SENSITIVITY OF PV SYSTEM DEGRADATION RATES TO DATA FILTRATION

YINGLI SOLAR

Kenneth J. Sauer

2011 PV MODULE RELIABILITY WORKSHOP, NREL, Colorado.
February 16 – 17, 2011

1. Introduction

PROBLEM STATEMENT

- Photovoltaic (PV) project size and financing are governed – and potentially limited – by the confidence investors have in the projected energy production of systems over the terms of their power purchase agreements or planned asset lifetimes.
- Production estimates are largely a function of anticipated annual system degradation rates, which are known to significantly vary both within and across technologies and systems.
- Objectively mining accurate downward trends in system performance over time (commonly referred to as annual system degradation rates) from large sets of historical meteorological and production data is an essential but challenging task, as each step in degradation analysis (summarized at right) carries the potential to contribute uncertainty and error.
- The frequent presence of noise in such data sets, stemming from any number of possible root causes, makes the task all the more challenging.

STEPS IN DEGRADATION ANALYSIS

1. Data input
   - Raw meteorological and production data
     - Quantity – observation interval (hourly or sub-hourly) and time (years)
     - Quality – equipment measurement uncertainties and their evolution over time
   - Integrated external data
     - Incorporation of data from multiple sources (e.g., ambient temperature or wind speed data from nearby RMIS)
   - PV module and system characteristics
     - Presence of light-induced degradation (extent and timeframe)
     - Agreement with module "datasheet" parameters (e.g., tolerances, temperature coefficients, NOCT, efficiency changes with temperature and irradiance)
     - Occurrence of downtime or climatic events
     - O&M procedures
2. Data transformation
   - Energy model
   - Thermal model (e.g., Martin Green "simple model", NOCT model, Fuentes, Sandia, CEC, BEW, IEC 61853-1, Dows)
3. Data cleansing
   - Statistical identification and processing of outliers and data shifts
   - Filtration method used to process noisy data that fail to meet user-specified upper and lower qualification limits
4. Data normalization
   - Power model (e.g., PVFORM, Myers, PVUSA, BEW, King, performance ratio)
5. Data evaluation
   - Statistical method (e.g., linear least squares fit, seasonal indexing, classical decomposition, ARIMA)

2. Approach

MODEL

- An Excel/VBA model has been built with a high degree of flexibility in all aspects of degradation rate analysis to enable research on the sensitivities of degradation rates to the numerous possible analytical approaches used to calculate them.
- In an effort to limit the potentially broad scope of investigation that the model allows, this study restricts each step of degradation analysis to one set of model configurations, as detailed in the sections that follow.

OBJECTIVE

- This particular study explores the application of the PVFORM power model as a possible data filtration tool for processing noisy raw input data that fail to meet user-specified upper and lower qualification limits. For global plane-of-array irradiances >125W/m², this model is more commonly referred to as either the single-point efficiency model (with temperature correction), or the power temperature coefficient model.

CASE STUDY

- A system with >1.5 years of continuous historical meteorological and production data is evaluated in this study.
- The chosen system experienced both inverter downtime and significant snow events, which are (thus) present in the raw input data set. Although these events contributed noise to the data that would influence the results of this sensitivity study, the occurrence of such events and presence of noise in raw data are not uncommon for large-scale PV systems. The selected system is therefore considered to be a realistic and not extreme case study.

DISCLAIMERS

- The effects of light-induced degradation, soiling, spectral shifts, sensor drift, and operations & maintenance (O&M) have not been analyzed at this time.
- Potentially sensitive project-specific details (e.g., equipment suppliers and owner information) are left undisclosed at this time.
- As this study is largely a work in progress, several future steps needed to further reduce analytical uncertainty and more objectively evaluate annual system degradation are outlined throughout the following sections.

3. Data Input

Raw Meteorological & Production Data

- 19 full months (1.6 years) of non-spectrally corrected 15-minute interval data were used in this case study. Longer-term (>4 year) data sets expected to yield more meaningful degradation rates will be explored in future work.
- Data for the chosen system included day, month, time, AC energy output (E_ac, kWh), global plane-of-array irradiance (G_poa, W/m²), module backsheet temperature (T_mod, °C), and ambient temperature (T_amb, °C).
- Initial uncertainties for G_poa, T_mod, and T_amb measurements were +/-5%, +/-0.5°C, and +/-0.3°C, respectively.
- Measured T_mod data (recorded for a small sample size of modules) were initially unavailable but eventually obtained.
- E_ac is a summed value, while G_poa, T_mod, and T_amb are instantaneous. It is hypothesized that the nature of the meteo readings as instantaneous (rather than averages or medians) may be contributing noise to the raw data.

