Journal of the Meteorological Society of Japan

EARLY ONLINE RELEASE

The DOI for this manuscript is
DOI:10.2151/jmsj.2020-048
J-STAGE Advance published date: June 12th 2020

Intra-day Forecast of Ground Horizontal Irradiance Using Long Short-Term Memory Network (LSTM)

Xiuhong Chen[1], Xianglei Huang[1*], Yifan Cai[2], Haoming Shen[3], Jiayue Lu[4]

1. Department of Climate and Space Sciences and Engineering, the University of Michigan, Ann Arbor, Michigan, USA
2. School of Software Engineering, the Shanghai Jiao Tong University, Shanghai, China
3. Department of Industrial and Operations Engineering, the University of Michigan, Ann Arbor, Michigan, USA
4. Department of Computer Science, University of Southern California, Los Angeles, CA, USA

* Corresponding author: Prof. Xianglei Huang, Dept. of Climate and Space Sciences Engineering, the University of Michigan, Ann Arbor, MI 48109-2143, USA.
xianglei@umich.edu, +1-734-936-0491.

Submitted to Journal of the Meteorological Society of Japan on March 8, 2020

The first revision on May 6, 2020

The second revision on May 19, 2020

Abstract

Accurate forecast of ground horizontal irradiance (GHI) is one of the key issues for power grid managements with large penetration of solar energy. A challenge for solar forecasting is to forecast the solar irradiance with a lead time of 1-8 hours, here termed as intra-day forecast. This study investigated an algorithm using a long short-term memory (LSTM) model to predict the GHI in 1-8 hours. The LSTM model has been applied before for inter-day (>24 hours) solar forecast but never for the intra-day forecast. Four years (2010 - 2013) of observations by the National Renewable Energy Laboratory (NREL) at Golden, Colorado were used to train the model. Observations in 2014 at the same site were used to test the model performance. The results show that, for a 1-4 hour lead time, the LSTM-based model can make predictions of GHIs with root-mean-square-errors (RMSE) ranging from 77 to 143 Wm^{-2}, and normalized RMSEs around 18.4~33.0 %. With 5-minute inputs, the forecast skill of LSTM with respect to smart persistence model is 0.34~0.42, better than random forest forecast (0.27) and the numerical weather forecast (-0.40) made by the Weather Research and Forecasting (WRF) model. The performance levels off beyond 4-hour lead time. The model performs better in fall and winter than in spring and summer, and better under clear-sky conditions than under cloudy conditions. Using adjacent information from the reanalysis as extra inputs can further improve the forecast performance.

Keywords: Atmospheric radiation; Solar energy forecast; Recurrent neural network

1. Introduction

Solar photovoltaic (PV) energy is a major source of renewable energy. Over the last decade it has been increasingly integrated into worldwide electricity grid. Electricity output by solar panels depends on solar irradiance reaching the horizontal surface, a.k.a. global horizontal irradiance (GHI). GHI includes three components: direct solar beam, diffusive solar radiation by atmosphere and clouds, and ground-reflected solar radiation (usually insignificant). The GHI is affected by several environmental factors such as solar irradiance at the top of the atmosphere, cloud fraction, aerosol loading, and relative humidity. As a result, the amount of electricity produced by a solar panel is subjected to fluctuations in such meteorological variables (Bae et al., 2017). This is distinctively different from traditional power generators that are stable over the entire operational period and insensitive to actual weather variations. The effect of such fluctuations becomes significant when PV power generation consists of 5 % of overall power used over a local grid and becomes critical when the PV penetration reaches 20 % (Matsura, 2009). As a result, some traditional generators must be used as a backup (a.k.a. operating reserve) to stabilize the power grid. However, conventional power generation systems (including nuclear) all have their own "warm-up" time, ramp rate, and minimal possible load (IEA, 2014). Therefore, improving the accuracy of solar irradiance forecasting can be critically important for efficient grid management with large penetration of solar energy.

Depending on the need for solar energy for different time horizons, different types of solar forecasting techniques can be used. For example, intra-hour forecasting (also known as nowcasting) refers to prediction of GHI from a few minutes up to one hour. It is usually done by simply extrapolating current wind and cloud information with an assumption that cloud property does not change and merely moves with the wind (Bosch and Kleissl, 2013; Chow et al., 2015;

Chu et al., 2015; Lorenz et al., 2014; Rana et al., 2016). Thus, only endogenous data are needed for such forecasting. For another example, inter-day forecasting refers to prediction of changes of GHI for one to several days ahead. For such a timescale, the evolution of weather system becomes the leading factor in determining the variations of solar irradiance. As a result, the most useful inter-day forecasts heavily rely on the numerical weather prediction (NWP) (e.g., Alessandrini et al., 2015; Mellit et al., 2014; Almeida et al., 2015). Between the above two ends of time horizon is the intra-day forecast, i.e., solar forecasting with a lead time of 1-8 hours. Many cloud systems can develop, mature, and dissipate within eight hours. The multiscale turbulence nature of cloud systems plays important roles within such timescale. State-of-the-art NWP models still have difficulties simulating such phenomena and, therefore, also have difficulty to well resolve the GHI variations within 8 hours (Cros et al., 2014).

The machine-learning approach recently emerges as a promising alternative to do the intra-day forecast since neither simple extrapolation technique nor pure physics-based NWP technique has proved to be fully successful for such type of forecast (Antonanzas et al., 2016; Nonnenmacher and Coimbra, 2014; Miller et al., 2018). A number of algorithms have been used to forecast the GHI over this timescale. For example, David et al. (2018) compared the intraday forecasting of deterministic GHI by three models: coupled autoregressive and dynamical system (CARDS), sequential neural network (NN), and more traditional recursive autoregressive and moving average (ARMA) model. Their results showed that the performance of the ARMA model is comparable to that of NN, but both of them outperform the CARDS.

Some studies have employed Long Short-Term Memory networks (LSTMs) to forecast photovoltaic (Gensler et al., 2016), total electron content (Sun et al., 2017), and day-ahead solar irradiance (Qing and Niu, 2018). LSTM is a particular kind of recurrent neural network (RNN;

refer to the appendix) introduced by Hochreiter & Schmidhuber (1997). Wang et al. (2018) and Fu et al. (2016) found that a LSTM network can achieve a better performance than the traditional time series prediction methods such as the ARMA and Support Vector Machine Regression. To the best of our knowledge, there has been no use of LSTM for intra-day solar forecasting yet. Therefore, inspired by the success of the LSTM for the inter-day solar forecasting, this study aims at assessing the performance of a LSTM-based model for intra-day solar forecasting. The remainder of this paper is organized as follows. Section 2 describes the structure of the LSTM model used in this study, the observational data used in this study, and the validation metrics used for assessing the performance of the forecast. Section 3 presents the forecasting results and sensitivity tests regarding the choices of input features, temporal resolutions, and some configurations of the LSTM model. Section 4 compares the performance of the LSTM model with that of several other forecast methods. Conclusions and further discussion are given in Section 5.

2. LSTM, Data, and Methodology

2.1 LSTM-based neural network structure

Figure 1 sketched the structure of LSTM-based time series prediction scheme used in this study. Mathematical description of the LSTM model can be found in the appendix. Unlike normal multi-layer perceptron framework, the LSTM has a special neuron structure called memory cell that can capture short-term memory and keep it for a long time (Graves, 2012). Three gates control the information flow into and out of the memory cell: input gate, output gate, and keep gate (also known as forget gate). If the forecast is made at time t_0 using m sets of observations prior to it, m LSTM models are then needed. The features (i.e. inputs) fed into each LSTM model through the input gate can be GHI only (i.e. one feature only) as well as a set of features including GHI. For our case, the set of features includes GHI, local time, and other observable meteorological

variables including relative humidity (RH; %), total cloud fraction (F_{cld}; unitless), near-surface temperature (T_s; K), and solar irradiance at the top of the atmosphere (R_{toa}; Wm^{-2}), as shown in Figure 1. These meteorological variables all have direct relevance to the solar radiative transfer in the atmosphere. All LSTM models were implemented using Keras neural network library (Chollet, 2015), which runs on top of TensorFlow. The output, which has a dimension of out_d, from the LSTM model, is then used to as the input to a dense layer to generate forecast of the GHI in the next n hours after time t_0. The final output for our case is hourly-averaged GHI for the n^{th} hours after time t_0. Previous studies such as Gensler et al. (2016) and Wang et al. (2018) showed that, in order to do a sequence of forecast, the input for such LSTM framework should be 2-5 times as long as the sequence of the forecast. Our targeted sequence is 8-hour GHI forecast, thus we have tested the length of input from 8 to 56 hours prior to the forecast time. It turns out that using 32-hour observations prior to the forecast time t_0 has the optimal performance. In the following discussions, all results are obtained using 32 hours of features prior to the time of the forecast.

