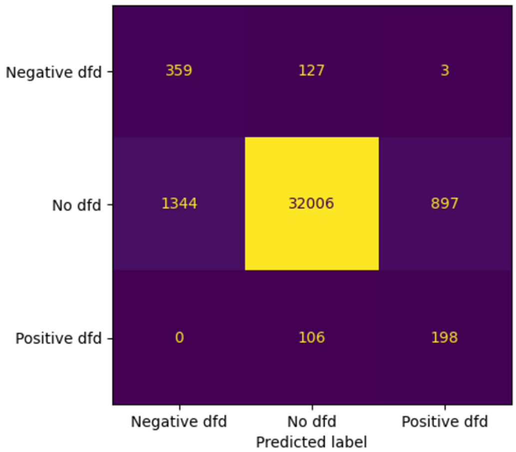
Naïve random oversampling



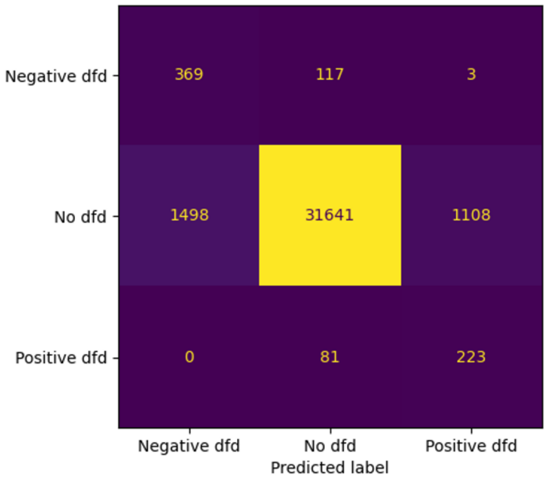
Naïve random oversampling with shrinkage



