Energy Consumption Data Analysis: Indonesia Perspective

Wahyu Andy Prastyabudi [1]* and Isa Hafidz [2]

[1] Department of Industrial Engineering, Institut Teknologi Telkom Surabaya; wahyu.andy@ittelkom-sby.ac.id

[2] Department of Electrical Engineering, Institut Teknologi Telkom Surabaya; isa@ ittelkom-sby.ac.id

* Correspondence: wahyu.andy@ittelkom-sby.ac.id;

Abstract: Energy consumption has been increasingly affected by many factors such as population growth, energy price, etc. This paper utilizes data analysis to examine energy consumption data and its correlation to GDP growth, particularly in the case of Indonesia. The data observed in this study comprises a record from 1961-2018. An exploratory data analysis was employed to comprehend the features of the data, while beforehand a pre-processing and testing were conducted to prepare the data. Data visualization was built then a Spearman correlation rank was also performed to see the association among variables. The result shows that energy consumption is dominated by fossil generated energy i.e. oil, coal, gas. Meanwhile, renewable energy was started to use since early 2000 with a huge leap of consumption on solar energy. The result of correlation ranking indicates that the energy consumption and GPD growth have no significant association as the rank value is negative and the p-value is greater than the chosen significance level. Moreover, Indonesia has the opportunity to switch to geothermal energy which contributes 3.6% to the national energy supply.

Keywords: energy; data analysis; GDP

1. Introduction

Energy has been always an issue to be encountered by many countries including Indonesia. As the fossil energy sources have decreased rapidly while its consumption is soaring gradually, there is a need to look forward the energy security and at the same time maintaining the consumption efficiency. Energy security is a such condition of energy availability secured and affordable for relatively a long period not being affected by the regional nor international fluctuation [1]. Global Energy Institute reported in the International Index of Energy Security Risk 2020, that the world will require the amount of energy 31% larger in 2040 than we produce today [2]. The report shows that Indonesia is classified as a large energy user group at ninth place among 25 countries.

By 2050, the demand of energy in Indonesia is expected to reach the average of 4.9% growth every year, in case of business as usual scenario is applied [3]. Population growth, economic growth, energy price, and development of technology are those among factors that affected the energy demand [3]. On the other hand, the growth of GDP may also be driven by the amount of energy consumption to boost up the industrial and economic activities. Thus, it is necessary to oversee the rate of energy consumption meanwhile the energy conservation program is carried out.

This study particularly attempts to unfold the correlation between energy consumption and GDP. The energy consumption data of Indonesia and the data of GDP growth were deserved to assess further by intention to give a broader view on how Indonesia spend most of their energy. Moreover, it may lead to provide insights for the government so that they are able to formulate an appropriate regulation. Exploratory data analysis approach was employed to analyze the data by which beforehand was pre-processed to prepare the data. Shapiro-Wilk test was performed to examine the

data distribution while another test so-called Spearman rank correlation was conducted as well to find out the relationship among observed variables.

In the recent study, we found some literature that investigates the energy consumption data for various objectives. Fatai et al. (2004) discussed the application of co-integration analysis, i.e. Granger causality, to study the relationship between energy consumption data and GDP using a case study in New Zealand, Australia, and four Asian countries including Indonesia [4]. It is highlighted in their study that there was a unidirectional link from energy to income, particularly in Indonesia. The same method was employed in a paper by Lise and Montfort whereas the study was specifically performed to analyze the energy consumption in Turkey [5]. In addition, another method to elucidate the relationship between energy and GDP is using panel data analysis approach by which it observes time-series cross sectional data [6], [7].

To complement those studies, this paper asserts the use of graphical representation to explore the salient features of energy consumption and GDP data. The paper is, then, organized into four sections. In addition to this introduction, Section 2 will discuss about the data and methods to analyze them by means of data exploratory analysis. In Section 3, the result of data analysis is presented. Spearman rank correlation was employed to show the association between the energy consumption and the GDP. Meanwhile, the Section 4 discusses the results.

2. Materials and Methods

2.1. Data

The data utilized in this study is available online taken from the website of world bank and British Petroleum (BP). From the world bank we collected the GDP data which comprises the record of country's GDP growth per capita from 1965-2017 [8]. Meanwhile, the energy consumption data provided by BP comprises annual data available from 1961-2018 [9]. The data contains energy consumption which can be categorized into two by its sources, i.e. non-renewable energy (oil, coal, gas, nuclear) and renewable energy (hydropower, wind, solar, others).

