Improve the quality of input data for congestion management

Today and possible evolutions

30/05/2022 | Workshop

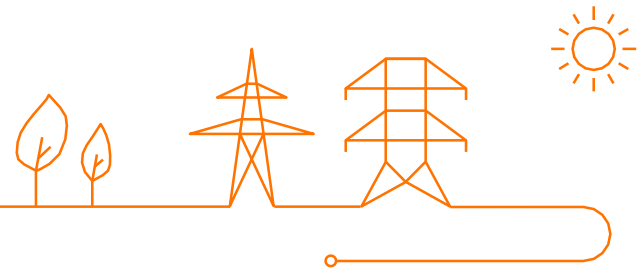
Agenda

1. Introduction

- Incentive 2022
- Congestions forecast today
- Models building
- Impact vs. accuracy

2. Deep dives

- Generic ideas
- Load forecast
- PV
- Wind
- CHPs forecast
- Smart IGMs & direct flows forecast
- Classical flexible units



Introduction

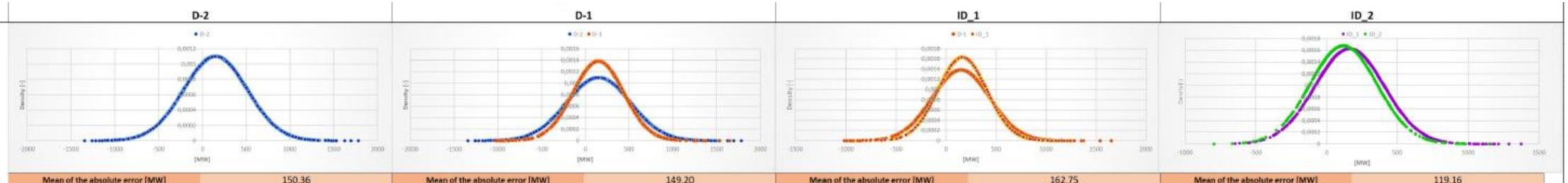
Incentive 2022: context

Incentive 2019: ‘Improvement of transparency with regards to the detection and management of Congestion’

- Information on the quality of the forecasts used as operational input data for the creation of the Individual Grid Models (IGM): wind, solar, CIPU, load, PST positions
- Information on the quality of output data: flows on 150kV – 380 kV network
- Information about the timing, power, location, and purpose for activations of Costly Remedial Actions by Elia

