Final Technical Report (FTR)

Agency/Office/Program	DOE/EERE/Solar Energy Technology Office	
Award Number	DE- EE0008216	
Project Title	HAIMOS Ensemble Forecasts for Intra-day and Day-Ahead GHI, DNI and Ramps	
Principal Investigator	Carlos F. M. Coimbra, Professor, ccoimbra@ucsd.edu, (858) 534-4285	
Business Contact	Susanna Pastell, spastell@ucsd.edu, (858) 534-4896	
Submission Date	9/30/2021	
DUNS Number	804355790	
Recipient Organization	University of California, San Diego	
Project Period	**Start**: 7/1/2018	**End**: 6/30/2021
Project Budget	Total $1,478,703 (DOE: $1,316,203; C/S: $162,500)	
Submitting Official Signature	Carlos FM Coimbra	

Acknowledgement:

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy, Solar Energy Technologies Office, under Award Number DE-EE0008216.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

Executive Summary

The objective of this research is to develop a hybrid physics-based/data-driven forecast model to improve direct normal and global horizontal irradiance (DNI and GHI) prediction for horizons ranging from 1 to 72 hours. Project objectives also address key gaps in state-of-the-art solar forecasting: accurate probabilistic solar forecasts and the forecasting of large irradiance ramps (ramp onset and magnitude). The proposed model ensembles Numerical Weather Prediction (NWP) forecasts, determinist physics-based algorithms, and new-generation cloud cover products (high-resolution rapid refresh satellite images and Large Eddy Simulations). The result is the Hybrid Adaptive Input Model Objective Selection (HAIMOS) ensemble model. HAIMOS blends state of the art machine learning methodologies with physics-based models for cloud cover and cloud optical depth forecasts. The technical activities followed a two-pronged strategy. First, the preprocessing of data, the selection of inputs to the nonlinear approximators, the type of approximator and objective functions, and post-processing ensembling techniques included in HAIMOS were all optimized adaptively to find the best model for a specific goal (reduce DNI/GHI forecast error, improve the prediction of ramp onset, etc.). Second, a large effort was put in improving cloud identification and the forecast of cloud cover and cloud optical depth. To this end, new-generation cloud parametrization products were developed in this work. These include improved algorithms to assist in cloud identification, cloud classification and cloud parametrization from satellite images – three key factors in the accuracy of 1 to 6-hours irradiance forecasts and prediction of ramp onset. Furthermore, we also included cloud information extracted from high-resolution rapid refresh satellite images (GOES-16) and Large Eddy Simulations (LES). LES was used to model the atmosphere in detail over locations of interest and produce cloud optical depth forecasts. Once these data streams were validated, they were used as input data to the HAIMOS forecast. The model was developed using data from several climatologically distinct locations with potential for high solar penetration. In the last year of the project, we conducted a validation campaign according to the guidelines stipulated by the Topic Area 1 project as described in the FOA.

This effort brings, for the first time, proven machine-learning methodologies for generating state-of-the-art solar forecasts interweaved with detailed physics-based models for cloud detection, and cloud optical depth forecasts. HAIMOS will generate accurate irradiance probabilistic forecast to assist in reducing solar generation prediction error. Globally optimized solar forecast models are more likely to impact solar energy stakeholders.

The goal of this project was to increase the state-of-the-art forecast skill from their present values of 10 to 35%. At the end of the project, we achieved between 30% and 50% forecast skill across a wide range of horizons for both GHI and DNI.

Contents

Background

Variability in solar output at high levels of penetration creates significant operational challenges for system managers. The objective of this project is to improve the accuracy of solar forecasting in support of more efficient energy markets. This objective enables high solar penetration without significant integration costs or power system performance degradation.

In this project, we developed innovative methods and tools to significantly improve both the certainty and accuracy of solar generation forecasts for utility and transmission system operators across intra-day and day(s)-ahead time horizons. Our team has deep knowledge of both solar forecasting as well as power system operation and economics and is experienced in leading high-profile, public stakeholder processes to build consensus. As a result, the approach developed is well-grounded and directly applicable for implementation by system operators.

The objective of this research is to develop a hybrid physics-based/data-driven forecast model to improve direct normal and global horizontal irradiance (DNI and GHI) prediction for horizons ranging from 1 to 72 hours. Project objectives also address key gaps in state-of-the-art solar forecasting: accurate probabilistic solar forecasts and the forecasting of large irradiance ramps (ramp onset and magnitude). The developed model ensembles Numerical Weather Prediction (NWP) forecasts, determinist physics-based algorithms, and new-generation cloud cover products (high-resolution rapid refresh satellite images and Large Eddy Simulations). The result is the Hybrid Adaptive Input Model Objective Selection (HAIMOS) ensemble model.

HAIMOS blends state of the art machine learning methodologies with physics-based models for cloud cover and cloud optical depth forecasts. The technical activities followed a two-pronged strategy. First, HAIMOS implies that the preprocessing of data, the selection of inputs to the nonlinear approximators, the type of approximator and objective functions, and post-processing ensembling techniques are all optimized adaptively to find the best model for a specific goal (reduce DNI/GHI forecast error, improve the prediction of ramp onset, etc.). Second, a large effort was put in improving cloud identification and the forecast of cloud cover and cloud optical depth. To this end, new-generation cloud parametrization products were developed in this work. These include improved algorithms to assist in cloud identification, cloud classification and cloud parametrization, cloud information extracted from high-resolution rapid refresh satellite images (GOES-16) and Large Eddy Simulations (LES).

Project Objectives

The goal of this project is to increase the state-of-the-art forecast skill from their present values of 10 to 35%. At the end of the project, we aim to achieve the forecast skills between 40 and 50% across a wide range of horizons for both GHI and DNI for periods of variable solar irradiance (i.e., excluding cloudless days).

The technical scope for HAIMOS is divided into six task groups summarized next. The seventh task consists in the validation set up by Topic Area 1 team.

Technical work plan summary

Selection of training sites and data collection (BP1: Task 1): The purpose of this task is to collect endogenous (GHI and DNI) and exogenous data (NPW forecasts, satellite images, etc.) data used in the development of HAIMOS. We selected data from 5 to 10 locations in the USA with potential for high solar penetration with the requirement that they are climatologically distinct. We used data collected at the UCSD solar laboratory, and high-quality, publicly available data such as the one provided by NOAA's SURFRAD network.

Commercial Irradiance Forecasts (BP1: Task 2, BP2: Task 5, BP3: Task 8): In these tasks we collaborated with a Commercial Irradiance Forecast Provider (Clean Power Research), CIFP for short, to obtain historical irradiance forecasts for the locations and time span selected in Task 1. CIFP forecasts cover 1 to 24 hours horizons with a temporal resolution of 30 minutes and a spatial resolution of 1 km. The CIFP also provided UCSD access to an application programming interface (API) for the purpose of obtaining on-demand DNI and GHI forecasts for the locations determined by the Area 1 test framework.

Improvement of cloud cover forecast (BP1: Task 3): The purpose of this task is twofold. First, we improved the accuracy of cloud detection models from satellite imagery. Current models that rely on image segmentation show low accuracy in discriminating between thin and thick clouds, for instance. We explored image processing algorithms to identify and classify clouds from satellite imagery. We used exogenous data such as detailed atmosphere data from NWP models to improve cloud classification. Second, we used physics-based models (i.e., Lagrangian cloud motion, convection-diffusion, WRF-Solar simulations) to improve the cloud cover forecast for time horizons ranging from 1 to 6 hours.

Set up HAIMOS framework (BP1: Task 4): This task consists in setting up the HAIMOS framework for the different forecasting tasks: accurate 1 to 72 hours DNI and GHI point forecasts, accurate probabilistic forecasts, and prediction of large irradiance ramps. The goal for this task consists in developing and testing adaptive training techniques that were tested using data from the previous three tasks. The successful completion of this task consisted in obtaining HAIMOS point forecasts for DNI and GHI 1 to 72 hours in the future, with skill over the smart persistence model, equal to or above 30%.

Optimization of HAIMOS framework (BP2: Task 6): This task consists in a detailed framework to optimize every component of the HAIMOS framework: input preprocessing and selection, machine-learning technique (regressive models, deep learning, nearest-neighbors, etc.), objective function, and post-processing ensembling. The optimality of the resulting model was assessed using data collected in Task 1.

New-generation cloud forecasting tools (BP2: Task 7): In this task we applied the HAIMOS model to new data sets in order to identify and correct problems such as training data overfitting. In this task we also included new-generation exogenous data as inputs to HAIMOS. The goal is to increase HAIMOS accuracy in the 1-to-6-hour horizon by using high-resolution imagery from the new GOES-16 satellite and improve cloud cover forecasts obtained from LES run at UCSD. LES simulations were nested over the areas of interest to simulate the local meteorology and solar forecasts up to a few hours ahead, at resolutions of seconds and 100 m. The WRF-Solar runs from Task 3 provided boundary conditions for LES.

Area 1 test framework validation (BP3: Task 9): The purpose of this task is to apply the HAIMOS forecast model developed in this project to the test framework developed in Area 1. This task entailed collecting the necessary exogenous data for the locations and time period determined by the test framework. HAIMOS was retrained using the endogenous data provided by the test framework and exogenous data produced by this project. In the last stage of this task, HAIMOS provided the forecasts according to the guidelines for model validation (variables, resolution, conditions, measurements, etc.) stipulated by Area 1 test framework.

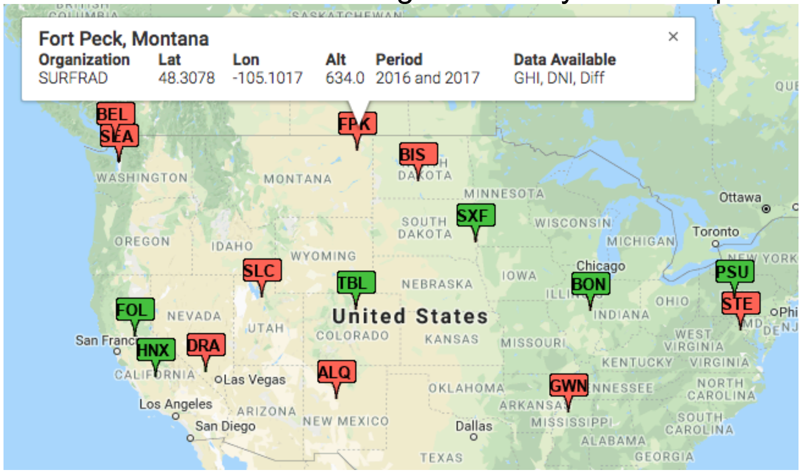
Project Results and Discussion

Selection of training sites and data collection

Forecasting models rely on quality-controlled data for proper training. HAIMOS was developed for GHI and DNI data from diverse locations with potential for high solar penetration. Thus, it was necessary to create a data set of historical irradiance data (DNI and GHI). This section describes the selection data from six locations in the USA with potential for high solar penetration with the requirement that they are climatological distinct. This task generated a quality-controlled database of irradiance and exogenous data that was used in the development of the forecast models.

This task started with the collection of data from several candidate locations in the CONUS. These data were collected from publicly available repositories (SURFRAD[1] and SOLRAD[2]) and from instruments owned and maintained by UCSD. Data from the candidate locations as pre-processed according to the following:

- Removed values flagged (SOLRAD and SURFRAD flag questionable data)
- Removed unrealistic values (negative values)
- Normalized GHI and DNI values using a clear sky model explained below.



Fort Peck, Montana

Organization	Lat	Lon	Alt	Period	Data Available
SURFRAD	48.3078	-105.1017	634.0	2016 and 2017	GHI, DNI, Diff

Figure 1: Location of candidate sites. Irradiance data (GHI and DNI) was collected for each site.

---

[1] https://www.esrl.noaa.gov/gmd/grad/surfrad/

[2] https://www.esrl.noaa.gov/gmd/grad/solrad/

Table 1: List of candidate sites with information about location and data availability.

Location	ID	Organization	Lat	Lon	Alt	Period	Data Available
Desert Rock, NV	DRA	SURFRAD	36.6237	-116.0195	1007	2016 to 2018	GHI, DNI, other
Bondville, Il	BON	SURFRAD	40.0519	-88.3731	230	2016 to 2018	GHI, DNI, other
Table Mountain, CO	TBL	SURFRAD	40.1250	-105.2368	1689	2016 to 2018	GHI, DNI, other
Fort Peck, MT	FPK	SURFRAD	48.3078	-105.1017	634	2016 to 2018	GHI, DNI, other
Sioux Falls, SD	SXF	SURFRAD	43.7340	-96.6233	473	2016 to 2018	GHI, DNI, other
Penn. State, PA	PSU	SURFRAD	40.7201	-77.9309	376	2016 to 2018	GHI, DNI, other
Goodwin Creek, MS	GWN	SURFRAD	34.2547	-89.8729	98	2016 to 2018	GHI, DNI, other
Folsom, CA	FOL	UCSD	38.6434	-121.1477	97	2016 to 2018	GHI, DNI, other
Bellingham, WA	BEL	UCSD	48.7540	-122.4890	7	2016 to 2018	GHI, DNI, other
Albuquerque, NM	ALQ	SOLRAD	35.0380	-106.6221	1617	2016 to 2018	GHI, DNI, other
Bismarck, ND	BIS	SOLRAD	46.7718	-100.7596	503	2016 to 2018	GHI, DNI, other
Hanford, CA	HNX	SOLRAD	36.3136	-119.6316	73	2016 to 2018	GHI, DNI, other
Salt Lake City, UT	SLC	SOLRAD	40.7722	-111.9550	1288	2016 to 2018	GHI, DNI, other
Seattle, WA	SEA	SOLRAD	47.6869	-122.2567	20	2016 to 2018	GHI, DNI, other
Sterling, VA	STE	SOLRAD	38.9767	-77.4869	85	2016 to 2018	GHI, DNI, other

Figure 1 shows the locations of 15 candidate sites, whose data was explored in this task. Table 1 lists the sites and provides additional site information.

Clear-sky model

The clear-sky model returns irradiance for a given geographical location and time when clouds are absent. However, even in this condition the broadband irradiance at ground level is influenced by other environmental conditions such as the aerosol content and precipitable water, primarily, and atmosphere ozone, nitrogen dioxide and site pressure secondarily.

The presence of aerosols and water in the atmosphere can be quantified through the atmospheric turbidity. Several turbidity parameters have been introduced with the most popular being the Linke turbidity factor. Many models have been developed to estimate the clear-sky irradiance and comprehensive comparisons of some of the most popular models can be found in Ineichen (2006); Gueymard (2012). In this work we used the clear-sky model proposed by Ineichen and Perez (Ineichen and Perez, 2002), as it is one of the best performing models, and one of the simplest. It requires as inputs the solar elevation and the Linke turbidity. The solar elevation for a given latitude and longitude is computed from the sun's orbital elements. The Linke turbidity is obtained from the worldwide monthly averaged maps available at Solar Energy Services for Professionals (2004)[3]. Figure 2 compares the GHI and the clear-sky GHI obtained with the Ineichen model for three days in diverse locations where the UCSD team has instruments: Ewa Beach (left) and San Diego (right). This figure demonstrates that the clear-sky model fits the GHI for cloudless periods very well.

---

[3] Solar Energy Services for Professionals, 2004. Climate: Averages, Normals, and Typical Years. http://www.soda-is.com/eng/services/climat_free_eng.php.

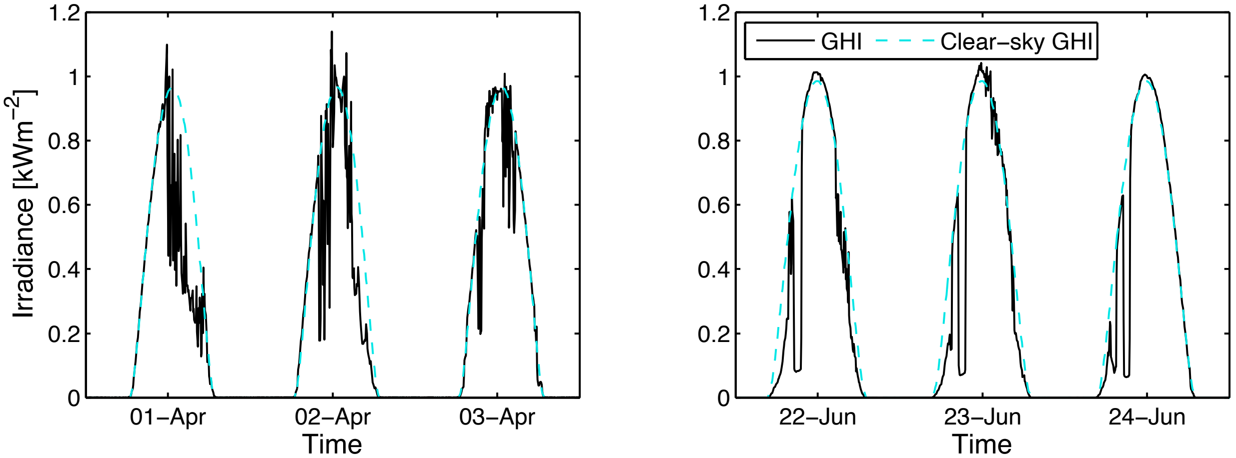


Figure 2: GHI and clear-sky GHI irradiance for a period of three days for Ewa Beach (left) and San Diego (right).

Data variability

Using the clear-sky index, k_G and k_B, for GHI and DNI, respectively, we computed the daily metrics to characterize the irradiance. In this step we computed the daily average

$$\text{AVG}(k_I)_d = \frac{1}{N}\sum_{i=1}^{N} k_{I,i} \text{ for } k_{I,i} \in \text{Day } d$$

and the daily standard deviation

$$\text{STD}(k_I)_d = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N} \left| k_{I,i}\text{-AVG}(k_I)_d \right|^2} \text{ for } k_{I,i} \in \text{Day } d$$

Where k_I is k_G or k_B. This procedure results in a list of pairs of $(\text{AVG}(k_I)_d, \text{STD}(k_I)_d)$ with as many elements as the number of days in the dataset. Figure 3 shows the respective scatter density plots for three of the candidate sites: Folsom, CA (FOL), Table Mountain, CO (TBL), and Seattle, WA (SEA).

These sites were selected since they represent cases of low, intermediate and large irradiance variability. In the figure, lighter and larger dots indicate large concentrations of $(\text{AVG}(k_I)_d, \text{STD}(k_I)_d)$ pairs. This data representation allows to identify the most common daily behavior for GHI and DNI in these locations. For instance, the large concentration of points near (AVG, STD) = (1,0) for Folsom, indicates that the irradiance at this location is dominated by clear-sky days with low variability. This observation applies to both GHI and DNI. The density plots for the other two locations show large densities at lower levels of AVG and higher levels of STD, indicating much more variability in the irradiance. In the case of DNI for Seattle (bottom-right panel in Fig. 2), it is notable that the highest density occurs near (AVG, STD) = (0,0) which reveals that DNI at this location is often and consistently close to 0. Given that it is not straightforward to compare density plots from all the 15 sites, Figure 4 plots the average $\text{AVG}(k_I)_d$ and average $\text{STD}(k_I)_d$ for all locations. These figures condense the information from the density plots into a single point. This figure allows to readily distinguish between low and high variability sites.

