Efficiency Evaluation and Comparisons of Solar Cell Technologies Based on Measurements from the Arabian Peninsula

Yannis Pantazis[1], Evangelia Kalligiannaki[1], Yiannis Kamarianakis[1], Konstantinos Kotsovos[2], Issam Gereige[2], Marwan Abdullah[2], Aqil Jamal[2], Athanasios Tzavaras[3], Theodoros Katsaounis[*1, 4]

[1]Institute of Applied and Computational Mathematics, FORTH, Heraklion (Greece)
[2]Renewable Energy, Carbon Management Division, ARAMCO, Dhahran (Saudi Arabia)
[3]Computer, Electrical and Math. Science & Engineering, KAUST, Thuwal (Saudi Arabia)
[4]Dept. of Mathematics & Applied Mathematics, University of Crete, Heraklion (Greece)

Abstract

In this work, we model the power output (a.k.a., energy yield output) of three photovoltaic (PV) cell technologies using statistical learning tools and compare their performance efficiency under real-field conditions. We introduce and train regression models to elucidate the relationship between irradiance and energy yield output. The training is performed on historical records obtained from an adverse-for-solar-panels location and a long period of time. We use standard and robust estimation approaches to compute the model's parameters. We explored various families of predictive models and found that the best-performing model includes an intraday variability factor. We then applied residual error analysis and related the models' coefficient with the efficiency of the solar cells. Our analysis showed that the efficiency decreases over time, and each PV technology has a different rate of deterioration. Moreover, we observe seasonal fluctuations in the efficiency of each PV technology which we have quantified. The decrease in efficiency during the summer months can reach up to 40% relative to the efficiency during the winter months.

Keywords: predictive modeling, performance evaluation, solar cell technologies comparison

1. Introduction

The great majority of Saudi Arabia's landscape is located inside the sun belt resulting in a significant opportunity for renewable energy from solar panels (Almasoud and Gandayh 2015). However, the development of solar plants is quite limited, with the contribution to the national energy mix being about 0.5% for 2020. Given the high potential and the future need for renewable energy, this study aims to highlight the performance differences between various PV cell technologies and elucidate their behavior over a long period of time in the distinct weather conditions of Saudi Arabia. More specifically, the location of the PV solar cells for which historical observations are collected is the New Energy Oasis (NEO) test field near the Red Sea coast (22.30 N, 39.10 E), KAUST, Thuwal, Saudi Arabia. This work presents the statistical analysis and performance evaluation of three photovoltaic (PV) cell technologies under real operating conditions and a five-year period of observations. The location (KAUST, Thuwal) is a challenging place for solar panels because they often operate at temperatures far beyond the Standard Test Conditions (STC) and well above the range observed in most common PV installation locations worldwide (Katsaounis et al. 2019). The historical data consists of irradiance and energy yield (EY) observations which are recorded every 10 minutes. The following three technologies are tested: Aluminum Back Surface Field (AlBSF), Hetero Junction (HJT), and Back Contact (BC). Data cleaning and manual outlier removal have been applied to eliminate any kind of statistical artifact.

The purpose of this paper is threefold: (a) to estimate the efficiency of a PV cell (e.g., reveal the performance over time), (b) to perform statistical comparisons between solar cell technologies, and (c) to quantify the deterioration of the solar cell performance over the years. Towards this purpose, we employ tools from regression analysis and machine learning. A single-parameter baseline regression model was used to linearly relate the EY output with the irradiance. Due to strong seasonal variability, the relationship between irradiance and EY is not stationary; therefore, the parameter estimation is performed on a monthly basis leading to a time-varying

