PUBLIC CONSULTATION ON

THE NEAR REAL TIME PUBLICATION OF THE DGO ALLOCATION PER BRP

Table of Contents

1. Introduction

In the past Elia has already taken initiatives to provide accurate information to the BRP's to support their balancing responsibility, such as the real time publication of the system imbalance and the infeed into the distribution grid. In addition, Elia also provides a forecast and a real time estimation of wind and solar production. This information facilitates the integration of decentralized and more volatile energy sources and can help Elia to balance the grid more efficiently. Another communication to the BRP's, which could facilitate the BRP's in taking their balancing responsibility is the near real time communication of the imbalance position.

On 22/12/2017 ELIA has submitted to the CREG a confidential study on the near real time communication of the individual ARP imbalance volume. This study was part of the discretionary objectives set for 2017 (cf. DECISION (B)160630-CDC-658E/38). In this study the "near real time communication of the imbalance position" was investigated at large and the results were the outcome of different analyses based on Elia's expertise and looking at the state of play on the field of modelling tools and communication technology. In the framework of this study an offline proof of concept (PoC) was built. This PoC concluded that it was possible to estimate the DGO Allocation volume, with acceptable accuracy, using statistical (based on machine learning) methods. Nevertheless, the study also concluded that the accuracy during the first months of the year was not up to par, due to changes of BRP portfolio happening at the beginning of the year (caused by switching of customers of the respective suppliers) that were not captured by the machine learning algorithm.

Based on the aforementioned study and PoC from 2017, CREG decided on 28/6/2018 to set a discretionary objective for 2019 on the implementation, by 1/1/2020, of the publication of the real-time volume allocation per DGO to the BRP (cf. DECISION (B)658E/52). This publication is based on the PoC from 2017. Moreover, the 2nd part of this discretionary objective for 2019 requires the submission, by 31/12/2019, of improvements of the calculation methodology for the real-time volume allocation per DGO to the BRP, in particular to enhance the accuracy of the estimation during the first two months of the year. Finally, a planning for the implementation of these improvements was requested as well. This submission to CREG is preceded by a public consultation of the market players.

This public consultation period runs **from 8 June 2019 to 8 July 2019 at 18h**. It consists of different parts:

In section 2, a description is provided on how the imbalance volume for a BRP for the imbalance settlement is currently being calculated.

In section 3 we look into the component of the DGO Allocation and how this component can be estimated on a near real time basis. We explain the model used in 2017, and we highlight the improvements embedded in the model developed in 2019 that will be used for the forecasting of the DGO Allocation. We zoom into the different variables that compose the DGO Allocation and identify the improvements in the proposed methodology.

In section 4 we set out the different forecasting processes that are used to calculate, process and publish the DGO Allocation to the BRP.

In section 5 we present the planning.

Finally, we conclude with some questions for feedback.

2. The imbalance settlement today

The imbalance settlement follows a monthly process in which Elia provides the result of the imbalance calculation for a month M at the latest at the end of month M+2 (although Elia usually succeeds in publishing by the end of month M+1). This lead-time is necessary to collect, process and validate all necessary data coming from other parties such as distribution grid operators (DGO's), and operators of closed distribution systems (CDS).

2.1. Imbalance components

The current imbalance calculation consists of different imbalance components that are all communicated separately to the BRP through the imbalance message at the end of the settlement process[1]. These components can be classified into 5 major blocks, each including several components:

1. DGO allocation result
2. Elia grid users (GU)
3. CDS allocation result
4. Ancillary Services
5. Market transactions

For every BRP the incoming and outgoing energy volumes per quarter-hour are registered or calculated for each component. The sum of all these components (taking into account the direction of the energy flow) results in the final imbalance for the BRP for a given quarter-hour.

The losses on the transmission grid are compensated by the BRP's based on a percentage of the offtake in their portfolio. This is the sum of the total offtake of the Elia GU and of the total allocation (including loop losses) for CDS and DGO. Thus, the volume of losses for a BRP depends directly of the volume of other imbalance components.

2.2. DGO allocation

The DGO plays an important role in the imbalance settlement process. Since Elia only has a view on the energy exchange on every border point between the Elia grid and the distribution grid (the so called infeed), it is the responsibility of the DGO to allocate the exchanged volumes to the BRP's. This exercise is called the allocation process, which is executed on a monthly basis and is described in the so-called MIG-rules (current version is MIG 4). These rules are the result of a mutual agreement between the DGO's and the involved market parties and are approved by the regional regulators. The calculations are executed by every DGO for his grid users and are provided to Elia. Elia performs the quarter-hourly comparison between the sum of all allocated volumes by the DGO and the total energy exchange between the Elia grid and the specific DGO, taking into account the volume exchanges between DGO's.

[1] For more information about the messages sent to the ARP, please consult the Elia metering manual: http://www.elia.be/~/media/files/Elia/Grid-data/Extranet/Metering-Manual.pdf

The quarter-hourly energy exchange on the border points is registered by and stored on the Elia meters that are installed in every transformer substation. Elia collects the metering data on a daily basis (overnight) which are then validated once a month (so-called 4.1 metering data).

For substations where several DGO's are connected to the Elia grid, the DGO's have to provide their metering data to Elia in order to split the volumes per DGO (so-called 4.2 metering data).

Based on these validated 4.2 metering data (taking into account the possible exchanges between DGO's) the DGO can start the allocation process. This process makes a distinction between distribution grid users with a quarter-hour meter and those with a monthly or annual meter reading (typically residential customers or SME's). For grid users with a quarter-hour meter (also called AMR) the injection and/or offtake for every quarter-hour is registered and stored in the meter and read once a day (overnight, as for the Elia meters). These volumes are directly allocated to the BRP having the access point in his portfolio (based on the access register of the DGO), taking into account the direction of the energy flow.

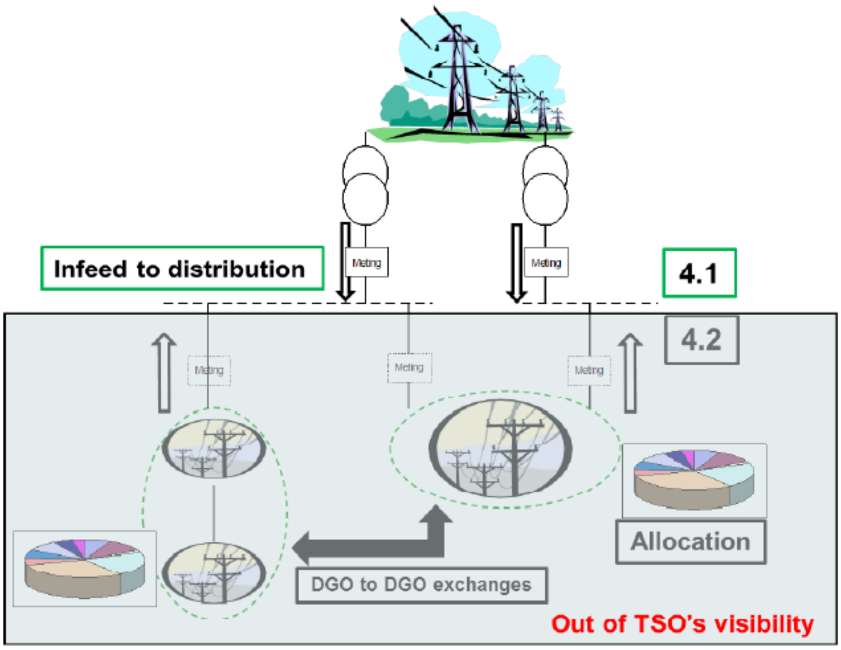

Figure 1: schematic view on the infeed metering

It is possible to appoint different BRP's for injection and offtake on the distribution grid. In the allocation messages published by the DGO afterwards, these volumes are indicated with "S10" or "SLP10" (cf. Figure 2). A distribution grid user that is able to consume and to inject energy in the grid has two access points: one for offtake and one for injection, with possibly a supplier and a different BRP.

The losses in the distribution grid are estimated as a percentage of the total infeed ("S70") and directly allocated to the BRP who is appointed by the DGO for supplying the energy for the grid losses.

For customers that do not have a quarter-hour meter (whose meter is only read once a year or once a month), the DGO makes an estimation of the quarter-hourly offtake (and injection) using the expected annual volume (EAV) for the current year on which a synthetic load profile (SLP) is applied, distributing the annual volume over all quarter-hours of the year.

There are 4 different profiles for electricity, one for each type of grid user (SLP 11, SLP 12, SLP 21 and SLP 22) which are available on the website of Synergrid.

Obviously this estimation deviates from the real injection and offtake by every single grid user, especially because the SLP for a given year is established before the beginning of the year. There is currently no correction for the temperature (having an effect on the residential energy consumption) neither for the solar production on residential level. On sunny days the estimation of the offtake will accordingly be too high and the estimation of the injection too low. This effect becomes more and more important, which is one of the main reasons why the DGO's have established a new set of rules, called MIG 6.

To overcome the estimation error and the consecutive difference between the sum of all allocated volumes compared to the infeed as determined by Elia, the DGO applies a correction factor (called the residu factor which can be higher or lower than 1) on all SLP-volumes. In this way, for every quarter-hour the sum of all allocated volumes (i.e. offtake minus injection) is equal to the net injected energy from the Elia grid to the distribution grid.

