



**Received:** 26 August, 2022

**Accepted:** 12 September, 2022

**Published:** 13 September, 2022

**\*Corresponding author:** Harry Ramenah, Senior Researcher, Physics, University of Lorraine, 7 rue de Marconi, 57070 Metz, France, Tel: + 33 3 72 74 93 25, 00 33 3 6 12 53 87 42, E-mail: [harry.ramenah@univ-lorraine.fr](mailto:harry.ramenah@univ-lorraine.fr)

**Keywords:** Photovoltaic; Regression; Engle & Granger; VECM; Cointegration; Correlogram

**Copyright License:** © 2022 Ramenah H. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

<https://www.peertechzpublications.com>



## Case Study

# From Engle & Granger model to Johansen model for a more accurate photovoltaic power output forecast

Harry Ramenah\*

Senior Researcher, Physics, University of Lorraine, France

## Abstract

The French government has recently decided to increase the Photovoltaic (PV) capacities to reach 35GW by 2028 in all french territories, the European territory, and overseas territories such as Reunion Island in the Indian Ocean. However, integrating growing numbers of PV power installations and microgrids onto the grid can result in larger-than-expected fluctuations in grid frequency. This is due to PV power output that is not only a function of the operating temperature and solar irradiation but also of other environmental parameters. In this paper, only two environmental parameters are considered in the European zone and when the Engle & Granger statistical method is used, a relationship between variables such as photovoltaic power output and solar irradiation at a different level is obtained. The final relationship without suspicious heteroscedasticity is determined. The model is formulated on the basis of photovoltaic real conditions statistical approach and is more realistic than steady approach models. The Engle & Granger method does not distinguish several cointegration relationships when more variables are considered. For the overseas zone, we added other measured environmental variables and applied a more robust statistical method known as the Johansen vector error correction model (VECM) cointegration approach. In the VECM model, for  $N$  explanatory variables and for  $N > 2$ , we established a long-run equilibrium relationship that has been tested and the outcome is more than reliable when comparing the model to measured data.

## Introduction

Renewable energies [1,2] are strongly developed to decarbonize our way of life in the energy sector and electricity production. The goal of this research is to contribute to the improvement of the short and medium-term predictability of Photovoltaic (PV) power production. The study is based on an analysis and prediction model (b) of PV production, involving spatial and temporal meteorological parameters. Power output (P) from PV systems in outdoor conditions is substantially influenced by climatic parameters such as solar irradiance (G) and module temperature (T). Integrating growing numbers of PV power installations and microgrids [2] onto the grid can result in larger-than-expected fluctuations in grid frequency. This is due to PV power output that is not the only function of the operating temperature and solar irradiation but also of other environmental parameters. In this paper, the geographical

distribution of PV output considering the module temperature effect and irradiation and other environmental parameters on PV system performance is considered in two different French territories: the European territory and an overseas territory such as Reunion Island in the Indian Ocean. The goal is to determine a linear relationship between PV power output and the environmental parameters from time series data.

For the European territory, we investigated rigorous statistical methods such as Engle & Granger (EG) [3,4], for stationary series to determine the estimate regression between variables P, G & T. In the EG method, when PV explanatory variables are nonstationary then the first difference study must be applied as the first difference transformation may make a nonstationary time series to become a stationary series to reach the final PV equation. Moreover, when outliers are suspected in the model, the EG method is put forward to determine the most

appropriate model. We investigated the dependent variable P on explanatory variables such as G and T. We first applied the Augmented Dickey & Fuller (ADF) [5,6], to our time series. The ADF test is a unit root test for stationarity. In this study, a visual diagnosis tool known as correlogram [7,8], is used to identify the first step for stationary test computing the autocorrelation function and the partial autocorrelation function. When serial correlation in residuals is detected, we first applied the Goldfeld-Quandt(GQ) [9], test when heteroscedastic variance is related in variables followed by the Durbin Watson (DW) [3] test in the regression model.

The downside of the EG method is that it does not distinguish several cointegration relationships for N simultaneous variables invoking up to N-1 cointegration relations. To overthrow such a situation, a more robust statistical method is suggested known as the Johansen vector error correction model (VECM) [10-12]. The Johansen test can be considered as a multivariate generalization of the ADF test, but the former is a strategic test that makes it possible to estimate all cointegrating vectors when more than two variables are considered. For the overseas and tropical zone, the Johansen VECM model is well adapted and we considered more explanatory variables such as wind speed(Wind) and humidity (Humi) that are added to G & T.