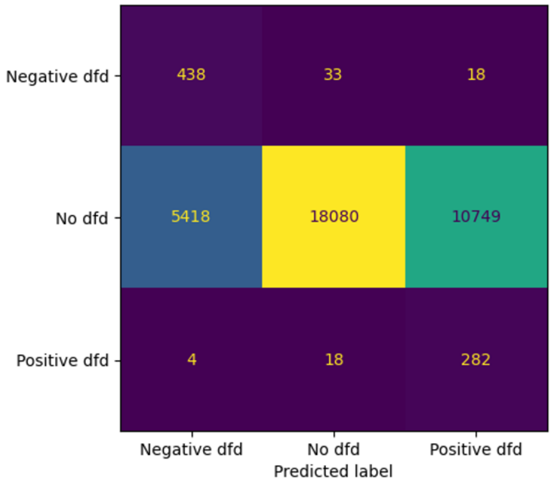
Smotenc



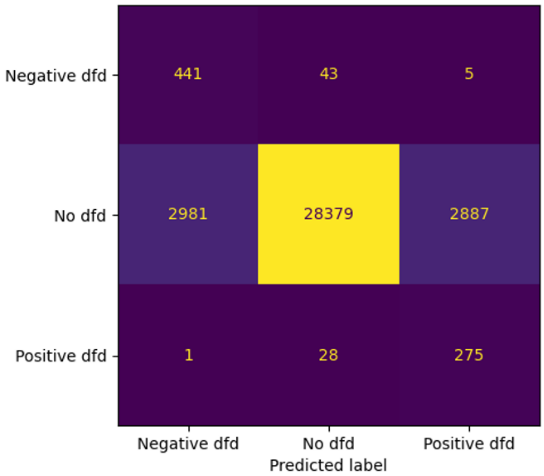
Adasyn



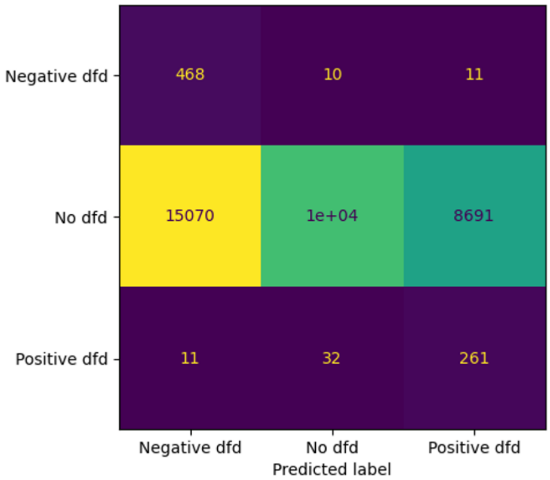
Undersampling clustercentroids



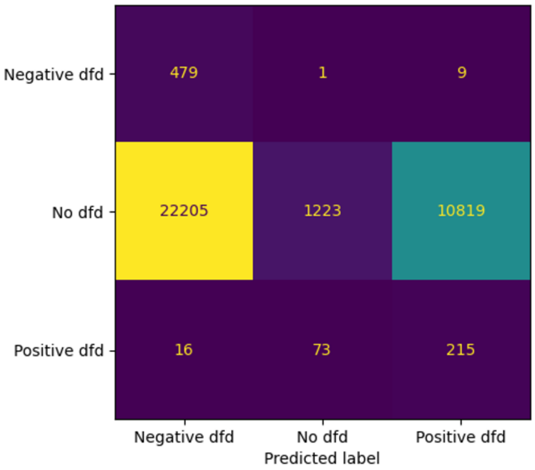
Random undersampling



Undersampling nearmiss v1



Undersampling nearmiss v2



Undersampling nearmiss v3

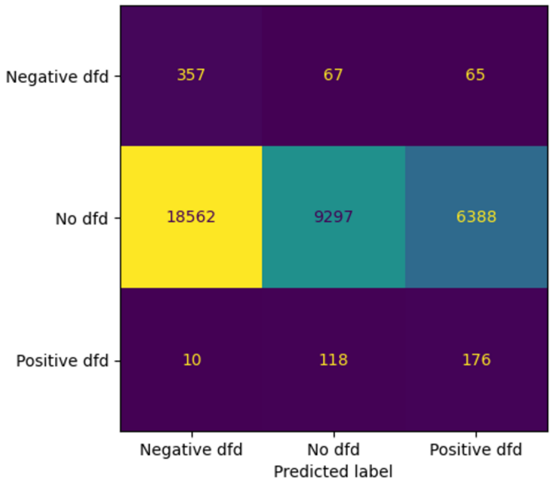


Table 7 - Sampling techniques metrics overview

	Precision					Recall					F1 score					
	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Overal accuracy
No resampling	0,404	0,980	0,410	0,598	0,967	0,117	0,995	0,164	0,426	0,976	0,181	0,988	0,235	0,468	0,970	0,976
Naïve random oversampling	0,231	0,992	0,226	0,483	0,975	0,673	0,948	0,671	0,764	0,941	0,343	0,969	0,338	0,550	0,955	0,941
Naïve random oversampling with shrinkage	0,456	0,981	0,427	0,621	0,969	0,157	0,995	0,201	0,451	0,976	0,234	0,988	0,273	0,498	0,971	0,976
Smotenc	0,211	0,993	0,180	0,461	0,975	0,734	0,935	0,651	0,773	0,929	0,328	0,963	0,282	0,524	0,948	0,929
Adasyn	0,198	0,994	0,167	0,453	0,975	0,755	0,924	0,734	0,804	0,920	0,313	0,958	0,272	0,514	0,943	0,920
Undersampling clustercentroids	0,075	0,997	0,026	0,366	0,976	0,896	0,528	0,928	0,784	0,537	0,138	0,690	0,050	0,293	0,677	0,537
Random undersampling	0,129	0,998	0,087	0,404	0,977	0,902	0,829	0,905	0,878	0,830	0,225	0,905	0,158	0,430	0,889	0,83
Undersampling nearmiss v1	0,030	0,996	0,029	0,352	0,974	0,957	0,306	0,859	0,707	0,320	0,058	0,468	0,056	0,194	0,459	0,32
Undersampling nearmiss v2	0,021	0,943	0,019	0,328	0,922	0,980	0,036	0,707	0,574	0,055	0,041	0,069	0,038	0,049	0,068	0,055
Undersampling nearmiss v3	0,019	0,980	0,027	0,342	0,959	0,730	0,271	0,579	0,527	0,281	0,037	0,425	0,051	0,171	0,417	0,281

