



# Evaluation of Time-Series Gap-Filling Methods for Solar Irradiance Applications

Alexis Denhard,<sup>1</sup> Soutir Bandyopadhyay,<sup>1</sup> Aron Habte,<sup>2</sup> and Manajit Sengupta<sup>2</sup>

*1 Colorado School of Mines*

*2 National Renewable Energy Laboratory*

**NREL is a national laboratory of the U.S. Department of Energy  
Office of Energy Efficiency & Renewable Energy  
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at [www.nrel.gov/publications](http://www.nrel.gov/publications).

Contract No. DE-AC36-08GO28308

**Technical Report**  
NREL/TP-5D00-79987  
October 2021



# Evaluation of Time-Series Gap-Filling Methods for Solar Irradiance Applications

Alexis Denhard,<sup>1</sup> Soutir Bandyopadhyay,<sup>1</sup> Aron Habte,<sup>2</sup> and Manajit Sengupta<sup>2</sup>

*1 Colorado School of Mines*

*2 National Renewable Energy Laboratory*

## **Suggested Citation**

Denhard, Alexis, Soutir Bandyopadhyay, Aron Habte, and Manajit Sengupta. 2021. *Evaluation of Time-Series Gap-Filling Methods for Solar Irradiance Applications*. Golden, CO: National Renewable Energy Laboratory. NREL/TP-5D00-79987. <https://www.nrel.gov/docs/fy22osti/79987.pdf>.

**NREL is a national laboratory of the U.S. Department of Energy  
Office of Energy Efficiency & Renewable Energy  
Operated by the Alliance for Sustainable Energy, LLC**

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at [www.nrel.gov/publications](http://www.nrel.gov/publications).

Contract No. DE-AC36-08GO28308

**Technical Report**  
NREL/TP-5D00-79987  
October 2021

National Renewable Energy Laboratory  
15013 Denver West Parkway  
Golden, CO 80401  
303-275-3000 • [www.nrel.gov](http://www.nrel.gov)

## NOTICE

This work was authored in part by the National Renewable Energy Laboratory, operated by Alliance for Sustainable Energy, LLC, for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office. The views expressed herein do not necessarily represent the views of the DOE or the U.S. Government.

This report is available at no cost from the National Renewable Energy Laboratory (NREL) at [www.nrel.gov/publications](http://www.nrel.gov/publications).

U.S. Department of Energy (DOE) reports produced after 1991 and a growing number of pre-1991 documents are available free via [www.OSTI.gov](http://www.OSTI.gov).

*Cover Photos by Dennis Schroeder: (clockwise, left to right) NREL 51934, NREL 45897, NREL 42160, NREL 45891, NREL 48097, NREL 46526.*

NREL prints on paper that contains recycled content.

## Acknowledgments

We are grateful to the U.S. Department of Energy Office of Energy Efficiency and Renewable Energy Solar Energy Technologies Office for supporting this research. Specifically, we acknowledge Dr. Tassos Golnas for the support and encouragement.

## List of Acronyms

ARIMA	autoregressive integrated moving average
BON	Bondville, Illinois
BSRN	Baseline Surface Radiation Network
DRA	Desert Rock, Nevada
FPK	Fort Peck, Montana
GHI	global horizontal irradiance
GWN	Goodwin Creek, Mississippi
MAE	mean absolute error
MBE	mean bias error
NA	null
NREL	National Renewable Energy Laboratory
NSRDB	National Solar Radiation Database
PSU	Pennsylvania State University, Pennsylvania
RMSE	root mean square error
SURFRAD	Surface Radiation Budget Network
SXF	Sioux Falls, South Dakota
TBL	Table Mountain, Boulder, Colorado

## Executive Summary

A complete solar resource data set is essential for any stage of a solar energy project—from feasibility studies to daily operations. But measured or modeled solar resource data are prone to data gaps and data quality issues. To mitigate these issues, a data imputation process should be implemented to obtain a complete and reliable temporal and spatial data series. This study focused on imputing temporal scales by applying random and artificial data gaps and then implementing eight imputation methods, including the Kalman filtering and smoothing and spline interpolations. These methods were implemented on 1-minute to half hourly irradiance data for 1 year using a few locations from the National Solar Radiation Database (NSRDB) and ground measurement data set. The results demonstrated that some of the simpler methods, such as the spline and linear interpolation methods, were the relatively best models based on the statistical metrics for imputing NSRDB and ground measurement data, respectively.

# Table of Contents

<b>1</b>	<b>Introduction</b> .....	<b>1</b>
<b>2</b>	<b>Methods</b> .....	<b>2</b>
2.1	Data Selection Criterion.....	3
2.2	Statistical Reporting Metrics.....	4
<b>3</b>	<b>Results</b> .....	<b>5</b>
3.1	Satellite Data.....	5
3.2	Ground-Based Measurements.....	13
	<b>Summary and Future work</b> .....	<b>17</b>
	<b>References</b> .....	<b>18</b>
	<b>Appendix</b> .....	<b>20</b>

## List of Figures

Figure 1. NOAA SURFRAD network and locations .....	2
Figure 2. Distribution of differing strings of consecutive NA values for the BON NSRDB data .....	5
Figure 3. Comparison of imputed $Kt$ values using Kalman filtering on a structural time-series model .....	6
Figure 4. Comparison of imputed $Kt$ values using Kalman filtering on an ARIMA model .....	7
Figure 5. Comparison of imputed $Kt$ values using linear interpolation .....	7
Figure 6. Comparison of imputed $Kt$ values using spline interpolation.....	8
Figure 7. Comparison of imputed $Kt$ values using stine interpolation.....	8
Figure 8. Comparison of imputed $Kt$ values using simple moving average .....	9
Figure 9. Comparison of imputed $Kt$ values using linear weighted moving average .....	9
Figure 10. Comparison of imputed $Kt$ values using exponential weighted moving average.....	10

## List of Tables

Table 1. RMSE for the Three Varying Bin Size Trials.....	11
Table 2. RMSE (%) for the Three Varying Bin Size Trials.....	11
Table 3. MAE for the Three Varying Bin Size Trials.....	12
Table 4. MAE (%) for the Three Varying Bin Size Trials.....	12
Table 5. MBE for the Three Varying Bin Size Trials.....	13
Table 6. MBE (%) for the Three Varying Bin Size Trials .....	13



# 1 Introduction

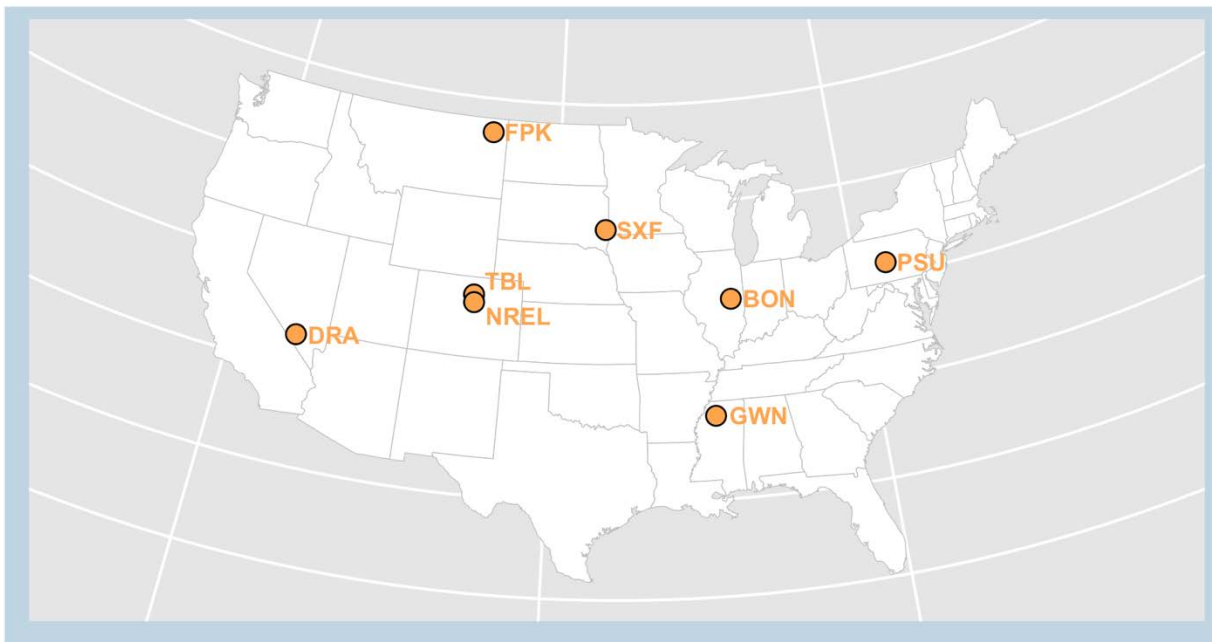
Photovoltaic systems are becoming more prevalent and cost-competitive. As these technologies continue to evolve, accurate and complete solar resource data are critical for predicting the solar energy output of these conversion systems and for reducing the expense associated with mitigating performance risks (Sengupta et al. 2017). Thus, both measured and modeled solar resource data are essential for various phases of solar energy projects—from making policy-level decisions, to checking the viability of these projects, to planning and implementing the final system design. Consequently, solar resource forecast data are essential for utility operators to cost-effectively plan for and manage solar variability and ramping events.

