2023 IEEE Symposium Series on Computational Intelligence (SSCI)
Mexico City, Mexico. December 5-8, 2023

Digital Twins for Creating Virtual Models of Solar Photovoltaic Plants

Deborah George, *Graduate Student*, IEEE
Real-Time Power and Intelligent Systems Laboratory
Department of Electrical and Computer Engineering
Clemson University, Clemson, SC 29634, USA
dgeorge@ieee.org

Ganesh K. Venayagamoorthy, *Fellow*, IEEE
Real-Time Power and Intelligent Systems Laboratory
Department of Electrical and Computer Engineering
Clemson University, Clemson, SC 29634, USA
gkumar@ieee.org

Abstract— **Amidst the challenges posed by the high penetration of distributed energy resources (DERs), particularly a number of distributed photovoltaic plants (DPVs), in modern electric power distribution systems (MEPDS), the integration of new technologies and frameworks become crucial for addressing operation, management, and planning challenges. Situational awareness (SA) and situational intelligence (SI) over multi-time scales is essential for enhanced and reliable PV power generation in MEPDS. In this paper, data-driven digital twins (DTs) are developed using AI paradigms to develop actual and/or virtual models of DPVs, These DTs are then applied for estimating and forecasting the power outputs of physical and virtual PV plants. Virtual weather stations are used to estimate solar irradiance and temperature at user-selected locations in a localized region, using inferences from physical weather stations. Three case studies are examined based on data availability: physical PV plant, hybrid PV plants, and virtual PV plants, generating real-time estimations and short-term forecasts of PV power production that can support distribution system studies and decision-making.**

Keywords -- AI, distributed energy resources, digital twin, power distribution systems, weather stations, virtual systems

I. Introduction

The high penetration of distributed energy resources (DERs), and in particular several distributed photovoltaics plants (DPVs) has increasingly created operation and management (O&M) challenges in the modern electric power distribution system (MEPDS), as well as complexities during distribution system planning and design [1]. High fidelity, high quality, real-time estimations and predictions of PV plant performance considering both environmental and spatial characteristics is valuable for analysis of complex behavioral dynamics in PV plants and thereby supporting O&M and planning and design tasks in distribution systems with high penetration of DPVs [2].

Data-driven digital twins (DTs) of DER sources such as PV plants can enable generation of realistic datasets over a wide range of scenarios. Such DTs can be valuable aids for estimating and predicting/forecasting the behavior of complex systems. The availability of DTs can enable MEPDS O&M support even with sparse/limited measurement infrastructure. These DTs can be further extended to facilitate enhanced monitoring and observability, condition monitoring, fault-detection, predictive analysis and forecasting, as well as other advanced functionality such as energy management, planning and design studies such as hosting capacity, cloud impact analysis and scenario testing of PV plants when integrated into testbeds such as real-time digital simulators (RTDS).

This research is funded by US NSF grant CNS 2131070, ECCS 2234032, CNS 2318612 and the Duke Energy Distinguished Professor Endowment Fund. Any opinions, findings and conclusions or recommendations expressed in this material are those of author(s) and do not necessarily reflect the views of NSF and Duke Energy.

The primary contributions of this paper are as follows:

- Data-driven DTs are developed for real-time estimation and prediction of a physical PV plant power output under different weather patterns. It is assumed a physical PV plant and a physical weather station is available to generate the data.
- Virtual weather stations are developed using a mutation approach based on data from physical weather stations, providing real-time estimated weather measurements at additional locations.
- A multi-DT methodology for real-time estimation and prediction of power outputs of physical and virtual PV power plants over the selected region is developed for supporting various scenarios and purposes, enabling SA and SI of PV power generation in a distribution system, as shown in Figure 1.

