Interactive Geographical Visualization of Multiple Statistical Forecasting Techniques for Solar Irradiance

Alex Niu, Garret Fick, Jitendra Rathour, Zhimen Shen

***Abstract*—We present and evaluate a novel approach to visualizing measured and forecast solar irradiation across a large timespan and geographic area. The methods shown enable insight into storage requirements that will be needed for a 100% renewable grid and how changes in climate affect the geographic availability of solar energy.**

I. Introduction

Solar energy is contributing an increasingly important part of the electrical energy composition. In some regions, instantaneous photovoltaic (PV) generation has exceeded demand [**?**]; in others, PV will soon participate firm generation [1]. Recent legislation, such as California Bill 100 [2], require 100% of energy from renewable sources by 2045.

Unlike conventional generation, solar is not dispatchable and availability depends on external controlled factors. Forecasting is therefore an important topic and the focus of much research, summarized in multiple recent review papers, such as [3] and [4].

Grid management utilizes forecasts with varying timescales in order to achieve multiple objectives. At the smallest time scale, forecasts enable the balancing authority, the agency that ensures supply matches demand, to dispatch conventional resources according to changes in the availability of PV. At longer time scales, on the order of 24 hours, forecasts enable solar-plus-storage to optimally schedule the battery charge-discharge cycle based on solar availability and financial market signals. On even longer scales, forecasts inform optimal sites for solar generation or planning for infrequent events, such as solar eclipses.

Much research has looked at forecasts in particular locations, and there are tools such as NREL's NSRDB viewer [**?**] that shows predictions over the US. Comparatively less effort has been spent visualizing wide area trends combined with what can be learned from historical data. This combination of wide area trends over time become increasingly important as solar penetration increases and replaces conventional generation. Developing this understanding requires two components:

- Measurements and models over statistically significant time scales and areas
- Capabilities to make sense of the data over large timescale and areas

This work is a novel approach for the visualization of historical and forecast solar irradiation over both a large timescale and area.

All authors contributed equally to this research.

II. Background

There are several approaches to predicting solar irradiance: statistical, image/satellite-based, numerical weather prediction (NWP), and hybrid. Statistical models focus on prediction using historical data, for example Autoregressive Integrated Moving Averages (ARIMA), neural networks, support vector machines, linear and multiple regression. These methods have a wide range of spatial coverage from 1 m to 10 km, and temporal horizon from 20 seconds to 1 month. Image and satellite driven analysis uses satellites and ground-based images of cloud coverage to predict radiance. NWP forecasts use differential equations to forecast the future state of the atmosphere based on weather observations. Finally, hybrid approaches combine two or more of the previous techniques [5], [6] and often benefit from a smoothing effect attributed to these ensemble methods.

The selection of a technique for a use depends on multiple factors, including time scale of interest short-term (intra-hour), medium-term (intra-day), and day ahead [7]. For example, grid operators normally prefer regional forecasts because these are more useful maintain the supply/demand balance in the electric system. The following sections highlight some important techniques.

A. ARIMA

Autoregressive Integrated Moving Average (ARIMA) normalizes and calculates moving averages with respect to time to provide highly accurate short term forecasts. Reikard [8] applied an ARIMA model to irradiance data achieving a high degree of accuracy for short term forecasts. He showed that publicly available data is sufficient for building a useful model.

B. Neural Networks

Neural networks have been used extensively to forecast solar irradiance. Kemmoku et al. [9] designed a multi-layer neural network with backpropagation which resulted in a 80 percent accuracy rate. Wang et al. [10] present a neural network for making short term predictions that uses relatively few parameters but takes into account environmental factors.

C. Hybrid

Hybrid models, such as Seul-Gi et al. [11] and Yuan-Kang et al. [12] use both observed and forecast data. These studies suggest that the observed data is more relevant for short term PV prediction horizon ($<$ 2 hrs) whereas forecast data improves the prediction for longer horizons ($>$ 2 hrs). The presented models perform well in general but few intermittent inflated predictions decrease the overall accuracy in some cases. They note the value of being able to describe the confidence interval for predictions.

