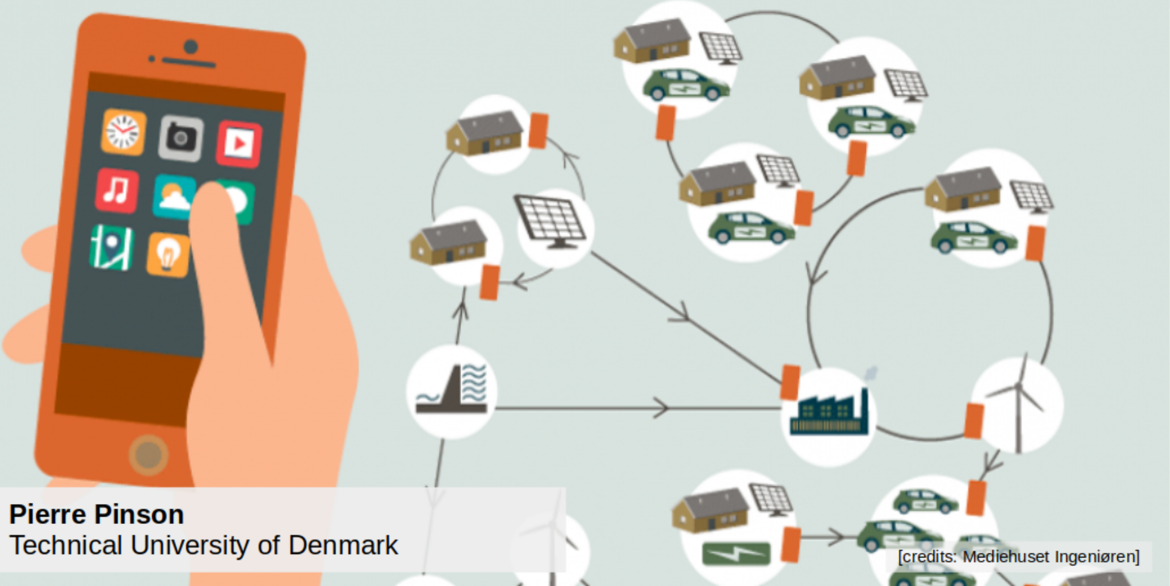
Module 8 – Verification of Renewable Energy Forecasts

8.2 Verification of deterministic forecasts



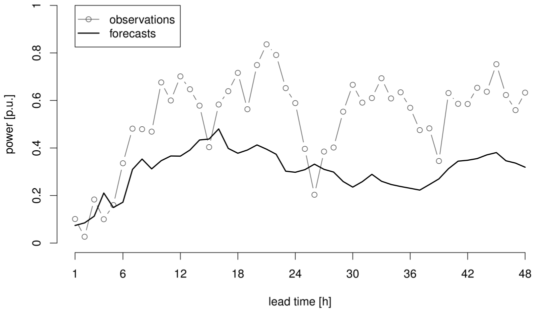
[credits: Mediehuset Ingeniøren]

Pierre Pinson
Technical University of Denmark

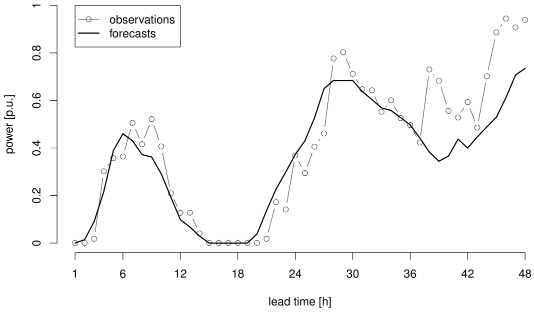
Visual inspection of forecasts

- Visual inspection allows you to develop susbtantial insight on forecast quality...
- This comprises a **qualitative** analysis only

- *What do you think of these two?*
  Are they good or bad?