The result of this monthly exercise which is done by the DGO for every single quarter-hour is communicated to the BRP (for the grid users in his portfolio) and to Elia (for every BRP) at the latest on the 15th working day of the following month (M+1)[2]. The DGO's provide the results both for offtake and injection volumes on their grid, which gives an additional insight into the total offtake and total injection on the distribution grid, while Elia metering data can only indicate the net infeed (i.e. offtake – injection) from the Elia grid towards the distribution grid.

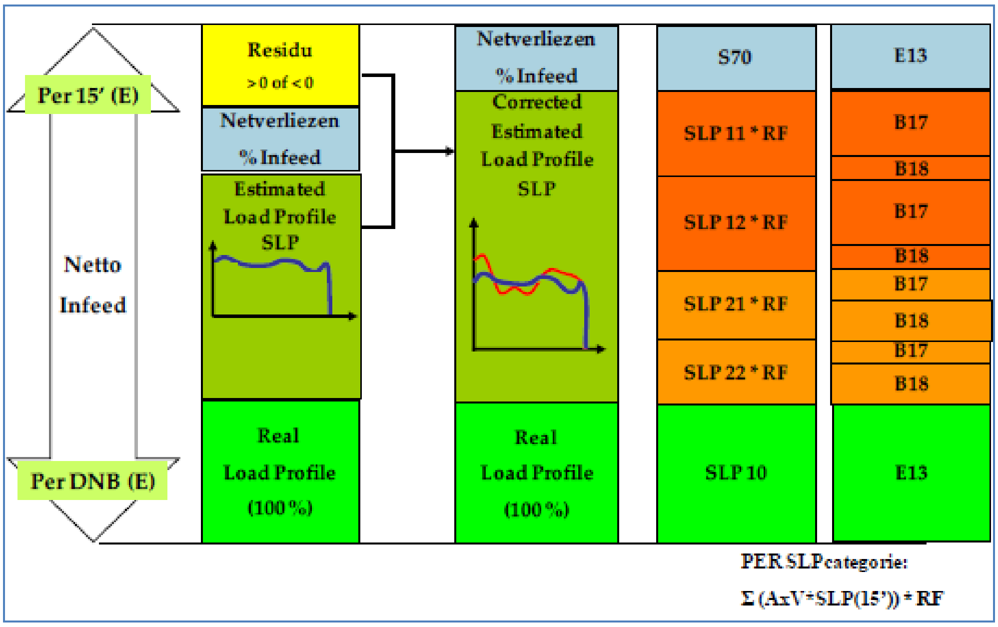

Figure 2: high-level process DGO allocation (source: MIG 4)

2 The allocation result is not the only communication from the DGO towards the ARP. On D+1 the ARP receives already the aggregated volume per quarter-hour for the offtake and injection of his AMR-customers, together with the Total infeed - ∑ AMR for all ARP's, which is an indication of the total SLP-volume for all ARP's. On a monthly basis the DGO provides the total EAV per type of SLP.

2.3. Elia grid users

For its own grid users, Elia disposes of quarter-hour meters on every access point. These meters register and store the volumes of offtake and injection by the grid user and are also read overnight once a day. Since an access point can be composed of different connection points (each having their own meter), the metering values of the different connection points are added up. Unlike the DGO, Elia only considers compensated values, meaning that for every quarter-hour we take into account the difference between offtake and injection. In this way for every quarter-hour one of the two values is zero. Elia can apply a correction factor on the registered volumes if the metering is executed on a different location than the actual access point (e.g. a meter on the secondary side of a transformer where the access point is on the primary side). The aforementioned data treatment is not done by the meter, but in the Elia tool for the management of metering data.

The net volume is then allocated to the designated BRP('s), based on the information in the Elia access register coming from the access contract for the concerned access point.

This allocation is done every day for the previous day, using the non-validated metering data, of which the results are sent to the BRP. Every month the exercise is repeated once the metering data have been validated. The result of this monthly exercise is then taken into account for the imbalance calculation of the BRP's. It should be noticed that Elia access holders have the possibility to appoint multiple BRP's for an access point, for instance: a different BRP for offtake and injection (as is possible on the distribution grid), but also a volume split between BRP's for the same direction can be requested.

2.4. CDS allocation

For the closed distribution systems connected to the Elia grid, the same principle applies as for the DGO's: Elia disposes of meters on the access point of the CDS (interface between the Elia grid and the CDS) and the CDS operator is responsible for the allocation of the registered volumes to the concerned BRP's, based on his own metering data for the CDS grid users. The CDS can request to delegate this task to Elia (including the management of the CDS access register). This is only important for CDS where the connected CDS grid users have a different supplier or BRP

2.5. Ancillary services

Elia procures ancillary services a.o. to maintain the system balance. From a high level perspective the balancing services that are offered by the market parties (BRP, grid users or aggregators) correspond to a modulation of the injection or offtake upon request by Elia. These activated energy volumes, apart from FCR, are accounted for in the imbalance calculation of the respective BRP.

A BRP offering balancing services will activate the requested volume by means of his own facilities (mostly power plants or storage) or facilities of his customers (production and load). As the requested volumes are remunerated by Elia, it should be avoided that the activated volumes are paid twice through the imbalance price. Therefore, the BRP perimeter is corrected with the requested volumes, implying also an incentive for the BRP to correctly deliver the requested volumes as any difference between the requested and delivered volume is exposed to the imbalance price. This "incentive correction" is currently applied for aFRR, reserved and non-reserved mFRR offered by CIPU units.

When an aggregator offers (upwards) balancing services, he can make use of flexibility of grid users within the portfolio of a (other) BRP. To that extent, Elia has developed the Transfer of Energy rules, in line with the amendment to the Electricity Act (13/7/2017).

These rules set the framework for the activation of demand side flexibility by independent flexibility service providers, the resulting imbalance correction of the BRP and the financial compensation between the concerned supplier of the delivery points used and the flexibility service provider. This allows to have a full competition and level playing field between all balancing service providers (BRP and independent aggregators). To properly correct the effect of such an activation, the imbalance adjustment is based on the actual delivered volumes. Hence the Transfer of Energy mechanism must be taken into account for the imbalance volume calculation.

For strategic generation reserves the activated the actual injection is not taken into account for the imbalance settlement since the injected volumes are replaced by 0, as determined in the BRP-contract. For strategic demand reserves the system is similar to that of mFRR-activations (with resp. without Transfer of Energy).

2.6. Market transactions

BRP’s can buy and sell energy on the power exchanges (day-ahead and intraday market). Furthermore BRP’s can trade energy on a bilateral basis (OTC). Transactions can happen between two Belgian BRP’s, but BRPs can also perform cross-border transactions. To be able to take these transactions into account for the imbalance settlement, all exchanges between BRP’s have to be communicated to Elia. Therefore Elia established the so called “Elia Hub” which allows the power exchanges (Central Counter Party) and the BRP to nominate their traded volumes.

In order to verify the correctness of this information all volumes must be nominated twice: once by the selling party and once by the buying party. For transactions on the power exchanges these volumes are nominated by Central Counter Parties on the one hand and by the buying or selling BRP on the other hand. Day ahead market results are known one day before delivery (since all BRP’s must have a balanced position on day-ahead (D-1)). BRP’s can nominate their intraday deals for a given day until 14h the day after (D+1).

Transactions which have been correctly nominated by both parties (buyer and seller) are taken into account for the eventual imbalance settlement. Inconsistent nominations are settled through a dedicated procedure.

3. Approach towards a DGO Allocation calculation in real time

3.1. Machine Learning technology model based on 2017 Proof of Concept

As explained in chapter 2.2 the DGO is responsible for the allocation of the infeed (energy that is exchanged between the Elia grid and the distribution grid) and the only means of verification for Elia is the comparison between the net infeed and the sum of all allocated volumes (offtake-injection). For this exercise the DGO uses the data coming from his meter installed on the access point of each distribution grid users. For the AMR-clients the allocation is quite straightforward, however these metering data are only available on D+1.

The only data available to the DGO in real time are its own measurements on the distribution substation and at the level of the connection of the most important grid users. The latter could provide some insight for a real time allocation, but would constitute only a fraction of the total volume and thus not useful for an accurate real time estimation of the allocation result which is provided only a month later.

Even if it would have been possible to elaborate an estimation for the large (AMR) distribution grid users, the question remains what to do with all SLP-customers. The ever growing penetration of PV-installations in the segment of the household customers creates more and more distortion of the allocation result: the application of a SLP differs more and more from the real consumption and injection, introducing more uncertainty and subsequently making it more difficult to make a real time estimation of the allocation result. Moreover, aside the problem of data availability, the applications that are currently used to run the allocation process do not permit to execute this calculation every quarter-hour.

Based on this Elia concluded that the only means to have a near real time estimation of the DGO allocation will consist of a model which is able to make such an estimation using the (limited amount of) data available in real time. The most relevant of these data is the net infeed, which is already published on the Elia website in real time. This net infeed uses the data coming from the measurements (in the SCADA) on every Elia transformer connecting the Elia grid to the distribution grid. As such the model calculates the correlation between the variables and the DGO Allocation using historical data. Once the main variables have been identified, the model elaborates a formula (by trial and error) leading to a minimal forecast error (in terms of Root Mean Square Error or RMSE).

