Multiple regression analysis on the HVAC energy consumption of railway passenger stations

Z Su¹ and X Li¹,*
¹ Tsinghua University, Beijing, China
* xfli@tsinghua.edu.cn

Abstract. With the rapid development of high-speed railway transport in the last decade, the number of railway passenger stations in China has surged. Distinguished by its large floor space, high quality of service, long operation time and high occupancy rate throughout the year, the electricity consumption of the heating, ventilation and air conditioning (HVAC) system for current railway stations is much higher than other types of public buildings. Despite of its huge amount of energy consumption, the assessment standard and benchmarking for the railway passenger stations is not completely established. Therefore, it is of great significance to study the energy use level and the associated factors of influence. In this paper, extensive field tests have been conducted to collect detailed energy data of stations in different climate zones in China. Multiple regression model is adopted to determine relevant impacting parameters. The results showed that building area tends to be the primary factor affecting the energy consumed by the HVAC system in each of the four climate zones. Furthermore, for each climate region, regression equations are developed, which serve as an effective standard to evaluate the energy consumption of stations.

1. Introduction

By the end of 2015, more than 19,000 km high-speed railway lines have been constructed in China [1]. The energy consumption of railway stations has gradually attracted substantial attention. It has been pointed out that large railway stations consume more energy-consuming than such public buildings as office buildings [2]. Hence, it is of great significance to study the energy use characteristics and the influencing factors, which is useful to evaluate and benchmark the energy consumption of stations. Previous studies have done numerous works to research energy use of railway passenger stations. For instance, it was revealed that a station located in north China consume almost nine times more energy than that of a station in south China [3]. The passenger density and the waiting time is related to the energy consumed by stations [4-5]. Another paper concluded that the cooling system consumes 35~60% of total energy in summer [6]. However, the major lack of these studies is that only a small number of stations were investigated, some are even one or two stations. Consequently, the results are not representative for the large number of stations in different climate zones in China.

The widely adopted approach to analyse buildings energy consumption is the linear regression method, which reveals the correlation between possible factors and the energy use. This model has been used to evaluate energy use of various kind of buildings in many former studies. In previous researches, gross floor area, number of guest rooms, occupancy rate, and building construction/retrofit year was found to have influence on the energy use amount of hotel buildings [7-9]. Similar researches were done to other types of buildings such as shopping malls, office buildings, etc. [10-12]. The multiple linear regression analysis is an advanced method, which can comprehensively reflect the relationship. However, it has not been adopted in the field of energy analysis of railway station buildings.

Compared to former studies, this study adopts the multiple linear regression method to analyse the significant influencing factors of energy consumption for railway passenger stations. Moreover, this
study has collected more data in each climate zones, making the results more representative. Based on the extensive data collection and multiple regression method, the energy equations are given to represent the typical energy levels of stations in each climate region.

2. Methodology

2.1 The studied stations

China has a vast territory with diverse climatic conditions. According to China's Code for the Design of Civil Buildings (GB 50352-2005), the territory is divided into 5 climate zones to make the buildings fully utilize and adapt to the local climatic conditions, including extreme cold region (EC), cold region (COLD), hot-summer and cold-winter region (HSCW), hot-summer and warm-winter region (HSWW), and mild region (MILD). This study includes stations located in each climate zones in China. Furthermore, the energy utilization level is also related to other factors such as the building area, the number of passengers, station level, etc. In order to make the results more representative, all of the aforementioned factors are taken into consideration in the process of the case selection and energy analysis. Eventually, in this study, 80 stations are selected as the studied cases. The fundamental information of stations in each climate zones has been shown in Figure1-3.

![Climate zones and studies stations](image1)

**Figure 1.** Climate zones and studies stations

![Station area of the studied stations](image2)

**Figure 2.** Station area of the studied stations
2.2 Data collection
Data collected in this study contains building area, building storey, station level, number of passengers, years of construction, heating and cooling equipment, lighting conditions energy consumption of each system, etc. Station information and energy consumption data were predominantly obtained from the unpublished Special Energy Consumption Investigation of Chinese Large Railway Station. Then the multiple linear regression method is applied to identify the important factors among a large set of possible variables.

