Analysis about factors affecting the degree of damage of buildings in earthquake

Jing JIA¹ and Jinghong YAN²

¹ Room 1406, Building 1 of College of Engineering, Ocean University of China, No.238 Songling Road, Qingdao, 266100 P.R.China
² Room A726, Building 25, Tianjin University, No.92 Weijin Road, Tianjin, 300072 P.R.China

E-mail: yjh_flyatdawn@163.com; TEL: 86-15122520952

Abstract: Earthquakes have been affecting human’s safety through human’s history. Previous studies on earthquake, mostly, focused on the performance of buildings or evaluating damages. This paper, however, compares different factors that have influence on the damage of buildings with a case study in Wenchuan earthquake, using multiple linear regression methodology, so as to identify to what extent this factors influence the buildings’ damages, then give the rank of importance of these factors. In this process, authors take the type of structure as a dummy variable to compare the degree of damages caused by different types of structure, which were barely studied before. Besides, Factor Analysis Methodology (FA) will be adapted to classify factors, the results of which will simplify later study. The outcome of this study would make a big difference in optimizing the seismic design and improving residential seismic quality.

1. Introduction

There is no need to emphasis damages caused by earthquakes, especially by buildings’ falling, which cause almost 95% of casualties. Therefore, it is weighty to study buildings’ damages in earthquakes. Scholars have been studying in buildings’ damages in earthquakes in terms of destruction mechanism of building or structure. Young-Ji Park and Alfredo H.-S. Ang(1985) proposed a method for evaluating structural damage of reinforced concrete buildings in which the damage can be expressed as a linear combination of the maximum deformation and absorbed hysteretic energy under random earthquake excitation[1]. They also proposed a model in which available monotonic and cyclic test data were analyzed to evaluate the statistics of the appropriate parameters[2].
Fajfar and Peter Gaspersic(1996) used N2 method to analyze seismic damage of reinforced concrete buildings, which characterized by the use of two separate mathematical models, application of the response spectrum approach and of the nonlinear static analysis, and the choice of a damage model which includes cumulative damage[3]. A.Ghobarah and H. Abou-Elfath(1999) discussed a number of available response-based damage indices and proposed a practical method based on the static pushover analysis to estimate the expected damage to structures[4]. Rob Shepherd(2000) reviewed the acceptability criteria, for vertical flexibility of residential building floor systems and concluded the typical timber frame residence is vulnerable to being loosened up by earthquake shaking[5]. Brendon A.Bradley(2010) deemed that conventionally fragility functions, defining the probability of incurring at least a specified level of damage for a given level of seismic demand, did not take epistemic uncertainties into consideration. Based on the situation, he presented and discussed the sources of epistemic uncertainty in fragility functions, their consideration, combination, and propagation[6].

Studies mentioned above are concerned with damage of particular structure, component and failure mechanism. There are few studies on factors and its’ influence degree affecting buildings’ damage from perspective of Multivariate Regression Analysis. By extracting information from Report of safety appraisal of Houses in Mianyang earthquake as data resource, this research try to rate factors that affect houses’ falling using Multivariate Regression Analysis and classify them into several groups according to results of Factor Analysis. In the end, suggestions are given based on results of regression coefficients and factor analysis, which may provide directional guidance to damage prevention in the future.

2. Data Collection and Processing
After Wenchuan earthquake in SiChuan province of China, faculties in department of civil engineering of Ocean University of China went to stricken areas to conduct houses’ safety appraisal and damage evaluation, thus got 231 Safety Appraisal Reports(SAR), including photograph of damage and digital data. 125 of them satisfying this research was selected as original samples.

2.1 Extraction of Independent Variables
First of all, independent variables are extracted from SAR, including types of structure, number of layer, year of construction. Secondly, Anti-seismic Design code was referred and three independent variables was selected: the degree of regularity of house, roof integrity and seismic fortification intensity. Besides, houses’ maintenance status can be determined by observing photographs and site grade by address of the house. Finally, eight independent variables affecting damage of houses in earthquake are selected. They are: Structure Type(ST), Number of Layer(NL), Years of Construction(YC), Degree of Regularity of House(DRH), Roof Integrity(RI), Seismic Fortification Intensity(SFI), Maintenance Status(MS) and Site Grade(SG).

2.2 Quantification of data
Original information should be quantified for the use of multiple regression analysis and factor analysis.
First of all, the value of independent variables “Structure Type” is of no order and restricted in three types in this research: brick-concrete structure, bottom-frame structure and frame structure. Namely, there are three value of independent variables ST. In order to analyzing difference of the three structure types, dummy variables D1, D2 are proposed in the study.

