Modelling the Efficiency of Paddy Production in Peninsular Malaysia Using Principal Component Analysis and Data Envelopment Analysis (PCA-DEA)

Aisyah Ismail[1], Zalina Zahid[2], Siti Aida Sheikh Hussin[3], Siti Shaliza Mohd Khairi[4]
[1,2,3,4]*Centre for Statistical and Decision Science Studies, Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, 40450 Shah Alam, Selangor Darul Ehsan, Malaysia.*
[1]aisyahsya.ismail@gmail.com
[2]zalina@tmsk.uitm.edu.my
[3]sitiaida@tmsk.uitm.edu.my
[4]shalizakhairi@tmsk.uitm.edu.my

Abstract- **Although the current paddy production in Peninsular Malaysia shows a satisfying self-sufficiency level, the country's annual rough paddy production, however, will still have to increase over the next 10 years to sustain up with the local population growth and the growing demand level for this staple food. Thus, immediate action needs to be done to evaluate the efficiency of paddy production, so that the information obtained can help the government to come out with strategies to maximize the outputs with the current existing inputs. The objective of this study is to measure the efficiency of paddy production of 11 states in Peninsular Malaysia using a hybrid of Principal Component Analysis and Data Envelopment Analysis. The variables used as inputs include planted paddy area, number of rice farmers, amount of seeds usage and total budget allocation whereas the variables used as outputs cover paddy output, rice output and rice income. As the input and output variables are highly correlated, this study proposes the combination of Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA) approaches to reduce the data dimensionality problem instead of eliminating the variables with multicollinearity problem from the analysis. PCA is applied separately on both sides of all inputs and all outputs, resulting in one principal component (PC) to represent the input and one PC to represent the output side. Both PCs were found to contribute greater than 70% of the data variation. The results from output-oriented DEA model under variable return to scale (VRS) indicate that Kedah is the most efficient state in producing the paddy output, leaving the rest of states with average efficiency scores ranging from 0.745 to 0.998. Further results show that Kelantan, Terengganu and Pahang were placed at the lowest three states in terms of efficiency levels. There is a need for the government to pay extra attention on these states in bringing off the significant factors that may disrupt the functioning of efficient paddy production.**

International Journal of Supply Chain Management
IJSCM, ISSN: 2050-7399 (Online), 2051-3771 (Print)

Keyword- *Efficiency, Paddy Production, Data Envelopment Analysis, Principal Component Analysis, multicollinearity*

1. Introduction

Generally, rice is placed second as the most widely grown cereal crop all around the world and contributes half of the world's population staple food. In a developing country, the people depend on rice for sources of food calories and protein. With the local population increasing in number day by day, the yield progress of the crop needs to grow at least as rapidly as the population, if not faster, to complement the demand. However, nowadays, the paddy production's level in Malaysia still have not managed to surpass the seventy per cent self-sufficiency marks (The Star, 2017), despite of the advances made in rice breeding research and development (R&D) where new high yielding varieties have been brought out to the farmers. Nevertheless, it is an ongoing effort for the government as well as paddy farmers to keep the pace in producing more paddy output with uncertainties of economic production and climate variability. This is because they need to be aware of such precise information regarding annual effective rainfall, runoff, water release policy and consumptive use [1]. Besides, it is crucial to effectively utilize the water resources and to efficiently practice a good management in paddy field to evade the cognitive hurdles in concern to increase the efficiency in paddy production.

Currently, the unfulfilled market local demand is mitigated by importing rice from the neighboring countries such as Thailand, Indonesia, Cambodia, and India. Many measures had been taken into considerations in achieving the primary agenda of most local paddy growers in which to enhance the paddy output. For examples, the rice farmers focus on the presentation of new cultivars, reviewing existing planting practices such as pesticide cycle and fertilizing, quantity and type of pesticide and fertilizer used as well as the frequency and intensity of such cycle [1]. According to the record from International Code Document Centre, the paddy plant places third as the most important crop in Malaysia after the oil palm plant with 2.3 million hectares and the rubber plant with 1.8 million hectares in terms of land utilization in 2002. In addition, there are about 116,000 full time paddy growers that entirely depend on the outcome of paddy yield as an income source [2].

As the projection of rice demand is increasing continuously over time, a sustainable approach of rice production has become essential. Even though paddy production in Malaysia exhibits an increased growth, the rice farmers still face many challenges which are connected to economic, social, technology, and the conformity to field infrastructure. The government of Malaysia targets to achieve the aim of a 100 per cent self-sufficiency level in paddy production in the following 10 years in advance, indicating that an immediate attention needs to be led into considerations to check on the efficiency of paddy production in Malaysia. Various studies [3], [4], [5], [6], [7] have been conducted to explore on the efficiency of paddy production, however, only a few studies [8], [9], [10] attempt to include all the inputs and outputs with high correlations. Instead of including them altogether in the analysis, these studies tend to exclude the variables with multicollinearity problem completely from the study, resulting in total loss of information from the original variables that are not chosen. Therefore, Principle Component Analysis (PCA) is utilized to aggregate data which have multicollinearity problem. This will further increase the discriminatory power of Data Envelopment Analysis (DEA) model in producing efficiency values. Therefore, this study attempts to measure the efficiency performance of 11 states in Peninsular Malaysia using the hybrid of Principal Component Analysis (PCA) and DEA taking consideration of reducing the data dimensionality that may occurs in selected data.

2. Literature Review

2.2 An Overview of Paddy Production in Malaysia

In the global market, Malaysia still cannot compete with the key players in rice producer as the total rice output in the country is only roughly about 0.4 per cent of total world rice output [10]. The decline in the paddy production is due to the impact of the above-average temperatures with the combination of unreasonable dry conditions resulting in reduction of water availability supplies to irrigation areas. In Malaysia, paddy is a commodity that is under protection of the government and is categorized as one of the main commodities in sustaining food production for country's food security. According to ASEAN Food Security Information System, Malaysia stands at the 8th place among the nine selected Asean countries which include Indonesia, Vietnam, Thailand, Myammar, Philipines, Cambodia, Lao

People's Democratic Republic and Brunei. The Malaysia's paddy production was 2 674.4 tonnes in 2015, i.e. equivalent to 1.25 per cent from the total paddy production among the aforementioned nine selected Asean countries. This shows that Malaysia is still in need of the enhancement of efficiency measures in producing more rice yield. As the demand from the population keeps increasing, there is an urge for Malaysia to enhance their production of rice in many ways to ensure the food security, thus can cover up the demand shortage in the country. Rice is generally grown on local granaries all around the Malaysia; taking for about 300 500 hectares land utilization in Malaysia Peninsular and 190 000 hectares in Borneo Islands. The main season paddy planting starts from August to February while harvesting paddy starts from January until end of May [11].