doi:10.18086/eurosun.2022.16.08 Available at http://proceedings.ises.org

coefficient. The obtained time-varying coefficient (i.e., the slope) serves as an estimate for the time-varying efficiency of a PV solar cell. The residual error analysis for the linear irradiance model revealed a systematic intraday variability of the EY output, which has also been quantified, and the best predictive model is determined using a k-fold cross-validation scheme. The observed temporal variation can also be attributed to the heat since the temperature in the afternoon (when the sun falls from its azimuth) is higher than in the morning (when the sun rises towards its azimuth).
Furthermore, we employed two approaches for cell technology comparisons. The first approach is non-parametric and depends only on the EY output. In this direct approach, we compute the average relative EY output difference between the technologies using only the measurements with common timestamps. The second approach utilizes the baseline model between EY and irradiance, and it is parametric. We compute the slope difference only for the common months since the non-stationarity of the efficiency could affect the comparison results. Indeed, an important advantage of having a parametric model with time-varying coefficients is that we can quantify both the deterioration of performance over time and the variations due to seasonal factors. The EY differences between cell technologies were small but distinctive. The highest, on average, power output was achieved by the BC technology. Indicatively, BC was, on average, 5.5-5.8% more efficient than AlBSF and 2.1-3.1% more efficient than HJT. Interestingly, on an annual comparison, there are years that the ranking is different. The deterioration over time is also evident; we estimated a 6-8% annual drop in power output efficiency depending on the cell technology. Data also revealed that the most dramatic performance difference was observed during seasonal alteration. The relative efficiency drop in summer relative to winter is occasionally more than 40%, eliminating the gains from summer's increased irradiance. Therefore, informed decision-making for PV installation projects should take into special account the temperature coefficient of the PV cells, while mitigation measures such as cooling may be worth considering and be financially viable.

2. Approach and Methodology

The available historical data consist of irradiance and EY observations from three cell technologies (AlBSF, HJT & BC) for the 2015-2019 calendar years. There are irradiance records for 60 months, and EY records for 33 months (AlBSF), 48 months (HJT), and 53 (BC). We initially performed an exploratory data analysis and visualization. We identified a tiny percent of the records which did not follow the overall sample distributions. Both irradiance and EY measurements had few problematic values, which we considered outliers, and thus all the necessary data cleaning and preprocessing steps were applied.

2.1. Predictive modeling and parameter estimation

A strong correlation between the irradiance and the EY is present in the data. Indeed, their association, as quantified by Pearson and Spearman correlation coefficients for each architecture, is above 0.9 (see Tab. 1). Recall that the Pearson correlation coefficient measures the linear correlation between the two variables while the Spearman correlation coefficient, which is a non-parametric measure of ranking, assesses how well the two variables can be described using a monotonic function. The fact that both correlation measures take similar values implies that the relationship between irradiance and EY is predominantly linear.

Tab. 1: Correlation coefficient for each of the three cell technologies using all available measurements.

	AlBSF	**HJT**	**BC**
Pearson	0.929	0.917	0.930
Spearman	0.927	0.920	0.928

Therefore, our baseline model is a single-parameter linear regression model. Letting x_i and y_i denote the *i*-th irradiance and EY output measurement, respectively, we model their association by

$$y_i = c_1 x_i + z_i \qquad \text{(eq. 1)}$$

where c_1 is the unknown coefficient while z_i corresponds to the residual error. We will also refer to c_1 as the slope, and it provides an estimate for the efficiency of the PV solar cell.

Our preliminary analysis also showed that the relationship between irradiance and energy yield is not stationary over the five-year period. Indeed, it is well documented that there are both long-term but also seasonal changes in the efficiency of a solar cell over time (Patel 2006, Dubey et al. 2013). Therefore, it is not optimal to compare the architectures on the coarse annual resolution; a more refined temporal resolution is required. Motivated by these observations, we estimate the linear irradiance predictive model month by month. Thus, we slice the data on a monthly basis and compute the slope using two standard estimation methods: Ordinary Least Squares (OLS) and Robust Least Squares (RLS) with Huber weights (Huber and Ronchetti 2009). We employ RLS because it is expected to be less sensitive to outlier values.