Set: $D_1 = \begin{cases} 1 & \text{Brick – concrete structure} \\ 0 & \text{Non – Brick – concrete structure} \end{cases}$

$D_2 = \begin{cases} 1 & \text{Bottom – frame structure} \\ 0 & \text{Non – Bottom – frame structure} \end{cases}$

Secondly, for independent variable “Years of Construction”, “Seismic Fortification Intensity”, “Number of Layer” and ‘Site Grade”, original data is adopted for convenience and authenticity.

For the value of independent variable “Maintenance Status”, “Roof Integrity”, “Degree of Regularity of House”, quantification is conducted according its original grade. Results of quantification are seen in table 1.

### Table 1 Quantification of Variables of Grade

| Grade | A | B | C | D |
|-------|---|---|---|---|
| Value | 1 | 2 | 3 | 4 |

Similarly, the independent variable “Degree of Damage of Houses” is quantified according to status of damage, as shown in table 2.(Noted: status of damage is recorded in the report.)

### Table 2 Quantification of Degree of Damage of Houses

| Status of damage | No damage | Slight damage | Moderate damage | Heavy damage | Collapse |
|------------------|-----------|---------------|-----------------|--------------|----------|
| quantification   | 1         | 2             | 3               | 4            | 5        |

### 3. Multiple Regression Analysis

#### 3.1 Model’s Building

Before building the multiple regression model, a series of variables should be set.

Set:
- Structure Type(ST)=X1, represented by dummy variable “D1” and “D2”.
- Years of Construction(YC)=X2
- Seismic Fortification Intensity(SFI)=X3
- Number of Layer(NL)=X4
- Maintenance Status(MS)=X5
- Roof Integrity(RI)=X6
- Degree of Regularity of House(DRH)=X7
- Site Grade(SG)=X8
- Degree of Damage of Houses(DDH)=Y

Then, the multiple regression model in this study is built as follows:
Regression Coefficients and Analysis

Quantified data of independent variables is the value of Xi and Y. After inputting them into SPSS and running, parameters are got as table 3.

| Model  | Non-standard Coefficient | Standard error | Standard coefficient | t   | Sig. |
|--------|--------------------------|----------------|----------------------|-----|------|
| (constant) | 45.830                    | 22.009         |                      | 2.082 | .040 |
| D1     | .321                      | .198           | .133                 | 1.619 | .108 |
| D2     | .642                      | .250           | .194                 | 2.564 | .012 |
| YC     | -.021                     | .011           | -.197                | -1.796 | .075 |
| SFI    | -.120                     | .156           | -.082                | -.770 | .443 |
| NL     | .054                      | .025           | .133                 | 2.139 | .035 |
| MS     | -.206                     | .078           | -.177                | -2.639 | .009 |
| RI     | -.415                     | .075           | -.328                | -5.546 | .000 |
| DRH    | -.425                     | .090           | -.306                | -4.707 | .000 |
| SG     | .782                      | .187           | .252                 | 4.188 | .000 |

A primary regression model is got as follow:

\[ Y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \beta_8 X_8 \]

3.2 Regression Coefficients and Analysis

Firstly, an analysis of type of structure is conducted. When D1=0, D2=0, the value of Y is the smallest when the value of other variables are unchanged, which means that frame structure cause to the least damage to house. When D1=1, D2=0, the value of Y is 0.321(0.642-0.321) more than that when D1=0 and D2=1. Therefore, conclusion could be reached that frame structure’s capacity of resisting earthquake is the best, followed by brick-concrete structure and bottom-frame structure, which is in consistence with the fact in theory and practice.

For independent variable “YC” and “SFI”, we could observe that the earlier the house was constructed and the higher SFI of a house, the slighter the damage of a house was. Similarly, the better the maintenance status was and the higher the RI and DRH was, the lighter the damage was; the higher the SG is (the first is the best), the heavier the damage is; the more the layers are, the heavier the damage is. Those deduction is correspond with common sense, therefore this model is authentic in experience.

3.3 Testing of The Model

First of all, goodness of fit is tested according to R^2. R^2 of this model is 0.655>0.5, suggesting a good fitness. Seen in table 4. In spite of this, the R^2 is far from 0.9 and indicates that some of the
independent variables are sightly related to Y, so a T-test is conducted next.