2.3 Data Envelopment Analysis (DEA) for Efficiency Measurement in Paddy Production

Previous researchers [3], [4], [5], [6], [7] defined efficiency in many different perspectives. In general, the ability to produce the lowest cost of output level is signified as efficiency [12]. According to [13], efficiency can be defined as the degree of production performance at maximum level using the similar level of input. On the other hand, the technical efficiency can also be defined as the difference in
the actual and potential output measured when the fixed and variable input is constant under the same level of observation [14].

DEA is a well-known non-parametric approach that has been applied in many areas [15], [16], [17], [18]. It uses a linear programming technique to evaluate the efficiency of homogenous units of assessment. Since the DEA model can adjust with the ratio of the outputs' weighted sum of outputs to the inputs' weighted sum, it can encounter the evaluation of the comparative efficiency of a group of DMUs that come from different scales of inputs. Hence, the outcomes could be interpreted clearly and can assist in making comparison of efficient and inefficient DMUs in data analysis. The empirical applications of DEA are also expanding at an alarming rate in diverse fields [19], [20], [21] and [22], amongst which the agriculture sector is included.

A study by [12] implements DEA method for efficiency measurement in paddy production in Myanmar. In this study, output-oriented variable return to scale model is assumed to be more appropriate to be applied as their study objective is to find the amount of maximum output that can be utilized based on the usage of current inputs. The inputs considered by the researchers are those representing land utilization, labor usage, material costs for production of rice and cost of operation category. The results show that there is a room for improvement in technical of paddy production in Myanmar

as the results drawn that the average technical efficiency is from 31.0 to 37.0 per cent range [12].

Meanwhile, another study by [23] also used DEA variable return to scale model to analyze the production efficiency of paddy in China from 1990 to 2008. Since paddy production can be categorized as a semi-open ecological system in which its paddy production is highly influenced by exogenous factors of environment, Variable Return to Scale (VRS) model is more suitable to come into practice as compared to Constant Ratio to Scale (CRS) model. Through the analysis of the results, it is found that the average paddy production in China scored only 0.806, with stable and significant values from 1990 to 2008. Besides, it also concluded that natural disasters, market price and government's agricultural policy affected the performance of paddy production in China.

2.4 Combination of Principal Component Analysis (PCA) and Data Envelopment Analysis (DEA)

PCA accomplishes the reduction of data dimensionality while keeping most of the variation in the data set. In its mathematical formulation, it identifies directions, known as principal components (PCs) to accomplish the linear combination of dimension reduction, along which the variation in the data is maximal. PCA is a widely used method of dimension reduction in many areas of research. Some previous researchers combined the application of DEA and PCA in their area of study. For an instance, [24] have developed a research framework that comprised of both combinations of PCA and DEA approaches to analyze and rank producing systems. The researchers assessed the producing systems, then ranked them with the proposed method. Their data validity was checked using Numerical Taxonomy (NT). It was found that high correlation existed between the results of DEA/ PCA and NT based on Pearson correlation analysis, thus strengthening the findings of this study.

[25] also implemented the merging application of PCA/ DEA model to check on the performance measurement of e-banking. The data used for this study were taken from 2007 annual reports of the main banks in the USA and the UK, which included both financial and non-financial variables. The DEA model produced the efficiency scores for each bank and PCA was used to rank the banks' performance. The results showed that most giant banks had good performance as compared to small banks. Meanwhile, [26] combined PCA and DEA in modeling the efficiency of e-banking for main players' banks in USA and UK banking sector. The results suggested that bank staffs were the most important variable in determining higher returns in e-banking services. Lastly, [27] employed PCA-DEA model that integrated different warehouse and transport indicators for a study to

assess the efficiency of distribution centers of a trading company in Serbia. It was found that model that emphasized with quality showed better results and smaller scale distribution centers tended to be more efficient than the large one. The findings from this study enabled the top managers group of the company to make a better decision as well as improving the company's performance.

3. Methodology

3.2 Research Framework

This study employs the PCA-DEA approach in modeling the efficiency of paddy production in Peninsular Malaysia. The calculation of efficiency measurement of the paddy production in Malaysia is executed using the controlled variables under the DEA method. Meanwhile, PCA method is used to aggregate input and output data which are highly correlated to increase the accuracy of the DEA results.

The data gathered from several reliable resources [11] are used as part of their input and output in the process of determining the efficiency score of paddy production. The range of efficiency scores are between 0 to1. The value of 1 indicates the state is efficient and the state is not efficient when its efficiency is less than 1. The research framework is shown in the Figure 1.

3.3 Identification of Decision Making Units

The number of DMUs follows the rule of thumbs which consider the total number of DMUs must be at least twice times with the total number of inputs and output [32]. This rule of thumbs attempts to make sure that the basic productivity models are more discriminatory. If the discriminatory power reduces due to the few numbers of DMUs, the number of input and output factors can either be reduced or turned them to a different productivity model that has more discriminatory power.

Firstly, 11 states under Peninsular Malaysia are chosen as the decision-making units (DMUs) of this study. Secondly, certain data are chosen to represent input and output variables from 2004 to 2017, based on theory of production and data availability. Thirdly, the inputs and outputs are aggregated using PCA, based on certain criteria. Next, DEA model is used to measure the efficiency level of paddy production of each state. Lastly, the states are ranked based on their efficiency measurements.