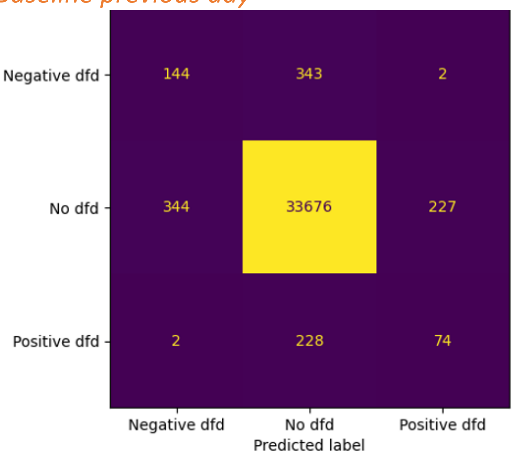
9.4 Details on model selection

9.4.1 DFD event forecast

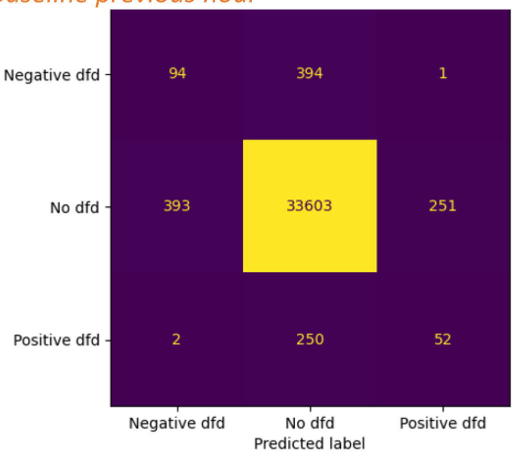
9.4.1.1 Confusion matrix and complete results

Model Comparison with naïve random oversampling:

