REPORT FOR CONSULTATION

System Imbalance forecast and its publication to stakeholders

Content

Elia Transmission Belgium SA/NV
Boulevard de l'Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium

Elia Transmission Belgium SA/NV
Boulevard de l'Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium

CREG incentive on System Imbalance forecast and evaluation of its publication

1. Introduction

The System Imbalance is a key indicator for the state of the system as it shows the difference between the Area Control Error (ACE) and the activated ancillary services (NRV). This makes that the SI is a key decision making factor – along with others - for the activation of (additional) balancing reserves. Hence, having the ability to accurately predict the SI of the future quarter-hours may improve the overall decision-making process for the activation of balancing reserves. This is especially relevant for the accession to the future balancing platform (MARI and PICASSO) where the activation lead-time will increase. Furthermore, this project allows a greater understanding of the different drivers of the System Imbalance.

The project consists in selecting, training and implementing a data mining type model (such as ARIMA, neural network, Support Vector Machine...) to predict the system imbalance and then testing it in parallel run in different system conditions. The project also includes an analysis of the relevance of making this representation of system imbalance (SI) available relevant stakeholders. The work presented in this report builds further on an earlier proof of concept (PoC), more information on this PoC is available in annex 3 of decision B658E[1].

This report presents the status of the project, describing the different data mining techniques and the process applied for the variable selection. In addition, it explains the selection of the linear regression model as the basis for the testing phase. It also includes an evaluation of the publication of the SI to the relevant stakeholders.

Throughout the report several question to stakeholders are raised, these are again summarized in the final section of the report.

1.1 Deliverables and timeline

The incentive as defined in decision B658E of the CREG sets out three deliverables. First, by January 31st 2021, the data set and model should be selected. The selection of the linear regression model was agreed by the CREG and presented in the WG balancing in March 2021.

Second, the consultation of draft report containing a description of the used method to select the data mining model on the basis of statistical indicators complemented by an analysis on the relevance of the publication of the system imbalance and a proposal for an implementation plan. In agreement with the CREG, this report will not include a proposal for an implementation plan. Rather it shall focus on identifying potential interdependencies with other evolutions

[1] https://www.creg.be/sites/default/files/assets/Publications/Decisions/B658E68Annex3.pdf

Elia Transmission Belgium SA/NV
Boulevard de l’Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium

on the balancing market and requesting feedback on the publication format. Following this consultation, Elia shall present an implementation plan in the WG Balancing later this year.

Finally, a final report should be published by the end of the year including test results and a potential implementation plan.

1.2 Public consultation

The goal of this draft report is to collect the feedback from stakeholders on the selected models and the assessment on the relevance of the publication. This report will be consulted from August 31st 2021 until September 30th 2021.

2. Forecasting models

This chapter describes the process that was followed to derive the relevant variables for the System Imbalance forecasting model. In addition, it assess the impact of increasing the so-called look back horizon and extending the number of variables considered.

Next, this chapter contains a high-level description of the different data mining models that were considered in the scope of this project.

2.1 Introduction

The forecast of the System Imbalance (SI) proof-of-concept of 2019 was developed using a multivariate linear autoregressive model. The SI estimate for quarter-hour (qh+k) is determined by the SI of the previous quarter-hours as well as other variables of the previous quarter-hours. In addition to these lagged variables, the model uses one-minute measurements happening during the current quarter-hour (qh), up to minute 8 in that quarter-hour.

A generic equation for such a model is given in Figure 1.

$$
\begin{aligned}
&\mathrm{SI}_{\mathrm{Forecast}}(\mathrm{qh}+\mathrm{k}) \\
&\quad = \mathrm{Intercept} + \sum_{n=1}^{\mathrm{lookback}} \mathrm{SI}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{SI}(\mathrm{n,k})} \\
&\quad + \sum_{n=1}^{\mathrm{lookback}} \mathrm{NRV}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{NRV}(\mathrm{n,k})} + \sum_{n=1}^{\mathrm{lookback}} \mathrm{GDV}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{GDV}(\mathrm{n,k})} \\
&\quad + \sum_{n=1}^{\mathrm{lookback}} \mathrm{GUV}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{GUV}(\mathrm{n,k})} + \sum_{n=1}^{\mathrm{lookback}} \mathrm{PPOS}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{PPOS}(\mathrm{n,k})} \\
&\quad + \sum_{n=1}^{\mathrm{lookback}} \mathrm{PNEG}(\mathrm{qh}-\mathrm{n}) * \mathrm{Weight}_{\mathrm{PNEG}(\mathrm{n,k})} + \ldots \\
&\quad + \sum_{n=1}^{8} \mathrm{SI}_{\mathrm{minute}}(\mathrm{qh},\mathrm{n}) * \mathrm{Weight}_{\mathrm{SI\ minute}(\mathrm{n,k})} \\
&\quad + \sum_{n=1}^{8} \mathrm{NRV}_{\mathrm{minute}}(\mathrm{qh},\mathrm{n}) * \mathrm{Weight}_{\mathrm{NRV\ minute}(\mathrm{n,k})} + \cdots
\end{aligned}
$$

Figure 1 – Multivariate Autoregressive model for SI forecasting.

In Figure 1, parameter k is called the forecasting horizon. Conventionally speaking, the estimate for current quarter-hour (qh+0) is named S0 (with k=0), the estimate for qh+1 is named S1 (with k=1) and the estimate for qh+2 is named S2 (with k=2).

Training such a model requires finding weights for each of the term in the equation. Thus even though the model is a relatively simple linear model, the number of terms in the model can grow quickly with the quarter-hours used in the past and the different variable types used (NRV, GUV…), as illustrated in Table 1.

Quarter-hours in the past						Minutes in qh			
96	**...**	**3**	**2**	**1**	Variable	**1**	**2**	**...**	**8**
Weight SI qh-96	..	Weight SI qh-3	Weight SI qh-2	Weight SI qh-1	SI	Weight SI min 1	Weight SI min 2	...	Weight SI min 8
Weight NRV qh-96	...	Weight NRV qh-3	Weight NRV qh-2	Weight NRV qh-1	NRV	Weight NRV min 1	Weight NRV min 2	...	Weight NRV min 8
Weight GUP qh-96	...	Weight GUP qh-3	Weight GUP qh-2	Weight GUP qh-1	GUP	Weight GUP min 1	Weight GUP min 2	...	Weight GUP min 8
Weight GDV qh-96	...	Weight GDV qh-3	Weight GDV qh-2	Weight GDV qh-1	GDV	Weight GDV min 1	Weight GDV min 2	...	Weight GDV min 8
Weight PPOS qh-96	...	Weight PPOS qh-3	Weight PPOS qh-2	Weight PPOS qh-1	PPOS	-	-	...	-
Weight PNEG qh-96	...	Weight PNEG qh-3	Weight PNEG qh-2	Weight PNEG qh-1	PNEG	-	-	...	-
Weight LOAD qh-96	...	Weight LOAD qh-3	Weight LOAD qh-2	Weight LOAD qh-1	LOAD	-	-	...	-
Weight WIND qh-96	...	Weight WIND qh-3	Weight WIND qh-2	Weight WIND qh-1	WIND	-	-	...	-

Table 1 Weights in the multivariate autoregressive model, note that Load and Wind include both intra-day forecast and actual production terms

Note that the model developed in 2019 considered a look-back horizon of 96 qh and variables families. Further research has been conducted in 2021 and is presented in this document section.

2.2 Variable selection

The model used in the 2019 proof-of-concept was an autoregressive model. This model included variables at quarter-hour granularity (SI, NRV, GDV….) and variables at minute granularity (SI, R2…). If we consider the values of each variable at different timeframes as individual model features, then we can consider that the model was using more than a thousand variables.

In order to use more sophisticated model, such as Support Vector Machines (SVM), or Artificial Neural Networks (ANN) it is necessary to select wisely the features to include in the model. Indeed, SVM and ANN are computationally intensive, and cannot be use practically with the thousands of variables used in the baseline autoregressive model. Of course, this feature selection can also benefit the baseline autoregressive model.

In this study, the variables at quarter-hour granularity have been included up to 96 quarter-hours back in the past (one day) with respect to the quarter-hour to forecast. The variables at minute granularity have been taken up to minute 8 in the quarter-hour, given this would be the most recent information available at the time an activation decision is made. When the activation lead-time would increase, less minute data would be available for the forecasting model.

2.2.1 Variable selection – correlation

In order to identify the feature importance, the correlation between each feature and the system imbalance has been calculated. The results are shown in Figure 2, it appears that the features at minute granularity present the highest correlation, in the vicinity or above 90%. They are followed by the System Imbalance at quarter-hour granularity.

