Module 8 – Verification of Renewable Energy Forecasts

8.1 What makes a good forecast?

Pierre Pinson
Technical University of Denmark
The nature of “goodness” in forecasting

Following Murphy (ref. and link below), the nature of “goodness” in weather forecasting (same goes for other types of forecasts) consists in:

- Forecast consistency: “Forecasts should correspond to the forecaster’s best judgement on future events, based on the knowledge available at the time of issuing the forecasts.”
- Forecast quality: “Forecasts should describe future events as good as possible, regardless of what these forecasts may be used for.”
- Forecast value: “Forecasts should bring additional benefits (monetary or others) when used as input to decision-making.”

The nature of “goodness” in forecasting

Following Murphy (ref. and link below), the nature of “goodness” in weather forecasting (same goes for other types of forecasts) consists in:

- **Forecast consistency:**

  “Forecasts should correspond to the forecaster’s best judgement on future events, based on the knowledge available at the time of issuing the forecasts”
The nature of “goodness” in forecasting

- Following Murphy (ref. and link below), the nature of “goodness” in weather forecasting (same goes for other types of forecasts) consists in:

  - **Forecast consistency:**
    
    “Forecasts should correspond to the forecaster’s best judgement on future events, based on the knowledge available at the time of issuing the forecasts”

  - **Forecast quality:**
    
    “Forecasts should describe future events as good as possible, regardless of what these forecasts may be used for”
The nature of “goodness” in forecasting

Following Murphy (ref. and link below), the nature of “goodness” in weather forecasting (same goes for other types of forecasts) consists in:

- **Forecast consistency:**

  “Forecasts should correspond to the forecaster’s best judgement on future events, based on the knowledge available at the time of issuing the forecasts”

- **Forecast quality:**

  “Forecasts should describe future events as good as possible, regardless of what these forecasts may be used for”

- **Forecast value:**

  “Forecasts should bring additional benefits (monetary or others) when used as input to decision-making”

Illustrative example (1)

- You are in charge of **optimal maintenance planning at Horns Rev**, and have booked both a vessel and an helicopter for onsite service (for a cost of 100,000€)

- The conditions for this to happen at time $t + k$ are
  - wind speed: $u_{t+k} \leq 15 \text{ m.s}^{-1}$
  - wave height: $h_{t+k} \leq 1.8 \text{ m}$

- 24 hours before service (time $t$), this is your last chance to cancel before huge financial penalties (another 100,000€)

- Your two forecasters (**Foresight** and **Blindspot**) tell you that:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{u}_{t+k</td>
<td>t}$</td>
<td>12.6 m.s$^{-1}$</td>
</tr>
<tr>
<td>$\hat{h}_{t+k</td>
<td>t}$</td>
<td>1.6 m</td>
</tr>
</tbody>
</table>

- In both cases, you go ahead with the planned service...
Illustrative example (1, continued)

- At time $t + k$, this is what actually happened:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{u}_{t+k</td>
<td>t}$</td>
<td>12.6 m.s$^{-1}$</td>
</tr>
<tr>
<td>$\hat{h}_{t+k</td>
<td>t}$</td>
<td>1.6 m</td>
</tr>
</tbody>
</table>

- In both cases, your overall cost is 100,000€,

- Both *Foresight* and *Blindspot* served their purpose, since you made the right decision... *Forecast value is good*

- You might want to have a chat with *Blindspot*, since its forecast quality appears to be far from good!
Illustrative example (2)

The boy who cried wolf (Tale from Ancient Greece) - revisited.

- Rogue Trading® made huge losses last year, due to expensive upregulation events...
- It is therefore decided to get a new forecaster that would be good at predicting them.
- Foresight and Blindspot are in competition for the job.

- The score is simple:
  \[ Sc = 100 \cdot \frac{\#\{\text{events leading to upregulation predicted}\}}{\#\{\text{events leading to upregulation}\}} \]
- the higher the better! (0 is worst, 100 is best)
Illustrative example (2, continued)

If you were *Foresight* and *Blindspot*, what would you do?

The two competitors have sharpened their strategy:

- **Foresight**
  - Always predict need for upregulation!
  - Do your best to find when upregulation will occur...

The results on the benchmarking exercise are such that:

\[
\begin{align*}
\text{market time units} &= 8760 \\
\text{events leading to upregulation} &= 3237 \\
\text{events leading to upregulation predicted by Foresight} &= 3237 \\
\text{events leading to upregulation predicted by Blindspot} &= 2500
\end{align*}
\]

Their scores:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score</td>
<td>100%</td>
<td>77.2%</td>
</tr>
</tbody>
</table>

Foresight gets the job!
If you were *Foresight* and *Blindspot*, what would you do?