In line with the discretionary incentive for 2017, ELIA has developed a methodology and a PoC to forecast the component of the DGO Allocation in near real time. This statistical model is able to make an estimation of the final DGO allocation outcome and is based on machine learning using a linear regression model. That is, the Real-Time DGO Allocation estimate can be calculated based the following formula:

$$RT\ DGO\ Alloc\ Estimate_{BRP_i}(qh) = Intercept_{BRP_i} + \sum_{j=1}^{N} Coeff_{BRP_i}(j) * Variable_j(qh)$$

Equation 1 Linear regression model

were the different coefficients in the equation ($Coeff_{BRP_i}(j)$) are calculated in order to minimize the RMSE of the prediction using machine learning techniques.

The variables selected for this estimate for the PoC built during the 2017 study was mainly based on the following variables:

- The individual infeed data, for each of the 360 substations, converted from 2‘ measurements into 360 quarter-hour time series
- Wind forecast and upscaled measurement, per wind farm
- Solar forecast and upscaled measurements, per region
- Solar forecast and upscaled measurements, per DGO
- Synthetic Load Profile (SLP) as published by Synergrid

At the end approximately 450 variables (input data) were tested in the model. Other variables were injected (DAM prices, Imbalance prices, Weather…) but were discarded as they had a marginal effect on the model accuracy.

3.2. Improvements to the estimation method introduced in 2019

As part of the improvements in the calculation methodology, Elia has identified the following improvements that have a significant contribution in decreasing the forecasting error:

- Introduction of extra variables in the model such as
  - DGO Nominations (day-ahead)
  - Total Load real-time estimation
  - Total Load day-ahead forecast
- Optimal variable selection per BRP
- Extension of machine learning training window

3.2.1. Introduction of extra variables

The **DGO Nomination** is provided by the BRP on D-1 in accordance with the BRP contract. These contain the expected net-offtake per DGO on a quarter-hour basis. These consist of the best estimated value based on the historical load figures and information provided by suppliers (eg. switching data). These DGO Nominations must correspond as close as possible with the actual Off-take and Injection at distribution level.

The **total load real-time estimation** incorporates all electrical loads on the Elia grid and in underlying distribution networks (and also includes electrical losses). This real-time estimate is calculated based on a combination of measurements and upscaled values of injections of power plants, including generation in the distribution networks, to which imports are added. Subsequently, exports and power used for energy storage are substracted, leading to an estimation of the actual total load in the Elia-grid and all underlying networks.

The **total load day-ahead forecast** is an estimate of the total load on day-ahead. This figure is calculated taking into account historical data on load as well as on forecasts of distributed production for day-ahead.

3.2.2. Optimal variable selection per BRP

During the PoC in 2017, the estimation of the RT DGO Allocation was done using all variables per BRP. However, given that BRP might have a different client base (residential, renewable, industrial clients…) not all variables might be relevant to all BRP.

For instance, as shown in Figure 3, the Infeed and DGO Allocation are highly correlated for a BRP with a large residential customer base. Thus, Infeed will be a good estimator of DGO Allocation in the regression model.

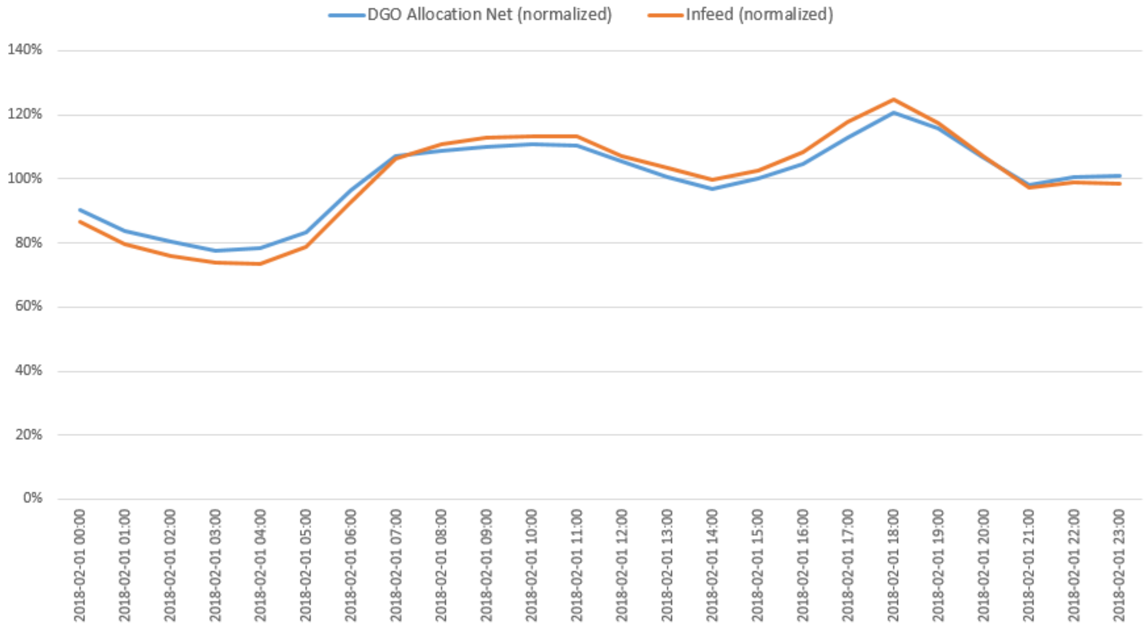

Figure 3: DGO Allocation and Infeed for BRP with large residential customer base

On the contrary, as shown in Figure 4, the Infeed and DGO Allocation are not correlated for a BRP with a large PV production customer base. In this case, Infeed does not bring much information for estimating the DGO Allocation. For this same BRP, as shown in Figure 5, the solar production and DGO Allocation are highly correlated, and thus, the solar production will be a good estimator of the DGO Allocation.

