A Moment in the Sun: Solar Nowcasting from Multispectral Satellite Data using Self-Supervised Learning

Akansha Singh Bansal*
akanshasinghbansal@colostate.edu
Cooperative Institute for Research in Atmosphere (CIRA)- CSU
Fort Collins, Colorado, USA

Trapit Bansal
tbansal@cs.umass.edu
University of Massachusetts Amherst
Amherst, Massachusetts, USA

David Irwin
deirwin@umass.edu
University of Massachusetts Amherst
Amherst, Massachusetts, USA

ABSTRACT

Solar energy is now the cheapest form of electricity in history. Unfortunately, significantly increasing the electric grid's fraction of solar energy remains challenging due to its variability, which makes balancing electricity's supply and demand more difficult. While thermal generators' ramp rate—the maximum rate at which they can change their energy generation—is finite, solar energy's ramp rate is essentially infinite. Thus, accurate near-term solar forecasting, or *nowcasting*, is important to provide advance warnings to adjust thermal generator output in response to variations in solar generation to ensure a balanced supply and demand. To address the problem, this paper develops a general model for solar nowcasting from abundant and readily available multispectral satellite data using self-supervised learning.

Specifically, we develop deep auto-regressive models using convolutional neural networks (CNN) and long short-term memory networks (LSTM) that are globally trained across multiple locations to predict raw future observations of the spatio-temporal spectral data collected by the recently launched GOES-R series of satellites. Our model estimates a location's near-term future solar irradiance based on satellite observations, which we feed to a regression model trained on smaller site-specific solar data to provide near-term solar photovoltaic (PV) forecasts that account for site-specific characteristics. We evaluate our approach for different coverage areas and forecast horizons across 25 solar sites and show that it yields errors close to that of a model using ground-truth observations.

CCS CONCEPTS

• **Computing methodologies** → *Machine learning approaches.*

KEYWORDS

Solar Nowcasting, Satellite Data, Machine Learning

*Work performed while the author was at the University of Massachusetts, Amherst.

e-Energy '22, June 28-July 1, 2022, Virtual Event, USA

ACM ISBN 978-1-4503-9397-3/22/06…$15.00
https://doi.org/10.1145/3538637.3538854

ACM Reference Format:
Akansha Singh Bansal, Trapit Bansal, and David Irwin. 2022. A Moment in the Sun: Solar Nowcasting from Multispectral Satellite Data using Self-Supervised Learning. In *The Thirteenth ACM International Conference on Future Energy Systems (e-Energy '22), June 28-July 1, 2022, Virtual Event, USA.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3538637.3538854

1 INTRODUCTION

Solar power is now the cheapest form of electricity in history. As a result, the U.S. Energy Information Administration (EIA) projects that the share of renewable energy from solar and wind in the electric grid will double to nearly 42% by 2050 with solar energy poised to account for nearly 80% of this increase [16]. This dramatic increase in solar energy generation is, of course, critical to mitigating the negative environmental and economic consequences of climate change due to carbon emissions from thermal generators, which generate energy from burning fossil fuels. Unfortunately, significantly increasing the grid's fraction of solar energy, e.g., beyond 50%, remains challenging due to its variability, which makes balancing electricity's supply and demand more difficult. While thermal generators' ramp rate—the maximum rate at which they can change their energy generation—is finite, solar energy's ramp rate is essentially infinite. As a result, to maintain the grid's frequency and voltage within a narrow range, utilities will require accurate near-term forecasts of solar energy generation that provide advance warning of significant changes in solar energy output to compensate for them, either by adjusting energy's supply, i.e., by altering thermal generator output, or its demand, i.e., by reducing load via demand response.

To address the problem, we develop a general model for near-term solar forecasting from multi-spectral satellite data using deep learning. Such near-term forecasting at temporal scales less than two hours is commonly called *nowcasting* [32], and has been focus of study in meteorology and precipitation for over three decades [12]. Many mobile weather apps, such as DarkSky, now include nowcasts of precipitation [3]. However, nowcasting solar generation is significantly more challenging than nowcasting precipitation, primarily because surface radars can directly sense precipitation but not clouds. Thus, our work instead leverages real-time satellite data from the recently launched GOES-R series of geostationary satellites, specifically GOES-16. The GOES-16 satellite observes the continental United States across sixteen spectral bands of light, and generates rich spatio-temporal data at an unprecedented temporal and spatial resolution: every five minutes for every 0.5-2km^2 area. As we discuss, we can accurately infer ground-level solar irradiance at any location using this spectral satellite data.

Real-time data from satellites, such as GOES-16, presents an untapped opportunity for solar nowcasting at global scales using data-intensive deep learning techniques. Prior work has primarily focused on solar nowcasting using surface-level sky-imagery data [30, 36, 43, 44], which predicts near-term solar generation from images of the sky taken at a solar site's location. Thus, this technique is "site-specific," and requires specialized hardware to be deployed at a particular location. In contrast, solar nowcasting from multi-spectral satellite data can be applied to any location.

Our approach utilizes the intuition that the first three spectral bands of light, corresponding to the visible region, capture information about the solar irradiance and cloud cover at any observed area. As a result, we propose to learn a global deep autoregressive model, i.e., a model that predicts the next observation in a sequence, directly from the raw satellite observations to capture the statistical patterns that are indicative of future solar irradiance.

Our intuition is similar to that of prior work on solar nowcasting using cloud motion vectors. Clouds are the primary reason solar sites' output drops from its clear sky potential, which is largely deterministic based on the ambient temperature, time-of-day, day-of-year, and location. While cloud movements are a function of complex non-linear atmospheric dynamics over long time periods, their movements are more predictable over short periods [15, 25]. Thus, solar forecasting over long periods, e.g., a few hours to days, requires Numerical Weather Prediction (NWP) models [34] that use physical atmospheric models to account for non-linear dynamics. In contrast, solar nowcasting over short time periods is simpler to model due to the larger influence of more recent changes.

Specifically, prior work on solar nowcasting has focused on programmatically identifying clouds in satellite or sky images to determine their size, direction, and velocity [15, 25]. Prior solar nowcasting techniques use such cloud motion vectors to forecast solar output based on the direction and velocity of clouds. Our approach has a similar intuition, but instead of directly identifying cloud motion vectors for which there is no training data, we train a deep learning model that takes as input historical spatio-temporal multi-spectral satellite data of a large region to infer how it changes over time and space. Changes in this spectral data implicitly capture cloud movements, as clouds reflect more light, which satellites implicitly capture as changes in their spectral data.

Based on our intuition above, we develop self-supervised deep learning models using convolutional neural networks (CNNs) and long short-term memory networks (LSTM) to forecast the next satellite spectral values at the solar sites of interest. These models require historical spectral satellite data over a region surrounding a particular site as input. We analyze and quantify model accuracy based on both the amount of temporal data, i.e., how far in the past, and the size of the region, i.e., how large of a region, used as input for forecasting 15 minutes in the future. As we show, the more distant the forecast horizon, the larger the historical data and spatial region required, and the lower the accuracy. However, there are rapidly diminishing returns in accuracy improvement, and significant increases in training time and resources, once the historical data and spatial region reach a certain size. After generating a spectral forecast, our approach then applies a simpler regression model to infer a specific site's solar output from its spectral forecast data obtained from the self-supervised CNN-LSTM model. We compare our solar nowcasting models with both the accuracy of this regression model, which infers solar output based on current conditions, as well as a persistence model that assumes that the future solar generation remains unchanged over the forecasting horizon, which also serves as the baseline for comparisons.

Importantly, we condition our analysis above based on the magnitude and frequency of changes in solar energy output at a given location. Put simply, if a location, such as San Diego, California, is rarely cloudy, then a simple persistence approach, which predicts near-term solar output never changes, will be highly accurate. Even in highly variable climates, solar output often does not change much over short time periods of 5-30 minutes, which makes simple persistence models appear highly effective. However, accurately forecasting "big" changes in solar energy output is much more important for grid operations, as these are the changes that require an active response. As a result, we specifically focus on the accuracy of forecasting "big" near-term changes in solar power. As we show, the larger the change in solar output, the larger the improvement in accuracy between our deep learning approach and others.

Ultimately, the novelty of our work lies in demonstrating how to leverage the increasing availability of rich satellite data for near-term solar forecasting by combining, adapting, and extending multiple existing machine learning approaches. Our hypothesis is that solar nowcasting using deep learning models trained on multispectral satellite data is both general and accurate, especially at forecasting large changes in solar output. In evaluating our hypothesis, we make the following contributions.