Public reports available on Elia website

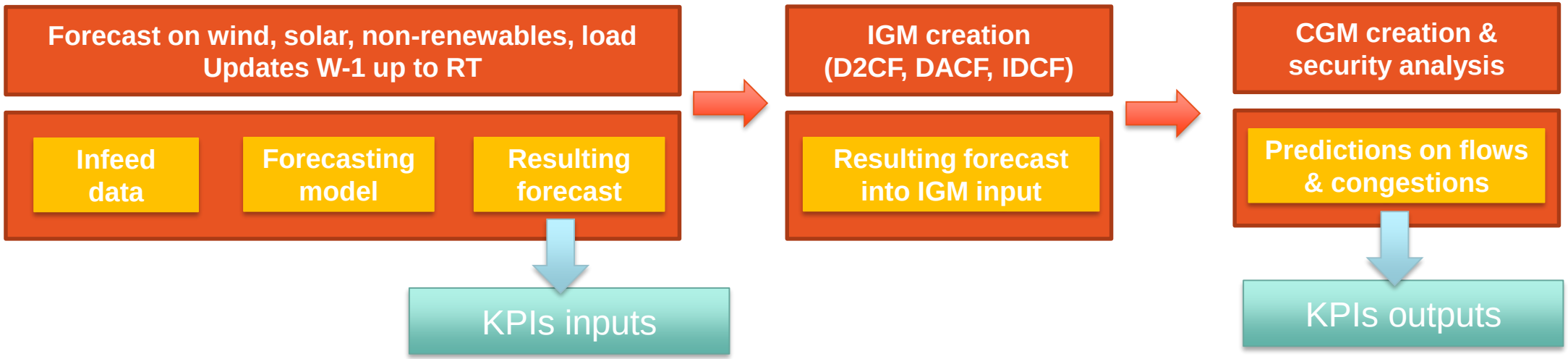
Offshore wind production



	D-2	D-1	ID_1	ID_2
Mean of the absolute error [MW]	150,36	149,20	162,75	119,16
Mean of the absolute error (% of installed capacity)	7%	7%	7%	5%
Mean of the relative error (% of actual production)	-64%	-200%	-139%	52%
Standard deviation [MW]	363,34	288,77	244,88	237,73
Maximum errors [MW]	1782,49	1652,65	1375,71	1148,07
Absolute P95 error [MW]	784,37	708,99	644,98	546,30
Mean of the installed capacity [MW]			2121,82	

Incentive 2022: goal

- **Transparency** on congestion forecast today (IGM) and modelling practices
- Transparency on forecast quality (insights in input, model and output data)
- Root-cause **analysis** on deviations in forecast compared to Real Time
- Look into **solutions** to improve the forecasts
  - Short-term implementation
  - Long-term roadmap



Incentive 2022: timing

Workshops with universities: presentation with specific questions to trigger discussion

- 24/05 (KU Leuven)
- 25/05 (ULB)

Workshop with market parties: presentation to introduce the report for public consultation

- 30/5

Launch of public consultation: report (focus on transparency, analysis and first indication of solutions)

- 10/6 (one month)
- Summer: Analyzing comments and responding to public consultation

Finalization of the report (including comments and finalizing solutions/roadmap)

- End of December

Congestions forecast today

Weekly process

Relevant variants calculated to cover extreme situations (“realistic worst cases”) → only **installed capacities** important

D2CF

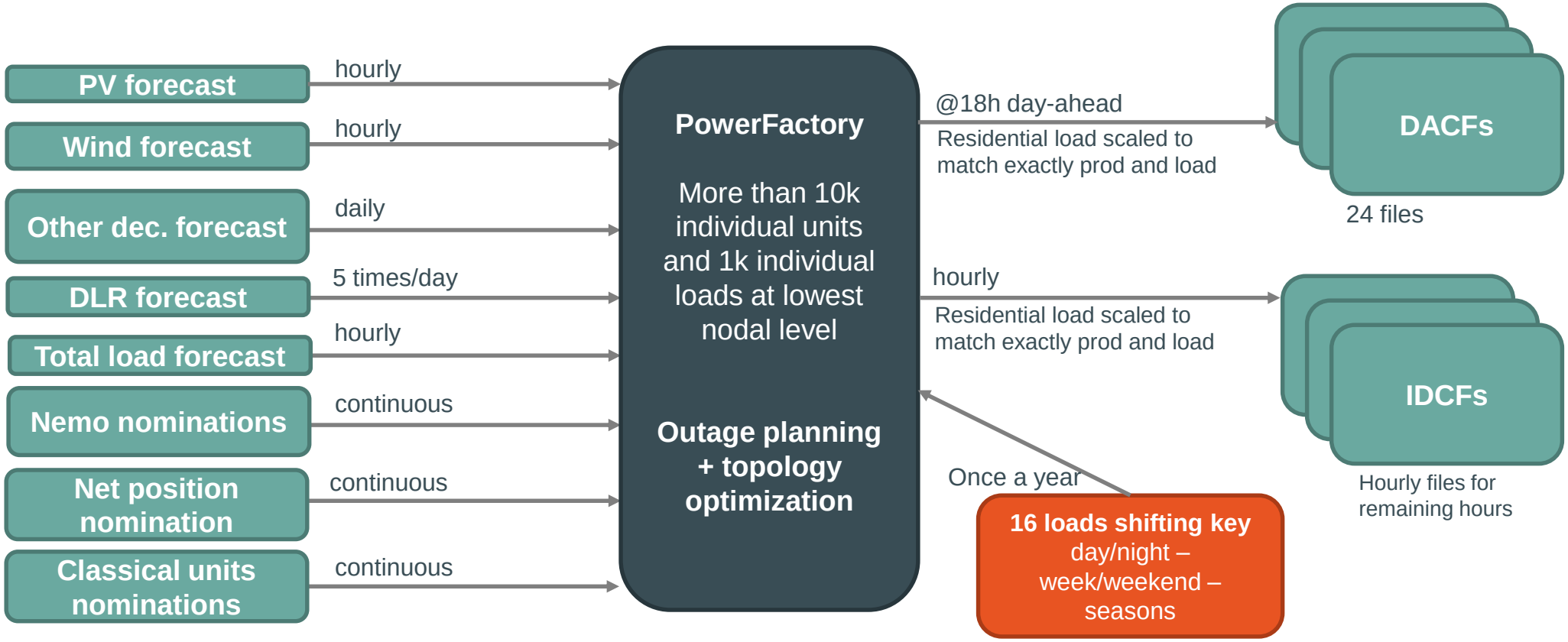
Input for day-ahead markets calculations (Flow-Based). Any error here may generate congestions. Improving the quality of D2CF will not improve the congestions management, but it will reduce the **risk to face congestions** (not exactly the scope of this incentive).