Additionally, we observed an almost-periodic intraday variability for the residual error of the linear model. This intraday variability is partially attributed to temperature differences between the morning and afternoon hours. We incorporate this behavior into the model with the addition of a periodic term with a period equal to one day. One day period corresponds to $T = 78$ data points; one data point for each ten-minute interval. Therefore, the new model is defined as

$$y_i = c_0 + c_1 x_i + \sum_{k=1}^{K} \quad a_k cos\left(\frac{2\pi k t_i}{T}\right) + b_k sin\left(\frac{2\pi k t_i}{T}\right) \qquad \text{(eq. 2)}$$

where $t_i \in \{1, \ldots, T\}$ corresponds to the time variable while K is the number of sinusoidal components to be determined. We also include the intercept parameter c_0.

2.2. Comparison metrics

We deploy two approaches for comparing the three solar cell technologies. The first approach is a straightforward calculation of the EY difference between two technologies measured at the same time points. This “synchronous” comparison is required because the variability in the EY values stemming from weather or maintenance factors or simply the existence of missing data could affect the statistics of the quantity of interest. Moreover, we further reduce the noise in the data by excluding measurements with EY values below 10 Wm^{-2}. The second approach is a parametric one, which is defined via the linear irradiance model (see eq. 1), motivated by the fact that the estimated correlation coefficient between irradiance and EY is above 92%, even when all records are considered. The linear irradiance model has only one parameter, the slope, and it is directly interpretable as the efficiency of the solar cell. We then compare the estimated slopes and utilize statistical testing to assess the significance of the differences.

3. Results and Discussion

Results on predictive modeling and residual error analysis, as well as the comparative results on the studied solar cell technologies, will be presented.

3.1. Quantify predictive performance and perform model selection

We determine the optimal number of sinusoids K of (eq. 2) using a standard cross-validation scheme. Tab. 2 reports the root mean square error (RMSE) for a series of predictive models determined by the number of sinusoids, K. The estimation method is OLS. In the table, we highlight in bold the best-performing model for each year and each architecture according to the RMSE and the minimum number of parameters in the case of ties. Based on the table, the best model has intercept and $K = 2$. Except for the months of 2015, where the mean RMSE is around 30 W/m^2 for all architectures, the mean RMSE is around 10 W/m^2 for all the remaining years. Moreover, it is worth noting that the intercept reduces the mean RMSE for 2015, but it does not improve the predictive performance in any of the other years.

Tab. 2: RMSE (in W/m^2) estimated on the test set of a 10-fold cross-validation and averaged over the months of each available year.

	Year	K = 0		K = 1		K = 2		K = 3		K = 4	
Intercept		Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
AlBSF	2015	30.0±2.2	27.8±2.2	28.30±2.8	20.4±2.7	27.8±2.7	**19.2±2.2**	27.4±2.5	19.2±2.3	26.9±2.9	19.2±2.5
	2016	10.7±1.5	10.7±1.3	8.9 ± 1.7	8.9 ± 1.6	8.5 ± 1.5	**8.4 ± 1.8**	8.4 ± 1.6	8.5 ± 1.4	8.4 ± 1.6	8.4 ± 1.5

