International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 446-450 www.ijaem.net ISSN: 2395-5252

Fault detection of a PV system by Principal Components Analysis

Todizara Andrianajaina[1], Sandaniaina S. A. Ranaivoson[2], Tsivalalaina David Razafimahefa[3], Eric Jean Roy Sambatra[4]

[14]Ecole Supérieure Polytechnique d'Antsiranana, University of Antsiranana, Madagascar
[2]Renewable Energy and Environment Thematic Doctoral School, University of Antsiranana,
[4]Higher Institute of Technology Antsiranana, Madagascar

Submitted: 05-11-2022 Accepted: 15-11-2022

ABSTRACT: PCA is often used to reduce the size of variables or to compress data. In this article, we use PCA to create a diagnostic for a solar-powered energy generation system. The idea is to represent the system's data in a graph with reduced dimensions to get a good representation of the data set and also to be able to interpret the behavior of the variables in this new representation.
The system has five potential variables that are not easy to plot. In this article, we plot the variables in two dimensions and interpret their behavior as a diagnosis..

KEYWORDS:PCA, fault detection, PV, photovoltaic.

I. INTRODUCTION

Principal component analysis (PCA) is a multivariate statistical technique [1]. Multivariate statistical techniques are powerful tools capable of compressing data and reducing their dimensionality [2] so that the essential information is preserved and easier to analyze than in the original dataset. These techniques can also manipulate noise and correlation to extract information effectively. The main function of this type of technique is to convert some correlated variables into a smaller set of uncorrelated variables through a mathematical procedure.

PCA is essentially based on an orthogonal decomposition of the covariance matrix of the process variables along the directions that explain the maximum variation of the data [3], i.e., this method looks for a projection of the observations on orthogonal axes. As a result, the first axis contains the largest variation. The second axis has the second largest variation orthogonal to the first.

The PCA's main goal is to find a set of factors (components) that are smaller in size than the original set of data and that can accurately describe the main trends [4].

In this article, we present the general principle of PCA, then how the variables are involved in constructing the axes, and then, see how to make the diagnosis.

II. PRINCIPLE OF PRINCIPAL COMPONENT ANALYSIS

The main interest of the ACP is to offer the best visualization of the multivariate data, by identifying the hyperplanes in which the dispersion is maximum, thus highlighting with the maximum precision the relations of proximity and distance between the variables [5]. The PCA consists of replacing a family of variables with new maximum variance variables uncorrelated and which are linear combinations of the original variables. These new variables, called principal components (CPs), define factorial plans that serve as a basis for a flat graphical representation of the initial variables.

Variables are usually expressed in units of measurement and scales. For this, it is preferable to carry out a PCA on a centered and reduced X matrix (columns of zero means and standard deviations units). The orthogonal space defined by the ACP is generated by the eigenvectors associated with the eigenvalues λ_a of the correlation matrix Σ of X. So, $x \in \mathbb{R}^m$ is a random data vector consisting of m variables. Let the data matrix $X \in \mathbb{R}^{nxm}$ of line vectors x_i^T which collects the n measurements on the m variables.

The ACP determines an optimal (versus a variance criterion) and linear transformation of the data matrix X as follows:

$$X = TP^T \qquad (1)$$

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 446-450 www.ijaem.net ISSN: 2395-5252

Where T = [t1 t2 . . . tm] $\in\mathbb{R}^{nxm}$, where the vectors ta are called scores or CPs and the matrix P = [p1 p2 . . . pm] $\in\mathbb{R}^{nxm}$, where the orthogonal vectors pa are the eigenvectors associated with the eigenvalues of the correlation matrix Σ of X:

$$\Sigma = P\Lambda P^T \text{avec } PP^T = P^T P = I_m \qquad (2)$$

Where Λ = diag(1 . . . m) is a diagonal matrix whose elements are put in decreasing order. Since the objective of the PCA is to reduce the size of the space, the first CPs (l m) is the most significant and sufficient to explain the variability of a process through its X database. Therefore, the partition into eigenvectors and principal components gives respectively:

$$P = [\hat{P}_l | \tilde{P}_{m-l}], T = [\hat{T}_l | \tilde{T}_{m-l}] \qquad (3)$$

