Research Article

Shoukai Chen, Yongqiwen Fu, Lei Guo*, Shifeng Yang, and Yajing Bie

Statistical Law and Predictive Analysis of Compressive Strength of Cemented Sand and Gravel

<https://doi.org/10.1515/secm-2020-0030>

Received Apr 13, 2020; accepted Jul 18, 2020

Abstract: A data set of cemented sand and gravel (CSG) mix proportion and 28-day compressive strength was established, with outliers determined and removed based on the Boxplot. Then, the distribution law of compressive strength of CSG was analyzed using the skewness kurtosis and single-sample Kolmogorov-Smirnov tests. And with the help of Python software, a model based on Back Propagation neural network was built to predict the compressive strength of CSG according to its mix proportion. The results showed that the compressive strength follows the normal distribution law, the expected value and variance were 5.471 MPa and 3.962 MPa respectively, and the average relative error was 7.16%, indicating the predictability of compressive strength of CSG and its correlation with the mix proportion.

Keywords: Cemented sand and gravel; Compressive strength; Boxplot; Normal distribution; BP neural network

1 Introduction

In contemporary society, the harmonious coexistence between man and nature is a prerequisite for healthy and sustainable development. The relationship between dams and ecological environment is receiving more and more attention, and the construction of environmentally-friendly dams has become a mainstream trend [1]. CSG dam is a new type of dam derived from the pursuit of high efficiency, low cost and minimal impact on the natural environment. CSG dam was first proposed by Raphael J M and Londe P [2, 3], which was developed based on Roller Compacted Concrete (RCC) dam and Concrete Face Rockfill (CFR) dam. The CSG dam building material is a new type of dam material obtained by mixing a small amount of cementing materials (mainly cement and fly ash), water, raw gravel of the riverbed or excavated waste materials with simple equipment [4, 5], in line with the new concept of “The dam structure is optimized in order to make better use of local materials” and “Proper materials can be selected for the different parts of the dam in order to realize better function of structures” [1, 6]. As a structural material, CSG needs to be strong enough to withstand different loads, thus ensuring the safety and reliability of the project.

Yang *et al.* [7] made a uniaxial compressive test and found that for every 10 kg/m³ increase in cement content, the compressive strength increased by 15% to 20%, and the economic and optimal amounts of fly ash were 40% and 50% of the total cementing materials (mainly cement and fly ash). Guo *et al.* [8] concluded a linear relationship between CSG cube compressive strength and split strength through strength tests, and the split strength was 7% to 12% of the compressive strength. Yang *et al.* [9] studied the effect of cement content on the deformation characteristics of CSG materials, and the results showed that with the increase of cement content, the failure strain of CSG materials decreases, the brittleness increases, and the initial modulus grows exponentially. Takashi *et al.* [10], based on the Nagashima sand blocking dam project, carried out a study on the effect of water consumption per unit on

***Corresponding Author: Lei Guo:** School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; Key Laboratory of Water Environment Simulation and Treatment of Henan Province, Zhengzhou 450002, China; Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Zhengzhou 450045, China; Email: glboss@126.com

Shoukai Chen: School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450045, China; Key Laboratory of Water Environment Simulation and Treatment of Henan Province, Zhengzhou 450002, China; Collaborative Innovation Center of Water Resources Efficient Utilization and Protection Engineering, Zhengzhou 450045, China

Yongqiwen Fu, Shifeng Yang, Yajing Bie: School of Water Conservancy, North China University of Water Resources and Electric Power, Zhengzhou 450045, China

CSG performance, and suggested that it was appropriate to control VC values from 5s to 20s in practical application. Feng *et al.* [11] conducted a research on the effect of sand ratio, water-binder ratio, cement content and fly ash content on the strength, and it was recommended that the sand ratio in actual projects should range from 18% to 32%, water-binder ratio should range from 0.7 to 1.3, and CSG minimum cement content should be not less than 30 kg/m^3 . Although these researches on CSG performance have achieved some positive results and provided theoretical guidance for the application of practical engineering, most of them are based on the analysis of the test results of variable factors, which does not reflect the potential law of CSG material performance from a statistical perspective, and lacks the multi-factor correlation analysis based on data prediction. To this end, based on the results of previous CSG tests, this paper conducted a study on CSG compressive strength performance using statistical and predictive methods, which provides reference for CSG's mix proportion design and innovative applications.

