Description of the onlineforecast R package

Peder Bacher
DTU Compute, Technical University of Denmark, DK-2800 Lyngby, Denmark

November 28, 2019

Contents

1 Statistical online forecast models

The aim of this chapter is to present a way to setup models, which can be fitted using weather forecasts as input – thus providing a setup for online forecasting. The general challenge is that weather forecasts are used, which leads to "overlapping" time series. For example, in many settings where hourly forecasts are needed, with a horizon up to 48 hours ahead, they must be updated every hour with the latest weather forecasts as input. This gives rise to two challenges: First the weather forecasts are not updated every hour, so we need ways to use the most recent weather forecasts, and second we need to set this up in order to fit the models in a way, which is both easy to use and effective, both in terms of forecast performance and computational resources.

Time series modeling is one of the major tools which has been used quite extensively both by engineering and scientific communities over last few decades to model the dynamic phenomenon or systems. The main aim of the time series modeling is to develop or apply rigorous statistical methods to capture the dynamical information present in the measured data. Time series models study the past observations of a predictor variables (also known as the feature variables of the time series) to develop an appropriate model which can describe the inherent structure of the time series as well as predict the response variable. This developed model is then finally used to make short or long term forecasts. Time series forecasting thus can be termed as the act of predicting the future by understanding the past (Box et al., 1976; Madsen, 2007). Time series forecasting is of indispensable importance to numerous practical fields such as business, economics, finance, science and engineering, etc. . Therefore, one must take proper care to identify a model with proper structure adequate to describe the dynamics of the underlying time series. Furthermore, an appropriate model estimation and validation criteria is of utmost importance for time series forecasting. A lot of efforts have been put by researchers over many years for the development of efficient model estimation algorithms to improve the forecasting accuracy. As a result, a plethora of methods for time series forecasting models have been reported in the literature.

Many approaches to load forecasting of different types of load are found in the literature. A good overview of references are given by (Mestekemper, 2011), who built load forecasting models using dynamic factor models. (Dotzauer, 2002) use a model based on the ambient temperature and a weekly pattern for forecasting of the heat load in district heating, i.e. the total heat load for many houses. (Zhou et al., 2008) use a grey-box model based on transfer functions for building thermal load prediction and validates it on a 50 floors multi purpose building.

Likewise, many approaches to solar power forecasting have been suggested during more than a decade. Some of the first literature is from Chowdhury and Rahman (1987) who make sub-hourly forecasts by normalizing with a clear sky model and using ARIMA models. Sfetsos and Coonick (2000) use NNs to make one-step predictions of hourly values of global irradiance and compare these with linear time series models that work by predicting clearness indexes. Heinemann et al. (2006) use satellite images for horizons below 6 hours, and in (Lorenz et al., 2007) numerical weather predictions (NWPs) for longer horizons, as input to NNs to predict global irradiance. This is transformed into solar power by a simulation

model of the PV system. Hocaoglu et al. (2008) investigate feed-forward NNs for one-step predictions of hourly values of global irradiance and compare these with seasonal AR models applied on solar power directly. Cao and Lin (2008) use NNs combined with wavelets to predict next day hourly values of global irradiance. Different types of meteorological observations are used as input to the models; among others the daily mean global irradiance and daily mean cloud cover of the day to be forecasted. Sharma et al. (2011) use weather forecasts to predict hourly solar intensity as a proxy for solar power generation. The study compares multiple regression techniques for generating prediction models, including linear least squares and Support Vector Machines (SVM) using multiple kernel functions. Furthermore, dimensionality reduction is explored using Principal Component Analysis (PCA). Similarly, Ragnacci et al. (2012) propose a model based on SVM for forecasting of PV power, where multiple techniques for dimensionality reduction of the input variables are exploited. These results clearly support the importance of proper feature and model selection. This paper takes advantage of the automated feature selection of the Gradient Boosted Regression Trees (GBRT). Marquez and Coimbra (2011) propose the use of Artificial Neural Networks (ANN) to forecast global horizontal irradiance (GHI) and direct normal irradiance (DNI) using weather forecasts as predictors. A preliminary feature selection is performed using a Genetic Algorithm (GA) and a Gamma Test. Pedro and Coimbra (2012) fit several forecasting models, which predict the hourly PV power generation for one and two hours ahead only using endogenous variables. The methods studied in the paper are among others Autoregressive Integrated Moving Average (ARIMA), k-Nearest-Neighbors (kNN), ANN and ANN optimized by Genetic Algorithms (ANN/GA). Tree based models have also been applied for probabilistic forecasting of solar power generation, e.g. Zamo et al. (2014) and Almeida et al. (2015) successfully apply Quantile Regression Forests to estimate quantiles of the PV power generation in order to produce probabilistic forecasts.

The forecasting model approach described in the present study are inspired by a setup used for forecasting of heat load in district heating, as described in (Nielsen et al., 2000) and (Nielsen and Madsen, 2006). It is applied in different settings for load forecasting by Bacher et al. (2009), Bacher et al. (2013b), Bacher et al. (2013a) and Vogler-Finck et al. (2017). The approach has been further developed to encompass the possibility to model many kind of functional relationships, both dynamics and non-linear. Thus now provides a sound framework for online forecasting of both load and generation in many settings. More specifically, the setup can also be used to fit the models suggested by Bacher et al. (2009) and extends even further, potentially also encompassing wind power forecast models presented by Nielsen et al. (2006) and Nielsen et al. (2011).

2 Notation

The notation in this section is following Madsen (2008) as much as possible.

2.1 Model notation

Generally, in statistical notation uppercase Latin letters are used to denote random variables, e.g. the simplest linear would model

$$Y_t = \beta_0 + \beta_1 u_t + \varepsilon_t \tag{1}$$

where

- the output Y_t is a random variable
- the input u_t is an input
- the error is a random variable $\varepsilon_t \sim N(0, \sigma^2)$ and i.i.d. (thus also Greek letters are denoting random variables, however for them upper- and lowercase are not distinguished)
- the parameters β_0 and β_1 are actually also random variables

The actual observations are denoted with small letters

$$y_t \tag{2}$$

and the predictions are

$$\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 u_t \tag{3}$$

where the hat ˆ indicates an estimated or predicted value.

Further, the realization of the error (ε_t) is the residual

$$\hat{\varepsilon}_t = e_t = y_t - \hat{y}_t \tag{4}$$

hence either denoted with a hat or with a lowercase Latin letter.

However, this is not always possible to keep this notation, firstly in physics uppercase Latin letters are used, e.g.

- T_t^{a} is the ambient temperature (°C)
- G_t is the global radiation (W m^{-2})
- W_t^{s} is the wind speed (m s^{-1})
- Q_t is the heat load (or heating energy or power (W) or energy (kW h))

as well as Greek letters

- Φ_t^{h} is the heating power (W)

Matrices and vectors are marked with bold.

2.2 Indexing using t and k

In the present document all time series considered are equidistant sampled and the sampling period is normalized to 1, hence the time t is simply an integer, which index the value of a variable at t. The same goes for horizon k, which index a prediction k steps ahead in time.

Further, in the generic setup a forecast at each time t is calculated for each horizon k up to n_k steps ahead. To achieve a notation which can deal with overlapping time series, a two dimensional index is needed. The most widespread notation is

$$u_{t+k|t} \tag{5}$$

which means: the value of variable x, at time $t+k$, and *conditional* the information available at time t. The bar conditional is indicated by the bar $|$.