The l first eigenvectors constitute the representation subspace or the main subspace (SP) defined by: Sp =span$\hat{P}$. While the residual subspace, denoted SR, is described by: Sr =span$\tilde{P}_{m-l}$. These two sub-spaces, Sp and Sr, are orthogonal. An observation vector x is projected onto the new space and decomposes on the two subspaces SP and SR respectively as follows:

$$\hat{x} = \hat{P}_l \hat{P}_l^T x = \hat{C}x \in S_p \qquad (4)$$

$$\tilde{x} = \tilde{P}_{m-l} \tilde{P}_{m-l}^T x = \tilde{C}x \in S_r \qquad (5)$$

$$\hat{x}^T \tilde{x} = \tilde{x}^T \hat{x} = 0 \text{ et } x = \hat{x} + \tilde{x} \qquad (6)$$

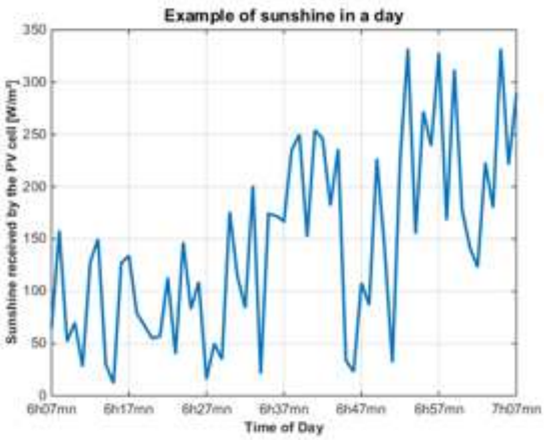
Where $\hat{x}$ and $\tilde{x}$ are respectively the projection of x on the two subspaces SP and SR respectively generated by the first CPs and the remaining (m-l).$\hat{C}$ and $\tilde{C}$ = (1- $\hat{C}$) represent the projection matrices respectively on the SP and the SR. Note that the matrix $\hat{C}$ is not equal to the identity matrix. If a value of this matrix is close to 1, this means that the corresponding variable is not correlated with the others, and therefore it is estimated from its measure (this variable is projected completely on the SP).

III. FAULT DETECTION

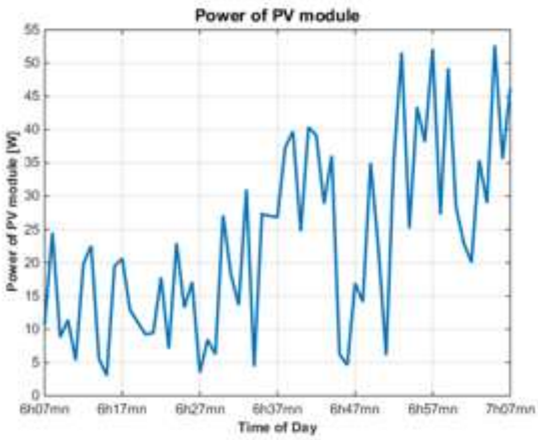
In this section, we discuss the implementation of the fault detection method for a solar power system (SPS). This requires an inventory of individuals described by quantitative variables to be studied. In fact, we are mainly interested in the output variables of SPS, including voltage, current and power, as well as the variables that mainly affect the behavior of the system, including temperature and irradiation.

The goal is to see, in a symmetrical way, the similarity of these variables with those that provide almost identical information. This leads us to see the connection between these variables by studying the correlation. And the most important thing is that we are able to make the data set formed by each one much more readable and interpretable in a graph.

Incidentally, let us consider the system subjected to the sunshine shown in Fig. 1a. The output variables are shown in Figs. 1b, 1c, and 1d, respectively

