

SEAMLESS PV POWER FORECASTING S1

RETROSPECTION

Louise Tiger

In this paper we will talk about the results obtained in the notebook and what is planned next.

Results:

On the first part of the notebook we treated the data merging datasets, dealing with missing values and creating a new column (Sirta_j+1). Then we had some visualization of the dataset. For the last part, different machine learning model has been tested to enhance forecast results: Linear Regression, Autoregression, SVR (linear, poly and RCB), Random Forest and GMDH. To make this prediction we did not use KC columns as they give us real time information about the real GHI. The only satellite value used was the further one (T0+360). For each model we found R^2 around 0.80 and DTW between 300 000 (SVR) and 415 000 (Linear Regression). This goes for every model except Autoregression which, for now, predict each point depending on the last 56 points which includes real results register 15 minutes before. It will need to be improved so that it does not have access to the first 24 points (prediction 360min before) to be compared to the over models.

A studied on input variables has been done for Random Forest, we can see that Arpege prediction is mainly used (80%) for the prediction. For our model to be really useful, we need to clearly outperform Arpege.

A cross validation shows that the more value we have the more Arpege is used to predict. It also shows a regularity in the prediction precision (the model is a little bit more accurate with more values).

Finally, a studied on the evolution of measurement depending on used satellite for Random Forest has been done. It shows that the more recent is the value the more it is accurate which is coherent.

Some Improvements for next Semester:

Autoregression:	Will need to be improved so that it does not have access to the first 24 points (prediction 360min before) to be compared to the over models.
Input variables:	Find a program to know the input variables on other models (not just Random Forest)
New Models:	Test more models (ELM, ANN, Neuro-fuzzy,...)
Year mean:	Use year mean as an input variable + maybe had a mean for the season or month
Deviation prediction:	Instead of predicting the GHI, predicting the deviation from the year mean .
Time horizon:	Try prediction on different time horizon (6h before, 1 day, 1 week)