The training was deemed finished when the root mean square error decreases to a small value (in our case, ~105 Wm^{-2} or less) and keeps essentially flat within 50 more training iterations. Coefficients derived from such training were then saved and used to do the 8-hour forecast of GHI.

2.2 Data

Data used in this study was from ground observations at Golden, Colorado (39.74°N, 105.18°W, elevation of 1.829 km) by the National Renewable Energy Laboratory (NREL; Andreas and Stoffel, 1981). Original observations were made every minute. Observed GHI, total cloud fraction, near-surface relative humidity and temperature were used. The GHIs were measured using well-maintained global horizontal pyranometers with a 3 % ~ 5 % measurement uncertainty (Wilcox 2012; Wilcox and Myers, 2008). Four years of observations from 2010 to 2013 were used

to train the LSTM network. The observations in the entire year of 2014 were then used for validation. The evaluation of the LSTM forecast performance with respect to different configurations of the network was also carried out using the same sets of training and validation data.

2.3 Metrics for evaluating the model performance: RMSE, nRMSE, and forecast skill

Multiple metrics exist for evaluating the forecasting performance (Orwig et al., 2015). The most commonly used in meteorological forecast is root mean square error (RMSE), which is defined as

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{\mathrm{N}}(\mathrm{GHI}_{\mathrm{predicted},i} - \mathrm{GHI}_{\mathrm{observed},i})^2} \quad (1),$$

where N is the number of observations used in the validation. RMSE measures how close the predicted GHIs are to the observed GHIs in the absolute unit of GHI. A closely related metric is normalized RMSE (nRMSE), which is defined as the RMSE divided by the standard deviation of observed GHI, which measures the prediction error with respect to the spread of the observed quantity. nRMSE can be a more useful metric when the value of $\mathrm{GHI}_{\mathrm{observed}}$ varies by several factors or even by orders of magnitude. nRMSE has been used in previous studies about solar forecasting (Deo et al., 2016). Both RMSE and nRMSE will be used here to evaluate the model performance.

In addition, we will perform a linear fit of the predicted hourly GHIs with respect to the observed counterparts. The R-square from such linear fit tells the fraction of total variance in observed GHIs that can be explained by a linear function of predicted GHIs. Meanwhile, the slope

of the linear fit tells the scaling of such linear function (with 1:1 being the ideal) and the intercept tells the bias of such linear fit.

Forecast skill (FS), which is introduced by Coimbra et al. (2013), is widely used as a comparison metric in solar forecast. It is defined as,

$$\mathrm{FS} = 1 - \frac{\mathrm{RMSE}}{\mathrm{RMSE_p}} \quad (2),$$

Where RMSE is from the model to be assessed. $\mathrm{RMSE_p}$ is the RMSE of the persistence model. In this study, the persistence model with clear-sky index is used following Kaur et al. (2016), also known as the smart persistence. The GHI for the n^{th} hour lead time at time t, GHI($t + n$), predicted by clear-sky persistence model is defined as,

$$\mathrm{GHI}(t+n) = k(t+n) \cdot \mathrm{GHI}(t+n)_{\mathrm{clr}} \quad (3),$$

$$k(t+n) = k(t) = \frac{\mathrm{GHI}(t)}{\mathrm{GHI}(t)_{\mathrm{clr}}} \quad (4),$$

Where $k(t + n)$ and $k(t)$ are the clear-sky index, they are the same for the current time and for the n^{th} hour lead time. $\mathrm{GHI}(t)_{\mathrm{clr}}$ and $\mathrm{GHI}(t + n)_{\mathrm{clr}}$ are the clear-sky solar radiations at surface for the time t and $t + n$. They are from solar radiation data (SoDa; Lefèvre et al., 2013). A forecast skill of 1.0 indicates a perfect forecasting. A forecast skill of 0 indicates an exact same performance as the persistence model. And a forecast skill of <0 indicates even worse performance than the persistence model.

3. Results

3.1 LSTM prediction Performance

First, 32 hours of high temporal resolution data (5 minutes per data point) was used as input to the LSTM model with out_d =100 to forecast GHIs in the next eight hours with also a temporal

174 resolution of five minutes. In other words, 384 sequential inputs (i.e. 32-hour observations) were
175 used to predict the next 96 sequential outputs (i.e. GHI forecast for the next eight hours). The input
176 data vectors have six features including GHIs, time, total cloud fraction, relative humidity, near
177 surface temperature, and solar irradiance at the top of atmosphere (TOA). Only daytime forecast
178 is examined further here, as the nighttime GHI is zero and has no relevance to solar power
179 generation.

180 Figure 2 shows the scatter plots between the predicated hourly-averaged GHIs and observed
181 counterparts for a 1-hour lead time (also referred to as 1-hour forecast horizon). Results from four
182 seasons are plotted separately. The largest deviation from 1:1 line is seen in spring, which has a
183 slope of 1.09. All slopes are larger than 1 and, correspondingly, the intercepts are all negative,
184 ranging from -2 Wm^{-2} in winter to -20.5 Wm^{-2} in spring. The R-squares range from 0.88 in the
185 summer to 0.93 and 0.92 in the fall and winter seasons, respectively. In general, the prediction
186 performance is the best in winter and the poorest in summer. This is consistent with our
187 understanding of the weather system. In winter large-scale, long-lived weather systems affect the
188 cloud evolutions the most. For example, when a synoptic frontal system passes the site, it usually
189 is featured with overcast sky over a few days. Then the high-pressure system comes with entire
190 clear sky or little cloud coverage over a few days. In summer, however, large-scale weather system
191 is less organized; diurnal cycle featured with local, short-lived convective storms dominates the
192 cloud variations and thus the GHI is less predictable than in the winter. This can also be seen from
193 the scatter plots in Figure 2: the linear model fit the wintertime data much better than the
194 summertime data.

195 The RMSE, nRMSE, and linear regression outcomes (slope and intercept) for a lead time
196 from one hour to eight hours are summarized in Figure 3. For annual average, the RMSE (nRMSE)

increases from ~77 Wm^{-2} (~18%) at one-hour lead time to ~143 Wm^{-2} (~33%) at four-hour lead time. Both RMSE and nRMSE suggest that the forecast performance levels off beyond four hours. As expected, the wintertime RMSE is always the smallest among all seasons and for all lead times. Except for winter, the regression slopes (intercepts) for other seasons decrease (increase) with the increase of lead time, which can be explained by the fact that randomness increases with respect to the lead time (e.g. a convective storm or shower happens after forecast time t_0 and lasts for only a hour or two), and such randomness has little dependence on what happens before t_0. As explained above, winter weather is mostly governed by large-scale synoptic system that could last several days, which can account for the seasonal contrast here.

The dependence of forecast performance with respect to cloud fraction is also examined. Figure 4 shows the performance metrics with respect to four groups of observed cloud fraction. The RMSE for the group with cloud fraction <1 % is significantly less than those for the rest three groups. The RMSE for such clear or nearly clear sky situation mainly rises from the aerosol absorption and scattering of sunlight and the water vapor absorption of solar radiation. While the input features such as surface temperature and relative humidity are related to water vapor, due to measurement limitation our input features do not include information about aerosols. When cloud fraction is less than 95 %, the RMSE and nRMSE both level off after the 4^{th} hour. However, this is not the case for overcast or nearly overcast situation, where both RMSE and nRMSE keep increasing with forecast horizon. The slope of regression also keeps dropping with increase of forecast horizon, dropping to as low as 0.6 for the 8^{th} hour, but the intercept of regression is always less than 10 Wm^{-2} and is nearly zero for the 8^{th} hour. These highlight the challenge of solar forecasting for overcast skies. From the meteorological perspective, overcast sky implies small GHI and even a small amount of reduction in cloud fraction can quickly increase the amount of

220 solar irradiance reaching the surface. Within eight hours, an overcast sky at t_0 can evolve to any
221 possibilities (e.g., still overcast, completely clear sky, and partly cloudy), which explains at least
222 partly why the forecast performance for this group is not as good as for the rest three groups and
223 why it has no apparent level-off forecast horizon.

224 **3.2 Sensitivity of the LSTM performance to different configurations**

225 We further examine forecast performance with respect to the inputs at different temporal
226 resolutions, to the different choices of input features, and to a key hyperparameter, the number of
227 the last LSTM outputs (out_d). The goal is to estimate an optimal configuration for the LSTM
228 forecast algorithm in terms of computational speed and accuracy. RMSE and nRMSE will be used
229 in the assessment of different configurations, as they are more useful than the linear regression
230 statistics for such purposes (Orwig et al., 2015).