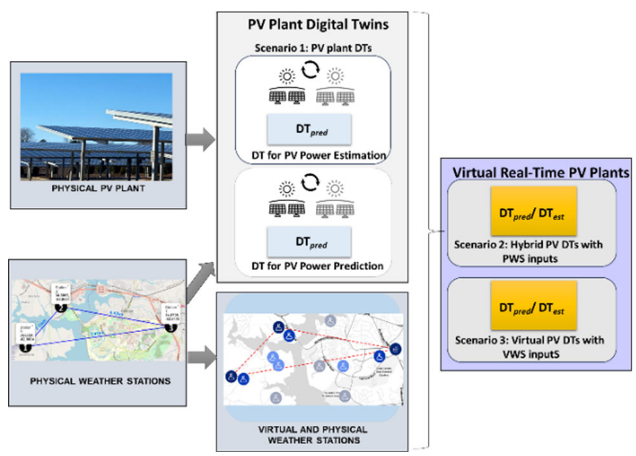

Figure 1. Scenarios for data generation using a physical PV plant, hybrid PV plants and virtual PV plants.

The rest of the paper is written as follows: Section II describes digital twins for PV plants. Section III presents the mutation methodology used to create virtual weather stations. Section IV describes the application of DTs to provide SA and SI of physical and virtual PV plant power outputs. Section V presents typical results and some discussions. Finally, the conclusion and some suggestions for future work is presented in Section VI.

II. DIGITAL TWINS FOR PV PLANTS

PV generation at a site is typically dependent on external environmental factors as well as internal factors. A mathematical equation defining typical PV power generated at time instant *t* is given in (1), where G_{std} and T_{std} are the standard test conditions for solar radiation and cell temperature, respectively, and α_T is the manufacturer temperature coefficient of a PV module.

$$PV(t) = P_{Peak}\frac{\mathrm{G(t)}}{\mathrm{Gstd}} - \alpha_T[Tc(t) - T_{std}] \qquad (1)$$

The following subsections describe digital twins, and DTs for PV plant power estimation and prediction.

A. Digital twins

A digital twin can be defined as a virtual model or representation of a physical system, asset or process enhanced with data connections enabling the transfer of data insights and process data from the virtual representation in real-time (RT) back to the physical system [3]. The DT has three major components: the physical system, the virtual model or "twin" and a constant real-time informational or control flow exchange, enabling "twinning" of the digital model with the changes in the physical system, as shown in Figure 2.

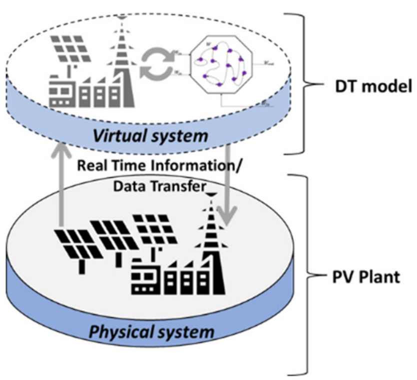

Figure 2. Digital twin components for a PV plant.

Traditionally, physics-based, mathematical approaches, data-driven or hybrid combinations of those methods can be used for creating DTs of DERs, especially for PV plants. In MEPDS with high penetrations of DPVs, physics-based or mathematical-based DT modeling will need to account for stochastic or nonlinear plant behavior dependent on environment processes. These include cloud cover variation, wind speed changes, shading losses, bundled behind-the-meter readings where load usage is to be predicted, seasonal variation, humidity, to PV panel degradation and aging, and other accompanying variables. Additionally, intensive modeling and setting up accompanying measurement equipment and management of data streams may not be economically viable. AI paradigms enable data-driven DT design through function approximation and enable reasonably accurate estimation of nonlinear and stochastic processes, require little maintenance and are easily adaptable and scalable compared to mathematical or physics-based DTs which may be system specific [4].

A data-driven PV plant DT is employed here to model PV power generation of an existing real-world PV-plant at a physical site, called the R06 site, which a 1 MW PV plant at Clemson University. The solar power generation data from this plant is streamed from a micro-PMU. A CR300 weather data logger that monitors and logs input solar irradiance, temperature, wind speed and wind direction data every second is utilized to capture the input variables. This data was archived and timestamped in an open Historian, a back-office No-SQL database used for streaming, archiving, and integrating process control and synchro phasor data. Weather and PV power data over 72 days in the spring 2023 was collected for this study.