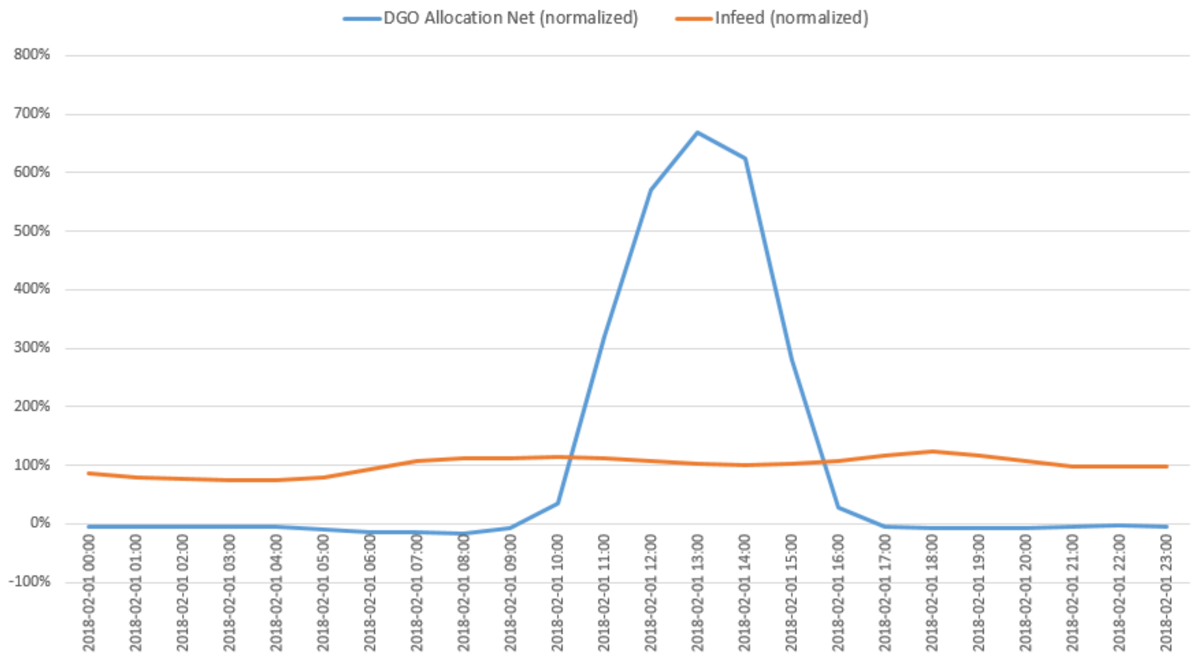

Figure 4: DGO Allocation and Infeed for BRP with aggregation of PV production

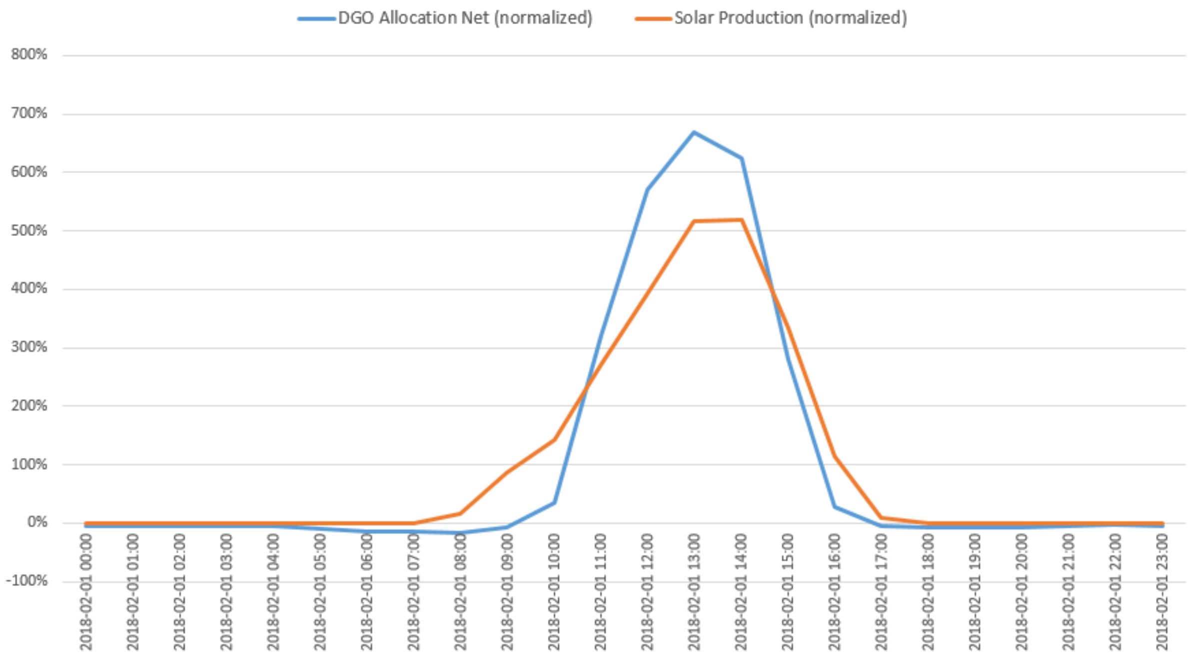

Figure 5: DGO Allocation and solar production for BRP with aggregation of PV production

Following this reasoning, in 2019, simulations have been performed to identify the optimal variable set per BRP, among the following families of variables:

- Infeed
- Wind forecast and upscaled measurement, per wind farm
- Solar forecast and upscaled measurements, per region
- Solar forecast and upscaled measurements, per DGO
- Synthetic Load Profile (SLP) as published by Synergrid
- DGO Nominations (day-ahead)
- Total Load Forecast and estimation

The simulations yield different variable combinations per BRP, for instance, the model for a BRP with a large residential customer base will use Infeed and SLP variables, while the model for a BRP aggregating renewable energy will use the solar and wind forecast and upscaled measurements variables.

Figure 6 below illustrates the families of variables that were tested in the 2017 PoC and added in the 2019 model.

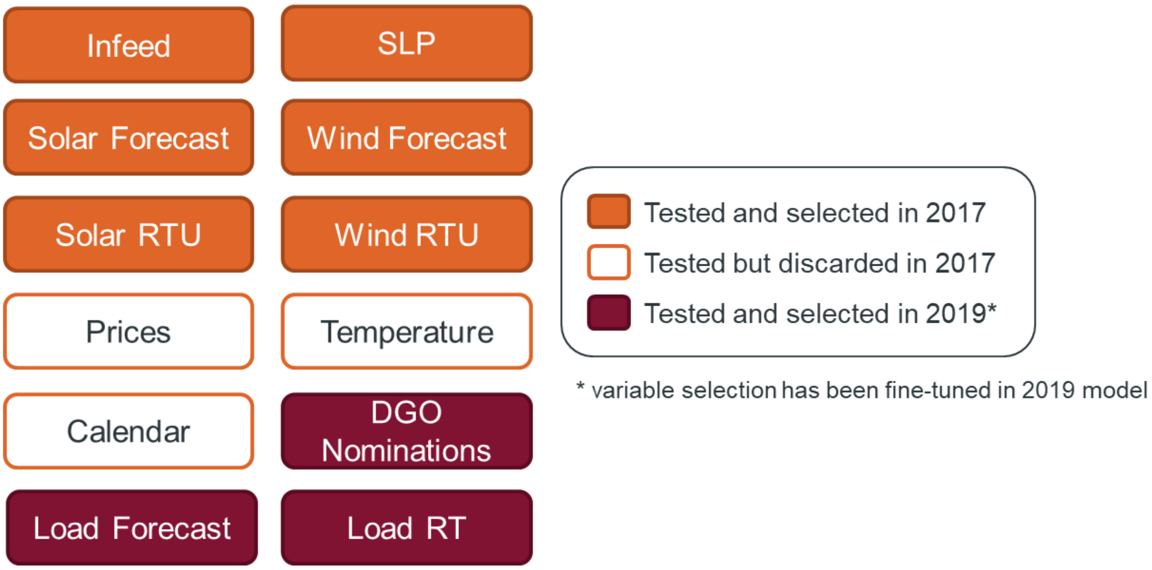

Figure 6: Variable families used in 2017 PoC and tested in 2019

The optimal combination of variables to be used for a given BRP will be defined in a so-called BRP-ID card, along with other information useful during the forecasting process.

3.2.3. Extension of machine learning training window

During the 2017 PoC, the training window for the machine learning algorithm was of 4 weeks. This training window has now been extended to 12 months. In this way, the machine learning algorithm training set comprises a full year of data and can thus capture seasonal effects.

In practice, the training of the algorithm will be using data up from month M- 14 to month M-2.

3.2.4. Reduction of forecast error with 2019 improvements

The three improvements explained in the previous sections have been implemented and tested in order to measure their effective contribution to improving the accuracy of the estimate.

The results are illustrated in Figure 7 where a forecast of DGO Allocation for January 2017 has been tested with the different improvements. The baseline (100%) is the standard deviation of the forecast error with the PoC from 2017. shows that the extension of the training window significantly cuts the forecast error for the large BRP, with an additional improvement from the extra variables and the BRP specific variable selection. For the small PV aggregator BRP, Figure 7 shows that the main error reduction comes from the specific variable selection.

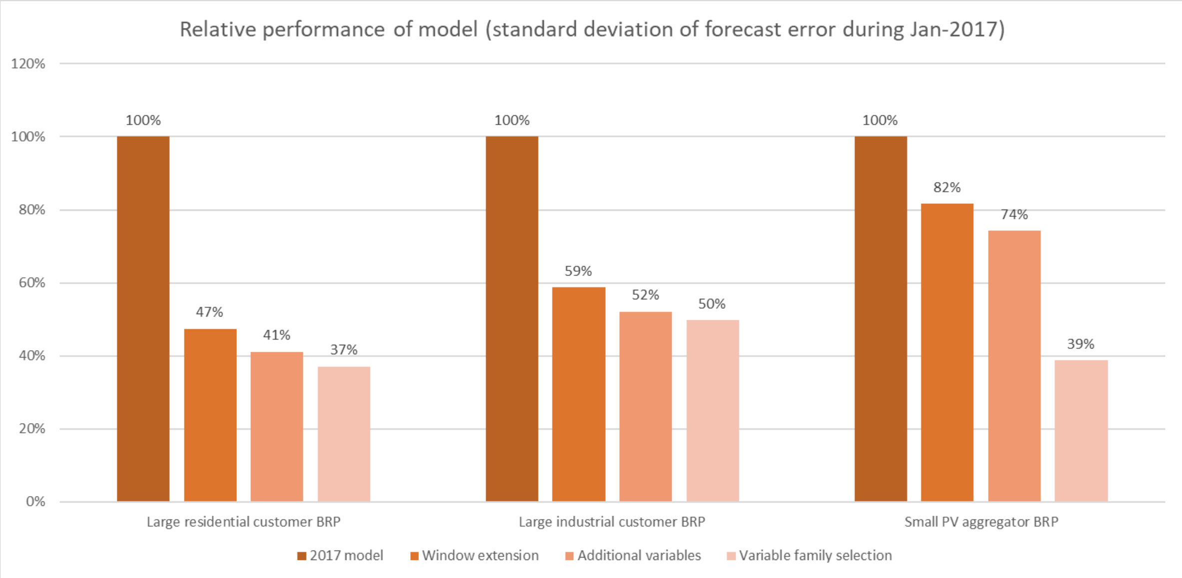

Figure 7: Reduction of forecast error with 2019 improvements

3.2.5. Additional improvements to be tested during 2019

In addition to the improvements already tested and mentioned in the previous sections there are other improvements that, **at the time of writing this document**, are not yet tested and validated. These extra improvements concern dealing with the collinearity between the different variables. Indeed, the model can use more than 600 variables and some of these variables might be highly correlated or have some linear dependency between each other. For instance, the infeed at two stations feeding mainly a large customer base will probably follow the same variations and thus present a high correlation between the two. The same can happen with the wind production of wind farms located close one to the other.

In general collinearity does not affect the prediction quality of a linear regression model, but it can lead to unstable models and yield difficulties in assessing the impact of missing data in the prediction quality (see 4.5.3).

Therefore, in addition to the variable selection per family explained in section 3.2.2, it is necessary to reduce the dimensionality of the regression model in order to mitigate the effect of collinearity in the input variables.

The following techniques for further refining the model are currently being investigated:

- Principal Component Analysis decomposition
- Stepwise regression
- Ridge regression

Note that in any case, the resulting model is still a linear regression model.