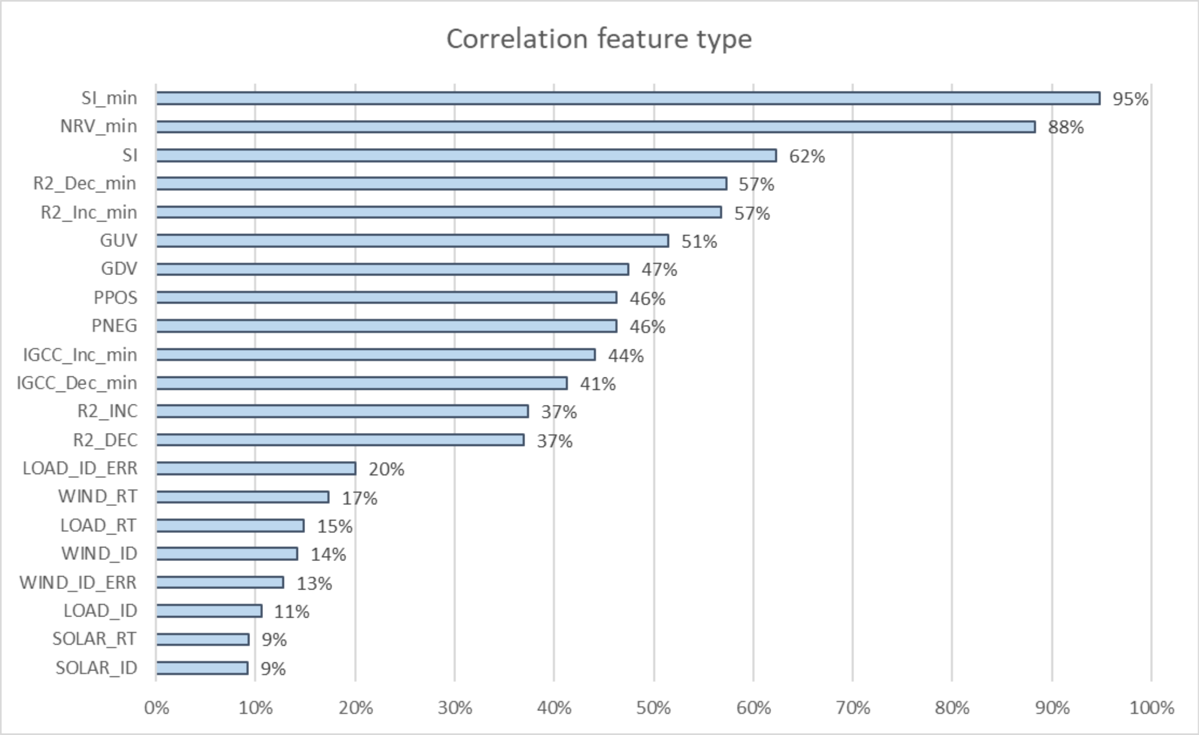


Figure 2 Correlation of S0, in absolute value, per feature

However, the correlation decreases significantly when looking further in the future (S1, S2 and further forecasts) as shown in Figure 3. This implies that it is possible to come with an accurate model for S0, but that for S1, S2 and further it will be more challenging.

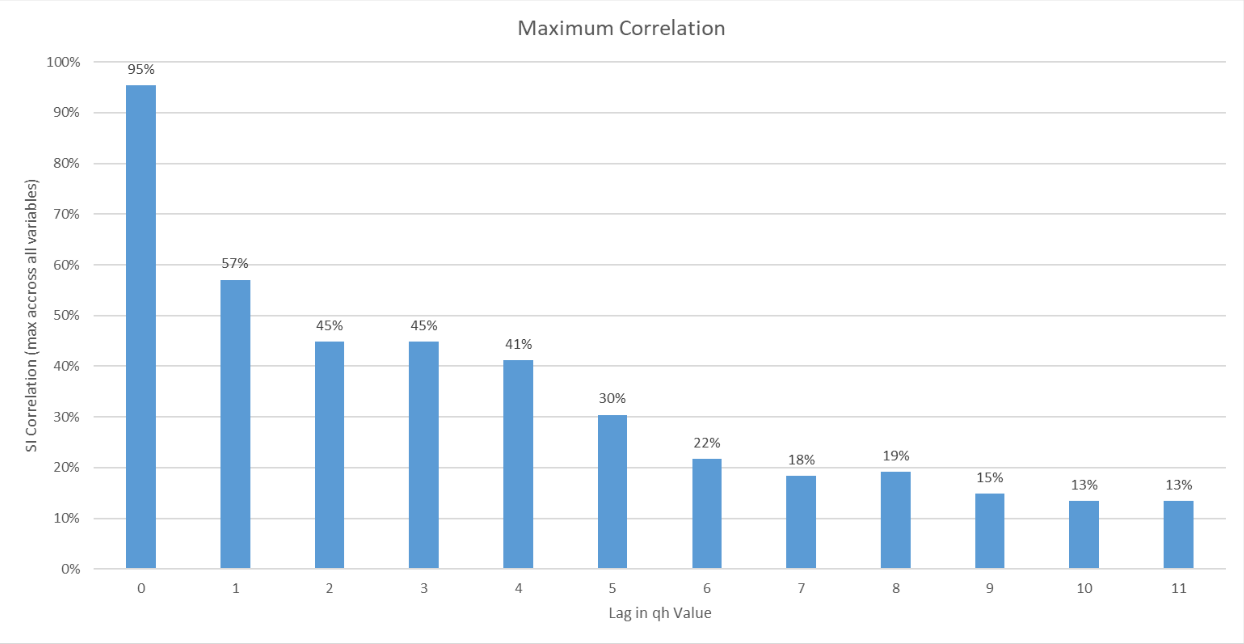


Figure 3 Correlation of SI at quarter-hour qh+k (with k from 0 to 11) with variables at quarter-hour qh

A description of the different variables used is provided in Table 2.

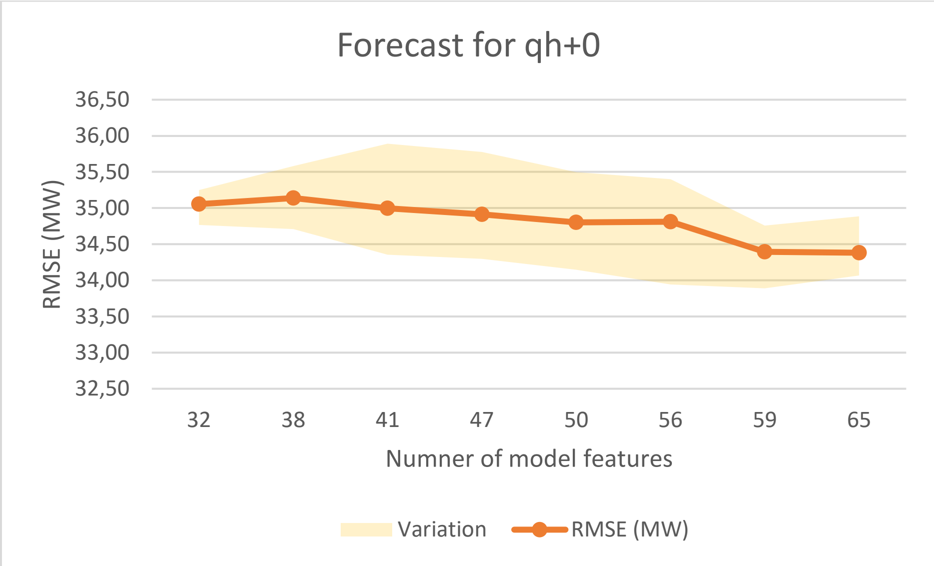
Variable Type	Description
SI	System Imbalance
GUV	Gross Upward Volume of activated reserves
GDV	Gross Downward Volume of activated reserves
PPOS	Price of Positive Imbalance (same as PNEG)
PNEG	Price of Negative Imbalance (same as PPOS)
LOAD_RT	Real-Time Total Load Estimation
LOAD_ID	Intraday Forecast of Total Load
LOAD_ERR	Error of intraday forecast of total load
WIND_RT	Real-Time total wind production estimation
WIND_ID	Intraday forecast of total wind production
WIND_ERR	Error of intraday forecast of total wind production
SOLAR_RT	Real-Time total Solar production estimation
SOLAR_ID	Intraday forecast of total solar production
SOLAR_ERR	Error of intraday forecast of total solar production
NRV	Net Regulating Volume
R2_Dec	Decremental R2 activation
R2_Inc	Incremental R2 activation
IGCC_Inc	International Grid Control Cooperation imports
IGCC_Dec	International Grid Control Cooperation exports
_min	Variable taken at minute granularity
NRV	Net Regulation Volume

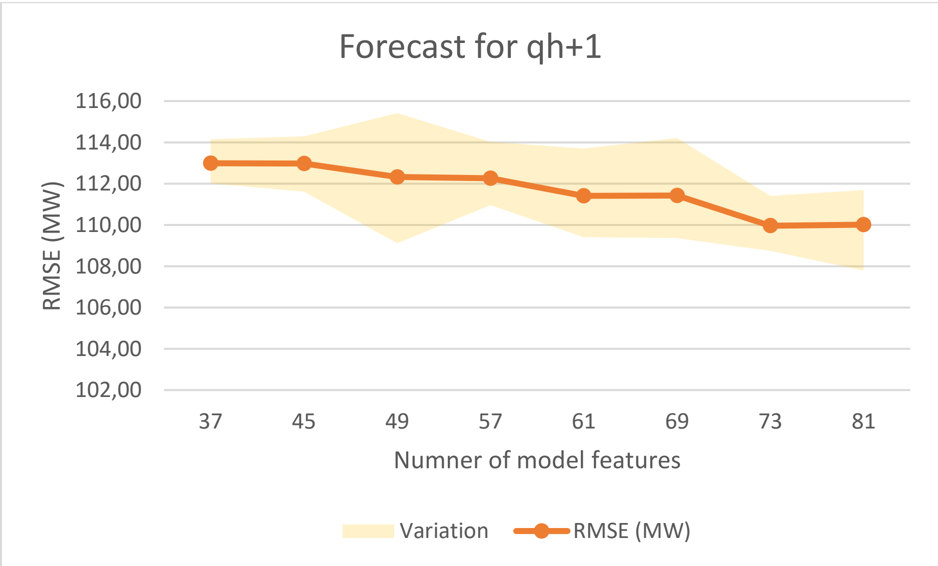
Table 2 Variable Types and description

2.2.2 Variable selection – combination of variables

In order to measure the influence of each variable type in the accuracy, and the impact of the look-back horizon, several simulations have been conducted.