DACF/IDCF

Sole input to detect and solve congestions. Any improvement in accuracy of the resulting flows on **congested** elements is crucial. Improving quality of non-congested elements has not a direct usefulness (e.g. some radial elements).

Day-ahead and intraday forecasts = main scope

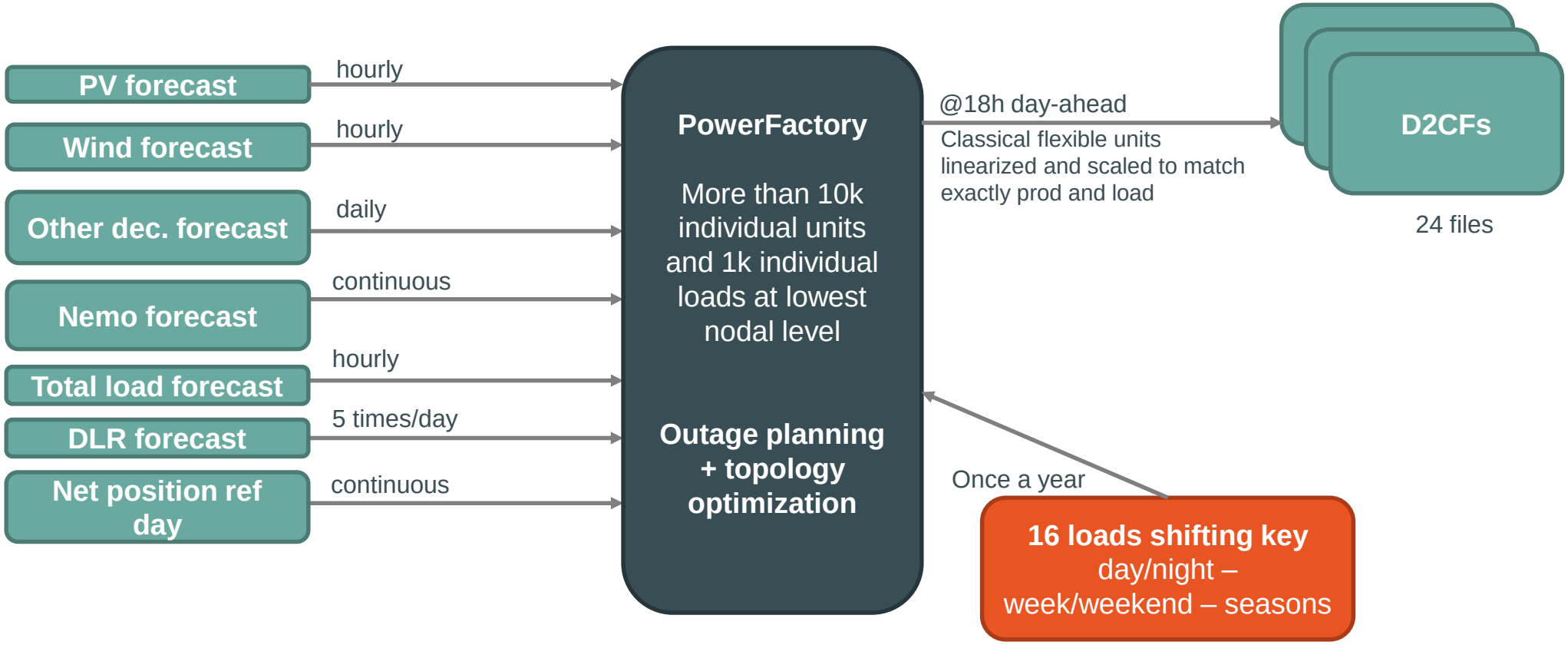
Models building – DACFs/IDCFs



Best estimate principle

Residential load scaling necessary to match forecasts Elia with nominations market parties

