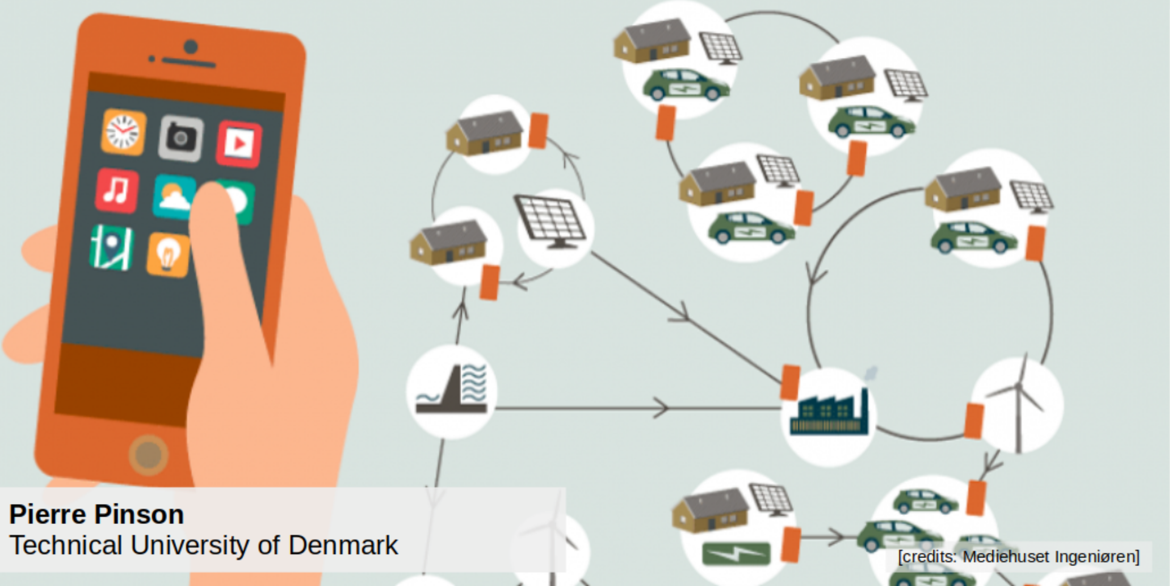
Module 9 – Renewable Energy Forecasting: First Steps

9.3 Digging in the data



[credits: Mediehuset Ingeniøren]

Pierre Pinson
Technical University of Denmark

What to look for in the data?

What do we have here...?

- **measurements**, i.e.,
  - power measurements (y_t) - Remember that only past measurements can be used!
- **weather forecasts**, i.e.,
  - wind speed ($\hat{u}_{t+k|t}$)
  - wind direction forecasts ($\hat{\theta}_{t+k|t}$)
  - temperature forecasts ($\hat{T}_{t+k|t}$)
- **different variations of those** could be used since the relationship between meteorological variables and power is nonlinear, e.g.,
  - power of wind speed: $\hat{u}^2_{t+k|t}$, $\hat{u}^3_{t+k|t}$, etc.
  - harmonics of wind direction: $\cos\left(\frac{2\pi\hat{\theta}_{t+k|t}}{360}\right)$, $\sin\left(\frac{2\pi\hat{\theta}_{t+k|t}}{360}\right)$, etc.
- we also know the **hour of the day** (h_t), or the lead time k, which could be useful... (though not used here)

Let us call all these variables x_j ($j = 1, \dots, m$), and also nickname them "features"

We can still write a linear regression...

- Remember that a linear relation between the x_j variables and y can be written as

$$y_i = \beta_0 + \sum_{j=1}^{m} \beta_j x_{j,i} + \varepsilon_i, \quad i = t-n, \ldots, t$$

(or, equivalently: $\quad y_i = \boldsymbol{\beta}^\top \mathbf{x}_i + \varepsilon_i$)

where

- y_i is still your response variable (say, wind power generation) observed at time i
- $x_{j,i}$ is the corresponding value for the j^{th} explanatory variable ($j = 1, \ldots, m$, example wind speed forecast used as input)
- β_j is the model parameter for the j^{th} explanatory variable
- ε_i is a noise term, which you may see as our forecast error we want to minimize

This linear regression model can be reformulated in a more compact form as

$$y_i = \boldsymbol{\beta}^\top \mathbf{x}_i + \varepsilon_i, \quad i = t-n, \ldots, t$$

with

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_m \end{bmatrix}, \qquad \mathbf{x}_i = \begin{bmatrix} 1 \\ x_{1,i} \\ \cdots \\ x_{m,i} \end{bmatrix}$$

Estimation and feature selection

- We need to find the best value of $\boldsymbol{\beta}$ that describes this cloud of point, **but also using a minimum of variables** (*parsimony principle*)
- LS-estimation is not very good for that, as the number of variables becomes high... the **LASSO version should be used instead**