2.1. Methods

2.1.1. Data Pre-Processing

Prior to harness the data for analytical purpose, the data is required to undergo a pre-processing. We filtered out the data, both GDP and energy consumption data, from Indonesia only as we only concern to analyze the data from this country in the present study. Then, both data were combined by intersecting its year vector values. So that the merged data has the same series of period. We also removed non-essential part of the data, particularly in the energy consumption data, we eliminated the column nuclear as Indonesia has not yet owned nuclear power.

2.1.2. Testing

Once the pre-processing data completed, the following stage is to conduct statistical testing, thus we can choose the appropriate means to analyze the data. Herein, we conduct Shapiro-Wilk test (see equation 1) to examine the normality of data.

$$W = \frac{(\sum_{i=1}^{n} a_i x_i)^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \tag{1}$$

The null hypothesis (H_0) assumed that the data is normally distributed. Otherwise, the alternative hypothesis (H_1) assumed that the data does not follow normal distribution. If the test result indicates *p*-value is less than the chosen alpha level, then the null hypothesis is rejected, and the alternative hypothesis is accepted. Thereby, we can be evidently confident that the data tested are not normally distributed. On the other hand, if the *p*-value is greater than the chosen alpha level, then the null hypothesis is accepted and reject the alternative hypothesis by which we can conclude that the data came from a normally distributed population.

Another statistical test that we conduct herein is Spearman rank correlation coefficient which measures the association between two observed variables X (independent variable) and Y (dependent variable). This rank, denoted as r_s, can be computed as follows:

$$r_s = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} d_i^2 \qquad (2)$$

where d_i is the deviation between the ranks assigned to x_i and y_i with the number of n pairs of data. The rank value (r_s) will range from -1 to +1 which indicates the direction of association of both variables. If the rank value r_s is equal to zero, it may be inferred that there is no tendency of those variables to be correlated. In addition, if the p-value is smaller than the chosen alpha level it means the correlation is considered to be significant.

2.1.3. Data Analysis Procedure

In order to gain understanding of the data including their relationship, an exploratory data analysis (EDA) approach was performed. This method was introduced by John W. Tukey by which it tries to look the data from as many aspects as possible [10]. Hence, we may be able to probe the underlying features reside in the data. The EDA offers an advantage as there is no model to be considered, rather it emphasizes the data visualization to comprehend its characteristics.

Numerous studies have employed EDA method to carry out analysis regarding the energy consumption data. Filipovic et al. studied about energy security index using EDA approach to determine a new index considering economic, environmental, political, and social dimensions [11]. Their observed 28 EU countries for ten years period of energy security index data. The study found that GDP per capita greatly affects the country's energy security. Another study by Baker and Rylatt (2008) tried to improve the prediction of energy demand by analyzing annual energy consumption data [12]. Meanwhile, Tyralis et al. used EDA to understand the time series electrical energy demand data and building a forecasting model [13].

In this study, analysis was performed by observing the statistical descriptive summary to see the range of energy consumption of each source. Thereafter, we can easily categorize the data which has a closed band of the amount of energy. The following step was to visualize the data using boxplot and time series line chart. Thus, it is possible to compare each energy source's statistical value, e.g median, maximum, and minimum. By means of time series line chart, we may examine the movement and trend over the period. The Spearman correlation rank test was performed to examine the associations between energy consumption, at each source, and the GDP. Finally, a correlogram correlation matrices was built to visualize the association of each variables.

3. Results

3.1. Data testing

The energy consumption data contains 5,066 records by which after cutting off to observe the data from Indonesia, we obtain 54 records. In the meantime, the GDP data by Indonesia only contains about 57 records whereas the raw data has 11,315 records. Since the energy consumption data and GDP data has the different range of period, both data are then intersected and obtained 54 records which range from the year of 1965-2017.