Forecast issued on 16 November 2001 (18:00)

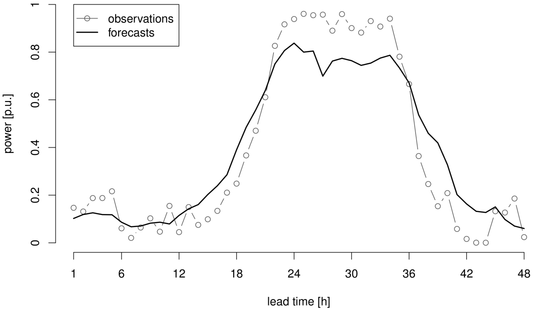


Forecast issued on 23 December 2003 (12:00)

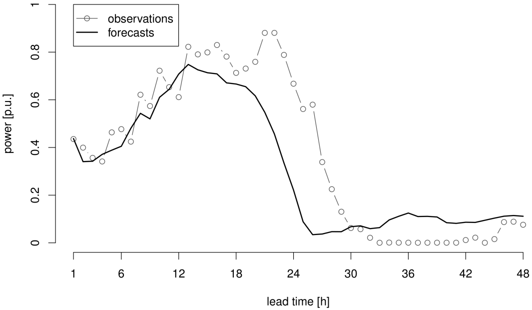
Various types of forecast error patterns

- Errors in renewable energy generation (but also load, price, etc.) are most often driven by weather forecasts errors

- Typical error patterns are:
  - **amplitude errors** (left, below)
  - **phase errors** (right, below)



Forecast issued on 29 March 2003 (12:00)



Forecast issued on 6 November 2002 (00:00)

Quantitative analysis and the forecast error

- For continuous variables such as renewable energy generation (but also electricity prices or electric load for instance)
  - **qualitative** analysis ought to be complemented by a **quantitative** analysis
  - these are based on *scores* and *diagnostic tools*

The base concept is that of the **forecast error**:

$$\varepsilon_{t+k|t} = y_{t+k} - \hat{y}_{t+k|t}, \qquad -\mathrm{P}_n \leq \varepsilon_{t+k|t} \leq \mathrm{P}_n$$

where

- $\hat{y}_{t+k|t}$ is the forecast issued at time t for time $t+k$
- y_{t+k} is the observation at time $t+k$
- P_n is the nominal capacity of the wind farm

- It can be calculated
  - directly for the quantity of interest
  - as a *normalized* version, for instance by dividing by the nominal capacity of the wind farm if evaluating wind power forecasts:

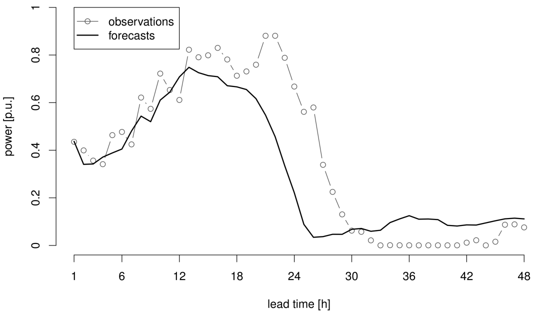
$$\varepsilon_{t+k|t} = \frac{y_{t+k} - \hat{y}_{t+k|t}}{\mathrm{P}_n}, \quad -1 \leq \varepsilon_{t+k|t} \leq 1$$

Forecast error: examples

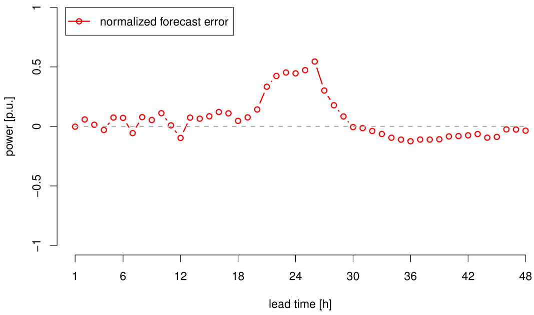
Example 1: If the 24-ahead prediction for Klim is of 18 MW, while the observation is 15.5MW

- $\varepsilon_{t+k|t} = -2.5$MW (if not normalized)
- $\varepsilon_{t+k|t} = -0.119$ (or, -11.9%, if normalized)

Example 2: forecast issued on the 6 November 2002 (00:00)



Forecast and observations



Corresponding forecast errors

(Note that we prefer to work with normalized errors from now on...)

