— FINAL REPORT

Improvement of the quality of input data for congestion management in the framework of CREG decision (B)658E/73

December 23, 2022

Table of Contents

1. Glossary

- **AI:** Artificial Intelligence
- **BRP:** Balance Responsible Party
- **CORE:** capacity calculation region in which Elia participates (16 TSOs)
- **DLR:** Dynamic Line Rating
- **EDW:** Electronic Data Warehouse (tool storing and distributing many operational data for ex-post analyses and reporting)
- **EFTool:** Elia Forecasting Tool, an internal tool acting as intermediary between different data sources and used for visualizing different forecasts
- **EMS:** Energy Management System, the tool used for real time grid operation at Elia
- **FB:** Flow Based
- **FBMC:** Flow Based Market Coupling
- **iCAROS:** Integrated Coordination of Assets for Redispatching and Operational Security, the project that designs and implements a new framework for the exchange of operational information related to outage planning, scheduling and congestion management from Y-1 to real-time
- **IGM**: Individual Grid Model. This is an extended description of the grid state for a specific target hour for a given TSO. Such hourly files are created for 3 times horizons:
    - **D2CF**: Day +2 Congestion Forecast
    - **DACF**: Day-Ahead Congestion Forecast
    - **IDCF**: IntraDay Congestion Forecast
- **CGM**[1]: Common Grid Model is the product of the combination of IGMs of many TSOs for a given hour
- **L&G:** Load & Generation is an ongoing internal project aiming at aligning the real time model (EMS) with the planning model (Power Factory)
- **ML:** Machine Learning
- **OPAL:** internal tool of Elia gathering all residential and industrial loads with their associated electrical localization as well as the installed capacities of decentralized units at each of those electrical localizations
- **PF:** Power Factory, commercial load flow tool used at Elia for operational planning and other grid calculations
- **PISA:** Elia database listing all known production units in Belgium (past, present and future units)
- **PST:** Phase Shifting Transformer
- **PTDF:** Power Transfer Distribution Factors
- **PV:** Photo Voltaic

[1] CGMM-v2-plus and CGMM-v3 approved, methodological framework for Common Grid Model process complete (entsoe.eu)

- **QH:** Quarter hour
- **RCC:** Regional Coordination Centre
- **RMSE:** Root Mean Square Error
- **RPN**: Referential Production Netcalc, internal centralized netcalc production database
- **TIC (Mnemonic):** Traitement Intégré des Comptages, used to manage metering data

2. Introduction

On December 9, 2021, CREG issued decision (B)658E/73 on the targets to be achieved by Elia in 2022 in the framework of the system balance as referred to in Article 27 of the tariff methodology. One of the incentives fixed in this decision is “Improvement of the quality of input data for congestion management”. The description indicates that the incentive primarily includes a report on the analysis of the most significant deviations between predictions and reality and an examination of possible short- and long-term solutions and secondly, includes a recommendation and proposal for implementation of concrete solutions in the form of a roadmap for the future.

The objective of this report is to give an overview of the current modelling practices (Individual and Common Grid Model) and to show transparency on the actual quality of the input data for congestion management. The report also covers a root-cause analysis of deviations in the input data (different forecasts f.e. wind, solar, load,…) and an implementation roadmap of a series of improvements.

3. Scope and approach

As proposed in execution of the incentive on "Improvement of transparency with regards to the detection and management of congestion" defined in CREG decision (B)658E/52 of 28 June 2018, Elia publishes since the beginning of 2020 a quarterly report[2] on congestion management covering a period of three months. This report includes:

- Information on the quality of forecasts used as operational input data for the creation of the Individual Grid Models (IGMs)
- Information on the quality of output data
- Information about the timing, power, location, and purpose for activations of Costly Remedial Actions by Elia.

The incentive defined in decision (B)658E/73 for 2022 builds further on the incentive realized in 2019. Based on the reports mentioned above, deviations can be detected between the input data used for the creation of the IGMs and the reality. On top, based on all the data stored in order to publish the reports for more than 2 years, it is possible to perform

[2] Congestion management (elia.be)

a more in-depth analysis. The scope of the incentive is to set up this in-depth analysis of the causes of the deviations in the different forecasts (wind, solar, load,...) and to study possible solutions to improve these forecasts. These solutions can be found in the infeed data, the forecasting model or the resulting forecast.

The different forecasts are analyzed on an individual basis and for each forecast the following aspects are studied and structured in the report (if relevant):

- Infeed data, forecasting model, resulting forecast (into IGM input): see figure below to make the link with the KPIs in the current quarterly reporting
- AS IS versus possible TO BE situation

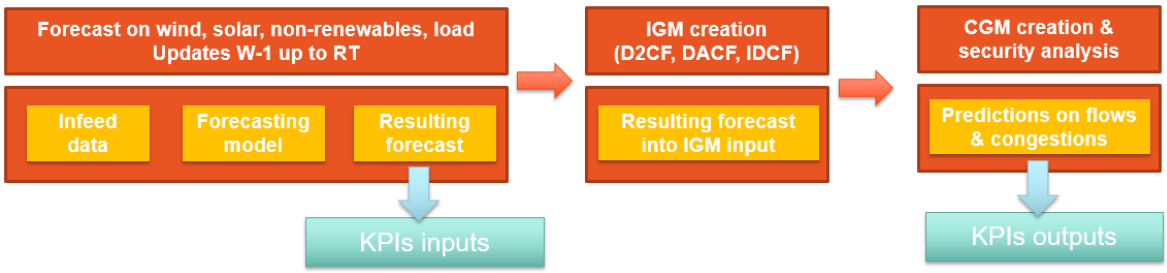


Figure 1: overview of the scope of the analysis performed in the report

It is part of the incentive to start an implementation plan for short-term solutions and to establish a roadmap for long-term solutions. The goal of the incentive is not as such to improve all the individual forecasts, but to focus on improvements and needed developments in order to keep a good level of congestion management decisions and avoid unnecessary costs.

The report is structured into the following sections, in line with the concrete deliverables as foreseen in the decision of CREG:

- Chapter 4: Transparency on congestion forecast today and current modelling practices. This chapter also includes a mapping of the opportunities for improvements and a benchmarking with other TSOs.
- Chapter 5: Transparency on forecast quality, root-cause analysis on deviations in forecast compared to real time and possible solutions (short- and long-term) to improve the forecasts.
- Chapter 6: Challenges and opportunities for the future. This chapter makes sure that the implementation roadmap anticipates the probable evolutions of the system.
- Chapter 7: Roadmap for implementing the selected improvements.

The content of this incentive has been discussed with several internal and external experts. A workshop has been organized to improve the transparency on the content and a public consultation has been organized[3].
The comments received during multiple occasions are used to elaborate the report and to develop the roadmap for implementation of the selected improvements. Since the roadmap covers multiple years, Elia can present on request at any time an implementation status in the coming years.

4. Congestion forecast, current modeling practices and performances

Every day Elia has to create the Individual Grid Model (IGM) for Belgium and this for 3 time horizons:

- **D2CF**: Day +2 Congestion Forecast
- **DACF**: Day-Ahead Congestion Forecast
- **IDCF**: IntraDay Congestion Forecast

The D2CF is one of the main inputs for the day ahead capacity calculation process. There is a strong link between DACFs/IDCFs and D2CFs and any error in the D2CF flow forecast may lead to unrealistic cross-border capacities given to the market. The congestions will only materialize if the market goes in this unrealistic position and if the Flow Reliability Margins (FRMs) are not large enough to cover those errors. Deviations in flow forecasts in day-ahead and intraday (DACFs and IDCFs) have a direct impact on congestion management, since the DACFs and IDCFs are the unique source detecting operational security violations on the grid and proposing coordinated remedial action to solve the identified constraints, which is the aim of the coordinated security analysis performed by the RCCs[4]. Of course deviations on uncongested elements are not relevant (e.g. on strong radial elements typically). The root-cause analysis of deviations in the report is focused on the day-ahead timeframe. Most of the proposed solutions will of course also have a positive impact on the D2CF and IDCF quality.

Note that the network calculations realized in week-ahead are not using forecasts. For this time horizon Elia calculates and solves several scenarios representing “realistic worst cases” ensuring that the outage plan and associated grid topology for the next week is acceptable. An accurate model at each node is of great importance in order to calculate realistic scenarios. However, improving week-ahead forecast would have no influence on the congestion management.

[3] Publieke consultatie over de verbetering van de kwaliteit van de inputgegevens voor congestiebeheer in het kader van een CREG incentive. (elia.be)
[4] Coordinated Security Analysis | Coreso

4.1 IGMs building process

IGMs are always built on the “best estimate” principle (i.e. most probable situation is the reference). The figure s below give an overview of the D2CF compared to the DACF/IDCF building process. More details are provided in the paragraphs below.

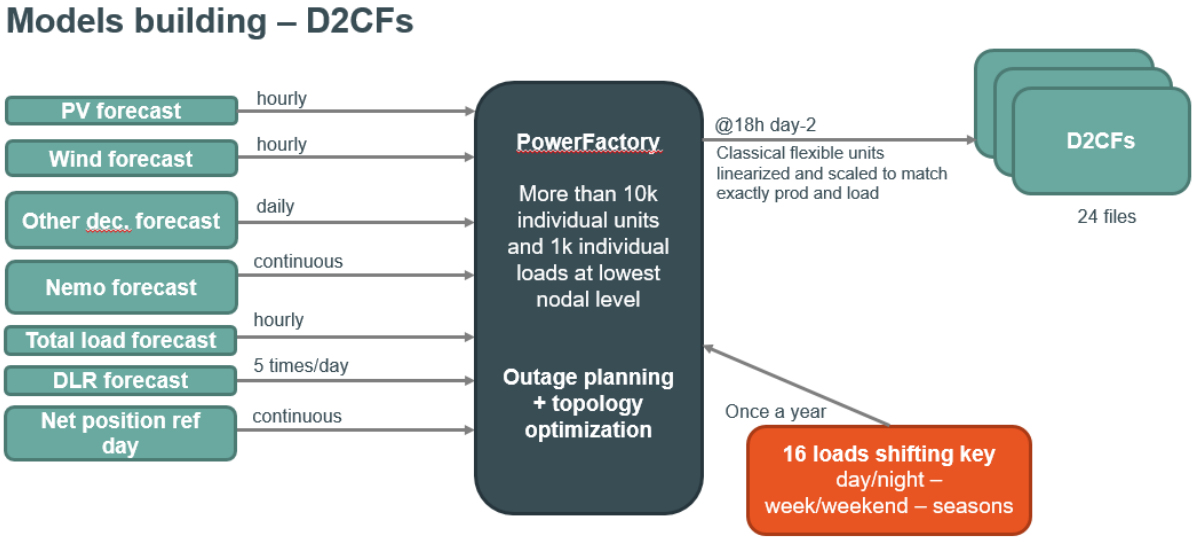


Figure 2: overview of D2CFs building process

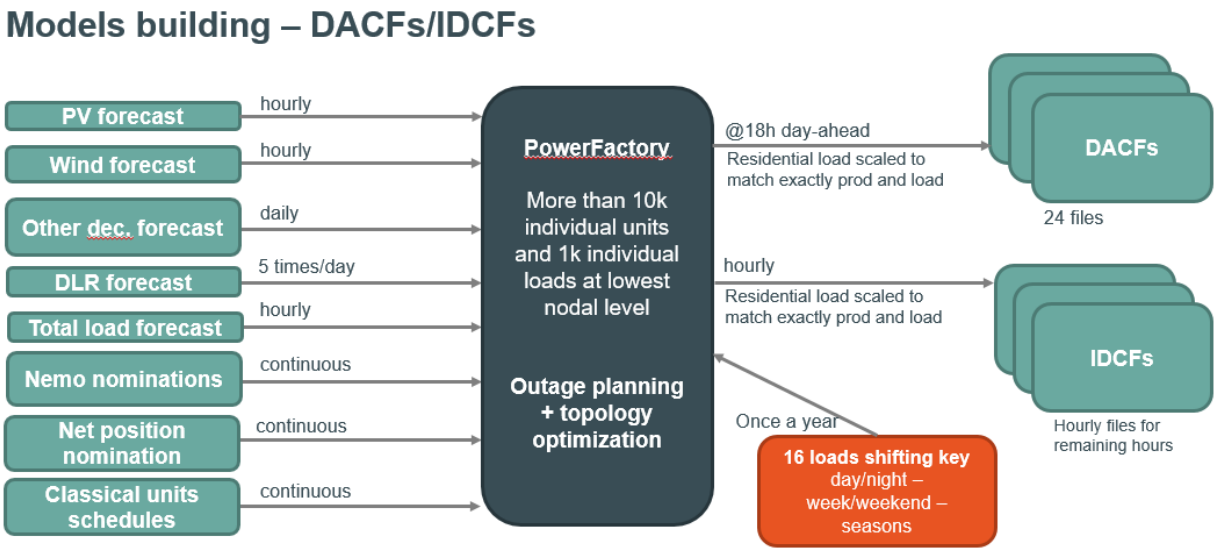


Figure 3: overview of DACFs/IDCFs building process

4.1.1 D2CF

The D2CF is one of the main inputs for the day ahead capacity calculation[5] process. Before 7 p.m., D2CF files for each hour of the day + 2 are created with the following contents and hypotheses:

- Latest (hourly updates mostly) **decentralized production forecasts** at power plant level (i.e. most accurate electrical localization as possible, > 10 000 individual units).
- Latest (hourly updates) **total load forecast** which is then spread over about 1000 individual loads by applying one of the 16 **repartition keys** depending on season, day/night and week/weekend categories.
- All **outages** down to the lowest voltage level and their preventive topology changes when necessary. Usually PSTs taps positions are neutral (will be optimized by the process itself, only north border often anticipated at 15/15/15/15 as a better best estimate position). With GO-live of FB CORE DA early June 2022, initial PSTs taps are pre-optimized (before first capacity calculation) in order to minimize loopflows.
- **Import/export net position** of the referential day imposed by the CGM process in order to have coherent IGMs from all countries (prerequisite for merging purposes). The referential day approach will evolve (expected in 2023) towards a central forecast of net positions for all CORE TSOs.
- Latest **NEMO** flow forecast (forecast bought by Elia from an external provider). Again, this will evolve towards the central forecast of net positions for all CORE TSOs (including NEMO flow).
- **Alegro** = Reference Day Schedule (but removed during capacity calculation, in order to be fully optimized by the FB process). With GO-live FB CORE DA, the net position forecast will also predict the initial set point of Alegro.
- **Element ratings**[6] based on forecasted temperature + dynamic line ratings for concerned lines (capped to 5% in order to ensure high reliability).
- **"Must run" conventional units** (essentially nuclear units). Such units are assumed to be running at their maximal capabilities.

  ➔ The only free parameter is the production of flexible conventional units. Total production of such units is imposed by the other fixed parameters and is then linearly (same % value w.r.t. units maximal production level) ventilated to all available flexible conventional units. In case there are not enough available flexible volumes to reach the imposed target, a scaling on residual load is done. This may happen when the referential day has a strong difference in renewable production level compared to the target day. The resulting imposed net position might be unrealistic in this case.

  ➔ With FB CORE DA go-live the net position CORE forecast ensures that the flexible conventional units are not saturated and no scaling of the residual load will be necessary.

[5] Annex%20I%20-%20ACER%20Decision%20on%20Core%20CCM.pdf (europa.eu)
[6] Dynamic Line Rating (elia.be)

4.1.2 DACF

The DACF is the main input for the coordinated security analysis performed by the RCCs. Before 18h, DACF files for each hour of the day + 1 are created with the following contents and hypotheses:

- Latest (hourly updates mostly) **decentralized production forecasts** at power plant level (i.e. most accurate electrical localization as possible, > 10 000 individual units).
- Latest (hourly updates) **total load forecast** which is then spread over about 1000 individual loads by applying one of the 16 **repartition keys** depending on season, day/night and week/weekend categories. Note that all the industrial loads are fixed but residential loads are the free parameter (see hereunder).
- All **outages** down to the lowest voltage level and their preventive topology changes when necessary. PSTs taps positions based on foreseen market flows, outages and recent taps positions. Topology will at any case evolve throughout the process based on identified congestions.
- **Net position** resulting from the international trades (nominations).
- **NEMO** nominations.
- **Alegro** set points defined by FB MC.
- **Element ratings** based on forecasted temperature + dynamic line ratings for concerned lines (capped to 5%).
- **Conventional units** schedules.

