The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

Multivariate Adaptive Regression Splines for Prediction of Wind Pressure Coefficients on Low-Rise Building Surfaces

Zhiyuan Fang [1], Zhisong Wang [2] and Zhengliang Li [3]

[1] *School of Civil Engineering, Chongqing University. Email: fangzhiyuan017@foxmail.com*
[2] *School of Civil Engineering, Chongqing University. Email: wangzhisong@cqu.edu.cn*
[3] *School of Civil Engineering, Chongqing University. Email: lizhengl@hotmail.com*

Abstract: The wind pressure coefficient (C_p) is a dimensionless quantity which describes the relative pressure on the building surface caused by wind. It has important influence on wind resistance design of structure and evaluation of natural ventilation. Wind pressure coefficients on building surfaces are generally determined by full-scale building tests, wind tunnel tests, computational fluid dynamics (CFD) simulation and analytical models derived from experiments. The main purpose of this paper is to establish accurate and easy-to-use analytical models for predicting average C_p on the surfaces of low-rise buildings. Based on a nonparametric regression algorithm known as multivariate adaptive regression splines (MARS) and data obtained from wind tunnel tests, a MARS model was developed for wall of low-rise buildings. By comparing the results obtained from the MARS model and other analytical models, the advantages of MARS model over other approaches in predictive accuracy and better interpretability have been demonstrated.

Keywords: Wind pressure coefficient; Low-rise building; Analytical model; Multivariate adaptive regression splines.

1. Introduction

Wind is a common natural phenomenon caused by air flow. The pressure on building surface caused by wind has an important influence on the force analysis of main structure and the wind resistance design of building envelopes. In addition, for building infiltration and ventilation, wind is also a dominant factor. Therefore, wind pressure is of great significance for establishment of analytical models about building structural analysis, environmental and energy consumption assessment. So far many analytical models have introduced wind pressure as an important input parameter, including building energy simulation (BES), airflow network (AFN), building component heat, air and moisture (HAM) transfer and so on. We usually use pressure coefficients (C_p) to characterize the wind pressure, and C_p are generally defined as:

$$\mathrm{C}_p = \frac{P - P_{ref}}{0.5\rho V_{\mathrm{h}}^2} \tag{1}$$

in which P is the static pressure at a given point on the building facade, P_{ref} is the static reference pressure, ρ is the air density and V_{h} is the wind speed at the reference height.

In order to obtain the wind pressure coefficients of a building, we can use wind tunnel test, computational fluid dynamics (CFD) simulation, and even full-scale measurement. Cóstola (2009) has summarized these methods as the primary sources for obtaining wind pressure coefficients. The primary sources can provide accurate C_p data for specific buildings, but the whole testing or analysis process is cumbersome and expensive. Therefore, these methods are only used to deal with building with complicated shape or for the establishment of wind pressure coefficient databases. Wind pressure coefficients can also be obtained from databases or analytical models, which are summarized as secondary sources. As for low-rise building, analytical models are often the first choice for predicting the wind pressure coefficients, which can provide ideal results if appropriately developed, and some analytical models have been widely used in various related projects.

The most commonly used analytical model was developed by Swami and Chandra (1988). They used step-wise regression analysis method to fit published studies about C_p and proposed equations for low-rise and high-rise buildings separately. The equation for low-rise buildings has been widely used in BES programs to predict surface average C_p, and it will be referred to as S&C equation in this paper. The S&C equation is mainly used to calculate C_p of rectangular low-rise buildings. By applying to C_p data from various sources, an acceptable correlation coefficient of 0.797 was obtained. With the improvement of test technique and measuring equipment, more precise pressure coefficients were obtained through advanced wind tunnel tests, which makes the S&C model have a need to be improved and upgraded accordingly.

The Tokyo Polytechnic University (TPU) has established a large database of wind tunnel tests for conventional buildings, in which very detailed wind pressure coefficients were given for low-rise buildings. Based on the TPU database, Muehleisen and Patrizi (2013) proposed an equation (M&P equation) by curve fitting to predict the surface-averaged C_p for low-rise buildings. The M&P equation fit the TPU data with a goodness-of-fit R^2 = 0.992, but there are still large errors over a wide range of wind directions ($90° < \theta < 165°$).

