Localized Online Weather Predictions with Overnight Adaption*

Michael Zauner[1], Michaela Killian[1], and Martin Kozek[1]

Vienna University of Technology,
Institute of Mechanics and Mechatronics,
Getreidemarkt 9, 1060 Vienna, Austria
michael.zauner@tuwien.ac.at

Abstract. This paper describes an approach to online forecasting of ambient temperature and solar irradiation. The proposed method creates a localized prediction with an improvement over the available weather predictions ranging from 52% to 92% in ambient temperature forecast and 8% to 42% for solar irradiation forecast. This localized forecast can be used for improved predictions in smart homes or PV power plants for a more efficient operation. A new method for adapting the parameters of the autoregressive model with external input (ARX) for the solar irradiation over the night is proposed. This allows the model to be tuned to changing weather conditions without relying on external inputs.

Keywords: Weather prediction · Local optimization · Parametric model

1 Introduction

With the ever growing energy demand and the exhaustion of non-renewable resources the efficient usage of renewable energy sources (wind, solar, tidal and biomass) gets more important [4]. Due to the weather dependent nature of those renewable energy sources it is challenging to balance energy production and consumption in global electrical grids and in decentralized smart grids with smart consumers (e.g. smart homes). Therefore it is vitally important to have accurate forecasting models for those renewable energy sources [6]. The most important factor influencing solar power production via PV (photovoltaic) systems is solar irradiation [9], followed by meteorological parameters like ambient temperature and relative humidity [4]. Another application where solar irradiation and ambient temperature are important factors are the heating and cooling tasks of residential buildings [5]. There a model predictive controller for an HVAC (heating, ventilation and air conditioning) system cannot only provide better comfort for the residents, but also save energy if accurate predictions are available.
The well-known modern numerical weather forecasting services (WFS) use discrete cells for simulating weather predictions. The initial conditions for those

* This work was supported by the project "intelliEE-Home" (FFG. No. 853663) in cooperation with evon GmbH.

simulations are gathered by land based weather stations and satellite images. This results in poor localized predictions as the forecast is valid for the whole cell. The idea of this paper is to create a localized weather prediction based on the forecasts of the WFS and the past and current local sensor data.
The WFS provide information up to several days ahead [9]. The temporal resolution of those WFS predictions is typically limited to 1 hour. This resolution is usually not sufficient for the performance of most applications [3]. A common approach for short term forecasting is based on sky imaging [2] and time-series models [1, 4, 7, 9].
While predictions based on sky imaging provide good results in the range up to a few minutes, they also suffer from drawbacks: the devices are expensive, they require a lot of maintenance and predictions are only usable when the cloud cover is not too high or too low.
The usage of localized sensors and global predictions provided by WFS allow for more accurate localized predictions and a higher temporal resolution. While many authors use non-linear methods like ANN (Artificial Neuronal Networks) [7], GMDH (Group Method of Data Handling) [4] or SVM (Support Vector Machines) [9] for forecasting, a linear autoregressive model with exogenous input (ARX) is proposed in this paper. The advantages of using ARX models over ANN, GMDH and SVM are that less parameters have to be optimized and the optimization can be done in real time. Bacher et.al [1] proposed a similar modeling approach in their work. The main differences to the proposed work are the usage of a diurnal component and a clear sky approximation via smoothing kernels. The method proposed here includes furthermore an overnight prediction scheme to accommodate unmeasured changes in weather conditions.
With the usage of ARX-models the proposed method can learn statistical differences between local conditions (provided via the sensors) and the WFS predicted conditions.
The remainder of the paper is structured as follows: Section 2 explains the methods and algorithms used for the ambient temperature forecast. Section 3 highlights the necessary changes for the solar irradiation forecast. Section 4 briefly explains the simulation setup as well as the results of the simulations for ambient temperature forecast and solar irradiation forecast. In the end Sect. 5 concludes this paper.

