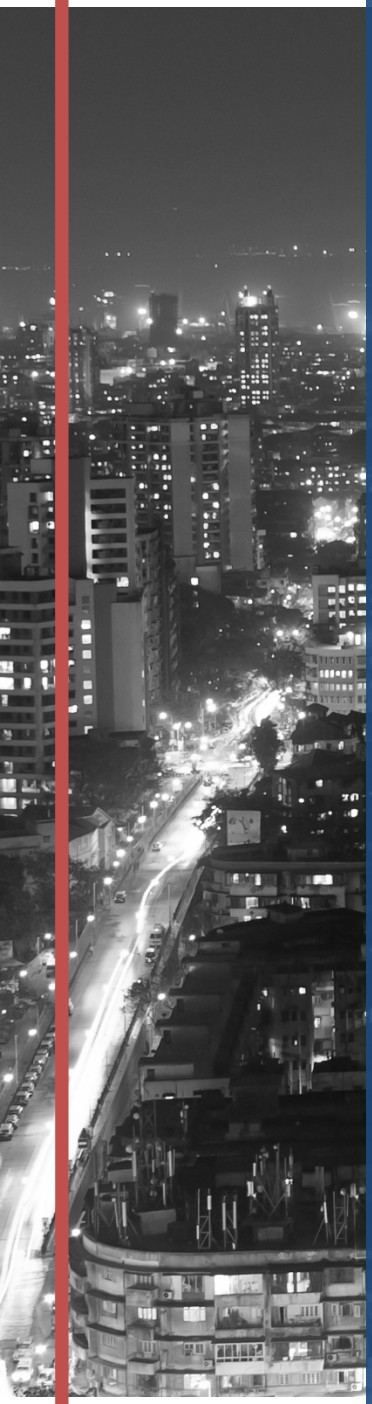

A SHORT-TERM SOLAR FORECASTING PLATFORM USING A PHYSICS-BASED SMART PERSISTENCE MODEL AND DATA IMPUTATION METHOD

Srinath Yelchuri,[1] A. G. Rangaraj,[1] Yu Xie,[2] Aron Habte,[2] Mohit Chandra Joshi,[2] K. Boopathi,[1] Manajit Sengupta,[2] and K. Balaraman[1]

[1]*National Institute of Wind Energy, Ministry of New and Renewable Energy, Government of India*

[2]*National Renewable Energy Laboratory*

December 2021

A product of the South Asia Group for Energy
Contract No. AID-386-T-16-00002

A SHORT-TERM SOLAR FORECASTING PLATFORM USING A PHYSICS-BASED SMART PERSISTENCE MODEL AND DATA IMPUTATION METHOD

Srinath Yelchuri,[1] A. G. Rangaraj,[1] Yu Xie,[2]
Aron Habte,[2] Mohit Chandra Joshi,[2]
K. Boopathi,[1] Manajit Sengupta,[2] and
K. Balaraman[1]

[1]National Institute of Wind Energy,
Ministry of New and Renewable Energy,
Government of India

[2]National Renewable Energy Laboratory

December 2021

A product of the South Asia Group for Energy
Contract No. AID-386-T-16-00002

NOTICE

This work was authored, in part, by the National Renewable Energy Laboratory (NREL), operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by the United States Agency for International Development (USAID) under Contract No. AID-386-T-16-00002. The views expressed in this report do not necessarily represent the views of the DOE or the U.S. Government, or any agency thereof, including USAID.

Acknowledgments

The authors thank Jaemo Yang, Yingchen Zhang, and Ben Kroposki of the National Renewable Energy Laboratory (NREL) and J.C. David Solomon from National Institute of Wind Energy (NIWE), Government of India, for their careful review and comments.

The authors would also like to acknowledge the sincere efforts of the directors, collaborators, group heads, guides, administrative, and other staff of Ministry of New and Renewable Energy (MNRE), Government of India and NIWE for their constant help and guidance throughout the project.

Finally, we are grateful for the editorial support from Isabel McCan and Terri Marshburn of NREL.

List of Acronyms

ARIMA	autoregressive integrated moving average
ARM	Atmospheric Radiation Measurement program network
BSRN	Baseline Surface Radiation Network
CERC	Central Electricity Regulatory Commission
DISC	Direct Insolation Simulation Code
ETR	extraterrestrial (top-of-atmosphere) GHI
GHI	global horizontal irradiance
MAE	mean absolute error
MNRE	Ministry of New and Renewable Energy
NIWE	National Institute of Wind Energy
NREL	National Renewable Energy Laboratory
PSPI	Physics-based Smart Persistence model for Intra-hour forecasting
PV	photovoltaic
SPA	solar position algorithm
SRRA	Solar Radiation Resource Assessment
WMO	World Meteorological Organization

Executive Summary

Electrical energy plays a vital role in our socio-economic activity; therefore, ensuring the reliability of the electric grid is critical—from generation to transmission to distribution. To maintain the power system parameter (i.e., frequency, voltage, and more), optimally, balancing generation and consumption is essential. However, solar energy is variable by nature because of cloud cover and other local phenomena. Hence, photovoltaic (PV) power generation brings a significant challenge to grid operators due to the inherent variability of the energy source. The complexity of this challenge—in terms of planning and dispatch ability of PV resources—aggravates the high-penetration capability of solar energy onto the electric grid. Therefore, to alleviate the solar energy variability and the effect on high-penetration capability, reliable solar radiation forecasting models based on accurate, high-quality input data become essential. To develop a suitable model for predicting solar radiation, high-quality, historical, real-time measurement is also needed.

In this study, researchers from India's National Institute of Wind Energy (NIWE) and the U.S. Department of Energy's National Renewable Energy Laboratory (NREL) jointly developed and tested short-term solar forecasting frameworks using Smart Persistence and the Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI). We also benchmarked nine data imputation techniques in 15 Solar Radiation Resource Assessment (SRRA) stations, located in different parts of India.

For various technical reasons, during any measurement campaign, researchers may miss a few observations. However, these missing observations often reduce the performance of any forecasting model. Therefore, a suitable data imputation method would assist researchers in obtaining continuous observation of solar radiation.

Station-by-station and method-by-method analyses were carried out to understand the performance of each model. Based on our analysis, among all the data imputation methods, the Kalman data imputation method is better for Indian weather conditions. In addition, the Kalman StructTS, Linear, Stine, and Arima methods yield slightly inferior accuracy compared to the Kalman method but outperform other methods.

Extended solar radiation data are used by solar forecasting models to predict solar radiation at 15 SRRA stations. For short-term forecasting, the PSPI model outperforms the Smart Persistence Model. However, the forecast error increases with the forecasting time-scale or horizon.

Table of Contents

List of Figures

List of Tables

1 Introduction

Solar energy systems require reliable solar resource data to predict the performance of a solar power plant. These data sets are prone to uncertainties and incompleteness. To increase the reliability and usability of such data sets for solar energy systems, this study developed a short-term solar forecasting framework using a physics-based solar forecasting model and data imputation methods to fill in missing data points.

Missing and erroneous data may occur for a variety of reasons, including communication and signal issues, sensor problems, equipment malfunctions, maintenance and calibration issues, and data outside of physical meteorological limit. These issues can range from few minutes to days, and they are unavoidable when measuring meteorological parameters, such as the solar resource. However, high-quality temporal resolution solar resource data are critical for various solar energy projects during the design, financing, and operation phases of solar energy systems (Sengupta et al. 2021). Therefore, it is important to assess data quality and estimate any incomplete data set using scientifically and statistically sound imputation methodologies. Currently, there are no consensus methodologies for assessing data quality and data imputation for solar irradiance measurement data. However, there are many methods developed over the years by various researchers. Some research organizations developed data quality assessment models that identify erroneous data, including QCRad, SERI-QC, and the Baseline Surface Radiation Network (BSRN). These were developed by the Atmospheric Radiation Measurement (ARM) program network (Long et al. 2006), the U.S. Department of Energy's National Renewable Energy Laboratory (NREL) (Maxwell et al. 1993), and the BSRN of the World Meteorological Organization (WMO) (Long and Dutton 2002), respectively.

This study first investigates the data imputation methodologies to fill erroneous and missing data. To find optimum data imputation methodologies, we implemented nine methods, such as linear and autoregressive integrated moving average (ARIMA) models, which are described in Denhard et al. (2021). Details of these methods can be found in various reports and publications (Denhard et al. 2021; Ekhosuehi and Dickson 2016; Grewal 2011; Harvey 1990; Johnston et al. 1999; Lyche and Schumaker 1973; Stineman 1980; Welch and Bishop 1995). The data imputation methods in this study will supply extended data for forecasting solar radiation.

Short-term solar radiation forecasting is often conducted using surface-based observations of global horizontal irradiance (GHI) (Kleissl 2013; Yang et al. 2018). Time series analysis or machine learning models are used to capture the features in the observed data to understand the underlying causes (David et al. 2016; Dong et al. 2013; Inman et al. 2013; Kleissl 2013; Lave et al. 2013; Reikard 2009). Predictions are often made with assumptions about the form of the data and the decomposition of the time series into several components, representing various statistical patterns. However, substantially well-investigated physics behind GHI have not been heeded by forecasting models. For example, extraterrestrial solar radiation is primarily dominated by the seasonal variation of the sun-earth distance that can be precisely predicted by physical models (Reda and Andreas 2004). Although it exhibits a strong correlation with the GHI forecast, the prediction of extraterrestrial solar radiation is usually absent in solar forecasting models that predict all unknown factors together in the form of GHI.

To overcome this uncertainty, Kumler et al. (2019) developed the Physics-based Smart Persistence model for Intra-hour forecasting of solar radiation (PSPI) based on a cloud retrieval technique (Xie and Liu 2013; Xie et al. 2014). PSPI deconstructs or decomposes solar forecasting into the computation of solar zenith angle, extraterrestrial solar radiation, and the prediction of cloud properties. The cloud fraction and cloud albedo are estimated using radiative transfer theory and surface-based observations of solar radiation. Their future variability is predicted using an exponentially weighted moving average with the assumption of persistent cloud structures. The predictions are combined to precisely compute solar radiation.

Because PSPI has shown consistently better performance than the persistence model and Smart Persistence model, it is employed in this study to provide short-term solar forecasting. With the advancements of satellite remote sensing, cloud forecasts, and radiative transfer computation (Sengupta et al., 2018; Xie and Sengupta, 2018; Xie et al., 2018; Xie et al., 2016; Xie et al., 2020; Xie et al., 2019), PSPI has the potential to be improved further.

2 Methodology

This Section describes the data imputation methods and short-term solar forecasting models used in this study.