Thus for $k > 0$ this is a forecast, usually an NWP of the input, however it can also be deterministic input, e.g. a Fourier series, which is known ahead in time (in this case the conditional notation ($|$) should not be kept).

For $k \leq 0$ some ambiguity is also possible, since observations can be used

$$u_{t+k|t} = u^{\text{obs}}_{t+k} \text{ for } k \leq 0 \tag{6}$$

however the NWP can also be used

$$u_{t+k|t} = u^{\text{nwp}}_{t+k|t} \text{ for } k \leq 0 \tag{7}$$

thus a value for the past time can come from an NWP, which was not available before time t.

These aspects will be more clear in later sections as it is explained how data is setup and updated in online settings.

3 Model input and output

Input, for which we have a forecast, e.g. NWP inputs, we set up, at time t, an input matrix for the variable with name **nm**. It holds for each time t *the latest available forecasts along the row*

$$\boldsymbol{u}^{\text{nm}}_t = \begin{array}{c} \begin{array}{cccccc} \textbf{k0} & \textbf{k1} & \textbf{k2} & \dots & \textbf{kxx} & \textbf{horizon/time} \end{array} \\ \begin{pmatrix} u^{\text{nm}}_{t_0|t_0} & u^{\text{nm}}_{t_0+1|t_0} & u^{\text{nm}}_{t_0+2|t_0} & \cdots & u^{\text{nm}}_{t_0+n_k|t_0} & t_0 \\ u^{\text{nm}}_{t_1|t_1} & u^{\text{nm}}_{t_1+1|t_1} & u^{\text{nm}}_{t_1+2|t_1} & \cdots & u^{\text{nm}}_{t_1+n_k|t_1} & t_1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ u^{\text{nm}}_{t-1|t-1} & u^{\text{nm}}_{t|t-1} & u^{\text{nm}}_{t+1|t-1} & \cdots & u^{\text{nm}}_{t-1+n_k|t-1} & t-1 \\ u^{\text{nm}}_{t|t} & u^{\text{nm}}_{t+1|t} & u^{\text{nm}}_{t+2|t} & \cdots & u^{\text{nm}}_{t+n_k|t} & t \end{pmatrix} \end{array} \tag{8}$$

where

- t is the counter of time for equidistant time points and the sampling period is 1 (note that it is not included in the matrix, it is simply the row number)
- t_0 is the first available time point
- n_k is the length of the forecasting horizon
- The column names are indicated above the matrix, they are simply a **k** concatenated with the value of k.

Hence, with a prediction horizon $n_k = 24$, having data from time $t = 1$, then at time $t = 100$ we would have the following matrix

$$\boldsymbol{u}_t^{\mathrm{nm}} = \begin{array}{c} \begin{array}{ccccc} \textbf{k0} & \textbf{k1} & \textbf{k2} & \dots & \textbf{k24} \end{array} \\ \begin{pmatrix} u^{\mathrm{nm}}_{1|1} & u^{\mathrm{nm}}_{2|1} & u^{\mathrm{nm}}_{3|1} & \dots & u^{\mathrm{nm}}_{25|1} \\ u^{\mathrm{nm}}_{2|2} & u^{\mathrm{nm}}_{3|2} & u^{\mathrm{nm}}_{4|2} & \dots & u^{\mathrm{nm}}_{26|2} \\ \vdots & \vdots & \vdots & & \vdots \\ u^{\mathrm{nm}}_{99|99} & u^{\mathrm{nm}}_{100|99} & u^{\mathrm{nm}}_{101|99} & \dots & u^{\mathrm{nm}}_{123|99} \\ u^{\mathrm{nm}}_{100|100} & u^{\mathrm{nm}}_{101|100} & u^{\mathrm{nm}}_{102|100} & \dots & u^{\mathrm{nm}}_{124|100} \end{pmatrix} \end{array} \begin{array}{c} \textbf{horizon/time} \\ 1 \\ 2 \\ \vdots \\ 99 \\ 100 \end{array} \tag{9}$$

This will actually allow for indexing using integers t indexing in the rows and k indexing in the columns.

This could for example be the forecasts of the ambient temperature

$$\boldsymbol{T}_t^{\mathrm{a}} = \begin{array}{c} \begin{array}{ccccc} \textbf{k0} & \textbf{k1} & \textbf{k2} & \dots & \textbf{kxx} \end{array} \\ \begin{pmatrix} T^{\mathrm{a}}_{t_0|t_0} & T^{\mathrm{a}}_{t_0+1|t_0} & T^{\mathrm{a}}_{t_0+2|t_0} & \dots & T^{\mathrm{a}}_{t_0+n_k|t_0} \\ T^{\mathrm{a}}_{t_1|t_1} & T^{\mathrm{a}}_{t_1+1|t_1} & T^{\mathrm{a}}_{t_1+2|t_1} & \dots & T^{\mathrm{a}}_{t_1+n_k|t_1} \\ \vdots & \vdots & \vdots & & \vdots \\ T^{\mathrm{a}}_{t-1|t-1} & T^{\mathrm{a}}_{t|t-1} & T^{\mathrm{a}}_{t+1|t-1} & \dots & T^{\mathrm{a}}_{t-1+n_k|t-1} \\ T^{\mathrm{a}}_{t|t} & T^{\mathrm{a}}_{t+1|t} & T^{\mathrm{a}}_{t+2|t} & \dots & T^{\mathrm{a}}_{t+n_k|t} \end{pmatrix} \end{array} \begin{array}{c} \textbf{horizon/time} \\ t_0 \\ t_1 \\ \vdots \\ t-1 \\ t \end{array} \tag{10}$$

All values where $k \geq 1$ naturally have to be forecasts of the inputs, e.g. for the ambient temperature these will come from a Numerical Weather Prediction (NWP).

If there are local observations available of the input the past values can be taken from the observations

$$\boldsymbol{T}_t^{\mathrm{a}} = \begin{array}{c} \begin{array}{cccccc} \textbf{k0} & \textbf{k1} & \textbf{k2} & \dots & \textbf{kxx} & \textbf{horizon/time} \end{array} \\ \begin{pmatrix} T^{\mathrm{a,obs}}_{t_0|t_0} & T^{\mathrm{a,nwp}}_{t_0+1|t_0} & T^{\mathrm{a,nwp}}_{t_0+2|t_0} & \dots & T^{\mathrm{a,nwp}}_{t_0+n_k|t_0} & t_0 \\ T^{\mathrm{a,obs}}_{t_1|t_1} & T^{\mathrm{a,nwp}}_{t_1+1|t_1} & T^{\mathrm{a,nwp}}_{t_1+2|t_1} & \dots & T^{\mathrm{a,nwp}}_{t_1+n_k|t_1} & t_1 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ T^{\mathrm{a,obs}}_{t-1|t-1} & T^{\mathrm{a,nwp}}_{t|t-1} & T^{\mathrm{a,nwp}}_{t+1|t-1} & \dots & T^{\mathrm{a,nwp}}_{t-1+n_k|t-1} & t-1 \\ T^{\mathrm{a,obs}}_{t|t} & T^{\mathrm{a,nwp}}_{t+1|t} & T^{\mathrm{a,nwp}}_{t+2|t} & \dots & T^{\mathrm{a,nwp}}_{t+n_k|t} & t \end{pmatrix} \end{array} \tag{11}$$

Note, that in most cases there is a bias between the NWPs and local observations, thus a model must be applied in a step before fitting the forecast models – such aspects are not dealt with in this report.