In each of these simulations, a different combination of variable types (as described in Table 2) has been tested. For the case of a look-back of 4 quarter-hours, the results are provided in Figure 4. These results show that adding extra features to the model improves the accuracy marginally.





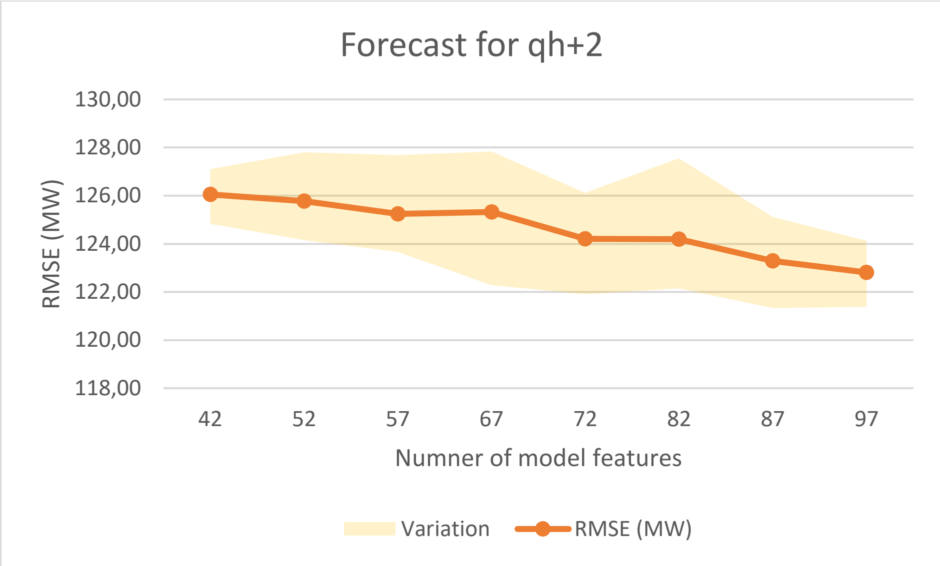


Figure 4 Model accuracy obtained for a look-back of 4qh and different combinations of variables

2.2.3 Variable selection – Look-back horizon

Too much historic data for training the model might decrease performance because trends of the past do not always persist. Too few historic data for training might also have a negative impact because the model might fail to pick up current trends. There is an optimum in the lookback horizon that we determined experimentally.

In the plot below the mean absolute percentage error of the linear model is plotted vs. the amount of training weeks it was given:

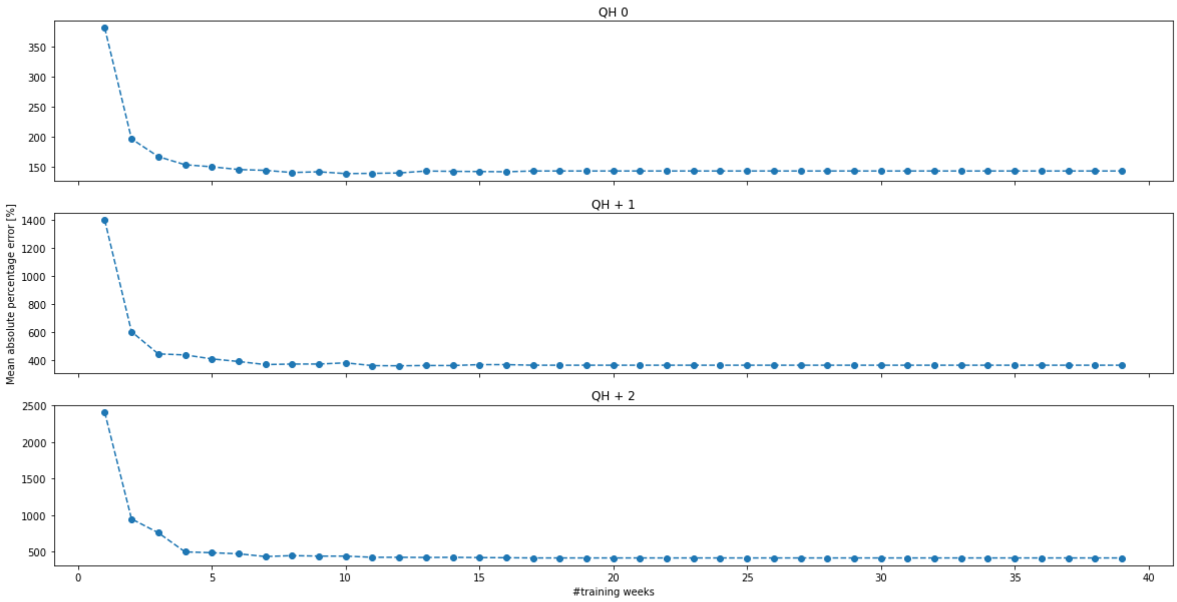


Figure 5 model error in function or training weeks

So the values on the x-axis are the size of the train set for the linear model, from 1 week of train data to 39 weeks of train data. The values on the y-axis are the performance of the model on the test set, which was 12 weeks of unknown data. We can see that after roughly 8 weeks of train data there is a plateau where the accuracy doesn't get better anymore if we add more weeks of training data. This is interesting, we can use only 8 weeks of train data and keep our accuracy while being computationally much less expensive than 39 weeks of train data. Therefore, the lookback horizon we will use for the historic data is 8 weeks.

There is a second horizon that is quite important for the models, which is the lookback horizon used during live predictions i.e. how many quarter hours do we look in the past while doing a live prediction?

In the graphs below you can see the results for some experiments with a linear model. On the x-axis the amount of quarter hours in the past is plotted, on the y-axis the RMSE:

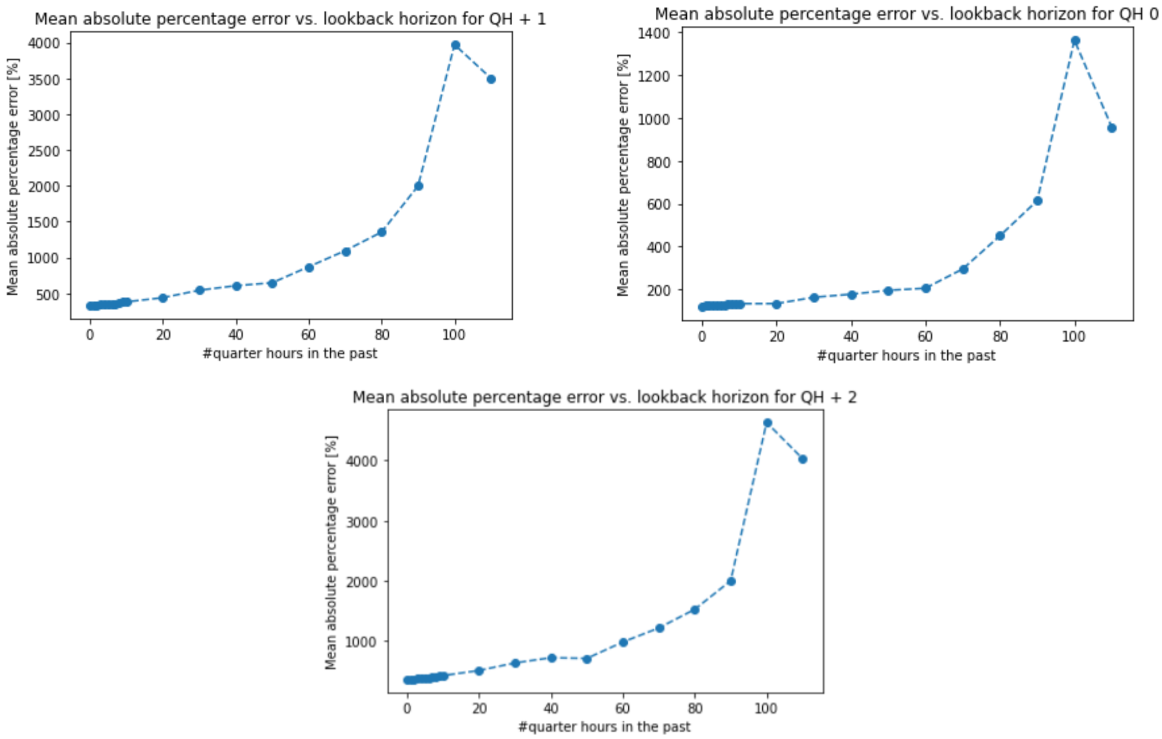


Figure 6 error in function of lookback horizon

The graphs show some interesting things. Firstly, the accuracy goes up until around 3 quarter hours of lookback horizon. Then the accuracy gets worse if we add more quarter hours in the past. The accuracy keeps going down until we add 96 quarter hours of data, then we see a dip in all three graphs. So adding data of exactly 1 day ago in the live predictions might increase accuracy. Otherwise looking only 3 quarter-hours in the past is the optimum for the accuracy of the model during live predictions.

