UNIVERSIDAD
NACIONAL
DE COLOMBIA
S E D E M E D E L L Í N
FACULTAD DE MINAS

Risk analysis using meteorological weather factors in solar energy conversion systems

David Cardoza [a], Jorge M. Uribe [b] & Jairo Palacios [a]

[a] Department of Electrical Engineering, Universidad del Valle, Cali, Colombia. david.cardoza@correounivalle.edu.co,
jairo.palacios@correounivalle.edu.co
[b] Department of Economics, Universidad del Valle, Cali, Colombia. jorge.uribe@correounivalle.edu.co

Abstract
We propose a methodology for conducting simulations of operational scenarios for energy projects based on photovoltaic generation systems. It considers several documented facts about time series of weather, such as strong seasonality and structural breaks, which the previous literature has not explored in depth. Our proposal uses public weather time series, which are usually recorded by meteorological observatories. This makes our approach a suitable strategy for any firm interested in applying it to its own data and projects. This approach also allows the user to set an appropriate level of confidence for the scenarios depending on her interests.

Keywords: photovoltaic systems; horizontal solar radiation; structural breakpoints; renewable energy sources; risk measurement.

Análisis de riesgo usando factores meteorológicos en sistemas de conversión de energía de solar

Resumen
Se propone una metodología para realizar simulaciones de escenarios de operación de sistemas solares fotovoltaicos. Esta aproximación tiene en cuenta características documentadas de las variables climáticas, como su estacionalidad y la presencia de quiebres estructurales, las cuales no han sido exploradas anteriormente con profundidad. La propuesta usa series de tiempo de datos de diferentes variables climáticas, las cuales son normalmente recolectadas por estaciones meteorológicas; esto le da utilidad a nuestra propuesta, ya que cualquier firma interesada pueda aplicar nuestro planteamiento a sus propios datos y proyectos. Este enfoque también les permite a los usuarios establecer un nivel de confianza adecuado para los escenarios de operación en los cuales se tenga interés.

Palabras clave: sistemas fotovoltaicos; radiación solar horizontal; quiebres estructurales; energías renovables; evaluación del riesgo.

1. Introduction

Forecasting and construction of operational scenarios is a key component of the appropriate functioning of any firm in the energy sector; it is crucial to project management and thus to decisions about the operational and financial viability of a given project. This is also true for companies with a business unit focused on exploiting renewable energy sources, such as photovoltaic energy conversion systems. Indeed, a correct assessment of operational risks, in terms of the energy that a firm could expect to offer to the market in a given day, is a crucial input when constructing cash flow scenarios and

estimating returns on assets, liquidity ratios, present values, and other financial indicators. One can even make the argument that the execution of an entire energy conversion project depends on the accuracy of this exercise, and hence, it warrants considerable research efforts.

For the above reasons, the literature has explored several alternatives to forecasting the level of radiation a user can expect to be recovered by a photovoltaic system in a given period. These alternatives generally take into account the crucial relationship that exists between weather conditions and collected radiation, including the stochastic nature of the weather variables. For instance, [1] aim at forecasting

radiation using *k*-th neighbor and interpolation methods, level of persistency, artificial neural networks, support vector regressions and time series analysis. These authors focus on the explanation of horizontal global radiation and use measures of radiation in a given location, alongside measures of weather variables in sparse locations around the radiation recording devices. [2] estimate an artificial neural network by genetic algorithms, aiming at forecasting daily radiation. They consider seven variables as inputs when training their network: average air temperature, minimum and maximum temperatures, relative humidity, atmospheric pressure, wind velocity and ground temperature. Using similar variables, [3] report that between three and four meteorological variables suffice to explain radiation, at least at the highest frequencies of their analysis. These authors base their conclusions on calculations of adjusted- R squared statistics, square residuals and forecasting errors. In contrast [4], employing similar methods and variables, finds that the best model includes all the available variables (seven in his case).

[5] estimate Multivariate GARCH models (time series models with autoregressive patterns in the conditional variance process), which include climatological variables such as solar brightness, precipitation, average minimum and maximum air temperatures, relative humidity and saturation deficit. [6] emphasize the importance of the range between the maximum and minimum temperatures when forecasting radiation, and they also highlight the changing nature of the relationship between radiation and weather variables, very likely driven by the strong seasonal patterns that characterize the latter.