Recently, using the artificial neural networks (ANN) mothed, Bre and Gimenez (2018) developed three analytical models to predict average C_p of low-rise buildings with different roof types, respectively. The analytical model for low-rise building with flat roof is named B&G FANN. These ANN models were trained and tested by TPU experimental database, and results from these models are in good agreement with the TPU database. However, the ANN model have complicated structure which comprises one or several layers of interconnected neurons, and neurons interact with each other via weights and bias, so it is difficult for ANN model to form a simple and easy-to-use equation.

This paper examines the application of multivariate adaptive regression splines (MARS) (Friedman 1991) to develop analytical model of the surface-average C_p of low-rise buildings. MARS can not only acquire complex data mapping in multi-variate data patterns, but also produce simpler and easy-to-interpret models. MARS has been applied in some engineering fields (Haghiabi 2017, Zhang et al. 2013, 2016, 2017, 2018). However, As far as I know, it has not been applied in the field of wind engineering and environmental engineering. In this paper, the details of MARS are introduced and a simple example is given to show the function approximation ability of this algorithm. Based on TPU database, the MARS model for the prediction of C_p of low-rise buildings has been developed. Comparisons between the current model and TPU experimental database and other previous analytical models have been made to illustrate the accuracy and effectiveness of the proposed MARS model.

2. Methodology

2.1 Multivariate adaptive regression splines(MARS)

MARS is a data-driven statistical method proposed by Friedman (1991). It is mainly used to deal with complex multi-variate data and to fit the mathematical relationship between input variables and output variables. The training data set is divided into piecewise linear splines with different gradients on the basis of divide-and-conquer strategy. There is no need to make specific assumptions about the underlying, intrinsic functional relationship between input and output variables. In the process of fitting, MARS divide the data into multiple segments, and of which the end are named knots. One knot symbolizes the end of one data region and the beginning of another. The piecewise linear splines are known as basis functions (BFs), and they connected by knots smoothly together to result in an effective mathematical model that has the capability to capture both linear and nonlinear characteristics flexibly and accurately.
The MARS model function $f(X)$ is established by linear combination of the BFs and their interactions. The expression is as follows:

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m \lambda_m(X) \tag{2}$$

in which $\lambda_m(X)$ is the basis function. Its concrete form may be a spline function, or the product of several spline functions. β are constant coefficients estimated by least squares method.

MARS fits the function f by using BFs which mainly involved piecewise linear functions and piecewise cubic functions. Considering that piecewise linear functions are simple and has sufficient fitting accuracy for most problems, so, just take piecewise linear functions for example, its typical form is max(0, x−t), and a knot appears at the value t. The function max(.) denotes that the positive part of (.) is used, or it is assigned to zero. Formally,

$$\max(0, x-t) = \begin{cases} x-t, if \ \ x \ge t \\ 0, otherwise \end{cases} \tag{3}$$

In order to explain the fitting process and effect of MARS more clearly, piecewise linear spline functions are used to fit some random data. The fitting result is shown in Fig. 1. And the mathematical equation obtained from MARS is expressed as

$$y = 20.344\text{-}4.042\times\text{BF1}\text{-}2.5893\times\text{BF2}+3.097\times\text{BF3}\text{-}4.2006 \times\text{BF4}+2.8776\times\text{BF5} \tag{4}$$

in which BF1 = max(0,15.5-x), BF2 = max(0, x-12.5), BF3 = max(0,12.5-x), BF4 = max(0,5-x), BF5 = max(0,8-x). x = 5, 8, 12.5 and 15.5 are the locations of the knots. They divide the scope of variables into five intervals in which different linear relationships are distinguished.