Therefore, this paper is divided into two parts comparing the model obtained to measured data for the two French territories: the European zone and the Tropical zone.

For the European zone, the EG model is determined for data from the GREEN platform of the Physics department of the University of Lorraine in Metz in northeast France. The PV design of the GREEN platform [13] is a grid-connected system. Six PV polycrystalline modules of SCHÜCO technologies are connected in a series wiring pattern and mounted on the south-southeast vertical wall of the platform building. Each module has a peak power of 205 Wp, a tilt angle of 60°, low ventilation, and is connected to an SCHÜCO inverter for a power level up to 1 kW. For the European zone, only data such as PV power out, solar radiation, and module temperature should be considered in the EG method to determine the estimated regression between these variables. The resulting model is then compared to real data to show a good agreement between the model and measured PV output data.

For the tropical zone on Reunion island in the Indian Ocean, the Johansen VECM model is performed to determine the regression relationship. The PV system in Reunion island is a grid-connected system. The modules that make up the PV plant are at a tilted angle of 21°, the same as Reunion Island latitude. The polycrystalline PV module of 180W each is equipped with solar irradiance, cell temperature, and wind and humidity sensors.

Other environmental parameters such as wind speed and humidity are considered in addition to solar radiation and module temperature for the tropical zone. The EG method is not suitable for the tropical zone as more than 2 explanatory

variables are considered, therefore the VECM model should be considered. The final long-term relationship obtained is compared to the recorded power output in real outdoor conditions.

This paper is organized as follows. Section 2 is the outline of correlograms for solar irradiance as well as its first difference to identify the stationary series. Still in that section, as the European zone is mainly focused on the Engle & Granger method, the EG principle is explained and applied to PV data from the Green platform in Metz. Section 3 describes the main approach of the Johansen VECM cointegration technique and the model is applied to compare PV power output on Reunion Island. Finally, a conclusion is proposed in section 4 where the perspective side is also discussed.

### Correlogram and engle & granger model

A correlogram is a visual diagnosis for the stationary test. For example, Table 1a is the solar irradiance series correlogram for data from the Green platform. In the autocorrelation column, spikes are outside the two lines indicating no stationarity. Moreover, the Q-stat was given as in 1:

$$Q = T(T + 2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{T - K} \tag{1}$$

Table 1a: Solar irradiance series correlogram for data from the Green platform.

	Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
1	0.498	0.498	90.872	0.000		
2	0.365	0.155	139.74	0.000		
3	0.355	0.169	186.19	0.000		
4	0.278	0.031	214.64	0.000		
5	0.218	0.012	232.26	0.000		
6	0.214	0.051	249.31	0.000		
7	0.244	0.102	271.53	0.000		
8	0.336	0.202	313.56	0.000		
9	0.318	0.073	351.43	0.000		
10	0.301	0.051	385.46	0.000		
11	0.295	0.032	418.28	0.000		
12	0.259	0.008	443.66	0.000		
13	0.225	0.012	462.83	0.000		
14	0.205	0.013	478.82	0.000		
15	0.194	0.012	493.19	0.000		
16	0.236	0.066	514.54	0.000		
17	0.275	0.082	543.41	0.000		
18	0.260	0.020	569.35	0.000		
19	0.236	-0.016	590.87	0.000		
20	0.244	0.021	613.84	0.000		
21	0.231	0.020	634.55	0.000		
22	0.166	-0.047	645.20	0.000		
23	0.150	-0.011	653.95	0.000		
24	0.233	0.110	675.16	0.000		
25	0.252	0.060	700.12	0.000		
26	0.234	0.012	721.59	0.000		
27	0.271	0.056	750.56	0.000		
28	0.289	0.044	783.52	0.000		
29	0.271	0.030	812.75	0.000		
30	0.225	-0.004	832.83	0.000		
31	0.188	-0.013	846.99	0.000		
32	0.177	-0.019	859.57	0.000		
33	0.181	0.004	872.73	0.000		
34	0.190	0.019	887.24	0.000		
35	0.207	0.013	904.53	0.000		
36	0.258	0.065	931.47	0.000		

Where  $k$  is the number of lags,  $T$  is the total number of observations,  $m$  is the lag length, and  $\hat{\rho}_k$  is the estimated values at lag  $k$ . Table 1b is the first difference solar correlogram. Similar diagrams have been obtained respectively for power output and module temperature but are not indicated here. Table 1b, indicates that stationary series is obtained for the first difference and the EG principle should be applied at a level as indicated below.

### Engle & granger model

The EG test for cointegration is a three-step procedure.