In this study we do not focus on the point-forecast of radiation in a given period, instead we emphasize the construction of operational scenarios for an energy project in terms of the radiation that a photovoltaic unit can expect to collect, under certain weather conditions, in a given period, with an appropriate level of confidence. This enriches the tools available for energy project administrators and allows the unit in charge to make more accurate decisions, based on richer information, about a project's realization or continuation.

Our methodology aims at providing thresholds of minimum radiation that one may expect to observe given certain weather parameters and for a desired confidence level. That is, we focus on the estimation of a specific quantile of the distribution of radiation conditional on weather variables, which increases the accuracy of our estimates. We explore several weather indicators in our empirical exercise and work with daily averages of those, as has been done previously (for instance in [7]). Given the impaired evidence regarding the number of variables required to explain radiation [3,4], we performed an automatic search algorithm to determine how many and which variables should be included in an optimal model. Importantly, our methodology takes into account the dynamic relationship that exists between weather and radiation. The changing nature of the relation between weather and radiation has been mainly ignored so far. That is, strong evolving weather conditions may change the way
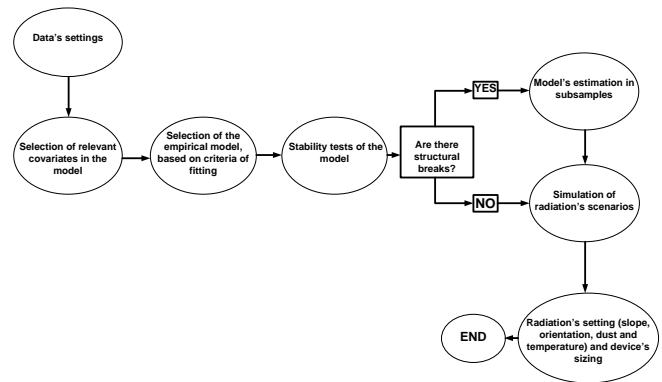

Figure 1. Graphical summary of the proposed methodology.
Source: The authors.

in which radiation interacts with temperature, wind, precipitation and other variables. Therefore, any model that is estimated under the assumption of a stable relationship among these variables may display spurious patterns, which may be avoided only by explicitly considering this possibility. Endogenous structural tests in time series analysis, such as the ones employed herein, are specially designed to address this challenge.

In what follows we present the proposed methodology, and we illustrate it using data from a company that manages photovoltaic energy conversion systems, located in Colombia. We present our empirical results in section 3, and we discuss the results and conclude in sections 4 and 5, respectively.

2. Materials and methods

Our methodology focuses on the problem of estimating the expected global radiation on a horizontal surface, given a set of weather-related covariates. The proposal consists of five steps: i) selection of relevant variables to be included in the model, ii) selection of a parsimonious empirical model, iii) stability tests of the model's parameters, iv) simulation of operation-scenarios, conditional on weather parameters and v) generation of the output variable to be included in traditional project management designs (see Fig. 1).

2.1. Selection of relevant covariates in the model

We assume a linear relationship[1] between the total amount of radiation y_t, and a set of weather-related covariates $x_{i,t}$ i=1,2,…k, . In our empirical implementation, we imposed a maximum possible number of covariates $k^{max}=7$, following data availability from regular records of a meteorological observatory. Working explanatory variables were: temperature, humidity, dew point, wind velocity, precipitation level, atmospheric pressure, and an indicator of cloudiness (the highest minus the lowest temperature within a day). In all cases, except for the last one, the recorded frequency of each variable was half an hour, and therefore we needed to transform the original series to daily averages to obtain standardized frequencies among all the covariates and

---

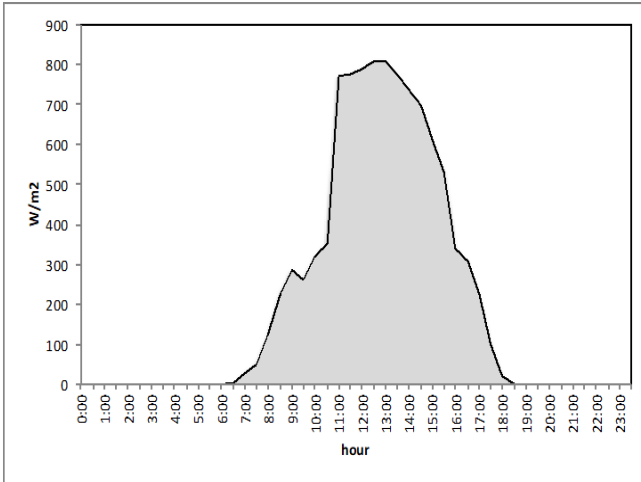[1] Perhaps non-constant in time, as will be explained in step 2.3.