Scores for point forecast verification

- One cannot look at all forecasts, observations, and forecasts errors over a long period of time
- *Scores* are to be used to summarize aspects of forecast accuracy...

The most common scores include, as function of the lead time k:

- **bias** (or Nbias, for the normalized version)

$$\text{bias}(k) = \frac{1}{T} \sum_{t=1}^{T} \varepsilon_{t+k|t}$$

- **Mean Absolute Error (MAE)** (or NMAE, for the normalized version)

$$\text{MAE}(k) = \frac{1}{T} \sum_{t=1}^{T} |\varepsilon_{t+k|t}|$$

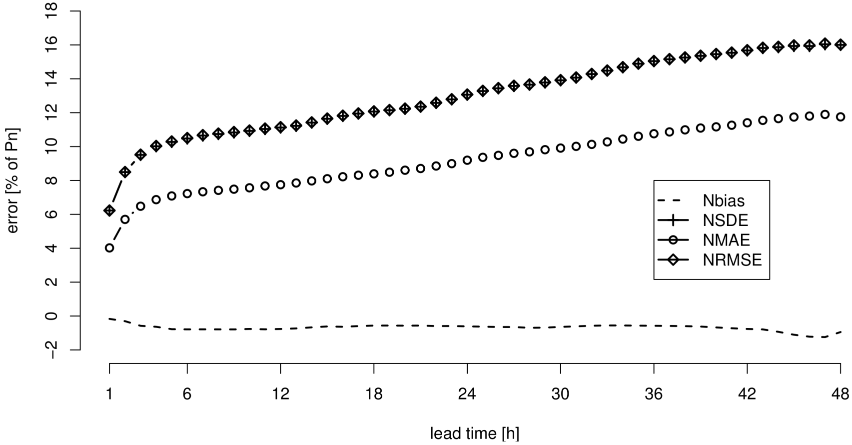
- **Root Mean Square Error (RMSE)** (or NRMSE, for the normalized version)

$$\text{RMSE}(k) = \left[\frac{1}{T} \sum_{t=1}^{T} \varepsilon_{t+k|t}^2 \right]^{\frac{1}{2}}$$

- MAE and RMSE are *negatively-oriented* (the lower, the better)
- Let us discuss their advantages and drawbacks...

Example: calculating a few scores at Klim

- Period: 1.7.2012 - 31.12.2012
- Forecats quality necessarily degrades with further lead times



- For instance, for 24-ahead forecasts:
  - bias is close to 0, while NMAE and NRMSE are of 8% and 12%, respectively
  - on average, there is ± 1.68 MW between forecasts and measurements

Comparing against benchmark approaches

- Forecasts from advanced methods are expected to outperform simple benchmarks!
- Two typical benchmarks are (to be further discussed in a further Module):
  - **Persistence** ("what you see is what you get"):

    $$\hat{y}_{t+k|t} = y_t, \quad k = 1, 2, \ldots$$

  - **Climatology** (the "once and for all" strategy):

    $$\hat{y}_{t+k|t} = \bar{y}_t, \quad k = 1, 2, \ldots$$

    where $\bar{y}_t$ is the average of all measurements available up to time t

A *skill score* informs of the relative quality of a method vs. a relevant benchmark, for a given lead time k:

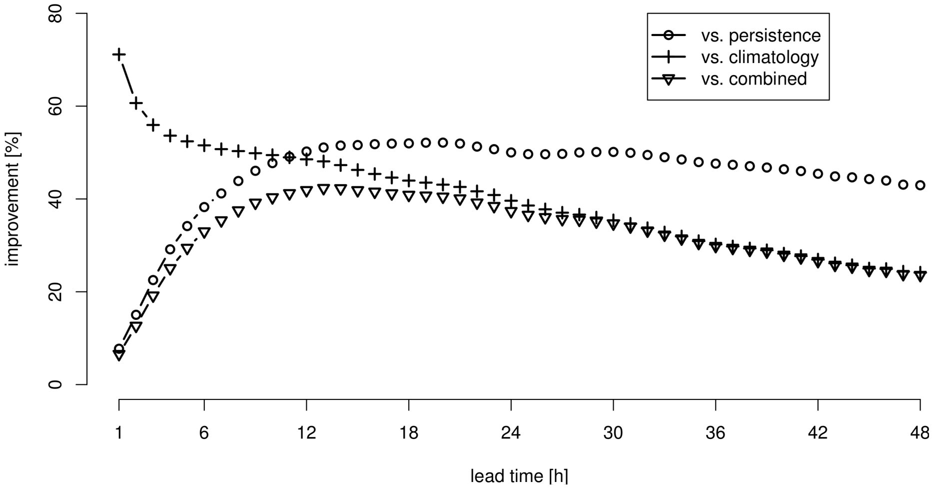
$$\text{SSc}(k) = 1 - \frac{\text{Sc}_{\text{adv}}(k)}{\text{Sc}_{\text{ref}}(k)}, \quad \text{SSc} \leq 1 \qquad \text{(possibly expressed in \%)}$$

where

- 'Sc' can be MAE, RMSE, etc.,
- 'Sc_{adv}' is score value for the advanced method, and
- 'Sc_{ref}' is for the benchmark

Example: benchmarking at Klim

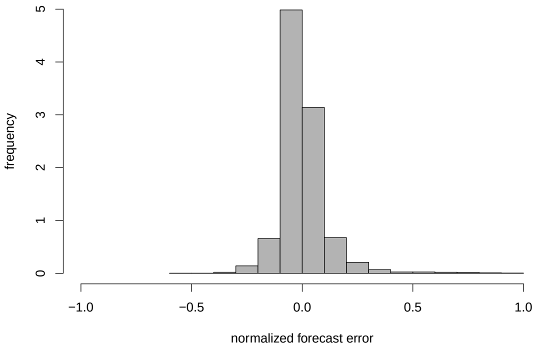
- Great! My forecasts are way better than the benchmarks considered (in terms of RMSE)



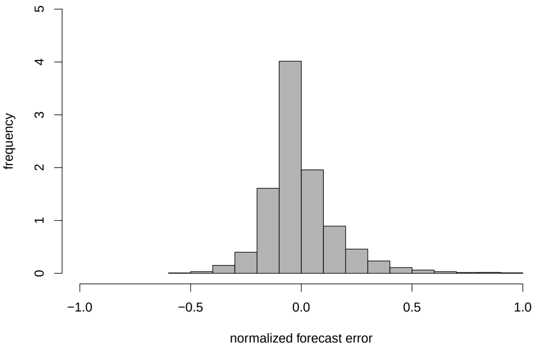
- Additional comments:
  - *persistence* is difficult to outperform for short lead times
  - *climatology* is difficult to outperform for longer lead times

Diagnostic tools based on error distributions

- Scores are *summary statistics*
- They only give a partial view of forecast quality
- **A full analysis of error distributions may tell you so much more!**