2 Data processing and analysis

The basic purpose of data processing is to extract valuable and meaningful data from the original data, form the data analysis style, and ensure the consistency and validity of the data, which is of great significance to the subsequent data analysis. Data analysis is to analyze the collected data and extract valuable information through appropriate analysis methods and tools, so as to get effective conclusions. In the established data set, the data is relatively multi-source, so it is hard to form a continuous whole with other data points due to the lack of data and the large dispersion of data points, which has an impact on the statistics and prediction analysis of data and could lead to wrong conclusions. Therefore, these data points are called “outliers” and eliminated in subsequent analysis.

2.1 Source and analysis of data

The test data was divided into two parts according to its source. One part was based on the CSG material performance test conducted in accordance with the *Test Code*

Figure 1: Specimen preparation process

Figure 2: Data set of CSG 28d compressive strength

for Hydraulic Concrete (SL352-2006) and the Specification of Soil Test (SL-237-1999), with the test process shown in Figure 1 and the specific method shown in the literature [2]. In order to ensure the universality and authenticity of the data while improving the processing capacity and prediction effect of statistical model analysis, another part cited the results of literature [11, 26, 27] to form a data set of the CSG's 28-day compressive strength (Figure 2). There were 99 sets of compressive strength data in the data set, in which the minimum value is 1.00 MPa and the maximum is 17.7 MPa. The data were mainly concentrated in 1-10 MPa, accounting for about 96% of the total sample number.

In addition, the CSG mix proportion was another basic data needed for the research in this article. In the selection of the mix proportion corresponding to the compressive strength data, the cement content ranged from 30 kg/m³ to 60 kg/m³, which met the requirements of CSG as a super lean cementing material although several mix proportions were slightly greater than 60 kg/m³. The range of fly ash content was 0-80 kg/m³, the range of water-binder ratio was 0.5-3.0, the range of sand-gravel content was 1,278-2,081 kg/m³, and the range of sand ratio was 0.10-0.45.

2.2 Identification and removal of outliers

Due to many data sources and differences in the mix proportion and test process corresponding to different strengths, some values may be “abnormal” and should be rejected. This article adopted Boxplot to identify and remove such abnormal values. Boxplot, as an effective data visualization tool, is not only intuitive, easy to understand, but also has no requirement for the distribution form, and no limitation for the data, which is more suitable for analyzing the CSG compressive strength with the distribution law unknown [12]. The implementation of this method was shown in Figure 3 [13]:

1. Set five statistics in the CSG compressive strength values: minimum value, first quartile Q_L (lower quartile), second quartile Q_M (median), third quartile Q_U (upper quartile) and maximum value, where Q_L , Q_M , and Q_U were the numbers arranged in ascending order at the 25%, 50%, and 75% of the sample data respectively, Q_L and Q_U were approximate estimates of the two quartiles of the data set.
2. Calculate the interquartile range (IQR), namely the distance between Q_U and Q_L .
3. Calculate the internal limits, namely $Q_L - 1.5IQR$ and $Q_U + 1.5IQR$.

Figure 3: Realization principle of boxplot method

Figure 4: Boxplot of 28d compressive strength

4. Identify outliers, namely values less than $Q_L - 1.5IQR$ or greater than $Q_U + 1.5IQR$.

The Boxplot of these 99 sets of compressive strength data was shown in Figure 4. The Boxplot of 28d compressive strength was flat in shape, and the square box was basically symmetrical around the median, indicating that the data distribution was concentrated and the skewness was weak. Three abnormal values were obtained and directly rejected, which were 17.7 MPa, 16.9 MPa and 15.6 MPa respectively.

3 Statistical analysis

According to the data processing results, the statistics of the remaining data was calculated, with the results shown in Table 1.