  ➔ The only free parameter is the residential loads. Any error in nominations or in decentralized production forecast will then propagate itself to the residential loads.

4.1.3 IDCF

IDCF files are regenerated automatically every hour for the remaining hours of the day with the latest available information following the same principles as DACF files.

4.2 Tooling and available input data at Elia

Before jumping into the detailed analysis of each specific forecast, we will shortly explain which type of data Elia has regarding production units and loads and how Elia manage them in its tools and processes.

Every month DSOs upload all new installed production capacities bigger than 400kW into the internal tool PISA. They have no obligation to include smaller units in this process but experience shows that they often still do it. Usually they bundle together the very small units of the same type and located at the same electrical localization (LGL). They often share, even if not mandatory units not yet commissioned. Elia encodes all units connected to its grid into the tool PISA as well. As such, PISA should contain a very exhaustive list of all production units bigger than 400kW located in Belgium but also many smaller units. Many attributes are available in the tool as well for example the electrical localization (LGL), the metering codes, the fuel type, the installed capacity etc.

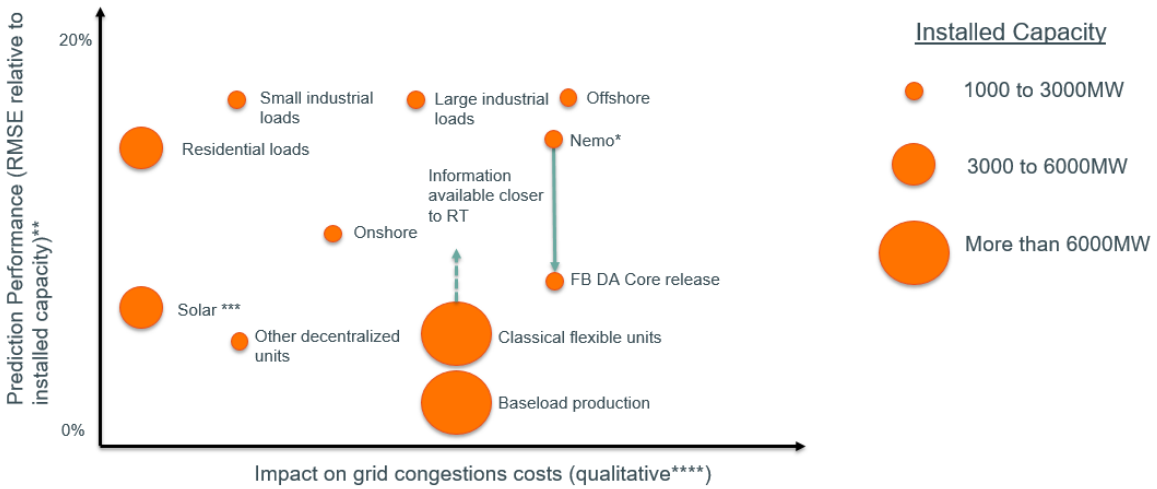
Elia generates an updated cadaster directly from the PISA tool but some manual steps are still required. For example for the wind cadaster (see the flow corresponding flow chart in the wind forecast section below) one must add the GPS location of the turbines and the real-time measurement code if any (see below in each corresponding sections the detailed explanation for the other data flows). Elia shares the cadaster with the external forecasting provider and one must implement the updated cadaster both in their tools and in Elia's tool in a synchronized way (application of the cadaster for the same date and time in the future). Elia Forecasting Tool (EFTool) takes care of data redistribution, ex-post processing and quality checks. EFTool receives continuously real-time measurements, uses them for upscaling the non-measured units and shares it all continuously with the external provider so that he can make use of latest data for improving the forecasts. Every day data from SOROBAN (the metering tool) are fetched into EFTool with a lag of 1 or 2 days. Again, it computes the upscalement on non-metered units and share it all with the external provider so that the best available historic data are available for generating the forecasts.

Each time that an operator wants to generate IGMs via Power Factory, RPN is automatically called. RPN asks EFTool the latest running values for each individual units for each hours or the IGMs to be created and push them into the Power Factory project.

A graphical representation of the data flow and use of tools can be found for each individuald forecast in the next chapter.

4.3 Mapping of opportunities

Before starting to reflect on possible solutions for forecast improvements, it is important to evaluate for which forecasts there is room for improvement, but also to evaluate which improvements have the largest impact on congestions.



*D2CF error
**Lower is better, RMSE day-ahead forecast 2021.
***For daylight hours only
****looking at PTDFs on grid elements might help to quantify the axis (work ongoing)

Figure 4: accuracy vs. impact on congestion for different forecasts

The position on the horizontal axis is strongly related to the concentration of the production (or load). If the units are concentrated in a restricted geographical area (such as the offshore wind parks), the impact on the congestions of a forecast error will be much stronger than a multitude of smaller units equally distributed on the territory (such as residential loads). The electrical localization and the associated costs of available remedial actions may also play a role in the x-axis determination.

From figure 5 it can be concluded that the segmentation of elements leads to interesting results: the total load would appear as a large circle on the bottom-left of the graph (low RMSE and low impact on the congestions). By splitting the loads into different categories and by computing the average of all the individual RMSEs of each category instead of the RMSE of the aggregated time-series, one can see that the large industrial loads have a stronger impact on the congestions (they are concentrated) and their forecasts errors are high.

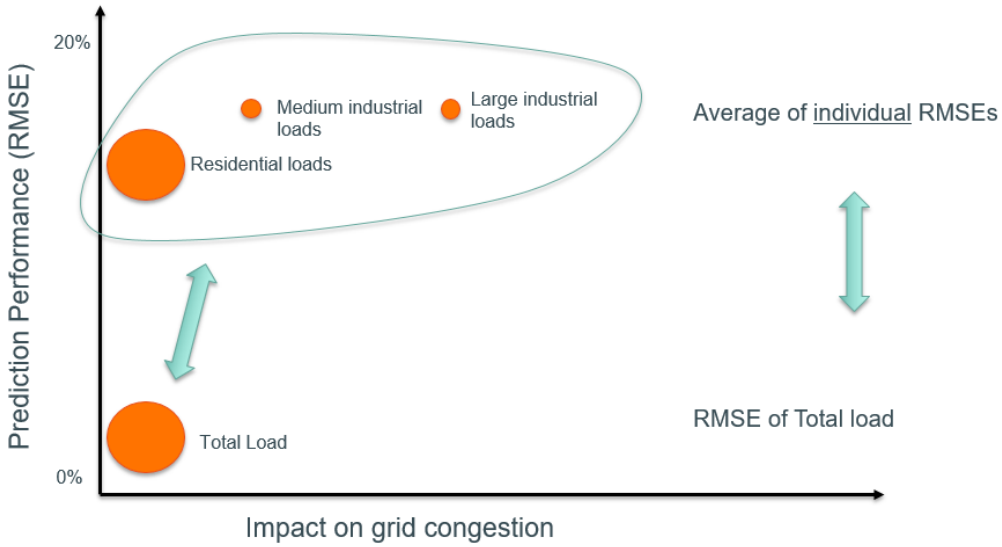


Figure 5: individual loads nodal accuracy comparison with aggregate forecast accuracy

Segmenting the “other decentralized units” leads to the same conclusion: the individual RMSE is higher and the influence on the grid is stronger for the large units (larger than 5MW).

A distinction must also be made regarding the RMSE of an aggregated category and the average of the individual RMSEs of the individuals from this category. In the congestion reduction paradigm, it is important to consider the individual RMSEs for the bigger elements since they are likely to influence the congestions.

4.4 Benchmarking

For some of the main forecasts (solar and wind), a benchmarking compared to different TSOs in Europe is made based on data available in the ENTSO-E database, see figure below.

However, comparing performances of forecasts among different TSOs is not as straightforward. Hereunder are some elements that can explain why a direct comparison might be misleading:

- The number of measuring points: not a single TSO has the exact historical measures of the total PV production of its bidding zone. At Elia this service is purchased from an external provider which uses about 80k measurement points to extrapolate the unmeasured ones. This approach is probably the most accurate way to compile such time-series but many TSOs are estimating it only based on realized weather data. Comparing RMSEs of two TSOs will certainly not determine which of them is the closest to the realized productions (the latter being unknown to both).
- Nodal versus total forecast: since the voltage levels operated by Elia are much lower than for most of the other TSOs (down to 36 kV) the information that Elia has is much more precise. It means that Elia must predict and model its grid in a much more detailed way.

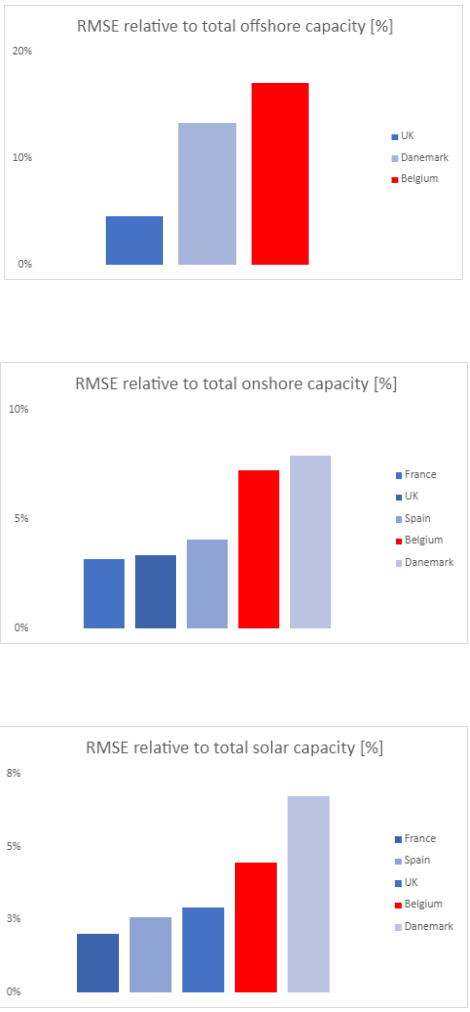


Figure 6: RMSE comparison for some forecast compared to other TSOs

The general trend seems to show that there is room for improvement for all renewable forecasts. The reader must keep in mind that the RMSE is just a partial indicator of the performance of a given forecast but for the sake of clarity this is the KPI that was compared for this benchmarking between several selected TSOs.

In order to improve the benchmarking, it would be needed to contact the individual TSOs in order to better understand their approaches and potentially learn from their methodology and evaluate whether it is compatible with Elia's forecasts.

4.5 Enablers

There are a few ongoing projects having an impact on the content and roadmap that will be elaborated in this report. The projects below are defined as enablers for some of the improvements proposed in the roadmap. The expected go-live of those enablers are mentioned in the roadmap in section 7.

4.5.1 Communication type B project

Elia has launched this project in order to offer economically viable solutions to type B (between 1 and 25MW) units for exchanging real-time measurement data. Today the only option is to lay a fiber optic to the closest remote terminal unit (RTU) of Elia. The proposed alternative is a kind of VPN connection. A couple of proof of concepts are foreseen early 2023 to validate the approach. In 2024, Elia will start regularizing the many existing type B units not sending any real-time measurement. Consequently, Elia expects to receive more real time data in the EMS, which is one of the enabler of better forecasts.

4.5.2 Power Factory and EMS alignment

Currently both tools are different in many aspects. Power Factory is a tool that contains a very detailed and exhaustive model of all Elia's assets. It contains many technical features of the assets, which is a prerequisite for an offline tool to best simulate complex grid behavior. Power Factory is used for offline grid modelling within Elia (from multiyear studies of grid development to intraday congestions management). On the other hand, EMS is a tool used for real-time grid management. It does not need as much details as Power Factory because it receives many real-time data meaning that it does not need to calculate all the equations describing the electrical behaviors but rather display the measurements. On top of the many real-time data acquisition, the tool must fill the gaps. Many items modelled in the tool, certainly in the lowest voltages, have no real-time measurements. An algorithm called the state estimator is continuously estimating all the non-measured elements by looking at the most probable flow repartition. As soon as there are enough real-time measurements in the tool the state estimator is performing well (accurate and quick). It is worth noticing that the EMS is most of the time reducing the lowest voltage levels to equivalent loads by lack of real-time measurements. An equivalent load is a fictive load encompassing all unknown loads and productions hidden behind a bus bar and the state estimator is then indirectly able to evaluate the loading of this item. Consequently, the EMS contains far less information as Power Factory. For example, in the EMS an overhead line is modelled as a single element with a single seasonal rating. In reality, an overhead line is constituted by many elements in its bays as well

as sometimes several conductors' types. Each of the sub-element has a different rating and a different dependency to weather conditions. If the EMS had all this information it could use more accurate ratings as today's ratings.

In the past it made sense to let both tools develop and optimize their grid structure in order to perform as good as possible. However, independently from all the studies realized for this incentive, Elia recently launched a project aiming at aligning as much as possible the assets definitions throughout the entire company, not only for Power Factory and EMS but also for asset maintenance purposes and tools, infrastructures project teams, etc.

4.5.3 Loads and generations project

On top of the alignment of the Elia assets in Power Factory and EMS, the "Loads and generations" project will ensure that the granularity of loads and generations are identical in both tools. Indeed, having the same grid structures in both tools does not ensure that the loads and generations will be aligned. The foreseen approach is shown in the graph below and is not yet final but the idea is aggregate loads (from OPAL) and generations (from PISA) according to the same rule in both PF and EMS models. The result will be equivalent productions and loads on each bus bar down to the lowest levels of Elia's grid (instead of a single virtual load). Any steps in that direction may only help to increase the quality of the IGMs allowing a better comparison at nodal level (even at individual units/loads sometimes) of measured data from EMS and forecasted data from PF. It will also allow to feed forecasting models with real time data (estimation or measurements). The project will also foresee automatic monitoring and sanity checks on measured and estimated data improving the general quality of measured data as well as model data.
Note that one must first align the grid modelling in both tools (lines, bus bars, etc.) in order to reach this goal.

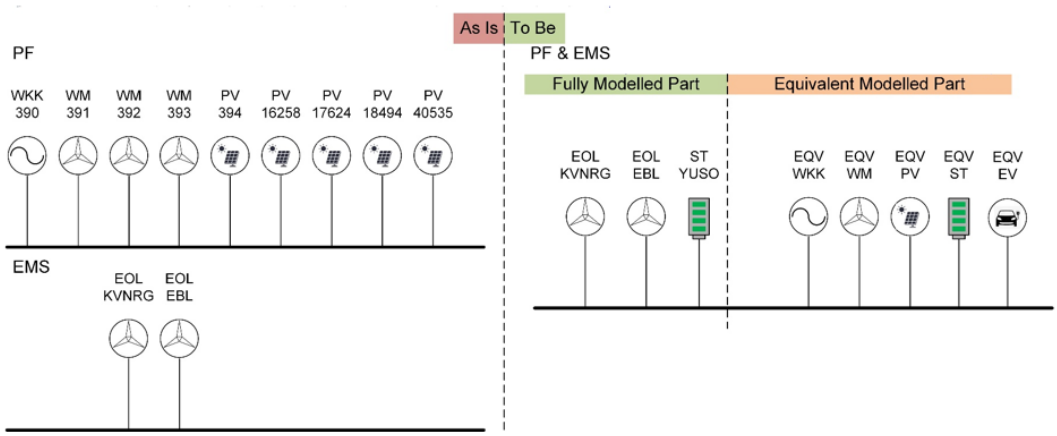


Figure 7: Evolution of the modelling of the real time tool for grid operation at Elia

4.5.4 iCAROS project

New regulation, a changing energy landscape and an evolution in operational needs call for a major evolution for the coordination of assets and congestion management. The iCAROS project, "Integrated Coordination of Assets for Re-dispatching and Operational Security", aims to redesign the whole process of operational data exchange between

Market Parties and Elia for outage planning, scheduling and congestion management. All type B-C-D units (>1MW) connected either to TSO and DSO will have to provide these data. The implementation of iCAROS is distributed in three phases:

- **Phase 1:** state of the art design but focus on system relevant assets ≥ 25 MW. No new forecasting data are expected in this phase.
- **Phase 2:** extension of state of the art design to all system relevant assets ≥ 1 MW & demand facilities (only TSO-connected). New data for outage planning (DSO & TSO connected), scheduling and congestion management (TSO connected assets and demand facilities).
- **Phase 3:** full extension of state of the art design to all system relevant assets ≥ 1 MW & demand facilities (only TSO-connected). New data for scheduling and congestion management (DSO connected assets).