Satellite Data Compilation. We compile a large-scale dataset for 25 solar sites that includes their average solar energy generation, ambient temperature, and satellite data across 16 spectral channels for their surrounding region (up to 10km away) every 5 minutes for a year-long period. We use this dataset to train and test our deep learning models, and plan to publicly release it.

Self-supervised Models on Satellite Data. We develop self-supervised deep learning models trained on spectral satellite observations that use convolutional neural networks (CNNs) and long short-term memory networks (LSTM) to forecast the future observations. These models utilize the spatio-temporal observations across all 25 sites (analyzed in this work) for large-scale training. We analyze the importance of both spatial and temporal components, and also compare with other simpler machine learning methods for such self-supervised modeling.

Solar Nowcasting Models. We demonstrate the utility of self-supervised models for solar nowcasting, which depend on numerous factors, including solar irradiance and cloud cover. Our approach uses the forecasted spectral data from a self-supervised model as input to a separate site-specific regression model that predicts a specific site's solar output 15 minutes in the future from current spectral satellite data. The regression model incorporates the effect of physical site characteristics, such as module area, tilt, orientation, and tree cover, on solar output.

Implementation and Evaluation. We implement our models above in python using Tensorflow [5], and train them on a GPU cluster. Given the size of the datasets and complexity of our models, training each model requires ~86 GPU-hours. We evaluate our approach for different coverage areas and forecast horizons across 25 solar sites, and show that our approach yields errors close to

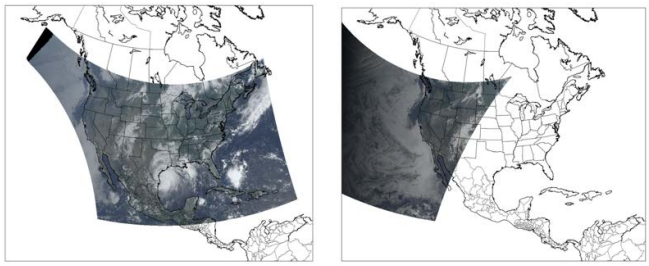

Figure 1: GOES-16 (left) and GOES-17 (right) coverage area. We use GOES-16 satellite data as it observes the continental U.S. covering all the 25 solar sites considered in this work.

those of a model that uses ground truth observations. We also qualitatively compare with prior approaches using ground-level sky imagery, and show that our approach yields similar or better results without requiring custom hardware. Finally, we show that our deep learning models are much more accurate at identifying "big" changes in near-term solar output.

2 BACKGROUND

The GOES-R series of geostationary satellites started launching in late 2017, and now provides remote sensing data from 16 spectral channels that comprise different ranges of wavelengths of light, as well as numerous secondary derived products, such as estimates of Downward Shortwave Radiation (DSR). Note that solar photovoltaic (PV) cells only generate power from the first 3 of these spectral channels, which are mostly in the visible range of light. As a result, our work only trains models on the first three spectral channels. As shown in Figure 1, GOES-16 covers the entire U.S., while GOES-17 provides additional coverage for the Western U.S. and the Pacific ocean. Our work uses data from GOES-16 as it covers all the solar sites in our evaluation dataset. GOES-R data has both high temporal and spatial resolution, including new spectral readings every 5 minutes for every 0.5-2km^2 area in the U.S., and is made publicly available in near real time. The data is a rich source of information about the environment that is useful for a wide range of applications. Solar nowcasting is a particularly compelling application, since solar energy output correlates directly with the amount of light (of certain wavelengths) that reaches the ground.

2.1 Prior Work

Traditionally, solar forecasts depend on some measure of cloud cover to assess the effect of clouds on solar output. Cloud cover is commonly measured by weather stations in units of "oktas," where 1 okta means that one-eighth of the sky is partially covered by clouds. Oktas are measured at ground-level using sky mirrors. Unfortunately, oktas are a coarse and imprecise measure of cloud cover that is typically released by weather stations every hour. In addition, not every solar site is located near a ground-level weather station that reports oktas. Thus, even though cloud cover measurements in oktas are widely available, this data remains an unreliable and inaccurate basis for solar forecasting.

A better basis for solar forecasting is direct ground-level readings of solar irradiance. The U.S. operates the Surface Radiation Budget Network (SURFRAD) [7] within the U.S., which measures and records ground-level solar irradiance at different monitoring sites. These monitoring sites operate in collaboration with the National Oceanic and Atmospheric Administration (NOAA). Unfortunately, while SURFRAD measurements are precise, they are not widely available, as there are only eight SURFRAD sites maintained in the entire U.S. As a result, we also cannot use SURFRAD data as a basis for solar forecasting models. Finally, NOAA also releases derived data products from raw GOES spectral data, including Downward Shortwave Radiation, or DSR [2], which is an estimate of ground-level solar irradiance.

Unfortunately, satellite-based estimates of DSR are only released every hour. In addition, DSR is computed using a complex physical model [1] that is highly sensitive to clouds, and thus inaccurate under significant cloud cover. As a result, DSR readings are often not even released during cloudy conditions, when they are most important for solar nowcasting [1]. Thus, satellite-based estimates of DSR are also not a reliable basis for solar forecasting. Instead, our work leverages a ML regression model that infers solar energy output directly from the spectral data, specifically using the channels in the visible range, as the basis for solar forecasting.

Recent work [6, 9–11, 13] has focused narrowly on solar performance modeling—inferring current solar output from current environmental conditions—but not forecasting. Solar forecasting is a much more challenging modeling problem, since it must infer, not only how spectral data correlates with a site's solar energy output, but also how the spectral data will change over time based on the movements of clouds. Accurately forecasting hours, or days, in the future is challenging because of non-linear atmospheric dynamics that affect cloud movements, and which are not directly captured by GOES data. Such long-term forecasting on the order of multiple hours or days requires Numerical Weather Prediction (NWP) algorithms [34], which leverage non-linear physical models of the atmosphere, and require more inputs beyond spectral satellite data.

On the other hand, near-term solar forecasting from satellite data is more tractable, since over short time periods, cloud movements are more heavily impacted by recent phenomenon. As a result, models that incorporate historical spectral data across a region have the potential to track changes in the data over time as they move across a region. Prior work on cloud motion vectors [25, 26] has taken this approach in identifying clouds and tracking cloud movements to assist solar nowcasting. However, they largely use physical models, and do not leverage either the latest multispectral data from GOES-R or recent advancements in forecasting using deep learning. The higher resolution data offered by GOES-R admits more accurate, localized, and near-term forecasts compared to prior work based on coarser and less precise data. Similarly, recent advancements in deep learning offer a more automated "black box" approach that does not require manually calibrating physical models for specific data sources or solar sites.

Finally, another line of research [28, 30, 36, 43, 44] utilizes sky-images collected from specialized cameras installed directly at a solar site, which continuously captures images of the sky at high resolution. Of course, this approach is not applicable to any site without such a specialized sky-imager. Indeed, prior work on such methods has only considered a very small number of solar sites, for instance, a maximum of only 2 solar sites in [28, 30, 36, 43, 44]. In contrast, our approach uses satellite data that is readily available for any location, and thus is widely applicable to any location

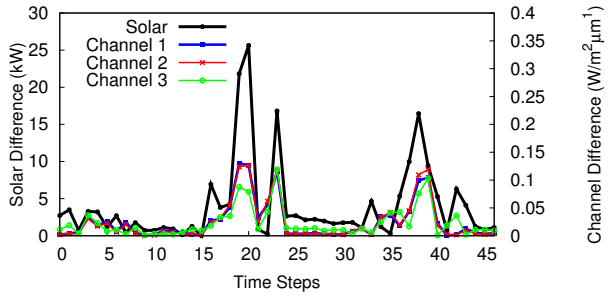

Figure 2: Change in solar output from a home and the first three spectral channels at the same location sampled every 15 minutes. The graph shows that changes in solar output align well with changes in spectral data.

within the GOES-R coverage area, i.e., much of North America. We evaluate our approach to solar nowcasting on 25 solar sites, which is an order of magnitude more sites than prior work.

2.2 Basic Methodology

As we discuss in §3, our approach divides the solar nowcasting problem into two steps: (1) learning a self-supervised model on raw satellite data across all sites of interest, and (2) modeling the future solar energy output at a particular solar site by utilizing predictions from the self-supervised model. The self-supervised model combines a convolutional neural network (CNN) with a long short-term memory (LSTM) for time-series forecasting of spatial multispectral satellite data. CNNs are commonly used for analyzing spatial imagery. Multi-spectral satellite data across a region of some size at any moment in time is akin to an image, where the spectral data is equivalent to a pixel value. In contrast, LSTMs have feedback connections that make them well-suited for forecasting temporal data, but cannot be directly applied to spatial data. Thus, combined CNN-LSTMs are generally useful when the input data has both a spatial and temporal structure, which is the case for satellite data.