B. PV power estimation

Real-time power estimation at a PV plant is useful for analytics, understanding of system dynamics, condition and health monitoring as well as a source of measurement data when µPMUs or other measurement devices at the site are absent, or bundled. A PV plant DT for estimation of power outputs is developed using Multi-Layer Perceptron feedforward neural network (MLP) that typically enables informational flow in one direction [5]. MLPs are used to model system behavior and estimate real-time PV power generation at the location. The MLP model inputs are real-time data and is re-trained appropriately to reflect in real-time changing PV plant conditions. The architecture of the MLP used is as shown in Figure 3.

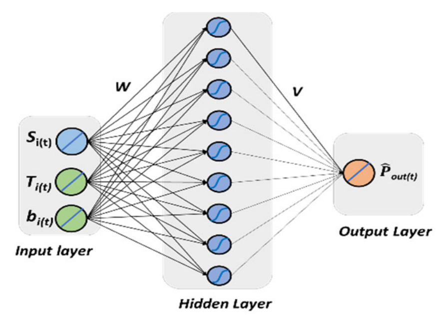

Figure 3. MLP for PV power estimation at current time interval.

The MLP equation is given in (2) [5], where *x*(*t*) and *y(t)* are the inputs and outputs of the MLP network, and *W* and *V* are input and output weight matrices, respectively.

$$y(t)=f(x(t), W, V) \qquad (2)$$

Python's scikit-learn module is used to define the MLP architecture, here "MLPRegressor", with hyperparameters defined as in [6]. Three MLPs are utilized, separately pre-trained on historical data categorized by amount of solar irradiance received per day, as "Cloudy", "Sunny" and "Moderately Cloudy". Real-time solar irradiance and temperature measurements obtained from the weather station at the PV plant are used as inputs to each MLP model to continuously estimate PV power output at the current time interval for each solar irradiance category. The mean squared error (MSE) of MLP estimation from the three models are compared every minute and the best performing model's output was recorded.

C. PV power prediction

Short term PV power prediction are useful for distribution system operation (DSO) needs and applications such as cloud cover impact identification, inverter voltage regulation and volt-var optimization and fault prediction. A PV plant DT for prediction of power output is developed using echo state networks (ESNs) to effectively predict PV plant performance in real-time.

An ESN is a form of a recurrent neural network (RNN), created by Herbert Jaeger as a paradigm falling under the category of reservoir computing [7]. The ESN dynamics can be represented as shown in (3) where f is an activation function, $x(n)$ is the n-dimensional reservoir state, W is the ($n \times n$) reservoir weight matrix, W_{in} is the ($n \times k$) input weight matrix, $u(n)$ is the k-dimensional input signal, W_{fb} is the ($n \times m$) output feedback matrix, and $y(n)$ is the m-dimensional output signal.

$$x(n+1)=f(x(n),\ W_{in},\ W,\ W_{fb,}\ y(n)) \tag{3}$$

For PV power prediction, three parameters are used to accurately model the PV array's power generation based on the approach used in [8]; the solar irradiance, temperature, and the output power of the PV array. Three ESNs are utilized, with hyperparameters as in [7], separately fitted on historical data categorized by amount of solar irradiance received per day, as “Cloudy”, “Sunny” and “Moderately Cloudy”. The DTs iteratively predict 60, 120 and 180 seconds ahead, and best performing ESN output is recorded.

III. Virtual Weather Stations

High quality, high resolution, spatiotemporal weather data is necessary for analyzing the spatial variances in PV and DPVs for further integration into MEPDS. A means to develop virtual weather stations (VWS) at any location, supported by locally available measurement sites through data re-analysis and strategic mutations are developed [9]. This approach enables the generation of real-time, multi-time scale weather data for a PV DT for fast, computationally efficient PV power estimation and prediction for an area with sparse weather measurements. The real-time behavior of the weather stations with respect to each other is observed and modeled for a particular day using three-month averages of historical data to capture long term trends and daily data to analyze short term trends between the stations.

A. Mutation procedure

VWS are developed to estimate solar irradiance, temperature, wind speed and direction for user selected locations in a region with less than 20km radius, based on real-time measurement data from neighboring physical weather stations (PWS). A minimum of three PWSs are necessary to create VWS in this method, with the mutation equations described below executed in real-time.