4.6 Analysis from congestions in the past

In order to properly assess the interest of an improvement of the forecasting in general, it is important to quantify the proportion of costly remedial action that was taken because of an error between the model and the reality. The first part will review the cause of a random sample of the remedial action from the last years. Then a segmentation is made to show where a forecast improvement would have changed the outcome. Finally, a quantification of the improvement is done for the load forecast, as an example.

4.6.1 General overview of the yearly costly remedial actions

The graph below shows a sample of 64 activations of costly remedial actions over the last three years (from 2020 to 2022). The activations are sorted according to three categories based on the explanations provided by the dispatchers for each of the occurrence as well as some investigation in the flows surrounding the congested element:

- *"Perfectible with better forecasts":* flows forecast accuracy of the congested element(s) is correlated with the forecasts accuracy from some elements in the IGM.
- *"No added value with better forecast"*: the origin of the redispatching was not linked to a flow forecast error for example trip of a line or any other element or the activation of aFRR from a power plant.
- *"Undefined"*: it was not possible to draw a clear conclusion on the origin of the redispatching. Note that this is among other linked to the misalignment between EMS and Power Factory models.

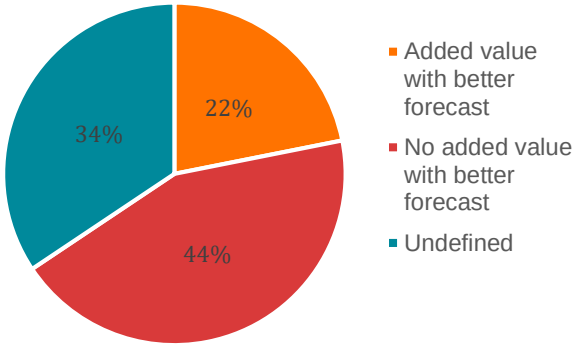


Figure 8: Segmentation of the redispatching cases (n=64)

Among the perfectible cases, an additional segmentation gives the following graph:

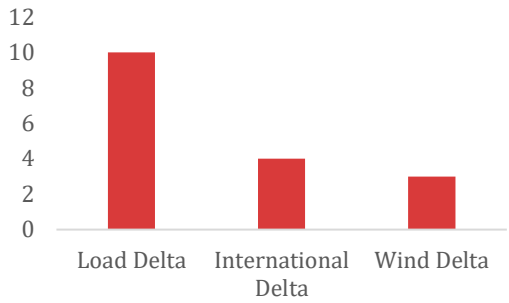


Figure 9: Reasons for the Deltas (n = 14)

More than half of the cases are related to the quality of the loads, the rest is divided between the incorrect forecast on the wind and on the error in the international flows (where the influence of Elia is somehow limited).
Errors on congested lines come from many sources and this is complicated to isolate those effects certainly with the current modelling differences between EMS and Power Factory. Nonetheless, this analysis will help Elia on the prioritization of items in its implementation roadmap.

5. Root-cause analysis and possible solutions

The different forecasts are analyzed on an individual basis, focusing on the following aspects:

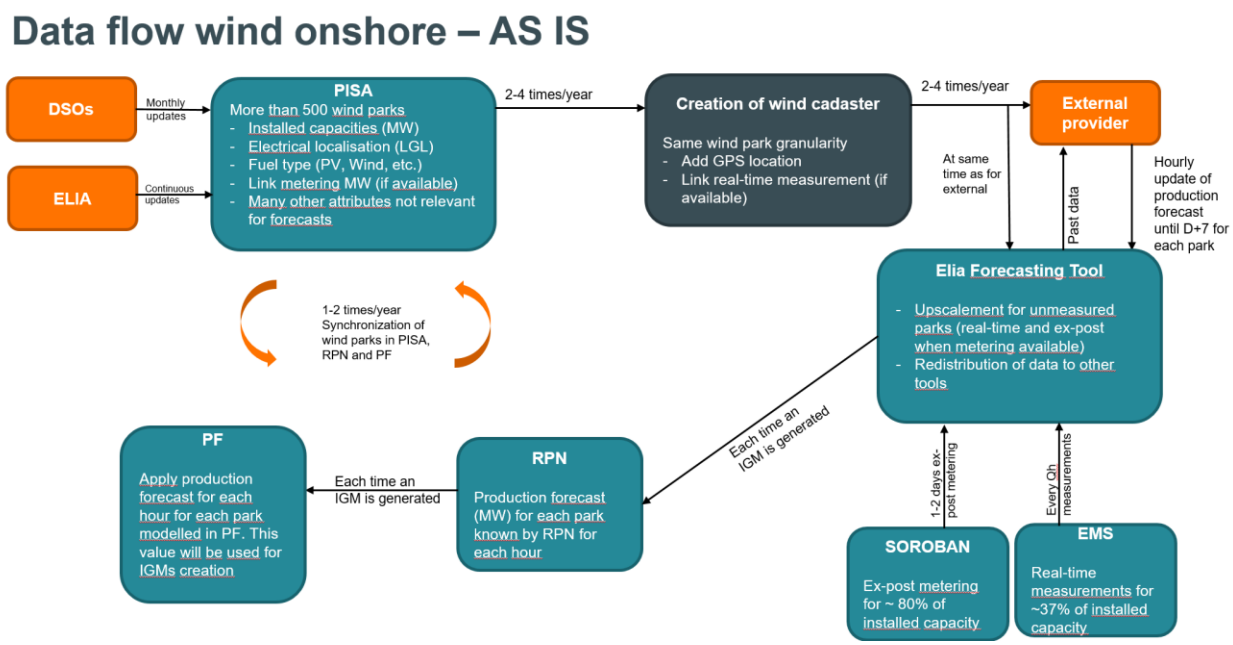
- Infeed data, forecasting model, resulting forecast
- AS IS versus possible TO BE situation

For the possible TO BE situations, a first indication is given on the effort (low, medium, high) and impact on congestions (low, medium, high). This information will be taken into account later on in the short-term implementation plan and long-term roadmap.

5.1 Wind Forecast

- More than 500 individual onshore parks amounting for 2900 MW listed in PISA database. Onshore parks smaller than 400kW are very rare in Belgium (a typical onshore wind turbine nowadays is 3,5GW) so we can assume that this represent very will the reality.
- 85% of those parks are lower than 10MW
- Each of the 500 parks in PISA is modelled and forecasted individually by an external provider
- Each park is geographically and electrically precisely located
- RMSEs for onshore forecast at individual level is ~10% (~7% at aggregated level) and ~18% for offshore (same at aggregated level because all at a similar location).
- Impact on congestions is also high (also for some onshore wind parks because they are sometimes concentrated on the same area, saturating the lower voltage grid).

Elia could not identify any game-changers to significantly improve this forecast but a structural data quality check could be realized on the onshore cadaster file. Fine-tuning GPS locations, installed capacities, measurement quality etc. could already improve the forecast accuracy at nodal level.



5.1.1 AS IS

INFEED DATA

- **Cadaster**: Installed capacities by GPS locations are necessary. More detailed data (number of turbines, turbine type/brand, hub height) were used in the past but with the evolution of forecasting algorithms (ML, AI, auto-tune etc.) those are not helping anymore and allowed to simplify the cadaster and concentrate on the inputs actually improving the prediction.
- **Historical data**:
    - **Offshore**: For every offshore park, the power output is directly measured with precision in real-time.
    - **Onshore**: For onshore parks only 37% of the total installed capacity is measured in real time, the rest being up-scaled based on the average power factor of the measured ones. A couple of days later the tool receives accurate metering data for about 80% of the onshore parks. Then the upscalement is redone with this new information and republished ex-post to improve the past data.
- **Wind curtailment ordered by Elia**: are shared with the external provider so that those events can be discarded from the training period for its predictive algorithm.

FORECASTING MODEL

- Currently, **two providers** are supplying data for the wind power forecast.
    - The first supplier (IRM) focuses only on the offshore wind parks and still uses “classical” method (i.e. non Machine Learning) to provide the forecast. This is also a forecast optimized for the power predictions during storms.
    - The second supplier (VITEC) is used for the onshore production (but delivers also the offshore production) and uses machine learning and data from several weather forecasters in order to provide an accurate forecast.

    ➔ IRM offshore forecast is supposedly better during storms but Elia realized that VITEC outperforms during normal situations. The performance gap between the two providers is around 3% (18% RMSE for IRM and 15% for VITEC).
- VITEC: hourly updates. Buys weather data from many global models on specific weather stations near wind parks. Model retrains itself automatically every day based on the 90 last days. It uses very recent data to correct the forecast of the next hours to come. It removes outliers from the training period.
- IRM: 5 updates/day. Model is not retrained except on demand and it takes a lot of effort. No outliers detection, no machine learning.

RESULTING FORECAST

- **RPN imports latest forecasted data** from EFTool and sends them to PF at the exact electrical node each time an IGM is created.
- **RPN/PISA/PF sync** -> see below on other decentralized productions (for AS IS and possible TO BE)

5.1.2 Possible TO BE

INFEED DATA

- **Compare the power factor:** The idea is to compare the power factor of the parks that are geographically close to each other. The main benefit of this analysis would be to detect any error in the installed capacity or a structural error of the provider. Moreover, Elia could improve the upscaling rule by applying a different upscaling factor based on the locations by using the power factor of the measured parks in the vicinity of unmeasured parks)
    - **Impact low:** the other checks proposed below should probably solve potential issues. Applying a more elaborate upscaling might make sense but this is still to be confirmed after the comparison work.
    - **Effort low:** The work should not be too hard to do and can be done with easily available data.

→ In the roadmap: "onshore cadaster quality check" includes this check in 2023

- **Using VITEC tool to spot inconsistencies** in the cadaster data (fine-tune GPS coordinates & installed capacities, issues on measurement data, etc.). This tool displays the RMSEs of each individual park and can identify some inconsistencies.
    - **Impact medium:** Elia has already spotted such issues in the past so a structural check for all wind parks would probably bring some benefit.
    - **Effort low/medium:** Elia could ask a student or starter to deep-dive into this relatively elaborate tool. The trouble remains that we regularly change providers and we do not want to invest too much time on learning their own tools.

→ Not in the roadmap: Elia will use internal data and tools for those inconsistency checks (see below).

- **Compare real time measurements with ex-post metering:** This data is only recently available but Elia should check the coherency on both data sources.
    - **Impact medium:** Elia has already spotted such issues in the past so a structural check for all wind parks would probably bring some benefit.
    - **Effort low:** Both data series will be available in an internal tool at park level. This is not yet the case but will be available by end 2022.

→ In the roadmap: "onshore cadaster quality check" includes this check in 2023

- **Increasing the real time measurement ratio**: all TSO-connected units do have such measurement but Elia still struggles to receive it for type B units because they need to lay an optic fiber to the closest RTU (remote terminal unit) of Elia.

- **Impact low:** it will greatly increase production estimation for the real time and the last hours in the past but as soon as Elia has metering data the added-value is gone. Considering the fact that forecast is mostly based on historical data (last 90 days) already including metering data for most of it, this improvement will only help for the short-term corrections of the forecast (the coming 2-3 hours). Increasing the "nowcasting" is interesting but still a bit late for being taken into congestion management.
- **Effort low:** project "type B measurement" will set up a new protocol of live data exchange in order to receive real-time data of smaller units at lower costs. Elia expects to receive a lot more live data. Thanks to the Load and Generation project the mapping with forecasted values in PF will be straightforward. It is worth noticing that in 2022 we already could increase the ration from 25% to 37% for onshore while installed capacity increases.

→ Not in the roadmap but we witness some improvements from the DSOs side. They are delivering more measurements the recent years so we can expect a continuous slow improvement until go live of project "Type B" by mid-2024. Then we expect measurement ratio to improve rapidly.

- **Including maintenance information**, certainly for offshore parks. Today Elia is using the forecast as input for the IGMs assuming that our forecast is in general better than those of the ARPs/SAs. Actually some owners might do a great job while others may not have an as elaborated forecast as Elia. Today Elia always supposes that the full park is available, the idea would be to scale the forecast based on the maintenance information.
  - **Impact low:** offshore parks are usually available when the wind blows and they usually stop a single turbine at a time for maintenance but sometimes they also have forced or planned outages of several turbines for several days (so also during high wind periods)
  - **Effort medium**: Using an internal tool to read and compare offshore nominations with offshore predictions at wind park level as well as the max available capacity provided by the BRP would allow the operators adapt the final forecast accordingly.

→ In the roadmap: "offshore maintenance" will ensure implementation in tools and procedure by end of 2024.

- **Using measurement data from other offshore parks in the surroundings:** in order to better forecast ramping events. When offshore wind changes rapidly it is a challenge for the prediction to be time-accurate.
  - **Impact low:** such event does not happen too often and IGMs granularity is hourly while the challenge is more on the 15min accuracy. Moreover, the errors only happen during the ramping events.
  - **Effort high:** need international collaboration and set up.

→ Not in the roadmap, this evolution might still be pushed for balancing purposes but it is out of scope of this incentive.

FORECASTING MODEL

- **Periodically challenge the suppliers:** The idea is to use existing in-house machine learning software and feed them with the data provided by the suppliers. Elia can then detect if there is room for improvement (or new strange behavior) for the forecast quality.
    - **Impact Low:** Elia organizes usually a call for tender every 3 years and selects the best candidates meaning that quality of such services is high already. Detecting errors with a long lead time is only relevant for structural errors while most of errors/bugs are sporadic events. Tracking performances on a shorter period remains valid to ensure that any eventual issue is detected and solved as quickly as possible.
    - **Effort Low:** some basic machine learning tools available on the market and easy to use so the check should be quite fast.

  Same idea could be pushed further with a live monitoring dashboard, notifying users in case of strange behavior in order to avoid any issue to propagate for a long period of time.
    - **Impact medium**: This would allow Elia to correct sporadic bugs right away. This effect will remain limited and will decrease through time because Elia is focusing on forecast quality since a long time already and each year the processes and tools are getting more robust. As such Elia can anticipate lower bug rates in the future.
    - **Effort high**: A live tool performing many automated checks and actions should be designed. On the other hand, such tool would benefit to all existing forecasts potentially.

➔ **In the roadmap: "Live monitoring" for aggregated forecasts will be implemented in 2025 when the ML/AI modeling tool will be integrated into operations.**

RESULTING FORECAST

- **Choose supplier in function of wind speed:** in order to keep IRM for storm situations (when wind speed is > 20m/s for example) and use VITEC otherwise.
    - **Impact High:** The expected outcome is high, Elia already tried to push IRM to use ML and other possible improvements but their tool is not easily configurable/editable.
    - **Effort Low:** VITEC could do the combination in their tool upfront.