3.1 Updating an NWP input matrix

One can choose either to update the input matrix in each time step or when a new NWP is received.

4 Two-stage modelling procedure

In order to model non-linear functional relations between inputs and output a two-stage modelling procedure is used. This is a widespread approach, see Hastie et al. (2009), since it allows to fit complex models with robust and fast estimation techniques. First, in the *transformation stage*, some function of the inputs, e.g. low-pass filtering, spline or Fourier basis, etc., is be applied. Second, in the *regression stage*, a regression model is applied to fit a function between the transformed inputs and the output – the parameters are fitted for each horizon.

Lets go through it with a simple example, first the transformation stage

$$\text{Intercept:} \quad \mu_{t+k|t} = 1 \tag{12}$$

$$\text{Regressor 1:} \quad x^{\mathrm{nm1}}_{t+k|t} = f_1(\boldsymbol{u}^{\mathrm{nm1}}_t; \boldsymbol{\theta}_{\mathrm{nm1}}) \tag{13}$$

$$\text{Regressor 2:} \quad x^{\mathrm{nm2}}_{t+k|t} = f_2(\boldsymbol{u}^{\mathrm{nm2}}_t; \boldsymbol{\theta}_{\mathrm{nm1}}) \tag{14}$$

and then the regression stage

$$Y_{t+k|t} = \beta_{0,k}\mu_{t+k|t} + \beta_{1,k}x^{\mathrm{nm1}}_{t+k|t} + \beta_{2,k}x^{\mathrm{nm2}}_{t+k|t} + \varepsilon_{t+k|t} \tag{15}$$

Thus, this model has:

- Intercept: $\mu_{t+k|t}$
- Two inputs (matrices): $\boldsymbol{u}^{\text{nm1}}_{t+k|t}$ and $\boldsymbol{u}^{\text{nm1}}_{t+k|t}$
- Output: $Y_{t+k|t}$
- Transformation parameters (vectors): $\boldsymbol{\theta}_{\text{nm1}}$ and $\boldsymbol{\theta}_{\text{nm2}}$
- Regression parameters for each horizon k: $\beta_{0,k}$, $\beta_{1,k}$ and $\beta_{2,k}$

The regression parameters are estimated with a closed form scheme, most simple is a linear least squares regression, however this can be extended to a recursive or local version, more on this later. The transformation parameters must be estimated with some heuristic optimization.

4.1 Transformations stage

In the first stage the inputs are transformed with some function, mostly either a recursive function (filtering to model a dynamical system) or some basis function (splines or Fourier series applied to model non-linear functions).

4.1.1 Filtering

When modeling the output of a (linear) dynamical system a "trick" is to apply a filter directly on the input (instead of applying an ARMAX model (Madsen, 2008)).

The input time series is filtered with a transfer function

$$x_t = H(B; a)u_t \tag{16}$$

where the simplest 1st order low-pass with stationary gain of 1 is

$$H(B; a) = \frac{1 - a}{1 - aB} \tag{17}$$

thus the low-pass filtered series becomes

$$x_t = \frac{(1 - a)u_t}{1 - ax_{t-1}} \tag{18}$$

thus we have a filter coefficient a between 0 and 1, which must be tuned to match the particular time constant of the linear system.

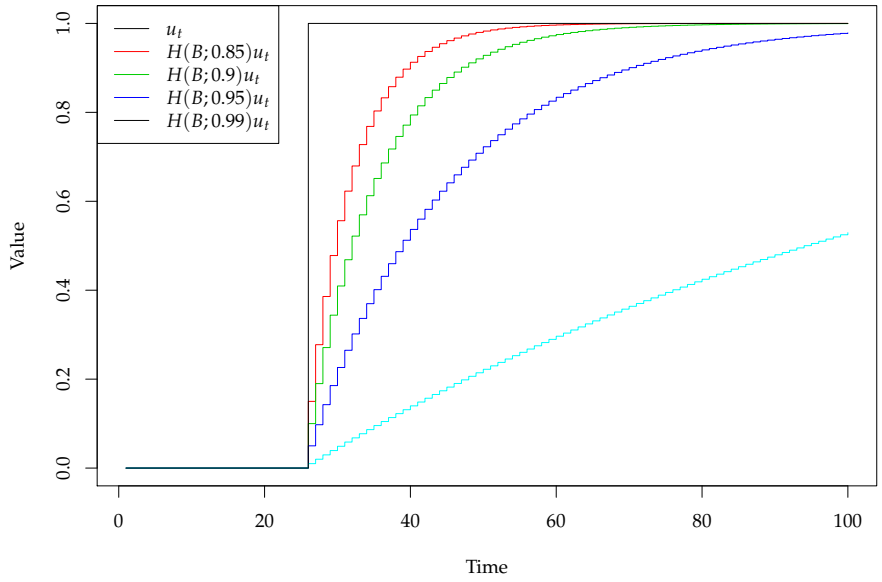
When filtering is applied in the transformation stage (see Equation (12) to input matrices ($\boldsymbol{u}_t$ defined in Eq. (8)), the filter can be applied in two ways: along the $\boldsymbol{u}_t$ columns

$$x_{t+k|t} = \frac{(1 - a)u_{t+k|t}}{1 - ax_{t-1+k|t-1}} \tag{19}$$

or along the rows

$$x_{t+k|t} = \frac{(1-a)u_{t+k|t}}{1-ax_{t+k-1|t}} \tag{20}$$

To illustrate the effect of low-pass filtering the following plot shows the response to a step function for different filter coefficients a:



It is clearly seen that the responses are exponential functions with different time constants depending on the value of a, such that a higher value results in a slower response. This filter is equivalent to the transfer function of a single resistor single capacitor system (a first order ARX model with a stationary gain of 1). If needed higher order filter can be selected for the transformation of data.

4.1.2 Base splines

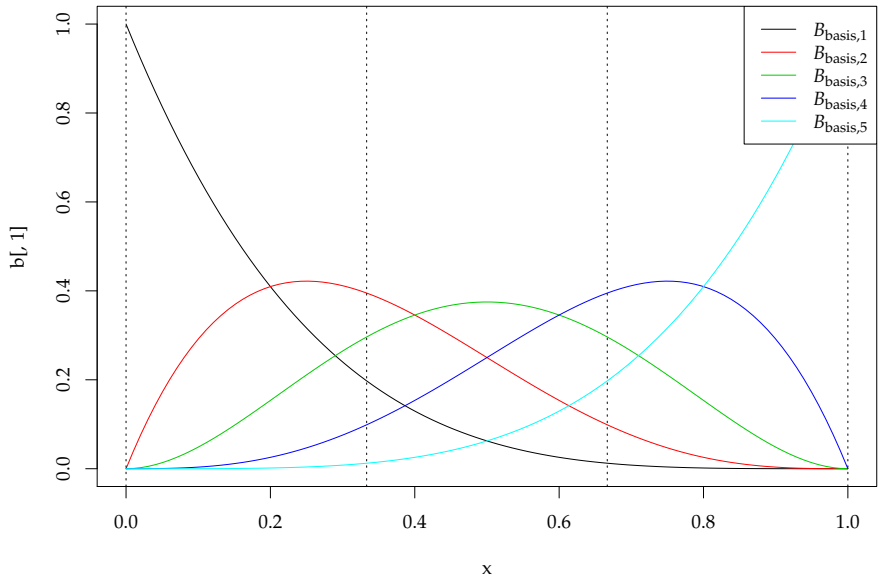
A wide spread approach to model non-linear functional relations is to apply spline basis functions (Hastie et al., 2009). The basic idea is to "resolve" a single input time series into several time series, which levels depends on the input time series. Thereafter a linear combination of the time series (fitted in the regression stage) results in a non-linear spline function of the input time series.