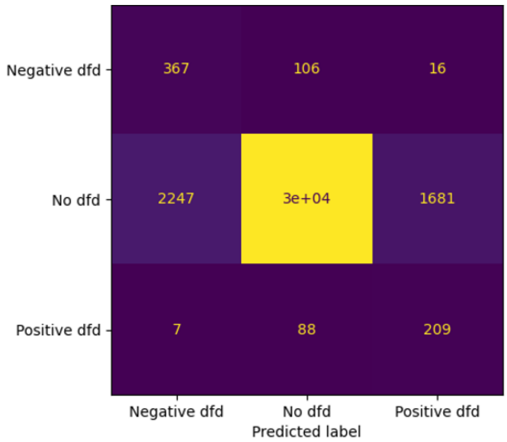
Baseline previous day



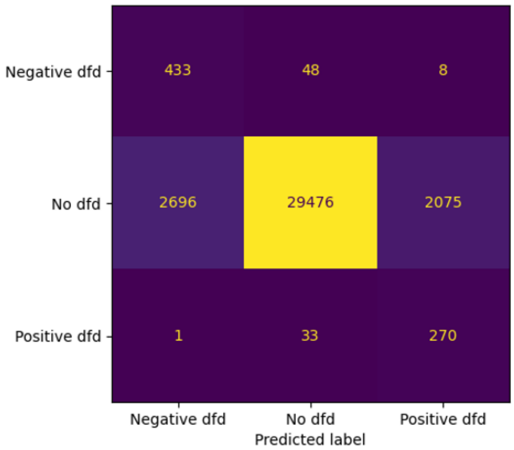
Baseline previous hour



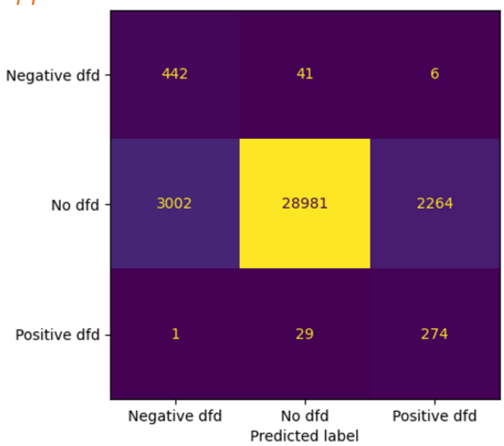
Neural Network



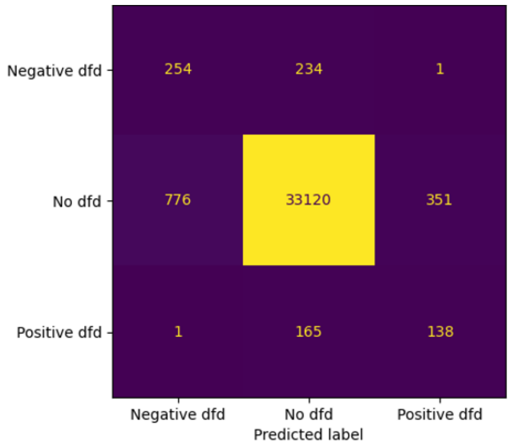
Logistic regression one vs rest



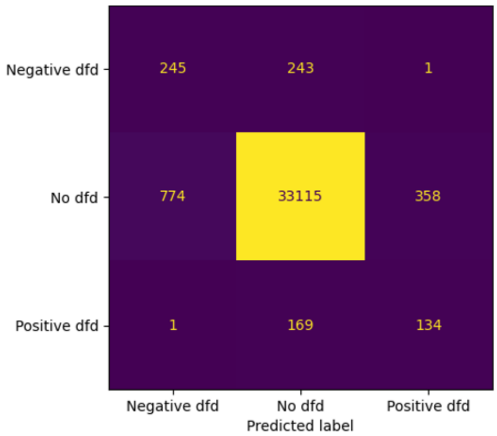
Support vector machine



Random forest



Ridge classifier



Hist gradient boosting classifier

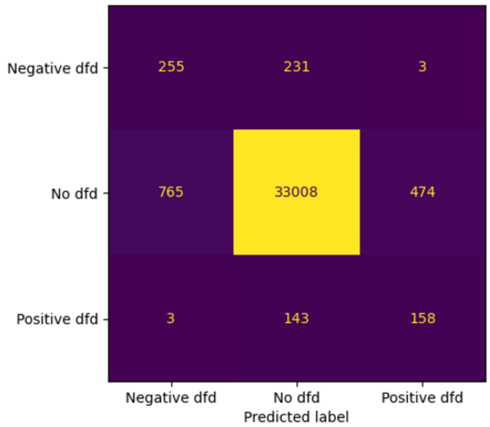


Table 8 - Models with naive random oversampling metrics overview

	Precision					Recall					F1 score					
	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Negative dfd	No dfd	Positive dfd	Macro average	Weigthed average	Overal accuracy
Baseline previous hour	0,192	0,981	0,171	0,448	0,963	0,192	0,981	0,171	0,448	0,963	0,192	0,981	0,171	0,448	0,963	0,963
Baseline previous day	0,294	0,983	0,244	0,507	0,967	0,294	0,983	0,243	0,507	0,967	0,294	0,983	0,244	0,507	0,967	0,967
Logistic regression	0,138	0,997	0,115	0,417	0,978	0,885	0,861	0,888	0,878	0,861	0,239	0,924	0,203	0,455	0,908	0,861
Neural network	0,140	0,994	0,110	0,414	0,974	0,751	0,885	0,388	0,774	0,882	0,236	0,936	0,189	0,454	0,920	0,882
Support vector machine	0,128	0,998	0,108	0,411	0,978	0,904	0,846	0,901	0,884	0,848	0,225	0,916	0,192	0,444	0,900	0,848
Random forest	0,246	0,988	0,282	0,505	0,972	0,519	0,967	0,454	0,647	0,956	0,334	0,977	0,348	0,553	0,963	0,956
Ridge classifier	0,240	0,988	0,272	0,500	0,971	0,501	0,967	0,441	0,636	0,956	0,325	0,977	0,336	0,546	0,963	0,956
Hist gradient boosting classifier	0,249	0,989	0,249	0,496	0,972	0,521	0,964	0,520	0,668	0,954	0,337	0,976	0,337	0,550	0,962	0,954

All these models were tested with their optimal parameters.

	Youden's index		
	Negative dfd	No dfd	Positive dfd
Baseline previous hour	-0,616	0,962	-0,658
Baseline previous day	-0,412	0,966	-0,513
Logistic regression	0,023	0,858	0,003
Neural network	-0,109	0,879	-0,502
Support vector machine	0,032	0,844	0,009
Random forest	-0,235	0,955	-0,264
Ridge classifier	-0,259	0,955	-0,287
Hist gradient boosting classifier	-0,230	0,953	-0,231

Figure 25 - Youden's index

9.4.1.2 Parameter sweep for dfd models

To derive the best parameters for the dfd models we used a randomized grid search on a time series split. The principle is that the model is evaluated over a time series split of a full year with a random hyperparameterset. Hundreds of these hyperparametersets are tried, and the best is kept. The chance that we have exactly the best hyperparameters is small, but we will be reasonably close to optimal parameters without having excessive computation time. Note that a lot of the hyperparameters are very similar to the ones used in the ace predictions. This is probably due to the fact that we use the same predictors.

