Renewable Energy Production Forecast

PROJ0016 - Big Data Project

Yann Claes, Gaspard Lambrechts and François Rozet

May 11, 2020

University of Liège

"How much renewable energy will be produced tomorrow in Liège ?"

Approach

We divided our approach in three branches :

1. Assessing the number and location of photovoltaic units;
2. Estimating the photovoltaic energy production;
3. Estimating the wind energy production.

Photovoltaic panels mapping

Approach

We couldn't find appropriate records of photovoltaic installations in Liège.

We decided to *estimate* the number and location of photovoltaic panels using remote sensing, in particular using aerial or satellite imagery.

Wallonia aerial images

Wallonia possesses a geographic information website WalOnMap. It uses the Web Map Tiles Service standard.

Using `owslib` Python library we could retrieve these tiles (512×512 images).

Unfortunately, the images have a quite low quality.

Detection model – U-Net

We decided to use a tweaked version of U-Net (Ronneberger, Fischer, and Brox 2015), a well known segmentation network.

We implemented it and trained it using the `PyTorch` Python library.

This part of the project was conducted in parallel with our project of *Deep Learning*. Therefore, we won't explain the implementation and training process (algorithm, loss function, etc.) in this presentation.

Training set

We found a dataset (Bradbury et al. 2016) of over 19 000 solar panels across 601 high resolution (5000×5000) images from four cities in California.

Using the coordinates (provided as polygon vertices), we produced segmentation masks in order to train our model.

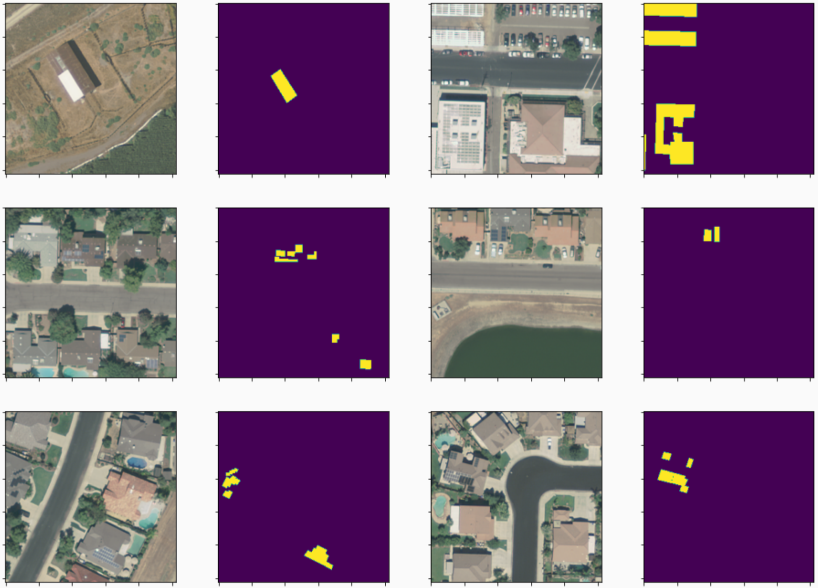


Figure 1: Training set images sample with their masks.

Data augmentation

The training images are very different (colorimetry, blurriness, etc.) from the images of WalOnMap.

We decided to augment our training set. Our approach is to randomly apply transformations to the images *while training*.

Transformations : rotations, flips, brightness, contrast, smoothing, sharpening, etc.

Figure 2: Training set augmentation sample.

Results – California

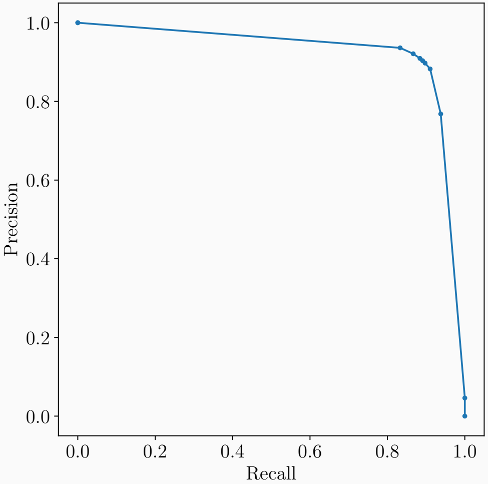


Figure 3: Precision-recall curve on the Californian test set.

