Identifying the Correlation between Temperature and Gas Consumption in a Local Energy System

Hantao Wang, Chenghong Gu, Xin Zhang, Furong Li, and Lihong Gu

***Abstract*—In order to understand energy consumption and assist precise load prediction, it is essential to identify the variation of gas consumption in response to temperature change. In this paper, the relationship is identified by using Empirical Mode Decomposition (EMD) and linear regression analysis together with outlier detection. EMD is a data processing tool that could divide original data into several Intrinsic Mode Functions (IMFs) with a lower frequency residue. By applying data mining technique-Mahalanobis distance measurement, some outliers from real-time gas consumption and temperature data points are detected, which are excluded from data sets to ensure accuracy. Correlation coefficients between gas load and temperature are calculated and denoted as an important index to quantify their relationship through regression analysis. By comparing such index on real-time data and EMD processed data, the weather-sensitive part of gas demand is identified. The methods are implemented on a local energy system and results reveal that the results after EMD present a higher level of correlation between gas load and temperature, compared to the results from directly using real-time gas load and temperature data.**

***Index Terms*— Correlation coefficient, load consumption, weather effect, Empirical Mode Decomposition, linear regression, outlier detection.**

I. Introduction

In the modern energy industry, the operational decisions such as power generation, power distribution, tariff design, and load dispatch always change in response to end-use load consumption [1]. Most traditional demand response schemes only focus on electricity consumption but ignore gas consumption [2]. However, natural gas also plays an important role our in energy landscape, for example in 2014, natural gas takes account for 63% of UK domestic energy consumption while electricity is only proportionating 25% [3]. Multi-carrier energy systems concept has become popular in recent years [4, 5]. Into the future, customers would be willing to interact with both natural gas and electricity networks, enabled by new technologies, such as Combined Heat and Power (CHP). For example, a UK University has utilised two gas-powered CHPs since 2011, which could generate roughly 2000MWh electricity, equivalent to 8% of the university's total annual electricity consumption. Additionally, extra heating is also provided by the CHPs for several academic buildings and a swimming pool. By optimal managing the CHP, it can help save £70k and 350 tonnes of CO_2 emission annually. In reality, the consumption of natural gas for heating or powering CHP is mainly affected by temperature.

Figure 1 visually shows the relationship between daily average temperature against the daily total gas consumption of a local system in 2011. Obviously, gas consumption changes dramatically with temperature: when the temperature drops continuously, gas consumption rises accordingly. This figure implies that temperature has huge impact on gas consumption.

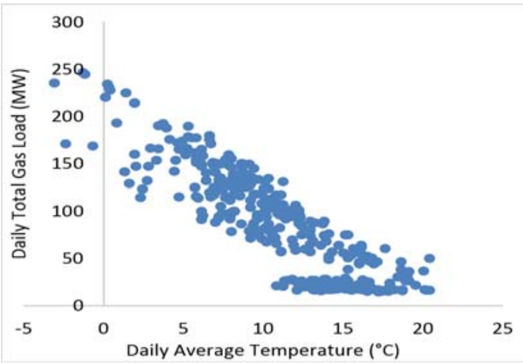


Fig. 1. Daily average temperature against daily total gas load of local system

By identifying the correlation between energy consumption and temperature, a forecasted changing trend represented by linear regression function of gas load might be drawn according to available weather data [6-9]. Therefore, more accurate load analysis and forecasting could be achieved, which enable optimal combined demand response opportunities.

Conventional gas consumption analysis does not take weather condition into account which could cause a high degree of inaccuracy, particularly considering that it has a great impact on gas consumption. In recent years, research starts to explore the relationship between weather and gas load [10-12], but there are some drawbacks:

- Firstly, most literature uses typical load data or maximum load with typical temperature or maximum temperature as input for mathematical modelling [13], resulting in a peak load demand estimation. Though it could help peak demand forecasting, it cannot reflect the load consumption apart from peak demand [14-16].

Hantao Wang (h.wang@bath.ac.uk), Chenghong Gu (cg277@bath.ac.uk) and Furong Li (eesfl@bath.ac.uk) are with the Department of Electronic and Electrical Eng., University of Bath, Bath, UK. Xin Zhang (xin.sam.zhang@gmail.com) is with the National Grid, UK. Lihong Gu (liongu9@126.com) is with the East Branch of State Grid Corporation, Shanghai China.

- Secondly, some papers focus on hourly analysis to identify load demand sensitive with temperature [17-20]. However, load demand is more volatile where human activity is inevitable such as special events. Such events may alter the way that people use energy, which is especially obvious in a short period. Consequently, lower accuracy degree of correlation is obtained [21].
- Thirdly, a few previous literature takes many types of customers including industrial and commercial users into consideration and conclude a general and comprehensive analysis of weather sensitivity. Nevertheless, no mathematical model is given as energy usage of the different type of customers varies dramatically [22].

In order to overcome the disadvantages of previous research, in this paper, regression is applied for data processing and quantifying the correlation between temperature and gas consumption in a mathematical way. Here, the temperature is the only considered weather element as it directly reflects gas consumption for heating. This work is achieved based on daily analysis on gas consumption and temperature. As daily analysis reflects smoother changes of gas consumption, it can eliminate the uncertainty in the hourly-scale analysis. Gas load in different time period, for example during weekday and weekend and during day and night in change with temperature are also explored to reflect influences from human activities. An outlier detection technique calculating the Mahalanobis distance (M-distance) is utilised to ensure the accuracy of used data, where some gas and temperature data points diverging greatly from the normal data set is defined as outliers. M-distance is widely used in cluster analysis and classification for detecting outliers, especially in the development of linear regression models. The real-time annual gas consumption of a real local energy system is used to demonstrate the analysis technique to directly identify the correlations.