Integrated External Data

- No data from external sources were incorporated at this time; however, wind speed data from a nearby reference meteorological and irradiance station (RMIS) will be incorporated in future work.

PV Module & System Characteristics

- The selected case study is a south-facing, ground-mounted, large-scale (>1MWac) system using multicrystalline silicon PV technology.
- The system is located in an area with snowfall and experienced snow events in months 6 (13.3"), 7 (1.7"), 8 (32.9"), 18 (24.5"), and 19 (37.4").
- One 250kWac inverter malfunctioned intermittently throughout month 1 and was offline during much of the latter half of month 12 and beginning half of month 13.

4. Data Transformation

Energy Model

Sandia Thermal Model

(1) $E_{AC} = \Delta T \cdot \sum_{i=1}^{n} P_{AC,i}$

(2) $T_{MOD} = G_{POA} \cdot \left(e^{a+b \cdot WS}\right) + T_{AMB}$

(3) $T_{CELL} = T_{MOD} + \frac{G_{POA}}{G_{STC}} \cdot \Delta T$

- Eq. (1) was used to convert between E_ac and AC power (P_ac, kW) as needed throughout the analysis [1].
- Cell temperature (T_cell, °C) values were required for input into the PVFORM power model (discussed below). As measured T_mod was initially unavailable for the chosen array, the Sandia thermal model was used to convert T_amb to T_mod (Eq. 2) and then T_mod to T_cell (Eq. 3) [2]. Coefficients "a", "b" and "ΔT" were derived using the PVModuleWizard tool provided by Maui Solar Energy Software Corporation, and based on EN 50380 module datasheet values. WS was assumed constant at 1 m/s, inevitably introducing some amount of error. T_mod data for a small sample size of (1) module were recently obtained and evaluated against the Sandia and also California Energy Commission (CEC) thermal models (see Fig. 1 and 2).
- In future work, measured T_mod values will be investigated; however, it is uncertain whether the sample size is large enough to adequately represent the module temperature distribution across the array.

Figure 1.

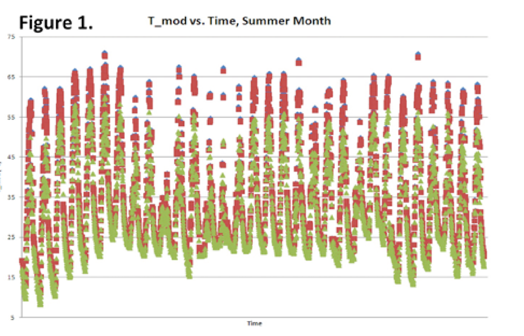
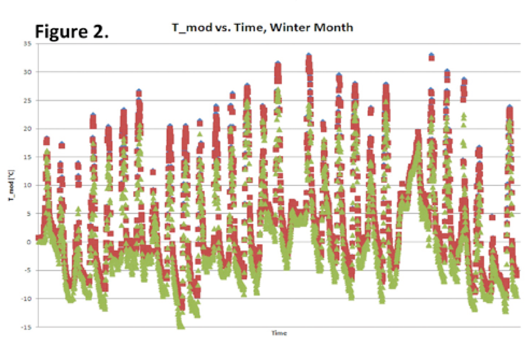


Figure 2.



5. Data Cleansing

Boxplot

(4) Lower limit = Q1 - 1.5 (Q3 - Q1)
(5) Upper limit = Q3 + 1.5 (Q3 - Q1)

PVFORM

For G_poa >125W/m²:

$$(6)\quad L.L. \leq \frac{P_{AC,t} \cdot \frac{G_{STC}}{G_{POA}}}{\left[1 + \lambda \cdot \left(T_{CELL} - T_{STC}\right)\right] \over P_{AC,t=0}} \leq U.L.$$

For G_poa <125W/m²:

$$(7)\quad L.L. \leq \frac{P_{AC,t} \cdot \frac{G_{STC}}{k \cdot G_{POA}^{2}}}{\left[1 + \lambda \cdot \left(T_{CELL} - T_{STC}\right)\right] \over P_{AC,t=0}} \leq U.L.$$

Table 1.

