COMPARISON OF A RECURRENT NEURAL NETWORK PV SYSTEM MODEL WITH A TRADITIONAL COMPONENT-BASED PV SYSTEM MODEL

Daniel M. Riley[1], and Ganesh K. Venayagamoorthy[2], *Senior Member, IEEE*
[1]Sandia National Laboratories, Albuquerque, New Mexico, USA
[2]Missouri University of Science and Technology, Rolla, Missouri, USA

ABSTRACT

Photovoltaic (PV) system modeling is used throughout the photovoltaic industry for the prediction of PV system output under a given set of weather conditions. PV system modeling has a wide range of uses including: pre-purchase comparisons of PV system components, system health monitoring, and estimation of payback (return on investment) times. In order to adequately model a PV system, the system must be characterized to establish the relationship between given weather inputs (e.g., irradiance, spectrum, temperature) and desired system outputs (e.g., AC power, module temperature). Traditional approaches to system characterization involve characterizing and modeling each component in a PV system and forming a system model by successively using component models. This paper compares a traditional modeling approach using the Sandia Photovoltaic Array Performance Model [1] to a new method of characterization using a recurrent neural network (RNN). The Sandia model predicts system performance from given weather data and individual component characterizations using a defined set of equations, while the RNN "learns" the input/output relationships by training on concurrent weather and performance data. The comparison of a traditional modeling technique and the new RNN method serves to validate the accuracy of the new method in comparison to a widely accepted modeling technique. Modeling using an RNN may be advantageous when component models are not available for the components in a PV system, when the components of a PV system are unknown to the modeler, or when system components are installed or altered in such a fashion that their model parameters are no longer applicable.

INTRODUCTION

The photovoltaic (PV) community frequently uses predictive system models to predict the output of a particular PV system under a given set of weather conditions. In a traditional PV modeling approach, the PV system components (e.g. PV modules, inverter) are characterized individually and sub-system models are developed for the components. The sub-system models are then used sequentially to determine a predicted PV system output due to given weather inputs (e.g. Typical Meteorological Year, Meteonorm). Thus the traditional model approach may be used without constructing the system. The model results have a wide variety of uses including pre-purchase comparisons of system components, predicted payback (return on investment) times, health monitoring of systems already in place, or they may be used by utility providers to determine expected-performance rebate incentives.

However, the traditional modeling approach requires that the user know a great deal about the PV system. The modeler must know what types of PV components are in the system, and how many of each component are present. Additionally, they must have information on the performance properties of each component, which may require extensive testing for use in some PV models. Lastly, many traditional models make assumptions regarding the performance of components in the system. For example, models may use general "derate" factors to account for unit-to-unit variation among components, variation of components in the system from the component(s) characterized for performance parameters, resistive wire losses, and shading.

Artificial neural networks have been used in many aspects of photovoltaic energy research including prediction of solar resource, prediction of module-level maximum power point, maximum power point tracking in inverters, and prediction of system output. A subset of artificial neural networks, the recurrent neural network (RNN), uses predicted values to predict future values. This feedback allows the RNN to capture any "memory" inherent in the characterized system (e.g., thermal mass).

A recurrent neural network has the ability to "learn" the relationship between a set of input and output data [2], and the relationship may then be used as a model to predict output system performance when given a set of input weather data. We have developed a method for modeling a PV system using an RNN. The RNN requires no information about the specific components of the modeled PV system. Instead, the RNN learns the relationship between input weather data and system performance by training itself on a data set with concurrent weather and performance data. The RNN may then make predictions about system performance when given weather data, even if the weather data was not in the training data set. Thus, the RNN method models the PV system as a whole, rather than modeling individual components, and includes system loss factors such as those described earlier. However, since a set of concurrent weather and performance data is required, the RNN technique may only be used to model systems which are already in operation.

PROCEDURE

Test System Description

The test photovoltaic system is a 1.05 kWP rated PV system in Albuquerque, New Mexico. The system uses five monocrystalline silicon (c-Si) PV modules, facing south, tilted at 35° from horizontal, and a 2 kW inverter. A monitoring system records plane of array (POA) irradiance via c-Si reference cell, ambient air temperature, wind speed, module backside temperatures, and AC power at two minute intervals. The monitoring system collected over 1 year of data while the PV system was in operation.