- The two competitors have sharpened their strategy:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>Always predict need for</td>
<td>Do your best to find when upregulation</td>
</tr>
<tr>
<td></td>
<td>upregulation!</td>
<td>will occur...</td>
</tr>
</tbody>
</table>

- The results on the benchmarking exercise are such that:
  - $\#\{\text{market time units}\} = 8760$
  - $\#\{\text{events leading to upregulation}\} = 3237$
  - $\#\{\text{events leading to upregulation predicted by Foresight}\} = 3237$
  - $\#\{\text{events leading to upregulation predicted by Blindspot}\} = 2500$

- Their scores:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc</td>
<td>100%</td>
<td>77.2%</td>
</tr>
</tbody>
</table>
If you were *Foresight* and *Blindspot*, what would you do?

- The two competitors have sharpened their strategy:
  
<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Strategy</td>
<td>Always predict need for upregulation!</td>
<td>Do your best to find when upregulation will occur...</td>
</tr>
</tbody>
</table>

- The results on the benchmarking exercise are such that:
  
  - $\#\{\text{market time units}\} = 8760$
  - $\#\{\text{events leading to upregulation}\} = 3237$
  - $\#\{\text{events leading to upregulation predicted by Foresight}\} = 3237$
  - $\#\{\text{events leading to upregulation predicted by Blindspot}\} = 2500$

- Their scores:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sc</td>
<td>100%</td>
<td>77.2%</td>
</tr>
</tbody>
</table>

- *Foresight* gets the job!
Illustrative example (2, continued)

- The consequences are:
  - even though never missing on upregulation events, Rogue Trading® will always miss the down regulation ones
  - eventually, the financial loss may still be there... and possibly much higher than expected
The consequences are:

- even though never missing on upregulation events, Rogue Trading® will always miss the down regulation ones
- eventually, the financial loss may still be there... and possibly much higher than expected

A more **consistent** way to evaluate these forecasters would be to consider:

<table>
<thead>
<tr>
<th>event predicted</th>
<th>event not predicted</th>
<th>event happens</th>
<th>no event</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>HIT</td>
<td>FALSE ALARM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>MISS</td>
<td>CORRECT REJECTION</td>
</tr>
</tbody>
</table>

And a **proper** score, ensuring forecast consistency, is:

\[
Sc = 100 \cdot \frac{\#\text{hits}}{\#\text{hits} + \#\text{misses} + \#\text{false alarms}}
\]

The higher the better! (0 is worst, 100 is best)

(This score is called the **Threat Score (TS)**)
Illustrative example (2, continued)

- In the present case:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>#{hits}</td>
<td>3237</td>
<td>2320</td>
</tr>
<tr>
<td>#{misses}</td>
<td>0</td>
<td>917</td>
</tr>
<tr>
<td>#{false alarms}</td>
<td>5523</td>
<td>180</td>
</tr>
<tr>
<td>#{correct rejections}</td>
<td>0</td>
<td>5343</td>
</tr>
</tbody>
</table>

- The resulting Threat Score (TS) values are:

<table>
<thead>
<tr>
<th></th>
<th>Foresight</th>
<th>Blindspot</th>
</tr>
</thead>
<tbody>
<tr>
<td>TS</td>
<td>36.9%</td>
<td>67.9%</td>
</tr>
</tbody>
</table>

- Conclusions: if using a proper score...
  - *Blindspot* should have gotten the job!
  - I can promise that *Rogue Trading*® would have lower financial losses
Test case: the Klim wind farm

- The wind farm:
  - full name: Klim Fjordholme
  - onshore/offshore: onshore
  - year of commissioning: 1996
  - nominal capacity ($P_n$): 21 MW
  - number of turbines in farm: 35
  - average annual electricity generation: 49 GWh
  - data available: 1999-2003 (for some researchers)
  - temporal resolution: 5 mins, and hourly averages
  - forecasts: deterministic and probabilistic

- A link to the online description: Vattenfall’s Klim wind farm

- The wind farm has been recommissioned recently: NordJyske online article
Splitting of available data

- Forecasting is about
  - being able to predict future events, in new situations
  - not only explain what happen in the past...

- **One need to verify forecasts on data that has not been used for the modelling!**

In this Module we focus on the last 6 months of 2002, with other examples for some other periods.
Use the self-assessment quizz to check your understanding!