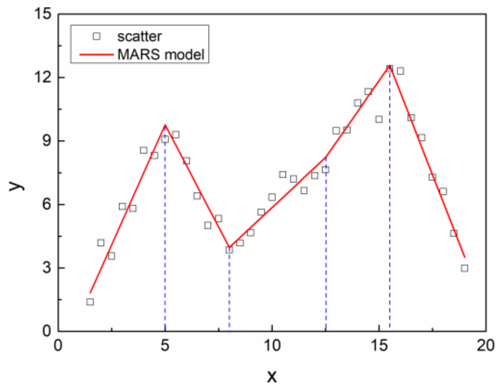

Figure 1. Knots and linear splines for a simple MARS example.

The MARS modeling process is divided into two steps: forward selection and backward deletion. The whole modeling process is data-driven, so it is easy to operate and highly automated.

In the forward phase, in order to make the model conform to Eq. (2), a forward selection program is first executed on the training data. An initial model with only intercept β_0 is established, and the basis pair with the greatest reduction in training error is added. Then, considering the present model with M basis functions, another next pair will be added to the model in form

$$\hat{\beta}_{M+1}\lambda_m(X)\max(0, X_j - t) + \hat{\beta}_{M+2}\lambda_m(X)\max(0, t - X_j) \tag{5}$$

and each β will be assessed by least square method. As a basis function added to the model space, the interactions between existing BFs in the model are also considered. Add BFs until the model gets a specified maximum number of terms, resulting in purposeful over-fitting of the model.

After the forward phase, the backward deletion will be performed to reduce the number of terms. The purpose of backward pruning is to find an approximate ideal model by deleting irrelevant variables. In order to get the final optimal model, backward pass deletes the basis function that contributes the least to the target model until the best sub-model is found. Model subsets are compared using Generalized Cross Validation (GCV) with lower computational cost. As a goodness-of-fit test, GCV equation punishes a lot of BFs and plays an important role in reducing the chance of over-fitting. Eq. (6) presents the calculation formula of GCV for training data with N observations.

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N}\left[y_i - f(x_i)\right]^2}{\left[1-\frac{M+d\times(M-1)/2}{N}\right]^2} \tag{6}$$

in which M is the number of BFs, N is the number of observations, d is the penalty parameter, and f(xi) represents the estimated values of the MARS model. The denominator indicates the change in variance as the complexity of the model increases. The numerator represents the mean squared error of the model evaluated in training data, which is punished by denominator. It should be noted that $(M - 1)/2$ denotes the quantity of function knots. Therefore, the GCV penalizes both the number of basis functions of as well as the number of knots. By default, the penalizing parameter d is set to value of 3.0. Each deletion step will prune a basis function, and the deletion step will be repeated until an optimal model is finally obtained.

After determining the final ideal MARS model, by combining all BFs containing one variable with another BFs containing pairwise interactions, a procedure named variance decomposition analysis (ANOVA) (Friedman 1991) can be utilized to evaluate contributions from input variables as well as BFs.

2.2 Database

Tokyo Polytechnic University recently established the TPU database on the basis of wind tunnel tests. It contains a variety of buildings, including high-rise buildings, low-rise buildings, and the effect of sheltering has been considered. The TPU database provides the times series of C_p, graphs of surfaced-averaged C_p, and contours of local C_p of wind directions from 0° to 90° per 15°. For rectangular floor-plan buildings, considering the geometric symmetry, the wind directions can easily be extended to 180° . Compared with other databases, TPU database uses more advanced instrumentations, and provides a plenty of data with a relatively smaller system error. Therefore, it may be considered as the most accurate database so far. In view of above advantages, in this paper, we decided to use TPU database to develop the MARS model.

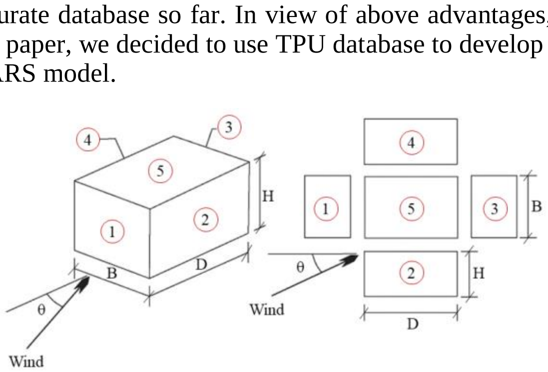

Figure 2. Building geometry identifying the building surfaces 1-5, the building dimensions *D*, *B*, *H*, and the wind angle *θ*.

