Renewable Energy Production Forecast

PROJ0016 - Big Data Project

Yann Claes, Gaspard Lambrechts and François Rozet

April 9, 2020

University of Liège

Photovoltaic production

Extension of the panel-wise model

Required inputs to the model:

- Time series
- Latitude and longitude of each panel
- Area of each panel
- Surface azimuth

Extension of the panel-wise model

Nice-to-have inputs:

- Panel efficiency η
- Panel tilt β

If not possible to obtain, we could investigate the use of PyStan.

Provincial model - Naive example

Sample example for the naive model $P = \eta IA$

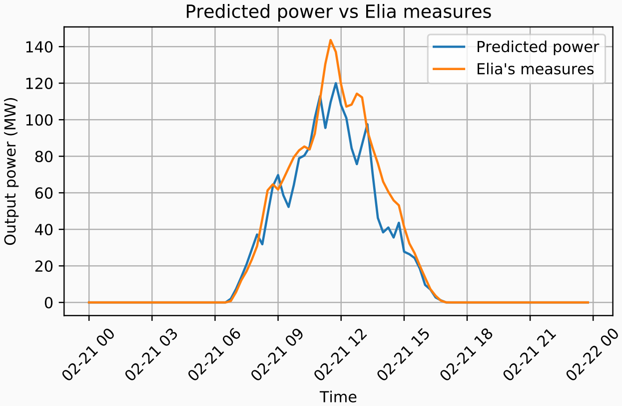


Figure 1: Naive physical model example.

Provincial model - Improvements

To improve this naive model, implement a PyStan model to define uncertainties on:

- The efficiency η
- The irradiance measures I
- The provincial panel area A

Provincial model - Posterior example

- Fit on period ranging from the 15th of February 2019 to the 23rd (excluded)
- Predict for the 23rd

The predictive model is defined as a normal distribution centered around $\mu = \eta IA$ with a standard deviation $\sigma = 100\,\mathrm{kW}$.

Provincial model - Posterior example

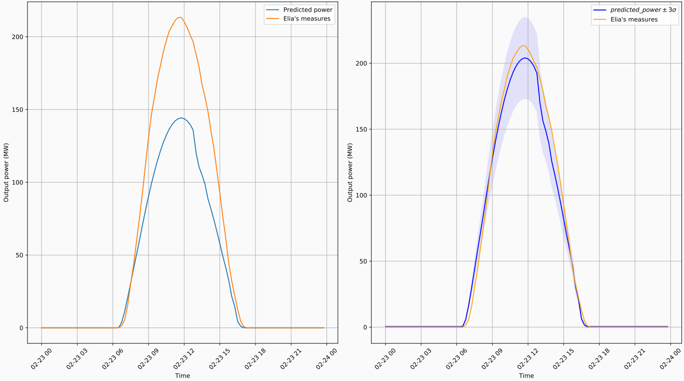


Figure 2: Comparison between prior and posterior predictive models.

Provincial model - Posterior example

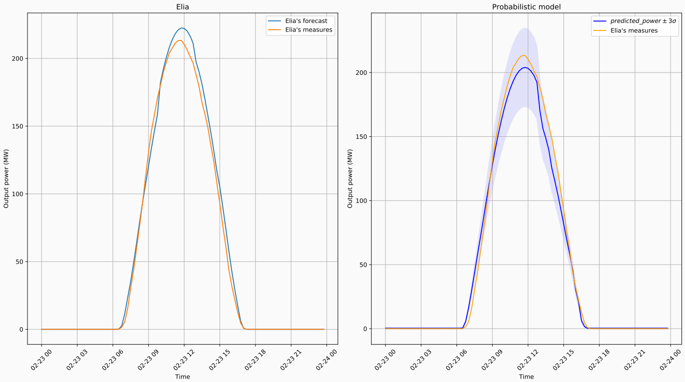


Figure 3: Comparison between Elia's forecast and the posterior model.

Provincial model - Posterior example

	MSE	RMSE
Elia	84.988	9.219
Naive	2211.522	47.027
Posterior	126.439	11.244

Table 1: MSE and RMSE for all three models [1].

Provincial model - Conclusions?

Same methodology for 15^{th} of March 2019 to the 23^{rd}.