24-ahead forecasts
1 July 2002 - 31 December 2002

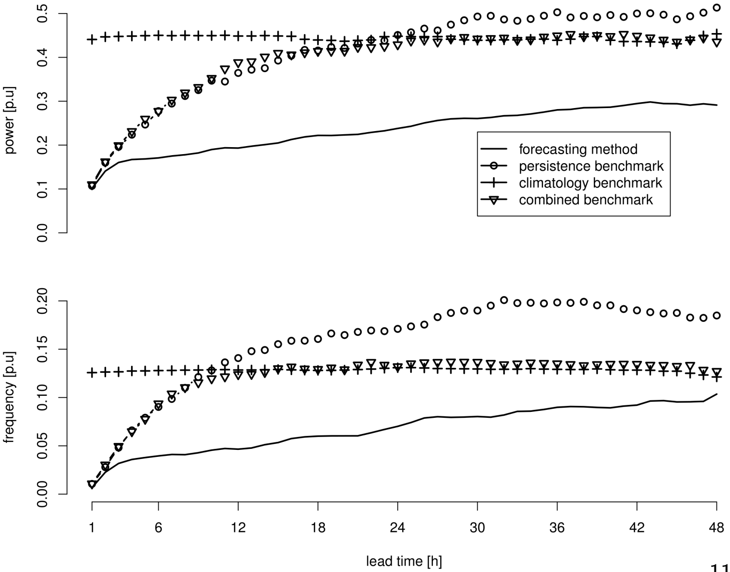


36-ahead forecasts
1 July 2002 - 31 December 2002

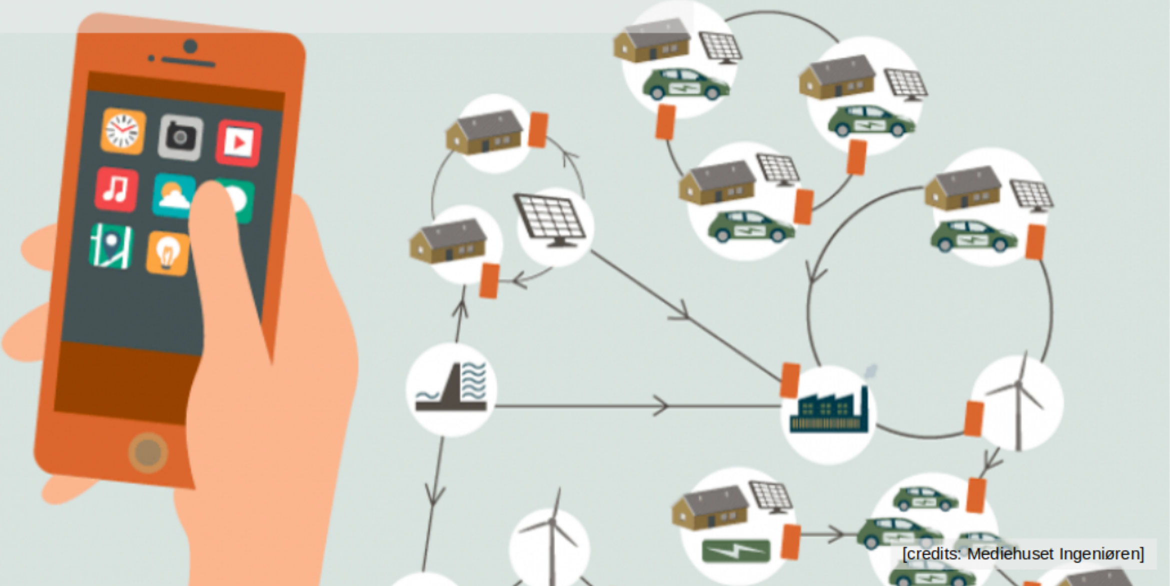
Analysis of "extreme" errors

- For risk management reason, you may be interested in knowing more about extreme forecast errors

- For the test case of Klim and the same period:
  - The upper plot informs of *the value X (in % of P_n) for which 95% of prediction errors are less than X*
  - The lower plot tells about *the percentage of prediction errors being greater than 0.2 P_n* (20% of the nominal capacity)



Use the self-assessment quizz to check your understanding!



[credits: Mediehuset Ingeniøren]