But solar resource data are prone to data gaps (missing data) because of instrument malfunctions, discarded data resulting from data quality problems and/or human error (Sengupta et al. 2017), or satellite issues from modeled satellite-derived solar resource data. These data gaps must be filled using gap-filling modeling techniques to obtain a complete time series data set. Many studies have recommended various gap-filling methods. For example, Roesch et al. (2011a, 2011b) analyzed various empirical gap-filling methods and recommended a mathematical approach for the Baseline Surface Radiation Network (BSRN) data. Similarly, Schwandt et al. (2014) implemented empirical gap-filling methods for measured data sets and, in one recommended approach, they used a satellite-derived data set to gap-fill missing data from measured data sets for overlapping time spans. These approaches are solely for temporal gap-filling methodologies, and they do not address spatial gap-filling approaches; however, there are many sophisticated spatiotemporal gap-filling methodologies for meteorological data, such as that recommended by Wang et al. (2012), which implemented a three-dimensional method based on discrete cosine transforms that use the spatiotemporal information.

In this study, we applied different statistical temporal gap-filling methods to both measured and modeled data sets. The ground measurement data were obtained from seven National Oceanic and Atmospheric Administration Surface Radiation Budget Network (SURFRAD) stations and modeled data from the National Solar Radiation Database (NSRDB V3). The NSRDB is a widely used public data set that provides modeled data of the three-component solar irradiance data set—i.e., global horizontal irradiance (GHI), direct normal irradiance, and diffuse horizontal irradiance—as well as related meteorological parameters (Sengupta et al. 2018). The method developed in this study considers the solar irradiance data and other meteorological inputs to gap-fill natural and artificial missing data using various statistical techniques.

## 2 Methods

In this work, several imputation methods were implemented to gap-fill the incomplete irradiance time series data from NSRDB and ground measurement data. These data were selected from among a few locations, such as the National Renewable Energy Laboratory (NREL) location and the National Oceanic and Atmospheric Administration's SURFRAD sites (Figure 1). The NSRDB data set contains measurements in 30-minute intervals, from which the clearness index,  $Kt$ , series was derived by a ratio method between GHI and extraterrestrial radiation.  $Kt$  is used as a normalization of the GHI by removing the effect of the low sun angle and reducing the scale of values to between 0 and 1. The same techniques were used to gap-fill synthetically removed missing values in the ground-based measurements, which occur in 1-minute intervals. Various imputation methods were tested, and a few were selected based on statistical metrics. The imputation methods' results were determined by the following criteria: mean absolute percentage error (MAE) and MAE (%), root mean square error (RMSE) and RMSE (%), and mean bias error (MBE) and MBE (%). (See Section 2.2 for further details on these criteria.)



**Figure 1. NOAA SURFRAD network and locations**

The following eight imputation methods were selected:

- 1. Kalman filtering and smoothing for structural time series:** Kalman filtering, also known as linear quadratic estimation, produces estimates of unknown variables from a series of measurements observed over time by estimating a joint probability distribution over the variables for each time frame. Structural time-series models are set up in terms of their components, which have a direct interpretation (Grewal 2011; Harvey 1990; Welch and Bishop 1995).
- 2. Kalman filtering and smoothing for the state-space representation of an ARIMA model:** Kalman filtering is applied to an autoregressive integrated moving average

(ARIMA) model, which uses several lagged observations of time series to forecast observations (Grewal 2011; Harvey 1990; Welch and Bishop 1995) .

- 3. Linear interpolation:** Linear interpolation is a curve-fitting method that uses linear polynomials to construct data points within the discrete range of known data points.
- 4. Spline interpolation:** Spline interpolation is a curve-fitting method that uses a piecewise polynomial interpolant to construct data points within a discrete range of known data points (Lyche and Schumaker 1973).
- 5. Stine interpolation:** Stine interpolation is a curve-fitting method that uses piecewise rational interpolation to replace missing values (Stineman 1980).
- 6. Simple moving average:** Simple moving average calculates an average of the last  $n$  observations (Johnston et al. 1999; Ekhsuehi and Dickson 2016).
- 7. Linear weighted moving average:** Linear weighted moving average calculates an average of the last  $n$  observations through applying weighting factors that decrease in a linear fashion (Ekhsuehi and Dickson 2016).
- 8. Exponential weighted moving average:** Exponential weighted moving average calculates of the last  $n$  observations through applying weighting factors that decrease exponentially, never reaching zero (Johnston et al. 1999; Ekhsuehi and Dickson 2016).

Additional methods examined include imputation by last observation carried forward, next observation carried backward, mean value, or random sample. These methods were excluded during the preliminary gap-filling trials because of applicability issues for time series and high deviations of the statistical metrics results. There also exist methods for seasonal data, which include seasonally decomposed and seasonally split missing value imputation. The  $Kt$  series did not exhibit seasonality; therefore, these imputation methods were excluded after preliminary gap-filling trials.

The eight imputation methods were calculated using the built-in functions in the *imputeTS* package in R (Moritz and Bartz-Beielstein 2017). These functions included *na\_interpolation*, *na\_kalman*, and *na\_ma*, where parameters values indicated the type of imputation model to run on the time series (Moritz and Bartz-Beielstein 2017).

## 2.1 Data Selection Criterion

For the Bondville, Illinois (BON) 2017 NSRDB data, data-filtering techniques were applied before performing imputations on the missing values in the  $Kt$  series via the following criterion:

1. Remove the first four rows of null (NA) values from the BON 2017 NSRDB data set, if applicable.
2. Set the observations of  $Kt$  to null where the fill flag variable does not equal zero. The fill flag indicates missing data for values other than zero. Imputation was implemented at these indices. Note: The NSRDB fill flags signify the following—0: no fill; 1: missing cloud type; 2: full time series missing cloud type; 3: missing cloud property; 4: full time series missing cloud property; 5: GHI exceeds clear sky; 6: missing irradiance.

For the ground-based measurements from the NREL data, the following criterion was used to filter the data:

- Remove the  $Kt$  observations where the solar zenith angle is greater than  $89.5^\circ$  to exclude nighttime observations.

## 2.2 Statistical Reporting Metrics

The following statistical measures were used to analyze the performance of the various imputation methods for both the NSRDB and ground measurement data.

MAE and percentage of reading (%) were calculated using the following equations:

$$MAE = \frac{1}{N} \sum_{i=1}^N |y_{true} - y_i|$$

$$MAE(\%) = \frac{\left(\frac{1}{N} \sum_{i=1}^N |y_{true} - y_i|\right)}{\left(\frac{1}{N} \sum_{i=1}^N (y_{true})\right)} * 100$$

where  $y_{true}$  represents the true  $Kt$  values from the data set, and  $y_i$  represents the imputed  $Kt$  values from each method.