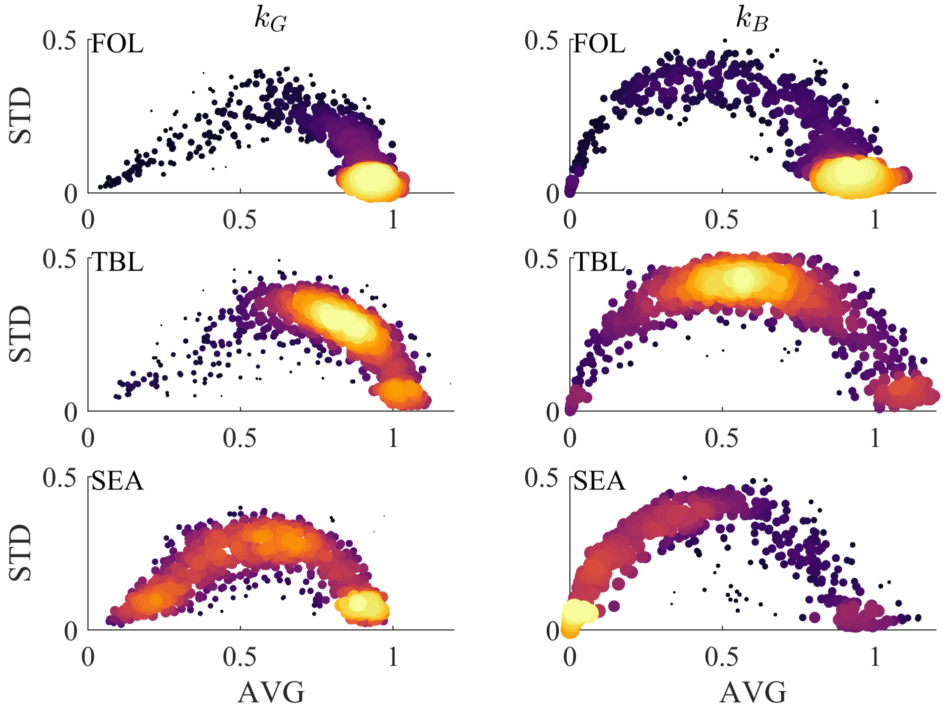


Figure 3: Plots of $(\text{AVG}(k_I)_d, \text{STD}(k_I)_d)$ for three locations in the initial dataset. From top to bottom: Folsom, CA (FOL), Table Mountain, CO (TBL), and Seattle, WA (SEA). The left column shows results for GHI and the right column the results for DNI. Lighter and larger dots indicate large concentrations of $(\text{AVG}(k_I)_d, \text{STD}(k_I)_d)$ pairs.

One of the aspects emphasized by the FOA is the need to quantify the uncertainty in irradiance and the onset, magnitude, and rate of ramps. Ramp characteristics depend on the location's weather patterns and are related to the variability analysis presented above. To compare quantitatively the different sites, hourly ramps were computed as:

$$r_I(t) = \bar{k}_I(t + 1\text{hr}) - \bar{k}_I(t)$$

where $\bar{k}_I(t)$ is the one-hour average of the clear-sky index in the interval $[t - 1\text{hr}, t]$. Using the clear-sky index in this equation limits ramps to deviations with respect to the clear-sky irradiance and removes ramps due to deterministic variations in the solar resource. This equation was applied to the 15 locations and two irradiance components. Figure 5 shows the cumulative density function (CDF) for the ramp magnitude (the absolute value of the equation above). The ramp CDFs for GHI and DNI are shown in the left and right panels, respectively. These plots illustrate clearly that the presence of large ramps varies from location to location. For instance, as the figure indicates, a ramp higher than 0.375 (indicated by the vertical dashed line) has a probability of 0.02 (=1-0.98) for Folsom, CA (FOL) and 0.07 (=1-0.93) for Table Mountain, CO (TBL). As expected, DNI shows large ramps more frequently, with the previous probabilities increased to 0.04 and 0.12, respectively.

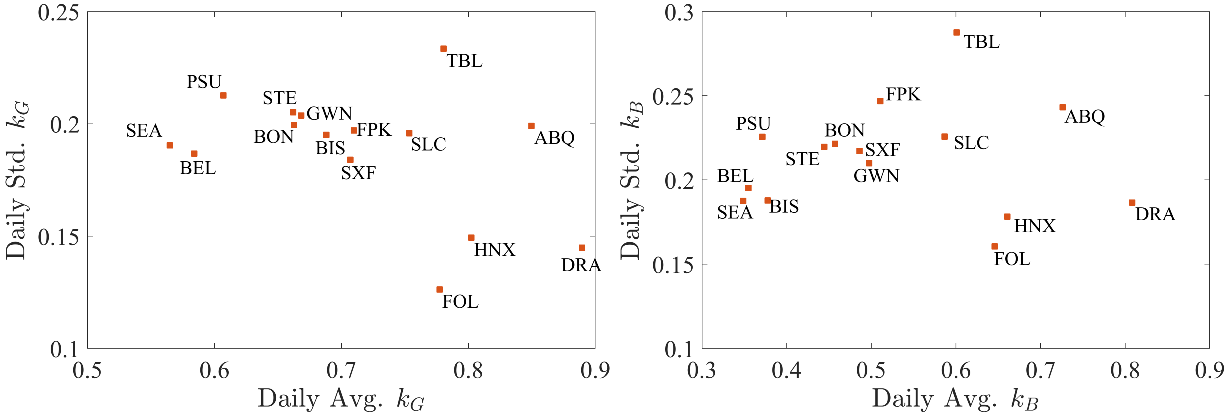


Figure 4: Average for the pairs of (AVG$(k_I)_d$, STD$(k_I)_d$) for all the 15 locations. GHI and DNI data are shown in the left and right panels, respectively.

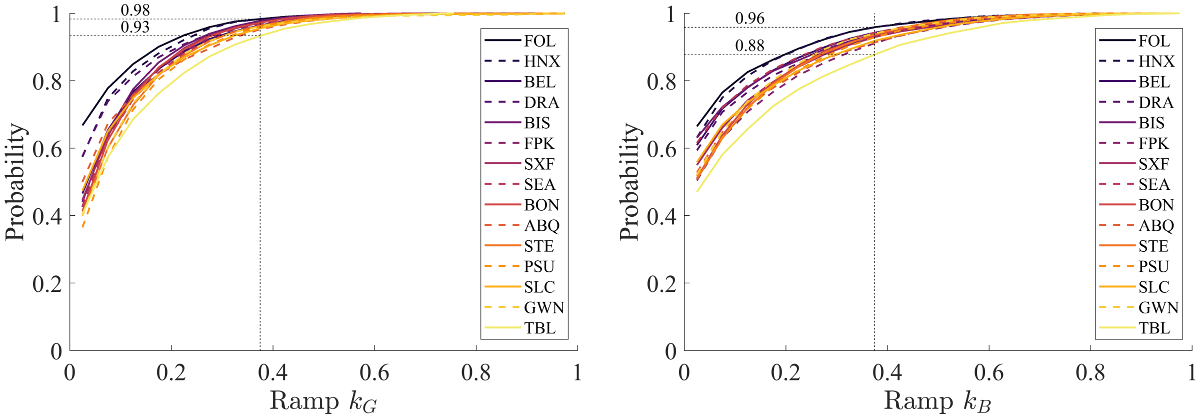


Figure 5: Cumulative density function for ramp distribution for GHI (left) and DNI (right). Each location is identified by a different combination of line color and style.

Again, to facilitate comparisons, the information in Figure 5 is condensed into a single number. In this case, a simple way to quantify the probability of the occurrence of large ramps is by computing

$$\varepsilon_I = 1 - \int_0^1 \mathrm{CDF}_I \, \mathrm{d}r_I$$

This value quantifies the area above the ramp magnitude CDF for each location. The larger the value the higher the probability of large ramps. For this reason, this value is denoted as *ramp density*. Applying this equation to GHI and DNI for all locations results in Figure 6, in which the ramp density for DNI is plotted against the ramp density for GHI for all locations. In this figure, points located towards the top-right corner indicated that large ramps in DNI and GHI are frequent.

An additional, and simple way to assess the irradiance variability consists in studying the persistence forecast. Here, GHI and DHI forecasts are produced according to

$$\hat{I}(t + \Delta) = \bar{k}_I(t) \times I_{cs}(t + \Delta)$$

where $\Delta = \{1, 2, 3, 4, 5, 6\}$ hours, and $\hat{I}$ denotes the irradiance forecast (either GHI or DNI). The forecast error is then analyzed in terms of bias:

$$\text{MBE}_\Delta = \frac{1}{N}\sum_i \hat{I}(t+\Delta) - I(t+\Delta)$$

and magnitude:

$$\text{RMSE}_\Delta = \sqrt{\frac{1}{N}\sum_i \left(\hat{I}(t+\Delta) - I(t+\Delta)\right)^2}$$

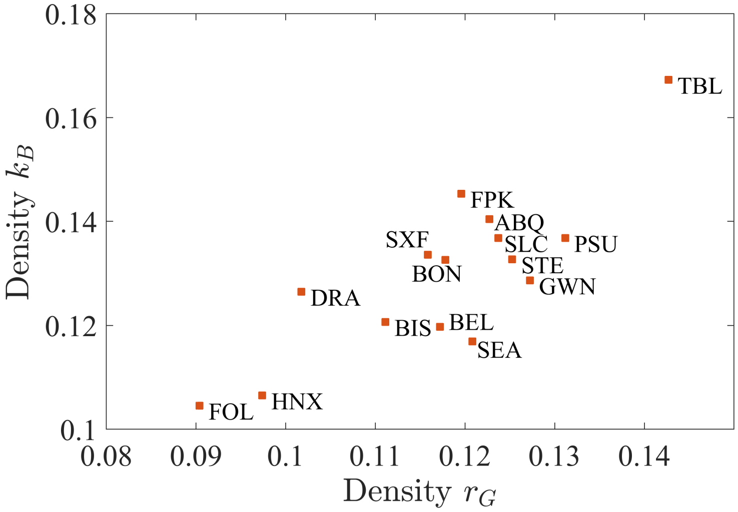


Figure 6: DNI ramp density versus GHI ramp density for all locations.

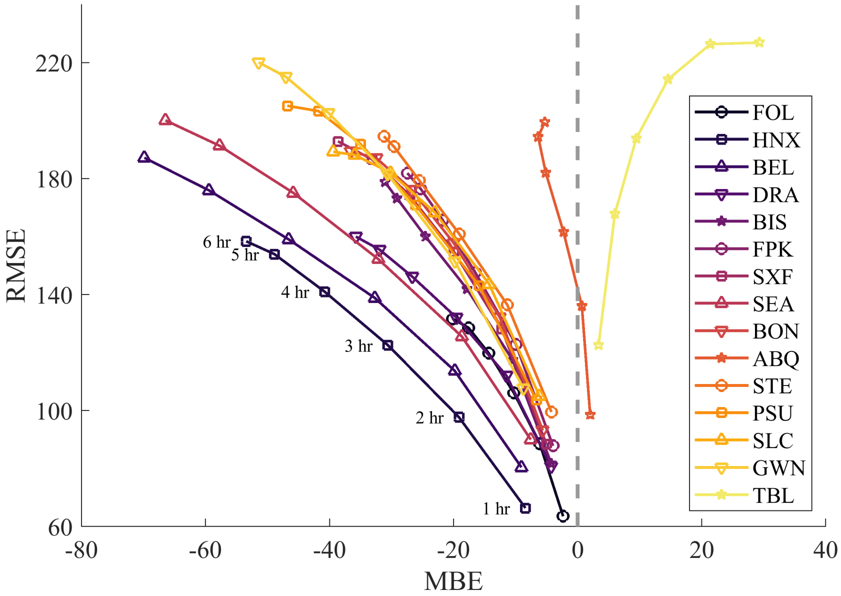


Figure 7: RMSE vs MBE for hourly GHI forecasts 1 to 6 hours ahead of time for the 15 candidate sites. The annotations in the leftmost curve indicate the forecasting horizon.

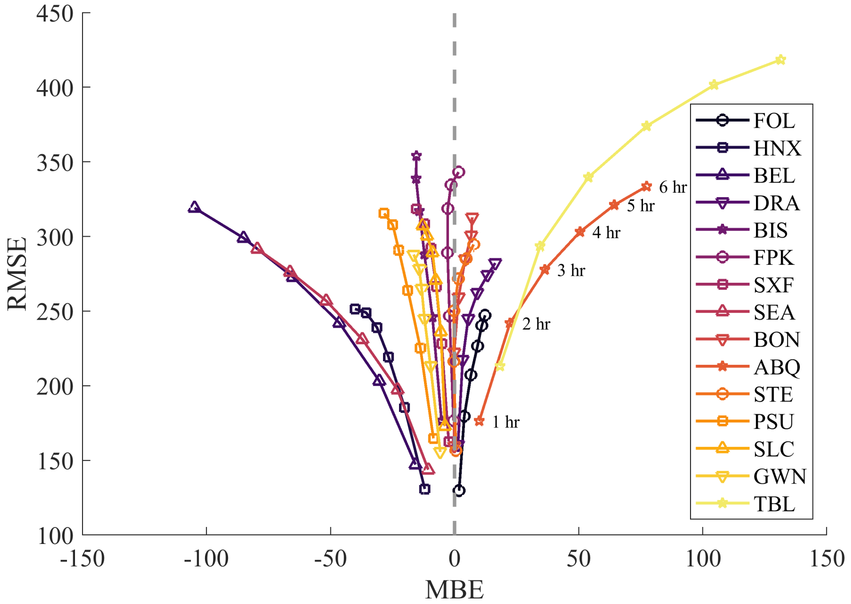


Figure 8: Same as Figure 7 but for DNI forecasts.

Pairs of $(\mathrm{MBE}_\Delta, \mathrm{RMSE}_\Delta)$ are plotted in Figures 7 and 8 for GHI and DNI forecasts, respectively. In these figures, deviations from the dashed gray vertical line indicate forecast error bias. As expected, in all cases the error magnitude increases with forecast horizon (Δ). On the other hand, the bias variation is not consistent across all 15 sites: some sites show a positive bias and others a negative bias with increasing Δ, most likely due to existing biases in the clear-sky models for these locations. Nevertheless, the variability characteristics seen above are imparted in the forecast errors: locations with low variability such as FOL and HNX are clearly “easier” to forecast (i.e., lower persistence forecast error) than locations with large irradiance variability such as TBL and GWN. These results also demonstrate that there is a group of four to six locations with very similar forecasting performances.

In this analysis, GHI and DNI for the 15 candidate sites were analyzed and several quantitative metrics were computed for each location:

1. Cluster centers for $(\mathrm{AVG}(k_I)_d, \mathrm{STD}(k_I)_d)$ pairs.
2. Ramp density
3. MBE_Δ and RMSE_Δ.

Based on these values, 6 sites are selected out of the 15 locations available. The goal in selecting a limited number of sites is to create a dataset of climatologically distinct locations based on the three values listed above. The six sites are listed in Table 2. They are also marked in Fig. 1 by a green marker. The development of HAIMOS was done with data from these locations, primarily. The ground data described above is augmented with publicly available Numerical Weather Prediction (NWP) forecasts produced and published by the National Centers for Environmental Prediction (NCEP)[4].

---

[4] https://www.ncdc.noaa.gov/data-access/model-data/model-datasets/north-american-mesoscale-forecast-system-nam

For this purpose, it was selected the North American Mesoscale Forecast System (NAM) model due to our previous knowledge of this forecasting product. We extracted GHI and cloud cover forecasts for the six selected locations identified above. Figure 9 compares the NAM forecasts against the ground data for three of the six locations. The NAM forecasts for all the six locations are available in the project website. Other forecast outputs, such as wind speeds and barometric pressure, for instance, are also of interest and available to the project team.

Table 2: List of selected sites and the respective selection metrics. All values are unitless except for MBE$_\Delta$ and RMSE$_\Delta$ which are given in Wm^{-2}.

Variability	**Site ID**	**AVG$(k_I)_d$**		**STD$(k_I)_d$**		**ε_I**		**MBE$_\Delta$**		**RMSE$_\Delta$**	
		GHI	DNI	GHI	DNI	GHI	DNI	GHI	DNI	GHI	DNI
Large	TBL	0.78	0.60	0.23	0.29	0.14	0.17	[3, 29]	[18, 131]	[123, 227]	[213, 418]
Large	PSU	0.61	0.37	0.21	0.23	0.13	0.14	[-47, -7]	[-28, -9]	[104, 205]	[165, 315]
Medium	BON	0.66	0.46	0.20	0.22	0.12	0.13	[-37, -6]	[-0, 7]	[93, 189]	[160, 313]
Medium	SXF	0.71	0.49	0.18	0.22	0.12	0.13	[-39, -5]	[-15, -2]	[88, 193]	[163, 318]
Low	FOL	0.78	0.65	0.13	0.16	0.09	0.10	[-20, -2]	[2, 12]	[64, 132]	[130, 247]
Low	HNX	0.80	0.66	0.15	0.18	0.10	0.11	[-53, -8]	[-40, -12]	[66, 158]	[131, 251]

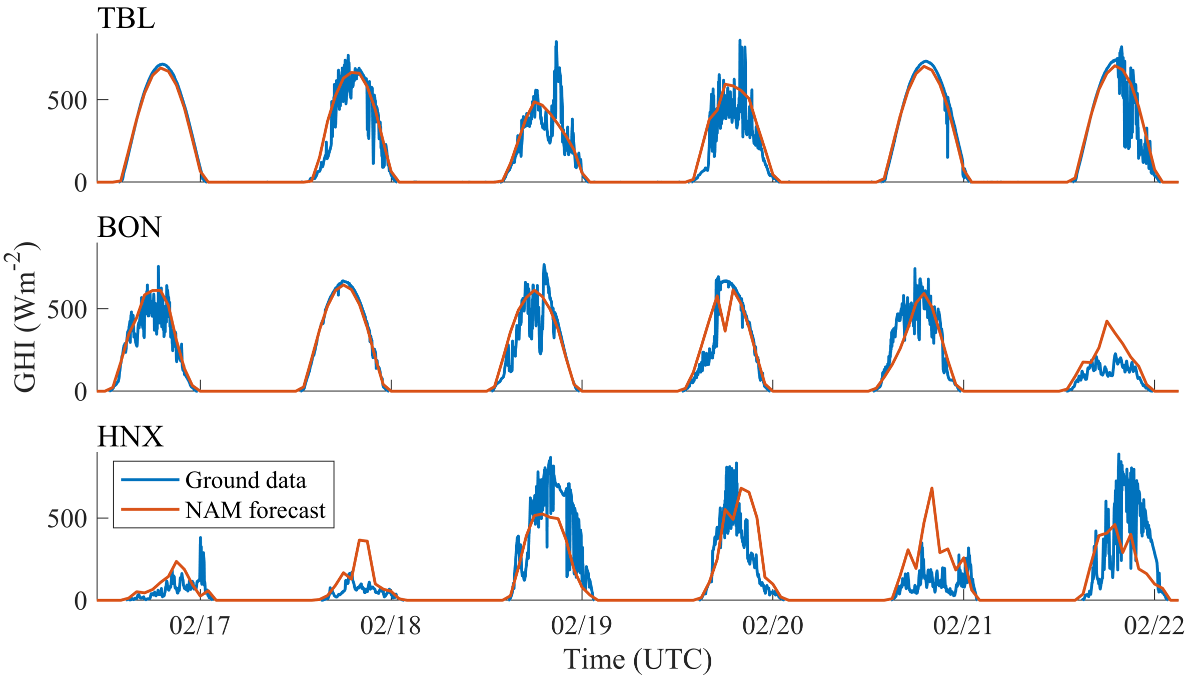


Figure 9: NAM GHI forecasts plotted against ground data for six days in February 2017. The figure shows the comparison for three of the six selected locations (TBL, BON, and HNX).