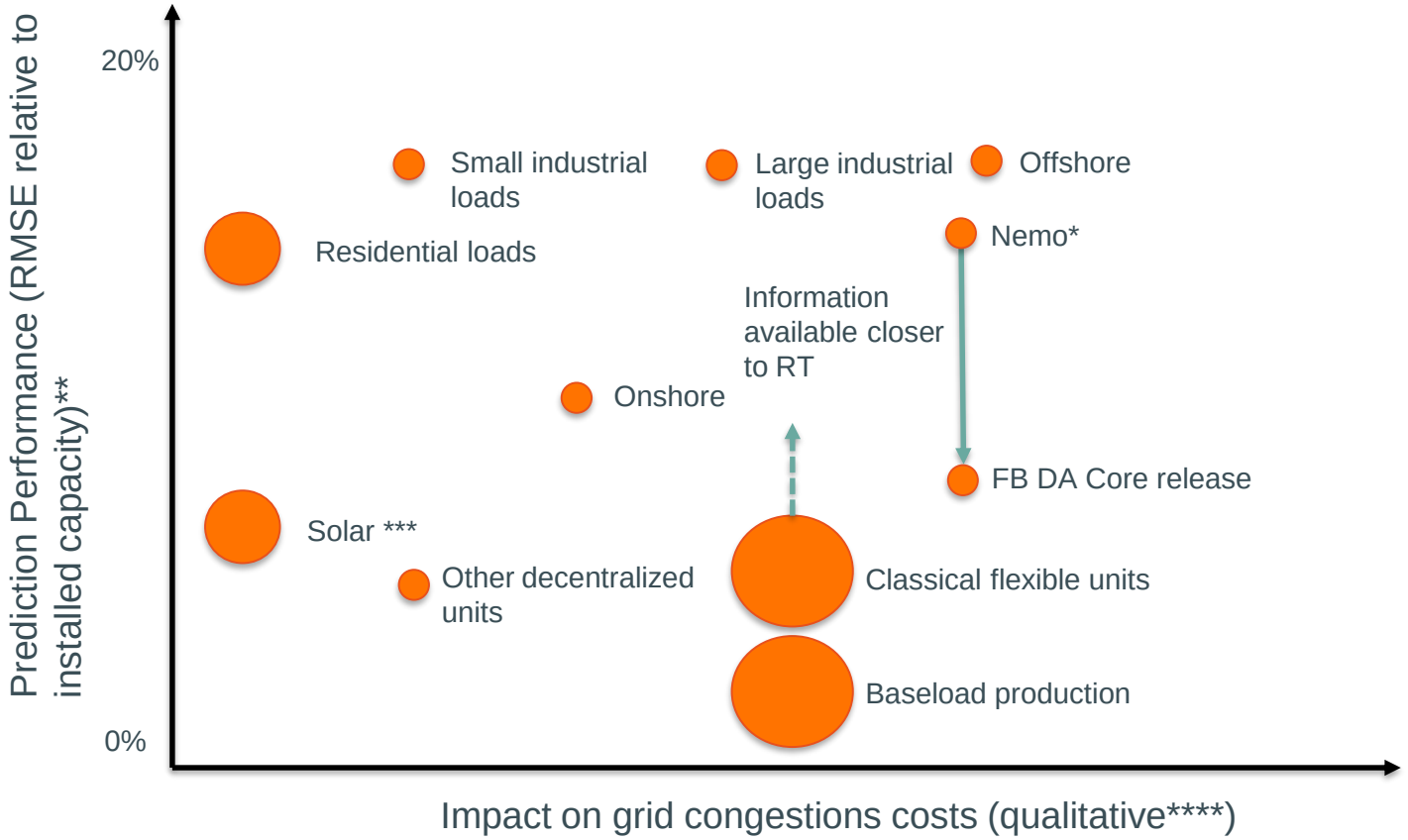
Models building – D2CFs

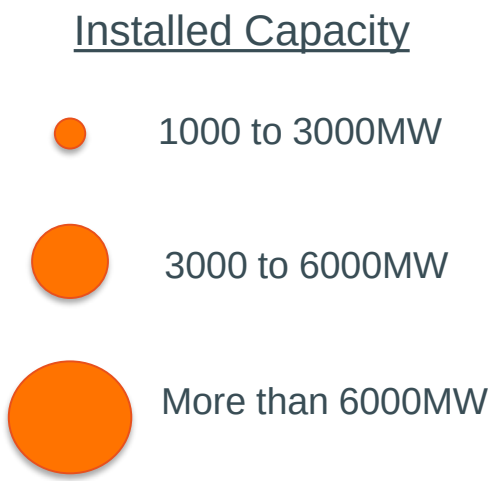


Best estimate principle

Classical flexible units scaling necessary to match forecasts Elia with imposed net position