Thus, the spline basis function

$$\boldsymbol{x}_t = f_{\text{bspline}}(u_t; n_{\text{deg}}) \tag{21}$$

where n_{deg} is the degree of the piecewise polynomial function, hence a higher n_{deg} results in a more "flexible" function. Note that $\boldsymbol{x}_t$ is a vector, since the transformation results in n_{deg} variable. The places where the piecewise polynomial meet

are known as knots. The key property of spline functions is that they and their derivatives may be continuous, depending on the multiplicities of the knots. The input is resolved with spline basis functions, e.g. for an input in the interval $[0, 1]$ the basis splines for $n_{\text{deg}} = 4$:

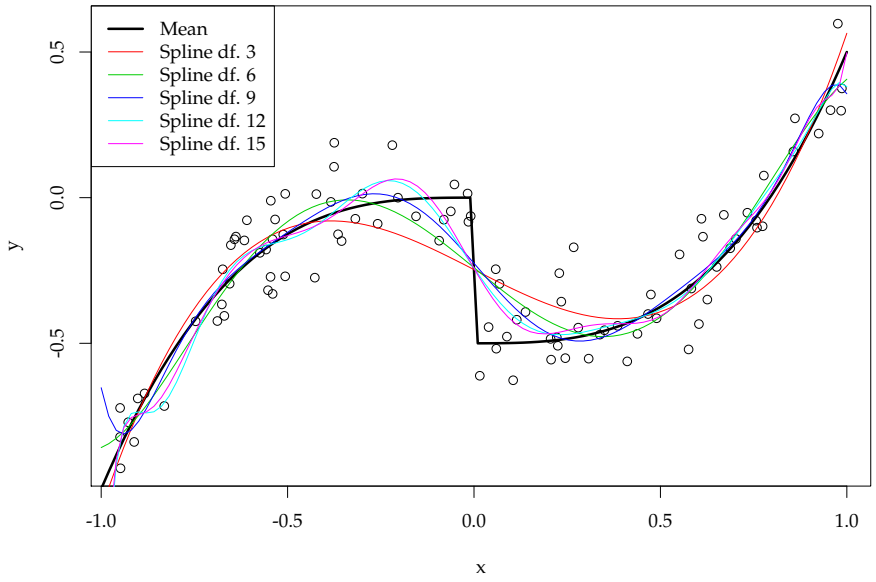


where the vertical dashed horizontal lines marks the knot points (must be set in some way, usually set as equidistant quantiles of values of u_t).

An example of resulting a spline functions, which can be fitted using spline basis functions is presented here. First a non-linear function with some added noise is simulated

$$Y_i = \begin{cases} u_i^3 + \varepsilon_i \text{ for } u_i \leq 0 \\ u_i^3 - 0.5 + \varepsilon_i \text{ for } u_i > 0 \end{cases} \tag{22}$$

where $\varepsilon \sim N(0, 0.1^2)$ and i.i.d.. A sample of 100 observations is simulated and spline basis functions of increasing degree is generated and used as input to a linear regression model. The resulting spline functions modeled is seen in the plot below. It is clear that there is a balance between bias and variance: A too low degree results in an under-fitted model (not able to "bend" enough), while a too high degree results in an over-fitted model (bends to much). The degrees of freedom can be optimized using a cross-validation approach.



4.1.3 Fourier series

In order to model periodic phenomena a linear combination of Fourier series is a very effective approach.

We use the notation

$$\boldsymbol{x}_t = f_{\text{fs}}(u_t; n_{\text{har}}) \tag{23}$$

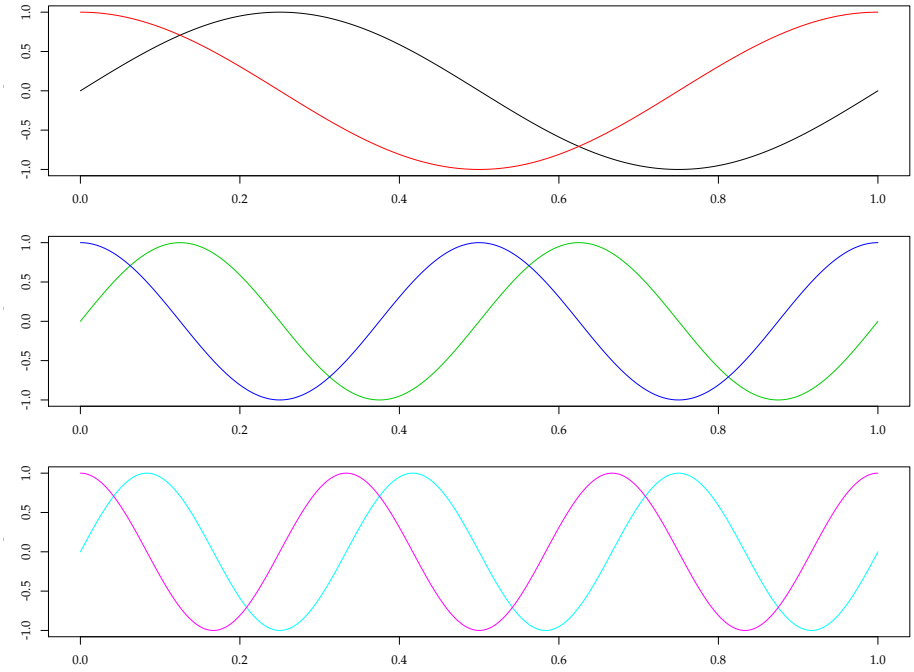
where n_{har} is the number of harmonic pairs included, hence $\boldsymbol{x}_t$ is vector of length $2n_{\text{har}}$, which is then linearly combined in the regression stage. Exemplified with time as input $u_t = t$ (which is often done when modelling period, e.g. diurnal or yearly phenomena)

$$\left[\sin\left(\frac{2\pi}{t_{\text{per}}}t\right) \quad \cos\left(\frac{2\pi}{t_{\text{per}}}t\right) \quad \sin\left(\frac{2\cdot 2\pi}{t_{\text{per}}}t\right) \quad \cos\left(\frac{2\cdot 2\pi}{t_{\text{per}}}t\right)\right. \tag{24}$$

$$\ldots \tag{25}$$

$$\left.\sin\left(\frac{n_{\text{har}}2\pi}{t_{\text{per}}}t\right) \quad \cos\left(\frac{n_{\text{har}}2\pi}{t_{\text{per}}}t\right)\right] \tag{26}$$

An example of Fourier basis functions is plotted below for $n_{\text{har}} = 3$, hence 3 pairs of harmonics and a period $t_{\text{per}} = 1$.



