Modelling the Power Production of a Photovoltaic System: Comparison of Sugeno-Type Fuzzy Logic and PVSAT-2 Models

Radu Platon*, Sophie Pelland and Yves Poissant
CanmetENERGY, Natural Resources Canada, 1615 blvd. Lionel-Boulet, J3X 1S6 Varennes, Québec, Canada
*rplaton@nrcan.gc.ca

Abstract

The Sugeno-type fuzzy logic inference method and the PVSAT-2 method were used to develop models of a photovoltaic system. The models use solar irradiance in the array plane and back-of-module temperature in order to predict alternating current (AC) power production. Measured data from a photovoltaic (PV) system in Quebec, Canada was used to train and validate the models. Global models, over all irradiance values, and models developed for different operating regimes dictated by irradiance values were computed. Model validation shows that the Sugeno-type fuzzy logic model outperforms the PVSAT-2 model when sufficient amounts of data are available, with validation root mean square errors of 5.6% and 6.5%, respectively, for the global models. Meanwhile, training separate models over different irradiance regimes yielded no improvement in accuracy with respect to global models. The predictive accuracy of the Sugeno model as a function of the number of inference rules and the effect of the training dataset size on the performance of both methods were studied. The results indicate that the models have relatively good prediction accuracy when trained on a small dataset (1 month), and that the PVSAT-2 is more robust to accommodating observations that are outside the training dataspace. The models developed will form the basis of a fault detection and diagnosis system for PV systems.

1. Introduction

The photovoltaic market and technology have grown rapidly over the past years, becoming a mature technology for power production from renewable energy sources and a common on-site electricity generation strategy. However, the growth of the number of installed PV systems has not been followed by an increase in the number of properly monitored PV systems. In an analysis of more than 300 grid-connected PV systems in Europe, it was noted that causes for low yields included partial shading, string failures and partial inverter failures [1]. These issues could be detected by simple performance monitoring and fault detection methods, leading to improved performance. Smaller systems, with an output power inferior to 25 kWp, are also often operated without a supervisory mechanism [2]. The goal of this study was to analyze and compare two PV modelling approaches that could serve as a basis for such a fault detection method, namely the PVSAT-2 model currently in use in Europe [3] and an approach based on Sugeno-type fuzzy logic inference.

Different methods can be used to develop predictive models for PV system power production. Parametric models employ a set of mathematical equations that use PV system and weather variables along with adjustable parameters to predict the electricity production. For example, the PVSAT-2 model uses the ambient temperature and solar irradiance, along with parameters derived from
manufacturer specifications or PV system component databases to estimate the electricity production
[4]. Artificial intelligence techniques, such as neural networks, fuzzy logic and expert systems,
constitute powerful alternatives to conventional modelling methods, as they can learn from examples,
and are able to efficiently deal with noisy data and nonlinearities; they have been applied in the
photovoltaic field for different applications, including analyzing PV system behaviour and
performance [5]. The majority of the reviewed publications present the performance of a single
modelling method; comparative studies of the accuracies of AI-based and conventional approaches are
rarely available. This paper presents the development of two predictive modelling approaches used to
estimate the electricity production of a PV system: the Sugeno-type fuzzy inference modelling method,
a fuzzy logic-based technique, and the PVSAT-2 parametric model. Field tests showed that the
PVSAT-2 model is accurate when calibrated using historical system data, and the Sugeno-type fuzzy
logic model has been proved to accurately model non-linear relations using input and output system
data [6].

The models used in this study use back-of-array temperature and solar irradiance in the array plane as
inputs. Two types of approaches were tested for both the Sugeno and PVSAT-2 models: a global
approach over all irradiance values and an approach involving a separate model for each of three
regimes dictated by the solar irradiance values: 50–300 W/m², 300–1000 W/m² and superior to 1000
W/m². The predictive performance of these models was compared. The accuracy of the Sugeno model
as a function of the number of inference rules, as well as the effect of the training dataset size on the
accuracy of both methods were studied.

2. Data collection and pre-processing

Historical system data was collected from a PV system located in the city of Varennes in Quebec,
Canada (near Montreal). This PV array is rack-mounted on a rooftop and supplies electricity to the
building directly at all times. It consists of 8 strings, each string containing 14 AstroPower APC 5103
modules connected in series; the array is connected to a 4 kW Omnium inverter. The nominal DC
power of the PV system at Standard Test Conditions (STC) is 5.376 kW. Measurements of the direct
current (DC) input power to the inverter, the AC power produced by the inverter, the back-of-module
temperature and solar irradiance in the PV array plane were collected on-site from January to
December 2008 and used for modelling. The solar irradiance was measured using a Kipp & Zonen
CMP21 pyranometer. Measurements were recorded every minute and averaged over each hour for the
purpose of this analysis.