**First step:** it is necessary to ensure that the first differences of the corresponding  $Z_t$  and  $X_t$  series are stationary series and where  $Z_t$  is regressed on  $X_t$ .

**Second step:** let  $e_t$  be the residual of the regression of  $Z_t$  with respect to  $X_t$  given as follows:

$$Z_t = aX_t + b + e_t \tag{2}$$

and if  $e_t$  is stationary then  $Z_t$  and  $X_t$  series are cointegrated and the relationship is usually called long-run equilibrium.

**Third step:** Error Correction Model (ECM) to reconcile the short-run behavior with long-run behavior.

We applied the first and second steps and computed the regression as each variable is stationary in difference and the resulting equation is given as follows :

$$\Delta P = 0.969 \Delta G + 0.500 \Delta T - 0.105 \tag{3}$$

From the resulting table of the regression difference data, we noted that the corresponding  $p$  - value of  $\Delta T$  and constant term  $C$  are not statistically significant because their corresponding probability value (0.4077, 0.9953) are greater than the usual significance level of 1%, 5%, and 10%.

$$\Delta P = 0.969 \Delta G \tag{4}$$

We thus removed  $\Delta T$  and 0.105. The  $R^2$  value is more than 94%. The corresponding plot of the  $\Delta P$  against  $\Delta G$  is illustrated in Figure 1.

From Figure 1, we deduced that scattered values and the slight intercept at  $\Delta P$  axis are indications of heteroscedasticity [14]. We, therefore, applied the Goldfeld quant test followed by the Durbin Watson D-Stat test but the regression model was still not appropriate and we deduced that the ECM must be applied.

The ECM which is the third step of the EG model is applied to equation 2 and is given as follows:

$$\Delta P = \alpha_0 + \alpha \Delta G + \beta e_{t-1} + \varepsilon_t \tag{5}$$

Where  $\Delta$  is the first difference operator,  $\varepsilon_t$  is a random error term,  $e_{t-1}$  is the equilibrium error term, and if  $\beta$  contains a negative sign meaning that a run long equilibrium exists among the variables. After computing, the characteristic terms of the ECM regression without the constant term are illustrated in Table 2.

The Residpoint(-1) is the  $e_{t-1}$  value with a negative sign validating the long-run relationship between  $P$  and  $G$ . The final relationship between variables is deduced from Table 2.

$$\Delta P = 0.966 \Delta G - 0.601 e_{t-1} \tag{6}$$

### Discussion

For the EG statistical method, we investigated the dependent variable  $P$  on explanatory variables such as  $G$  and  $T$  and showed that the first difference of  $P$  is a function of only solar irradiance at the level. We showed that the model equation is in agreement with experimental measurements. However, when more explanatory variables are considered the EG is no longer suitable and another model such as the Johansen VECM cointegration model must be applied. This is discussed in the next section.

### Johansen VECM cointegration

In this section, we are considering the Johansen VECM cointegration applied to data from a grid-connected PV

**Table 1b:** Stationary series is obtained for first difference and the EG principle should be applied.

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1 -0.370	-0.370	50.000	0.000
		2 -0.119	-0.297	55.185	0.000
		3 0.066	-0.128	56.780	0.000
		4 -0.022	-0.100	56.952	0.000
		5 -0.055	-0.130	58.088	0.000
		6 -0.032	-0.165	58.478	0.000
		7 -0.065	-0.253	60.027	0.000
		8 0.109	-0.117	64.418	0.000
		9 0.003	-0.087	64.422	0.000
		10 -0.007	-0.060	64.439	0.000
		11 0.030	-0.033	64.772	0.000
		12 -0.002	-0.034	64.773	0.000
		13 -0.014	-0.033	64.847	0.000
		14 -0.013	-0.035	64.909	0.000
		15 -0.056	-0.091	66.097	0.000
		16 0.004	-0.106	66.104	0.000
		17 0.055	-0.039	67.269	0.000
		18 0.008	-0.003	67.292	0.000
		19 -0.027	-0.037	67.572	0.000
		20 0.020	-0.038	67.722	0.000
		21 0.054	0.032	68.855	0.000
		22 -0.050	-0.003	69.830	0.000
		23 -0.100	-0.125	73.751	0.000
		24 0.062	-0.076	75.254	0.000
		25 0.039	-0.023	75.863	0.000
		26 -0.057	-0.068	77.145	0.000
		27 0.025	-0.052	77.382	0.000
		28 0.030	-0.044	77.735	0.000
		29 0.034	-0.004	78.195	0.000
		30 -0.014	-0.002	78.278	0.000
		31 -0.024	0.008	78.502	0.000
		32 -0.016	-0.012	78.600	0.000
		33 -0.006	-0.027	78.616	0.000
		34 -0.004	-0.015	78.623	0.000
		35 -0.039	-0.074	79.225	0.000
		36 0.023	-0.068	79.433	0.000