4.1.4 Combined transformations

It is perfectly possible to combine transformations to create models involving complicated functions. E.g. first a low-pass filter and then a spline basis

$$\boldsymbol{x}_t = f_{\text{bspline}}(H(B;a)u_t; n_{\text{deg}}) \tag{27}$$

or basis splines and which are then low-pass filtered

$$\boldsymbol{x}_t = f_{\text{bspline}}(u_t; n_{\text{deg}}) \tag{28}$$

Thus in both cases the result is multiple variables ($\boldsymbol{x}_t$ is a vector) and transformation parameters are $\boldsymbol{\theta} = [a, n_{\text{deg}}]$.

4.2 Regression stage

In the regression stage the coefficients are estimated for each horizon k. For a linear least squares regression, this is written by

$$Y_{t+k|t} = \beta_{0,k} + \beta_{1,k} x_{t+k|t} + \varepsilon_{t+k|t} \tag{29}$$

where

- $Y_{t+k|t}$ is a random variable (in total there will be $t \cdot n_k$ of them at time t when first time was $t = 1$)
- $u_{t+k|t}$ is the input variable
- $\varepsilon_t \sim N(0, \sigma^2)$ and i.i.d. is a random variable (again there will be $t \cdot n_k$ of them at time t when first time was $t = 1$)

- $\beta_{0,k}$ and $\beta_{1,k}$ are the parameters for horizon k

The model is fitted separately for each horizon k using only past data.

Regarding the normal assumption of the error, is not very important, since first of all the least squares method ensures the best estimation of the conditional mean, which is often the wanted and optimal point prediction, see Madsen (2008). Regarding the i.i.d. assumption of the errors, this should be checked with the Auto-Correlation Function (ACF) for the one-step ahead residuals, as well as the Cross-Correlation Function (CCF) between the one-step ahead residuals and the inputs, as done by Bacher et al. (2013b). textcolorblueOn of the basic problems is the selction of the optimal model from competing linear regression models. One of the frequently used criteria for model selection is K-cross cross-validation, where $K \in 1 \cdots N_K$, and N_K .

The k step predictions are

$$\hat{y}_{t+k|t} = \hat{\beta}_{0,k} + \hat{\beta}_{1,k} x_{t+k|t} \tag{30}$$

Note, that the hat is reserved for predictions and estimates calculated using the statistical model, thus the hat is not on the inputs, which are however often predictions (NWPs). If models are fitted in several stages, e.g. the inputs in the final model are actually first predicted using another model, then the hat is removed on the inputs in the final model.

In the above model there are in total $t \cdot n_k$ output (Y) random variables, as well as there are $t \cdot n_k$ error random variables.

Following regression methods are available:

- Least squares
- Recursive least squares (RLS) (ref)

however it is planned to also implement:

- Kernel regression (local fitting)
- Quantile regression (estimate quantiles)

The latter opens up the possibilities to calculate probabilistic forecasts (ref), as well as carry out normalization and Copula transformations, which can be very useful for spatio-temporal forecast models (ref, se irpwind raport).

One note is that when using a recursive update scheme, e.g. RLS, then the parameters are changing over time, which will be indicated with a t on the parameters

$$Y_{t+k|t} = \beta_{0,k,t} + \beta_{1,k,t} x_{t+k|t} + \varepsilon_{t+k|t} \tag{31}$$

4.3 Forecast model notation

In this section a suggestion on how to write full description of a forecast model is presented, together with simplified notation, which can be very useful in articles and presentations. Note that some variables are noted in bold font indicating that they are vectors.

Full notation of the transformation stage

$$\text{Intercept:} \quad \mu_{t+k|t} = 1 \tag{32}$$

$$\text{Periodic:} \quad \boldsymbol{x}^{\text{per}}_{t+k|t} = f_{\text{fs}}(t; n_{\text{har}}) \tag{33}$$

$$\text{Part 1:} \quad x^{\text{nm1}}_{t+k|t} = H(B; a) u^{\text{nm1}}_{t+k|t} \tag{34}$$

$$\text{Part 2:} \quad \boldsymbol{x}^{\text{nm23}}_{t+k|t} = f_{\text{bspline}}(u^{\text{nm2}}_{t+k}; n_{\text{deg}}) u^{\text{nm3}}_{t+k|t} \tag{35}$$

$$\text{Part 3:} \quad \boldsymbol{x}^{\text{nm4}}_{t+k|t} = u^{\text{nm4}}_{t} \tag{36}$$

and the regression stage

$$Y_{t+k|t} = \beta_{0,k}\mu_{t+k|t} + \boldsymbol{\beta}_{1,k}\boldsymbol{x}^{\text{per}}_{t+k|t} + \beta_{2,k}x^{\text{nm1}}_{t+k|t} + \boldsymbol{\beta}_{3,k}\boldsymbol{x}^{\text{nm23}}_{t+k|t} + \varepsilon_{t+k|t} \tag{37}$$

The model inputs are:

- t is simply the time value
- $u^{\text{nm1}}_{t+k|t}$ some forecast input (e.g. NWP variable)
- u^{nm2}_{t+k} some calculated value (e.g. time of day)
- $u^{\text{nm3}}_{t+k|t}$ some forecast input (e.g. NWP variable)
- u^{nm4}_{t} some value known at time t (e.g. an observed variable)

The transformation parameters are

$$\boldsymbol{\theta} = (n_{\text{har}}, a, n_{\text{deg}}) \tag{38}$$

which must be set or optimized (heuristically). In practice, in order to find a good set of parameters, different initializations can be tried along with different optimazation algorithms.

The regression coefficients are

$$\boldsymbol{\beta}_k = \left[\beta_{0,k} \;\; \beta_{1,1,k} \;\; \beta_{1,2,k} \;\; \cdots \;\; \beta_{1,2n_{\text{har}},k} \;\; \beta_{2,k} \;\; \beta_{3,1,k} \;\; \beta_{3,2,k} \;\; \cdots \;\; \beta_{3,n_{\text{deg}},k} \;\; \beta_{4,k}\right] \tag{39}$$

The full notation where a model is specified in all details can be cumbersome and thus it can be simplified by writing it in a single equation. In all simplifications suggested below, it should then be clearly stated, what is implicit or referenced to a full notation of the model.

The first suggested simplified form is

$$Y_{t+k|t} = \mu_k + f_k^{\text{fs}}(t; n_{\text{har}}) + H_k(B; a)u_{t+k|t}^{\text{nm1}} + f_k^{\text{bspline}}(u_{t+k}^{\text{nm2}}; n_{\text{deg}})u_{t+k|t}^{\text{nm3}} + \beta_k u_t^{\text{nm4}} + \varepsilon_{t+k|t} \tag{40}$$

thus the transformation and regression stage is written implicitly.

To simplify this even more it is suggested to write

$$Y = \mu + f_{\text{fs}}(t; n_{\text{har}}) + H(B; a)u_{\text{nm1}} + f_{\text{bspline}}(u_{\text{nm2}}; n_{\text{deg}})u_{\text{nm3}} + \beta u_{\text{nm4}} + \varepsilon \tag{41}$$

hence removing time and horizon indexing.

Finally, the most simplified notation suggested is

$$Y = \mu + f_{\text{fs}}(t) + H(B)u_{\text{nm1}} + f_{\text{bspline}}(u_{\text{nm2}})u_{\text{nm3}} + \beta u_{\text{nm4}} + \varepsilon \tag{42}$$

hence removing also the time and horizon indexing.