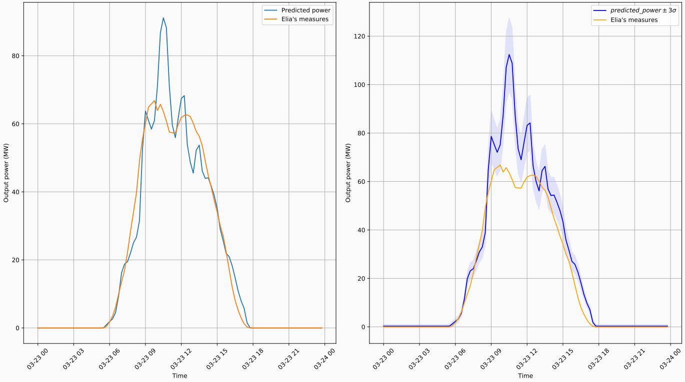


Figure 4: Same example for March 2019 (naive vs posterior model).

Provincial model - Conclusions ?

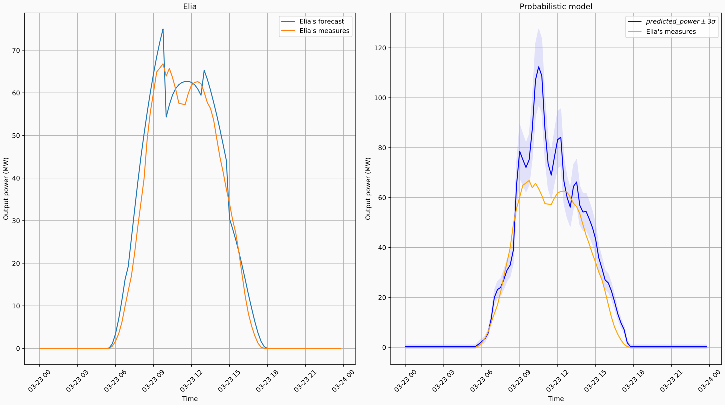


Figure 5: Same example for March 2019 (Elia vs posterior model).

Next objectives

Two main aspects need to be completed:

- Finding a (reliable) historical-forecast API
- Complete the analysis of the provincial model (baseline models, representative period)

Photovoltaic panels enumeration

Reminder

We have a suitable dataset : a large collection of annotated high resolution aerial imagery.

Our goal for this review was to design and/or train a neural network able to detect photovoltaic panels in satellite images, hence the name *Automatic Detection Of Photovoltaic Panels Through Remote Sensing*[1] or ADOPPTRS[2].

[1] Inspired from the paper "Automatic Detection of Solar Photovoltaic Arrays in High Resolution Aerial Imagery" [2].

[2] Repository : `https://github.com/francois-rozet/adopptrs`

DeepSolar

Since our goal are very similar, we looked after DeepSolar [3] model.

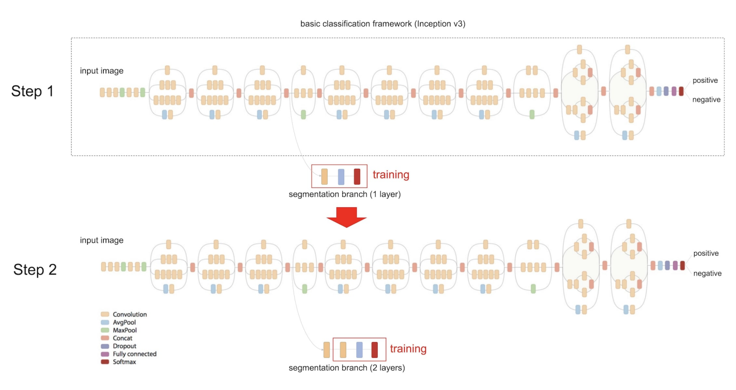


Figure 6: DeepSolar's network representation. [3]

DeepSolar

Fairly complicated. It incorporates both image classification (Google Inception V3 [4]) and semantic segmentation in a single convolutional neural network.

Classification branch is used to localize the panels. The segmentation branch is used to estimate their size.

We believe that reproducing such network is not within our capabilities.

One of the most famous segmentation model is U-Net [5].

