Online Short-term Solar Power Forecasting

Peder Bacher, Henrik Madsen, Henrik Aalborg Nielsen

***Abstract*—This paper describes two methods for online forecasting of power production from PV systems. The methods are suited for online forecasting in many applications and in this paper they are used to predict hourly values of solar power for horizons up to 32 hours. The data used is hourly observations of solar power from a single PV system located on a rooftop in a small village in Denmark. One approach is a two-stage method in which a statistical normalization of the solar power is obtained using a clear sky model. The clear sky model is found using statistical smoothing techniques, which ensure that local phenomena are directly modelled from data, as opposed to applying a deterministically derived clear sky model. In the second stage forecasts of the normalized solar power are calculated using adaptive linear time series models. A second approach is to apply conditional parametric models with both autoregressive input and NWPs exogenous input. The results indicate that for forecasts up to two hours ahead the most important input is the available observations of solar power, while for longer horizons NWPs are the most important input. A root mean square error improvement over a persistence model around 40 % is achieved for 1 and 2 hour horizons and around 35 % for longer horizons.**

***Index Terms*—Solar power, prediction, forecasting, time series, photovoltaic, numerical weather predictions, clear sky model**

I. Introduction

The increasing installed solar power capacity rises the challenges of grid integration. The need for efficient forecasting methods is evident and the research activities within the topic is increasing, see for example [1], [2], [3], and [4]. In this paper methods for online forecasting are presented. The methods are suited for forecasting of solar power for different systems and here they are applied to forecast the power production of a single 4 kW-peak PV-system installed on a rooftop of a single family house. Due to the fluctuating nature of solar power such forecasts are essential for optimal grid integration and will be essential for solar power smart grid technology. The applications include energy trading for large solar power producers, and diurnal peak-shaving and cost optimization for smaller systems with storage capacity in battery packs (e.g. provided in an electrical car). Two approaches are considered. One is based on a two-stages approach: first the systematic dependency of the position of the sun relative to the PV panel are removed with a clear sky model, and secondly the resulting process is forecasted with time-adaptive linear time series methods. The clear sky model is calculated with non-linear statistical techniques, which will also model the local conditions, such as e.g. shadows from elements in the surrounding environment and snow cover. In the second approach numerical weather predictions (NWPs) are used as input to conditional parametric non-linear models [5] to forecast the solar power. Finally, the two approaches are combined by normalizing the forecast with the clear sky model, and finally using this as input to the linear forecasting model, such that an ARX model is formed.

The paper is organized as follows. First the data and how it is preprocessed is described. The next section contains an outline of the clear sky model, followed by a section where all the forecasting models are described. Then an evaluation is given and the results are presented, followed by a discussion of the results and ideas for further work. Finally, the paper ends with a conclusion.

P. Bacher and H. Madsen are with the Department of Informatics, Technical University of Denmark, DK-2800 Lyngby, Denmark (e-mail: pb@imm.dtu.dk)
H.A. Nielsen is with ENFOR A/S, Lyngsø Alle 3, DK-2970 Hørsholm, Denmark (URL: www.enfor.eu)

II. Data

The data used in this study consist of hourly mean values of solar power from a 4 kW-peak PV-system and NWPs of global irradiance. The NWPs are provided by the Danish Meteorological Institute using the HIRLAM mesoscale NWP model. The data covers the entire year 2006.

The time series of hourly observed solar power is

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

where $N = 8760$. The NWPs have a calculation time of 4 hours, which is taken into consideration, such that e.g. the forecast from 2009-01-01 00:00 are only available from 2009-01-01 04:00. The NWPs are provided in a time resolution of 3 hours. They are pre-processed into time series of hourly values, such that the most recent available forecast k hours ahead is selected each hour. The time series for a given k of the direct radiation is

$$\left\{G^{\text{nwp}}_{t+k|t};\ t = 1, \ldots, N\right\} \tag{2}$$

A. Pre-processing

The solar power data is plotted for each hour of the day in Figure 1. The solar radiation is zero at night, hence the observed solar power is also zero. For the current data set only periods, for a given hour of the day longer than 40 days in which the solar power is different from zero, are included for evaluation of the model performance. This is illustrated in Figure 1, where the non-included periods are grayed out.