In order to test the normality of each data, Shapiro-Wilk test was performed using $\alpha = 0.05$ and **Table 1** below shows the test result. As shown from the table, most of the data do not follow the normal distribution as its p-value is less than the given α value, except for the Wind and Other data that closely follow the normal distribution. In addition, the energy consumption data has been classified into three different groups according to its consumption level. The first group contains the data of energy consumption produced by oil, coal, and gas whereby they produce the biggest energy consumption with average more than 150 terawatt-hours. The second group consists of the energy consumption data produced by hydropower and other sources with both have roughly the same

median to 17 terawatt-hours. Meanwhile, the energy consumption data produced by wind and solar which almost share the same median of 0.01 terawatt-hours are put together into the third group.

Table 1. Shapiro-Wilk normality test for each energy consumption by its source

Variables	W	p-value
Oil	0.91104	0.0007855
Coal	0.74532	3.117e-08
Gas	0.87809	6.268e-05
Hydropower	0.90803	0.0006137
Solar	0.73971	0.002674
Wind	0.83857	0.05568
Other	0.93561	0.08551
GDP	0.67043	1.233e-09
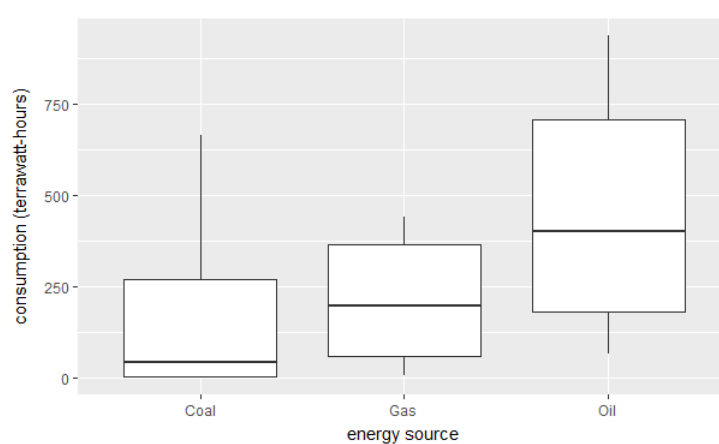
Figure 1. Box plot of energy consumption produced by oil, coal, and gas
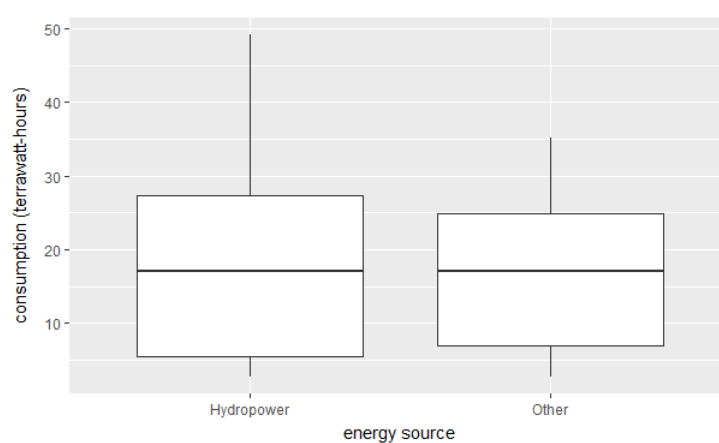
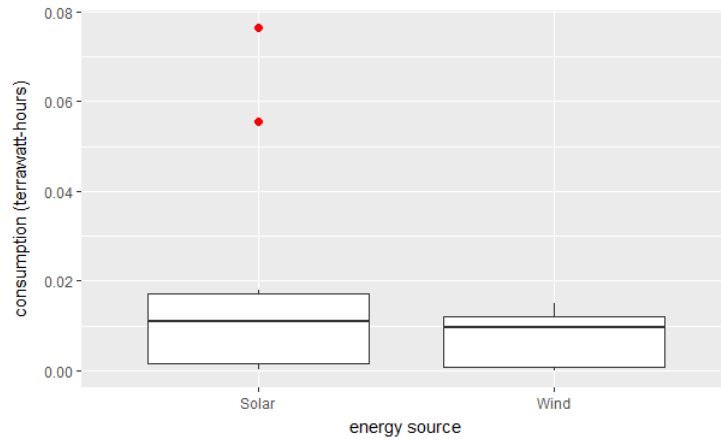**Figure 2.** Box plot of energy consumption produced by hydropower and other sources

Figure 3. Box plot of energy consumption produced by solar and wind

3.2. Exploratory data analysis

Figure 1 presents the statistical summary of energy consumption data produced by oil, coal, and gas. During the period energy produced by oil has a lot of usages compared to other sources by which it is confirmed by the median surpass the other two data. On the other hand, however, the energy consumption produced by oil has also more variability. The figure also reveals that the coal data and gas data have relatively similar variability, however the coal data has more skewness to the right. Over the period, the energy consumption produced by the coal has been surpass the gas and nearly reach the 0.75 percentile of the oil data.