	2017	11.7±2.6	11.7±2.4	10.7 ± 2.6	10.7±3.2	10.4±2.6	10.4±2.5	10.4±2.8	**10.2±2.8**	10.4±2.9	10.4±2.8
	2018	9.8 ± 1.3	9.7 ± 1.3	8.4 ± 1.4	8.3 ± 1.6	**8.1 ± 1.6**	8.1 ± 1.5	8.1 ± 1.4	8.1 ± 1.4	8.1 ± 1.5	8.1 ± 1.6
	2019	9.6 ± 2.0	9.5 ± 1.8	8.4 ± 2.0	8.4 ± 2.1	**7.9 ± 2.5**	8.0 ± 1.9	7.9 ± 2.7	8.0 ± 2.2	7.9 ± 2.4	8.0 ± 2.2
HJT	2015	30.0±2.9	27.9±2.6	28.2 ± 2.6	19.7±2.4	27.8±2.7	**18.7±2.5**	27.6±2.6	18.7±2.4	27.0±2.3	18.7±2.1
	2016	11.4±2.0	11.3±1.9	9.4 ± 2.2	9.3 ± 2.3	9.0 ± 2.4	9.0 ± 2.5	9.0 ± 2.4	**8.9 ± 2.5**	9.0 ± 2.3	9.0 ± 2.4
	2017	10.7±3.4	10.7±3.0	9.3 ± 3.4	9.4 ± 3.2	9.2 ± 3.5	**9.1 ± 3.7**	9.4 ± 3.3	9.3 ± 3.3	9.4 ± 3.2	9.4 ± 4.1
	2018	13.9±1.6	13.8±1.9	13.7 ± 1.4	**13.5±1.6**	13.6±2.0	13.5±1.8	13.6±2.0	13.5±1.4	13.6±1.9	13.5±1.7
	2019	14.3±2.1	14.0±2.1	12.5 ± 2.5	11.3±1.8	12.4±2.2	**11.1±2.1**	12.3±2.1	11.1±2.2	12.2±2.1	11.1±2.3
BC	2015	29.8±2.5	27.2±2.3	27.9 ± 2.9	19.4±2.4	27.4±2.7	**18.2±2.4**	27.1±2.2	18.2±2.5	26.6±2.3	18.2±2.4
	2016	11.2±1.5	10.9±1.5	9.1 ± 1.7	9.1 ± 1.5	8.4 ± 1.8	8.4 ± 1.7	8.5 ± 1.6	8.4 ± 1.7	8.4 ± 1.8	8.4 ± 1.5
	2017	11.2±1.9	11.2±1.6	9.6 ± 1.9	9.6 ± 1.8	9.3 ± 1.9	**9.2 ± 1.9**	9.2 ± 2.0	9.2 ± 1.8	9.3 ± 1.8	9.2 ± 1.9
	2018	12.7±2.0	12.7±2.1	12.2 ± 2.3	12.2±2.4	12.1±2.3	12.1±2.2	12.1±2.2	12.1±2.5	12.2±2.0	12.2±2.0
	2019	11.3±1.8	11.2±1.9	9.7 ± 2.0	9.7 ± 1.9	9.6 ± 2.1	9.6 ± 2.0	9.6 ± 2.1	9.6 ± 2.0	9.6 ± 2.0	9.6 ± 2.0

Fig. 1 shows the residual error in terms of RMSE for the single parameter regression model (eq. 1), (slope only), and the regression model (eq. 2), the best model, according to Tab. 2 on a monthly basis.

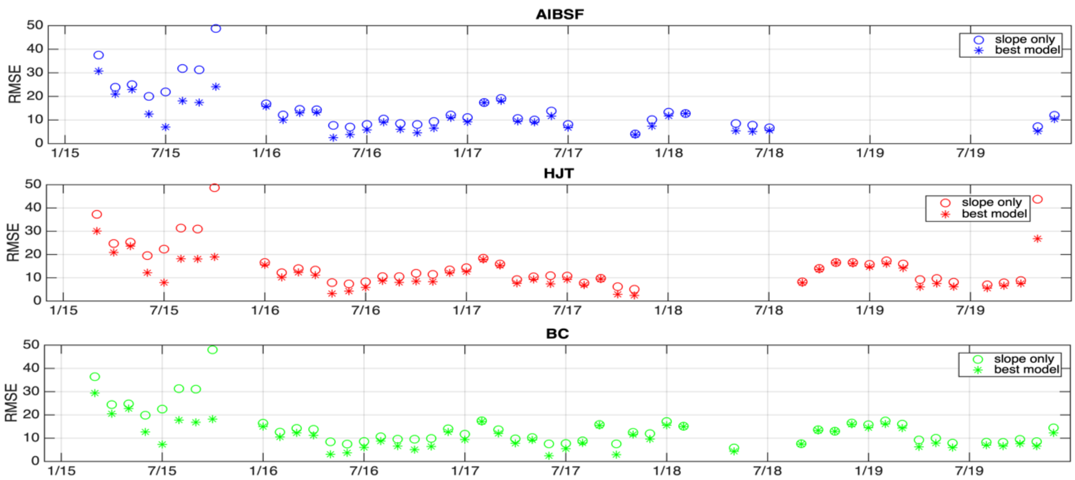

Fig. 1: RMSE for the baseline model (just one parameter, the slope) and for the best model (*K*=2 with intercept). The best model which quantifies the intraday oscillations has 6 parameters in total.