It can be seen from Table 1 that the expected value, variance, and standard deviation of the 28-day compressive strength sample data were: 5.471 MPa, 3.962 MPa, and

Table 1: Statistics of 28d compressive strength sample data

		Statistics							
Number of cases	96	Standard deviation	1.990	Skewness	−0.032	Kurtosis	−0.232	Minimum	1.00
Average value	5.471	Variance	3.962	Standard deviation of skewness	0.246	Standard deviation of kurtosis	0.488	Maximum	10.49

Table 2: Single-sample K-S test results

Summary of hypothesis testing			
Null hypothesis	Test	Significance	Decision makers
The 28d compressive strength data showed normal distribution, with an average value of 5.471 and a standard deviation of 1.990.	Single-sample Kolmogorov-Smirnov test	0.200	Keep the null hypothesis

Note: It showed progressive significance with a significance level of 0.05.

1.990 MPa respectively. According to the skewness kurtosis test, $|-0.032 \div 0.246| = 0.130$, and $|-0.232 \div 0.488| = 0.475$, which were all less than 1.96, indicating that the 28d compressive strength sample data was normalized [14]. In order to further prove that it obeys the law of normal distribution, a single-sample Kolmogorov-Smirnov (K-S) test was performed, with results shown in Table 2.

According to the results of skewness kurtosis test and K-S test, it can be considered that the compressive strength data of CSG material obeys the normal distribution law.

The normal distribution probability density function is [15]:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (1)$$

Where, $f(x)$ is the probability density function, σ is the standard deviation, and μ is the sample mean.

According to the normal “3 σ ” principle, the normal distribution function can be applied to obtain the probability P of the samples falling into the interval (a, b) . The distribution function is:

$$P(a < x \leq b) = \frac{1}{\sigma\sqrt{2\pi}} \int_a^b e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx \quad (2)$$

According to formula (2), the probability that the 28d compressive strength sample data fell in the interval $(\mu - \sigma, \mu + \sigma)$, $(\mu - 2\sigma, \mu + 2\sigma)$, and $(\mu - 3\sigma, \mu + 3\sigma)$ was: 68.26%, 95.44%, and 99.74% respectively. According to formulas (1) and (2), the statistical distribution characteristics of 28d compressive strength data were obtained, the normal distribution curve was drawn, and the interval frequency was determined, as shown in Figure 5.

Figure 5: Statistical characteristics of 28d compressive strength data

4 Predictive analysis

4.1 BP neural network

In the 1980s, D. E. Rumelhart *et al.* [16, 17] proposed a multi-layer feedforward network trained by the error back-propagation algorithm (Back-Propagation Network, BP network for short), which is currently the most widely used and mature artificial neural network algorithm. BP neural network is a supervised learning algorithm based on the least square method, and features high accuracy and high versatility. Its topology includes three parts: the

Table 3: BP neural network training sample data

Training group	Cement (kg/m ³)	Fly ash (kg/m ³)	Water consumption (kg/m ³)	Sand (kg/m ³)	Gravel material (kg/m ³)	The biggest size (mm)	Sand ratio (%)	Water-binder ratio	Compressive strength (MPa)
1	40	40	56	594	1847	150	0.25	0.70	9.20
.....									
3	40	40	88	573	1783	150	0.25	1.10	4.20
.....									
5	50	0	65	478	1982	150	0.20	1.30	8.90
.....									
78	60	30	90	434	1736	40	0.20	1.00	4.60

Table 4: BP neural network prediction sample data

Prediction group	Cement (kg/m ³)	Fly ash (kg/m ³)	Water consumption (kg/m ³)	Sand (kg/m ³)	Gravel material (kg/m ³)	The biggest size (mm)	Sand ratio (%)	Water-binder ratio	Compressive strength (MPa)
1	40	40	82	462	1915	150	0.194	1.025	4.50
.....									
3	50	40	100	864	1296	40	0.4	1.11	3.75
.....									
5	30	0	60	484	2007	150	0.2	2.00	5.20
.....									
18	40	40	90	436	1744	40	0.2	1.13	5.36

input layer, the hidden layer, and the output layer. All neurons in adjacent layers are interconnected, while each neuron in the same layer is independent with each other. When the output value of the output layer does not match the expected value, the weights and thresholds of the network will be modified through error backpropagation, to ensure the squared error between the output value and the expected value within the threshold range [18, 19].

