Research on the Sensitivity of Thermal Comfort Using Sensitivity Algorithms Based on Variance and Stochastic Expansion

Peng Wang*, Yuting Zhou, Liangyu Wang, Chengkai Miao, Qikun Wang

College of Electronic Information and Automation, Tianjin University of Science and Technology, 300222, China

*E-mail: *2324365941@qq.com*

www.tust.edu.cn
(Tianjin University of Science and Technology, Tianjin, China)

Abstract

In the research field of modern architectural environment, the sensitivity research of human thermal comfort factors is of crucial significance. In this paper, first of all, common algorithms in the field of sensitivity analysis and data sets for thermal comfort research are elaborated in detail. Secondly. The variance method is taken into consideration. It has the capacity to reflect the fluctuation of the influence that different factors exert on the results. Meanwhile, the stochastic expansion method is also regarded. It is capable of handling complex non-linear relationships. A decision is made to combine these two methods. And the combined methods will be applied to conduct the sensitivity analysis of thermal comfort factors. Finally, the most critical factors for thermal comfort are successfully identified, providing an important basis for the construction and optimization of the thermal comfort prediction model.

Keywords: Sensitivity analysis, variance, stochastic expansion, thermal comfort

1. Introduction

With the continuous improvement of people 's requirements for the quality of building environment, human thermal comfort has become the key content of building environment research[1]. The physical and mental health of users is not only affected by thermal comfort, but it is also closely related to building energy consumption. Identifying the key factors affecting human thermal comfort accurately is of great significance for optimizing architectural design and improving energy efficiency. Sensitivity analysis, as an effective tool, is capable of quantifying the impact of various factors on thermal comfort, while the variance method and the random expansion method each have their own advantages in dealing with complex data relationships. The combination of these two methods is expected to provide more accurate results for the analysis of thermal comfort factors.

The rest of this article is organized as follows. The second chapter introduces the specific algorithms and data sets used. In the third chapter, we conduct sensitivity analysis on the data sets using two algorithms respectively and draw conclusions. In the last chapter, we summarize the full text.

2. Related Algorithms and Data Sets in Thermal Comfort Research

In the field of statistics, there are numerous studies on sensitivity algorithms. In this paper, the global sensitivity analysis based on variance and the sensitivity analysis based on the stochastic expansion method are mainly adopted.

2.1. Variance-based sobol global sensitivity analysis

The Sobol global sensitivity analysis was proposed by Ilya M. Sobol. It has been widely applied in various fields. For instance, in environmental science, it is used to evaluate the impacts of meteorological factors and pollution source emission parameters on the spatial distribution of pollutant concentrations; in engineering system design, it helps to identify crucial design parameters; in financial risk assessment, it is employed to analyze the effects of economic factors on the returns of investment portfolios.

The main indices of this method include the Sobol total effect index (ST), which measures the contribution of an input factor (including itself and all its interactions with other factors) to the total variance of the output variable, and the Sobol first-order effect index (S1), which reflects the contribution of the change of a single input factor itself to the variance of the output variable.

The advantages of the Sobol global sensitivity analysis are as follows: Firstly, it has strong comprehensiveness as it can consider the entire range of values of input factors rather than just local changes, thus comprehensively assessing the impacts of factors on the output results. Secondly, it can effectively quantify the impacts of interactions among input factors on the output results. In many complex systems, the interactions among factors may have significant influences on the system performance, and this method is capable of quantifying these influences. Thirdly, it is based on the variance decomposition theory, possessing a solid theoretical foundation, and the results are highly interpretable. That is to say, researchers can accurately judge the importance of each factor in the system according to the calculated results of the Sobol indices.

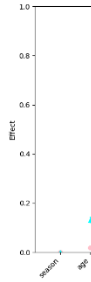
2.2. Sensitivity analysis based on stochastic xxpansion method (PCE Model)

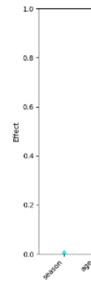
The sensitivity analysis based on the stochastic expansion method (PCE model) has been gradually developed and refined by scholars in relevant research. It is widely applied in various fields such as engineering and physics. For instance, it is used to evaluate the impacts of uncertain factors in structural reliability analysis.The main indicators include the contribution degrees of various coefficients to the results. Its advantages lie in the fact that it can effectively handle complex nonlinear relationships and approximate complex models with relatively fewer computational resources. It can also take into account multiple uncertain factors and their interactions.However, for high-dimensional and strongly nonlinear problems, there may be insufficient accuracy. There are certain difficulties in determining appropriate basis functions and solving coefficients when constructing the model. Moreover, it has certain requirements on the data distribution, and the accuracy of the results may be affected when the data characteristics are not favorable.

