

Introduction: position of the problem

Usually speaking, Forecast systems are classified along several axes:

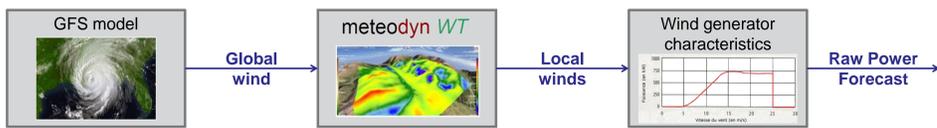
- **Intraday/Extraday**, or **Short Term/ Very Short Term**. This has to do with the look ahead time (horizon): from 0 to several hours for Very short term (Intraday) , from some hours to some days for Short term (Extraday).
- **Deterministic/Stochastic**. These two opposite concepts use different scientific tools: physics and mechanics for deterministic approach, statistical tools and machine learning for stochastic approach.
- **Numerical Weather Prediction / Online Data**: there are basically two sources of data in order to perform a power forecasting. The first one is meteorological predictions, based on global measurements and heavy numerical calculations, the second is online measurement (for instance from SCADA system).

For each axis, one concept generally excludes the other. Intraday (Very Short term) is commonly Stochastic with online measurements while Extraday (Short term) is usually Deterministic based on NWP data.

This work aims to breakdown these classifications, proposing a unique tool based on the unification of all these techniques.

Deterministic Forecast: physical approach

Our deterministic approach starts from a Numerical Weather Prediction, which predicts the "meteorological" wind near the wind farm area. Then, a Computational Fluid Dynamics tool makes a micro down scaling to provide the future wind characteristics at the exact location of each wind turbine. Then, thanks to the power curves of the machines, the air density provided by the NWP, and a planning maintenance given by the user, the final Power Output is calculated. This step includes the interaction between the wind turbines, through out the Jensen wake model.

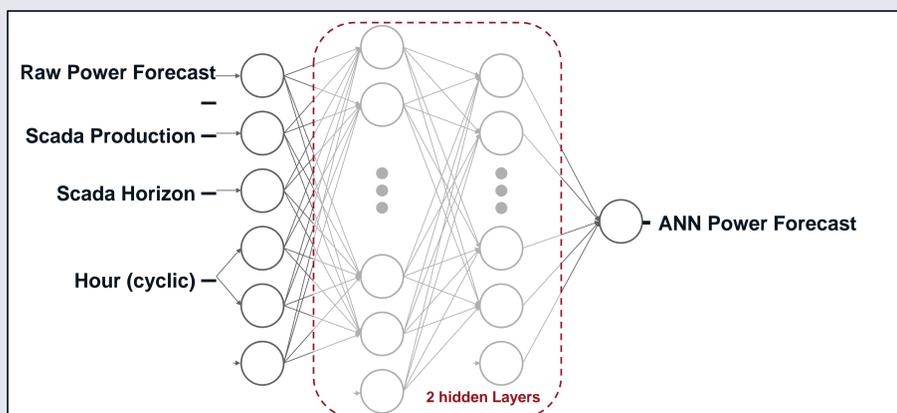


Stochastic Approach: Machine Learning

Our Machine Learning procedure is based on an Artificial Neural Network. The architecture is a supervised feed-forward fully connected network. It is trained using a backpropagation learning process, and selected thanks to a genetic algorithm.

The optimization of input variables brought the following final set:

- **Raw Power Forecast**: the power provided by the deterministic approach
- **Scada Production**: the instant output power measured in the real wind farm
- **Scada Horizon**: the time lag between the last measurement and the forecasted time
- **Hour**: the forecast time within the day (in order to take into account the night/day effects, maybe omitted in the deterministic approach)



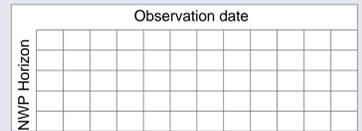
References

1. Micro-Scale Modeling combined with Statistical Learning for Short-Term Wind Power Forecasting, **Jeremie JUBAN**, EWEA 2013, Vienna
2. NWP Research and Its Applications for Power Systems in China, **Dr Rong Zhu**, ICEM 2013, Toulouse
3. Implementation of a Fast Artificial Neural Network Library (FANN), **S. Nissen**, University of Copenhagen (DIKU), 2003.