2.2.4 Recursive Feature Extraction (RFE)

In previous subsection, an optimal variable selection has been provided by testing different combinations of variable types and different look-back time horizons.

In order to further optimize the variable selection, a statistical technique has been tested. This technique is called Recursive Feature Extraction (RFE). With this technique, different models are tested, starting with a model with all variables and then recursively extracting variables in order to find an optimum.

RFE was tested to check if further model improvement could be obtained, but did not provide any significant improvement to the variable selection procedure.

2.3 Machine Learning

2.3.1 Model Families

Different models have been used in this study to compare the performance obtained with each of them.
The first model type used is a **linear regression model**, this model basically implements the equation depicted in Figure 1 and is quite straightforward to use.

The second model tested are **Artificial Neural Network (ANN) model**. This model represents a significant complexity leap from linear models. ANN require several choices (such as the number of layers, neuron per layer, activation functions…). For this specific example, an Artificial Neural Network model, a single hidden layer architecture was chosen. The input layer has a neuron for each input variable) with an Exponential Linear Unit (ELU)[2] activation function. The hidden layer has half of the neurons of the input layer, using also an ELU activation function and the output layer has a linear activation function. This architecture is depicted in Figure 7

[2] Fast and accurate deep network learning by Exponential Linear Units (ELUS), Djork-Arne Clevert, Thomas Unterthiner & Sepp Hochreiter, 2016, 1511.07289.pdf (arxiv.org)

Elia Transmission Belgium SA/NV
Boulevard de l'Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium

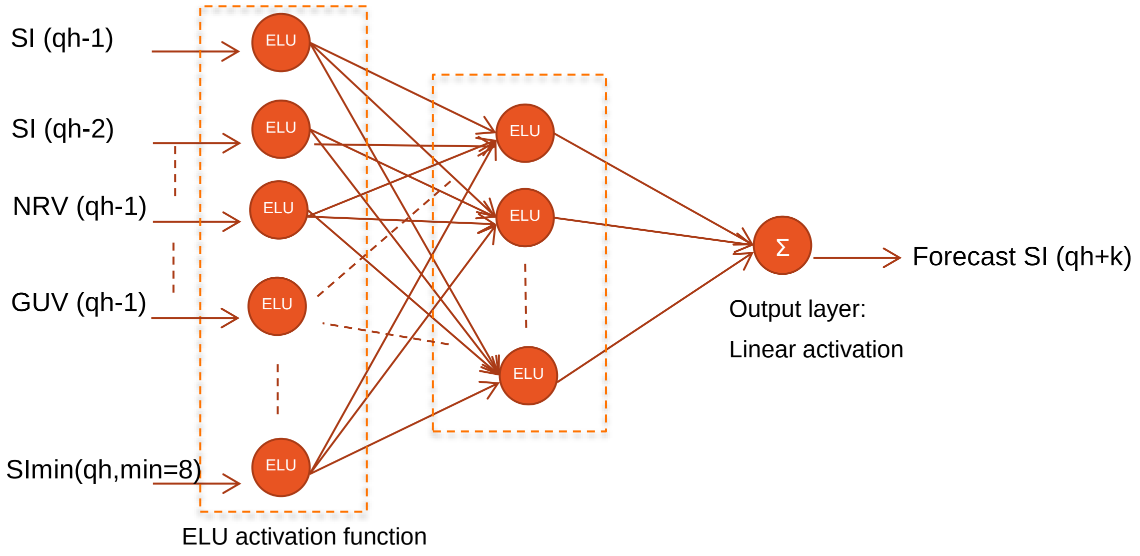


Figure 7 Artificial Neural Network Architecture

The third model type tested consist of a regression based on **Support Vector Machines (SVM).** These models are also very demanding in terms of computational power and require different choices to be made (radial function for instance). For this exercise, a kernel function was used as radial function.

A last model type was tested, consisting of **regression trees**, these models are rather simple in nature, comparable to linear regression.

2.3.2 Training techniques

In machine learning, a dataset is provided for training the model and estimating its accuracy. It is recommended, to avoid overfitting, to split the dataset into two distinct sets, a train set and a test set. The train set is used to train the model and the test set is used to evaluate the model performance on a set that was not used for the training.

In the machine learning literature, it is often recommended to follow a k-fold cross validation approach for producing statistically meaningful results on model performance. In a k-fold cross-validation, several train-test exercises are performed on random splits of the dataset into train and test. This is illustrated in Figure 8.

In this manner, k-fold cross-validation can highlight if the model performance varies according to the train/test split.

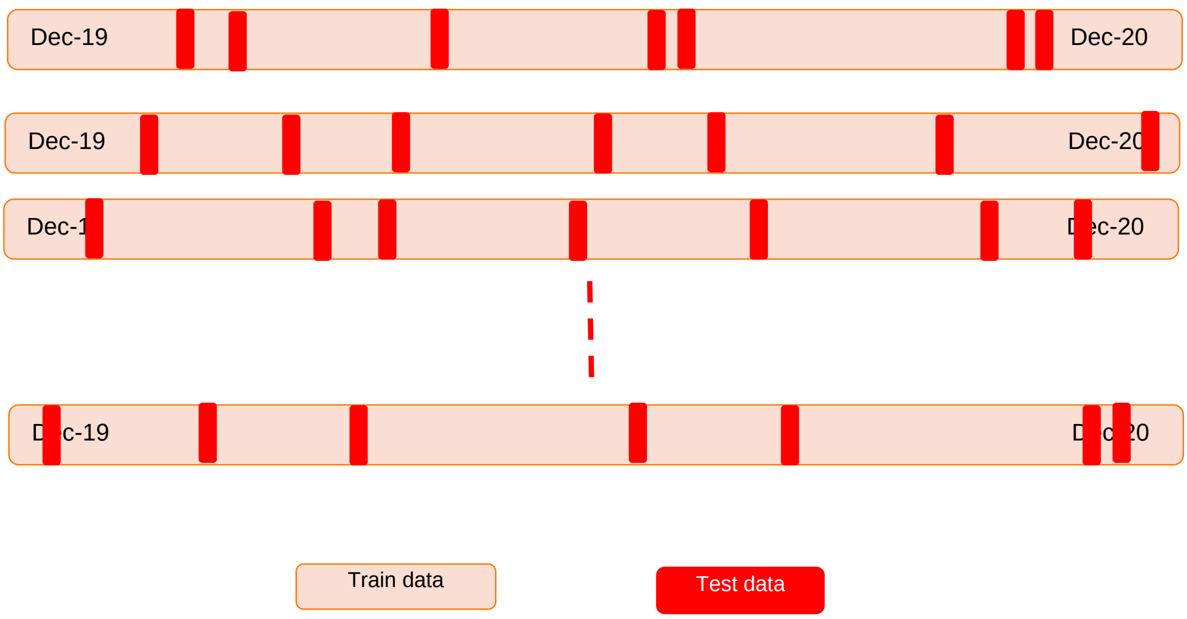


Figure 8 K-fold cross-validation on a yearly dataset

However, when it comes to time-series estimation, it is necessary to evaluate the model performance in circumstances close to operations. During operations, the model will obviously use data that is known that the moment the forecast is made. Therefore, a time-series estimation model has to be trained in such a way that only known data is used. Thus, using a k-fold cross-validation as-is, would train the model with data in the future and test it with data in the past. For instance, using 12 months of consecutive data as train data and a month of test data, just after the train data. This is shown in Figure 9:

12/19	01/20	02/20	03/20	04/20	05/20	06/20	07/20	08/20	09/20	10/20	11/20	12/20
Train Set												Test Set

Figure 9 Train/Test split using consecutive data for train and following data for test

It must also be noted that k-fold cross-validation is computationally intensive, as the train/test operation is repeated several times.

In order two asses on one hand the achievable model performance into operations and the statistical meaningfulness of the results, the two approaches have been followed in this study:

- Model evaluation with consecutive train and test periods
- Model evaluation with k-fold cross-validation, with 10 repetitions.