231 **3.2.1. Temporal resolutions**

232 Five temporal resolutions for the input features are tested, i.e., 5-minute, 15-minute, 20-
233 minute, 30-minute, and 1-hour inputs. Other configurations are identical to what is used in Section
234 3.1. For each temporal resolution, the LSTM network is trained separately. The performances are
235 shown in Figure 5. After the 4th hour, the RMSE and nRMSE both show little dependence on the
236 temporal resolutions of the input features. Within four hours, a finer temporal resolution can
237 improve the performance but the dependence is stepwise instead of linear. For example, the 5-
238 minute and 15-minute inputs have nearly identical performance, so do the 20-minute and half-hour
239 inputs. The improvements from one-hour to half-hour resolution, in terms of RMSE and nRMSE,
240 are ~ 20 % while from half-hour to five-minute resolution the improvement is only ~ 10 %. On the
241 same computer platform, it takes 5, 22, 50, 88, and 790 seconds to finish one iteration of training
242 for the inputs with 1-hour, 30-minute, 20-minute, 15-minute and five-minute resolutions,

243 respectively. The prediction takes 37, 134, 306, 480 and 4848 seconds for the inputs with 1-hour,
244 30-minute, 20-minute, 15-minute and five-minute resolution, respectively. The computational cost
245 from 15-minute to five-minute resolution increases by a factor of ~10 while the forecast
246 performance is comparable. Given the RMSE dependence and the computing cost, inputs with 15-
247 minute resolution are recommended among all temporal resolutions examined here.

248 **3.2.2 Choices of input features**

249 Among six input features, the most important one is observed GHI. It is meaningful to look
250 a LSTM network using the GHI as the only input feature and to compare its performance with the
251 counterpart that uses all six features as input to train the LSTM network. Figure 6 summarizes the
252 comparisons of performance between six features and only the GHI feature. For the 1-hour inputs,
253 though the 6-feature prediction always performs better than the one-feature prediction, the
254 difference is small for the first two hours (~ 2 %) and then increases with the forecast horizon to
255 ~5 %. For the 15-minute inputs, the difference is only 1 % or less. On the same platform, one
256 iteration in training of 6-feature and one-feature networks takes nearly the same amount of time.
257 Given no difference in training, six features should be advocated instead of one feature. But if the
258 data availability is an issue, GHI observation with high temporal resolution (15-minute or even
259 higher) alone can be used for such forecast.

260 **3.2.3 Choice of out_d**

261 Five configurations of out_d are tested. The performances are shown in Figure 7. For 1-hour
262 inputs (Figure 7a and 7c), when out_d varies from 50 to 100, the performance is essentially the
263 same. When out_d increases from 10 to 50, the performance is significantly improved for the first-
264 hour prediction. For the case of 15-minute (Figure 7b and 7d), however, changing out_d from 10

to 100 nearly leads to no improvement in the performance for all forecast horizons. The trade-off between out_d and the temporal resolution of the inputs can be clearly seen by contrasting left panels and right panels in Figure 7. For a 1-5 hour lead time, the performances with any out_d using 15-minute inputs are better than the performances with out_d =100 and 1-hour inputs. Beyond the 5th hour, however, the performance using 15-minute inputs becomes comparable to the performance using 1-hour inputs with out_d ≥50. The time to train a LSTM network with out_d = 100 is 2.5 times as much as the time needed to train a LSTM network with out_d = 10. Based on these facts, 15-minute inputs with out_d =10 is an affordable option with performance no worse than any other configurations examined here.

3.2.4 With additional neighboring features from reanalysis

This section is to test whether the features from adjacent locations can help improve the forecast. Given the actual GHI observations at neighboring locations to this NREL site are not available, we use hourly GHIs from the latest European Centre for Medium-Range Weather Forecasts (ECMWF) ERA5 reanalysis (C3S, 2017) over a 0.75-degree by 0.75-degree domain centered at the grid box encompassing NREL location. The hourly GHIs from ERA5 is at a resolution of 0.25-degree by 0.25-degree. The ERA5 GHIs on 3-by-3 grids are thus used as input features to the LSTM to predict the GHI at NREL, together with other features used in the previous subsections. Figure 8 shows the differences between with and without GHIs from the ERA5 reanalysis, for RMSE and nRMSE, respectively. The differences are negative up to a 7-hour lead time, with maximum difference being 5 Wm^{-2} for the RMSE and 1% for the nRMSE. Therefore, using neighboring GHI features near the observed site can help the prediction performance.

4. The forecast skill and comparisons with other forecast methods

Using the forecast skill (FS) defined in subsection 2.3 as a metric, we compare the performance of LSTM with that of the random forest method (Breiman, 2001) and of numerical weather prediction made by the NCAR's Weather Research and Forecasting version 3.9 (WRF; Skamarock et al., 2008), as well as the FS from a physical method published by Miller et al. (2018). The same 5-minute data used for the LSTM model training is used for the random-forest training. The WRF model forecast domain is centered at NREL Golden site with an outside domain of 192 km by 192 km (a 3-km grid on each direction) nested with an inner domain of 64 km by 64 km (1-km grid on each direction). The model used NCEP operational Global Forecast System (GFS) analysis as initializations and lateral boundary conditions for a 8-hourly daytime forecast every day in 2015, and then the GHI forecast were evaluated against observation.

The overall comparisons are summarized in Table 1. The LSTM algorithm outperforms the random forest, with ~2 % reduction for nRMSE and 8-10 Wm^{-2} reduction in RMSE. Compared to numerical weather forecast by the WRF model, the LSTM algorithm can reduce RMSE by 54~68 Wm^{-2} and nRMSE by ~12%. Miller et al. (2018) used a NWP model in combination with satellite-based cloud advection techniques to predict GHIs with a lead time of 1-3 hours. They evaluated their method at four different locations including the Table Mountain in Boulder, Colorado (40.125° N, 105.237° W), which is close to the NREL site. They found that, for this site over a three-year period, the 1-hour and 3-hour forecast RMSEs are 194.7 and 237.9 Wm^{-2}, respectively. The forecast skill is negative. These findings are comparable to our WRF model results (Table 1). In contrast, the 3-hour and 1-hour forecast skills of LSTM are 0.22~0.33 for hourly inputs and 0.37~0.42 for 5-minute inputs, better than those of NWP forecast. Forecast skill for all 1-8 hour lead time is shown in Figure 9. Overall, the forecast skill in 2014 ranges from 0.22 to 0.39 for hourly inputs and 0.34 to 0.42 for 5-minute inputs.

5. Conclusion and Discussion

This study designed a LSTM-based model to perform intra-day GHI forecast and evaluated its performance using observations from a single station at Golden, Colorado by NREL. Four years of observations from 2010 to 2013 were used to train the model, and the evaluation is based on the prediction for the entire year of 2014. The normalized RMSE is ~20 % for the 1st-hour prediction. The performance is better in the winter (RMSE = 58.0 Wm^{-2}) than in the summer (RMSE = 109.1 Wm^{-2}), which is consistent with our physical understanding of the traits of different weather systems governing the GHI variations in different seasons. The forecast performance levels off after the fourth hour with a nRMSE around 30%. The forecast skill to smart persistence model is 0.22 to 0.39 for hourly inputs and 0.34 to 0.42 for 5-minute inputs. Such performance is better than that of forecast based on random forest method (0.27), and is much better than that of numerical weather forecast using the WRF model (-0.40). The performance within first four hours show dependence on the temporal resolution of the input features: inputs with high temporal resolution, in general, perform better than those with low temporal resolution but the performance shows step-wise dependence with temporal resolution. Using multiple features with physical relevance to GHI in the context of atmospheric radiation outperforms a single GHI feature, especially for forecast beyond two hours. But there is a trade-off between the temporal resolution of the inputs and the number of input features, as well as a trade-off between the temporal resolution of the inputs and the choice of a key hyperparameter, i.e., the number of outputs from the last cell of the LSTM (out_d). With computational cost and accuracy both considered, an optimal LSTM configuration for the intra-day GHI forecasting can be 15-minute GHI as the only input to LSTMs with out_d =10. We also found that using the GHI features from neighbor locations (such as GHIs derived from the ERA5 reanalysis) can improve the LSTM model forecast performance.

This study explores the potential of LSTM in solar forecasting with the use of only indigenous observations. Limited by observational data, some key physical factors such as aerosol loadings were not available for this study. In principle introducing such features that are directly related to GHI can have promise to further improve the performance of machine-learning based algorithm, which is one pathway that we are interested in pursuing further. Meanwhile, introducing features from other data sources, such as simultaneous satellite observations and information about large-scale weather patterns from NOAA National Weather Service, also have a chance to improve the performance of such algorithms further. If the LSTM, as one type of recurrent neural network (RNN), can be successfully employed for solar forecasting, it is encouraging to further explore other RNNs in such forecast applications.