Californian results

Our model :

$$\text{average precision} = 92.51\,\%$$
$$\text{precision}(0.5) = 90.3\,\%$$
$$\text{recall}(0.5) = 89.1\,\%$$

Californian results

Our model :

$$\text{average precision} = 92.51\,\%$$
$$\text{precision}(0.5) = 90.3\,\%$$
$$\text{recall}(0.5) = 89.1\,\%$$

Malof et al. (2016) :

$$\text{precision}(?) = 80\,\%$$
$$\text{recall}(?) = 72\,\%$$

Fine tuning

However, the results on WalOnMap were not satisfactory.

Therefore, we fine-tuned our model for a few more epochs on 550 WalOnMap images we hand-annotated.

111 other images are kept for metric evaluation.

Fine tuning

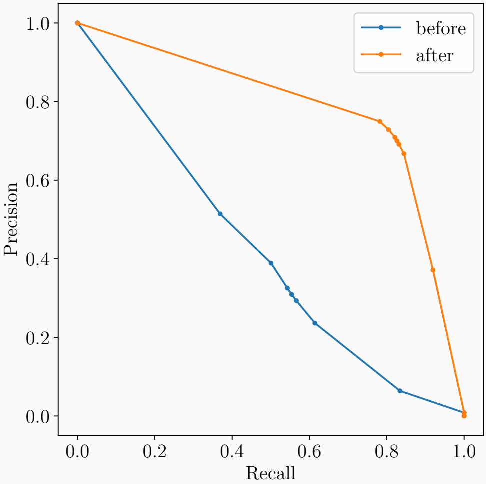


Figure 4: Precision-recall curve before and after fine-tuning.

Fine tuning

$$\text{average precision} = 78.3\,\%$$
$$\text{precision}(0.5) = 71.3\,\%$$
$$\text{recall}(0.5) = 82.6\,\%$$

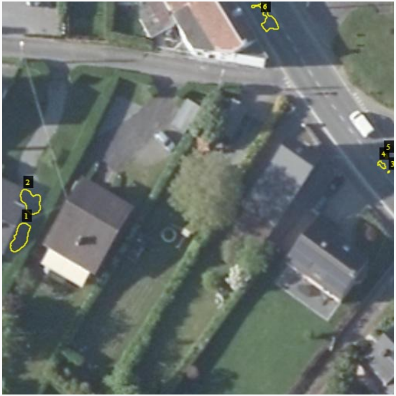
Mapping

Finally, we applied our fine-tuned model on the 846 703 tiles of the Province of Liège. After 1.5 d we obtained the location and shape (polygons) of 64 463 PV installations covering 2 554 505 m^2 of land.


1
2
3
4
5
6


1
2
3
4


2
1


6
2
1
5
4
3

Binding

As such, the locations and shapes are not usable for the power production model. Therefore, we summarized each detected installation into 3 quantities :

- its geospatial location (latitude and longitude);
- its surface;
- its azimuth angle.

Photovoltaic production

Approach

As specified by the guidelines, we opted for a short-term prediction task.

3 possibilities (Nespoli et al. 2019):

- Statistical methods: regression, artificial neural networks
- Physical methods
- Hybrid methods: mix of statistical and physical

Approach

As specified by the guidelines, we opted for a short-term prediction task.

3 possibilities:

- Statistical methods: regression, artificial neural networks
- Physical methods
- Hybrid methods: mix of statistical and physical

Our main idea was to rely on physical intuition.

$\Rightarrow$ Build simple physical models.

Approach

Build a *panelwise* model that we would scale up $\Rightarrow$ focus on a small scale first (ULiège) and assess the obtained results.