➔ **In the roadmap: IT developments are ongoing, go-live expected in 2023.**

5.2 PV Forecast

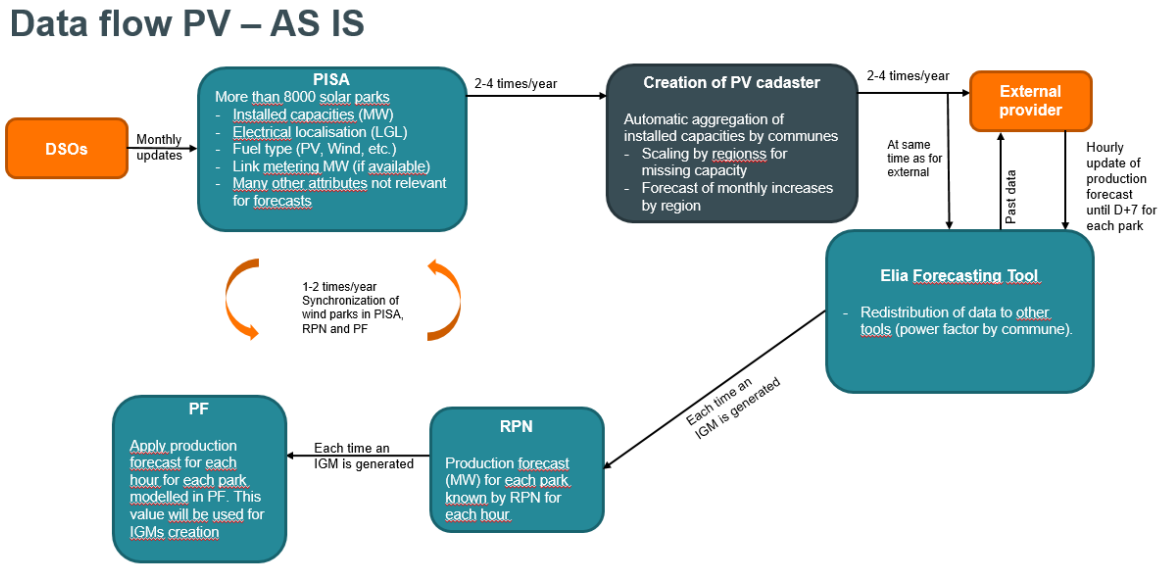
- In January 2022, 8800 individual[7] units amounting for 4690 MWp[8] were listed in PISA tool while the estimated power was more around 5900MWp. Even if the figure 10 shows that Elia has many very small units in PISA database, the total capacity in PISA is far behind the expected reality. Elia makes use of 500 virtual power

[7] Those units are often an aggregation of even smaller PV installations. In reality, there are hundreds of thousands of individual units on private houses but in PISA all those are aggregated.

[8] Elia uses MWpeak (MWp) for the time being because there are no indications for the time being that inverters would limit power outputs in Belgium.

plants spread over the grid to reach the probable total installed capacity (see below 5.2.1). The location and capacity of the virtual units are determined by Elia by looking at the estimation of the remaining growth potential of such small PV units on the Belgian territory.

- All those units are connected on lowest voltage levels available in our model (low influence on flow forecast on Elia's grid) and the DA RMSE for 2021 is ~4% (if computed only for daylight hours) and ~1% (if computed for all hours).
- Elia buys state of the art estimated measures and forecasts from an external provider.



There seems to be little room for improvement but considering the upcoming increasing of PV units Elia should ensure the same level of accuracy in the coming years.

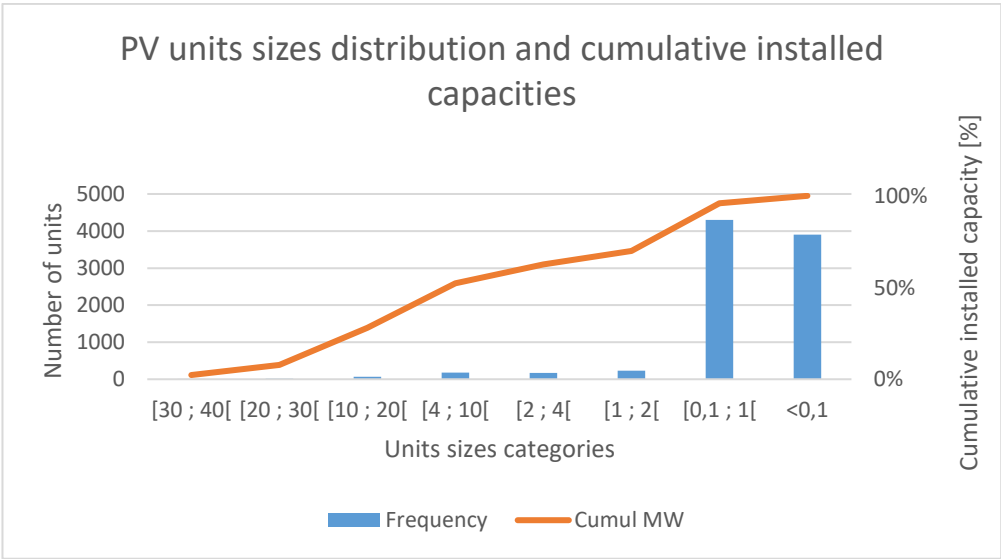


Figure 10: Installed capacities of PV units in PISA

5.2.1 AS IS

INFEED DATA

- **Cadaster management**: Installed capacities at communal-level (~400 communes) + filter TSO/DSO-connected[9] are required[10]. Cadaster is generated by PISA on demand. Based on information of authorities[11] Elia can estimate missing capacities and scale existing units to match with a regional total target. Estimation of monthly capacity increases by region is then applied to anticipate monthly evolution of the cadaster. This is predefined in the cadaster file sent to the external provider and EFTool meaning that updates will be applied automatically at the predefined dates (no need for cumbersome coordination with external provider and our own tool). Twice a year Elia updates the cadaster for the 12 months in the future based on an extract from PISA + reevaluation of best estimates from the authorities. Hereunder some numbers for January 2022 are listed as an example (**gap of 1200MW = 20%** between PISA and best estimate).

	PISA (JAN - 2022)	ASIS JAN 2021	Best estimate JAN 2022	Best estimate JAN 2023
Flanders	3598	3470	4300	4600
Wallonia	980	1263	1500	1700
Brussels	110	55	130	140
TOTAL	4689	4788	5930	6440

- **Historical values**: are themselves a forecast based on a subset of measured installations. This is provided by an external provider[12]. The upscaling is done based on 80k measurement points amounting for more than 1GW of installed capacities. They are covering all regions in Belgium and all installations' sizes. On paper this is the best Elia can hope for as methodology to recreate the estimated produced energy. The only downside is the black-box effect and our incapacity to challenge the received values.

[9] TSO/DSO information does not improve in any ways the quality of the flows forecast, relevant for publication & settlement purposes.

[10] In 2021 it was still necessary to distinguish <10kVA units from the rest (need of Synergrid). This is finished and make it easier for Elia to upgrade cadaster. We can now use PISA as main source.

[11] **Carte dynamique (solaire et éolien) de la Wallonie - Site énergie du Service public de Wallonie** and **Cijfers energiekaart - Energiesparen**

[12] Energy Meteo and Services

FORECASTING MODEL

- **External provider**[13], selected among 6 pre-selected candidates on a live trial run of 3 months. RMSE is typically used to evaluate forecasts performances in such selection. Elia shares the real time data (in this case from same provider) and the cadaster file as inputs. The external provider buys weather data near the installed units (in this case for all units in Belgium). Usually such provider buys those weather data from several providers/weather models and launch many forecasting methods in parallel with different approaches. Then the provider combines all those forecasts in order to minimize RMSE (with a special attention to DA forecast).

RESULTING FORECAST

- Same as for other decentralized productions (see hereunder) except that Elia can easily upscale the installed capacities of some 500 virtual units in RPN and PF in order to match with total target of installed capacities. Those virtual units are spread in function of the potential of future installations (Elia can assume a good hypothesis). This way IGMs are following correctly the total capacities with a relatively good repartition even if a full synchronization of PISA/RPN/PF is not performed.

5.2.2 Possible TO BE

INFEED DATA

- **Reducing the gap from PISA** by identifying the DSOs not providing enough of the units lower than 400kW and simplifying the data exchange procedure for the DSOs.
    - **Impact low:** current upscaling method is not shocking for such types of units but it would be valuable to have at least no medium/big units missing in our models because those generate clear local disturbances. Elia can anticipate that results will not be perfect considering the fact that this level of information is probably not always known by the DSOs themselves.
    - **Effort high:** Advocating for DSOs to increase installed capacity coverage is not really on Elia's hand but it will not be a big effort for Elia to identify the DSOs not providing enough of the units lower than 400kW and improve the collaboration with them. Changing the data format exchange and simplifying is challenging though. Indeed this format is used and industrialized by each DSOs so it is never simple to make modifications. This should certainly be possible considering the fact that the regional authorities do have more information than what PISA contains.

→ Partially in the roadmap: "industrialization of quality checks in PISA" will allow Elia to better track contents quality and exhaustively but it is not foreseen to change the data exchange format or procedure.

[13] Energy Meteo and Services

- **Smart meter data:** could increase coverage of measured installations to improve estimation of realized production.
    - **Impact low:** current coverage is already great thanks to the 80000 measurement points from the external provider.
    - **Effort high:** smart meter data have not always information on the PV production isolated from the auto-consumption.

→ Not in the roadmap: current way of working is already very advanced, smart meters will not improve quality or even reduce costs.

FORECASTING MODEL

- **Combining several providers**: is something Elia already tried without success. Nowadays external providers are already doing this upfront. They buy many independent weather data from different weather models and they use many forecasting algorithms in parallel. They combine all their different predictions with ML and AI in order to optimize RMSE (or other parameters on demand).
    - **Impact very low:** give a backup in case of issue but providers have high reliability standards (by selection).
    - **Effort high**: at least doubling the costs for buying the data + IT/Process effort for combination.

→ Not in the roadmap: we will continue with regular tenders to make sure we have a performant provider offering a good quality/price ratio.

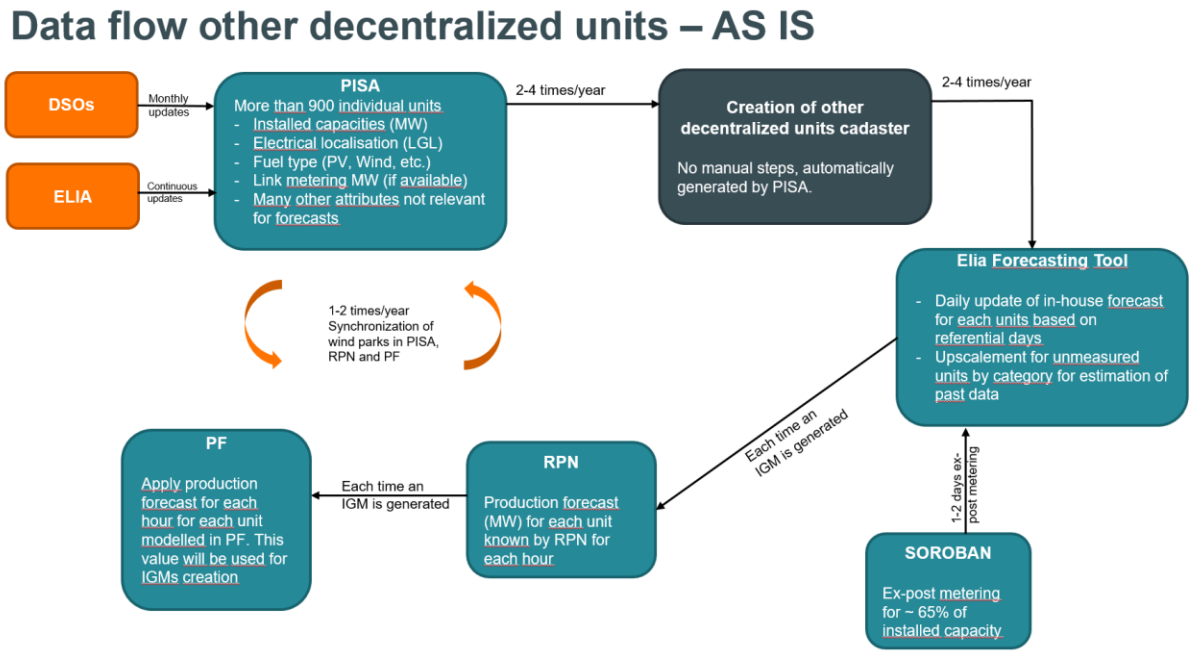
RESULTING FORECAST

- See other decentralized productions

5.3 Other decentralized productions (non-wind, non-PV)

- 2152MW installed capacity by 2022 in the PISA tool. As always units smaller than 400kW are shared on a best effort basis by the DSOs but we see in figure 11 that PISA contains still many very small units.
- 900 individual units (each of them modeled in Power Factory) and 1750MW of CHPs.
- This category is slowly but continuously increasing, so it becomes more and more interesting to invest efforts to improve this forecast. This forecasting model has not been changed since many years.
- Battery storages are progressively entering into this category, Elia needs to anticipate how to best forecast them by looking at behavior the few batteries recently connected to the grid. Elia lacks of data to pinpoint the best forecasting method already but keeping an eye on those newcomers will enable Elia to anticipate the probable massive penetration of such technologies in the coming years.
- RMSE ~5% for the aggregated category, for the units with a power output larger than 5 MW (the ones with the strongest impact on grid congestion) the average individual RMSE is 15%. Considering the size and numbers of such units, it seems that building individual models for the most problematic units would make sense. In parallel, Elia could improve the current methodology for the other units. Of course, nothing

prevent a very small units to be at the source of a congestion issue, especially in the lowest grid voltage levels. In such cases individual forecasts could still be relevant, it will only be a matter of trade-off between complexity of the solution and measurable impact.



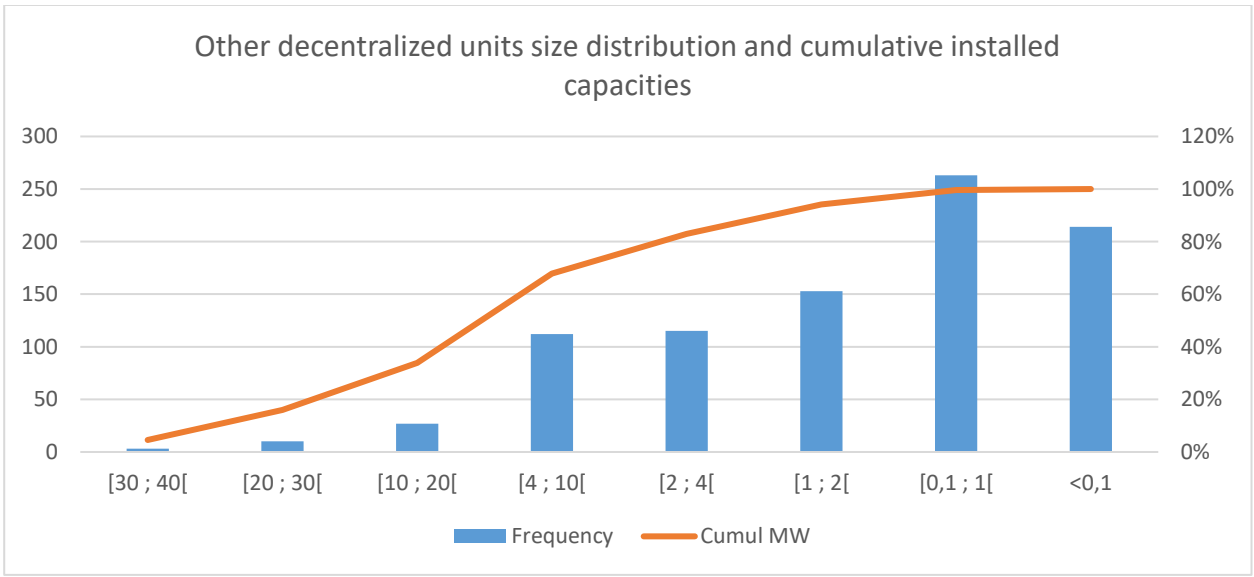


Figure 11: Installed capacities of other decentralized (non-wind, non-PV) units in Elia grid model

5.3.1 AS IS

INFEED DATA

- **Cadaster:** installed capacities + production profile + electrical location + Mnemonic TIC (metering point) are necessary and are updated typically once a year based on PISA reference. PISA contains a very good reference for those type of units; Elia assumes it is almost complete.
- ➔ **NEW:** this has been improved since early 2022 with updates possible at any frequency thanks to IT projects and process optimization. One button to generate cadaster file by filtering "InService", not PV, not Wind, type A or B. The idea is to trigger an update each time there is a delta of 50 MW to avoid unnecessary workload.
- **Historical values for each individual unit:** ~65% of total installed capacity is effectively metered. Reminder: metering is not a real time measurement; the data is accessible in our systems with typically 1-day lag for non-validated data and 1-month lag for validated data. The metering data for such units comes from the DSOs on request of Elia[14].
- **Production profile**: the historical measurement estimations of unmetered units are an upscaling based on the power factor of all metered units from the same production type. Here the production profile currently in use: batteries, CHP large, CHP small, Run of river, incinerators.

$$UM_{unit} = \frac{\sum_{Profile} P_{meas}}{\sum_{Profile} P_{installed}} * P_{inst,unit}$$

FORECASTING MODEL

- **Internal tool** with basic algorithm. The tool looks at the **3 most recent days** with metered data available from the same calendar, category and for each quarter of an hour to be forecasted it computes the average of the power factor for each profile and then apply it to each unit.

$$Pforecast_{unit,QH} = \frac{\sum_{For\ each\ refday} \sum_{Machines\ with\ same\ profile} Pmeas_{QH}}{\sum_{For\ each\ refday} \sum_{Machines\ with\ same\ profile} Pinst_{QH}} * Pinst_{unit,QH}$$

- **Update** is done once at 6h24 for values extending from "Now" to day + 7.
- **Timers** for data imports of the most recent metered data (metering data are not real time, they are mostly available 1 or 2 days ex-post).
- **Calendar days** (categories used: Weekdays, Saturdays and Sundays).