Specifically, given a sequence of spectral data covering some area over some previous time steps, our self-supervised model is trained to predict the value of the spectral data at the center location, which corresponds to a solar site's location, in the next time step. In our case, we focus on time-steps of 15 minutes. We focus specifically on 15 minutes because currently ~11% of the grid's generating capacity can be brought online in under 15 minutes [4] to offset renewable variations. In addition, this fraction of fast-response "peaking" generation capacity will need to increase as more solar and wind are integrated into the grid.

Note that our model is "self-supervised," since it only makes use of the raw spectral data to predict subsequent samples in that datastream, which is already captured by the GOES-16 satellite and does not require any site-specific solar generation data. This is akin to self-supervised models in machine learning literature, such as language models [31] that predict the next word in a sentence or general auto-regressive models that predict future samples in a sequence [29]. These models can be further specialized to specific supervised tasks of interest.

To enable site-specific solar nowcasting, we feed predictions of the future spectral data at a site's location from the self-supervised model to a regression model that infers a site's solar output from

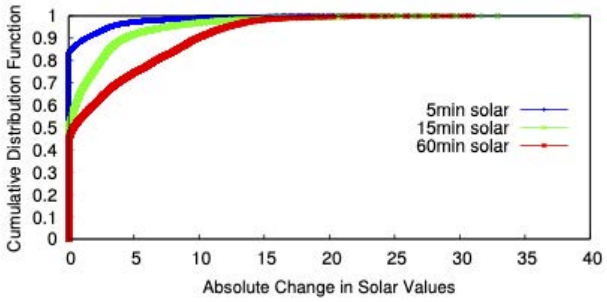

Figure 3: Cumulative Distribution Function of changes in solar energy output at 5, 15 and 60 minute frequencies over one year for a representative solar site.

given spectral data. This regression model incorporates site-specific information that affects solar output from ground-level solar irradiance, such as a site's size, efficiency, orientation, tilt, and shading from obstructions, as we discuss in §3.

We evaluate our solar nowcasting models above in §5 across 25 solar sites over a year. Our evaluation particularly focuses on the accuracy of our models to predict large changes in solar energy output over short time periods. These are the changes that are most disruptive to the grid, and other solar-powered systems. In addition, evaluating solar nowcasting over all time periods obscures the problem, since solar energy output often does not change much within a 5-15 minute period. As a result, a simple persistence model that predicts solar output never changes over 5-15 minutes are highly accurate when averaged over many time periods, even though they are highly inaccurate, by definition, when any change in solar energy output occurs. Figure 2 illustrates this point by showing the change in solar output every 15 minutes, as well as the first three spectral channel values, over a day for a particular solar site. As shown, most of the time, neither the solar output nor the channel value changes significantly. However, there are a few times within the day that experience significant changes. These significant changes are the ones that are most important to accurately predict. This graph also demonstrates the correlation between the spectral channel values sensed by the satellite and a site's solar output: they tend to rise and fall in tandem, although the magnitude of the increase and decrease varies over time.

Figure 3 then shows a Cumulative Distribution Function (CDF) of the change in solar energy generation at 5, 15, and 60 minute periods. This graph shows that, as expected, there are fewer large changes in solar output at small intervals, and the size of the changes are generally larger over longer periods. Thus, accurately predicting the few large changes can be a challenging problem. As a result, we condition our evaluation in §5 on the accuracy of predicting changes in near-term solar output above a specified magnitude. In addition, note that at a 5-minute resolution, close to 80% of the data (value at 0) exhibit no changes in solar in subsequent time steps. Thus, we choose a 15-minute interval for our study, which has more instances (~50%) with non-trivial changes in solar.

3 SOLAR NOWCASTING MODEL DESIGN

In this section, we present our methodology for developing a solar nowcasting model using multispectral satellite data. We first describe our neural network for modelling spatio-temporal satellite

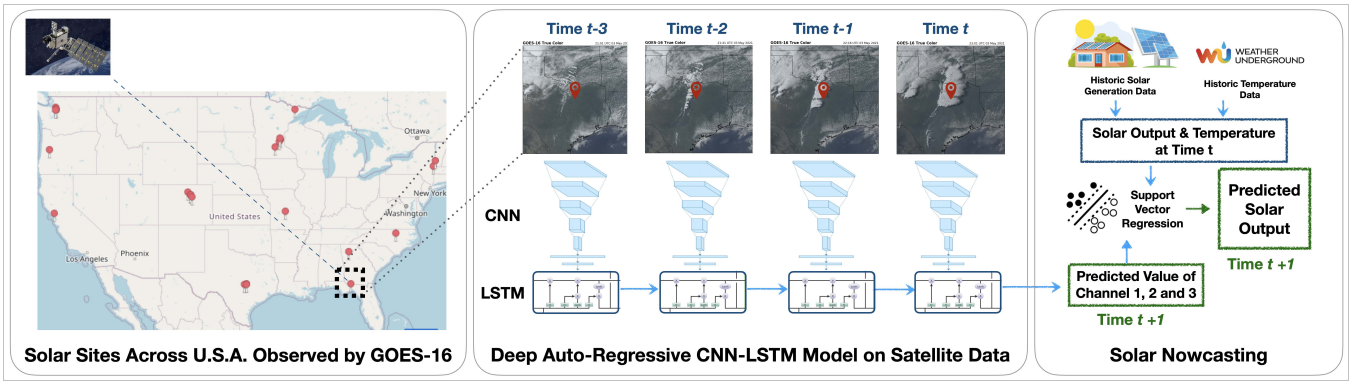

Figure 4: Overview of our modeling pipeline. We collect multispectral satellite data from 25 U.S. sites. We use sequences of observations at each site, specifically the first 3 spectral channels (visualized in the middle), as input to a CNN-LSTM model, which we train to predict the observation at the center (red pin) for the next time step ($t+1$). An auto-regressive SVR then takes the predicted channels and previous solar output, as well as previous temperature, to predict the solar output at next step.

observations, and then describe how to train this model using self-supervised learning on raw satellite data for forecasting satellite channel values at any location. Finally, we feed these forecasted channel values into a regression model that infers a site's solar output. Figure 4 provides an overview of our modeling approach and its two primary stages, which we summarize below.

- **Stage 1**. Train a general model to forecast the three spectral channel values–equivalent to a pixel value—in the future for a site's location using the three spectral channel values from the surrounding area, both currently and in the past (§3.1-3.3).
- **Stage 2**. Train a separate site-specific model to infer solar output from the three spectral channel values, as well as the current solar output and temperature, and feed the forecasted channel values from Stage 1 to forecast solar output (§3.4).

Given a set of solar sites of interest, we consider an area (or matrix) of $w \times w$ around each site, and extract the 3 visible channels of satellite data from the GOES-16 satellite for each element of the area's matrix. Note that each location l within the area is described by a 3-tuple of channel values. We denote each such 3D image of dimensions $(w \times w \times 3)$ at a location l observed at instant t as $I_t^{(l)}$. We then extract a temporal sequence of these images over time from the satellite, with the target site l always at the center of the image as this is a geostationary satellite. This data effectively has four dimensions described by the 2D area (length and width), three channel values, and time. We vary both the area and the amount of historical data (or time resolution) that we use for training our models, as the optimal values depend on a location's climate and the target forecast horizon. For example, we could use the previous 4 satellite images for a region of size w with an interval of 15 minutes between images, or the last hour of changes in the spectral data. Larger areas and longer historical time periods increase the training data size, which increases the computational overhead of training. As we discuss, however, there are diminishing returns with respect to improvements in accuracy as these values increase.

3.1 Spatial Modeling using CNNs

Data from the first three spectral channels over an area forms a 3D image $I_t^{(l)}$ that we first process using a Convolution Neural Network (CNN) to extract spatial features from the image. A CNN model [24] is the standard neural network architecture used for modeling visual imagery and extracting visual features. CNN models are comprised of trainable convolution filters and pooling operations that together extract spatially invariant features from images. We use multiple layers of convolution filters followed by max pooling layers. The exact CNN architecture is described in §4.2. The output of processing the image with the CNN model is a k-dimensional feature vector:

$$v_t^{(l)} = CNN(I_t^{(l)}; \theta)$$

where θ represents trainable parameters of the CNN model and $v_t^{(l)}$ is the extracted d-dimensional spatial feature vector for location l at time t. Note that the CNN is trained to extract spatial features that help model the temporal dynamics of the satellite data as explained below.