Recurrent Neural Network Model

POA irradiance, ambient temperature, and wind speed were used as input data for the RNN model; the model estimated (predicted) AC power and module temperature.

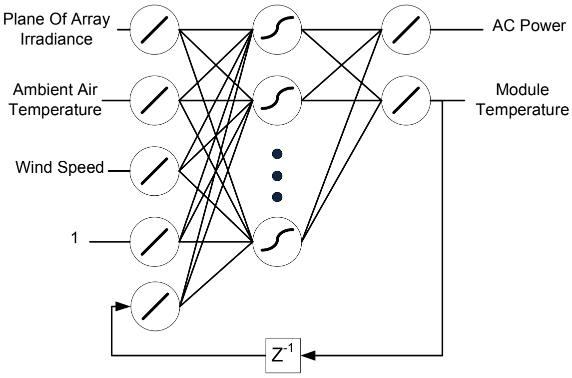

Figure 1: Recurrent neural network used

The RNN is diagrammed in Figure 1, note that the RNN size is (4 + 1) × 70 × 2 and all 70 hidden-layer neurons are not shown in Figure 1 for brevity. Each line in the figure denotes a synaptic weight, w, and each circle denotes a single artificial neuron with transfer equation as shown in (1). Activation functions, f, are either linear (unity gain) or sigmoidal as shown in the figure. The use of a recurrent neural network allows for the time-delayed prediction of module temperature to be used as an input for the next time prediction. Thus, future predictions about module temperature and power include information from prior module temperature predictions; which should allow the RNN to correctly account for the thermal mass of the PV system.

$$y = f\left(\sum_{i=1}^{n} x_i w_i\right) \qquad (1)$$

where:
y = output of the neuron
f = the activation function of the neuron
n = number of inputs to the neuron
x_i = the *ith* input to the neuron
w_i = the *ith* synaptic weight to the neuron

The RNN training data set is approximately 30,000 data points sampled at two minute intervals (41.6 days) from late February through mid-April. The remaining 197,455 data points are reserved for testing the trained neural network. The training process for the RNN minimized the sum of mean absolute error (MAE) across both output parameters over the 30,000 training points, as shown in (2), by modifying the 490 synaptic weights within the RNN. In this case, a particle swarm optimization (PSO) [2] modified the network weights to train the network and minimize MAE.

$$\text{Fitness} = \sum_{j} \text{MAE}_j = \sum_{j=1}^{m}\sum_{i=1}^{n} \frac{1}{n}\left|\hat{X}_{i,j} - X_{i,j}\right| \qquad (2)$$

where:
n = number of training data points or input data patterns
m = number of outputs
$\hat{X}_{i,j}$ = estimated (predicted) value for output j, point i
$X_{i,j}$ = measured value for output j, point i

Sandia Component Models

The application of the Sandia Photovoltaic Array Performance Model and Performance Model for Grid-Connected Photovoltaic Inverters [3] represents a more typical PV modeling approach. Using these component models requires the same 3 inputs as the RNN method, but also requires 3 PV module thermal parameters, 11 module electrical parameters, 2 array configuration parameters, 7 inverter electrical parameters, and a set of typical “derate” factors (mismatch, diodes and connections, DC wiring, and AC wiring) obtained from PVWatts default derate factors webpage [4]. A data flow diagram is presented in Figure 2.