2 Ambient Temperature Forecast

The WFS is assumed to provide an ambient temperature prediction for the next 80 hours. In the first 65 hours hourly prediction values are available. After that the WFS only provides predictions in 3 hour intervals. An example WFS prediction for the ambient temperature can be seen in Fig. 1.
The WFS prediction is linearly interpolated and a new time series $\vartheta_{\text{pred}}(k)$ with the uniform sampling time of $T_s = 0.25$ h is constructed. Where $k = \{1, ..., T\}$ with T being the final time step where predictions are available.
It is assumed that a local weather station is measuring the local ambient tem-

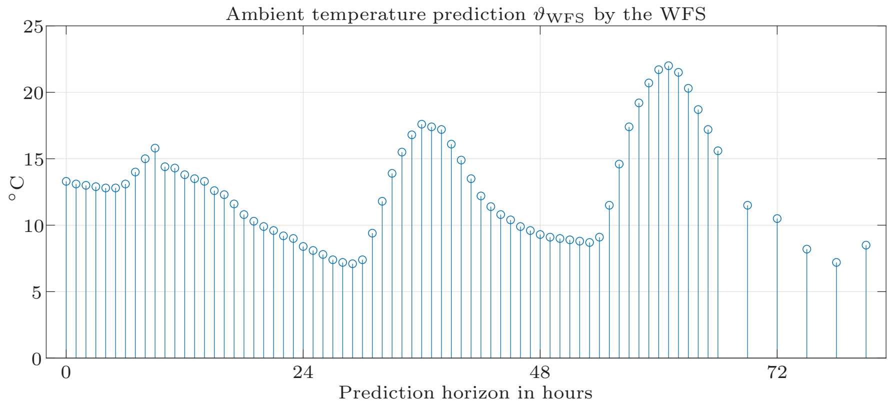

Fig. 1. Temperature prediction from the WFS. The hourly ambient temperature predictions are provided for the next 65 hours and after that in 3 h intervals for an additional 15 hours.

perature $\vartheta_{\mathrm{amb}}(k)$ every 0.25 hours. The last n measurements of the local temperature are saved in the system.
At every time step the vector

$$\boldsymbol{x}^T(k) = [\vartheta_{\mathrm{amb}}(k-n+1), \ldots, \vartheta_{\mathrm{amb}}(k), \vartheta_{\mathrm{WFS}}(k), \ldots, \vartheta_{\mathrm{WFS}}(k+m-1)] \quad (1)$$

is constructed, where $\vartheta_{\mathrm{WFS}}(k)$ is the latest WFS prediction for the current time step and $\vartheta_{\mathrm{amb}}(k)$ is the current measured ambient temperature. The variables $n \in \mathbb{N}^+$ and $m \in \mathbb{N}^+$ represent the order of the denominator and nominator in the resulting ARX model.
Using the **W**eighted **R**ecursive **L**east **S**quares algorithm (WRLS) shown in (2a)-(2c),

$$\boldsymbol{\gamma}(k) = \frac{\boldsymbol{P}(k)\boldsymbol{x}(k)}{\boldsymbol{x}^T(k)\boldsymbol{P}(k)\boldsymbol{x}(k)+\lambda}, \quad (2a)$$

$$\hat{\boldsymbol{\theta}}(k+1) = \hat{\boldsymbol{\theta}}(k) + \boldsymbol{\gamma}(k)[\vartheta_{\mathrm{amb}}(k+1) - \boldsymbol{x}^T(k)\hat{\boldsymbol{\theta}}(k)], \quad (2b)$$

$$\boldsymbol{P}(k+1) = \frac{1}{\lambda}[\boldsymbol{I} - \boldsymbol{\gamma}(k)\boldsymbol{x}^T(k)]\boldsymbol{P}(k), \quad (2c)$$

with $\boldsymbol{P}(k) \in \mathbb{R}^{(n+m)\times(n+m)}$ being the parameter-covariance matrix, $\hat{\boldsymbol{\theta}}(k) \in \mathbb{R}^{(n+m)}$ representing the estimated parameter vector, and $\boldsymbol{\gamma}(k) \in \mathbb{R}^{(n+m)}$ the correction vector. The scalar value $\lambda \leq 1$ represents the sensitivity of the algorithm to more recent values. Choosing a λ closer to 1 will increase the weight of past samples that are significant to the current parameter estimation. With $\lambda = 1$ the WRLS algorithm will behave like a regular recursive least squares algorithm. Furthermore $\boldsymbol{I} \in \mathbb{R}^{(n+m)\times(n+m)}$ is defined as the unity matrix.
The initial value for the parameter-covariance matrix $\boldsymbol{P}$ is chosen as $\boldsymbol{P}(0) = \alpha\boldsymbol{I}$ where $\alpha \gg 1$. The initial value for $\hat{\boldsymbol{\theta}}(0)$ is chosen as a random $(n+m)\times 1$ vector.