4. Forecasting process

4.1. Overview

The real-time allocation forecast has to be supported by processes that occur prior to making the actual forecast in order to prepare the data and the models. Also after making the actual forecast, a process is needed in order to assess the quality of the models ex-post. Hence, the real-time allocation forecast process is situated in between these different pre- and post-supporting processes.

To that extent, we differentiate the following processes:

1. Ex-ante historical data quality assessment
2. Training of allocation forecasting model
3. Ex-ante real-time data quality assessment
4. Real-time allocation forecast
5. Ex-post assessment of forecasting quality
6. Model hyper parameter tuning

The ex-ante historical data quality assessment, model training and ex-post accuracy assessment are monthly processes, triggered by the completion of a DGO Allocation process.

During the training of the allocation forecasting model, the optimum weighted values or coefficients of the variables (this optimum being defined in terms of error minimization and model robustness) per BRP in the DGO Allocation estimation equation will be calculated using machine learning techniques. Only the variables selected in the BRP ID-card are used. Also during this process, the impact of each variable in data quality will be estimated and provided as an output. This will be used in real-time to assess the impact of a missing variable on the quality of the DGO Allocation estimation.

The ex-ante real-time data quality assessment is performed in real-time and allows evaluating the impact of missing data in the quality on the allocation forecast.

Finally, the allocation forecasting is a real-time process taking place every quarter hour and estimates the allocation volume per BRP for the preceding quarter hour.

The ex-post assessment of forecasting quality compares the result of the forecast with the actual allocation values provided by the DGO.

The model hyper parameter tuning is a process that will take place 1-2 times per year. During this process, the variable set providing the best estimation performance, per BRP, will be identified and recorded in the so-called BRP ID-card. Indeed, it has been shown that a BRP specific model, taking into account only the variables relevant for the BRP, works better than a generic model taking into account all variables (as explained in section 3.2.2). In this BRP ID-card, also the minimum quality threshold for estimation will be defined. This parameter will be used in case of missing data in real-time to assess if an estimation should be made or not. During this hyper parametrization, some technical parameters might be reviewed (duration of training window, cost function…) in order to fine-tune the performance of the estimation algorithm

Figure 8 shows the various processes in the sequence they are run through as well as their frequency. Note that all these process will be run by Elia, and that BRP input might be required for the hyper parametrization of the model, by providing input to Elia about the relevant variables to include.

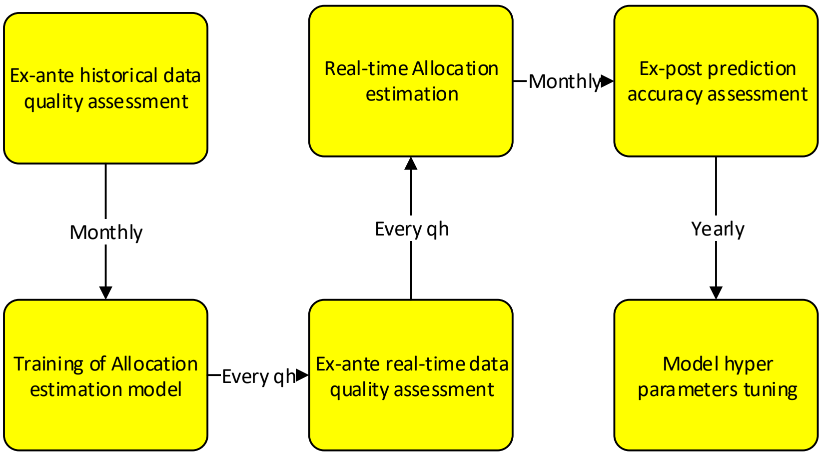

Figure 8: High level view of business process

The following sections will provide more details about each of these steps in the business process.

4.2. Ex-ante historical data quality assessment

The training of the machine learning algorithm will use significant amount of historical data. These data include, and may not be limited to:

- DGO Infeed
- Wind Production Day-Ahead Forecast
- Wind Production Real-time Estimation
- Solar Production Day-Ahead Forecast
- Solar Production Real-time Estimation
- DAM prices (day-ahead)
- Imbalance prices
- Weather
- DGO Nominations (day-ahead)
- Total Load Day-Ahead Forecast
- Total Load Real-time Estimation

Before launching a training, it is essential to make sure that these data are available and contain sufficient quality for the selected training period.

4.3. Training of allocation estimation model

4.3.1. General process

The real time estimation of the DGO Allocation volume per BRP is based on a model obtained by machine learning. The machine learning algorithm uses historical data to identify the model that fits in the most optimal way the DGO Allocation volume per BRP based on the variables selected for each BRP (defined in the BRP ID-card).

This fitting must be done regularly, in order to capture changes that might affect the estimation of the DGO Allocation volume for the BRP, such as the evolution of the market (more installed PV or wind), the portfolio composition of the BRP (client switching).

Ideally, this re-fitting, or re-training, should be done at least on a monthly basis, once Elia has received the latest DGO Allocation volumes. The performance of the new model has to be measured and benchmarked against the previous model on a common dataset.

This performance and comparison will be used to decide whether the new model is deployed in operations or if it is discarded, this latter option implying that the previous model remains in operations. It is important to stress that the model is BRP specific, that is, there is only one estimation model per BRP.

In this process, it is required to select the period of data on which the model has to be trained. Afterwards it has to be decided whether the standard variable selection in the BRP ID-card will be used or if specific variables have to be selected to train the model. This manual variable selection will depend on different aspects such as:

- Actual availability of variables in real-time against predefined availability thresholds (cf. 4.3)
- Benchmarking of models with different variables set
- Period of the year: it might be necessary to fit a particular (adjustment of the) model for January and February (to grasp the switching in BRP portfolio)

The process chart in Figure 9 gives a high-level overview of the different steps.

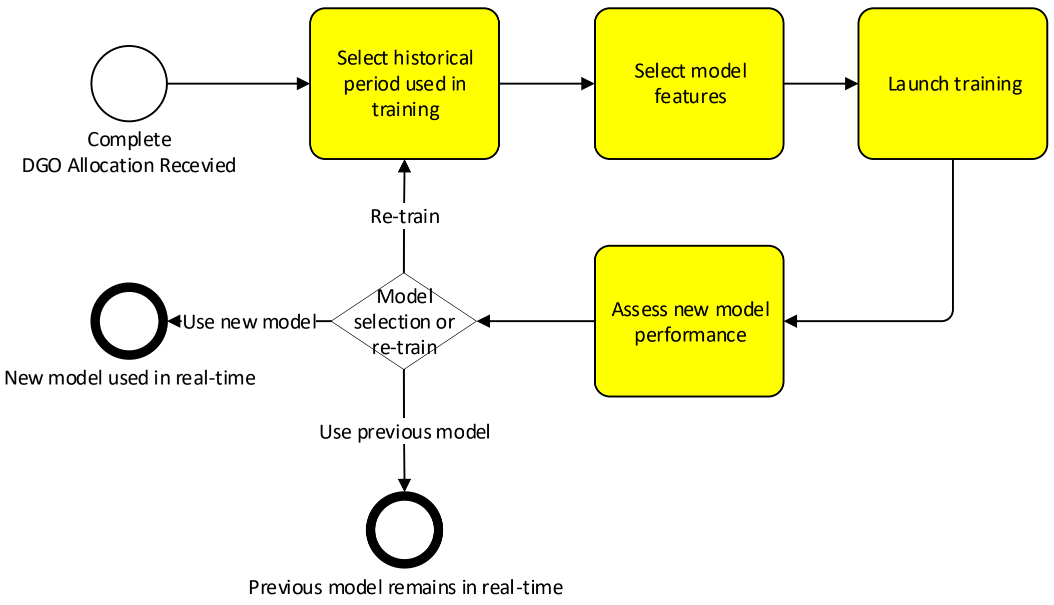

Figure 9: Machine learning training process

The obtained model is an equation that allows calculating the DGO Allocation volumes based on different variables.

The modelling approach is based on linear regression models, implying that DGO Allocation at instant qh is calculated based on features (variables) at instant qh:

$$RT\ DGO\ Alloc\ Estimate_{BRP_i}(qh) = Intercept_{BRP_i} + \sum_{j=1}^{N} Coeff_{BRP_i}(j) * Variable_j(qh)$$

Equation 2 Linear regression model

where

$RT\ DGO\ Alloc\ Estimate_{BRP_i}(qh)$ equals the DGO Allocation estimate at qh

$Intercept_{BRP_i}$is the bias of DGO Allocation for BRP_i

$Coeff_{BRP_i}(j)$ is the weight of feature (variable) j for BRP_i

$Variable_j(qh)$ represents the variable j at instant qh. Note that these variables can include:

- Telemeasurements: Infeed, per substation
- Day ahead production forecasts: Wind (per farm) and Solar (per geographical area or per DGO)
- Real-time production estimation: Wind (per farm) and Solar (per region or per DGO)
- Nominations: DGO Nominations made by BRP (per BRP and DGO)
- Day ahead total load forecast
- Real-time total load estimation
- Synthetic Load Profiles (SLP)
- Date time characteristics: Hour of the day, month, season…

4.3.2. Output from training process

The training process will provide the following deliverables:

- An estimation model, per BRP, of the DGO Allocation. This model consists of the terms *Intercept* and *Coeff*. Shown in Equation 1 and Equation 2.
- A performance report (document) with accuracy measurements of the model and benchmark with previous model that will be used by Elia to decide upon selecting the newly trained model for usage in real-time estimation.
- A weighting of the variables in the accuracy of the prediction. This is a per-variable-per BRP table showing the impact of each variable in the quality of the prediction. This will be used on real-time to assess if there is sufficient data to perform a prediction

A simplified example of variable weighting table is provided in Table 1. It illustrates that for each BRP, the impact of a variable in the quality of the estimation is different.