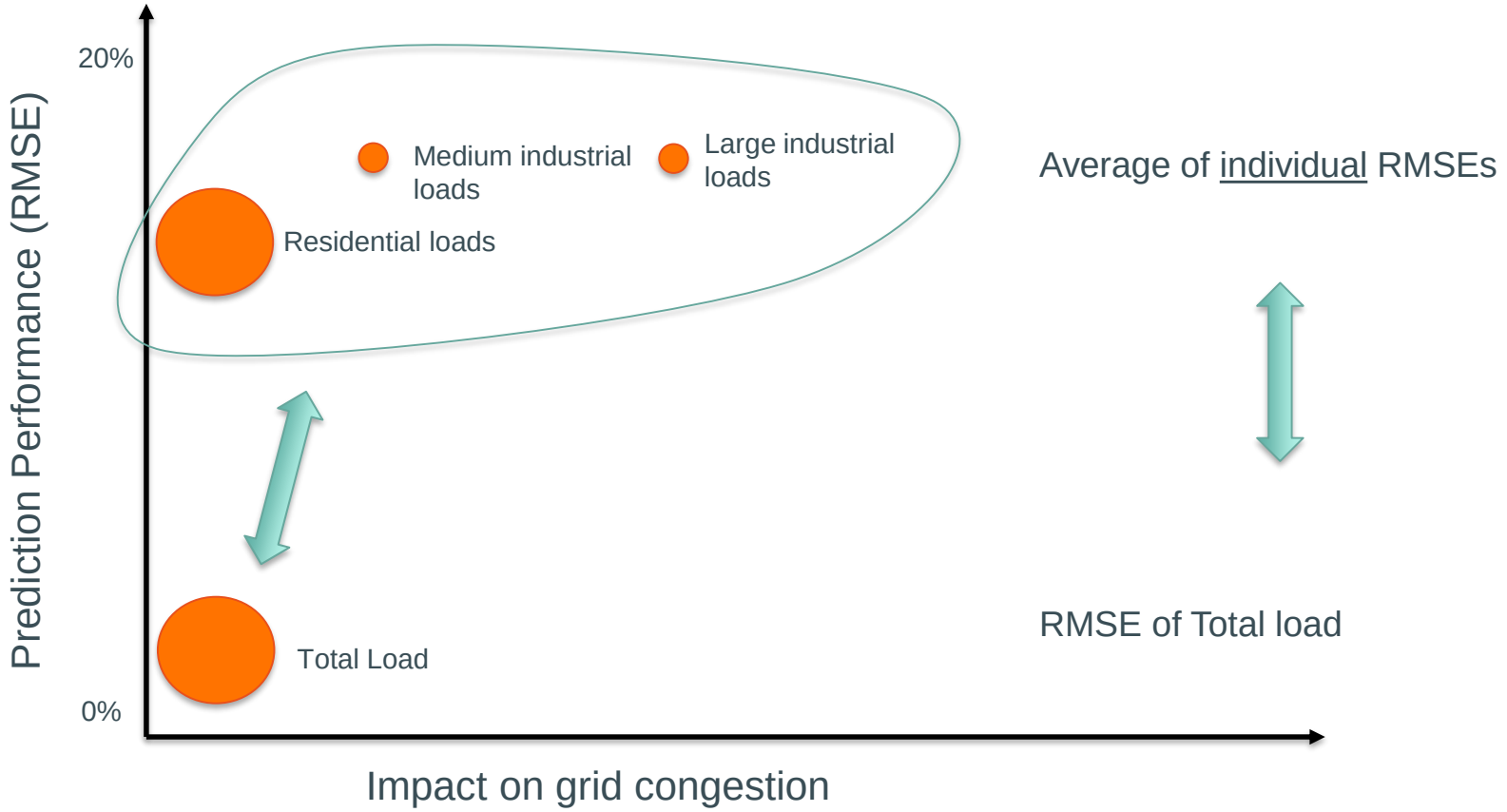
Mapping the opportunities





*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)

Aggregation hides the errors



Possible improvements

Some generic ideas to improve current situation

AS IS

Detailed info on most of installed capacities available (>10k units), some discrepancies for smallest units (PVs typically). Some forecast still done at aggregated level. Monitoring of forecasts quality mostly ex-post and not systematic.

Tomorrow:	More and more decentralized units → need to keep high quality
[Infeed]	Ensure completeness by maintain/improve data exchange DSOs → ELIA
[Infeed]	Align real-time and planning modelization for loads and production (access to recent historic for forecasting, possibility to compare forecasts and reality -> sanity checks, model improvements)
[Model]	Apply individual forecast with machine learning (only when relevant because > 10k units)
[Resulting Forecast]	Pushing updates in installed capacities into operational tools more rapidly. Heavy process today, done once every 2 years, we should aim at 4 times /year.
[Resulting Forecast]	Continuous and automatic monitoring of forecasts quality (avoid propagation of errors)

Having a good model but poor infeed quality or poor data processing to be avoided.
Crucial to focus on individual loads/production having a real impact on congestions

Total load forecast

AS IS

External provider (equivalent T°, historic, calendar), forecast aggregate total load (cleaned from dec. prod)

10GW year average, **RMSE ~3%**, split into ~1000 individual loads → **low impact congestions**

Tomorrow: EVs, HPs, consumers to grid, electrification, demand response → will **deteriorate** RMSE and **increase** impact

[Infeed] Latest market prices ? Today no measurable prices response but tools and processes need to be ready

