

Analysis on the Correlation Between Energy Consumption and Weather Parameters

Bowei Wang

*School of Mathematical Sciences, University of Southampton, Southampton, UK
bw6u23@soton.ac.uk*

Abstract. After the industrial revolution, and entering the 21st century, energy consumption has grown more rapidly than expected as the usage of technology and machines for both commercial and residential purposes continuously increase. An inevitable consequence is the artificial damage of natural environment which could lead to climate change in certain areas. On the other hand, the variation of climate and weather conditions could also influence energy use. Hence, how do they correlate with each other becomes a more noticeable issue nowadays. In this paper, a regional weather and energy dataset in London is chosen, and by using Pearson Coefficient, a correlation analysis is implemented between weather parameters and energy consumption. The result show that out of various weather variables, sunshine, global radiation and temperature take the highest weight in influencing the household energy demand in London. These three factors are then used to fit a linear regression model to further emphasize their sensitivity level to energy consumption.

Keywords: Weather and energy, Linear regression, Pearson coefficient

1. Introduction

Climate change has become a global issue in today's era, according to the United Nations, "climate change refers to long term shift in temperature and weather patterns, since the 1800s, human activities have been the main driver of climate change, primarily due to the burning of fossil fuels like coal, oil and gas" [1]. In the 21st century, electricity has now taken the main part in energy consumption, and the transformation of energy in between is changing the weather and temperature. Taking the heating and cooling system as an example, the major substance in air conditioning is Freon. The process of air conditioning will first compress the cool Freon gas into hot high-pressure Freon gas and dissipate outside by the window unit, then the condensed liquid freon would go through a tunnel and evaporate to absorb heat inside [2]. This process will decrease the indoor temperature, but on the other hand, the massive hot gases released outside will influence the outdoor temperature in the long term. From this specific example, the temperature is dependent on energy consumption, hence it makes sense to ask does the energy consumption also correlated to the entire weather system, and mathematically, what kind of relationship is between these two variables. These questions are going to be answered accordingly in the following sections.

2. Literature review

Over the past 21st century, many researchers have done various work in this area from different perspectives. In the field of the relation between weather conditions and electricity, Aisyah and Simaremare conducted research in Central Java Province and Bali Island. The results show that, out of different weather variables, temperature, solar radiation and wind speed play a major role in influencing energy consumption. The paper also points out that the key influential factors could vary geographically. For example, rainfall has a high correlation with energy consumption in Java Province, but not in Bali Island [3]. Another research was done by Fikru and Gautier, who analyzed the impact of weather variations on two types of houses in Tyler, Texas, USA. For house 1, it is a conventional house, while for house 2, it uses solar energy. The results show that house 2's energy consumption has less sensitivity to the variation of weather parameters than house 1. "On average, one unit increase in heating and cooling degree minutes increases energy use by about 9% and 5% respectively for house 1 and 5% and 4% respectively for house 2." The overall influencing factors are temperature, solar radiation and humidity. Moreover, the paper shows that energy use is also highly dependent on the seasonality and the specific time in day and night [4]. From these two researches, it can be seen that the relationship between weather and energy is complicated and depends on a lot of different factors, but out of these factors, some parameters such as temperature or solar radiation consistently influence the change of energy consumption no matter how the other external variables change. In this paper, the author will continue analyzing the correlation between weather and electricity demand, considering all the possible weather variables, finding the correlation coefficient between them, and hence fit a multivariate linear regression model between energy and weather.

3. Methodology

3.1. Dataset description

This paper specifically analyzes a weather dataset recorded in London by the European Climate Assessment (ECA) and a London Energy dataset recorded by UK Power Networks. The two datasets are downloaded from open data source Kaggle, named London Weather Data (<https://www.kaggle.com/datasets/emmanuelwerr/london-weather-data>) and London Energy Data (<https://www.kaggle.com/datasets/emmanuelwerr/london-homes-energy-data>) respectively. They will be described separately.

3.1.1. London weather data

As shown in Table 1, this is a time series dataset from Jan 1st, 1979, to Dec 31st, 2020. The dataset contains 10 variables in total:

Table 1. London weather data description

Variab le	Date	cloud_co ver	sunshin e	global_radiat ion	max_te mp	mean_te mp	min_te mp	precipitati on	pressure	Snow_dep th
Unit	yyyy/mm/ dd	per oktas	per hours	W/m2	°C	°C	°C	mm	Pa (Pascals)	cm

3.1.2. London energy data

As shown in Table 2, this is a time series dataset that keeps track of 5567 randomly selected households in London from November 2011 to February 2014. The dataset contains three variables:

Table 2. London energy data description

Variable	LCLid	Date	KWH
Unit	individual household unique identifier - (str)	yyyy/mm/dd	kwh

By calculating the mean value of all recorded households' energy consumption, the daily average energy usage in London could be obtained.