The future predictions for the ambient temperature, denoted by $\hat{\vartheta}_{\mathrm{amb}}(k|j+1)$ where $j+1$ represents any given future time step and k the current time step are given by (3). Note that $j \geq k$ and $j < T$ must hold.

$$\hat{\vartheta}_{\mathrm{amb}}(k|j+1) = \hat{\boldsymbol{x}}^T(k|j)\hat{\boldsymbol{\theta}}(k), \tag{3}$$

where

$$\hat{\boldsymbol{x}}^T(k|j) = [\tilde{\vartheta}_{\mathrm{amb}}(k|j-n+1), \ldots, \tilde{\vartheta}_{\mathrm{amb}}(k|j), \vartheta_{\mathrm{WFS}}(j), \ldots, \vartheta_{\mathrm{WFS}}(j+m-1)], \tag{4}$$

with $\tilde{\vartheta}_{\mathrm{amb}}(k|i)$ being defined as

$$\tilde{\vartheta}_{\mathrm{amb}}(k|i) = \begin{cases} \vartheta_{\mathrm{amb}}(i) & \text{if } \ i \leq k \\ \hat{\vartheta}_{\mathrm{amb}}(k|i) & \text{else.} \end{cases} \tag{5}$$

In (4) $\vartheta_{\mathrm{WFS}}(j)$ is the most recent prediction for the time step j. Equation (5) recursively calculates predictions by (3) until only current or past measurements are needed for the formulation of $\hat{\boldsymbol{x}}^T$ in (4).
The predicted future values have to be recalculated after every new measurement since the parameter vector $\hat{\boldsymbol{\theta}}(k)$ is updated in (2b).

3 Solar Irradiation

The WFS provides the hourly solar irradiation predictions for the next 43 hours. Because of the diurnal and annual periodicity of the sun the scheme presented in Sect. 2 cannot be used without modifications. The absence of measurements during the night does not allow for parameter adaption during the night. This is problematic since the weather conditions could change drastically overnight.
In a first step the global horizontal irradiance (GHI), provided by the sensors, is transformed into the clear sky index. The clear sky index τ_{cs} is defined by

$$G = G_{\mathrm{cs}} \cdot \tau_{\mathrm{cs}}, \tag{6}$$

where G is the current global horizontal irradiation (in W/m^2) and G_{cs} is the clear sky global horizontal irradiation (in W/m^2). The clear sky index τ_{cs} is an indication for the transmissivity of the atmosphere. The GHI for clear sky conditions is calculated via the toolbox provided by Sandia National Laboratories [8].
As previously the vector

$$\boldsymbol{x}^T(k) = [\tau_{\mathrm{cs}}(k-n+1), \ldots, \tau_{\mathrm{cs}}(k), \tau_{\mathrm{WFS}}(k+1), \ldots, \tau_{\mathrm{WFS}}(k+m)], \tag{7}$$

is created at every time step $k = \{1, ..., T\}$, where T is the final time step where predictions are available and $\tau_{\mathrm{WFS}}(k+1)$ the next clear sky index calculated with the WFS data. It is important to note that the current prediction $\tau_{\mathrm{WFS}}(k)$ is not used, instead the next future prediction $\tau_{\mathrm{WFS}}(k+1)$ is included. This corresponds to a negative input dead time. The variables $n \in \mathbb{N}^+$ and $m \in \mathbb{N}^+$

represent again the order of the denominator and nominator in the resulting ARX model.
During the day the WRLS algorithm shown in (2a) - (5) can be applied to compute the predictions for the clear sky index $\hat{\tau}_{\text{cs}}$.
During the night no calculations are possible due to the lack of measurements. In the morning new initial values for $\boldsymbol{P}$ and $\hat{\boldsymbol{\theta}}$ are needed since the weather conditions could have changed significantly overnight.
To calculate the new initial values the latest predictions from sunrise to sunset are queried before sunrise and concentrated into the vector $\boldsymbol{\tau}_{\text{pred}}$. These predictions are then compared against measured solar irradiation time-series of past days in a database.