3.2. Residual error analysis

Next, we study the residual error statistics and quantify the differences between the models and the estimation approaches. More specifically, we plot in Fig. 2 the residual error for three selected months for AlBSF technology as a function of the irradiance (left column of panels) and its histogram (right column of panels). The first observation is that the residual error is quite different depending on the studied month. The months of 2015 contain large residual error values, both positive and negative, for the baseline model (blue dots). On the other hand, the best model (red dots) has only large negative residual error values. Given that OLS has zero mean residual error, the large negative residual error values imply a positive bias evident in Fig. 2 (upper panels). Using RLS, this positive bias is alleviated, as we observe in Fig. 3 (upper panels).

The months of 2016 contain intraday variability (blue dots in mid panel) which results in a bimodal distribution for the residual error of the baseline model. In contrast, the residual error's distribution of the best model is almost a normal distribution with zero mean implying that the best model predicts almost perfectly the energy yield output from the irradiance and the time variable. The modeling of the intraday variability resulted in improved predictive performance for the remaining years, however, the performance improvement is smaller as it is also evident from Fig. 3 (lower panels). Using RLS instead of OLS had a minimal effect on the residual error from 2016 and onward. We note that the same conclusions hold for the other two architectures.

As a final remark, the statistical modeling of the intraday variability does not affect the prediction of the total energy power produced during the course of the day. Thus, the baseline model will be as accurate as the best model for the coarse daily energy yield power output. However, modeling of the intraday variability is necessary when the hourly energy yield output is asked. Given that several power energy markets (e.g., various European countries) provide prices on an hourly basis, the bias due to intraday variability needs to be removed.

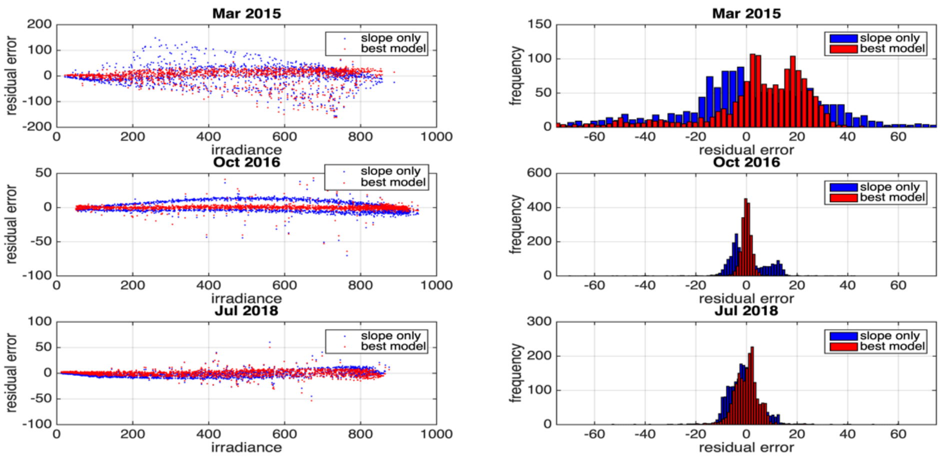

Fig. 2: Left column: the residual error as a function of the irradiance for three indicative months. OLS is the estimation method and measurements come from the AlBSF technology. Right column: the respective histograms for the residual error.