Need panels characteristics:

- Location (latitude - longitude)
- Area
- Surface azimuth

Build a second model, relying on available solar data:

- Installed power
- Peak area

Collecting data

We needed 2 types of data:

- Photovoltaic production data
- Irradiance data

Collecting data

We obtained production and irradiance data for ULiège

$\Rightarrow$ evaluate our panelwise model.

We needed the same, at a larger scale. Only publicly available production data: Elia (measurements, forecasts, etc.), with a granularity of 15 minutes.

Collecting data

However, Elia provides photovoltaic production data at a provincial scale.

Hence, to get reference data to compare to, we opted for predicting the photovoltaic production of the province of Liège.

As far as irradiance data is concerned, we discovered Solcast's API (historical and forecast data). It provides data at a granularity of 30 minutes.

Pre-processing data

To match the irradiance granularity, we simply averaged Elia's measures for each 30-minutes period.

Dealing with missing data was done using linear interpolation.

Panelwise model

Our panelwise model relies on the computation of two main solar angles:

- Incidence angle θ between the sun rays and the normal vector of the panel installation (depends on other solar angles)
- Solar altitude *alt*

The cosine of θ relies on:

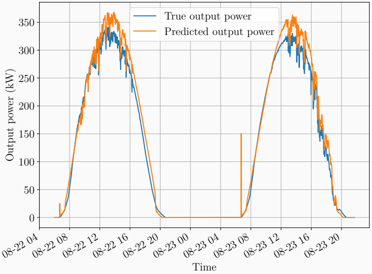
- Surface azimuth
- Declination angle δ
- Latitude and longitude
- Hour angle ω
- Tilt angle β

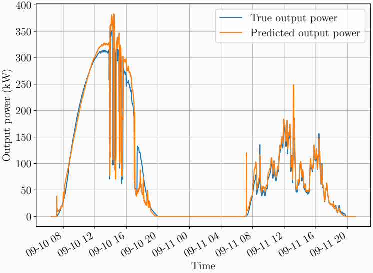
Panelwise model

We derived the following model

$$P = \frac{\eta I \cos(\theta) A}{\sin(alt)} \quad [\mathrm{W}]$$

Panelwise model





Panelwise model

With fixed parameters (η, β), our model is very rigid and dependent on the quality of our panels mapping and on the irradiance forecasts.

$\Rightarrow$ introduce uncertainties.

Panelwise model

We considered the quantity

$$\mathit{naive\_power} = \frac{I \cos(\theta) A}{\sin(\mathit{alt})}$$

as fixed and modeled uncertainties on η, using PyStan as probabilistic programming language.

To model the posterior distribution of η, we *fit* using Elia's production measurements over the past 7 days.

Panelwise model

The predicted photovoltaic power is modeled with a normal distribution centered around

$$\eta \times \textit{forecast\_power}$$

where *forecast_power* is computed using irradiance forecasts, and η is drawn from its posterior distribution.

Results will be shown later.

Provincial model

We decided to call our second model the *provincial* model, although both will be used at a provincial scale.

Here, we model our photovoltaic production using solar data, such as:

- Installed power in the province of Liège
- Estimate of the peak area

Provincial model

We derived the following model

$$P = \eta I A \quad [\mathrm{W}]$$

where η is the panel efficiency, I is the irradiance and $A = peak\_area \times \mathrm{kVA}$ is the area of panels in the province (with kVA being the installed power).

Provincial model

Similarly to the panelwise model, we also modeled uncertainties to make the model more flexible.

Again, each posterior is built using Elia's production measurements over the past 7 days, and the predicted power is modeled with a normal distribution.

Analysis of both models

To evaluate the quality of both models, we applied the following procedure: for a given day of prediction, we construct both posterior models using Elia's production measurements over the past 7 days and predict for the considered day.

We compare the predictions of our models with respect to Elia's measurements, using the MSE and RMSE. Beforehand, we remove night observations to only evaluate when there is a nonzero produced power.

Analysis of both models

We will compare our posterior models with their prior (rigid) counterparts, along with two simple baselines:

- Predict for day D the measured production of day $D-1$
- Predict for day D the average measured production over the last 7 days.