3.2 Temporal Modeling using LSTMs

Short-term forecasting of solar energy generation should also account for the recent history of changes to solar irradiance at the surface, and how it will evolve in the near term. We use long short-term memory networks (LSTM) [20] to capture the evolution of the per-instant spatial features extracted from the CNN over time, which is crucial for predicting future satellite channel values. The LSTM is a prominent neural architecture used for modeling sequences of data and is also often employed in time-series forecasting. LSTMs make use of both a cell state, which is an internal memory summarizing the previous history at a given time, and a hidden state, which is the output of the current time step. Multiple gating mechanisms update the cell state by combining it with the current input and the previous hidden state. In our case, the LSTM update at step t is summarized as:

$$s_t^{(l)}, h_t^{(l)} = LSTM(s_{t-1}^{(l)}, h_{t-1}^{(l)}, v_t^{(l)}; \phi)$$

where ϕ are the LSTM trainable parameters, s_t is the LSTM cell state, h_t is the hidden state, and $v_t^{(l)}$ is the CNN spatial feature vector at time t for location l. Thus, recursively reapplying the same function at every time-step, the LSTM models the evolution of the input features over time.

After processing a sequence of T images through the CNN and LSTM, for instance the satellite imagery over the previous one hour, the final hidden state h_T of the LSTM summarizes the entire sequence. This state is passed through a dense layer with sigmoid output units to predict the value of the visible channels at the site's location for the next step:

$$\hat{C}_{T+1}^{(l)} = \sigma\left(W h_T^{(l)}\right)$$

where W is a $(3 \times k)$ matrix with k being the hidden state dimension and $\hat{C}_{T+1}^{(l)}$ are the predicted values of the 3 channels of satellite at the next time instant.

3.3 CNN-LSTM Model Training

We train the CNN-LSTM model end-to-end in an auto-regressive manner. That is, given the sequence of past images we use the model to compute the predicted values for the next instant and use mean-squared error with respect to the true future satellite values as the loss function:

$$\min_{\theta,\phi} \sum_{l,T} \left\| \hat{C}_T^{(l)} - C_T^{(l)} \right\|^2$$

where C_T^l is the ground-truth satellite observations for all 3 channels at time T and location l, $\hat{C}_T^l$ is the prediction from the CNN-LSTM model, as described above, and $\|\cdot\|$ represents the euclidean norm. Note that the prediction is a function of both the CNN spatial extraction model and the LSTM temporal extraction model, such that backpropagation optimizes the parameters of these models to extract features that can predict future observations well. Thus, using widely available satellite data, we train CNN spatial extraction and LSTM temporal models to capture the dynamics of the multi-spectral satellite data. Note that, in our evaluation, we only train one global model by combining satellite data across multiple sites. This enables modeling of shared statistical properties across sites rather than overfitting to the peculiar characteristics of any single site. In addition, as discussed below, this approach also provides a large amount of data for learning a useful CNN-LSTM model, which typically are not accurate when trained on small datasets.

The drawback of global modeling is that it does not account for unique aspects of any specific location's climate. A local model trained only on data from a specific location is capable of identifying unique attributes of a location's climate to improve accuracy. For example, in some locations, winds may typically move west to east, while in others, they may typically go in the other direction. However, training our models requires a significant amount of data, and, since the GOES-R satellites only began releasing data a few years ago, there is not yet a large volume of data available for training and testing on any single location. A global model is also beneficial because it does not require re-training, and can be applied to any location.

3.4 Nowcasting Solar Energy Output

As shown in Figure 2, GOES spectral data highly correlates with solar irradiance at the surface, which enables accurate inference of solar energy output via machine learning models trained on historic generation data using GOES spectral data as input. We can leverage this relationship for solar forecasting by inferring future solar generation from the forecasted spectral values by our CNN-LSTM model.

That is, given a trained CNN-LSTM model that can generate future satellite observations at a given site, we leverage these predictions in a model for solar energy forecasting at any solar installation site of interest. We leverage this relationship between visible bands and solar irradiance by considering the following auto-regressive model for forecasting near-term solar output:

$$P_{t+1}^{(l)} = f(P_t^{(l)}, C_{t+1}^{(l)}, T_t^{(l)}) \quad (1)$$

where P_t is the solar energy generated, $C_t^{(l)}$ are the satellite channel values and $T_t^{(l)}$ is the temperature at time t. $f(\cdot)$ is a regression model, such as support vector regression (SVR), that models the relationship between the input and output variables using historical data. Temperature is an important component of solar generation as solar efficiency is sensitive to temperature [13]. Note that we use $C_{t+1}^{(l)}$ instead of $C_t^{(l)}$ in (1). Note that we do not include a forecasted temperature as an input since i) temperature does not change significantly over short time periods, e.g., 15 minutes, and ii) temperature forecasts are typically not released at 15 minute resolutions. A major component of change in P_{t+1} from P_t is captured in the change in C_{t+1} from C_t. This complex relationship is modeled using our CNN-LSTM model, described above, which predicts an estimate $\hat{C}_{t+1}^{(l)}$, which is an estimate of true channel values at $t+1$ for the auto-regressive model in (1).

The regression model for solar nowcasting in (1) is trained using current satellite observations. That is, $f(\cdot)$ is trained using the ground-truth satellite observations at $t+1$ time-step ($C_{t+1}^{(l)}$), historic solar output at the location l, and the historic temperature data at location l. This step does not require the use of the CNN-LSTM model, and is necessary to train an accurate auto-regressive model that, given true satellite observations, can accurately infer future solar energy output. Once the regression model is trained, instead of using true future satellite observations, which are unavailable, we replace them with estimates from the CNN-LSTM model ($\hat{C}_{t+1}^{(l)}$) to compute the forecast.

4 IMPLEMENTATION

4.1 Satellite Data and Solar Sites

GOES-16 multispectral data is made publicly available by NOAA as netCDF files hosted on Amazon S3 buckets. We recursively download the data for each date each year along with the description of the data product, bucket, domain, and the satellite name. The size of each 5 minute netCDF file is in the range of ~75MB, which requires nearly 16 terabytes to store two years of data from a single GOES-R satellite. Each 5 minute file includes data across 16 spectral bands covering the entire American subcontinent. To minimize storage requirements, we filter each file as we download it to extract only

the relevant spectral channel data for the specific area around our site locations of interest, and discard the rest.

The netCDF files for our multispectral data require some processing to filter out the data for the location of interest. Specifically, we implemented python modules to read the *goes_imager_projection* variable to convert (x, y) degree coordinates for latitude and longitude to radians. We then search the file for the latitude-longitude pair that is closest to our location of interest. Since the GOES satellites are geostationary, their rotation matches that of the Earth, enabling us to look at the same part of the file each time. Thus, we read a single netCDF file and first create a list of the closest latitude-longitude pairs using the Vincenty formula [40], which calculates the distance between two points on the surface of a spheroid. This step reduces computational resources, since it eliminates the need to repeat this process for each 5 minute file.

Our evaluation uses data from 25 U.S. solar sites across two years. Specifically, we extract satellite data for the continental U.S. in 2019. We restrict our modeling to a 10×10 window around each of the 25 solar sites, which constitutes the training data for the CNN-LSTM model for computational efficiency. The 10×10 window covers an area of approximately 10km^2. We average observations across a 15-minute window, which reduces the sequence length for modeling and the noise in the data by reducing the number of missing observations and sensor errors. The solar sites we use in this work are shown in Figure 4, and are uniformly spread across the continental U.S., including both coasts and the central regions. Since solar forecasting is only relevant during the daytime, we restrict our satellite data to be from 9am in the morning to 5pm in the evening based on the local time of each solar site. This yields more than 300,000 5-step sequences of 10×10 images (with 3 channels) at intervals of 15 minutes.

