Fitting Analysis for Flow Coefficient Predictive Model of Residential Buildings with High Airtightness

Wenqian Zhou, Xiangli Li and Lin Duanmu

Abstract The flow coefficient is an important parameter of air infiltration, affected by many factors, such as building parameters, wind speed and temperature difference between indoor and outdoor. And its modeling is used to adopt the method, multiple linear regression, ignoring the problem caused by characteristic of small sample size. Stepwise regression, least squares and partial least squares (PLSs) with corresponding independent variable screening methods are adopted to obtain the predictive models which are compared with F-test, goodness of fit and CV (RMSE). The result shows that the goodness of fit of apartment's model and villa's model received by PLS is increased by 25.9% and 2.2%, respectively, compared with least squares and the PLS models' CV (RMSE) which are 9.4% and 18.3%. Considering the essential characteristics and fitting results of the three methods, it can be concluded that PLS which is suitable for small sample size is the preferred choice for the establishment of flow coefficient model.

Keywords Air infiltration · Flow coefficient · Data fitting

1 Introduction

Building air tightness, as an index of ultra-low energy consumption building design, is usually measured by a "blower door" test or tracer gas test. And the principle of air infiltration calculation is used to be presented by rule of thumb, power law or quadratic equation [7]. Some parameters involved in the above equations are usually expressed with empirical values, such as flow coefficient and pressure exponent. But, because of the complexity of air infiltration research, there are few test data on building air tightness and the distribution of test sites is scattered in China. Therefore, on account of the existing small amount of data, it is necessary to establish a reasonable predictive model of air infiltration. Based on the experimental data in the research of air tightness of newly built residential buildings in cold regions in China [6],

W. Zhou (✉) · X. Li · L. Duanmu
Dalian University of Technology, Dalian 116024, Liaoning, China
e-mail: wqzhou@mail.dlut.edu.cn

Z. Wang et al. (eds.), *Proceedings of the 11th International Symposium on Heating, Ventilation and Air Conditioning (ISHVAC 2019)*, Environmental Science and Engineering, https://doi.org/10.1007/978-981-13-9528-4_66

this chapter studies the regression model of flow coefficient. The power law will be used and empirical values are selected for pressure exponent. Each regression analysis corresponds to an independent variable selection method, and then, the flow coefficient models are determined by fitting. In addition, the results of different fitting methods are tested and compared according to goodness of fit and coefficient of variation of the root-mean-square error.

The models obtained in this chapter can predict the flow coefficient with building characteristic parameters or meteorological parameters conveniently when ΔP is 50 Pa, thus avoiding complicated test, which greatly saves manpower and material resources. The data analysis method involved will provide some guidance for subsequent air tightness data processing.

2 Methods

2.1 Influencing Factors of Flow Coefficient

The basic research of air infiltration is to determine the relationship between air flow rate Q and pressure difference ΔP. Power law (Eq. 1) and quadratic equation (Eq. 2) are the most common, which were used to analyze building airtightness as early as 1949 [4].

$$Q = C(\Delta P)^n \tag{1}$$

$$\Delta P = aQ^2 + bQ \tag{2}$$

In the 1970s and 1980s, there were many studies on the applicability of power law and quadratic equation [5]. The research shows that the power law is widely used because it can evaluate the air tightness of buildings with unknown characteristics of crack opening, and it is easier to obtain the better statistical results [2, 3]. Based on power law, the chapter refers to the flow coefficient model of Ji [6] (Eq. 3), in which n is 0.65 [7] and ΔP_{f} often takes 50 Pa.

$$C_{\mathrm{ave}} = \frac{C_{\mathrm{dep}} \cdot (\Delta P_{\mathrm{f}})^{|n_{\mathrm{dep}} - 0.65|} + C_{\mathrm{pre}} \cdot \left(\Delta P_f\right)^{|n_{\mathrm{pre}} - 0.65|}}{2} \tag{3}$$

Air infiltration occurs in the gaps of building walls, roofs, exterior windows, etc. The permeability of building envelope is influenced by many factors, such as building materials, building structure and installation technology. Infiltration as a form of natural ventilation must have the same characteristics. It can be seen that the stack-induced ventilation is mainly caused by the difference of air density between indoor and outdoor. Its strength is related to the height of the building and the temperature difference between inside and outside. The wind-induced ventilation is mainly due to the increase in air pressure on the windward side and the decrease in air pressure on the leeward side, resulting in the phenomenon of air flow from the windward side to the leeward side. The ventilation intensity under wind pressure is related to the opening area and the wind force.

$$\Delta P_{\mathrm{h}} = (\rho_{\mathrm{o}} - \rho_{\mathrm{i}})gh = K_{\mathrm{s}}h\left(\frac{1}{T_{\mathrm{o}}} - \frac{1}{T_{\mathrm{i}}}\right) \tag{4}$$

$$\Delta P_{\mathrm{w}} = K_{\mathrm{w}}\frac{U^2}{2}\rho_{\mathrm{o}} \tag{5}$$

where ΔP_{h} is the pressure difference of stack-induced ventilation, Pa, ΔP_{w} pressure difference of wind-induced ventilation, Pa, ρ_{o} outdoor air density, kg/m^3, ρ_{i} indoor air density, kg/m^3, T_{o} outdoor air temperature, °C, T_{i} indoor air temperature, °C, K_{s} stack factor and K_{w} wind factor.