Require: $\exists\ \tau_{\text{pred},i} \in \mathbb{R}^{1\times 1}, \tau_{\text{datab}} \in \mathbb{R}^{n_d \times 48}$
1: normalize τ_{pred} to 12h day $\Rightarrow \tilde{\tau}_{\text{pred}} \in \mathbb{R}^{1\times 48}$
2: **for** $j = 1$ to n_d **do**
3: $\quad e_j = \sqrt{\frac{1}{48}\sum_i^{48}(\tilde{\tau}_{\text{pred},i} - \tau_{\text{datab},j,i})^2}$
4: **end for**
5: sort e_j ascending
6: **return** $\tau_{\text{datab},j}$ of $[e_1, e_2, ..., e_K]$

Algorithm 1: Find the K most similar solar days

Algorithm 1 showcases an example on how to search a database with n_d normalized entries. For a normalized solar day the time between sunrise and sunset is defined as 12 hours. Therefore a single normalized solar day consists of 48 entries when sampled at 15 minutes intervals. Line 3 of the algorithm uses the euclidean distance to calculate the similarity, but other distances could be considered too. Algorithm 1 returns the K most similar solar days in the database.
With the usage of the clear sky index and normalizing the solar days to 12 hours, comparisons between the daily solar conditions can be drawn regardless of the time of the year.
Algorithm 2 describes the overnight prediction process. The prediction for the next day and the K most similar solar days from the database along with λ are the inputs. After normalizing the database entries and initializing $\boldsymbol{P}$ and $\hat{\boldsymbol{\theta}}$ the WRLS algorithm shown in (2a) - (5) is executed. In this WRLS algorithm $\tau_{\text{WFS}} = \tau_{\text{pred}}$ and $\tau_{\text{CS}} = \tilde{\tau}_{\text{datab},j}$ according to (7). The overnight prediction algorithm then returns the new initial values for $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{P}$ for the WRLS algorithm that is active during the next day.
The choice of K should be large enough for the parameters to settle during the overnight prediction.

Require: $\exists\ \tau_{\text{pred},i} \in \mathbb{R}^{1\times 1}, \tau_{\text{datab}} \in \mathbb{R}^{K\times 48}, \lambda \in \mathbb{R}^{1\times 1}$
1: normalize τ_{datab} to size(τ_{pred}) $\Rightarrow \tilde{\tau}_{\text{datab}} \in \mathbb{R}^{K\times I}$
2: $\hat{\boldsymbol{\theta}}(0) = \alpha * \boldsymbol{I}$
3: $\boldsymbol{P}(0) = \text{rand}$
4: **for** $j = 1$ to K **do**
5: $[\hat{\boldsymbol{\theta}}(j), \boldsymbol{P}(j)] = \text{WRLS}(\tau_{\text{pred}}, \tilde{\tau}_{\text{datab},j}, \hat{\boldsymbol{\theta}}(j-1), \boldsymbol{P}(j-1), \lambda)$
6: **end for**
7: **return** $\hat{\boldsymbol{\theta}}(K), \boldsymbol{P}(K)$

Algorithm 2: Overnight prediction

4 Results

In this section the simulation setup as well as the results of the proposed localized weather prediction algorithm described in Sec. 2 and Sec. 3 are presented. Both ambient temperature and solar irradiation forecasting simulations use data collected over a period of 36 days. The local weather station collects measurements for ambient temperature (in deg C) and solar irradiation (in W/m^2) with a common sampling time of 15 minutes. The database for the overnight prediction consists of 181 collected daily solar irradiation values from a different weather station from a different time-frame. The dataset entries were already normalized to the normalized solar day with the correct sampling time of 15 minutes.