We use 5-fold validation in all of our experiments, splitting by day so that test sets include entire days held out for evaluation. This is done for both types of evaluation: evaluating channel prediction models and evaluating end-to-end solar nowcasting. Table 1 shows the training, validation, and test split of the satellite observations used in this work. Solar generation data from the energy meters for the same 25 sites and temperature data from the weather station are obtained for years 2018-19. We also restrict our generation data to be from 9am to 3pm each day, which is the peak duration of solar generation. Finally, given our dataset's scale, training the full CNN-LSTM model on a $10 \times 10\text{km}^2$ area with 4 previous time-steps of historical data requires ~86 GPU-hours. However, the inference time for prediction is much less than the 5 minute interval between satellite data readings. Specifically, for our model implementation which is not expressly optimized for latency, the inference time is only 72ms. Thus, while our model takes significant resources to train, it can be used for predictions in *real time*.

4.2 Model Hyper-parameters and Metrics

Our CNN model includes 2 blocks of convolutions, where each block contains 2 convolution layers with 32 filters of size 3×3 and ReLU activation followed by a max-pooling layer of size 2×2. These layers are followed by two dense layers with hidden dimension $k = 256$ and ReLU non-linearity between layers. We use a one layer LSTM that takes these 256 dimensional inputs and has a hidden state dimension of 64. Hyper-parameters for this and the other ML models we consider, specifically decision tree and random forest, were determined using the validation set.

Data Sets	*Number of points*	*Number of days in a year*
Training	236375	262
Validation	27040	30
Testing	65258	73

Table 1: Total number of sequences as well as the number of days of the year that comprise training, validation and testing for the CNN-LSTM model.

We use two metrics to evaluate the performance of our channel forecasting models and end-to-end solar forecasting models. Our first metric is the Mean Absolute Percentage Error (MAPE), which quantifies the average percentage across time.

$$MAPE = \frac{1}{n}\sum_{t=0}^{n}\left|\frac{A_t - P_t}{A_t}\right| \qquad MAE = \frac{1}{n}\sum_{t=0}^{n}|A_t - P_t|$$

Here, A_t and P_t represents the actual and predicted values. MAPE, which is often used to quantify the performance in prior work [42], is an intuitive metric and is comparable across solar sites of different installation sizes and configurations, which each have a different maximum generation. However, MAPE is highly sensitive to periods of low absolute solar generation and thus can be significantly affected by small absolute errors during periods of low generation. Thus, we also use mean absolute error (MAE) to quantify the error in channel modeling given that the first three channels are reflectance values strictly in the range of 0 to 1. In this case, similar to MAPE, the MAE of channel predictions are also comparable across locations and time. Specifically, when assessing the accuracy of spectral channel predictions, MAE is similar to MAPE, since the channel values are bounded between 0 and 1. That is, an MAE of 0.15 essentially means an error of 15% of the total range. In our graphs, we plot MAE×100 unless otherwise stated.

5 EVALUATION

In this section, we evaluate our proposed CNN-LSTM for deep auto-regressive modeling on satellite data and its accuracy for end-to-end solar nowcasting. First, we evaluate the efficacy of the CNN-LSTM model for predicting future values of satellite channels for a given location. We consider our evaluation along spatial and temporal axes, as well as consider alternative ML models. Then, we utilize our trained CNN-LSTM model for solar nowcasting at each site, and quantify its accuracy. We use the following terminology throughout the evaluation:

Persistence Model: Since all of our models predict values for the next instant, typically 15 minutes in the future, a natural baseline is one that assumes there will be no change in the predicted quantity, which is generally called a persistence model. As discussed earlier, solar output often changes in small, abrupt bursts and thus a large fraction of the time there is negligible change in near term solar output (see Figure 3). Thus, improving the persistence model's prediction's is challenging, and serves as an important baseline.

Note that weather forecasts, i.e., released by the National Weather Service in the U.S., are based on numerical weather

Tolerance	*% of points C01*	*% of points C02*	*% of points C03*
0	100	100	100
0.01	65.24	68.33	80.56
0.02	47.32	49.95	64.54
0.05	22.0	24.34	35.39
0.10	8.38	9.63	13.29

Table 2: Variation in number of data points with respect to different tolerances for different channels.

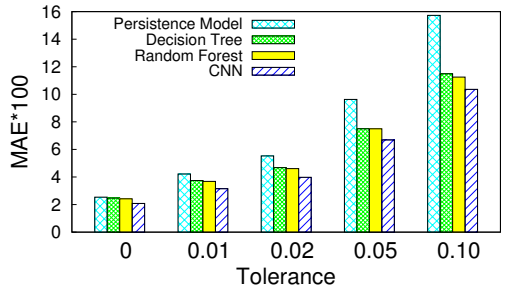

Figure 5: MAE for the different ML models we evaluate in predicting the next spectral channel value 15 minutes in the future. The CNN model yields the lowest error.

prediction models that have temporal and spatial resolutions that are too coarse for direct comparison. Specifically, typical weather forecasts are released at hourly (or coarser) resolutions and typically cover regions larger than $10 \times 10\text{km}^2$. As a result, using these forecasts for 15-minute-ahead predictions is equivalent to using the persistence model above.

Tolerance: Since, over the year, changes in solar energy potential (and hence satellite observations) over short periods, such as 15 minutes, are often negligible, we conduct our analyses subject to varying thresholds of changes in solar. That is, we define a tolerance δ and consider only points x_i where subsequent changes were at least δ: $\{x_i | \ |x_{i-1} - x_i| \geq \delta\}$. We evaluate all models over a range of different values of δ to provide a sense of how they perform over both small and large sudden changes in solar. Table 2 lists the fraction of points in the validation data for each tolerance value considered.

Forecast Skill Score: We also use "forecast skill score" (SS) to compare the performance between various methods, which is commonly used in prior work [42, 43] and given by:

$$SS = \left(1 - \frac{\mathcal{E}_{\text{prediction}}}{\mathcal{E}_{\text{baseline}}}\right) * 100\%$$

Here, $\mathcal{E}$ is the error metric used to evaluate the performance for every model. If the *prediction model* performs equally well as the *baseline model*, the skill score will be 0. A higher skill score thus means that the prediction model outperforms the baseline model. We will use the skill score to compare the performance between different models. In this work, for the skill score, our baseline is always the persistence model discussed above.

5.1 Evaluating ML models for spatial modeling

In this section, we consider different choices of ML models for spatial modeling and evaluate their utility compared to using a CNN.

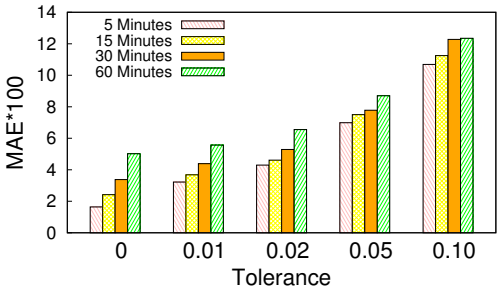

Figure 6: MAE for models trained and evaluated at different time intervals. At larger intervals (30-60 minutes) forecasting becomes increasingly more challenging.

We also consider the size of the spatial area and its effect on predicting future channel values. For this purpose, we consider three standard ML models: (1) Decision Tree, (2) Random Forest, and (3) Convolution Neural Networks (CNNs). Note that we only analyze next step predictions given current data, i.e., a lag-1 time-series models, and consider longer history temporal modeling in the next section. We only show results for channel 1 to avoid repetition as results for the other two channels are qualitatively and quantitatively similar. To train our decision tree and random forest models, we flatten the $w \times w$ spatial satellite observation into a vector of size $w * w$ that is input to the model.

Figure 5 shows the mean absolute error (MAE) for all three models at different tolerances. As mentioned earlier, all models are trained over a 10×10 area around each solar site and predict the satellite observations in the next 15 minutes. The point of this graph is to show how the accuracy of a deep learning approach improves relative to that of simpler non-spatial models as the size of the subsequent change increases. At 0, which represents all of the data points, the CNN model is only marginally better than the other models. However, this occurs primarily because most of the time there are only small changes in solar over short time periods, as evident from the low error of the persistence model. As we increase the size of the changes we examine, we see that the persistence model's predictions, which assume the past is the same as the future, become increasingly worse, while the CNN model remains the best and improves over the others by a large margin.

Figure 6 then shows the effect on forecast error for models trained and evaluated at intervals of 5, 15, 30 and 60 minutes. We use the random forest model for this evaluation, since training a CNN for every setting is expensive. As expected, predicting further into the future is less accurate, since more changes occur. This discrepancy in accuracy is most evident at 0 tolerance when we include all the data points. This occurs because there are few changes in solar output over 5 minutes on average, while on average there are much more significant changes over 60 minutes, including changes due to movement of the sun in the sky. As we increase the tolerance to assess the accuracy of predicting larger changes, as expected, the error increases. However, interestingly, the discrepancy in error actually decreases. That is, the error in predicting a large change 30-60 minutes in the future is more similar to predicting a large change 5-15 minutes in the future. This result highlights that accurately predicting large changes in solar output is challenging even over small forecast horizons.

