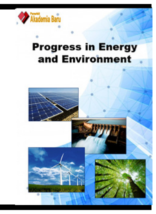
Penerbit Akademia Baru

Progress in Energy and Environment

Journal homepage: http://www.akademiabaru.com/ProgEE.html
ISSN:



Principal Component Analysis on Meteorological Data in UTM KL

Research Article

Tan Lit-Ken*,[1], Ong Sie-Meng[1], Lee Kee-Quen[1], Gan Yee-Siang[2], Goh Chien-Yong[2], Tey Wah-Yen [1,3], Su Kong-Ngien[4]

[1] Department of Mechanical Precision Engineering, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia, UTM KL, Jalan Sultan Yahya Petra 54100 Kuala Lumpur, Malaysia
[2] Department of Mathematics, Xiamen University Malaysia, 43900 Sepang, Selangor Darul Ehsan, Malaysia
[3] Department of Mechanical Engineering, Faculty of Engineering, UCSI University, 56000 Cheras Kuala Lumpur, Malaysia
[4] Faculty of Civil Engineering and Earth Resources, Universiti Malaysia Pahang, Lebuhraya Tun Razak, 26300 Gambang Kuantan, Pahang Darul Makmur, Malaysia

ARTICLE INFO

Article history:
Received 3 April 2017
Received in revised form 8 May 2017
Accepted 9 May 2017
Available online 5 June 2017

ABSTRACT

The high usage of fossil fuel to produce energy for the increasing demand of energy has been the primary culprit behind global warming. Renewable energies such as solar energy can be a solution in preventing the situation from worsening. Solar energy can be harnessed using available system such as solar thermal cogeneration systems. However, for the system to function smoothly and continuously, knowledge on solar radiation's intensity several minutes in advance are required. Though there exist various solar radiation forecast models, most of the existing models requires high computational time. In this research, principal component analysis were applied on the meteorological data collected in Universiti Teknologi Malaysia Kuala Lumpur to reduce the dimension of the data. Dominant factors obtained from the analysis is expected to be useful for the development of solar radiation forecast model.

Keywords:
Principal component analysis, Meteorological data, Solar radiation

1. Introduction

Global warming is a top environmental issue concerned by most of the people in recent. It brought effects not only to the environments but to human as well. Among the victims, human population are affected the most. Looking on the negative impacts brought by global warming to human and environments, reducing the demand on fossil fuel is deemed necessary. Indeed, there is no single solution of stopping people from using it. However, switching the demand from fossil fuel to other renewable energy such as solar, wind and marine energy is one of the consensuses from the scientists.

Renewable energy is generally characterized as clean and durable energy. There are many forms of renewable energy, and most of these energies are inter-related to sunlight. For instances, solar thermal

*Corresponding author.
E-mail address: tlken@utm.my

system harnesses solar energy to generate electrical energy while hydroelectricity generator generates hydropower via water movement. Since there are much renewable energies have been found in nowadays, some researches had been done to compare among these found renewable energies. Most of the scientists believed that solar energy could be one of the future potential powering source in the world.

Located near the equator, Malaysia receives abundant solar radiation annually, which indirectly caused to dramatically increment of solar energy usage in the past decade. However, power fluctuation is one of the many challenges faced by the implementation of solar energy. This is because of the intermittency and variability of solar radiation (see **Fig. 1**) received in Malaysia. To overcome such obstacle, establishment of an accurate solar radiation prediction model is important.

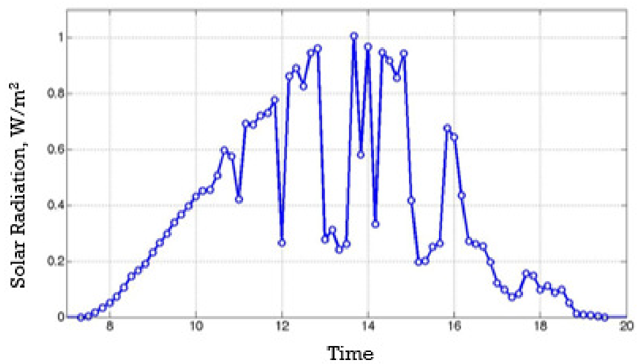


Fig. 1. Measured solar radiation during daytime

2. Literature Review

Solar energy can be harvested and transformed into useable electricity through the use of two main technologies such as solar thermal cogeneration system (STCS). STCS is a complete energy system, converting radiant energy into a useable energy form, providing both electrical and thermal energy for a detached residential or commercial structure. Basic components of STSC include a solar collector, working fluid, thermal storage, turbine, and generator as shown in **Fig. 2**.