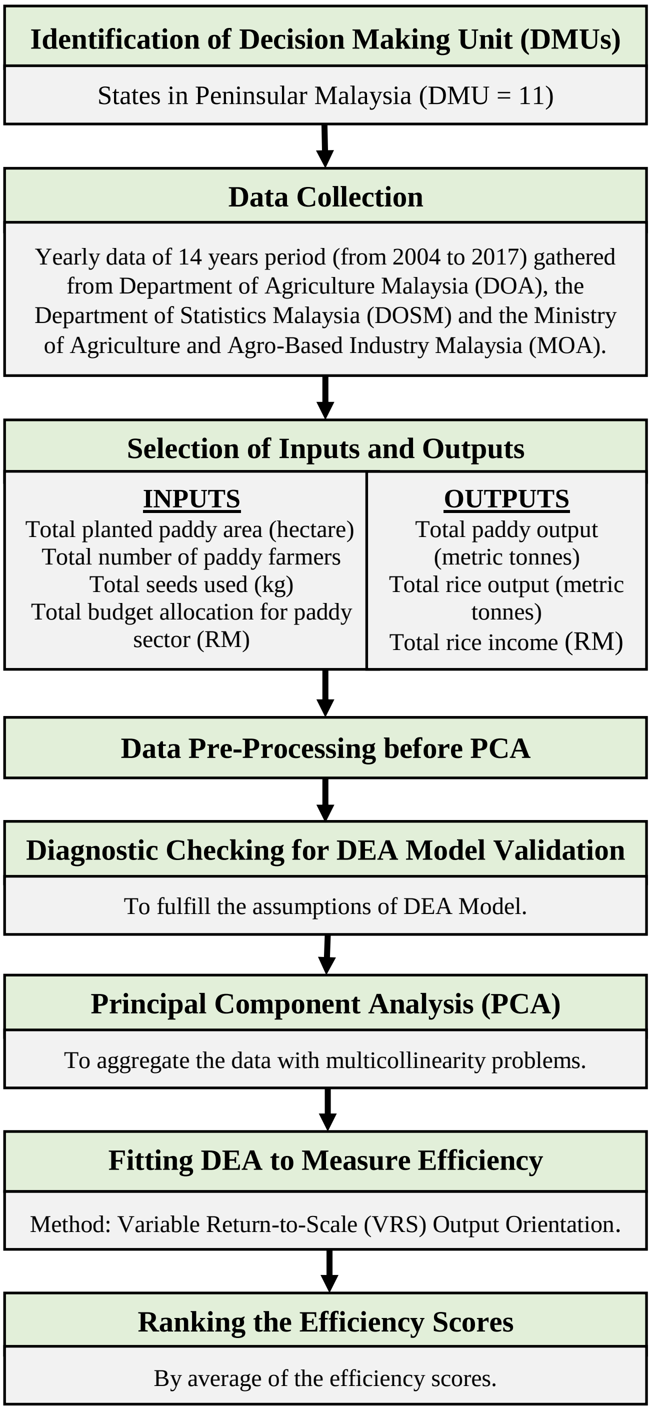

Figure 1. Research Framework

Secondary data used in this are obtained from the database resources referring to several government agencies such as the Department of Agriculture Malaysia (DOA), the Department of Statistics Malaysia (DOSM) and the Ministry of Agriculture and Agro-Based Industry Malaysia (MOA). These data cover 11 states in Peninsular Malaysia from period of 2004 till 2017. Yearly data for paddy refers to data reported in two seasons, i.e. Main Season and Off Season. For example, data for 2014 refers to the Main Season 2013/2014 and the Off Season 2014. The descriptions of inputs and outputs are portrayed in the Table 1.

3.4 Data Collection

This research is using secondary data which are generated from the database resources from several government agencies such as the Department of Agriculture Malaysia (DOA), the Department of Statistics Malaysia (DOSM) and the Ministry of Agriculture and Agro-Based Industry Malaysia (MOA). All the data collections are based on the 11 states in Peninsular Malaysia for 14 years from period of 2004 until 2017.

Yearly data for paddy refers to data reported in both seasons which are the Main Season and the Off Season. For example, data for 2014 refers to the Main Season 2013/2014 and the Off Season 2014. This concept applies to the data related to paddy production; total planted paddy area. In addition, the planting is defined as the process of sowing for direct seeding system and transplanting for nursery system.

Table 1. Descriptions of Inputs and Outputs

Output Variables	**Unit**	**Description**
Output Variables		
Total paddy output, Y_1	metric tonnes (mt)	Total paddy output (raw) from the overall paddy production for all seasons throughout the year.
Total rice output, Y_2	metric tonnes (mt)	Total rice quantity after processing of all the production throughout the year.
Total rice income, Y_3	Ringgit Malaysia (RM)	Total rice income from the overall paddy production for all seasons throughout the year.
Output Variables	**Unit**	**Description**
Input Variables		
Total planted paddy area, X_1	hectare (ha)	Total agriculture land that is used for paddy growing; main seasons and off season throughout the year.
Total no. of paddy farmers, X_2	Person	Total farmers who are engaged in paddy plantation sector throughout the year.
Total seeds used, X_3	kilogram (kg)	Total number of rice seeds sown in the paddy field throughout the year.
Budget allocation for paddy sectors, X_4	Ringgit Malaysia (RM)	Total budget and incentive allocation from federal government to the paddy farmers based on the wide area of paddy plantation.

3.5 Selection of Inputs and Outputs

The selection of inputs and outputs is being done based on the Theory of Production [28] and data availability. Under the production industry, economists described the factors of productions as the use of resources or inputs along the production processes. Four group categories are identified in this study, which are raw materials, labor services, capital goods and land as according to the Theory of Production.

Planted paddy area is defined as variables in land category [29], [30], [12] while amount of seeds used is part of raw materials category [30]. Next, for annual budget allocation in capital goods category, this variable is also supported by [29] as one of the input variables to analyze the efficiency of paddy production in Malaysia. Lastly, the number of workers is used to represent labor services category become one of the controlled variables in assessing the paddy production efficiency [29], [30], [31].

3.6 Data Pre-Processing before Principal Component Analysis (PCA)

Data pre-processing involves data cleaning, data integration and transformation and data reduction. Data pre-processing is needed to check missing attribute values in the data, identification of outliers or errors and inconsistency in the raw data so that incorrect or even misleading statistics can be avoiding. Since PCA requires data to be free from outlier, outlier detection is conducted followed by outlier treatment using winsorization technique if outlier exits in any of the input and output variables. To solve the outlier problem, one of the ways is implement winsorization technique also known as trimming which is a method to reduce the effect of outliers in data (Traskin *et al*.,2012). This technique is applicable when there exist outliers in unitivariate data as it can ensure that the mean estimator is unbiased. Based on the Real Statistic, the outliers that are highest or lowest value will be trimmed by replacing it with value of selected percentile, for example 5th percentile, 10th percentiles, etc. Winsorization can be applied as much or as little as seems appropriate.

Then, Correlation Analysis is performed to check on the correlations of the variables involved. Correlation analysis is a technique to assess the relationship between an outcome variable and one or more confounding variables. Sample correlation coefficient or more specifically Spearmann's Rank Correlation coefficient which is also denoted as r is used in correlation analysis. The value ranges from -1 to +1 between the two variables. The correlation between two variables can be positive or negative which indicates the direction of the association which mean negative or positive relationship between two variables. There is a positive strong relationship between the two variables if the value of r is near +1.

3.6.1 Data Descriptive of Input and Output Variables

Table 2 displays the descriptive statistics of input and output variables from 2004 to 2017. The values of mean, median, standard deviation, minimum and maximum are being tabulated for every variable as in Table 2. Within the period of analysis, the averages for total paddy outputs and total rice outputs are 205,697.77 and 133,750.66 metric tonnes respectively. As for rice income, its average value is equivalent to RM2,262,052,122.37. Meanwhile, on the input sides, the average for the planted paddy area is 46,846.86 hectares which is cultivated by the average number of 11,577.28 rice farmers. On average, the number of seeds used in paddy planting is 27,028,294.23, under the average amount of budget allocation equals to RM98,858,580.29.