3. Energy utilization status
The energy consumption has been converted to electricity consumption using the electricity-equivalent conversion method [13]. Then the energy breakdown of each sub-system can be compared, as presented in Figure 4. Considering the low energy use amount of the stations located in the mild region, the energy consumption of buildings in this region is not discussed in this study. It can be concluded that the total energy consumption tends to increase as the climate is colder. As far as the cooling/heating energy use, it is closely related to the climate condition. In the extreme cold region, heating energy consumption takes up 68% of the total amount, with the cooling energy consumption accounting for merely 6%. As a result, carrying out the energy-saving technologies or operation patterns to renovate the heating system is of great significance, the opposite of which is true for the hot-summer and warm-winter region.

4. Multiple regression model
4.1 Total energy consumption
According to former researches, various factors may influence the differences among the energy consumptions, such as passenger flow density, climate difference, station scale and local economic development level [4,7-9]. Considering the results of relative former studies and characteristics of station buildings, in this study, the building area (A), number of passengers (N), years after the construction (Y), years after the last major retrofit (Y’), height of the building (H), layers of the building (L) and regional Gross Domestic Product (GDP) are selected in the multiple regression model for the energy consumption of the stations (E). The regression results are listed in Table 1. The adjusted R² ranges from 0.609 to 0.898, indicating that the model is reliable and applicable. And the values of β₀~β₇ is different for the four climate zones. Judging by the P-Value, the building area is the important influencing parameter for stations in all of the four climate zones. Apart from that, the number of passengers and regional GDP affect the energy consumption for stations in some climate zones.

\[ E = \beta_0 + \beta_1 A + \beta_2 N + \beta_3 Y + \beta_4 Y' + \beta_5 H + \beta_6 L + \beta_7 GDP \]

**Table 1.** Multiple regression results for the total energy consumption in each climate zones

| (a) EC Region Regression Outline | Variable | Coefficients | Standard Error | P-value | Confidence Interval |
|---------------------------------|----------|--------------|----------------|---------|---------------------|
| Confidence Level 90%            | β₀       | -2523.80     | 3147.39        | 0.481   | -9930.76 4883.16    |
| Multiple R                     | β₁       | 90.77        | 23.08          | 0.029   | 36.45 145.09        |
| R Square                       | β₂       | -207.58      | 349.55         | 0.594   | -1030.20 615.04     |
| Adjusted R Square              | β₃       | 49.48        | 58.70          | 0.461   | -88.66 187.62       |
| Standard Error                 | β₄       | -234.78      | 359.68         | 0.560   | -1081.24 611.67     |
| F                              | β₅       | -37.87       | 133.54         | 0.795   | -352.14 276.39      |
| Significance F                 | β₆       | 763.88       | 987.95         | 0.496   | -1561.13 3088.88    |
| Number of Cases                | β₇       | 16.63        | 10.17          | 0.201   | -7.31 40.57         |

| (b) COLD region Regression Outline | Variable | Coefficients | Standard Error | P-value | Confidence Interval |
|-----------------------------------|----------|--------------|----------------|---------|---------------------|
| Confidence Level 90%              | β₀       | -4797.68     | 3623.18        | 0.205   | -11149.30 1553.94   |
| Multiple R                       | β₁       | 47.26        | 19.61          | 0.029   | 12.88 81.63         |
| R Square                         | β₂       | 467.74       | 161.58         | 0.011   | 184.48 751.00       |
| Adjusted R Square                | β₃       | -22.80       | 39.03          | 0.568   | -91.21 45.62        |
| Standard Error                   | β₄       | 87.21        | 169.85         | 0.615   | -210.55 384.97      |
| F                                 | β₅       | 88.00        | 153.04         | 0.574   | -180.29 356.29      |
| Significance F                   | β₆       | 1039.33      | 792.33         | 0.209   | -349.67 2428.34     |
| Number of Cases                  | β₇       | 0.46         | 3.66           | 0.902   | -5.96 6.88          |