2.1 Data Imputation Methods

As stated earlier, many solar energy project phases require high-quality and continuous solar resource data. Consequently, it is important to impute missing data to improve the modeling. In this study, we used nine methods to test 15 well-maintained SRRA stations with 15 minutes of temporal resolution. Various data imputation methodologies can be applied based on the extent and type of missing data. To find the optimum data imputation methodology, researchers implemented these nine methods, described in Denhard et al.(2021):

- Kalman filtering and smoothing for structural time series fitted by maximum likelihood (Section 3, labeled as Kalman)
- Kalman filtering and smoothing for structural time series with additional parameters (Section 3, labeled as Kalman_StructTS)
- Kalman filtering and smoothing for the state-space representation of an ARIMA model (Section 3, labeled as ARIMA)
- Linear interpolation (Section 3, labeled as Linear)
- Spline interpolation (Section 3, labeled as Spline)
- Stine interpolation (Section 3, labeled as Stine)
- Simple moving average (Section 3, labeled as Simple_mvg_avg)
- Linear weighted moving average (Section 3, labeled as Linear_wted_mvg_avg)
- Exponential weighted moving average (Section 3, labeled as Exp_wted_mvg_avg).

The data imputation methods consider available clearness indices, Kt. Then, for the data gaps, Kt were estimated using the nine data imputation methods. The clearness index of GHI is defined as the ratio, between measured GHI and extraterrestrial (top-of-atmosphere) GHI (ETR).

$$ETR = \mathrm{ETRN} * \cos(\mathrm{Z}) \quad (1)$$

where ETRN = direct extraterrestrial, and Z = solar zenith angle.

Six artificial data gaps were created by following the procedure described in Denhard et al. (2021)—containing singular missing data (denoted by bin1), strings of one to two consecutive missing data (bin2), strings of one to three consecutive missing data (bin3), strings of one to four consecutive missing data (bin4), strings of one to five consecutive missing data (bin5), and strings of one to six consecutive missing data (bin6). The artificial data gaps correspond to the measurement data, except for the original gaps. Consequently, these synthetic gaps were generated by removing corresponding measured data. Creating these six artificial data gap scenarios helps to better evaluate the performance of the nine methods in each gap. The gap-filled data evaluation using various statistical metrics was carried out by applying the actual site measurement data without gaps. The methodology of the gap-fill approach implemented in this study is shown in Figure 1.

Furthermore, the data sets that were filled using the data imputation statistical models, were checked for "physically plausible limits." If the data were outside of these limits, these outliers were filled using a Bird clear sky model (Bird and Hulstrom, 1981), using criteria described in Figure 1.

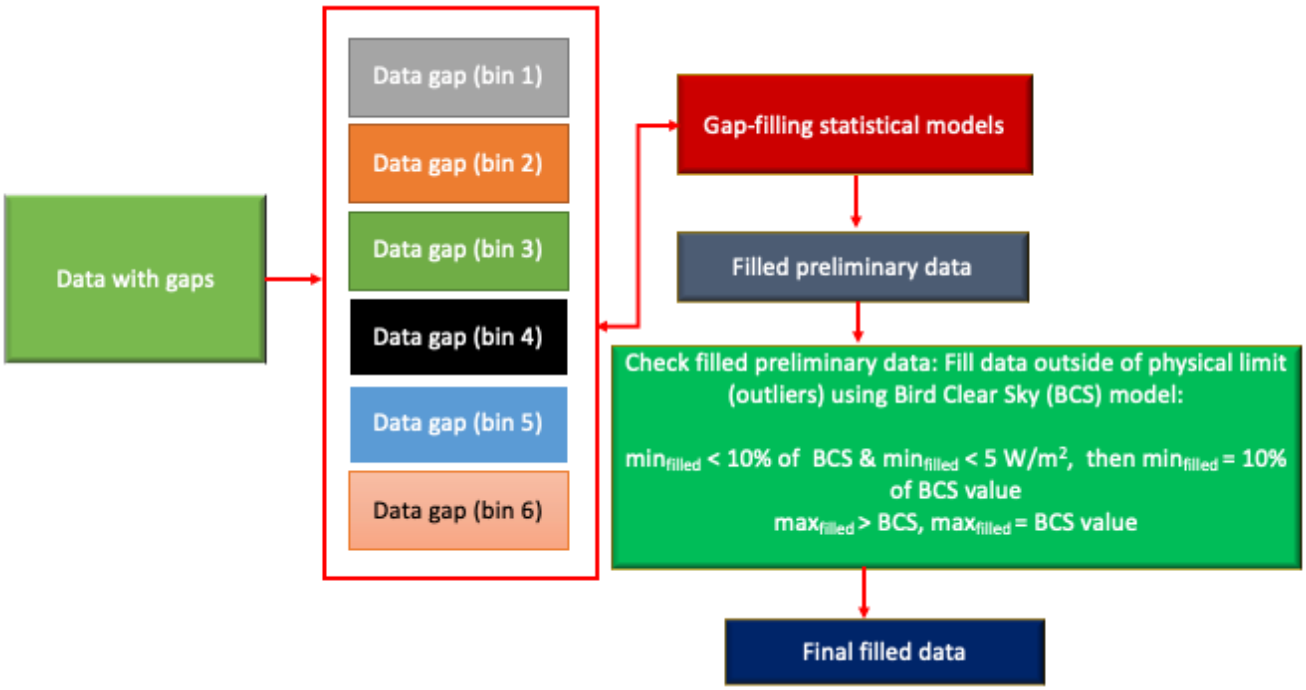

Figure 1 : Flow chart of the data imputation process for solar irradiance data.

2.2 Short-term Forecasting of Solar Radiation

In this study, we predict short-term variation of solar radiation using PSPI (Kumler et al. 2019) and surface-based observations of solar radiation. In contrast to the conventional solar forecasting models using time-series analysis or machine learning techniques, PSPI is based on the simultaneous retrieval of cloud fraction and cloud albedo using the theory of atmospheric radiation (Liu et al. 2021; Xie and Liu 2013; Xie et al. 2014).

Following the simulation of GHI contributed from direct radiation, diffuse radiation from the first-order cloud scattering and that related to multiple scattering between cloud and land surface, the cloud fraction for a single cloud layer atmosphere can be derived as:

$$f = \frac{F_{clr} - F_{all}}{\alpha F_{clr} - \alpha_r F^{\uparrow} T^2} \tag{2a}$$

where F_{clr} is clear-sky solar radiation, F_{all} is the solar radiation for the surface observation, α_r is cloud albedo, α is the total of cloud albedo and cloud absorptance, $F^{\uparrow}$ is the upwelling solar radiation reflected by the land surface, and T is the transmittance of the atmosphere for diffuse radiation. With the observation or simulation of direct solar radiation, cloud fraction can be also derived as:

$$f = \frac{F_{clr,d} - F_{all,d}}{F_{clr,d}(1 - e^{-\tau/\mu_0})} \tag{2b}$$

where $F_{clr,d}$ is the direct solar radiation in the clear-sky condition, $F_{all,d}$ is the direct solar radiation observed as the land surface, τ is the optical thickness of the cloud, and μ_0 represents the consine value of the solar zenith angle. Following a two-stream approximation suggested by Sagan and Pollack (1967), cloud optical thickness can be linked with an expression of cloud albedo, the asymmetry factor of the cloud particles, and solar zenith angle:

$$\tau = \frac{2\alpha_r \mu_0}{(1 - \alpha_r)(1 - g)} \tag{3}$$

where g denotes the asymmetry factor of the cloud particles. Therefore, cloud fraction and cloud albedo are simultaneously solved from Eqs. (2 and 3). More details on solving the equations and determining the required parameters can be found in Xie and Liu (2013).

Following the cloud retrieval technique, PSPI decomposes the forecasting of solar radiation into the computation of solar zenith angle and extraterrestrial solar radiation, the forecasting of cloud fraction and cloud albedo, and the computation of GHI. It first computes the solar zenith angle and extraterrestrial solar radiation from the solar position algorithm (SPA) (Reda and Andreas 2004). It then retrieves the cloud fraction and cloud albedo using the observations of GHI. Direct solar radiation is estimated using the Direct Insolation Simulation Code (DISC) model (Maxwell 1987) and the GHI observations. For short-term forecasting, the shape and structure of the clouds are assumed as persistent; thus, the cloud optical thickness is persistent in the future time steps. The predicted cloud albedo is computed using the cloud optical thickness and the projected solar zenith angle. The cloud fraction is predicted using an exponential weighted moving average given by:

$$f' = \frac{\sum_{i=0}^{t}(1-a)^i f_{t-i}}{\sum_{i=0}^{t}(1-a)^i} \tag{4}$$

where *i* represents the time steps in the observations, f_{t-i} is the retrieved cloud fraction from the observations, and *a* is a constant smoothing factor. The GHIs in future time steps are finally given from the predicted cloud fraction and cloud albedo:

$$F_{all} = \frac{F_1}{1-\alpha_s \alpha_r f T^2} \tag{5}$$

where α_s is surface albedo, and F_1 is the first-order downwelling irradiance that is given by the cloud fraction, cloud albedo, and clear-sky solar irradiance.

According to the discussion from Kumler et al. (2019), PSPI is combined with the Smart Persistence model to achieve better overall performance in all scenarios. For clear-sky conditions and cloud overcast conditions, the Smart Persistence model is used, while PSPI is employed to forecast the GHIs related to broken clouds. Kumler et al. (2019) also demonstrated that PSPI has a consistently better performance than the persistence model and the Smart Persistence model, although an assumption of persistent cloud structures is made. Thus, there is a potential to further improve the solar forecasting model when advanced techniques (e.g., machine learning), are used in the cloud forecasting.

3 Results

3.1 Data Imputation Methods

A data imputation analysis was performed using nine data imputation methods to fill gaps in ground-measured data. A bin-by-bin analysis was conducted in 15 SRRA stations for 2019. The data imputation methods are plotted as subplots from least to greatest error in the order that can be seen in Figures 2–7. In addition, for each method, a line plot displaying the station-wise mean absolute error (MAE), in descending order, has been plotted. In Figure 2, the Stine data imputation method outperformed the other methods, while the simple moving average method underperformed. In Figure 3, the ARIMA method performed better than the other methods, and the maximum error in Bin2 was higher than in Bin1.