Figure 2 summarizes the statistical features of energy consumption data produced by hydropower and other sources. Both data share the quite identical median that is 17 terawatt-hours, however, the hydropower data has more variability and its maximum value is higher as well than the other sources do. Moreover, **Figure 3** depicts the statistical summary of the rest group data that is energy consumption produced by solar and wind. Both data has relatively low energy consumption under 0.02 terawatt-hours and the wind data has more skewness to the left.
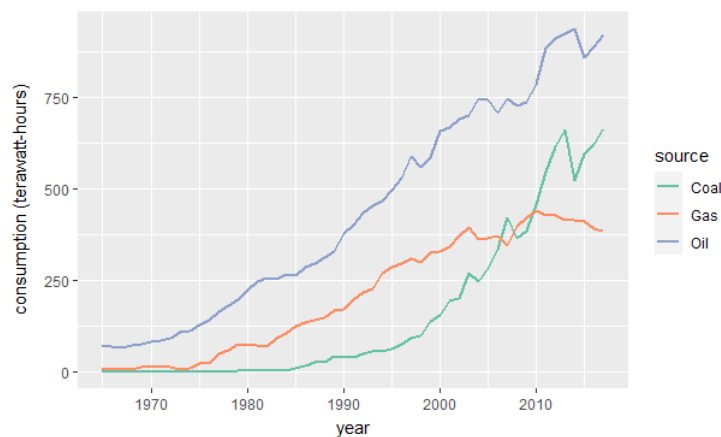

Figure 4. Energy consumption of the first group per year during 1965-2017

Figure 4 depicts the change of energy consumption produced by oil, coal, and gas. Meanwhile, **Figure 5** and **Figure 6** show the change of energy consumption in the second and third group, respectively. The consumption of most of energy sources are soaring over period. The energy consumption from coal has surpassed the gas by the year of 2010. In the meantime, energy generated by hydropower was utilized relatively low during 15 years in early period. Then, the consumption increased rapidly by 1985 and fluctuated after 1990. On the other hand, the utilization of other

renewable energy was begun at 1990 with relatively low variance, however, the usage yet lower than those produced by hydropower. In fact, it is shown that the cultivation of renewable energy from solar and wind was just started at 2007. Apparently, the consumption from wind energy was somehow quite low under 0.02 terawatt over the period. From **Figure 6**, it is also shown that the usage of solar energy was escalate incisively since 2015.
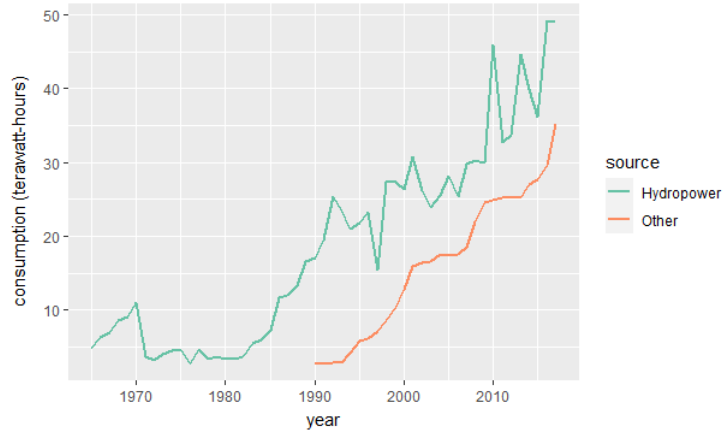