Similarly, the RMSE and MBE were computed using the following equations:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_{true})^2}$$

$$RMSE(\%) = \sqrt{\frac{\left(\frac{1}{N} \sum_{i=1}^N (y_i - y_{true})^2\right)}{\left(\frac{1}{N} \sum_{i=1}^N (y_{true})^2\right)}} * 100$$

$$MBE = \frac{1}{N} \sum_{i=1}^N (y_i - y_{true})$$

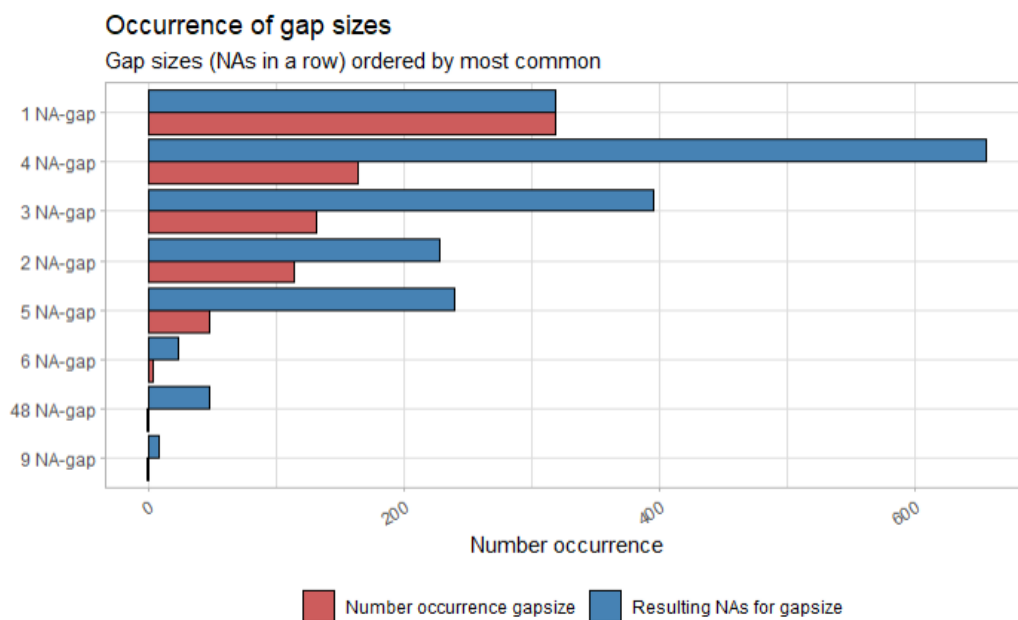
$$MBE(\%) = \frac{\left(\frac{1}{N} \sum_{i=1}^N (y_i - y_{true})\right)}{\left(\frac{1}{N} \sum_{i=1}^N (y_{true})\right)} * 100$$

where  $y_{true}$  represents the true  $Kt$  values from the data set, and  $y_i$  represents the imputed  $Kt$  values from each method. These metrics allowed us to measure the imputation method performance for the NSDRB and ground measurement data.

## 3 Results

### 3.1 Satellite Data

The  $Kt$  time series for the 2017 BON NSRDB data was synthetically modified to show the missing data. To calculate the statistical reporting metrics to measure the method’s performance, synthetic missing values were created and added to the series. The original  $Kt$  series contains 1,920 missing observations. An additional 1,768 values were removed from the series by random sampling of the distribution of the original missing values, shown in Figure 2. This new  $Kt$  series, which was imputed on, contains a total of 3,688 missing values. The synthetically removed values allow for the calculation of the statistical reporting metrics mentioned previously.

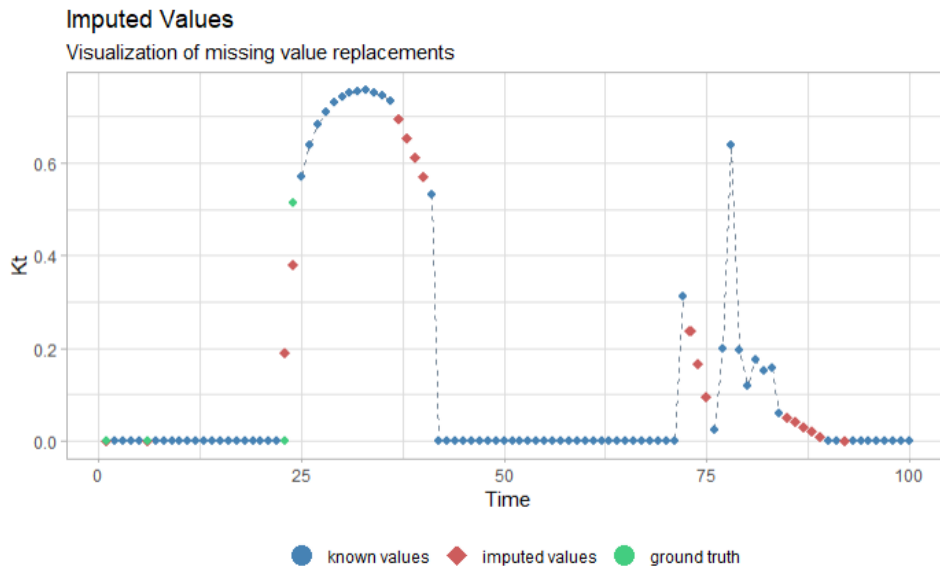


**Figure 2. Distribution of differing strings of consecutive NA values for the BON NSRDB data**

Previous publications have assessed gap-filling procedures for solar radiation data—including “Analysis and Evaluation of Gap Filling Procedures for Solar Radiation Data” (Pereira et al. 2018), “Gap-Filling Techniques for Solar Radiation Data and Their Role in Solar Resource Assessment” (Kumar and Ravindra 2020), and “Global Solar Radiation Modelling Using Polynomial Fitting” (Karim and Singh 2014)—and polynomial curve fitting (Kumar and Ravindra 2020; Karim and Singh 2014). Pereira et al. (2018) proposed the application of both an algebraic equation of the irradiation composition and linear interpolation for filling solarimetric data in southeast Brazil. Both methods were successful in filling the missing data with minor processing time. Kumar and Ravindra (2020) used transmittance ratio, meteorological values, curve-fitting, and comparable satellite data to gap-fill missing solar radiation data at Jodhpur, Rajasthan, a station northwest of India. Polynomial curve fitting and the Centre for Wind Energy Technology - Deutsche Gesellschaft für Internationale Zusammenarbeit (CWET-GIZ) approaches gave the best results for the region while predicting the missing values for various cloud conditions. The CWET-GIZ approach is described in Kumar et al. (2014). Karim and Singh (2014) assessed the performance of the first- to fourth-order-degree polynomial curve

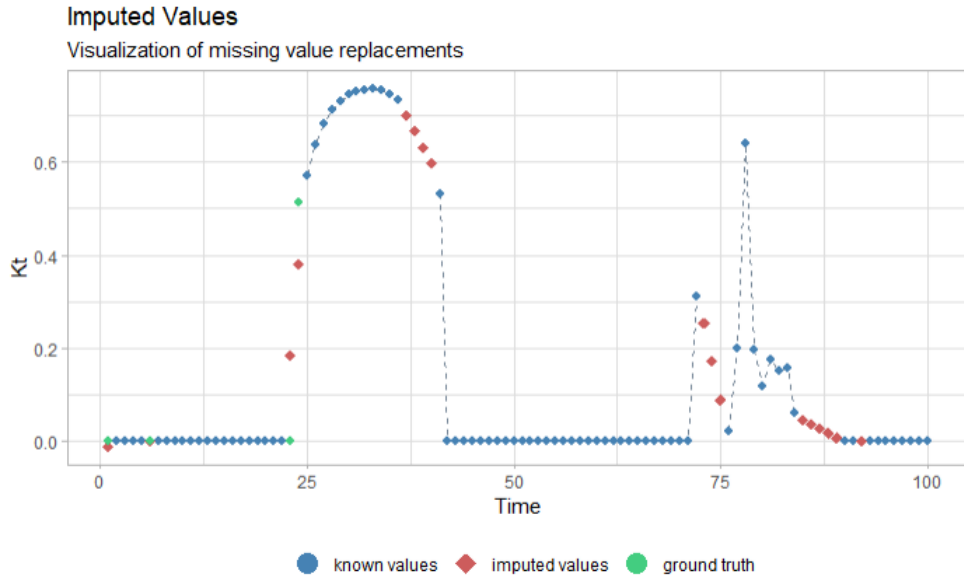
fitting to model global solar radiation by using a solarimeter at Universiti Teknologi PETRONAS, located in Perak, Malaysia. A fitting model with second-degree order gave the best RMSE and R-squared values (Karim and Singh 2014).

The following figures (Figure 3–Figure 10) show the eight imputation methods’ imputed values (red) for the first 100 observations in the gap-filled  $Kt$  series as well as the known values (blue) and the real values (green). The real values are the original  $Kt$  observations that were replaced with null values during the synthetic removal of additional points in the series. The dashed line connects the previous known value to the next known value. There is no dashed line between the imputed or the real values.