This paper will aggregate the two datasets for analysis, to find the correlation between weather and energy consumption in London. The choice of this specific geographic location here is due to the special climate in London-- Temperate Marine Climate. It is a type of mild and humid climate with no severe cold in winter, no scorching heat in summer and the precipitation is evenly distributed around the year. Therefore, there is not going to be any extreme weather which heavily influences the data and analysis results.

3.2. Correlation analysis

To conduct correlation analysis between weather variables and energy consumption (KWH), the Pearson Coefficient is used here to determine whether the two variables are linearly dependent. "Pearson correlation coefficient is specifically used to describe the linear relationship between two normally distributed variables" [5]. It has the formula:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (1)$$

By calculating the person correlation coefficient using R for each variable in the two datasets piecewisely, the strength of the linear dependency could be known. The range of this number is between -1 and 1, where -1 means a complete negative linear relationship and 1 means a completely positive linear relationship. If two variables have a coefficient with absolute value above 0.7, it is considered a strong linear relationship, if it is between 0.5 and 0.7, it has a medium linear relationship, and if it is below 0.5, it has a weak linear relationship.

3.3. Model fitting

After finding the correlation coefficient between energy and other variables in weather dataset, the weather parameter with the highest correlation coefficient will be chosen to fit a multivariate linear regression model with energy as the response variable. The estimation method used here is least squares estimation. If the final model has the form:

$$y_i = ax_i + b \quad (2)$$

then the associated error term is:

$$E(a, b) = \sum_{i=1}^n (y_i - (ax_i + b))^2 \quad (3)$$

The least square estimation aims to find the value of a and b that minimizes the error using differentiation [6]. By obtaining the model, part of the weather variables would be used to predict energy consumption.

4. Results and discussion

By using Pearson Coefficient, the correlation coefficient between variable KWH in London Energy data and other variables in London Weather dataset is shown in Table 3:

Table 3. Pearson coefficient between KWH and other factors in the London weather dataset

	KWH	cloud_cover	sunshine	global_radiation	mean_temp	precipitation	pressure	snow_depth
KWH	1.0000	0.4642	-0.7423	-0.8139	-0.9415	0.1691	0.0019	0.4412

The Table 3 shows that, out of the 7 factors, sunshine, global_radiation and mean_temp play a major role in influencing KWH. All these three variables have a negative linear relationship with energy consumption. It is probably due to the lack of usage of cooling and heating systems in the London area in summer and winter, which makes the energy consumption lowest when the weather is hottest/coldest. But since this field doesn't cover much research, the accuracy of this conclusion remains uncertain.

After choosing these three variables as the highest influencing factor here, they're then used to fit a multivariate linear regression model using the lm function in R, the parameters of this model are shown in Table 4:

Table 4. Linear regression model summary

	Coefficient	Standard Error	t value	Pr(> t)	Signif. Codes
Intercept	-4.254e-16	1.446e-02	0.00	1	n/a
sunshine	2.172e-01	2.881e-02	7.54	1.24e-13	***
global_radiation	-4.479e-01	3.520e-02	-12.72	<2e-16	***
mean_temp	-6.785e-01	2.088e-02	-32.49	<2e-16	***

Notice that the coefficients of the three variables are on the same level, this is because the units of three factors are standardized with mean=0 and standard deviation =1. After this process, the coefficients will genuinely represent the weight of each factor's influence towards the KWH in this linear regression model. Here, mean_temp has the highest weight, following by global_radiation and sunshine. Even though all three variables are inversely proportional to energy, sunshine still has positive coefficients here. It is probably used to neutralize the effect of global_radiation and mean_temp since sunshine only has the lowest correlation coefficient with energy usage. Nevertheless, three factors are all significant with a low p-value. The overall R-squared of this model is 0.8276 and the residual plot of it is shown in Figure 1:

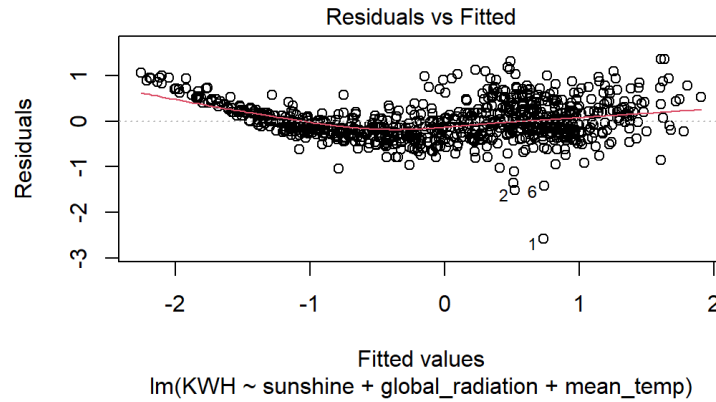


Figure 1. Residual plot of the fitted model

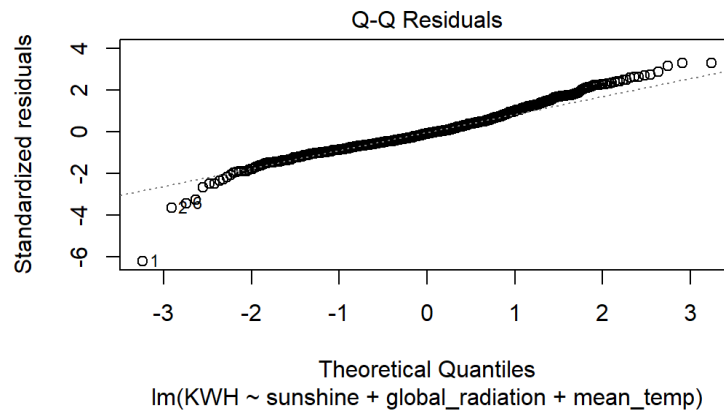


Figure 2. Q-Q plot of the fitted model

R squared and adjusted R squared are statistics parameters that are used to describe the accuracy of a model, it represents what percentage does the variance of the response variable is explained by the predictors [7]. Here, the R squared value of 0.8276 means 82.76% of the variation in energy consumption is explained by the three weather factors. Moreover, by looking at the two plots, the accuracy of the model can be further examined. Linear and squared residual plot are specifically used to check the nonlinearity and heteroscedasticity of a model. The graph plots the residual against the fitted value, and by looking at the certain shape (curvature or fan-shape), the nonlinearity of a fitted model could be detected [8]. From the residual plot, it can be observed that there is no obvious curvature on the residual plot, and all the points are distributed around the origin line. As shown in Figure 2, from the Q-Q plot, all the points are close to the fitted line. By using these diagnostic tests, it could be concluded that the model is adequate, and the final multivariate linear regression model between KWH and sunshine, global_radiation and mean_temp is:

$$KWH = -4.254 \cdot 10^{-16} + 0.2172s - 0.4479gr - 0.6785T \quad (4)$$

Additionally, by dividing the original dataset into four quarters, it is possible to analyze the seasonality effect on energy consumption. The coefficients of the three explanatory variables for the four quarters individually are shown in Table 5:

Table 5. Seasonality analysis on energy consumption

coefficient of...	Q1 (spring)	Q2 (summer)	Q3 (autumn)	Q4 (winter)
Sunshine	0.0344	0.5371	0.1204	0.2973
Global Radiation	-0.3512	-0.8103	-0.2485	-1.1647
Temperature	-0.6734	-0.4416	-0.1743	-0.5671

Looking at the result, although the temperature has the highest weight in the whole year, it is not necessarily true for all seasons. For both summer and winter, the global radiation has a higher influence on energy consumption. And overall, spring and autumn's energy usage is less sensitive to the weather conditions. This potentially shows the sophistication and complexity in the modelling between weather and energy consumption.

5. Conclusion

Overall, this article uses Pearson Coefficient to calculate correlation coefficients between weather parameters in London Weather dataset and KWH in London Energy data. It is discovered that out of various weather variables, sunshine, global radiation and temperature is the most influencing factors. This is consistent with the work done by Aisyah and Simaremare. It further emphasizes that these variables' correlation with energy consumption is not influenced by the geographical location that the research is done. Secondly, the article uses the most correlated factors to fit a multivariate linear regression model. The model could be used to predict energy consumption in London if these three weather parameters' data can be known by the local climate bureau. Lastly, the article does a seasonal analysis which shows the difference of the weight of the effect between weather and energy in all four seasons. Hopefully, these results could provide potential advice for local climate center or related enterprises. Under the essential topic of climate change and energy usage, there is still a long way to go.

References

- [1] United Nations. (n.d.). What is climate change? United Nations. <https://www.un.org/en/climatechange/what-is-climate-change>
- [2] Brain, M., Bryant, C. W., & Elliott, S. (2000). How air conditioners work.
- [3] Aisyah, S., & Simaremare, A. A. (2021, December). Correlation between weather variables and electricity demand. In IOP Conference Series: Earth and Environmental Science (Vol. 927, No. 1, p. 012015). IOP Publishing.
- [4] Fikru, M. G., & Gautier, L. (2015). The impact of weather variation on energy consumption in residential houses. *Applied Energy*, 144, 19-30.
- [5] Xu, H., & Deng, Y. (2017). Dependent evidence combination based on shearman coefficient and pearson coefficient. *Ieee Access*, 6, 11634-11640.
- [6] Miller, Steven J. "The method of least squares." *Mathematics Department Brown University* 8.1 (2006): 5-11.
- [7] Miles, J. (2005). R-squared, adjusted R-squared. *Encyclopedia of statistics in behavioral science*.
- [8] Tsai, C. L., Cai, Z., & Wu, X. (1998). The examination of residual plots. *Statistica Sinica*, 445-465.