At present, this method has been successfully applied to the study of the cement-based materials. For example, Liang *et al.* [20] applied the BP neural network to build a multi-factor model and predict the compressive strength of concrete in dry and wet environments, with an average error of 1.09%; Chen *et al.* [21] applied the BP neural network to establish a prediction model of the performance of the recycled aggregate permeable concrete based on the permeability and strength properties, with the average relative error within 10%. This article attempted to use this method to predict the compressive strength of CSG.

4.2 Sample data

After removing outliers, the 96 sets of sample data involved in the prediction were divided into training group (78 sets, accounting for 71.25%) and prediction group (18 sets, accounting for 18.75%). The data were randomly allocated to

ensure that the predictive analysis is representative. The training and prediction data were shown in Tables 3 and 4.

4.3 Construction of network model

A prediction model of CSG compressive strength was established with 8 parameters namely cement content, fly ash content, water content, sand content, sand-gravel content, biggest size of sand and gravel, sand ratio and water-binder ratio as the input layer, and 28d compressive strength as the output layer. In addition, there was also one hidden layer, with the number of hidden nodes determined by cut-and-try method [22]:

$$h = \sqrt{m + n} + a \quad (3)$$

Where, h is the number of nodes in the hidden layer; m is the number of nodes in the input layer; n is the number of nodes in the output layer; a is an adjustment constant taken from 1 to 10.

After several trials, when a was 10, the training effect reached the best; thus an 8-13-1 prediction model was constituted. The topology of the network model was shown in Figure 6.

The network model training was achieved with the help of Python software. The *logsig* function was used as the hidden layer transfer function, the *purelin* function

Figure 6: Schematic diagram of BP network topology

was used as the output layer function, and the *traingdm* function in the BP algorithm was used as the training function. The number of iterations, the learning rate, and the correction coefficient were set to 15,000 times, 0.05, and 0.10 respectively.

A large difference in the input parameter range may affect the initialization of the network, and cause imbalance in the pattern classification. Similarly, if data with a large range was directly input into the network, the weight would become abnormally large after an accumulator, making it difficult for the network to converge. In addition, considering that the range of the non-linear activation function in the BP neural network was limited to $[-1, 1]$ or $[0, 1]$, the target data for network training should also be mapped to the range of the activation function [23] so that the discrimination of the activation function can be used fully to achieve a better prediction effect. Therefore, in order that the network can converge faster and more accurate, the data was normalized according to the formula (4).

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (4)$$

Where, y is the normalized value; $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the data column; x is the original sample data.

4.4 Analysis of prediction results

As seen from Table 5, the absolute error range of predicted values of CSG 28d compressive strength was from 0.024 to 1.096 MPa, with the average value of 0.39 MPa and the relative error range was from 0.52 to 20.45%, with the average value of 7.16%. The measured value of the CSG compressive

Figure 7: Comparison of measured and predicted values of 28d compressive strength

Figure 8: Histogram of relative errors of 28d compressive strength

strength was in good agreement with the predicted value, indicating that the prediction effect was good.

Figure 7 showed the comparison between the measured values and predicted values of the CSG 28d compressive strength, and Figure 8 showed the relative errors. From them, the error of most prediction results was less than or slightly larger than the average value, but there were still data points with a large error, such as Groups 5 and 18, whose average relative errors reached 13.73% and 20.45% respectively, about 2 times and 3 times of the average relative error.