Solar irradiance, S_i in W/m^2, is queried in real-time to form arrays of real-time solar irradiance measurements S_p for a period of time, t, for example, 60,120 or 180 seconds from each station as shown in (4). Inverse distance weighting (IDW) is then used to estimate base solar irradiance values S_{pred}, wind speed and wind direction at VWS based on the approach used in [10].

The formula for IDW can be expressed as shown in (5) and (6) where $V(u)$ is the estimated value at the unknown location, V_i is the value at the known location i and w_i is the weight assigned to the known location i, calculated based on the distance between the known and unknown locations as a squared or cubic of the distance. Herein, three locations are known and p in (6) is set to two. After IDW is applied, ten S_{pred} and W_{msi} arrays with t lengths are identified for each VWS.

$$S_p = [S_1, S_2, \dots S_i] \tag{4}$$

$$V(u) = \frac{\sum (Vi * Wi)}{\sum Vi} \tag{5}$$

$$Vp = 1 \,/\, (x1 - x2)^p \tag{6}$$

Based on the random number generator model referenced in [11], a scaled random distribution based on estimated distance from the plant is used to mimic C_m (7) over geographical locations within a 20 kms radius. C_m is obtained from combinations of controlled random sampling of t seconds IDW solar irradiance estimations, S_{pred} and the addition of up to 5% Gaussian noise as seen in (7). From this, S_m or mutated solar irradiance streams are obtained, as shown in (8).

$$C_m = rand(S_{pred}) + \frac{0.05}{\sigma\sqrt{2\pi}} e^{-(S_{pred}-\mu)^2/2\sigma^2} \tag{7}$$

$$S_m = \begin{cases} [S_{pred}] & S_{pred} > 0 \ and \ Cd \le 10\% \\ [C_m] & S_{pred} > 0 \ and \ Cd > 10\% \\ 0 & S_{pred} = 0 \end{cases} \tag{8}$$

Wind speed and direction is a separate parameter that is seen to be very weakly correlated to cloud cover and thereby does not directly impact solar irradiance patterns affected by cloud coverage [12]. A weak correlation is seen, where for both rural and urban areas, highly windy days are slightly more likely to be clear [13]. However, wind speed affects the temperature at the weather station. A cooling effect factor, W_e, is obtained to vary temperature data obtained at each VWS based on estimated wind speed effects from IDW calculation. W_{diff} is the wind speed difference at the VWS station from the average wind speeds at the three PWS at instant t, divided by the average wind speeds at the PWS. W_{diff} is used to look up temperature variation values from Table I, where each range in W_{diff} corresponds to a cooling effect factor W_e. If W_{diff} is positive at time t, the W_e is added to temperature T_i, and if negative, it is subtracted. In this manner, real time temperature data is estimated at the VWS locations.

IV. Situational awareness and Intelligence of PV Plant Power Outputs

A. Situational Awareness

A significant challenge in MEPDs is the lack of sufficient observability and difficulty in forecasting and planning in active distribution networks [14]. This means that SA of real-time behavior of various DERs, particularly DPVs in the system is often difficult to monitor, especially due to lack of measurement data.

In this study, PV Plant DTs that obtain real-time data from physical and virtual weather stations are used to facilitate both real-time monitoring and DSO operator SA of PV system behaviors. Data collection over a day can further allow for PV integration into distribution system studies or time series analysis to be performed in the system.

TABLE I. LOOK UP TABLE FOR REAL-TIME TEMPERATURE ESTIMATION

	W_{diff} scale (+/-)	Temperature Variation (degrees Celsius) T_i (+/-)
1	(0.1, 0.4]	-0.332
2	(0.4, 0.7]	-0.202
3	(0.7, 1.0]	-0.125
4	(1.0, 1.3]	0.062
6	(1.3, 1.8]	0.332
7	(1.8, 2.1]	0.396
8	(2.1, 2.4]	0.440
9	(2.4, 2.6]	0.486
10	(2.6, 2.9]	0.544
11	(2.9, 3.2]	0.594