Improvement of cloud cover identification

For cloud identification, we make use of GOES-16 irradiance products. Initially we explored the longwave channels 7-16 (wavelength > 4 μm). The longwave channels are favored over shortwave channels for the following reasons:

- the longwave irradiance in the atmosphere can be approximated as diffuse, so there is no need to consider the satellite view angles, solar angles and relative angles between the satellite, the sun and the ground stations.
- because of the diffuse nature of longwave irradiance, a two-flux radiative model (Li et. al. 2018) can be used to estimate spectral longwave irradiance in the entire atmosphere at minimal computation cost (takes less than 1 minute to run with a spectral resolution of 0.1 cm^{-1}).
- most surfaces (grassland, desert, ocean) can be approximated as black surface in the longwave spectrum, while their shortwave albedo varies.
- the outgoing shortwave irradiance measures the albedo of the atmosphere-Earth system; therefore, cloud identification becomes hard to distinguish from highly reflective snow/ice.
- the outgoing longwave irradiance also provides information about cloud top temperature (i.e. cloud location) while outgoing shortwave irradiance is independent of cloud temperature.

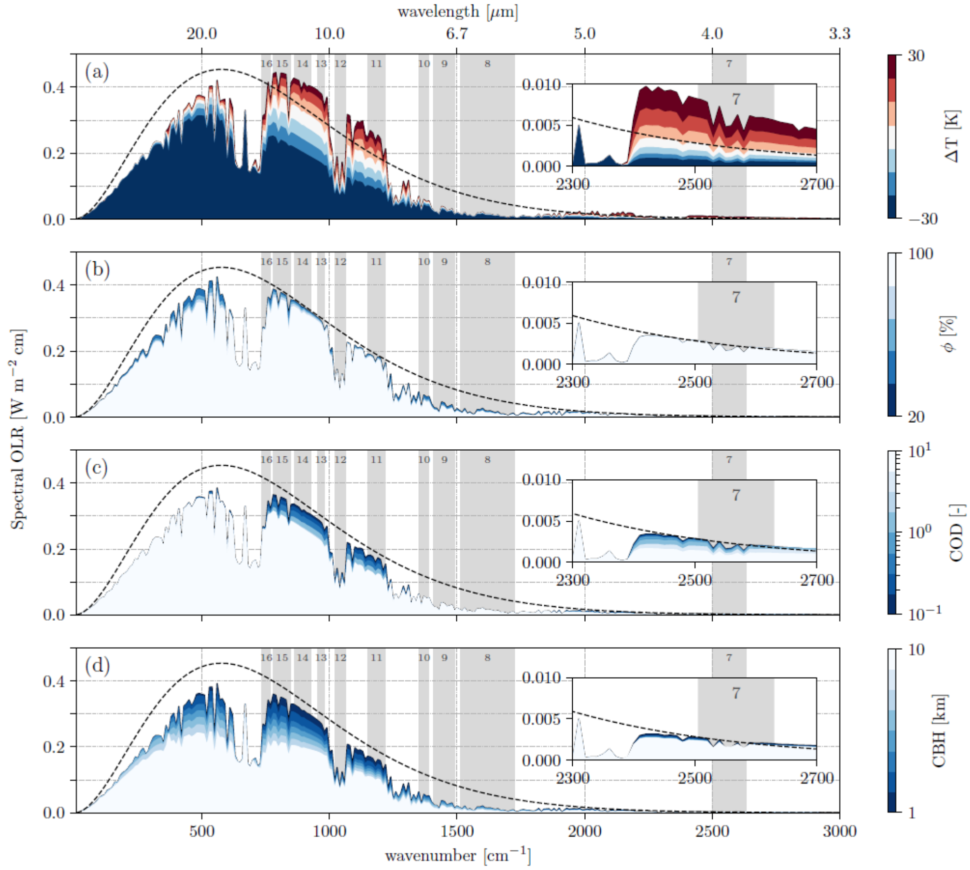


Figure 10: The effects of changing the model parameters on the spectral OLR [W m^{-2} cm] for (a)--(b) clear-sky conditions (COD = 0) and (c)--(d) cloudy conditions (COD > 0).

Figure 10 plots modeled Outgoing Longwave Radiation (OLR) responses with surface temperature, relative humidity, cloud optical depth at 497.5 nm (COD) and cloud base height (CBH). For all plots, the location and bandwidth of the 10 longwave Advanced Baseline Imager (ABI) channels are highlighted, with inset plots added to better show the effects on the spectral range that contains channel 7 (2300--2700 cm^{-1}). Also, the spectral OLR is shown with a resolution of 10 cm^{-1} for improved readability, whereas the results in the rest of the study are based on 0.1 cm^{-1}.
The black dashed line in each subplot shows the blackbody radiation predicted by Planck's distribution at T = 294 K, the surface temperature of the Air Force Geophysics Laboratory (AFGL) midlatitude summer profile.

The figure also shows

a) The effect of changing the surface temperature by ΔT while keeping Φ fixed at 50%.
b) The effect of varying Φ, while keeping T fixed at 294 K (ΔT = 0 K).
c) The effect of varying the COD of a cloud layer (base height: 2.72 km, thickness: 0.5 km, cloud average temperature: 279 K), while keeping T and Φ fixed at 294 K (ΔT = 0 K) and 50%, respectively.
d) The effect of the varying the cloud base height (CBH) of a single cloud layer with COD = 1.0, T = 294 K, and Φ = 50%. Increasing the CBH is equivalent to decreasing the cloud temperature from ~289 K (CBH = 1 km) to ~231 K (CBH = 10 km) for the AFGL midlatitude summer profile.

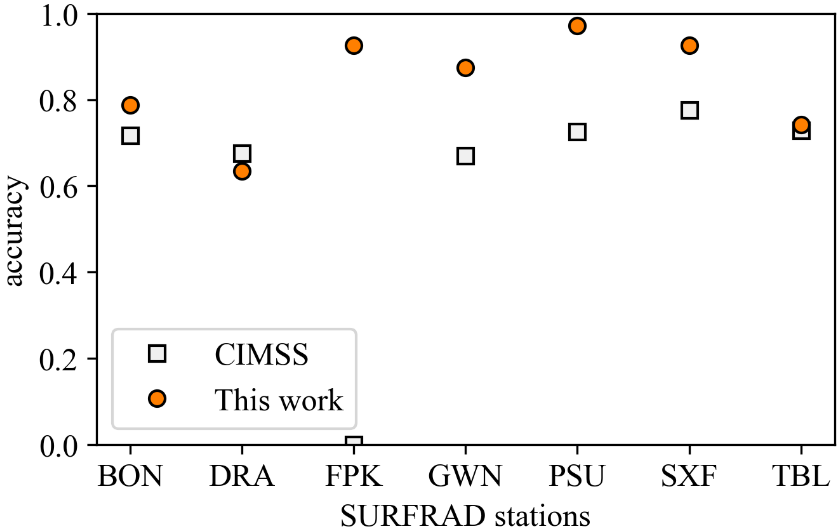


Figure 11: Accuracy of cloud detection of our method and CIMSS cloud mask over 7 SURFRAD stations.

Using GOES-R longwave ABI irradiance per channel, together with the measurement of local air temperature and relative humidity and the radiative model, the cloud optical depth and base height per pixel can be estimated. The pixel is deemed ‘clear’ if the normalized ABI irradiance in channels {11, 13, 14, 15} are within 5% of modeled clear sky OLR. If local irradiance measurements are available, clear sky periods are identified by analyzing GHI and DNI time series, as detailed in (Inman et al. (2015)). Figure 11 shows the accuracy of our satellite-model clear sky detection for seven SURFRAD stations over the entire year of 2018. The ‘ground truth’ clear sky is derived by GHI and DNI time series analysis. For comparison, the accuracy of clear sky detection from the Cooperative Institute for Meteorological Satellite Studies (CIMSS) cloud mask[5] are plotted in Figure 11. The accuracy is defined as:

$$\text{Accuracy} = \frac{(\#\text{ of correctly identified as clear } + \#\text{ of correctly identified as cloudy}}{\#\text{ of data points}}$$

For CIMSS, data from March 12th to April 02th, 2019 are used. For 6 out of 7 stations, our method outperforms CIMSS cloud mask. Note that CIMSS has no data for station FPK.

Improvement of cloud optical depth estimation

Initially we used all 10 longwave ABI channels to select optimal pairs of (COD, CBH). With further analysis, we found that channels 8, 9, 10 do not respond to COD variation, therefore are used to select CBH (see Figure 10). Thus, the COD and CBH are now selected in a two-step manner: (i) selects CBH based on channels {8, 9, 10}; (ii) selects COD using channels {7, 11, 13-15} based on the CBH from (i).

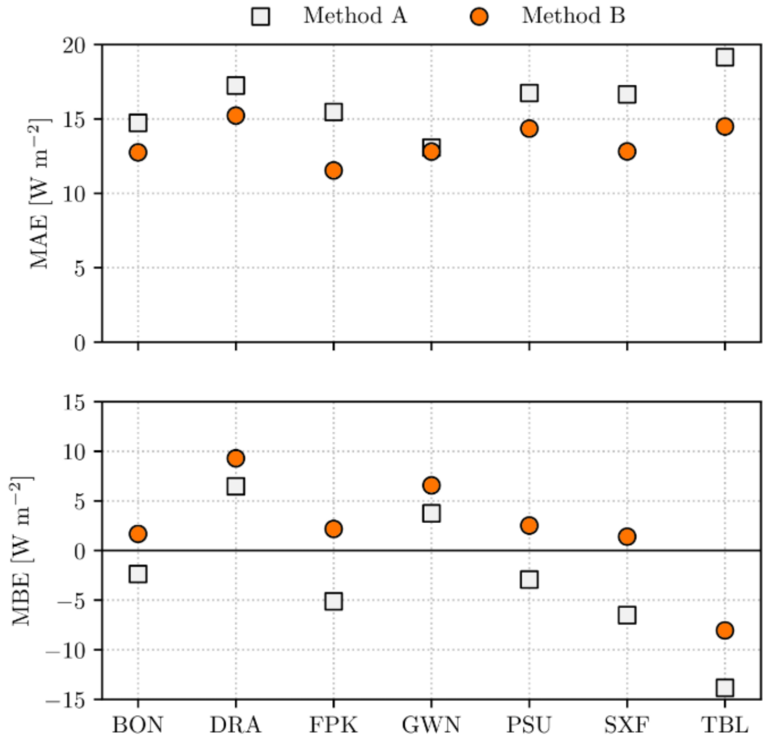


Figure 12: Comparison of DLW prediction error during daytime cloudy periods using two variations of the cloudy property estimation method. Method A: estimate COD and CBH at the same time using all ten LW channels. Method B: first estimate CBH based

---

[5] http://cimss.ssec.wisc.edu/clavrx/

on channels 8--10, then use the estimated CBH value with channels 7, 11, and 13--15 to estimate COD. Positive MBE values correspond to the model over-predicting DLW.

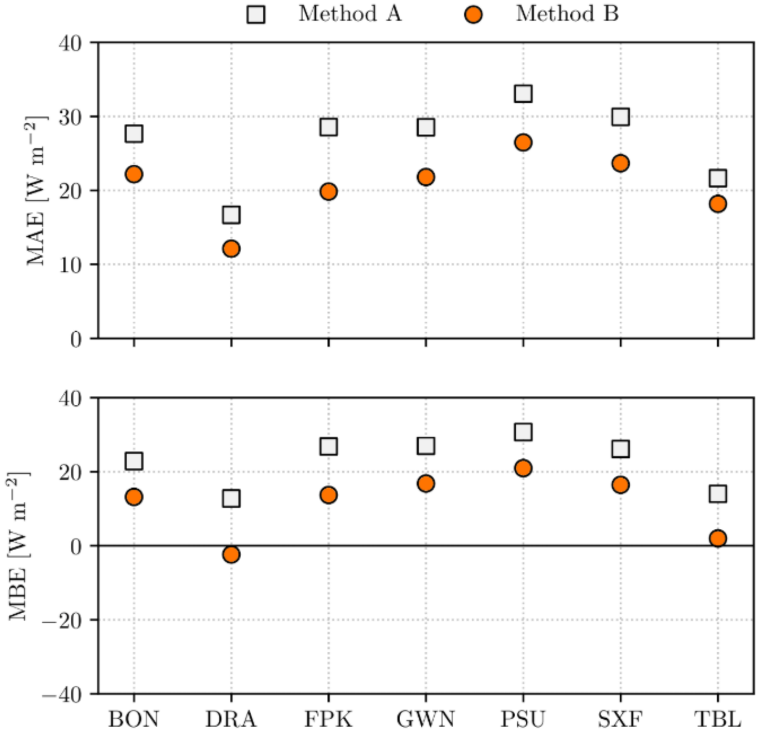


Figure 13: Comparison of DLW prediction error during nighttime without a temperature inversion (Method A) and with a temperature inversion at 1 km (Method B).

Table 3: Summary statistics for the DLW validation during (1) daytime clear, (2) daytime cloudy, and (3) nighttime all-sky conditions. The error metrics are computed against ground measurements of DLW from Eppley PIR instruments, which have measurement uncertainty of ±5 W m^{-2}. For the daytime cloudy and nighttime all-sky cases, the DLW is predicted using COD and CBH values estimated solely from the OLR, with no model output statistics or other corrections applied. Negative MBE values correspond to the model under-predicting DLW, either due to a lower value for COD, a higher value for CBH or a combination of the two.

	MAE [W m^{-2}]			MBE [W m^{-2}]			RMSE [W m^{-2}]		
Site	Clear	Cloudy	Night	Clear	Cloudy	Night	Clear	Cloudy	Night
BON	8.4	12.8	22.2	-5.8	1.7	13.2	10.0	17.1	29.2
DRA	14.7	15.2	12.1	11.1	9.3	-2.4	19.7	20.8	15.4
FPK	6.5	11.5	19.8	-2.9	2.2	13.7	8.2	16.2	26.5
GWN	6.6	12.8	21.8	-3.4	6.6	16.8	7.7	16.8	29.5
PSU	9.5	14.4	26.5	-9.0	2.5	21.0	11.0	19.3	33.1
SXF	7.6	12.8	23.7	-5.4	1.4	16.5	9.1	17.9	31.8
TBL	16.1	14.5	18.2	-16.0	-8.0	2.0	17.6	17.8	23.8

Figure 12 shows that by applying the two-step selection methods, the (COD, CBH) pairs generate smaller error when estimating surface downwelling irradiance (DLW) for all seven SURFRAD stations, indicating more accurate COD and CBH estimations.

To improve solar irradiance forecasts in the early morning, cloud properties before dawn are necessary. Therefore, we expand our method of COD and CBH selection to the nighttime. A temperature profile with inversion at 1 km is used to during the nighttime. Figure 13 shows that by applying the nighttime inversion, the method is more accurately estimating downwelling longwave (DLW) radiation during the night, indicating more accurate COD and CBH estimations.

The DLW estimation errors (as an indicator of the accuracy of COD and CBH estimations) for daytime clear sky periods, daytime cloudy sky periods, daytime all sky periods (without prior knowledge of knowing whether the sky is clear) and nighttime all sky periods are shown in Table 3. The relative error is smaller than 10% for all cases.

Set up HAIMOS framework

A key aspect in improving the forecast skill and especially the prediction during large variability periods resides in the selection of the proper inputs available in the search space. In proposal we tested input selection based on techniques such as clustering analysis and time series correlation analysis.

The input data available to HAIMOS was collected from UCSD resources, CIFP, irradiance networks (SURFRAD and SOLRAD), and other publicly available sources.

Table 4: Data used in for HAIMOS input selection

Data	Description
Measured data	Irradiance data measured
Modeled irradiance	Satellite-derived irradiance data for the target locations and neighboring nodes (49 in total)
CIF_i	Forecasted irradiance from the CIFP. Several forecasts are available, denoted by the subscript *i*, depending on data used (e.g., HRRR).
NAM	GHI from the NAM model NAM
NAMcc	Total Cloud cover from NWP model NAM

The data-driven nature on HAIMOS explores all these inputs to obtain a more accurate irradiance forecast with focus on periods of large ramps and variability.

Clustering:

HAIMOS seeks to optimize the input selection depending on the current and future irradiance variability. Thus, clusters are formed based on the average and standard deviation of the clear-sky index k_t at the forecast issuing time t, and the average and standard deviation of the clear-sky index at the future time $t + \Delta t$. Values at time t are defined using the measured data. At time $t + \Delta t$, there are several options from the data listed in Table 4. In this work, we opted to used NAM forecast given that these are publicly available.

Using this strategy, three clusters are defined. In general, the resulting three clusters are characterized by the following weather conditions: (i) cloudy periods (low irradiance), (2) partially cloudy periods (medium irradiance), and (iii) clear-sky (high irradiance).

HAIMOS input selection

The HAIMOS optimization algorithm selects the best set of inputs for each cluster and forecast horizon with the goal to minimize both the forecasting error bias and variance.

The algorithm first tests adding all available inputs, then ranks the input addition in terms of bias-variance metrics (MAE, MBE, BMSE). The best input is the one that ranks highest in all the metrics. The algorithm proceeds then to test the addition of the remaining, unselected inputs until no improvements are observed in the error metrics.

Results

In this section we present a representative subset of the results obtained in this task for GHI forecasting. The results shown below correspond to the locations of Bondville (BON) and Penn. State University (PSU). These locations are characterized by medium and large irradiance variability, respectively, as documented previously. Furthermore, this task concerns the anticipation of periods of large variability, thus it is focused on forecast horizons ranging from 1 to 5 hours ahead of time.

Figures 14 and 15 compare HAIMOS normalized MBE and RMSE against the competing models. The MAE comparison and improvement is very similar to RMSE, thus omitted from this report. Besides CIF and NAM forecasts we also include in the comparison the persistence model (Pers.). The bottom panel in the figures indicates HAIMOS performance relative to the best of the other models for a given horizon (indicated in the x-axis). In general, HAIMOS reduces bias and variance errors, the only exception happens for MBE, where at times HAIMOS is worse than one or more competing models (see middle figure in the bottom panels). This issue impacts little the forecast accuracy since the bias error is low, as evidenced by the figures in the top panels.

Also, as expected, forecasts error metrics for all models are higher for PSU than BON. A clear indication of the higher level of irradiance variability at this location.

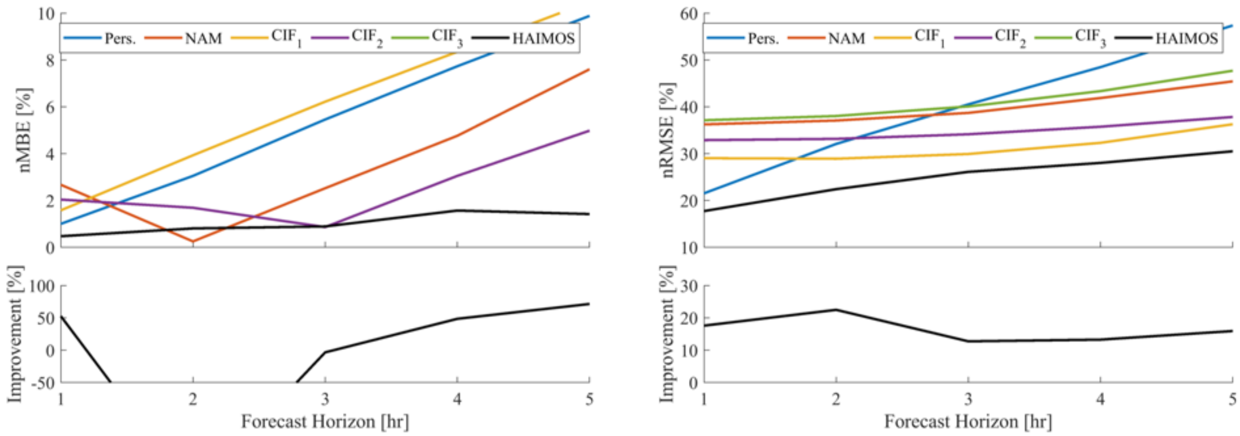


Figure 14: Normalized MBE (left), and RMSE (right) for the validation testing set for BON as a function of the forecasting horizon. HAIMOS forecast is in black and competing forecasts are in color. The top panel shown the error metric and the bottom panel shows HAIMOS improvement relative for the best competing model the horizon in the x-axis. Error metrics are normalized by the average irradiance in the validation set.