Detyniecki et al. [13] explored adding Fuzzy decision trees to the model to improve performance. They showed that this hybrid approach reduces the forecast inaccuracies observed in other hybrid approaches.

III. METHOD

In this work, we introduce VIGSI, an approach and set of tools to *VI*sualize *G*raphically *S*olar *I*rradiance data sets. The VIGSI tools are available as open source software [14].

A. Data Source

Key to our objective is irradiance data that is available over a long period of time. Availability of data ensures that we are able to show differences between measured and modelled data. We focused on United States and the Wind Integration National Dataset Toolkit [15]. This 50 TB data set contains, in addition to wind, measured and modelled GHI on a uniform 2-km grid for the time period beginning 1-Jan-2007 and ending 31-Dec-2013 (7 years). As will be see, we believe this span is sufficient to reveal important insights.

B. Modelling

We desire, as an objective, to clearly indicate the differences between measured and modelled data and differences between modelling approaches. We implemented two forecasting models that were previously described in the literature.

The ARIMA forecast model is built based on historical irradiance data. We conducted the following exploratory data analysis:

- Dickey-Fuller (DF) test on the time-series data to check for stationarity.
- Auto-Correlation function (ACF) and Partial Auto-correlation (PACF) plots to help visualize and understand required degrees of Moving Averages (q) and Auto-regression (p) components of ARIMA.
- Plot the data decomposition to visualize the Original Data, Trend, Seasonality and Randomness components separately.

We started with manually selecting a model based on exploratory tests (DF,ACF,PACF), which suggested ARIMA(3,1,1)$(0,1,0)_{365}$. Then we used Auto ARIMA, which evaluates models with different combinations of (p,d,q) & $(P,D,Q)_{seasonal}$ values and picks the best performing model with lowest AIC/BIC. This analysis recommended ARIMA(1,0,0)$(0,1,0)_{365}$. Results of both models are very similar with the later being fractionally better in accuracy (with 53% accuracy & RMSE=1672). With both models producing similar results, selected go for ARIMA(1,0,0)$(0,1,0)_{365}$ for its simplicity & inexpensive processing.

LSTM (Long Short-Term Memory) model of recurrent neural network was selected in Keras for this solar irradiance forecast study. Hourly weather data for Albany NY from 2014 to 2017 was extracted from NOAA. The model was trained on the first 3 years of data points, and then tested on year 2017. Previous hour steps (n-hour) were tested incrementally from hour 1 to hour 12. It was found RMSE is at its minimum (84) when n-hour is set around 5-8.

The following parameters were set with the neural network design, and they need to be further tuned: number of neurons in hidden layer- 50, epoch - 50, and batch size -72.

We plan to expand the model onto other weather stations across US. The current chosen weather features are temperature, dew point, relative humidity, pressure, precipitable water, wind direction, and wind speed. Dimensionality reduction is also planned if time permits.

C. Data Processing, Preparation, and Delivery

The NREL solar data from [15] is stored as a set of HDF5 files in Amazon S3. Each file contains a year's worth of data and is 1.6 TB in aggregate. NREL has developed Highly Scalable Data Service (HSDS) which will allow API access to relevant parts. Data is stored in approximately 2 million 4 km^2 rectangular areas for each half-hour increment.

To perform our experiments, we filter and transform the data in a data processing pipeline. The filter restricts data for data points that lie within the contiguous United States. Second, we transform the HDF5 data into GeoJSON and upload it to S3 so that it can be queried by timestamp. For simplicity, we use an bucket open to the public as the means of access.