[Infeed/model] Split into independently forecasted subcategories (EVs, HPs, residential, industries)? Today not enough EVs/HPs to study and impossible to maintain live data with split residential/industries today

Need to keep an eye on that and anticipate tools and processes compatibility

Individual loads forecast

AS IS

16 load repartition keys (day/night – weekday/weekend – seasons) → average of last year data per catergories

6 GW year average residential load, **RMSE ~15%, P95 ~25%**, evenly split in the grid → **low** impact congestions

1,5 GW year average large industrial loads (40MW up to 300MW), **RMSE ~17%, P95 ~ 30%** unevenly split → **medium** impact congestions

Tomorrow: Electrification of big industries → impact on congestions expected to **increase**

[Infeed] Use offtake forecasts of biggest variable demand (f.e. industries)

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	/

[Model] Individual residential loads forecasting (KNN model, predictors = D-1 total load forecast, temperature forecast and datetime info= ML1 + recent (48h lag) measurements available = ML2).

- **Need aligning loads models to tap full potential of individual load forecast with ML**
- **Offtake forecasts still better than ML2 for large industries**
- **ML1 for residential will deteriorate with EV/HP penetration → real-time data feeding necessary in the long run?**
- **Preliminary results, probably possible to further increase ML performances**

PV

AS IS

- External provider for data measurements (80k measurement points) and forecast. Nodal capacities 80% complete, the gap is homothetically derived from existing units (NEW from March 2022). Gap is estimated by using regional official data.
- We do use MWp for cadaster because inverters not yet limiting
- Errors on PVs units are **small** and on very small units spread on lowest voltage levels -> **low** impact on congestion

Tomorrow: if assuming 20% gap remain -> will increase in absolute value
PVs curtailments to be anticipated?