The **LASSO estimate** $\hat{\boldsymbol{\beta}}$ of the linear regression model parameters is given by

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \frac{1}{\sqrt{n}\lambda} \sum_i {\varepsilon_i}^2 + \sum_j |\beta_j|$$

with λ a so-called *regularization parameter*, and

$$\hat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \cdots \\ \beta_m \end{bmatrix}, \quad \mathbf{X} = \begin{bmatrix} 1 & x_{1,t-n} & \dots & x_{j,t-n} \\ 1 & x_{1,t-n+1} & \dots & x_{j,t-n+1} \\ \vdots & \vdots & \dots & \vdots \\ 1 & x_{1,t} & \dots & x_{j,t} \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_{t-n} \\ y_{t-n+1} \\ \vdots \\ y_t \end{bmatrix}$$

- As before, some functions in R/Matlab can do it for you!

Example based on a set of features

- Our proposal model has the following form, for a given lead time k,

$$\begin{aligned} y_i = & \; \beta_{k,0} + \beta_{k,1}\hat{u}_{i|i-k} + \beta_{k,2}\hat{u}^2_{i|i-k} + \beta_{k,3}\hat{u}^3_{i|i-k} \\ & + \beta_{k,4}\cos\left(\frac{2\pi\hat{\theta}_{t+k|t}}{360}\right) + \beta_{k,5}\sin\left(\frac{2\pi\hat{\theta}_{t+k|t}}{360}\right) \\ & + \beta_{k,6}\hat{T}_{i|i-k} + \beta_{k,7}y_{i|i-k} + \varepsilon_i \end{aligned}$$

where

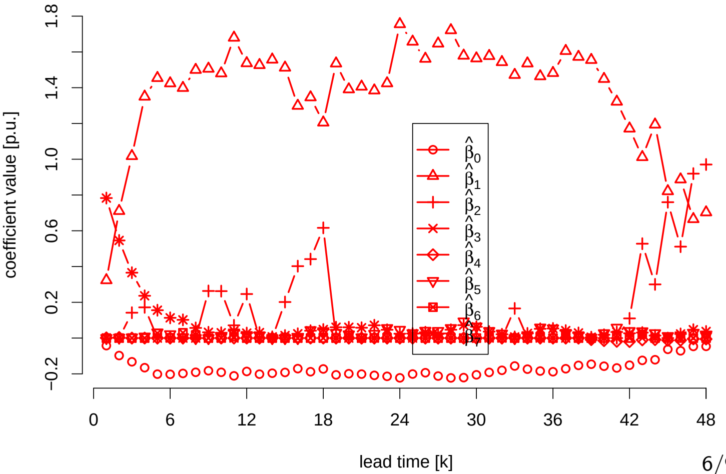
- we have 8 model parameters to estimate, for each lead time k
- the weight given to each of these features therefore varies with the lead time k
- ε_i is a noise term, which you may see as our forecast error we want to minimize

Estimation of the model coefficients

- For the example of the 28 April 2002 (as for the other examples),
  - the necessary vectors and matrices are formed, with $n = 600$ last values
  - LASSO estimates of the estimates $\hat{\beta}_{k,j}$s are computed for every lead time k ($k = 1, \ldots, 48$)

- *Right:*
  Evolution of the estimated model parameters as a function of the lead time k

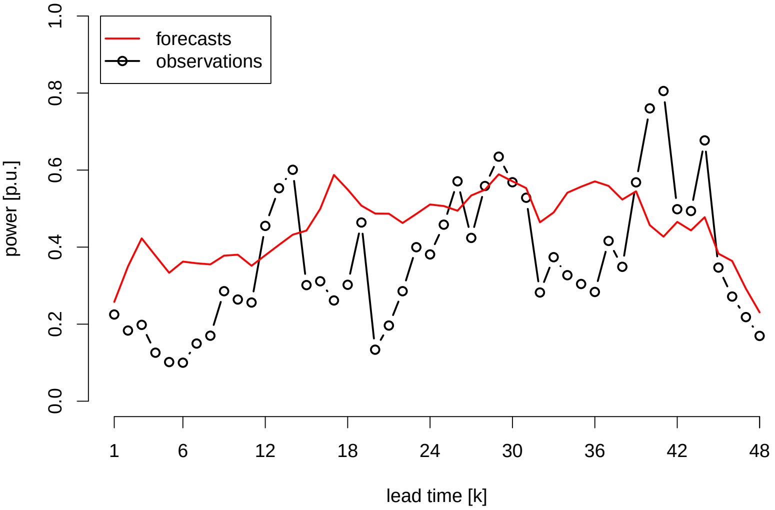
- Only a few features have parameters significantly different from 0: $\hat{u}_{t+k|t}$, $\hat{u}^2_{t+k|t}$ and y_t