Acknowledgments

We want to thank two anonymous reviewers for their comments and suggestions. We also thank the Editorial Committee members for their comments and careful proofreads. The National Renewable Energy Laboratory (NREL) ground measurements are downloadable at https://midcdmz.nrel.gov/apps/go2url.pl?site=BMS. SoDa data is downloaded from http://www.soda-pro.com/web-services/radiation/cams-mcclear. GFS forecast and analysis data is downloaded from https://rda.ucar.edu/datasets/ds084.1/. Upon acceptance, the code used in this study will be available at https://github.com/xiuhongchen/Solar_forecast_LSTM. This work was supported by Office of Research at the University of Michigan.

Appendix

A description of LSTM neural network cell

Details of LSTM neural network can be found in Olah (2015) and Sun et al. (2017). Below is a brief summary of the essence of LSTM. Figure A1 shows the typical LSTM neural network cell. The data flow in and out of three type of gates, namely $\boldsymbol{h}(t\text{-}1)$ and $\boldsymbol{x}(t)$ as two input gates, $\boldsymbol{C}(t)$ as a keep gate, and $\boldsymbol{h}(t)$ as an output gate. $\boldsymbol{h}(t\text{-}1)$ contains outputs from the previous LSTM cell state and $\boldsymbol{x}(t)$ are the current inputs directly from the input layer. $\boldsymbol{C}(t)$ and $\boldsymbol{h}(t)$ are expressed as :

$$\boldsymbol{C}(t) = \boldsymbol{f}_t * \boldsymbol{C}(t-1) + \boldsymbol{i}_t * \widehat{\boldsymbol{C}}_t \quad \text{(A.1)},$$

$$\boldsymbol{h}(t) = \boldsymbol{o}_t * \tanh[\boldsymbol{C}(t)] \quad \text{(A.2)},$$

where * stands for the scalar product of two vectors or matrices, and

$$\boldsymbol{f}_t = \sigma(\boldsymbol{W}_{f,x} \cdot \boldsymbol{x}(t) + \boldsymbol{W}_{f,h} \cdot \boldsymbol{h}(t-1) + b_f) \quad \text{(A.3)},$$

$$\boldsymbol{i}_t = \sigma(\boldsymbol{W}_{i,x} \cdot \boldsymbol{x}(t) + \boldsymbol{W}_{i,h} \cdot \boldsymbol{h}(t-1) + b_i) \quad \text{(A.4)},$$

$$\boldsymbol{o}_t = \sigma(\boldsymbol{W}_{o,x} \cdot \boldsymbol{x}(t) + \boldsymbol{W}_{o,h} \cdot \boldsymbol{h}(t-1) + b_o) \quad \text{(A.5)},$$

$$\widehat{\boldsymbol{C}}_t = \tanh(\boldsymbol{W}_{\hat{C},x} \cdot \boldsymbol{x}(t) + \boldsymbol{W}_{\hat{C},h} \cdot \boldsymbol{h}(t-1) + b_{\hat{C}}) \quad \text{(A.6)},$$

where σ denotes the sigmoid function and tanh denotes hyperbolic tangent function. The parameters to be determined by training are weight matrices $\boldsymbol{W}_{f,x}$, $\boldsymbol{W}_{f,h}$, $\boldsymbol{W}_{i,x}$, $\boldsymbol{W}_{i,h}$, $\boldsymbol{W}_{o,x}$, $\boldsymbol{W}_{o,h}$, $\boldsymbol{W}_{\hat{C},x}$,$\boldsymbol{W}_{\hat{C},h}$ and bias vectors b_f, b_i, b_o and $b_{\hat{C}}$.

Based on above formulas, the interpretation of each LSTM cell can be given as follow. $\boldsymbol{f}_t$ is between 0 and 1, with one being completely keeping the previous cell state, $\boldsymbol{C}(t\text{-}1)$, and zero being

completely discarding it. $\widehat{\boldsymbol{C}}_t$ are new candidate values derived from tanh function with input of both outputs from the previous LSTM cell, $\boldsymbol{h}(t\text{-}1)$, and the current input, $\boldsymbol{x}(t)$. $\boldsymbol{i}_t$ determines to what extent $\widehat{\boldsymbol{C}}_t$ is used for updating the current state $\boldsymbol{C}(t)$. $\boldsymbol{o}_t$ decides the fraction of tanh($\boldsymbol{C}(t)$) to be used for computing the current output $\boldsymbol{h}(t)$.

References:

Alessandrini, S., L. Delle Monache, S. Sperati, G. Cervone, 2015: An analog ensemble for short-term probabilistic solar power forecast. *Appl. Energy*, **157**, 95–110.

Almeida, M. P., O. Perpinan, L. Narvarte, 2015: PV power forecasting using a nonparametric PV model. *Sol. Energy*, **115**, 354–368.

Andreas, A. and T. Stoffel, 1981: NREL Solar Radiation Research Laboratory (SRRL): Baseline Measurement System (BMS); Golden, Colorado (Data); NREL Report No. DA-5500-56488. http://dx.doi.org/10.5439/1052221.

Antonanzas, J., N. Osorio, R. Escobar, R. Urraca, F.J. Martinez-de-Pison, F. Antonanzas-Torres, 2016: Review of photovoltaic power forecasting. *Sol. Energy*, **136**, 78-111.

Bae, K.Y. and Co-authors, 2017: Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis. IEEE Trans. *Power Systems*, **32**, 935-945.

Bosch, J. and J. Kleissl, 2013: Cloud motion vectors from a network of ground sensors in a solar power plant. *Sol. Energy*, **95**, 13–20.

Breiman, L., 2001: Random forests, Mach. Learn. 45, 5-32.

Chollet , F., 2015: Keras. Accessed 6 November 2018, https://github.com/fchollet/keras.

Chow, C., S. Belongie, J. Kleissl, 2015: Cloud motion and stability estimation for intra-hour solar

forecasting. *Sol. Energy*, **115**, 645–655.

Chu, Y., B. Urguhart, S. Gohari, H. Pedro, J. Kleissl, C. Coimbra, 2015: Short-term reforecasting of power output from a 48MWe solar PV plant. *Sol. Energy*, **112**, 68–77.

Coimbra, C.F.M., J. Kleissl, R. Marquez, 2013: Overview of solar forecasting methods and a metric for accuracy evaluation. In Solar Energy Forecasting and Resource Assessment, 1st ed.; Kleissl, J., Ed.; Elsevier Academic Press: Boston, MA, USA, 2013; Chapter 8, pp. 171–194, ISBN 9780123971777.

Copernicus Climate Change Service (C3S), 2017: ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate . Copernicus Climate Change Service Climate Data Store (CDS). Accessed 20 November 2019, https://cds.climate.copernicus.eu/cdsapp#!/home.

Cros, S., S. Jolivet, N. Sébastien, N. Schmutz, 2014: Using Satellite Data for Intraday Solar Energy Forecasts: Benchmarks of Two Cloud Trajectory Retrieval Techniques. 29th European Photovoltaic Solar Energy Conference and Exhibition, 2220 - 2224, https://doi.org/10.4229/EUPVSEC20142014-5AO.8.4.

David, M., L. Mazorra, P. Lauret, 2018: Comparison of intraday probabilistic forecasting of solar irradiance using only endogenous data. *Int. J. Forecast*, **34**, 529–547.

Deo, R.C., X.H. Wen, F. Qi, 2016: A wavelet-coupled support vector machine model for forecasting global incident solar radiation using limited meteorological dataset. *Appl. Energy*, **168**, 568–593.

Fu, R., Z. Zhang, L. Li, 2016: Using LSTM and GRU neural network methods for traffic flow prediction. In *Proceedings of the Youth Academic Annual Conference of Chinese Association*

416 *of Automation (YAC)*, Wuhan, China, 11–13 November, pp. 324–328.

417 Gensler, A., J. Henze, B. Sick, and N. Raabe, 2016: Deep learning for solar power forecasting-An
418 approach using AutoEncoder and LSTM Neural Networks. In *Systems, Man, and Cybernetics*
419 *(SMC)*, 2016 IEEE International Conference on (pp. 002858-002865).

420 Graves, A., 2012: Supervised sequence labelling with recurrent neural network. Studies in
421 Computational Intelligence, Chapter 4: Long short-term memory, P31-38. Springer.

422 Hochreiter, S. and J. Schmidhuber, 1997: Long short-term memory. *Neural Computation*, **9(8)**,
423 1735-1780.

424 IEA (International Energy Agency), 2014: The power of transformation - Wind, Sun and the
425 Economics of Flexible Power Systems. Edition, 238pp, Paris, OECD/IEA.

426 Kaur, A., L. Nonnenmacher, H.T. Pedro, and C.F. Coimbra, 2016: Benefits of Solar Forecasting
427 for Energy Imbalance Markets. *Renew. Energy*, **86**, 819–830.