To get representative results, we conducted the procedure on the whole 2019 year (using irradiance measures since past irradiances forecasts were not available).

Results

	MSE	RMSE
Prior panelwise model	2852.73	49.25 ± 20.68
Posterior panelwise model	510.84	20.06 ± 10.42
Prior provincial model	608.98	21.94 ± 11.31
Posterior provincial model	431.23	18.49 ± 9.46
Baseline 1	2195.60	38.87 ± 26.19
Baseline 2	1816.74	37.34 ± 20.56
Elia's forecast	381.85	16.79 ± 10.01

Table 1: Mean MSE and mean RMSE, along with standard deviation, for 2019 (using irradiance measurements), in MW.

Analysis of both models

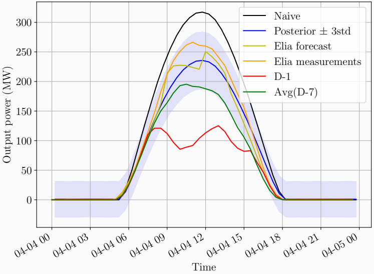
To get more *real-life* representative results, we conducted the same procedure considering April 2020. This time, we use irradiance forecasts.

Results

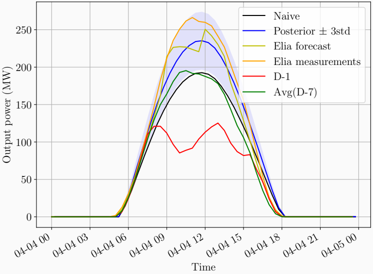
	MSE	RMSE
Prior panelwise model	5316.57	62.85 ± 37.66
Posterior panelwise model	1563.30	32.23 ± 23.33
Prior provincial model	2210.83	42.44 ± 20.62
Posterior provincial model	1595.59	32.63 ± 23.48
Baseline 1	2616.99	36.34 ± 36.68
Baseline 2	2339.52	40.87 ± 26.34
Elia's forecast	324.85	15.16 ± 10.11

Table 2: Mean MSE and mean RMSE along with standard deviation, for April 2020 (using irradiance forecasts), in MW.

Results

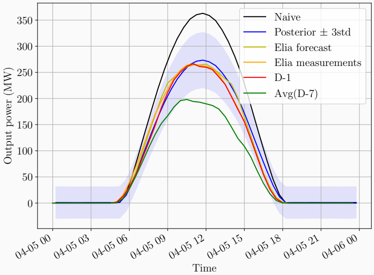


(a) Panelwise model (using irradiance forecasts).

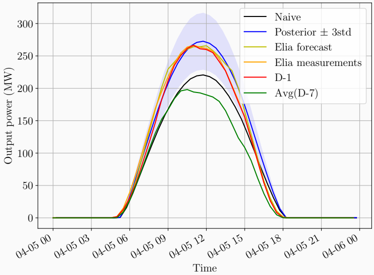


(b) Provincial model (using irradiance forecasts).

Results

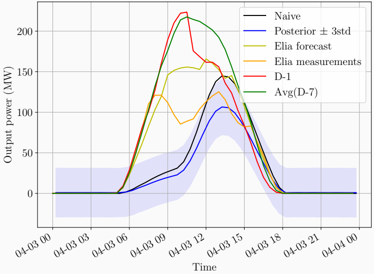


(a) Panelwise model (using irradiance forecasts).

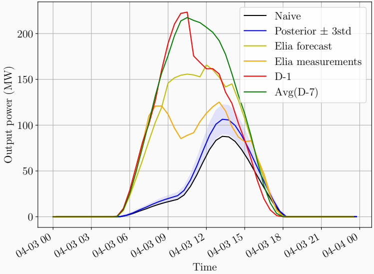


(b) Provincial model (using irradiance forecasts).

Results

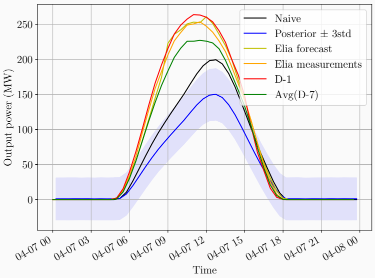


(a) Panelwise model (using irradiance forecasts).