L.L.	Qual Designation	U.L.
98%	2	102%
95%	5	105%
90%	10	110%
80%	20	120%
70%	30	130%
0%	100	∞

- Boxplot analyses on P_ac values and PVUSA regression residuals were used throughout to statistically identify outliers per Eq. (4) and (5).
- Outliers in P_ac were present (and therefore processed) only when the PVFORM model (described below) was not used to filter out data.
- The primary objective of this study was to test the sensitivity of degradation rates to filtration, using the PVFORM power model as a possible tool for removing noisy data that failed to meet user-specified upper and lower qualification limits (UL and LL). A successful tool would remove data recorded when inverters were malfunctioning or offline or when snow was covering all or part of the array. If the instantaneous nature of the irradiance and temperature readings truly added noise to the raw input data, this data would likewise be processed.
- For all data above a G_poa threshold of 500W/m² (as defined in the PVUSA section below), P_ac was normalized to Standard Test Conditions using the PVFORM model along with the EN 50380 module datasheet values for the temperature coefficient of P_max (γ) and k coefficient [3,4]. The normalized values were then divided by the total system size (kWp AC STC) and evaluated against the user-specified qualification (qual) limits summarized in Table 1. Whenever the criteria were not fulfilled, the data would be filtered out. The PVFORM-based filtration methods evaluated in this study are summarized in Eq. (6) and (7). Note that the investigated qual limits mostly fall well outside a range of what might be expected for annual system degradation.
- The PVFORM power model is customarily applied to DC (typically module) power. Because only AC data was available, this study is also used to investigate the suitability of the PVFORM model to the AC power of the selected system. It is often difficult to avoid the effects of, e.g., inverter peak power tracking when evaluating long-term system performance or degradation using production data from operational PV systems.

6. Data Normalization

PVUSA

$$(8)\quad P_{PTC} = G_{POA} \cdot \begin{pmatrix} A + \\ B \cdot G_{POA} + \\ C \cdot T_{AMB} + \\ D \cdot WS \end{pmatrix}$$

- The multiple regression-based PVUSA power model (Eq. 8) was used to derive monthly system AC power ratings at PVUSA Test Conditions (PTC) from the filtered data set [5,6]. WS was assumed constant in this study as an approximation but will be integrated from a nearby RMIS in future work.
- PVUSA power ratings at PTC (P_ptc, kW) partly eliminate seasonal effects, as new sets of regression coefficients (A, B, C, and D) are generated for each month throughout the observation time. Variations in monthly PTC ratings can be used to indicate module, inverter, or climatic events if they exist in the raw data.
- Following Dows et al., G_poa values below a 500W/m² threshold were removed from the input data set. The effect of using other G_poa thresholds reported in literature (600, 700, 750, 800W/m²) – as well as time of day, P_ac, T_cell, T_amb, and WS thresholds – can conveniently be explored using the model that has been built.
- Additional regression-based models, including the simplified Dows, Myers, BEW, and King models, will be investigated in future work.

PVFORM

- The previously described PVFORM power model was used to derive monthly average P_stc values from the same post-filtered data set described above.

Performance Ratio

- The same data set was also used to derive monthly average performance ratio (PR) values. While the actual values of PR may be arbitrary due to filtration, decreasing trends may be useful for indicating system degradation.
- Although PR normalizes values with respect to solar radiation, monthly PRs are typically highest in winter months and lowest in summer months, resulting from module efficiency reductions at higher temperatures. Due to the pronounced seasonality effects present in monthly PR values, longer time spans are generally needed for deriving meaningful degradation analyses [7].

7. Data Evaluation

Linear Least Squares Fit

(9) $y = m \cdot x + b$

(10) $R_{DEG}\left(\frac{\%}{yr}\right) = \frac{(m \cdot 12)}{b} \cdot 100$

- The linear least squares fit method was used to generate trend lines described by Eq. (9) from normalized monthly values of P_ptc, P_stc, and PR.
- The y-axis intercept ("b"), indicating the normalized system power [kWp AC] at the beginning of the observation time, and the slope ("m") of the trend line were entered into Eq. (10) to yield degradation rates (R_deg) [8].
- A system experiencing no measureable annual degradation is expected to have a straight, horizontal trend line. Note that when evaluating the output of a linear regression model in degradation rate analysis, a low R^2 value can be expected due to the low signal-to-noise ratio caused by other sources of normal variation (e.g., spectral shifts). This is particularly true for shorter observation times.

