Final project topic proposal:
Power time series forecasting using Machine Learning

1. Introduction

Forecasting of the future power output is an important task in wind energy, as it allows planning the operation of the wind power plants and of the energy system balancing in general. It is also useful if there is an estimate of how uncertain the forecast is, because this allows for assessing the risks of the operator not being able to deliver the power specified in the bid.

The main purpose of the project is to develop and show examples of a procedure for wind power forecasting, evaluate the model performance against alternatives such as e.g. the "persistence" approach. In addition, the newly developed model will be assessed in terms of uncertainty, as well as sensitivity to the choice of hyperparameters such as the forecast length and the amount of historic data used in the forecast.

2. Approach

The data set with measurements from a V52 turbine can be used to demonstrate a power output-forecasting tool. The dataset contains the following data channels:

1) Time
2) Wind speed
3) Wind direction
4) Active power
5) Reactive power

The first three channels are based on a met mast positioned 2.5 rotor diameters westward of the turbine, while the power channels are part of the turbine's SCADA records.

The key element of the task is to build a model function $f()$ of the following form:

$$P(t + \tau) = f\big(\mathbf{X}(t, t - \Delta t, \dots t - N\Delta t)\big)$$

where P is the power output, τ is the forecast horizon, Δt is the sampling rate of the time series, and $\mathbf{X}$ is a set of input features (variables) that are used to predict the power. Depending on the type of model $f()$, $\mathbf{X}$ may or may not contain the power signal.

Several modelling approaches might provide useful results. Some of the potential approaches are:

- Implement a univariate autoregressive model (i.e., predicting the power based on just the historical power). Could be done with e.g. LSTM (Long Short-Term Memory) model, or with ARIMA (Auto-Regressive Integrated Moving Average) model.
- Multivariate sequence model (i.e., using several time series inputs like wind speed, power, wind direction, to predict future power). Could be implemented with an LSTM model.

- Multi-lagged discrete model.

3. Implementation specifics

When running forecasts, we are dealing with time series and capturing/modelling the time dependence properly is of high importance. We need to make sure our time series are continuous - i.e., adjacent rows in the data are exactly one time step apart, otherwise we lose the autocorrelation and time dependence information. Keeping track of time dependence is facilitated by using the time stamp as an index of the data – e.g. by using data frames (the pandas library in Python). Then one can either fill missing data points, or limit the analysis to periods where there's no gaps in the data.

Another thing to keep in mind is the data filtering. The Risø site where the measurements are taken is flat, but there are different wind conditions for different directions due to the fjord to the west, and the wakes of turbines to the east and southeast.

4. Formal requirements

The final project gives the students the possibility for applying the knowledge gained on the course to a problem relevant to their own technical interests. The topic and goals of the project are agreed individually between the students and the teacher. The solution should involve methods within the scope of the course, and the results of the project should demonstrate that the student has to a significant degree achieved the learning objectives.

The report on the project should reflect this purpose, and should contain at least the following information:

- Description of the problem/topic of the study.
- A brief description of the methods used, including the most important equations and supported with references.
- A description of the solution procedure with examples and demonstration of results in a suitable form.

Jupyter notebooks are a useful tool for running Python code along with including description and documentation. However, it is not recommended to directly submit final reports in Jupyter format (or a pdf printout of the Jupyter notebook), due to the usually low quality of the formatting and sometimes a lot of unnecessary items. It is still allowed to submit a Jupyter notebook but only in case the formatting quality is comparable to that of a report written in e.g. Word.

When defining the project details, care should be taken so that the expected workload will correspond to what is expected of a student for achieving 5 ECTS points. It is allowed for students to work individually as well as in groups of two or three. A reasonable project ambition corresponds to individual efforts of each participant comparable to the efforts required for completing e.g. Assignment 2 and 3 in the course.

The students have to submit a report describing the final project and its outcomes. The report should be submitted as a single pdf file with the name of the student(s) indicated on the title page. In case more than one student has worked on the report, the specific contribution of each student has to be explicitly described in the report. Just stating that everyone has equal contribution is not desirable as it does not allow distinguishing the work of individual students. The total length of the project report should be less than 6000 words.