See the following section for use of the notation in a specific context.

5 Examples

Two examples of forecasting models are presented: One for heat load in single family houses and one for observed global radiation.

5.1 Data

Data for the forecasting examples is taken from a data set collected in Sønderborg, Denmark. It comprises heat load measurements for sixteen houses, together with local climate observations and weather forecasts (NWPs). The houses are generally built in the sixties and seventies, with a floor plan in the range of 85 to 170 m^2, and constructed in bricks. For each house only the total heat load, including both space heating and hot tab water heating, is available. The climate observations are measured at the local district heating plant within 10 kilometers from the houses. The NWPs are from the HIRLAM-S05 model and provided by the Danish Meteorological Institute. All times are in UTC and the time stamp for average values are set to the end of the time interval.

5.1.1 Heat load measurements

The houses are typical Danish single family houses from the sixties and seventies. Only houses with radiator heating is considered. A single signal for each house is used, which consist of both the energy for space heating and hot water heating. The heat load measurements consist of 10 minutes average values. Time series plots over the entire period, spanning nearly two and a half years, for four of the houses are shown in Figure 1. Also shown, with red lines, is the distribution

Figure 1: The heat load for four selected houses over the entire period, which is nearly spanning two and a half years. The red lines are estimates of the $0\%, 2\%, \ldots, 98\%, 100\%$ quantiles, which indicate the distribution of the heat load at a given time.

over time, which are estimates of the $0\%, 2\%, \ldots, 98\%, 100\%$ quantiles. They are estimated using local quantile regression (Koenker, 2005), where the weighting is local in time. They clearly indicate that the distribution of the heat load is heavily skewed, for example only two percent of the values are between the two upper lines, which cover more than half of the range. The reason for this skewness is that the heat load for water heating consists of high frequency spikes added to the more slowly varying space heating signal. The highest peaks are from showers and cause the high skewness.

The time series are resampled into hourly average values. The hourly heat load for a single house is denoted by

$$\{Q_t;\ t = 1, \ldots, N\} \tag{43}$$

where $N = 21144$ and the unit is kW. Notice that no distinguishment in between the houses is used in the notation, but when the results are presented the house number, ranging from 1 to 16, is clearly stated.

5.1.2 Local climate observations

The local climate observations are from a weather station at the district heating plant in Sønderborg, which is less than 10 kilometers from the houses. The data is resampled to hourly average values and the following time series are used:

$$\begin{aligned} \text{Ambient temperature: } & \left\{T_t^{\text{a,obs}}; t = 1, \ldots, N\right\} \\ \text{Global radiation: } & \left\{G_t^{\text{obs}}; t = 1, \ldots, N\right\} \\ \text{Wind speed: } & \left\{W_t^{\text{s,obs}}; t = 1, \ldots, N\right\} \end{aligned} \tag{44}$$

The local climate observations are not used as input in the present examples, only the Global radiation is used as model output in the solar forecasting example presented later. They could be setup and combined with the NWPs as described around Equation (11).

5.1.3 Numerical weather predictions

The numerical weather predictions (NWPs) used for the forecasting are provided by the Danish Meteorological Institute. The NWP model used is DMI-HIRLAM-S05, which has a 5 kilometer grid and 40 vertical layers (DMI, 2011). The NWPs consist of time series of hourly values for climate variables, which are updated four times per day and have a 4 hour calculation delay (e.g. the forecast starting at 00:00 is available at 04:00). Since a new two day heat load forecast is calculated

every hour, then - in order to use the latest available information - every hour the latest available NWP value for the k'th horizon at time t is picked as

$$\begin{aligned} \text{Ambient temperature (°C):} \quad & T^{\text{a,nwp}}_{t+k|t} \\ \text{Global radiation (W/m}^2\text{):} \quad & G^{\text{nwp}}_{t+k|t} \\ \text{Wind speed (m/s):} \quad & W^{\text{s,nwp}}_{t+k|t} \\ \text{Wind direction (°azimuth):} \quad & W^{\text{d,nwp}}_{t+k|t} \end{aligned} \tag{45}$$

These are setup in input matrices as described in Section 3. Similarly as for the ambient temperature in Equation (11), only using purely NWPs (no local observations). Thus the two input matrices T^{a}_t and G_t are formed and used in the following examples.

5.2 Load forecasting

Lets go through it with a simple example, first the transformation stage

Full notation

$$\text{Intercept:} \quad \mu_{t+k|t} = 1 \tag{46}$$

$$\text{Periodic:} \quad \boldsymbol{x}^{\text{per}}_{t+k|t} = f_{\text{fs}}(t; n_{\text{har}}) \tag{47}$$

$$\text{Temperature:} \quad x^{T_{\text{a}}}_{t+k|t} = H(B; a_{T_{\text{a}}}) T^{\text{a}}_{t+k|t} \tag{48}$$

$$\text{Solar:} \quad x^{G}_{t+k|t} = H(B; a_G) G_{t+k|t} \tag{49}$$

$$\tag{50}$$

and then the regression stage

$$Q_{t+k|t} = \beta_{0,k,t}\mu_{t+k|t} + \boldsymbol{\beta}_{1,k,t}\boldsymbol{x}^{\text{per}}_{t+k|t} + \beta_{2,k,t}x^{T_{\text{a}}}_{t+k|t} + \beta_{3,k,t}x^{G}_{t+k|t} + \varepsilon_{t+k|t} \tag{51}$$

Note the additional t on the coefficients indicating that they are fitted with Recursive Least Squares (RLS).

Simplified form

$$Q_{t+k|t} = \mu + f^{\text{fs}}_{k,t}(t; n_{\text{har}}) + H_{k,t}(B; a_{T_{\text{a}}}) T^{\text{a}}_{t+k|t} + H_{k,t}(B; a_G) G_{t+k|t} + \varepsilon_{t+k|t} \tag{52}$$

or more simplified

$$Q = \mu_k + f_{\text{fs}}(t; n_{\text{har}}) + H_k(B; a_{T_{\text{a}}}) T_{\text{a}} + H_k(B; a_G) G + \varepsilon \tag{53}$$

or finally

$$Q = \mu + f_{\text{fs}}(t) + H(B) T_{\text{a}} + H(B) G + \varepsilon \tag{54}$$

5.3 Solar forecasting

The forecast method can be used for solar forecasting, and in this section it is used to forecast first the global radiation and then for solar power . Forecast horizons from 1 to 36 hours are calculated. The algorithm uses the basis splines to form a conditional parametric model, in which the functional relationship between the NWPs of global radiation are conditional on the time of day.

The model is the same for all the four setups, but of course it is fitted using the output series, just denoted by Y in the following. Model notation of the transformation stage

$$\text{Global rad. input:} \quad \boldsymbol{x}^{\mathrm{G}}_{t+k|t} = f_{\mathrm{bspline}}(t^{\mathrm{day}}_{t+k}; n_{\mathrm{deg}} = 5) G^{\mathrm{nwp}}_{t+k|t} \tag{55}$$

and the regression stage

$$Y_{t+k|t} = \boldsymbol{\beta}_{1,k} \boldsymbol{x}^{\mathrm{per}}_{t+k|t} + \varepsilon_{t+k|t} \tag{56}$$

The model inputs are:

- t^{day}_{t+k} the time of day, i.e. simply the hour in this case
- $G^{\mathrm{nwp}}_{t+k|t}$ the global radiation NWP

The regression is fitted using the recursive least squares scheme, which thus has the forgetting factor parameter λ. It is optimized in an off-line setting as described in Bacher et al. (2013b).