B. *Situational Intelligence*

Operator situational awareness can be established through behavioral modeling of DPVs in various data availability scenarios. However, situational intelligence, or predictive analysis is often needed in the system, especially for short term PV production generation forecast, feeder voltage analysis, scenario analysis or DSO decision-making wherein multiple DT models can be run in parallel for predictive analysis and volt-var algorithm optimization or control, among other use cases. Here, short term forecasts of PV power production support DER management system (DERMS) applications and other distribution system analyses involved in resiliency, reliability and planning such as hosting capacity studies, voltage profile analyses, distribution system behavior in peak load and light generation, light load and peak generation, and other long-term or time-series analysis performed on DER/ DPV sources integrated into the grid.

The creation of virtual real-time PV plants (V-RT-PVPs) for DPV situational awareness and situational intelligence is broken down into three scenarios as follows:

- *Scenario 1*: Both PWS and PMU data is available, called the *Physical PV source*.
- *Scenario 2*: Only PWS may be available, or *Hybrid PV sources*.
- *Scenario 3*: Neither PWS or PMU data is available, called *Virtual PV sources*, and DPV power generation needs to be estimated or forecasted in the region.

V. RESULTS AND DISCUSSION

Section V.A presents typical results obtained for the virtual weather stations. Typical SI (prediction) results for the three scenarios mentioned above are presented in Sections V.B to V.D.

A. *Virtual Weather Stations*

VWS are generated at the PWS site locations to validate the approach defined in Section III, comparing estimated vs actual solar irradiance and temperature results shown in Figure 4. Mutated temperature differences are due to utilization of two physical weather stations in our limited experimental setup.

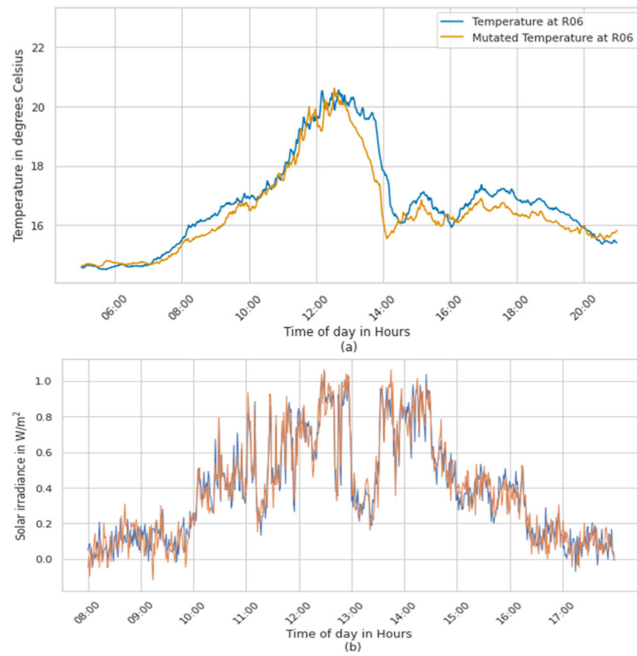

Figure 4. (a) Estimated vs. Actual Temperature and (b) Solar Irradiance (b) at R06.

B. *Scenario 1*

This case study is used to validate the proposed DT based approach. Here, measurements are present for both PWS streams as well as the PV plant power data (from the micro-PMU). The developed PV DTs can be used similarly to traditional DTs and utilized for prognostics, condition monitoring and fault detection and faster-than-real-time operation analysis. Results of the PV DT for estimation and predication are shown in Tables II and III. These results include mean absolute error (MAE), mean squared error (MSE), root-mean squared error (RMSE) and mean absolute percentage error (MAPE).

C. *Scenario 2*

There is often a need to estimate PV power or predict its impact over the distribution system for locations where micro-PMU, separate meter readings or other measurement data is unavailable or where a PV plant has not been installed yet. In this scenario, PWS may be present at the site, or other weather models may be available for estimation of solar irradiance and temperature at the site. The results of the PV prediction DT at PWS sites, called "Ravenel" and the "Airport" for a cloudy day are shown in Figures 5 (a) and (b).