3. Selection of the data mining model

3.1 Introduction

The model performance evaluation is done using different statistical indicators, such as the RMSE, Max. Error or R-squared, these are described in Figure 10:

$$Absolute\ Error_i\ (MW) = \left| SI(qh+k)_i - SI\_Forecast(qh+k)_i \right|$$

$$Mean\ Absolute\ Error\ (MAE) = Average\ (Absolute\ Error)$$

$$Root\ Mean\ Squared\ Error\ (RMSE) = \sqrt{\sum_i Absolute\ Error_i\ (MW)}$$

$$P99 = 99th\ quantile\ of\ Absolute\ Error$$

$$Max\ Error = Max_i(Absolute\ Error_i)$$

$$R\ squared, \qquad R2 = \frac{covariance\ (SI(qh+k)\ \ , SI_{Forecast(qh+k)}\ )}{\sigma(SI(qh+k)\ \ \sigma(SI_{Forecast(qh+k)})}$$

$$Mean\ absolute\ percentage\ error\ (MAPE) = \frac{100}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right| \quad where\ At\ is\ the\ actual\ value\ and\ Ft\ the\ forecast$$

Figure 10 – Statistical indicators used to assess model performance

3.2 KPI evaluation

3.2.1 S0 Forecast: Model comparison with 3 quarter-hour lookback

The forecast of S0 (SI at the end of current quarter hour) is done in this case by feeding variables up to 3 quarter-hours in the past. The models are trained using consecutive train and test set.

In this case, it appears that the linear model is providing the best performance for most indicators, followed closely by the ANN model. The SVM and Tree model have significantly worse performance.

Model Type	MAE (MW)	RMSE (MW)	P99 (MW)	Max Error (MW)	R2
Linear	26,45	34,44	99,46	178,28	0,962
ANN	26,66	35,09	100,66	168,49	0,959
Regression Tree	30,86	42,45	129,57	457,15	0,941
SVM	32,71	69,70	172,42	1283,25	0,840

Figure 11 qh+0 Forecast with consecutive train/test and 4qh lookback

3.2.2 S0 Forecast: Model comparison at different lookback horizons

The forecast of S0 (SI at the end of current quarter hour) is done in this case with different lookback horizons, 3 and 12 quarter-hours. The models are trained using consecutive train and test set.

Two conclusions can be drawn from the results, first is that increasing the lookback horizon does not improve the results (for most indicators), and that in both cases (3 or 12 quarter-hour lookback) the linear model performance is better than the one from the other models.

Model Type	Look-back (qh)	MAE (MW)	RMSE (MW)	P99 (MW)	Max Error (MW)	R2
Linear	3	**26,45**	**34,44**	**99,46**	178,28	**0,962**
Linear	12	26,83	34,80	**99,46**	**163,77**	**0,962**
ANN	3	**26,66**	**35,09**	**100,66**	**168,49**	**0,959**
ANN	12	30,69	40,26	116,95	215,10	0,948

Figure 12 qh+0 Forecast with consecutive train/test for 4qh and 12qh lookback

3.2.3 S0, S1 and S2 Forecast: Model comparison at different forecast horizon

In this case, a model with 3 quarter-hour lookback is trained for different forecasting horizons:

- S0: SI at end of current quarter-hour
- S1: SI at next quarter-hour
- S2: SI two quarter-hours in the future.

The results in Figure 13 show different outcomes. First, the linear model remains the model with the best performance at the different forecasting horizons, with the ANN model presenting marginally worse performance. In this case, the SVM and Tree model performance is omitted for clarity. Second, the forecast performance worsens when looking further into the future. For instance, the RMSE goes from 34,44MW to 122,91MW in the case of the linear model when changing the forecast horizon from S0 to S1.

Forecast Horizon qh in the future	Model	MAE (MW)	RMSE (MW)	P99 (MW)	Max Error (MW)	R2
0	Linear	26,45	34,44	99,46	178,28	0,962
0	ANN	26,66	35,09	100,66	168,49	0,959
1	Linear	88,82	122,91	382,34	1209,87	0,518
1	ANN	90,26	124,90	385,47	1256,21	0,502
2	Linear	102,34	141,14	411,62	1322,37	0,365

2	ANN	103,26	142,44	426,87	1363,32	0,353

Figure 13 Forecast with consecutive train/test and 4qh lookback for different forecasting horizons

3.2.4 S0, S1 and S2 Forecast: Model comparison at different forecast horizon and different variables. K-fold cross validation

In this case, a model with 3 quarter-hour lookback is trained for different forecasting horizons:

- S0: SI at end of current quarter-hour
- S1: SI at next quarter-hour
- S2: SI two quarter-hours in the future.

Furthermore, different variables combinations (Figure 2) are tested and each combination is trained using a k-fold cross-validation. As explained previously, a k-fold cross-validation has to be interpreted carefully, as technically speaking, the models are being trained with data in the future from the test set.

The results in Figure 13 show different outcomes. First, the linear model remains the model with the best performance at the different forecasting horizons, with the ANN model presenting marginally worse performance. In this case, the SVM and Tree model performance is omitted for clarity. Second, the forecast performance worsens when looking further into the future. For instance, the RMSE goes from around 35MW to around 105MW in the case of the linear model when changing the forecast horizon from S0 to S1.Note that these results are different from the ones in Figure 13. This difference can be explained by the fact that the train and test set are different from the one used to produce the results in Figure 13

In Figure 14, Figure 15 and Figure 16 the results for S0, S1 and S2 are given respectively. The conclusions from previous sections remain valid: the linear model is providing the best forecasting performance, and the forecasting performance decreases significantly with the forecasting horizon. However, these plots show that the performance can vary depending on the variable set used and the fold performed, in particular the max error. This difference with the max error might be due to particular situations in the test set.

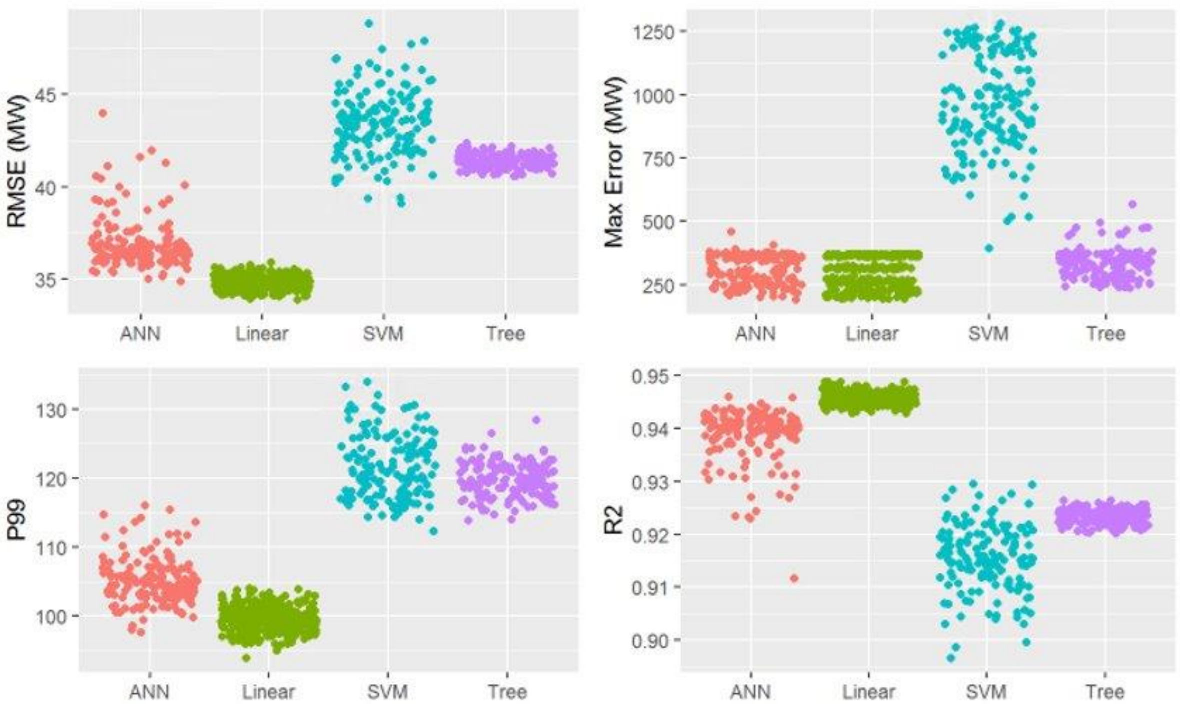


Figure 14 qh+0 Forecast with k-fold cross-validation and different variable combination