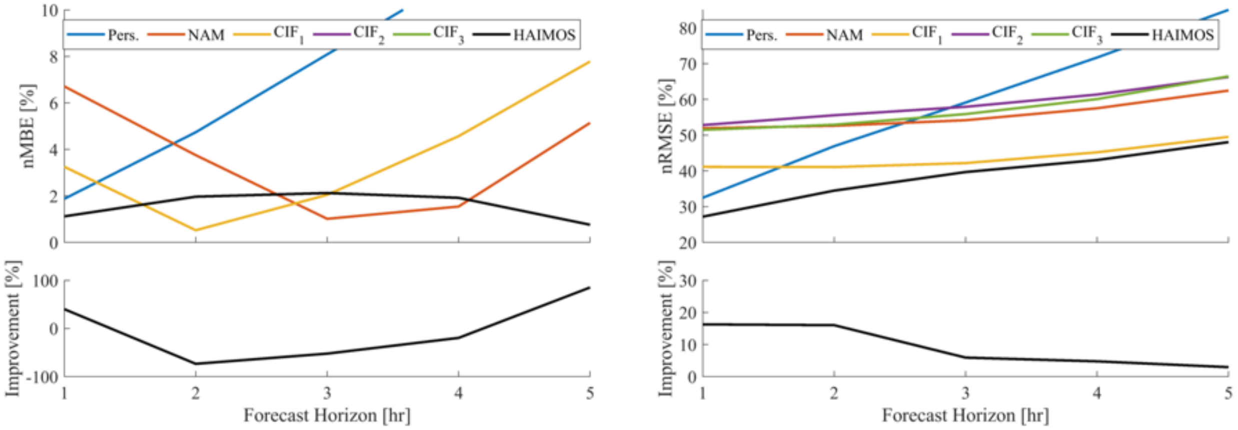


Figure 15: Same as Fig. 14 but for PSU.

Figures 16 and 17 indicate the inputs selected for the HAIMOS used to obtain the results shown above, for BON and PSU, respectively. For each forecast horizon identified by the label on the left-hand side the figure indicate which inputs are selected per cluster. The background colors indicate the origin of the inputs, and the color of selected inputs varies from black (most important) to white (least important).

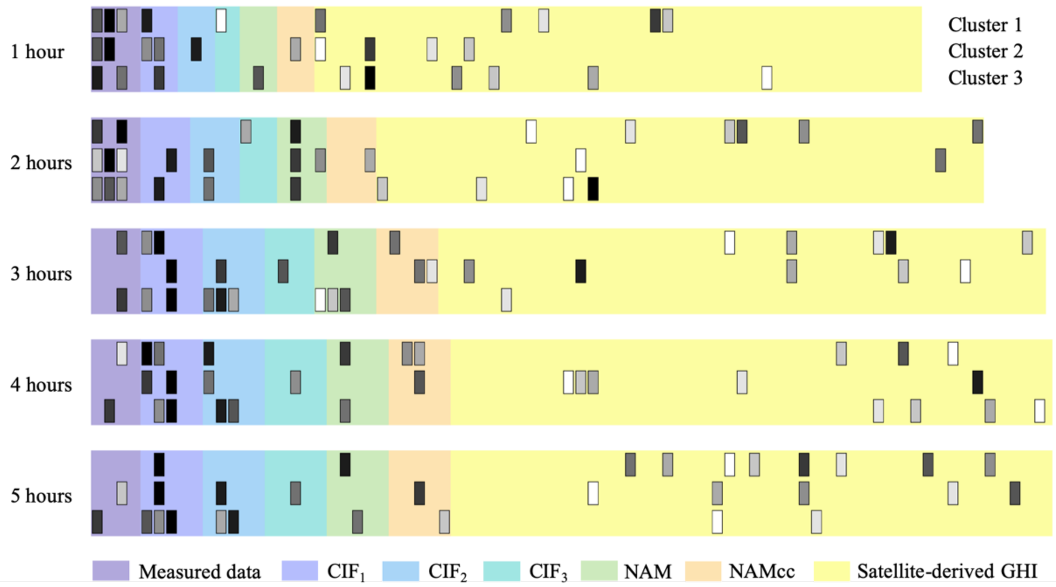


Figure 16: Input selection as a function of cluster and forecasting horizon for BON. The horizon is indicated in the left. The colored background indicates the provenance of the input, and the gray scaled rectangles indicate that the input was selected by HAIMOS.

Darker gray color indicates higher importance, form black (most important) to white (least important).

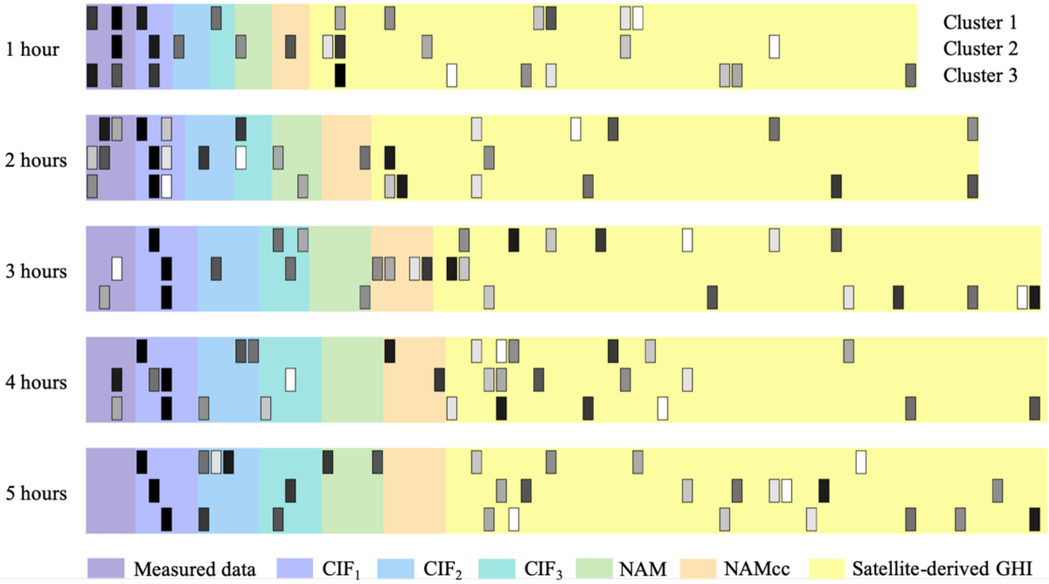


Figure 17: Same as Fig. 16 but for PSU.

A few insights can be discerned from these results:

1. As time horizons increase the importance of measured data decreases.
2. From the CIFP forecasts CIF$_1$ is clearly the most important and reliable input (it's also the one with the lowest error as shown in figures 14 and 15).
3. Spatially resolved data (Satellite-derived GHI) becomes more important for longer horizons.
4. For locations with higher irradiance variability (PSU in this case – Fig. 17) the satellite-derived GHI is selected more often than for BON.
5. There is no obvious pattern in the input selection with respect to the clusters.

These insights and the input selection algorithm were applied in the subsequent development of HAIMOS. One area that requires further study is the effect of clustering in the input selection. The fact that no patterns emerge in the input selection as a function the data variability associated to the clusters, suggest that new clustering approaches could lead to better forecasts.

As mentioned above, in the following iterations we augmented the input search space with new data. These will include data from subtasks 3.1 and 3.2 as reported above, as well as other data such as WRF-Solar (subtask 3.3).

Training and optimization of machine-learning models for HAIMOS

In this task we identified the best type of machine-learning (ML) approximator to map the input data into the target irradiance data. We explored mainly stochastic models as they offer many advantages over purely deterministic models. Algorithms such as deep machine learning, of which artificial neural networks (ANN) are a subset, k-nearest-neighbors (kNN), support vector regression (SVR) and non-linear least squares were included in the pool of model candidates.

Initially, these two subtasks were addressed separately. In that strategy, input selection and ML algorithm selection are done independently and sequentially. If input selection is done in the first place, that requires the use of a reference ML algorithm (kNN in this case), which may result in an input selection optimized for that reference algorithm. Thus, we opted to tackle the two subtasks simultaneously via the use of a genetic algorithm (GA). The GA is also used to optimize the clustering (number of clusters and clustering variables).

The GA minimizes the problem:

$$\underset{x}{\operatorname{argmin}} \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(I_i-\hat{I}(x)_i\right)^2}$$

where I_i and $\hat{I}(x)_i$ denote measured and forecasted irradiance for a training data set, respectively. The argument x is a vector that describes a specific HAIMOS instance. The GA operates over x to return the lowest forecasting error as indicated in the previous equation. The elements of x control:

1. The variables used in the clustering. The pool of variables used to define clusters include all the data collected in subtask 4.1. GA chooses which of those to include in the clustering along with the number of clusters.
2. The ML algorithm to be used to produce $\hat{I}(x)_i$.

Once clusters and ML algorithm are defined, the input selection is performed with a greedy search algorithm. The algorithm is initialized by producing a forecast for all the available inputs, separately. Then ranks the inputs in terms of bias-variance metrics (RMSE, MBE). The best input is the one that ranks highest in all the metrics. The algorithm proceeds then to test the addition of the remaining, unselected inputs until no improvements are observed in the error metrics.

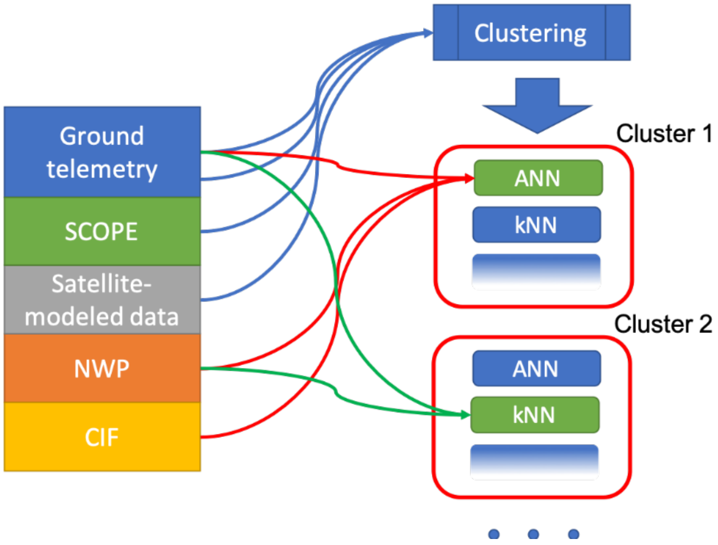


Figure 18: Schematic o the optimization framework used in this work. The blank ellipses in the figure indicate that additional clusters may be defined.

Figure 18 illustrate the GA optimization. The blue arrows indicate the input selection for clustering from the variables shown in the left (grouped by data origin). Then several clusters are produced and for each cluster the GA selects a specific ML algorithm (indicated in green in Fig. 18). The greedy input selection then selects which variables to include, indicated by the green and read arrows in this illustration.

The algorithm outlined above is deployed to every location for all the horizons and for GHI and DNI separately. Figure 19 shows a typical convergence plot for a GA run, in this case the run for GHI from TBL for the 1-hour horizon.

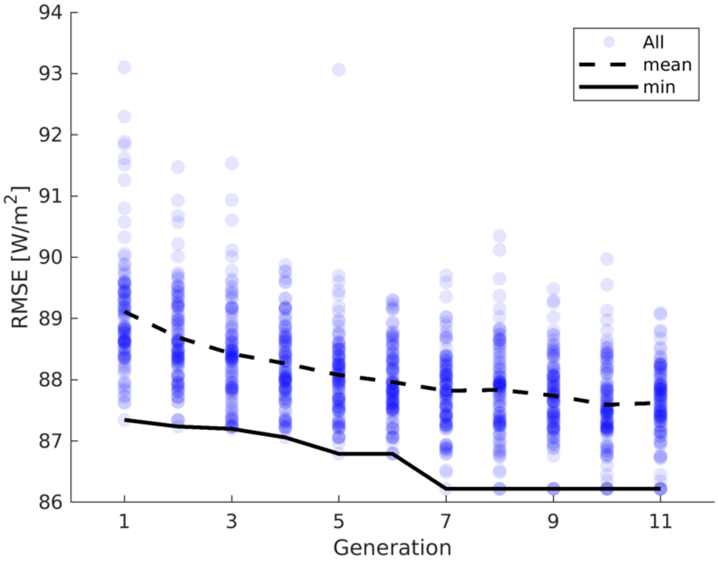


Figure 19: Convergence of the GA optimization for the HAIMOS GHI model for TBL for the 1-hour horizon and for the first 11 generations. The scatter dots represent all HAIMOS configurations analyzed, the dashed line the average fitness (RMSE) across a generation and the solid line the best fitness for the respective generation.

Results for HAIMOS optimization using GA

The framework described above was applied to data from the six selected locations with hourly data from Jan 2016 to June 2017 (1.5 years). The forecasting accuracy is computed with an independent testing data set that ranges from July 2017 to Dec 2018. HAIMOS' performance is compared against a baseline smart persistence model

$$\hat{I}_{sp}(t+\Delta t) = k_t(t)\, I_{cs}(t+\Delta t)$$

where k_t is the clear-sky index at forecasting issuing time t, $I_{cs}(t+\Delta t)$ is the clear sky irradiance (DNI or GHI) at time $(t+\Delta t)$ and $\hat{I}_{sp}$ the resulting forecasted value. The forecasting performance is evaluated in terms of bulk error metrics

$$\text{nMBE} = \frac{1}{N}\sum_{i=1}^{N}\left(I_i - \hat{I}_i\right)/\bar{I}$$

and forecasting skill

$$\text{s} = 1 - \frac{RMSE\left(\hat{\mathrm{I}}_{\mathrm{HAIMOS}}\right)}{RMSE\left(\hat{\mathrm{I}}_{\mathrm{sp}}\right)}$$

where $\bar{I}$ denotes the average irradiance. The error metrics are computed for daytime only, defined in terms of solar zenith angle $\theta_z(t + \Delta t) < 85°$.

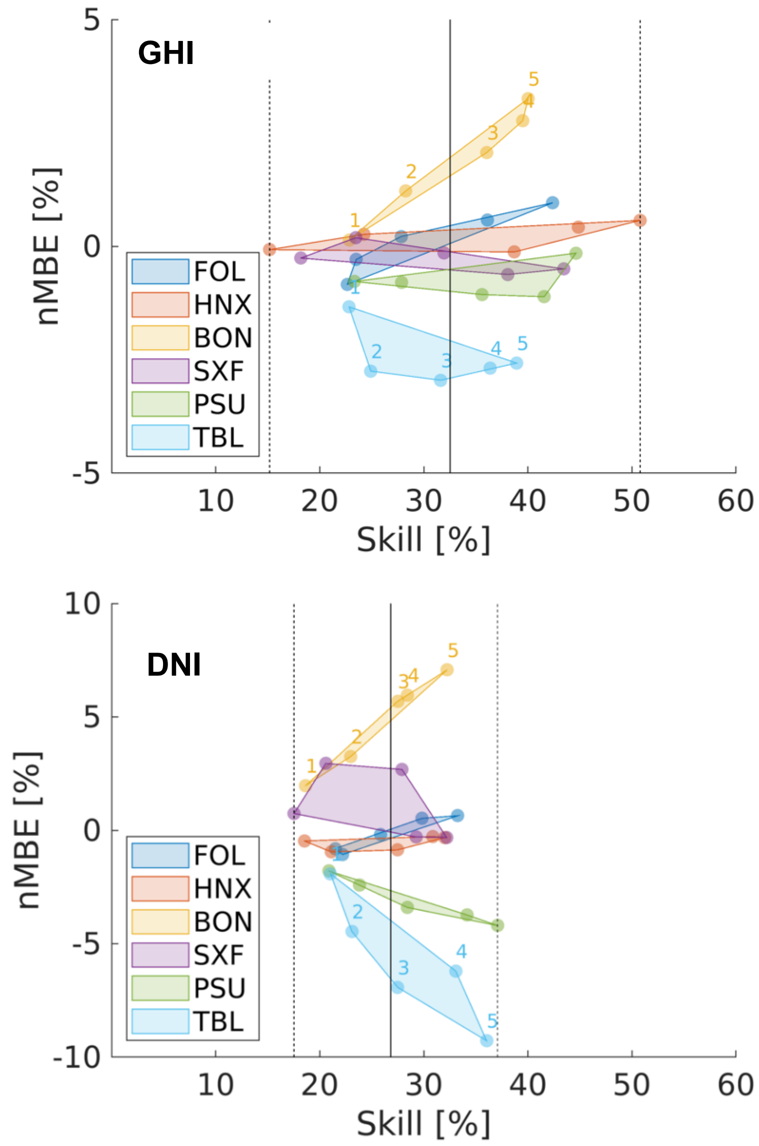


Figure 20: $nMBE$ and forecast skill for GHI (top) and DNI (bottom) for the six locations and forecast horizons ranging from 1 to 5 hours. The minimum and maximum skill values for all locations and horizons are indicated by the vertical dashed lines. The average skill across location and horizon is indicated by the vertical solid line.

Figure 20 summarizes the forecast $nMBE$ and s for GHI (top) and DNI (bottom). Results are shown for the six locations and five forecast horizons $\Delta t = \{1,2,3,4,5\}$ hours. The pairs of $(nMBE, s)$ for each location and different Δt values are grouped by the shaded patches. The horizons Δt are identified explicitly for TBL and BON by the numeral 1 to 5. For clarity these annotations are omitted for other locations, but their placement would follow the same trend as shown for BON and TBL. The vertical dashed lines indicate the minimum and maximum skills attained. The vertical solid line indicates the average skill across all locations and horizons.

Given that this project focuses on improving forecasting for periods of larger irradiance variability, the same error metrics are computed for a data subset representative of cloudy weather. For this analysis, we consider the simple filter

$$\begin{cases} \text{cloudy if } k_t(t) < 0.9 \; and \; k_t(t + \Delta t) < 0.9 \\ \\ \text{clear otherwise} \end{cases}$$

Figure 21 shows the resulting error metric after removing all clear periods from the testing dataset.