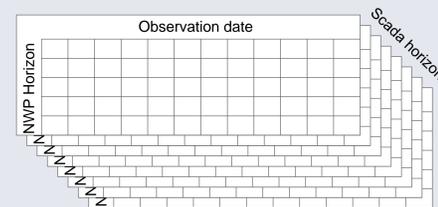
Data sets

A usual question raised in Machine Learning concerns the historical data necessary to train the network. In this case, the historical NWP meteorological data has to be collected, as well as long term measurements on site (Observations). An automatic batch for re-running the deterministic approach provides the Raw Power Forecast historical data.

The NWP runs are provided twice a day and they cover several days. This provides several forecasts for each time step, with distinct NWP-horizons. In order to match with observations (measured data), a two dimensional vision of time has to be adopted: Observation time vs NWP-horizon.



With the addition of online measurements, with their own horizons, this two dimensional vision of time has to be changed in a 3D time vision: Observation time vs NWP-horizon vs Scada-horizon. A single observation can be predicted with different combinations of NWP and Scada horizons.



This process replicates several times the same data (a single measurement is duplicated along NWP and Scada horizon axes), which makes a strong correlation between different data points. Thus, constructing the learning/test/validation by pure randomization results in data snooping (data within different sets are strongly correlated). A way to avoid this failure is to regroup the data by day before splitting them into different sets randomly.

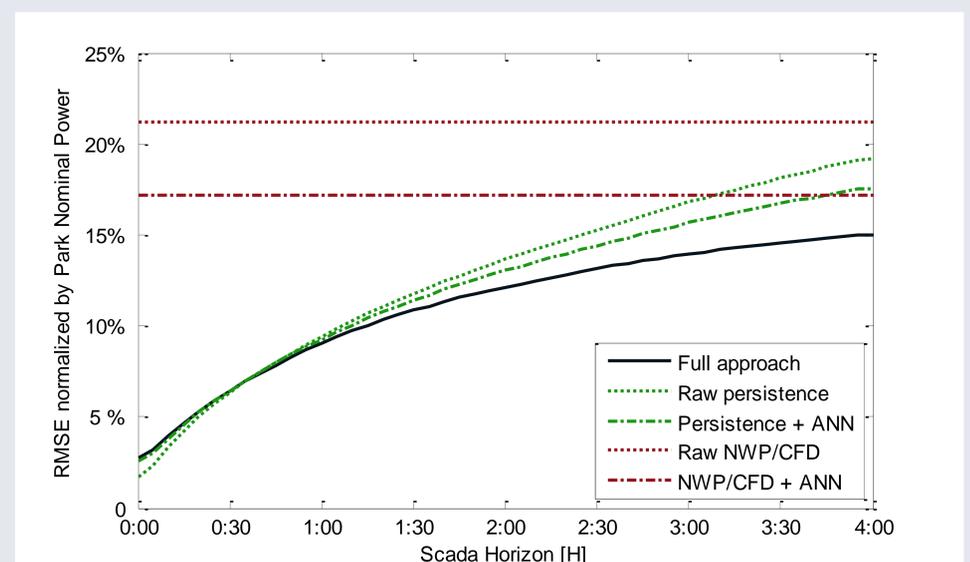
Benchmark

This approach is tested on a large wind farm (99 MW, 66 WTG) on a complex terrain. We have about 1 year historical data. There are 4 NWP runs a day, with a 15 min time step, leading to 2,5 millions of data points (250k=37 days for validation, 750k=108 days in test set, 1M= 216 days in the learning set). The results are based on the Normalized Root Mean Square Error, binned by Scada horizons.



We compare the results of the whole process, with those from simpler approaches:

- pure persistence (forecasted power = measured power)
- persistence + ANN Learning,
- pure NWP/CFD (with no calibration at all)
- NWP/CFD + ANN.



For horizon less than 30min, Raw persistence is still preferred. For all longer horizon, the full model has a smaller error than other solutions. Numerical solutions (NWP/CFD) beats online measurements after 3~4 hours.

Presenters



María Bullido García
Forecasting Expert



Daniel Dias De Oliveira
Sales department manager



Céline Bezault
Wind Power Expert