Figure 2. Radiation recorded on January 1st 2015
Source: The authors.
Note: the Figure shows the level of radiation collected by a solar panel during each half- hour on a given day. The area below the curve corresponds to the total level of radiation collected by the solar panel. The data shown in the Figure was selected just for illustration purposes.

the dependent variable[2]. This dependent variable was constructed as the cumulated level of radiation collected by the panel during the course of a day. We had measures of the level of radiation collected at half hourly intervals. These measures were aggregated by integrating out the area below the curve in Fig. 2 for each day within the sample.

In this case, the population model that describes the relation between the collected radiation and the weather regressors can be written as follows:

$$y_t = \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t, \quad (1)$$

In practice, we do not know coefficients β_i, nor do we know the value of k. That is, we must estimate the marginal effects of each covariate on the total amount of collected radiation, and, at the same time, we must decide how many and which variables may appear in eq. (1). Once this is done we end up with a sample-counterpart of eq. (1):

$$y_t = \hat{\beta}_1 x_{1,t} + \hat{\beta}_2 x_{2,t} + \cdots + \hat{\beta}_{\hat{k}} x_{\hat{k},t} + e_t, \quad (2)$$

In eq.(2), marks over each of the coefficients indicate that those coefficients are related to the sample, instead of the population, and therefore they can be obtained from the data.

In the process of deciding which variables to include in eq. (2), we consider that redundant variables will conduce to inefficient estimations of the model's parameters, while omitted (and relevant) variables will likely conduce to biased and non-consistent estimations. Hence, lacking a generally accepted theory regarding the shape and composition of eq.

(2) led us to undertake a data-driven model selection process using ordinary least squares (OLS) and subsets of regressions. We conduced our exhaustive search for relevant covariates among subsets of variables typically recorded by meteorological observatories, many of which have been previously explored in the empirical literature.

Formally, given a maximum number of variables recorded by the meteorological observatory, k^{max}, we estimated different models fitted on subsets of variables, each of cardinality $\bar{k} \leq k^{max}$. For each value of $\bar{k}$, which must be chosen a *priori* by the researcher, it was necessary to evaluate all possible combinations of covariates. Then, for each $\bar{k}$, we selected the model that minimized the sum of squared residuals, $SSR = \sum_{t=1}^{T} e_t^2$ [3]. This strategy is equivalent to selecting the subset of variables for a given $\bar{k}$, that maximize the coefficient of determination, R^2, of the linear projection of y_t onto $x_{1,t}, x_{2,t}, ...,$ defined as $R^2 = 1 - SSR/STS$, where $STS = \sum_{t=1}^{T} y_t^2$. There are several alternative approaches that may be used to accomplish this task. One that is widely accepted in the statistical literature, which, moreover, is available in the statistical software R, consists of conducing an exhaustive (forward or backward) search for the best subsets as explained in [8]. In practical terms, this can be done using the function 'regsubsets()', from the package 'leaps'. In our implementation, we set $\bar{k} = 1 ... 7$ because our maximum number of variables was constrained by the number of variables available from the original observatory data.

2.2. Selection of the empirical model, based on criteria of fitting and parsimony

Notice that the preceding step gives us the best linear model for each value $\bar{k}$. Nevertheless, in practice, the researcher also needs to estimate the optimum value of $\bar{k}$. That is, we must select the best model among those available, taking into account model's parsimony considerations, by penalizing the inclusion of redundant variables in the estimations. That is, if a variable is to be included in a model, it must add to the model in terms of explanatory power more than it takes from it in terms of efficiency. To this end, we used two different criteria: Akaike's Information Criterion (AIC) and a Bayesian Information Criterion (BIC)[9], defined as:

$$AIC = 2\bar{k} - ln(L), \quad (3)$$

$$BIC = ln(T)\bar{k} - ln(L), \quad (4)$$

where L is the likelihood of the model, $ln()$ stands for the natural logarithm, and T is the number of time series observations in the sample. The best model minimizes either of these two criteria.