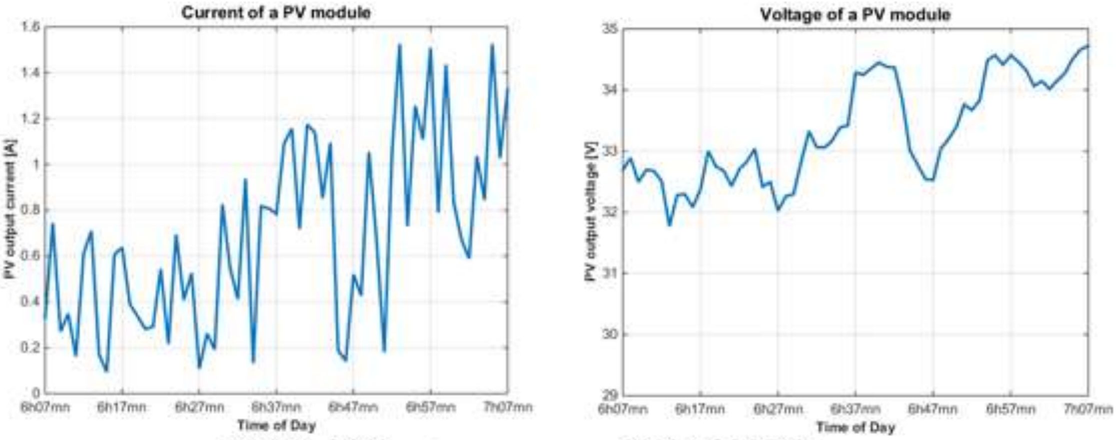


(a) Example of sunshine



(b) PV Module Power

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 446-450 www.ijaem.net ISSN: 2395-5252



(c) PV Module Current (d) PV Module Voltage

Fig. 1: Output of the PV module for a given sunshine

We studied all the data collected for one hour with a sampling interval of one minute.

The eigenvalue graph with the percentage of inertia associated with each dimension is given in Fig. 2.

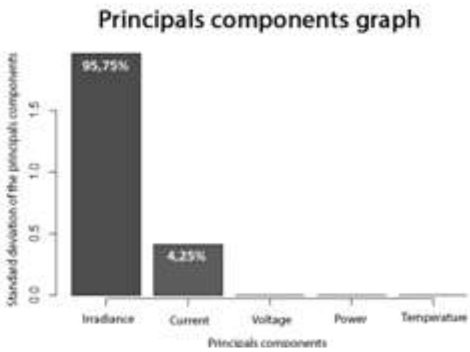


Fig. 2: Graph of the eigenvalues

From this figure, it can be said that the projections of the individuals on the two axes are highly significant (at 100%). Indeed, the percentage of inertia explained by the first dimension is 94.88%, which means that 94.88% of the information is represented by axis 1, and the second dimension expresses 5.12% of the information. Since the axes are ordered, it is thus possible to add the percentage of inertia of these axes, and then axis 1 and axis 2 express all the continuous information in the data set.

This means that if the initial variables of the dataset are summarized by two dimensions, all the information contained in the dataset is still recovered. We consider subsequently the two main components whose irradiance as the first component (denoted by Dim1 later) and the current as the second component (denoted Dim2 hereinafter).

The result of the PCA of all these data is given in the following figures (Fig. 3 and 4).

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 446-450 www.ijaem.net ISSN: 2395-5252

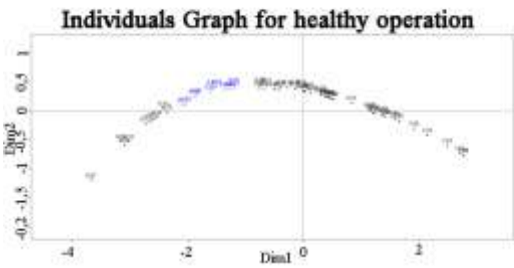


Fig.3:Individuals Graph for healthyoperation

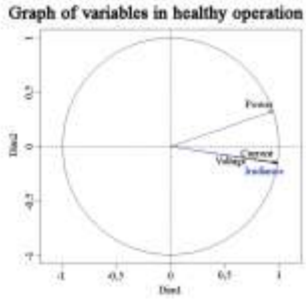


Fig.4: Graph of variables in healthyoperation

Individuals are the data points of measurement, here, from 6:07 to 7:07. We are particularly interested in the behaviour of individuals marked in blue. In fact, it is during this interval of time that we inserted, the fault to be monitored, is the subject of the following paragraph.

We have implemented several scenarios of defects (shading) to see how the system will react. The following figure 5 shows the projection of individuals according to the ACP method for each scenario.