III. Clear sky model

Forecasting effectively using linear time series methods calls for stationarity of the underlying process [6]. The process that generates the solar power is not stationary, which is seen by plotting quantiles of the distribution of solar power conditioned on the time of day, see Figure 2. Clearly the distribution of solar power is not independent of the time of day.

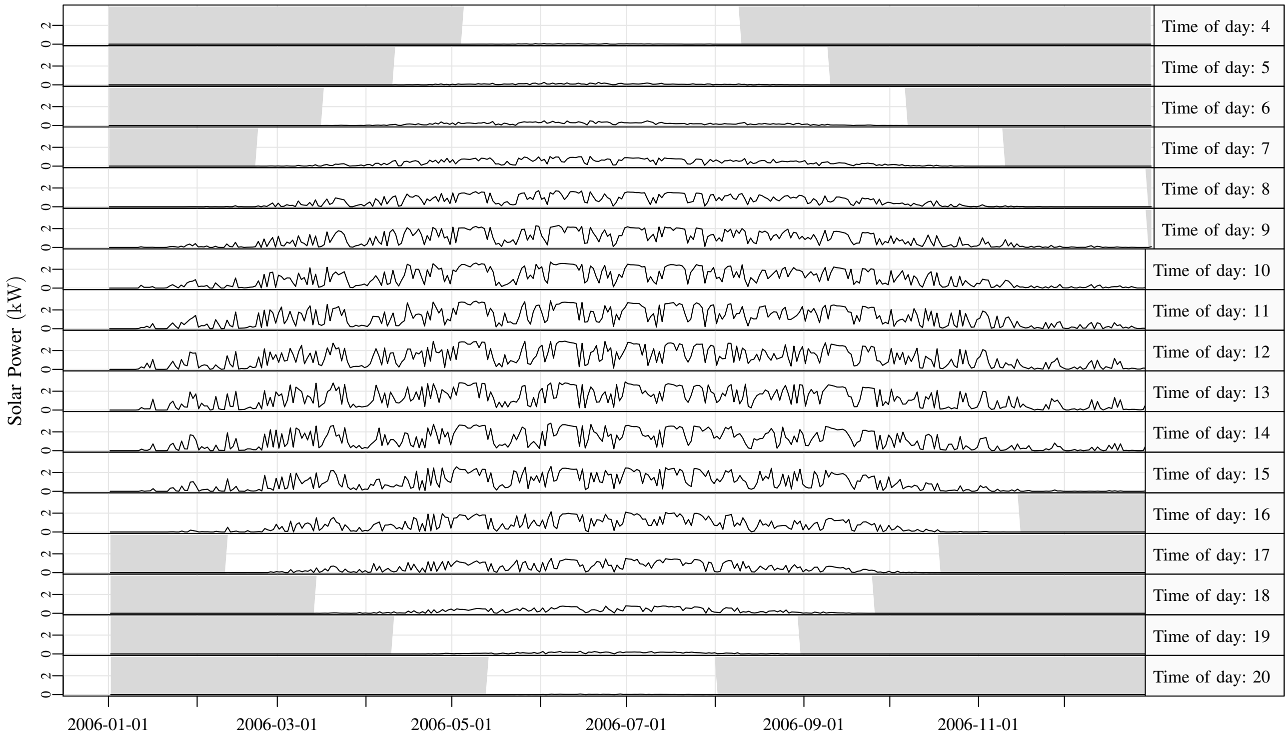

Fig. 1. The solar power data. The greyed area are not included in the evaluation of the model performance.

Most of this dependency can be removed by a normalization using a clear sky model

$$\tau_t = \frac{P_t}{P_t^{\mathrm{cs}}} \tag{3}$$

where P_t is the observed solar power, P_t^{cs} is the estimated clear sky solar power, and τ_t is the normalized solar power.

A. Statistically estimated clear sky solar power

The clear sky solar power is estimated using a statistical non-linear and adaptive model. Quantile regression [7] locally weighted in the day of year and time of day dimension is applied. This is carried out fully causal, i.e. only past values are used. The clear sky model is

$$\hat{P}_t^{\mathrm{cs}} = q_{0.99}(P_1, P_2, \ldots, P_t, h_{\mathrm{y}}, h_{\mathrm{tod}}) \tag{4}$$

where $q_{0.99}$ is the 99% quantile based on the solar power values up to time t. The bandwidths h_{day} and h_{tod}, are in the day of year and time of day dimension, respectively. The bandwidths control how "locally" the model is fitted, i.e. a lower bandwidth puts more emphasis on data which is close in the two dimensions. The local weighting function is an Epanechnikov kernel. The applied bandwidths are

$$h_{\mathrm{day}} = 100 \text{ days}, \quad h_{\mathrm{tod}} = 3 \text{ hours} \tag{5}$$

which were found by visual inspection of the fitted clear sky curve. Finally, it is noted that second-order polynomials were applied in the time of day dimension to include curvature into the model. The estimated clear sky solar power is shown in Figure 3.