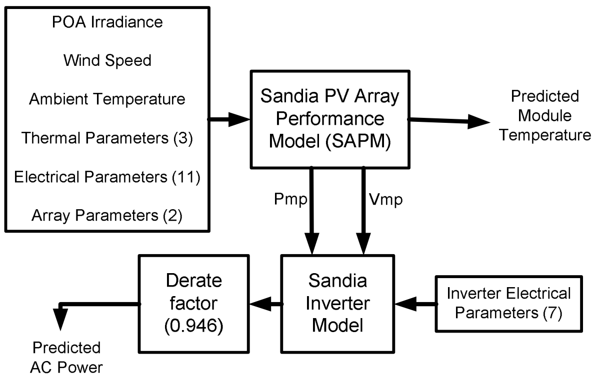

Figure 2: Data flow for component-based model

The Sandia PV Array Performance Model (SAPM) and the Sandia inverter model are widely recognized models for predicting PV system output based upon a given set of meteorological and irradiance data. Unlike the RNN model, these models can predict PV system output without requiring prior performance and weather data, and thus may be used prior to system installation. However, as noted above, the performance parameters for the particular system components (modules and inverters) must be determined from empirical testing and the number of components within the system must be known.

RESULTS

We trained the neural network to find the 490 weights to minimize the mean absolute error of both the module temperature and AC power over only the 30,000 point training data set. After establishing the synaptic weights, the entire 227,455 point data set (training data and test data) is processed to establish predictions of AC power and module temperature from the RNN. Figure 3 and Figure 4 show two sample days of predictions from the RNN and indicate that the RNN model is appropriately determining the relationships on both clear and partly cloudy days. Figure 5 and Figure 6 show the modeled vs. measured performance of the RNN on the aggregate 227,455 point data set. The coefficients of determination (R^2) shown indicate that the recurrent neural network is capable of accounting for over 99% of variation in AC power, and over 95% of variation in module temperature. Calculation of the coefficient of determination is shown in (3).

$$R^2 = 1 - \frac{\sum_i \left(X_i - \hat{X}_i\right)^2}{\sum_i \left(X_i - \overline{X}\right)^2} \qquad (3)$$

where:

X_i = the i^th measured data point

$\hat{X}_i$ = the i^th modeled output data point

$\overline{X}$ = the mean value of all measured data points

As the training data consisted of only data from late February to mid-April, the ambient temperatures experienced during training are lower than the ambient temperatures included during the summer of the testing data, requiring the RNN to extrapolate the performance of the system in high ambient temperatures. The irradiance and wind speed during included in the training data nearly cover the full range of irradiances and wind speeds in the test data.

Note that during the operation of the system there were approximately 2-3 days during which the PV system was not performing correctly and the AC power measured 0 watts. These data were nonetheless included in the evaluation of both models.

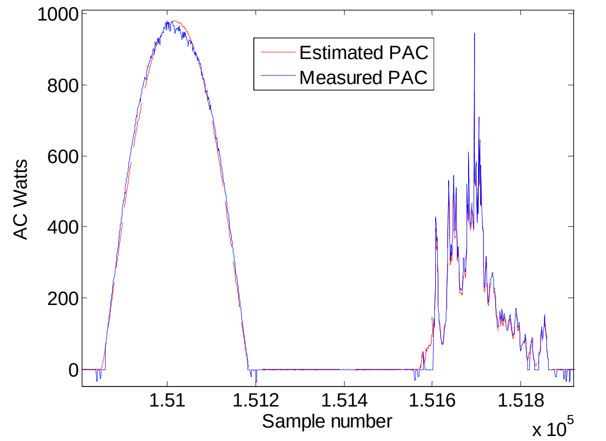

Figure 3: Estimated and Measured AC Power for a clear and cloudy day using an RNN model

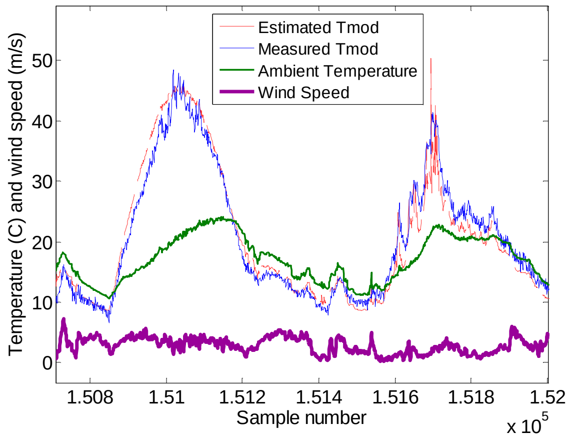

Figure 4: Estimated and measured module temperature for a clear and cloudy day using an RNN