Compared with previous work in [13-22], this paper has the following merits. Firstly, all gas load and temperature data are real data recorded, thus the results by using the proposed method can truly reflect gas and temperature correlations in the local energy system. By using the proposed method of EMD and linear regression considering various conditions under different time resolutions, a comprehensive correlation study between gas load and temperature is carried out. Secondly, the accuracy of the results is validated by using a whole-year data and compared with other studiesss by conducting the study with data collected from only a single day or specific days. Therefore the results in this paper are more representative. Many scenarios also consider different seasons and different human activities in weekdays and weekends. It could be concluded that this paper conducts a range of scenarios and cases to explore the the correlation between gas load and temperature. This paper provides strong evidence to prove temperature weight significantly in gas load consumption and potential prediction.

The remainder of the paper is organized as follows: In Section II, outlier detection, EMD process, and linear regression are introduced. Results and analysis of case studies are in Section III. A brief conclusion is given in Section IV.

II. Methodology

This section introduces the method to quantify the correlation between gas demand and temperature, including outlier analysis, Empirical Mode Decomposition (EMD), and regression. Outlier analysis is used to detect faulty or error data due to measurement mistakes or data absence. Mahalanobis distance is applied for outlier identification. Empirical Mode Decomposition is a process to decompose both gas and temperature data into several intrinsic mode functions. Results from this process will be used to study the correlation between gas and temperature, and determine the correlation by using linear regression analysis.

A. Outlier detection

In statistics, an outlier of a set of data is defined as data points that diverge greatly from the other data points [23]. This concept is introduced in this study as there might be some temperature and load data that diverges greatly from other general patterns. Such outliers will be detected and excluded in linear regression analysis. By using statistical methods [24] for outlier detection, those data points that locate relatively far from the whole data set distribution could be identified.

The Mahalanobis distance [25] is a very popular criterion in outlier identification, which defines the distance from observation point to the major data set distribution [26]. For given n observation points from a p -dimensional data set, by denoting the sample covariance matrix by $\mathbf{V_n}$ and the mean of the sample data by $\bar{x}_n$, $\mathbf{V_n}$ could be calculated as [27]:

$$\mathbf{V_n} = \frac{1}{n-1}\sum_{i=1}^{n}(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})^\mathbf{T} \tag{1}$$

where $(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})^\mathbf{T}$ indicates the transpose matrix of $(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})$.

In this study, gas consumption and temperature are considered and thus the sample vector x_i is a 2-dimensional vector. The Mahalanobis distance is represented by:

$$M_i = (\sum_{i=1}^{n}(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})^\mathbf{T}\mathbf{V_n^{-1}}(\mathbf{x_i}-\bar{\mathbf{x}}_\mathbf{n})) \tag{2}$$

where M_i is the Mahalanobis distance for each multivariate data i=1,2,…,n, and $\mathbf{V_n^{-1}}$ is the inverse matrix of $\mathbf{V_n}$.

The Mahalanobis distance for each data point changes according to the size of the whole data set. Usually, there are no specific thresholds to define an outlier. Outliers are quantified as those observation points with relatively large Mahalanobis distance compared with other data points [28]. The implementation of this outlier detection is demonstrated in Case B in the case study section. It is primarily suitable for low-dimensional data [29]. In this paper, the original gas and temperature data pairs of half-hour temporal resolution are used. Sometimes there are data outliers due to mismeasurement or data absence of either gas load or temperature. Thus, it is necessary to discover and exclude data outliers before EMD and linear regression applied. Without identifying and removing bad data, the analysis of the following steps and the accuracy of results can be jeopardized.

B. Empirical Mode Decomposition (EMD)

EMD is used for nonlinear and non-stable signal processing. EMD process is adaptive and highly efficient. In this study, by using EMD, gas consumption and temperature data pairs can be decomposed into several Intrinsic Mode Functions (IMFs) and a residue with lower frequency. All IMFs must satisfy two conditions:: i) during the whole data set, the number of the extrema and the number of the zero crossing point differ no more than one; ii) the average value of the upper envelope and the lower envelope must equal zero. IMFs are usually generated by the following three steps:

- By applying cubic spline to calculate the mean value of the upper and lower envelopes of the original signal *s(t),* represented by *m(t).*
- To obtain the difference between *s(t)* and *m(t)*, *h(t)*:
  $$h(t) = s(t) - m(t) \tag{3}$$
- To treat *h(t)* as new *s(t)* and rerun previous two steps until the new function meets stopping criteria that the mean value of the upper and lower envelopes of *h(t)* is nearly zero. Otherwise rerun the first two steps. When EMD process is finished, the last function is regarded as residue.

C. Linear Regression

In this case, after using EMD, a set of IMFs for both gas load and temperature will be obtained respectively. Each of IMF from gas load will be in pair with the corresponding IMF from temperature to calculate the correlation coefficient introduced in Section In this study, linear regression technique is applied as it could directly present the correlation between gas consumption and temperature with a numerical index. Such index can simply reflect the level of how load corresponds to the temperature change. Linear regression is widely used to analyse the relationship between two quantitative variables by measuring two discriminative coefficients: correlation coefficient r_{xy} and coefficient of determination R^2.