The collector in STCS consists of a long rectangular curved mirror (concentrator) and a central heat pipe (receiver) located at the focal line of the curved mirror. The sunlight which enters the curved mirror is parallel to its plane of symmetry and will then focus along the focal line, where working fluid within the pipe is intended to be heated [1]. A tracking mechanism maintains concentrator focus as the sun traverses the sky is used to efficient heating up the working fluid within a pipe. The working fluid plays an important role to drive STCS's performance under changeable solar radiation. One of the common used working fluid iswater. This is because water can easily be boiled and converted into steam phase within a short period [1].

Another main component that is incorporated in STCS is thermal storage. It behaves as a container to store thermal energy (steam) in molten salt for later use. Thermal storage is controlled by a valve which manipulates the in-and-out of steam flow [1]. Generally, any thermal storage that can be found in nowadays have three main functions:

i.	Charge	: A heat source is used to provide heat to the storage medium
ii.	Store	: The storage medium is used to store the heat for later use
iii.	Discharge	: The heat leaves the storage medium in a controlled manner to be used
for		another purpose

The three main functions mentioned in above are switched among each other. This switching processes depend on the solar radiation that will be collected in a day by STCS. For instances, the thermal storage will switch from discharging mode to charging mode when the demand of electricity is higher than the solar radiation received and vice versa.

The switching between modes of the thermal storage takes time to stabilize the steam pressure within the storage. To ensure thermal storage in STCS is able to perform efficiently, an accurate short-term solar radiation forecasting several minutes in advance is needed.

There exist numerous numbers of methodologies to forecast solar radiation. The common methodologies found are either sky image based or by meteorological data based methodology. Sky image based approach (e.g. [2] and [3]) uses thousands of pixel information obtained from the image to make prediction, resulting in high computational time.

For meteorological data based approach (e.g., [4] and [5]), meteorological factors information such as ambient temperature, wind speed, humidity and etc. are used to make prediction. This is because meteorological factors have effect on solar irradiance intensity. There also exists research that uses lower number of variables, such as in [4] which uses one meteorological factor to predict solar radiation. However, the obtained results were not accurate (see **Fig. 3**).

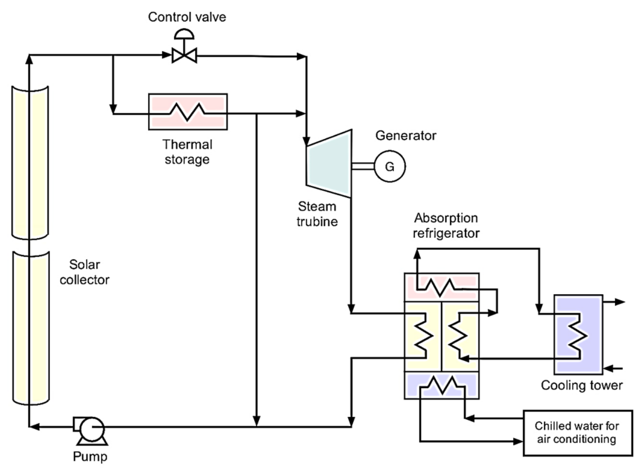


Fig. 2. Schematic diagram of solar thermal cogeneration system

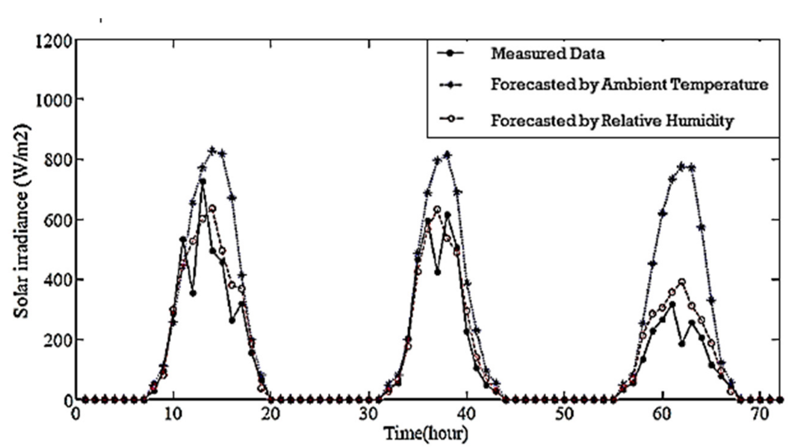


Fig. 3. Solar radiation predictions with different meteorological variables (ambient temperature and relative humidity) [4]

Increasing the number of meteorological data can improve the accuracy as demonstrated in [6]. However, accounting higher number of variables will often complicate the data structure and leads to a longer computational time due to the redundancy of certain variables. For examples of related work in using principal component analysis for solar irradiance forecasting readers can refer to [9] and [10].