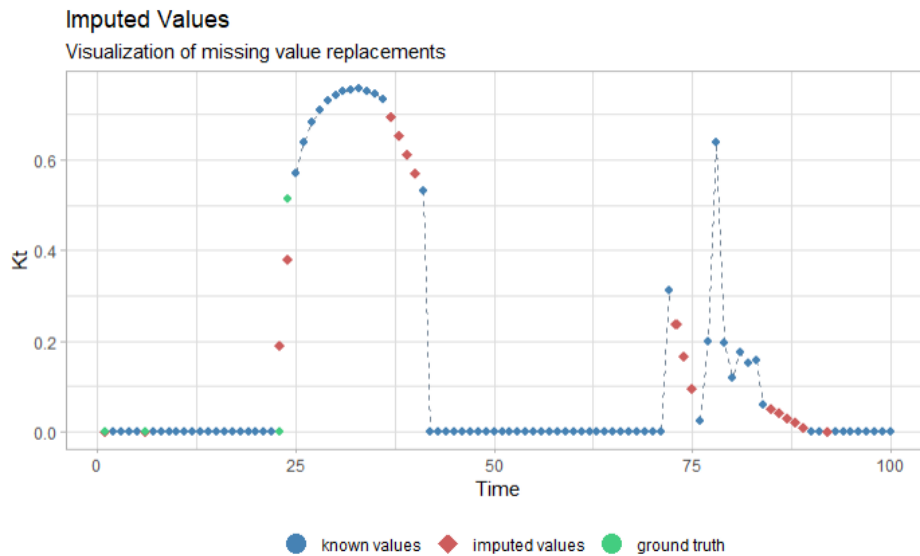
**Figure 3. Comparison of imputed  $Kt$  values using Kalman filtering on a structural time-series model**

Kalman filtering on a structural time series performed well; however, there are discrepancies between the imputed value at approximately the 24th observation point and the real  $Kt$  value.



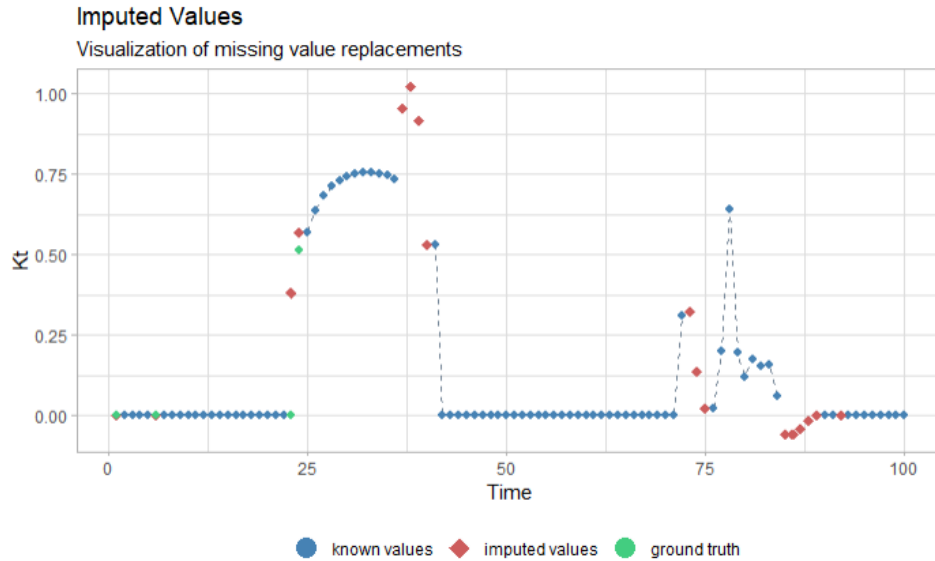
**Figure 4. Comparison of imputed  $K_t$  values using Kalman filtering on an ARIMA model**

The performance of the Kalman filtering on an ARIMA model is similar to the Kalman filtering of a structural time series. The same discrepancy between the imputed and real values occurs near the 24th observation point.



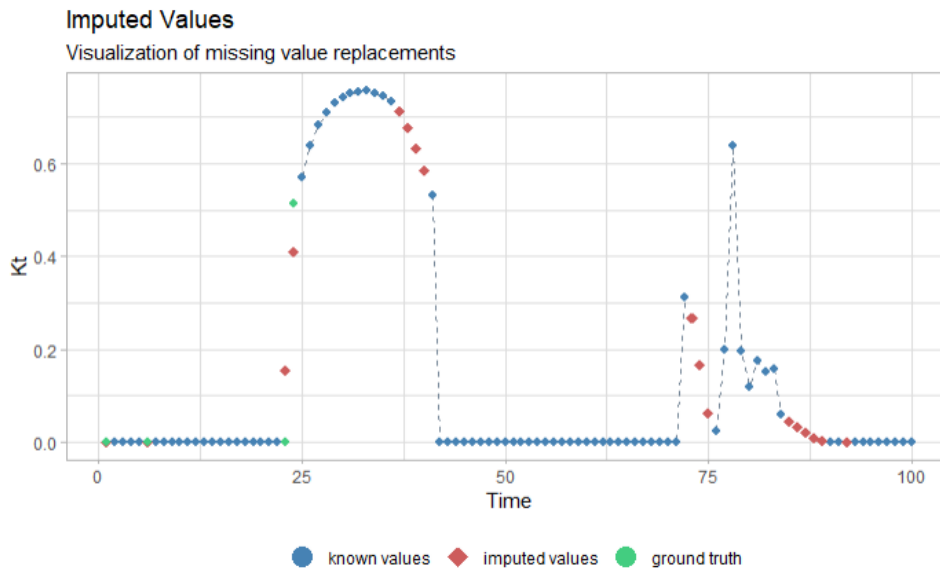
**Figure 5. Comparison of imputed  $K_t$  values using linear interpolation**

Linear interpolation performed well to impute missing values; however, note that the imputed values do not fit the curve in the series well near the 37th to 41st observation points compared to other imputation methods. This is because of the linear nature of the interpolation.



**Figure 6. Comparison of imputed  $K_t$  values using spline interpolation**

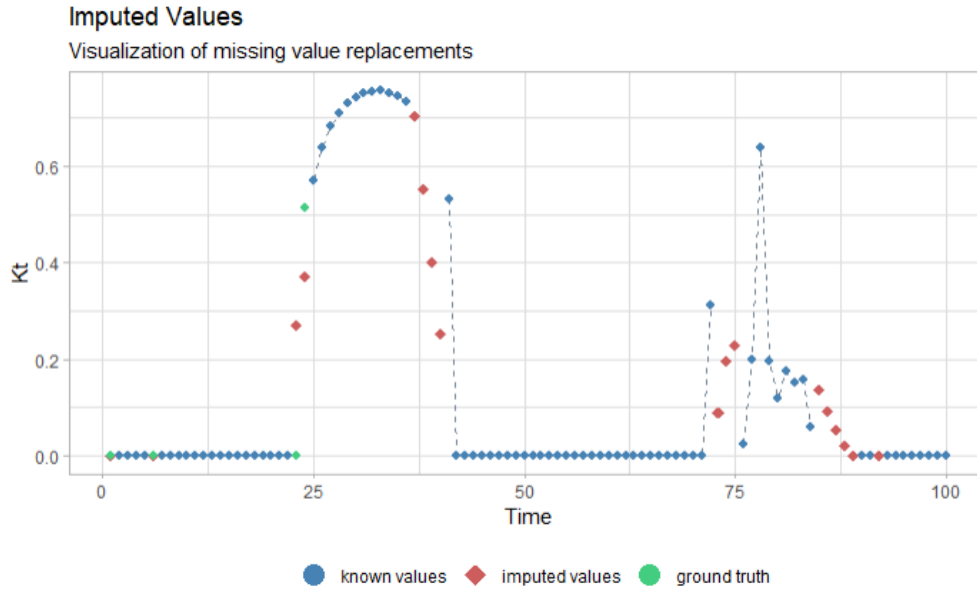
Spline interpolation performed poorly. The imputed values do not closely match the curves of the series, extending well above the maximum of the series near the 37th to 41st observation points.



**Figure 7. Comparison of imputed  $K_t$  values using stine interpolation**

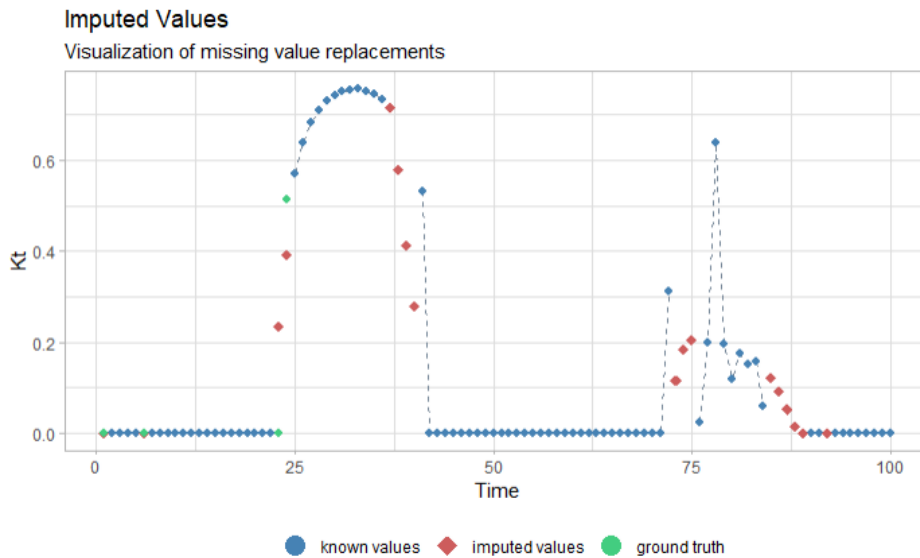
Stine interpolation performed well and captured the curvature in the series that the linear interpolation fails to model. Like the previous imputation models, there is a discrepancy between the imputed and real values near the 24th observation point.