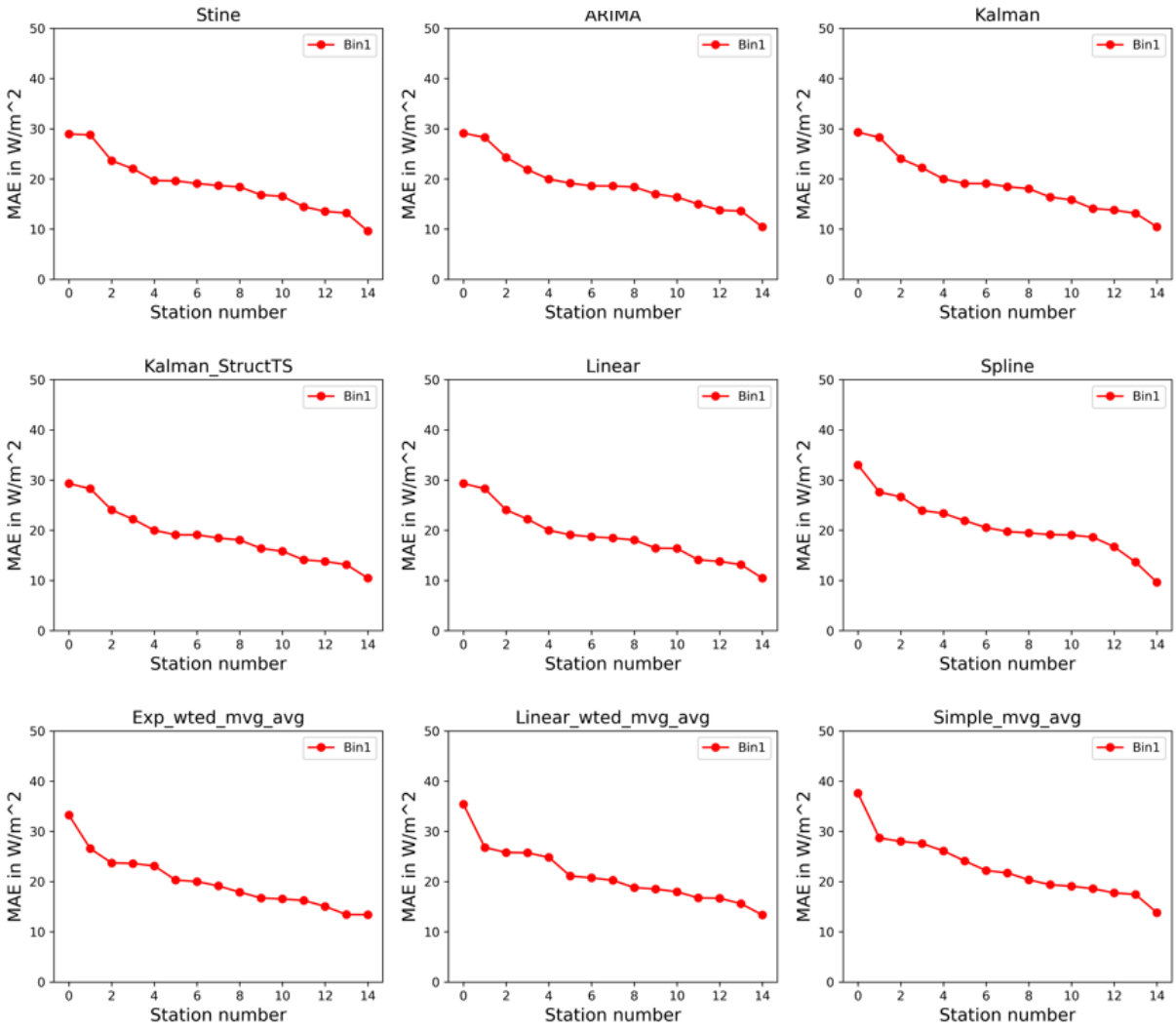

Figure 2 : Result of data imputation methods for Bin1.

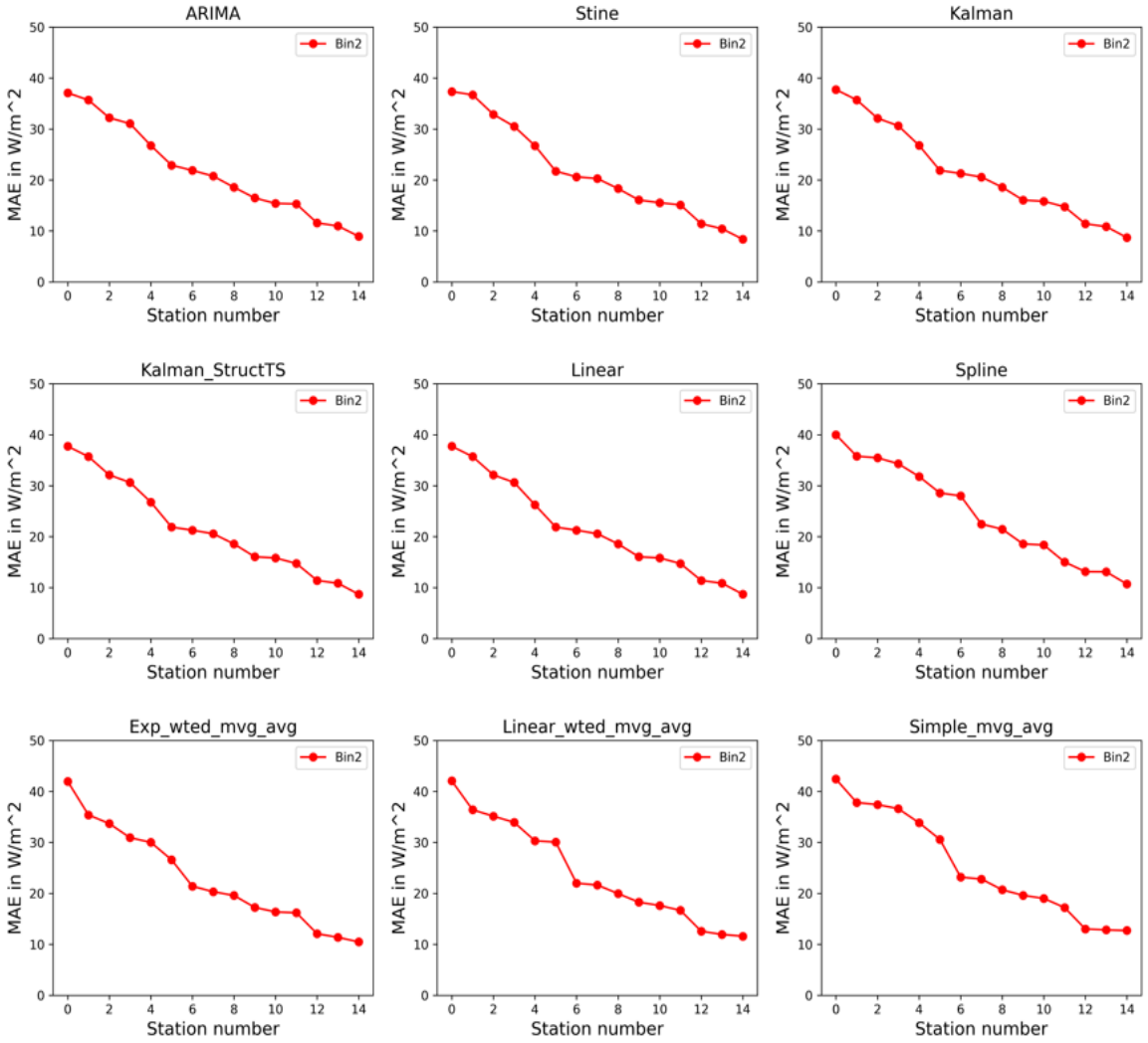

Figure 3 : Result of gap-filling methods for Bin2.

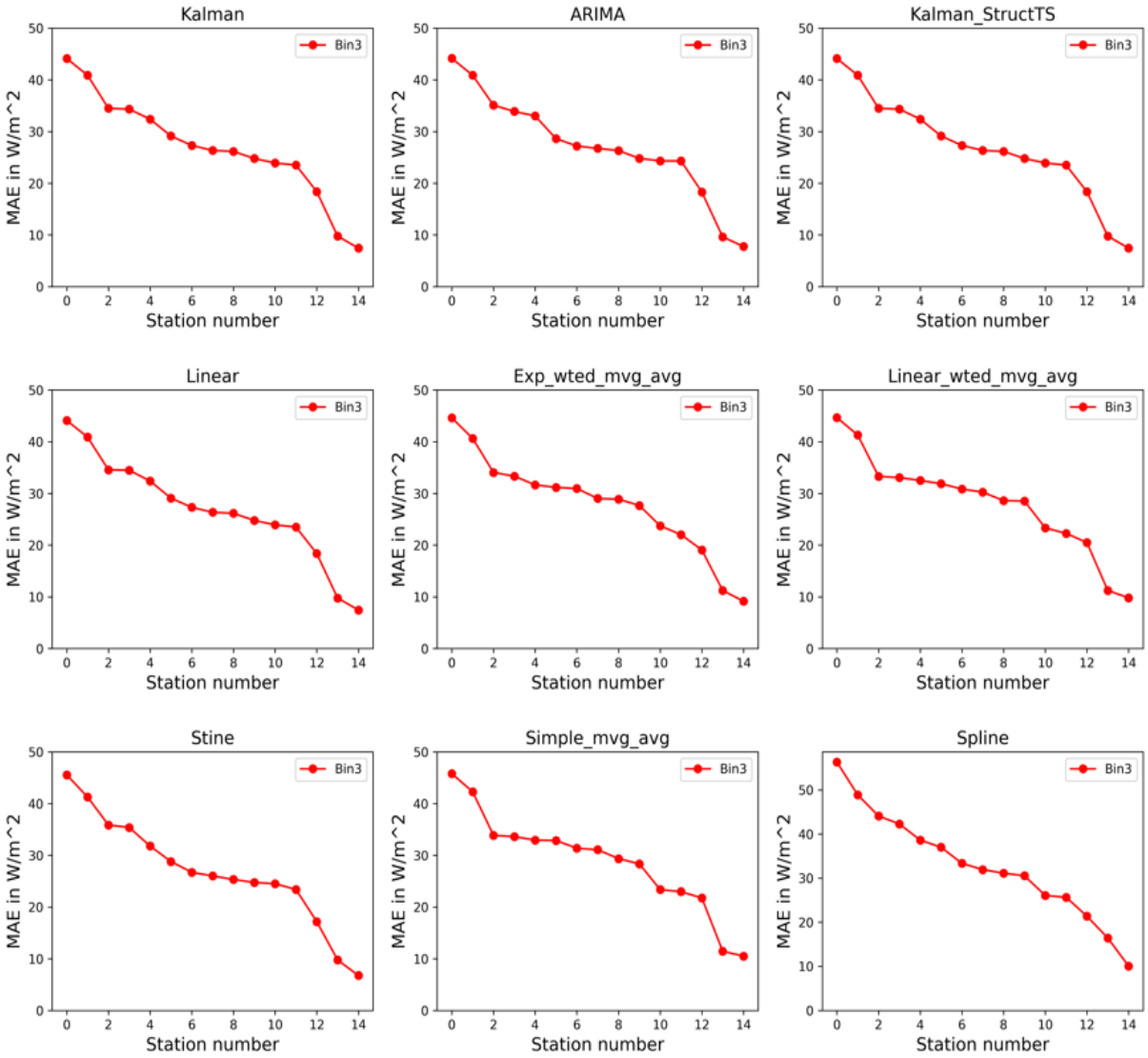

Figure 4 : Result of gap-filling methods for Bin3.