[14] **Standard Metering Services (elia.be)**

RESULTING DATA

- **Availability/centralization**: forecasted values, per quarter of an hour, from real time to D+7, per individual units are made available for all operational tools.
- **RPN imports latest forecasted data** from EFTool and sends them to PF at the exact electrical node each time an IGM is created.
- RPN & PF production units[15] cadasters are **synchronized with PISA** once every year. There are many IT issues for this process, this is well-known and efforts are being done to smooth this process in the future. In the meantime, when updating the cadaster in EFTool, new units are not pushed to IGMs until next synchronization with PISA/RPN/PF. Nonetheless, it remains pertinent to regularly update EFTool cadaster for other processes.

5.3.2 Possible TO BE

INFEED DATA

- **Directly link ("1-to-1") between meteringand PISA** units to avoid double counting (also true for onshore winds). This is under discussion with metering services to evaluate feasibility.
    - **Impact medium:** not so many units concerned and most of them are manually corrected.
    - **Effort low:** depending on metering services possibilities, alternative is to add and maintain a new field in PISA with weighting factors.

→ In the roadmap, "increase metering ratio" will notably ensure that the metering granularity is identical to the PISA granularity.

- **Add missing metering** for the biggest units (e.g. add TIC for 30 biggest missing units -> coverage goes > 90% having a TIC) + ensure that existing TIC are functioning correctly + ensure that metering services are systematically requesting metering data for all units big enough.
    - **Impact medium:** already 65% coverage but biggest units have sensible effect on the grid because concentrated in one electrical location).
    - **Effort low/medium:** depending were we put the threshold.

→ In the roadmap, "increase metering ratio" will make sure that all units with more than 5MW will have a metering.

[15] All units are concerned (PV, Wind, other RES, centralized)

- **Using real time data** (measured or estimated by the state estimator) instead of metering data to make the forecast.
    - **Impact medium**: will improve the short-term forecast but also allows hourly updates of all time-horizons forecasts.
    - **Effort medium**: thanks to Load and Generation project, PF and EMS granularity for decentralized units will be aligned. Moreover, the project "communication type B" will ensure that more real-time measurements are configured in EMS. Consequently, real-time data (either from a measurement or from an estimation of the state estimator) will be available.

→ In the roadmap, "real-time measurements as input" will do exactly that by end 2025 when the Load and Generation project will be live.

- **Improve production type categories:** CHP baseload vs CHP stopping during weekends and nights, units usually not producing (typically used as manual Frequency Restoration Reserve, mFRR), …
    - **Impact low**: only valuable for unmetered units (35% today but will hopefully reduce in the future).
    - **Effort medium**: Need to adapt PISA accordingly (no IT development) and verify that all users are fine with this change or apply the categorization-logic into EFTool (IT development)

→ Not in the roadmap: Elia esteems that improving the measurement ratios will cover this issue.

- **Schedules of B type units:** this will be available by the implementation of iCAROS phase 2.
    - **Impact low**: quality of those nominations still to be demonstrated and can be expected to be relatively poor considering their number and sizes. If data is of good quality then it is certainly a valuable input data because OFF status are practically impossible to forecast in most cases.
    - **Effort medium**: no IT structure today to feed model with this input.

→ In the roadmap as an attention point, "Schedules small units" make sure that we will investigate it in function of iCAROS project progress.

- **Add market data:** typically the spot price of DA/ID market could be relevant for a better prediction of such units' behavior. If prices are very high, some might produce more than usual and vice versa. A quick view on the graph below shows the influence of several factors on the production of such units. One can see that the price is a strong driver but there is a multitude of other parameters. A more detailed analysis is required in order to be able to improve the forecast.
    - **Impact medium:** quite marginal effect today but probably progressively bigger in the future.
    - **Effort low**: quite a basic data easily accessible.

→ In the roadmap by end of 2024: "IT integration" of AI/ML model into operations will for sure be fed with a lot of input data, notably market data.

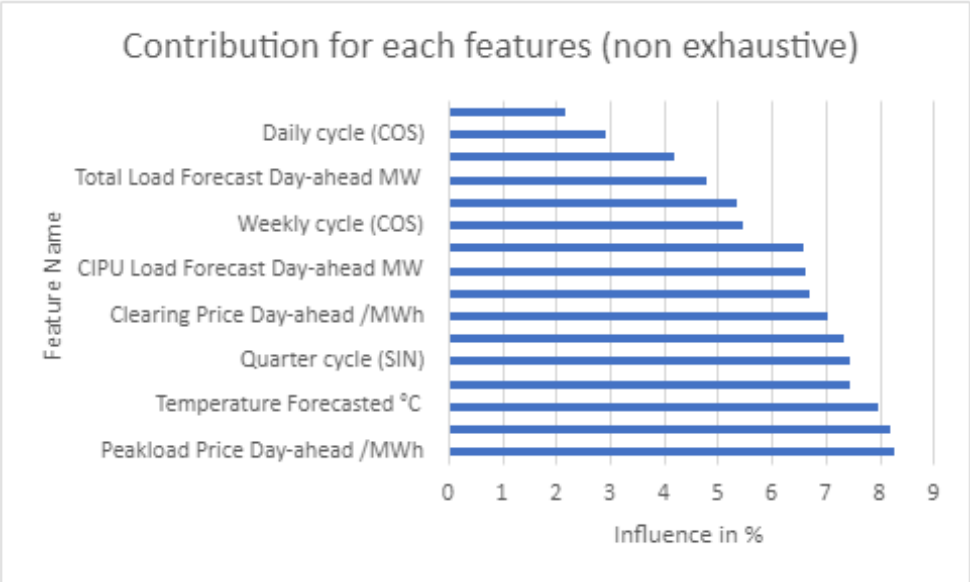


Figure 12: Identified drivers for forecasting other decentralized (non-wind, non-PV) units production levels

FORECASTING MODEL

- **Switch from a forecast per profile to an individual forecast.** The tool linearizes the individual forecasts based on their profile, which makes little sense. The tool could simply apply the same logic at an individual unit level :
    - Pforecast_unit,QH = sum_for each ref day (unit Pmeas_QH)/number of ref days
    - **Impact high**: this would greatly improve the forecast at nodal level because for sure all units within any given profile do not behave the same.
    - **Effort low:** keeping the same tools and main logic, just adapting some formulas.

→ Already done: actually, when Elia decided to implement this as soon as possible, it discovered that it was already applied. It was a mistake in our description here above.

- Look for **external** partners
    - **Impact high:** for sure external providers will be able to meet our need and improve significantly current model.
    - **Effort high:** high costs w.r.t. todays' + high workload to organize tender and selection + lot of changes in current processes. Considering the fact that Elia has all the main drivers in-house it would make little sense to look for external providers like Elia typically does when complex weather data are required.

→ Not in the roadmap: Elia chooses to implement in-house ML/AI models in operations because this will benefit many other forecasts types.

- Adding **ML/AI** modules to existing logic
    - **Impact high:** this will most probably correct many issues notably from the inputs perspectives.
    - **Effort high**: not feasible in current tool.

→ Not in the roadmap: Elia prefers to have a dedicated tool usable for the entire company instead of doing it for a single forecast type.

- Add **market data** to existing logic: correction of forecast based on spot prices in the past and next day.
    - **Impact low**: probably not a good way to apply the same correction per production profile, this is more specific to each unit.
    - **Effort high**: current tool is basic and has no built-in forecasting modules.

→ Not in the roadmap: Considering the fact that by end of 2024 Elia will have ML/AI models in operation including market data, it make little sense to add this features in EFTool in the meantime (sunk costs).

- Add historical values and nominations of **pump-turbine units and big batteries (scheduled ones)**. They could be good drivers/proxies for small batteries production forecast but this is for the time being impossible to study by lack of data. A mitigation measure could be to foresee the possibility in the tools and processes to feed on those inputs when batteries penetration will be higher.

→ Partially in the roadmap: "accumulate XP and knowledge on batteries' behavior" will make sure that Elia is proactively looking for best ways to anticipate how such assets will behave. By end of 2024, the ML/AI modeling tool will for sur be a very good start to simulate them.

- Building **individual forecast** with off-the-shelf forecasting modules and some basic inputs (market prices, total load, historical data) and stop with current referential days basic approach. Hybrid situation could be also imagined were only biggest units are individually modelled (e.g. 40 biggest units = 35% of installed capacity) while the rest would still be forecasted with current method (with still other identified improvements).
    - **Impact medium:** this will most probably correct many issues notably from the inputs perspectives. Concrete numbers are not yet known, certainly not if other identified possible improvement of the current logic are also taken into account. However, our findings on load forecasts showed already good results while such units are strongly linked to load behaviors. A first evaluation of the possible improvement was done during the summer of 2022. A summer school challenge done at the KUL, with the cooperation of Elia, with anonymized production data (for mainly small and medium cogeneration units) has shown a reduction of 30% of the RMSE on the day-ahead forecast on individual loads.
    - **Effort high**: Internal tool is more a data handling tool than a forecasting tool and it would be probably necessary to make this module on a separate tool. Of course such a tool would be beneficial for other forecasts as well.

→ **In the roadmap: "IT integration" of AI/ML models in operations will do exactly that by end 2024. Increase updates rate:** ideally hourly updates but only if this makes sense (in this case also needed to have input data on hourly frequency).

- **Impact very low**: TIC values are not updated more than once a day. Only relevant if model uses other inputs with smaller time granularity (real time data, market data, etc.).
- **Effort low:** just a tool configuration.

→ **Partially in the roadmap, implementation of ML/AI models for the largest units of this category is foreseen by 2025 and will have an hourly update rate. Increasing the update rate before that is not relevant because TIC values are imported once a day anyways.**

- **Change the weighting factors of the 3 last reference days rule** (weighted average with more weights on more recent days, using only reference day, …)
  - **Impact low**: still to be assessed, but impact expected to be low.
  - **Effort low:** minor changes of the code.

→ **Not in the roadmap: current logic makes sense and Elia could not find better weighting factors for this formula. The fact that the calendars days will be enlarged (see below) will make the change of the weighting factors even less relevant.**

- **Fine-tune timers** in order to have more recent data available (today 2 days delay) even if not all metered data are yet available.
  - **Impact low:** current times = 2-days delay between availability of metered data vs. real time while lot of data available in day+1.
  - **Effort low:** just changing configuration of current tool (no IT development) but some analyses to be done to find optimum solution.

→ **In the roadmap, "upgrade existing logic" will change the timers to day+1 instead of day+2 in 2023.**

- **Improve calendar days categories**: bridge days = Saturdays, Bank holidays = Sundays, 1st of January = special day, Christmas holidays = special holidays, …
  - **Impact medium**: current logic is very basic, e.g. 1st of Jan = normal day while very special day in practice. In this case improvement only valid for those special days (~10% of the years).
  - **Effort medium**: small analysis to be done but probably very similar to current load categorization (CHPs are strongly linked to industries/loads). Then some IT developments (minor code changes) to be foreseen.

→ **In the roadmap: "upgrade existing logic" will improve in 2023 the available categories.**

RESULTING FORECAST

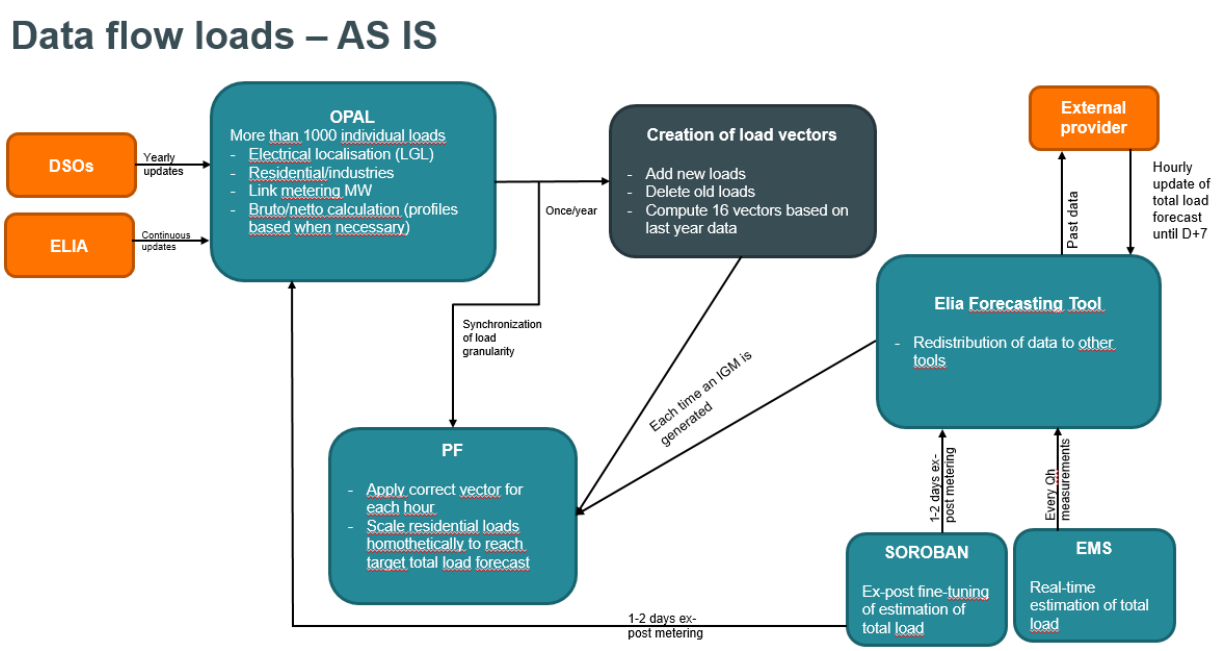
- Smooth process of **synchronizing PISA with RPN and PF** and increase frequency of synchronizations.
  - **Impact high:** structural errors due to this lack of synchronism.

- **Effort high**: IT challenge is not to be underestimated, notably because the long-term planning department is using the same tool and models but for all future horizons and such update should be validated for all the future years too.

→ In the roadmap: "Automating and smoothing of process" is already ongoing and will allow Elia to perform 2 or 3 full synchronization a year by end of 2023. By end of 2025, on demand synchronizations should be possible thanks to the alignment of load and generation granularity.

5.4 Load Forecast

As explained in section 4, DSOs are feeding every year OPAL tool with information of their present loads as well as future evolutions. Elia maintains itself in OPAL the loads connected to its grid. OPAL contains about 1000 individual loads with their necessary attributes (residential/industrial, electrical localization (LGL), historical data. Currently these 1000 loads are mostly defined by the following rule: an individual load is foreseen for each direct client (by LGL) and for each interface point with a DSO (by LGL). The OPAL tool contains also the installed capacities of all units located at the same LGLs. This allows the tool to compute the gross and net loads. Indeed, the available metering data are rarely "pure" loads, most of the time the signal is polluted by decentralized production (typically the small PVs).