[Infeed] Market prices to predict curtailments but need to know localization of curtailable PV?

Today little room for improvement while low impact on grid. Need to maintain quality.

Wind

AS IS

- One provider for onshore (400 individual forecasted time-series)
- Two providers for offshore. Today we favor provider the most accurate in storm.
- Errors are **high** while some units are large w.r.t. their connection voltage -> **high** impact on congestion

Tomorrow:	offshore critical due to concentration and lack of non costly remedial actions
[Resulting Forecast]	switch offshore provider based on weather condition (if wind > 20m/s switch to 'storm' forecast)
[Infeed]	Sanity checks from onshore cadaster (GPS coordinates, measurements quality etc.)
[Infeed]	Use other offshore parks in the vicinity (other countries) to improve ramping event predictions

Errors being high does not necessarily means there is room for improvements
Regular tenders ensure that best of the class is selected

Other decentralized (mainly CHPs)

AS IS

Internal tool, RMSE **~5%** aggregated (**quid nodal?)**

35% of capacity are > 10MW (**medium** impacting)

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

Tomorrow: Low growth, batteries without scheduling obligations?

[Model] Apply current logic of ref days but at power plant level (**short-term** improvement)

→ *Pforecast_unit,QH = sum_for each ref day (unit Pmeas_QH)/number of ref days*

[Model] Individual power plants forecast (ML) for relevant ones (same approach as for large industries)

[Infeed] Use ON/OFF status when available (2025?) to capture unpredictable behavior

Need to identifies units causing congestions and apply individual ML model on those only

Smart IGMs & direct flows forecast

AS IS

Not available for the moment, mission is launched with external consultant to explore the opportunities

Tomorrow: growing activation of redispatching -> top down approach could help in detection of congestions and decision making (no replacement of IGM building process)

[Infeed] Explore if dynamic and/or weighted combinations off available forecasts could lead to better IGM

[Model] If analysis on infeed is successful, develop a prediction model for flows

R&D, to be followed

Classical flexible units

AS IS

- DA/ID nominations of good quality. No forecast of such units on DA/ID timeframe today.

Tomorrow:	Market goes more and more close to real time, DA vision becomes blurier. Maybe need to forecast ID moves?
[Infeed]	Using latest prices and volumes, open position of Elia zone, ATCs
[Model]	Unit commitment module PowerFactory to solve constraint (gap) while minimizing costs
[Resulting Forecast]	Somehow ensure that results are indeed a best estimate (e.g. only push those changes if gap was big enough)

PoC has been done, in some extreme situations forecast works partially (not great but still better)
Missing part : open positions from neighboring countries

Thank you.

Back-up slides

Exploring the AI landscape

“Simple” Machine Learning

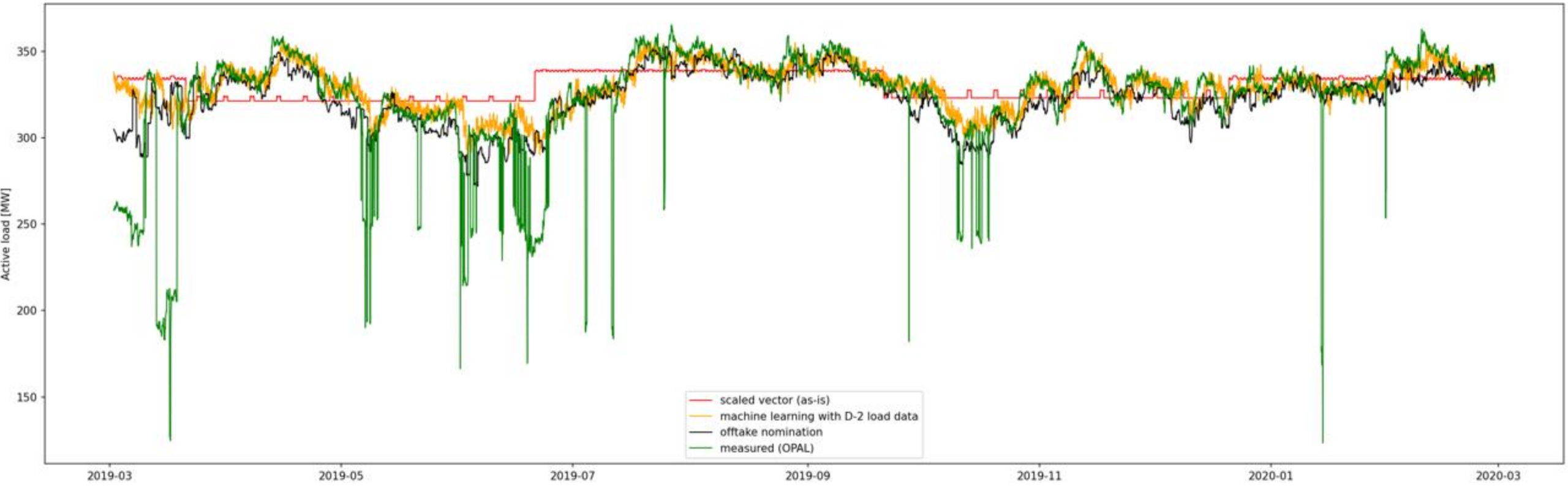
- Relatively fast to train
- Hyper parameters tuning is easy
- Able to compare easily different algorithm
- Gradient Boosting Regressor, Random Forest, KNN,…