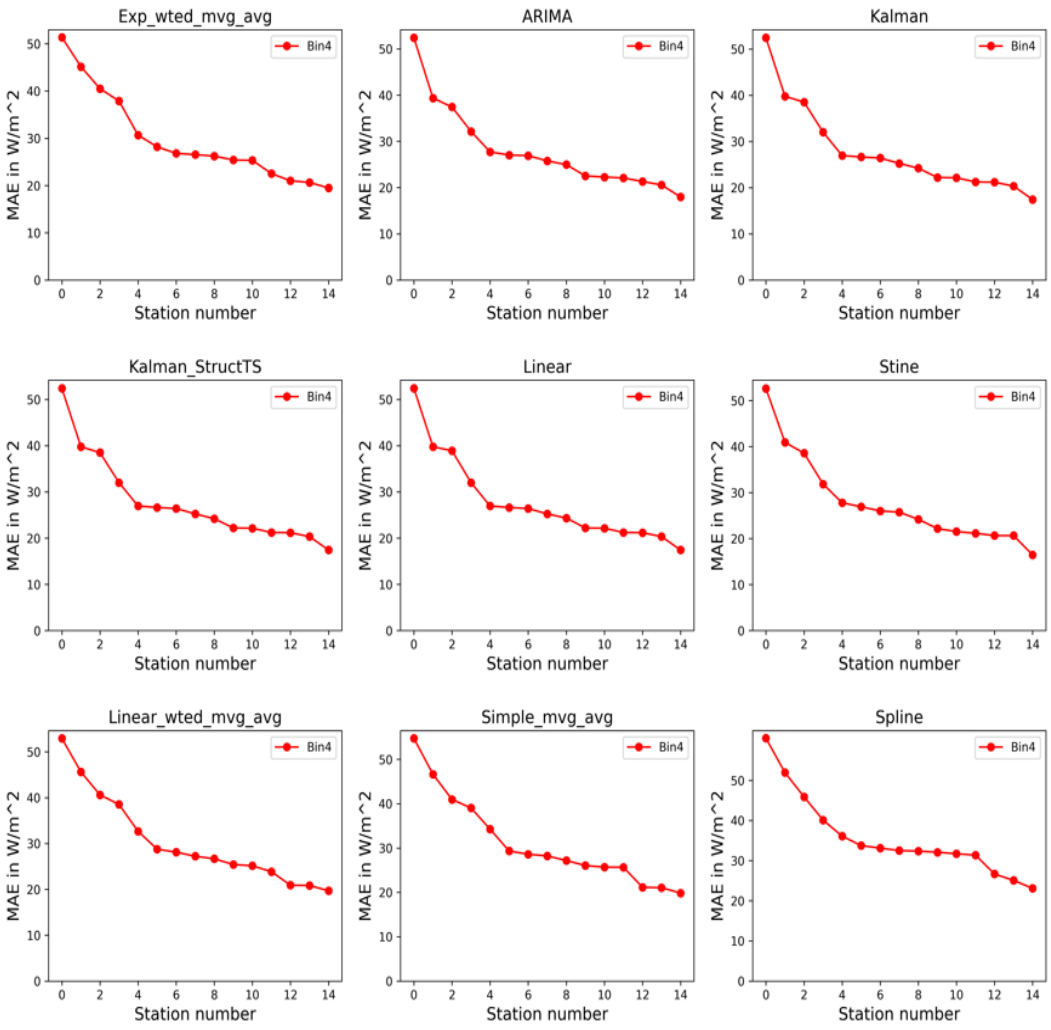

Figure 5 : Result of gap-filling methods for Bin4.

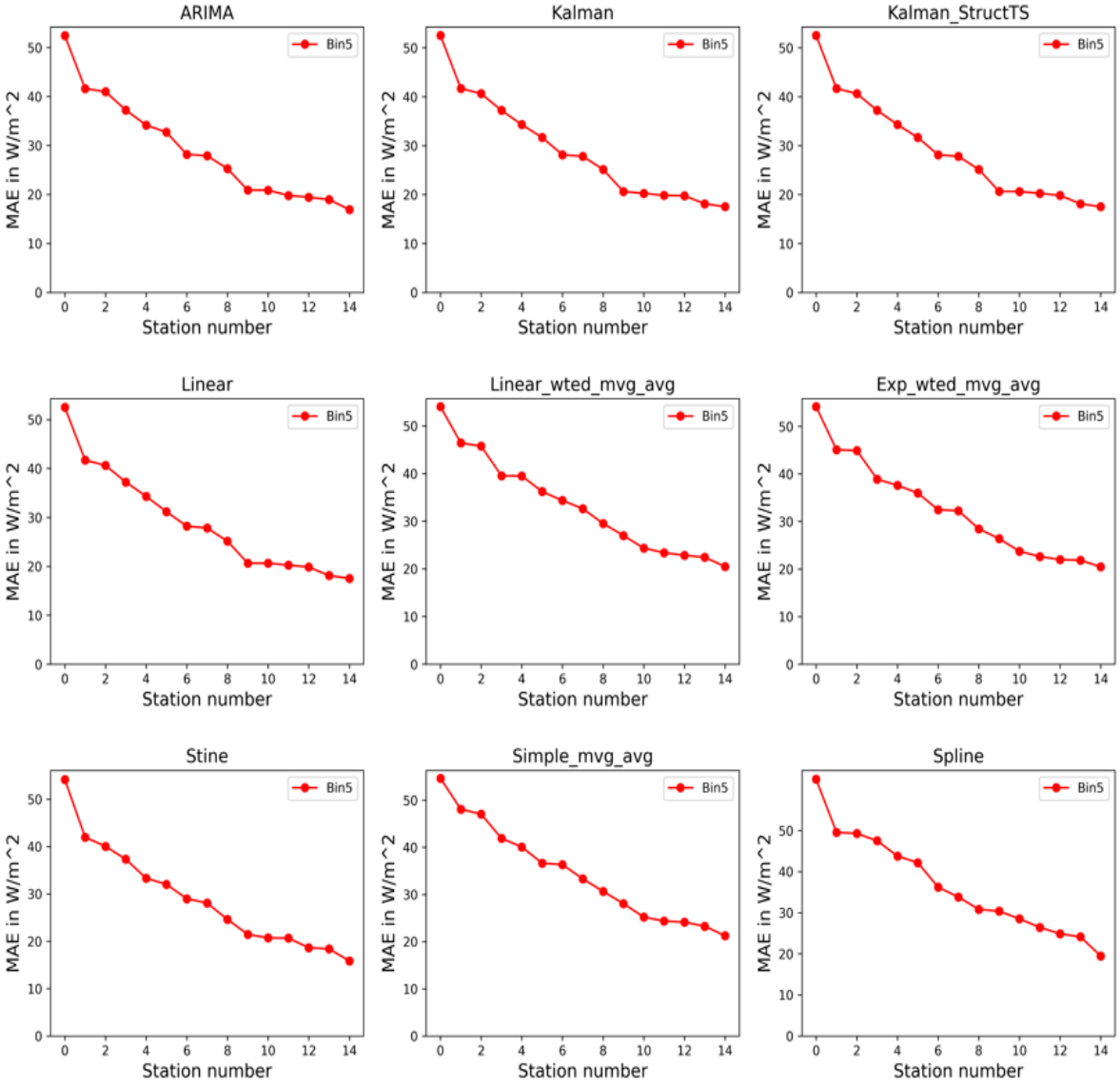

Figure 6 : Result of gap-filling methods for Bin5.

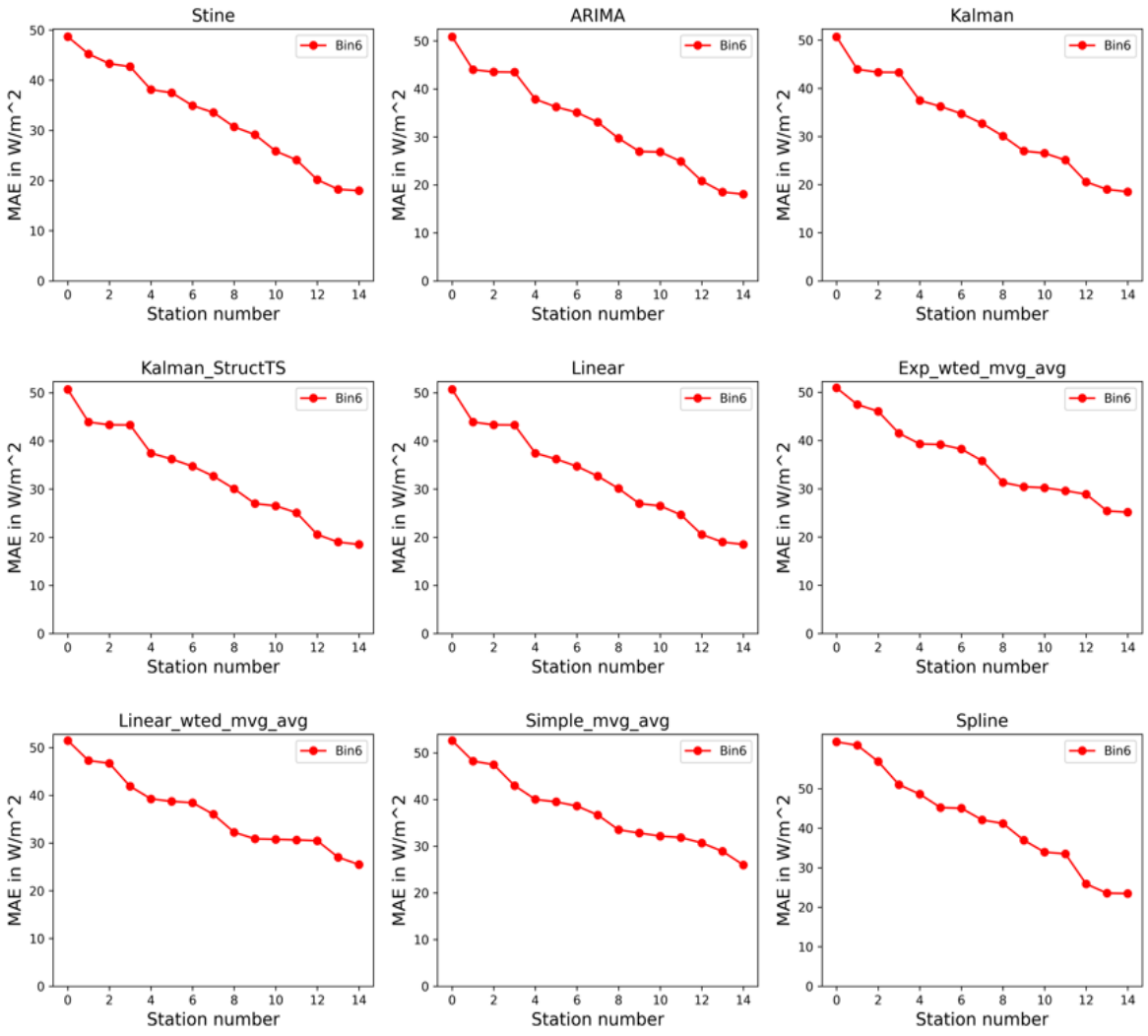

Figure 7 : Result of gap-filling methods for Bin6.