| (c) HSCW region Regression Outline | Variable | Coefficients | Standard Error | P-value | Confidence Interval |
|-----------------------------------|----------|--------------|----------------|---------|---------------------|
| Confidence Level 90%              | β₀       | -756.36      | 3436.76        | 0.828   | -6664.52 5133.80    |
| Multiple R                       | β₁       | 56.35        | 19.62          | 0.009   | 22.72 89.97         |
| R Square                         | β₂       | 385.04       | 190.62         | 0.055   | 58.34 711.74        |
| Adjusted R Square                | β₃       | 94.22        | 77.07          | 0.234   | -37.86 226.30       |
| Standard Error                   | β₄       | -371.64      | 266.83         | 0.177   | -828.95 85.67       |
| F                                 | β₅       | -114.73      | 121.45         | 0.355   | -322.88 93.42       |
| Significance F                   | β₆       | -186.81      | 615.26         | 0.764   | -1241.29 867.68     |
| Number of Cases                  | β₇       | 7.29         | 2.84           | 0.017   | 2.43 12.16          |

| (d) HSWW region Regression Outline | Variable | Coefficients | Standard Error | P-value | Confidence Interval |
|-----------------------------------|----------|--------------|----------------|---------|---------------------|
| Confidence Level 90%              | β₀       | 1102.45      | 5922.29        | 0.857   | -9910.33 12115.23   |
| Multiple R                       | β₁       | 133.08       | 26.86          | 0.001   | 83.13 183.03        |
| R Square                         | β₂       | -403.41      | 563.51         | 0.494   | -1451.29 644.47     |
| Adjusted R Square                | β₃       | 76.11        | 124.88         | 0.559   | -1561.11 308.34     |
| Standard Error                   | β₄       | -364.75      | 625.35         | 0.576   | -1527.63 798.12     |
| F                                 | β₅       | -437.97      | 242.31         | 0.108   | -888.56 12.62       |
| Significance F                   | β₆       | 1360.38      | 1280.23        | 0.319   | -1020.27 3741.03    |
| Number of Cases                  | β₇       | 16.56        | 9.53           | 0.121   | -1.17 34.29         |
4.2 Energy consumption for the heating system

When analyzing the energy use of the heating system, the heating sources need to be considered, which can be divided into centralized and decentralized heating. For the extreme cold and cold region, centralized heating takes up 64% and 45% respectively, serving as the most commonly used heating source. Whereas for the hot-summer and cold-winter region, decentralized heating takes up the majority. And for the hot-summer and warm-winter region, the heating system is not equipped in most of the stations, as a result of which, it is not included in this part.

Figure 5. Proportion of heating source

Multiple regression model with qualitative variable is established and the results are shown in Table 2. The building area is the primary influencing factor on the heating energy consumption for stations in all the four climate zones. From the value of the fitting coefficient $\beta_{h1}$, it can be seen that the heating energy use index decreases from the extreme cold region to the hot-summer and cold-winter region, which is related to the average air temperature of each zones.

Moreover, the number of passengers and the regional GDP also matter. The reasons are as follows: (1) The number of passengers mainly affects the demand of fresh air, and therefore influence the heating energy consumption indirectly. However, in the extreme cold region, the outdoor environment is harsh in winter. Consequently, the fresh air supply volume is strict, and therefore the increase of passengers will not arise the great increase of fresh air supply. As for the hot-summer and cold-winter zone, outdoor environment is more moderate in winter. With the increase of fresh air supply, the heating load will not increase significantly. Thus, heating energy use is not sensitive to the number of passengers in these two zones. (2) According to the P-value, GDP affects heating energy consumption for stations in hot-summer and cold-winter region. For stations in this region, the heating system is optional in that the weather in winter is not so cold. Therefore, the energy use of the heating system is related to the regional GDP.

$$E_h = \beta_{h0} + \beta_{h1}A + \beta_{h2}N + \beta_{h3}GDP + \beta_{h4}D_h$$

Table 2. Multiple regression for the heating energy consumption

| Climate Region | EC  | COLD | HSCW |
|----------------|-----|------|------|
| Adjusted R Square | 0.734 | 0.802 | 0.531 |
| Significance F | 0.014 | <0.001 | <0.001 |

| Regression Results | Coefficients | P-value | Coefficients | P-value | Coefficients | P-value |
|-------------------|--------------|---------|--------------|---------|--------------|---------|
| $\beta_{h0}$ | 932.86 | 0.397 | -47.99 | 0.971 | -366.34 | 0.617 |
| $\beta_{h1}$ | 68.85 | 0.008 | 21.66 | 0.091 | 11.01 | 0.080 |
| $\beta_{h2}$ | -89.29 | 0.411 | 370.78 | 0.000 | 87.43 | 0.125 |
| $\beta_{h3}$ | 10.39 | 0.181 | 1.22 | 0.623 | 2.14 | 0.033 |
| $\beta_{h4}$ | -2414.43 | 0.076 | -2374.76 | 0.113 | 1275.53 | 0.321 |

4.3 Energy consumption for the cooling system

As for the cooling energy, the same process is conducted. The cooling modes are divided into the central air-conditioning (CAC) and room air-conditioning (RAC). Most of stations in the extreme cold region maintain indoor thermal comfort by natural ventilation (NV) in summer. Thus, the multiple regression
analysis is not applied in this region.