TABLE II, PV DT ESTIMATION RESULTS

Day Category	n	MAE	MSE	RMSE	MAPE
Sunny	10	0.0003	0.0002	0.0141	0.0064
Moderately Cloudy	20	0.0075	0.0087	0.0933	0.0112
Cloudy	20	0.009	0.0095	0.0975	0.0142

D. *Scenario 3*

PV power also can be estimated or predicted at VWS sites, where no PV plant has yet been installed or where micro-

PMU, separate meter readings or other measurement data is unavailable. These virtual PV sources can be easily adapted to be integrated into a distribution testbed for short term PV forecasts, data collection and general situational intelligence of PV power generation in the area. The results of the study for virtual PV sites (“Ravenal” and “Airport”) for moderately cloudy days are shown in Figures 6 (a) and (b).

Figure 7 shows the three scenarios (physical PV plant, hybrid PV plant and virtual PV plant). In each case, the solar irradiance and solar PV power predicted 180 seconds ahead is shown.

TABLE III. PV DT PREDICTION RESULTS FOR DIFFERENT PREDICTION INTERVALS

Day Category	Reservoir neurons (n)	Metric	60 s	120 s	180 s
Sunny	200	MAE	**0.0005**	0.0006	0.0011
	200	MAPE	**0.023**	0.0353	0.0423
	200	MSE	**0.001**	0.0013	0.0014
	200	RMSE	**0.0316**	0.0361	0.0374
Moderately Cloudy	550	MAE	**0.0063**	0.0089	0.012
	550	MAPE	**0.2456**	0.3571	0.3345
	550	MSE	**0.0079**	0.0123	0.0149
	550	RMSE	**0.0889**	0.1109	0.1221
Cloudy	450	MAE	**0.016**	0.0168	0.0194
	450	MAPE	**0.2124**	0.2232	0.2361
	450	MSE	**0.0183**	0.019	0.0204
	450	RMSE	**0.1353**	0.1378	0.1428

VI. CONCLUSION

Multi-time scale situational awareness and situational intelligence is necessary to deal with operation, management, and planning and design complexities in modern electric power distribution systems with high penetration of distributed solar PV systems. Measurements for existing and/or potential PV sites to support MEPDS O&M are sparse. Installation of measurement devices may be costly, create additional maintenance needs or may not be useful to add during the planning stages. The methodology presented herein provides a cost-effective, scalable and computationally efficient way to develop virtual real-time PV plants based on available data.

Data-driven DTs that enable PV power estimation and prediction at a given site have been developed. Virtual weather stations that estimate solar irradiance and temperature at specific user-selected locations within a localized region, utilizing insights from physical weather stations in the region have also been developed. Finally, these DTs and VWS have been combined to illustrate three possible scenarios for PV power estimation and prediction based on spare measurements. The developed models provide real-time estimations and short-term forecasts of PV power generation, offering valuable O&M support for distribution systems.

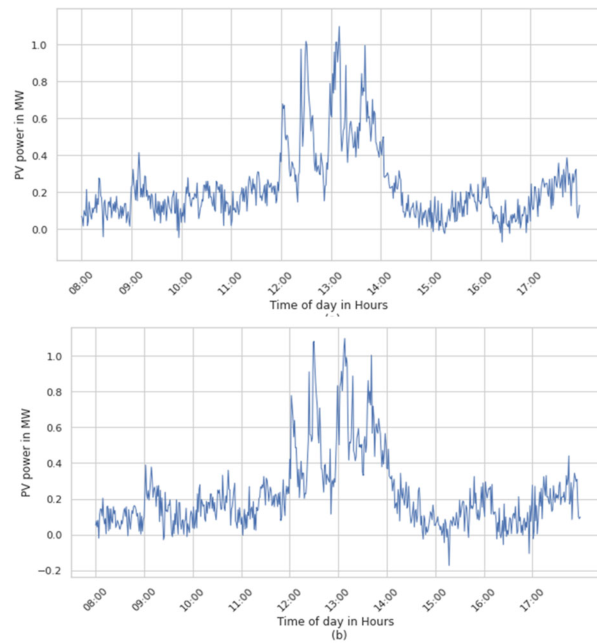

Figure 5. Predicted PV power using PWS 180 seconds ahead for (a) Ravenel amd (b) Airport on a cloudy day.