D. Visualization

The VIGSI user interface is a browser-based application split into 4 primary regions. The map region displays a color-coded map of solar irradiance and related statistics for the area of interest, according to the users selection. Within the map, the user may inspect a point of interest (by dragging a pointer), which shows details related to that point in a second area at the sides of the map. The details are shown as two graphs - one showing information for all points along the same latitude and one showing information for all points along the same longitude. These graphs depict all measured and predicted data in a single plot

The application shows 5 different data sets accounting for different timescales and data predictions using statistical approaches. These data sets were produced to demonstrate particular characteristics in the data.

- Measured solar instantaneous irradiance (1 hour time steps)
- Total daily measured solar energy
- Total monthly measured solar energy
- Total yearly measured solar energy
- ARIMA predicted total daily measured solar energy

The third region provides standard playback controls for navigating through the data over time. The controls permit effective zooming into a particular time window of interest. This is important due to the large available time window.

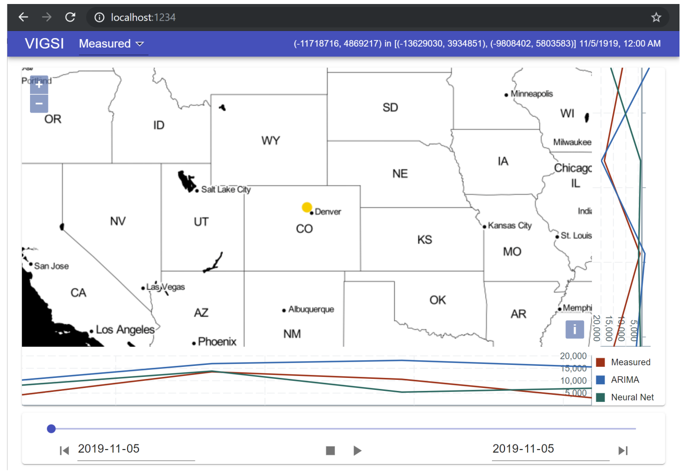

Figure 1. VIGSI displaying irradiance predicted by a Neural Network model with differences shown in the side charts. The time control shows along the bottom and allows the user to explore predictions at selected points in time. The dot near Denver, CO reveals the location for which the charts are based.

The final header shows essential information about the selected data set and allows selecting different data sets.

IV. Evaluation

Our hypothesis is that an innovative approach to visualization of solar irradiance large spatial and time dimensions can enable insights. Within the application framework described above, we evaluated multiple choices and iterated over the design.

Figure 1 shows an early VIGSI prototype, prior to integrating a data set. This early prototype aimed to implement the Material design language [16]. We see large margins, including generous separation from the browser edge.

The map was two-tone, land areas were drawn in white and water areas drawn in black, and was selected to give emphasis to measurements rather than geographic boundaries. This design also introduced the focal point as a large yellow marker that controls two associated charts.

The same initial design with added measurements is shows in Figure 2. (The hatching pattern in the image is due to a defect in this version). In this version, the water the geographic elements are drawn with lower contrast as it was shown that these geographic elements changes the color associated with solar irradiance. However, this version still suffered from color problems, such as those exhibited along the west coast border. It is further unclear whether saturated colors represent high or low solar irradiance.

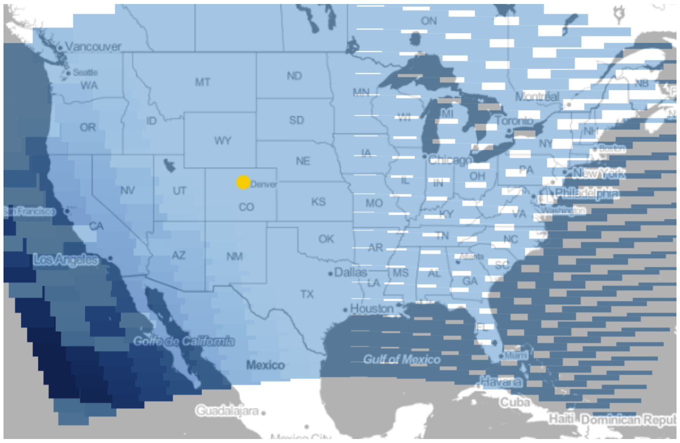

Figure 2. Blah blah blah

Figure 3. Blah blah blah

The experiment shown in Figure 3 attempts the solve the contrast mapping issue. In this version, white is mapped to maximum solar irradiance and black is mapped no solar irradiance This results in a natural mapping to day and night, however the background contrast continues to have a significant impact as can be seen by the clear east coast boundary.