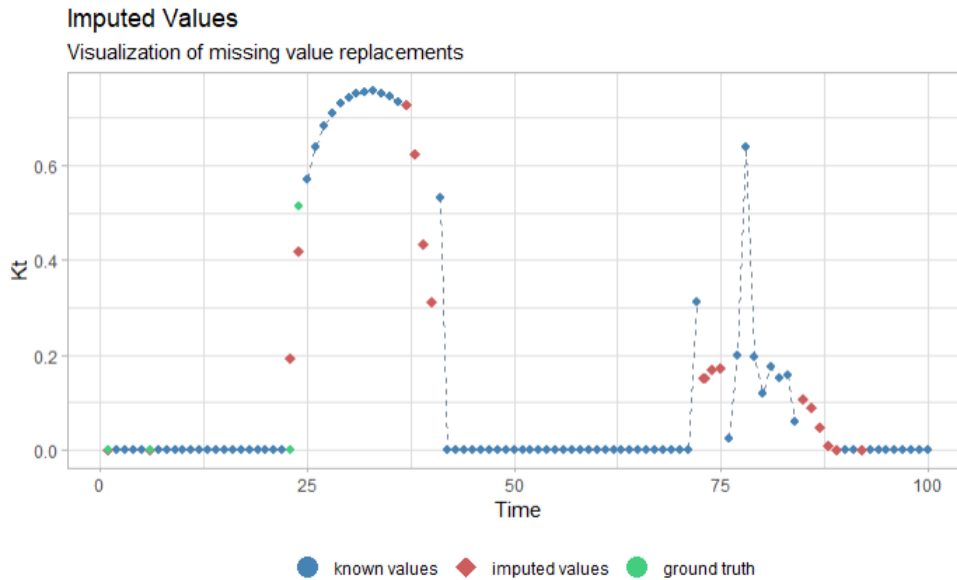
**Figure 8. Comparison of imputed  $Kt$  values using simple moving average**

A simple moving average performed poorly. The imputed values do not closely match the real  $Kt$  values. Additionally, the imputed values do not follow the curvature of the series well, as shown in the 73rd to 75th observation points.



**Figure 9. Comparison of imputed  $Kt$  values using linear weighted moving average**

As shown in Figure 9, a linear weighted average performed poorly. The imputed values do not closely match the real  $Kt$  values. The imputed values near the 73rd to 75th observation points are not well predicted.



**Figure 10. Comparison of imputed  $Kt$  values using exponential weighted moving average**

An exponential weighted moving average also performs poorly. The imputed values near the 73rd to 75th observation points, as shown in Figure 10 fail to capture the nature of the series. It is concluded that moving averages do not work well as imputation methods for the  $Kt$  series.

The imputation methods for the NSRDB  $Kt$  time series were repeated using partitioning tactics on varying string sizes of consecutive NA values. The varying sizes of the consecutive NA values were fit into different bin sizes as follows:

- Trial 1:
  - Bin 1: singular NAs and strings of two consecutive NAs
  - Bin 2: strings of three or more consecutive NAs.
- Trial 2:
  - Bin 3: strings of one to three consecutive NAs
  - Bin 4: strings of four or more consecutive NAs.
- Trial 3:
  - Bin 5: singular NAs
  - Bin 6: two or more consecutive NAs.

The purpose of these partitioning tactics was to differentiate between methods that perform well with larger strings of consecutive NA values compared to methods that only perform well with singular NA values and/or small strings of consecutive NA values. The following tables (Table 1–Table 6) provide the statistical reporting metrics for all three trials.

**Table 1. RMSE for the Three Varying Bin Size Trials**

Imputation Method	RMSE					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	0.0815	0.141	0.0854	0.156	0.0711	0.141
KalmanRun and ARIMA model	0.0816	0.141	0.0856	0.157	0.0712	0.142
Linear interpolation	0.0815	0.141	0.0854	0.156	0.0711	0.141
Spline interpolation	0.147	0.345	0.156	0.394	0.0940	0.351
Stine interpolation	<b>0.0812</b>	<b>0.139</b>	<b>0.0839</b>	<b>0.1554</b>	<b>0.0709</b>	<b>0.0140</b>
Simple moving average	0.107	0.152	0.106	0.167	0.0873	0.153
Linear weighted moving average	0.0969	0.147	0.0973	0.163	0.0796	0.148
Exponential weighted moving average	0.0900	0.145	0.0917	0.162	0.0752	0.146

The bold font represents the best-performing method for each bin size. For all six bin sizes, stine interpolation outperformed the other imputation methods when measuring RMSE.

**Table 2. RMSE (%) for the Three Varying Bin Size Trials**

Imputation Method	RMSPE (%)					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	<b>20.4</b>	34.9	21.5	38.5	17.3	35.0
KalmanRun and ARIMA model	<b>20.4</b>	35.0	21.5	38.7	17.3	53.1
Linear interpolation	<b>20.4</b>	34.9	21.5	38.5	17.3	35.0
Spline interpolation	36.9	85.7	39.2	97.2	22.9	87.1
Stine interpolation	<b>20.4</b>	<b>34.6</b>	<b>21.2</b>	<b>38.3</b>	<b>17.2</b>	<b>34.7</b>
Simple moving average	26.7	37.8	26.5	41.3	21.3	38.0
Linear weighted moving average	24.3	36.6	24.5	40.2	19.4	36.8
Exponential weighted moving average	22.6	36.1	23.0	39.8	18.3	36.3

For the bin size of the singular and two-length consecutive NA values, the Kalman filtering for both the structural time series and the ARIMA model as well as the linear and stine interpolations performed the best; however, stine interpolation outperformed these and the remaining methods for the other bin sizes.

**Table 3. MAE for the Three Varying Bin Size Trials**

Imputation Method	MAE					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	0.0390	0.0646	0.0416	0.0723	0.0320	0.0643
KalmanRun and ARIMA model	0.0392	0.0648	0.0417	0.0726	0.0320	0.0645
Linear interpolation	0.0390	0.0646	0.0416	0.0723	0.0320	0.0643
Spline interpolation	0.0586	0.131	0.0644	0.155	0.0416	0.133
Stine interpolation	<b>0.0371</b>	<b>0.0600</b>	<b>0.0385</b>	<b>0.0678</b>	<b>0.0297</b>	<b>0.0597</b>
Simple moving average	0.0577	0.0768	0.0577	0.0844	0.0450	0.0771
Linear weighted moving average	0.0515	0.0722	0.0521	0.0800	0.0397	0.0725
Exponential weighted moving average	0.0457	0.0678	0.0466	0.0758	0.0353	0.0678

For all six bin sizes, stine interpolation was the best-performing imputation method.

**Table 4. MAE (%) for the Three Varying Bin Size Trials**

Imputation Method	MAE (%)					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	15.3	25.5	16.6	28.0	12.1	25.3
KalmanRun and ARIMA model	15.3	25.6	16.6	28.2	12.1	25.4
Linear interpolation	15.3	25.5	16.6	28.0	12.1	25.3
Spline interpolation	22.9	51.7	25.7	59.9	15.7	52.5
Stine interpolation	<b>14.5</b>	<b>23.7</b>	<b>15.4</b>	<b>26.3</b>	<b>11.2</b>	<b>23.5</b>
Simple moving average	22.5	30.3	23.1	32.7	17.0	30.3
Linear weighted moving average	20.1	28.5	20.8	31.0	15.0	28.5
Exponential weighted moving average	17.9	26.7	18.6	29.4	13.3	26.7

Stine interpolation also performed the best when considering MAE in terms of percentage of reading.

**Table 5. MBE for the Three Varying Bin Size Trials**

Imputation Method	MBE					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	-0.00379	0.00855	-0.00447	<b>0.0142</b>	-0.00208	0.00766
KalmanRun and ARIMA model	-0.00422	0.00870	0.00468	0.0143	-0.00247	0.00786
Linear interpolation	-0.00379	0.00855	-0.00447	<b>0.0142</b>	-0.00208	0.00766
Spline interpolation	-0.00487	0.0493	<b>0.00022</b>	0.0659	-0.0115	0.0509
Stine interpolation	<b>-0.00176</b>	0.00943	-0.00311	0.0153	-0.00102	0.00881
Simple moving average	0.00253	<b>0.00744</b>	-0.00269	0.0146	-0.00068	0.00640
Linear weighted moving average	0.00221	0.00776	-0.00250	0.0146	<b>0.00023</b>	<b>0.00670</b>
Exponential weighted moving average	0.00179	0.00844	-0.00227	0.0151	0.00079	0.00737

No imputation method performed the best for each bin size in terms of prediction bias. Stine interpolation had the smallest average bias for Bin 1 but did not hold the smallest bias for any other bin.