While various solar radiation forecast models can be found from the literature, most of the methodologies requires high computational time due to the large number of data used. In order to get the optimum number of meteorological factors, under the constraints of low computational time and quality of information used, Principal Component Analysis (PCA) are used to reduce the dimension of the data and to obtain the dominant meteorological factors affecting solar radiation. The obtained results are expected to be useful for future research work on solar radiation prediction.

3. Methodology

All types of meteorological data available were collected and used. They are relative humidity, wind speed, air temperature, wind gust, and wind. These data were collected from October 2015 to November 2015 from the rooftop of Malaysia-Japan International Institute of Technology (MJIIT) building in Universiti Teknologi Malaysia Kuala Lumpur (UTM KL).

Then, PCA were implemented on these data to reduce the dimension of the data set and identify the principal components. This is done by computing the covariance and correlation between every pair of variables (meteorological factors) using Eq. (1) and (2) respectively [7].

$$\operatorname{cov}(x, y) = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{1}$$

where,

$\operatorname{cov}(x, y)$ = covariance of the variables x and y

x_i = independent variable of sample data

y_i = dependent variable of sample data

n = number of data points in the sample

$$r_{(x,y)} = \frac{\operatorname{cov}(x, y)}{s_x s_y} \tag{2}$$

where,

$r_{(x,y)}$ = correlation of the variables x and y

$\operatorname{cov}(x, y)$ = covariance of the variables x and y

s_x sample standard deviation of the random variable x

s_y sample standard deviation of the random variable y

Basically, the correlation between pairs of variables listed as follow will be calculated

i. Relative humidity versus wind speed
ii. Relative humidity versus air temperature
iii. Relative humidity versus wind gust
iv. Relative humidity versus wind direction
v. Wind speed versus air temperature
vi. Wind speed versus wind gust
vii. Wind speed versus wind direction
viii. Air temperature versus wind gust
ix. Air temperature versus wind direction

x. Wind gust versus wind direction

Upon summarizing the obtained correlation values in a matrix (Eq. 3), the eigenvector and eigenvalue of the matrix are then computed using Eq. (4) and (5).

$$A = \begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{bmatrix} \tag{3}$$

where $A_{i,j}$ is the correction between variable i and variable j computed using Eq. (2).

$$\begin{bmatrix} A_{1,1} & A_{1,2} & \cdots & A_{1,n} \\ A_{2,1} & A_{2,2} & \cdots & A_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n,1} & A_{n,2} & \cdots & A_{n,n} \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \tag{4}$$

Or equivalently,

$$AV = \lambda V \tag{5}$$

The eigenvector V of each variable can be determined by solving the determinant of its characteristic polynomial which can be written as follow

$$(A - \lambda I)V = 0 \tag{6}$$

For eigenvalue (λ), it can be obtained via Eq. (7).

$$p(\lambda) = |A - \lambda I| \tag{7}$$

Through these steps, the principal components can be identified. The obtained eigenvalues are then used to determine the percentage of variability accounted and reduce the dimension of the data. In general, an inclusion of variables that accounts a total of 90 % and above variability are considered reasonably good [8].

4. Results and Discussions

A sample of corrrelation plot of a pair of varible (relative humidity and air temperature) is shown in **Fig. 4**.

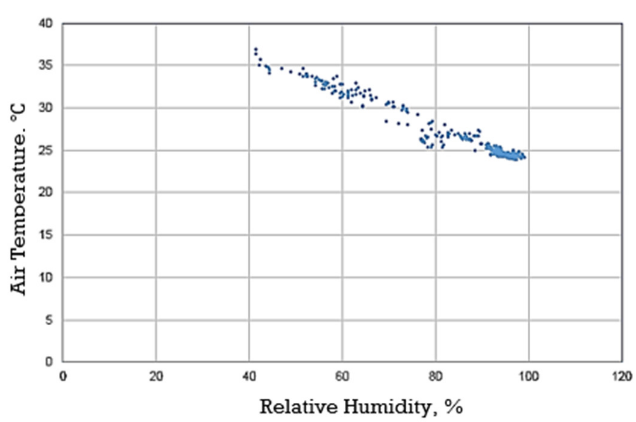


Fig. 4. Correlation plot of relative humidity versus air temperature (correlation = -0.9514)

Based on the correlation analysis between every pairs of variables the following results are obtained as in **Table 1**, while the principal component of multivariate time series meteorological data is illustrated in **Table 2**.