---

[2] In our empirical implementation, other alternatives to averages were explored. Specifically, we conducted some preliminary exercises using principal components instead of averages, with no significant gains in terms of model fitting.

[3] Since our algorithm returns a best model of each $\bar{k}$ the results do not depend on a penalty model for model size such as AIC or BIC.

2.3. Stability tests of the model's estimates due to weather conditions

With the best model determined, after steps 1 and 2 we test for parameters' stability. In the case of photovoltaic energy conversion systems, this is a crucial step of the methodology because weather-related variables are known to exhibit strong seasonality patterns, which in principle may have a significant effect on the relationship between the covariates in the model and the collected radiation during the year. This is even more important when modeling weather variables for energy companies located in countries characterized by higher latitudes and markedly contrasting seasons.

To this end, we used endogenous structural break tests following [10,11]. These procedures offer a reliable strategy to identify changes in the relationship between the covariates and the explained variable. The multiple-breaks statistic is intuitively a set of Chow's statistics [12], calculated using recursive regressions over subsamples of increasing lengths. Several candidates for breaks are selected using the biggest F-statistics for which the null hypothesis in the Chow tests (i.e., parameters stability) is rejected. Then, asymptotic (corrected) critical values are used to contrast the null of no-breaks.

If we define $\hat{\theta} = \{\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_{\hat{k}}\}$ and $X_t = \{x_{1,t}, x_{2,t} \dots x_{\hat{k},t}\}$ eq.(2) can be rewritten as:

$$y_t = X_t'\theta_j + e_t, \tag{5}$$

where $j = 1, \dots, m + 1$ represent a given model regime. That is, if we have m structural breaks in the model's parameters, it allows us to estimate $m + 1$ regimes (or 'seasons'). To identify and estimate the model in eq. (5), it is necessary first to estimate the number and location of breaks. The general idea underlying Bai and Perron's [10] algorithm consists of the following steps:

i. First, we estimate a model through recursive regressions, increasing the sample from $h * T$ to $T * (1 - h)$, where $h \in (0,1)$. The sample is divided into two sub-samples, in the period where R^2 of the global regression is maximized. We set $h = 0.05$.

ii. Once this break point is determined, the same searching procedure is applied to each of the two sub-samples. The process continues to the point where either it reaches a maximum number of breaks determined a priori ($m^{max} = 5$ in our case) or to the point at which new break candidates are no longer statistically significant.

iii. To conduct hypothesis testing about the significance of each break point, critical values provided by Andrews [13] and Hanssen [14] must be used to avoid the identification of spurious break estimates.

Once the numbers and locations of the endogenous breaks are known, it is possible to estimate eq.(6):

$$y_t = X_t'\hat{\theta}_j + e_t. \tag{6}$$

Notice that the model has different coefficients for each regime or season. Such coefficients are determined endogenously, provided $j > 1$. Estimations like those described above are performed using the package 'strucchange' in the statistical software R.

2.4. Simulation of scenarios

The simulation exercises in our methodology take into account, by construction, the most relevant variables explaining radiation, such as structural changes associated with different weather regimes. Eq.(6) for each regime can be rewritten as:

$$y_t = \hat{y}_t + e_t, \tag{7}$$

where $\hat{y}_t = X_t'\hat{\theta}$ for a given j, which has been supressed to simplify notation. The sources of uncertainty in this case are $\hat{y}_t$ and e_t. In the former case, we may observe different arrangements of the meteorological variables relevant to the model, which may conduce to distinct scenarios of operation. In the latter case, uncertainty may arise from other variables not included in the model, which may potentially affect the operation. To take into account both sources of uncertainty, we generated different scenarios of operation S, conducting re-sampling with replacement in the series of $\hat{y}_t$ and in the series of e_t. Then, we aggregated the two generated series into a single series to formulate comprehensive scenarios of operation for y. This procedure is justified as long as $\hat{y}_t$ is orthogonal to e_t, which holds by construction in our case.

The final idea consists of estimating some order statistics of the simulated scenarios for y_i, $\{y_i\}_{i=1}^S$. These are percentiles of interest for the administrator in charge of the project decisions. For example, percentiles $\alpha \in (0,1)$ close to 0 ($\alpha = 0.01, 0.05, 0.10$) may be interpreted as conservative scenarios for decision making.