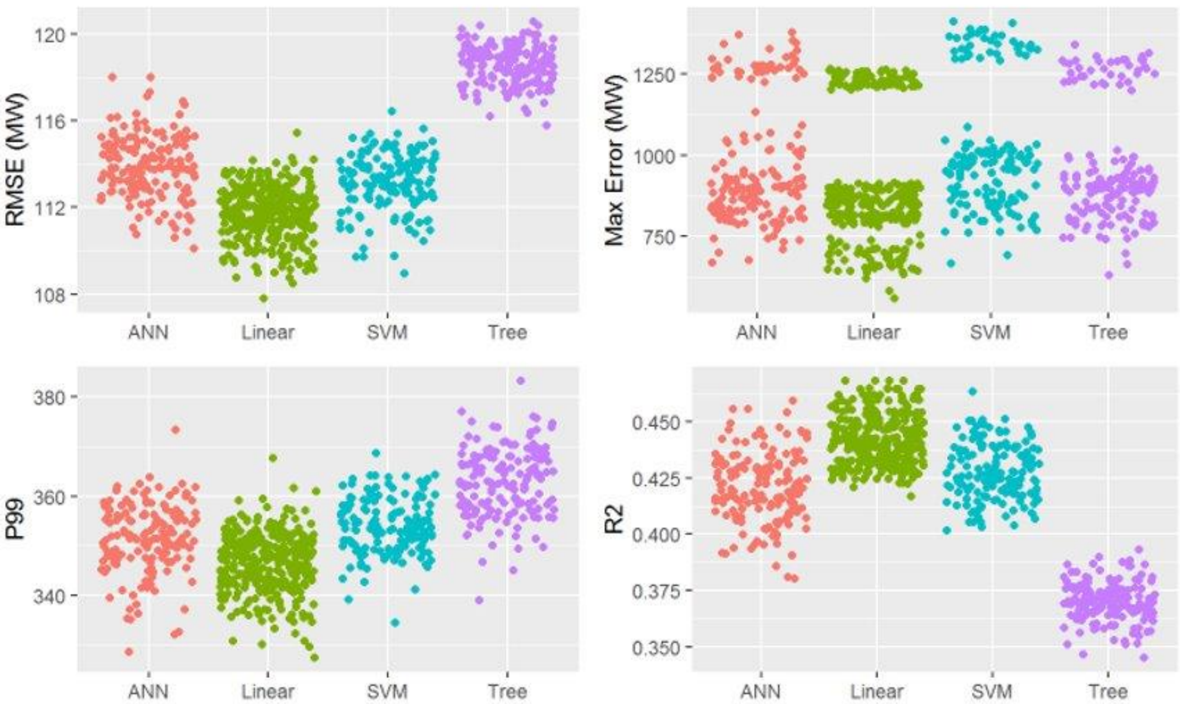


Figure 15 qh+1 Forecast with k-fold cross-validation and different variable combination

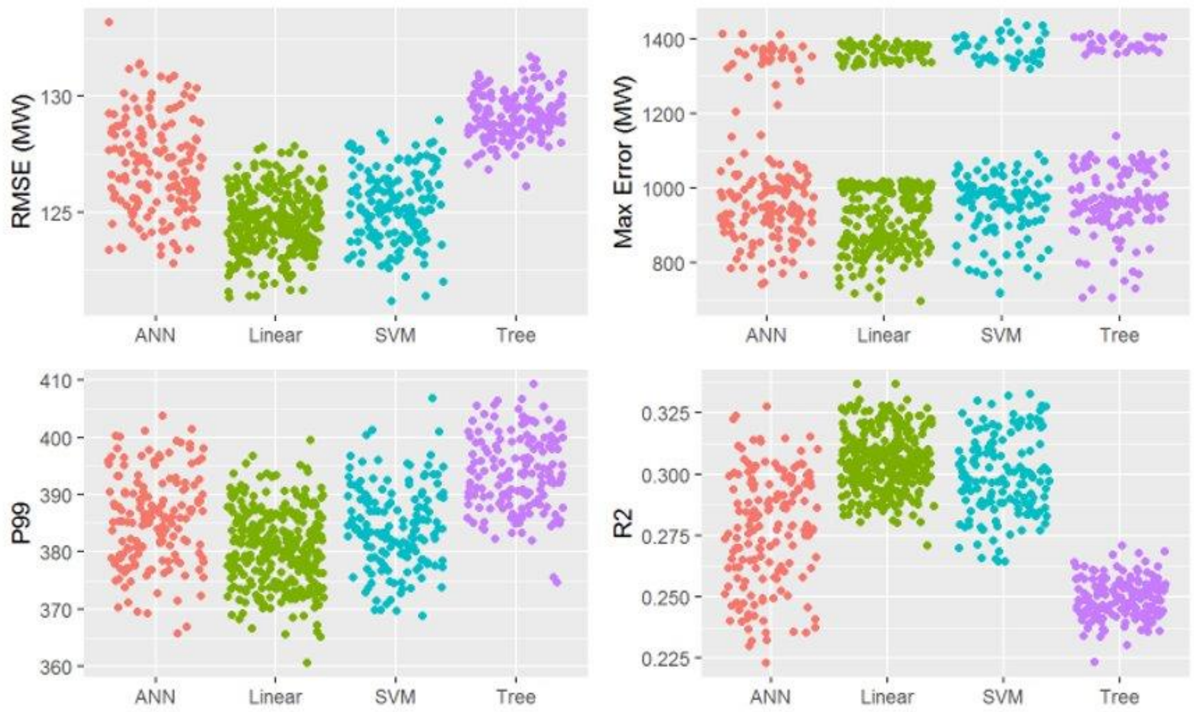


Figure 16 qh+2 Forecast with k-fold cross-validation and different variable combination

3.3 Model performance conclusions

Significant research has been done in the area of System Imbalance forecasting. The following topics have been addressed:

- Variable selection
- Look-back horizon analysis
- Machine learning model type comparison
- Train/Test split using consecutive data
- Train/Test split with k-fold cross-validation

The results show that the linear model has the best performance, despite its simplicity, and this at all forecasting horizons (S0, S1 and S2). It also appears that extending the look back horizon does not improve the model performance. The k-fold cross-validation has shown that the model performance is contained in a limited range, with the exception of the max error that can present significant variations (probably due to particular quarter-hours present in the test set).

3.3.1 Future improvements

Future improvements to the model are currently being discussed. For instance, the model could use all available intraday information, not only the forecast but also the nominations, in particular the cross-border nominations. Furthermore, these intraday nominations are known in advance, at quarter-hour qh, the nominations of at least qh+3 are known. This implies that it would be feasible to include these intraday nominations for following quarter-hours in the model.

3.4 Evaluation of the selected model under different conditions

To ensure the validity of the model under all circumstance, the performance is evaluated under specific conditions, which might occur in real time (such as storm situation, outages, large SI). In order to capture a sufficient amount of events the evaluated period was extended to cover the period of August 2020 until May 2021.

3.4.1 Storm events

A notification is generated when the forecasting tool detects a Storm event in the North Sea in the next 36 hours. The total storm impact in terms of wind power generation drop and the timing of the storm are published. Over the course of the evaluation period 14 days with storm events were observed. The analysis focusses on the day of the storm event as published by Elia and not only the forecasted time interval to cover potential forecasting error (change in timing).

Figure 17 shows the difference between the forecasted and actual wind production on September 25 and 26 2020. Two elements might impact the quality of the SI forecast, first the significant forecasting error and second the large swings in wind generation. Especially the second effect translated into large swings of the system imbalance as illustrated in Figure 18.