4.1 Ambient Temperature Forecast

The values for n and m in (1) represent the order of the ARX model. The denominator-order is set by n and represents how many past measurements are used in the model. The nominator order is defined by m and corresponds to the amount of future predictions used.
Since the ARX model represents a stochastic system rather than a physical system only an optimal model order can be determined. This optimal order was determined by calculating the global significance of the model and the significance of the individual parameters and choosing the values that offer the best trade-off. The parameters used for the simulation can be found in Table 1. There Np represents the maximal prediction horizon.

Table 1. Parameters used for the ambient temperature forecast

Variable	Value
Np	96 Samples = 24 hours
n	1
m	2
λ	0.996

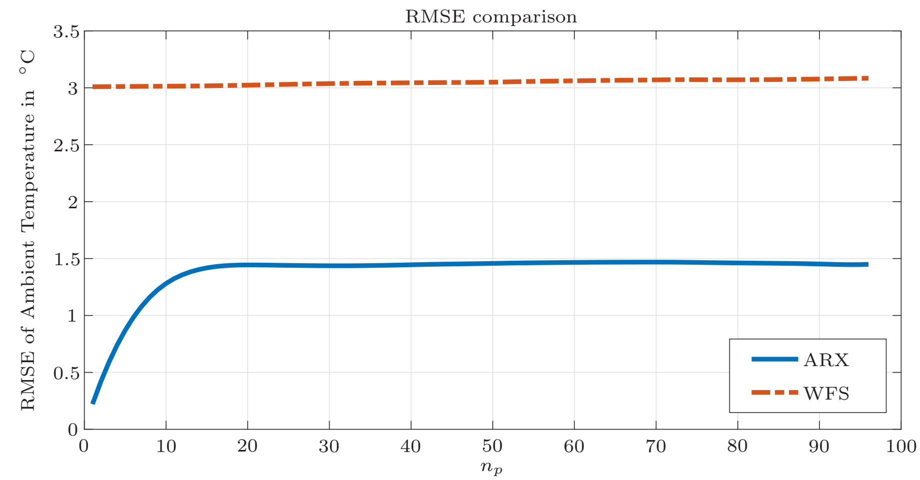

Fig. 2. RMSE between the ARX-Model output and the WFS for ambient temperature forecasting

In Fig. 2 the root mean square error (RMSE) between ϑ_{WFS} and ϑ_{amb} are shown. The RMSE for the WFS is fairly constant over the whole prediction horizon with 3°C, while the RMSE between $\hat{\vartheta}_{\mathrm{amb}}$ and ϑ_{amb} is only 1.5°C at the most and significant lower for short term predictions. This yields a forecast skill [9] ranging from 52% to 92% over the available weather predictions.
In this dataset the ARX model is primarily acting as offset compensation. The ARX model output and the WFS forecast can be seen in Fig. 3. Note that the depicted interval requires a sufficient run-in period for the parameters to settle.

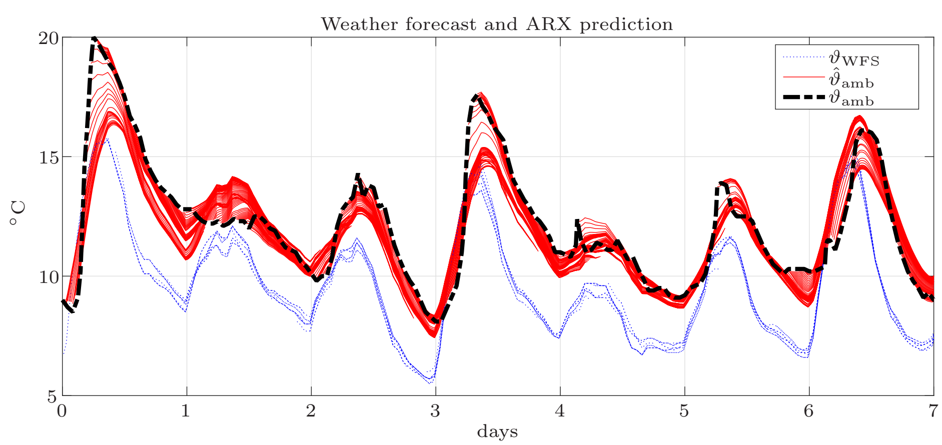

Fig. 3. Sample 7 day interval of WFS and ARX-Model output for ambient forecasting