Table 2. Summary of Descriptive Statistics of Input and Output Variables (2004-2017)

Variables	**Mean**	**Median**	**Std. Deviation**	**Minimum**	**Maximum**
Output Variables					
Paddy output	205,697.77	142,598.00	264,367.33	2,791.00	1,280,759.00
Rice output	133,750.66	92,689.00	171,825.81	1,814.00	832,494.00
Rice income	2,262,052,122.37	1,227,531,800.00	3,173,591,860.89	16,634,360.00	18,858,293,657.51
Input Variables					
Planted paddy area	46,846.86	25,564.00	59,590.35	1,095.00	228,868.00
Paddy farmers	11,577.28	7,311.50	12,273.60	150.00	58,476.00
Seeds used	27,028,294.23	14,749,149.80	34,380,653.09	631,760.25	132,045,392.60
Budget allocation	98,858,580.29	53,946,431.00	125,750,538.50	2,310,723.75	482,968,697.00

3.7 Principal Component Analysis

In this study, PCA is applied separately on both sides of inputs and inputs, as the intercorrelations within the input variables and output variables are high. Through the use of

PCA, the loss of information is minimized as the decision maker does not require to remove the variables completely from the analysis or specify which variables are more important.

Let the random vector $\boldsymbol{X}' = [X_1, X_2, \dots, X_p]$ have the covariance matrix $\boldsymbol{\Sigma}$ with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$. Consider the i^{th} principal component can be presented as follows:

$$Y_1 = \boldsymbol{a'}_1\boldsymbol{X} = a_{11}X_1 + a_{12}X_2 + \cdots + a_{1p}X_p \quad (1)$$
$$Y_2 = \boldsymbol{a'}_2\boldsymbol{X} = a_{21}X_1 + a_{22}X_2 + \cdots + a_{2p}X_p$$
$$\dots$$
$$Y_p = \boldsymbol{a'}_p\boldsymbol{X} = a_{p1}X_1 + a_{p2}X_2 + \cdots + a_{pp}X_p$$

where Y_i is the linear combinations for every i^{th} component, a_{ik} is the coefficient of i^{th} variable for the k^{th} principal component and X_p is the indicators under input and output dimension. Consider these principal components have:

$$Var(Y_i) = \boldsymbol{a'}_1\boldsymbol{\Sigma}\boldsymbol{a}_i = \lambda_i \text{ and } Cov\ (Y_i, Y_k) = \boldsymbol{a'}_i\boldsymbol{\Sigma}\boldsymbol{a}_k = 0 \quad (2)$$

where $= 1,2, \dots, p$, $i, k = 1, 2, \dots, p$ and $i \neq k$. The principal components are the uncorrelated linear combinations $Y_1, \dots, Y_p$ where their importance are ranked according to their variances following decreasing manner. In this study, the PCA analysis is computed using XLSTAT Excel add-in software.

3.7.1 *Data Pre-Processing for PCA*

Before PCA implementation, there are list of assumptions that need to be fulfilled first. Further discussions on the calculation for every assumption will be included in the next subsections. The data pre-processing steps before PCA application are summarized in the Table 3.

(a) Outliers Treatment

Table 4 shows the details on the specification of winsorization technique for every variable involved. Since the aim is to get the least number of trimmed data with fairly minor corrections for each variable, the data set is arranged in ascending order to see how much values that need to be corrected. By doing that, it is observed that the bottom part of the data need to be modified following the closest value from reference state and year as stated in the table.

The percentage of winsorization referred to the number of observations that are being modified per overall observation in that column variables. In this study, the total

observations per column are 156, such that the data consist of 11 states with 14 years period of study duration.

Table 3. Data Pre-Processing for PCA Validation

Validated Aspects	**Descriptions**
Outliers detection	There must be no outliers before conducting PCA as the presence of outliers can affect interpretations arising from PCA [32]. After separately checked for inputs and inputs, there is no outlier exists for inputs meanwhile, winsorization technique is applied to treat outliers [33] in output variables.
Adequate correlation and linearity	Data reduction is conducted if there are high intercorrelations between input and output variables, since only a few components are required to represent the original variables [32]. Using both Pearson Correlation and Spearman's Rank Order Correlation on inputs and outputs, all variables indicate positive linear correlation, therefore validating the assumptions of PCA.
Data reliability	Cronbach's Alphas based on Standardized Items must be high for both inputs and outputs, then the internal consistency of them are relatively acceptable for all years of study period.

Table 4. Winsorization Technique to Treat Outliers

Variables	**Winsorization (%)**	**Year of Modified Data**	**Reference State/Year**	**Reference Value**
Total paddy output (mt)	2.60	2014, 2015, 2016, 2017	Kedah/ 2009	923,666
Total rice output (mt)	2.60	2014, 2015, 2016, 2017	Kedah/ 2009	600,383
Total rice income (RM)	1.95	2015, 2016, 2017	Kedah/ 2009	989,0199,000

(b) Correlation Analysis for Input and Output Variables

Correlation analysis under this section explains the relationship between the variables within inputs and within outputs using both Pearson Correlation and Spearman's Rank Order Correlation after the outliers' treatment. Table 5 and Table 6 show the Pearson's Correlation and Spearman's Rank Order Correlation respectively for input variables while Table 7 and Table 8 for output variables. All variables

below indicate high positive correlations within the variables validating the assumptions of PCA method.

Table 5. Pearson's Correlation for Input Variables

Variable		Input			
		X_1	X_2	X_3	X_4
Input	X_1	1			
	X_2	0.841	1		
	X_3	1.0000	0.841	1	
	X_4	1.0000	0.841	1.0000	1

Table 6. Spearman's Rank Order Correlation for Input Variables

Variable		Input			
		X_1	X_2	X_3	X_4
Input	X_1	1			
	X_2	0.912	1		
	X_3	1.0000	0.912	1	
	X_4	1.0000	0.912	1.0000	1

Table 7. Pearson's Correlation for Output Variables

Variable		Output		
		Y_1	Y_2	Y_3
Output	Y_1	1		
	Y_2	1.00	1	
	Y_3	0.905	0.905	1

Table 8. Spearman's Rank Order Correlation for Output Variables

Variable		Output		
		Y_1	Y_2	Y_3
Output	Y_1	1		
	Y_2	1.00	1	
	Y_3	0.947	0.946	1

(c) Consistency Checking

Table 9 shows the overall reliability statistics for the input and output variables from the year 2004 to 2017, traced out by output result from SPSS. The Cronbach's Alpha value that is below 0.70 is considered as having a low internal consistency. Since the Cronbach's Alphas Based on Standardized Items are 0.979 and 0.978 respectively for both inputs and outputs, the internal consistency of them are relatively acceptable for all years of study period. However, the maximum expected value for Cronbach's Alpha to test for internal reliability is 0.90, reflecting that any value above 0.90 perceived as redundancy or duplication. This indicates that these two data sets may face dimensionality issues because of too high Cronbach's Alpha values that will be catered by the application of PCA.