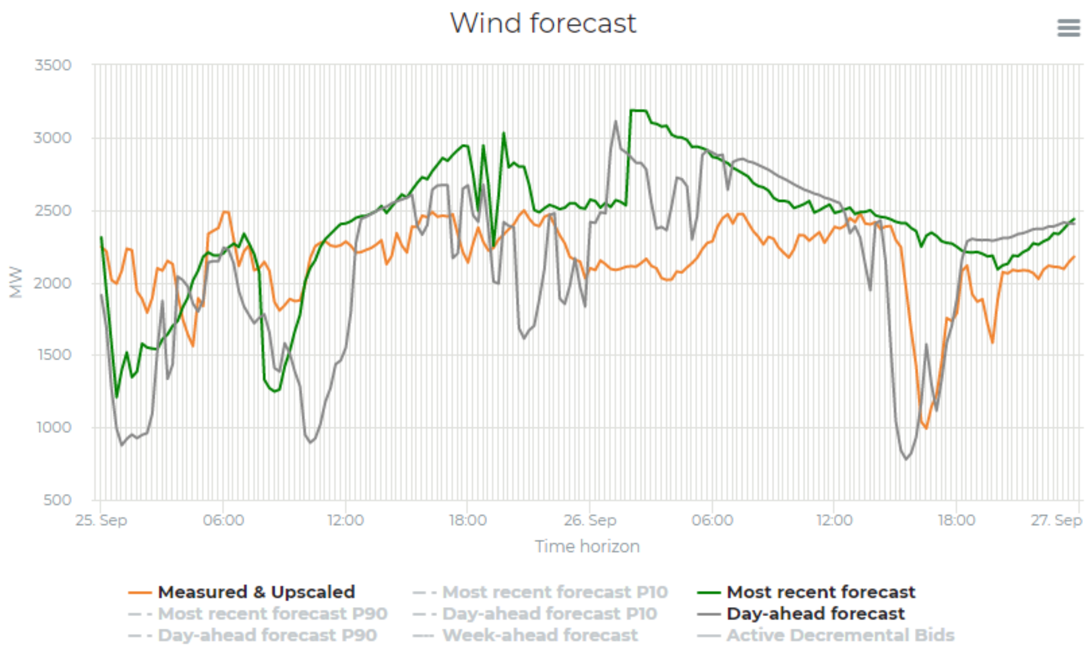


Figure 17 wind forecast during storm event on 25 and 26/09/2020

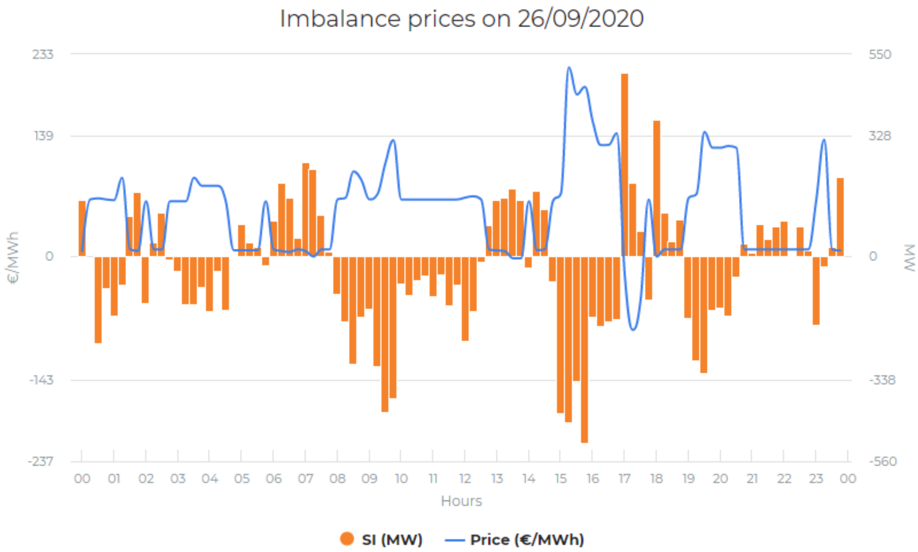


Figure 18 System imbalance and prices for September 26th 2020

By calculating the KPIs on the days where a storm event was detected we see that for the ongoing Qh (Qh0) the performance indicators are very similar. For the future quarter hours, we see a slight deterioration of the performance indicators, especially for the R^2 and RMSE. Given the limited amount of timestamps, comparing the Max error and the P99 is less relevant. Overall, the linear model is able to handle storm events comparatively well. We see no strong deterioration of the relevant performance indicators.

	All timestamps				During storm events		
	QH0	QH1	QH2		QH0	QH1	QH2
MAE (MW)	23	91	104	MAE (MW)	26	104	121
RMSE	30	124	140	RMSE	34	140	159
P99 (MW)	86	389	431	P99 (MW)	95	423	503
Max Error (MW	252	1274	1456	Max Error (MW)	252	605	725
R2	0,97	0,47	0,32	R2	0,97	0,42	0,26
Count	28603	28604	28603	Count	768	768	768

Table 3 Comparison of KPI during storm events

3.4.2 High system imbalance and system imbalance ramps

A second situation where the model might fail to predict the system imbalance accurately is during high system imbalances or during strong change of system imbalance. To evaluate these situations the dataset is reduced to events where the absolute system imbalance exceeds 300 MW or where the difference between two subsequent quarter-hours exceeds 300 MW. The threshold of 300 MW was selected to ensure sufficient data points were available for the analysis.

Comparing the KPIs of Table 3 and Table 4 it is clear that indeed the relative error increases in high SI and high SI ramp situation (as seen by an augmentation of the MAE and RMSE). However, the R^2 value improve significantly for all quarter-hours.

SI > 300 MW	QH0	QH1	QH2
MAE (MW)	30	219	280
RMSE	39	256	312
P99 (MW)	115	608	685
Max Error (MW)	252	1274	1456
R2	0,99	0,72	0,59
Count	2328	2329	2329

High SI ramp (>300 MW between Qhs)	QH0	QH1	QH2
MAE (MW)	31	298	239
RMSE	40	327	285
P99 (MW)	113	656	663
Max Error (MW)	167	1274	1456
R2	0,98	0,70	0,38
Count	536	536	536

Table 4 Performance KPIs during high SI and high SI ramp situations

Further exploring the distribution of forecasting error vs the observed SI in Figure 19 shows a relatively stable minimum, average and maximum error over all observed SI values (keeping in mind the few observations at the tails) for Qh0. Looking at the same distribution for Qh2, the spread of error remains similar. However, a bias of the SI forecasting error in the direction of the SI can be observed. This would indicate that the linear model is not able to well forecast large SI for future quarter-hours. This observation is confirmed in Figure 20 where the SI forecasts are plotted against a load duration curve of the SI.

Figure 19 Forecasting error vs distribution of SI

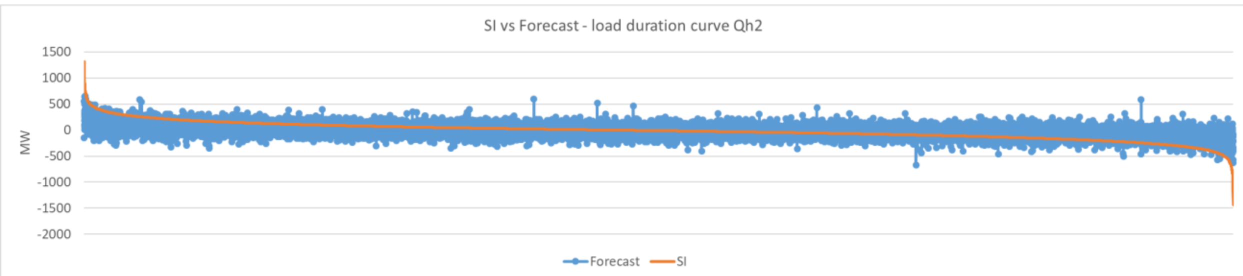


Figure 20 SI load duration curve and SI forecast

3.4.3 Impact of outages on SI forecast quality

A final element in the evaluation of the model was the performance during an outage of a production unit. This assessment was done using the data published (for the considered time period) on the Elia website[3] and only considering forced outages of generation units.

Figure 21 shows the distribution of the SI for Qh0 (current quarter-hour) and indicates whether are not a Forced Outage (FO) has taken place during that quarter hour. From the graph, one could not derive the volume of generation that was lost during the outage. Although the share of quarter hours with forced outages is relatively higher at the tails of the SI distribution curve, the absolute number is low. Hence, high SI events may happen together with a FO, but this situation is rather rare. Furthermore, outages are observed in almost the entire distribution of the SI.

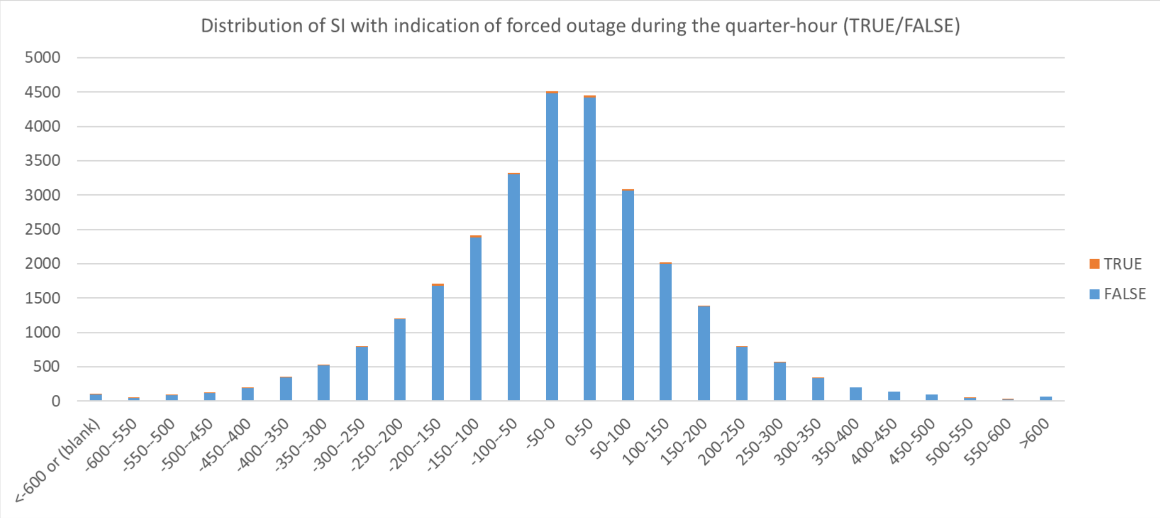


Figure 21 Distribution of SI for QH0 indicating a Forced Outage during the Qh

The number of outages is quite limited (only 257 quarter-hours were observed in the dataset), hence the KPIs shown in Figure 22 are of questionable statistical relevance. When evaluating the model using the KPI, no strong change in performance is observed from splitting the outage situation from the no-outage situation when it comes to the forecasting of the SI. Even considering whether or not an outage has taken place in the previous 4 quarter-hours (to capture the effects of an outage at the end of the quarter hour or long lasting effects), no strong changes in KPIs can be observed.