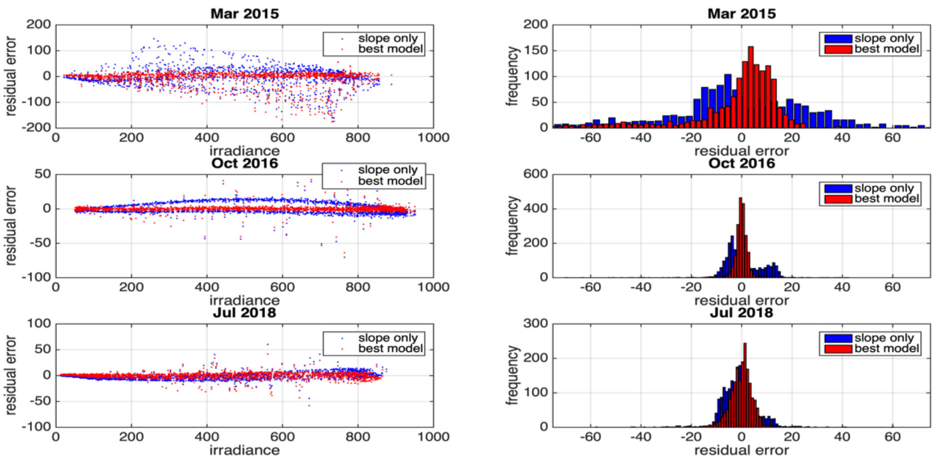

Fig. 3: Same as Fig. 2 using RLS with Huber weights as the estimation method.

3.3. Efficiency estimation

Next, we study the efficiency of each technology. The efficiency is defined as the percentage of irradiance that becomes energy power from the PV solar cell. Thus, the coefficient of the baseline model is a direct estimate of the efficiency. Fig. 4 depicts the estimated slope for every available month and for the three studied technologies using both OLS (circles) and RLS (stars). The differences are small and statistically insignificant (p-value between 0.4 and 0.6) therefore both can be used as reliable estimates for the efficiency factor.

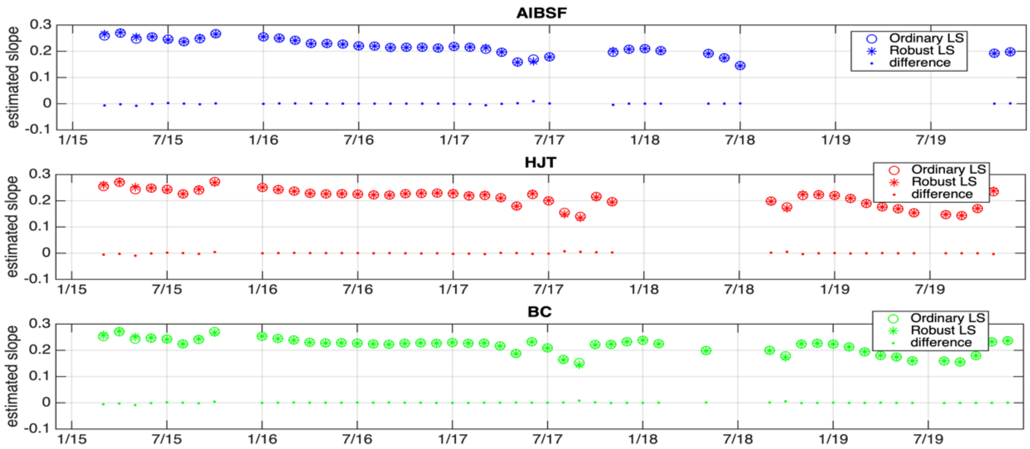

Fig. 4: Estimated slopes per month for the three technologies. The comparison between OLS and RLS reveals that there is no statistically significant difference in the estimation of the slope parameter.

Interestingly, the slope comparison between the baseline model and the best model (Fig. 5) showed that for 2015 the best model is significantly different than the baseline model. An ablation analysis revealed that the intercept is the parameter that makes the slope estimates so different between the two models. Therefore, in the subsequent section, the indirect (i.e., parametric) comparisons between the three architectures will be performed using the baseline model with the slope being estimated with OLS.