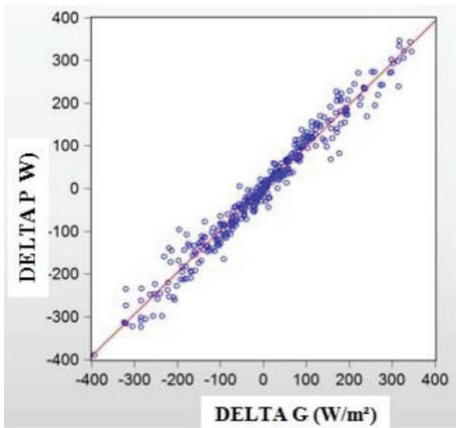


Figure 1: Regression of  $\Delta P$  against  $\Delta G$ .

Table 2: ECM regression data without constant.

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ResidCoInt(-1) -	0.601050	0.049178	-12.22193	0.0000
DELTA G	0.966657	0.008066	119.8493	0.0000
R-squared	0.977102	Mean dependent var 0.949972		
Adjusted R-squared	0.977037	S.D. dependent var 140.1690		
S.E. of regression	21.24035	Akaike info criterion 8.955253		
Sum squared resid	160610.3	Schwarz criterion 8.976932		
Log likelihood -	1600.990	Hannan-Quinn criter. 8.963875		
Durbin-Watson stat	2.062846	F-statistic 7578.573		
Prob(F-statistic)	0.000000			

system on Reunion island in the Indian Ocean. The Johansen VECM cointegration test can be considered a multivariate generalization of the ADF test and makes it possible to estimate all cointegrating vectors when more than two variables are considered.

### The Johansen approach

The general form of the VAR (p) model, without drift, is given as in equation 7

$$X_t = \pi + A_1 X_{t-1} + \dots + A_p X_{t-p} + \varepsilon_t \quad (7)$$

Where  $\pi$  is the vector-valued mean of the series,  $A_i$  is the coefficient matrices for each lag, and  $\varepsilon_t$  is a multivariate white noise term. The vector error correction model (VECM) is obtained by differencing the series as given in equation 8.

$$X_t = \pi + A X_{t-1} + \gamma_1 \Delta X_{t-1} + \dots + \gamma_p \Delta X_{t-p} + \varepsilon_t \quad (8)$$

Where  $\Delta X_t = X_t - X_{t-1}$  is the differencing operator,  $A$  is the coefficient matrix for the first lag, and  $\gamma_j$  are the matrices for each differenced lag. The matrix rank is respectively,  $X_t (N,1)$ ,  $A (N, N)$ , and  $X_{t-1} (N,1) \dots A_p (N, N)$ ,  $X_t (N,1)$ , and  $\varepsilon_t (N,1)$ , where  $N$  is the number of variables of nonstationary  $I(1)$ . When matrix  $A = 0$  there is no cointegration but for multiple linear combinations of time series the eigenvalue decomposition of  $A$  is carried out.

This first difference VAR (2) model can be written in a vector error correction model (VECM) as a function of only  $P_{t-1}$  as in equation 9:

$$\Delta X_t = -A_2 \Delta X_{t-1} + \Pi X_{t-1} + \varepsilon_t \quad (9)$$

Where  $\Pi = A_1 + A_2 - I$  and  $I$  is the unit matrix. Equation 9 can also be written as a function of  $P_{t-1}$  and  $P_{t-2}$  as given in Eq.10.

$$\Delta P_t = (A_1 - I) \Delta X_{t-1} + \Pi X_{t-2} + \varepsilon_t \quad (10)$$

If the coefficient matrix  $\Pi$  has reduced rank  $r < k$ , where  $k$  is the vector variables of  $I(1)$ ,  $r$  is the number of cointegration equations. The matrix  $\Pi$  can be written in terms of a vector of adjustment parameters  $\alpha$  and a matrix of cointegration vectors  $\beta'$  given by equation 11.

$$\Pi = \alpha \beta', \text{ where } \beta' P_t \text{ is } I(0) \quad (11)$$

Where  $\alpha$  is an  $(N,r)$  matrix with  $r < N$ , and  $\beta'$  has  $r$  cointegration vectors such that  $0 < r < N$  as to highlight the VECM model.