One advantage of the normalization is that it will automatically adapt to changes in the system, such as degraded performance or changes in the surroundings e.g. snow cover and shadowing effects. It can as well be used for monitoring of the solar system, since degraded performance from the same time of year will result in a lower clear sky solar power curve. Plots of the quantiles of the distribution of normalized solar power conditional on the time of day are shown in Figure 2, from which it is seen that the normalized solar power process is considerably less dependent on the time of day and therefore a much more stationary process. It is noted that further work could include physical considerations into the clear sky model.

[t]

IV. Forecasting models

In this section a description of the applied forecasting models is given. The models can be divided into models using linear time series models to forecast the normalized solar power: autoregressive (AR) and autoregressive with exogenous inputs (ARX) models - and models which forecast in a single stage: conditional parametric (CP) models. Each model is fitted separately for each horizon, such that the same model structure is used, but the parameters are estimated separately for each horizon.

A. Reference model

To compare the performance of prediction models, and especially when making comparisons between different studies, a common reference model is essential. The reference model for solar power used in this study is the best performing naive

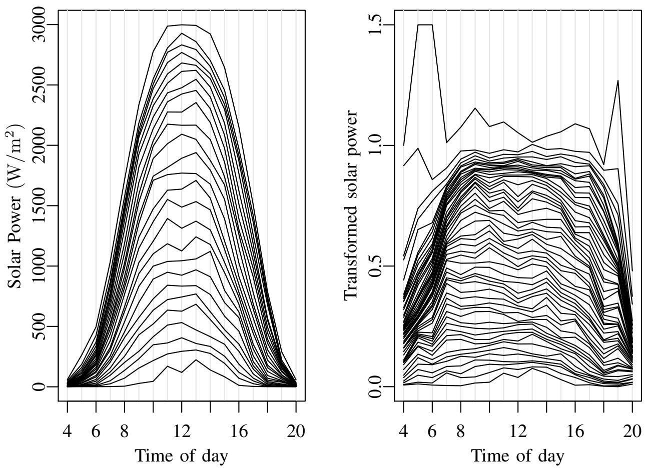

Fig. 2. The $0, 4\%, ..., 100\%$ quantiles of the distribution of the solar power and the normalized solar power conditioned on the time of day. Values above 1.5 has been clipped, which was the case for 6 values.

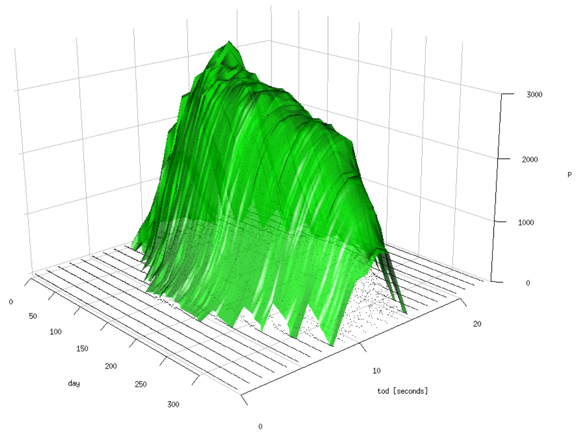

Fig. 3. The estimated clear sky solar power.

predictor for a given horizon. Two naive predictors of solar power are found to be relevant. Persistence

$$p_{t+k|t} = p_t + e_{t+k}, \tag{6}$$

and diurnal persistence

$$p_{t+k|t} = p_{t-s(k)} + e_{t+k} \tag{7}$$
$$s(k) = 24 + k \bmod 24 \tag{8}$$

where $s(k)$ ensures that the latest diurnal observation is used, i.e. the value which, depending on the horizon, is either 24 or 48 hours before the time point that is to be forecasted.