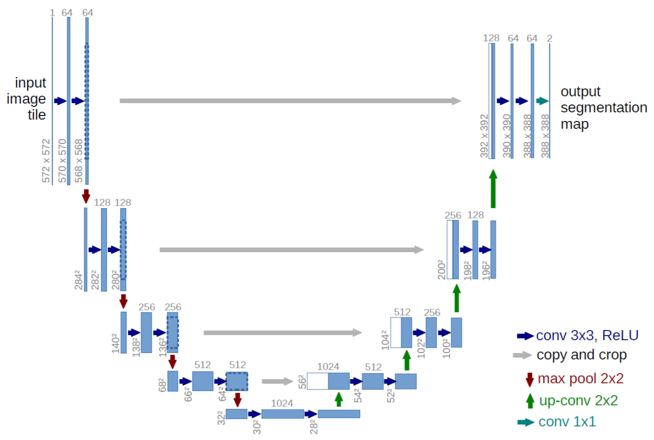


Figure 7: U-net architecture. [5]

U-Net

Initially designed for biomedical image segmentation, it actually works very well for a lot of applications and is easy to train thanks to the pass-through mechanism preventing vanishing gradients.

We have implemented U-Net using `PyTorch`.

Training

We divided our dataset in 3 subsets :

1. A training set (75 %) for training the models
2. A validation set (12.5 %) for selecting the best model(s)
3. A testing set (12.5 %) to evaluate our final model(s)

Training - Loss

We looked after a *loss function for imbalanced segmentation* and found the dice coefficient and intersection over union.

$$IoU(A, B) = \frac{|A \cap B|}{|A \cup B|} \qquad Dice(A, B) = \frac{2\,|A \cap B|}{|A| + |B|}$$

Both are similar, but we chose to implement dice.

Training - Loss

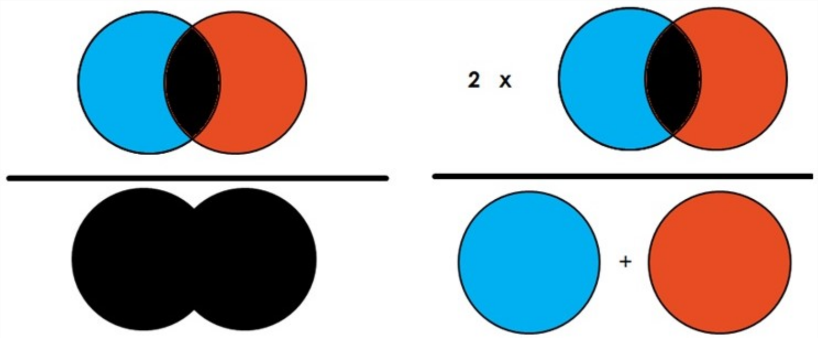


Figure 8: Illustration of Intersection over Union (left) and Dice Coefficient (right). [6]

Training - Data augmentation

Our approach to data augmentation is to randomly apply some transformation(s) while training. At each epoch, the training set is slightly different.

Training - Data augmentation

Our approach to data augmentation is to randomly apply some transformation(s) while training. At each epoch, the training set is slightly different.

- Rotations : 90°, 180°, 270°
- Flipping horizontally or vertically
- Brightness alteration
- Saturation alteration
- Contrast alteration

Training - Data augmentation

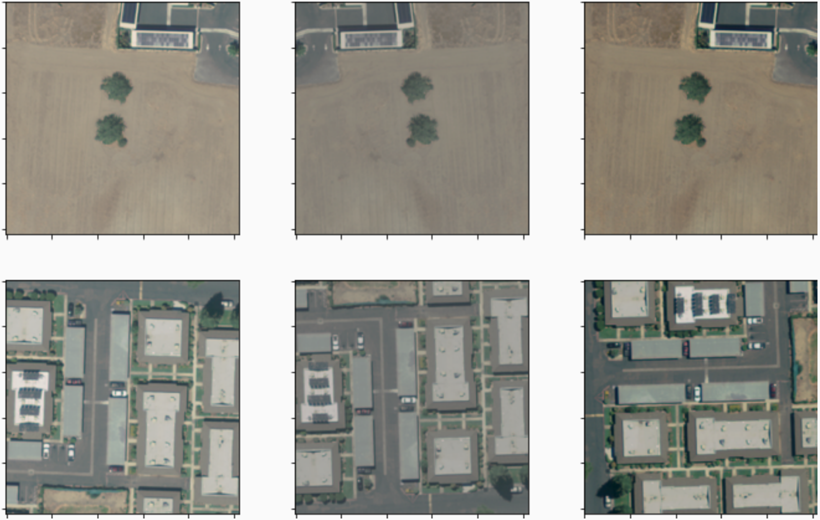


Figure 9: Some images transformations.