The maximum error increases from Bin1 to Bin6 as the number of consecutive gaps in each bin increases from Figure 2 to Figure 7. The MAE of nine gap-filling methods is also observed to range from 6– 62 W/m^2. In Bin1, Bin2, and Bin6, the Stine method has a lower error. In other bins, the ARIMA, Kalman, and exponential weighted average methods have lower errors. Table 1 displays the method-wise bin-wise average error of all 15 stations.

According to these results, all of the gap-filling methods have good performance. The Kalman gap-filling method, in particular, outperforms the others, while the Kalman_StructTS, Linear, Stine, and ARIMA methods are nearly as good. Figure 8 shows the outcome of gap-filled methods and measured data on a typical day.

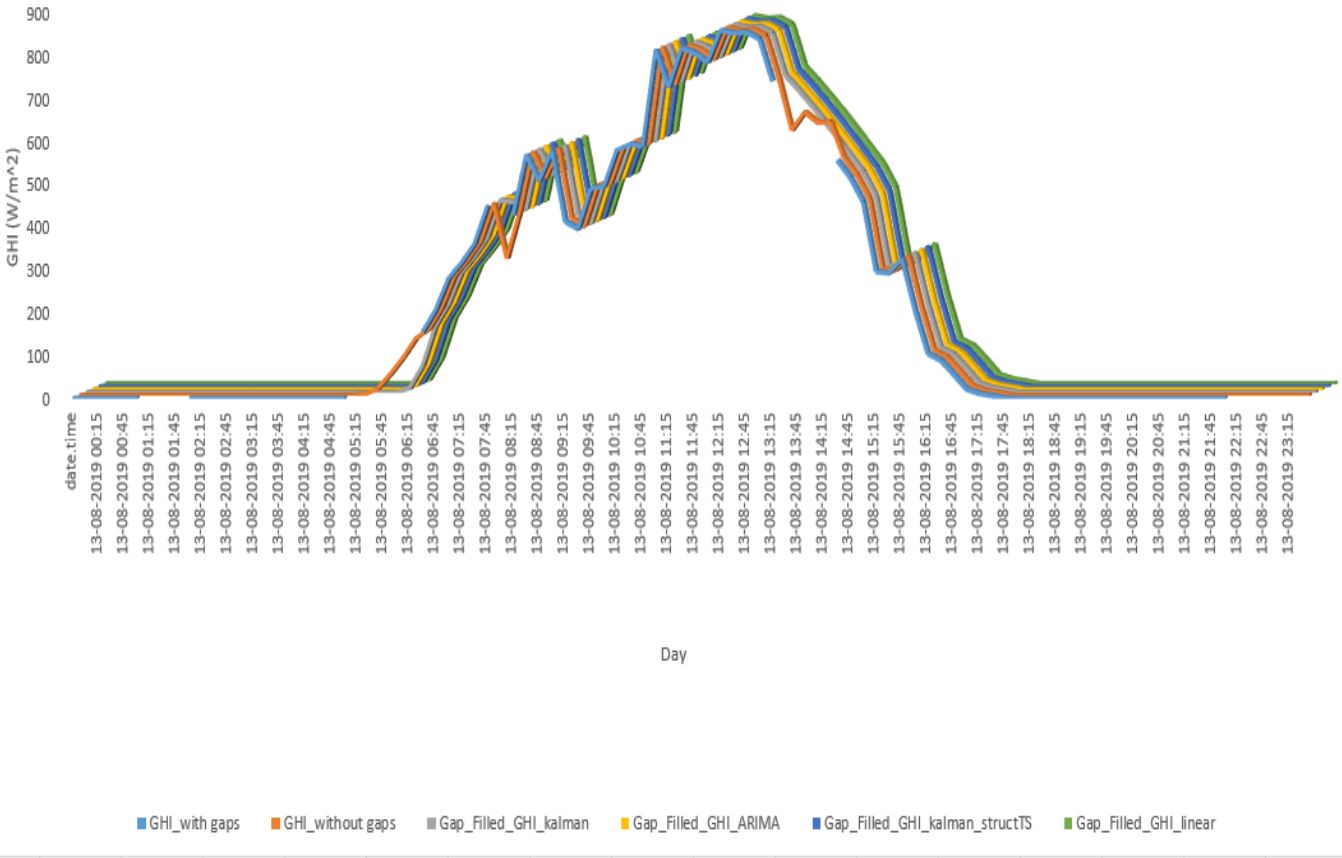

Figure 8 : Result of gap-filling methods on a typical day.

Table 1 : Result of Gap-Filling Methods in Terms of MAE

Method	Bin1	Bin2	Bin3	Bin4	Bin5	Bin6	Overall
Kalman	18.82	21.52	26.86	27.78	29.02	32.61	26.10
Kalman_StructTS	18.82	21.52	26.86	27.78	29.08	32.61	26.11
Linear	18.83	21.49	26.88	27.82	29.05	32.58	26.11
Stine	18.87	21.47	26.87	27.81	29.08	32.68	26.13
ARIMA	18.96	21.70	27.00	28.04	29.16	32.66	26.25
Exp_wted_mvg_avg	19.93	22.91	27.82	29.85	32.46	35.95	28.15
Linear_wted_mvg_avg	21.22	24.01	28.14	30.46	33.24	36.48	28.93
Simple_mvg_avg	22.83	25.32	28.77	31.23	34.33	37.47	29.99
Spline	20.87	24.45	32.92	35.75	36.63	42.01	32.11

Table 2 : Result of Method-wise Ranking in each Station

S. No	Station Name	ARIMA	Exp_wted_ mvg_avg	Kalman	Kalman _structTS	Linear	Linear_wted _mvg_avg	Simple_ mvg_avg	Spline	Stine
1	AbuRoad	2	6	3	4	5	7	9	8	1
2	Amarsagar	5	7	1	3	4	8	9	6	2
3	Bhubaneshwar	5	9	2	3	4	6	7	8	1
4	ChitraDurga	1	6	2	3	4	7	8	9	5
5	Erode	5	6	2	3	4	7	8	9	1
6	Jabalpur	5	6	2	3	4	7	8	9	1
7	Kadiri	6	5	1	2	3	7	8	9	4
8	KotadaPitha	5	6	1	2	3	7	8	9	4
9	Leh	1	6	2	3	4	7	8	9	5
10	Mahabubnagar	5	6	1	2	3	7	8	9	4
11	Pandharpur	5	6	2	3	4	7	8	9	1
12	PortBlair	1	6	3	4	2	7	8	9	5
13	Rajahmundry	5	6	2	3	4	8	9	7	1
14	Shegaon	5	6	2	3	4	7	8	9	1
15	Trichy	5	6	2	3	4	7	8	9	1

Table 2 shows the ranking of gap-filling methods for all stations. The ranking number is shown in ascending order from least to greatest error. It is evident from Table 2 that the Stine method in eight stations, the Kalman method in four stations, and the ARIMA method in three stations performed the best. In nine stations, the Kalman method is the second-best method; in 10 stations, the Kalman strucTS method is third best. It was also observed that Linear, Stine, Spline ARIMA performed almost equally good.

Table 3 : Result of Station-wise Ranking of all Gap-Filling Methods for Bin6

S. No	Station Name	ARIMA	Exp_wted_ mvg_avg	Kalman	Kalman _structTS	Linear	Linear_wted _mvg_avg	Simple_ mvg_avg	Spline	Stine
1	AbuRoad	2	1	1	1	1	1	1	1	1
2	Amarsagar	1	3	3	3	3	3	3	3	3
3	Bhubaneshwar	3	4	4	4	4	4	4	4	4
4	ChitraDurga	4	11	11	11	11	11	11	11	11
5	Erode	6	10	10	10	10	10	10	10	10
6	Jabalpur	5	6	6	6	6	6	6	6	6
7	Kadiri	7	13	13	13	13	13	13	13	13
8	KotadaPitha	8	8	8	8	8	8	8	8	8
9	Leh	9	14	14	14	14	14	14	14	14
10	Mahabubnagar	11	12	12	12	12	12	12	12	12
11	Pandharpur	10	9	9	9	9	9	9	9	9
12	PortBlair	13	7	7	7	7	7	7	7	7
13	Rajahmundry	12	5	5	5	5	5	5	5	5
14	Shegaon	14	2	2	2	2	2	2	2	2
15	Trichy	15	15	15	15	15	15	15	15	15

Table 3 shows the station-wise ranking of all data imputation methods for Bin6. Because we're interested in larger bin gaps, the data imputation methods have been evaluated in Bin6. In each method, we ranked the station performance. The ranking number is shown in ascending order from least to greatest error. It is evident from Table 3, in Bin6, that all data imputation methods—except ARIMA—performed similarly in all stations. From the selected stations, AbuRoad has the least error, and Trichy has the highest error.

According to these results, the Exp wted mvg avg, Linear wted mvg avg, Simple mvg avg, and Spline methods performed poorly in all stations when compared to the remaining methods.

3.2 Solar Forecasting

A study was conducted to evaluate the performance of the PSPI and Smart Persistence forecasting models in 15 SRRA stations for 2019. Figures 9–14 depict the models' performance over various forecast horizons. The hourly performance of the forecasting models was plotted as subplots in each figure, and a line graph for station versus error metrics (MAE) was plotted in each subplot. The figures show that the PSPI model outperforms the smart persistence model and the error of radiation forecast increases as the forecast horizon increases.

According to Central Electricity Regulatory Commission (CERC) of India the guidelines were, intraday revisions are permitted every 1.5 hours, and schedule revisions must be submitted 1 hour before the intraday revision. As a result, in this analysis, we are interested in forecasting up to 2.5 hours ahead (150 mins). Table 4 shows the average error for each of the 15 SRRA stations.

Table 4 : Result of Solar Radiation Forecast in Terms of MAE

	15 mins ahead		60 mins ahead		150 mins ahead	
Hour	PSPI	Smart Persistence	PSPI	Smart Persistence	PSPI	Smart Persistence
7	47.80	43.85				
8	52.31	54.92	77.43	129.27		
9	70.44	70.58	93.78	108.35	139.27	246.88
10	88.24	86.64	112.35	119.54	144.14	204.55
11	104.40	101.47	127.85	132.52	161.04	179.11
12	112.78	110.20	140.84	144.44	173.23	179.22
13	105.30	103.23	139.24	141.71	172.95	176.61
14	87.99	86.47	124.34	127.15	165.92	170.26
15	62.95	63.42	100.60	104.57	142.09	149.59
16	38.74	40.05	73.37	77.24	107.95	118.47
17	19.69	20.51	41.58	45.43	63.69	72.60
18	9.82	9.70	16.03	18.99	23.63	29.75
Average	66.70	65.92	95.22	104.47	130.10	153.52

The PSPI and Smart Persistence models outperformed the other forecast horizons in the 15-min forecast horizon. As shown in Table 4, the forecasting model's error increases as the forecasting horizon lengthens.