B. Autoregressive models

Autoregressive (AR) models are applied to forecast the normalized solar power. These models can include either the latest available observation or the latest available diurnal observation, or both, as input. The models are fitted with k-step recursive least squares with forgetting factor [8]. The model formulated as a k-step AR model

$$\tau_{t+k|t} = m + a_1\tau_t + a_{24}\tau_{t-s(k)} + e_{t+k} \tag{9}$$
$$s(k) = 24 + k \bmod 24 \tag{10}$$

where the function $s(k)$ ensures that the latest observation of the diurnal component is included. The model without the diurnal component, denoted AR, performs best on short horizons

$$\tau_{t+k|t} = m + a_1\tau_t + e_{t+k} \tag{11}$$

and is included in the evaluation. The AR model with only the diurnal performs better on longer horizons, but is inferior to the models including the NWPs.

C. Conditional parametric models

Conditional parametric (CP) models where the coefficients are conditional on the time of day and time of year are applied with both past solar power observations and NWPs as inputs. The CP model with the latest solar power observation as input is

$$P_{t+k} = m + a\big(t_{\text{day}}, t_{\text{tod}}, P_t\big)P_t + e_{t+k} \tag{12}$$

where the coefficient function is a non-linear function of the solar power. It is denoted as CP_P. The CP model with NWPs of global radiation as input is

$$P_{t+k} = m + b\Big(t_{\text{day}}, t_{\text{tod}}, G^{\text{nwp}}_{t+k|t}\Big)G^{\text{nwp}}_{t+k|t} + e_{t+k} \tag{13}$$

where $G^{\text{nwp}}_{t+k|t}$ is the k-hour ahead NWP of global radiation. This model is denoted CP_{NWP}. Finally, the model with both inputs

$$P_{t+k} = m + a\big(t_{\text{day}}, t_{\text{tod}}, P_t\big)P_t \tag{14}$$
$$+ b\Big(t_{\text{day}}, t_{\text{tod}}, G^{\text{nwp}}_{t+k|t}\Big)G^{\text{nwp}}_{t+k|t} + e_{t+k} \tag{15}$$

is denoted $CP_{NWP,P}$.

In the following the coefficients dependency of the time of day for CP_{NWP} is elaborated on. It is noted that the bandwidths are optimized for each horizon. Plots of the fitted forecasting function $b\Big(t_{\text{day}}, t_{\text{tod}}, G^{\text{nwp}}_{t+k|t}\Big)$ for $k = 24$ hours are shown in Figure 4. It is seen how the slope of the function is lower in the morning, than in the middle of the day. This is naturally caused by the higher angle of incidence in the morning, which cause less horizontal radiation to be absorbed due to reflection. Likewise for the afternoon. Finally, non-linearity in the fitted function is seen.

D. Autoregressive model with exogenous input

The AR model is be expanded to include the forecast of the CP models, thus combining information in past observed solar power and NWPs. The solar power forecasts from the