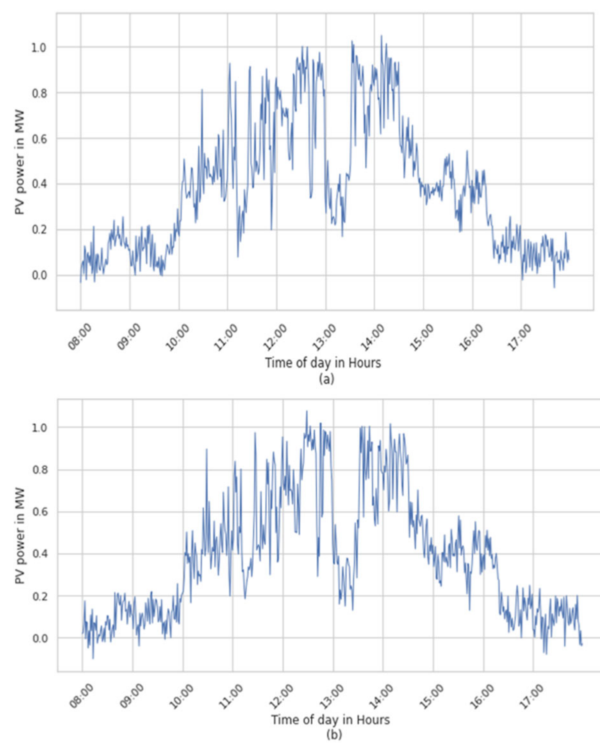

Figure 6. Predicted PV power using VWS 180 seconds ahead for for (a) Ravenel amd (b) Airport on a moderately cloudy day.