The Johansen test and estimation strategy which is a maximum likelihood test make it possible to estimate all cointegrating vectors for  $N$  variables, which all have unit roots and there are at most  $N-1$  cointegrating vectors. The Johansen test provides estimates of all cointegrating vectors if a cointegration relationship does exist, and a rank test is useful. Thereby, if:

Rank ( $\Pi$ ) = 0, then  $r = 0$  meaning that none cointegration relationship and VECM cannot be applied,

Rank ( $\Pi$ ) =  $r$ , meaning that variables are cointegrated and the number of cointegration relationships is equal to  $r$ . VECM model can be estimated.

Rank ( $\Pi$ ) =  $N$ , meaning that none cointegration relationship.

Johansen's procedure is based on the maximum Eigenvalue and Trace tests that are conducted on the error correction model foundation. For both test statistics, the initial Johansen test is a null hypothesis test of no cointegration against the alternative of cointegration.

The first test of maximum Eigenvalues is to determine whether the rank of the matrix is zero, and the null hypothesis is rank ( $\Pi$ ) = 0 whereas the alternative hypothesis is rank ( $\Pi$ ) = 1.

The second test of Trace is to determine whether the rank of the matrix is  $r_0$ , the null hypothesis is rank ( $\Pi$ ) =  $r_0$  and the alternative hypothesis is that  $r_0 < \text{rank}(\Pi) \leq r$ , where  $r$  is the maximum number of possible cointegration vectors.

For the VECM model in this study, variables such as wind speed(Wind), and humidity (Humi) are two more explanatory variables in addition to  $G$  &  $T$  in the EG section and noted respectively as Irra and Temp in this section. Different



forecasting classifications have been proposed and in our study, we should consider the short-term forecast, that is hourly, several hours up to a day ahead to guarantee system commitment and scheduling.

The Johansen VECM test for cointegration is a five-step procedure.

- Step 1: Performing series stationarity (correlogram & ADF) tests to determine whether there is a cointegration relationship or not.

- Step 2: If the step1 is true, meaning that series are of the same order of integration and cointegration is likely, therefore VECM model can be estimated. Determining the lag length using Akaike and Schwarz criteria [3,15].

- Step 3: Implementing the Johansen test to determine the number of cointegration relationships.

- Step 4: Identifying the cointegration relationships or long-term relationships between variables.

- Step 5: Estimating the VECM model by maximum likelihood method, test validations by visual diagnostic or correlogram, and checking that residuals from the model are white noise.

A vector autoregression  $P_t$  of 5 variables lagged 2 is given in equation 12:

$$P_t = A_1 P_{t-1} + A_2 P_{t-2} + \varepsilon_t \tag{12}$$

The matrix form is illustrated as follows :

$$\begin{bmatrix} P_{1t} \\ P_{2t} \\ P_{3t} \\ P_{4t} \\ P_{5t} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \\ a_{31} & a_{32} & a_{33} & a_{34} & a_{35} \\ a_{41} & a_{42} & a_{43} & a_{44} & a_{45} \\ a_{51} & a_{52} & a_{53} & a_{54} & a_{55} \end{bmatrix} \begin{bmatrix} P_{1t-1} \\ P_{2t-1} \\ P_{3t-1} \\ P_{4t-1} \\ P_{5t-1} \end{bmatrix} + \begin{bmatrix} a_{16} & a_{17} & a_{18} & a_{19} & a_{110} \\ a_{26} & a_{27} & a_{28} & a_{29} & a_{210} \\ a_{36} & a_{37} & a_{38} & a_{39} & a_{310} \\ a_{46} & a_{47} & a_{48} & a_{49} & a_{410} \\ a_{56} & a_{57} & a_{58} & a_{59} & a_{510} \end{bmatrix} \begin{bmatrix} P_{1t-2} \\ P_{2t-2} \\ P_{3t-2} \\ P_{4t-2} \\ P_{5t-2} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \varepsilon_{3t} \\ \varepsilon_{4t} \\ \varepsilon_{5t} \end{bmatrix} \tag{13}$$

We computed the different steps up to step 3 where the number of cointegration relationships is based on Trace and Eigenvalues [14,16], tests. For this study, this test is performed with the deterministic trend assumption, that is no intercept or trend in the cointegration equation or VAR test. The result after computing the Johansen VECM test with one lagged indicates four cointegration equations. These are indicated in Tables 3a and table 3b are the error correction coefficients.