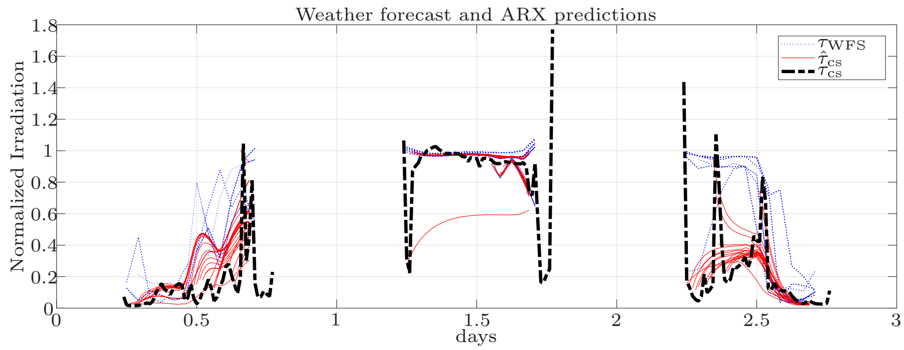

Fig. 4. Comparison of normalized WFS and ARX-Model output for solar forecasting for 3 selected days

4.2 Solar Irradiation

As seen in Sec. 4.1, the optimal values for n and m were evaluated by examining the parameter significance and the global model significance. The chosen simulation parameters can be found in Tab. 2 alongside with the parameter K which is used for the overnight prediction.

Table 2. Parameters used for the solar forecast

Variable	Value
Np	till sunset
n	1
m	3
λ	0.98
K	5

In Fig. 4 the ARX model output $\hat{\tau}_{\mathrm{cs}}$ and the WFS predictions τ_{WFS} can be seen for 3 selected days. The black dash-dotted line represents the real (measured) clear sky index τ_{cs}. During the start/end of the sunny days the value for the clear sky index τ_{cs} shows erratic changes. This is due to the small magnitudes of solar irradiation measured and also small magnitudes of clear sky solar irradiation G_{cs} which leads to ill-conditioned normalization. During the day, when the clear sky index settles, the predictions are reliable. In Fig. 5 the outputs have been converted to GHI in $\mathrm{W/m}^2$ and the clear sky global horizontal irradiation G_{cs} is plotted for reference.

Figure 6 gives a better overview on the accuracy of the ARX model compared to the predictions provided by the WFS. The development of the RMSE error over the prediction horizon is plotted for the ARX model output and the WFS predictions. Both predictions have an increasing RMSE over the prediction horizon with the ARX predictions displaying a significantly better RMSE for short

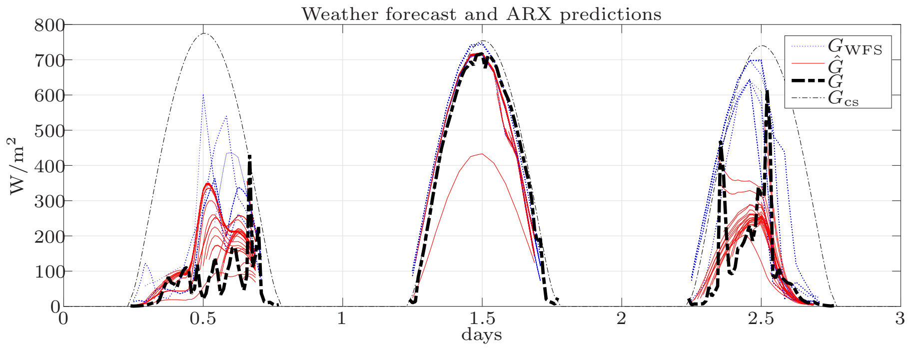

Fig. 5. Comparison of WFS and ARX-Model output for solar forecasting for 3 selected days