Results

For now, our best model achieves an average dice loss ($1 - dice$) of 26.5 % on the validation set. But the confusion matrix is more informative :

		Truth	
		0	1
Prediction	0	3.236×10^5	1.054×10^3
	1	3.096×10^2	2.690×10^3

Table 2: Average confusion matrix on the validation set.

Results

$$
\begin{aligned}
accuracy &= \frac{TP + TN}{TP + TN + FP + FN} = 99.58\,\% \\
precision &= \frac{TP}{TP + FP} = 89.69\,\% \\
recall &= \frac{TP}{TP + FN} = 71.92\,\%
\end{aligned}
$$

Our model rarely classifies something else as a photovoltaic panel but it often fails to recognize a photovoltaic panel[3].

[3] Accuracy isn't relevant because of class imbalance

Samples

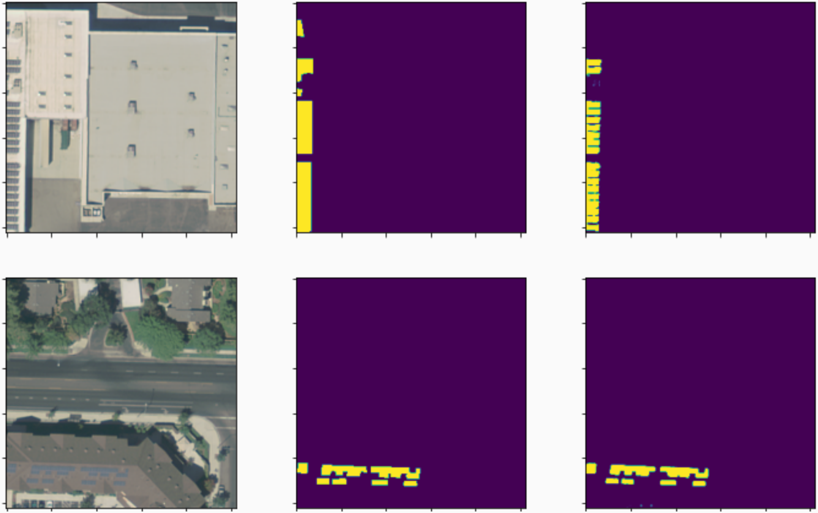


Figure 10: Representative behavior

Samples

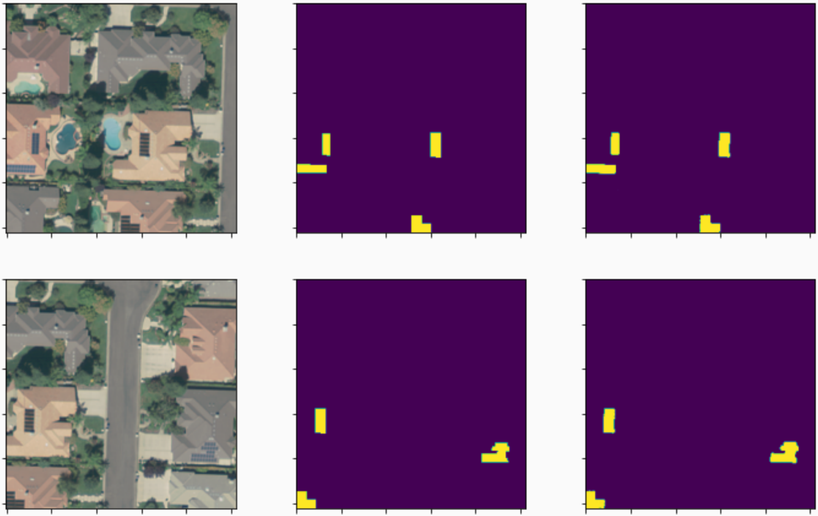


Figure 11: Representative behavior

Samples

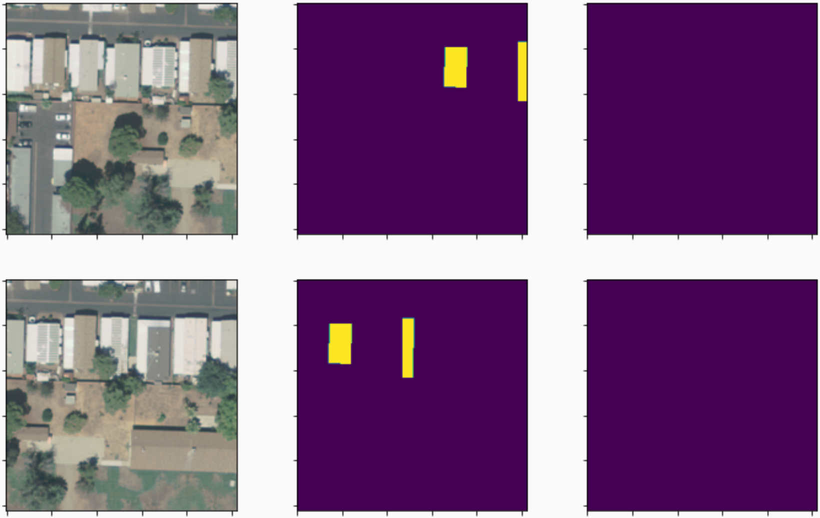


Figure 12: Abnormal panel color

Samples

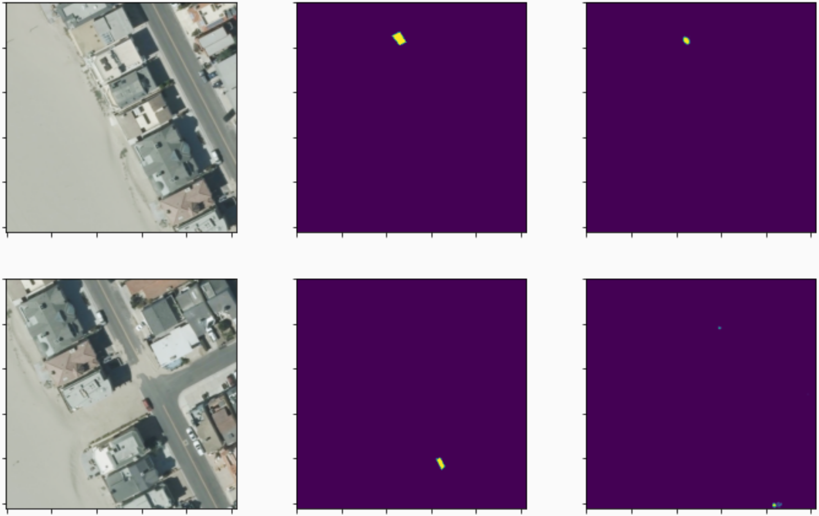


Figure 13: Abnormal panel size

Samples

Figure 14: Predictions better than annotations

Such *inaccurate* annotations could be one of the causes of the relatively bad recall of our model.

Next objectives

- We still want to improve our model and try a few others before applying it to the detection in Liège.
- There is still to do a few post-processing to convert our predictions into usable photovoltaic panel locations.

Wind production

Classification of wind forecasting problems[4]

- Very-short-term or immediate-short-term (a few hours ahead)
- Short-term (few hours to several days)
- Long-term (multiple days ahead)

[4] Wang et. al, A Review of Wind Power Forecasting Models, 2011 [7]

Classification of wind forecasting problems[4]

- Very-short-term or immediate-short-term (a few hours ahead)
- Short-term (few hours to several days)
- Long-term (multiple days ahead)

[4] Wang et. al, A Review of Wind Power Forecasting Models, 2011 [7]

Classification of forecasting techniques[5]