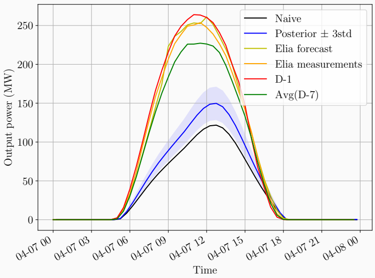


(b) Provincial model (using irradiance forecasts).

Results



(a) Panelwise model (using irradiance forecasts).



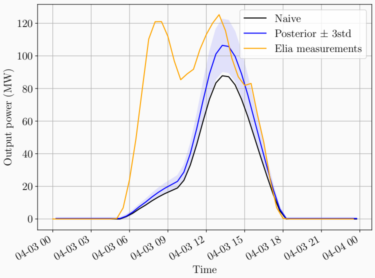
(b) Provincial model (using irradiance forecasts).

Limitations

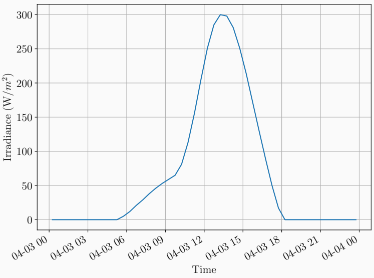
Although we obtain satisfactory results on average, some limitations can be pointed out:

- Simplicity of the models
- Using a single source of irradiance data
- Strong influence of the irradiance curve
- No real uncertainty on our panels mapping

Limitations



(a) Provincial model.



(b) Irradiance forecasted.

Wind power production

Wind forecasting problem

Wind power forecasting problems classification:

- Immediate-short-term (a few hour ahead);
- Medium-term (few hours to several days);
- Long-term (multiple days ahead).

Wind forecasting problem

Wind power forecasting problems classification:

- Immediate-short-term (a few hour ahead);
- Medium-term (few hours to several days);
- Long-term (multiple days ahead).

Wind forecasting problem

Wind power forecasting approaches classification

- **Physical modelling**: from physical parameters and NWP[1];
- **Statistical modelling**:
  - **Time series forecasting**: from power time series;
  - **NWP post-processing**: from NWP.
- **Hybrid approaches**: various approaches (constrained regression, physical model as input feature, etc. *e.g.* WPPT (Croonenbroeck and Dahl 2014)).

[1] Numerical Weather Prediction

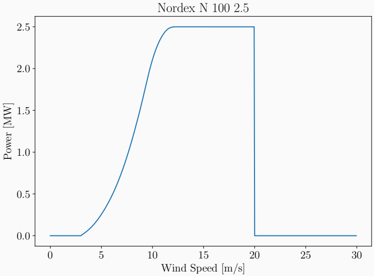
Wind forecasting problem

Wind power forecasting approaches classification

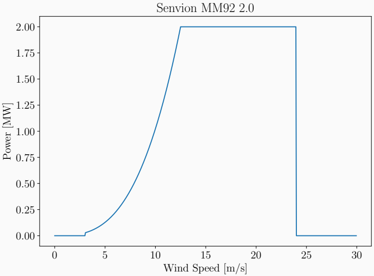
- **Physical modelling**: from physical parameters and NWP[1];
- **Statistical modelling**:
  - **Time series forecasting**: from power time series;
  - **NWP post-processing**: from NWP.
- **Hybrid approaches**: various approaches (constrained regression, physical model as input feature, etc. *e.g.* WPPT (Croonenbroeck and Dahl 2014)).