2.5. Data

We use data from February 4th 2014 to October 26th 2015. The variable of main interest is radiation, which was constructed as explained in section 2.1 and in Fig. 2. As explanatory variables we used average daily temperature, humidity, dew-point temperature, wind velocity, precipitation level, atmospheric pressure, and the difference between maximum and minimum temperatures each day. All variables were collected at half-hour intervals, and all but the last one were averaged to work with a daily version of each. All data were taken from a meteorological observatory located in Colombia, latitude 3.4383 and longitude -76.5161. After transforming each variable, according to our requirements we ended up with a total of 580 days of observations. Fig. 3 depicts our variables, and in Table 1 we show the summary statistics of our sample. Fig. 3 also highlights two regimes in the sample, which were detected using structural breaks tests as explained above.
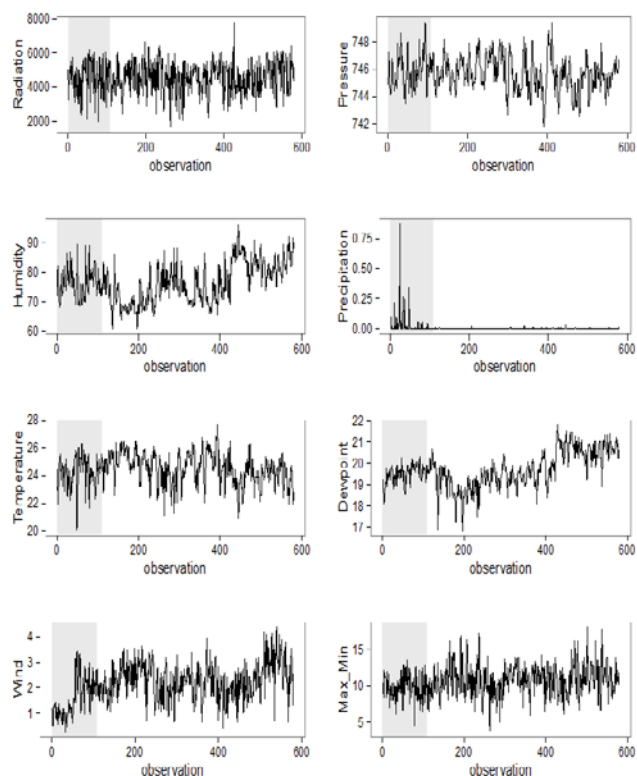
Figure 2. Radiation and Explanatory Variables Used in The Analysis.
Source: The authors.
Note: The Figure shows the daily averages of the variables employed in our estimations. Radiation is measured in kWh/m2, temperature is measured in °C, humidity is measured in %, dew-point temperature is measured in °C, wind velocity is measured in m/s, precipitation is measured in mm, atmospheric pressure is measured in Bar. Max-Min corresponds to the difference between the highest and the lowest temperatures registered each day. The shadowed area corresponds to the first regime in our sample, detected by using structural breaks tests.

Table 1.
Summary statistics of the variables used in the estimations.

	Radiation	Temperature	Humidity	Dew Point
Min.	1646	20.03	60.81	16.86
1st Quartile	3938	23.78	71.21	19.17
Median	4569	24.51	76.30	19.67
Mean	4559	24.44	76.49	19.70
3rd Quartile	5274	25.24	81.40	20.24
Max.	7812	27.64	96.54	21.81

	Wind Velocity	Precipitation	Atmospheric Pressure	Max-Min Temperature
Min.	0.275	0.0000000	741.7	3.667
1st Quartile	1.621	0.0000000	744.8	9.278
Median	2.229	0.0000000	745.7	10.667
Mean	2.167	0.0066074	745.7	10.703
3rd Quartile	2.688	0.0002083	746.4	12.111
Max.	4.396	0.8750000	749.5	18.167

Source: The authors.
Note: Minimum, 1st and 3rd quartiles, median, mean and maximum for each variable.

Table 2.
Selected variables given k, from k=2 to k=7.

k	Temp.	Dew Point	Humidity	Wind Velocity	Precip.	Atmosp. Pressure	Max-Min Temp.
1							*
2	*						*
3	*			*			*
4	*			*		*	*
5	*	*		*		*	*
6	*	*		*	*	*	*
7	*	*	*	*	*	*	*

Source: The authors.
Note: An asterisk indicates that a given variable is included in the model with the number of variables indicated in the first column. The variables were selected using exhaustive (forward or backward) searches for the best subsets.