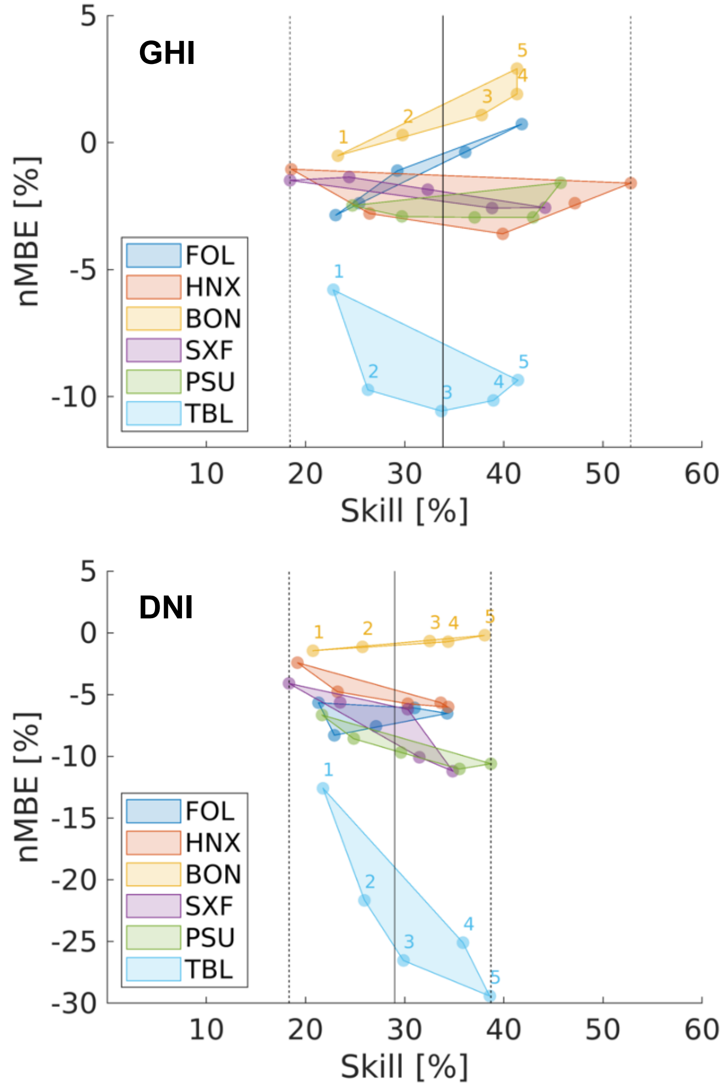


Figure 21: Same as Fig. 20 but for cloudy periods.

The locations listed in the figures' legend are sorted from top to bottom according to the irradiance variability, with FOL as the lowest and TBL the highest. The variability analysis is based on the ground data, and it was presented previously.

From figures we can conclude the following general aspects about the HAIMOS forecasts:

- On average HAIMOS can achieve a forecasting skill ~30% for both GHI and DNI across different climate zones and forecasting horizons. For GHI, the skill values range from just under 20% for 1-hour forecasts to over 40% for the 4- and 5-hour ahead forecasts. In the case of DNI the larger skills are near 40%.
- As expected, the forecasting skill increases when considering only cloudy periods.

- The small forecasting bias increases two-fold or more when considering only cloudy periods.
- The lowest skills are obtained for the shortest horizons. This is expected since those are the horizons for which smart persistence works best, and for which other types of telemetry not used in this work are best suited
- HAIMOS prediction skill is not strongly dependent on the level of variability specific to each location. This is indicated by the overlapping of the projections of the colored patches upon the x-axis.
- For cloudy periods (Fig. 21), HAIMOS forecasts tend to overpredict the irradiance (denoted by a negative $nMBE$). This fact is more pronounced in the case of DNI.
- The impact of variability is more noticeable on the forecast bias. Figure 21 shows that, in general, forecasting for locations with higher variability (PSU and TBL) is more biased than the forecasting for other locations.

The figures also allow for some additional location-dependent findings:

- HAIMOS' GHI forecasting skill for HNX (a location with low irradiance variability) covers a wider range, almost twice as wide as the range of other locations. The same does not occur in the case of the DNI forecasts.
- Bias for TBL is much larger than the bias for any other location: twice as much for GHI and up to three times for DNI.

Developing adaptive protocols for HAIMOS

In this task we developed adaptive protocols for HAIMOS as a function of the forecast goal (GHI or DNI) and the forecast horizon. We tested two approaches in this task:

1. Continuous training in which the ML models identified by the GA optimization are trained daily using data from the training set plus all new data from the testing dataset (up to the updating time).
2. Analog training in which the ML models are retrained as in 1, however, the training is preceded by a data selection process that keeps only data in the historical record that is analogous to current data.

From the two approaches it was concluded that the first one yielded more robust results given that the conditions for the preceding day(s) are typically very well correlated with the present day. The results are shown in Fig. 22 for cloudy periods only.

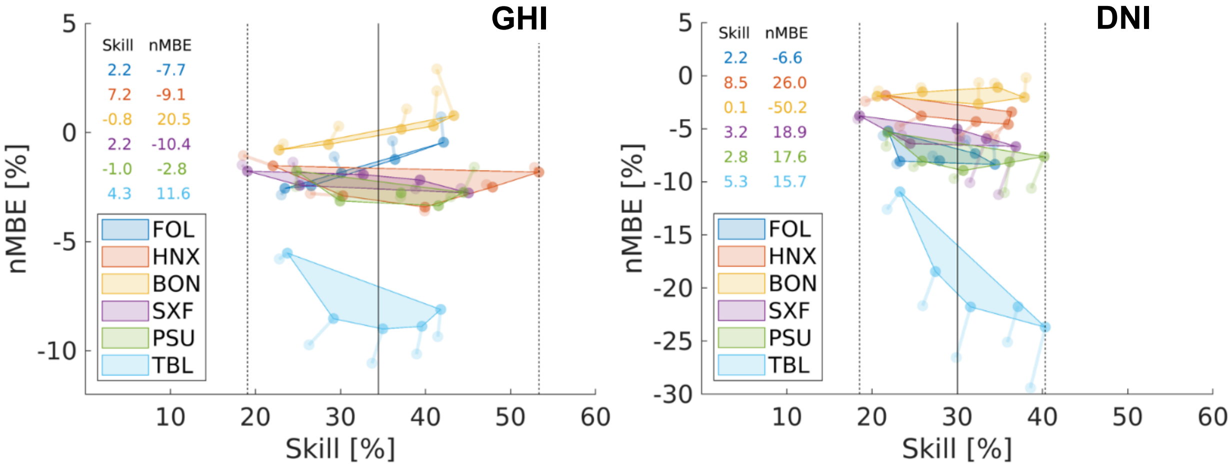


Figure 22: $nMBE$ and forecast skill for GHI (top) and DNI (bottom) for the six locations and forecast horizons ranging from 1 to 5 hours. These results are for the adaptive HAIMOS forecasts for cloudy periods. For both DNI and GHI, the mean value of the forecast skills is on or above 30%, a result never reported before for these time horizons and for a wide range of solar microclimates.

The figure shows the non-adaptive HAIMOS results (the ones from Fig. 21) as light-colored dots. The new skill and $nMBE$ are shown as darker dots. The (non-adaptive, adaptive) pairs are connected by a light line. Pairs that show a trajectory towards the ideal point (0,100), that is 0 $nMBE$ and 100% skill, indicate that the adaptive algorithm improved the HAIMOS predictions. The annotations in the top-left corner of the figures indicate the change (in percentage) over the non-adaptive results for each location averaged across all forecast horizons. A positive value indicates an improvement (higher skill or lower $nMBE$). Negative values indicate worse predictions with respect to those metrics.

From the two plots it is possible to draw the following conclusions:

- The adaptive framework implemented was more successful in improving the DNI forecasts than GHI forecasts. For DNI, skills improved between 0.1% for BON to 8.5% for HNX; $nMBE$ shows changes between -50% and 26%. Note that relative changes for $nMBE$ can show large numbers since the denominator is close to 0.
- For GHI the adaptive HAIMOS shows improvements in both metrics for TBL (4.3% increase in skill and 11.6% reduction in bias). For SXF it results in worse skill and $nMBE$, although by small margins. For the other locations it improves one of the metrics at the expense of the other. For example, for BON, bias improved by 20% but skill is also reduced by 0.8%.
- The largest improvement is observed for TBL in terms of bias as denoted by large positive displacement in the y-axis.

Optimization of HAIMOS framework

Virtually all solar forecasting models are optimized to minimize some bulk error metric for the irradiance over a large training set. That is a poor approach when optimizing models to predict ramps since in a large training set large ramps comprise a very small subset and have little weight in the fitness of the forecast model. For this reason, we will test in this project model optimizations subject to new objective functions that penalize the model's fitness as function of ramps. Thus, we advanced this task by:

1. Setting up the framework for ramp forecasting evaluation.
2. Evaluating the ramp error for HAIMOS and benchmarking models.

In this task we follow the recommendation for ramp forecasting evaluation set up by the team in Topic Area 1[6] where irradiance ramps are defined as

$$r(t,n,\Delta t) = \left|\frac{I(t+n\Delta t) \; - I(t- \; (n-1)\Delta t)}{\Delta t}\right|$$

where I is the measured irradiance, t the forecast issuing time and $n\Delta t$ the forecasting horizon, given in multiples of the forecasting resolution (Δt). The predicted ramps are computed the same way but replacing I by $\hat{I}$ (the forecasted irradiance). Figure 23 shows measured and forecasted ramps in a graphical form.

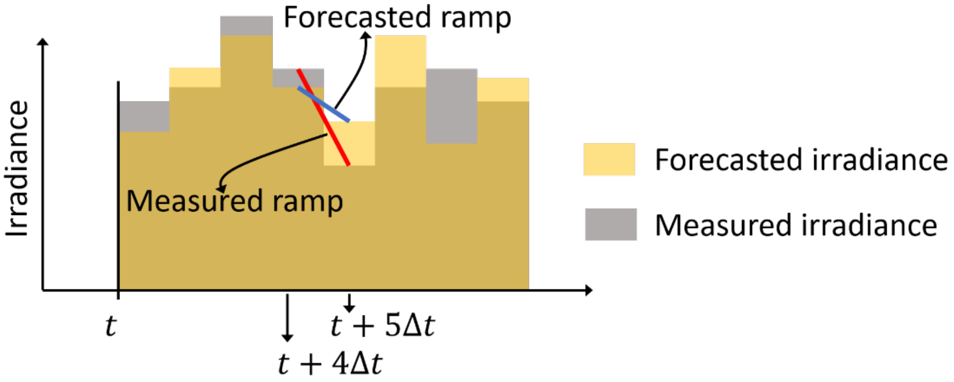


Figure 23: Irradiance ramp definitions.

We assessed the HAIMOS forecasting ability in predicting ramps and compared it against the smart persistence model. That was done by computing the predicted HAIMOS ramps:

$$r_{HAIMOS} = \left|\frac{\hat{I}_{HAIMOS}(t+n\Delta t) \; - \hat{I}_{HAIMOS}(t- \; (n-1)\Delta t)}{\Delta t}\right|$$

and smart persistence (SP) ramps

$$r_{SP} = \left|\frac{\hat{I}_{SP}(t+n\Delta t) \; - \hat{I}_{SP}(t- \; (n-1)\Delta t)}{\Delta t}\right|$$

The models are then evaluated in terms of bias and variance by computing the MBE and RMSE, respectively. Finally, HAIMOS forecasts are compared to the SP model via forecasting skill for ramps

$$s = 1 - \frac{\text{RMSE}(r, r_{HAIMOS})}{\text{RMSE}(r, r_{SP})}$$

---

[6] https://solarforecastarbiter.org/metrics/

The results for this analysis are shown in Fig. 24 for GHI (left) and DNI (right) ramps.

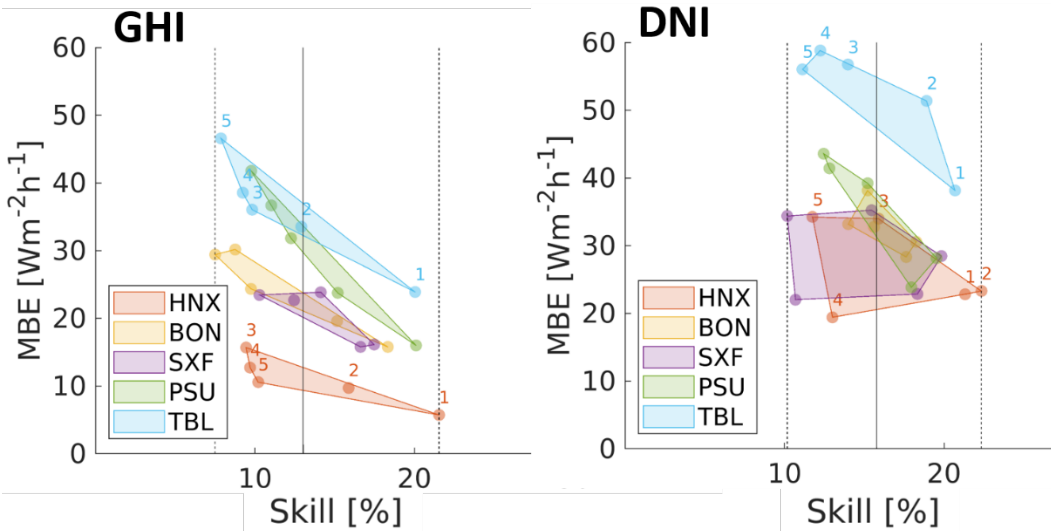


Figure 24: MBE and forecast skill for GHI (left) and DNI (right) ramps for five locations and forecast horizons ranging from 1 to 5 hours. The minimum and maximum skill values for all locations and horizons are indicated by the vertical dashed lines. The average skill across location and horizon is indicated by the vertical solid line.

From these results it's possible to observe that:

- Forecast skills vary between 10 and 20% for all locations studied.
- Higher skills are observed for shorter horizons. This behavior is the opposite of what was observed for the irradiance forecast, as expected given that the reference forecast for irradiance (persistence) decays faster for longer horizons.
- Bias across locations varies as expected with larger errors for locations with more variable weather patterns (e.g. PSU and TBL).

The accuracy in detecting ramps can also be measured using event detection metrics such as probability of detection (POD), false alarm ration (FAR), etc.

In this task we ensembled HAIMOS point forecasts to create prediction intervals for GHI and DNI forecasts. The quality of the prediction intervals is measured via the Prediction Interval Normalized Averaged Width (PINAW) and the Probability Interval Coverage Probability (PICP). In this task we aim to achieve PICP > 90% and PINAW<20% for all the forecast horizons and target locations. These two metrics can be computed from the predicted and measured data (GHI or DNI) as

$$\text{PINAW} = \frac{1}{I_{\text{max}}}\frac{1}{N}\sum_{i=1}^{N}\hat{U}_i - \hat{L}_i$$

and

$$\text{PICP} = \frac{1}{N}\sum_{i=1}^{N}\varepsilon_i$$

where

$$\varepsilon_i = \begin{cases} 1 & \text{if } I_i \in [\hat{L}_i, \hat{U}_i] \\ 0 & \text{otherwise} \end{cases}$$

and $[\hat{L}_i, \hat{U}_i]$ is the predicted Prediction Interval, I_i is the measured irradiance (GHI or DNI), $I_{\max}$ the maximum measured irradiance, and N is the number of forecast instances in the validation dataset.

The results for the probabilistic HAIMOS forecasts are shown in Fig. 25 for horizons 1 to 5 hours and all locations. These results show that PICP and PINAW are clearly related with the variability of the target location and variable. Also, given that DNI is much more variable than GHI, it's harder to obtain low PINAW and large PICP for this variable.

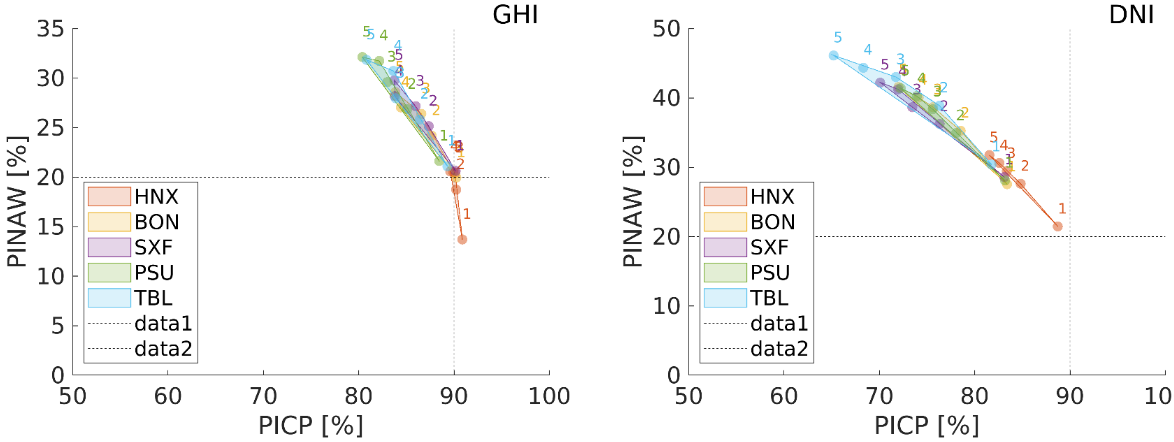


Figure 25: Irradiance (GHI and DNI) probabilistic forecast PINAW vs PICP for all locations and horizons from 1 to 5 h. Values computed using the validation dataset (2018). Models trained with data from 2017. Forecasts lower and upper bounds computed using a KNN model that uses as features the best set of inputs for each location and horizon. Set of inputs are identified by HAIMOS as shown previously. Note that values for PINAW and PICP are comparable to the best values found for short-term forecasts based on sky imaging (Chu et al., 2015, and Chu and Coimbra, 2017. The target values for ramps should be adjusted to reflect the fact that 80-90% PICP for GHI and 65-90% PICP for DNI are likely to be best targeting.

New-generation cloud forecasting tools

GOES-16 high-resolution cloud cover identification

In this task we explored the potential of the CNN-derived cloud mask model to generalize to new locations and test the advantages of combining locations for training. Data was processed and partitioned. The processing allowed classification of clear and cloudy based on the model from Reno and Hansen (2016) with a 20-min sliding window on 5-min resampled data. This was shown to produce similar classification metrics among the 12 locations and, when evaluated against the ABI cloud mask, produced accuracy metrics matching those reported in the data product technical document (Heidinger, A. and Straka (2012)). The repartitioning provided three distinct sets from the data: training, validation, and test. The partition was made randomly on dates, first at 80% training and 20% test from the full set of post-processed data. Then the training set was further partitioned at 85% training and 15% validation. The validation set was used for model tuning, while all model evaluation was performed using the test set. Results from the CNN cloud mask (CCM) are compared to the ABI cloud mask (ACM) in Table 5.

$$TPR = \frac{TP}{TP + FN}$$

$$MCC = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

Table 5: Numerical values of the ACM and CCM TPR and MCC scores. For each CCM score, a standard deviation from 10 iterations is given.

	TPR		MCC	
Location	ACM	CCM	ACM	CCM
BON	0.800	0.880 ± 0.023	0.390	0.710 ± 0.023
DRA	0.899	0.920 ± 0.007	0.591	0.751 ± 0.007
FPK	0.920	0.910 ± 0.015	0.631	0.723 ± 0.015
GWN	0.994	0.929 ± 0.009	0.537	0.721 ± 0.009
PSU	0.811	0.946 ± 0.014	0.517	0.707 ± 0.014
SXF	0.808	0.868 ± 0.051	0.560	0.685 ± 0.051
TBL	0.832	0.916 ± 0.013	0.596	0.768 ± 0.013
ABQ	0.940	0.895 ± 0.012	0.655	0.800 ± 0.012
BIS	0.871	0.904 ± 0.018	0.515	0.666 ± 0.018
HNX	0.951	0.945 ± 0.015	0.636	0.735 ± 0.015
SLC	0.959	0.854 ± 0.022	0.661	0.742 ± 0.022
STE	0.862	0.889 ± 0.027	0.474	0.639 ± 0.027
AVG	0.887	0.905 ± 0.076	0.564	0.721 ± 0.058

Table 5 also shows true positivity rate (TPR) and Matthews correlation coefficient (MCC) scores per location for the ACM and CCM. Both terms are defined below in terms of confusion matrix components: true positive (TP), false positive (FP), false negative (FN), and true negative (TN). The CCM outperforms the ACM in every location when evaluated by the MCC score, which is the less biased of the two for binary classification of unbalanced data classes. The average percent improvement in MCC is 30.0%.

