

## Study on the calculation of surface subsidence coefficient based on principal component analysis and neural networks\*

Ming-Hua Wu and Xiao-Gang Xia

*School of science, Xi'an University of Science and Technology  
Xi'an, Shaanxi, 710054, China*

*†mhWu\_1978@126.com, xiayu1978@126.com*

In order to study the effect of geological factors on the surface subsidence coefficient. Based on the analysis of all influential geological factors of subsidence coefficient and the typical mobile surface observation station data of China, the influence of each factor and the influence difference were analyzed comprehensively by means of principal component analysis. Then the first and second principal component values, as inputting parameters, were used to build up a calculation model of subsidence coefficient on the basis of a principal component and artificial neural networks. And the calculation results and measured values were also compared. The results show that calculation model, which is based on principal component analysis and artificial neural networks, takes every factors into account comprehensively and produces reliable results which are much closer the reality. This new calculation model provides a new attempt to the calculation of surface subsidence coefficient.

*Keywords:* subsidence coefficient; principal component analysis; artificial neural networks

### 1. Introduction

The surface subsidence coefficient is an important parameter to characterize the mining subsidence and surface movement rule; it also is the key parameter for the prediction of surface movement and deformation in the three-underground mining research [1, 2]. The degree of accuracy of the surface subsidence coefficient prediction, directly affects the prediction of the surface movement deformation and the damage degree of the mining influence area buildings, it also has an important significance to the architectural design of town layout and the resistance to deformation.

Some valuable research results have been obtained for the prediction of surface subsidence deformation over the past years. Many modern mathematical and mechanical methods have also been introduced into the prediction and

---

\* The authors receive the financial support from NSFC(71473194, 51174156)

research of the surface deflection, such as mathematical statistics, fuzzy mathematics, continuum mechanics and so on[3,4]. A number of mathematical models, such as gray system and GIS model, which can be used to predict the surface deflection, are established[5,6].

The principal component analysis of multivariate statistical analysis, which is characterized by its simplicity, the objectivity of empowerment, has been widely used to the evaluation and ranking in many fields, such as economy, society, environmental protection, engineering computation and so on[7]. Neural network has the properties of self-organization, self-learning, strong fault tolerance and processing of deterministic and uncertain dynamic nonlinear information. So it has been widely used in rock mechanics, mining engineering and other fields. This paper established a calculation model of surface subsidence coefficient based on principal component and artificial neural network. The research also explores a new method for the theoretical calculation of surface subsidence coefficient.

## **2. The Analysis of Factors Influencing the Surface Subsidence Coefficient**

A large number of measured data and theoretical studies have confirmed that the surface subsidence coefficient is mainly related to the following factors [1-3, 5]: the overburden lithology, the mining depth  $H$  and mining thickness  $M$ , the loose layer thickness  $\delta$ , the proportion of hard stratum among the overburden, whether to repeat mining and the dimensions of workface

## **3. The Principal Component Analysis of Influencing Factors of Surface Subsidence Coefficient**

### **3.1. *The establishment of principal component analysis model***

In practice, in order to make the principal component analysis more effective, the selected sample should be larger than the number of indicators (factors) and the greater the better. However, the large sample will undoubtedly increase the amount of computation, while the requirement of large samples increases difficulty for data collection and measurement. So it is impossible to select too many samples in actual calculation. Based on the surface movement data of 208 typical observation stations in the literature [10], 25 measured data are selected as the samples.

For the selected observation data, MATLAB is used to carry on principal component analysis [7], the corresponding principal components and the cumulative contribution rate are shown as table 1.