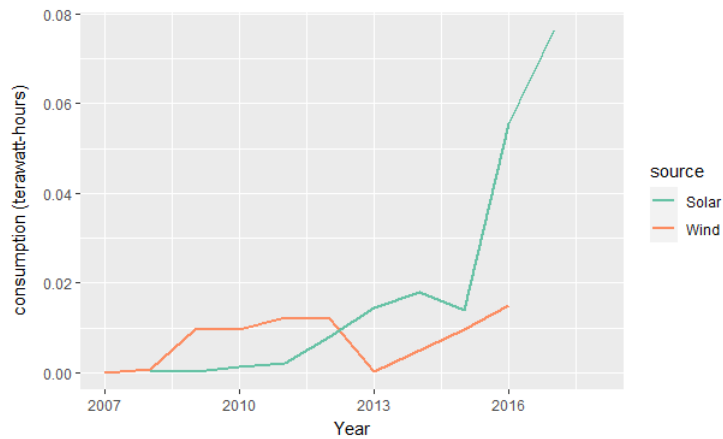**Figure 5.** Energy consumption data of the second group during 1965-2017

Figure 6. Energy consumption data of the third group during 2007-2017

Considering the growth of GDP along with the energy consumption over the period, a bubble chart visualization is presented in **Figure 7, Figure 8,** and **Figure 9** for each group of energy sources, respectively. The bubble size represents the amount of energy consumption in terawatt-hours. Note that each chart is depicted an outlier of the GDP in an extremely condition which occurred in crisis period 1998 with only -14.34% growth of GDP. It is shown from **Figure 7** and **Figure 8** that the amount of energy consumption relatively getting increased since 1990, though there was still no implication in a positive way to the growth of GDP. In addition, **Figure 9** reveals another outlier of GDP growth with only about 3.2% that occurred in 2009. The energy consumption from solar increased gradually since 2013. Meanwhile, the energy consumption produced by wind was vacant at 2014 and 2017 as no bubble shown in the figure.
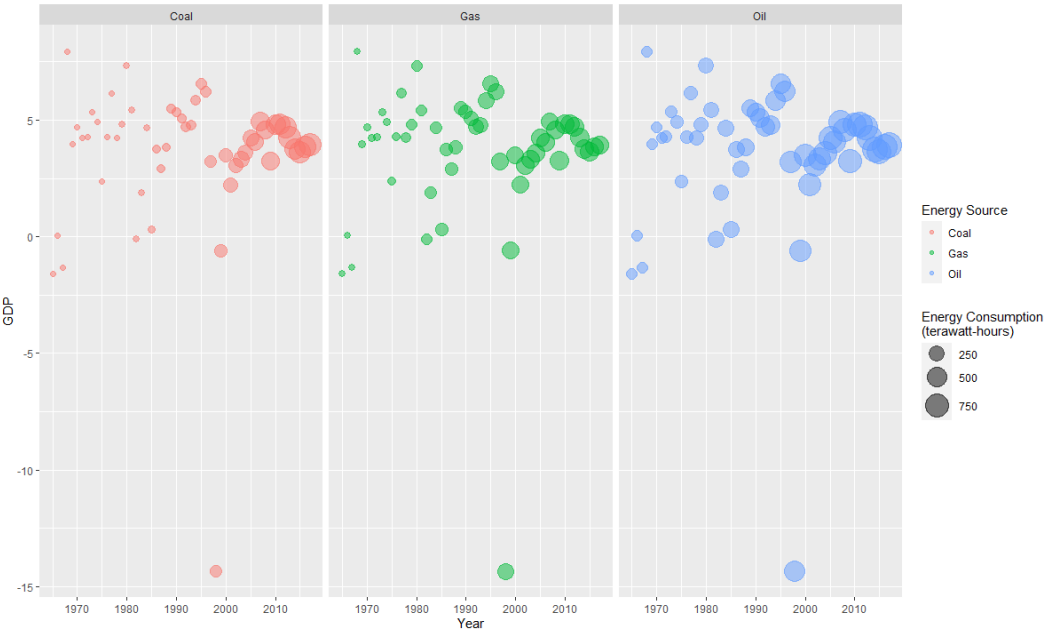
Figure 7. Change of energy consumption (oil, gas, coal) and GPD over period 1965-2017
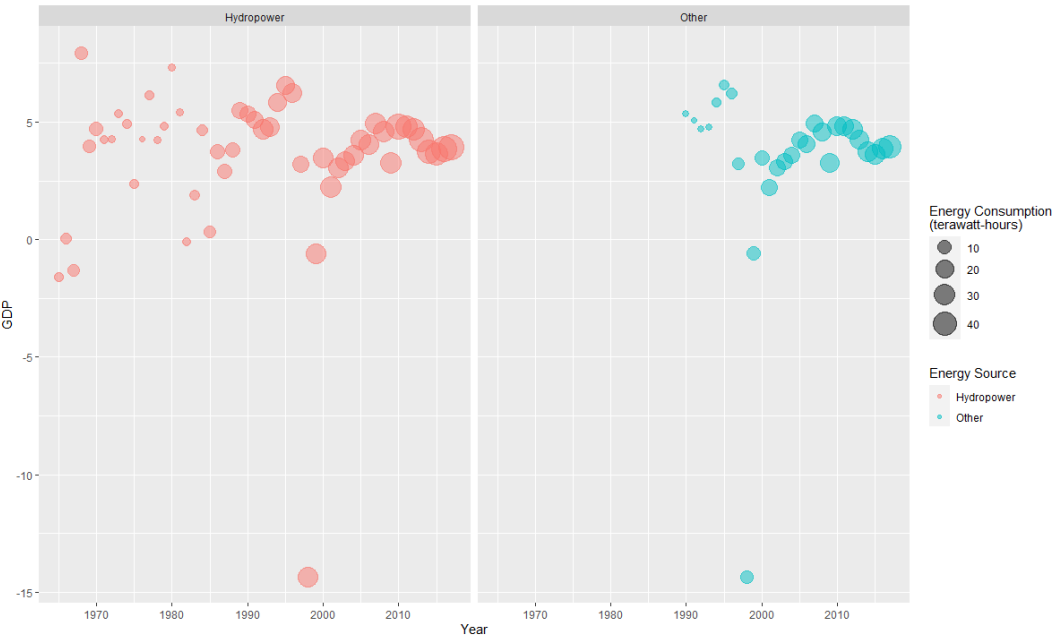