428 Lefèvre, M., A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L.Wald, M. Schroedter-
429 Homscheidt, C. Hoyer-Klick, A. Arola, A. Benedetti, J. W. Kaiser, and J.-J. Morcrette, 2013:
430 McClear: a new model estimating downwelling solar radiation at ground level in clear-sky
431 conditions. *Atmos. Meas. Tech.,* **6**, 2403-2418.

432 Lorenz, E. and Co-authors, 2014: PV power predictions on different spatial and temporal scales
433 integrating PV measurements, satellite data and numerical weather predictions. In: *29th*
434 *EUPVSEC*, Amsterdam, Netherland.

435 Matsura, M., 2009: Island Breezes: Renewable Energy Integration from a Hawaiian Perspective.
436 *IEEE Power Energy Mag*., **7(6)**, 59-64.

437 Mellit, A., A. Massi Pavan, V. Lughi, 2014: Short-term forecasting of power production in a large-

scale photovoltaic plant. *Sol. Energy*, **105**, 401–413.

Miller, S.D., A.R. Matthew, J.M. Haynes, M. Sengupta, 2018: Short-term solar irradiance forecasting via satellite/model coupling. *Sol. Energy*, **168**, 102-117.

Nonnenmacher, L. and C. F. M. Coimbra, 2014: Streamline-based method for intra-day solar forecasting through remote sensing. *Sol. Energy*, **108**, 447–459.

Olah, C., 2015. Accessed 8 Feburary 2019, http://colah.github.io/posts/2015-08-Understanding-LSTMs/.

Orwig, K.D., M.L. Ahlstrom, V. Banunarayanan, J. Sharp, J.M. Wiczak, J. Freedman, S.E. Haupt, J. Cline, O. Bartholomy, H.F. Hamann, B-M Hodge, C. Finley, D. Nakafuji, J.L. Peterson, D. Maggio, M. Marquis, 2015: Recent Trends in Variable Generation Forecasting and Its Value to the Power System. *IEEE Trans. Sustain. Energy*, **6 (3)**, 924–933.

Qing, X., Y. Niu, 2018: Hourly day-ahead solar irradiance prediction using weather forecasts by lstm. *Energy*, **148**, 461–468.

Rana, M., I. Koprinska, V. Agelidis, 2016: Univariate and multivariate methods for very short-term solar photovoltaic power forecasting. *Energy Convers. Manage*, **121**, 380–390.

Skamarock, W., J. Klemp, J. Dudhia, et al., 2008: A description of the advanced research wrf version 3. Technical Note NCAR/TN-475+STR. 125 pp. University Corporation for Atmospheric Research. http://dx.doi.org/10.5065/D68S4MVH.

Sun, W.Q., L. Xu, X. Huang, W.Q. Zhang, T.J. Yuan, Z. Chen, Y.H. Yan, 2017: Forecasting of ionospheric vertical total electron content (TEC) using LSTM networks. In: *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, pp. 340–344.

Wang, Y., W. Liao, Y. Chang, 2018: Gated Recurrent Unit Network-Based Short-Term Photovoltaic Forecasting. *Energies*, **11**, 2163.

Wilcox, S.M., 2012: National Solar Radiation Database 1991–2010 Update: User's Manual (Technical Report NREL/TP-5500-54824). Golden, CO: National Renewable Energy Laboratory. http://www.nrel.gov/docs/fy12osti/54824.pdf.

Wilcox, S.M., and D. Myers, 2008: Evaluation of Radiometers in Full-Time Use at the National Renewable Energy Laboratory Solar Radiation Research Laboratory (Technical Report NREL/TP-550-44627). Golden, CO: National Renewable Energy Laboratory. http://www.nrel.gov/docs/fy09osti/44627.pdf.

Table 1 Comparison of the performance of the LSTM with the performance by other machine learning model, Random Forest algorithms and by other numerical models. RMSE refers to root-mean-square error and nRMSE refers to normalized RMSE as defined in the text. Forecast skill for different years is with respect to the results of the persistence model in the corresponding year(s). Note these results are the statistics of daytime GHIs. Random Forest is done only for 1 hour lead time.

	Time period	RMSE (Wm^{-2})		nRMSE(%)		Forecast skill	
		1 hour	3 hour	1 hour	3 hour	1 hour	3 hour
LSTM(5min-input)	2014	78.2	134.2	17.4	29.9	0.42	0.37
LSTM(1hour-input) + ERA5	2014	106.0	140.8	23.2	30.8	0.22	0.33
	2015	100.6	129.8	22.0	28.5	0.20	0.33
Random Forest (5min-input)	2014	98.7	NA	21.0	NA	0.27	NA
WRF	2015	176.3	197.5	35.9	43.0	-0.40	-0.01
Satellite+model (Miller et al., 2018)	2014-2016	194.7	237.9	24.9	28.8	-0.57	-0.21
Smart persistence	2014	135.6	211.4	28.7	41.7		
	2015	126.1	194.9	26.8	38.6		
	2014-2016*	123.9	196.0	26.0	38.3		

*The persistence results for these 3 years are calculated using SURFRAD GHIs and SoDa clear-sky irradiances at Table mountain, CO (40.125° N, 105.237° W). SURFRAD GHIs are downloaded at ftp://aftp.cmdl.noaa.gov/data/radiation/surfrad/.

Figure Captions

Figure 1. LSTM-based time series prediction framework. The input feature(s) is denoted as x and (m+1) time steps in the past are used for input to predict m time steps in the future. Three gates control the information flow in and out of each LSTM cell, namely input gate, output gate and keep gate.

Figure 2. Scatter plot of the predicted verse observed hourly-averaged GHI. The prediction is for one-hour lead time. The entire year of 2014 data were used. Solid black lines are the 1:1 reference lines. Linear regression results are shown as red line. Each panel is for one season.

Figure 3. The LSTM performance for hourly averaged GHI prediction with different lead time from one to eight hours. (a) regression slope, (b) intercept, (c) RMSE, and (d) nRMSE. Results from different seasons are shown in different colors. The performances for the entire year are shown in black curves.

Figure 4. Performance of the LSTM forecast composite with respect to four groups of cloud fraction at forecast time zero: cloud fraction <1 % in red, 1-50 % in red, 50-95 % in blue and >95 % in green. Four panels are the regression slope (a), regression intercept (b), RMSE (c) and nRMSE (d).

Figure 5. Sensitivity of RMSE (a) and nRMSE (b) to the temporal resolution of the inputs. The same LSTM configuration used for Figures 2-4 is used here.

Figure 6. Sensitivity of RMSE (a) and nRMSE (b) to number of features used as the inputs. Results predicted using 1-hour inputs and 15-minute inputs are in solid and dashed lines, respectively. One feature is to use GHI only. Six features include time, GHI, total cloud fraction, solar irradiance at

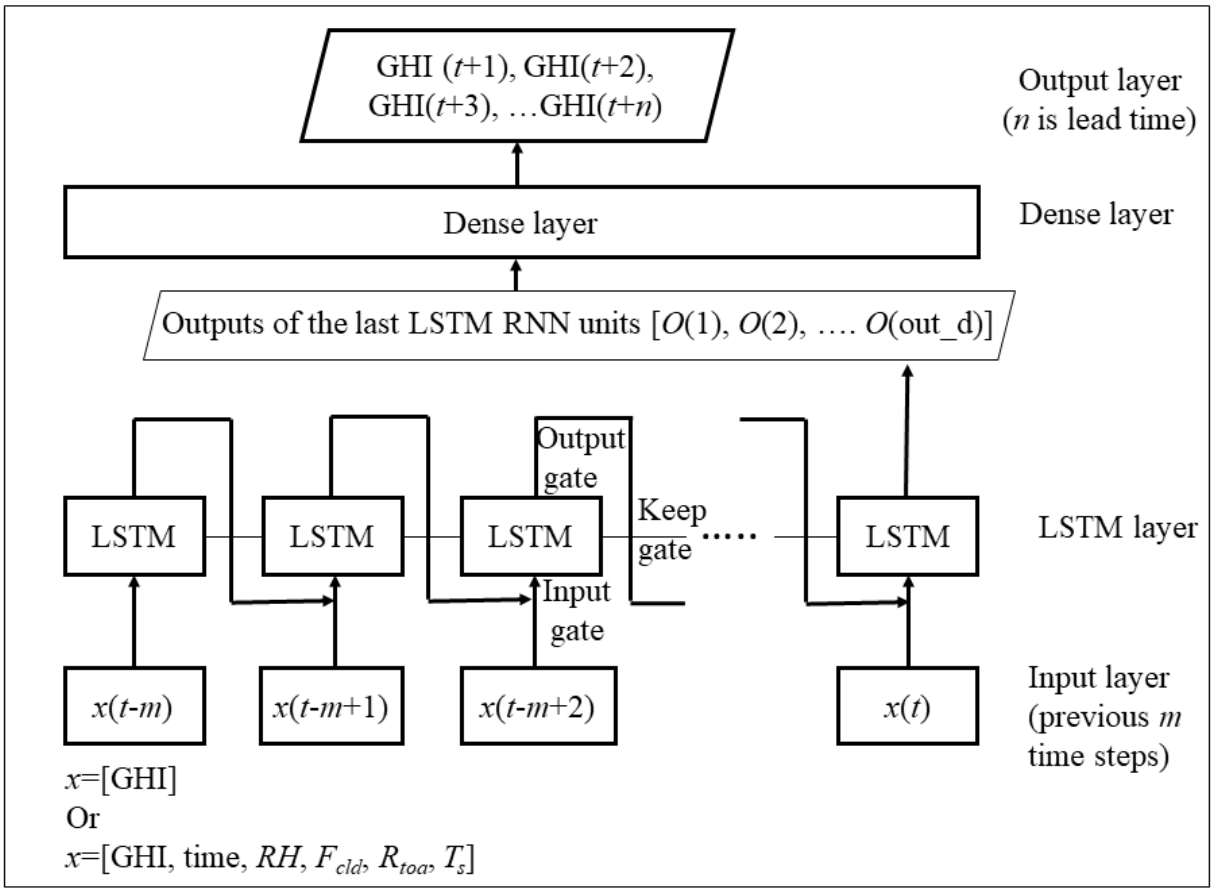
TOA, relative humidity, and air temperature. The same LSTM configuration used for Figures 2-4 is used here.

