31761 - Renewables in Electricity Markets

Assignment 3: Renewable energy forecasting - Let's compete!

General considerations

Assignment 3 concentrates on renewable energy forecasting, more precisely wind energy, mainly based on Modules 7-10 of the course. The assignment takes the form of a forecasting competition inspired by the recent Global Energy Forecasting Competitions held in 2012 and 2014 (GEFCom2012 and GEFCom2014) on kaggle.com and crowdanalytix.com. The data for Assignment 3 includes wind power data (normalized by the wind farm nominal capacity, to simplify things) for a wind farm in Western Denmark, Horns Rev with a nominal capacity of 160MW, while the weather forecast input information is from the European Centre for Medium-range Weather Forecasts (ECMWF - ecmwf.int), the world-leading research and operational weather forecasting centre.

The competition is organized in 4 stages, with the teams being ranked at every stages, even though only ranking at stages 2-4 will be used for the final ranking of the various teams. Stage 1 concentrates on a single day while stages 2-4 are for an entire month. The dates for the delivery of the forecasts at various stages, as well as the template format, are detailed in the description of the assignment hereafter.

The aim of Assignment 3 is to evaluate

- your understanding of the basics of renewable energy analytics,
- your ability to evaluate forecasts and have a critical eye on forecast quality,
- your ability to generate forecasts,
- your creativity in proposing approaches to forecasting.

The expected outcome of Assignment 3 includes:

- a report of maximum 10 pages (excluding appendices),
- the 4 deliveries of forecasts, at regular intervals.

Assignment 3 is to be performed in groups, where a group can consist in 1, 2 or 3 students. It is not recommended to do it alone, though it is possible. For the data handling and computation, you may use your favorite programming environment, e.g. Matlab, R, etc.

The **evaluation of Assignment 3 will count for 26.7% of the final grade**.
All groups will be ranked based on their forecast performance at the last 3 stages.
A part of the grade for the assignment will be directly linked to that ranking. Good luck!

The evaluation of the assigment will equally give weight to:

- Description of forecast strategies (background and motivations, type of methods and tools employed, process of making the forecast strategies evolves through the weeks, etc.)
- Rigour and creativity (originality and relevance of ideas being proposed, quality of formulation, etc.)
- Evaluation of forecasts and illustration (use of forecast verification metrics and tools, analysis of performance, deduction of potential ways to improve forecasts, etc.)
- Discussion (what went well, what went wrong, how the forecasts could be improved - in your opinion, etc.)
- Ranking (0 for the last team, max. grade for the first team - app. 20% of the grade)

- General presentation of the assignment report (flow of information, quality of English, quality of illustrations, etc.)

Description of the Assignment

Wind power forecasts are essential inputs to a number of decision-making problems in electricity markets. It is therefore a good idea to understand how these forecasts can be generated. You will have to generate wind power forecasts for given periods, with weather forecasts as input, and with a long history of past cases (2 years) to learn from. Such historical data includes the weather forecasts themselves, and the power observations that were eventually collected at the wind farm. Therefore, what can be learned from these data is the relationship between weather forecasts and the power eventually produced at the wind farm. Based on this relationship learned from the past, one is to apply it to input weather forecasts to issue genuine wind power forecasts over the evaluation periods defined for the various stages of the forecast competition.

In the present case, the input data consists in wind forecasts at 2 heights (10m and 100m above ground level). Wind forecasts are given in terms of their zonal and meridional components (u and v), which correspond to the projection of the wind vector on West-East and South-North axes, respectively. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead. The data and the various files provided for the assignment are described in detail in the Appendix.

The various stages of Assignment 3 include:

Stage 1. Getting familiar with the data, and giving it a go!

During the first session for Assignment 3, we will use an hour or so to get familiar with the forecasting competition and its setup, the data provided, and the expectations in terms of forecast delivery.

Historical data (weather forecasts and wind power measurements) will be provided for a period between 1.1.2015 and 31.12.2017. Based on these data, as well as input weather forecasts issued on the 1.1.2018 at 00:00, you will be asked to issue your best wind power forecast for the 24 hours of 1.1.2018. At this stage you also need to pick a name for your team!