Figure 8. Change of energy consumption (hydropower and other) and GPD over period 1965-2017
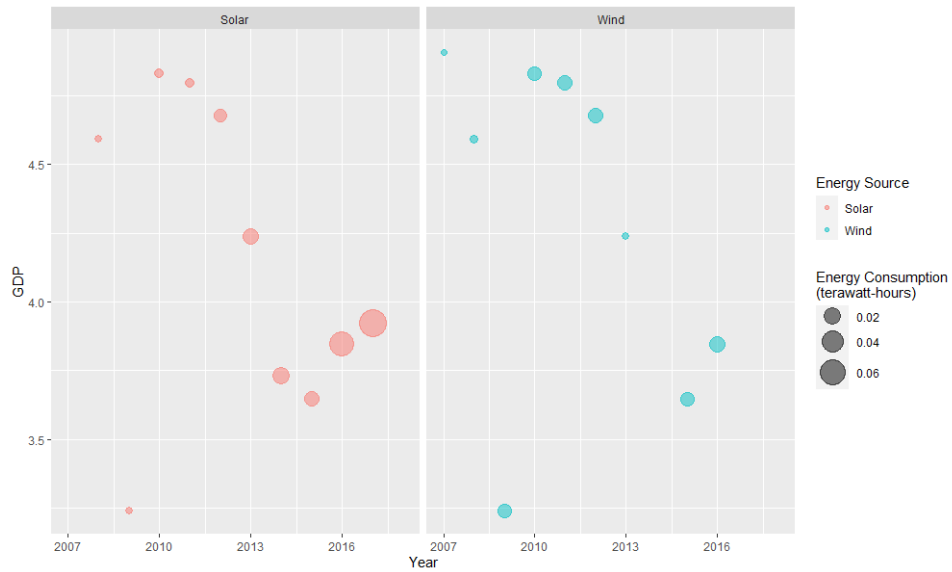
Figure 9. Change of energy consumption (solar and wind) and GPD over period 2007-2017