Table1 Principal Component and their accumulative contributive proportion

| Number of principal components           | $y_1$   | $y_2$   | $y_3$   | $y_4$   | $y_5$    |
|--|---------|---------|---------|---------|----------|
| Eigen value                              | 12380   | 1675    | 242     | 1       | 0.65     |
| Contributive Proportion (%)              | 86.5844 | 11.7126 | 1.6915  | 0.0075  | 0.0040   |
| Accumulative contributive Proportion (%) | 86.5844 | 98.2970 | 99.9885 | 99.9960 | 100.0000 |

The calculating formulas for the first two principal components are following:

$$y_1 = \hat{e}_1' X^* = -0.0035x_1^* + 0.9264x_2^* + 0.0099x_3^* + 0.3760x_4^* - 0.0161x_5^* \quad (1)$$

$$y_2 = \hat{e}_2' X^* = -0.0267x_1^* - 0.3749x_2^* - 0.0078x_3^* + 0.9182x_4^* - 0.1246x_5^* \quad (2)$$

Where:

$X^* = (x_1^*, x_2^*, \dots, x_5^*) = \hat{W}(X - \bar{X}) = [\frac{x_1 - \bar{x}_1}{\sqrt{s_{11}}}, \frac{x_2 - \bar{x}_2}{\sqrt{s_{22}}}, \dots, \frac{x_5 - \bar{x}_5}{\sqrt{s_{55}}}]$  is the standardized variable of the sample matrix  $(x_1, x_2, \dots, x_5)'$

### 3.2. Meaning of the principal component analysis

The actual meaning of the two principal components can be found in the field of mining practices. As  $y_1$  and  $y_2$  are linear function of the comprehensive index (factors)  $x_1, x_2, \dots, x_5$ , and the coefficient of each variable is positive and negative, they reflect the influence on each geological factor of surface subsidence coefficient.

Further analyzing the coefficient of the two principal components, it can be seen that the significance of the two main components is obvious. For the first principal component  $y_1$  (see formula 1), the coefficients of each index are positive and negative, but the absolute value of each value is not obviously different (except the coefficients of  $x_2$  and  $x_4$ ). Of course, the larger coefficient on the index  $x_2$  and  $x_4$ , that is, the two indicators have the greatest impact on the surface subsidence coefficient.

From the second principal component (see formula 2), it can be seen that the second principal component has a large positive value on the index  $x_4$ , and has a large negative value on the index  $x_2$ . As other influencing factors have been fixed, the influence of mining depth and the loose layer thickness on the surface subsidence coefficient are relatively large, which is consistent with actual situation.

#### 4. The Calculation of Surface Subsidence Coefficient Based on the Principal Component and Neural Network

##### 4.1. The establishment of three layer BP network computing model

Because the BP neural network model has the advantages of self-learning, self-organization, strong fault tolerance, simple calculation, fast processing speed and other advantages [8, 9], it has a wide range of applications. Robert-Nielson proves that the three layers BP network model with a hidden layer can effectively approach any nonlinear continuous function. So, the network model of the surface subsidence coefficient was established by using three layer BP network structure in this paper.

##### 4.2. The structure and parameters of network

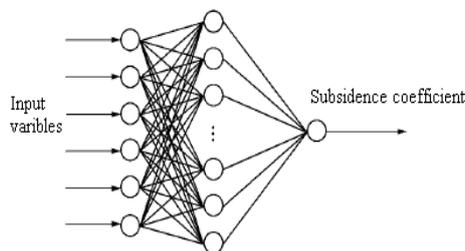


Fig. 1. ANN model for calculation of surface subsidence coefficient

In the system, the surface subsidence coefficient, as a value, determines the single output network structure. Based on the above analysis, it can be seen that the subsidence coefficient is the comprehensive result of all influencing factors. In this paper, the influence factor is simplified as two comprehensive indexes  $y_1$  and  $y_2$ , the first principal component and the second principal component. Here, the two principal components values are used as the network input parameters and network structure as shown in figure 1.

##### 4.3. The test of computational model

According to the observation data given in table 1, the two principal component values of each observation station are calculated out with the principal component calculation formula obtained in this paper, and the calculation results of the two principal component values are normalized as the value of the [0,1] interval. At last, the improved BP network algorithm is used to train the network, and the average error less than 0.01 of the network system is as the ending conditions.