**Table 6. MBE (%) for the Three Varying Bin Size Trials**

Imputation Method	MBE (%)					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
KalmanSmoother and StructTS model	-1.48	3.37	-1.78	<b>5.49</b>	-0.784	3.01
KalmanRun and ARIMA model	-1.65	3.43	-1.87	5.56	-0.934	3.09
Linear interpolation	-1.48	3.37	-1.78	<b>5.49</b>	-0.784	3.01
Spline interpolation	-1.91	19.4	<b>0.088</b>	25.6	-4.36	20.0
Stine interpolation	<b>-0.687</b>	3.72	-1.24	5.91	-0.387	3.46
Simple moving average	0.988	<b>2.94</b>	-1.07	5.64	-0.259	<b>2.52</b>
Linear weighted moving average	0.863	3.06	-0.999	5.67	<b>0.085</b>	2.63
Exponential weighted moving average	0.701	3.33	-0.907	5.86	0.300	2.90

Looking at the calculations in terms of percentage of reading for MBE, linear interpolation has the smallest average bias for Bin 4; however, there is no clear best-performing method in terms of bias.

From the graphs and tables, we conclude that for the BON NSRDB (2017) data, stine interpolation is the best-performing imputation method; however, the results of the stine interpolation do not significantly differ from those of linear interpolation.

### 3.2 Ground-Based Measurements

The clearness index ( $Kt$ ) was calculated by using the extraterrestrial irradiance. The  $Kt$  time series for the NREL ground measurements is complete; therefore, all the missing values for the

series were synthetically removed. To create values to remove from the series, we used the same distribution of the original missing values (NA) in the NSRDB *Kt* series. This distribution is shown in Figure 2 and is listed as follows:

- 1 NA in a row: 319 times
- 2 NA in a row: 114 times
- 3 NA in a row: 132 times
- 4 NA in a row: 164 times
- 5 NA in a row: 48 times
- 6 NA in a row: 4 times
- 9 NA in a row: 1 times.

The NSRDB data are given in 30-minute intervals, whereas the ground based-measurements are given in 1-minute intervals. To account for this in the manual removal of the ground measurements, each NA row size ranged from 1 to 30 multiplied by its row size. For example, strings of one NA had potential values between 1 and 30. Strings of two consecutive NAs had potential values between 31 and 60, and so on. This factors in the 30-minute to 1-minute conversion. These ranged values based on the NSRDB NA distribution where then randomly removed from the ground-based measurements.

Similar to the NSRDB data trials, partitioning tactics were applied to the calculation criteria by creating various bin sizes in the ground-based measurements as follows:

- Bin 1: strings of 1–30 consecutive NAs
- Bin 2: strings of 31–60 consecutive NAs
- Bin 3: strings of 61–90 consecutive NAs
- Bin 4: strings of 91–120 consecutive NAs
- Bin 5: strings of 121–150 consecutive NAs
- Bin 6: strings of 151+ consecutive NAs.

The purpose of the partitioning tactics was to differentiate among methods that perform well with larger strings of consecutive NA values compared to methods that only perform well with singular NA values and/or small strings of consecutive NA values. In the case of the ground-based measurements, we compared only the linear and stine interpolations because of their high performance shown in Section 3.1 compared to the other imputation methods. The results are summarized in Table 7–Table 11.

**Table 7. RMSE for the Six Varying Bin Sizes**

Imputation Method	RMSE					
	Bin 1	Bin 2	Bin 3	Bin4	Bin 5	Bin 6
Linear interpolation	<b>0.158</b>	<b>0.194</b>	<b>0.208</b>	<b>0.226</b>	<b>0.251</b>	<b>0.223</b>
Stine interpolation	0.164	0.205	0.216	0.258	0.263	0.240

The bold font represents the best-performing method for each bin size. Linear interpolation outperformed stine interpolation in the case of the ground-based measurements analyzing RMSE.

**Table 8. RMSE (%) for the Six Varying Bin Sizes**

RMSE (%)						
Imputation Method	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	<b>27.7</b>	<b>33.3</b>	<b>35.9</b>	<b>39.1</b>	<b>43.4</b>	<b>38.6</b>
Stine interpolation	28.8	35.2	37.3	44.7	45.5	41.6

For all of the bin sizes, linear interpolation had a better percentage of reading for RMSE than stine interpolation. This is opposite of the NSRDB data trials.

**Table 9. MAE for the Six Varying Bin Sizes**

MAE						
Imputation Method	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0915	0.130	<b>0.144</b>	<b>0.168</b>	<b>0.188</b>	<b>0.159</b>
Stine interpolation	<b>0.0194</b>	<b>0.100</b>	0.145	0.173	0.192	0.163

For bins 1 and 2, stine interpolation performed better than linear interpolation; however, for the remaining bins, linear interpolation had better MAE values.

**Table 10. MAE (%) for the Six Varying Bin Sizes**

MAE (%)						
Imputation Method	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	<b>18.2</b>	<b>25.3</b>	<b>27.8</b>	<b>32.7</b>	<b>36.6</b>	<b>31.3</b>
Stine interpolation	<b>18.2</b>	26.1	28.1	33.6	37.4	31.7

Linear interpolation was the better imputation method for the ground-based measurements when looking at the percentage of reading for MAE (%) for the other bin sizes; however, the methods return similar MAE% percentages for Bin 1.

**Table 11. MBE for the Six Varying Bin Sizes**

MBE						
Imputation Method	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00511	0.00617	<b>0.000782</b>	0.00936	<b>0.00749</b>	<b>0.00642</b>
Stine interpolation	<b>0.00299</b>	<b>0.000210</b>	-0.00251	<b>0.00286</b>	0.0201	0.00735

For bins 3, 5, and 6, linear interpolation had a lower average bias than stine interpolation; however, for bins 1, 2, and 4, stine interpolation had the lower average bias. This is consistent with the NSRDB results, where stine interpolation did not consistently have the lowest bias among the other imputation methods.

**Table 12. MBE (%) for the Six Varying Bin Sizes**

Imputation Method	MBE (%)					
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	1.02	1.20	<b>0.151</b>	1.82	<b>1.46</b>	<b>1.25</b>
Stine interpolation	<b>0.594</b>	<b>0.0410</b>	-0.484	<b>0.556</b>	3.91	1.43

For the percentage of reading of MBE, we obtained the same results for each method and bin size as shown for MBE in Table 11.

For the NSRDB data, stine interpolation performed the best, whereas for the ground-based measurements, linear interpolation performed the best. In the NSRDB case, the stine interpolation results did not significantly differ from those of linear interpolation. Computationally, linear interpolation is more appropriate because of its simplicity compared to stine interpolation; therefore, we can conclude that linear interpolation is the most appropriate choice of imputation method.

Further, we performed both methods on the NSRDB and ground-based measurements for the other seven locations (see Appendix). Overall, the methods performed similarly. We did not see differing results in the other locations.

These differences in method performance might lie in the ratio of the sum of clear  $Kt$  values divided by the sum of cloudy-sky  $Kt$  values. The ratio for the NSRDB data was 0.4592721, whereas the ratio for the ground measurements was 0.7159485. This means that the NSRDB data contained more cloudy points than the ground-measurement data, potentially affecting the choice of interpolation method. This ratio and its effect on the interpolation method can be examined in a future project.



## Summary and Future work

Missing data exist in measured or modeled irradiance data, which cause problems in various phases or solar energy system projects. This study investigated eight imputation methods. Based on the statistical metrics, the sine and linear interpolation imputation methodologies performed better for NSRDB and ground measurements, respectively. These results provide a baseline models toward realistic reconstruction of time series in the presence of significant amounts of missing data as well as data with inferior data quality. Future work will encompass machine learning imputation techniques with forecasting ability. Ensemble models of these machine learning algorithms for forecasting big data time series will be investigated.