As shown on the scheme above, there are two important levels:

- the **aggregated** total load of Belgium and
- how to ventilate this total load into about **1000 individual loads** at lowest possible nodal level.

10GW total load on Christmas Eve is most probably different from 10GW on Monday morning in April. The contributions of each individual load are not the same. Currently Elia is using 16 different load repartition keys to capture the main behaviors at play. A recent study with an external data scientist confirmed that those 16 vectors are capturing most of the behaviors and constitute a good compromise between accuracy and operability for all 1000 loads for a top-down approach. The consultant could not find a better way to spread the total load by creating other repartition keys. Of course, it does not mean that any improvement is not possible as explain below, it just shows that Elia reached the limit of the potential of such approach. The 16 vectors are the result of the following categorization: day/night, week-day/weekend and the 4 seasons (2 x 2 x 4 = 16 vectors).

Today the total load is relatively easy to predict (RMSE ~3% in 2021) based on some simple inputs and the impact on congestions of the total load forecast errors is low due to the diffusion of this error into all individual residential loads (~6GW residential load on yearly average). Indeed, the selected repartition key fixes industrial loads while the residential loads are scaled to meet the total load target during IGMs creation.
However, in the upcoming years, Elia expects important changes in the load contributors: electric vehicles penetration, the increase in heat pumps, the load flexibility development and adaptation of consumption based on market prices, ... All those factors will probably deteriorate the RMSE of the total load forecast if the same methodology is kept while at the same time the total load will only increase in the future, increasing its impact on congestions altogether. Nonetheless, by the lack of data, those future evolutions are hard to anticipate accurately. Elia preconize to keep an eye on those evolutions in upcoming tenders for external forecast providers and to make sure that our tools and processes are ready when needed.

Regarding the individual load forecast Elia calculated that RMSEs are on average around 15% for residential loads and around 17% for large industries (>40MW). Applying individual forecasts for the highest loads seems promising w.r.t. influence on congestion management. One of the findings is that for industries it would only make sense if Elia has access to very recent past data (day-ahead typically) to feed the model (ML 2 in the table below). For residential loads Elia sees that this prerequisite is less important. Results also indicate that schedules received for large industries still outperform the ML 2. For these industries, it would make perfect sense to receive and use the schedules, having a demonstrated impact on congestions. All the results displayed in the Figure 13 for large industries are calculated for a sample of about 20 industries for which Elia could map the DA schedules information (Offtakes) currently available with the Power Factory granularity. This sample represents about 1,5GW of load (i.e. 30-40% of industrial loads). Note that even if those DA offtakes nominations are well available every day in Elia's tools, Elia does not apply them in its IGMs (see below 5.4.2).

RMSE [%]	Large ind	Residential
AS IS	16,9	15,1
ML 1	16,4	12,6
ML 2	13,4	12,1
Offtakes	12,1	/

Figure 13: RMSEs for several forecasting methods

5.4.1 AS IS

INFEED DATA

- **Historical data** of total load are computed by Elia and shared with the external provider via a live stream.
- **Equivalent temperature** is computed by the external provider based on some simple weather data.
- **Calendar days**: some special events are highlighted and categorized by the external provider in order to best cope with special days (bridge days, holidays etc.).
- **16 repartition keys** calculated once a year based on last year metering data. The average values of a given individual load for the selected timestamp in the past is computed and used in the repartition key. For some special industries a worst case load is chosen to best represent and anticipate the potential congestions (typically for electric oven when the average would be much lower than their peaks while they typically reach their peaks on a daily basis).

FORECASTING MODEL

- Elia buys the total load forecast from an external provider, the trial phase allows to make a good selection based on performances and prices while the 16 repartition keys are well performing for a top-down approach (i.e. forecast of total load then ventilation among individual loads). Consequently, there is little room for improvement if we do not change the paradigm altogether.

RESULTING FORECAST

- Industrial loads are fixed by the selected repartition key while the residential loads are scaled to meet the total load forecast.

5.4.2 Possible TO BE

INFEED DATA

- **Market prices** will certainly become a crucial driver for total load prediction even if today no such correlation exists.
    - **Impact unknown** (lack of data)**.**
    - **Effort low:** during next selection of external provider Elia will impose that such input is taken into consideration.

→ In the roadmap, "Tender total load forecast" foreseen in 2024 will help Elia to assess need for such evolutions as well as to gain experience from experts in the field.

- **Recent past data** can clearly help individual load models when unpredictable behavior on loads occurs. This could be demonstrated with ML2 model on industrial forecast accuracy.

- **Impact medium/high:** depends if Elia has access to latest QH data or if Elia has only access to the data with some time lag.
- **Effort medium/high:** to have data from the latest QH requires a perfect alignment between loads definitions in PF and EMS models. This is foreseen in the Load and Generation project. Having the data with some time lag would still require many adaptations in our tools and processes.

→ In the roadmap, "real-time data as input" foresees it by end of 2025.

- **Historical data of different categories** as electric vehicles, heat pumps, residential batteries etc. It could improve the total load forecast. Specific models could separately forecast each of those categories and then combined by a total load forecast model.
  - **Impact unknown** (lack of data).
  - **Effort high:** this represent many data while the latter are not directly connected to Elia grid.

→ Not in the roadmap but Elia will certainly keep an eye on available data and relevancy for forecast purposes. The tender on total load forecast might help Elia to better understand the importance of such data.

- **Schedules of biggest industries** are certainly helpful when some unpredictable behaviors are at stake as typically revision periods, change of typical consumption patterns due to supply chain issues or other externalities. The ML/AI models Elia has tested could not beat most of the schedules received for industries with a load superior to 40MW. This indicates that for the biggest industries, those schedules could be either used as final forecasts or eventually used as input parameter for a ML/AI model.
  - **Impact high**: Some industries have already load levels of several hundreds of MWs, with the progressive electrification of more industrial processes such loads will only increase in number.
  - **Effort high**: requesting such schedules represents a non-negligible effort for the asset owners and should be set in place only in case if a clear link exists between congested elements and the set point of those industries. Such selection would make sure we only ask schedules for the biggest industries (the bigger it is the greater its influence on flows) located in congested areas.

→ In the roadmap, "Large ind. load schedules" will implement existing DA schedules (those on the right granularity and with high quality) on a short-term view while "Using DA schedules" will enlarge this concept to all industrial loads via implementation of iCAROS 2.2 in 2027[16].

FORECASTING MODEL

- Use **machine learning to forecast the loads individually**. With such approach, RMSE at nodal level can be reduced for both industrial and residential loads as shown in figure 13.

[16] For more information on nomination: **Public consultation of the study on the evolution of the BRP nominations (elia.be)**

- **Impact medium/high:** For the biggest loads, the impact will be tangible already today. More importantly, the new types of loads (e.g. electric vehicles) will introduce more exotic behaviors in the future while those are typically well detected by machine learning (there is no need for complex modelling if this is confirmed). With the electrification of industries, mitigating the error on such forecast might be more and more valuable.
- **Effort Medium:** The main challenge is to integrate this in the existing IT infrastructure. For industrial loads, it seems that live stream of recent data is key to really improve the forecast quality. This requires much more efforts in terms of model alignments and data acquisitions.

→ In the roadmap, "IT integration" of ML/AI model in operation will allow a go live in 2025.

- **Launch tender** to find a new provider and explore the possibility to do an **in-house** load forecast with AI Center of Excellence and 50Hz.
  - **Impact Medium/High:** Elia might see an improvement in the performances but Elia could also ask the provider to offer new services such as categorization of the loads in order to prepare the future (increase in electric car, heat pumps, load elasticity with respect to prices, etc.)
  - **Effort Medium:** Enlarging the scope of the forecast requires more data while those are not directly available at Elia for the time being.

→ Partially in the roadmap, "Tender total load" is foreseen in 2024. Elia hopes that by then some detailed data on subcategories of the load components will be accessible for the assessment.

- **Use categorization** in order to improve the forecast (Proof of Concept with an external data scientist). This approach is another way to tackle the problem of different types of loads compared to the machine learning approach. Forecasting each load individually with the additional information that they belong to the same cluster (i.e. they show similar patterns), may sometimes further improve the final individual forecast.
  - **Impact medium:** It is hard to quantify the gain for such a method before doing it but with the development of new usage, this might be more important in the future.
  - **Effort medium/high:** The effort seems harder than for ML because it requires a more in-depth analysis in order to categorize the usages (less automatic than ML).

→ Already done as first step from the "studies & PoCs smart IGMs". It showed that our current categorization is already close to optimum. The only way forward seems to be forecasts at individual level.

RESULTING FORECAST

- **Combining total load forecast** with the total load forecast reconstructed with all the day-ahead & intraday market positions. A machine learning algorithm could be used, in addition with other infeed parameters to best combine both time-series.
  - **Impact low**: if total load forecast is well optimized upfront, there is a fair chance that the total load based on the market positions will not contain much additional information.
  - **Effort low**: an off-the-shelf ML/AI model could set this up.

→ In the roadmap, "In-house total load forecast" will be set up as soon as the integration of AI/ML in operations is realized.

5.5 Conventional units

Nominations are directly used for DACF and IDCF files (no forecasting model). The quality of those nominations has been high in the past but this should be closely monitored in the framework of the progressive balancing obligation relaxation and higher uncertainty on DA/ID horizon for BRPs (intermittent prod etc.) meaning that decision making are pushed closer and closer to real time.

5.5.1 Possible TO BE

FORECASTING MODEL

- **Forecast of intraday market moves** based on open position of Belgium, DA market prices and offer & demand curves, nominated Belgian redispatching bids (volumes and prices). A proof of concept has been set-up in 2021 to evaluate feasibility and performances.
    - **Impact Low:** Nowadays nominated volumes remain most of the time very accurate. For some extreme situations the proof of concept could anticipate some moves correctly but changing the official nominated power cannot easily happen in an official IGM (net position is imposed by the merging process).
    - **Effort High:** many new inputs and scripts are required to make it work in daily operations but feasibility has been proven. If impact is starting to grow due to increasing uncertainty in day-ahead horizon, Elia can always reopen this initiative.

→ Not in the roadmap, might be re-evaluated as soon as open positions would become large enough to generate errors in the congestions identification. Elia will for sure continue to monitor the open position of Belgium.

5.6 Net position

Net positions resulting from nominated power exchanges between European countries is used as a fixed parameter for DACF/IDCF files. For D2CF files the net position from a reference day is applied but this will evolve with the use of the CORE centralized net positions forecasts in 2023. This forecast already exists but its results arrive too late to be introduced into D2CFs.Elia will apply the resulting net positions of Belgium directly into its D2CFs as soon as the results are available earlier (around 3pm). RCCs are managing this forecast and they foresees to deliver daily forecasts around 3 pm by 2023.

5.6.1 Possible TO BE

FORECASTING MODEL

- **Net Position correction** for DA process based on DA open positions and on latest market prices curves from all European countries could be done at RCC level. Some most probable exchanges could be anticipated in order to reduce the open positions and as a result improve the net positions imposed by the CGMs process.
    - **Impact medium:** nowadays, open positions of Belgium remains acceptable but this is not always the case for neighboring countries. Moreover, the situation in Belgium could worsen with time considering the fact that there are no obligations for market player to provide balanced schedules.
    - **Effort high**: this should be done at RCC level with a close collaboration of as many as possible TSOs. A proof of concept could be done to assess feasibility and gains.

➔ **In the roadmap, "Open positions influence on net positions" foresees that Elia will raise awareness on the issue and push the proposed solution on CSO (Coordinated Security Operations) meetings.**

5.7 Grid topology

Outages (planned and unplanned) are all included in the IGMs of Elia as well as their necessary preventive topological actions. PSTs taps are mainly at neutral taps for D2CF (actually 15/15/15/15 to anticipate structural loopflows) but this is of course only an initial state and optimized PSTs taps position will be taken to increase Flow-based domain in estimated market direction.

For the initial version of DACF, PSTs are already optimized by looking at the market flows ("inputs") resulting from FB MC. Of course this is a rough estimation ("model") considering the fact that the loopflows are not yet known at this stage. During the DACF process iterations, PSTs will be adapted to relieve eventual congestions on Elia's grid or even abroad.

5.8 Nemo Forecast

The forecast on the Nemo flow is only relevant for D-2 time horizon because later on there are deterministic nominations. Nemo forecast comes from an external provider and will be replaced by the CORE centralized net position forecast in 2023. The new model is outperforming the previous one mostly because it is based on a broader range of infeed data. This forecast is not anymore into the hands of Elia and it seems that there is not much room for improvement.

5.9 Dynamic Line Rating

Dynamic Line Rating (DLR) allows adapting the rating of equipped lines based on the local weather conditions (i.e. cooling of the line). On average, this means an increase of capacity w.r.t. the more conservative static line rating. Of course, such technology never ensures that the capacity will be increased when needed. Using DLR is actually not improving the congestion management since it adds a new uncertainty into the system. Increasing ratings reduces the probability of congestions occurrences as well as their impact because most of the time at equivalent risk the dynamic ratings increases when one approaches real-time.

5.9.1 AS IS

INFEED DATA

- **Real time measurements:** About 30 lines are currently equipped with Ampacimon modules and a dozen underground cables have the real time thermal rating installed. This represents all the typically congested lines for which the bus bay elements are not limiting the DLR.
- **Lines/bottlenecks selection**: selection of element to be equipped, is based on a techno-economic study. Estimated avoided costs and installation costs are the main drivers but Elia also looks at the benefit of such investments in terms of risk mitigation. Typical candidates are lines monitored by the market coupling. Long infrastructure works, typically for High Temperature Low Sag (HTLS) conductors reinforcements, are also good candidates during the outages period. The long-term planning department also identifies future candidates many years in advance based on their security analyses results.
- **Temperature forecasts**: are used to anticipate which reference rating should be applied for each IGM. Indeed Elia changes the reference ratings of its grid based on the measured temperature increasing in average the available ratings w.r.t. the previous static rating approach (same reference rating applied by fixed predefined periods).

FORECASTING MODEL

- **Ampacimon forecast:** uses the past measured ratings, the past local weather measures and the past global weather predictions to build a predictive model. The provider trains the model on a fixed 3 months period manually selected. Only 14 elements, those being monitored in the Flow-based, have such forecast due to economic optimization. For the remaining lines without this forecast, it is up to the operators to do it by hand based on recent data and weather predictions. Consequently, a margin needs to be applied in order to ensure that the used rating will not be too optimistic w.r.t. the rating effectively available in real time.

RESULTING FORECAST

- **Capping rules:** in order to ensure a 99% reliability of the forecasted values in the IGMs, forecast results provided by Ampacimon are capped to 105% of the applied reference rating, before putting it automatically in the IGMs. This limits the potential of the DLR, but is mandatory due to the weak reliability of the current results.

5.9.2 Possible TO BE

INFEED DATA

- **Improving the long-term detection criteria's:** DLR never ensures that capacity will be increased when needed except when the congestion is clearly linked to high wind production. It is then rarely possible to consider such technology to release an identified bottleneck, it is only a good (but costly) mitigation measure for temporary situations. Consequently, DLR installation adds costs without ensuring that the congestions will be released meaning that most of the installations of DLR are validated quite late with the consequence that

sometimes bus bays must first be reinforced delaying the installation by many years. It could be more optimal to set up a process identifying possible candidates far in advance and making sure that their bays will not be limiting when it is estimated that there is a risk of congestion on this element. By doing this Elia will make sure that if the bottleneck materialize itself, Elia will be able to quickly install DLR (could be done in 6 months).

  - **Impact high**: this could be very helpful to anticipate the Ampacimon installation, certainly in the case of limiting bus bays. Changing bus bays elements may take years and should be planned coherently in the infrastructure 4 years-ahead plan.
  - **Effort medium**: Ideally, long-term planning could generate a list of most loaded expected elements for each year in the future. Those would be natural candidates for DLR and at least evaluation of need and feasibility could be launched well ahead of time.

→ Already done, grid development is already proposing many installations several years in advance. It has become a standard mitigation measure for grid development when standard infrastructure reinforcements take too much time or when congestions are exclusively caused by too high wind infeed.