The correlation coefficient r_{xy} is introduced to show the degree and direction of the linear relationship between two variables, in this paper, which are gas consumption and temperature. It is generally defined as the covariance of two variables divided by the product of separate standard deviations [30], represented by:

$$r_{xy} = corr(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{4}$$

where, cov (X,Y) is the covariance and $\sigma_X \sigma_Y$ are the standard deviations of X and Y respectively.

In general, the population mean μ_X and μ_Y are unknown. In order to calculate the sample correlation coefficient r_{xy}, the unbiased estimation of covariance is used:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

Where, x_i and y_i are samples from the population X and Y, and $\bar{x}$ and $\bar{y}$ are the sample means.

Typically, r_{xy} ranges between +1 and -1. Table I shows the level of correlation values of r_{xy}.

TABLE I
LEVEL OF CORRELATIONS OF R_{XY}

$\lvert r_{xy} \rvert$ range	Demonstration
1	Perfectly positive or negative linear correlation
0	No correlation
(0,0.3]	Barely linear correlation
(0.3,0.5]	Weakly linear correlation
(0.5,0.7]	Moderately linear correlation
(0.7,0.1)	Strongly linear correlation

Another discriminative coefficient is the coefficient of determination denoted as R^2. It indicates the degree of scatter points related to the regression line [31]

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \tag{6}$$

where, SS_{res} is the sum of squares of residual and SS_{tot} is the total sum of squares.

In this case, R^2 can be calculated as square of correlation coefficient r_{xy}

$$R^2 = r_{xy}^2 \tag{7}$$

In general, the closer of R^2 to 1, the higher degree of sample scatter point is close to the trend line, indicating better data fitting from the statistical regression functions. In this study, as there are several IMFs pairs from gas load and temperature after using EMD, several correlation coefficients will also be obtained. The final correlation coefficient is defined as the average value of correlation coefficient from each IMF pair.

D. Implementation

The major steps are as follows and the flowchart methodology is shown in Fig.2.

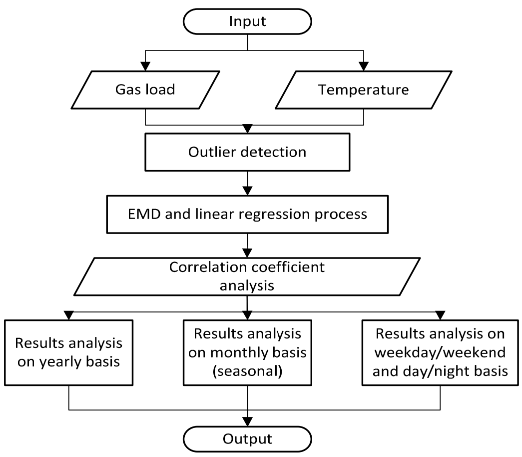


Fig.2. Methodology flowchart

- Half-hourly temperature data is extracted along with corresponding gas load used as original inputs and then to calculate daily average temperature and daily total gas load to form gas load- temperature data pairs.

- By calculating the Mahalanobis distance of each pair of gas load-temperature data point, the outliers that diverge far from the main data set are detected and excluded.
- Linear regression is then used to calculate the correlation coefficient and EMD is applied to improve the correlation identification. In this way, the relationship between gas consumption and temperature could be obtained.

III. Case Study

Different case studies on a local UK energy system are presented to demonstrate the results. Results are shown according to daily time intervals and with respect to various temperature intervals. This section is organised as follows:

- Section A shows the utilisation of outlier detection by measuring Mahalanobis distance of each gas-temperature data pair and those pairs with relatively high Mahalanobis distance are removed.
- Section B discusses how EMD improves the correlation between gas and temperature compared with the case without EMD processed data.
- Sections C, D present the correlation identification between gas consumption and temperature from yearly and monthly perspectives.
- Section E and F discuss different gas-temperature scenarios during weekday/weekend and day/night to explore how such correlation changes in regards to different customer activities.
- Section G is an additional section that uses the proposed method on gas load and temperature data from a different country to exam its performance.

Each of the section focuses on the results of different methods proposed in section II. Section A mainly shows how outlier detection works before applying EMD and linear regression. Section B compares the correlation between gas load and temperature without and with EMD process. Sections C to G discuss different gas load-temperature relationship during different time intervals, from half-hour to year.

As the UK locates at the very west of Europe and east of Atlantic Ocean, the weather conditions in spring and autumn are relatively mild. Additionally, the high latitude of UK makes the summer moderate for human activity and a great amount of heating demand appears in winter. Therefore, winter is a perfect season to explore the relationship between gas load and temperature. Nevertheless, other seasons and annual performance will also be analysed to comprehensively understand how consumption changes with weather conditions.

In this study, the temperature data are collected from Paul Wilman Bath Weather [32] and Weather Underground [33]. The gas load is collected from 7 meters around Bath University campus. All gas load data is provided by the Department of Estates of the university [34].