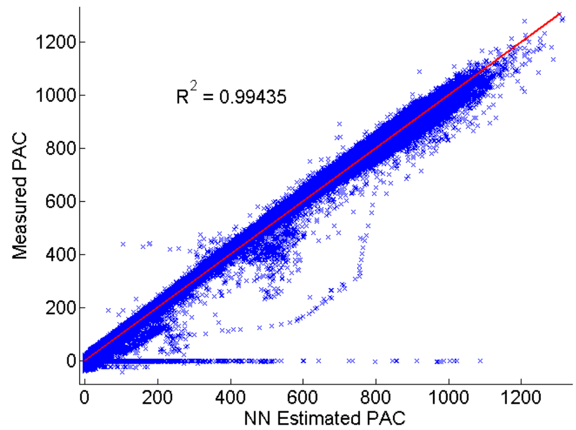

Figure 5: Measured AC power vs. RNN estimated AC power with 1:1 line

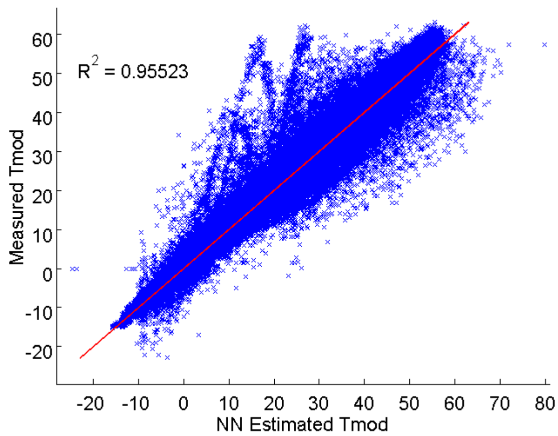

Figure 6: Measured module temperature vs. RNN estimated module temperature with 1:1 line

The Sandia component based models perform similarly as shown in Figure 7 and Figure 8. The component-based models clearly perform better in predicting module back temperature with a tighter distribution around the 1:1 line.

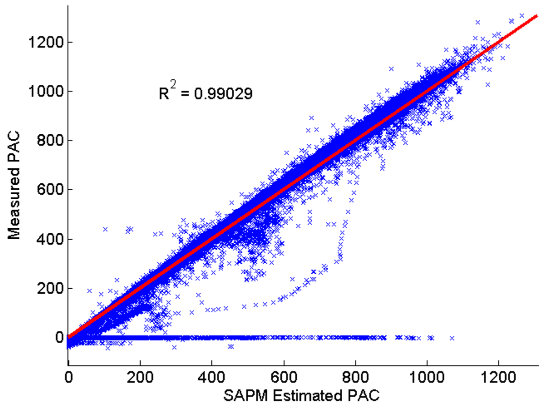

Figure 7: Measured AC power vs. component-model estimated AC power with 1:1 line

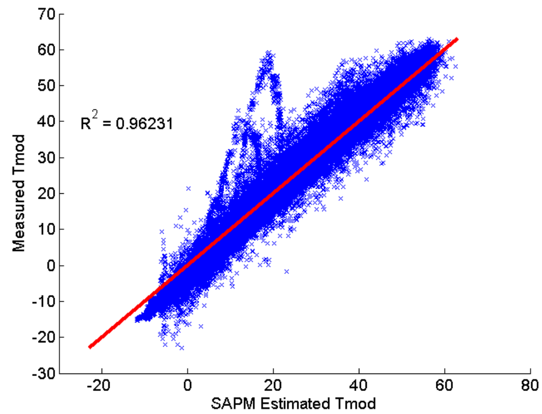

Figure 8: Measured module temperature vs. component-model estimate with 1:1 line

The mean bias error (MBE), mean absolute error (MAE), and root mean squared error (RMSE) for each model can provide some insight on the overall performance of the models.

