

## Empirical study on the outliers of compressed natural gas (CNG) refueling behaviors

LI Yang<sup>1,2,\*</sup>, JIN Deng Chao<sup>1</sup>, ZHAO Yu Lian<sup>1</sup>, JIN Hao<sup>1</sup>, BAO Zhen Bo<sup>1</sup>, ZHANG Wei Yu<sup>1</sup>, WEI Yong<sup>1</sup>, YANG Di<sup>2</sup> and SHAO Jian<sup>2</sup>

<sup>1</sup>College of Engineering and Technology, Tianjin Agricultural University, Tianjin, China

<sup>2</sup>Tianjin Huabei Gas & Heating Engineering Design Co. LTD, Tianjin, China

\*[liyong005@163.com](mailto:liyong005@163.com)

**Keywords:** CNG, Refueling behaviors, CNG filling station, Outlier, Python

**Abstract.** To improve a secondary CNG filling station's refueling effectiveness and reduce its operation costs, the abnormal CNG refueling behaviors or outliers are empirically studied. Firstly, the central tendency and dispersion of volume and pressure change is empirically studied, and the simple regression model is adopted to test the positive correlation between the two variables. Secondly, the abnormal behaviors are detected by parametric methods and non-parametric methods of statistical approaches based on volume or pressure change. Finally, these outliers are found mainly from the repeated filling behaviors by Python language. Therefore, the way to improve the refueling effectiveness and reduce operation costs is to reduce repeated filling actions.

### Introduction

Natural gas is mostly composed of methane, CH<sub>4</sub>, and it first was used as a transportation fuel in 1930s in Italy. But the reservoir tanks could not reserve enough gas. To solve this problem, the technology of compressed natural gas (CNG) was developed in 1980s. CNG is made by compressing natural gas, and is stored in high-pressure (20MPa in China) tanks on the vehicle. A compressed natural gas vehicle (CNGV) is an alternative fuel vehicle using CNG as a cleaner alternative to other fossil fuels. A CNGV connects to the high-pressure reservoir tanks at a secondary CNG filling station during refueling process by refueling guns.

According to the 17th China international NGVs and gas station equipment exhibition & summit forum[1], there are 5,030,000 NGVs on the road in China, with over 4,800,000 CNGVs, and over 4,700 CNG filling stations. Both the numbers of CNGV and CNG filling stations are top of the world. Obviously, for CNG's cheaper and fewer emission than gasoline, there is room for significant growth worldwide and in China. However, CNGVs require a greater amount of space for fuel storage[2,3]. The on board storage capacity of NGV is limited. In China, a fully filled tank of CNGV can run about 200 km. Therefore, a CNGV driver has to refuel the natural gas powered vehicle frequently. But the number of CNG secondary filling stations is limited and its effectiveness is need to be improved. In Tianjin, CNGV drivers need to wait more than 2 hours to fill up the natural gas vehicles.

Therefore, to improve the CNG refueling effectiveness is necessary and important, and a lot of studies have been done[2-6]. A bibliography[7] reviews the potential of CNG as transportation fuel, at the end of which containing 822 references from 111 scientific journals to papers, conference proceedings and dissertations on the subject that are published between 2001 and 2015. These references about CNG and NGV are mainly based on theoretical or numerical study. To improve the CNG secondary filling station's effectiveness, the abnormal behaviors need to be analyzed. As far as we known, there is no empirical studies on behavioral outliers of the CNG secondary filling station. In this study, the abnormal refueling behaviors of a secondary fueling station, characterized by the amount of gas filled (in volume), the initial and final pressure of the refueled NGV, are empirically studied by Python language based on statistical approaches.

The paper is arranged as follows: In Sec. 2 we briefly introduce the data studied in this paper, and related methods are also introduced. In Sec. 3 we present the empirical results. And in Sec. 4 we conclude this paper.

## Data and Methods

**Data description.** The data set that we empirically analyzed contains 125,021 records from a CNG secondary fueling station in Tianjin, over a period of 3 months from May to July in 2014. A refueling record logs the refueling behavior's attributes. Spatial attentions are paid on attributes including starting time, the initial and final pressure of the NGV, and the volume added into the NGV during its refueling process. The first ten records (the names of attributes are in Chinese) of these attributes are shown in Fig.1.

The first column of the table is the refueling time, referenced as  $t$  in this paper. The second column is the add gun number, and the third column is the total volume of CNG injected into the CNGV during this refueling process, referenced as  $V$ . The fourth column logs the CNGV's identification number (For privacy, the last number of ID is replaced by \*.). The fifth and sixth columns record the initial and final pressure of the NGV in this refueling process, and are referenced as  $P_{initial}$ ,  $P_{final}$  respectively.