A. Outlier Detection

By measuring Mahalanobis distance for each gas-temperature data pair, the outliers that are far from the main data set distribution could be identified. Take the sample data of November 2011 as an example and the results are shown in Figure.3. The colour bar on the right indicates the value of Mahalanobis distance. From blue to green to red and then to dark brown, the Mahalanobis distance increases continuously. There is one gas-temperature data point that diverges greatly from others coloured in dark brown. Its Mahalanobis distance is M=5.9810. Such point is therefore identified as an outlier. Physically, it indicates at a relatively low temperature (8.6 °C), there is a low gas consumption of 87,152kWh. This is abnormal compared with the general gas-temperature pattern, which should be excluded from further analysis.

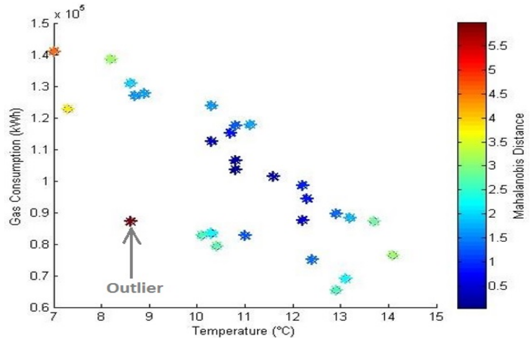


Fig. 3. Mahalanobis distance of gas-temperature data of November 2011

By excluding the outlier from the data set, the accuracy of data is achieved. The wrong data recorded, misreading by both meter and human-made could be eliminated. It provides a tool to handle the complexity and errors of the gas-temperature data. The data in the following case studies are all processed with this outlier detection technique to ensure the accuracy of the data.

B. EMD data processing

Table II lists some correlation results between gas load and temperature in different time scales. The first two measure daily total gas load and daily average temperature for a month and last two measure 1-month-average value of every half hour gas load and temperature data pairs. The correlation of both real time data before and after using EMD is relatively high between gas load and temperature in different time scales. After EMD, their absolute values increase, indicating the IMFs of gas load and temperature are more correlated than before EMD.

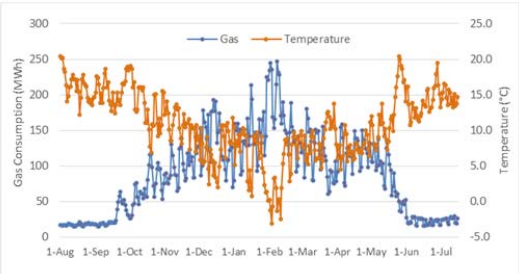
TABLE II
COEFFICIENT OF DIFFERENT CASES

Temperature (X)	Gas Load (Y)	Correlation coefficient r_{xy}
Real Daily Average	Real Daily Total	-0.6049
EMD Daily Average	EMD Daily Total	-0.8143
Real Half Hour Average	Real Half Hour Total	0.4525
EMD Half Hour Average	EMD Half Hour Total	0.9385

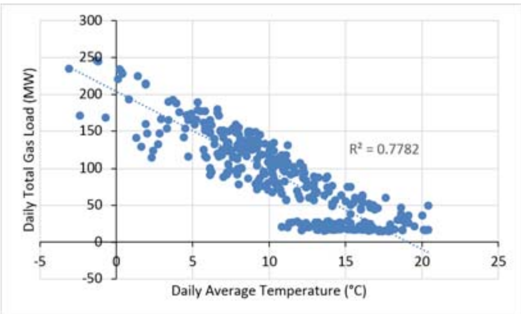
In the first two rows of Table II, the correlation between gas and temperature changes from -0.6049 using real-time data to -0.8143 by using EMD technique. It means that on daily basis, gas demand is negatively related with temperature. The last two rows of Table II shows that gas demand is positively correlated with temperature when the time is scaled down to half hour basis. After applying EMD on the real-time data, such correlation increased greatly from 0.4526 to 0.9385. It is prominent that EMD has good performance for improving the identification of the correlation between gas and temperature while applying linear regression

C. Annual Gas and Temperature Correlation

In this case, every gas consumption data point is the total daily consumption and temperature point is the average temperature of the same day. Figure.4 illustrates the total daily gas consumption with an average temperature of the days from August 2011 to July 2012. Figure.4a is the real-time daily gas consumption against the temperature in 2011-2012 and Figure.4b is their scatter point graph with the trendline (linear regression equation) and correlation of determination R^2.



a. Gas consumption and temperature of 2011-2012



b. Gas-temperature scatters 2011-2012

Fig. 4. Gas-temperature change in 2011-2012

Figure.4a clearly shows that daily gas consumption always changes conversely against temperature annually. While it is hot summer at the beginning of the academic year 2011-2012, gas demand remains at a relatively low level. With climate becoming colder, the gas consumption begins to rise continuously and reached its peak in February 2012. Then the weather becomes warmer, and the gas demand reduces again.

Figure.4b presents the scatter points of the gas-temperature dataset and results from linear regression with temperature as input and gas demand as output. The annual correlation of determination R^2 is 0.7782 and the correlation coefficient r_{xy} is -0.8822, indicating that a strong correlation is between daily gas consumption and temperature. The regression coefficients obtained by using coefficients in Figure.4b is helpful for understanding the impact of temperature on gas consumption.

However, such index obtained from annual data is not accurate enough for short-time understanding of the correlation as it presents a general pattern of the correlation over a long-term. A more appropriate equation according to different seasons and daily time periods can achieve better accuracy of correlation and will be demonstrated next.