Table 9. Reliability Analysis for Input and Output Variables

Reliability Statistics		Variables	
		Input	Output
Measure	Cronbach's Alpha	0.450	0.000
	Cronbach's Alpha Based on Standardized Items	0.979	0.978

3.7.2 Data standardization before PCA computation

Standardization is a process of variables' transformation to make it easier for data interpretation and to ensure that all variables will contribute evenly to a scale for analysis purposes since there are many different units of analysis used in this study. Otherwise, the results may become infeasible due to the substantial differences in the unit of measurement, for examples; metric tonnes, hectare and kilogram. According to [34], standardization is the process where the data are standardized with respect to its mean (μ_P) and standard deviation (σ_P). The new value of standardize (Z) of the original data is calculated as:

$$Z_p = \frac{(X_P - \mu_P)}{\sqrt{\sigma_{PP}}} \quad (3)$$

3.8 Data Envelopment Analysis

One of the most well-known methods in estimating the efficiency measurement is by using the Data Envelopment Analysis (DEA) as it is suitable for small sample size and requires a minimal number of data [33]. Moreover, DEA is a non-parametric approach which is compatible for multiple uses of input and output as part of consideration in comparing the efficiency level between models [35]. The efficiency measurement in this study is to indicate the efficiency in paddy production by finding the total input contributing towards the production against the output received. For DEA modelling, the technical efficiency score estimates by using the controlled variable input and output. The general model of efficiency as in Eq. (4).

$$\text{Efficiency} = \frac{\text{sum of weighted outputs}}{\text{sum of weighted inputs}} \quad (4)$$

This equation can be written as:

$$\text{Efficiency} = \frac{\sum_{r=1}^{n} u_r * y_r}{\sum_{i=1}^{m} v_i * x_i} \quad (5)$$

where n is the number of outputs for each dimension, m is the number of inputs for each dimension, y_r is output gained by paddy production r, u_r is the weight of output for paddy production r, x_i is the quantity used by input, i and v_i is the weight of input used, i.

3.8.1 Diagnostic Checking for DEA Model Validation

The assumptions of variables' correlations and intercorrelations as well as for the DMUs were checked to validate the DEA model. Table 10 summarizes the validation done for DEA model application.

Table 10. Summary of Validation for DEA Model

Validated Aspects	Descriptions
Isotonic assumption	There must be a positive correlation between inputs and outputs [36].
Multi-collinearity assumption	There must be less than 0.8 for correlation within inputs and within outputs or otherwise, one of the variables need to be removed since it represented similar dimension [37]. Instead of eliminating the variables, Partial Component Analysis is applied to replace the original variables with minimal loss of information [9].
Number of DMUs	After the PCA application separately to both sides; inputs and outputs, DMUs number followed the rule of thumb [38]; Number of DMUs ≥ 2 (Number of inputs + outputs); 11 ≥ 2 (1+1).

(a) Isotonicity Assumption of Correlation between Inputs and Outputs

The results of the correlation between all the variables of input and output showed that there exist positive correlations since all the rho values are greater than 1. Moreover, the high r values indicates that there exists high correlation between inputs and outputs thus consequently validating the isotonicity assumption in this study. High correlations between input and output variables are also crucial in finding the significant relationships between inputs and outputs in DEA model. Table 11 shows the correlation, *r* value for each of the correlated variables involved.

Table 11. Spearman's Rank Order Correlation between Input and Output Variables

Variable		**Input**			
		X_1	**X_2**	**X_3**	**X_4**
Output	**Y_1**	0.9808	0.9055	0.9808	0.9808
	Y_2	0.9809	0.9056	0.9808	0.9808
	Y_3	0.9072	0.8664	0.9072	0.9072

(b) Multicollinearity Assumption of Correlation between Inputs and Inputs and between Outputs and Outputs

Table 12 shows the correlation within the input variables meanwhile Table 13 presents the rho values within the variables in input. The results of Spearman's rank order correlation test portray high correlations within variables from both inputs and outputs. This is because of the nature of the data themselves that depended on the respective paddy
area in that particular place and are each variable are extremely related to each other in terms of environmental and geographical perspective. This indicated that there exist a serious multicollinearity problem in the analyzed data.

Table 12. Spearman's Rank Order Correlation within Input Variables

Variable		**Input**			
		X_1	**X_2**	**X_3**	**X_4**
Input	**X_1**	1			
	X_2	0.9117	1		
	X_3	1.0000	0.9117	1	
	X_4	1.0000	0.9117	1.0000	1

Table 13. Spearman's Rank Order Correlation within Output Variables

Variable		**Output**		
		Y_1	**Y_2**	**Y_3**
Output	**Y_1**	1		
	Y_2	0.9999	1	
	Y_3	0.9466	0.9465	1

3.8.2 Fitting DEA VRS Output Orientation to Measure Efficiency

In this study, VRS output orientation DEA model is developed in modeling the efficiency of paddy production

[12], [23] across 11 states in Peninsular Malaysia for 14 years period. The output orientation can be defined as input level is being kept fixed while finding the maximum proportion of output production level [39]. The output orientation can also be called as output maximization as the output level will be varied with constant number of inputs.

Under the specification of VRS, the model can tolerate with an environment which having an operation below optimal scale [9]. For an instance, the paddy production will be affected when the weather is not in good condition, thus VRS model will give some modification for the constraint mentioned above. Furthermore, VRS is known as the non-constant returns to scale result in a non-proportionate change which is increase or decrease in the outputs [40]. Hence the VRS for DEA model can be written as follows:

$$min\left[\theta - \varepsilon\left(\sum_{i=1}^{m} s_i^- + \sum_{k=1}^{r} s_k^+\right)\right] \quad (6)$$

$$subject\ to$$

$$\sum_{j=1}^{n} \lambda_j x_{ij} + s_i^- = \theta x_{iq}$$

$$\sum_{j=1}^{n} \lambda_j x_{kj} + s_i^+ = y_{kq}$$

$$\sum_{j=1}^{n} \lambda_j = 1$$

$$\lambda_j \geq 0, s_k^+ \geq 0, s_k^- \geq 0,$$

where θ is the efficiency score $(0 < \theta < 1)$, j is the states in Peninsular Malaysia, x and y refer to input and output

variables respectively, n is the DMUs' number $(n = 1, 2, \ldots, 11)$, λ_j is the weight attached to inputs and outputs, s^- is a slack for input and s^+ is a slack for output. A high efficiency score will be as close to 1, meanwhile, a lower score of θ can be said that the regions have low efficiency in producing paddy. The calculation of technical efficiency of paddy production in this paper is conducted by using DEAP 2.0 software.