After analysis, it is found that: ① Some data was estimated or missed in the process of collection, causing a certain random error. ② The limitation of the BP network model itself led to a certain system error. ③ Due to different sources of test samples, there was a large difference in

Table 5: Error table of 28d compressive strength

SN	Measured value/ Predicted value (MPa)	Absolute error (MPa)	Relative error %	SN	Measured value/ Predicted value (MPa)	Absolute error (MPa)	Relative error %
1	4.50/4.82	0.321	7.12	10	3.80/3.59	0.213	5.62
2	5.49/5.05	0.436	7.94	11	10.49/10.09	0.399	3.81
3	3.75/3.39	0.361	9.63	12	6.32/6.37	0.047	0.74
4	5.79/5.61	0.179	3.09	13	5.97/5.26	0.712	11.92
5	5.20/4.49	0.714	13.73	14	5.14/4.6	0.545	10.60
6	6.80/6.62	0.185	2.72	15	4.56/4.58	0.024	0.52
7	6.20/5.58	0.622	10.03	16	6.32/6.25	0.066	1.05
8	6.50/6.91	0.406	6.24	17	4.06/3.75	0.312	7.67
9	6.30/5.92	0.382	6.06	18	5.36/4.26	1.096	20.45

production processes and material ratios. For example, the same compressive strength value may have different material ratios, which brought about certain difficulties to the training of BP neural network. ④ The training sample data and factors considered may not be comprehensive enough. If data or factors were added, this error may be reduced.

In general, the CSG compressive strength prediction model based on the BP neural network has higher accuracy.

5 Conclusions

In this paper, the data sets of the mix ratio and compressive strength of CSG were established through different data sources, and the box plot method was used to test and eliminate possible outliers in them. On this basis, the non-parametric test method and BP neural network were adopted to analyze and study the distribution and predictability of the compressive strength of CSG, and the conclusions are as follows:

1. A data set of CSG mix proportion and 28d compressive strength was established, with the range of cement content is 30-60kg/m³ and the range of compressive strength is 1-10MPa. Based on the Boxplot method, it was concluded that the CSG 28d compressive strength data was relatively concentratively distributed and weakly skewed, and three outliers were excluded, namely 15.6 MPa, 16.9 MPa, and 17.7 MPa.
2. With the help of skewness kurtosis test and single-sample K-S test, it was founded that the compressive strength of CSG follows the normal distribution law while the expected value, variance and standard de-

viation were: 5.471MPa, 3.962MPa, and 1.990MPa respectively.

3. A CSG compressive strength prediction model was established based on the BP neural network to predict the compressive strength using its mix proportion, and results showed that the predicted value was in good agreement with the measured value as the average relative error and absolute error were 7.16% and 0.39 MPa respectively, which proved the predictability of CSG compressive strength and a high correlation between the compressive strength and the mix proportion. In addition, it also showed that the BP neural network model can be used as a method for predicting the compressive strength of CSG, providing a reference for the design of the CSG mix proportion.

Acknowledgement: This work was supported by the North China University of Water Resources and Electric Power. The authors also would like to acknowledge the support of the National Key R & D Plan (2018YFC0406830).

References

- [1] Jia JS, Michel L, Jin F, Zheng CY. The Cemented Material Dam: A New, Environmentally Friendly Type of Dam. *Engineering*. 2016; 2(4):490-497.
- [2] Sun MQ. Study on Mechanical Properties, Durability and Dam Type of Cemented Sand and Gravel. 1st ed. China Water & Power Press; 2016.
- [3] Deng ZQ, Zhang L, Chen Y, Dong JH, Yang BQ. Contrastive Analysis on Structural Characteristic of Hardfill Dam and Gravity Dam. *Advanced Engineering Sciences*. 2014;46 S1:63-68.
- [4] Research on the Form of CSG Dam [dissertation]. Zheng Zhou: North China University of Water Resources and Electric Power;