D. Monthly Correlation in Four Seasons

This subsection identifies correlation in seasons based on daily total gas consumption against daily average temperature. In Aug and Nov 2011 the average monthly temperature were 16.2 ℃ and 10.8 ℃ respectively, while that of Feb and May 2012 were only 4.5 ℃ and 12℃. Climates of each month are typical climates to represent every season.

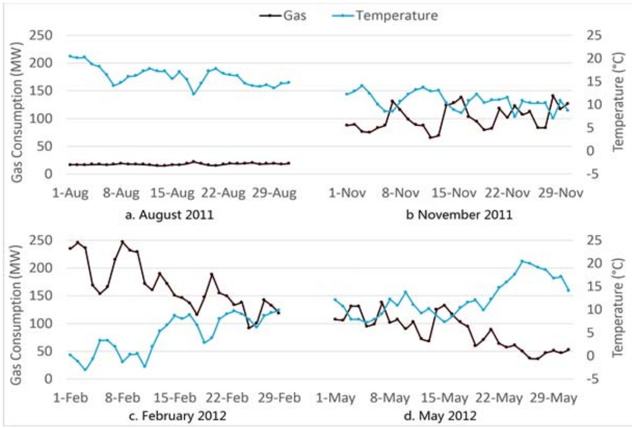


Fig. 5 Monthly gas consumption and temperature in 2011-2012.

In Figure.5, the curve labelled on the left represents the total daily gas consumption and the curve with unit labelled on the right represents the trend of average daily temperature. It is very clear that while the average daily temperature changes, the total daily gas load changes in the other way around. This is especially obvious in Figure.6c (February 2012), which is the coldest month of each year. For the first one-third of February 2012, the weather was relatively cold and fluctuated and the daily gas demand corresponded to it conversely. Then, temperature continuously climbed and the gas load decreased accordingly. During the last period of that month, temperature fluctuated again and gas consumption went up and down accordingly in an opposite direction.

TABLE III
COEFFICIENT OF FOUR SEASONS

Time	Correlation coefficient r_{xy} before EMD	Correlation coefficient r_{xy} after EMD
Spring (2012.05)	-0.7646	-0.9375
Summer (2011.08)	-0.6836	-0.9195
Autumn (2011.11)	-0.7209	-0.9365
Winter (2012.02)	-0.8309	-0.9406

According to the methodology in Section II, the coefficient R^2 and r_{xy} could be calculated, where x represents daily average temperature and y represents the total daily gas consumption. The results are shown in Table III.

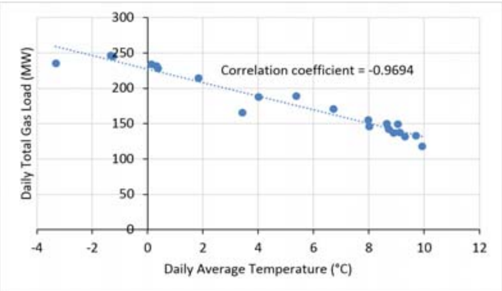
Compared to other seasons, summer time (Aug 2011) has the lowest correlation level of -0.6836. According to the correlation coefficient margins in Table I, it indicates that the daily gas consumption has a moderately linear correlation with daily average temperature. This is reasonable as during summertime, as there is a small amount of heating demand consuming gas. Therefore, no obvious correlation could be revealed.

On the other hand, the correlation from the data in winter (Feb 2012) illustrates more obvious correlation between gas load and temperature with r_{xy} of -0.8309. The negative coefficients are very close to -1, indicating a strong negative linear regression. The regression results of Feb and May 2012 are more obvious than those of Aug and Nov 2011, mainly due to their different temperature levels. When the temperature is lower, more gas is consumed for heating and thus, the regression sensitivity between load and temperature is significantly affected.

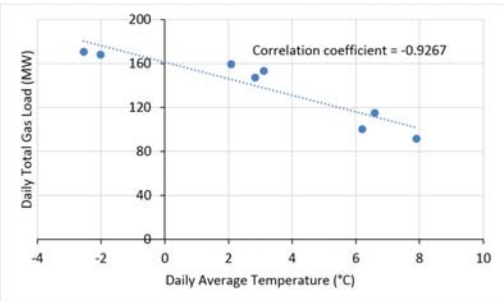
By using EMD on the data and calculating the correlation coefficient of each month, it could be seen that the correlation between gas load and temperature have been improved significantly. Even the least correlation level in summer time (August 2011) has been increased to -0.9195.

E. Correlation Change during Weekday and Weekend

In this section, correlation change during weekday and weekend is analysed to explore how the temperature-sensitive gas load is affected by different customer behaviours. Figure.6 shows the scatter point of gas-temperature data pairs of weekday (Figure.6a) and weekend (Figure.6b) using the data of February 2012 and Table IV lists correlation results.



a. Gas-temperature scatter on weekday of February 2012



b. Gas-temperature scatter on weekend of February 2012

Fig. 6. Absolute value of r_{xy} of every half hour in February 2012

Figure.6 divides the data into weekday and weekend. While calculating the correlation coefficient without clustering them, the result is -0.8309 before using EMD. In Figure.6a, the correlation coefficient during the weekday is -0.9694 and in Figure.6b, that coefficient during the weekend is -0.9267. Both correlation levels are improved. This indicates that gas load during weekdays and weekends is sensitive to temperature and shares different pattern according to different customer activities. Through both correlations are very high, the one during the weekday is slightly higher than that during the weekend. This implies that gas load during working days are more related to temperature than that during weekends.