In all forecast horizons, the error of the PSPI model is higher in the 12^{th} and 13^{th} time horizons.

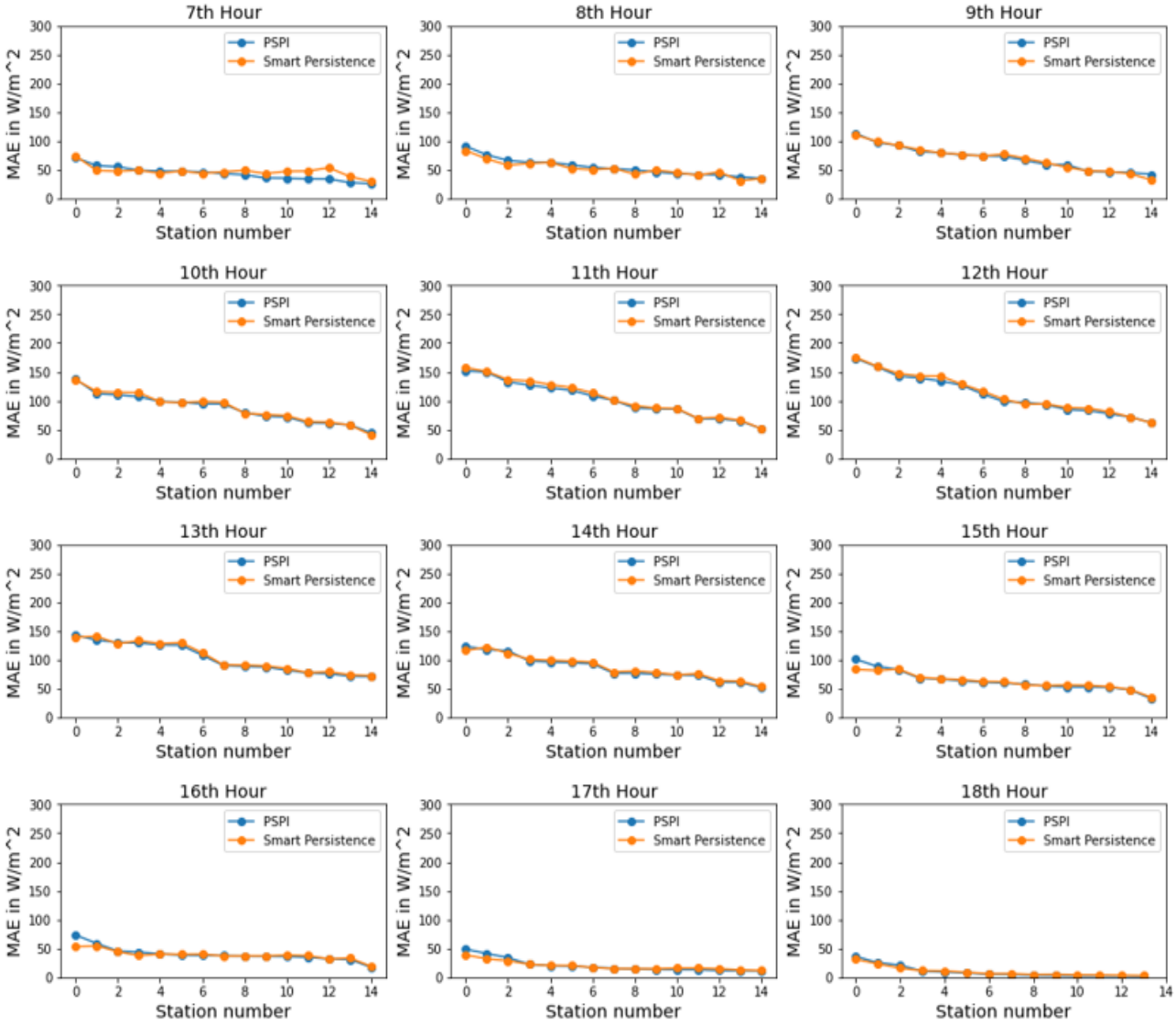

Figure 9 : Result of 15-minute-ahead solar radiation forecast.

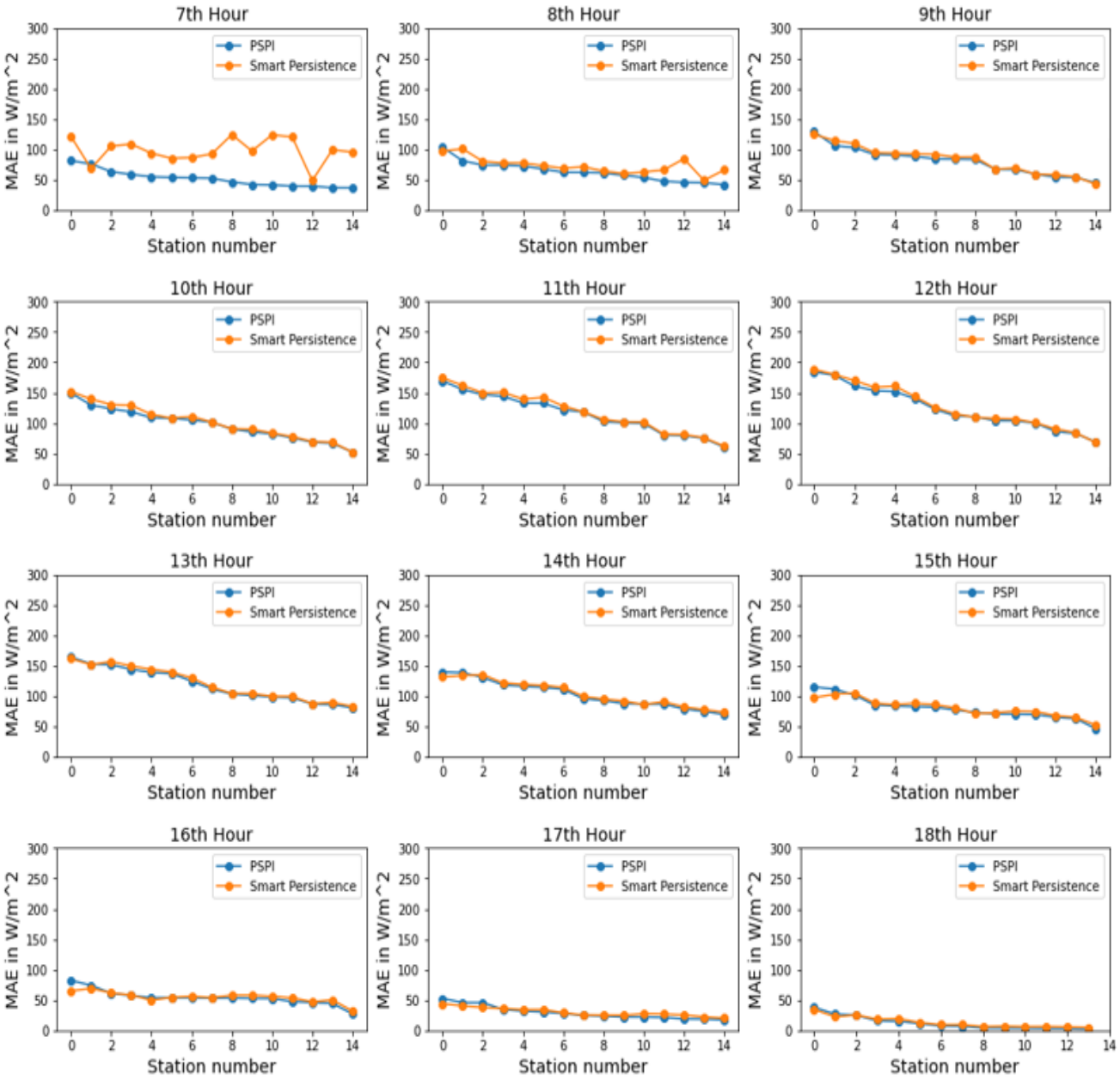

Figure 10 : Result of 30-minute-ahead solar radiation forecast.

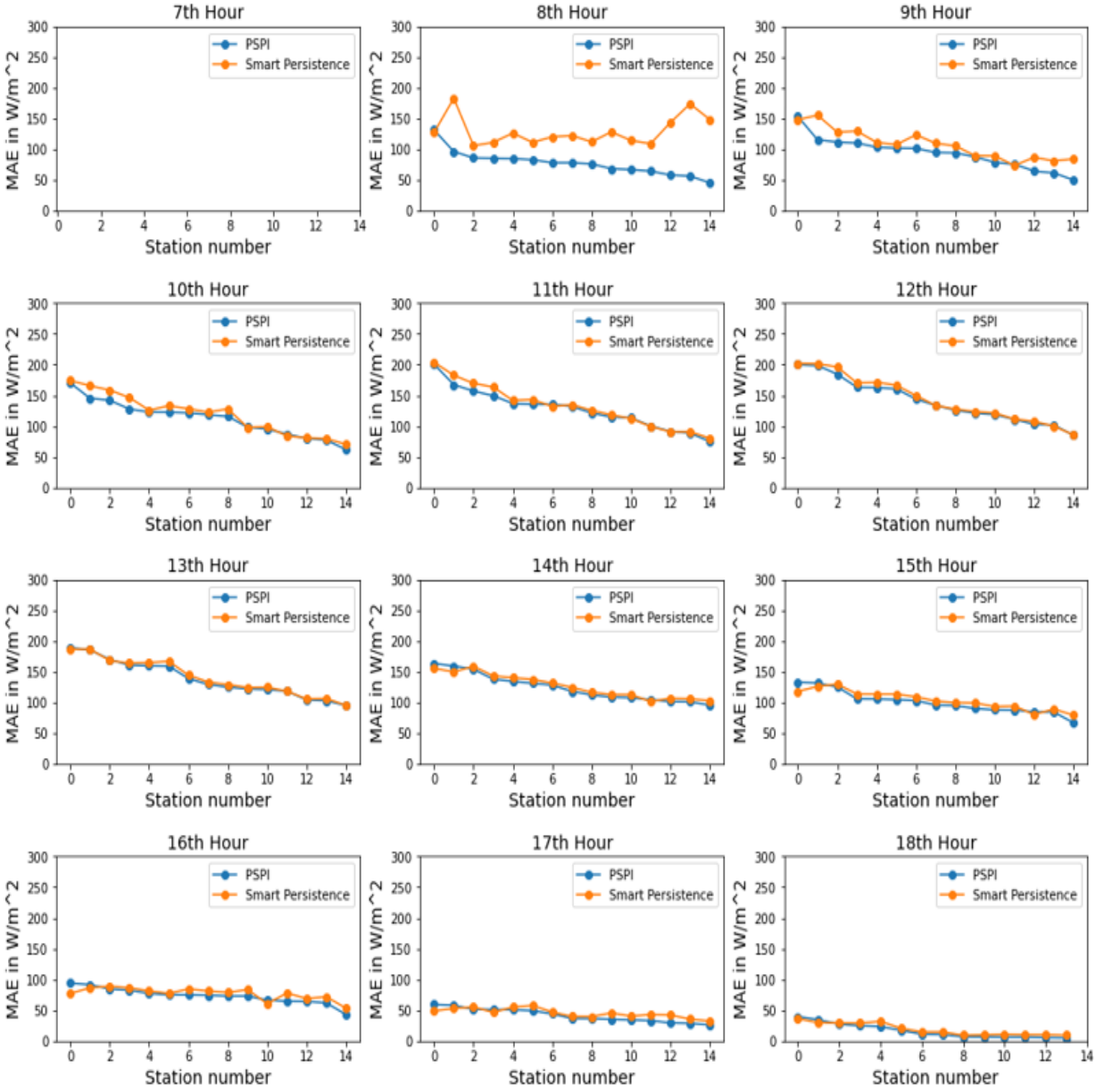

Figure 11 : Result of 60-minute-ahead solar radiation forecast.