Figure 7 shows the a number of essential changes that were the result of experiments with the user interface. Use of the Material design language is largely abandoned, and instead maximizes the use of available space. Geographic boundaries are drawn as a single color overlay with not distinction between land and water regions. The background color is dark and similar in color to black, which for the selected data set, represents zero solar energy on the selected day. The color gradient is fixed based on the shown quantity, as opposed to the displayed data.

A second dimension of evaluation whether the visualization approach of VIGSI effectively shows meaningful differences in the data set. For this, we

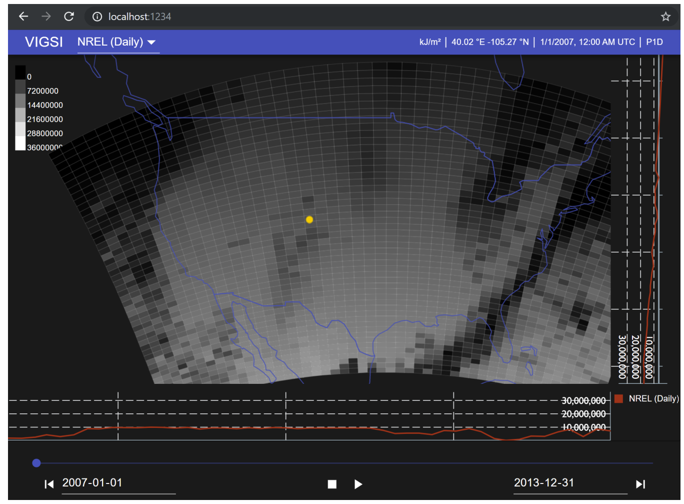

Figure 4. Blah blah blah

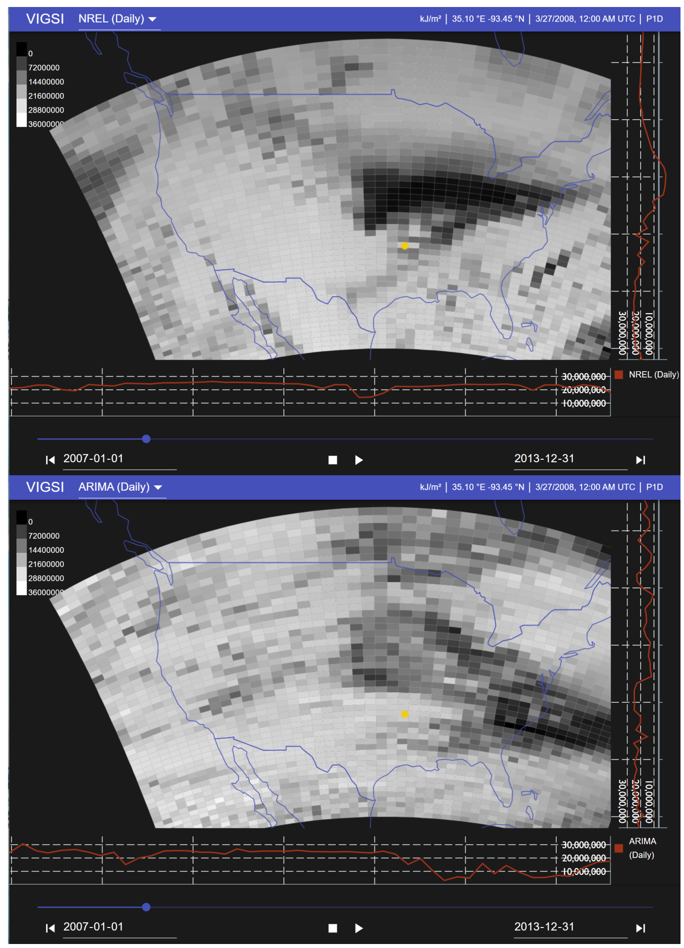

Figure 5. Blah blah blah

make several comparisons.