Classical Decomposition

- Although the total observation time was less than the recommended two seasonal cycles, the classical time series decomposition (CD) method was applied for interest and comparison.
- Using the monthly normalized data sets as inputs, the CD method was used to extract the trend component – of most immediate interest in this investigation – from the seasonality and remaining irregular components [9].
- In CD, the trend component is obtained with a centered 12-month moving average applied to the normalized data. An additive model was chosen with a period of 12.
- Following CD, extracted trend values were evaluated using the linear least squares fit-based method described in Eq. (9) and (10) above.
- Future work (following at least two seasonal cycles) will further investigate the resultant seasonality and irregular components further.

8. Results: Data Filtration

Figure 3.

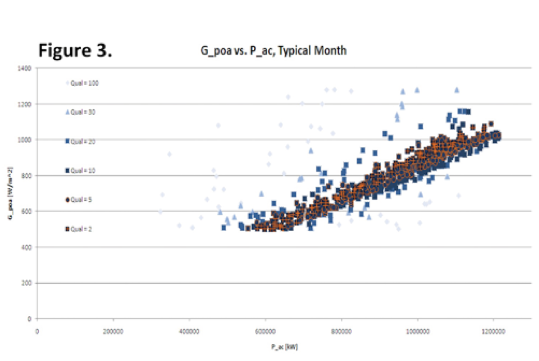
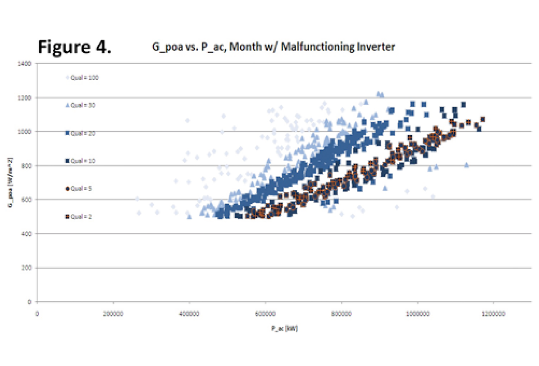


Figure 4.



- For typical months, such as that shown in Fig. 3, it was observed that tightening the qual limits mostly yielded a higher R^2 in the G_poa vs. P_ac trend line while making little or no difference to the fitted trend.
- Interestingly, the slope of the trend lines began to slightly increase for qual limits <5 (e.g., 98-102, 99-101). This may have been due to an overcorrection for T_cell in the PVFORM power model, stemming from the use of derived, rather than measured, T_mod values (where derived values exceeded measured, as previously shown). The relatively small effect was avoided by not exceeding a qual limit of 5 in the subsequent degradation rate evaluation.
- Note that a shift in the opposite direction for tighter qual limits could possibly indicate degradation. For official degradation rate reporting, qual limits and any data that are removed from the analysis should be statistically justified.
- Having applied a traditionally DC (module) power model to AC system power, there appears to be a strong correlation between module and system performance, at least during times of uninterrupted production. This may be indicative of proper inverter sizing and effective maximum power point tracking.
- Months with inverter downtime or snow events showed effects similar to that displayed in Fig. 4. As the qual limits tightened from 100 to 5, the trend line expectedly followed the cluster of data corresponding to higher P_ac values (for given values of G_poa).

9. Results: Data Normalization

Figure 5.

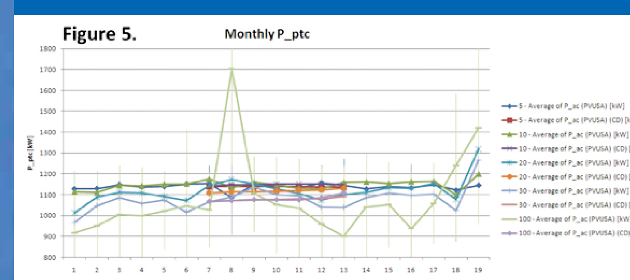


Figure 6.

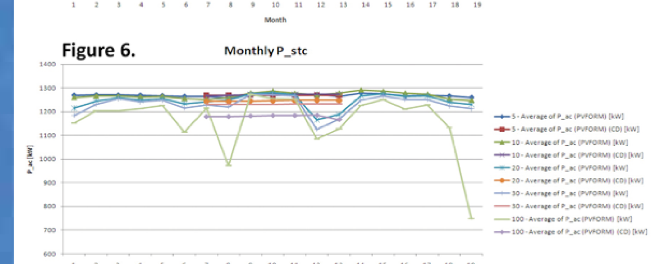
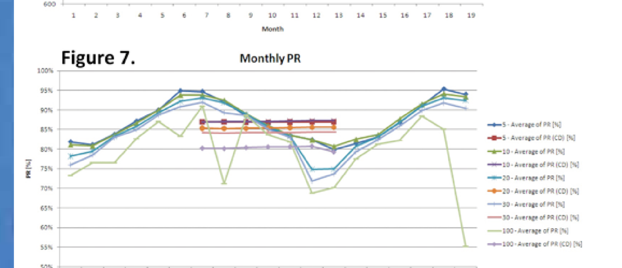


Figure 7.