Similarly, building area is the primary factor affecting the cooling energy use. And the number of passengers exerts significant influence on the cold region and the hot-summer and cold-winter region. This can be illustrated by the fact that the heat gain from the fresh air accounts for the majority of the cooling load. Nevertheless, in the hot-summer and warm-winter region, the heat gain from the solar radiation takes up the majority. Thus, the influence of passengers is subtle in this region.

\[ E_c = \beta_{c0} + \beta_{c1}A + \beta_{c2}N + \beta_{c3}GDP + \beta_{c4}D_c \]

Table 3. Multiple regression for the cooling energy consumption

| Climate Region | COLD  | HSCW  | HSWW  |
|----------------|-------|-------|-------|
| Adjusted R Square | 0.478 | 0.622 | 0.830 |
| Significance F  | 0.005 | <0.001| <0.001|
| Regression Results | Coefficients | P-value | Coefficients | P-value | Coefficients | P-value |
| βh0          | -624.63 | 0.634 | -2665.18 | 0.152 | -2380.25 | 0.139 |
| βh1          | 18.95   | 0.082 | 13.56   | 0.053 | 27.14   | 0.001 |
| βh2          | 153.00  | 0.042 | 142.88  | 0.023 | 28.61   | 0.815 |
| βh3          | -2.15   | 0.320 | 3.24    | 0.004 | 4.09    | 0.140 |
| βh4          | 1444.55 | 0.397 | 1278.23 | 0.473 | 959.40  | 0.570 |

5. Conclusion

This study is an attempt to apply the multiple linear regression model to the energy consumption analyses of railway passenger stations. The energy breakdown of each sub-system, including heating, cooling, lighting, elevator and other, is proposed for each climate region. The energy use pattern is closely related to the climate conditions, especially for the cooling and heating energy use. In the extreme cold region, heating energy consumption takes up 68% of the total amount, with the cooling energy consumption accounting for merely 6%. Thus, stations located in different climate regions are expected to concentrate on the energy-saving projects for different energy-use systems. Furthermore, the results of regression show that building area is the primary factor for heating/cooling energy use in all the four climate zones in China. The number of passengers and the regional GDP also exert influence on the heating/cooling energy use for some of the climate zones, the basic principles of which have been discussed in this paper. Then the regression equations are developed for each climate region, representing the average energy consumption level of stations in each climate zones. In future, further study will concentrate on the energy-saving potentials of stations in each climate zones and with different heating sources or cooling modes.
References

[1] Chen S, Li H 2016. *J Railway Economics Research* **05** 37-41
[2] Zhou H, Yang K 2012. *J Sichuan Architecture* **03** 43-44+47
[3] Han Y and Wang KP 2012. *J Railway Econ Res* **5** 8–13
[4] Song L, Wang Y, Li X 2016. *J Indoor & Built Environment* **0** 1-12
[5] Ma W, Liu X, Li L, et al 2009. *J Energy & Buildings* **41** 1313-1318
[6] Song G, Liu Y, Zhu D, Yan D and Peng C 2013. *J HV&AC* **4** 85–90
[7] Priyadarsini R, Wu X, Eang L 2009. *J Energy & Buildings* **41** 1319-1324
[8] Deng S M, Burnett J 2000. *J Energy & Buildings* **31** 7-12
[9] Wang J C 2012. *J Energy & Buildings* **49** 268-275
[10] Lam J, Li D 2003. *J Energy Conversion & Management* **44** 1391-1398
[11] Lam J, Wan K, Liu D, et al. 2010. *J Energy Conversion & Management* **51** 2692-2697
[12] Chung W, Hui Y 2009. *J Energy & Buildings* **41** 696-701
[13] Jiang Y, Yang X 2010. *J Energy of China* **05** 5-11