Forecasts are to be uploaded through campusnet (Assignment 3.1) before 27.4.2020 at 5pm, using the template provided (in csv format). The name of the file should be "31761-teamname-stage1.csv", where "teamname" is your team name. Note that Stage 1 is there to give it a try only, though it does not mean it is allowed not to deliver forecasts.

3 files will be provided at that stage:

- TrainData1.csv: file containing past cases (1.1.2015 to 31.12.2017) to learn from
- WeatherForecastInput1.csv: file containing input weather forecasts for the period for which we aim at generating wind power forecasts (issued on 1.1.2018 at 00:00)
- ForecastTemplate1.csv: file giving the template for how the wind power forecasts should be delivered

Here are a few hints when one is to generate forecasts rapidly, and with limited knowledge/data:

- When recent power measurements are available, one can always use the last power measurement as the forecasts for any lead time in the future (so-called persistence);
- Some averaging of past measurements can also be a good idea, for instance over the last m hours;
- One can also think that wind power generation, for a given hour of the day, could be quite similar from day to day. Therefore, if aiming to issue a forecast for a certain hour of the day, the average of what was observed for that hour for the last m days could actually be a good first prediction;
- Expert rules can also work. E.g., you could set your basic rules such that: (rule 1) if predicted wind speed is less than 4 m/s, let's predict power to be 0; (rule 2) if predicted wind speed is greater than 12 m/s, let's predict power to be 1; (rule 3) Let's then draw a straight line between these two, to predict power generation as a linear function of predicted wind speed, for wind speeds betwen 4 and 12 m/s.

Stage 2. The true competition begins...

During the second session, we will clarify all potential misunderstandings and remaining questions about the forecast setup that could still be present after deliveries of the forecasts for stage 1. Besides, the power measurements for the 1.1.2018 (day for which forecasts issued at stage 1) will be shared, for you to have an idea of how well you performed.

Historical data (weather forecasts and wind power measurements) will be provided for a period between 1.1.2015 and 1.1.2018. Based on these data, as well as input weather forecasts issued at 00:00 for every day of 1.2018 (except 1.1.2018), you will be asked to issue your best wind power forecasts (with lead times up 24 hours) for every day of that month. At that stage, you may want to dig more into advanced model-based approaches to learn from the past and issue your forecasts.

Forecasts are to be uploaded through campusnet (Assignment 3.2) before 4.5.20120 at midnight, using the template provided (in csv format). The name of the file should be "31761-teamname-stage2.csv", where "teamname" is your team name.

4 files will be provided at that stage:

- Solution1.csv: file containing the power measurements for 1.1.2018, to evaluate your forecasts issued at stage 1
- TrainData2.csv: file containing past cases (1.1.2015 to 1.1.2018) to learn from
- WeatherForecastInput2.csv: file containing input weather forecasts for the period for which we aim at generating wind power forecasts (from 2.1.2018 to 31.1.2018)
- ForecastTemplate2.csv: file giving the template for how the wind power forecasts should be delivered

The teacher and teaching assistants will there act as consultants to help you build your best models and strategies to issue competitive forecasts.

Stage 3. Time to improve your forecast strategies...

During the third session, you will be given more time to verify the performance of the forecasts issued previously, and to improve your forecasting methods and strategy. Indeed, the power measurements for the whole month of 1.2018 (day for which forecasts issued at stages 1 and 2) will be shared, for you to assess how well you performed.

Historical data (weather forecasts and wind power measurements) will be provided for a period between 1.1.2015 and 31.1.2018. Based on these data, as well as input weather forecasts issued at 00:00 for every day of 2.2018, you will be asked to issue your best wind power forecasts (with lead times up 24 hours) for every day of 2.2018.

Forecasts are to be uploaded through campusnet (Assignment 3.3) before 11.5.2020 at midnight, using the template provided (in csv format). The name of the file should be "31761-teamname-stage3.csv", where "teamname" is your team name.

4 files will be provided at that stage:

- Solution2.csv: file containing the power measurements for 1.2018 (except for 1.1.2018), to evaluate your forecasts issued at stage 2
- TrainData3.csv: file containing past cases (1.1.2015 to 31.1.2018) to learn from
- WeatherForecastInput3.csv: file containing input weather forecasts for the period for which we aim at generating wind power forecasts (from 1.2.2018 to 28.2.2018)
- ForecastTemplate3.csv: file giving the template for how the wind power forecasts should be delivered

The teacher and teaching assistants will there act as consultants to help you build your best models and strategies to issue competitive forecasts.