[1] Numerical Weather Prediction

Physical model



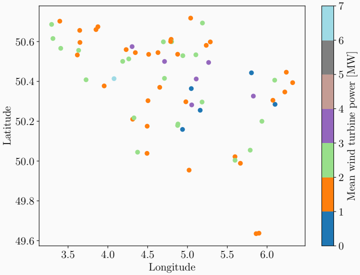
(a) Power curve



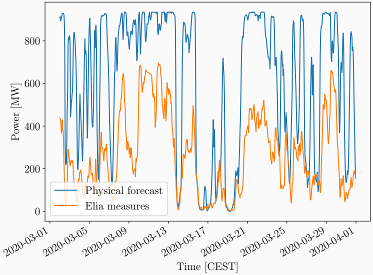
(b) Theoretic power curve

Figure 13: Examples of power curves

Physical model



(a) Wind power plants in Wallonia

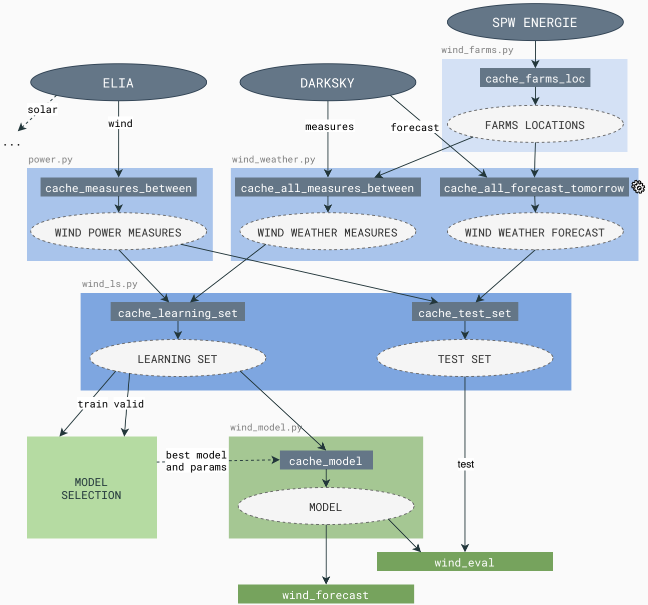


(b) Physical model forecast example

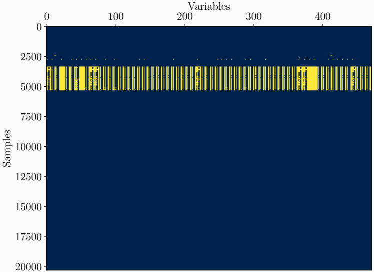
Figure 14: Wind power plants location and physical model

MAE: 355.16 MW (maximum measured power of 727 MW)

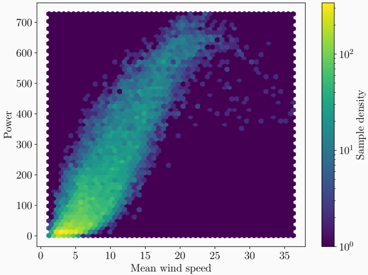
Data and computation flow



Feasibility of the regression



(a) Missingness in the learning set (2069 NA corresponding to 0.024 %)



(b) Learning set sample density with respect to the total power and the mean wind speed at the 67 wind power plants location

Feature selection

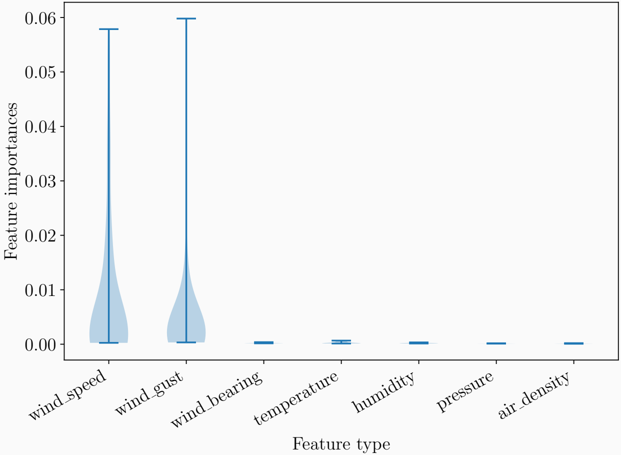


Figure 16: Feature importances distribution with respect to the type of feature

Supervised learning methods

The following supervised learning methods have been extensively experimented

- *Quantile* Extra Trees: variant of the extra trees where each leaf contains all learning samples that ended up in it, instead of only the mean. Quantiles estimated at prediction time.
- *Quantile* Gradient Boosting: gradient boosting is performed on the quantile loss:

$$\alpha(y, q) = \begin{cases} |y - q|\,\alpha & \text{if } y > q \\ |y - q|\,(1 - \alpha) & \text{else} \end{cases}$$