- 2016.
- [5] Feng W, Jia JS, Ma FL. Study on durability of dam materials and new-type protective materials for cemented sand and gravel dam. *J Hydraul Eng.* 2013; 4(44):500-504.
 - [6] Jia JS, Zhao C, Miao L, Zhang DQ, Luo ZK. Development and application of the CSG dam construction quality monitoring system. *J Chin Inst Water Resour Hydr Res.* 2018;1(16):1-8.
 - [7] Yang SF, Chai QH, Sun MQ. Study on the Effect of the Amount of Cementitious Materials on the Compressive Strength of CSG Material. *Yellow River.* 2016; 7(38):92-94.
 - [8] Guo L, Liu GS, Chai QH, Sun MQ. Experimental Study on the Relationship Between Compressive Strength and Splitting Tensile Strength of CSG Material. *Yellow River.* 2016;9(38):86-87.
 - [9] Yang J, Cai X, Guo XW, Zhao JL. Effect of Cement Content on the Deformation Properties of Cemented Sand and Gravel Material. *Appl Sci.* 2019; 9(11): 2369.
 - [10] Application of CSG Method to Construction of Gravity Dam. In: Takashi Y, Yoshio O, Mikio K. *Proceedings of 20th ICOLD Congress.* Beijing, China: China Society for Hydropower Engineering; 2000:989-1007.
 - [11] Feng W, Jia JS, Ma FL. Study on design parameters of mix proportion for cemented sand and gravel (CSG). *Water Resour Hydr Eng.* 2013; 2(44):55-58.
 - [12] Walker ML, Dovoedo YH, Chakraborti S, Hilton CW. An Improved Boxplot for Univariate Data. *STATISTICS & PROBABILITY.* 2018; 4(72):348-353.
 - [13] Babura BI, Adam MB, Abdul Samad AR, Anwar F, Bashir Y. Analysis and Assessment of Boxplot Characters for Extreme Data. *J. Phys.: Conf. Ser.* 2018;1(1132):1-9.
 - [14] Qiu HZ. *Quantitative Research and Statistical Analysis.* 5th ed. Chongqing University Press; 2013.
 - [15] Xiao JZ, Lin ZB, Zhu J. Effects of Recycled Aggregates' Gradation on Compressive Strength of Concrete. *Advanced Engineering Sciences.* 2014; 4(46):154-160.
 - [16] Zhang NY, Yan PF. *Neural Network and Fuzzy Control.* 1st ed. Tsinghua University Press; 1998.
 - [17] Studies on simultaneous determination of multi-component by chemiluminescence and spectrophotometry with ANN [dissertation]. Sichuan: Sichuan University; 2005.
 - [18] Xue HZ, Cui HW. Research on image restoration algorithms based on BP neural network. *J. Vis. Commun. Image R.* 2019; 59:204-209.
 - [19] Zhou X, Zhu XY, Wen CY, Chen S. Application of ANN to predict the drainage in mine. *J Hydraul Eng.* 2000; 12(31):59-64.
 - [20] Liang CY, Qian CX, Chen HC, Kang WC. Prediction of Compressive Strength of Concrete in Wet-Dry Environment by BP Artificial Neural Networks. *ADV MATER SCI ENG.* 2018;(2018):1-11.
 - [21] Chen SK, Chen JL, Wang LY, Li HR, Guo L. Key Performance Statistics of Recycled Aggregate Pervious Concrete and Prediction Analysis. *Journal of Building Materials.* 2019; 2(22):214-221.
 - [22] Performance Prediction of Recycled Concrete based on BP neural network [dissertation]. Zhengzhou: Zhengzhou University; 2015.
 - [23] Prediction of the recycled concrete strength based on BP neural network [dissertation]. Qingdao: Qingdao Technological University; 2012.
 - [24] Study on the Influence of Fly Ash on the Mechanical Properties of CSG Material [dissertation]. Zhengzhou: North China University of Water Resources and Electric Power; 2016.
 - [25] Study on the Relationship Between Water Consumption and other Factors Based on the Compressive Strength of CSG Material [dissertation]. Zhengzhou: North China University of Water Resources and Electric Power; 2017.
 - [26] Study on Failure Mechanism of Cement Sandy Gravel Material Based on Particle Flow Code [dissertation]. Zhengzhou: North China University of Water Resources and Electric Power; 2018.
 - [27] Wang CX, Zhang DX, Hu W, Wen YX. Study on the influence of mix ratio parameters of CSG materials on basic mechanical properties. *Hunan Hydro & Power.* 2018;6:46-48+61.