2.3. Introduction to data sets

The ASHRAE Global Thermal Comfort Database II (ASHRAE GTDB-II) is sourced from the American Society of Heating, Refrigerating and Air-Conditioning Engineers (ASHRAE)[2]. This database is a global thermal comfort database that has collected data on human perception of ambient temperature on a global scale. It is designed to provide a large amount of data regarding human perception of ambient temperature to support thermal comfort research and development. The ASHRAE GTDB-II database encompasses different environmental factors such as temperature, humidity, wind speed, illuminance, etc., as well as human evaluations of the environment. These data can be utilized to assess and optimize the thermal comfort performance of buildings, improve building design, and develop new thermal comfort assessment tools and techniques. The ASHRAE GTDB-II database is an important resource for thermal comfort research and development and is one of the significant tools in the fields of building design, system selection, and operational assessment. It is of great significance for ensuring the comfort and energy efficiency of buildings and helps to enhance the energy-saving efficiency of buildings and the satisfaction of occupants.

The ASHRAE GTDB-II encompasses climate data on a global scale, including data from North America, South America, Europe, Asia, Africa, Australia, and Antarctica. These regions were selected by ASHRAE for data collection because they cover a wide range of climate types, such as temperate, tropical, and subtropical climates. The extensiveness and richness of the climate data from these regions can better meet the requirements of thermal comfort prediction. Additionally, factors such as building types, building heights, and building materials in these regions all have an impact on thermal comfort prediction, and the diversity of the data from these regions can better illustrate the influence of different influencing factors on thermal comfort.





3. Experimental Design for Sensitivity Analysis

In this chapter, we first preprocessed and separated the data. Secondly, we analyzed the preprocessed data by using two sensitivity methods.

3.1. Data preprocessing

In the research field of building environment and human thermal comfort, considering that significant differences exist in the age structure, activity characteristics, and specific thermal comfort requirements among occupants in different building types, a detailed and systematic data processing procedure was carried out in this study. Firstly, the data were divided into three categories, namely office, classroom, and multi-family residential, based on building types, thereby constructing three independent data subsets. After the classification was successfully completed, null value filling was immediately performed on each sub-table. For data columns closely related to climate types, such as temperature, humidity, and clothing thermal resistance, when the season information could be clearly determined, the mode of the data in that season was used for filling to best match the typical characteristics of the season. If the season information could not be determined, the mean value was used for filling to ensure the relative integrity and rationality of the data. For data columns that have no direct correlation with climate seasons, such as age, height, weight, and metabolic rate, a unified mean value filling strategy was adopted to compensate for the information gap caused by missing data. In addition, for columns marked with non-numeric labels, such as thermal sensation evaluations (e.g., cold, hot) and gender information, they were all converted into digital codes to facilitate subsequent data analysis and model construction. Through the above rigorous and comprehensive data processing procedures, a substantial and reliable data resource was finally obtained. Specifically, 77,260 records of office data, 11,704 records of classroom data, and 13,984 records of residential data are available, as shown in Table 1. A solid data foundation is laid for further exploring the characteristics related to thermal comfort in different building types.

Table 1 data classification

type of construction	Num
office	77,260
classroom	11,704
multifamily houses	13,984

3.2. Sensitivity analysis based on variance

In this experiment, the machine learning model is first trained, and 18 input features such as season, age, gender, ht, wt, ta, top, tr, tg, rh, vel, met, clo, t_out, rh_out, fan, window, door and thermal_sensation are selected as output features. The processed data is divided into training set and test set, and the proportion of 80 %-20 % is divided. Then, the training set data is used to train the Cubist model, and the parameters of the model are adjusted to optimize the performance of the model. The optimal parameter combination is determined by cross-validation, which makes the model have better fitting effect on the training set and avoids over-fitting. The trained Cubist model is evaluated using the test set data, and the mean square error (MSE) of the model is calculated as the evaluation index，as shown in Table 2.