4. Results

PCA is applied separately on both sides of all inputs and outputs, as there exist high correlations within the groups of input variables and output variables. Table 14 displays the values of PCs under the inputs. Meanwhile, Table 15 shows the details of PCs for the outputs. The results from these tables indicate that only one PC are required in each case to explain data variation as its percentage of contribution in explaining the data variation is more than 70%. Thus, PCs with low percentage of contribution for correlation explained are dropped from the analysis, resulting in minimal loss of information for the reduced model. However, some of the PCs are negative, indicating that a modification need to be done before executing DEA model. This is because the input and output data cannot be negative in DEA application. Therefore, these negative of PCs have to be converted into positive data before DEA is applied. The following equations [41] are used to convert the data from Table 14 and Table 15:

$$Z_j = PC_j + Q \quad (7)$$

$$\text{where } Q = -min\{PC_j\} + 1 \quad (8)$$

Table 14. Principal Component Analysis for Inputs

INPUTS	Year													
	2004		2005		2006		2007		2008		2009		2010	
	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2
Correlation explained (%)	**93.132**	6.868	**93.597**	6.403	**93.955**	6.045	**94.006**	5.994	**90.719**	9.281	**94.020**	5.980	**91.345**	8.655
Planted area	0.513	-0.265	0.512	-0.267	0.511	-0.268	0.511	-0.268	0.519	-0.254	0.511	-0.268	0.517	-0.257
Farmers	0.458	0.889	0.462	0.887	0.464	0.886	0.465	0.885	0.439	0.898	0.465	0.885	0.445	0.896
Seeds	0.513	-0.265	0.512	-0.267	0.511	-0.268	0.511	-0.268	0.519	-0.254	0.511	-0.268	0.517	-0.257
Budget	0.513	-0.265	0.512	-0.267	0.511	-0.268	0.511	-0.268	0.519	-0.254	0.511	-0.268	0.517	-0.257

INPUTS	Year													
	2011		2012		2013		2014		2015		2016		2017	
	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2
Correlation explained (%)	**93.691**	6.309	**90.726**	9.274	**92.616**	7.384	**91.184**	8.816	**99.819**	0.181	**99.899**	0.101	**99.795**	0.205
Planted area	0.512	-0.267	0.519	-0.254	0.514	-0.262	0.518	-0.256	0.500	-0.288	0.500	-0.288	0.500	-0.288
Farmers	0.462	0.887	0.439	0.898	0.455	0.891	0.443	0.896	0.499	0.867	0.499	0.866	0.499	0.866
Seeds	0.512	-0.267	0.519	-0.254	0.514	-0.262	0.518	-0.256	0.500	-0.288	0.500	-0.288	0.500	-0.288
Budget	0.512	-0.267	0.519	-0.254	0.514	-0.262	0.518	-0.256	0.500	-0.288	0.500	-0.288	0.500	-0.288

Table 15. Principal Component Analysis for Outputs

OUTPUTS	Year													
	2004		2005		2006		2007		2008			2009		
	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_3	PC_1	PC_2	PC_3
Correlation explained (%)	**99.981**	0.019	**99.778**	0.222	**99.180**	0.820	**99.684**	0.316	**99.615**	0.382	0.002	**99.438**	0.561	0.000
Paddy output	0.577	-0.408	0.578	-0.408	0.579	-0.407	0.578	-0.408	0.578	-0.454	-0.678	0.578	-0.409	-0.706
Rice output	0.577	-0.408	0.578	-0.408	0.579	-0.407	0.578	-0.408	0.578	-0.359	0.733	0.578	-0.406	0.708
Rice income	0.577	0.817	0.577	0.817	0.575	0.818	0.576	0.817	0.576	0.815	-0.055	0.576	0.818	-0.002

OUTPUTS	Year															
	2010		2011		2012		2013		2014		2015		2016		2017	
	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2	PC_1	PC_2
Correlation explained (%)	**97.882**	2.118	**99.932**	0.068	**97.389**	2.611	**99.823**	0.177	**97.697**	2.303	**95.045**	4.955	**99.180**	0.820	**94.342**	5.658
Paddy output	0.581	-0.404	0.577	-0.408	0.581	-0.403	0.578	-0.408	0.581	-0.403	0.585	-0.397	0.579	-0.407	0.586	-0.395
Rice output	0.581	-0.404	0.577	-0.408	0.581	-0.402	0.578	-0.408	0.581	-0.403	0.585	-0.397	0.579	-0.407	0.586	-0.395
Rice income	0.571	0.821	0.577	0.817	0.569	0.822	0.577	0.817	0.570	0.821	0.561	0.828	0.575	0.818	0.559	0.829

Table 16 presents the efficiency scores of Peninsular Malaysia states across 4 years (from 2004 to 2017) in terms of paddy production. Using the modified PC values for both inputs and outputs, the PC data are evaluated by DEA to produce the efficiency values. Efficiency scores are ranged from 0 to 1, with less than 0 represents inefficient states and 1 represents efficient state. It appears that Kedah seems to be efficient across 14 years of observations, implying that it is the most efficient state as compared to other states. The rankings of all eleven states in Peninsular Malaysia are based on the average of the efficiency scores across 14 years. Table 7 shows that Kedah has the highest average efficiency scores, followed by Negeri Sembilan that is placed second with thirteen times of appearance with efficiency scores of 1 throughout the 14 years period. Meanwhile, Pulau Pinang has an average efficiency score of 0.989, appeared with an efficiency score of 1 eleven times, leaving Selangor with only five times appearances. Another group of states, Melaka, Johor and Perlis have the average efficiency scores of 0.976, 0.975 and 0.932 respectively. Finally, the last four inefficient states namely, Perak, Pahang, Terengganu and Kelantan share the same place in term of no appearance of perfect efficiency scores at all. It seems that Kelantan exhibited the lowest overall average efficiency level across fourteen years as it consistently has low efficiency performance as compared to other states.

Table 16. Efficiency Scores for Paddy Production of Peninsular Malaysia states for the 14 years (2004 – 2017)

DMU	Efficiency Score (VRS)														Average Efficiency Score	Appearance of '1' Efficiency Score (times)	Rank
	Year																
	2004	**2005**	**2006**	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**	**2017**			
JOHOR	0.975	0.966	0.941	0.959	**1.000**	0.964	0.966	0.969	0.965	0.972	0.964	0.983	0.956	0.990	0.975	1	5
KEDAH	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	1.000	14	1
KELANTAN	0.688	0.646	0.717	0.696	0.981	0.706	0.708	0.747	0.704	0.681	0.665	0.812	0.753	0.931	0.745	0	11
MELAKA	**1.000**	**1.000**	**1.000**	0.981	**1.000**	0.971	0.967	0.993	0.972	0.976	0.973	0.940	0.960	0.932	0.976	4	4
N. SEMBILAN	**1.000**	**1.000**	**1.000**	**1.000**	0.978	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.998	13	2
PAHANG	0.901	0.888	0.847	0.816	0.886	0.818	0.829	0.807	0.802	0.768	0.755	0.723	0.705	0.762	0.808	0	9
PERAK	0.743	0.746	0.749	0.778	0.788	0.843	0.807	0.860	0.921	0.921	0.857	0.751	0.741	0.828	0.810	0	8
PERLIS	0.958	0.979	0.856	0.898	**1.000**	0.941	0.827	0.969	0.950	**1.000**	0.927	0.994	0.754	**1.000**	0.932	3	6
PULAU PINANG	0.980	0.991	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	**1.000**	0.879	**1.000**	0.989	11	3
SELANGOR	0.898	0.910	**1.000**	0.831	0.759	0.818	**1.000**	0.907	0.950	0.911	0.939	**1.000**	**1.000**	**1.000**	0.923	5	7
TERENGGANU	0.843	0.767	0.773	0.724	0.820	0.709	0.735	0.736	0.721	0.734	0.723	0.759	0.790	0.735	0.755	0	10