For instance, for BRP 1 the DGO Infeed accounts for 54% of the accuracy, while that same variable only impacts the accuracy of BRP 2 estimation by 17%. This means that for BRP 1, the lack of Infeed data will severely affect the quality of the estimation, while for BRP 2 the impact is more restrained.

For the Solar Forecast it is the other way around, lack of Solar Forecast data will severely affect the estimation of DGO Allocation for BRP 2 and BRP 3, while for BRP 1 the impact will be limited.

This justifies the creation of a specific pattern of parameters (weighted value of parameters reflected in an ID-card per BRP), setting the most fit profile per BRP to estimate the DGO Allocation volume.

	BRP 1(%)	BRP 2 (%)	BRP 3 (%)
DGO Infeed	54%	17%	27%
Solar Forecast	22%	77%	66%
SLP	15%	6%	2%
DGO Nominations	9%	0%	5%

Table 1: Variable weighting in estimation quality table (**simplified mock data**)

Please note that Table 1 is a simplified example. The actual table might have hundreds of variables and instead of just one DGO Infeed or one Solar Forecast variable, it will have many variables of such kind eg. Infeed per Border Point, sloar forecast per DGO,…(DGO_Infeed_PU_1, DGO_Infeed_PU_2,…, DGO_Infeed_PU_N, Solar_Forecast_DGO_1, Solar_Forecast_DGO_2…, Solar_Forecast_DGO_M,…)

4.4. Ex-ante real-time data quality assessment

The trained model will use following data, including, and not limited to:

- DGO Infeed
- Wind Production Day Ahead Forecast
- Wind Production Real-Time Estimation
- Solar Production Day-Ahead Forecast
- Solar Production Real-time Estimation
- DAM prices (day-ahead)
- Imbalance prices
- Weather
- DGO Nominations (day-ahead)
- Total Load day-ahead Forecast
- Total Load real-time Estimation

The effective availability of these data must be monitored in order to screen and assess the minimal availability that has to be taken into account by the machine learning model. This implies that even if a variable appears to significantly contribute to the quality of the prediction, it might be decided to discard the variable from the model due to non-sufficient availability in real-time.

During the training of the machine learning algorithm, the weighting of each of the variables in the quality of the estimation was calculated. These weights will be used in real-time to assess the degradation of the DGO Allocation estimation due to non-available variables in real-time. The minimum availability shall be defined at parameter level and per BRP, in the BRP ID-card.

Aside the minimum availability requirements, also an actual availability, in time, is required. Actual availability means that the data for quarter-hour qh is available within 5 minutes[3] after the end of the respective quarter-hour qh.

For instance, if it appears that the DGO Infeed data does help obtaining a very good estimate of the DGO allocation, but that the DGO Infeed is only 50% of the time available on time, one might consider training a new model that does not take into account the DGO Infeed.

This monitoring should be done regularly, i.e; daily on D+1, so that it can be used before each training to assess the data to be used.

Table 2 provides a mock up example of the observed monthly availability per variable in real-time. Based on this information, a decision can be taken regarding the variables to be used for training, depending on the month. For instance, for the training to be done end of January 2019 it might be decide to drop the wind variable, given this variable does not appear to be readily on time. The same choice could be made end of February, with the observed actual availability. At end of March the situation is different, now the Infeed appears to miss around 50% of the real-time availability. Hence in this case, it might be decided to include Wind, but to exclude Infeed. And at end of April, the observed availability appears to be acceptable for both variables and thus the recommendation would be to include all variables for the training.

	2019-01	2019-02	2019-03	2019-04
INFEED	90%	90%	50%	90%
SOLAR	90%	90%	90%	90%
WIND	50%	60%	90%	90%
Imbalance Price	90%	90%	90%	90%

Table 2: Observed average availability of variable on real-time **(simplified mock data)**

Obviously, dropping a variable from the model might have an impact on the accuracy of the estimation. But if a variable is not regularly available in real-time and has a non-significant contribution/impact in the overall estimation, it might be more appropriate to build a model that can estimate without this variable. It is a trade-off between estimation accuracy and estimation feasibility.

[3] Current proposal is 5 minutes, but this value should be reviewed once there are more accurate values about the required time to do the data acquisition, data quality check, estimation and publication.

4.5. Real-time estimation and publication

4.5.1. General process

The overall process is illustrated in Figure 10. It is an automatic process triggered at the end of quarter-hour QH.

- First step is to acquire the data from the concerned QH necessary for the estimation of the DGO Allocation.
- Second step is to make sure that the acquired data meets the required data quality threshold for making an estimate. From this point on, the process might be different for each BRP. Thus, the process is **BRP dependent**. The data quality checks are explained more in detail in subsection 4.5.3
- If data quality thresholds are not met, a number of automatic retrials will take place, from the data acquisition step. These could be limited to the acquisition of the erroneous data only.
- If sufficient data quality is reached, the process continues with the estimation of the DGO allocation volume per BRP.

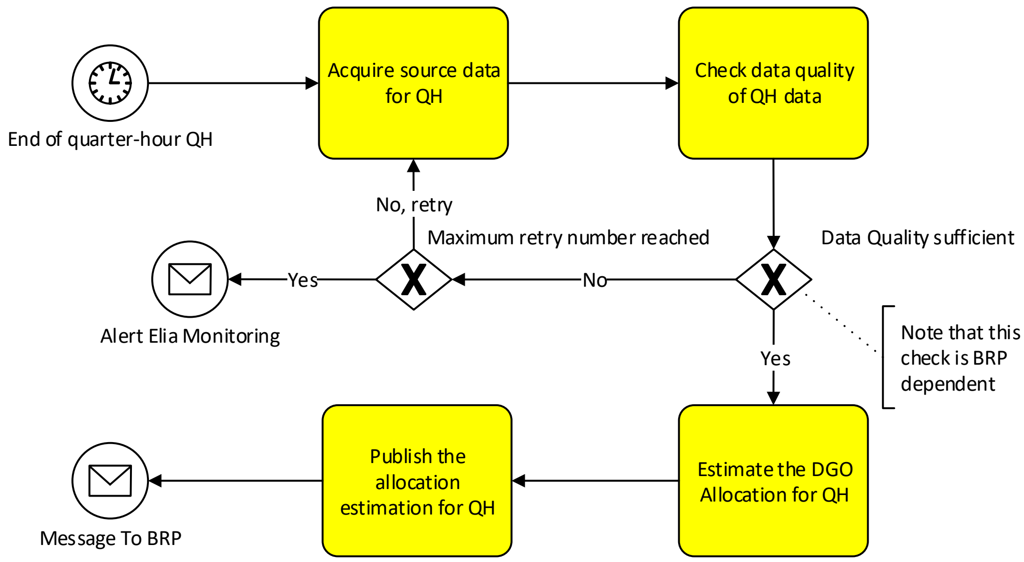

Figure 10: Real-Time estimation and publication process

Given that the process is BRP dependent, i.e. for some BRP the quality checks might pass while for others they will fail, it is necessary to keep track of the status of each BRP in the process. For instance, some BRP might be more dependent on DGO Infeed than others, thus in case that DGO Infeed is not available in real-time, for a given quarter hour, this might result in the DGO Allocation estimation being calculated for some BRP but not for others. Indeed, for those BRP with low correlation between DGO Allocation and DGO Infeed, the minimum data quality threshold might be reached even without the DGO Infeed.

4.5.2. Process time-constraints

Note that this process should be executed within 15 minutes, so that the estimation of the DGO Allocation for quarter-hour QH is published no later than the end of quarter-hour QH+1, as illustrated in Figure 11.

Note that the **data acquisition done during QH+1 concerns the data up to the end of QH**. Hence not all data might be available as of the beginning of QH+1.

Note as well that some data can already be pre-fetched on Day Ahead, such as DGO Nominations, Belpex DAM Prices and SLP. This might lessen a bit the stress on the real-time data acquisition.

In order of magnitude with respect to the process execution time, the data acquisition phase should be fully executed within 5 minutes, depending on the moment the data from previous quarter hour is effectively available. Ex ante quality check should execute within 3 minutes, and estimation should execute within 3 minutes leaving 4 minutes for the publication.

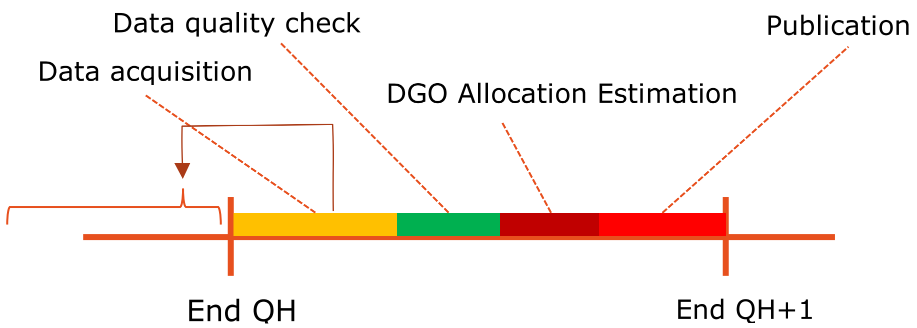

Figure 11: Time-line for real-time estimation and publication

4.5.3. Data Quality checks

The data quality checks verify if the necessary data are available for making an accurate estimate. We differentiate between data that might just be missing because the source application has not yet made the data available, as could be the case with real-time sources such as wind-forecast, and data that might not be of sufficient quality due to the nature of the acquisition, for instance EMS might send telemeasures with bad quality indicator.