Table 2 Cubist model indicators

type of construction	Mean Squared Error
office	1.03288586766953
classroom	1.06793067981422
multifamily houses	0.789635208548701

Secondly, the Sobol global sensitivity analysis is carried out by using the constructed Cubist model. In this paper, Sobol sequence sampling is used to generate a large number of input feature sample combinations, which will be used for subsequent sensitivity analysis and calculation. For the input feature sample combination obtained by each sampling, the corresponding output prediction value is calculated by the Cubist model. According to the variance decomposition theory, the main effect index (S1) and the total effect index (ST) of each input feature are calculated. The main effect index measures the contribution of a single input feature 's own change to the output variance, and the full effect index takes into account the contribution of the feature and all its interactions with other features to the total output variance. The main effect and full effect index results of each input feature in three different data sets are sorted out and recorded, so as to analyze and discuss the results in the future, and further understand the influence degree and interaction relationship of each input feature on the thermal _ sensation output feature. Fig. 1, Fig. 2 and Fig. 3 show the main effects and full effects of the three types of data input features such as office, classroom, and multi-family residential.

The data show that in the office building type, age, gender, ta, top, te, tg, clo, including t _ out, rh _ out and other factors have a greater impact on the results, and the basic total effect is greater than the main effect. Among the clssroom building types, only ta, top, tr, rh and t _ out have a greater impact on thermal sensation. In the multifamily houses building type, season, age, ta, top and t _ out have a greater impact. On the whole, the influence of temperature factors on thermal sensation in all buildings is relatively large, while for other factors, different building types have different sensitivities.

3.3. Sensitivity analysis ased on the PCE model

In this experiment, season, age, gender, ht, wt, ta, top, tr, tg, rh, vel, met, clo, t _ out, rh _ out, fan, window, door and other input features and thermal _ sensation are still selected as output features.

Firstly, the PCE model is constructed. Since we are dealing with nonlinear problems, we choose the second-order model, and the basis function is the Legendre polynomial. By substituting the sampled data into the expression of the PCE model, a linear equation group is obtained. The coefficients of the PCE model are obtained by solving the linear equations by the least square method. The objective of the least squares method is to minimize the sum of squared errors between the model predictions and the actual observations, namely，

$$\min \sum_{k=1}^{n} (y_k - \hat{y}_k)^2 \quad (1)$$

where y_k is the actual observed value, $\hat{y}_k$ is the predicted value of the model, and n is the number of sample points. Independent validation data sets were used to evaluate the accuracy of the PCE model. The verification data is

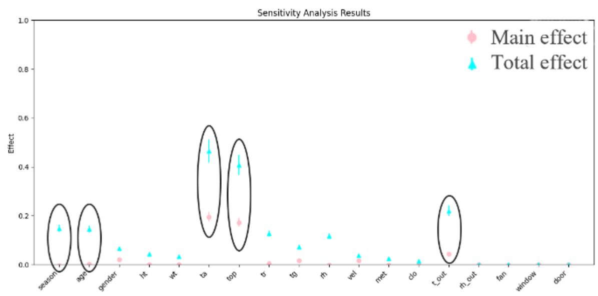


Fig. 3 multifamily houses- Sensitivity Analysis

substituted into the PCE model to calculate the prediction error.

Secondly, we conduct sensitivity analysis. The first-order sensitivity index (S1) measures the contribution of a single input variable to the variance of the output variable.

$$S1_i = \frac{V\left[E\left(y \middle| x_i\right)\right]}{V[y]} \quad (2)$$

Where $V\left[E\left(y \middle| x_i\right)\right]$ is the variance of the conditional y expection under the given conditions of x_i , and $V[y]$ is the total variance of y , The total sensitivity index (ST) measures the contribution of the input variable and all its interactions with other variables to the variance of the output variable.