- **Irradiation forecast:** could be used to further fine-tune the applied rating of each element of the grid. The simplest way to implement it, could be a day / night forecast as input. This would typically increase the rating every night while keeping the current rating during daylight. More advanced models could be envisaged based on a real irradiation forecast during days.
  - **Impact medium**: such approach has the advantage to increase ratings of the entire grid at once and we know for sure that this would only mean higher or equivalent ratings
  - **Effort medium**: a simple rule could be applied on the short run by using the same approach as the one done with temperature-based ratings.

→ In the roadmap, "irradiation as input" will implement it by 2024.

- **Temperature and irradiation behavior by conductor type:** instead of by element type and voltage. Today Elia clusters elements by their types (cables, lines, transformers etc.) and by voltage levels. For each of these categories Elia takes the less performing element w.r.t. temperature and irradiation behavior as the reference for the entire category. By doing this Elia underrates most of the grid. To tap all the potential Elia should enrich the models (Power Factory for planning and EMS for real time grid operation) with all those necessary information.
  - **Impact high**: again this will boost the capacities of almost all elements of the grid.
  - **Effort high**: this represent a huge work in terms of data handling and tools improvements.

→ In the roadmap, "ratings by conductors type" will be possible as soon as the alignment of EMS and PF is finished.

FORECASTING MODEL

- **Revamping the Ampacimon forecast**: Ampacimon is presently working on an all-new forecasting model with:

1. machine learning and artificial intelligence embedded,
2. improved learning periods (either the last 3 years of available data or maybe a rolling window of 90 days in order to better follow the seasonality)
3. merging their 2 days-ahead forecast with their short-term forecast (now up to H+6).

    - **Impact high**: this new license is expected to provide decent gains, certainly for DA and ID timeframes.
    - **Effort high**: this is a complete rework of their product but of course the work for Elia is very limited.

→ In the roadmap, "revamping Ampacimon forecast" is ongoing at Ampacimon side with a strong collaboration with Elia. Commissioning in 2023 is expected.

- **Installation of new short-term forecast (H+6 and less, ongoing):** Operators might use the forecast 6h as a reliable source for their decisions making in terms of congestions management. Present version stops at 4h-ahead and is not reliable enough to be used.
    - **Impact medium**: this manual use of a reliable forecast 6h can be an interesting decision support, but cannot be easily implemented automatically into our IGMs without IT developments. Elia will probably wait for the 2 days-ahead forecast improvement for the automatic inclusions of the forecasts into the IGMs.
    - **Effort medium**: new version of this forecast is still to be deployed, and then a statistical study is needed to determine its reliability.

→ In the roadmap, "revamping Ampacimon forecast" will further improve the forecast 6h horizon and make it available directly into Elia IGMs in 2023.

RESULTING FORECAST

- **Increase of the capping factors**: this improvement is linked to the Horizon forecast improvement mentioned above but Elia can already anticipate that the ID caps will be greatly released based on some internal studies. Note that Elia will still probably have to cap the forecasts because it is calibrated on real-time ampacity (which make sense from physics point of view) while Elia makes use in real-time of a degraded "1h forecast" value as a proxy in order to smooth the signal (which makes sense from an operation point of view)[17].
    - **Impact high**: this improvement will allow to better align the real time ratings and the forecasted ones used in IGMs, facilitating a lot the decision making when preparing costly remedial actions.
    - **Effort medium**: once the new version of the 2 days-ahead forecast is available, a statistical study is needed to determine the new capping factors while conserving the same reliability level.

[17] **Dynamic Line Rating (elia.be)**

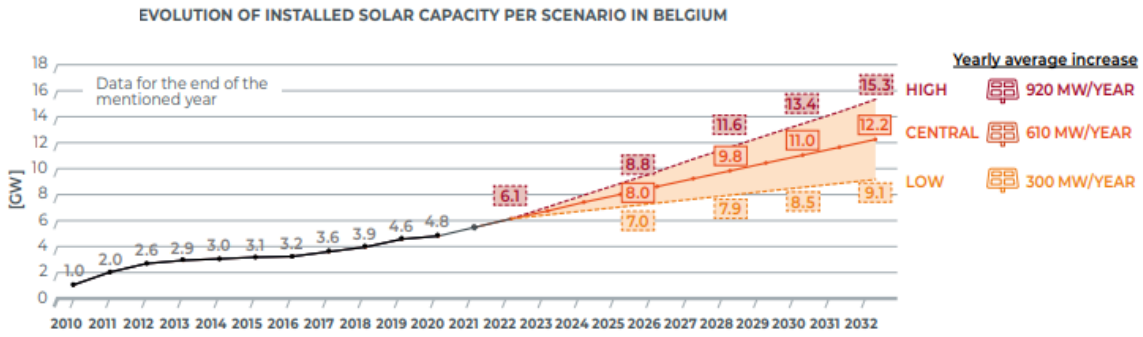
→ In the roadmap, during the "revamping Ampacimon forecast", Elia will also re-evaluate the capping rules.

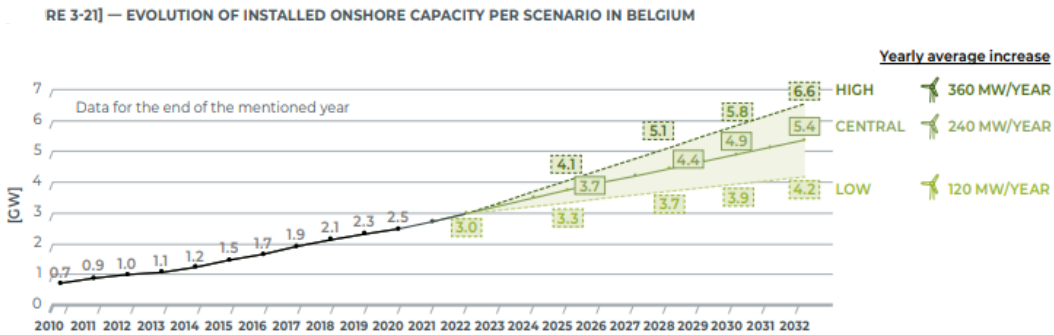
6. Challenges and opportunities for the future

In order to build a short-term implementation plan and long-term roadmap it is important to keep in mind future challenges and opportunities and to make the link with ongoing projects. Some challenges trigger improvement of forecasts on short-term, while others indicate that it is better to wait before starting any implementation.

6.1 Increase of installed capacity of decentralized production

The expected growth of each forecasted data can help to anticipate the future troublemakers in the flow forecasts from IGMs. It is well-known that all PV, onshore and offshore wind productions will increase dramatically in the coming years (see figures below from the adequacy and flexibility study for Belgium[18]).





[18]https://www.elia.be/-/media/project/elia/shared/documents/elia-group/publications/studies-and-reports/20210701_adequacy-flexibility-study-2021_en_v2.pdf

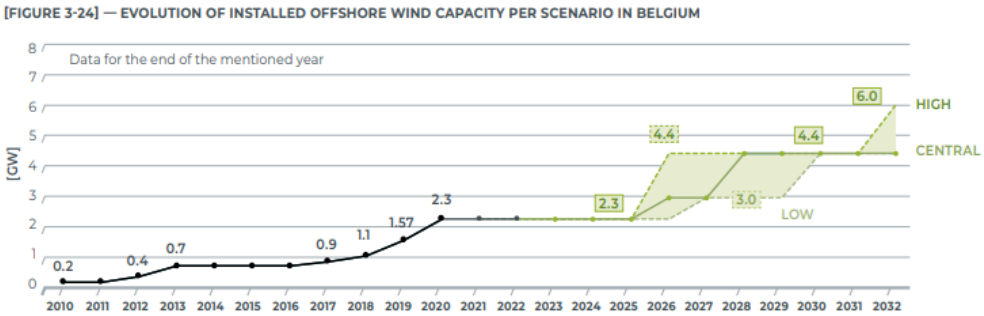


Figure 14: Evolution of solar and wind capacity per scenario in Belgium

6.2 Regional Operational Security Coordination (ROSC)

By mid-2025, the ROSC v1 will be live for the CORE region. A centralized and partly automatized security analysis will be realized notably for the costly remedial action optimization. Any error in the IGMs flow predictions on the monitored elements will directly impact the volumes and prices of redispatching. There will be less room for human assessment in case of an error in the IGMs files, the importance of a qualitative IGMs will consequently increase.

6.3 Update of information closer to real time

Updates on production information might become available closer to real time. BRPs are progressively authorized to nominate unbalanced positions to cope with the intrinsic growing uncertainty of the intermittent productions notably. Up to now, Elia did not detect an impact on congestion management and Elia does not predict a negative impact towards the near future. Elia should keep monitoring that the current practices (i.e. load scaling for DACFs and IDCFs) is a good approach.

6.4 Consumer centricity

The energy landscape is changing fast and in a deep manner, the consumer will be more active and aware about his consumption. Elia is already working with medium/large sized consumers in order to develop interactive tools that are able to forecast the individual consumption of the consumer. The idea is to have a platform that allows our customer to know what is their forecasted individual load and be able to improve this forecast for their use.

This is important because the behavior of the load will change soon, from a relatively inelastic load (relative to pricing) to a much more volatile load. A big chunk of this volatility will come from electric cars charging, heat pumps and batteries (charging, discharging, or holding) and the price will, likely, be a strong driver.

Elia must be agile in the way it forecast the load, more interactions with the customers means that the needs are better understood and anticipated. These profound changes will also occurs for the smaller customers so the knowledge of consumption dynamic profile will have to be even stronger.

6.5 Innovation projects and incentive

The innovation department of Elia is deeply involved in the improvement of the forecasting in general (not only for congestion management). The department is contributing to the task on several axes. The first is that the resources of the AI CoE (Artificial Intelligence Center of Excellence) are available to help the business to improve the forecasts. The data scientists of the center have a deep knowledge in machine learning and data handling, which is very valuable for the work done now regarding the reduction of the congestion costs through improving forecast. Moreover, the center is collaborating across business units inside the Elia Group (in Belgium and in Germany). This means that a broad range of experience and competence are put together increasing the general quality of the center. We have been exchanging extensively with this pool of experts during our analysis. Their strong knowledge about AI/ML modelling helped us to validate the potential of such approaches with respect to current way of working.

One of the most challenging part of AI/ML is the integration into operational (critical) processes and tools. The first concrete pilot of this AI CoE was the grid losses forecast, which was already developed in-house in Germany and then made in Belgium using the expertise from the German colleagues. The next project will be working together, across both Germany and Belgium with the AI CoE, on the forecast for Deterministic Frequency Deviations (DfDs). Those experiences will be very valuable for a successful integration of the proposed approaches in the roadmap.

7. Elia's implementation roadmap

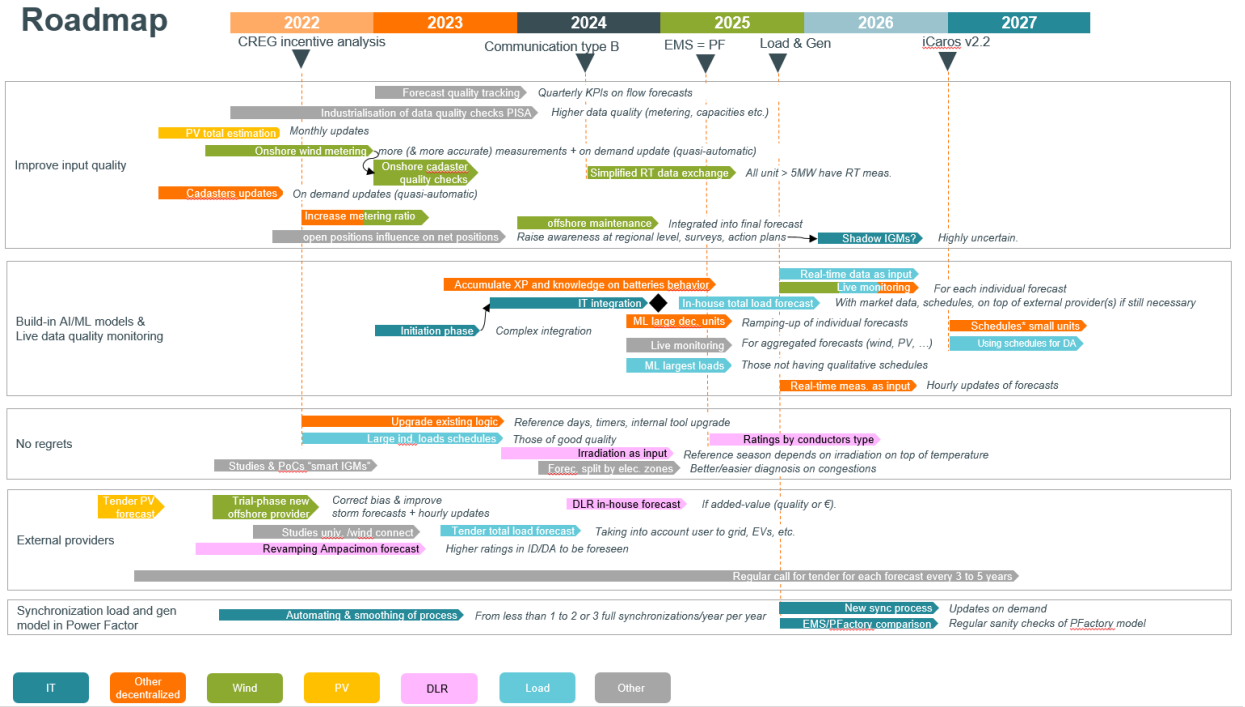
Elia took into account external dependencies as available budget and resources as well as some crucial enablers to build this implementation roadmap. Elia also tried to identify the root-causes of past congestions in order to set the right priorities in the roadmap. The analysis showed that sometimes short-term improvements made sense but Elia focused principally on structural, qualitative improvements with a long-term vision. The two main improvement categories are:

- **Improving the input data quality**: even if the present category is relatively good, possible improvements have been identified. Solving issues at the source is according to Elia the best way forward in improving the IGMs' quality.
- **Integrating AI/ML models into critical operation environment**: analysis showed little room for improvement if Elia does not migrate from a top-down forecasting approach to a bottom-up approach where each individual entity (load or generation) would be individually forecasted when it make sense (only the biggest or those located in potentially congested area's). It does not mean that Elia will completely stop current practices altogether. Elia will sometimes keep the current way of working as backup or as pragmatic approach for insignificant production or load. ML/AI models have the drawback to be less understandable as classical models. It is then very complicated for the end-user to understand what is happening and fix it. Nonetheless, with the alignment of EMS and Power Factory, operators will quickly identify divergences with reality and will act accordingly to fix it.

On top of those, Elia will implement some "no regret" items as soon as budget and resources allow it. Elia will also continue to collaborate with external providers where it makes sense. Analysis showed that for some aspects, at least in the coming years, the effort to internalize some forecasts is much bigger than the estimated gains. Finally, a stream

on model synchronization has been launched to ensure that the latest available data is continuously available in the grid models.

Below a visual summary of all initiatives validated for the coming years with a color code helping to grasp which kind of forecast will be impacted (“Other” being sometimes several forecasts at once or studies and other generalities). It also highlight the necessary IT developments, the latter being often a demanding and structural step.
Please bear in mind that implementation timings are ambitious and estimations are based on a “happy flow” hypothesis. Elia did not foresee any buffer in this roadmap to cope with potential hurdles it could encounter along the road.



7.1 Improving infeed data quality

7.1.1 Short-term

- Elia will capitalize on the existing flow forecast by building a quarterly quality tracking. In the current quarterly congestion management, the output data forecast part presents some KPIs on flow forecast on relevant lines in the IGMs. Elia will develop further KPIs in the same reporting. For each timeframe and relevant line, Elia will also include:
    - KPI on relative error of the forecast with hourly granularity on the reported period.
    - KPI on relative error of the forecast on the previous year to visualize the trend of the forecasting improvement on each line.

This will help Elia spotting eventual errors in its Power Factory model or in one of its forecasts used to feed IGMs. An example of proposal is included in Annex.