Fig. 2 shows the geometry of the low-rise building utilized in this study. The TPU database used for MARS modeling comprises of surface-averaged C_p with D/B=1/2.5, 1/1.5, 1/1, 1.5/1, 2.5/1 for θ from 0° to 180° in 15° increments. Table 1 lists a summary of the input variables, outputs and parameter statistics. In this paper, 200 data sets (about 77% of the overall data set) were randomly selected as the training patterns in the 260 data sets, and the remaining 60 were used for testing purposes.

Table 1. Summary of input variables and outputs.

Inputs and outputs	Parameters
Inputs	D/B, θ
Outputs	C_p
C_p range	-0.8277~ 0.6925
No. of training data	200
No. of testing data	60

2.3 Evaluation metrics

In the C_p analyzed using MARS in the section 3, the same data were also analyzed using other analytical models. Table 2 shows the various evaluation metrics utilized for prediction comparison of these models.

Table 2. Summary of performance measures.

Measure	Calculation
Coefficient of determination (R2)	$R^2 = 1-\frac{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}{\sum_{i=1}^{N}(\mathrm{y}_i-\bar{y})^2}$
Coefficient of correlation (r)	$r = \frac{\sum_{i=1}^{N}(Y_i-\bar{Y})(\mathrm{y}_i-\bar{y})}{\sqrt{\sum_{i=1}^{N}(Y_i-\bar{Y})^2}\sqrt{\sum_{i=1}^{N}(\mathrm{y}_i-\bar{y})^2}}$
Mean square error (MSE)	$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i-Y_i)^2$
Mean absolute error (MAE)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert y_i-Y_i\rvert$

$\bar{y}$ is the mean of the target values of y_i; $\bar{Y}$ is the mean of the predicted Y_i; N denotes the number of data points in the used set, training set, testing set or the overall set.

3. Results and discussion

The prediction model of wind pressure coefficients for low-rise building is established by using MARS method. The details of MARS model are shown in Eq. (7) and Table 3. Fig. 3 shows the comparison between the predicted results and measured data. It can be seen that the fitting is excellent: R^2=0.994, r=0.997, MSE=0.001, MAE=0.026 (Training data), R^2=0.984, r=0.992, MSE=0.004, MAE=0.045 (Testing data). Obviously, the MARS model can capture the complex relationship between C_p and the influential initial parameters, and it can be a useful tool for wind pressure prediction on building surfaces.

$$\begin{aligned}Cp = &-0.696+0.0162\times BF1-0.00758\times BF2-0.0081\times BF3\\ &+0.00893\times BF4+0.00184\times BF5-0.00831\times BF6\\ &-0.0251\times BF7-0.00167\times BF8+0.0254\times BF9\\ &+0.0136\times BF10-0.0193\times BF11+0.00759\times BF12\\ &-0.00637\times BF13-0.00276\times BF14-0.00322\times BF15\\ &-0.00502\times BF16+0.00459\times BF17+0.0101\times BF18\\ &+0.384\times BF19+0.00411\times BF20-0.0199\times BF21\\ &+0.00784\times BF22-0.00295\times BF23-0.00207\times BF24\\ &+0.0104\times BF25-0.0156\times BF26+0.267\times BF27\end{aligned} \tag{7}$$

Table 3. BFs and corresponding equations of MARS model.