[3] https://www.elia.be/en/grid-data/power-generation/planned-and-unplanned-outages

25

All timestamps

	QH0	QH1	QH2
MAE (MW)	23	91	104
RMSE	30	124	140
P99 (MW)	86	389	431
Max Error (MW)	252	1274	1456
R2	0,97	0,47	0,32
Count	28603	28604	28601

Outage during current Qh

	QH0	QH1	QH2
MAE (MW)	22	110	131
RMSE	29	149	176
P99 (MW)	82	488	544
Max Error (MW)	121	552	814
R2	0,98	0,43	0,21
Count	257	257	257

Outage during past h

	QH0	QH1	QH2
MAE (MW)	24	104	123
RMSE	31	140	162
P99 (MW)	86	472	518
Max Error (MW)	161	763	814
R2	0,98	0,49	0,33
Count	1002	1002	1002

no outage during current Qh

	QH0	QH1	QH2
MAE (MW)	23	91	104
RMSE	30	124	140
P99 (MW)	86	388	429
Max Error (MW)	252	1274	1456
R2	0,97	0,47	0,32
Count	28346	28347	28344

no outage during past h

	QH0	QH1	QH2
MAE (MW)	23	91	103
RMSE	30	123	140
P99 (MW)	86	387	426
Max Error (MW)	252	1274	1456
R2	0,97	0,47	0,32
Count	27601	27602	27599

Figure 22 KPIs considering Forced Outages

3.5 Conclusion

Elia proposes to use the linear model for the forecasting of the System Imbalance. The model has proven to be the most accurate of the different models. It relies on simple concepts avoiding any potential black box effect. The linear model was evaluated in different situations such as storm events and was proven robust.

Q1: The analysis resulted in the selection of the linear regression model. Can you support this choice? Please motivate your answer.

Q2: Do you see other elements which could increase the performance of the model?

4. Qualitative assessment of publishing a SI forecast

4.1 Publication proposal

Elia proposes to publish the forecast of the SI for Qh0, Qh+1 and Qh+2 under the condition that the quality of the model is sufficient (e.g. obtaining an RMSE below 100 MW). The publication would be accompanied with the necessary quality indicators and legal disclaimers on the quality and liability of the publication. It will be made clear that Elia puts this information at the disposal of the market parties in an effort to increase market's ability to better balance the energy system, but that under no circumstances this information implies a shift in responsibility or liability towards Elia.

Given the current performance of the model, the forecast for Qh0 is sufficient to merit the publication of the forecasted value. The publication of the forecast for Qh1 and Qh2 could either consist of an exact value (expressed in MW) or as a range (e.g. [150 MW-300 MW]) combined with a confidence interval. Elia would make available the relevant information such that the model can be reproduced by external stakeholders.

The data will not be included in the balancing dashboard but made available via the Elia open data platform in the form of an API. Data used in the model, but available via other API's would not be included in the publication. The publication will thus contain the following information:

- Date
- Quarter-hour
- SI (actual): Ex-post measured SI (when available)
- SI Forecast qh+0: Forecast of SI made during current quarter-hour (exact value)
- SI Forecast qh+1: Forecast of SI made during quarter-hour-1 (exact value)
- SI Forecast qh+2: Forecast of SI made during quarter-hour-2 (exact value)

Depending on feedback from market parties the publication of categorical predictions (with confidence interval) could be added. In such a publication, the model would not predict an exact value, but rather it would indicate a certain range of SI (e.g. 50-150MW). A first analysis has shown that these categorical predictions are mainly relevant for Qh+1 and Qh+2.

> Q3: Do you agree with the proposed publication horizon and publication format
> Q4: Would you prefer the publication of the exact forecasted value or categorical predictions for Qh+1 and Qh+2.

4.2 Elia evaluation of publication

Elia sees a number of advantages for the publication of the SI forecast when the model is considered to be sufficiently accurate. First, the publication of an accurate system imbalance forecast might enable a higher implicit reaction from market parties in the balancing markets. Secondly, we see that the forecasting quality of the current quarter hour and the use of a simple linear model will allow a higher transparency on the main drivers of the SI, this would in turn allow market parties to better balance their portfolio. Finally, after an evaluation period the publication would allow an ex-post analysis on the implicit reaction of market parties on this publication.

At this time, Elia does not see strong concerns against the publication of the SI forecast, but a number of smaller points were identified. First, we can observe that the reactive balancing model in Belgium seems to work well given that the NRV and SI has remained relatively stable even with the high increase in RES. This is somewhat confirmed by the lack of strong correlation of RES forecasts and the SI observed in this study. Hence, it is yet to be seen whether market parties will react to the proposed publication. Second, assuming an accurate forecast for the future quarter-hours – in combination with the available ATCs, a market party could derive whether or not they will have a domination market position in the balancing timeframe.

In terms of REMIT impact, Elia does not see the publication of the forecasted SI as an insider information. We only see the obligation to inform the market when the publication fails (outage of the tool).

Q5: Do you believe the publication of the SI forecast is relevant? Please motivate your answer.
Q6: Should Elia withhold the publication if a certain quality level cannot be reached? What do you believe is the right threshold (e.g. RMSE < 100 MW).
Q7: Elia sees no strong concerns for the publication of the SI forecast, do you agree with this evaluation.
Q8: Would you see an impact on the market as result of the publication, and which one?

4.3 Approach on the implementation plan proposal

The definition of the incentive requires the inclusion of a draft implementation plan in this public consultation. Following an alignment with the CREG, Elia would rather await the feedback on the relevance and format of the publication before drafting an implementation plan. Hence, Elia proposes to present a draft implementation plan later this year during the WG balancing.

In 2022 and 2023 several significant changes to the balancing timeframe are foreseen (dates are indicative and subject to change):

- Test phase for the relaxation of the Day Ahead balancing obligation (end of 2021 until mid-2023)
- Go-live of the PICASSO project (Q2 2022)
- Go-live of the MARI project (Q4 2022)

Any of the above constitutes a paradigm shift in how the balancing timeframe is operated; this in turn might affect the SI quality and behaviour of market parties. If not chosen well, the start of the SI forecast publication could overlap with the go-live or implementation of the before mentioned projects, making it difficult to assess the impact of the publication on the SI.

Q9: Do you recognize the dependency between the publication of the SI forecast and other changes to the balancing timeframe? Do you see others? Q10: Elia believes the start of the publication should not overlap with other major changes, do you agree?

Elia Transmission Belgium SA/NV
Boulevard de l'Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium

Annex 1: Description of the model used during the testing phase

During the testing phase scheduled for later this year, the following indicators will be used for the prediction of the System Imbalance using the linear model described in this report. In preparation of this testing phase, additional indicators might be added. The model used during the testing phase will be included in the final report.

	Wind power	Wind powerIntraDay	Solar IntraDayForecast	Solar Measurement	SI_Cumulative	SI_Instantaneous	NRV_Cumulative	NRV_Instantaneous	IGCC_I	IGCC_D	SI	GUV	GDV	PPOS / PNEG	R2_I	R2_D
Wind power	100%	99%	-19%	-18%	5%	6%	-5%	-5%	2%	-1%	12%	-3%	14%	-11%	-3%	17%
Wind powerIntraDay	99%	100%	-19%	-18%	5%	5%	-4%	-5%	2%	-1%	11%	-3%	14%	-11%	-2%	16%
Solar IntraDayForecast	-19%	-19%	100%	99%	5%	5%	-5%	-6%	0%	1%	6%	3%	13%	-3%	0%	10%
Solar Measurement	-18%	-18%	99%	100%	7%	8%	-7%	-8%	-1%	2%	9%	1%	16%	-6%	0%	11%
SI_Cumulative	5%	5%	5%	7%	100%	92%	-92%	-84%	-48%	42%	43%	-32%	34%	-31%	-14%	22%
SI_Instantaneous	6%	5%	5%	8%	92%	100%	-90%	-90%	-52%	46%	40%	-30%	33%	-30%	-13%	21%
NRV_Cumulative	-5%	-4%	-5%	-7%	-92%	-90%	100%	95%	54%	-48%	-39%	32%	-34%	31%	11%	-18%
NRV_Instantaneous	-5%	-5%	-6%	-8%	-84%	-90%	95%	100%	57%	-51%	-36%	29%	-33%	30%	10%	-17%
IGCC_I	2%	2%	0%	-1%	-48%	-52%	54%	57%	100%	-25%	-24%	27%	-11%	13%	4%	-8%
IGCC_D	-1%	-1%	1%	2%	42%	46%	-48%	-51%	-25%	100%	19%	-7%	26%	-8%	-2%	4%
SI	12%	11%	6%	9%	43%	40%	-39%	-36%	-24%	19%	100%	-74%	68%	-73%	-51%	51%
GUV	-3%	-3%	3%	1%	-32%	-30%	32%	29%	27%	-7%	-74%	100%	-18%	71%	65%	-17%
GDV	14%	14%	13%	16%	34%	33%	-34%	-33%	-11%	26%	68%	-18%	100%	-42%	-12%	64%
PPOS / PNEG	-11%	-11%	-3%	-6%	-31%	-30%	31%	30%	13%	-8%	-73%	71%	-42%	100%	52%	-29%
R2_I	-3%	-2%	0%	0%	-14%	-13%	11%	10%	4%	-2%	-51%	65%	-12%	52%	100%	-14%
R2_D	17%	16%	10%	11%	22%	21%	-18%	-17%	-8%	4%	51%	-17%	64%	-29%	-14%	100%

Elia Transmission Belgium SA/NV
Boulevard de l'Empereur 20 | Keizerslaan 20 | 1000 Brussels | Belgium