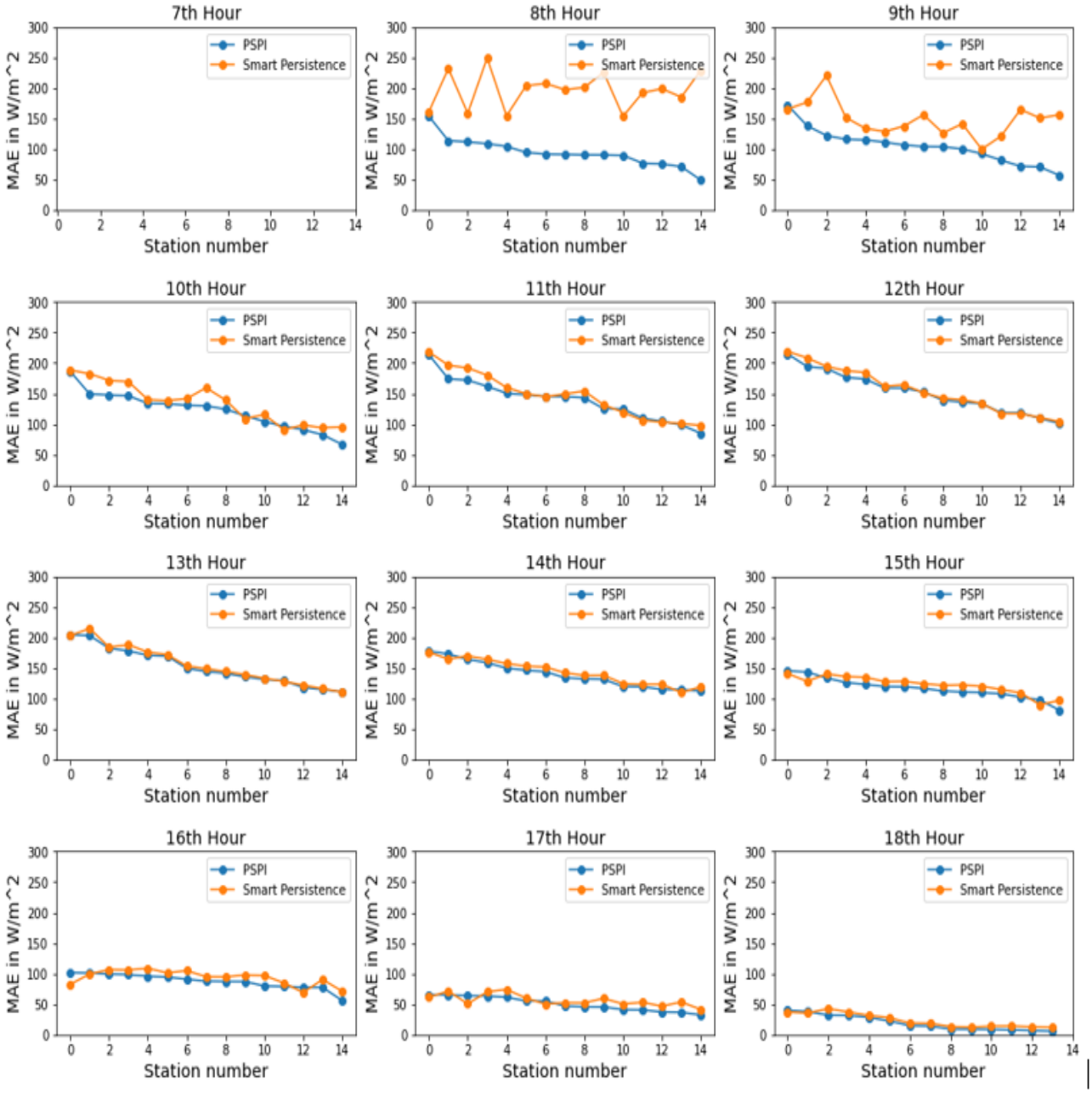

Figure 12 : Result of 90-minute-ahead solar radiation forecast.

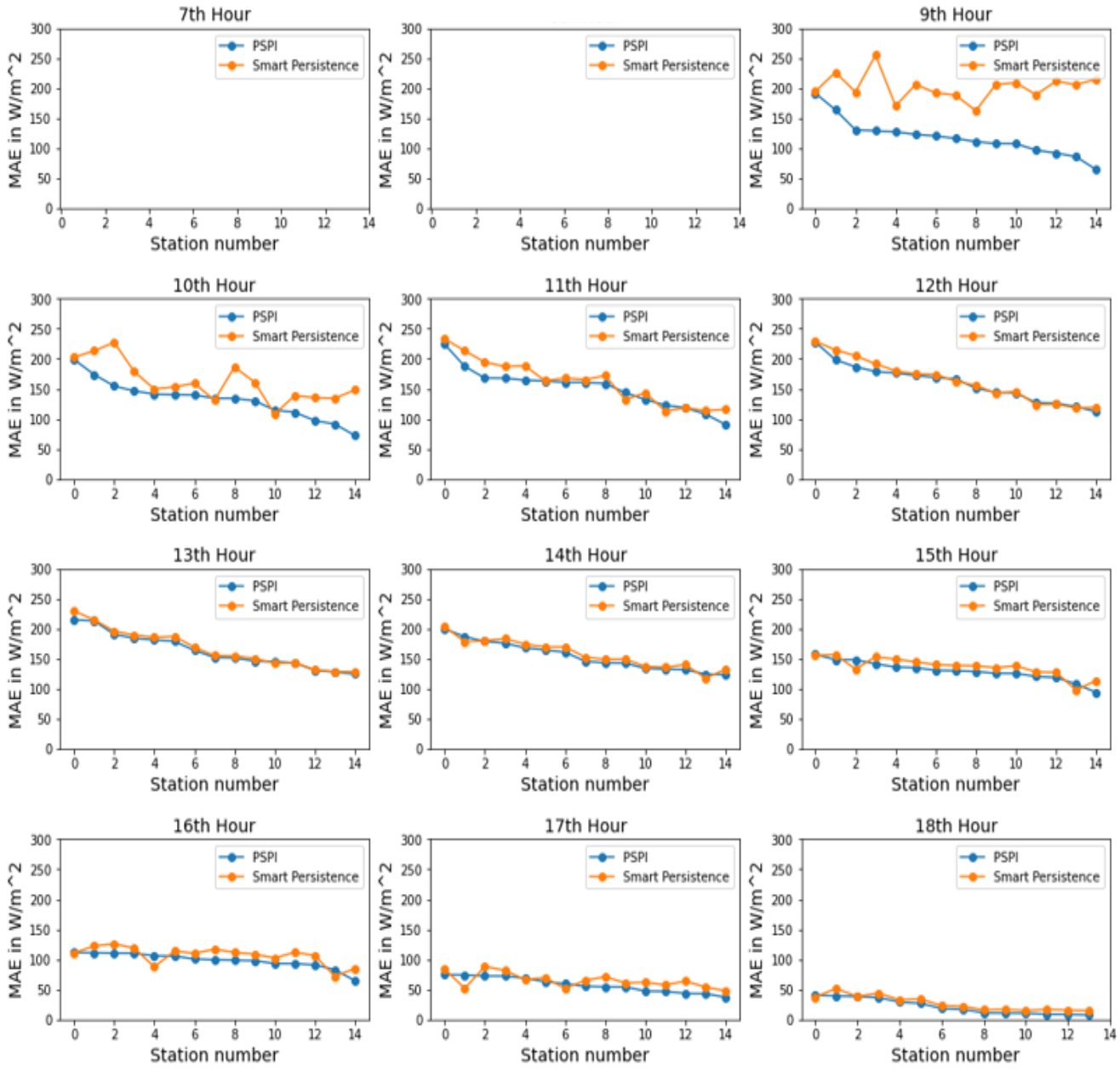

Figure 13 : Result of 120-minute-ahead solar radiation forecast.