Hyper-parameters

Hyper-parameters have been slightly tuned to:

- *Quantile Extra Trees*: 1000 estimators, minimum number of samples required to split: 10, number of features to consider when looking for the best split: all.
- *Quantile Gradient Boosting*: 1000 estimators, learning rate: 0.1, number of features to consider when looking for the best split: all, subsample: 0.7.

Metrics

- MAE (mean absolute error);
- nMAE (normalized mean absolute error);
- nRMSE (normalized root mean squared error);
- MQL10 (mean quantile loss for 10^{th} quantile);
- MQL90 (mean quantile loss for 90^{th} quantile).

Test sets

- Shuffled sample split (25 %)
- Last 25 % of the data ($\approx$ six months).
- *Weather forecast* test set: weather forecast for day D has been requested on day $D - 1$ at 11 o'clock. (only between April 5 and April 27, 2020)
- *Weather measures* test set: same period as the *weather forecast* test set.

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	31.14	4.28 %	5.97 %	53.65	57.78	752
	Gradient Boosting	32.90	4.52 %	5.99 %	52.54	51.90	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

NB: On LS + TS, train time are 188 s and 995 s respectively.

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	31.14	4.28 %	5.97 %	53.65	57.78	752
	Gradient Boosting	32.90	4.52 %	5.99 %	52.54	51.90	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

1. First test set is invalid: information leakage between train and test set due to temporal correlation

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

1. First test set is invalid: information leakage between train and test set due to temporal correlation

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

2. *Weather forecast* test set isn't representative and certainly lead to optimistic results

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

3. Gradient Boosting and Extra Trees are equivalent in terms of MAE and RMSE;

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

4. Gradient Boosting outperforms in terms of mean quantile losses;

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

5. Considering train and prediction time, Gradient Boosting is profitable while we train less than twice a year;

Results

Test set	Method	MAE	nMAE	nRMSE	MQL10	MQL90	Time[1] [s]
Shuffled TS (Elia MAE: 34.87)	Extra Trees	-	-	-	-	-	752
	Gradient Boosting	-	-	-	-	-	0
Last 6 months (Elia MAE: 40.16)	Extra Trees	40.43	5.56 %	7.45 %	64.58	55.58	710
	Gradient Boosting	41.91	5.76 %	7.69 %	59.25	44.58	0
Weather forecast (Elia MAE: 21.51)	Extra Trees	36.86	5.07 %	7.51 %	52.51	52.71	95
	Gradient Boosting	35.71	4.91 %	7.05 %	51.33	47.14	0
Weather measures (Elia MAE: 21.51)	Extra Trees	33.26	4.57 %	6.44 %	55.66	47.26	115
	Gradient Boosting	31.62	4.35 %	5.87 %	56.18	41.77	0

Table 3: Results of the two algorithms on the learning sets (MAE and MQL are expressed in MW)

6. nRMSE results are comparable to results obtained in the literature (Croonenbroeck and Dahl 2014), but we are working on weather measures.