- Figures 5, 6, and 7 display (as univariate time series) the monthly normalized P_ac values following PVUSA (P_ptc), PVFORM (P_stc), and PR analyses, respectively.
- The graphs seem to indicate the utility of the PVFORM power model as a filtration tool, at least in eliminating noisy data collected during inverter downtime (months 12 and 13) and significant (>12") snow events (months 6, 8, 18, 19). Note that normalized values converged for qual limits well outside an expected range of degradation.
- As expected, the relative standard error around P_ptc values (graphed in Fig. 5) decreased dramatically (from >30% to <2%) as the qual limits tightened from 100 to 5. It is important to note that model error – for all power models – is only one contributor to the total uncertainty associated with the normalized power values.
- As expected, the months with the lowest R^2 coefficients in the pre-filtered G_poa vs. P_ac plots (months 6, 8, 13, 18, and 19) shifted the most following filtration, while the months with the highest monthly R^2 (months 9, 10, 11, 15, 17) shifted the least (see Table 2 below). This helps statistically reinforce the utility of the PVFORM power model as a possible tool for removing noisy data.

10. Results: Data Evaluation

Table 2.

Month	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
R^2	0.37	0.43	0.51	0.40	0.45	0.21	0.51	0.09	0.70	0.56	0.56	0.35	0.25	0.42	0.62	0.29	0.52	0.13	0.00
Rank	12	9	6	11	8	16	7	18	1	3	4	13	15	10	2	14	5	17	19

Table 3.

Qual	PVUSA	ΔR_deg	PVUSA - CD	ΔR_deg
100	10.7%		6.3%	
30	6.4%	4.3%	3.8%	2.5%
20	6.1%	0.3%	3.6%	0.2%
10	1.7%	4.4%	1.1%	2.5%
5	0.1%	1.6%	-0.4%	1.5%
Qual	**PVFORM**	**ΔR_deg**	**PVFORM - CD**	**ΔR_deg**
100	-6.0%		-0.9%	
30	0.1%	6.2%	0.6%	1.5%
20	0.3%	0.1%	0.9%	0.3%
10	0.5%	0.2%	1.1%	0.2%
5	0.0%	0.5%	0.1%	1.0%
Qual	**PR**	**ΔR_deg**	**PR - CD**	**ΔR_deg**
100	-3.6%		-0.7%	
30	3.0%	6.6%	0.5%	1.2%
20	3.2%	0.2%	0.8%	0.4%
10	3.4%	0.1%	0.9%	0.1%
5	2.8%	0.5%	0.0%	0.9%

- Table 3 reports the "degradation" rates calculated using both linear least squares fit and CD and following the PVUSA, PVFORM, and PR models. In Tables 3 and 4, differences in rates between the various investigated qual limits and statistical methods are displayed (ΔR_deg) for relative comparisons between the different approaches.
- The near-zero or slightly positive rates suggest that the signal-to-noise ratio is too low for a meaningful assessment of, and therefore sensitivity analysis on, the system degradation rate at this time. Any general observations made here will continue to be monitored in future work, given that the results are subject to change as additional data is collected. An evaluation of monitoring equipment may also be conducted to identify any possible lurking variables (e.g., soiling or drift of sensors).
- From the results in Table 4, it appears as if degradation rates calculated using the multiple regression-based PVUSA power model are most sensitive to data filtration, followed by PR and PVFORM.
- Sensitivities to filtration are considerably reduced between qual limits and across power models through the use of CD.
- It appears as if PR is most affected by the use of classical decomposition, followed by PVFORM and PVUSA.

Table 4.