Table IV shows the correlation change before and after using EMD. For both weekdays and weekends, the correlation levels are increased by using EMD.

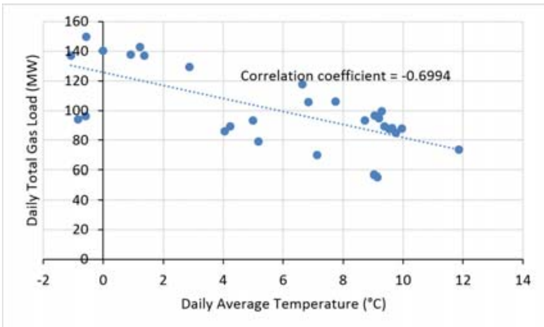
TABLE IV
COEFFICIENT OF WEEKDAY AND WEEKEND

Time	Correlation coefficient before EMD	Correlation coefficient after EMD
Weekday	-0.9694	-0.9993
Weekend	-0.9267	-0.9960

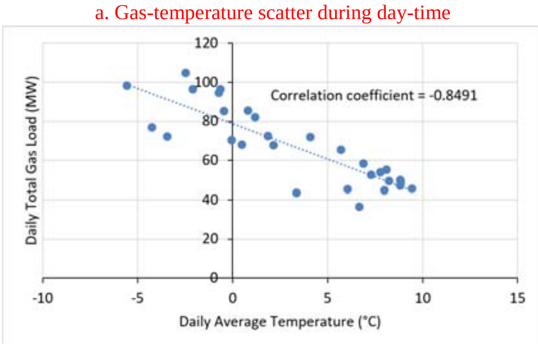
From Figure.6 and Table IV, it could be concluded that the gas-temperature correlation experiences much higher values when the data is spilt into weekdays and weekends than without considering date type. This indicates that human activities can significantly affect gas consumption and another case will be analysed in the next section.

F. Correlation Change during Day and Night

This section studies the correlation change between day (08:00 20:00) and night and Figure.7 depict the points of gas consumption and temperature.



a. Gas-temperature scatter during day-time



b. Gas-temperature scatter during night

Fig. 7. Correlation during day and night (February 2012)

As seen, during day-time, the correlation coefficient is -0.6994 and at night, it is -0.8491. This reveals a signal that gas load at nights are generally more closely correlated to temperature than that during day time. After applying EMD on the data, the correlation coefficient for both day time and night time are increased greatly, shown in Table V.

TABLE V
COEFFICIENT OF DAY AND NIGHT

Time	Correlation coefficient before EMD	Correlation coefficient after EMD
Day (08:00 – 20:00)	-0.6994	-0.8464
Night (other time)	-0.8491	-0.9634

In Table V, the correlations of day and night are negatively increased from -0.6994 and -0.8491 to -0.8464 to -0.9634, respectively. On one hand, this result indicates that during day and night, the temperature has different impacts on gas load and this could have positive influence to identify the weighting of temperature in future load forecasting for different time period. On the other hand, EMD is proven to be a promising tool to increase the correlation level between gas load and temperature.

G. Correlation Analysis and Comparisons in Australia

In this section, another data set from Australia is used to validate the proposed method. The daily total gas load data is collected form [35] and the daily average temperature data is from[36]. Results are shown in Table VI.

TABLE VI
COEFFICIENT OF DATA IN AUSTRALIA

Time	Coefficient before EMD	Coefficient after EMD
2016.09.28 to 2016.10.27	-0.7310	-0.9582

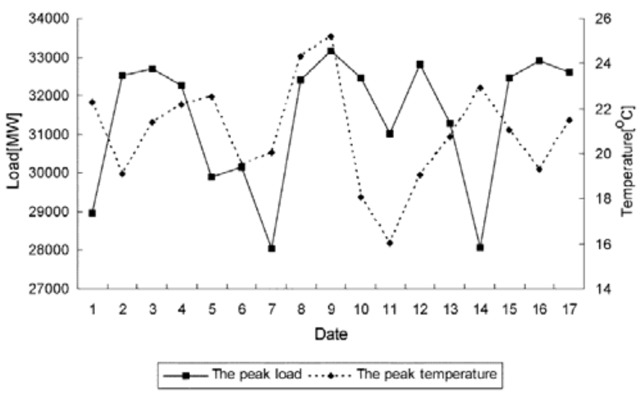


Fig. 7. Correlation during day and night (February 2012)[14]

TABLE VII
COEFFICIENT OF TWO PAPERS

Data	Correlation coefficient using method from [14]	Correlation coefficient using proposed method
May 2000 from [14]	-0.51	-0.8466

As seen, the correlation coefficient between gas load and temperature is -0.7310 before using EMD. After applying EMD, the correlation increases to -0.9582. This result is similar to the results in Section D, indicating that the proposed method using outlier detection by calculating M-distance and using EMD to identify the correlation level between gas load and the temperature is not only valid for local energy system in the UK, but also valid in other places.