1. **Physical modeling**: based on Numerical Weather Prediction (NWP)
2. **Statistical modeling**
   2.1 On the very short term: *time series based forecasting*: the NWP data is not used, only the power time series is used to make the prediction.
   2.2 On the longer term: *post processing of NPW*: the temporal aspect is not considered
3. **Hybrid method**: uses *post processing of NPW* **and** *of physical modeling*

[5] Sweeney et. al, The Future of Forecasting for Renewable Energy, 2019 [8]

Classification of forecasting techniques[5]

1. **Physical modeling**: based on Numerical Weather Prediction (NWP)
2. **Statistical modeling**
   2.1 On the very short term: *time series based forecasting*: the NWP data is not used, only the power time series is used to make the prediction.
   2.2 On the longer term: *post processing of NPW*: the temporal aspect is not considered
3. **Hybrid method**: uses *post processing of NPW* **and** *of physical modeling*

[5] Sweeney et. al, The Future of Forecasting for Renewable Energy, 2019 [8]

Possible improvements on physical model

- Take hub height wind into account using the log wind profile:

$$u(z_2) = u(z_1)\frac{\ln(z_2/z_0)}{\ln(z_1/z_0)}$$

- Model in a simple way the self-wake effects according to the wind speed and direction

Statistical model: *weather to power*

The inputs of our model are (for 15 different locations):

windSpeed	windGust	windBearing	temperature	humidity	pressure	airDensity
m/s	m/s	deg	K	%	Pa	kg/m^3

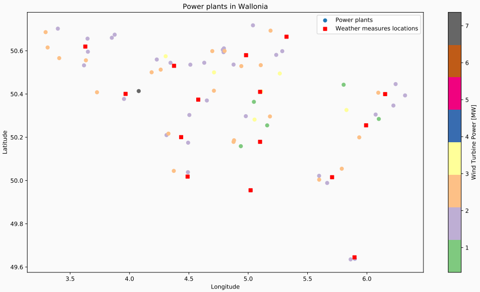
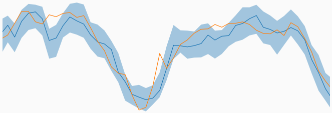


Figure 15: Power plants and weather measures in Wallonia

Metrics[7]

- MAE: Mean Absolute Error
- sMAE / nMAE: standardized / normalized Mean Absolute Error
- MQL: Mean Quantile Loss[6]

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



[6] Meinshausen, Quantile Regression Forests, 2006 [9]

[7] Messner et. al, Evaluation of Wind Power Forecasts – An up-to-date view, 2020 [10]

Models: assessment and selection protocols

Models:

- Tree bagging method: Extra Trees
- Tree boosting method: Gradient Boosting

Protocols:

- Protocol 1: train on 2019, test on February 2020: allows visualizing the prediction
- Protocol 2: 1.5 year shuffled using 30×24 samples as test set: reduces the bias in the test set.

Results

MAE	Extra Trees	Gradient Boosting
Protocol 1	36.35	36.93
Protocol 2		

Table 3: Train set CV scores

Results

MAE	Extra Trees	Gradient Boosting
Protocol 1	36.35	36.93
Protocol 2	25.37	28.18

Table 3: Train set CV scores

Results

Protocol	Method	MAE	sMAE[8]	MQL10	MQL90
Protocol 1	Extra Trees	42.38	6.00%	69.86	60.51
	Gradient Boosting	43.08	6.10%	63.62	50.94
Protocol 2					

Table 4: Test set scores

[8] To be compared with Croonenbroeck and Dahl, Accurate Medium-Term Wind Power Forecasting in a Censored Classification Framework, 2014 [11]

Results

Protocol	Method	MAE	sMAE[8]	MQL10	MQL90
Protocol 1	Extra Trees	42.38	6.00%	69.86	60.51
	Gradient Boosting	43.08	6.10%	63.62	50.94
Protocol 2	Extra Trees	28.13	4.03%	44.67	51.91
	Gradient Boosting	31.11	4.46%	47.24	50.24

Table 4: Test set scores

[8] To be compared with Croonenbroeck and Dahl, Accurate Medium-Term Wind Power Forecasting in a Censored Classification Framework, 2014 [11]