3.3. Correlation between GDP and energy consumption

Herein, the association of GDP and energy consumption by means of Spearman correlation rank. The data was divided, for the sake of simplicity, into three groups. The first group comprises the energy consumption data of oil, gas, coal, and hydropower by which the usage period was started at the same year. The second group consist the data of energy powered by solar and wind by which the usage of these power was started relatively the similar period since 2007. Thus, the observed GDP growth was tailored to the same period by 2007. The third data is left only for the energy consumption of other generated power as it was firstly utilized at 1990 which had the different period to the other sources.
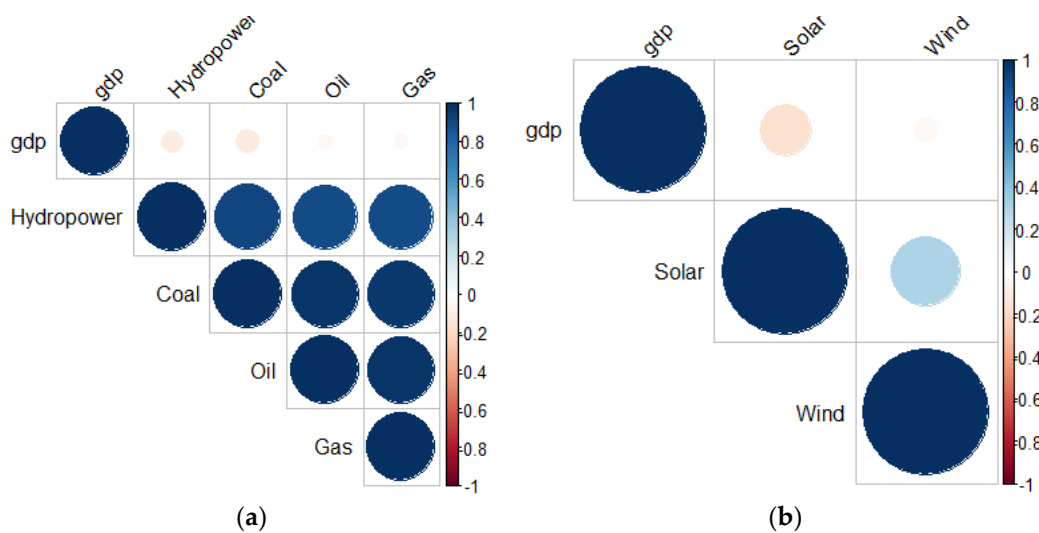

(a)   (b)

Figure 10. Correlogram of energy consumption and GDP. (a) The correlation matrices between GDP and energy consumption produced by oil, gas, coal, and hydropower. (b) The correlation matrices between GDP and energy consumption produced by solar and wind.

Table 2 and

Table 3 show the correlation rank for the first and second group, respectively. As depicted from the tables that the correlation between GDP and most of energy sources has negative association. In addition, the correlation rank between the consumption of other powered energy and GPD yields the rank $r = -0.2364$. **Figure 10** illustrates the correlogram matrices which visualize the result of correlation rank in **Table 2** and **Table 3** into figure (a) and (b), respectively. Positive correlation is indicated by the blue color, otherwise the negative correlation is indicated by the red color. Meanwhile, the color intensity and the size of circle imply the proportion of the correlation coefficient value. As we can see that the consumption energy generated by wind and gas are the highest rank while the solar energy is the lowest rank. The p-value (see **Table 4**), by which larger than the significant level (alpha) 0.05, also indicates that correlation between energy consumption and GDP growth is considered as insignificant.

Table 2. Spearman correlation rank of GDP and energy consumption produced by oil, gas, coal, and hydropower

	Oil	Gas	Coal	Hydropower	GDP
Oil	1.0	0.9797	0.9726	0.8941	-0.0420
Gas	0.9797	1.0	0.9639	0.8903	-0.0411
Coal	0.9726	0.9639	1.0	0.9159	-0.1151
Hydropower	0.8941	0.8903	0.9159	1.0	-0.1098
GDP	-0.0420	-0.0411	-0.1151	-0.1098	1.0

Table 3. Spearman correlation rank of GDP and energy consumption produced by solar and wind

	Solar	Wind	GDP
Solar	1.0	0.3072	-0.1676
Wind	0.3072	1.0	-0.0359
GDP	-0.1676	-0.0359	1.0

Table 4. p-value of Spearman correlation test

Energy source/GDP	p-value
Oil	0.7647
Gas	0.7700
Coal	0.4116
Hydropower	0.4333
Solar	0.6914
Wind	0.9327
Other	0.2248

4. Conclusion

The paper discusses pertaining energy consumption in Indonesia during the period between 1965 and 2017. As well as trying to unfold the relationship of energy consumption and the growth of GDP. The energy consumption data examined in this study was divided by its sources into seven i.e. oil, gas, coal, hydropower, solar, wind, and other. An exploratory data analysis approach was employed to reveal the features of the data. The method emphasizes on the graphical representation analysis. Two testing were conducted prior to the analysis i.e. Shapiro-Wilk test and Spearman correlation rank.