Since the model output should represent the expected power production of the PV system under
normal operation conditions, data associated with faulty operation were identified and eliminated. The
presence of faulty data can be caused by instrument or equipment malfunction, environmental
conditions such as snow or dust coverage of the PV modules and factors related to the installation
location, such as shadowing caused by trees or buildings. The first step of data pre-processing was the
elimination of observations corresponding to very low solar irradiance values – less than 50 W/m² –
for which pyranometer accuracy is significantly reduced. In order to detect potential observations
related to inverter malfunction, the inverter efficiency curve was examined; this efficiency is defined
as the ratio of the AC output to the DC input of the inverter. A few inverter efficiency values were
superior to 100%, which is evidently impossible and must be attributed to inaccurate AC or DC power
readings for these data points. Also, an inspection of the inverter efficiency plot and analysis of data revealed the presence of a few outliers; they were removed from the dataset. Figure 1 (left) shows the inverter efficiency plot; the DC power is shown on the x-axis of the plot.

The PV array efficiency, defined as the ratio of the AC energy output from the inverter to the in-plane solar energy reaching the array, was also analyzed in order to detect observations not representative of normal operation. Figure 1 (right) shows the efficiency plotted against the solar irradiance; it is seen that the system efficiency decreases at higher irradiance values, from an average of 0.071 in the irradiance interval of 300-1000 W/m² to an average of 0.069 for irradiance values superior to 1000 W/m²; this is partly caused by an increase of the array temperature at high solar irradiances. A few system efficiency outliers were identified and removed from the dataset.

![Inverter efficiency plot](image1.png)

**Fig. 1.** Inverter and array efficiency plots.

Given the relatively low number of values not representative of the general trend of the data, as it can be seen in Figure 1, the outlier detection was performed manually, by visually inspecting the plots and the corresponding numerical values in the dataset. The outlier detection procedure can be automated, by using thresholds for defining the normal operation levels; for example, system efficiency values can be placed in solar irradiance bins and the fault detection threshold can be set according to a 95% confidence interval [7].

### 3. Sugeno-type fuzzy logic and PVSAT-2 modelling

#### 3.1. Sugeno-type fuzzy logic modelling

Traditional logic allows for only two truth values – the binary 1 or 0, or “true” or “false”. These crisp values only afford two levels of set membership – either an element is a member of the set or it is not a member. Fuzzy logic is an approach to computing based on "degrees of truth" rather than the usual "true or false" (1 or 0) Boolean logic. With fuzzy logic, an element could partially belong to a set or to multiple fuzzy sets and this is represented by the set membership [8]. In the Sugeno-type fuzzy methodology (also known as Takagi-Sugeno-Kang), the model contains a set of fuzzy rules generated from a given input-output data set. These if-then rules represent the knowledge base of the system. In Sugeno-type models, the output of a rule is expressed as a polynomial function of the inputs; a typical first-order multi-input one-output Sugeno-type fuzzy inference model consists of a set of rules of the form:
If \( x_1 \) is \( A_1 \) and \( x_2 \) is \( A_2 \) ... and \( x_i \) is \( A_i \) Then \( z(u) = p_1u_1 + p_2u_2 + \ldots + p_nu_n + k \),  
\( (1) \)

where \( n \) is the number of input dimensions, \( A_i \) is the rule antecedent associated with the \( i^{th} \) input, \( u_i \) is the value of input \( x_i \), \( z \) is the rule output, \( p_i \) is the rule consequent associated with the \( i^{th} \) input dimension and \( k \) is a numerical constant. The system modelling is done by grouping the input-output data into clusters of similar behaviour and evaluating system behaviour from these clusters. The subtractive clustering algorithm was used to develop the models presented in this paper [9]. Each cluster is represented by a polynomial function. For a given input, each rule has a certain weight on determining the model result for that input. The output of each rule is multiplied by the rule weight, and the final output of the Sugeno system is the weighted rule average of all rule outputs. Detailed information regarding Sugeno-type fuzzy inference can be readily found in the literature [10], [11].

### 3.2. PVSAT-2 modelling

The PVSAT-2 method is a parametric approach to PV system modelling, which has been used to develop an automated failure detection routine [3]. In this approach, the array DC output power at the maximum power point, \( P_{DC} \), is given by:

\[
P_{DC} = G(A_1 + A_2G + A_3ln(G))(1 + A_4(T - 25))
\]

where \( G \) is irradiance in the array plane, \( T \) is back-of-module temperature and the \( A_i \) are adjustable parameters. The PVSAT-2 fit for DC power described in [3] involves a fit on efficiency rather than power, and multiplies this by wiring and generic loss terms. The approach used in this study is therefore a slight variation on PVSAT-2.