Before conducting our estimations, we had to stabilize the mean of the atmospheric pressure variable using a dummy variable regression. We detected two changes in the scale of the series on March 31st 2014 and on October 5th 2015. These changes were unrelated to weather, but they were caused by settings in the measurement device.

2.6. Target

The procedure described above aims at sizing a photovoltaic system in an accurate fashion. We seek a suitable radiation value that could be used for sizing solar photovoltaic and photo-thermal panels, batteries, battery chargers and inverters. Applying the proposed methodology, we provide an option for performing sizing of these devices, which uses statistical tools for the selection of the level of radiation to be employed.

3. Results and discussion

Here we present the main results of our empirical implementation. First, we set the relevant variables to be included in our model. Given an optimal number of variables, the selected combinations are presented in Table 2. As can be seen, the difference between maximum and minimum temperatures appears in all selected models, using one to seven variables. The second most recurring variable is average temperature and the third is wind velocity. The least relevant variables, in terms of their significance in the models, are humidity and level of precipitation.

We estimated six models using the indications provided in Table 2. The models included between two and seven variables (we omitted the model with only one variable). Following the AIC criterion, the optimal model should include five variables, namely, average temperature, dew point temperature, wind velocity, atmospheric pressure and the differences between the max and min temperatures. According to BIC criteria, the optimal model should include only four variables, namely the same as those listed above except for dew-point temperature. In the interest of parsimony we followed BIC in the final specification (Table 3).

Table 3.
Criteria for model selection.

k	2	3	4	5
AIC	8,987.542	8,959.928	8,952.471	8,951.929
BIC	9,004.994	8,981.743	8,978.649	8,982.471
k	6		7	
AIC	8,953.383		8,955.363	
BIC	8,988.288		8,994.630	

Source: The authors.
Note: Two criteria were used to determine the optimal number of variables in the model, the Bayesian Information Criterion (BIC) and the Akaike's Information Criterion (AIC). The best models are those with lower values and are highlighted using bold type in each case.

Having selected the best model, we test for the stability of its parameters. We detected one statistically significant structural break in the sample that minimizes the BIC criterion (8914.26) when including partitions. The break is located on May 25th 2014 (day 110) with an associated 95% confidence interval between May 16th 2014 and June 1st 2014. Thus, we documented two different regimes in our sample, the first one from February 4th 2014 to May 25th of the same year, and the second one from May 26th until the end of the sample (see Fig. 1).

We present the estimated coefficients corresponding to the two detected regimes in Table 4. In the first regime, only two variables - temperature and the highest minus the lowest temperatures in a day - were statistically significant. Both of them exhibited positive signs in the regressions, meaning that an increase of one degree in the average temperature recorded in a given day increases the radiation collected by 342.44 kWh/m2. In the same line, an increase in the difference between the max and the min temperatures (i.e., indicating a less cloudy day) increases the supply of radiation by 364.68 kWh/m2, all other variables remaining constant. Although only two variables were significantly different from zero during this regime, the model is highly significant on a joint basis, as shown by an associated F statistic of 85.81, which is also significant at a 99% confidence level. The coefficient of determination, which is usually interpreted as a measure of a model's goodness of fit, is relatively high, indicating that the estimated model accounted for 76.57% of the total variation in the collected radiation series.

In the second regime, all the variables were statistically significant. Although, the two variables (temperature and max-min) continued to explain radiation in a significant fashion, their average effect on radiation was considerably less pronounced compared to the first regime. That is, with an increase of one degree in the average temperature, collected radiation increases by 139.15 kWh/m2 during the second regime (in contrast with 342.44 w/m2 in the first regime). The same holds true for the level of cloudiness. This means that during the second period of the sample, other variables, such as wind velocity and atmospheric pressure, were also highly relevant to explain the total level of radiation collected by a panel. Both variables have positive signs. These changes in the model were only possible to detect by testing the model's stability by means of structural break statistics. This is a key point when modeling energy generation devices that depend on weather variables,

especially when modeling renewable energies, which are characterized by strong seasonality patterns and are, therefore, very sensitive to structural instabilities associated with weather.