**Table 4 Information of The Model**

|          | R   | $R^2$ | Adjusted $R^2$ | Error |
|----------|-----|-------|---------------|-------|
|          | .809| .655  | .628          | .465  |

The last volume of table 3 shows that when the significance level is 0.05, significance of variable “D1”, “YC” and “SFI” is higher than the level. That is, these variables in the model may be collinear. The reality implies the impossibility of collinearity between structure type(D1 and D2) and other variables. The reason why significance of D1 is lower than 0.05 may be the shortage of the data. Collinearity does exist between year of construction and seismic fortification intensity in reality because SFI is 5 before 1993 and 6 after that. Under this condition, this study still chose variable “YC” and “SFI” to explore the impact of year of construction on the buildings’ destruction when the SFI is the same.

Next, colinearity is analyzed again by the value of tolerance and VIF(Variance Inflation Factor). Seen in table 5.

**Table 5 Statistics of Model’s Colinearity**

| Statistics | D1 | D2 | YC | SFI | NL | MS | RI | DRH | SG |
|------------|----|----|----|-----|----|----|----|-----|----|
| Tolerancea | .442 | .521 | .250 | .267 | .780 | .664 | .856 | .709 | .826 |
| VIFb       | 2.263 | 1.918 | 4.007 | 3.741 | 1.282 | 1.506 | 1.168 | 1.410 | 1.210 |

a: The value of tolerance is between 0 and 1. The closer to 1 the tolerance is, the less colinear the variable is.
b: The value of VIF is between 1 and 10. The closer to 1 the tolerance is, the less colinear the variable is.

Table 5 demonstrates that all variables’ VIF is close to 1, except variable “YC” and “SFI”. That is, significant colinearity exists between the two variables. This may because the scale of sample is small and most of the value of Y is “2” or “3”, being unable to reflect the impact of YC on buildings’ destruction, which leads to a significant correlation between YC and SFI.

### 3.4 Modification of The Model

The result of tests above suggest a strong correlation between YC and SFI. Considering that SFI is a crucial factor affecting damage, variable “YC” is deleted in the modification.

After the modification, a series of tests are conducted the same as before. All the variables’ VIF are close to 1. Significance of all variables are less than 0.05, excluding “NL”. On the condition that the number of houses’ layer is less than 7 in the study, variable “NL” has no direct relation with the damage of house[7]. Therefore, “NL” is deleted in the following factor analysis. And, regression function after deleting “YC” and “NL” is shown below.

$$Y = 6.633 + 0.349D_1 + 0.671D_2 - 0.350X_3 - 0.186X_5 - 0.434X_6 - 0.429X_7 + 0.606X_8$$

Normalized function is shown next.

$$Y = 0.145D_1 + 0.203D_2 - 0.238X_3 - 0.161X_5 - 0.343X_6 - 0.309X_7 + 0.195X_8$$

### 4. Factor Analysis

Factors should reflect sample information comprehensively, so all variables except “ST” and “NL” are
qualified to factor analysis. The reason why “ST” is excluded is that “ST” is a classified variable, which makes no sense to be used to conduct factor analysis. Given the fact that it is an important variable to affect houses’ damages, we take it as an individual factor in later analysis. “NL” has no direct relationship with the damage of house according to analysis above, so it is excluded.

4.1 Feasibility Testing

At the beginning, correlation test is conducted and correlation coefficient matrix is shown in table 6.

|       | SFI   | MS    | RI    | DRH   | SG    | YC    |
|-------|-------|-------|-------|-------|-------|-------|
| SFI   | 1.000 | .413  | .274  | .264  | -.126 | .836  |
| MS    | .413  | 1.000 | .320  | .390  | -.184 | .418  |
| RI    | .274  | .320  | 1.000 | .210  | -.105 | .269  |
| DRH   | .264  | .390  | .210  | 1.000 | -.119 | .288  |
| SG    | -.126 | -.184 | -.105 | -.119 | 1.000 | -.022 |
| YC    | .836  | .418  | .269  | .288  | -.022 | 1.000 |

The coefficient is lower than 0.5 excluding coefficient between “YC” and “SFI”, suggesting a slight correlation of the variables.

Next, test of KMO and Bartlett is conducted, as shown in table 7.

|       | KMO   | Bartlett sphericity test |
|-------|-------|--------------------------|
|       | .661  | Approximate Chi-square   |
|       |       | df                      |
|       |       | Sig.                    |
|       |       | 220.026                 |
|       |       | 15                      |
|       |       | .000                    |

The value of KMO is higher than 0.5 and significance of Bartlett test is 0.000, all indicating a feasibility of factor analysis.