Neural Net

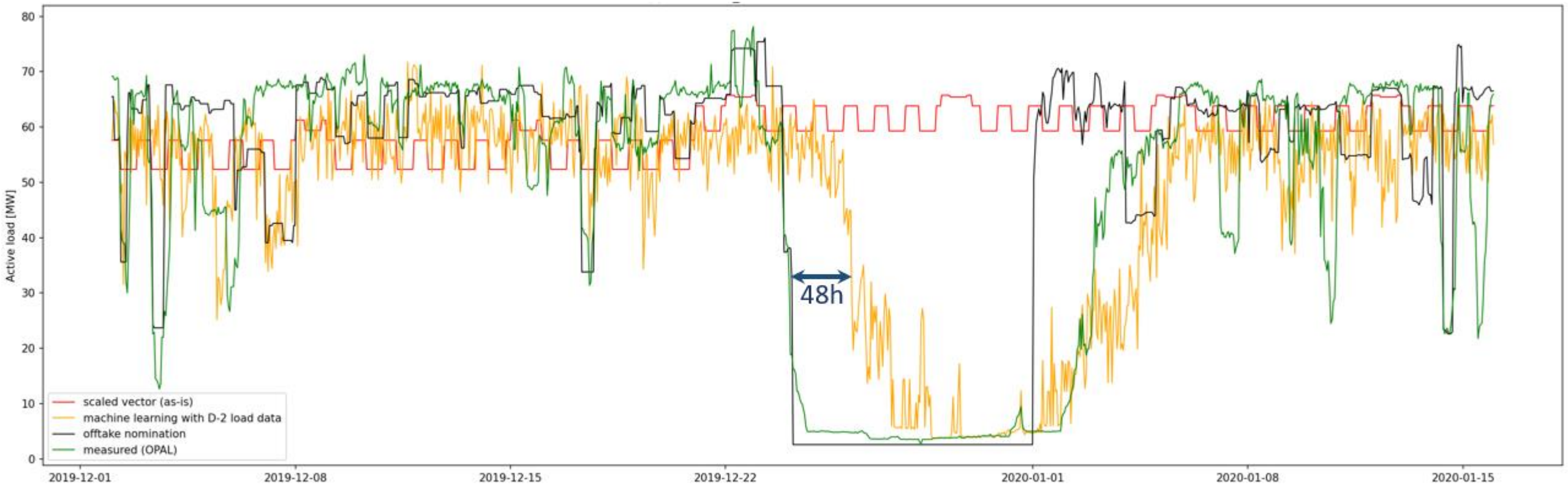
- Training needs a huge dataset
- Training is slow
- Able to grasp subtleties not seen by simple ML?
- Convolutional Neural Net (CNN) using LSTM (Long Short Term Memory), Transformers

Which approach to use for each forecast?

Typical example of industrial load



Machine learning over-fitting to the noise?
Revisions/stops are unpredictable -> recent data or nominations to solve it



Typical example scaled load vector versus machine learning (domestic load):