## References

- 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. Berlin, Heidelberg: Springer. [https://doi.org/10.1007/978-3-642-04898-2\\_321](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>.
- Johnston, F., J. Boyland, M. Meadows, et al. 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>.
- Karim, S., and B. Singh. 2014. “Global Solar Radiation Modeling Using Polynomial Fitting.” *Applied Mathematical Sciences* 8: 367–78.
- Kumar, D., and B. Ravindra. 2020. “Gap-Filling Techniques for Solar Radiation Data and Their Role in Solar Resource Assessment.” In *Advances in Energy Research, Vol. 1*, edited by S. Singh and V. Ramadesigan. Singapore: Springer. [https://doi.org/10.1007/978-981-15-2666-4\\_53](https://doi.org/10.1007/978-981-15-2666-4_53).
- Kumar, S., G. Gomathinayagam, I. Giridhar, R. Mitra, R. Vashistha, R. Meyer, M. Schwandt, and K. Chhatbar. 2014. “Field Experiences with the Operation of Solar Radiation Resource Assessment Stations in India.” *Energy Procedia* 49: 2,351–61. ISSN 1876-6102. <https://doi.org/10.1016/j.egypro.2014.03.249>.
- 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. Birkhäuser.
- Moritz, S., and T. Bartz-Beielstein. 2017. “imputeTS: Time Series Missing Value Imputation in R.” *The R Journal* 9 (1): 207–18. <https://doi.org/10.32614/RJ-2017-009>.
- Pereira, Gessica, Rafael Stonoga, Daniel Detzel, Kristi Kuster, Renato Neto, and Luis Paschoalotto. 2018. “Analysis and Evaluation of Gap Filling Procedures for Solar Radiation Data.” *2018 IEEE 9th Power, Instrumentation and Measurement Meeting (EPIM)*: 1–6. <https://doi.org/10.1109/EPIM.2018.8756358>.
- Roesch, A., M. Wild, A. Ohmura, E. G. Dutton, C. N. Long, and T. Zhang. 2011a. “Assessment of BSRN Radiation Records for the Computation of Monthly Means.” *Atmospheric Measurement Techniques* 4: 339–54. <https://doi.org/10.5194/amt-4-339-2011>.
- Roesch, A., M. Wild, A. Ohmura, E. G. Dutton, C. N. Long, and T. Zhang. 2011b. “Assessment of BSRN Radiation Records for the Computation of Monthly Means.” *Atmospheric Measurement Techniques* 4: 339–54. Corrigendum, *Atmospheric Measurement Techniques* 4 (973), 2011.

Schwandt, Marko, Kaushal Chhatbar, Richard Meyer, Katharina Fross, Indradip Mitra, Ramadhan Vashistha, Godugunur Giridhar, S. Gomathinayagam, and Ashvini Kumar. 2014. “Development and Test of Gap Filling Procedures for Solar Radiation Data of the Indian SRRA Measurement Network.” *Energy Procedia* 57: 1100–9. <http://dx.doi.org/10.1016/j.egypro.2014.10.096>.

Sengupta, Manajit, Aron Habte, Stefan Wilbert, Christian Gueymard, and Jan 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, Manajit, Yu Xie, Anthony Lopez, Aron Habte, Galen Maclaurin, and James Shelby. 2018. “The National Solar Radiation Data Base (NSRDB).” *Renewable and Sustainable Energy Reviews* 89: 51–60. ISSN 1364-0321. <https://doi.org/10.1016/j.rser.2018.03.003>. <http://www.sciencedirect.com/science/article/pii/S136403211830087X>.

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

Wang, Guojie, Damien Garcia, Yi Liu, Richard de Jeu, and A. Johannes Dolman. 2012. “A Three-Dimensional Gap Filling Method for Large Geophysical Datasets: Application to Global Satellite Soil Moisture Observations.” *Environmental Modelling & Software* 30: 139–42. ISSN 1364-8152. <https://doi.org/10.1016/j.envsoft.2011.10.015>. <http://www.sciencedirect.com/science/article/pii/S1364815211002453>.

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

# Appendix

## A.1 BON: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.108	0.149	0.170	0.191	0.198	0.176
Stine interpolation	0.113	0.159	0.190	0.227	0.227	0.202
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	21.5	27.8	32.4	37.3	39.0	34.0
Stine interpolation	22.5	29.6	36.1	44.2	44.6	39.0
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0679	0.0978	0.124	0.138	0.148	0.124
Stine interpolation	0.0671	0.100	0.134	0.158	0.158	0.136
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	15.5	20.3	26.5	30.9	33.0	27.3
Stine interpolation	15.4	20.9	28.7	35.5	35.3	29.9
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.00152	-0.00670	0.0113	-0.0123	0.0266	0.00153
Stine interpolation	0.00135	-0.00577	0.0110	-0.0232	0.0496	0.000905
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.348	-1.39	2.42	-2.76	6.68	0.338
Stine interpolation	0.309	-1.20	2.34	-5.20	11.1	0.199

## A.2 NREL: NSRDB Results

<b>RMSE</b>						
<b>imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.110	0.132	0.121	0.131	0.101	0.133
Stine interpolation	0.111	0.133	0.123	0.130	0.101	0.134
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	28.6	31.3	29.6	31.5	25.3	31.4
Stine interpolation	28.8	31.4	30.1	31.3	25.2	31.5
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0488	0.0713	0.0595	0.0697	0.0423	0.0724
Stine interpolation	0.0459	0.0653	0.0568	0.0625	0.0392	0.0664
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	21.9	27.3	23.7	28.0	17.7	27.3
Stine interpolation	20.6	25.0	22.6	25.1	16.4	25.0
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00168	-0.00609	-0.000803	-0.00661	-0.00376	-0.00732
Stine interpolation	0.00114	-0.00478	-0.00133	-0.00467	-0.00333	-0.00573
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.757	-2.33	-0.319	-2.65	-1.57	-2.76
Stine interpolation	0.513	-1.83	-0.528	-1.87	-1.39	-2.16

### A.3 FPK: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0874	0.137	0.0954	0.140	0.0922	0.137
Stine interpolation	0.0868	0.133	0.0956	0.135	0.0929	0.133
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	22.1	36.9	24.9	37.3	22.9	36.6
Stine interpolation	21.9	35.7	24.9	36.0	23.1	35.5
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0388	0.0778	0.0428	0.0809	0.0409	0.0788
Stine interpolation	0.0354	0.0677	0.0391	0.0701	0.0382	0.0684
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	16.2	36.4	19.0	37.3	16.4	36.2
Stine interpolation	14.8	31.7	17.3	32.3	15.4	31.4
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00182	0.0140	0.00419	0.0144	0.00322	0.0131
Stine interpolation	0.00272	0.0135	0.00500	0.0138	0.00369	0.0124
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.759	6.56	1.86	6.64	1.30	6.02
Stine interpolation	1.14	6.33	2.21	6.36	1.48	5.71

## A.4 FPK: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.137	0.161	0.174	0.193	0.217	0.185
Stine interpolation	0.144	0.170	0.208	0.221	0.244	0.210
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	25.2	30.7	31.7	35.4	39.9	34.2
Stine interpolation	26.3	32.3	37.7	40.6	44.9	38.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0812	0.108	0.122	0.138	0.163	0.130
Stine interpolation	0.0825	0.100	0.138	0.153	0.180	0.142
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	16.6	22.6	24.8	28.3	33.2	26.6
Stine interpolation	16.8	23.1	28.1	31.4	36.7	29.1
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.00991	0.0111	0.0159	-0.00214	-0.0231	-0.00322
Stine interpolation	-0.00712	0.0210	0.0336	0.0155	-0.0160	0.00950
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-2.02	2.33	3.24	-0.439	-4.70	-0.659
Stine interpolation	-1.45	4.41	6.83	3.17	-3.25	1.95

## A.5 TBL: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.102	0.123	0.105	0.128	0.0997	0.120
Stine interpolation	0.0991	0.122	0.103	0.127	0.0992	0.119
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	24.5	30.3	26.4	30.7	25.4	29.7
Stine interpolation	23.9	30.0	26.0	30.4	25.3	29.5
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0507	0.0646	0.0515	0.0701	0.0436	0.0624
Stine interpolation	0.0462	0.0592	0.0470	0.0642	0.0404	0.0571
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	19.3	25.8	21.4	26.3	18.4	25.3
Stine interpolation	17.6	23.6	19.5	24.1	17.1	23.1
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.00386	-0.000425	-0.00462	0.00186	-0.00308	0.000142
Stine interpolation	-0.00257	-0.000385	-0.00425	0.00225	-0.00219	0.000107
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-1.47	-0.170	-1.92	0.698	-1.30	0.0578
Stine interpolation	-0.978	-0.154	-1.77	0.842	-0.923	0.0435