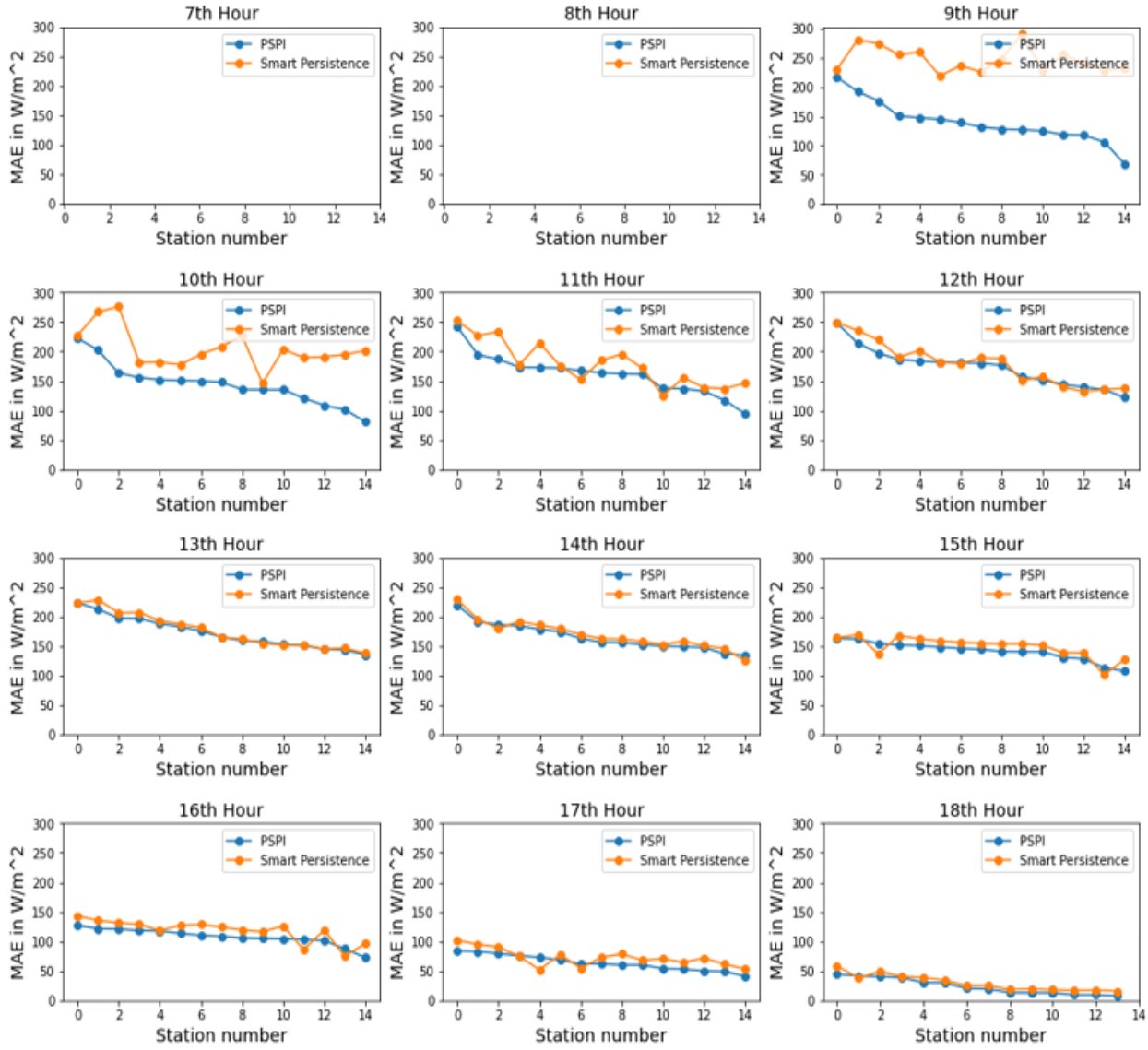

Figure 14 : Result of 150-minute-ahead solar radiation forecast.

When compared to other time horizons, the error of the Smart Persistence radiation forecast model is higher in the early hours. The average error of the PSPI and Smart Persistence models in the forecast horizons shown is between 65 and 153 W/m^2.

4 Conclusions

A short-term solar forecasting platform was developed using a physics-based solar forecasting model and data imputation methods, which was tested using data from 15 solar radiation stations from the SRRA network. Nine data imputation methods were validated at all stations. In a method-by-method comparison, the Kalman method outperformed other methods. In a station-by-station analysis, the Stine method performed best in eight stations. Overall, the Kalman method outperforms others. The Kalman strucTS, Linear, Stine, and ARIMA methods all performed equally well. After data imputation, it is now possible to produce continuous solar forecasts with the extended observations. According to the results and analysis, the PSPI forecast method outperforms the Smart Persistence model at all stations. The forecast model's error increases as the forecast horizons lengthen. The maximum error occurs between the 12^{th} and 13^{th} hours.

5 References

Bird, R. E., and R. L. Hulstrom. 1981. "A Simplified Clear-Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces." Solar Energy Research Institute (SERI). SERI/TR-642-761. Golden, CO. https://www.nrel.gov/grid/solar-resource/assets/data/tr-642-761.pdf

David, M., F. Ramahatana, and P. Trombe. 2016. "Probabilistic Forecasting of the Solar Irradiance with Recursive ARMA and GARCH Models." *Sol. Energy* 133: 55–72.

Denhard, L., S. Bandyopadhyay, A. Habte, and M. Sengupta. 2021. "Evaluation of Time-Series Gap-Filling Methods for Solar Irradiance Applications." National Renewable Energy Laboratory (NREL) technical report. Forthcoming.

Dong, Z., D. Yang, T. Reindl, and W. Walsh. 2013. "Short-term Solar Irradiance Forecasting Using Exponential Smoothing State Space Model." *Energy* 55: 1104–1113.

Ekhosuehi, N., and E. Dickson. 2016. "On Forecast Performance Using a Class of Weighted Moving Average Processes for Time series." *Journal of Natural Sciences Research* 6 (13). Accessed March 28, 2021. https://iiste.org/Journals/index.php/JNSR/article/view/31793.

Grewal, M. S. 2011. "Kalman Filtering." In *International Encyclopedia of Statistical Science*, edited by M. Lovric. Heidelberg: Springer. Berlin, Germany. https://doi.org/10.1007/978-3-642-04898-2_321.

Harvey, A. 1990. *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781107049994.

Inman, R., H. Pedro, and C. Coimbra. 2013. "Solar Forecasting Methods for Renewable Energy Integration." *Prog. Energ. Combust* 39: 535–567.

Johnston, F., J. Boyland, M. Meadows, and E. Shale. 1999. "Some Properties of a Simple Moving Average When Applied to Forecasting a Time Series." *Journal of the Operational Research Society* 50: 1,267–71. https://doi.org/10.1057/palgrave.jors.2600823.

Kleissl, J., 2013. Solar Energy Forecasting and Resource Assessment. Academic Press.

Kumler, A., Y. Xie, Y. Zhang. 2019. "A Physics-based Smart Persistence Model for Intra-hour Forecasting of Solar Radiation (PSPI) Using GHI Measurements and a Cloud Retrieval Technique." *Sol. Energy* 177: 494–500.

Lave, M., J. Kleissl, and J. Stein. 2013. A Wavelet-based Variability Model (WVM) for Solar PV Power Plants. *IEEE T. Sustain. Energ.* 4: 501–509.

Liu, W., Y. Liu, X. Zhou, Y. Xie, Y. Han, S. Yoo, and M. Sengupta. 2021. "Use of Physics To Improve Solar Forecast: Physics-informed Persistence Models for Simultaneously Forecasting GHI, DNI, and DHI." *Sol. Energy* 215: 252–265.

Long, C. N., and E. G. Dutton. 2002. "BSRN Global Network Recommended QC tests, V2.0." http://hdl.handle.net/10013/epic.38770.d001.

Long, C. N., and Y. Shi. 2006. *The QCRad Value Added Product: Surface Radiation Measurement Quality Control Testing, Including Climatology Configurable Limits*. United States: N. p. https://www.osti.gov/servlets/purl/1019540.

Lyche, T., and L. L. Schumaker. 1973. "On the Convergence of Cubic Interpolating Splines." In Spline Functions and Approximation Theory, edited by A. Meir and A. Sharma, 169–189.

Maxwell, E. 1987. "A Quasi-physical Model for Converting Hourly Global Horizontal to Direct Normal Insolation." Solar Energy Research Institute (SERI). Golden, CO. https://www.nrel.gov/docs/legosti/old/3087.pdf

Maxwell, E., S. Wilcox, and M. Rymes. 1993. "Users Manual for SERI_QC Software: Assessing the Quality of Solar Radiation Data." National Renewable Energy Laboratory (NREL). NREL/TP-463-5608. Golden, CO. https://www.nrel.gov/docs/legosti/old/5608.pdf.

Reda, I., and A. Andreas. 2004. "Solar Position Algorithm for Solar Radiation Applications." *Sol. Energy* 76: 577–589.

Reikard, G. 2009. "Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts." *Sol. Energy* 83: 342–349.

Sagan, C., and J. B. Pollack. 1967. "Anisotropic Nonconservative Scattering and the Clouds of Venus." *J. Geophys. Res.* 72: 469–477.

Sengupta, M., A.Habte, S. Wilbert, C. Gueymard, and J. Remund. 2021. "Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications: Third Edition." Golden, CO: National Renewable Energy Laboratory. NREL/TP-5D00-77635. https://www.nrel.gov/docs/fy21osti/77635.pdf.

Sengupta, M., Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby. 2018. "The National Solar Radiation Data Base (NSRDB)." *Renew. Sustain. Energy Rev.* 89: 51–60.

Stineman, R. W. 1980. "A Consistently Well-Behaved Method of Interpolation." *Creative Computing* 6 (7): 54–57.

Welch, G., and G. Bishop. 1995. *An Introduction to the Kalman Filter* (SIGGRAPH 2001, Course 8). Chapel Hill, NC: ACM, Inc.

Xie, Y., and Y. G. Liu. 2013. "A New Approach for Simultaneously Retrieving Cloud Albedo and Cloud Fraction from Surface-based Shortwave Radiation Measurements." *Environ. Res. Lett.* 8. doi:10.1088/1748-9326/1088/1084/044023.

Xie, Y., Y. G. Liu, C. N. Long, and Q. L. Min. 2014. "Retrievals of Cloud Fraction and Cloud Albedo from Surface-based Shortwave Radiation Measurements: A Comparison of 16 Year Measurements." *J. Geophys. Res. Atmos.* 119(14): 8925–8940.

Xie, Y., and M. Sengupta. 2018. "A Fast All-sky Radiation Model for Solar Applications with Narrowband Irradiances on Tilted surfaces (FARMS-NIT): Part I. The Clear-sky Model." *Sol. Energy* 174: 691–702.

Xie, Y., M. Sengupta, and M. Dooraghi. 2018. "Assessment of Uncertainty in the Numerical Simulation of Solar Irradiance Over Inclined PV Panels: New Algorithms Using Measurements and Modeling Tools." *Sol. Energy* 165: 55–64.

Xie, Y., M. Sengupta, and J. Dudhia. 2016. "A Fast All-sky Radiation Model for Solar applications (FARMS): Algorithm and Performance Evaluation." *Sol. Energy* 135: 435–445.

Xie, Y., M. Sengupta, Y. Liu, H. Long, Q. Min, W. Liu, and A. Habte. 2020. "A Physics-based DNI Model Assessing All-sky Circumsolar Radiation." *iScience* 22. doi.org/10.1016/j.isci.2020.100893.

Xie, Y., M. Sengupta, and C. Wang. 2019. A Fast All-sky Radiation Model for Solar Applications with Narrowband Irradiances on Tilted surfaces (FARMS-NIT): Part II. The cloudy-Sky Model." *Sol. Energy* 188: 799–812.

Yang, D., J. Kleissl, C. Gueymard, H. Pedro, and C. Coimbra. 2018. "History and Trends in Solar Irradiance and PV Power Forecasting: A Preliminary Assessment and Review Using Text Mining." *Sol. Energy* 168: 60–101.

Monali Hazra
U.S. Agency for International Development
Email: mhazra@usaid.gov

David Palchak
National Renewable Energy Laboratory
Email: david.palchak@nrel.gov

Meredydd Evans
Pacific Northwest National Laboratory
Email: m.evans@pnnl.gov

Shruti Deorah
Lawrence Berkeley National Laboratory
Email: smdeorah@lbl.gov

NREL/TP-5D00-81421 | December 2021

The South Asia Group for Energy (SAGE) is a consortium comprising USAID, the U.S. Department of Energy and three national laboratories: the Lawrence Berkeley National Laboratory (LBNL), the National Renewable Energy Laboratory (NREL), and the Pacific Northwest National Laboratory (PNNL). The consortium represents excellence in research and international development in the energy sector to advance the Asia Enhancing Development and Growth through Energy (Asia EDGE) priorities in the South Asia region.