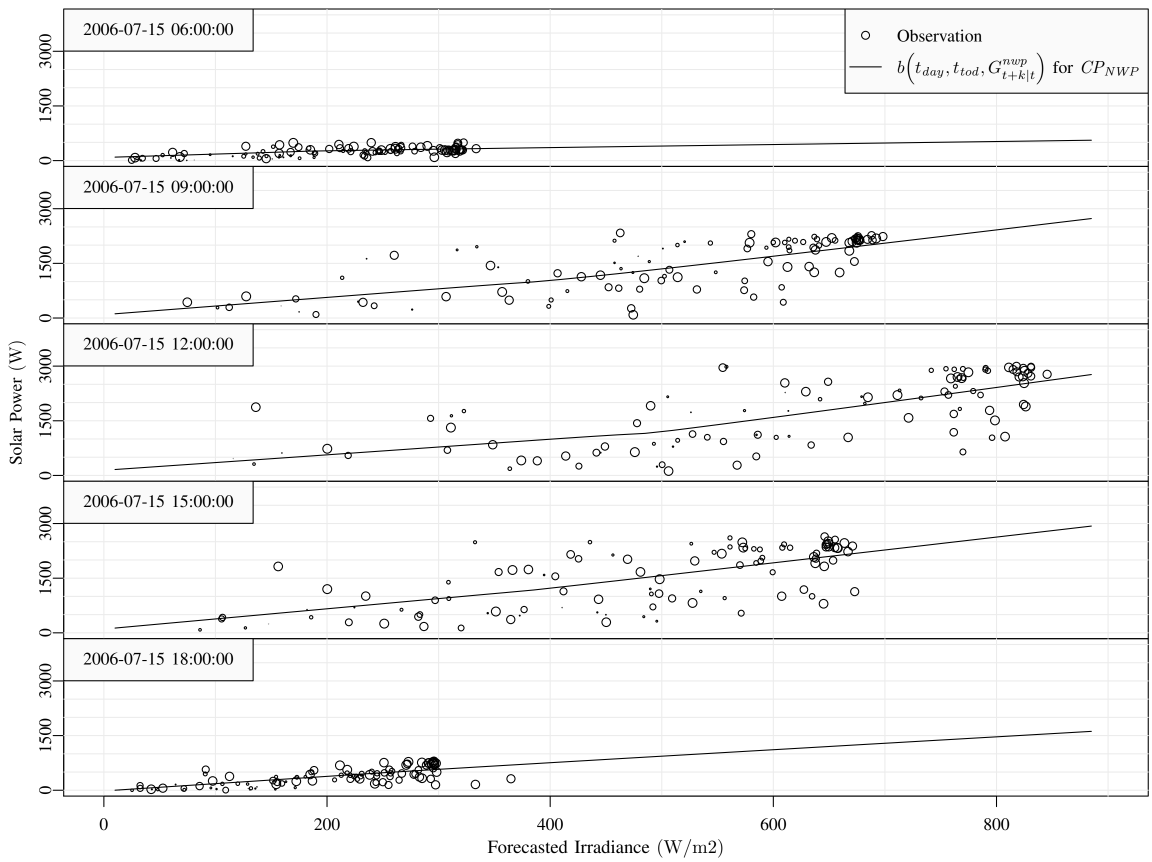

Fig. 4. Examples of the function fitted for $k = 24$ hours forecasting with the NWPs of global radiation at different times of the day on the 15'th of July 2010 with the CP_{NWP} model. For each observation the size of circle indicates the weighting of the observation in the CP models. Thus observations with a larger circle have more influence on the fitted function.

CP is normalized with the clear sky model by

$$\hat{\tau}^{\text{nwp}}_{t+k|t} = \frac{\hat{P}^{\text{nwp}}_{t+k|t}}{P^{\text{cs}}_{t-s(k)}} \tag{16}$$

$$s(k) = f_{\text{spd}} + k \bmod f_{\text{spd}} \tag{17}$$

where $f_{\text{spd}} = 24$ is the sample frequency in number of samples per day. The *ARX1* model is

$$\tau_{t+k} = m + a_1\tau_t + b_1\tau^{\text{nwp}}_{t+k|t} + e_{t+k} \tag{18}$$

V. Evaluation

The methods used for evaluating the prediction models are inspired by [9]. The clear sky model, RLS, and CP fitting do not use any degrees of freedom and the data set is therefore not divided into a training set and a test set. It is only for the optimization of the kernel bandwidths and the forgetting factor that the entire data set is used. The period before 2006-03-01 is considered as a burn-in period and not used for calculating the error measures.

A. Error measures

The Root Mean Square Error for the k'th horizon is

$$RMSE_k = \left(\frac{1}{N}\sum_{t=1}^{N} e^2_{t+k}\right)^{\frac{1}{2}} \tag{19}$$

where e_{t+k} is the k-hourly prediction error. The $RMSE_k$ is used as the main evaluation criterion (EC) for the performance of the models. The Normalized Root Mean Square Error is found by

$$NRMSE_k = \frac{RMSE_k}{p_{\max}} \tag{20}$$

where $p_{\max}$ is the maximum observed solar power output. The mean value of the $RMSE_k$ for a range of horizons

$$\overline{RMSE}_{k_{\text{start}},k_{\text{end}}} = \frac{1}{k_{\text{end}} - k_{\text{start}} + 1}\sum_{k=k_{\text{start}}}^{k_{\text{end}}} RMSE_k \tag{21}$$

is used as a summary error measure. When comparing the performance of two models the improvement

$$I_{\text{EC}} = 100 \cdot \frac{EC_{\text{ref}} - EC}{EC_{\text{ref}}}(\%) \tag{22}$$

is used, where EC is the considered evaluation criterion. When calculating the error measures it is important to consider how to handle missing values for the solar power forecasts. The problem is handled by replacing missing forecast values with forecast values from the reference model *Ref*.