AC power, \( P_{AC} \), was modelled as linear in the DC power, with a linear fit performed over the training dataset. The AC power is given by:

\[
P_{AC} = B_1P_{DC} + B_2
\]

where the \( B_i \) are adjustable parameters. Note that while PVSAT-2 uses a 3-parameter model for the inverter efficiency, a linear (2-parameter) fit for AC power was judged to be sufficient for the purposes of this study, since it yielded a coefficient of determination of 0.9997.

### 4. Models developed using one year of data

Measurements collected from January to December 2008 were used to build predictive models of the AC power production using the solar irradiance in the array plane and array temperature.

#### 4.1. Operating regimes

Solar irradiance is the parameter that affects the most the performance of a PV module. As shown in Figure 1 (right), the efficiency of the PV system changes rapidly at irradiance levels less than about 300 W/m\(^2\), is fairly constant under irradiances between 300 W/m\(^2\) and 1000 W/m\(^2\), and starts to decrease at higher irradiances. In order to determine if the predictive accuracy increases if the dataset is divided into operating regimes dictated by the solar irradiance values, models corresponding to the 50-300 W/m\(^2\), 300-1000 W/m\(^2\) and superior to 1000 W/m\(^2\) irradiance values were computed. Global models, using the entire dataset, irrespective of the irradiance, were also developed.
4.2. Training and validation datasets

Prior to computing the models, the dataset was separated into training and validation subsets. The training dataset was used for developing the model, while the validation dataset was used to validate the model performance. The validation dataset was obtained by setting aside a portion of the original dataset that was not used during the training process. After the model was fitted on the training data, its performance was tested on the validation data. For each dataset, 20% of the observations were randomly selected, in a uniformly distributed manner, as validation data. This selection procedure ensured that no systematic pattern was present in the validation dataset, and that the validation data covered the entire training dataset. Overall, the training dataset contained 2314 observations, while the validation dataset contained 567 observations.

4.3. Model results

The models were scored in terms of the training and validation Root Mean Square Error (RMSE %). This error is calculated as the square root of the average of the squares of the error for each observation, divided by the average value of the dataset.

In the case of Sugeno models, different clustering parameters can lead to models with the same number of rules; therefore, for the same number of rules, only the models with the lowest training and validation RMSE, respectively, were kept. Models displaying the best modelling and validation error trade-off are considered as most accurate; at a certain point during the modelling process, the training accuracy continues to increase and the validation accuracy starts to decrease, as the model becomes too precise on the training data and is not able to generalize on unseen data. The best model in terms of training and validation errors was selected just before this phenomenon occurs.

Sugeno and PVSAT-2 models were computed for each of the three operating regimes corresponding to the irradiance levels mentioned in Section 4.1. Global models that use the entire dataset – irrespective of the irradiance – were also computed. The modelling results are presented in Table 1. The results indicate that the 3 regime approach and the global approach perform comparably, with the global model performing slightly better in the Sugeno case. This indicates that dividing the data into different irradiance regimes does not yield any benefits in terms of overall model accuracy. The Sugeno model outperforms the PVSAT-2 model, for both the different operating regimes and global models. In the case of different operating regimes, the model accuracy increases as the irradiance levels increase. The models fare relatively well: for the global models, the RMSE values are 6% or less for the Sugeno model, and 6.5% for the PVSAT-2 model.

Model accuracy of a 2 rule model is close to that of the 18-rule model, considered as the best model, and is superior to that of the PVSAT-2 model. Both the Sugeno 2-rule and the PVSAT-2 models have 6 parameters; therefore, for the same number of model parameters, the Sugeno approach proves to be more accurate given sufficient amounts of training data.
Table 1. Model performance.

<table>
<thead>
<tr>
<th>Model</th>
<th>Training RMSE (%)</th>
<th>Validation RMSE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 regimes, Sugeno</td>
<td>5.63</td>
<td>5.72</td>
</tr>
<tr>
<td>3 regimes, PVSAT-2</td>
<td>6.61</td>
<td>6.47</td>
</tr>
<tr>
<td>Global Sugeno – 18 rules</td>
<td>5.78</td>
<td>5.63</td>
</tr>
<tr>
<td>Global Sugeno – 2 rules</td>
<td>6.18</td>
<td>5.96</td>
</tr>
<tr>
<td>Global PVSAT-2</td>
<td>6.72</td>
<td>6.50</td>
</tr>
</tbody>
</table>

The results analysis revealed that the Sugeno model accuracy is fairly insensitive to the number of fuzzy rules, provided that the model is not over-fitted. The validation RMSE starts at 13.00% for a 1-rule model, decreases to 5.96% for a 2-rule model and reaches its lowest value of 5.63% for the 18-rule model. The evolution of the training and validation errors for the models having 1 to 35 rules is shown in Figure 2; this figure also shows that the most accurate global Sugeno models has 18 rules in this case.