5. Discussion and Conclusion

This study is focused on the computation of efficiency level of paddy production for each Peninsular Malaysian state using the combination of PCA-DEA method. PCA is applied to reduce the curse of dimensionality when the variables considered as inputs and outputs are highly correlated within each other. With the use of four inputs and three inputs, PCA summarized them parsimoniously to one PC for each input side and output side. Even though when the number of principal components is less than the number of original variables may cause in the loss of some of the explanatory powers of the data, PCA is able to improve the discriminatory power of the model instead. This is done by looking for a few linear combinations which can be used to summarize the data with losing as little information as possible along the process. Throughout the implementation of PCA-DEA method in this study, it is clearly noticeable that this method can improve the discriminatory power of the model to evaluate the efficiency scores. This is because all available inputs and outputs are being taken into considerations without eliminating neither one of them.

The findings also show that Kedah is the most efficient state in paddy production efficiency. This is supported by the fact of MADA; the largest granary area in Malaysia, is located there in Alor Setar, Kedah. It is known that MADA implements the reticulation system, enabling the farmers to obtain a sufficient supply of water throughout the year. This factor might be one of the reason the state of Kedah can maintain a satisfactory level of outputs with the current inputs utilization. As for Kelantan, which was found to be the least efficient state in terms of paddy production, its paddy fields may have been affected by the climatic conditions. For an instance, in 2014, Kelantan encountered the worst flood ever in history (The Star, 2014), which has caused the disruption in the agriculture sector especially to the paddy crop.

It is worthwhile to be concerned that the three lowest states in efficiency level are states from East Coast Region of Peninsular Malaysia; i.e. Kelantan, Terengganu and Pahang. This is probably due to the impact of the East Coast monsoon, which occurs between mid-October and the end of March every year. The weather during these periods of time are usually very rough, with heavy rain and affect badly on the paddy farmers in the region involved. Thus, the government as well as farmers can take the best action to adapt with uncertain weathers that strikes certain states in our country from this study findings. Moreover, at farm level, farmers must always take all the precautions and be aware about the uncertainty of low and heavy rainfall besides maintaining the financial management to maximize the rice yield in that particular farm. It is widely acceptable that Malaysia urgently requires a proper mitigation strategy to hinder the risks of unexpected environmental factors such as

devastating deluge or extreme rising in temperature in the country, hence securing the 100.00 per cent self-sufficient level of rice output in the future.

6. Recommendations

In this research, only controllable inputs are considered under the application of PCA-DEA method. Future researchers may include uncontrollable factors in the second stage of DEA to get a better understanding of the factors affecting the efficiency level of paddy production in Malaysia. In addition, the researchers can also use other performance evaluation methods and compare the results. This will give more confirmation in the available findings and provide more information to the government to strategize a plan to increase the efficiency of paddy cultivation in our country. From the findings in this study, it is advisable to take serious measure on the flood events that occurred in the East Coast Region of Peninsular Malaysia. Continuous research needs to be done to find the solution in this matter so that the irrigation areas in our country can keep increasing the paddy yield. With the expectation of facing the deleterious global effect of climate in the future, adaptation strategies should be undertaken to improve the irrigation efficiency and to increase the cultivation output.

Acknowledgments

The authors would like to thank Universiti Teknologi MARA (UiTM) for financially supporting this research under the LESTARI Research Grant Scheme, Code: 600-IRMI/MyRA 5/3/LESTARI (0151/2016).

References

[1] Toriman, M. E., Er, A. C., Lee, Q. Y., Mastura, S. S., Jali, F. M., Mokhtar, M., & Ahmah, H. Paddy production and climate change variation in Selangor, Malaysia. *Asian Social Science*, Vol 9, No. 14, pp. 55-62, 2013.

[2] Alam, M. M., Siwar, C., Molla, R. I., & Talib, B. Socioeconomic impacts of climatic change on paddy cultivation: an empirical investigation in Malaysia. *Journal of Knowledge Globalization*, Vol 3. No.2, pp. 71-84, 2010.

[3] Dhungana, B.R., P.L. Nuthall, & G.V. Nartea, Measuring the economic inefficiency of Nepalese rice farms using data envelopment analysis. *The Australian Journal of Agricultural and resource Economics*, Vol 48, No. 2, pp. 347-369, 2004.

[4] Rahman, K. M. M. Measuring efficiency of producing rice in Bangladesh: a stochastic frontier analysis. *Wissenschaftsverlag Vauk Kiel KG*, 2002.

[5] Wadud, A., & B. White, Farm household efficiency in Bangladesh: a comparison of stochastic frontier and DEA methods. *Applied Economics*, Vol 32, pp. 1665-1673, 2000

[6] Abdulai, A., & W. Huffman, Structural adjustment and economic efficiency of rice farmers in Northern Ghana. *Economic Development and Cultural Change*, Vol 48, No. 3, pp. 503-520, 2000.

[7] Audibert, M, Technical inefficiency effects among paddy farmers in the village of the "Office du Niger", Mali, West Africa. *Journal of Productivity Analysis*, Vol 8, pp. 379-394, 1997.

[8] Ueda, T., & Hoshiai, Y, Application of principal component analysis for parsimonious summarization of DEA inputs and/or outputs. *Journal of the Operations Research Society of Japan*, Vol 40, No. 4, pp. 466-478, 1997.

[9] Adler, N., & Golany, B, Pca-Dea. In *Modeling data irregularities and structural complexities in data envelopment analysis*, Springer, Boston, MA, pp. 139-153. 2007.

[10] Papademetriou, M. K, Rice production in the Asia-Pacific region: issues and perspectives. *Bridging the rice yield gap in the Asia-Pacific region*, pp. 4-26, 2000.

[11] Department of Statistics Malaysia, Official Portal, https://www.dosm.gov.my/v1/, Last access of visit (23-03-2017).