V. Conclusion

Our project is innovative in a couple of different ways. First, if we can accomplish our goals, it will represent the first comprehensive, wide area

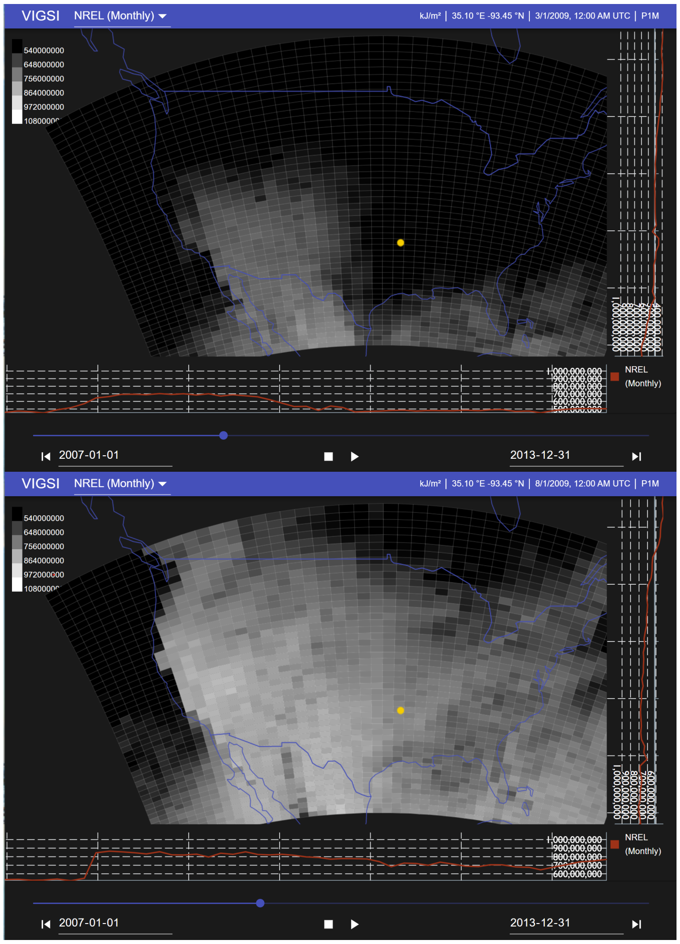

Figure 6. Blah blah blah

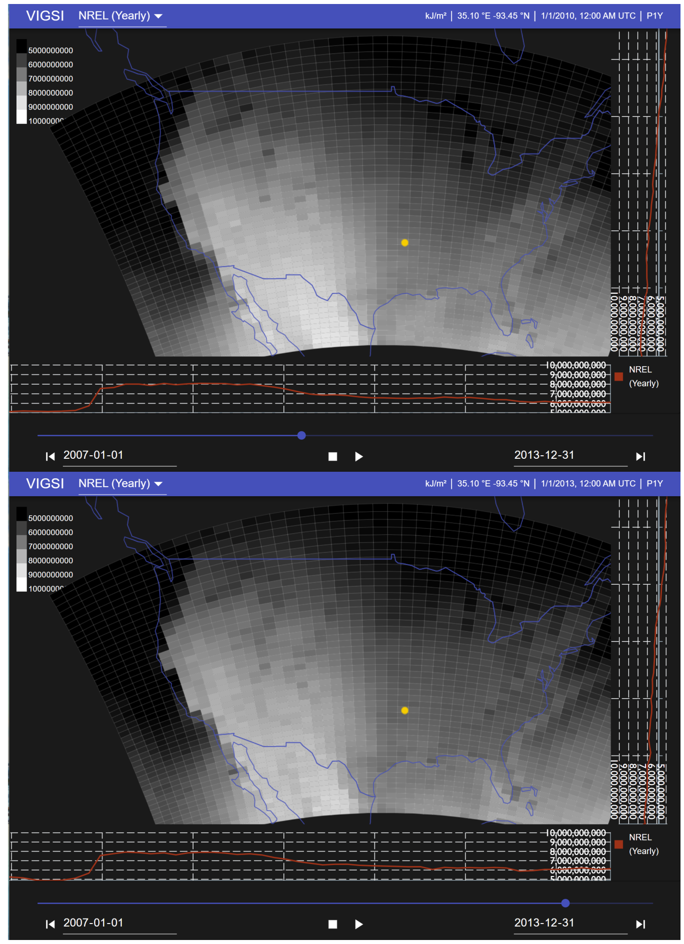

Figure 7. Blah blah blah

implementation of forecasting algorithms. While these algorithms have been run against single sites or small geographic areas in the past, this will be the first time they are used on a large scale. Second, current visualizations for the NREL dataset are limited to static, point in time representations. Our visualization will present a time series representation of NREL's data. Third, the transformation work we are doing to produce GeoJSON files will allow for large scale consumption by web applications in a way that is not possible with HDF5 files.

References

[1] Sunrun wins big in new england capacity auction with home solar and batteries. https://www.greentechmedia.com/articles/read/sunrun-wins-new-england-capacity-auction-with-home-solar-and-batteries, Accessed 2019-10-05.

[2] Senate Bill No. 100. https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=201720180SB100, Accessed 2019-10-05.

[3] J. Antonanzasa, N. Osoriob, R. Escobarb, R. Urracaa, F.J. Martinez de Pisona, and F. Antonanzas-Torresa. Review of photovoltaic power forecasting. *Solar Energy*, 136:78–111, 2016.

[4] R. H. Inman, H.T.C. Pedro, and C.F.M. Coimbra. Solar forecasting methods for renewable energy integration. *Progress in Energy and Combustion Science*, 39:535–576, 2013.

[5] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27:65–76, 2013.