Power Model - Stat Method:	PVUSA	PVUSA - CD	Δ	PVFORM	PVFORM - CD	Δ	PR	PR - CD	Δ
Max ΔR_deg:	4.4%	2.5%	1.9%	6.2%	1.5%	4.7%	6.6%	1.2%	5.5%
Average ΔR_deg:	2.1%	1.5%	0.6%	1.4%	0.6%	0.8%	1.5%	0.5%	1.0%
St. dev. ΔR_deg:	2.1%	1.0%	1.1%	2.7%	0.6%	2.1%	2.9%	0.5%	2.3%
Overall (Qual: 100 -> 2) ΔR_deg:	10.6%	5.9%	4.7%	6.0%	0.9%	5.0%	6.4%	0.7%	5.6%

11. Summary

- Determining accurate PV system degradation rates is essential in yielding more favorable system production exceedance probabilities and improving the overall financial viability of PV. Such efforts can be challenging, however, particularly when handling large and oftentimes noisy sets of historical and meteorological data. Nearly every step in degradation analysis has the potential to contribute uncertainty and error to the calculated rates.
- An Excel/VBA model was built with a high degree of flexibility in all aspects of degradation analysis. The potentially broad scope of investigation was limited to focusing on the sensitivity of degradation rates to the quantity and quality of data that gets filtered out prior to data evaluation. Given the convenience of the model that was built, however, a number of additional aspects of degradation analysis can and will be explored in future work.
- It is important to note that the intention of this study was not to advocate any specific method of data filtration, but rather to gain experience with one possible approach (PVFORM) as applied to a real, operating PV system. In fact, careful consideration of qual limits would be needed when following this method for official degradation rate reporting. As a general rule of thumb, chosen qual limits and any filtration of data should be statistically justified.
- In this study, a system with >1.5 years of continuous historical meteorological and production data was selected for evaluating the sensitivity of degradation to filtration, using the PVFORM power model as a potential tool for eliminating noise from data sets prior to rate evaluation. The selected system had experienced both inverter downtime and significant snow events that contributed noise to the raw input data set.
- The PVUSA, PVFORM, and performance ratio models were used to normalize P_ac, after which the linear least-squares fit and classical decomposition methods were applied to extract degradation trends.
- As measured T_mod was initially unavailable, the Sandia model (with assumptions) was used to derive module and cell temperatures. Although the recently acquired measured T_mod data show a measureable difference from derived values, the difference seems to have had a relatively small effect following PVFORM-based filtration.
- The results indicated that the PVFORM power model effectively filtered noise (at least that which was caused by inverter downtime and snow events) out of the raw input data set for the selected case study.
- The PVUSA power model (no CD) appeared to be most sensitive to filtration. In all cases, sensitivities to filtration were reduced between qual limits and across power models through the use of CD, particularly for the PR model.
- The chosen system will be continuously monitored to see if and how additional data input affects the calculated degradation rates, as well as the general observations made on the sensitivity of those rates to data filtration.

12. References

1. C. Whitaker and J. Newmiller, Photovoltaic module energy rating procedure: Final subcontract report, National Renewable Energy Laboratory, Golden, CO NREL/SR-520-23942, 1998.
2. D. L. King, W. E. Boyson, J.A. Kratochvil, "Photovoltaic Array Performance Model", Sandia National Laboratories Publication #SAND2004-3535, August 2004.
3. B. Marion, "Comparison of Predictive Models for Photovoltaic Module Performance," in *33rd IEEE Photovoltaic Specialists Conference, NREL, 2008.*
4. STC: 1,000W/m² global POA irradiance, 25°C T_mod, AM1.5g spectrum according to EN 60904-3.
5. R. Dows, et al. PVUSA Procurement, Acceptance, and Rating Practices for Photovoltaic Power Plants, September 1995, PG&E R&D Report #9530910000.1.
6. PTC: 1,000W/m² global POA irradiance, 20°C T_amb, 1m/s wind speed.
7. Marion, B.; Adelstein, J.; Boyle, K.; Hayden, H.; Hammond, B.; Fletcher, T.; Canada, B.; Narang, D.; Kimber, A.; Mitchell, L.; Rich, G.; Townsend, T.; , "Performance parameters for grid-connected PV systems," *Photovoltaic Specialists Conference, 2005. Conference Record of the Thirty-first IEEE* , vol., no., pp. 1601- 1606, 3-7 Jan. 2005.
8. Dhere, N.G.; Pethe, S.A.; Kaul, A.; , "Comparative study of various PV technologies in terms of energy yield and annual degradation," *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE* , vol., no., pp. 002899-002903, 20-25 June 2010.
9. Jordan, D.C.; Kurtz, S.R.; , "Analytical improvements in PV degradation rate determination," *Photovoltaic Specialists Conference (PVSC), 2010 35th IEEE* , vol., no., pp.002688-002693, 20-25 June 2010.

Yingli Americas