交易日期	枪号	气量	车牌号	开始压力	结束压力
2014-5-1 0:01	4	14.11	津E2469*	4.5	18.6
2014-5-1 0:01	6	18.95	津E1186*	0.6	18.9
2014-5-1 0:01	5	17.89	津E0525*	1.5	18.7
2014-5-1 0:01	1	14.12	津E2045*	2.4	18.8
2014-5-1 0:01	3	15	津E1314*	4.4	18.7
2014-5-1 0:04	4	12.66	津E2614*	6.4	19
2014-5-1 0:04	2	18.66	津E1814*	1.6	19.5
2014-5-1 0:04	1	13.12	津E0479*	1.5	19.2
2014-5-1 0:04	3	15.34	津E1235*	4.3	19.4
2014-5-1 0:05	5	11.59	津E2235*	7.6	19.9

Fig.1: The first ten records of the data set used in this paper.

**Outlier detection and graphic displays of basic statistical descriptions.** In this paper, a common rule of thumb for identifying suspected outliers[8] is used to select out values not falling in the following range

$$(Q_1 - 1.5 \times IQR, Q_3 + 1.5 \times IQR) \quad (1)$$

Where interquartile range (IQR) is defined as

$$IQR = Q_3 - Q_1 \quad (2)$$

$Q_1 - 1.5 \times IQR$  is referenced as Min, while  $Q_3 + 1.5 \times IQR$  referenced as Max. And  $Q_1$ ,  $Q_3$  are the first and third quartile respectively. Quartiles are measures of the dispersion of a data set[8-10], and can be easily calculated by percentile functions from Python language.

As discussed in reference [8], this outlier detection method is like to using  $3 \times S$  as the threshold for normal distribution. Rather than assuming the data set is normal, kernel density estimation, a non-parametric method, estimates the probability density distribution from the input data only. The probability density function of a distribution is estimated by `kde.KDE1d()`, the kde function from Python, with a way taking into account the boundary condition by *renormalizing* the kernel[11].

To make the outliers easy to understand, graphics are helpful for visualizing inspection of data.

## Data Analysis and Results.

**Outlier Detection and Data Visualization.** For each record, there are three attributes,  $V$ ,  $P_{initial}$ ,  $P_{final}$ . The outliers we need to detect are in a three-dimensional attribute space. One approach for outlier detection in high dimensional data is to search for outliers in various subspaces[8]. Following this method, Li[12] detects the abnormal refueling behaviors by considering these actions in two-dimensional subspaces of the three attributes, especially in volume and initial pressure. But this

method has some drawback without considering the final pressure. For this reason, we provide another approach for outlier detection by considering all the three attributes and still in two-dimensional space.

It is known that a normal refueling behavior has two results. One is a certain volume of CNG is fill into the CNGV. Another is that the pressure in the tank of CNGV increases a certain number. We introduce one variable representing the change of tank's pressure in the refueling process, which is defined as

$$\Delta p = P_{final} - P_{initial} \quad (3)$$

Therefore, a normal refueling action is one with big  $V$  and  $\Delta p$ . In other word, a refueling behavior with small  $V$  or  $\Delta p$  would be considered as outlier. Intuition tells us that a small pressure change follows a small volume. To test this, we use the outlier method introduced in Sec.2. to detect outliers only in one dimension, by volume or pressure change.

The outlier searching process is realized by a recursive algorithm for each attribute. This searching action is stopped when there is not outlier left. Only by volume, if a refueling action with volume smaller than the smallest normal value (Min), this action is outlier. After 4 rounds of iteration, the Min of normal volume value of refueling behavior is 1.395 m<sup>3</sup>. And the number of outliers is 4227, account for about 3.38 percent of the total 125021 records. Only by pressure change, if a refueling action with pressure change smaller than the smallest normal value (Min), this action is outlier. There are 4719 outliers with pressure change less than the smallest normal value (Min) of 3.850 MPa, account for about 3.77 percent of the total 125021 records, after 4 rounds of iteration. There are 4136 records being outliers if considering both pressure change and volume.