In Table 5 we report the expected operational scenarios at different levels of confidence from 5% to 90%. This information is valuable for project managers who seek to determine how much of their energy supply can be compromised for future delivery depending on the weather conditions. For example, during regime 2, the 5th percentile of the generated scenarios corresponds to 3,058.46 kWh/m2, this means that on only 5% of the days during a year (approximately 18 days) will the collected level of radiation be lower than 3,058.46 kWh/m2. Analogously, the 90th percentile is 5,685.84 kWh/m2, which tells us that on only 10% of the days will the collected radiation be higher than this level. Those scenarios can be interpreted as corresponding to 'optimistic' or 'pessimistic' expectations. The lowest quantiles are associated with very conservative scenarios of operation, and they can be used as forecasts of the worst scenarios a firm might face, during a particular weather regime, in terms of energy supply.

Notice that the percentiles vary considerably from one regime to another. Specifically, they are almost always higher for the second regime than for the first regime. This seems quite natural because, during the first period of the sample, the observatory recorded more precipitation, lower dew-point temperatures, fewer differences between the highest and the lowest temperatures in a day (cloudier days) and a lower wind velocity (see Fig. 1). Nevertheless, when

Table 4.
Coefficient estimates that relate each explanatory variable to radiation in each regime of the sample.

First Regime		
Dependent Variable: Radiation		
Degrees of Freedom 105		
	Estimate	Robust S.E.
Intercept	-64684.2	43926.6
Temperature	342.4***	84.4
Wind Velocity	32.7	69.5
Max-Min Temp.	364.7***	31.2
Atmospheric Pre.	76.8	56.9
R-squared	0.77	F-statistic 85.8***
Adjusted R-squared	0.76	(4 and 105 DF)
Second Regime		
Degrees of Freedom 465		
	Estimate	Robust S.E.
Intercept	-68581.1***	16099.6
Temperature	139.2***	28.4
Wind Velocity	338.4***	37.9
Max-Min Temp.	274.0***	13.9
Atmospheric Pre.	88.5	21.2
R-squared	0.70	F-statistic 272.9***
Adjusted R-squared	0.70	(4 and 465 DF)

Source: The authors.
Note: * significant at 90% of confidence, ** significant at 95% of confidence, *** significant at 99% of confidence.

Table 5.
Operation's Scenarios

	5%	10%	15%
Regime 1	2969.396	3342.907	3579.546
Regime 2	3058.455	3401.363	3627.674
	20%	50%	90%
Regime 1	3768.835	4570.654	5774.753
Regime 2	3808.448	4572.227	5685.841

Source: The authors.
Note: The Table shows how much radiation we can expect the panel to collect, on a given day, during each regime in the sample, under certain levels of confidence. For example, in the second regime, we can expect that the collected level of radiation will surpass 5,685.841 w/m2 in a day only 10% of the time, while we can expect it to be lower than 3,401.363 w/m2 only 10% of the time.

we focus on the 90th percentile during the first regime, the average level of collected radiation was lower compared to the second regime, but it was also more volatile. Therefore, we can conclude that the first regime of operation is also more risky in terms of the level of radiation that one may expect to receive in photovoltaic panels.

5. Conclusions

We propose a methodology for conducting simulations of operational scenarios for energy projects based on photovoltaic generation systems. It considers several documented facts about time series of weather, such as strong seasonality and structural breaks. Our proposal uses public weather time series, which are usually recorded by meteorological observatories. This makes our approach a suitable strategy for any firm interested in applying it to its own data and projects. We illustrate our methodology using data for a Colombian energy firm and we document that the difference between maximum and minimum temperatures appears in all selected models, using one to seven variables. The second most recurring variable is average temperature and the third is wind velocity. The least relevant variables, in terms of their significance in the models, are humidity and level of precipitation.

Bibliography

[1] Lazzaroni, M., Ferrari, S., Piuri, V., Salman, A., Cristaldi, L. and Faifer, M., Models for solar radiation prediction based on different measurement sites, Measurement, 63, pp. 346-363, 2015. DOI: 10.1016/j.measurement.2014.11.037

[2] Mousavi, S.M., Mostafavi, E.S., Jaafari, A., Jaafari, A. and Hosseinpour, F., Using measured daily meteorological parameters to predict daily solar radiation, Measurement, 76, pp. 148-155, 2015. DOI: 10.1016/j.measurement.2015.08.004

[3] Ododo, J. and Usman, A., Correlation of total solar radiation with common meteorological parameters for Yola and Calabar, Nigeria, Energy Conversion and Management, 37(5), pp. 521-530, 1996. DOI: 10.1016/0196-8904(95)00217-0