The formula of calculating air flow rate can be obtained (Eq. 6). It can be concluded that the factors affecting the pressure difference can be regarded as the ones of flow coefficient. On the other hand, flow coefficient is a physical quantity related to the crack characteristics of building envelope, so it is necessary to consider the relevant building parameters, such as the building envelope area and building height.

$$Q = A \cdot C_{\mathrm{ave}} \cdot (\Delta P_{\mathrm{h}} + \Delta P_{\mathrm{w}})^n \tag{6}$$

$$C_{\mathrm{ave}} = \frac{Q}{A \cdot (\Delta P_{\mathrm{h}} + \Delta P_{\mathrm{w}})^n} \tag{7}$$

In summary, considering the difficulty of data collection, the chapter mainly considers the following factors of apartments. They are floor area (A_{F}), internal volume (V), envelope area (A_{E}), residential height (H), open gap length (L_{wd}), outdoor wind speed (U) and indoor and outdoor temperature difference (ΔT). The influencing factors of villas are somewhat different from apartments, so the model should be analyzed separately.

2.2 *Correlation Analysis of Experimental Data*

Because of the many interference factors in the test, there are only 11 groups and 10 groups of complete test data for apartments and villas, respectively, after eliminating the unreliable data. With the characteristics and limitations of the small sample, the chapter adopts three regression methods and corresponding independent variable screening methods to analyze the data.

The results of correlation analysis of influencing factors of apartments and villas are listed in Tables 1 and 2, respectively. For apartments, when the confidence level is 0.01 and 0.05, there is no significant correlation between the listed factors. Then,

Table 1 Correlation analysis of influencing factors of apartments

Correlation coefficient	A_F	V	A_E	H	L_{wd}	U	ΔT
C_{ave}	0.117	0.117	0.481	0.570	0.481	0.116	−0.168

Table 2 Correlation analysis of influencing factors of villas

Correlation coefficient	A_F	V	A_E	H	N	L_{wd}	U	ΔT
C_{ave}	0.562	0.552	0.543	−0.782**	−0.453	0.534	−0.351	−0.618*

Note **When confidence level (double test) was 0.01, the correlation was significant
*When confidence level (double test) was 0.05, the correlation was significant

the confidence level is set to 0.1, and A_E, H and L_{wd} can be considered. For villas, in the case of confidence level of 0.05, the influencing factors are H and ΔT which are significantly correlated.

In addition, it is found that there are significant correlations among variables, that is, multicollinearity problem which is not in line with the requirements in the multiple linear regression model. It can be solved by stepwise regression analysis, removing the secondary factors and partial least squares regression.

3 Results

The independent variable selection methods adopted are correlation analysis, stepwise regression and variable importance in the projection (VIP). The above methods correspond to the following fitting process one by one.

3.1 Ordinary Least Squares (OLSs)

OLS has been widely used in various fields since it was proved in 1930s [9]. Its main principle is to minimize the sum of squares of errors to obtain the most appropriate functional equation. According to the correlation analysis in Sect. 2.2, the influencing factors of apartments and villas are A_E, H, L_{wd} and H, ΔT, respectively. Taking apartments as an example, OLS is used to assume that the independent variables (A_E, H, L_{wd}) have the following linear functions.

$$C_{ave}^i = a_0 + a_1 A_E^i + a_2 H^i + a_3 L_{wd}^i, \quad i = 1, \ldots, 11 \tag{8}$$

Independent variables can be recorded as data matrix X, and there are also coefficient matrix a and dependent matrix Y.

$$Xa = Y \tag{9}$$

$$\begin{bmatrix} 1 & A_{\mathrm{E}}^{1} & H^{1} & L_{\mathrm{wd}}^{1} \\ 1 & A_{\mathrm{E}}^{2} & H^{2} & L_{\mathrm{wd}}^{2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & A_{\mathrm{E}}^{11} & H^{11} & L_{\mathrm{wd}}^{11} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} C_{\mathrm{ave}}^{1} \\ C_{\mathrm{ave}}^{2} \\ \vdots \\ C_{\mathrm{ave}}^{11} \end{bmatrix} \tag{10}$$