Figure 2 shows different examples of the global radiation forecast. In the upper plot it is seen how the $k = 36$ hours ahead forecasts match the observations for a 2 weeks period. Naturally, it is not perfect, but it is clearly seen how the general pattern is very well modeled. The largest errors occur when the NWP have large errors, hence much of the accuracy of the forecasts depends on the quality of the NWPs.

In the middle plot it is seen how the forecasts, which are updated every hour using the latest available NWPs, do change for particular time points. Thus, as new NWPs become available the forecasts will use this new information. In general the NWPs of shorter horizon are better, however it must be noted that some NWPs used for $k = 1$ hour forecasts are actually 10 hours old, due to NWP calculation time (4 hours in this case) and the fact that the NWPs are only updated every 6 hours. If the latest available observation is also used as an input (an auto-regressive model part), then on shorter horizons the forecasts will become more accurate, as demonstrated in for example Bacher et al. (2013b).

In the lower plot of Figure 2 the $k = 8$ hours forecast is shown for the five day period. It can be seen how there is pattern during the late morning hours, which tend to be lower than in the afternoon, and how the model actually adapts to this pattern.

In Figure 3 the $k = 1$ hour and $k = 36$ hours forecasts are plotted for the solar power for PV panels pointing in different directions. It is clearly seen how the

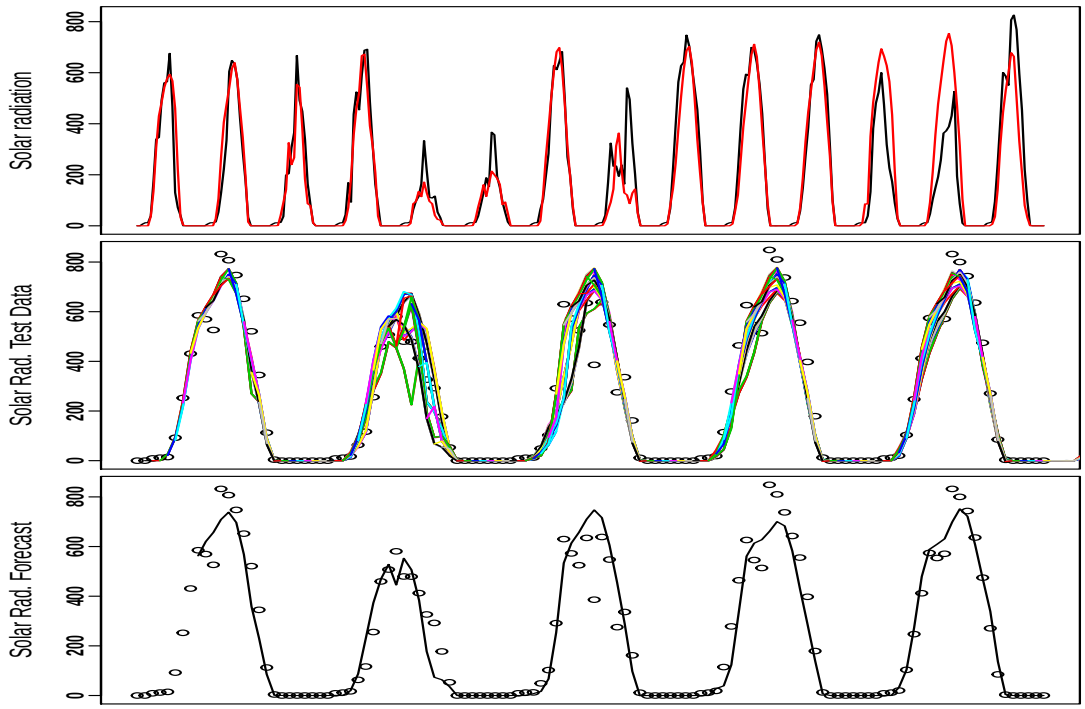


Figure 2: Upper plot: The observed solar radiations in black and the $k = 36$ hours solar radiation forecast in red for the two first weeks of April 2011. Middle plot: All forecasts calculated during 16. to 20. April. Lower plot: The $k = 8$ hours ahead forecast for the same period as in the middle plot.

model adapts to the different patterns of the signals, hence taking automatically into account, the different directions of the panels. Comparing the $k = 36$ hours forecasts to the same forecasts of global radiation (upper plot of Figure 2), it is seen that the larger errors occur exactly at the same time points, simply because the same NWPs are used and since the errors are propagating through the model.

6 Conclusion

In this report, it is shown that both heat load and solar radiation can be modelled accurately and reliable forecast can be generated. The basic methodology about how a forecasting problem can be modelled mathematically is explained and all the stages of such a mathematical formulation are discussed in details. It is also shown that the proposed algorithm is very flexible to model and forecast different variables due to its modular structure. Finally, the use of the proposed algorithm to forecast two different variables related to Flexcoop project i.e. head load forecasting and solar forecasting respectively has been demonstrated using the real world data.

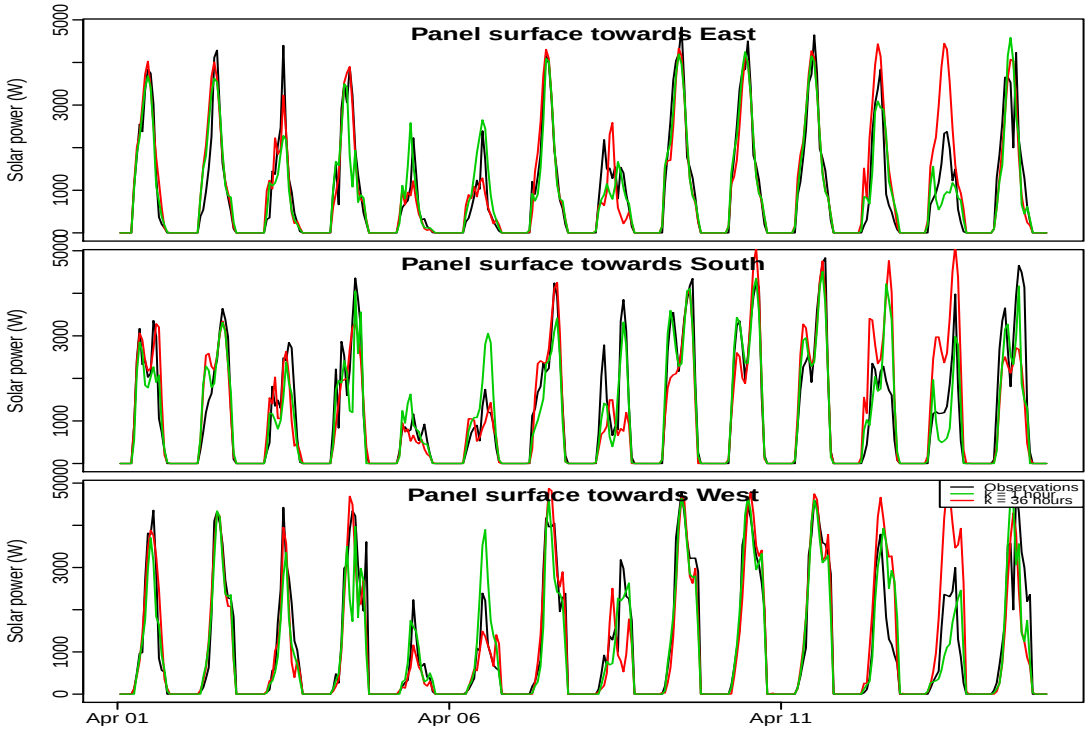


Figure 3: Solar power forecasts for panels pointing in three different directions. The period is the first two weeks of April 2011.