Results - Extra Trees

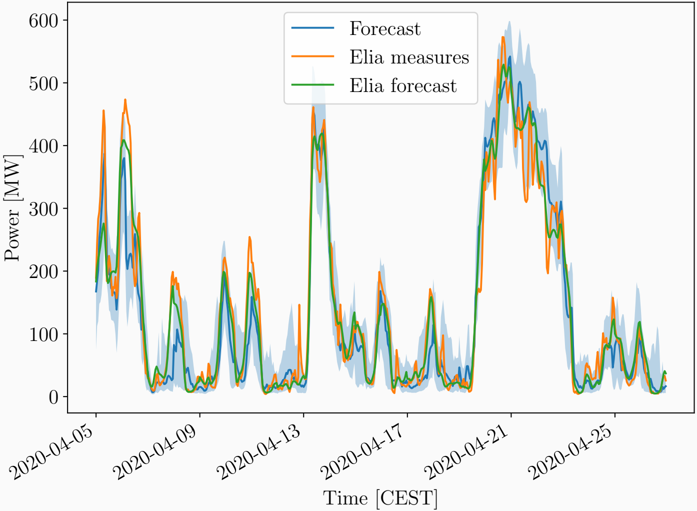


Figure 17: Quantile Extra Trees forecast

Results - Extra Trees

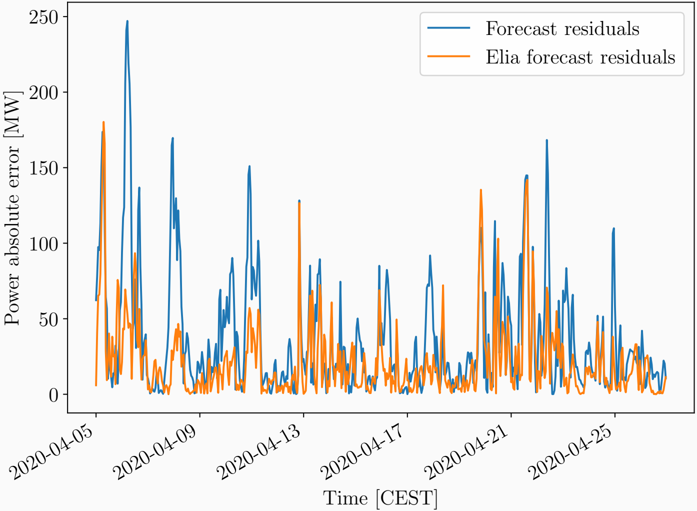


Figure 18: Quantile Extra Trees forecast residuals

Limitations and improvements

- Even if we can't assess the performance in real conditions (forecasting), we are certainly outperformed by Elia;
- We worked at the regional level instead of provincial level, because of the unavailability of data.

- We could consider a hybrid approach where we post process the output of the physical model (with or without NWP);
- We could consider a hybrid approach that mix the two statistical approaches: taking the temporal correlation into account to predict, even on the day-ahead task;
- Considering more complex or fine-tuned SL models, especially in the case of a time-series hybrid approach.

Conclusion

Conclusion

- We may regret the unavailability of data;
- No past weather forecast (wind or solar) is available, preventing us from assessing our models;
- The photovoltaic power mapping has yielded very good accuracy compared to the literature;
- The photovoltaic mapping was able to adapt to Belgian satellite imagery;
- Solar and wind models perform as well as Elia when based on weather measures instead of weather forecasts;
- Further improvements should be possible for the three parts, if more data was available.

Demonstration

znkvzr.com/apps/big-data

References

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical image computing and computer-assisted intervention*. Springer. 2015, pp. 234–241.

Kyle Bradbury, Raghav Saboo, Timothy L Johnson, Jordan M Malof, Arjun Devarajan, Wuming Zhang, Leslie M Collins, and Richard G Newell. “Distributed solar photovoltaic array location and extent dataset for remote sensing object identification”. In: *Scientific data* 3.1 (2016), pp. 1–9.

Jordan M Malof, Leslie M Collins, Kyle Bradbury, and Richard G Newell. “A deep convolutional neural network and a random forest classifier for solar photovoltaic array detection in aerial imagery”. In: *2016 IEEE International Conference on Renewable Energy Research and Applications (ICRERA)*. IEEE. 2016, pp. 650–654.

Alfredo Nespoli, Emanuele Ogliari, Sonia Leva, Alessandro Massi Pavan, Adel Mellit, Vanni Lughi, and Alberto Dolara. "Day-ahead photovoltaic forecasting: A comparison of the most effective techniques". In: *Energies* 12.9 (2019), p. 1621.

Carsten Croonenbroeck and Christian Møller Dahl. "Accurate medium-term wind power forecasting in a censored classification framework". In: *Energy* 73 (2014), pp. 221–232.