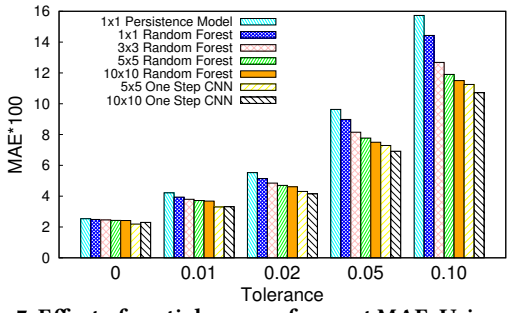

Figure 7: Effect of spatial area on forecast MAE. Using a larger area improves the forecast for all models with CNN trained using a (10×10) area yielding the lowest error.

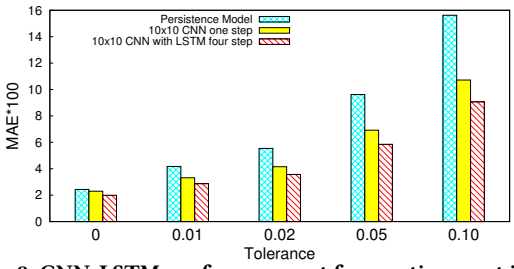

Figure 8: CNN-LSTM performance at forecasting next instant channel values. Compared to not using previous temporal history (CNN one step), CNN-LSTM leads to significant error reduction on predicting large changes (tolerance>0) and retains overall better performance compared to persistence model (tolerance$=0$).

Figure 7 then compares the effect of using different sizes of the spatial area, focusing specifically on the two best models from Figure 5. We train models with areas of 1×1 (i.e. just the site), 3×3, 5×5 and 10×10. The graph shows that increasing the spatial area around the site used by the model results in a large improvement in accuracy. We see that using a 10×10 area yields the best result, and is more accurate than not considering any surrounding area from the site (1×1). Moreover, using the CNN model results in much better spatial processing and improved results over variants of random forests.

Key Points. *Our CNN model is more accurate than the other models with the difference in accuracy becoming much more pronounced as the magnitude of the changes in channel value increases (Figure 5). We also show that the model error is sensitive to both the forecast horizon (Figure 6) and the spatial area (Figure 7).*

5.2 Evaluating Joint CNN-LSTM Models for Spatio-Temporal Modeling

We next add an LSTM model on top of the 10×10 area CNN model and utilize the previous timesteps as input to the CNN-LSTM model. In this case, we use 4 steps, which means we train the model on a dataset that includes the 4 previous 10×10 spatial regions, corresponding to the past 1 hour of spatial observations. The overall results are shown in Figure 8, comparing the CNN-LSTM model with a single step CNN model using a 10×10 spatial area. The use of the LSTM model significantly improves the results in terms of accurately predicting large changes. Similar to the previous graph, when evaluating across all of the data where the tolerance is 0, the improvements over a single step CNN are not significant, since most of the time there are only small changes in solar output. However, the advantage of the CNN-LSTM becomes apparent when we focus on predicting any significant change larger than 0, i.e., a tolerance ≥ 0.01. As we increase the tolerance threshold, we observe that our model that covers a 10×10 area and combines a CNN-LSTM leads to increasingly larger reductions in errors.

We next evaluate the performance of CNN-LSTM variants in forecasting next time instant channel values. We explore the following temporal variants: CNN using a 1-step static image, CNN-LSTM using a 1-step static image, and a CNN-LSTM using 2, 3, and 4 steps of images in the past. Figure 9 shows our results compared with the persistence model predictions. Incorporating multiple steps of information in our CNN-LSTM is better than using the current static image for forecasting, showing the utility of a deep auto-regressive approach. We find that using 3 or 4 steps, i.e., 45 minutes or 60 minutes in the past, perform comparably. This is likely due to the fact that clouds from outside our 10×10 area have time to move over our sites within 45-60 minutes. Incorporating more historical data likely requires increasing our spatial area. Notably, though, using even just 2 past time steps leads to a marked reduction in error. This signifies that our model is able to infer temporal changes in the satellite data, such as cloud movement, for better predictions.

Key Points. *Extending our CNN model with an LSTM further improves accuracy by considering historical data, especially for larger changes (Figure 8). The marginal improvement in accuracy diminishes when using more than 30 minutes (2 steps) of historical data (Figure 9), likely because clouds from outside our $10 \times 10 km^2$ spatial area move into it within 30 minutes.*

5.3 End-to-End Solar Nowcasting

Our primary goal is to leverage the spectral channel predictions above to perform end-to-end solar nowcasting at a solar site. In this section, we evaluate the utility of our prediction models for this purpose. Since we want a clear comparison of the benefit of using the self-supervised CNN-LSTM model for solar nowcasting, we use the CNN-LSTM model as a fixed model for solar nowcasting. That is, after the self-supervised learning on raw satellite observations, this model is fixed and not trained further on any site-specific data from a solar site. This enables us to decouple the contribution of the predictions from the self-supervised model in solar nowcasting–if the predictions are useful, it will improve nowcasting results over using a persistence model's predictions. Moreover, this enables faster computation and cheaper memory overhead as the expensive CNN-LSTM model is not trained on each of the many solar sites.

We use the SVR auto-regressive model, discussed in 3.4, to forecast 15-minute ahead solar energy generation. We consider 4 different models to evaluate our forecast at time t:

- *Solar Persistence Model*: a simple past-predicts-future baseline that predicts the solar energy output in the next time step remains the same and does not change;
- *CNN-LSTM-SVR*: an SVR model using the predictions of our CNN-LSTM, which include the forecasted channel values from

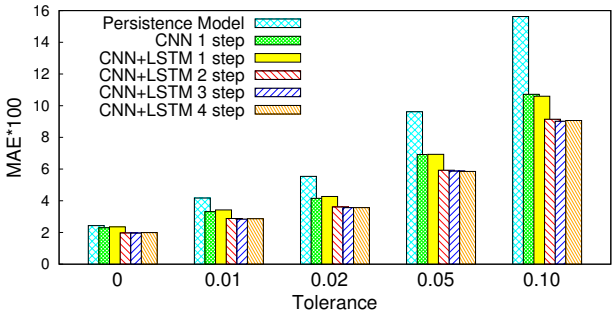

Figure 9: Comparison of CNN-LSTM models using varying amounts of previous temporal information. Note that using 3-4 steps is better than using lesser steps. There is a marked reduction in error when using more than 1 step in the model.

past values (C_{t-1},C_{t-2},C_{t-3}) using the self-supervised CNN-LSTM model;

- *SVR(C_{t-1})*: using the persistence model on satellite observations, C_{t-1}, as the forecast input to SVR instead of CNN-LSTM forecast, this should be an upper-bound on the error only if the CNN-LSTM model produces useful forecasts;
- *SVR(C_t)*: this is a lower-bound on the error that uses the ground-truth satellite observation at the *future* instant and is not a feasible forecast as C_t is unavailable ahead of time.

Note that SVR(C_t) uses the current satellite observations to make predictions. This is a lower-bound on the error of the model given a particular site with some historical data when using SVR auto-regressive models. Estimates of satellite channel values $\hat{C}_t$ from a model using a previous time instant's observations will be useful if they lead to an accuracy that is closer to the performance of using the actual observations. Similarly, SVR(C_{t-1}) corresponds to using a persistence model's predictions as the future satellite channel estimates, assuming no change in values, and serves as an upper bound on the error. A model's error should be below this value for it to be useful for solar nowcasting.

Figure 10 shows our result for two scenarios: only over summer months and over the whole year. We include both scenarios, as typically forecasting is less challenging over summer months, due to largely sunny days, and more difficult over the rest of the year, due to a higher frequency of rain, clouds, and snow. As shown, the performance of forecasting solar using our CNN-LSTM is close to using the ground truth channel values from the future in the model, an upper-bound, and hence shows that the approach is useful and accurate for solar forecasting. We have further split the performance of these models into the percent changes between successive solar generation values, as shown on the x-axis, where 0 means any change and includes all the values, whereas N% means a change of at least N% in subsequent values. We can also compare the results in the left and right plots of Figure 10 in that they both show similar trends but only differ in the MAPE, which is higher for a full year and slightly lower across the summer months. Interestingly, we find these models are not significantly worse over non-summer months, which indicates that they capture rich spatio-temporal phenomenon from the satellite data for accurate modeling.