Model	**Possible parameters**	**Optimal parameters**
Ridge classifier	Alpha: 0 to 10 000 in 1000 steps evenly spaced on a log scale	Alpha: 1390
Neural network: Multilayer perceptron	Hidden layer sizes: [(54,25,1), (54,20,20,20,1), (10,10,10), (50,30,30,10), (4,4,4,4)]	Hidden layer sizes: (4,4,4,4)
Random forest	Depth: 1 to 30 and full depth	Depth: 15
Logistic regression	No parameters	/

SVM	No parameters	/
Hist gradient boosting regressor	Depth: 1 to 20 and full depth L2 regularization: 0 to 1000	Depth: 6 L2 regularization: 14

9.4.1.3 Sensitivity analysis

To assess the sensitivity of the model we use the permutation importance as indicator. It consists of the following steps:

1. Train a model on the training dataset
2. Evaluate the model on the test set, we will call this result the base_result. We construct the metric in a way that higher = better. (For MAE, RMSE etc this simply means taking the negative of it.)
3. Permute (=shuffle) a column of the test set and evaluate the model on new test set. Call this the new score.
4. The performance will probably drop since one of the columns is totally shuffled. The permutation score is the base_line – new_score.
5. Redo steps 3 and 4 with the original test set, but for another column.

All steps 1 to 5 get repeated a lot of couple of times, in our case 10, because a variable could always get a lucky shuffle if you only do it once.

The results from the permutation test are model dependent but seem to be largely similar between different models. For computational feasibility we used the histogram gradient boosting classifier since it trains fast and can catch non linearities without problems.

Permuted variable	**Weighted F1-score decrease**
IT_Gen_first_derivative	0.014
Hour of the day	0.012
Minute of the day	0.011
IT_load_first_derivative	0.008
AT_gen_second_derivative	0.005
Month of the year	0.004
...	<0.001

9.4.2 ACE forecast

9.4.2.1 Parameters sweep

Optimal parameters are found doing a time series cross validation on 1 year of data. The model uses 3 months of train data and 1 day of test data. Everything is then shifted 1 day and the model is retrained. Doing this for all parameter combinations would be computationally very expensive, so we used a randomized grid search approach for 1000 combinations. This means 1000 combinations of random hyperparameters were tested and the best were kept. The possibility is high that these are not the actual best hyperparameters, but they will be close enough. If better ones are found in the future, they will of course be used.

Model	Possible parameters	Optimal parameters
K-nearest neighbors	Scaler: no / standard / robust Neighbors: 1 to 500 with 50 steps that are evenly spaced on a log scale Weights: uniform / distance	Scaler: robust Neighbors: 81 Weights: distance
Linear regression	Scaler: no / robust	-
Neural network: Multilayer perceptron	Hidden layer sizes: [(54,25,1), (54,20,20,20,1), (10,10,10), (50,30,30,10), (4,4,4,4)]	Hidden layer sizes: (4,4,4,4)
Random forest	Depth: 1 to 30 and full depth	Depth: 15
Ridge regression	Alpha: 0.1 to 10000 in 1000 steps evenly spaced on a log scale Scaler: no / standard / robust	Alpha: 1390.4
Hist gradient boosting regressor	Depth: 1 to 20 and full depth Scaler: no / standard / robust	Depth: 15

9.4.2.2 Sensitivity analysis

Similarly to DFD occurrence model, the sensitivity analysis is performed by permutating the columns one by one. The permutations are repeated 50 times and the performance averaged to ensure consistency of the results. This is done with the histogram based gradient boosting model and using only the BorutaSHAP variables. This model gave the best performance.