Small pressure change is in connection with small volume as expected. To test the two variables' positive relationship, the simple regression model is adopted to analyze them. By taking pressure change as the regressand, volume as regressor, the ordinary least squares (OLS) is used to estimate the model's slope and intercept. This estimation process is provided by Python function stats.linregress( $V, \Delta p$ ). The slope is 0.893, and the intercept is 2.473. Both the r value (about 0.939) and p value (0.0) show that the regression results are acceptable. Pressure change has obviously positive correlation with volume.

The recursive process and measures of the central tendency and dispersion of the two variables are listed in Table 1. X means that there is no outlier for  $\Delta p$  at its upper boundary.

Table 1. The statistical values of central tendency and dispersion of the CNG refueling behaviors.

Measures	Mean	Median	Min	Q1	Q3	Max	S
$\Delta p$	13.072	13.5	3.00	10.80	16.00	X	4.087
	13.520	13.7	3.60	11.10	16.10	X	3.396
	13.541	13.7	3.85	11.20	16.10	X	3.369
	13.548	13.7	3.85	11.20	16.10	X	3.360
$V$	12.077	12.34	1.045	9.43	15.02	23.405	4.032
	12.248	12.42	1.360	9.58	15.06	23.405	3.805
	12.267	12.43	1.395	9.60	15.07	23.405	3.780
	12.269	12.43	1.395	9.60	15.07	23.405	3.778

For each variable, there are four sets of values in Table 1. The first set is calculated without deleting any record. The second set is obtained by removing the behaviors with value smaller than the Min, which taking as this attribute's smallest normal value. The fourth set is the final results, because there is no outlier left for each of the two variables.

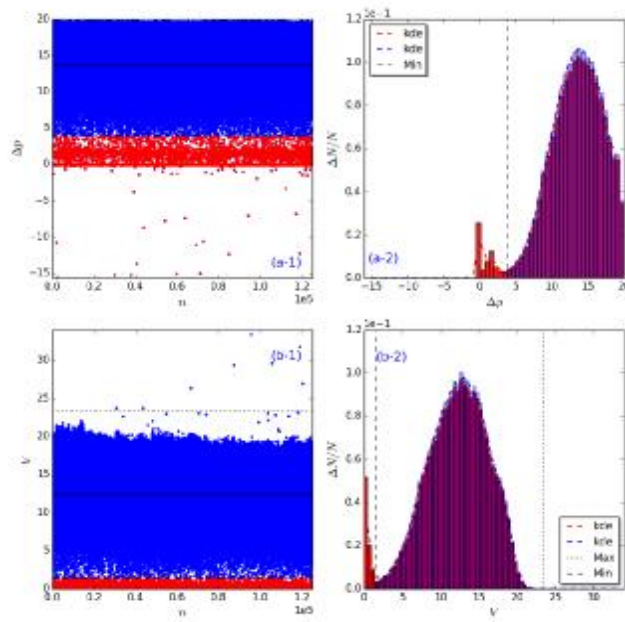


Fig.2:(color online)Scatter, histograms and its kernel density estimation functions of volume,  $\Delta p$ . Sites with red color represents the outliers.

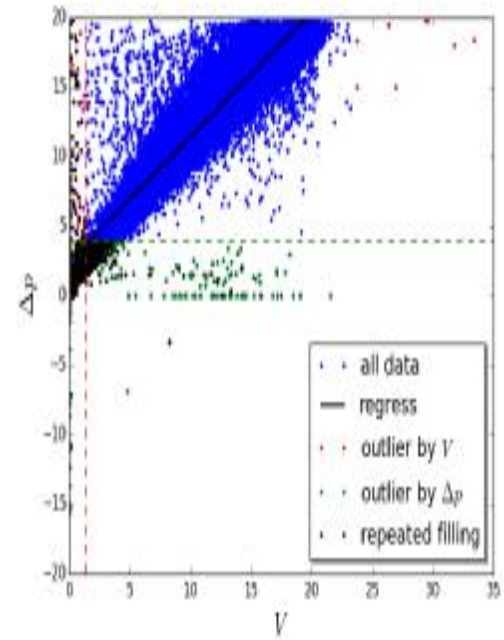


Fig.3: (color online)Outliers with  $V$ ,  $\Delta p$  less than certain values of the CNG refueling behaviors are mainly repeatedly filling actions.

Fig.2 includes the two variables' time series, histograms and their estimated density functions of all data set in red color. And the same thing done with eliminating the outliers in blue color.

From the estimated density function of volume, the probability of value of volume less than 1.395 is about 2.86 percent. And the probability of value of pressure change less than 3.85 is 3.78 percent according to its estimated density function. Sites or histograms with red color are outliers.