It is important to remark that it is not because a particular data item is missing that the whole estimation process has to be aborted. In fact, the impact of a missing or bad data item is BRP-dependent. This is related to the so-called BRP ID-card or variable weighting table, as illustrated in Table 3.

Based on this variable weighting table, the total quality of the estimation will be calculated, and if this quality is above a given threshold, the estimation will proceed. To illustrate this, two different examples are shown:

- An example in which the solar forecast data is missing, illustrated by Table 3
- An example in which the DGO infeed data is missing, illustrated by Table 4

In the first example, the result of the calculation shows that for BRP 1, the quality of the estimation would be 78%, while for BRP 2 and BRP 3 it would be 23% and 34% respectively. In such situation, with a threshold of minimum 70% for estimation, the process would continue for BRP 1 but it will be aborted for BRP 2 and BRP 3.

	Available	BRP 1(%)	BRP 2 (%)	BRP 3 (%)
DGO Infeed	Yes	54%	17%	27%
Solar Forecast	No	~~22%~~	~~77%~~	~~66%~~
SLP	Yes	15%	6%	2%
DGO Nominations	Yes	9%	0%	5%
Quality of estimation		**78%**	**23%**	**34%**
Proceed with estimation		**Yes**	**No**	**No**

Table 3: Scenario A: Missing Solar Forecast data – go/no go for estimation. Weighting of variable in estimation quality is determined during training.

In the second example, with DGO Infeed missing, the result of the calculation shows that for BRP 1, the quality of the estimation would be 46%, while for BRP 2 and BRP 3 it would be of 73% and 83% respectively. In such situation, with a threshold of minimum 70% for estimation, the process would continue for BRP 3 and BRP 2 but it would be aborted for BRP 1.

Note that, alternatively to an abortion of the prediction process in case of bad data quality, it would be possible to estimate at all situations and provide, in addition to the real time DGO estimate, the quality of estimation, so that each BRP can assess whether the estimate has to be taken into account or not. This preference is part of the questions in this public consultation (see 6.3)

	Available	BRP 1(%)	BRP 2 (%)	BRP 3 (%)
DGO Infeed	No	~~54%~~	~~27%~~	~~17%~~
Solar Forecast	Yes	22%	66%	77%
SLP	Yes	15%	2%	6%
DGO Nominations	Yes	9%	5%	0%
Quality of estimation		**46%**	**73%**	**83%**
Proceed with estimation		**No**	**Yes**	**Yes**

Table 4: Scenario A: Missing DGO Infeed data – go/no go for estimation. Weighting of variable in estimation quality is determined during training.

Please note that the examples in Table 3 and Table 4 are simplified. In the actual calculation there might be hundreds of variables, the DGO Infeed will be provided not as a single value but as a value per border point (363 as of writing this document), the Solar forecast will be provided per region and/or per DGO… etc

For some BRP, the solar forecast data might be essential for making an estimate of the DGO Allocation volume, while for others the solar forecast does not have significant influence in the estimate and hence are less critical for data quality. There could be different reasons for which this check is not met, such as:

- Missing data at the source
- Invalid data (for instance, from EMS)

The minimum quality threshold can be specified per BRP, in this way it is possible to take into account differences between the preferences of different BRP. Some BRP might prefer receiving an estimation even with a low quality of estimation, whereas other might prefer receiving an estimation of the DGO Allocation only with high quality of estimation.

Please note that a quality of estimation of 100% does not mean that the estimate is 100% accurate, it means that the estimate was calculated with all variables impacting the estimate being available.

4.5.4. Publication

The publication is a rolling publication, meaning that the complete estimated DGO Allocation volumes that are available for the current day will be published, along with the estimated forecast quality (100% if all variables were available on real-time, if not, the weight of missing variables in data-quality is subtracted). In this way, if the publication fails at a given quarter-hour, it may be recovered by the publication at the next, or a later, quarter-hour.

For the last quarter-hour of the day, we foresee a full publication on the first quarter-hour of the next day.

Thus the rules for publication, as illustrated in Figure 12, are:

- At the first quarter-hour of the day: Publish the complete estimates of previous day
- From the second quarter-hour of the day: Publish all the available estimates of the current day

Published Value

Published values / Qh of publication	D Qh 0	D Qh 1	D Qh 2	D Qh 3	…	D Qh 93	D Qh 94	D Qh 95
D Qh 1								
D Qh 2								
D Qh 3								
D Qh 4								
…								
D Qh 94								
D Qh 95								
D+1 Qh 0								

Figure 12: Publication timing

Please note that the publication of the DGO Allocation estimate will be done using existing Elia applications for communication of data to the market, in order to ease the access to this new publication. For this specific case, Elia will be using EVMS, the application used to publish metering and imbalance data.

4.6. Ex-post accuracy assessment of the DGO Allocation estimation

After reception of actual DGO Allocations volumes, it is possible to compare these to the estimated DGO Allocations for the respective quarter hour. This comparison provides information about the accuracy and performance of the estimation algorithm and can also help fine-tune the machine learning algorithm (cf. variable selection and period exclusion for training)

This quarter hourly comparison is performed per BRP and provides the following indicators:

- MAE: Mean Absolute Error – This is the average of the absolute value of the estimation error (DGO Allocation minus estimated DGO Allocation)
- MAPE: Mean Absolute Percent Error – This is the average of the absolute value of the relative error (estimation error divided by DGO Allocation)
- SAE: Standard Deviation of Absolute Error
- RMSE: Root Mean Square of error
- P90 and P10 of error

Figure 13 shows an example of such a quarter hourly comparison. The actual DGO Allocation is shown in blue while the estimate is shown in dashed orange. Some error indicators (RMSE, MAE and MAPE) are provided.

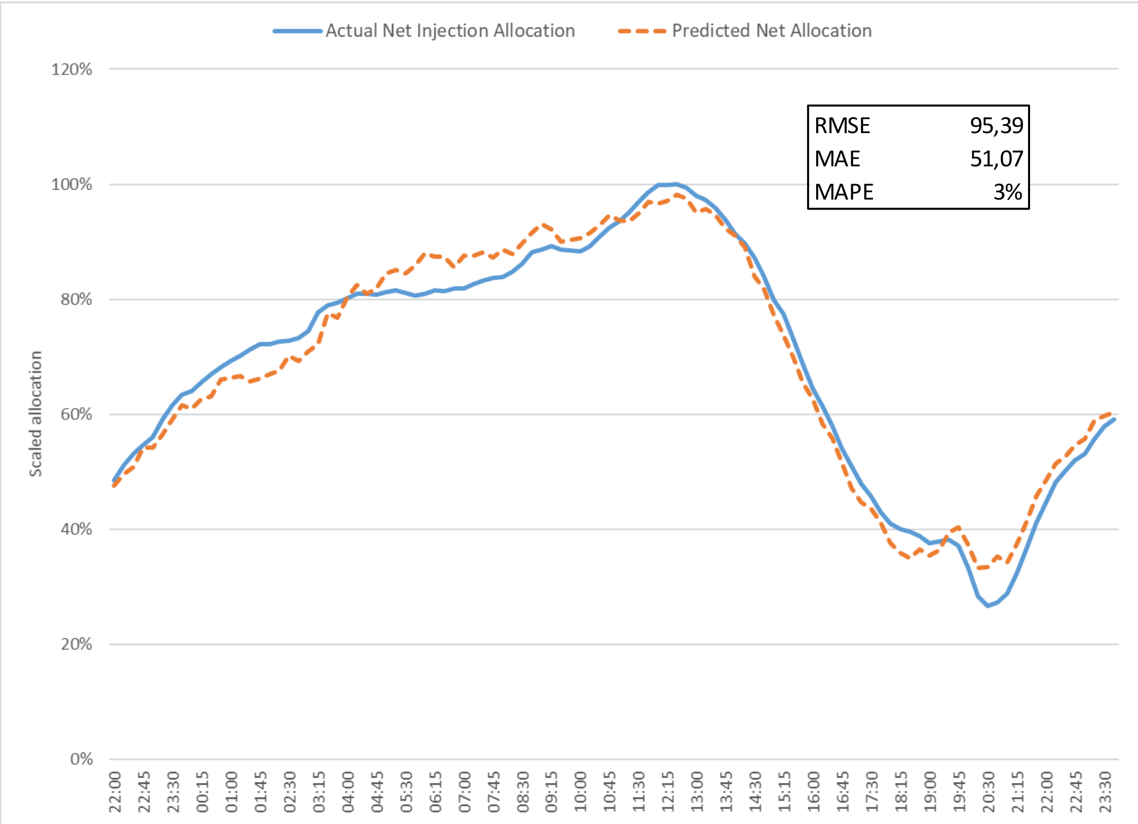

Figure 13: Comparison of prediction (dashed orange) and actual value (blue).