B. Completeness

In order to evaluate a model for its performance regarding missing forecast values a measure is defined. It is denoted

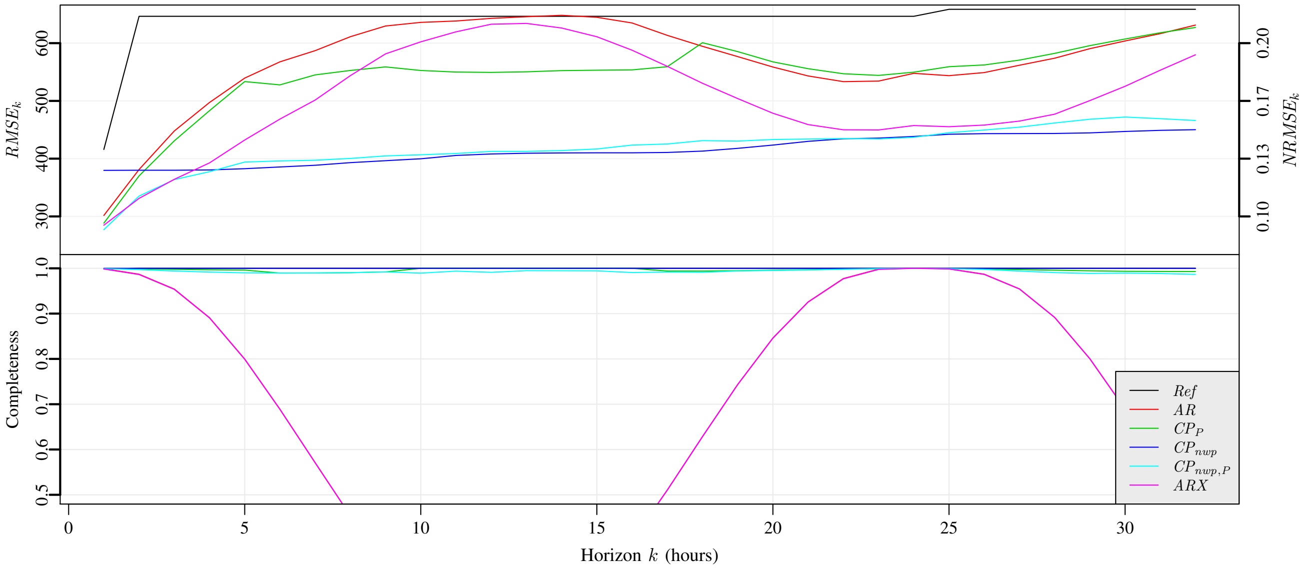

Fig. 5. The upper plot is $\boldsymbol{RMSE_k}$ for the forecasting models. On the right side the $\boldsymbol{NRMSE_k}$ is indicated. The lower plot is completeness $\boldsymbol{C_k}$.

completeness. The completeness of a forecast for horizon k, is the ratio of the total sum of solar power and the summed solar power for time points where the forecasts are not missing

$$C_k = \frac{\sum_{t=1}^{N} P_t \; I\big(\hat{P}_{t|t-k}\big)}{\sum_{t=1}^{N} P_t} \tag{23}$$

where $I()$ is the indicator function which is 0 if $\hat{P}_{t|t-k}$ is missing, and 1 if not. Only the included values are used, i.e. not values during nighttime.

VI. Results

In this section the results are presented and evaluated. The $\overline{RMSE}_{k_{\text{start}},k_{\text{end}}}$ improvements for relevant ranges of horizons are listed in Table I. For selected models the $RMSE_k$ is shown in the upper plot of Figure 5 and the completeness in the lower.

Considering the improvements it is seen that most of the models perform very well on either the short horizons or the longer horizons. Starting with short horizons (1 to 2 hours) the four models using the latest observed solar power have better performance than CP_{NWP}, which only uses the NWPs. Using the combination of observed solar power and NWPs improves the performance, except on longer horizons where using only NWPs are slightly better. Considering the performance of AR, CP_P, and ARX it is seen that the $RMSE_k$ increase really fast as the horizon increases and reach the reference model around a horizon of 10 hours. This is simply because the models are using night values (which are missing) to forecast day values. This is also seen in the completeness of the AR and ARX model.

TABLE I
Improvements in percent for selected ranges of horizons.