**Results and Discussion.** From Table 1 and Fig.2, as expected, the refueling behaviors characterized by the three attributes, volume, initial pressure and final pressure, have abnormal behaviors, or outliers. Since volume has positively correlation with pressure change, there are two ways to detect outliers. One is according to volume. CNG refueling behaviors with volume less than 1.395 are considered as outliers. Another is based on pressure change. Refueling actions with pressure change less than 3.85 are token to be abnormal behaviors.

After detailed analyzing the outliers in the original records, we find that these abnormal behaviors with less volume or less pressure change are mainly the refilling behaviors, that is the same CNGV is refueled repeatedly. By Python language, there are 4354 repeatedly filling behaviors (black sites in Fig.3) with volume less than 1.395  $\text{m}^3$  or with pressure change less 3.85 MPa. There are 4227 refueling behaviors (red sites in Fig.3) of the all 125021 records with volume less than 1.395  $\text{m}^3$ , of which there are 4101 behaviors are repeatedly filling actions. And there are 4719 with pressure change less than 3.85 MPa (green sites in Fig.3), of which there are 4281 repeatedly filling behaviors. There are 4136 refueling behaviors with both volume less than 1.395  $\text{m}^3$ , and pressure change less than 3.85 MPa, of which there are 4028 behaviors are repeated ones.

Therefore, to reduce the outliers is to reduce repeating filling behaviors.

## Conclusions

By analyzing the 125,021 records from a CNG secondary filling station in Tianjin, the empirical results of volume and pressure change are obtained.

Firstly, the positive correlation between volume and pressure change are empirically studied by the simple regression model. The results show that the coefficient of correlation of the two variables is 0.893. A reasonable way to detect outliers is to detect it by volume or by pressure change.

Secondly, the central tendency and dispersion of the data set are analyzed. The mean and median of pressure change are 13.548 and 13.7 MPa after deleting outlier, which are behaviors with pressure change less than 3.85 MPa. The IQR and standard deviation of pressure change 4.90 and 3.36 MPa. For volume, the same measures are 12.269, 12.43, 5.47 and 3.778 m<sup>3</sup>. And refueling behaviors with volume less than 1.395 may be taken as outliers.

Finally, the reasons of outliers are analyzed. By Python language, we find that outliers detected by volume or pressure change are mainly from repeatedly filling behaviors. Since the outliers are behaviors with less volume or less pressure change, one way to improve a CNG secondary filling station's effectiveness is to reduce repeatedly filling behaviors.

## Acknowledgments

This work was supported by the Innovation Fund for Small Technology-based Firms under Grant No. 14ZXCXGX00392.

## References

- [1] Information on <http://www.ngvchina.com/exhibition-overview/industry-news/830.html>
- [2] M. Farzaneh-Gord, M. Deymi-Dashtebayaz, Polish journal of chemical technology Vol. 15(1)(2013), p.88
- [3] M. Farzaneh-Gord, M. Deymi-Dashtebayaz, H. R. Rahbari, Journal of natural gas science and engineering, Vol.3(2011), p.334.
- [4] M. Baratta, H. Kheshtinejad, D. Laurenzano, D. Misul, S. Brunetti, Journal of natural gas science and engineering, Vol.24(2015), p.52.
- [5] N. N. Nahavandi, M. Farzaneh-Gord, Journal of Brazilian society of mechanical science and engineering, Vol. 36(4)(2014), p.837.
- [6] M. Saadat-Targhi, J. Khadem, M. Farzaneh-Gord, Journal of natural gas science and engineering, Vol.29(2016), p.453.
- [7] M. I. Khan, T. Yasmeen, A. Shakoore, N. B. Khan, M. Wakeel, B. Chen, Journal of natural gas science and engineering, Vol. 31(2016), p.352.
- [8] J. W. Han, M. Kamber, J. Pei, Data mining: concepts and techniques(3rd edition) (Elsevier Pte Ltd, Singapore 2012).
- [9] G. Casella, R. L. Berger, Statistical Inference(2nd edition) (Wadsworth Group 2002).
- [10] X. R. Chen, Probability and mathematical statistics (Science Press, Beijing 2000).
- [11] Information on [https://pythonhosted.org/PyQt-Fit/KDE\\_tut.html](https://pythonhosted.org/PyQt-Fit/KDE_tut.html)
- [12] Y. Li, D. C. Jin, Z. B. Bao, H. Jin, J. W. Guo, Y. L. Zhao, D. Yang, J. Shao, The 2nd international conference on energy, environment and materials science, Singapore, 2016.