![Sugeno models training and validation errors vs. number of fuzzy rules](image)

Fig. 2. Training and validation errors as a function of the number of fuzzy inference rules for the global model.

### 5. Models developed using one month of data

The effect of the training dataset size on the predictive accuracy of the models using the global dataset was investigated. The models were trained using only a month of data, and they were validated using the remaining 11 months. One of the limitations of data-driven modelling approaches is that a model’s predictive accuracy decreases for validation values that are outside the training data space. In order to determine the amount of the validation observations located outside the training dataspace, the minimum and maximum values of the array efficiencies for the training observations were calculated; then, the array efficiency values for the validation data – the remaining 11 months – were also
calculated, in order to determine how many of them fall outside the numerical range of the training minimum and maximum values. The results are presented in Table 2, and they indicate that the PVSAT-2 model extrapolates outside the training dataspace better than the Sugeno model, as it is more accurate despite the large number of validation observations outside the training range. Also, if the training data set is chosen between April and August, the PVSAT-2 model accuracy with one month of training is as good as the accuracy using one year of training, while this is never the case for the Sugeno model. The Sugeno model proves to be more accurate than the PVSAT-2 model provided that few validation observations are outside the training dataspace. The validation Mean Bias Error (bias %) was also calculated, as it indicates the total AC production difference between the measured and model-computed AC values for the 11 months used as validation data.

Table 2. Models developed using one month of data.

<table>
<thead>
<tr>
<th>Training month</th>
<th>PVSAT-2 validation errors</th>
<th>Sugeno validation errors</th>
<th>% of validation array efficiency values outside the training range</th>
<th>Most accurate model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE %</td>
<td>Bias %</td>
<td>RMSE %</td>
<td>Bias %</td>
</tr>
<tr>
<td>January</td>
<td>11.59</td>
<td>7.00</td>
<td>7.81</td>
<td>-4.13</td>
</tr>
<tr>
<td>February</td>
<td>11.31</td>
<td>6.22</td>
<td>8.47</td>
<td>-1.55</td>
</tr>
<tr>
<td>March</td>
<td>10.17</td>
<td>5.59</td>
<td>7.14</td>
<td>-0.95</td>
</tr>
<tr>
<td>April</td>
<td>5.99</td>
<td>-0.19</td>
<td>7.06</td>
<td>1.37</td>
</tr>
<tr>
<td>May</td>
<td>5.72</td>
<td>-0.41</td>
<td>8.87</td>
<td>2.69</td>
</tr>
<tr>
<td>June</td>
<td>5.94</td>
<td>-0.61</td>
<td>8.01</td>
<td>3.25</td>
</tr>
<tr>
<td>July</td>
<td>6.16</td>
<td>0.36</td>
<td>16.61</td>
<td>9.09</td>
</tr>
<tr>
<td>August</td>
<td>5.94</td>
<td>1.30</td>
<td>7.99</td>
<td>-0.35</td>
</tr>
<tr>
<td>September</td>
<td>7.31</td>
<td>3.37</td>
<td>7.55</td>
<td>-3.27</td>
</tr>
<tr>
<td>October</td>
<td>6.69</td>
<td>3.57</td>
<td>7.27</td>
<td>-2.75</td>
</tr>
<tr>
<td>November</td>
<td>9.18</td>
<td>4.96</td>
<td>8.23</td>
<td>-3.59</td>
</tr>
<tr>
<td>December</td>
<td>10.84</td>
<td>6.54</td>
<td>7.64</td>
<td>-2.69</td>
</tr>
</tbody>
</table>

5. Conclusion

Sugeno fuzzy logic and PVSAT-2 models were developed to predict the AC power production using the solar irradiance in the array plane and the back-of-module temperature as inputs. One year of data – 2008 – was used to train the models. Prior to computing the models, outliers and abnormal values were eliminated. Models using data separated into 3 different operating regimes, according to the irradiance levels: 50-300 W/m², 300-1000 W/m² and superior to 1000 W/m² were developed, and their predictive performances were compared to those of global models with no irradiance regime separation. This comparison showed that dividing data into irradiance regimes does not yield benefits in terms of overall model accuracy.
The results revealed that the Sugeno models are more accurate than the PVSAT-2 models, provided that the training dataset covers the whole range of operating conditions. However, the PVSAT-2 model extrapolates outside the training dataspace better than the Sugeno model; this very beneficial when only a small set of measured data is available as is the case for new installations.

The strengths of each method can be combined to adapt the model to the amount of data available: when little data is available, the PVSAT-2 model can be used since it performs better; however, as more data becomes available, the Sugeno model can be used since it out-performs the PVSAT-2 model.

It was also found Sugeno models with very few rules are accurate and still outperform the PVSAT-2 models when sufficient amounts of training data are available; this allows for the development of the model in a short period of time.

References