4.2 Identification of Factor

First of all, by observing the sample scree plot (Seen in fig.1, the horizontal ordinate represents the number of factors, the longitudinal ordinate represents the factors’ characteristic value. It is a plot reflecting to which degree the factor could explain the variables), the number of factors could be determined. Normally, factors whose characteristic value is higher than 1 will be retained. But according to research of Nyaradzo H.Mvududu and Christopher A.Sink. The point at which the line begins to show a clear bend indicates the actual number of factors that should be retained[8]. Therefore, three factors could be extracted in the study.
Fig. 1 Sample Scree Plot

Next, total variance explained by the former three factors is 74.819% and higher than 0.7, suggesting a good reflecting of all factors. Seen in table 8.

| Component | Extraction Sums of Squared Loadings | Total Variance Explained: Cumulative % |
|-----------|------------------------------------|----------------------------------------|
|           | Eigenvalue | % of Variance |                           |
| 1         | 2.571      | 42.847        | 42.847                    |
| 2         | 1.061      | 17.688        | 60.536                    |
| 3         | .857       | 14.284        | **74.819**                |
| 4         | .796       | 13.261        | 88.081                    |
| 5         | .559       | 9.311         | 97.391                    |
| 6         | .157       | 2.609         | 100.000                   |

In the end, the six variables are classified into three factors according to their factor loading to every factor (That is, component in table 9).

| Component     | Rotated Component Matrix |
|---------------|--------------------------|
|               | Component 1 | Component 2 | Component 3 |
| SFI           | .924        | .234        | -.072       |
| YC            | .919        | .260        | .055        |
| MS            | .059        | .771        | -.012       |
| DRH           | .096        | .687        | .024        |
| RI            | .330        | .678        | -.163       |
| SG            | -.029       | -.121       | **.986**    |

From table 9, the first component has a bigger loading on variable “SFI” and “YC”, so it could be named as “Time Factor”; the second component has a bigger loading on variable “MS”, “DRH”, “RI”,...
and it could be named as “House status Factor”; the third component has a bigger loading on variable “SG” and be named using variable name “Site grade Factor”.

5. Conclusion
According to the normalized regression function got in the second part in the paper, conclusion could be reached that the rank of degree of variables affecting houses’ damage is “Roof Integrity”, “Degree of Regularity of House”, “Seismic Fortification Intensity”, “Site Grade”, “Maintenance Status”. As for “Structure Type”, frame structure causes the least damages to house, followed by brick-concrete structure. Bottom-frame structure causes the biggest damages, so this type of structure should be avoided especially in the town in the future.

In the process of factor analysis, there factors are extracted: “Time Factor”, “House status Factor”, “Site grade Factor”. Cumulative variance explained by the there factors is up to 74.819%, suggesting a comprehensive reflection of variables studied. The outcome of factor analysis could be used in damage prediction and study in the future.

The research has limitations. Restricted by difficulties in data collection and the need for integrity of data, samples in the study are not enough and limited in urban district of Mianyang, so most of the degrees of damage are “slight” or “moderate”, which fails to represent all the degrees of damage greatly.

6. Acknowledgments
This research is supported by Safety Appraisal Reports, which are the outcome of teachers’ hard-working. We thank their work. We also appreciate the anonymous reviewers for taking the effort and time to review an earlier version of this paper, which will enhance it significantly.

Reference
[1] Young-Ji P. Alfredo H.S 1985 Seismic damage analysis of reinforced concrete buildings J. Struct. Eng. 111 pp 740-757
[2] Young-Ji P. Alfredo H.S. 1985 Mechanistic seismic damage model for reinforced concrete J. Struct. Eng. 111 pp 722-739
[3] Peter F. Peter G. 1996 The N2 method for the seismic damage analysis of reinforced concrete buildings Earthquake Eng. and Struc. 25 pp 31-46
[4] Rob S. 2000 Unacceptable floor flexibility resulting from earthquake damage Forensic Eng. pp 200-208
[5] Brendon A. Bradley. 2010 Epistemic uncertainties in component fragility functions Earthquake Spectra. 26 pp 41-62
[6] Guoqing L. Qingshan Y. 2009 Detailed investigation and analysis of damage of structures in Dujiangyan city in Wenchuan Earthquake J. Bldg. Struc. 39 pp227-231(in Chinese)
[7] Nyaradzo H. Christopher A. 2013 Factor analysis in counseling research and practice CORE(Counseling Outcome Research. and Evaluation). 4.2 pp 75-98