Skill score is a popular metric used to understand the performance of solar forecasting models. For solar nowcasting, the persistence model is the default baseline model that is used in prior work [28, 37, 42, 43]. Figure 11 shows the average of the skill score across all the 25 sites at different tolerances. In addition, we show the distribution of the number of data points available for evaluation at the various tolerances through a histogram. We can see that the solar nowcasting model improves over the persistence baseline, yielding an average skill score in the range of 14-19%. We see that over a full year, the skill score improves as we increase the tolerance of subsequent changes and then drops a bit at predicting very large changes of more than 5%. Interestingly, skill score is consistently high at predicting very large changes during summer months.

Figure 12 then shows the distribution of forecast skill across the 25 solar sites. We can see that the skill varies widely across solar sites, from 14%-27% across the 25 sites. These variations are expected as different sites have different characteristics that contribute to inaccuracy, including differences in installation capacities, shading from nearby buildings or trees, and widely different climates, e.g., sunny versus rainy and snowy. In particular, note that prior work on solar nowcasting using sky-camera imagery has found that state-of-the-art nowcasting models using deep learning have a skill in the range of 10%-20% [28, 37], as evaluated on only 1 or 2 solar sites, which is typical in prior research on sky-camera nowcasting. The results in Figure 12 indicate that our results are competitive or better, while our approach is much more scalable and cost-efficient, as it does not require installing specialized hardware at every solar site.

Key Points. *Our approach for end-to-end solar nowcasting has i) a low error overall and is increasingly more accurate at predicting large changes compared to our baseline methods (Figure 10) and ii) a forecast skill score that is near or better than prior work that performs solar nowcasting by analyzing ground-level sky-imagery (Figure 11).*

6 RELATED WORK

Forecasting solar energy output is akin to forecasting solar irradiance, since the former strongly correlates with the latter [33]. Numerical Weather Predictions (NWP) algorithms [14, 17, 27, 38], which mostly leverage physics-based modeling, are often used for solar irradiance forecasting. These physics-based models are most appropriate for forecast horizons on the scale of hours to days, and not near-term forecasts on the scale of minutes to an hour [19, 42]. Over long-term horizons, the complex and non-linear evolution of climate patterns can be difficult to model, requiring knowledge of climate processes and the history of many atmospheric events over time that can cause subtle changes in their movement and intensity.

In contrast, at shorter time scales of 5 to 60 minutes, machine learning approaches have the potential to implicitly model local changes directly from observational data [35, 42]. While there has been recent work on analyzing images from ground-based sky cameras [30, 36, 43, 44] for near-term solar forecasting, it requires installing additional infrastructure at each site. Another alternative is based on estimating cloud motion vectors [15, 25, 26] from satellite images, however ML approaches that more directly model solar irradiance tend to perform better [9, 23]. Our approach differs from recent approaches in solar nowcasting by forecasting solar

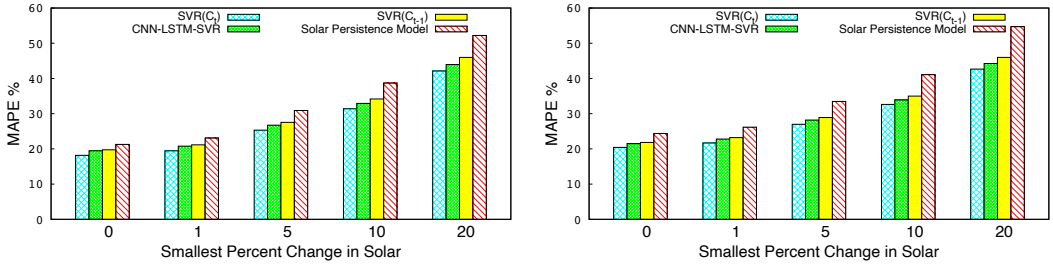

Figure 10: End to end solar forecasting on 10x10 km area, averaged over 25 solar sites over 15 mins. Performance for summer months (May-September) is shown on left and for the full year on the right. Using the predictions from the CNN-LSTM model leads to solar output forecasting with error close to that of using the current satellite observations. Compared to precision model forecasts, this approach is consistently better, especially at predicting when there will be large changes in solar ($\geq$ 5%).

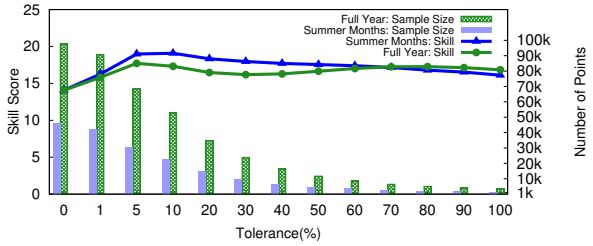

Figure 11: Forecast skill score distribution over summer months and full year, along with the distribution of the number of instances at different tolerances.

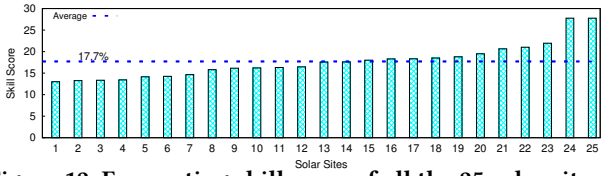

Figure 12: Forecasting skill score of all the 25 solar sites at 5% tolerance for the full year.

irradiance values from multispectral satellite data using a combined CNN-LSTM, which can forecast changes in spatial features over time. We then combine these solar irradiance forecasts with a model that predicts a site's solar output from solar irradiance.

Our approach is self-supervised in that we directly use abundant satellite data for modeling. Such methods have gained increasing popularity in computer vision recently [22]. While self-supervised methods have been widely successful, their application to remote sensing have been limited and their application specifically to solar modeling have not been explored in prior work. Jean et al. [21] similarly uses self-supervised learning over Landsat images, although their approach is designed for classifying geographical regions and not directly applicable to solar nowcasting. Vincenzi et al. [41] reconstructs visible bands from other bands in a colorization task to learn useful representations for land cover classification. Related to our work, Ayush et al. [8] uses temporal information for constructing positive-negative pairs for classification of remote sensing data. However, unlike our approach, the authors ignore complex spatio-temporal dynamics and auto-regressive modeling, which are more crucial for forecasting. Recently, parallel work [32] utilized generative modeling on radar data for precipitation nowcasting using generative adversarial networks. Compared to this, we focus on a different application of solar nowcasting and are able to demonstrate the utility of a simpler model for this application where directly predicting the next instant values is sufficient, as they directly correlate with solar irradiance [10], and do not necessitate requiring a complex discriminator for learning, as in GAN models [18]. Finally, our work relates to auto-regressive language models, which predict the next word given previous words and have been highly successful for natural language processing [31].

7 CONCLUSION

Our work shows how to apply deep learning to multispectral satellite data to forecast short-term changes in solar output. We propose deep auto-regressive models that combine CNN and LSTMs trained in a self-supervised manner on abundant satellite data from GOES-R satellites. Such self-supervised training captures rich spatio-temporal dynamics that help improve solar nowcasting and is readily applicable to any solar site of interest, that is captured by the GOES-R satellites, without requiring any specialized hardware, as in prior work. We evaluate our approach for different coverage areas and forecast horizons across 25 solar sites. Our results demonstrate that 15 minute forecasts using our approach have an error near that of a solar model using current weather and have forecast skill that is comparable with highly localized methods that require the installation of specialized sky cameras.

While this is a promising first step, we believe there is much progress to be made in this area, as self-supervised learning is a promising approach for rapid progress in this domain due to the abundant availability of rich satellite data. Our self-supervised learning approach itself can be improved by exploring longer contexts – both in input and output, spatially and temporally – through more sophisticated recent neural networks, such as Transformers [39] which can model longer range dependencies. In addition, including other features, such as time-of-day and cloud location, and training local models based on the unique weather patterns of specific locations has the potential to further improve accuracy.

Finally, while solar nowcasting is our primary focus, the self-supervised models we develop are also generally useful for many other applications, such detecting anomalies, e.g., wildfires, forecasting cloud cover, and precipitation nowcasting.

Acknowledgements. This work is funded, in part, by NSF grants #2021693 and #2020888.

REFERENCES

[1] 2018. *GOES-R Advanced Baseline Imager (ABI) Algorithm Theoretical Basis Document for Downward Shortwave Radiation (Surface), and Reflected Shortwave Radiation (TOA).* Technical Report. NOAA NESDIS Center for Satellite Applications and Research.