[12] Tun, Y., & Kang, H. J, An analysis on the factors affecting rice production efficiency in Myanmar. *Journal of East Asian Economic Integration*, Vol. 19, No. 2, pp. 167-188, 2005.

[13] Pascoe, S., & Tingley, D, Capacity and technical efficiency estimation in fisheries: parametric and non-parametric techniques. In *Handbook of Operations Research in Natural Resources* (pp. 273-294). Springer US, 2007.

[14] Tsitsika, E. V., Maravelias, C. D., Wattage, P., & Haralabous, J, Fishing capacity and capacity utilization of purse seiners using data envelopment analysis. *Fisheries Science*, Vol 74, No. 4, pp. 730-735, 2008.

[15] Rahman, R., Zahid, Z., Khairi, S. S. M., & Hussin, S. A. S. Modeling technical efficiency of inshore fishery using data envelopment analysis. In *AIP Conference Proceedings*, Vol. 1782, No. 1, AIP Publishing, 2016.

[16] Zahid, Z., Saharizan, N. S., Hamzah, P., Hussin, S. A. S., & Khairi, S. S. M. Multi-dimensional flood vulnerability assessment using data envelopment analysis. In *AIP Conference Proceedings*, Vol. 1905, No. 1, AIP Publishing, 2017.

[17] Fadzer, S. A. M., Zahid, Z., & Hussin, S. A. S, Flood vulnerability assessment using data envelopment analysis. *Menemui Matematik*, Vol 39, No. 1, pp. 1-7, 2017.

[18] Saharizan, N. S., Zahid, Z., Hamzah & Khairi, S. S. M. (2018). Spatial flood vulnerability assessment in Peninsular Malaysia using data envelopment analysis. *Journal of Technology Management and Business*, Vol 5, No. 1.

[19] Farzipoor Saen, R, A decision model for selecting the best entry modes via data envelopment analysis. *International Journal of Applied Decision Sciences*, Vol 4, No. 3, pp. 213–229, 2011.

[20] Kthiri, W., Emrouznejad, A., Boujelbene, Y. and Ouertani, M.N. A framework for performance evaluation of employment offices: a case of Tunisia. *International Journal of Applied Decision Sciences*, Vol 4, No. 1, pp. 16–33, 2011.

[21] Meenakumari, R., Kamaraj, N. and Thakur, T, Measurement of relative operational efficiency of SOEUs in India using data envelopment analysis. *International Journal of Applied Decision Sciences*, Vol 2, No. 1, pp. 87–104, 2009.

[22] Sufian, F, The evolution of Malaysian banking sector's efficiency during financial duress: consequences, concerns, and policy implications. *International Journal of Applied Decision Sciences*, Vol 3, No. 4, pp. 366–389, 2010.

[23] Xiao, J., & Li, D, An analysis on technical efficiency of paddy production in China. *Asian Social Science*, Vol 7, No. 6, pp. 170, 2011.

[24] Azadeh, A., Haghighi, S. M., Zarrin, M., & Khaefi, S. Performance evaluation of Iranian electricity distribution units by using stochastic data envelopment analysis. *International Journal of Electrical Power & Energy Systems*, Vol 73, pp. 919-931, 2015.

[25] Ho, C. T. B., & Wu, D. D. Online banking performance evaluation using data envelopment analysis and principal component analysis. *Computers & Operations Research*, Vol 36, No. 6, pp. 1835-1842, 2009.

[26] Wu, D., & Dash Wu, D, Performance evaluation and risk analysis of online banking service. *Kybernetes*, Vol 39, No.5, pp. 723-734, 2010.

[27] Andrejić, M., Bojović, N., & Kilibarda, M, Benchmarking distribution centres using Principal Component Analysis and Data Envelopment Analysis: a case study of Serbia. *Expert Systems with applications*, Vol 40, No.10, pp. 3926-3933, 2013.

[28] Frisch, R, *Theory of production*. Dordrech D. Reidel Publishing Co, 1965.

[29] Ibrahim, A. Z., & Alam, M. M, Climatic changes, government interventions, and paddy production: an empirical study of the Muda irrigation area in Malaysia. *International Journal of Agricultural Resources, Governance and Ecology*, Vol 12, No. 3, pp. 292-304, 2016.

[30] Parichatnon, S., Maichum, K., & Peng, K. C, Evaluating technical efficiency of rice production by using a modified three-stage data envelopment analysis approach: a case study in Thailand. *International Journal of Scientific & Technology Research*, Vol 4, No. 8, pp. 152-159, 2015.

[31] Nassiri, S. M., & Singh, S, Study on energy use efficiency for paddy crop using data envelopment analysis (DEA) technique. *Applied Energy*, Vol 86, No.7, pp. 1320-1325, 2009.

[32] Golany, B., & Roll, Y, An application procedure for DEA. *Omega*, Vol 17, No. 3, pp. 237-250, 1989.

[33] Traskin, M., Wang, W., Ten Have, T. R., & Small, D. S. (2012). Efficient estimation of the attributable fraction when there are monotonicity constraints and interactions. *Biostatistics*, Vol 14, No.1, pp. 173-188, 2012.

[34] Dinç, İ., Sigdel, M., Dinç, S., Sigdel, M. S., Pusey, M. L., & Aygün, R. S, Evaluation of normalization and PCA on the performance of classifiers for protein crystallization images. In *Southeastcon 2014, IEEE* (pp. 1-6). IEEE, 2014.

[35] Aisyah, N., Arumugam, N., Hussein, M. A., & Latiff, I, Factors affecting the technical efficiency level of inshore fisheries in Kuala Terengganu; Malaysia. *International Journal of Agricluture Management & Development*, Vol 2, pp. 49-56, 2012.

[36] Bowlin, W. F., Charnes, A., Cooper, W. W., & Sherman, H. D, Data envelopment analysis and regression approaches to efficiency estimation and evaluation. *Annals of Operations Research*, Vol 2, No. 1, pp. 113-138, 1984.

[37] Cooper, W. W., Seiford, L. M., & Zhu, J. (Eds.). *Handbook on data envelopment analysis* (Vol. 164). Springer Science & Business Media, 2011.

[38] OECD (2008). Handbook on Constructing Composite Indicators–Methodology and User Guide. Organisation for Economic Cooperation and Development, Paris, pp. 66-67. ISBN 978-92-64-04345-9.

[39] Coelli, T. J., & Rao, D. S, Total factor productivity growth in agriculture: a Malmquist index analysis of 93 countries, 1980–2000. *Working Paper Series No. 02/2003*, Centre for Efficiency and Productivity Analysis, 2003.

[40] Pasupathy, K. S, *Modeling undesirable outputs in data envelopment analysis: various approaches.* Doctoral Dissertation, Virginia Tech, 2002.

[41] Pastor, J. T, ranslation invariance in data envelopment analysis: a generalization. *Annals of Operations Research*, Vol 66, No. 2, pp. 91-102, 1996.