Table 1 Correlation between Input Parameters

	Relative Humidity	Air Temperature	Wind Speed	Wind Direction	Wind Gust
Relative Humidity	1	-0.9514	0.5727	-0.0715	0.6439
Air Temperature	-0.9514	1	-0.5839	0.1052	-0.5908
Wind Speed	0.5727	-0.5839	1	-0.0305	0.6197
Wind Direction	-0.0715	0.1052	-0.0305	1	0.1064
Wind Gust	0.6439	-0.5908	0.6197	0.1064	1

Table 2 Principal component of multivariate time series meteorological data

Variable	Eigenvalue	Percentage of Variability Accounted (%)	Principal Component
Relative Humidity	2.9936	59.872	1st
Air Temperature	0.5513	11.026	3rd
Wind Speed	1.0707	21.414	2nd
Wind Direction	0.0448	0.896	5th
Wind Gust	0.3396	6.792	4th

Graphically, the principal components can be displayed as in **Fig. 5**. **Table 1** and **Fig. 5** shows the percentage of variability accounted for each variable. Variable with the highest percent of variability accounted is the first principal component. Based on **Fig. 4**, it can be seen that the first three components is relative humidity (59.872 %), wind speed (21.414 %), and air temperature (11.026 %). Furthermore, the total variability accounted by these three variables is 92.312 %. Since these first three principal components have accounted for more than 90 %, the rest of two variables can be regarded as redundant variables. The elimination of these two variables accounts a total loss of 7.688 % variability.

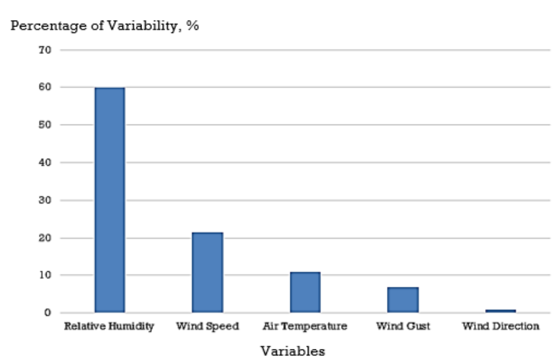


Fig. 5. Principal components

5. Conclusions

PCA were implemented on the meteorological data collected in Universiti Teknologi Malaysia Kuala Lumpur. Based on the analysis, it was found that relative humidity has the most significant contribution on affecting solar radiation. Although using higher number of variables can results in a better forecast accuracy, the findings in this paper found that three variables is sufficient to describe the whole data structure. The results from this research will serve as the starting point for solar radiation prediction using only dominant meteorological factors.

Acknowledgement

The authors would like to thank their family members and friends, for their continuous support and assistance. The authors would also like to express his appreciation to Universiti Teknologi Malaysia and Takasago Thermal Engineering research grant, Vot No. 4B262 for financial support.

References

[1] Volker, Q. “Technology fundamentals of solar thermal power plants.” *Journal of Renewable Energy World* 6 (2003): 109-113.
[2] Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J. and Washom, B. “Intra-Hour forecasting with a total sky imager at the UC San Diego Solar Energy Testbed.” *Solar Energy* 85 (2011): 2881-2893.
[3] Reda, I., Andreas, A. “Solar position algorithm for solar radiation applications.” *Solar Energy* 5 (2004): 577-589.
[4] Taylor, J.W. and Buizza, R. “Using weather ensemble predictions in electricity demand forecasting.” *International Journal of Forecasting* 19 (2003): 57-70.
[5] Ghanbarzadeh, A., Noghrehabadi, A., Assareh, E. and Behrang, M. “Solar radiation forecasting based on meteorological data using artificial neural networks.” In: 7th*IEEE International Conference on Industrial Informatics*, pp. 227-231, 2009.
[6] Mubiru, J. and Banda, E.J.K.B. “Estimation of monthly average daily global solar irradiation using artificial neural networks.” *Solar Energy* 82 (2008): 181-187.
[7] Khattree, R. and Naik, D.N. Multivariate Data Reduction and Discrimination with SAS Software. Cary, N.C., SAS Institute, 2000.
[8] Tymvios, F.S., Jacovides, C.P., Michaelides, S.C., and Scouteli, C. “Comparative study of angstrom’s and artificial neural networks methodologies in estimating global solar radiation.” *Solar Energy* 78 (2005): 752-762.
[9] Federica, D., Stefano, A., Simon, S., Luca, D.M., Davide, A. and Maria, T.V. “Post-processing techniques and principal component analysis for regional wind power and solar irradiance forecasting.” *Solar Energy* 134 (2016): 327-338.
[10] Christina, S. and Wold-Gerrit, F. “Wind forecasting using principal component analysis. *Renewable Energy* 69 (2014): 365-374.