The relationships ($S(a) = \min\|Xa - Y\|_2$) need to be met. Finally, the coefficient matrix is obtained. The fitting results of apartments and villas are shown in Eqs. (11) and (12).

$$C_{\mathrm{ave}} = 8.802 - 0.015A_{\mathrm{E}} + 0.016H - 0.017L_{\mathrm{wd}} \tag{11}$$

$$C_{\mathrm{ave}} = 22.552 - 0.498H - 0.749\Delta T \tag{12}$$

3.2 Stepwise Regression

The basic principle of stepwise regression is to introduce all the influencing factors into the model one by one according to the significant degree of their influence on flow coefficient. Next, for each new factor introduced, F-test is needed for the existing variables. If the existing significant variables are no longer significant, the new variables will be eliminated until all the factors are considered, so as to ensure that all the selected factors have the best interpretation effect [10].

There are two methods of stepwise regression, forward method and backward method. The chapter adopts the former one that is to increase the number of variables from small to large and one by one, until all variables have been introduced into the test [8]. Finally, the models of apartments and villas can be obtained (Eqs. 13 and 14).

$$C_{\mathrm{ave}} = 7.606 - 0.524A_{\mathrm{F}} + 0.202V - 0.142L_{\mathrm{wd}} \tag{13}$$

$$C_{\mathrm{ave}} = 5.347 + 0.016A_{\mathrm{F}} - 0.051L_{\mathrm{wd}} \tag{14}$$

3.3 Partial Least Squares (PLSs)

PLS [11] was first proposed in 1983. It combines multivariate linear regression, principal component analysis and canonical correlation analysis and can solve the problems of small sample size, large independent variables and serious multicollinearity problem. Even when the number of variables is more than the number of samples, it can also meet the requirements of fitting.

Step 1 Standardize the sample data and find out the correlation coefficient matrix.
Step 2 Extract the principal component.
Step 3 Establish the regression model related to the principal component vector.
Step4. Transform the regression coefficient to get the regression equation corresponding to the original variable.

According to the principle mentioned above, select the variables whose VIP > 1 and the fitting model of apartments and villas can be presented as follows:

$$C_{\text{ave}} = 7.729 - 0.004A_{\text{E}} - 0.016L_{\text{wd}} - 0.048\Delta T \tag{15}$$

$$C_{\text{ave}} = 7.194 - 0.005A_{\text{E}} - 0.451H - 0.437U \tag{16}$$

4 Discussion

The fitting formulas of stepwise regression, OLS and PLS can be obtained through the above analysis. The goodness of fit of each model can be explained by the coefficient of determination (R^2), and the regression effect of models can be considered with F-test. The distribution of experimental data and predicted data is provided (Fig. 1). Table 3 shows that the goodness of fit of apartments varies from 0.340 to 0.620 and it is ranked as stepwise regression > PLS > OLS, while the goodness of fit of villas is

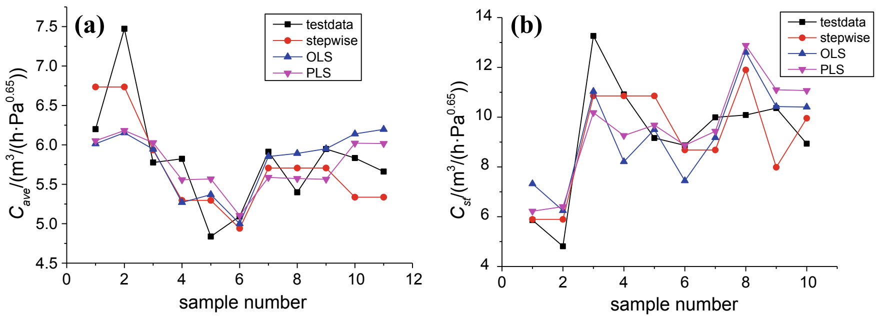

Fig. 1 Comparison of prediction model: **a** apartments; **b** villas

Table 3 Test for all models

Building	Methods	R^2	F	P	CV (RMSE) (%)	Δ%
Apartments	OLS	0.340	4.641	0.06	9.5	−0.4
	Stepwise	0.620	14.675	0.004	12.8	1.6
	PLS	0.428	6.735	0.029	9.4	0.3
Villas	OLS	0.486	7.555	0.025	18.1	−3.5
	Stepwise	0.602	12.082	0.008	15.9	−1.4
	PLS	0.497	7.909	0.023	18.3	−6.0

between 0.486 and 0.602 and it can be ranked as stepwise regression > PLS > OLS. Besides, F-test illustrates that the obtained models have statistical significance, and the mean relative error (△%) is no more than 10% which can be regarded acceptable.