<u>Results:</u>

variable	RMSE increase in MW if variable is permuted
IT_Gen_first_derivative	5,28
hour	2,01
aFRR_previous_qh	1,67
minute	1,35
HVDC Nemo_first_derivative	1,23
previous_ace_belgium	0,90
Load Forecast	0,88
BE NET Position_first_derivative	0,85
AT_Gen_first_derivative	0,65
HVDC Nemo_second_derivative	0,58
CH_Gen_first_derivative	0,46
Specific Power Unit (Anonymized)	0,35
BE NET Position_second_derivative	0,28
Load Forecast_first_derivative	0,27
mFRR_previous_qh_first_derivative	0,26
DK_Load_first_derivative	0,23
aFRR_previous_qh_first_derivative	0,20
NL_Load_first_derivative	0,16
aFRR_previous_qh_second_derivative	0,16
Specific Power Unit (Anonymized)	0,11
DE_Load	0,10
Specific Power Unit (Anonymized)	0,10
Specific Power Unit (Anonymized)	0,09
CH_Gen	0,08
GCC_previous_qh	0,08
CZ_Load_second_derivative	0,08
HVDC Alegro	0,07
FR_Load_second_derivative	0,07
Specific Power Unit (Anonymized)	0,07
DK_Gen	0,06

Figure 26 Sensitivity analysis ace

From the results it seems like only a couple of variables are really having a big impact. The rest are still improving the model but only marginally.

9.5 Mitigation measure: complete tables on sensitivity analysis

9.5.1 aFRR mitigation measure

	Upwards DFD									
	SQ	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR
Start		-7	-7	-7	-5	-5	-5	-2	-2	-2
End		7	5	2	7	5	2	7	5	2
# violations	63	24	24	28	35	35	39	54	54	55
% of resolution		62%	62%	56%	44%	44%	38%	14%	14%	13%
aFRR_costs	€ 584.558	€ 412.602	€ 421.999	€ 432.385	€ 455.846	€ 467.213	€ 478.960	€ 530.582	€ 536.673	€ 546.194
mFRR_costs	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806	€ 246.806
Total costs	€ 831.364	€ 659.408	€ 668.806	€ 679.191	€ 702.652	€ 714.019	€ 725.766	€ 777.388	€ 783.479	€ 793.000
Costs per avoided violation		-€ 4.409	-€ 4.168	-€ 4.348	-€ 4.597	-€ 4.191	-€ 4.400	-€ 5.997	-€ 5.321	-€ 4.796
Additional cost per action taken		-€ 2.729	-€ 2.580	-€ 2.415	-€ 2.043	-€ 1.863	-€ 1.676	-€ 857	-€ 760	-€ 609

	Downwards DFD									
	SQ	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR	aFRR
Start		-7	-7	-7	-5	-5	-5	-2	-2	-2
End		7	5	2	7	5	2	7	5	2
# violations	91	48	48	55	54	54	61	69	69	76
% of resolution		47%	47%	40%	41%	41%	33%	24%	24%	16%
aFRR_costs	€ 1.419.815	€ 1.840.176	€ 1.794.549	€ 1.704.670	€ 1.712.070	€ 1.669.095	€ 1.584.245	€ 1.514.140	€ 1.482.904	€ 1.439.492
mFRR_costs	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061	€ 844.061
Total costs	€ 2.263.877	€ 2.684.237	€ 2.638.610	€ 2.548.731	€ 2.556.132	€ 2.513.157	€ 2.428.306	€ 2.358.201	€ 2.326.966	€ 2.283.553
Costs per avoided violation		€ 9.776	€ 8.715	€ 7.913	€ 7.899	€ 6.737	€ 5.481	€ 4.287	€ 2.868	€ 1.312
Additional cost per action taken		€ 4.619	€ 4.118	€ 3.130	€ 3.212	€ 2.739	€ 1.807	€ 1.037	€ 693	€ 216

9.5.2 aFRR and mFRR mitigation measure

	Upwards DFD				Downwards DFD			
	SQ	aFRR_mFRR			SQ	aFRR_mFRR		
aFRR		70%	50%	30%		70%	50%	30%
mFRR		30%	50%	70%		30%	50%	70%
# violations	63	30	31	34	91	20	27	34
% of resolution		52%	51%	46%		78%	70%	63%
aFRR_costs	€ 584.558	€ 582.202	€ 533.417	€ 501.275	€ 1.419.815	€ 1.338.823	€ 1.448.908	€ 1.546.968
mFRR_costs	€ 246.806	-€ 1.168	-€ 11.825	€ 41.533	€ 844.061	€ 5.214.209	€ 3.892.358	€ 2.559.846
Total costs	€ 831.364	€ 581.034	€ 521.592	€ 542.808	€ 2.263.877	€ 6.553.032	€ 5.341.265	€ 4.106.814
Costs per avoided violation		-€ 7.586	-€ 9.680	-€ 9.950		€ 60.411	€ 48.084	€ 32.332
Additional cost per action taken		-€ 3.974	-€ 4.917	-€ 4.580		€ 47.134	€ 33.817	€ 20.252