#### 4.4. The measurement of network performance

Values of the No.20 to No.25 observation station are used to test the trained network model with the two principal components, and the results are shown in table 2. It can be seen that the maximum absolute error between the surface subsidence coefficient calculated by the network model and the measured value is 0.0178, and the relative error is 3.07%. Compared with other theoretical calculation results which only consider the part influencing factors, the error of the calculation results of the network model is small, which has high precision, and can meet the needs of engineering practice.

Table 2 Comparison of calculated results of ANN with observed values

| Number of Stations | The first principal component | The second principal component | Calculated value | Measured value | Absolute error | Relative error |
|--------------------|-------------------------------|--------------------------------|------------------|----------------|----------------|----------------|
| 20                 | 0.0995                        | -0.4674                        | 0.7318           | 0.72           | 0.0118         | 1.64%          |
| 21                 | 2.5471                        | -8939                          | 0.7022           | 0.72           | -0.0178        | 2.47%          |
| 22                 | -0.7338                       | -0.4661                        | 0.1443           | 0.14           | 0.0043         | 3.07%          |
| 23                 | -0.2995                       | -0.0206                        | 0.8327           | 0.85           | -0.0173        | 2.03%          |
| 24                 | 3.2034                        | 1.6590                         | 0.8471           | 0.83           | 0.0171         | 2.06%          |
| 25                 | 1.3323                        | 0.7522                         | 0.9591           | 0.95           | 0.0091         | 0.96%          |

#### 5. Conclusions

(1) The influence difference of each factor on the surface subsidence coefficient is analyzed comprehensively through the principal component analysis model. From the model, we can know that the mining depth has the maximum influence on the surface subsidence coefficient, and the thickness of the loose layer is the second influence factor. And the others are shown as the follows: proportion of hard stratum, sturdiness coefficient and the mining thickness.

(2) Using the first and second principal component values of each observation station as the input parameters, the calculation model of surface subsidence coefficient based on principal component analysis and artificial neural network is established, developed for the theoretical calculation of surface subsidence coefficient.

(3) Through the training and performance testing of the network model, the feasibility, accuracy and science of the calculation model based on the principal component and artificial neural network used to calculate the surface subsidence coefficient has been proved.

## References

1. G. Q. He, L. Yang and G. D. Ling. "Mining Subsidence Research". China Univ. of Min and Technol Press, 1994:118-123,125-143
2. Zh. X. Tan and K. Zh. Deng. "Review and expectation of study on mining under buildings". *Journal of Liaoning Technical University*, 2006, 25(4): 485-488.
3. W. B. Guo, K. Zh. Deng and Y. F. Zou. "Study on artificial neural network method for calculation of subsidence coefficient". *Chinese Journal of Geotechnical Engineering*, 2003, 25(2):212-215.
4. J. J. Chen, Y. F. Zou and W. B. Guo. "Study on the Relationship between Subsidence Coefficient and Mining Degree under A Thick Alluvium Stratum". *Journal of Mining & Safety Engineering*, 2012, 29(2):250-254.
5. J. Zh. Wang, Zh. Q. Chang and Y. Chen. "Study on mining degree and patterns of ground subsidence in condition of mining under thick unconsolidated layers". *Journal of China Coal Society*, 2003, 28(3):232-234.
6. X. D. Zhao and Zh. Q. Song "GIS raster digitized method with face element in predicting subsidence and its application". *Journal of China Coal Society*, 2000, 25(5):473-477.
7. Sh. K. Qin. "Theory and Application of Synthesis Appraise. Publishing." *House of Electronics Industry*, 2003.
8. X. T. Feng and Y. J. Wang. "Mining intelligent system - artificial intelligence and neural network application in mining engineering". *Metallurgical Industry Press*, 1994:164-213.
9. Q. MIAO, Sh. Q. YANG and X. Y. OUI. "Establishment and Application of Grey-Neural Network Forecasting Model of Coal and Gas Outburst". *Journal of Mining & Safety Engineering*, 2008,25(3):309-312
10. China Coal Board. "Rules for the establishment and exploitation of coal pillars in buildings, water bodies, railways and main shaft Lane". *Coal Industry Press*, 2000:131-156.