- PISA being at the core of all necessary inputs for production forecasts it is worthy to invest on more automatic quality checks. The status on the following attributes will particularly retain the attention of Elia: installed capacity, “InService” status, electrical localization, metering values should be complete and coherent with installed capacity.
- Elia will align the onshore wind cadaster granularity with PISA’s granularity allowing a lot more consistency checks by comparing real-time measurements, ex-post metering, and forecast for each individual park.
- Elia will make sure that all big enough units have a metering data of good quality improving the quality of the estimation of the effective production from the past.
- Finally, Elia will raise awareness on the open position situation in DA process to neighboring TSOs and Elia will advocate if the need is confirmed to set up a better net position DA forecast for at least CORE TSOs.

Elia expects to have cleaned as much as possible its infeed data by end of 2023.

7.1.2 Long-term

- In 2024, the project “Communication type B” will be live and Elia will enrich the EMS with a lot more real-time data from decentralized units. Upscalement of unmeasured units will be more accurate.
- In 2024, Elia will make it possible in EFTool for operators to adapt the forecasted output based on curtailment or maintenance information on offshore wind.
- If the net position forecast DA is put in place at regional level Elia might reconsider the need for a forecast of centralized units to cover the eventual remaining open positions. Indeed the idea is that first the open positions of Elia in DA should be partially reduced by the net position forecast DA, the leftovers should be covered by ID trades on flexible units in Belgium. Elia has tested and showed that it has little meaning without the suggested net position forecast DA.
- Finally, new types of input data might emerge in the coming years (EVs, batteries, etc) and Elia will continue to evaluate their eventual pertinence for congestion management.

7.2 Integration ML/AI modelling into operations

7.2.1 Short-term

- Elia has successfully tested offline AI/ML modelling for individual loads and decentralized productions. In 2023, Elia starts an ambitious implementation project. Many off-the-shelf tools exist but they must satisfy demanding requirements to be continuously communicating with critical applications (EMS, PF, EFTool, etc.). By end of 2023, we should have a better idea on the possible IT architecture, implementation would then start in 2024.

7.2.2 Long-term

- When the modelling tool will be available in operations Elia will be able to progressively configure individual forecasts by starting with the most critical (based on size and location).

- Elia will progressively configure live monitoring for the different forecasts from the external providers.
- Elia will set-up an in-house total load forecast to try to challenge the external provider and test other eventual predictors.
- By end of 2025, loads and generations in PF and EMS should be perfectly aligned allowing Elia to enrich the individual ML/AI models with real-time data as well as live monitoring.
- When iCAROS 2.2 will be live, Elia will receive more information on small units. Elia will then asses the quality and pertinence of those new inputs for further improving the individual models.
- Elia will also receive more information for loads, the latter having being proved already useful Elia will certainly feed the individual models with them.

7.3 No regrets

7.3.1 Short-term

- During the extensive analysis realized for this incentive Elia realizes that the current logic of the other decentralized units could be improved by adding some better representative type of days categories. EFTool will be upgraded in 2023 accordingly and the metering acquisition timer will be shortened. This combined with the increase of metering data described above will make sure that the existing logic is optimized. This is important to upgrade the existing logic because even if Elia wants to go for individual ML/AI modelling it will remain for a subset of units. Indeed there are about 900 units and a lot are very small.
- Elia also realizes that some available offtakes DA schedules were a good quality while not implemented into Elia's IGMs. The IT integration are foreseen in 2023 to capitalize as soon as possible on the biggest loads (about 1GW equivalent).
- Finally, Elia worked with an external expert (Cepsis NV – Data Scientist Expert) to deep dive on the flow forecast errors from its IGMs, the drivers of those errors and potential underlying dynamics. The idea was to better understand the forces at play and eventually set the right priorities in this implementation roadmap. Conclusions showed that the higher the voltage, the lower the autocorrelation of the errors on the IGM flows. It means that except for few lines on 150 & 220 kV, using errors older than 2-3 days in the past to build the next IGMs will not help much. Moreover, the study of the correlation between many different drivers and the error on flow forecasts showed that only the drivers "price indication, wind and consumption of gas power plants" could have a positive impact to improve the error made on flow forecast. The forecast improvement of those drivers is already foreseen in the above roadmap. Consequently, Elia will stop looking further to such approach and will use the study outcomes to improve the few elements pointed out by the study and keep improving all individual forecasts to make sure that its IGMs remains of good quality despite the increasing share of volatility and decentralization in its grid.

7.3.2 Long-term

- End of 2021 Elia implemented a change of reference ratings for all grid elements based on temperature forecasts (for IGMs) and measures (for real-time grid management). This increases on average the entire grid

ratings by about 4-5 %. Elia wants to continue in the same direction by adding in 2024 the irradiation component. Concretely this would change the ratings in the IGMs based on temperature and irradiation. Meaning that every night higher ratings will be applied.

- In 2026 Elia will capitalize on the alignment of the entire assets data base in all our systems to apply those dynamic ratings (temperature and irradiation based) at a conductor type level instead of the current way of working (by voltage levels). Again, this feature will increase on average the grid ratings because for the moment the least performant conductor of a given voltage is dragging the other ratings down.
- In 2024, Elia will also make some IT investment in the EFTool for increasing the performances, layout and users experiences. In this package, Elia foresees to add an electrical zones filter, which will help operators to understand the root-cause of some congestions (typically those linked to inter-zonal 150kV lines).

7.4 External providers management

7.4.1 Short-term

- During the winter 2022-2023, Elia is testing the offshore forecast of the onshore wind production provider (VITEC). Elia help them to improve their forecast during storms. Elia already knows that they perform better than the current provider (IRM/KMI) when there is no storm. If they perform well enough during storms Elia will activate VITEC instead of IRM/KMI. If they do not perform well enough Elia will apply a more complex solution which is the activation of VITEC when wind speed is lower than 20m/s and IRM/KMI otherwise.
- In 2023, Elia will integrate a new version of the forecast of DLR with hopefully higher capacities but also with all equipped lines while today only a subset have a forecast (due to cost optimization).
- In 2024, Elia will launch a tender on total load forecast. Elia will have a strong focus on EVs, batteries and other possible game-changers for such forecast. Elia expects to learn a lot from this exercise on top of to select the best candidate.

7.4.2 Long-term

- Elia will set up an AI/ML forecast for the rating of each lines equipped with DLR. Maybe it will outperform or at least be competitive with the forecast bought from Ampacimon. This is clearly one of the benefits to do this ambitious integration of ML/AI capability in operations: building a forecast will become relatively easy even if configuring a good forecast might remain challenging of course.
- Elia will continue to look for the top of the class providers for wind, solar and total load forecasts. Tenders are always a moment where Elia learns a lot and for those forecast Elia cannot be competitive in terms of costs and quality.

7.5 Synchronization Power Factory load and generations model

7.5.1 Short-term

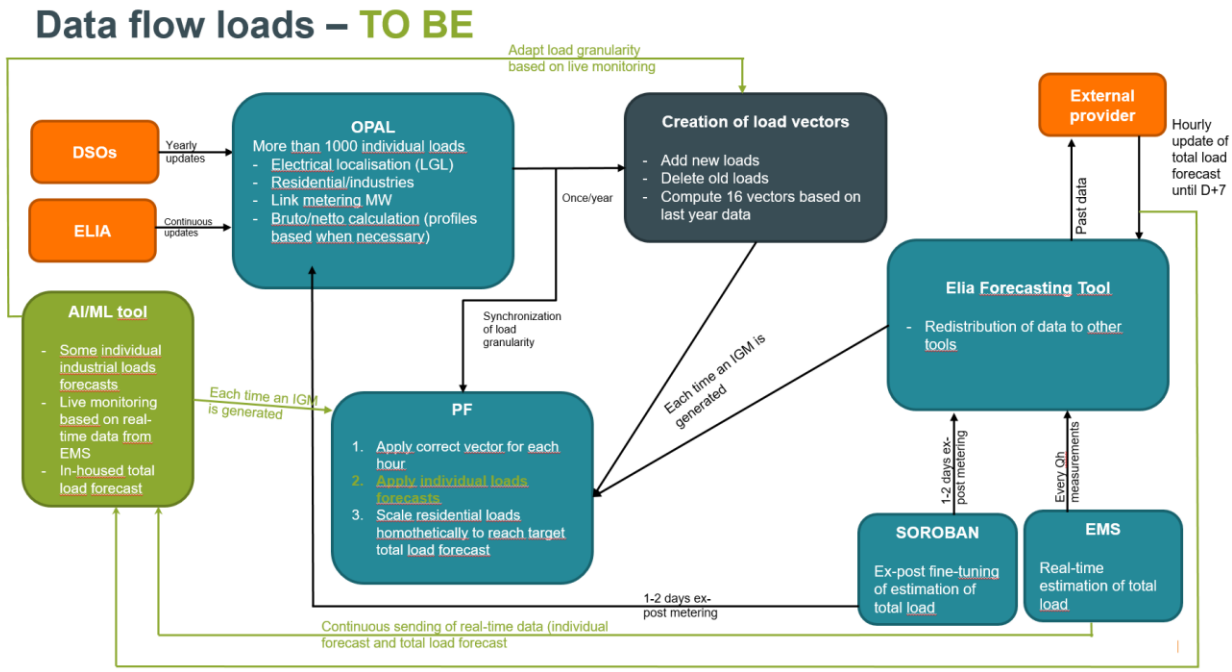
- In 2023, Elia will automatize many of the steps thanks to the help of python experts. It will already grandly ease the process. Elia will be able to synchronize PISA/RPN/PF up to 2-3 times a year. Increase the frequency

would be too much effort in comparison to the benefits but Elia is always able to adapt the model manually by adding or changing a load or a production when the latter have impact on potential congestions.

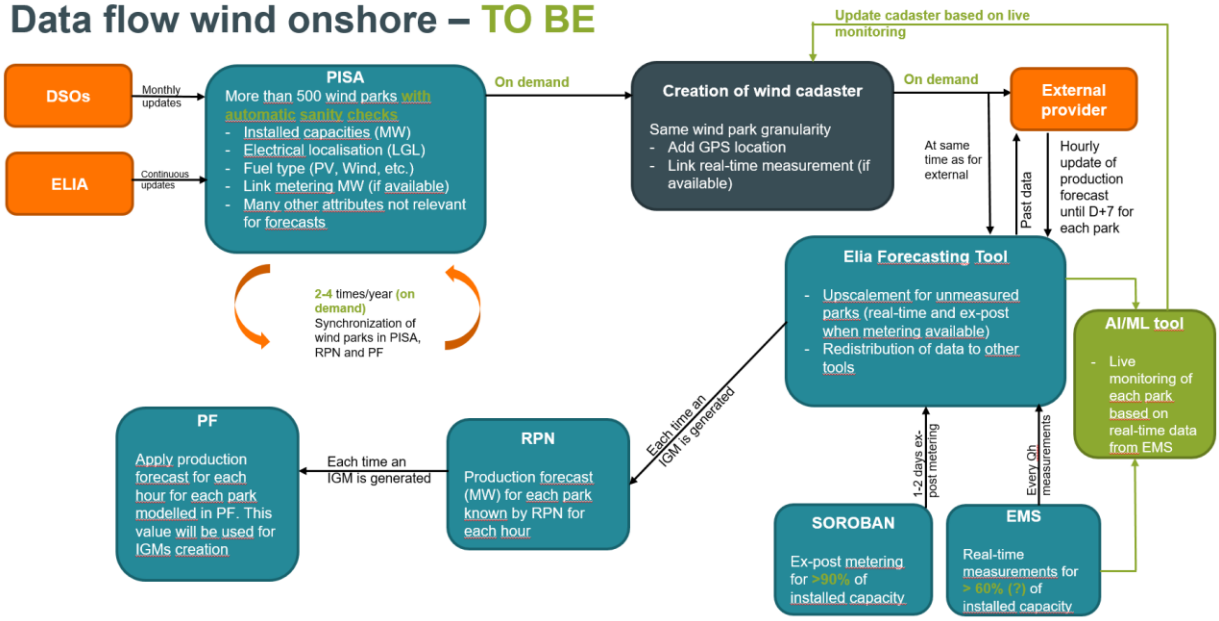
7.5.2 Long-term

- With the go-live of load and generation project one may expect an even smoother process. Elia does not have yet a clear view on the possibilities but for sure, Elia will capitalize on this alignment in PF/EMS.

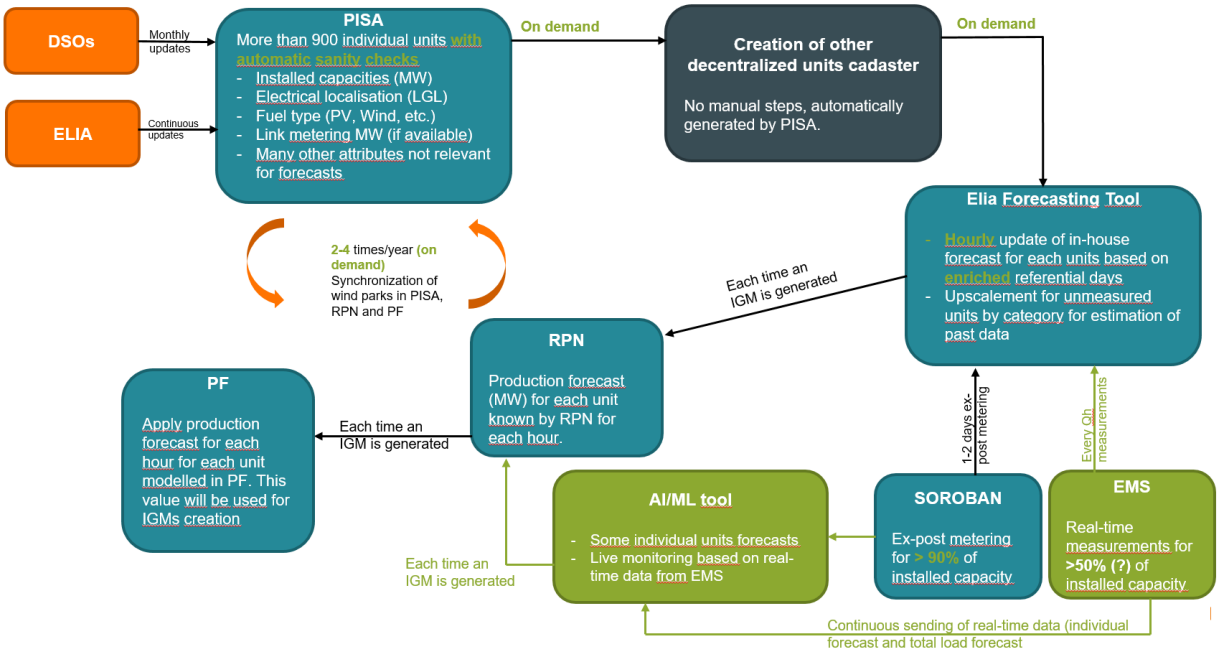
7.6 Data flow charts TO BE



Data flow wind onshore – TO BE



Data flow other decentralized units – TO BE



8. Conclusion and next steps

This report covers an in-depth analysis of the causes of the deviations in the different forecasts (wind, solar, load,...) and a roadmap to improve these forecasts. Elia wants to thank all market parties and academics for their inputs to come to this result. Elia commits the coming years to implement the proposed roadmap in order to improve the forecasts having an impact on congestion management. Elia also commits to implement new KPIs in case this is needed for a proper follow-up of the forecasts. Due to changing circumstances, the timings and priorities may change in which case Elia can repeat the analysis performed and adapt the roadmap. At any time in the coming years Elia can present a status overview.

9. Annex – Improved KPI on output data forecast

	D-2			D-1			ID_3h			IDCF_15h		
line name	P95 relative hourly error (%)	Average relative error on Y-1 period	Average hourly relative error on the current period	P95 relative hourly error (%)	Average relative error on Y-1 period	Average hourly relative error on the current period	P95 relative hourly error (%)	Average relative error on Y-1 period	Average hourly relative error on the current period	P95 relative hourly error (%)	Average relative error on Y-1 period	Average hourly relative error on the current period