Figure 7. Sensitivity of RMSE and nRMSE to the number of LSTM output (out_d). Left column: 1-hour inputs are used. Right column: 15-minute inputs are used. Upper row is for RMSE and lower row is for nRMSE. Results from different out_ds are shown in different colors, as labeled in (a) for all the panels.

Figure 8. Difference in RMSE and nRMSE between with and without hourly ERA5 reanalysis used. Hourly inputs are used, other configurations are the same as in Figures 2-4. For black line, data in 2010-2013 are used to predict GHIs in 2014. For red line, data in 2011-2014 are used to predict GHIs in 2015.

Figure 9. Forecast skill in 2014 compared to smart persistence model.

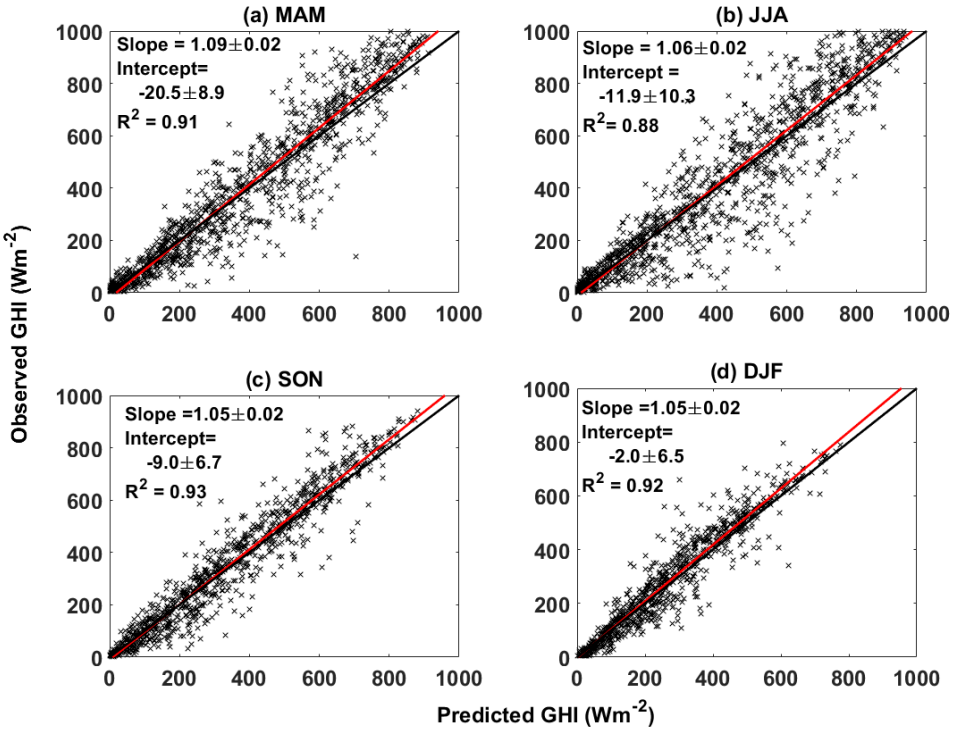
Figure A.1 Sketch of the structure of LSTM neural network cell based on illustration from Olah (2015) and Sun et al. (2017).



514

515 Figure 1. LSTM-based time series prediction framework. The input feature(s) is denoted as $\boldsymbol{x}$ and
516 (m+1) time steps in the past are used for input to predict m time steps in the future. Three gates
517 control the information flow in and out of each LSTM cell, namely input gate, output gate and
518 keep gate.

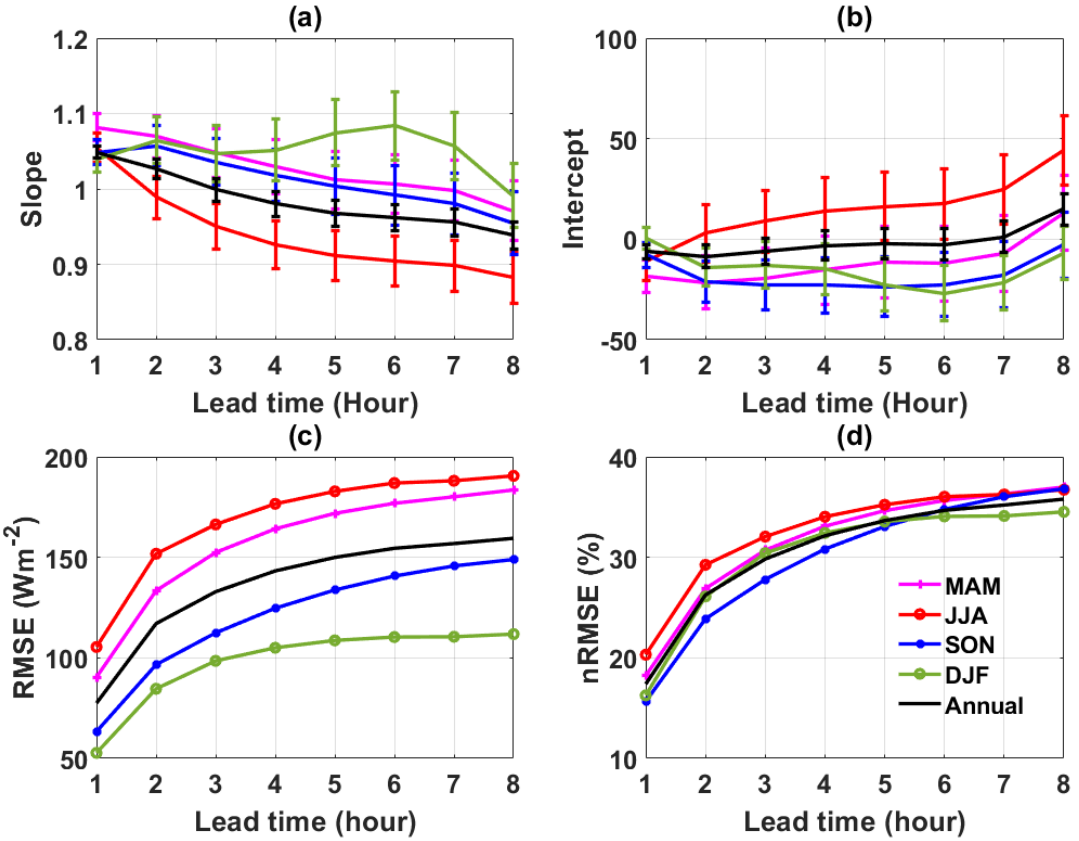
519



520

521 Figure 2. Scatter plot of the predicted verse observed hourly-averaged GHI. The prediction is for
522 one-hour lead time. The entire year of 2014 data were used. Solid black lines are the 1:1 reference
523 lines. Linear regression results are shown as red line. Each panel is for one season.

524



525

526 Figure 3. The LSTM performance for hourly averaged GHI prediction with different lead time
527 from one to eight hours. (a) regression slope, (b) intercept, (c) RMSE, and (d) nRMSE. Results
528 from different seasons are shown in different colors. The performances for the entire year are
529 shown in black curves.