BF	Equation	BF	Equation
BF1	max(0,105-x3)	BF15	BF12 × max(0,0.5-x1)
BF2	max(0, x3-30)	BF16	BF2 × max(0, x2-1.5)
BF3	BF2 × max(0, x1-0.75)	BF17	BF2 × max(0,1.5-x2)
BF4	BF2 × max(0,0.75-x1)	BF18	max(0, x2-1) × max(0, x3-75)
BF5	BF1 × max(0, x2-1.5)	BF19	max(0,1-x2) × max(0,0.5-x1)
BF6	max(0, x3-105) × max(0, x2-1)	BF20	max(0, x3-90) × max(0, x2-0.67)
BF7	max(0, x3-105) × max(0,1-x2)	BF21	max(0, x3-90) × max(0,0.67-x2)
BF8	max(0, x2-1) × max(0,90-x3)	BF22	BF10 × max(0,0.67-x2)
BF9	max(0,1-x2) × max(0, x3-120)	BF23	max(0,60-x3)
BF10	max(0, x3-45)	BF24	max(0, x2-1) × max(0, x3-150)
BF11	max(0,45-x3)	BF25	max(0, x3-90) × max(0, x1-0.5)
BF12	max(0,90-x3)	BF26	max(0, x3-90) × max(0,0.5-x1)
BF13	max(0, x3-15)	BF27	max(0, x1-0.5)
BF14	BF12 × max(0, x1-0.5)		

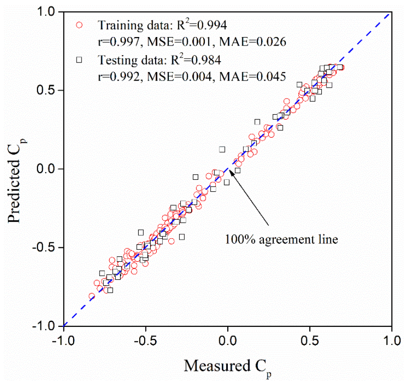

Figure 3. Comparison between measured targets and MARS predicted C_p

Wind pressure coefficients from the TPU data, MARS model, S&C equation, M&P equation and B&G FANN model are shown in Fig. 4 for the wall with *D*/*B* as a function of *θ*. As mentioned in the previous literatures (Muehleisen et al. 2013, Bre et al. 2018), the S&C equation could not fit the TPU data well for *D*/*B* ≠ 1. M&P equation and B&G FANN model achieve relatively good results for these cases, but in some wind angle range, the accuracy is still not ideal enough. As we can see, the MARS model has the best fitting results with TPU experimental data for almost any *D*/*B* and *θ*.

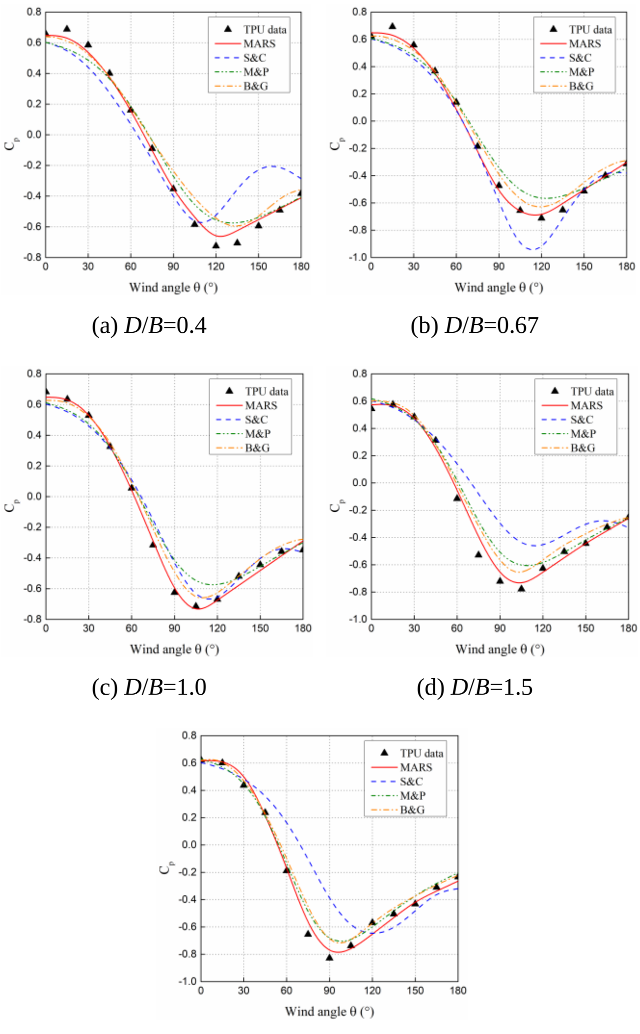

Figure 4. Building with side ratio D/B: TPU measurements vs. predictions of MARS model, the S&C and the M&P equations and the B&G FANN model of the surface-averaged pressure coefficient C_p with respect to the wind attack angle *θ*.