TPR describes the fraction of correctly identified positive cases, clear sky states in this case, from total positive samples in the test set. In TPR, the CCM shows 2.47% improvement across locations, or roughly comparable performance. Confusion matrix metrics provide a more complete picture of model performance by capturing each combination of true and false positives or negatives.

While some model tuning was completed to optimize convolutional layer width, learning rate, fully connected (linear) layers, and dropout rate, a full hyperparameter optimization was outside the scope of this project, which aims to study the potential of a CCM and its ability to generalize to new locations. While it may certainly be possible to further optimize model hyperparameters, we determined the performance achieved here represents sufficient skill, without needing to specify different hyperparameter sets for different locations. For reference, desert locations like Albuquerque and Desert Rock tend to achieve higher performance with deeper and more complex networks, while locations with more varied cloud conditions like Bondville and Penn State saw diminishing or even negative returns on increasing complexity. The basic structure of the CNN is described in Table 6 and is built on two convolutional-pooling layers plus two fully connected layers with use of batch normalization and drop out to improve regularization.

Table 6: CNN architecture used in this project.

Block	Layer	Parameter [size, stride] or weight	Output size
	input		11x11x16
conv-pool	Conv2d	[3,1]	11x11x12
	BatchNorm2d		11x11x12
	ReLU		11x11x12
	MaxPool2d	[1,1]	11x11x12
conv-pool	Conv2d	[3,1]	11x11x24
	BatchNorm2d		11x11x24
	ReLU		11x11x24
	MaxPool2d	[2,0]	5x5x24
reshape			
FC1	BatchNorm1d		600
	dropout	0.1	600
	Linear		256
	ReLU		256
FC2	BatchNorm1d		256
	dropout	0.1	256
	Linear		2

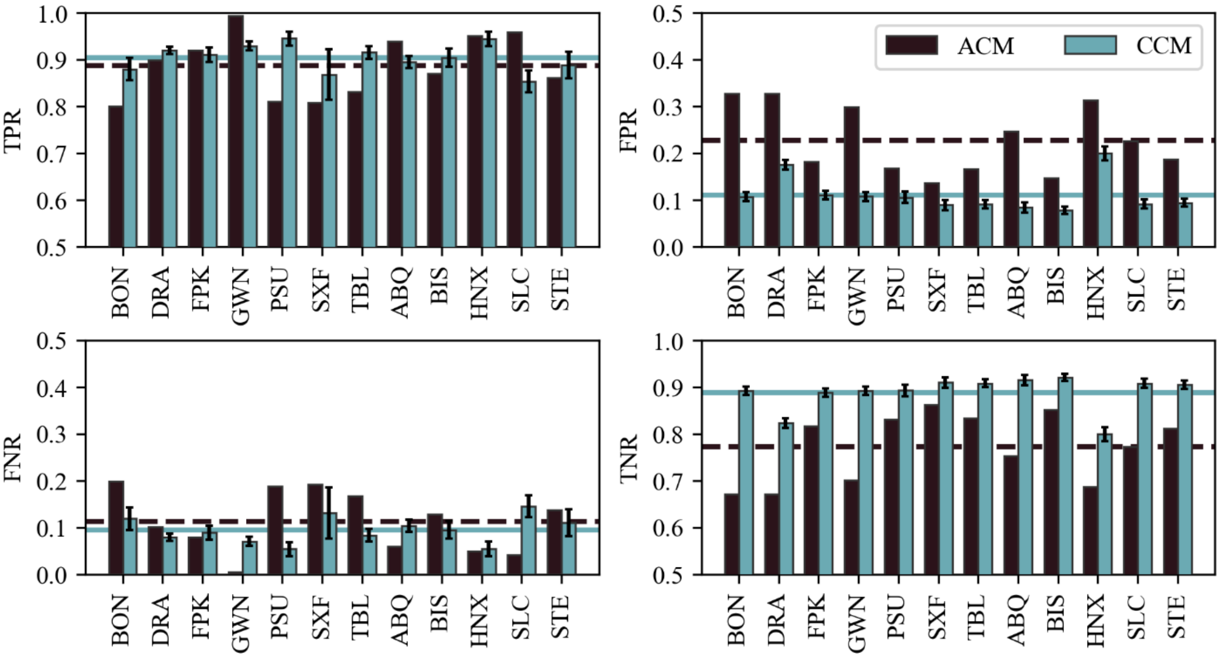


Figure 26: Comparison of confusion matrix values of the ACM and CCM per location. The CCM shows slight improvement in TPR and FNR and significant improvement in

FPR and TNR. The CCM maximizes performance in all four categories of the confusion matrix, rather than only maximizing accuracy in clear sky detection.

Figure 26 shows the confusion matrix metrics per location. For perfect performance, TPR and TNR should be unity while FPR and FNR are zero. Error bars on the CCM values represent standard deviation in evaluation of 10 identically trained CCMs.

In Figure 26 we see that CCM shows comparable to slight improvement in TPR and FNR with significant improvement in FPR and TNR across locations. This demonstrates that the CCM improves correct detection of cloudy states, which contributes to overall performance improvement without detrimental impact to correct detection of clear states.

Transfer learning and the improvement potential of training on multiple locations are explored through two scenarios. The CCM is trained on either (1) one source location or (2) a combination of two source locations, then applied to a new target location. Performance scores and other metrics are indexed to identify the transfer learning scenario. Case (1) is identified by the subscript *i:k* where the model is trained on data from the source location (*i*) and applied to data from a target location (*k*), where $i \neq k$. Case (2) is denoted by *ij:k* where $i \neq j \neq k$.

Results show that transfer learning performance is asymmetric. In general, $\mathrm{mcc}_{i:k}$ or the MCC score of a CCM trained on *i* and applied to *k* will show better performance if the mean value of ABI channel 1 is greater in the source location than in the target location and if the standard deviation of ABI channel 1 is smaller in the source location than in the target location. To put this more succinctly: $\mathrm{mcc}_{i:k} > \mathrm{mcc}_{k:i}$ when $\mu_i > \mu_k$ and $\sigma_i < \sigma_k$.

For each location, a set of 10,000 samples was selected, uniformly spaced throughout the training set. Per channel, each 11x11 pixel image was reduced to an average value. A second average and a standard deviation was then calculated from the sample set. From these per channel mean and standard deviation values, Pearson correlation coefficients were calculated with respect to the baseline MCC score, $\mathrm{mcc}_{i:i}$, resulting in an evaluation of the predictive potential of each channel. Channel 1 had the highest positive correlation coefficient at 0.740. Subsequent analysis thus used channel 1 mean and standard deviation to characterize closeness of input data characteristics between source and target locations. Channels 2 and 3 also showed high positive correlation with coefficients of 0.734 and 0.714, respectively. This suggests, similar results could be shown with respect to channels 2 and 3.

As noted before, transfer learning shows significant asymmetry, exemplified in Figure 27 for *i:k* models and in Figure 28 for *ij:k* models. The χ values on the x-axis capture a percent difference in channel 1 means between the source and target location, as:

$$\chi_{\mathrm{i:k,ch1}} = \frac{\mu_{\mathrm{i,ch1}} - \mu_{\mathrm{k,ch1}}}{\mu_{\mathrm{k,ch1}}}$$

$$\varphi_{\mathrm{i:k,ch1}} = \frac{\sigma_{\mathrm{i,ch1}} - \sigma_{\mathrm{k,ch1}}}{\sigma_{\mathrm{k,ch1}}}$$

Similarly, the metric φ is introduced on the color bar to show percent differences in standard deviation in channel 1. The non-uniform color mapping is meant to highlight regions of interest, namely standard deviations less than and greater than 10% apart in both the positive and negative directions.

The (a) subplots show MCC scores, with standard deviations from sets of 10 models shown by vertical error bars. The (b) subplots show a percent penalty score that can be interpreted as the performance percentage lost by using a transfer learning model, *i:k* or *ij:k*, relative to a local model like *k:k*.

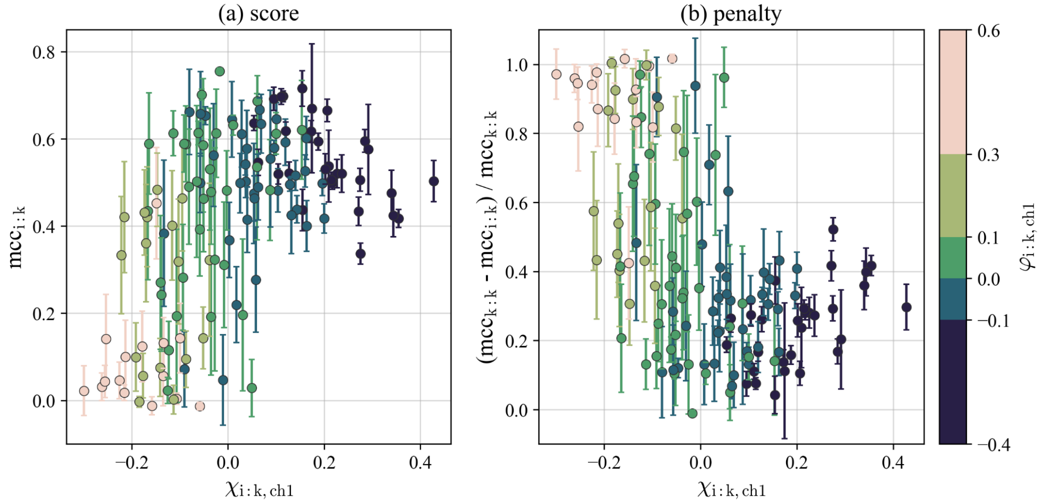


Figure 27: Showing asymmetry of model transferability. (a) MCC scores from *i:k* models and (b) Penalty of *i:k* models in comparison to *k:k* models.

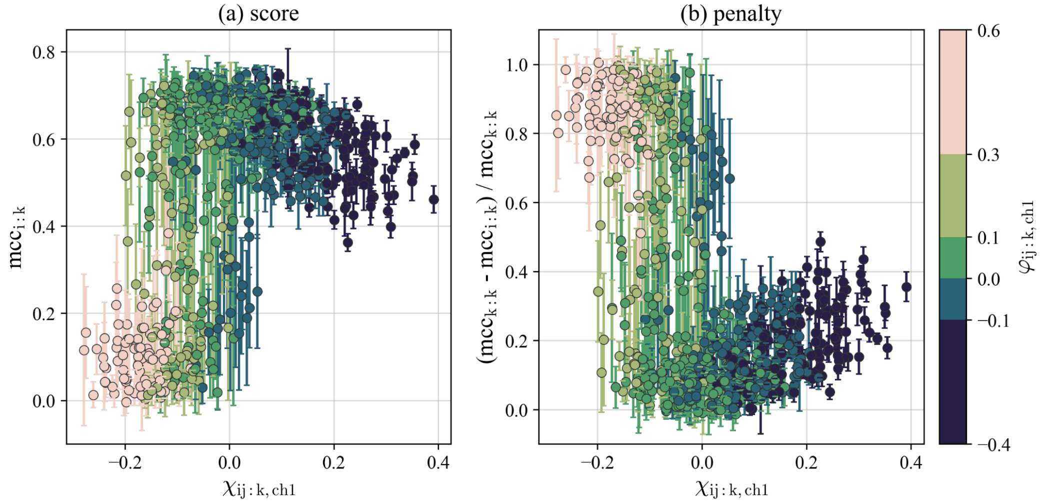


Figure 28: Same as figure 27 but for *ij:k* models.

HAIMOS refinement, ingestion of new-generation cloud data

In this project, we also explored merging GOES images and HAIMOS data. The model implemented is described in Fig. 29. The HAIMOS data produced previously is formatted into a 11x11 matrix and concatenated with the 11x11 GOES images.

The model was trained 10 times for each location and forecast horizons in order to account for the random initialization of the CNNs. Figure 30 compares the forecasting accuracy for the validation set, in terms of RMSE, for all locations and horizons, for the models without GOES data (dashed lines) against models with GOES data (solid lines and shaded band). The solid lines indicate the average RMSE for the 10 CNNs. The shaded band indicates the minimum and maximum RMSE obtained out of those 10 models. By merging HAIMOS data obtained from ground stations and CIF forecasts with GOES data RMSE is consistently reduced.

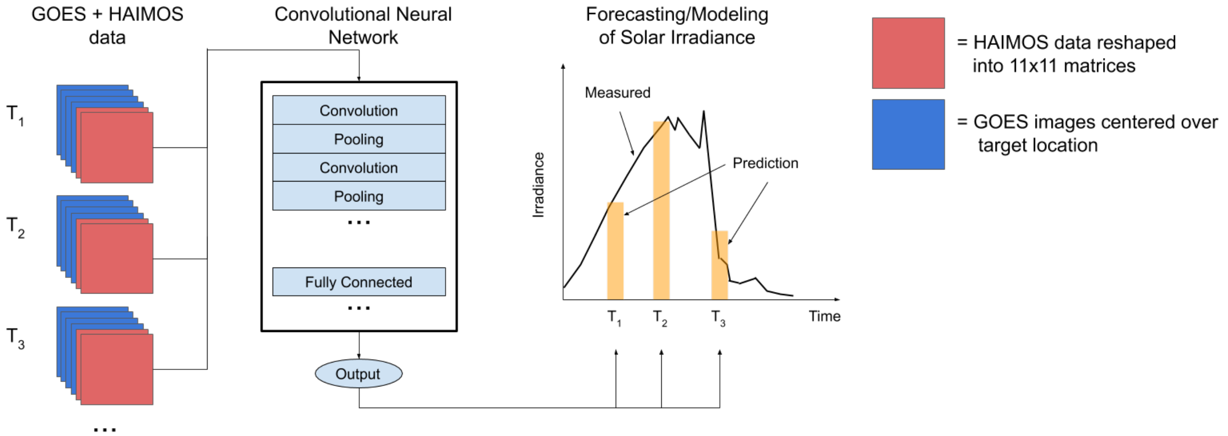


Figure 29: CNN model that ingest HAIMOS data and GOES imagery.

The model improvement, in terms of RMSE reduction, is shown in Fig. 31. As in figure 30 the solid line shows the average improvement and the band the minimum and maximum values. The figure indicates that models for HNX show the largest (horizon = 4 hours) and lowest improvement (horizon = 5 hours), as well as a more unstable evolution across horizons.
In all cases the models are improved with RMSE reductions ranging from 2% to 18%. On average this task reduced RMSE by ~7%.

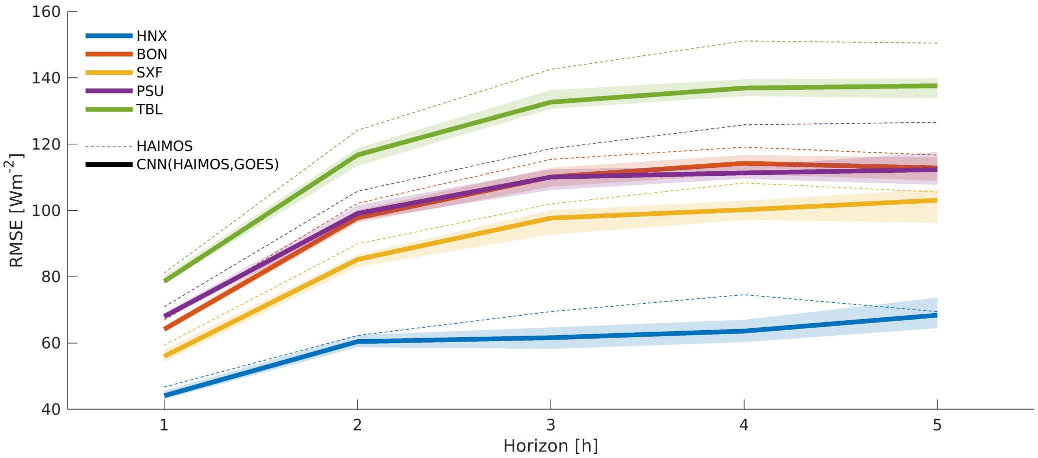


Figure 30: RMSE for GHI forecasting as a function of the horizon. The solid lines represent results that merge GOES data and HAIMOS. The dashed lines represent HAIMOS results without GOES data.

The results in Figs. 30 and 31 were obtained with data for 2018. The training and validation sets are depicted in the figure below. Both sets contain data from every season. The whole month of September is included in the validation set. This results from the comparison against data for GHI modeling from Laszlo et al (2020).

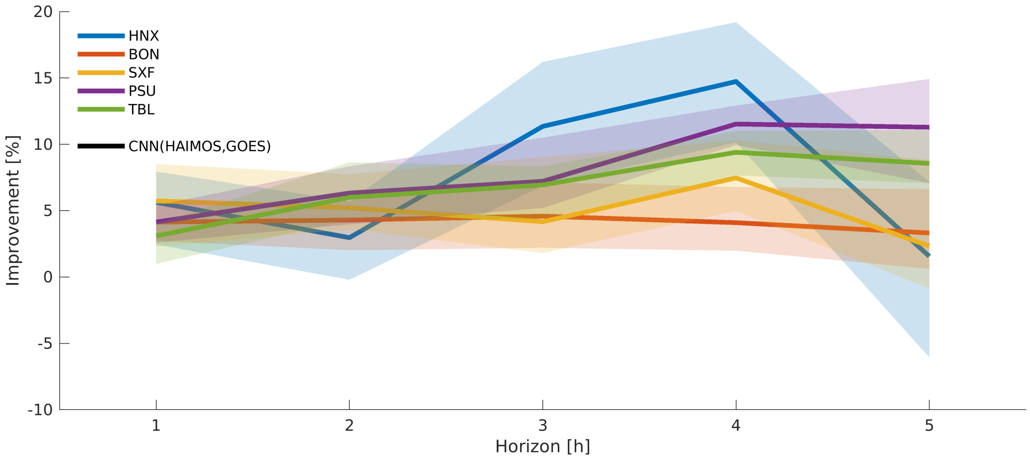
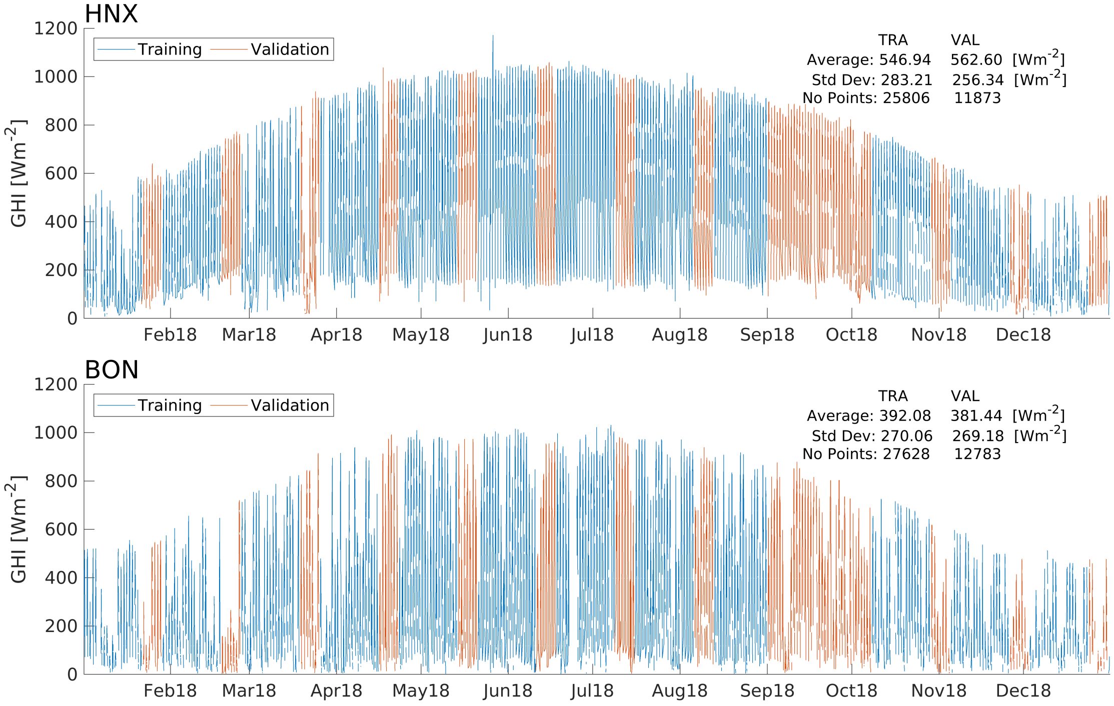


Figure 31: RMSE Improvement of the HAIMOS+GOES forecasts relative to the previous HAIMOS forecasts (without GOES data).