[4] Skeiker, K., Correlation of global solar radiation with common geographical and meteorological parameters for Damascus province, Syria, Energy Conversion and Management, 47(4), pp. 331-345, 2006. DOI: 10.1016/j.enconman.2005.04.012

[5] Sun, H., Zhao, N. Zeng, X. and Yan, D., Study of solar radiation prediction and modeling of relationships between solar radiation and meteorological variables, Energy Conversion and Management,105, pp. 880-890, 2015. DOI: 10.1016/j.enconman.2015.08.045

[6] Shamshirband, S., Mohammadi, K., Chen, H., Narayana-Samy, G., Petković, D. and Ma, C., Daily global solar radiation prediction from air temperatures using kernel extreme learning machine: A case study for Iran, Journal of Atmospheric and Solar-Terrestrial Physics, 134, pp. 109-117, 2015. DOI: 10.1016/j.jastp.2015.09.014

[7] Ouali, K. and Alkama, R., A new model of global solar radiation based on meteorological data in Bejaia City (Algeria), Energy Procedia, 50, pp. 670-676, 2014. DOI: 10.1016/j.egypro.2014.06.082

[8] Wooldridge, J.M., Introductory econometrics a modern approach, 5th ed. Michigan: Cengage Learning, 2013.

[9] Bai, J. and Perron, P., Computation and analysis of multiple structural change models, Journal of Applied Econometrics, 18(1), pp. 1-22, 2003. DOI: 10.1002/jae.659

[10] Perron, P., Dealing with structural breaks, in Hassani, H. et al. Palgrave Handbook of Econometrics Vol. 1: Econometric Theory, UK, Palgrave Macmillan, 2006, pp. 278-352.

[11] Chow, G,. Tests of equality between sets of coefficients in two linear regressions, Econometrica, 28, pp. 591-605, 1960. . DOI: 10.2307/2288492

[12] Andrews, W.K.D., Test for parameter instability and structural change with unknown change point, Econometrica, 61(4), pp. 821-856, 1993. DOI: 10.2307/2951764

[13] Hansen, B.E., Approximate asymptotic P values for structural change tests, Journal of Business & Economic Statistics, 15(1), pp. 60-67, 1997. DOI: 10.1080/07350015.1997.10524687

D. Cardoza-Montalvo, has a MSc.in Engineering with Sp. in Electric Engineering from Universidad del Valle, Cali, Colombia, and a BSc. in Electric Engineer from the same institution. Currently he works as a professional in regulatory analysis at the division of Regulatory Analysis of EPSA E.S.P. S.A. David belongs to the research group in Energy Conversion (Convergía) attached to the School of Electric and Electronic Engineering at Universidad del Valle, Colombia. His research areas include the study of renewable energies, efficient energy usage, electric sector regulation and power systems analysis.
ORCID: 0000-0003-2319-981X

J.M. Uribe, has a Ph.D. in Economics, Universitat de Barcelona, Spain, a MSc. of Research in Economics, European University Institute, and a MSc. in Economics, Universidad de los Andes, Colombia. He is an Economist from Universidad del Valle, Colombia and a BSc. in Finance from Santiago de Cali University. Currently he is an associate professor in the Economics Department at Universidad del Valle and director of the research group in Applied Macroeconomics and Finance. He is author of several publications in energy and finance fields in international peer-review academic journals and has participated with contributed papers and presentations in many international congresses and conferences about energy markets, finance, and risk. His research has focused on the econometrics of risk and uncertainty, so as on asset pricing in energy and stock markets.
ORCID: 0000-0002-5844-2771

J.A. Palacios-Peñaranda, has a PhD. in Energy Generations Systems and is a Specialist in Rational Use of Energy, both from Universidad Politécnica de Madrid, Spain. He is also an BSc. Electrical Engineer from Universidad del Valle, Colombia. Currently he is full professor at Universidad del Valle, and he belongs to the Group in Energy Conversion at the School of Electric and Electronic Engineering at Universidad del Valle. He has been part of Universidad del Valle since 1984 and has been in charge of numerous courses both at undergraduate and graduate levels. He also has served as director and committee member at numerous master and PhD thesis, in the research lines of: alternative sources of energy, rational use of energy, electric power generation and electric machines. He has participated as speaker in numerous national and international professional and academic meetings.
ORCID: 0000-0002-9742-5608