Model	RMSE	MAE	MBE
SAPM	3.050 %	1.052 %	-0.069 %
RNN	3.198 %	1.178 %	0.321 %

Table I: Error measurements for AC power for both models, measured in % of rated power (W/W_P)

Model	RMSE	MAE	MBE
SAPM	3.1	2.3	1.0
RNN	3.2	1.9	-0.4

Table II: Error measurements for module backside temperature for both models, measured in ℃

While the overall performance predictions of a model are important, they do not show the model's shortcomings with respect to its input parameters. It is important to examine the model residuals (measured value – modeled value) with respect to the model inputs in order to evaluate the tendency of a model to incorrectly predict output across a range of input values [5]. Results of such an analysis are shown.

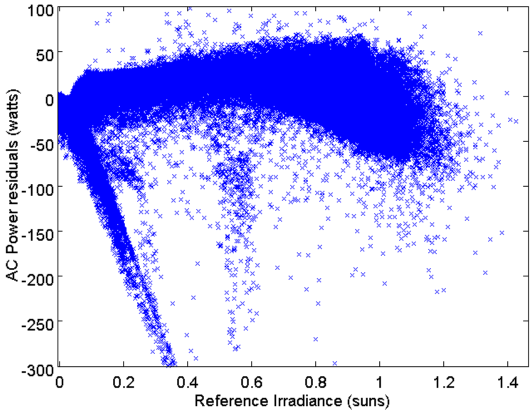

Figure 9: RNN modeled AC power residuals vs. reference POA irradiance

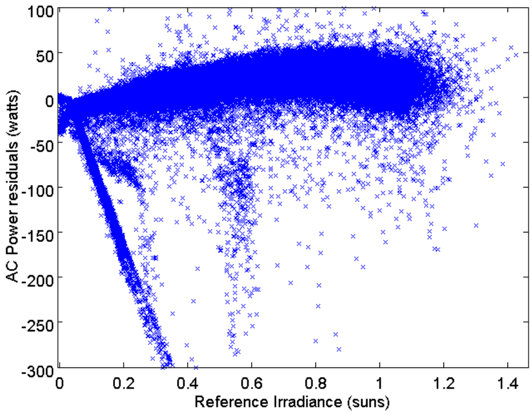

Figure 10: Component-based modeled AC power residuals vs. reference POA irradiance

Observing the model residuals of AC power with respect to reference irradiance (see Figs. 9 and 10) shows that the RNN model is able to capture much of the same effects as the component-model, but has a wider distribution of residuals at high irradiance (an undesirable trait). Again, the data shown includes times where the PV system was disconnected (producing no power), but the monitoring remained active; explaining the straight data “tail” where the models predict the system should be producing power.

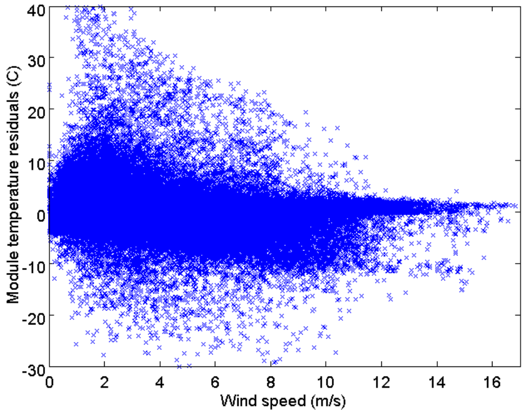

Figure 11: RNN modeled module temperature residuals vs. wind speed

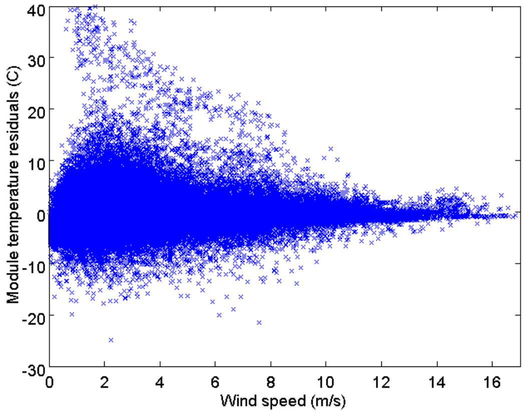

Figure 12: Component-based modeled module temperature residuals vs. wind speed

The residuals for predicted module backside temperatures with respect to wind speed are the most notable of the temperature residuals. Note that the component-based model improves with increasing wind speed, but the RNN based model does not show similar improvement.