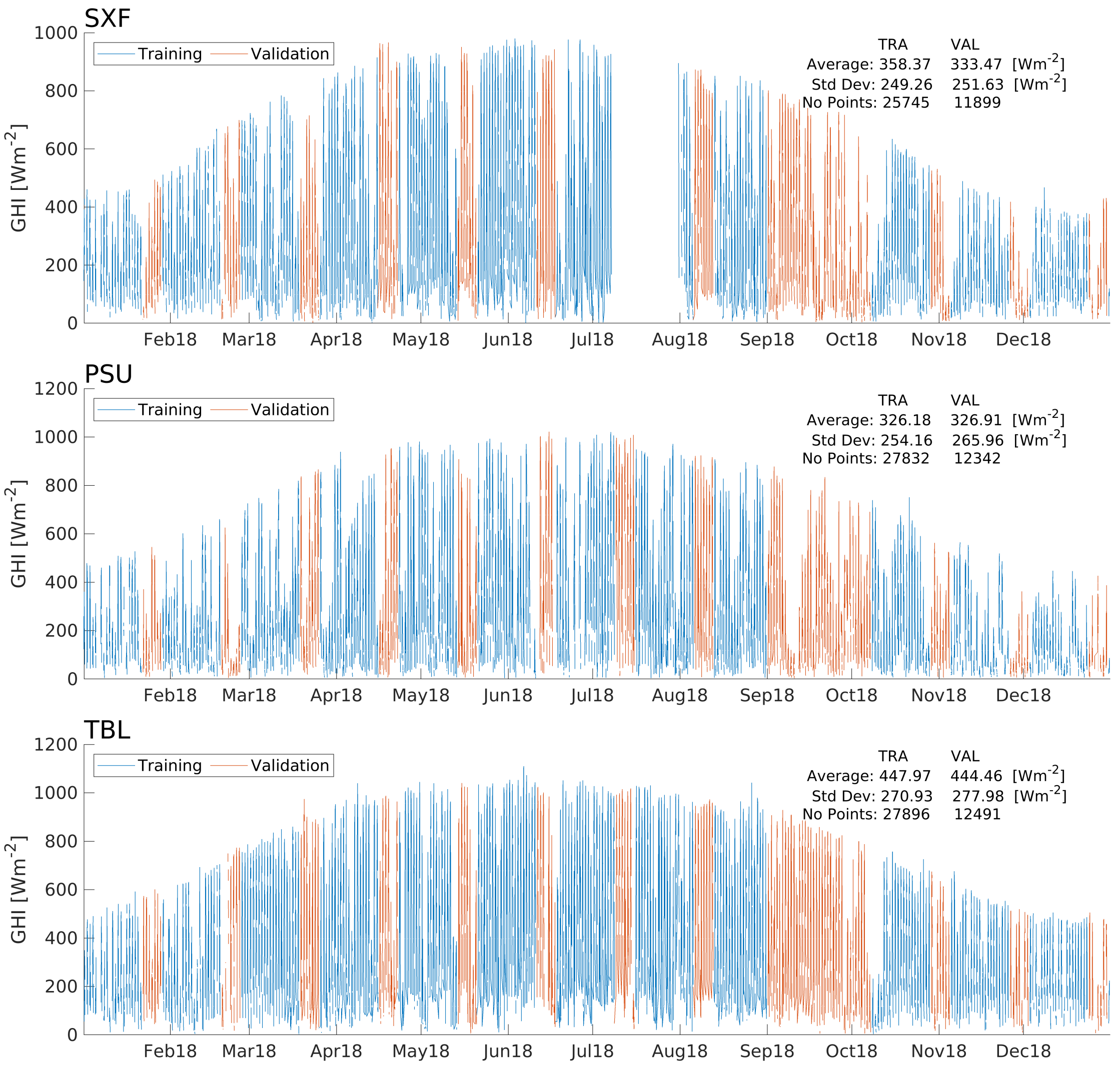


Figure 32: Training and validation sets for the 5 locations. The annotations indicate the average, standard deviation, and the number of data points for each dataset. The gap in SXF data for July 2018 is due to the lack of ground data.

Significant Accomplishments and Conclusions

Significant accomplishments obtained by HAIMOS are measured according to the test framework devise by the Topic Area 1 team. In this framework the forecasts by the different teams must follow the requirements:

- Variables
  - GHI
  - DNI
- Evaluation time range
  - Start: 2018-01-01 00:00 local time at each site.
  - End: 2018-12-31 23:59 local time at each site.
- Forecast parameters
  - Forecast issue time: 10 am local standard time

  - Forecast lead time: 14 hours
  - Forecast run length: 24 hours
  - Value type: interval mean
  - Interval length: 1 hour
  - Interval label: beginning

These requirements do not match the set up used to develop HAIMOS, which was focused on intra-day horizons and multiple aggregations (15, 30 minutes, etc.). For this reason, we retrained HAIMOS under these guidelines.

Quality controlled GHI and DNI data were collected from the Solar Forecast Arbiter (SFA) for the years 2017 to 2020 for the 10 locations selected by Area 1 team. Data for the year of 2018 were used as validation set. The remaining data were used as training. Figures 33 and 34 show the data for Humboldt State and Langley, respectively

The 1-minute data was augmented with clear-sky irradiances and solar geometry using the pvlib[7] library and then averaged into 1-hour bins according to the Area 1 testing guidelines (Figs. 35 and 36).

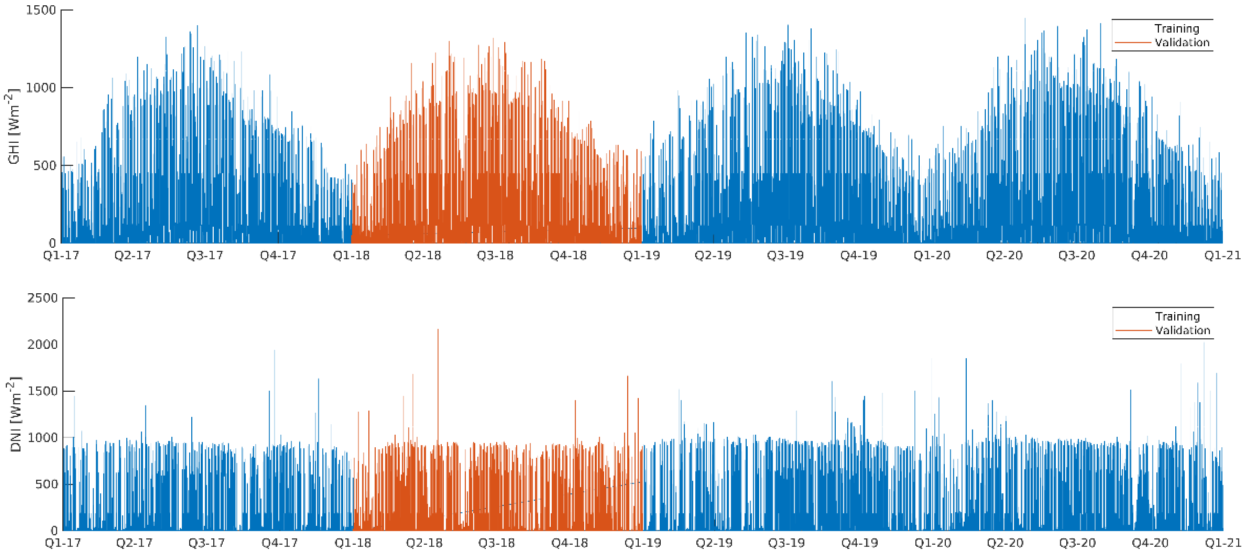
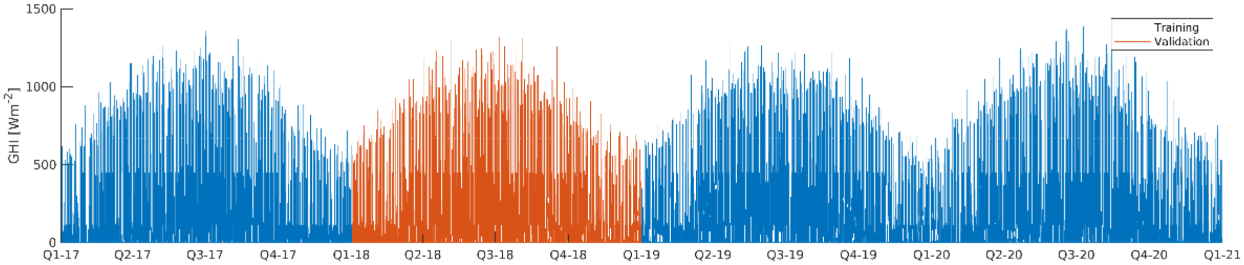


Figure 33: Quality-controlled GHI and DNI data for Humboldt State, California (NREL MIDC).



---

[7] https://pvlib-python.readthedocs.io/en/stable/index.html

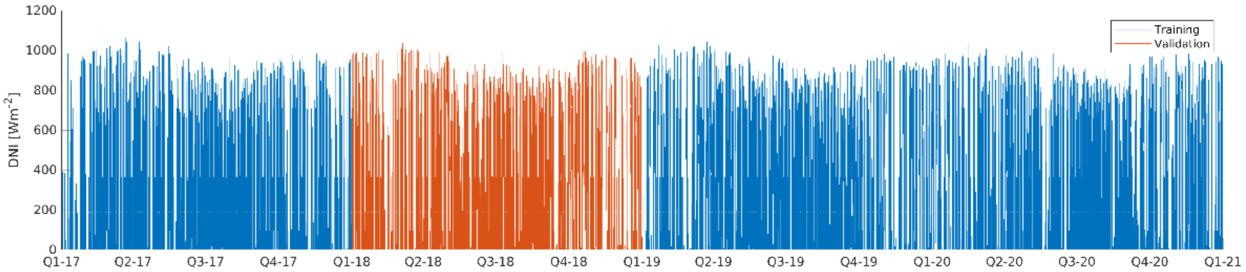


Figure 34: Quality-controlled GHI and DNI data for Langley (NASA, BSRN).

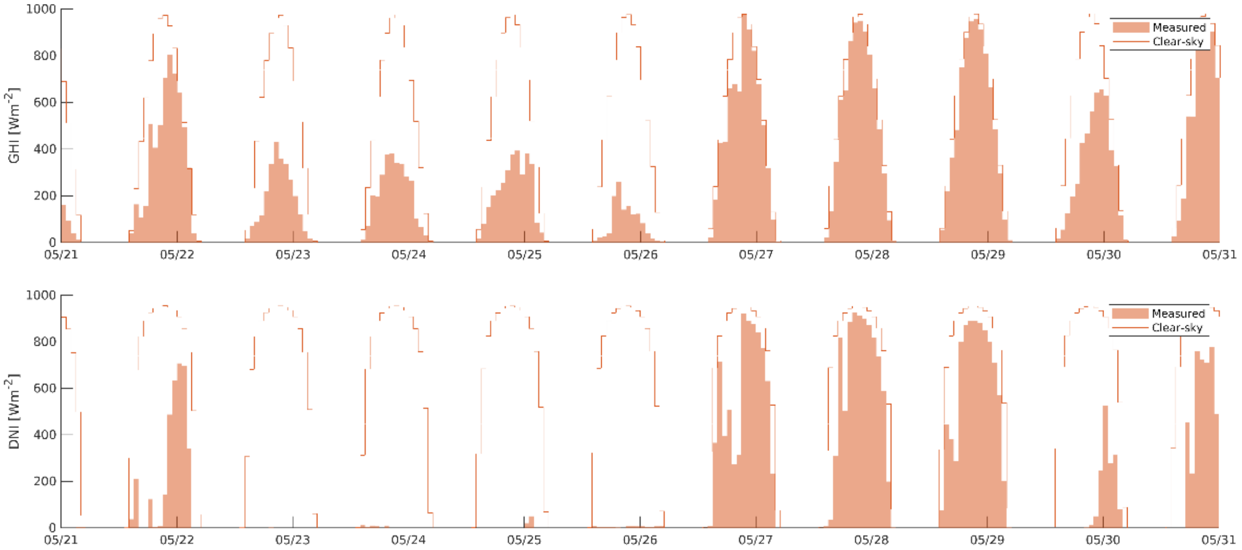


Figure 35: Hourly averaged GHI and DNI data for Humboldt State, California (NREL MIDC) for a 10-day window in the validation set.

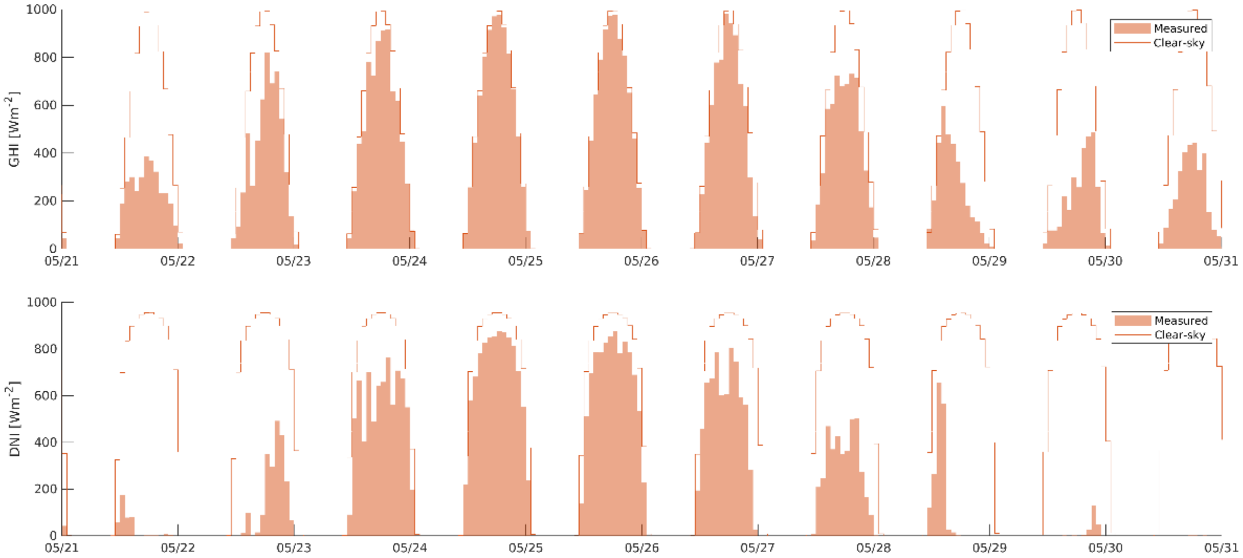


Figure 36: Hourly averaged GHI and DNI data for Langley (NASA, BSRN) for a 10-day window in the validation set.

Exogenous data for this task includes several products from NWP models. We downloaded and processed data from the NAM and HRRR models for the period 2017 to 2020. For each location, several variables were extracted and synchronized with ground data. The variables used include:

- DSWRF: Downward Short-Wave Radiation Flux [W/m^2]
- TCDC: Total Cloud Cover [%]
- LCDC: Low Cloud Cover [%]
- MCDC: Medium Cloud Cover [%]
- HCDC: High Cloud Cover [%]

Cloud cover variables (CC) were converted to GHI or DNI using a simple model:

$$I_{cc}(t) = \left(1 - \frac{CC(t)}{100}\right) I_{cs}(t)$$

were I_{cs} denotes the clear-sky irradiance and I_{cc} the resulting irradiance.

To avoid data leakage, that is, the use of data that would not be known if HAIMOS was run in real time, all the NWP-derived data must the available at the forecasting issue time of 10 am local standard time. That is ensured by using only NWP data whose reference time (time when the data is produced) precedes the forecasting issue time. Figure 37 shows some the NWP-derived data used as input to HAIMOS for Humboldt and Langley. This process was run for all locations and for DNI.

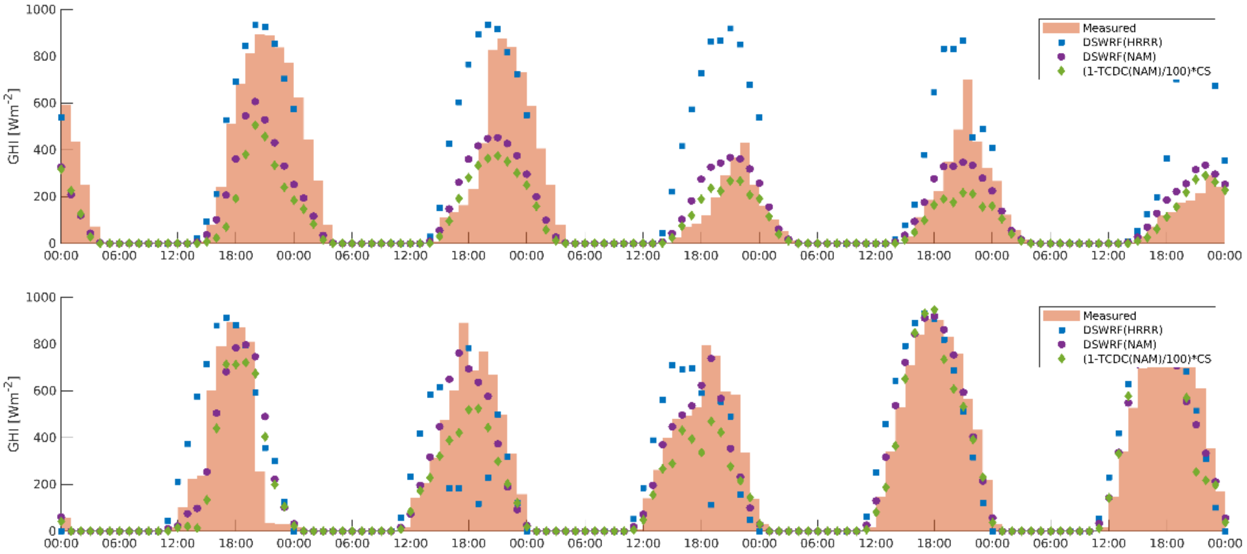
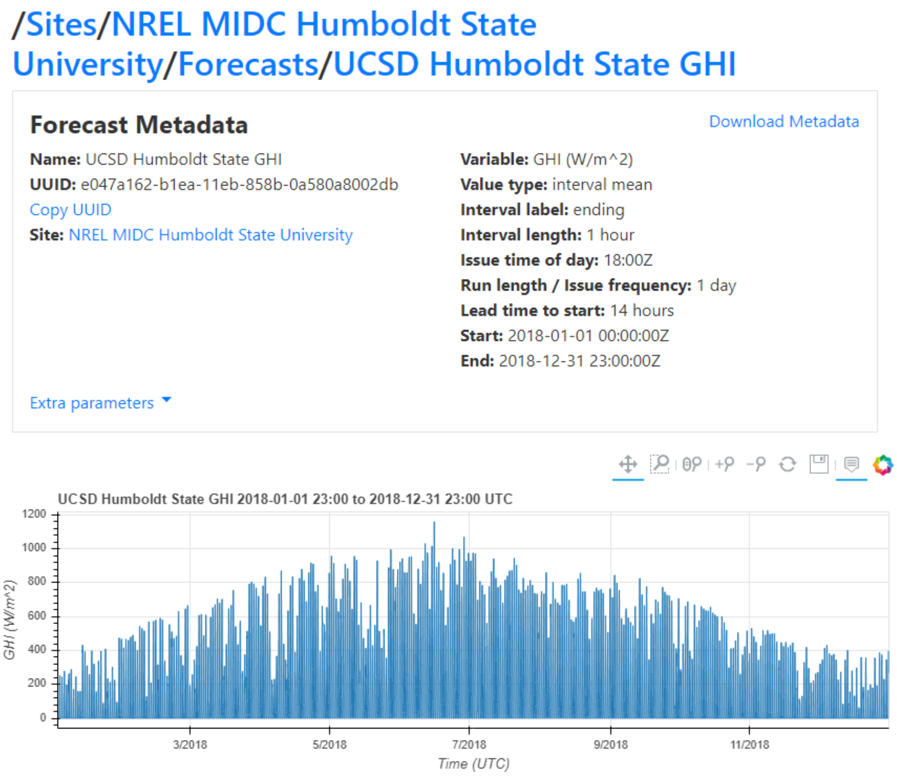


Figure 37: Some of the NWP-derived GHI for Humboldt (top) and Langley (bottom).