Stage 4. A last chance to improve...

At the occasion of the last session, the power measurements for the whole month of 2.2018 (period for which forecasts were issued at stage 3) will be shared, for you to have an idea of how well you performed and to learn from it. The official ranking of the teams will also be revealed.

Historical data (weather forecasts and wind power measurements) will be provided for a period between 1.1.2015 and 28.2.2018. Based on these data, as well as input weather forecasts issued at 00:00 for every day of 3.2018, you will be asked to issue your best wind power forecasts (with lead times up 24 hours) for every day of 3.2018. This is your last chance to tune and modify the models and forecast strategies that you developed at stages 1-3.

Forecasts are to be uploaded through campusnet (Assignment 3.4) before 18.5.2020 at midnight, using the template provided (in csv format). The name of the file should be "31761-teamname-stage4.csv", where "teamname" is your team name.

4 files will be provided at that stage:

- Solution3.csv: file containing the power measurements for 2.2018, to evaluate your forecasts issued at stage 2
- TrainData4.csv: file containing past cases (1.1.2015 to 28.2.2018) to learn from
- WeatherForecastInput4.csv: file containing input weather forecasts for the period for which we aim at generating wind power forecasts (from 1.3.2018 to 31.3.2018)
- ForecastTemplate4.csv: file giving the template for how the wind power forecasts should be delivered

The teacher and teaching assistants will there act as consultants to help you refine your models and strategies to issue competitive forecasts.

Structure and contents of the report to be delivered

The report for the assignment should include:

- A description of the models and forecast strategies developed through the various stages of the competition
- An evaluation of the forecasts issued at stages 1-3. This may be based on, e.g., calculation of basic scores (bias, MAE, RMSE) overall and as a function of the lead time (between 1 and 24 hours ahead)
- 2 or 3 plots of some of the forecasts (i.e., for some days of your choice) issued at stages 1 and/or 2, along with corresponding power measurements, with a critical visual assessment of the forecasts
- Some proposals for how you would improve your forecast strategies if you were given more time and/or more data
- The code in an Appendix (if not provided separately)

Delivery of the Assignment

Assignment 3 is to be uploaded through campusnet before the **final deadline of 18.5.2020** (all day included). It should take the form of a zip or tar.gz archive with naming convention "31761-Assignment3-Student1-Student2.zip" (or ".tar.gz"), where "Student1", "Student2" are the names or DTU student numbers (e.g. s093482) of the students in the group. More of less students in a group obviously means more of less student names/ids to be used in the naming convention.

Suggested readings

Morales JM, Conejo A, Madsen H, Pinson P, Zugno M (2014). Renewable energy sources - Modelling and forecasting. Chapter 2 in *Integrating Renewables in Electricity Markets - Operational Problems*, Springer Verlag: New York (pdf) - also the references therein

Appendix: short description of data and various files

The files to be used for this assignment have all the necessary input weather forecasts, past wind power measurements, as well as templates for submitting the forecasts, for the assignment. The various files include:

ForecastTemplate1.csv

This file gives the template for submitting the forecasts for 1.1.2018. This file can be used as a basis for submitting the forecasts at stage 1 of the competition.

Variables (columns):

- TIMESTAMP: time stamps for the wind power forecast values to be generated
- FORECAST: forecast values themselves

ForecastTemplate2.csv

This file gives the template for submitting the forecasts for the whole month of 1.2018 (except for the first day). This file can be used as a basis for submitting the forecasts at stage 2 of the competition.

Variables (columns):

- TIMESTAMP: time stamps for the wind power forecast values to be generated
- FORECAST: forecast values themselves

ForecastTemplate3.csv

This file gives the template for submitting the forecasts for the whole month of 2.2018. This file can be used as a basis for submitting the forecasts at stage 3 of the competition.

Variables (columns):

- TIMESTAMP: time stamps for the wind power forecast values to be generated
- FORECAST: forecast values themselves

ForecastTemplate4.csv

This file gives the template for submitting the forecasts for the whole month of 3.2018. This file can be used as a basis for submitting the forecasts at stage 3 of the competition.