Model	$I_{\overline{RMSE}_{1,2}}$	$I_{\overline{RMSE}_{3,17}}$	$I_{\overline{RMSE}_{18,32}}$
AR	34.3	7.4	12.6
CP_P	36.7	17	11.5
CP_{NWP}	25	38.4	33.1
$CP_{NWP,P}$	40.8	37.6	31.4
ARX	40.1	15.9	25

VII. Discussion and applications

This section contains a short discussion of the results and ideas for further work, and ends with an outline of applications.

Considering the improvement achieved over the reference model the forecasting models are found to perform very well. Clearly the quality of the NWPs of solar radiation is the most influential source of error, hence improved NWPs will improve the performance. Especially using NWPs of direct and diffuse radiation should be tried. Regarding further improvement of the forecasting models, it is suggested that the following should be considered:

- Application of regime models and hidden Markov models to handle different aspects of forecasting for e.g. low and high radiation values, and it might be useful to use different forecasting models for different types of cloud conditions. This is ideal to apply in the setting of the CP models.
- For the CP models using higher order polynomials in the day of year and time of day dimensions should improve the models. It was tried but didn't improve the performance, but as the NWPs are getting better this will most likely be important.
- A thorough evaluation of the forecast errors to find ideas for how the models can be improved.

The applications for solar power forecasting include the integration of PV systems into the electricity grid, especially for smart grids. The solar power forecasts can be used as input to model predictive control to optimize the operation of the PV system. This will enable diurnal peak-shaving and cost optimization for smaller systems with storage capacity in battery packs (e.g. provided in an electrical car). For large solar power producers forecasting is essential for optimized energy trading.

The method is furthermore well suited for monitoring the performance of PV systems. Measures of the performance can be derived from the CP models, with which systems can be compared on an absolute scale. Sudden high deviation from the CP forecasting model will allow for very fast detection of failures in the system. For an individual system the change in performance over time can also be assessed by monitoring the clear sky curve for unusual behavior, and compare the change from year to year.

VIII. Conclusion

Two approaches for solar power forecasting are presented and applied to forecast hourly values for horizons up to 32 hours. Both a method based on a two-stage approach, where first the solar power is normalized with a statistical clear-sky model, and a method in which the solar power is forecasted in a single step. The normalization with a clear sky model removes most of the non-stationarity caused by the changing position of the sun relative to the PV panel. This a pre-requisite for optimal application of linear time series models. Conditional parametric models are used to include NWPs of global radiation, and a one-stage approach, solely based on conditional parametric models, is presented. A root mean square improvement over a persistence reference model on short horizons (1 to 2 hours) is in average 40%, and in average 35% on the longer horizons. The method can furthermore be applied to monitor and check the performance of PV systems.

References

[1] A. Sfetsos and A. Coonick, "Univariate and multivariate forecasting of hourly solar radiation with artificial intelligence techniques," *Solar Energy*, vol. 68, no. 2, pp. 169–178, 2000.

[2] F. O. Hocaoglu, Ö. N. Gerek, and M. Kurban, "Hourly solar radiation forecasting using optimal coefficient 2-d linear filters and feed-forward neural networks," *Solar Energy*, vol. 82, no. 8, pp. 714–726, 2008.

[3] E. Lorenz, J. Hurka, D. Heinemann, and H. Beyer, "Irradiance forecasting for the power prediction of grid-connected photovoltaic systems," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 2, no. 1, pp. 2–10, 2009.

[4] W. Ji and K. C. Chee, "Prediction of hourly solar radiation using a novel hybrid model of arma and tdnn," *Solar Energy*, vol. 85, no. 5, pp. 808–817, 2011.

[5] T. S. Nielsen, H. Madsen, and H. A. Nielsen, "Prediction of wind power using time-varying coefficient-functions," *Proceedings of the 15th IFAC World Congress on Automatic Control*, 2002.

[6] H. Madsen, *Time Series Analysis*. Chapman & Hall, 2007.

[7] R. Koenker, *Quantile Regression*. Cambridge University Press, 2005.

[8] P. Bacher, H. Madsen, and H. A. Nielsen, "Online short-term solar power forecasting," *Solar Energy*, vol. 83, no. 10, pp. 1772–1783, 2009.

[9] H. Madsen, P. Pinson, G. Kariniotakis, H. A. Nielsen, and T. S. Nielsen, "Standardizing the performance evaluation of shortterm wind power prediction models," *Wind Engineering*, vol. 29, no. 6, p. 475, 2005.