HAIMOS forecasting for Area 1 test framework

We prepared and uploaded HAIMOS forecasts the SFA according to the guidelines from Area 1 test framework. The HAIMOS forecasts were submitted to the SFA platform under the “TA2 evaluations” forecast metadata/endpoints.

Figure 38 shows the HAIMOS forecast for GHI for Humboldt and Fig. 39 shows the DNI forecast for Langley.



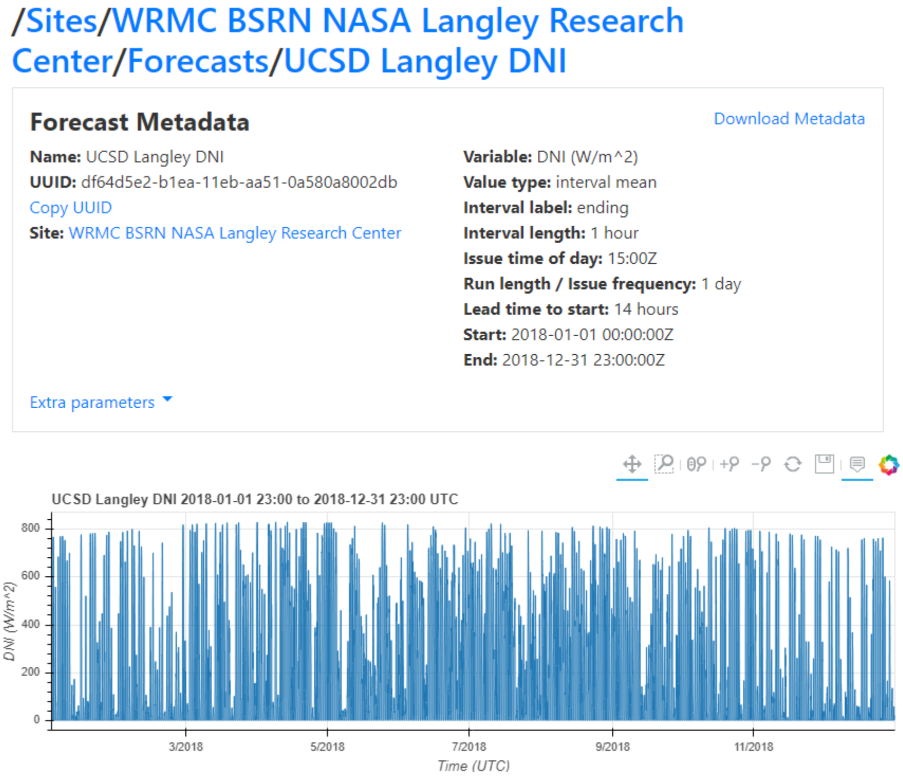
/Sites/NREL MIDC Humboldt State University/Forecasts/UCSD Humboldt State GHI

Forecast Metadata

Name: UCSD Humboldt State GHI
UUID: e047a162-b1ea-11eb-858b-0a580a8002db
Site: NREL MIDC Humboldt State University

Variable: GHI (W/m^2)
Value type: interval mean
Interval label: ending
Interval length: 1 hour
Issue time of day: 18:00Z
Run length / Issue frequency: 1 day
Lead time to start: 14 hours
Start: 2018-01-01 00:00:00Z
End: 2018-12-31 23:00:00Z

Figure 38: GHI forecasts for Humboldt.



/Sites/WRMC BSRN NASA Langley Research Center/Forecasts/UCSD Langley DNI

Forecast Metadata

Name: UCSD Langley DNI
UUID: df64d5e2-b1ea-11eb-aa51-0a580a8002db
Site: WRMC BSRN NASA Langley Research Center

Variable: DNI (W/m^2)
Value type: interval mean
Interval label: ending
Interval length: 1 hour
Issue time of day: 15:00Z
Run length / Issue frequency: 1 day
Lead time to start: 14 hours
Start: 2018-01-01 00:00:00Z
End: 2018-12-31 23:00:00Z

Figure 39: DNI forecast for Langley.

Once the forecasts were uploaded, we used the report functionality in SFA to validate HAIMOS under the Area 1 framework. Figures 40 and 41 show snapshots of the table that lists the forecast error metrics computed by SFA and the forecasting skill of HAIMOS against the reference NAM forecasts prepared for the Area 1 team. NREL's forecasts are also included in this report. Such reports are produced for GHI and DNI for all locations.

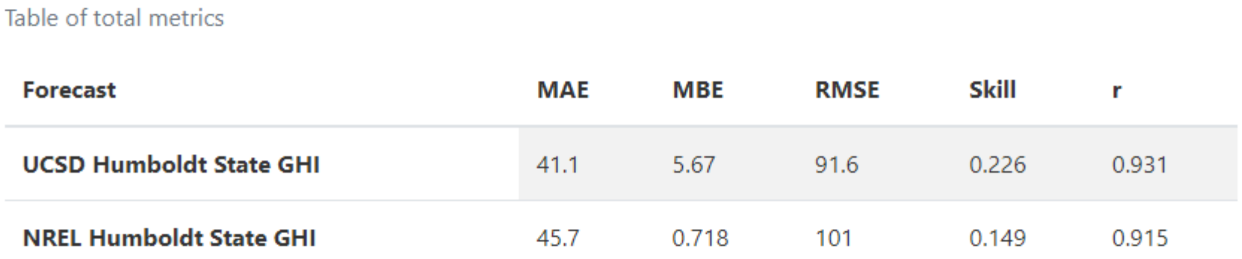


Table of total metrics

Forecast	MAE	MBE	RMSE	Skill	r
UCSD Humboldt State GHI	41.1	5.67	91.6	0.226	0.931
NREL Humboldt State GHI	45.7	0.718	101	0.149	0.915

Figure 40: Error metrics computed by SFA for Humboldt's GHI. Skill is computed against the reference NAM forecasts.

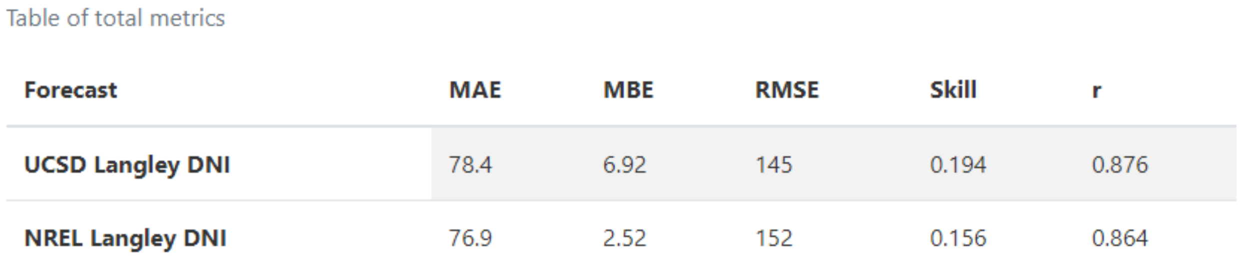


Table of total metrics

Forecast	MAE	MBE	RMSE	Skill	r
UCSD Langley DNI	78.4	6.92	145	0.194	0.876
NREL Langley DNI	76.9	2.52	152	0.156	0.864

Figure 41: Error metrics computed by SFA for Langley's DNI. Skill is computed against the reference NAM forecasts.

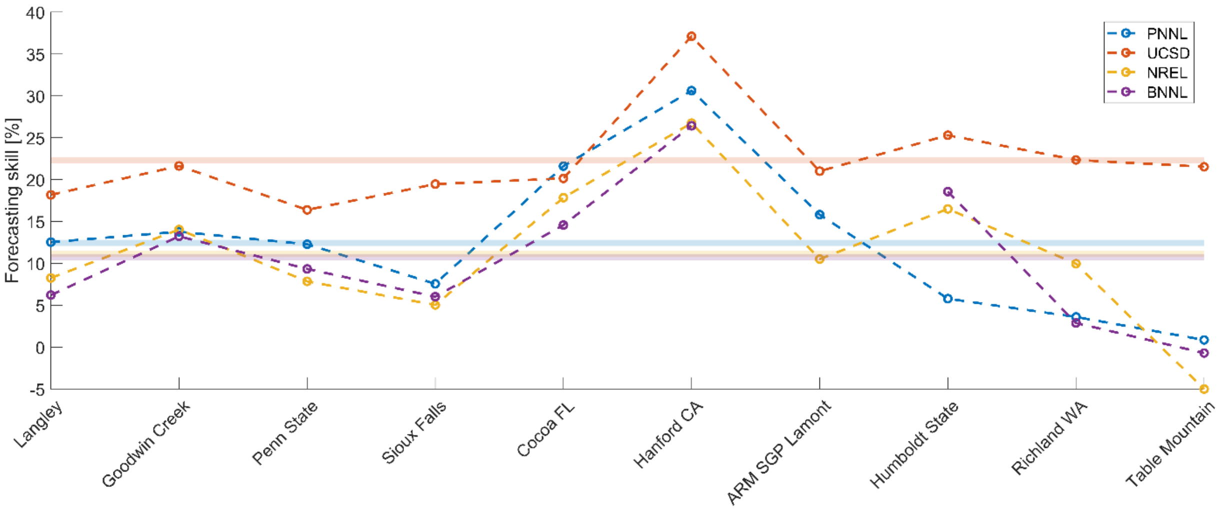


Figure 42: GHI forecasting skill relative to NAM's reference forecasts for all the teams in topic area 2. Dots show the skills for each location. Horizontal lines show the average skill across locations.

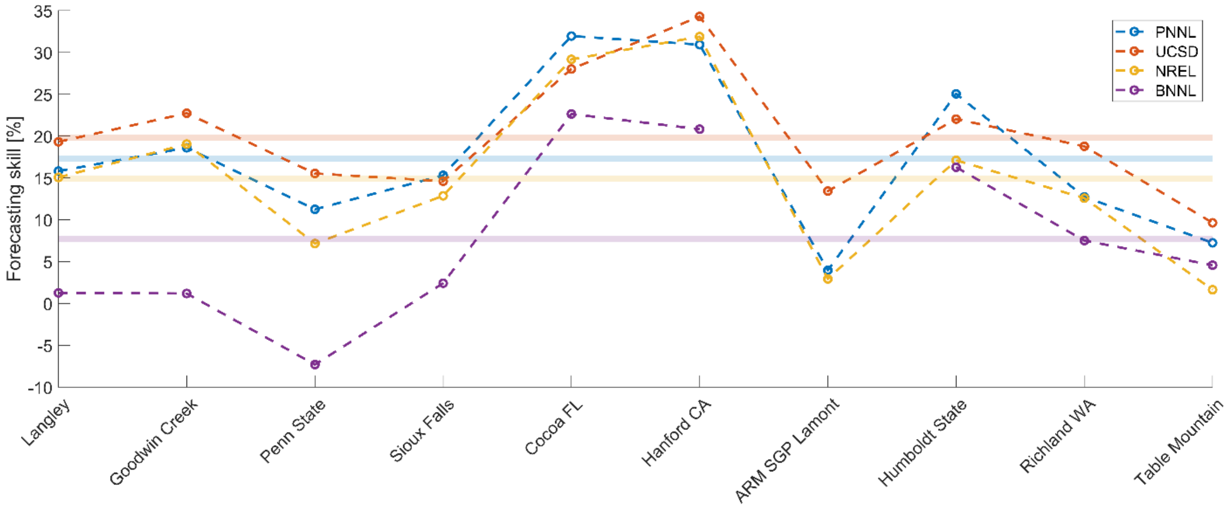


Figure 43: Same as figure 42 but for DNI.

A direct comparison between the HAIMOS methodology and the other day-ahead forecasts produced by Area 2 participants is depicted in figures 42 and 43. Both the robustness and the effectiveness of the HAIMOS methodology are clear, as the forecasting skill of the HAIMOS method is consistently higher across the multiple locations, even though most of the HAIMOS methodology has been primarily optimized for intra-day forecasts.

Budget and Schedule

The budget and schedule originally approved by the DOE has been followed closely throughout the duration of this project. There were no significant delays in any phase or budget period despite the disruptions caused by the ongoing pandemic. Starting in early 2020, minor changes in budget categories have been adjusted to reflect changes in personnel and workload due to the pandemic. Overall, both the project budget and the project schedule have been proposed and managed properly since the beginning of the project. There were minor delays in invoicing by a subcontractor, but none that compromised the progress of the project. UC San Diego also went through a massive financial system overhaul in 2020, which unfortunately affected some of the financial reporting from central campus to DOE. All technical and financial reports generated from our research group have been submitted on time and in agreement with procedures outlined by the DOE and by campus policies.

Path Forward

The HAIMOS project was unique among the Area 2 projects in the sense that it did not rely on WRF-Solar improvement but offered a comprehensive methodology for producing improved forecasts based on best practices in physical modeling and machine learning. The project demonstrates the effectiveness and robustness of the methodology used (see Figs. 42 and 43). Although not included in the SOPO, this project also allowed us to study the effectiveness of CNNs for cloud masking and their potential for producing forecasts that utilize an entirely different set of inputs from WRF-Solar. As such, we submitted two (2) proposals to DOE in order to continue this effort. One proposal was for the development of CNN-based forecasts as a new reference forecast for the entire CONUS, and the other was a collaboration with EPRI on long-term solar forecasts for improved grid operations at high penetration of renewables. These proposals were submitted in 2020 in response to DE-FOA-0002243. Both proposals were declined by the DOE, and as a consequence of the COVID-19 pandemic impact on personnel combined with the loss of continued DOE funding, the UC San Diego solar forecasting effort will likely end with this project. There is a remote possibility of continuing this research effort with funding from the California Energy Commission or the National Science Foundation, but no direct applicable funding opportunities by either agency have been announced at this point.

As for technology transfer and commercialization, we will continue to openly publish our methods, data and results in technical journals and conferences. We will also engage with interested parties that show interest in high-fidelity solar forecasts. That said, the ability to commercialize best practice methods for solar forecasting has always been difficult due to the conservative nature of the energy industry. With few exceptions, power generators show little interest in better understanding solar resourcing, much less forecasting. Utility companies and system operators also see little reward in high-fidelity forecasts given that they can continue to dispatch carbon-intensive reserve units in response to solar and wind variability. As a result, operational forecasts show substantially lower skills than the ones reported here. By publishing and presenting our research findings and data in journals and conferences, we hope that eventually the methodologies presented in this project will make their way into the operational forecasting systems used by the energy industry.

Acknowledgement

This material is based upon work supported by the Department of Energy, Office of Energy Efficiency and Renewable Energy, Solar Energy Technologies Office, under Award Number DE-EE0008216.

Publications

Peer-Reviewed Journal Articles

M. Li and C. F. M. Coimbra (2019) “On the Effective Spectral Emissivity of Clear Skies and the Radiative Cooling Potential of Selectively Designed Materials,” International Journal of Heat and Mass Transfer (135), pp. 1053–1062

M. Li, H. B. Peterson, and C. F. M. Coimbra (2019) “Radiative cooling resource maps for the contiguous United States,” Journal of Renewable and Sustainable Energy (11), 036501

H. T. C. Pedro, D. P. Larson, and C. F. M. Coimbra (2019) "A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods," Journal of Renewable and Sustainable Energy (11), 036102

D. P. Larson, M. Li and C. F. M. Coimbra (2020) “SCOPE: Spectral cloud optical property estimation using real-time GOES-R longwave imagery,” Journal of Renewable and Sustainable Energy (12), 26501

M. Li, Z. Liao, and C. F. M. Coimbra (2020) “Spectral solar irradiance on inclined surfaces: A fast Monte Carlo approach,” Journal of Renewable and Sustainable Energy (12), 053705

Conference Presentations

M. Li and C. F. M. Coimbra (2019) “Spectral Modeling of the Radiative Interactions Between Large Scale Solar Farms and the Atmosphere,” 100th AGU Annual Meeting

Mengying Li and Carlos F. M. Coimbra (2019) “Effects of Local Atmospheric Conditions on Plane-of-array Solar Irradiance”, AGU Fall Meeting 2019

David P. Larson, Mengying Li, and Carlos F. M. Coimbra (2019) “SCOPE: Spectral Cloud Optical Property Estimation from GOES-R Imagery”, AGU Fall Meeting 2019

Hugo Pedro, Carlos Coimbra, David Larson and Mengying Li “Physics-based, Data-driven Solar Irradiance Forecasting”, AGU Fall Meeting 2019

References

Y. Chu, M. Li, H. T. C. Pedro and C. F. M. Coimbra (2015) “Real-Time Prediction Intervals for Intra-Hour DNI Forecasts,” Renewable Energy (83) pp. 234–244

Y. Chu and C. F. M. Coimbra (2017) “Short-Term Probabilistic Forecasts for Direct Normal Irradiance,” Renewable Energy (101), pp. 526–536

Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. Solar Energy 86, 2145–2169.

Heidinger, A. and Straka, W. C. III, “Algorithm Theoretical Basis Document: ABI Cloud Mask.” Technical report; NOAA NESDIS Center for Satellite Applications and Research: College Park, MD, USA. 2012

Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. Solar Energy 80, 468–478.

Ineichen, P., Perez, R., 2002. A new airmass independent formulation for the Linke turbidity coefficient. Solar Energy 73, 151–157.

R. H. Inman, J. G. Edson and C. F. M. Coimbra (2015) “Impact of Local Broadband Turbidity Estimation on Clear Sky Models for Direct Normal Irradiance,” Solar Energy (117) pp. 125–138

Laszlo, I., Liu, H., Kim, H.-Y., Pinker, R.T., 2020. Chapter 15 - Shortwave Radiation from ABI on the GOES-R Series, in: Goodman, S.J., Schmit, T.J., Daniels, J., Redmon, R.J. (Eds.), The GOES-R Series. Elsevier, pp. 179–191. https://doi.org/10.1016/B978-0-12-814327-8.00015-9

M. Li, Z. Liao and C. F. M. Coimbra (2018) “Spectral Model for Clear Sky Atmospheric Longwave Radiation,” Journal of Quantitative Spectroscopy and Radiative Transfer (209), pp. 196–211

Reno, M. J. and Hansen, C. W. “Identification of Periods of Clear Sky Irradiance in Time Series of GHI Measurements.” Renewable Energy, v90. 2016.