The testing result of normality test indicates that most of the data did not follow normal distribution. Hence, to figure out the association between energy consumption and GDP growth Spearman correlation rank was employed. The result implies that in general energy consumption is increasing gradually despite its source. In regard of renewable energy, Indonesia uses a numerous amount of energy generated by solar power since 2015 with a huge leap on the following year. On the other hand, the result of correlation test implies there was no significant relationship between energy consumption and GDP growth indicated by the negative rank value. Notwithstanding, as the growth of energy consumption, moreover, the energy reserve is decreasing. It is expected that, in case of Indonesia, the reservation of oil, gas, and coal reserve will last at 2024, 2052, and 2091, respectively [1]. Hence, to maintain the energy security index Indonesia needs to switch to the alternative renewable energy, such as geothermal energy which contributes 3.6% to the national energy supply [14].

References

1. R. E. Caraka, "Simulasi Kalkulator Energi Baru Terbarukan (EBT) Guna Memenuhi Ketahanan Energi di Indonesia," *Stat. J. Theor. Stat. Its Appl.*, vol. 16, no. 2, pp. 77–88, 2016.
2. U.S. Chamber of Commerce, "International Index of Energy Security Risk: 2020 Edition," 2020. Accessed: May 11, 2020. [Online]. Available: https://www.globalenergyinstitute.org/international-index-energy-security-risk-2020-edition.
3. D. E. Nasional, "Indonesia energy outlook 2016," *Jkt. Sekr. Jenderal Dewan Energi Nas.*, 2016.
4. K. Fatai, L. Oxley, and F. G. Scrimgeour, "Modelling the causal relationship between energy consumption and GDP in New Zealand, Australia, India, Indonesia, The Philippines and Thailand," *Math. Comput. Simul.*, vol. 64, no. 3–4, pp. 431–445, 2004.
5. W. Lise and K. Van Montfort, "Energy consumption and GDP in Turkey: Is there a co-integration relationship?," *Energy Econ.*, vol. 29, no. 6, pp. 1166–1178, 2007.
6. B.-N. Huang, M. J. Hwang, and C. W. Yang, "Causal relationship between energy consumption and GDP growth revisited: a dynamic panel data approach," *Ecol. Econ.*, vol. 67, no. 1, pp. 41–54, 2008.
7. C.-C. Lee and C.-P. Chang, "Energy consumption and economic growth in Asian economies: a more comprehensive analysis using panel data," *Resour. Energy Econ.*, vol. 30, no. 1, pp. 50–65, 2008.
8. World Bank, "GDP per capita growth (annual %)." https://data.worldbank.org/indicator/NY.GDP.PCAP.KD.ZG (accessed May 11, 2020).
9. BP Global, "Statistical Review of World Energy," *BP global*. https://www.bp.com/en/global/corporate/energy-economics/statistical-review-of-world-energy.html (accessed May 11, 2020).
10. S. Morgenthaler, "Exploratory data analysis," *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 1, no. 1, pp. 33–44, 2009.
11. S. Filipović, M. Radovanović, and V. Golušin, "Macroeconomic and political aspects of energy security–Exploratory data analysis," *Renew. Sustain. Energy Rev.*, vol. 97, pp. 428–435, 2018.
12. K. J. Baker and R. M. Rylatt, "Improving the prediction of UK domestic energy-demand using annual consumption-data," *Appl. Energy*, vol. 85, no. 6, pp. 475–482, 2008.
13. H. Tyralis, G. Karakatsanis, K. Tzouka, and N. Mamassis, "Exploratory data analysis of the electrical energy demand in the time domain in Greece," *Energy*, vol. 134, pp. 902–918, 2017.
14. S. Setiawan, "Energi panas bumi dalam kerangka MP3EI: Analisis terhadap prospek, kendala, dan dukungan kebijakan," *J. Ekon. Dan Pembang.*, vol. 20, no. 1, pp. 1–28, 2012.