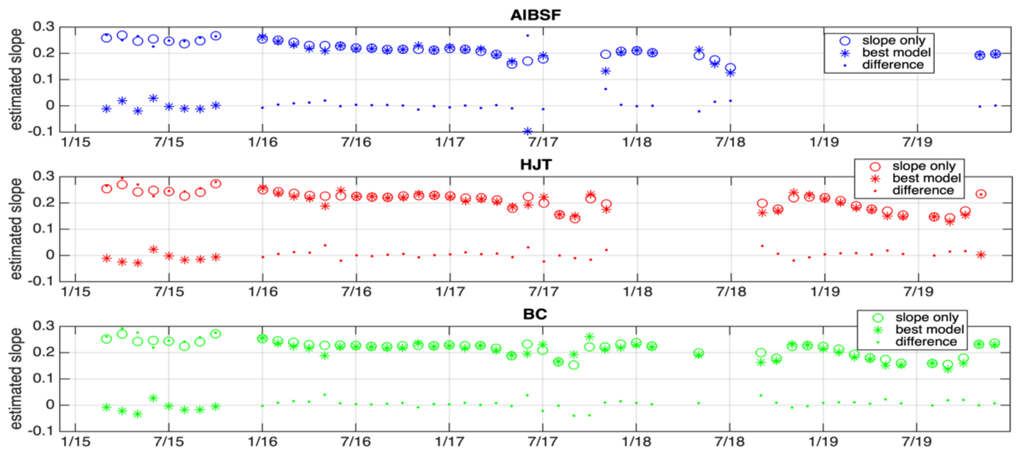

Fig. 5: Estimated slopes per month for the three technologies. The comparison is between the baseline model and the best model. A significant difference in the estimation of the slope is evident for 2015 implying that the best model's slope loses its physical interpretation.

3.4. Solar cell technologies comparisons

The upper plot of Fig. 6 shows the estimated slopes for the three cell technologies. The time-varying behavior of the slope, hence, of the solar cell efficiency, is evident. The seasonal variations of the estimated slope are rather striking (see also the lower plot of Fig. 6). The difference in the performance, as quantified by the slope difference, can be as large as 50%. We anticipate that those seasonal variations mainly stem from the differences in the air surface temperature (Patel 2006, Dubey et. al. 2013). Moreover, there is an annual deterioration of performance in all cell technologies (see thick dashed lines in Fig. 6). Using a linear regression model with two parameters (decay and intercept) and OLS we quantified the rate of deterioration for each technology. For BC, the average rate of deterioration was 5.8%, for HJT the rate was 6.9% while for AlBSF the average rate of deterioration was 7.5%.

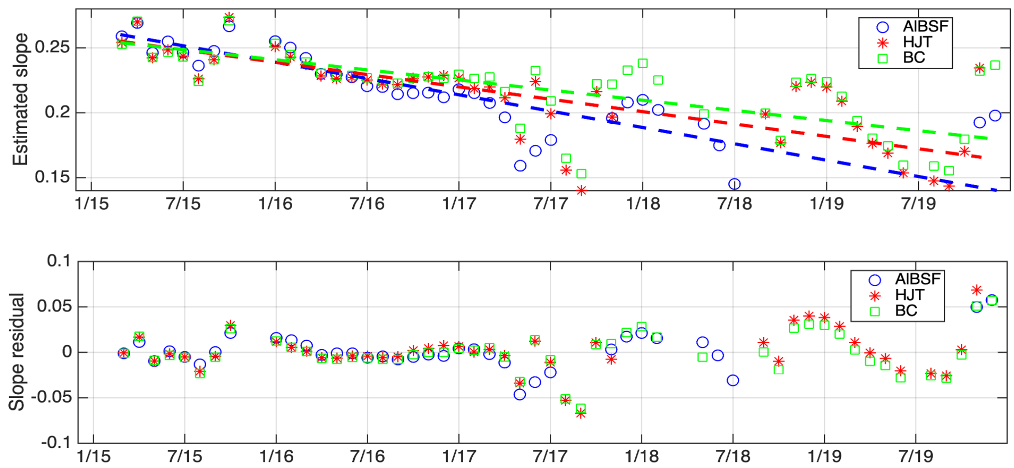

Fig. 6: Upper plot: Estimated slope on a monthly basis for the three cell technologies. The slope is a direct estimator of cell efficiency. Thick dashed lines correspond to the linear trend of the slopes for each cell type and reveal the drop of efficiency over time. Lower plot: The residual error of the slope after removing the linear trend. A seasonal variability is evident.