530

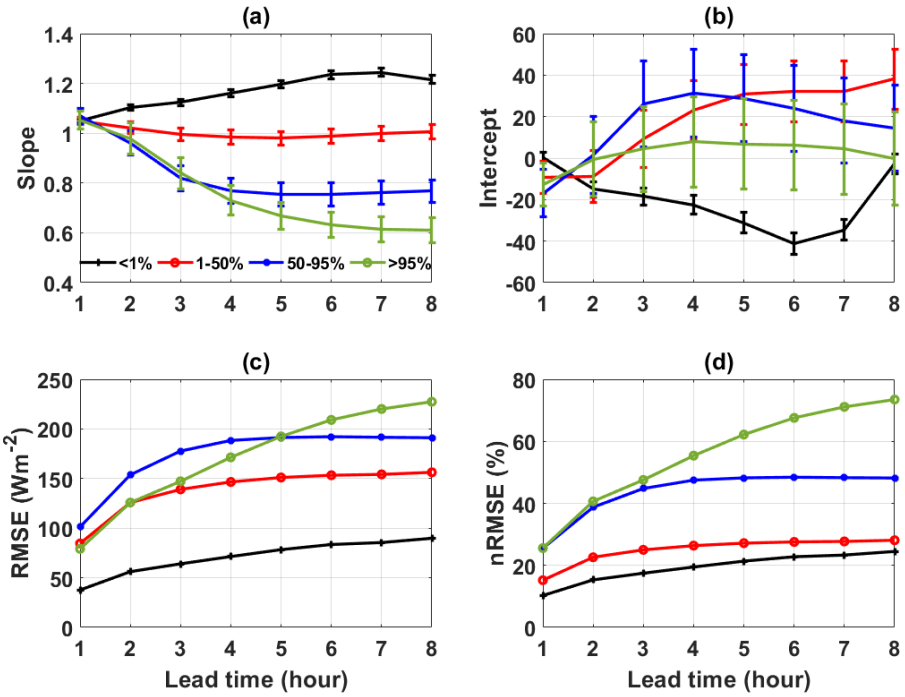
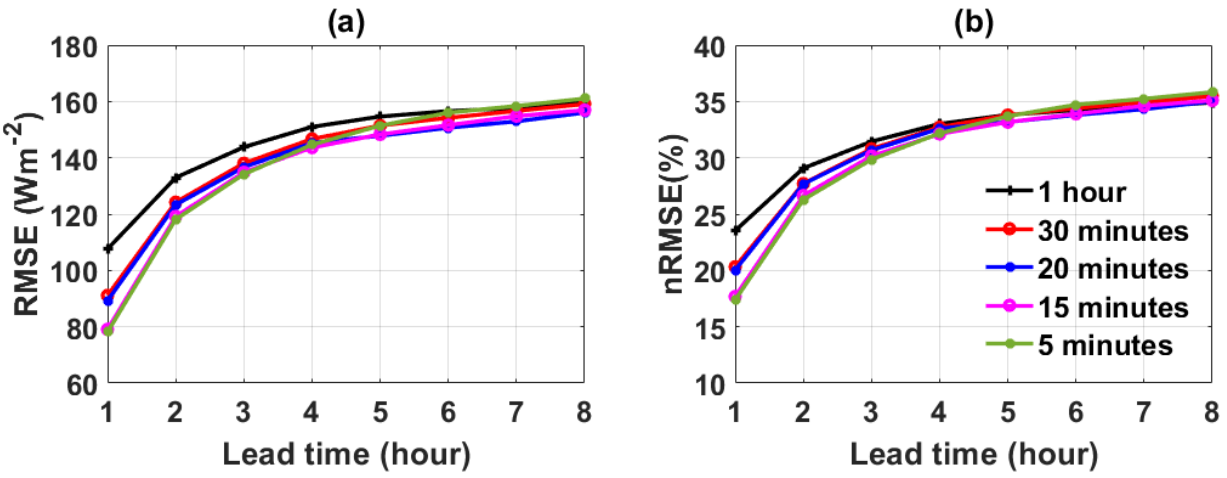


Figure 4. Performance of the LSTM forecast composite with respect to four groups of cloud fraction at forecast time zero: cloud fraction <1 % in red, 1-50 % in red, 50-95 % in blue and >95 % in green. Four panels are the regression slope (a), regression intercept (b), RMSE (c) and nRMSE (d).



537

538 Figure 5. Sensitivity of RMSE (a) and nRMSE (b) to the temporal resolution of the inputs. The

539 same LSTM configuration used for Figures 2-4 is used here.

540

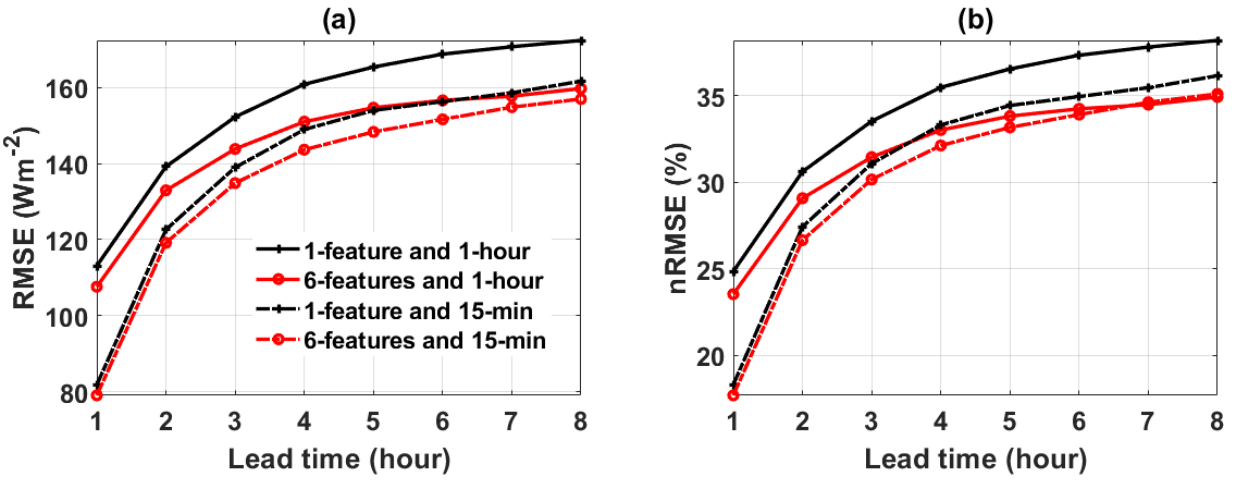
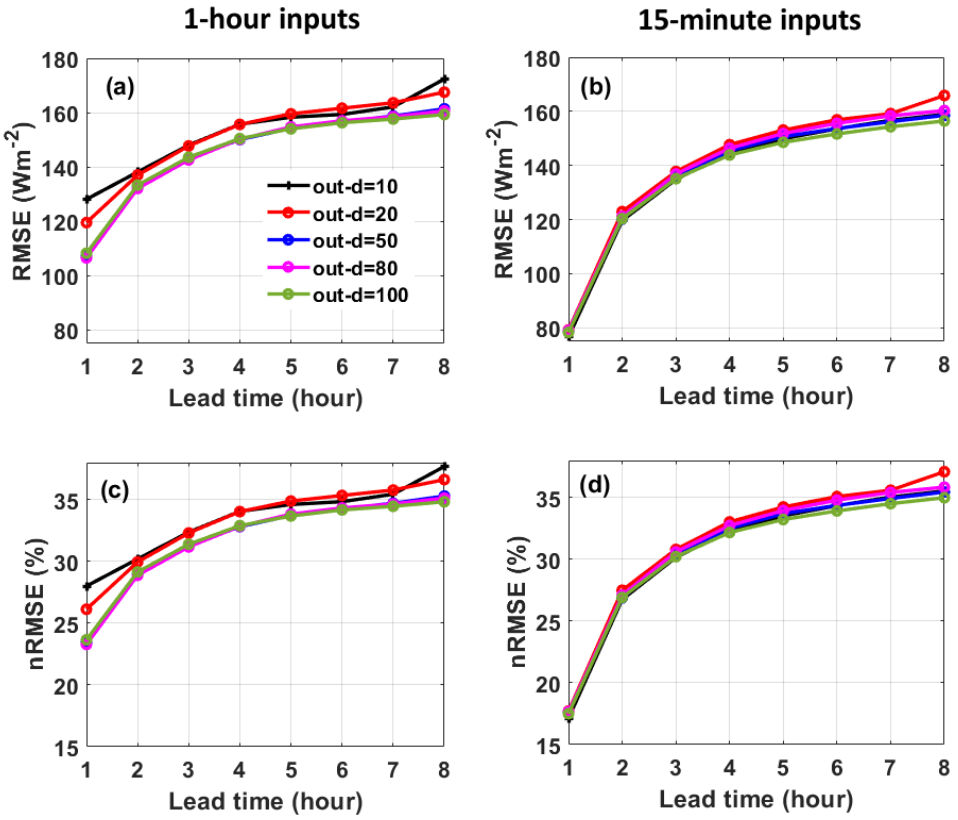