With data, respectively from Tables 3a and 3b, we identified the 4 equations as follows :

$$\text{CoinEq1} = (\text{Power}_{t-1} - 3521,54 \text{Wind}_{t-1}) \tag{14a}$$

$$\text{CoinEq2} = (\text{Irra}_{t-1} - 189,05) \tag{14b}$$

$$\text{CoinEq3} = (\text{Temp}_{t-1} - 16.52 \text{Wind}_{t-1}) \tag{14c}$$

$$\text{CoinEq4} = (\text{Himi}_{t-1} - 0.289 \text{Wind}_{t-1}) \tag{14d}$$

The  $\Delta P$  equation is given as follows :

Table 3a: Four Cointegration Equations.

Cointegration Eq:	CointEq1	CointEq2	CointEq3	CointEq4
Power(-1)	1.000000	0.000000	0.000000	0.000000
IRRA(-1)	0.000000	1.000000	0.000000	0.000000
TEMP(-1)	0.000000	0.000000	1.000000	0.000000
HUMI(-1)	0.000000	0.000000	0.000000	1.000000
WIND(-1)	-352.536	-189.051	-16.5261	-0.28973
	(115.133)	(6.10357)	(0.43571)	(0.00856)
	[-30.5867]	[-30.9738]	[-37.9295]	[-33.8617]

Table 3b: Error correction coefficients.

Error Correction	D(POWER)	D(IRRA)	D(TEMP)	D(HUMI)	D(WIND)
CointEq1	-0.738061 (0.44968) [-1.64130]	0.003671 (0.02418) [0.15180]	0.000668 (0.00069) [0.96372]	-6.30E-06 (9.4E-06) [-0.66729]	-0.000101 (0.00011) [-0.90932]
CointEq2	2.764963 (8.49770) [0.32538]	-0.633466 (0.45697) [-1.38624]	-0.020299 (0.01309) [-1.55056]	-6.30E06 0.00018 [-0.35314]	0.001919 (0.00211) [0.91023]
CointEq3	29.00405 (57.8867) [0.50105]	1.202475 (3011288) [0.38629]	-0.168783 (0.08918) [-1.89266]	0.007478 (0.00122) [6.15305]	0.014007 (0.56441) [0.97515]
CointEq4	3493.349 (2274.49) [1.53588]	170.2571 (122.312) [3.50397]	10.64644 (12.312) [3.03840]	-0.318148 (0.04776) [-6.66201]	1.006675 (0.56441) [1.78360]
D(POWER(1))	-0.061329 (0.34563) [-0.17744]	0.000548 (0.01859) [0.02947]	-0.000373 (0.00053) [-0.70012]	8.67E-06 (7.3E06) [1.19434]	0.000102 (8.6E-05) [1.18555]
D(IRRA(1))	-0.385590 (6.98093) [-0.05523]	-0.119264 (0.37540) [-0.31770]	0.009972 (0.01075) [0.92728]	-4.70E-05 (0.00015) [-0.32055]	-0.0011840 (0.00173) [-1.06198]
D(TEMP(1))	-41.98734 (79.2692) [-0.52968]	-1.471655 (4.26273) [0.34524]	-0.312499 (0.12212) [-2.55899]	-0.0004109 (0.00166) [-2.46904]	-0.012250 (0.01967) [-0.62278]
D(HUM(-1))	-2443.284 (2670.53) [-0.91491]	-9171841 (143.609) [-0.63867]	-4715188 (4.11409) [-1.14611]	-0.009332 (0.05607) [-0.16643]	-0.646033 (0.66268) [0.97487]
D(WIND(-1))	-388.2601 (219.111) [-1.77198]	-22.43381 (11.7828) [-1.90395]	-0.629913 (0.33755) [-1.86612]	-0.003435 (0.0046) [-0.74676]	-0.025043 (0.05437) [-0.46059]

$$\Delta P = -0.738(\text{Power}_{t-1} - 3521,54 \text{Wind}_{t-1}) + 2.764(\text{Irra}_{t-1} - 189,05) + 29.004(\text{Temp}_{t-1} - 16.52 \text{Wind}_{t-1}) + 3493.349(\text{Humi}_{t-1} - 0.289 \text{Wind}_{t-1}) + \varepsilon_{1t} \tag{15}$$

For  $\Delta P = P_t - P_{t-1}$ , and as each variable at (t-1) is equal to each variable at t , we deduced the long-term relationship as in equation 16:

$$P_t = 3.74 \text{Irra}_t + 36.38 \text{Temp}_t + 833.33 \text{Wind}_t + 4758.96 \text{Humi}_t + \varepsilon_{1t} \tag{16}$$

To determine the residual of the cointegration equation we performed distinct tests such as the Wald test, the Lagrange multiplier test, the jarque bera statistic and finally the CUSUM test indicating that the residual is a random or white noise process. These tests can be easily understood in literature [3].