Results

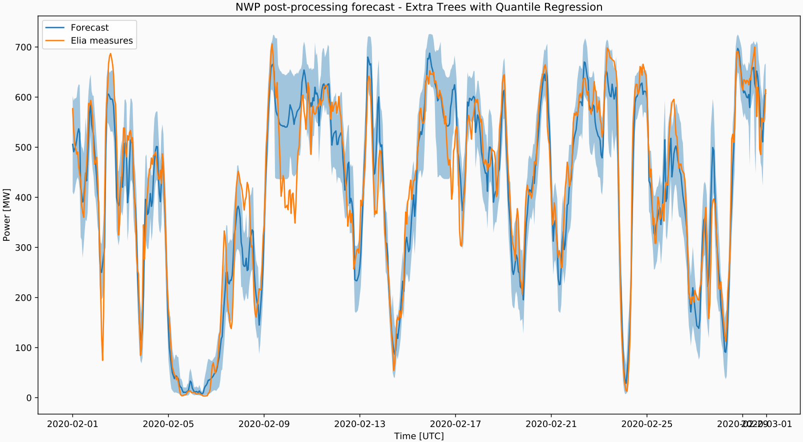


Figure 16: Extra Trees with Quantile Regression - Forecasting for Protocol 1

Results

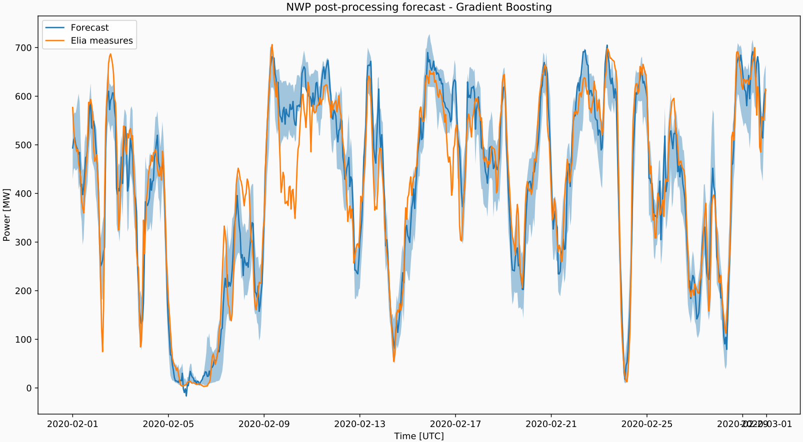


Figure 17: Gradient Boosting with Quantile Regression - Forecasting for Protocol 1

Next objectives

- Adding 3 new variables in the learning set
  - one-day before measurements
  - total wind power in Wallonia over time
  - total monitored power by Elia over time
- Testing a MLP as SL learning method
- Having a look at feature importances and considering feature selection
- Testing and assessing the quality of the forecast on actual weather prediction (NWP) instead of weather measurements

References

Alexis Bocquet et al. "Assessment of probabilistic PV production forecasts performance in an operational context.". In: *6th Solar Integration Workshop - International Workshop on Integration of Solar Power into Power Systems.* (2016).

Jordan M Malof, Kyle Bradbury, Leslie M Collins, and Richard G Newell. "Automatic detection of solar photovoltaic arrays in high resolution aerial imagery". In: *Applied energy* 183 (2016), pp. 229–240.

Jiafan Yu, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal. “DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States”. In: *Joule* 2.12 (2018), pp. 2605–2617.

Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. “Rethinking the inception architecture for computer vision”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 2818–2826.

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.

Ekin Tiu. *Towards data science*. 2019. URL: https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2.

Xiaochen Wang, Peng Guo, and Xiaobin Huang. "A review of wind power forecasting models". In: *Energy procedia* 12 (2011), pp. 770–778.

Conor Sweeney, Ricardo J Bessa, Jethro Browell, and Pierre Pinson. "The future of forecasting for renewable energy". In: *Wiley Interdisciplinary Reviews: Energy and Environment* (2019), e365.

Nicolai Meinshausen. "Quantile regression forests". In: *Journal of Machine Learning Research* 7.Jun (2006), pp. 983–999.

Jakob W Messner, Pierre Pinson, Jethro Browell, Mathias B Bjerregård, and Irene Schicker. "Evaluation of wind power forecasts—An up-to-date view". In: *Wind Energy* (2020).

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