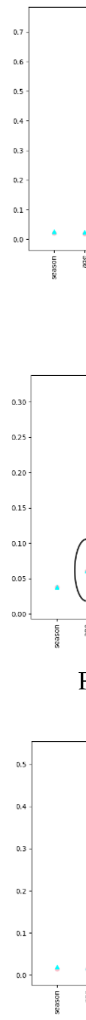


Fig. 6

$$ST_i = 1 - \frac{V\left[E\left(y|x_{-i}\right)\right]}{V\left[y\right]} \quad (3)$$

Where $V\left[E\left(y|x_{-i}\right)\right]$ is the conditional expectation variance of y under all other variable conditions except x_i. For the PCE model, S1 can be calculated by mathematical derivation of the coefficients and basis functions. According to the calculated S1 and ST indicators, the input variables are sorted. Variables with larger S1 and ST values have a greater impact on the output variables and are key sensitive factors. Compare the values of S1 and ST to understand the intensity of interaction between variables. If ST is much larger than S1, it shows that there is a strong interaction between the variable and other variables. The impact on the output variable is not only its own individual effect, but also the synergistic effect with other variables.

Fig4, Fig5 and Fig6 show the main effects and full effects of the three types of data input features based on the PCE model, such as office, classroom, and multi-family residential.

Through a detailed analysis of the data, we have obtained some important findings. Among all building types, the factor of "ta" has a relatively significant impact on the final results. Further observation reveals that the difference between its total effect and main effect is not obvious. When focusing on the "classroom" building type, factors such as age, gender, "top" and "tr" also have a considerable influence on the results. Among them, "age" may have an impact on the results related to the building due to the differences in usage habits among people of different age groups. "Gender" may affect the results because of the differences in spatial needs and usage preferences between men and women. The characteristics related to the top of the building represented by "top" and the specific building attributes or conditions represented by "tr" all play important roles among the factors influencing the results of the "classroom" building type.

4. Conclusion

This paper is mainly in the field of modern building environment research, and it is of great significance to study the sensitivity of human thermal comfort factors. Through research, it can be seen that common algorithms in the field of sensitivity analysis and data sets for thermal comfort research are crucial for in-depth exploration. In this study, in view of the fact that the variance method can show the fluctuation of the influence of different factors on the results, and the random expansion method can deal with the complex nonlinear relationship, these two methods have achieved remarkable results in the sensitivity analysis of thermal comfort factors. Finally, the most critical factors for thermal comfort are successfully identified. These results provide an important basis for the construction and optimization of thermal comfort prediction models. It is helpful to consider the influence of key factors more accurately in the subsequent design and regulation of building environment, so as to effectively improve the thermal comfort experience of human body in building environment and promote the development of thermal comfort in building environment to a more scientific and reasonable direction.

Acknowledgements

The research is supported by the major special project of carbon neutralization technology in Tianjin, China in 2024 (24ZXTKSN00070).

References

1. Cheng Chenxi, Xiangtian Deng, Xiaoyong Zhao, et al., Multi-Occupant Dynamic Thermal Comfort Monitoring Robot System. *Building and Environment* 234: 110137.
2. Földváry Ličina, Veronika, Toby Cheung, et al.Thomas Parkinson, Edward Arens, Chungyoon Chun, Development of the ASHRAE Global Thermal Comfort Database II. *Building and Environment* 142:502–12.

Authors Introduction

Ms. Peng Wang

She is a postgraduate tutor of Tianjin University of Science and Technology. In 2014, she received a doctorate from North China Electric Power University. The research direction is the functional safety assessment of safety instrumented systems.

Ms.Yuting Zhou

In 2023, she entered Tianjin University of Science and Technology.She is pursuing a Bachelor of Engineering degree in the School of Electronic Information and Automation.

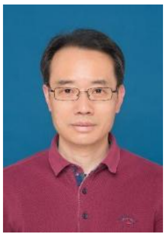
Mr. Liangyu Wang

He is currently pursuing a Bachelor of Engineering degree in the College of Electronic Information and Automation at Tianjin University of Science and Technology.

Mr. Chengkai Miao

He is a sophomore in the School of Electronic Information and Automation at Tianjin University of Science and Technology. He was admitted to Tianjin University of Science and Technology in 2023. He is currently engaged in his undergraduate studies in this university.

Mr. Qikun Wang



In 1996, he received his Bachelor of Engineering degree from the School of Electronic Information and Automation, Tianjin University of Science and Technology, China. He is a senior engineer, research direction is