9.6 Decision Tree: Methodology of the Quality Parameter

9.6.1 Initial revision framework for the Y parameter

ENSTO-E evaluates the percentage of violation every quarter of a year. This means that for each quarter of a year, Elia should remain below the 30% of violations. So, the process to figure out the percentage of violation compared to the 30% ENTSOe threshold will be performed per quarter as well.

During the first month of the quarter, the "Y" parameter will be set to 15%[13]. As a results, according to the decision tree, during the first month of a quarter, Elia will apply a mitigation measure every time a DFD is forecasted and the probability of having an ACE larger than 217 MW is above the threshold, under the hypothesis that the cost for action is smaller than the cost for no action.

Then Elia will review the "Y" parameter for the second month and the third one according to the real percentage of violation with regards to the 30% ENTSOe threshold.

> *E.g: After the first month, Elia only has 10% of violations, the Y parameter is set to 10% for month 2. This means that during the second month of the quarter, Elia will never apply any mitigation measure. For the third month, Elia will review its position again. By not applying any mitigation measure for one month, Elia reached 22% of violations so Y is equal to 22 for the whole third month and Elia will apply mitigation measures according to the decision tree every time there is a DFD forecasted and the Elia contribution exceeds the threshold.*

9.6.2 Target revision framework for the Quality Parameter "Y"

The target model for the revision of the "Y" parameter would be to still have it reset to a certain value every quarter but to have it revised in real time and not only once a month. This means that each time Elia would have a DFD forecasted and an ACE contribution over the threshold, Elia would take a mitigation measure only if we are above the 15% of violations for the quarter.

[13] This 15% parameter can be reviewed over time. If we notice that it is too high and do not allow to remain below the 30% threshold then we could lower it. Would it be too low and would it lead us to systematically solve way too many violations, it could also be relaxed.

10. Bibliography

Anon., n.d. [Online]
Available at: https://imbalanced-learn.org/stable/index.html

CREG, 2022. [Online]
Available at: https://www.creg.be/sites/default/files/assets/Consult/PRD658E79FR.pdf

CREG, 2022. [Online]
Available at: https://www.creg.be/sites/default/files/assets/Consult/PRD658E79FR.pdf

Elia - Auction Results, n.d. [Online]
Available at: https://www.elia.be/en/grid-data/balancing/capacity-auction-results
[Accessed 2023].

ELIA, 2020. [Online]
Available at:
https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&ved=2ahUKEwjWq72d9Pb8AhXH57sIHf7JCjYQFnoECA8QAQ&url=https%3A%2F%2Fwww.elia.be%2F-%2Fmedia%2Fproject%2Felia%2Felia-site%2Fpublic-consultations%2F2020%2F20200701_deterministic_frequency_deviatio

entso-e, 2019. [Online]
Available at: https://consultations.entsoe.eu/system-development/deterministic_frequency_deviations_report/user_uploads/report_deterministic_frequency_deviations_final-draft-for-consultation.pdf

entso-e, n.d. [Online]
Available at:
https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Feepublicdownloads.entsoe.eu%2Fclean-documents%2FSOC%2520documents%2FSAFA_for_RG_CE%2FSAFA_for_RG_CE_-_03_-_Annex_01_-_Policy_on_Load-Frequency_Control_and_Reserves_220720_ABC.docx&wdOrigin=BR

Johannes Kruse, B. S. D. W., 2021. [Online]
Available at: https://arxiv.org/pdf/2106.09538.pdf

Regelleistung, n.d. *Regelleistung.net - Data Center.* [Online]
Available at:
https://www.regelleistung.net/apps/datacenter/tenders/?productTypes=SRL,MRL,PRL&markets=BALANCING_CAPACITY,BALANCING_ENERGY&tenderTab=PRL$CAPACITY$1
[Accessed 2023].