In the next section, the model obtained is compared to the measured data.

### Comparing Johansen model to PV output in reunion island

The Johansen VECM model is applied to forecast PV power output by comparing the model data to measured data in real outdoor conditions. This was done on a yearly basis from 2013 to 2016 for a PV grid connected system in Reunion Island in the Indian Ocean but in this study, we are putting forward a daily comparison upon one month. This is represented in Figure 2, where the blue (series 2) and green (series 1) colors of the bar chart are respectively measured power output and Johansen model power output for each day of the month of march 2016.

The model was also applied for an hourly short-term forecast as illustrated in Figure 3.

The orange and blue (series1) colors are respectively the Johansen model power output and measured power output.

### Conclusion

Power output from photovoltaic (PV) systems in outdoor conditions are substantially influenced by climatic parameters such as solar irradiance and module temperature. One of the objectives of this paper consisted of applying a statistical method of time series data to identify the most important on-site climatic and environmental parameters that influence PV output variability. It is strongly hard to estimate the impact of PV output variability on the power grid stability unless a good comprehension of the parameters controlling this variability. Many OLS regression models relating to PV

variables have been proposed in the literature but most of them had led to spurious results. In this paper, two robust statistical models have been used in two French territories. The Engle & Granger statistical method is applied in the European zone and the VECM Johansen method is applied to the tropical zone on Reunion island in the Indian Ocean. In this study, we showed that if only these two parameters are considered in the European zone and mainly in the East of France, then the Engle & Granger method is a rigorous method for the long term-term forecast of PV power output. However, when more explanatory variables such as wind speed and humidity are considered in addition to solar irradiance and temperature, the EG method is unable to distinguish several cointegration relationships. We showed that a more robust technique such as the Johansen VECM cointegration must be applied to the Reunion island and can be an accurate technique for short-term forecast of PV power output. From the perspective side in the future, a Spatio-temporal model using recurrent neural networks with persistent short-term memory (LSTM) should be developed to produce efficient forecasts over the whole Reunion island. The developing methodologies should eventually offer an opportunity to provide additional guarantees to the network manager. If in the future efficient forecasting solutions become widespread, this opportunity should open up the market beyond the current regulatory threshold of 35% renewable energy as expected in Reunion Island. Moreover, the mathematical aspects behind the statistical theories should be computed inline code using Python 3.10 and integrated on an FPGA chip in order to be applied at minute sampling time to make more accurate daily predictions.

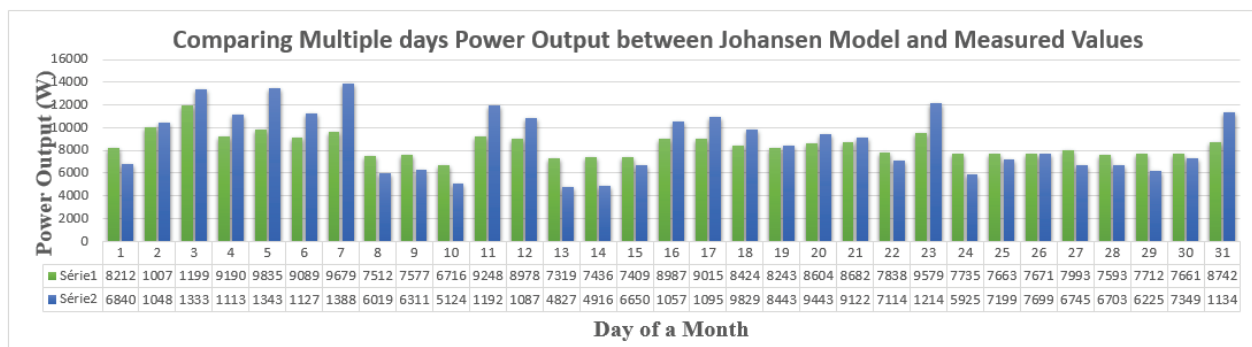


Figure 2: Comparing power daily power output between model and measured data.