## A.6 TBL: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.158	0.194	0.208	0.226	0.251	0.223
Stine interpolation	0.164	0.205	0.216	0.258	0.263	0.240
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	27.7	33.3	35.9	39.1	43.4	38.6
Stine interpolation	28.8	35.2	37.3	44.7	45.5	41.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0915	0.130	0.144	0.168	0.188	0.160
Stine interpolation	0.0914	0.134	0.146	0.173	0.192	0.163
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	18.2	25.3	27.8	32.7	36.6	31.1
Stine interpolation	18.2	26.1	28.1	33.6	37.4	31.7
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00511	0.00617	0.000782	0.00936	0.00749	0.00642
Stine interpolation	0.00299	0.000210	-0.00251	0.00286	0.0201	0.00735
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	1.02	1.20	0.151	1.82	1.46	1.25
Stine interpolation	0.594	0.0410	-0.484	0.556	3.91	1.43

## A.7 SXF: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0833	0.124	0.0937	0.128	0.0789	0.124
Stine interpolation	0.0824	0.122	0.0916	0.128	0.0775	0.122
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	20.5	30.2	23.4	31.0	19.8	30.3
Stine interpolation	20.3	29.9	22.9	30.8	19.4	29.9
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0381	0.0613	0.0442	0.0637	0.0342	0.0613
Stine interpolation	0.0356	0.0559	0.0405	0.0584	0.0317	0.0557
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	14.8	23.6	17.6	24.0	13.8	23.7
Stine interpolation	13.8	21.5	16.1	22.0	12.8	21.6
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00791	-0.00926	0.00194	-0.00986	0.00197	-0.00901
Stine interpolation	0.00888	-0.00879	0.00257	-0.00926	0.00293	-0.00836
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	3.08	-3.57	0.774	-3.71	0.795	-3.49
Stine interpolation	3.45	-3.38	1.03	-3.48	1.18	-3.23

## A.8 SXF: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.115	0.165	0.162	0.186	0.212	0.178
Stine interpolation	0.125	0.172	0.186	0.208	0.223	0.195
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	21.4	31.7	31.4	34.3	39.6	33.4
Stine interpolation	23.3	32.9	36.1	38.2	41.7	36.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0688	0.110	0.115	0.135	0.159	0.125
Stine interpolation	0.0717	0.100	0.123	0.140	0.159	0.129
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	14.4	24.0	25.5	27.7	33.7	26.6
Stine interpolation	15.0	24.2	27.2	28.8	33.6	27.4
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.000138	-0.00386	0.00239	0.00717	-0.0240	-0.00246
Stine interpolation	0.00117	-0.00946	0.0116	0.0165	-0.0171	0.00363
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.0289	-0.845	0.530	1.47	-5.08	-0.523
Stine interpolation	0.245	-2.07	2.57	3.40	-3.61	0.770

## A.9 PSU: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0897	0.109	0.0931	0.114	0.0843	0.110
Stine interpolation	0.0899	0.110	0.0928	0.115	0.0842	0.111
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	23.8	29.0	24.9	29.9	21.3	29.1
Stine interpolation	23.8	29.1	24.9	30.2	21.3	29.3
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0443	0.0593	0.0490	0.0615	0.0417	0.0598
Stine interpolation	0.0422	0.0554	0.0454	0.0580	0.0398	0.0558
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	18.8	25.5	21.2	26.2	16.6	25.6
Stine interpolation	17.9	23.8	19.6	24.7	15.9	23.9
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.00413	-0.00426	-0.00442	-0.00407	0.00249	-0.00478
Stine interpolation	-0.00276	-0.00408	-0.00332	-0.00414	0.00332	-0.00451
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-1.76	-1.83	-1.91	-1.73	0.991	-2.05
Stine interpolation	-1.17	-1.75	-1.44	-1.73	1.32	-1.93

## A.10 PSU: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.167	0.223	0.187	0.234	0.253	0.220
Stine interpolation	0.211	0.242	0.236	0.324	0.354	0.293
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	34.4	43.7	38.5	49.5	52.2	45.5
Stine interpolation	43.3	47.5	48.5	68.6	73.0	60.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.104	0.152	0.132	0.171	0.192	0.157
Stine interpolation	0.122	0.200	0.155	0.218	0.259	0.195
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	25.2	34.4	31.6	43.2	48.1	38.4
Stine interpolation	29.8	36.1	37.2	55.1	64.9	47.7
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.0142	-0.0153	0.00625	0.0256	-0.0153	0.00590
Stine interpolation	-0.0143	-0.00721	0.0119	0.0332	-0.0399	0.00685
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-3.45	-3.45	1.50	6.45	-3.84	1.45
Stine interpolation	-3.49	-1.63	2.86	8.38	-10.0	1.68

## A.11 DRA: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0861	0.0948	0.0863	0.102	0.0794	0.0944
Stine interpolation	0.0867	0.0925	0.0860	0.0992	0.0787	0.0920
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	18.2	20.8	18.4	22.7	16.8	20.8
Stine interpolation	18.3	20.3	18.3	22.1	16.6	20.3
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0350	0.0431	0.0373	0.0450	0.0320	0.0428
Stine interpolation	0.0321	0.0372	0.0331	0.0395	0.0298	0.0369
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	11.0	14.7	11.9	16.1	10.2	14.8
Stine interpolation	10.1	12.7	10.6	14.2	9.49	12.8
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.000267	-0.000946	0.000282	-0.00214	0.00292	0.000069
Stine interpolation	0.00180	0.000763	0.00199	-0.000702	0.00404	0.00172
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0844	-0.324	0.0902	-0.768	0.931	0.0238
Stine interpolation	0.569	0.261	0.635	-0.252	1.29	0.596

## A.12 DRA: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.120	0.140	0.179	0.196	0.228	0.172
Stine interpolation	0.121	0.140	0.182	0.206	0.229	0.176
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	18.2	20.9	27.1	29.7	33.3	25.9
Stine interpolation	18.3	20.9	27.7	31.2	33.5	26.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0630	0.0834	0.115	0.139	0.167	0.109
Stine interpolation	0.0607	0.100	0.120	0.138	0.157	0.106
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	10.1	13.2	18.5	22.5	25.6	17.5
Stine interpolation	9.74	12.5	17.7	22.3	24.1	16.9
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00719	-0.0156	0.00826	-0.00691	-0.0110	-0.00127
Stine interpolation	0.00963	-0.0110	0.0190	0.00540	0.00819	0.00779
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	1.15	-2.47	1.33	-1.12	-1.69	-0.204
Stine interpolation	1.55	-1.74	3.06	0.875	1.26	1.25

## A.13 GCM: NSRDB Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0921	0.119	0.0958	0.130	0.0857	0.120
Stine interpolation	0.0908	0.118	0.0948	0.128	0.0846	0.118
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	22.9	30.9	23.8	34.6	22.5	31.1
Stine interpolation	22.6	30.5	23.6	34.1	22.2	30.6
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0471	0.0604	0.0478	0.0678	0.0420	0.0607
Stine interpolation	0.0435	0.0555	0.0440	0.0623	0.0394	0.0557
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	18.1	25.1	18.4	29.6	17.5	25.4
Stine interpolation	16.7	23.1	17.0	27.1	16.4	23.3
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.00423	-0.00818	-0.000823	-0.00874	-0.000339	-0.00901
Stine interpolation	0.00432	-0.00731	-0.000147	-0.00823	-0.000315	-0.00824
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	1.63	-3.40	-0.318	-3.81	-0.141	-3.77
Stine interpolation	1.66	-3.04	-0.0567	-3.58	-0.131	-3.44



## A.14 GCM: Ground-Based Measurement Results

<b>RMSE</b>						
<b>Imputation Method</b>	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.128	0.150	0.157	0.180	0.219	0.168
Stine interpolation	0.132	0.165	0.165	0.194	0.235	0.180
<b>RMSE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	23.5	28.2	29.5	32.4	40.3	31.0
Stine interpolation	24.2	31.0	31.0	35.0	43.4	33.2
<b>MAE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	0.0749	0.0986	0.110	0.131	0.170	0.118
Stine interpolation	0.0749	0.100	0.113	0.134	0.173	0.120
<b>MAE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	15.3	21.1	23.2	26.5	34.9	24.3
Stine interpolation	15.3	22.2	23.9	27.0	35.4	24.9
<b>MBE</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-0.00745	-0.0154	-0.00395	-0.00638	-0.0213	-0.00864
Stine interpolation	-0.00177	-0.00563	0.00378	0.00861	-0.00864	0.00209
<b>MBE (%)</b>						
	Bin 1	Bin 2	Bin 3	Bin 4	Bin 5	Bin 6
Linear interpolation	-1.52	-3.30	-0.835	-1.29	-4.36	-1.79
Stine interpolation	-0.363	-1.20	0.799	1.74	-1.77	0.433