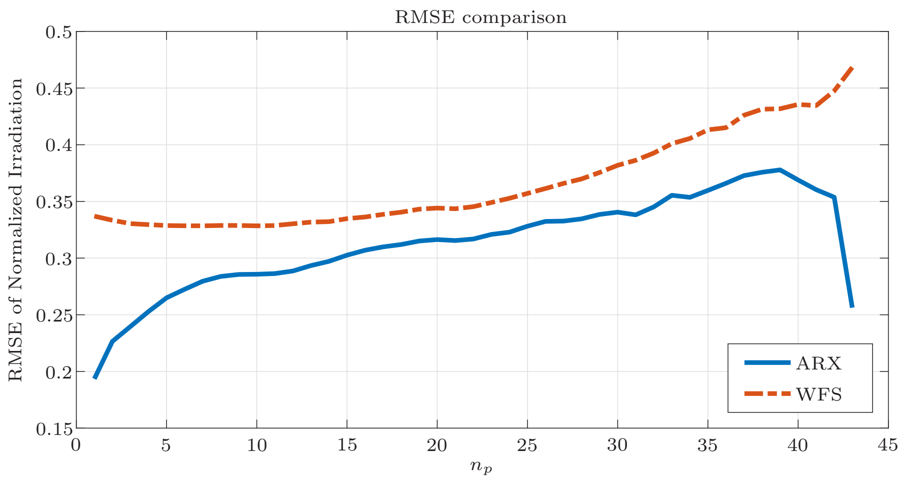

Fig. 6. RMSE between the ARX-Model output and the WFS for solar forecasting

term predictions and at the end of n_p. The forecast skill over the WFS model is 8% to 42% for the solar irradiation forecast.

5 Conclusion

Inspired by previous studies a forecasting method for ambient temperature and solar irradiation has been developed. The proposed approach localizes the numerical weather prediction provided by WFS to increase the local accuracy and reduce the forecasting errors. The overnight adaption of the ARX model parameters allow the model to accommodate to unmeasured changes and ensures a well tuned start into the next day.
Simulation results with real data showcase the benefits of the proposed methods. The method is not computational intensive and can easily be run online

on low-cost CPUs, for example in home automation systems. There the localized forecasts could provide better predictions for smart home controllers and therefore increase the comfort, save money and energy.

References

1. Bacher, P., Madsen, H., Nielsen, H.A.: Online short-term solar power forecasting. Solar Energy **83**(10), 1772–1783 (2009)
2. Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B.: Intra-hour forecasting with a total sky imager at the uc san diego solar energy testbed. Solar Energy **85**(11), 2881–2893 (2011)
3. Fernndez-Peruchena, C.M., Gastn, M.: A simple and efficient procedure for increasing the temporal resolution of global horizontal solar irradiance series. Renewable Energy **86**, 375 – 383 (2016). https://doi.org/https://doi.org/10.1016/j.renene.2015.08.004, http://www.sciencedirect.com/science/article/pii/S0960148115302044
4. Giorgi, M.D., Malvoni, M., Congedo, P.: Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine. Energy **107**, 360 – 373 (2016). https://doi.org/https://doi.org/10.1016/j.energy.2016.04.020, http://www.sciencedirect.com/science/article/pii/S0360544216304261
5. Killian, M., Zauner, M., Kozek, M.: Comprehensive smart home energy management system using mixed-integer quadratic-programming. Applied Energy **222**, 662 – 672 (2018). https://doi.org/https://doi.org/10.1016/j.apenergy.2018.03.179, http://www.sciencedirect.com/science/article/pii/S0306261918305282
6. Larson, D.P., Nonnenmacher, L., Coimbra, C.F.: Day-ahead forecasting of solar power output from photovoltaic plants in the american southwest. Renewable Energy **91**, 11 – 20 (2016). https://doi.org/https://doi.org/10.1016/j.renene.2016.01.039, http://www.sciencedirect.com/science/article/pii/S0960148116300398
7. Marquez, R., Coimbra, C.F.: Forecasting of global and direct solar irradiance using stochastic learning methods, ground experiments and the nws database. Solar Energy **85**(5), 746–756 (2011)
8. Stein, J.S.: The photovoltaic performance modeling collaborative (pvpmc). 2012 38th IEEE Photovoltaic Specialists Conference (2012). https://doi.org/10.1109/pvsc.2012.6318225, https://pvpmc.sandia.gov/
9. Wang, Z., Tian, C., Zhu, Q., Huang, M.: Hourly solar radiation forecasting using a volterra-least squares support vector machine model combined with signal decomposition. Energies **11**(1), 68 (2018)