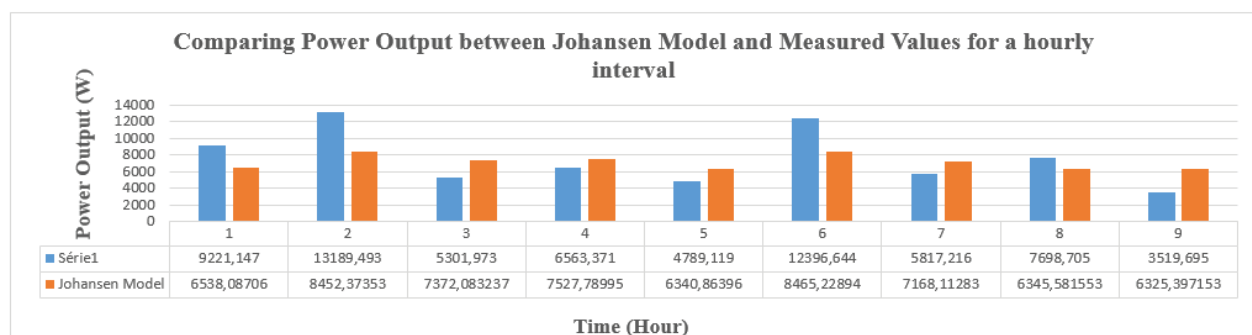


Figure 3: Comparing hourly power output between model and measured data.



## References

1. Coban HH, Rehman A, Mohamed A. Technical and economical investigation of a centralized and decentralized hybrid renewable energy system in cadaado, somalia, processes. 2022; 10(4):667. [HTTPS://DOI.ORG/10.3390/PR10040667](https://doi.org/10.3390/PR10040667)
2. Coban HH. While covid-19 outbreak affects economies and societies; exploring the energy demand in turkey. year. 2021; 11:126–135. [HTTPS://DOI.ORG/10.36222/EJT.894463](https://doi.org/10.36222/EJT.894463)
3. Engle RF, Granger CWJ. co-integration and error correction: representation, estimation, and testing. essays in econometrics. harvard university press 1987; 251–276.
4. Granger CWJ, Weiss AA. Time series analysis of error-correction. karlin s, amemiya t, goodman la, eds. studies in econometrics, time series, and multivariate statistics, academic press. 1983; 255–278.
5. Gujarati DN. Basic of econometric, fourth edition. the mcgraw-hill econometrics, fourth companies. 2004.
6. applied economic time series wiley series in probability and statistics. 1995; 420.
7. Hassani H, Yeganegi MR. Sum of squared acf and the ljung–box statistics. Phys a: Stat Mech Appl. 2019; 520: 81–86.
8. Ljung GM. Box Gep on a measure of lack of fit in time series models. Biometrika. 1978; 65:297–303.
9. Ramenah H, Casin P, Ba M, Benne M, Tanougast C. accurate determination of parameters relationship for photovoltaic power output by augmented dickey fuller test and engle granger method, AIMS-energy. 2018; DOI: 10.3934/ENERGY.2018.1.19
10. Johansen S. Statistical analysis of cointegration vectors. J Econ Dyn Control .1988; 12: 231–254.
11. Johansen S. Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. econometrica. 1991; 59: 1551–1580.
12. Johansen S. Likelihood-based inference in cointegrated vector autoregressive models. new york: Ooxford University Press. 1995.
13. Cornillon PA. Regression theorie & application, springer, 2007.
14. Hamilton JD. Time series analysis. princeton university press. 1994; 767–768.
15. Ramenah H, Tanougast C. Techniques of the green platform for research activities. (IOSR Journal of environmental science 2019. DOI:10.9790/2402-1312023241.
16. Long JS, Ervin LH. Using heteroscedasticity consistent standard errors in the linear regression model. the american statistician. 2000; 54: 217-24.

### Discover a bigger Impact and Visibility of your article publication with Peertechz Publications

#### Highlights

- ❖ Signatory publisher of ORCID
- ❖ Signatory Publisher of DORA (San Francisco Declaration on Research Assessment)
- ❖ Articles archived in worlds' renowned service providers such as Portico, CNKI, AGRIS, TDNet, Base (Bielefeld University Library), CrossRef, Scilit, J-Gate etc.
- ❖ Journals indexed in ICMJE, SHERPA/ROME0, Google Scholar etc.
- ❖ OAI-PMH (Open Archives Initiative Protocol for Metadata Harvesting)
- ❖ Dedicated Editorial Board for every journal
- ❖ Accurate and rapid peer-review process
- ❖ Increased citations of published articles through promotions
- ❖ Reduced timeline for article publication

Submit your articles and experience a new surge in publication services (<https://www.peertechz.com/submission>).

*Peertechz journals wishes everlasting success in your every endeavours.*