All performance measures of MARS model, the S&C equation, the M&P equation and the B&G FANN model are listed in Table 4. It can be observed that the MARS model could predict the C_p better than other methods.

Table 4. Comparison of performance measures.

Method	R^2	r	MSE	MAE
MARS	0.9919	0.9959	0.0017	0.0306
S&C	0.8837	0.9415	0.0244	0.1114
M&P	0.9750	0.9874	0.0053	0.0536
B&G	0.9297	0.9682	0.0148	0.0907

4. Conclusions

The main purpose of this paper is to explore the application of multivariate adaptive regression splines algorithm in the prediction of C_p on low-rise building surfaces. Based on this algorithm, a new analytical model was proposed which can accurately predict the surface-average C_p on wall of low-rise buildings. Firstly, the MARS algorithm was introduced in detail, and the accuracy of the algorithm was verified by a simple example. Then, based on the TPU database, MARS algorithm was used to establish the analytical model of the surface-average C_p on wall of low-rise buildings with side ratios of D/B=1/2.5, 1/1.5, 1/1, 1.5/1, 2.5/1 over a range of wind direction from 0° to 180°. Previous popular S&C equation and M&P equation are simple in form, but their prediction accuracy is relatively low. The application of ANN model given by B&G is difficult to use for people without relevant knowledge background. Compared with the previous analytical models, MARS model have higher prediction accuracy and good interpretability, and they are easy to implement in hand, spreadsheet, and other calculation software.

Acknowledgement

The research reported in this paper was conducted with the support of the National Key R&D Program of China (2018YFC0809400).

References

Haghiabi, A.H. 2017. Prediction of River Pipeline Scour Depth Using Multivariate Adaptive Regression Splines. *Journal of Pipeline Systems Engineering and Practice*, 8(1): 04016015.

Cóstola, D., Blocken, B. and Hensen, J. L. M. 2009. Overview of pressure coefficient data in building energy simulation and airflow network programs. *Building and Environment*, 44(10): 2027-2036.

Bre, F., Gimenez, J. M. and Fachinotti, Víctor D. 2018. Prediction of wind pressure coefficients on building surfaces using artificial neural networks. *Energy and Buildings*, 158: 1429-1441.

Friedman, J. H. 1991. Multivariate Adaptive Regression Splines. *The Annals of Statistics* 19(1): 1-141.

Swami, M. V., and Chandra, S. 1988. Correlations for pressure distribution on buildings and calculation of natural-ventilation airflow. *ASHRAE Transactions*, 94(1): 243-266.

Muehleisen, R. T. and Patrizi, S. 2013. A new parametric equation for the wind pressure coefficient for low-rise buildings. *Energy and Buildings*, 57(2): 245-249.

Zhang, W. and Goh, A. T. C. 2016. Multivariate adaptive regression splines and neural network models for prediction of pile drivability. *Geoscience Frontiers*, 7(1): 45-52.

Zhang, W., Zhang, R. and Goh, A. T. C. 2018. Multivariate Adaptive Regression Splines Approach to Estimate Lateral Wall Deflection Profiles Caused by Braced Excavations in Clays. *Geotechnical and Geological Engineering*, 36(2): 1349-1363.

Zhang, W., Zhang, Y. and Goh, A. T. C. 2017. Multivariate adaptive regression splines for inverse analysis of soil and wall properties in braced excavation. *Tunnelling and Underground Space Technology*, 64(4): 24-33.

Zhang, W. and Goh, A. T. C. 2013. Multivariate adaptive regression splines for analysis of geotechnical engineering systems. *Computers and Geotechnics*, 48 (3): 82-95.