Figure. 7 is a peak load and temperature from [14] and the correlation coefficient between peak load and the temperature is only -0. 51 (Shown in Table VII), indicating that peak load is less sensitive to temperature. By using the proposed method and decomposing this same set of data in Figure.7 and calculating the average correlation coefficient of each IMF between gas load and temperature, the final correlation coefficient increases to -0.8466. This result indicates that after EMD process, each IMF component of gas demand and temperature respectively is more related to each other by revealing a much higher level of the correlation coefficient.

IV. CONCLUSION

In this paper, a novel method is proposed to analysis the correlation between gas consumption and temperature. Mahalanobis distance is a promising mean for outlier detection and EMD substantially improves the correlation level between gas load and temperature. By finding correlation coefficient between real-time gas consumption and temperature, it is seen that gas load is negatively related to temperature. In addition, such correlation changes with the overall temperature level. In winter, gas and temperature are in highest genitivally correlated and this correlation drops when the temperature is higher in spring and autumn. By dividing the data into weekday/weekend and day/night, it is found that during different time periods, the correlation is stronger during weekday and night than that during weekend and day time, respectively. After applying EMD technique, correlation levels are significantly increased in all different case studies. To summarise, the novel method proposed in this paper identifies the temperature sensitive part of the gas load and measures the correlation between gas load and temperature as most of the research focus the relationship between electricity load and temperature and very few focus on the gas load and temperature. It provides the reference for weight temperature as a variable in gas load forecasting.

In future work, polynomial regression techniques will be applied to analysis with weather factors such as humidity, wind speed considered. The new concept 'feel like [37]' temperature might replace actual temperature as it directly reveals real feelings against temperature change and would have a direct influence on gas consumption.

REFERENCES

[1] E. A. Feinberg and D. Genethliou, "Load forecasting," *Applied Mathematics for Restructured Electric Power Systems: Optimization, Control, and Computational Intelligence,* p. 17, 2005.
[2] S. Li, D. Zhang, A. B. Roget, and Z. O'Neill, "Integrating Home Energy Simulation and Dynamic Electricity Price for Demand Response Study," *Smart Grid,* vol. vol.5, p. 9, March 2014.
[3] L. Waters, V. Goodright, and E. Wilkes, "Energy Consumption in the UK (2015) Chapter 3: Domestic energy consumption in the UK between 1970 and 2014," D. o. E. C. Change, Ed., ed. London, UK, 30.07.2015.
[4] C. Shao, X. Wang, M. Shahidehpour, X. Wang, and B. Wang, "An MILP-Based Optimal Power Flow in Multicarrier Energy Systems," *IEEE Transactions on Sustainable Energy,* vol. 8, pp. 239-248, 2017.

[5] A. Shabanpour-Haghighi and A. R. Seifi, "Energy Flow Optimization in Multicarrier Systems," *IEEE Transactions on Industrial Informatics,* vol. 11, pp. 1067-1077, 2015.

[6] Z. Yu, "A temperature match based optimization method for daily load prediction considering DLC effect," *IEEE Transactions on Power Systems,* vol. 11, pp. 728-733, 1996.

[7] S. Fan and R. J. Hyndman, "Short-Term Load Forecasting Based on a Semi-Parametric Additive Model," *IEEE Transactions on Power Systems,* vol. 27, pp. 134-141, 2012.

[8] Y. Goude, R. Nedellec, and N. Kong, "Local Short and Middle Term Electricity Load Forecasting With Semi-Parametric Additive Models," *IEEE Transactions on Smart Grid,* vol. 5, pp. 440-446, 2014.

[9] P. Mirowski, S. Chen, T. K. Ho, and C. N. Yu, "Demand forecasting in smart grids," *Bell Labs Technical Journal,* vol. 18, pp. 135-158, 2014.

[10] T. Karimi, M. R. Sadeghimoghadam, and H. G. Shakouri, "Investigation of Temperature Variation Effects on Natural Gas Consumption in Iran," presented at the Power and Energy Engineering Conference (APPEEC), 2011 Asia-Pacific, 25-28 March 2011.

[11] M. Simunek and E. Pelikán, "Temperatures data preprocessing for short-term gas consumption forecast," presented at the Industrial Electronics, ISIE 2008, June 30 2008.

[12] M. Fagiani, S. Squartini, R. Bonfigli, and F. Piazza, "Short-term load forecasting for smart water and gas grids: A comparative evaluation: A comparative evaluation," presented at the Environment and Electrical Engineering (EEEIC), 2015 IEEE 15th International Conference on 10-13 June 2015.

[13] A. P. Douglas, A. M. Breipohl, F. N. Lee, and R. Adapa, "The impacts of temperature forecast uncertainty on Bayesian load forecasting," *Power Systems, IEEE Transactions on,* vol. 13, no.4, p. 6, Nov 1998.

[14] K.-B. Song, S.-K. Ha, J.-W. Park, D.-J. Kweon, and K.-H. Kim, "Hybrid load forecasting method with analysis of temperature sensitivities," *Power Systems,* vol. vol.21, no.2, p. 7, May 2006.

[15] H. S. Hippert and C. E. Pedreira, "Estimating temperature profiles for short-term load forecasting: neural networks compared to linear models," *Generation, Transmission and Distribution, IEE Proceedings-* vol. 151, no.4, p. 4, 11 July 2004.

[16] J. Valenzuela and M. Mazumdar, "Influence of temperature and load forecast uncertainty on estimates of power generation production costs," *Power Systems, IEEE Transactions on* vol. 15, no.2, p. 6, May 2000.