Variables (columns):

- TIMESTAMP: time stamps for the wind power forecast values to be generated
- FORECAST: forecast values themselves

Solution1.csv

This file gives the true wind power measurements for 1.1.2018, which will be used to calculate your scores and give you an idea of how good you did with your first try. This file will only be provided after stage 2 of the competition, i.e., after you have submitted your final forecasts for 1.1.2018.

Variables (columns):

- TIMESTAMP: time stamps for the wind power measurements, corresponding to the forecasts to be compiled in *ForecastTemplate1.csv*
- POWER: measured power values themselves (normalized)

Solution2.csv

This file gives the true wind power measurements for the whole month of 1.2018 (except 1.1.2018), which will be used to calculate your scores and also to give a chance to improve your forecasts. This file will only be provided after stage 2 of the competition, i.e., after you have submitted your final forecasts for the month of 1.2018.

Variables (columns):

- TIMESTAMP: time stamps for the wind power measurements, corresponding to the forecasts to be compiled in *ForecastTemplate2.csv*
- POWER: measured power values themselves (normalized)

Solution3.csv

This file gives the true wind power measurements for the whole month of 2.2018, which will be used to calculate your scores and also to give a chance to improve your forecasts. This file will only be provided after stage 2 of the competition, i.e., after you have submitted your final forecasts for the month of 2.2018.

Variables (columns):

- TIMESTAMP: time stamps for the wind power measurements, corresponding to the forecasts to be compiled in *ForecastTemplate3.csv*
- POWER: measured power values themselves (normalized)

TrainData1.csv

This file gives the set of data that can be used to find a relationship between weather forecasts inputs (wind speed forecasts at 10m and 100m above ground level) and observed power generation. These data cover the period from 1.1.2015 to 31.12.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts, and corresponding power measurements
- POWER: measured power values (normalized)
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

TrainData2.csv

This file gives the set of data that can be used to find a relationship between weather forecasts inputs (wind speed forecasts at 10m and 100m above ground level) and observed power generation. These data cover the period from 1.1.2015 to 1.1.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts, and corresponding power measurements
- POWER: measured power values (normalized)
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

TrainData3.csv

This file gives the set of data that can be used to find a relationship between weather forecasts inputs (wind speed forecasts at 10m and 100m above ground level) and observed power generation. These data cover the period from 1.1.2015 to 31.1.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts, and corresponding power measurements
- POWER: measured power values (normalized)
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

TrainData4.csv

This file gives the set of data that can be used to find a relationship between weather forecasts inputs (wind speed forecasts at 10m and 100m above ground level) and observed power generation. These data cover the period from 1.1.2015 to 28.2.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts, and corresponding power measurements
- POWER: measured power values (normalized)
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

WeatherForecastInput1.csv

This file gives the set of input weather forecasts that can be used as input to predict wind power generation for 1.1.2018, i.e., at stage 1 of the competition. These include wind speed forecasts at 10m

and 100m above ground level. These weather forecasts are issued on 1.1.2018 at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

WeatherForecastInput2.csv

This file gives the set of input weather forecasts that can be used as input to predict wind power generation for the whole month of 1.2018 (except for 1.1.2018), at stage 2 of the competition. These include wind speed forecasts at 10m and 100m above ground level. These data cover the period from 2.1.2018 to 31.1.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

WeatherForecastInput3.csv

This file gives the set of input weather forecasts that can be used as input to predict wind power generation for the whole month of 2.2018, at stage 3 of the competition. These include wind speed forecasts at 10m and 100m above ground level. These data cover the period from 1.2.2018 to 28.2.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level

WeatherForecastInput4.csv

This file gives the set of input weather forecasts that can be used as input to predict wind power generation for the whole month of 3.2018, at stage 4 of the competition. These include wind speed forecasts at 10m and 100m above ground level. These data cover the period from 1.3.2018 to 31.3.2018. Weather forecasts are issued once a day at 00:00, while they have lead times between 1 and 24 hours ahead.

Variables (columns):

- TIMESTAMP: time stamps for the wind forecasts
- U10: zonal component of the wind forecast (West-East projection) at 10m above ground level
- V10: meridional component of the wind forecast (South-North projection) at 10m above ground level
- U100: zonal component of the wind forecast (West-East projection) at 100m above ground level
- V100: meridional component of the wind forecast (South-North projection) at 100m above ground level