The resulting forecast

- For the example of the 28 April 2002 (as before),
  - the necessary vectors and matrices are formed, with $n = 600$ last values

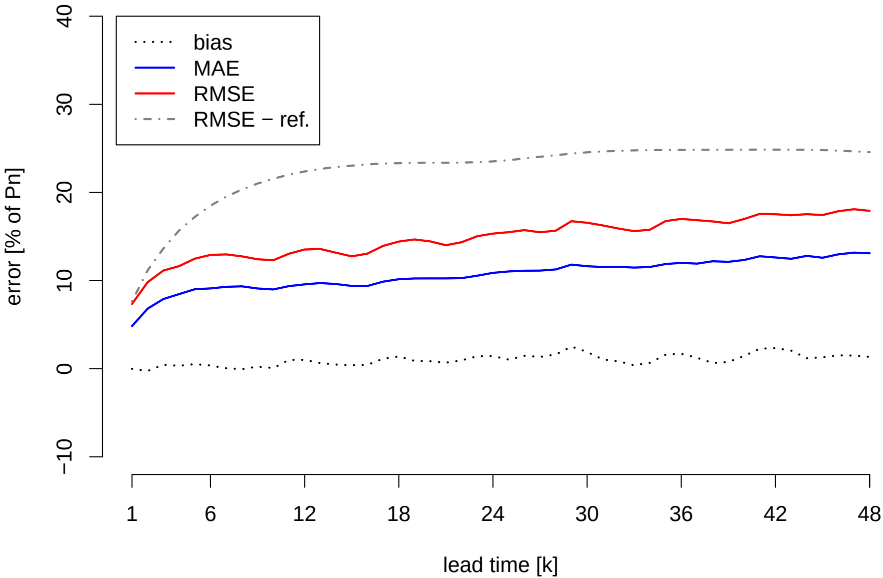
- *Right:*
  Example of a forecast for Klim with our more advanced model, issued on 28 April 2002, 00:00UTC



- Let us similarly apply that strategy for a whole sample year (2002), and analyse its performance

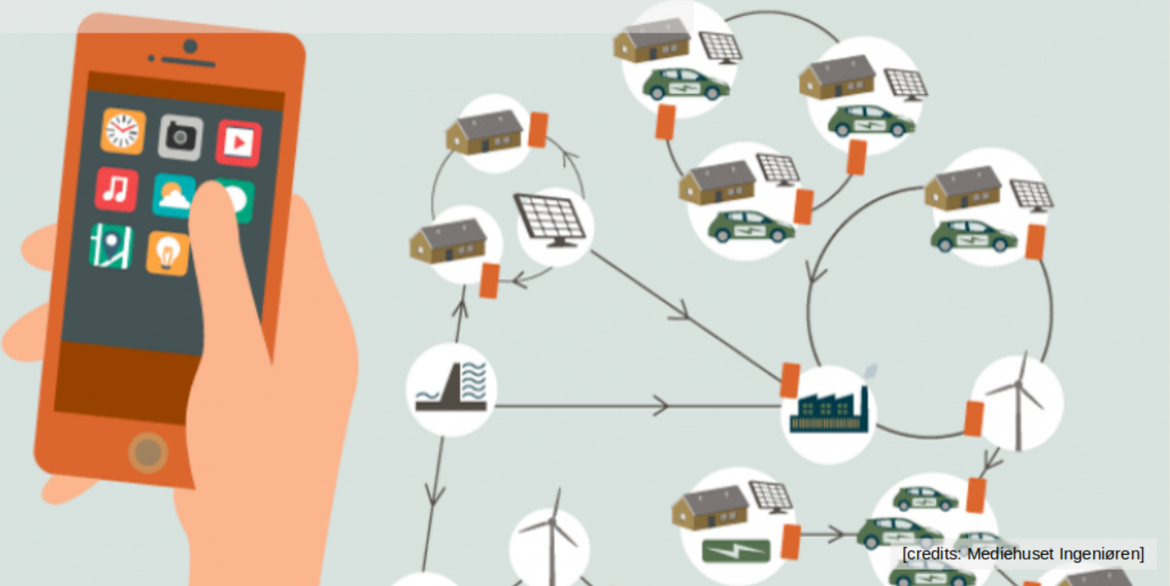
Evaluation of more advanced forecasts

- Various criteria: **bias**, **MAE**, **RMSE** (and RMSE comparison with our previous best forecasts)



- It seems we have substantially improved... Could still do better!!

Use the self-assessment quizz to check your understanding!



[credits: Mediehuset Ingeniøren]