Tab. 3 reports the relative performance gain (or loss) between the three solar cell technology pairs averaged per each year and the total average. The relative performance results for both non-parametric (NP) and parametric (P) approaches are given. We stress once again that for a fair comparison the differences are evaluated at points for which data for both technologies are available.

Clearly, both approaches report qualitatively similar results. However, the parametric indirect approach is more conservative relative to the direct non-parametric approach. One possible explanation for this difference is that the parametric approach puts more weight on measurements with high irradiance value while the non-parametric approach treats all measurements equally. Given that more power output is produced when irradiance is higher, it is expected that the parametric approach provides more consistent comparisons. Overall, and despite being indecisive for 2015, it is evident for the subsequent calendar years that BC technology outperforms the other two.

Tab. 3: Relative performance difference between the three cell technologies. Positive values imply that the first cell technology is better. NP refers to the non-parametric approach while P to the parametric one.

	2015		**2016**		**2017**		**2018**		**2019**		**Average**		**Annual Average**	
	NP	**P**	**NP**	**P**	**NP**	**P**	**NP**	**P**	**NP**	**P**	**NP**	**P**	**NP**	**P**
BC vs AlBSF	0.5	-1.8	4.4	1.9	1 5.6	12.8	13.2	9.0	18.8	18.3	5.8	5.5	10.5	8.0
BC vs HJT	1.5	-0.4	2.4	0.6	5.4	4.8	3.0	1.0	4.8	3.5	3.1	2.1	3.4	1.9
HTJ vs AlBSF	-1.0	-1.4	2.1	1.3	12.4	8.6	-	-	-	-	2.5	3.2	4.5	2.8

4. Conclusions

In this study, we estimated and then analyzed the efficiency of three solar cell technologies over a five-year span from observations collected at KAUST, Thuwal, Saudi Arabia. The weather conditions of this particular location and specifically the high temperature have a quite adverse effect on solar panels operation. We first observed that efficiency varies over time; therefore, we estimated it on a monthly basis with regression. Our analysis of the time-varying efficiency revealed that solar cell performance deteriorates over the years with a rate between 5% and 8% per year. The observed annual deterioration rate is sizably higher than the rate provided by the manufacturers of the PV cell technologies. The time-varying efficiency also revealed that large performance differences during seasonal alteration are evident, which may reach up to more than 40% in some years (efficiency in winter is higher than in summer). Finally, we performed statistical comparisons between the three solar cell technologies. Our analysis showed that newer technologies (i.e., BC and HJT) are more efficient than older technologies (i.e., AlBSF) mainly due to a lower deterioration rate over time.

5. Acknowledgments

The authors acknowledge the support of KAUST and Saudi Aramco R&D Center - Carbon Management Division for their financial support in developing this work. This work was partially supported by grant *#OSR–2020-4433.02* from KAUST and Saudi Aramco.

6. References

Almasoud, A.H, Gandayh, H.M, 2015, Future of solar energy in Saudi Arabia, J of King Saud University - Engineering Sciences 27, 153-157

Dubey, S., Sarvaiya, J.N, Seshadri, B., 2013, Temperature Dependent Photovoltaic (PV) Efficiency and Its Effect on PV Production in the World – A Review, Energy Procedia 33, 311-321

Huber, P.J, Ronchetti, E., 2009, Robust Statistics; 2nd Ed., John Wiley & Sons New Jersey, USA

Katsaounis, T., Kotsovos, K., Gereige, I., Basaheeh, A., Abdullah, M., Khayat, A.,Tzavaras, A. E. (2019). Performance assessment of bifacial c-Si PV modules through device simulations and outdoor measurements. Renewable Energy, 143, 1285–1298. doi:10.1016/j.renene.2019.05.057

Patel, M.R., 2006, Wind and Solar Power Systems, 2nd Ed., CRC Press, Florida, USA