[17] C. Genyong and S. Jingtian, "Study on the methodology of short-term load forecasting considering the accumulation effect of temperature," *Sustainable Power Generation and Supply, 2009. SUPERGEN '09. International Conference on,* p. 6, April 2009.

[18] A. Khotanzad, M. H. Davis, A. Abaye, and D. J. Maratukulam, "An artificial neural network hourly temperature forecaster with applications in load forecasting," *Power Systems, IEEE Transactions on,* vol. 11, no.2, p. 6, May 1996.

[19] T. Senjyu, H. Takara, K. Uezato, and T. Funabashi, "One-hour-ahead load forecasting using neural network," *Power Systems, IEEE Transactions on* vol. 17, no.1, p. 5, Feb 2002.

[20] T. Haida and S. Muto, "Regression based peak load forecasting using a transformation technique," *Power Systems, IEEE Transactions on,* vol. 9, no.4, p. 6, Nov 1994.

[21] J. P. Rothe, A. K. Wadhwani, and S. Wadhwani, "Short Term Load Forecasting Using Multi Parameter Regression," *International Journal of Computer Science and Information Security, IJCSIS, Vol. 6, No. 2, pp. 303-306, November 2009, USA,* 2009.

[22] C. S. Chen, M. S. Kang, J. C. Hwang, and C. W. Huang, "Temperature effect to distribution system load profiles and feeder losses," *Power Systems,* vol. vol.16, no.4, p. 5, Nov 2001.

[23] F. Angiulli and C. Pizzuti, "Outlier mining in large high-dimensional data sets," *Knowledge and Data Engineering, IEEE Transactions on,* vol. 17 no.2, pp. 203,215, Feb. 2005.

[24] T. S. Arthanari, *Mahalanobis Distance in Multivariate Measurement Systems*. University of Auckland. Department of Management Science and Information Systems: University of Auckland, 2003.

[25] Y. Washizawa and S. Hotta, "Mahalanobis Distance on Extended Grassmann Manifolds for Variational Pattern Analysis," *Neural Networks and Learning Systems, IEEE Transactions on* vol. 25, no.11, p. 10, Nov. 2014.

[26] G. Verdier and A. Ferreira, "Adaptive Mahalanobis Distance and Nearest Neighbor Rule for Fault Detection in Semiconductor Manufacturing," *Semiconductor Manufacturing, IEEE Transactions on,* vol. 24, no.1, p. 9, Feb. 2011.

[27] I. Ben-Gal, *Outlier detection, In: Maimon O. and Rockach L. (Eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*: Kluwer Academic Publishers, 2005.

[28] C. Shen, J. Kim, and L. Wang, "Scalable Large-Margin Mahalanobis Distance Metric Learning," *Neural Networks, IEEE Transactions on* vol. 21, no.9, p. 6, Sept. 2010.

[29] J. Zhang, "Advancements of Outlier Detection: A Survey," *Scalable Information Systems, ICST Transactions on,* vol. 13, Jan 2013.

[30] A. Molisch, *Diversity*: Wiley-IEEE Press, 2011.

[31] D. Larose, *Regression Modeling*: Wiley-IEEE Press, 2006.

[32] P. B. Weather, "Bath Weather Historic Data," ed. Bath, 2015.

[33] I. The Weather Channel Interactive, "Bath, United Kingdom," ed, 2015.

[34] P. Phelps, "Energy and Environment Report 2012," Department of Estates, University of Bath, Bath, UKJune, 2012.

[35] (2016, 10.27). *Daily Gas Consumption*. Available: https://www.aemo.com.au/Gas/Declared-Wholesale-Gas-Market-DWGM/Data

[36] (2016, 10.27). *Weather Underground*. Available: https://www.wunderground.com/

[37] M. Office, in *What is Feel-like Temperature*, M. O. N. Blog, Ed., ed: Met Office News, 2015.

Hantao Wang is a PhD student with the department of Electronic and Electrical Engineering at the University of Bath. His research is on multi-vector energy system modelling and optimisation.

Chenghong Gu (M'14) received his Bachelor degree and Master degree in electrical engineering from Shanghai University of Electric Power and Shanghai Jiao Tong University, Shanghai, China, in 2003 and 2007 respectively. In 2010, he obtained his PhD from University of Bath, U.K. Now, he is a Lecturer and EPSRC Research Fellow at the University of Bath. His research is in power system economics and multi-vector energy systems.

Xin Zhang received the B.Eng. degree in automation from Shandong University, China, in 2007; the M.Sc. and Ph.D. degrees in electrical power engineering from The University of Manchester, U.K., in 2007 and 2010 respectively. He is currently a Chartered Engineer with National Grid UK. His research interests include power system planning, operation and control.

Furong Li (SM'09). She received her B.Eng. degree in electrical engineering from Hohai University, China, in 1990, and her Ph.D. degree from Liverpool John Moores University, U.K., in 1997. She took up a lectureship in 1997 in the Power and Energy Systems Group at the University of Bath, U.K., where she is now Professor. Her major research interests are in the areas of power system planning, operation, automation and power system economics.

Lihong Gu is an Electrical Engineering the East Branch of State Grid Corporation, Shanghai China. He obtained the MSc and PhD degrees from Shanghai Jiao Tong University in 2006 and 2010 respectively. His research area is in power system stability and control.