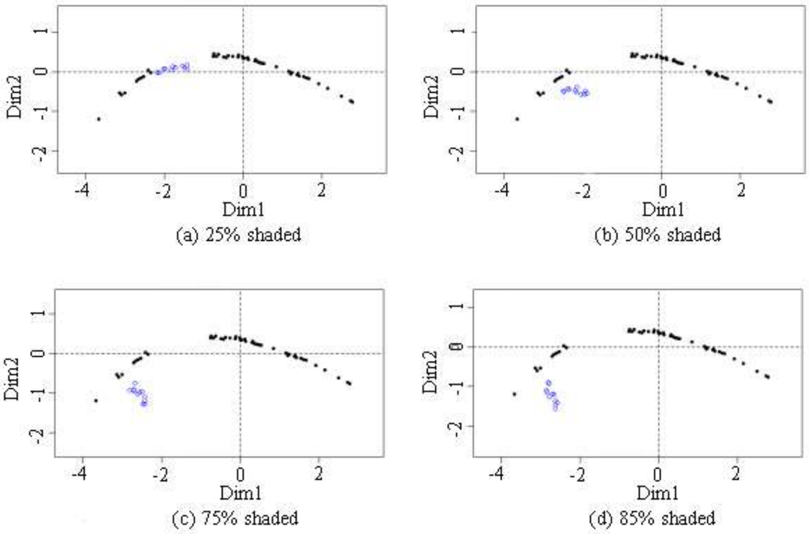


Fig. 5: Graph of individuals in faulty operation

To interpret these results, we need to know each axis is chosen and each cloud of variables that participates most in the formation of the axis. The participation of a variable in the formation of the axis results in the square of the correlation between the variable and the axis divided by the sum of the correlations between the variables and the axis as equation 7 shows [3, 4].

$$\mathrm{CTR_s(k)} = \frac{\mathrm{r(x_k, v_k)}^2}{\sum_{\mathrm{k=1}}^{\mathrm{K}} \mathrm{r(x_k, v_k)}^2} \qquad (7)$$

The summary of the results of the PCA we have done is given in Table 1.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 4, Issue 11 Nov. 2022, pp: 446-450 www.ijaem.net ISSN: 2395-5252

Table 1: Contribution of variables to the constitution of the axis and quality of representation

Variables	Dim.1	cos2	Dim.2	cos2
Irradiation	0.994	0.988	-0.110	0.012
Current	0.994	0.988	-0.111	0.012
Voltage	0.993	0.986	-0.119	0.014
Power	0.932	0.968	0.363	0.132
Temperature	0	NaN	0	Inf

With:

- Dim.1: coordinate of the variable on axis 1.
- Dim.2: coordinate of the variable on axis 2.
- cos2: The quality of the representation of a variable along a main axis is given by the square of its correlation coefficient with this axis and represents the squared cosine

$$\cos 2 = \mathrm{cor} \times \mathrm{cor} \qquad (8)$$

For each PCA done, the result is always the same, ie, the power has a significant value in the formation of the first component and it contributes to the formation of the second component; While current and voltage contribute strongly to the formation of the first component, but they contribute less to the formation of a second.

So, the loss of power in the presence of defects results in the translation of the individuals with respect to axis 1 and also slightly with respect to axis 2. Thus, this phenomenon will help us to detect the manifestation of a defect translating into a loss of power in the system.

IV. CONCLUSION

This paper presents a tool for fault detection using principal component analysis. The variables observed are the current and the temperature, since these are the two most weighty variables. The approach presented is to observe the sliding of the individuals on the principal axes, and it was found that the more consistent the error, the more the individuals move away. We can conclude that with ACP we were able to quantify the occurrence of defects.

REFERENCES

[1]. de Carvalho Michalski, M. A., & de Souza, G. F. M. (2022). Comparing PCA-based fault detection methods for dynamic processes with correlated and Non-Gaussian variables. Expert Systems with Applications, 207, 117989.

[2]. Prusty, M. R., Jayanthi, T., Chakraborty, J., &Velusamy, K. (2017). Feasibility of ANFIS towards multiclass event classification in PFBR considering dimensionality reduction using PCA. Annals of Nuclear Energy, 99, 311-320.

[3]. Kherif, F., &Latypova, A. (2020). Principal component analysis. In Machine Learning (pp. 209-225). Academic Press.

[4]. Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. D. F. (2021). Principal component analysis: A natural approach to data exploration. ACM Computing Surveys (CSUR), 54(4), 1-34.

[5]. Berger, J. L. (2021). Analyse factorielleexploratoire et analyse encomposantesprincipales: guide pratique.