4.7. Hyper parametrization of the estimation model

4.7.1. Purpose and frequency

Model hyper parametrization has two purposes

- Identify the optimum variable set per BRP
- Fine tune the model technical parameters

Hyper parametrization requires a lot of computational time (1 week during first experiments in 2019) and time for analyzing the results (1 week). As it is a heavy process, it will only take place 1-2 times per year.

Note that hyper parametrization might imply changing the IT solution in order to:

- Include new variables that are not yet integrated in the data flows
- Change the way the estimation is calculated (linear model, ARIMA model…), although evolution this is not foreseen in the IT solution before 2020.

4.7.2. Optimization of variable set per BRP

As explained in previous section, the estimation of the DGO Allocation performs better with a BRP specific model, selecting the variables in the model per BRP, rather with a generic model with all variables.

During the hyper parametrization process, the optimum set of variables per BRP can be reviewed. This can be achieved by combining knowledge of the BRP portfolio and machine learning techniques.

Pure machine learning techniques try different variable set combinations until identifying the variable set providing a model with best performance. Given the number of variables, trying all combinations can take a long time. Thus reducing the search space, by using knowledge of BRP portfolio can reduce the time required for finding the optimum variable set.

For instance, some BRP have a large residential customer base, in that case the DGO Infeed variable should be selected. Other BRP have a smaller residential customer base and a higher part from PV aggregators, for these BRP it might be more efficient to use a model without DGO Infeed but with solar production day-ahead forecast and real-time estimation.

4.7.3. Fine tuning of technical parameters

During the hyper parametrization process, other technical parameters of the algorithm might be reviewed:

- Training window length: What is the length of the training window that provides the best performance?
- Cost function: What is the cost function (MAE, RMSE, combination…) that provides the most suitable model?
- Model family: Linear regression, ARIMA… Note that this will not be reviewed before 2020 as in the current IT implementation it is only possible to use linear regression models.
- Managing collinearity: Which approach provides better results, in reasonable computational time, for dealing with collinearity?

5. Planning

The 2019 CREG incentive on real-time DGO allocation included two requests:

1. Set-up an IT-solution for the real-time estimation and publication of the DGO Allocation
2. Propose improvements to the estimation methodology

In order to meet these two requests, Elia has planned the necessary IT-developments for being able to estimate and publish on real-time the DGO Allocation. In parallel, Elia has also elaborated a methodology improvement track, in which this public consultation is included.

The scope of the IT-development covers:

- The operationalization of the regression model from the PoC in 2017
- The improvement to the 2017 model in terms of
  - Inclusion of extra variable
  - Optimization of variable selection
  - Optimization of training window length

Note that the IT development **will not cover**:

- The implementation of methods other than linear regression model
- The inclusion of remarks and suggestions from this public consultation.

It is important to stress that the remarks and suggestions from this consultation will not be incorporated in this year IT-implementation, but they will be investigated and planned for the next development phase. The latter is not foreseen prior to 2020.

In the planning given in Figure 14, the IT-developments have already started and will end in mid-November, were a pilot phase with volunteering BRP is planned. The methodology track has also started and part of the improvements are described in this document (extra variables, improved variable selection and training window extension)

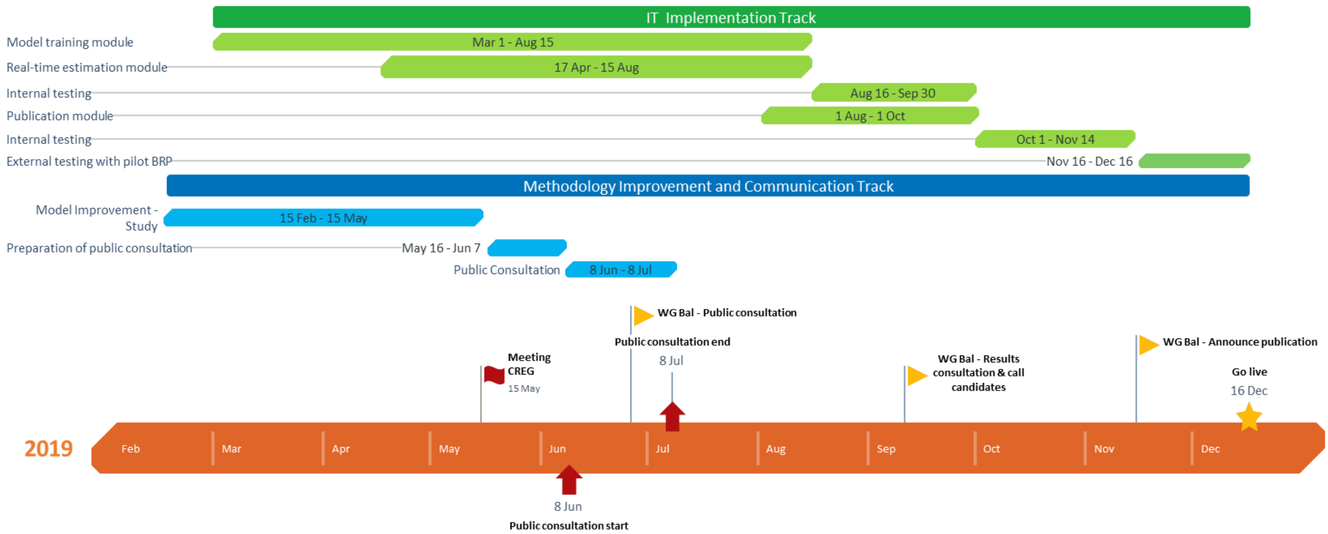

Figure 14: Planning for implementation of real-time DGO Allocation estimation and publication

6. Specific questions of this public consultation

6.1. General remarks on methodology

In this document the approach for calculating an estimate of the real time DGO Allocation has been provided. This method is based on a linear regression model that can take into account hundreds of variables. Elia is conscious that there might be other types of model with increased accuracy (such auto-regression model or neural networks) but given the time constraints for the 2019 implementation, these will not be considered at this stage.

These variables can be fine-tuned for each BRP and this has shown a significant improvement in the estimation accuracy, with respect to the PoC of 2017. Elia is currently investigating further refinement in variable selection using well-known techniques in statistical modelling (such as principal component analysis, ridge regression or stepwise regression).

We would like to know if market participants have additional information to share regarding the estimation process of DGO Allocation in real time, such as

- Are there any particular remarks regarding the modelling approach used by Elia?
- Are there other variables that might be worth taking into account?

6.2. Call for pilot BRP

In section 5 the planning for this 2019 project was presented. As part of this project, there is a testing phase foreseen, first internal within Elia and then externally, as from November 2019, with volunteering BRPs (pilot phase)

These BRPs that have volunteered for the pilot phase will receive the estimates from Elia test system. Note that these will not be using actual data and might be using test or even mock data. The purpose of this phase is to test end-to-end the complete IT-infrastructure, from data acquisition to publication to the BRP. The purpose of this pilot will be using test data, it is not relevant to measure the accuracy of the estimate.

Elia expects that the BRP actively participate in this pilot phase, by taking part in specific tests with Elia and to provide feedback to Elia about the availability of the publication service.

6.3. Preferences in case of data quality issue

In section 4.5.3 the process for dealing with data quality issues on real time was explained. Each variable is assigned a weight in the quality of the estimation of the DGO Allocation. This weight is assigned during training. It is important to stress that by variable, it is mean each individual variable and not the variable families. That is, the weight in quality will be assigned for each variable (up to 650 variables) and not to the main families of variables (Solar, Wind…).

In case of missing data, the weighting is used to determine the quality of the estimation. Should this quality be below a given threshold, the estimate will not be published. Alternatively, it would be possible to always calculate an estimate, whether there is missing data or not, and publish, along with the estimate, the calculated quality of the estimate.

It is important to understand if the market participants want to receive an estimate, even if the data quality has worsen or if they do not want to receive an estimate at all.

Therefore, we have the following specific questions:

- In case of data quality issue, do you prefer to receive an estimation (with a quality indicator)?
- Alternatively, do you prefer receiving no estimation at all if the data quality is below the threshold?

6.4. Elaboration of BRP ID Card

In section 3.2.2 the concept of BRP-ID card was introduced. The BRP-ID card allows specifying which variable families (Solar, Wind, Infeed…) have to be taken into account in the regression model used for estimating the DGO Allocation.

The optimum variable set can be determined during the algorithm hyper-parametrization, but prior knowledge of the client base (residential, industrial, renewable…) can help reducing the search space. Furthermore, it is possible to manually specify the variable combination for a given BRP without conducting a hyper-parametrization for that BRP.

The BRP-ID card also records the BRP preferences in terms of data quality. The BRP-ID card specifies the minimum level of data quality required for publishing an estimate to the BRP (cf. section 4.5.3). If the BRP preference is to receive an estimate no matter the data quality, this threshold will be set to 0%. Should the BRP prefer to receive only above a given quality factor, this will be then set in the threshold (cf. question in 6.3)

In that sense, Elia is interested in the level of involvement of the market parties in the elaboration of the BRP-ID card. In particular,

- Would you like to participate in the selection of the optimum variable set for your BRP-specific estimation model ?
- Would you like to define the minimum data quality threshold for receiving an estimate?