[6] E. Lorenz, D. Heinemann, H. Wickramarathne, H.G Beyer, and S. Bofinger. Forecast of ensemble power production by grid-connected pv systems. *Proceedings of the 20th European PV Converence*, pages 1–6, 2007.

[7] R. Urraca, J. Antonanzas, M. Alia-Martínez, F.J. Martínez de Pisón, and F.Antonanzas-Torres. Smart baseline models for solar irradiation forecasting. *Energy Conversion and Management*, 108:539–548, 2016.

[8] G. Reikard. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83:342–349, 2009.

[9] Y. Kemmoku, , S. Orita, S. Nakagawa, and T. Sakakibara. Daily insolation forecasting using a multi-stage neural network. *Solar Energy*, 66:193–199, 2013.

[10] F. Wang, Z. Mi, S. Su, and H. Zhao. Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies*, 5:1355–1370, 2012.

[11] Seul-Gi Kim, Jae-Yoon Jung, and Min Kyu Sim. A two-step approach to solar power generation prediction based on weather data using machine learning. *Sustainabililty*, 11:1501, 2019.

[12] Yuan-Kang Wu, Chao-Rong Chen, and Hasimah Abdul Rahman. A novel hybrid model for short-term forecasting in pv power generation. *International Journal of Photoenergy*, 2014:1–9, 2014.

[13] Marcin Detyniecki, Christophe Marsala, Ashwati Krishnan, and Mel Siegel. Weather-based solar energy prediction. *IEEE International Conference on Fuzzy Systems*, June 2012.

[14] VIGSI. https://github.com/vigsi, Accessed 2019-10-29.

[15] Wind Integration National Dataset Toolkit - NSRDB Data Viewer. https://maps.nrel.gov/nsrdb-viewer, Accessed 2019-11-04.

[16] Material design. https://material.io/, Accessed 2019-11-23.