[2] 2020. Data Products: Downward Shortwave Radiation (Surface). https://www.goes-r.gov/products/baseline-DSR.html.

[3] 2022. DarkSky for iOS. https://darksky.net/app.

[4] 2022. U.S. Energy Information Administration, About 25% of U.S. power plants can start up within an hour. https://www.eia.gov/todayinenergy/detail.php?id=45956.

[5] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*. 265–283.

[6] R.W. Andrews, J.S. Stein, C. Hansen, and D. Riley. 2014. Introduction to the Open Source pvlib for Python Photovoltaic System Modelling Package. In *IEEE Photovoltaic Specialist Conference*.

[7] John A. Augustine, John J. DeLuisi, and Charles N. Long. 2000. SURFRAD - A National Surface Radiation Budget Network for Atmospheric Research. *Bulletin of the American Meteorological Society* 81, 10 (2000), 2341 – 2358. https://doi.org/10.1175/1520-0477(2000)081<2341:SANSRB>2.3.CO;2

[8] Kumar Ayush, Burak Uzkent, Chenlin Meng, Kumar Tanmay, Marshall Burke, David Lobell, and Stefano Ermon. 2020. Geography-aware self-supervised learning. *arXiv preprint arXiv:2011.09980* (2020).

[9] A. S. Bansal and D. Irwin. 2020. Exploiting Satellite Data for Solar Performance Modeling. In *2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm)*. 1–7. https://doi.org/10.1109/SmartGridComm47815.2020.9302984

[10] A. S. Bansal and D. Irwin. 2020. See the Light: Modeling Solar Performance Using Multispectral Satellite Data. In *Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys)*. 1–10. https://doi.org/10.1145/3408308.3427610

[11] Noman Bashir, Dong Chen, David Irwin, and Prashant Shenoy. 2019. Solar-TK: A Data-driven Toolkit for Solar PV Performance Modeling and Forecasting. In *2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*. IEEE, 456–466.

[12] K.A. Browning and C. Collier. [n. d.]. Nowcasting of Precipitating Systems. *Reviews of Geophysics* 27, 3 ([n. d.]), 345–370.

[13] D. Chen, J. Breda, and D. Irwin. 2018. Staring at the Sun: A Physical Black-box Solar Performance Model. In *BuildSys*.

[14] Kunjin Chen, Ziyu He, Kunlong Chen, Jun Hu, and Jinliang He. 2017. Solar energy forecasting with numerical weather predictions on a grid and convolutional networks. In *2017 IEEE Conference on Energy Internet and Energy System Integration (EI2)*. 1–5. https://doi.org/10.1109/EI2.2017.8245549

[15] Sylvain Cros, Olivier Liandrat, Nicolas Sébastien, and Nicolas Schmutz. 2014. Extracting cloud motion vectors from satellite images for solar power forecasting. *International Geoscience and Remote Sensing Symposium (IGARSS)*, 4123–4126. https://doi.org/10.1109/IGARSS.2014.6947394

[16] EIA. 2021. EIA projects renewables share of U.S. electricity generation mix will double by 2050. https://www.eia.gov/todayinenergy/detail.php?id=46676.

[17] Harold Gamarro, Jorge E. Gonzalez, and Luis E. Ortiz. 2019. On the Assessment of a Numerical Weather Prediction Model for Solar Photovoltaic Power Forecasts in Cities. *Journal of Energy Resources Technology* 141, 6 (03 2019). https://doi.org/10.1115/1.4042972 arXiv:https://asmedigitalcollection.asme.org/energyresources/article-pdf/141/6/061203/6395250/jert_141_06_061203.pdf 061203.

[18] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. *Advances in neural information processing systems* 27 (2014).

[19] Yan Hao and Chengshi Tian. 2019. A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting. *Applied energy* 238 (2019), 368–383.

[20] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[21] Neal Jean, Sherrie Wang, Anshul Samar, George Azzari, David Lobell, and Stefano Ermon. 2018. Tile2Vec: Unsupervised representation learning for spatially distributed data. arXiv:1805.02855 [cs.CV]

[22] Longlong Jing and Yingli Tian. 2020. Self-supervised visual feature learning with deep neural networks: A survey. *IEEE transactions on pattern analysis and machine intelligence* (2020).

[23] Jesus Lago, Karel De Brabandere, Fjo De Ridder, and Bart De Schutter. 2018. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. *Solar Energy* 173 (2018), 566–577.

[24] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation* 1, 4 (1989), 541–551. https://doi.org/10.1162/neco.1989.1.4.541

[25] Elke Lorenz, Annette Hammer, and Detlev Heinemann. 2004. Short term forecasting of solar radiation based on satellite data. *EUROSUN2004 (ISES Europe Solar Congress)* (01 2004).

[26] Elke Lorenz and Detlev Heinemann. 2012. Prediction of solar irradiance and photovoltaic power. (2012).

[27] Patrick Mathiesen and Jan Kleissl. 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Solar Energy* 85, 5 (2011), 967–977. https://doi.org/10.1016/j.solener.2011.02.013

[28] Yuhao Nie, Yuchi Sun, Yuanlei Chen, Rachel Orsini, and Adam Brandt. 2020. PV power output prediction from sky images using convolutional neural network: The comparison of sky-condition-specific sub-models and an end-to-end model. *Journal of Renewable and Sustainable Energy* 12, 4 (2020), 046101.

[29] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. 2018. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748* (2018).

[30] Quentin Paletta and Joan Lasenby. 2020. Convolutional Neural Networks applied to sky images for short-term solar irradiance forecasting. arXiv:2005.11246 [cs.CV]

[31] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1, 8 (2019), 9.

[32] Suman Ravuri, Karel Lenc, Matthew Willson, Dmitry Kangin, Remi Lam, Piotr Mirowski, Megan Fitzsimons, Maria Athanassiadou, Sheleem Kashem, Sam Madge, et al. 2021. Skilful precipitation nowcasting using deep generative models of radar. *Nature* 597, 7878 (2021), 672–677.

[33] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. 2016. On recent advances in PV output power forecast. *Solar Energy* 136 (2016), 125–144.

[34] Lewis Fry Richardson. 1922. *Weather prediction by numerical process*. Cambridge university press.

[35] David Rolnick, Priya L Donti, Lynn H Kaack, Kelly Kochanski, Alexandre Lacoste, Kris Sankaran, Andrew Slavin Ross, Nikola Milojevic-Dupont, Natasha Jaques, Anna Waldman-Brown, et al. 2019. Tackling climate change with machine learning. *arXiv preprint arXiv:1906.05433* (2019).

[36] Talha A. Siddiqui, Samarth Bharadwaj, and Shivkumar Kalyanaraman. 2019. A deep learning approach to solar-irradiance forecasting in sky-videos. arXiv:1901.04881 [cs.CV]

[37] Yuchi Sun, Vignesh Venugopal, and Adam Brandt. 2019. Short-term solar power forecast with deep learning: Exploring optimal input and output configuration. *Solar Energy* 188 (08 2019), 730–741. https://doi.org/10.1016/j.solener.2019.06.041

[38] Soumya Tiwari, Reza Sabzehgar, and Mohammad Rasouli. 2018. Short Term Solar Irradiance Forecast Using Numerical Weather Prediction (NWP) with Gradient Boost Regression. In *2018 9th IEEE International Symposium on Power Electronics for Distributed Generation Systems (PEDG)*. 1–8. https://doi.org/10.1109/PEDG.2018.8447751

[39] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008.

[40] Thaddeus Vincenty. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review* 23, 176 (1975), 88–93.

[41] Stefano Vincenzi, Angelo Porrello, Pietro Buzzega, Marco Cipriano, Pietro Fronte, Roberto Cuccu, Carla Ippoliti, Annamaria Conte, and Simone Calderara. 2021. The color out of space: learning self-supervised representations for Earth Observation imagery. In *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 3034–3041.

[42] Huaizhi Wang, Zhenxing Lei, Xian Zhang, Bin Zhou, and Jianchun Peng. 2019. A review of deep learning for renewable energy forecasting. *Energy Conversion and Management* 198 (2019), 111799.

[43] Jinsong Zhang, Rodrigo Verschae, Shohei Nobuhara, and Jean-François Lalonde. 2018. Deep photovoltaic nowcasting. *Solar Energy* 176 (2018), 267–276.

[44] Xin Zhao, Haikun Wei, Hai Wang, Tingting Zhu, and Kanjian Zhang. 2019. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Solar Energy* 181 (2019), 510–518.