The coefficient of variation (CV) of the root-mean-square error (RMSE) is an indication of how much variation or randomness there is between the data and the model can be acceptable with the CV (RMSE) within ±30% [1]. CV (RMSE) can be calculated by Eq. (14), and the results are listed in Table 3. The variation degree of residential flow coefficient model obtained by PLS is the smallest. For villas, CV (RMSE) is not much different.

$$\mathrm{CV(RMSE)} = \frac{\sqrt{\frac{\sum_{i=1}^{m}(y_i - \hat{y}_i)}{m}}}{\bar{y}} \tag{17}$$

Compared with OLS, the goodness of fit of apartment's model and villa's model received by PLS is increased by 25.9% and 2.2%, respectively, while the goodness of fit of the two models obtained by stepwise regression could be increased by 80% and 19.2%, respectively. The results show that stepwise regression is more suitable for the selection of independent variables and model fitting than OLS combined with correlation analysis. Although the goodness of fit of PLS is inferior to stepwise regression, it has better fitting effect than OLS. Considering the characteristics of data quantity, PLS combined with VIP is the preference of flow coefficient model fitting. Finally, Table 4 shows the characteristics of three methods, including sample size, elimination of multicollinearity, etc.

5 Conclusions

The fitting of flow coefficient model is studied in this chapter, including the screening of independent variables, data fitting and testing. There are three forms, OLS combined with correlation analysis, stepwise regression and PLS combined with VIP.

Table 4 Characteristic comparison of three methods

Methods	OLS	Stepwise regression	PLS
Describe	Establish the quantitative relationship of linear mathematical model among multiple variables and analyze it with sample data		The combination of multiple linear regression analysis, principal component analysis of variables and canonical correlation analysis of variables
Sample size	Minimum sample size is at least 1 plus the number of explanatory variables; in general, the sample size is at least three times that of independent variables		Suitable for small sample data
Software tool	SPSS, MATLAB, Mathematica		MATLAB, SAS
Elimination of multicollinearity	No	Yes	Yes

(1) According to the correlation analysis, the relevant independent variables of apartments and villas are envelope area (A_E), building height (H), open gap length (L_{wd}) and building height (H), temperature difference (ΔT), respectively.

(2) The goodness of fit of apartments' models varies from 0.340 to 0.620, while the goodness of fit of villas' is between 0.486 and 0.602. F-test shows that the confidence level of all the models can reach above 90%, and the mean relative error ($\triangle\%$) is below 10%. Compared with OLS, the goodness of fit of apartment's model and villa's model received by PLS is increased by 25.9% and 2.2%, respectively, and CV (RMSE) can be satisfied, so PLS combined with VIP can not only guarantee the goodness of fit for small sample data, but also deal with multicollinearity problems.

Although PLS can obtain better results of small sample with many independent variables, the experimental data should be increased to ensure the accuracy of fitting models. So future work should collect more data of air tightness test and further improve and validate the flow coefficient models.

Acknowledgements This chapter is supported by the National Key Research and Development Program of China "Near-Zero Energy Building Technology System and Key Technology Development" (number 2017YFC0702600).

References

1. Ashrae.: ANSI/ASHRAE guideline 14-2014 measurement of energy and demand savings, Atlanta (2014)

2. Chiu, Y.H., Etheridge, D.W.: Calculations and notes on the quadratic and power law equations for modelling infiltration. Int. J. Vent. **1**(1), 65–77 (2002)
3. Colliver D.G., et al.: ASHRAE research project report RP-438: evaluation of the techniques for the measurement of air leakage of building components (1993)
4. Dick, J.B.: Experimental studies in natural ventilation of houses. J. Inst. Heating Ventilating Eng. (1949)
5. Etheridge, D.: A perspective on fifty years of natural ventilation research. Build. Environ. **91**, 51–60 (2015)
6. Ji, Y.: Study on Air Infiltration Prediction Model for Residential Buildings with High Airtightness Envelope in Cold Area. Dalian University of Technology (2018)
7. Sherman, M.H.: Building Airtightness: Research and Practice. Lawrence Berkeley National Laboratory, Berkeley (2004)
8. Wang, S., et al.: Error Theory and Surveying Adjustment. Tongji University Press (2010)
9. Wang, Y., et al.: Dictionary of Mathematics. Science Press (2010)
10. Xian, Y., et al.: Application of multiple-stepwise and Kalman Filtering in haze forecast. J. Syst. Simul. **30**(04), 278–285 (2018)
11. Zhang, H., et al.: Analysis and Method of Small Sample Multivariate Data. Northwestern Polytechnical University Press (2002)