CONCLUSIONS

The results presented show that modeling and characterizing an existing PV system with a recurrent neural network may provide adequate results for existing PV systems, although in this case, the RNN model did not perform as well as the component-based model. Thus, it seems that in the case where component parameters are known, a traditional PV modeling approach may yield more accurate model results.

The RNN model correctly learned the relationships between the weather data and performance data; the characterization required only concurrent input weather and output performance data. As such, a characterization may be performed when standard modeling parameters are unknown or may not be applicable due to abnormalities in the system (e.g. location, mounting). Furthermore, the RNN was able to learn the "derate" factors associated with the system. However, it is important to remember that the required performance data limits the use of an RNN characterization to existing systems which have been monitored for several weeks. This limitation will not allow an RNN model to provide energy predictions prior to installation, but it should allow for degradation or soiling detection, detection of performance-affecting failures on larger systems, or calculation of energy rebates by utilities.

The capability of an RNN to accurately characterize a system is entirely based upon the set of training data provided to train the RNN. As such, an optimal training data set should include data which is representative of the full range of expected conditions. Extrapolation outside the range of training data is possible, but may not be as accurate. Of course, the same caveat may apply to testing components for generation of performance parameters in component-based models.

The characterization of a PV system using a recurrent neural network is clearly a possibility in circumstances which do not allow for traditional modeling techniques. The use of a neural network also makes it trivial to add input and output data fields which are not included in this study. For example if POA irradiance from a spectrally similar reference cell is not available, it may be possible to use information such as sun position and airmass (as a proxy for detailed spectral information).

Further work on the subject of RNN characterization and modeling should extend the modeling capabilities to more difficult data sets (e.g., without spectrally corrected POA irradiance), include more input parameters, and attempt to learn using a smaller training data set.

ACKNOWLEDGMENT

Some work performed with funding from the National Science Foundation, USA under the grants – CAREER ECCS # 0348221 and ECCS # 0802047. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.

REFERENCES

[1] D. King, W. Boyson, J. Kratochvil, "Photovoltaic Array Performance Model". SAND2004-3535. Sandia National Laboratories 2004.
[2] A. P. Engelbrecht, "The Artificial Neuron" and "Particle Swarm Optimization" in *Computational Intelligence*, 2nd ed. Chichester, West Sussex, England: Wiley, 2008, pp. 17-25, 289-309.
[3] D. King, S. Gonzalez, G. Galbraith, W. Boyson, "Performance Model for Grid-Connected Photovoltaic Inverters". SAND2007-5036. Sandia National Laboratories 2007.
[4] National Renewable Energy Laboratory. PVWATTS Calculator for Overall DC to AC Derate Factor. http://rredc.nrel.gov/solar/calculators/PVWATTS/version1/derate.cgi (accessed 6/8/2011)
[5] J. Stein, *et al,* "A Standardized Approach to PV System Performance Model Validation" in IEEE Photovoltaic Specialist Conference, Honolulu, HI, 2010
[6] Almonacid F, Rus C, Pérez P J, Hontoria L, "Estimation of the Energy of a PV Generator Using Artificial Neural Network". *Renewable Energy*, vol. 24, pp. 2743-2750, Dec. 2009.
[7] Kennedy J, Eberhart R, "Particle Swarm Optimization" *Proc. IEEE International Conference on Neural Networks*, vol. 4, pp 1942-1948, Dec. 1995
[8] Mellit A, Pavan A, "Performance prediction of a 20 kWp grid-connected photovoltaic plant at Trieste (Italy) using artificial neural network". *Energy Conversion and Management*, vol. 51, pp. 2431-2441, Dec. 2010.