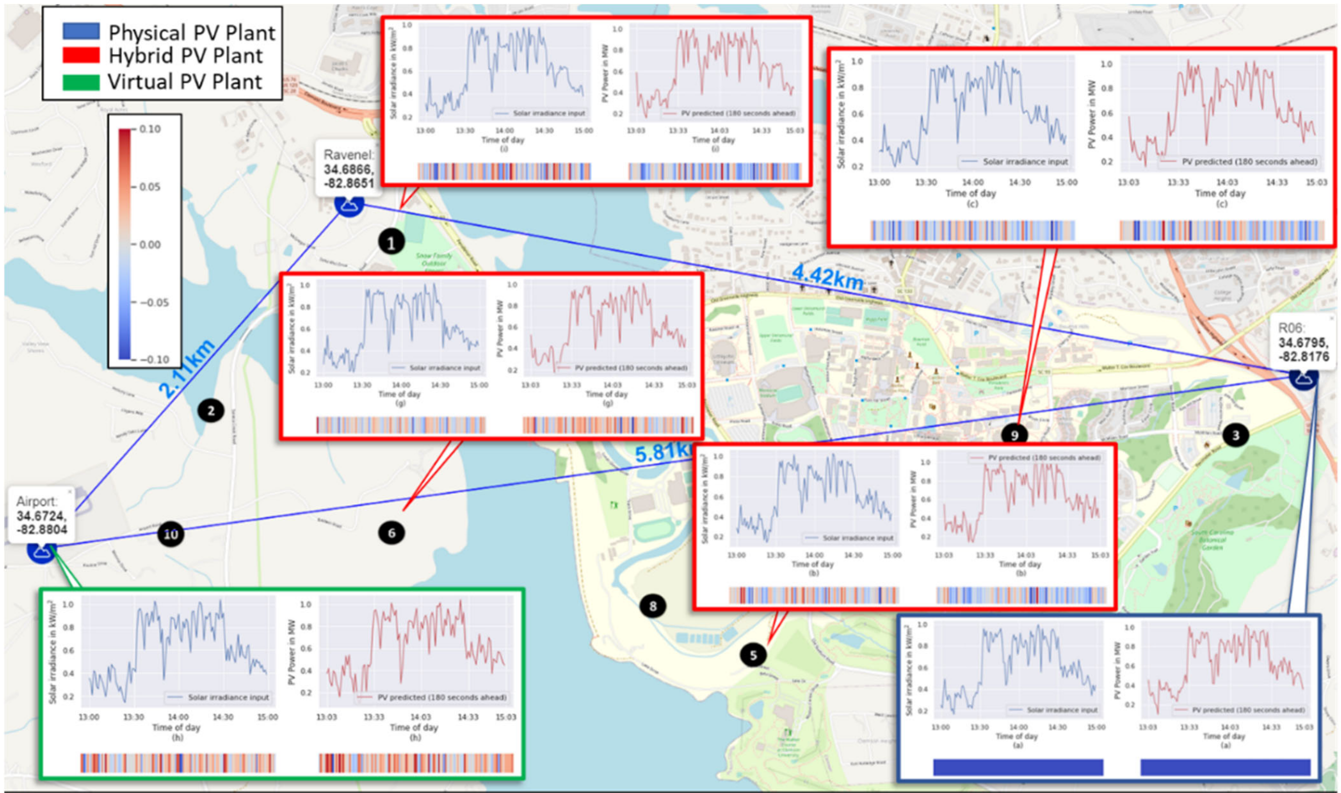

Figure 7. Predicted PV power at 180 seconds ahead for selected locations in the region for a two hour window.

REFERENCES

[1] "IEEE Standard for Interconnection and Interoperability of Distributed Energy Resources with Associated Electric Power Systems Interfaces," in IEEE Std 1547-2018 (Revision of IEEE Std 1547-2003), vol., no., pp.1-138, 6 April 2018, doi: 10.1109/IEEESTD.2018.8332112.

[2] Olowu, T O., et al, "Future challenges and mitigation methods for high photovoltaic penetration: A survey", *Energies*, 11.7 (2018): 1782.

[3] Grieves, M and Vickers, J, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems." In *Transdisciplinary perspectives on complex systems*, Springer, Cham, pp. 85-113, 2017.

[4] Zhaoyun, Z and Lv Linjun, "Application status and prospects of digital twin technology in distribution grid", *Energy Reports*, 8 (2022): 14170-14182.

[5] Pinkus, A, "Approximation theory of the MLP model in neural networks", *Acta numerica*, 8 (1999): 143-195.

[6] Walters, M, Yonce J, Venayagamoorthy GK, "Data-driven digital twins for power estimations of a solar photovoltaic plant", 2023 *IEEE Symposium Series on Computational Intelligence (SSCI)*, Mexico City, Mexico, December 4-7, 2023.

[7] Jaeger, H, "Echo state network", *Scholarpedia,* 2.9 (2007): 2330.

[8] Jayawardene, I and Venayagamoorthy, GK, "Comparison of adaptive neuro-fuzzy inference systems and echo state networks for PV power prediction", *Procedia Computer Science*, 53 (2015): 92-102.

[9] Guzman R, Dorian E, et al, "A genetic algorithm approach as a self-learning and optimization tool for PV power simulation and digital twinning", *Energies*, 13.24 (2020): 6712.

[10] Pathiravasam, C Arunagirinathan, P, Jayawardene, I, Venayagamoorthy GK, Wang Y, "Spatio-temporal distributed solar irradiance and temperature forecasting", *2020 International Joint Conference on Neural Networks (IJCNN)*, 2020, pp. 1-8, doi:10.1109/IJCNN48605.2020.9206936.

[11] Kafka, J. L. Miller, M. A, "A climatology of solar irradiance and its controls across the United States: Implications for solar panel orientation." *Renewable Energy*, 135 (2019): 897-907.

[12] Morris, C. J. G, Simmonds, I and Plummer, N, "Quantification of the influences of wind and cloud on the nocturnal urban heat island of a large city", *Journal of Applied Meteorology and Climatology*, 40.2 (2001): 169-182.

[13] Bett, P. E., and Thornton, H. E, "The climatological relationships between wind and solar energy supply in Britain", *Renewable Energy*, 87 (2016): 96-110.

[14] Obi, M, and Bass R, "Trends and challenges of grid-connected photovoltaic systems–A review", *Renewable and Sustainable Energy Reviews*, 58 (2016): 1082-1094.