7 Bibliography

References

M. P. Almeida, O. Perpion, and L. Narvarte. PV power forecast using a nonparametric PV model. *Solar Energy*, 115:pp. 354–368, 2015. ISSN 0038092x, 14711257.

P. Bacher, H. Madsen, and H. A. Nielsen. Online short-term solar power forecasting. *Solar Energy*, 83(10):1772–1783, 2009. ISSN 0038092x.

P. Bacher, H. Madsen, and H. Aalborg Nielsen. Load forecasting for supermarket refrigeration. Technical report, DTU Compute, 2013a.

P. Bacher, H. Madsen, H. A. Nielsen, and B. Perers. Short-term heat load forecasting for single family houses. *Energy and Buildings*, 65(0):101–112, 2013b. ISSN 0378-7788. doi: http://dx.doi.org/10.1016/j.enbuild.2013.04.022. URL `http://www.sciencedirect.com/science/article/pii/S0378778813002752`.

G. Box, G. Jenkins, and G. Reinsel. *Time series analysis*. Holden-day San Francisco, 1976.

J. Cao and X. Lin. Study of hourly and daily solar irradiation forecast using diagonal recurrent wavelet neural networks. *Energy Conversion and Management*, 49 (6):1396–1406, 2008. ISSN 01968904.

B. Chowdhury and S. Rahman. Forecasting sub-hourly solar irradiance for prediction of photovoltaic output. In *IEEE Photovoltaic Specialists Conference, 19th, New Orleans, LA, May 4-8, 1987, Proceedings (A88-34226 13-44). New York, Institute of Electrical and Electronics Engineers, Inc., 1987, p. 171-176.*, pages 171–176, 1987.

DMI. Danish Meteorological Institute, DMI-HIRLAM-S05, 2011. URL `http://www.dmi.dk/eng/index/research_and_development/dmi-hirlam-2009.htm`.

E. Dotzauer. Simple model for prediction of loads in district-heating systems. *Applied Energy*, 73(3–4):277–284, 2002. ISSN 0306-2619. doi: 10.1016/S0306-2619(02)00078-8.

T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, and R. Tibshirani. *The elements of statistical learning*, volume 2. Springer, 2009.

D. Heinemann, E. Lorenz, and M. Girodo. Forecasting of solar radiation. In E. Dunlop, L. Wald, and M. Suri, editors, *Solar Resource Management for Electricity Generation from Local Level to Global Scale*, pages 83–94, New York, 2006. Nova Science Publishers.

F. O. Hocaoglu, O. N. Gerek, and M. Kurban. Hourly solar radiation forecasting using optimal coefficient 2-d linear filters and feed-forward neural networks. *Solar Energy*, 82(8):714–726, 2008. ISSN 0038092x.

R. Koenker. *Quantile Regression*. Cambridge University Press, 2005.

E. Lorenz, D. Heinemann, H. Wickramarathne, H. Beyer, and S. Bofinger. Forecast of ensemble power production by grid-connected pv systems. In *Proc. 20th European PV Conference, September 3-7, 2007, Milano*, 2007.

H. Madsen. *Time Series Analysis*. Chapman & Hall/CRC, 2007.

H. Madsen. *Time series analysis*. Chapman and Hall/CRC, 2008. ISBN 142005967x, 9781420059670.

R. Marquez and C. F. Coimbra. Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database. *Solar Energy*, 85(5):pp. 746–756, 2011. ISSN 0038-092X. doi: http://dx.doi.org/10.1016/j.solener.2011.01.007. URL `http://www.sciencedirect.com/science/article/pii/S0038092X11000193`.

T. Mestekemper. *Energy demand forecasting and dynamic water temperature management*. PhD thesis, Bielefeld University, 2011.

H. Nielsen, H. Madsen, and D. E. F. P. og Fordeling af El og Varme. Predicting the heat consumption in district heating systems using meteorological forecasts. Technical report, DTU IMM, 2000.

H. Nielsen, T. Nielsen, and H. Madsen. An overview of wind power forecasts types and their use in large-scale integration of wind power. In *Proceedings of the 10th International Workshop on Large-Scale Integration of Wind Power into Power Systems*, 2011.

H. A. Nielsen and H. Madsen. Modelling the heat consumption in district heating systems using a grey-box approach. *Energy & Buildings*, 38(1):63–71, 2006. ISSN 03787788. doi: 10.1016/j.enbuild.2005.05.002.

T. Nielsen, H. Madsen, H. A. Nielsen, P. Pinson, G. Kariniotakis, N. Siebert, I. Marti, M. Lange, U. Focken, L. V. Bremen, et al. Short-term wind power forecasting using advanced statistical methods. In *Proceedings of The European Wind Energy Conference, EWEC 2006*, 2006.

H. T. Pedro and C. F. Coimbra. Assessment of forecasting techniques for solar power production with no exogenous inputs. *Solar Energy*, 86(7): pp. 2017–2028, 2012. ISSN 0038-092X. doi: http://dx.doi.org/10.1016/j.solener.2012.04.004. URL `http://www.sciencedirect.com/science/article/pii/S0038092X12001429`.

A. Ragnacci, M. Pastorelli, P. Valigi, and E. Ricci. Exploiting dimensionality reduction techniques for photovoltaic power forecasting. In *Energy Conference and Exhibition (ENERGYCON), 2012 IEEE International*, pages 867–872, Sept 2012. doi: 10.1109/EnergyCon.2012.6348273.

A. Sfetsos and A. Coonick. Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques. *Solar Energy*, 68(2):169–178, 2000. ISSN 0038092x.

N. Sharma, P. Sharma, D. Irwin, and P. Shenoy. Predicting solar generation from weather forecasts using machine learning. In *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pages 528–533, Oct 2011. doi: 10.1109/SmartGridComm.2011.6102379.

P. Vogler-Finck, P. Bacher, and H. Madsen. Online short-term forecast of greenhouse heat load using a weather forecast service. *Applied Energy*, 205:1298 – 1310, 2017. ISSN 0306-2619. doi: https://doi.org/10.1016/j.apenergy.2017.08.013. URL http://www.sciencedirect.com/science/article/pii/S0306261917310292.

M. Zamo, O. Mestre, P. Arbogast, and O. Pannekoucke. A benchmark of statistical regression methods for short-term forecasting of photovoltaic electricity production. part II: Probabilistic forecast of daily production. *Solar Energy*, 105:pp. 804–816, 2014. ISSN 0038-092X. doi: http://dx.doi.org/10.1016/j.solener.2014.03.026. URL http://www.sciencedirect.com/science/article/pii/S0038092X14001601.

Q. Zhou, S. Wang, X. Xu, and F. Xiao. A grey-box model of next-day building thermal load prediction for energy-efficient control. *International Journal of Energy Research*, 32(15):1418–1431, 2008. ISSN 1099-114X. doi: 10.1002/er.1458.