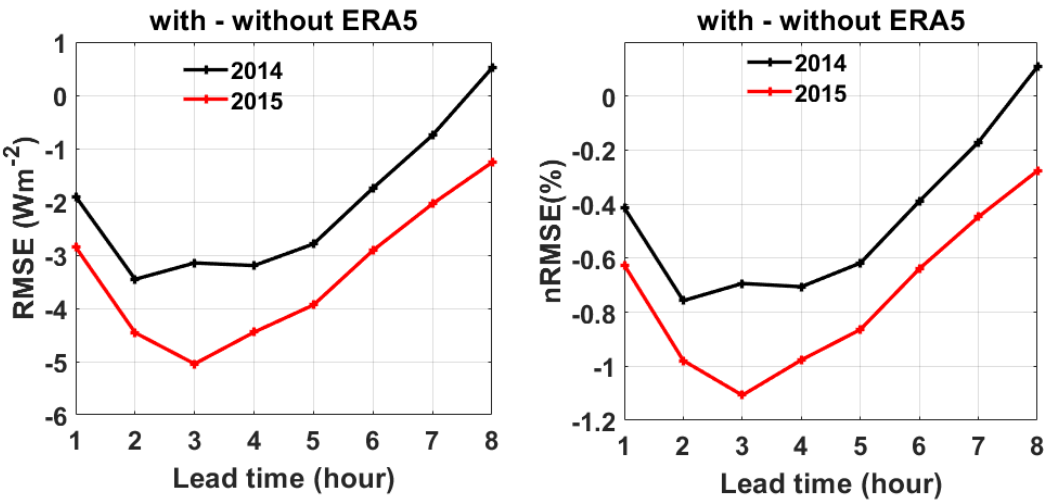
Figure 6. Sensitivity of RMSE (a) and nRMSE (b) to number of features used as the inputs. Results predicted using 1-hour inputs and 15-minute inputs are in solid and dashed lines, respectively. One feature is to use GHI only. Six features include time, GHI, total cloud fraction, solar irradiance at TOA, relative humidity, and air temperature. The same LSTM configuration used for Figures 2-4 is used here.



548

549 Figure 7. Sensitivity of RMSE and nRMSE to the number of LSTM output (out_d). Left column:
550 1-hour inputs are used. Right column: 15-minute inputs are used. Upper row is for RMSE and
551 lower row is for nRMSE. Results from different out_ds are shown in different colors, as labeled
552 in (a) for all the panels.

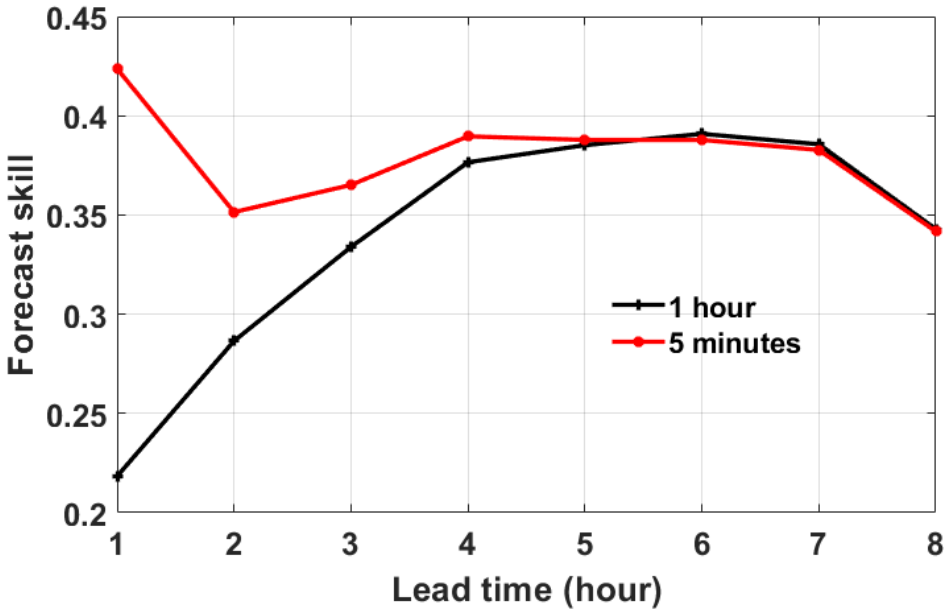
553



554

555 Figure 8. Difference in RMSE and nRMSE between with and without hourly ERA5 reanalysis
556 used. Hourly inputs are used, other configurations are the same as in Figures 2-4. For black line,
557 data in 2010-2013 are used to predict GHIs in 2014. For red line, data in 2011-2014 are used to
558 predict GHIs in 2015.

559



560

561 Figure 9. Forecast skill in 2014 compared to smart persistence model.

562

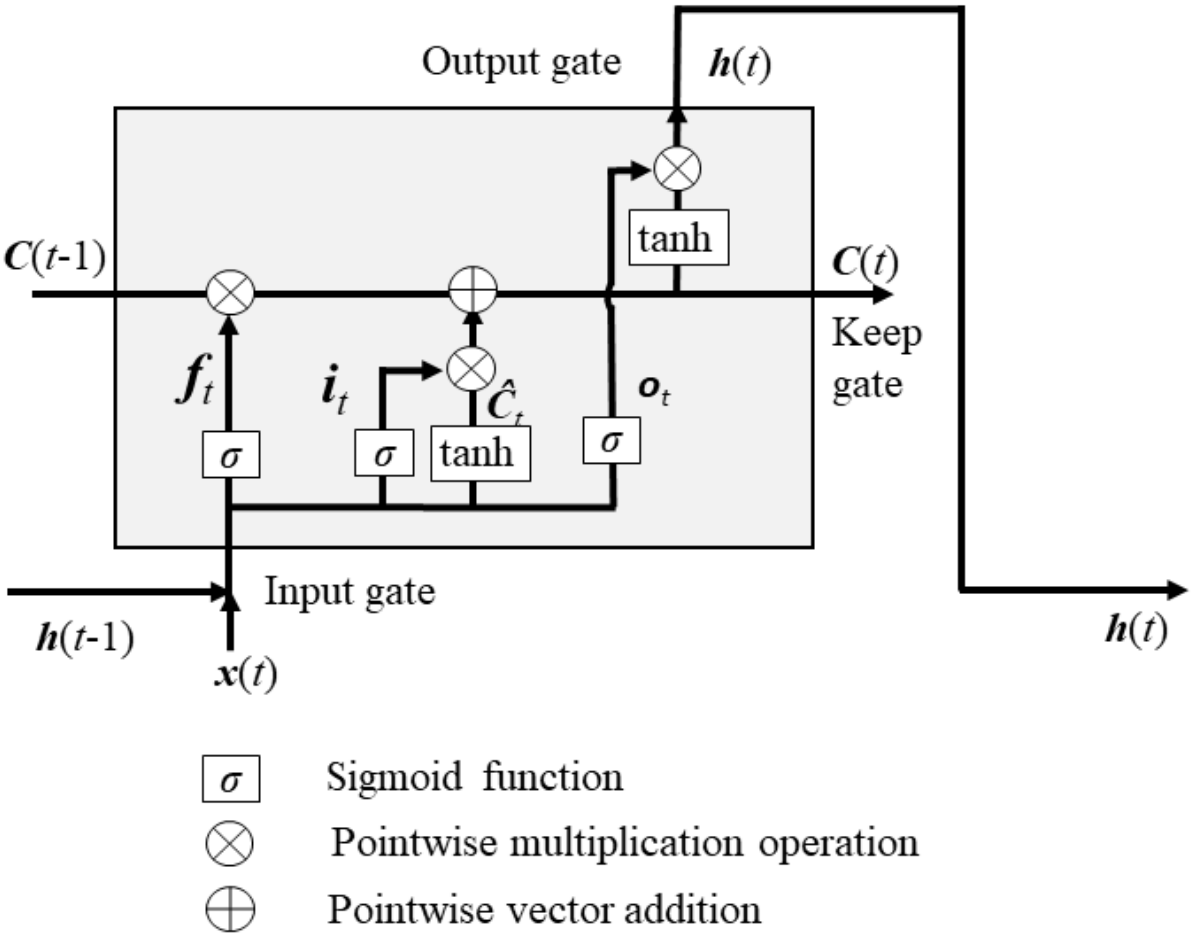


Figure A.1 Sketch of the structure of LSTM neural network cell based on illustration from Olah (2015) and Sun et al. (2017).