A real estate data based online survey method on the live load change of residential buildings

Chi Xu
College of Civil Engineering, Tongji University

Jun Chen
College of Civil Engineering, Tongji University

Jie Li (Corresponding author)
College of Civil Engineering, Tongji University

ABSTRACT: The change interval is a key factor for modelling the live load, which is a basis for the rational structural design and analyses. Being a random variable, the change interval requires a large number of samples for its statistical study. However, due to the lack of a survey method combining the feasibility, accuracy and efficiency, the available interval samples are considerably limited. In this study, a new online survey method for collecting the change interval samples of residential buildings is proposed. The new method decomposes the load change interval into two independent variables: the occupied period and sales period, whose samples can be easily obtained from the well-developed and open-access data of real state websites. Statistical properties of the load change interval can then be derived using these samples. The new method is applied to a live load survey in Shanghai, with more than 30000 samples being collected. Lifetime maximum live loads are derived from the survey results. The survey process demonstrates the feasibility of the new method, which is compatible with the current Big Data era, high efficient, low cost and disturbance free.

1. INTRODUCTION

The live load of structures is mainly induced by human activities (Li et al. 1990). Based on behaviors of the duration and occurrence rate, the live load of building structures can be decomposed into the sustained load and extraordinary load (Peir 1972). The sustained load is of interest in this study. For simplicity, the sustained live load is designated live load hereafter.

The live load is typically modelled as a Poisson square process (Peir 1972), thereby two random variables are required to describe the live load: arbitrary point-in-time load and load change intervals, as shown in Figure 1. The main objectives of live load surveys are the probability statistics of the above two random variables, with the load change interval being focused in this study.

In 1960s, the temporal change of the live load was first considered using a simple model in which a fixed number of changes were assumed during the structure lifetime (Karman 1965, Karman 1969). In 1970s, the process of load change occurrences is assumed to be a Poisson process (Peir 1972). Since then, the load change rate has become an essential parameter for the probabilistic live load description.

The background of the live load change is the occupant change. Genarally, it is not feasible to carry out decades of observations on the occupant change process, thereby four categories of solutions have been proposed in historical surveys, which includes: (1) Occupant oriented sampling (Niu J and Niu D 2006, Ge et al. 2008): The occupants are interviewed for their previous moving experience and the period during which the respondent lives in the same place is considered as a sample. (2) Archive investigation: Some available archives, such as telephone directories (Mitchell and Woodgate 1971) and official annual housing survey (Harris et al. 1981), are utilized to obtain the interval samples. (3) Subjective inference: In some surveys, the statistics are simply determined by personal experience or engineering judgement (Ruiz and Soriano 1997). (4) Using substitutions: The samples of load change intervals are substituted by the samples of the occupancy duration, defined from the time that the present occupant moved in to the survey implementation moment (Kumar 2002a, Kumar 2002b, Andam 1986).

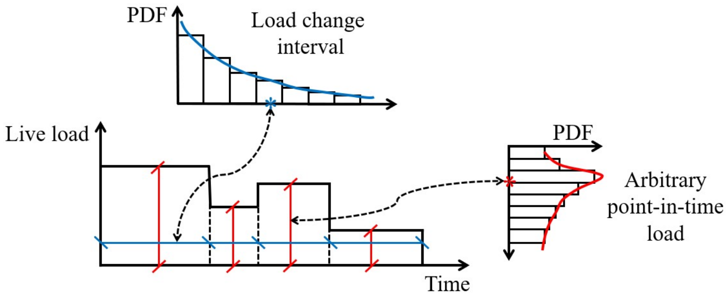

Figure 1 Two random variables required to describe a typical live load process.

In Table 1, the above four types of solutions are compared in feasibility, efficiency and accuracy. Due to the lack of a survey method combining all the above aspects, the data accumulation of load change intervals is slow throughout the last decades, especially for residential buildings.

Table 1 Comparisons between the four traditional survey methods

Survey methods	Evaluation factor		
	Feasibility	Efficiency	Accuracy
Occupant oriented sampling	High	Low	Middle
Archive investigation	Low	Middle	High
Subjective inference	High	High	Low
Using substitutions	High	Low	Low

The arrival of the Big Data era brings new hope to traditional problems of the live load research. Based on the unlimited property transaction data, a new online survey method on the live load change of residential buildings is proposed.

The remainders of the paper are organized as follows. In Section 2, the probabilistic model of the live load is introduced. In Section 3, the fundamentals and implementation process of the new survey method is proposed. Subsequently, in Section 4 the new method is applied to a survey on live load change intervals. More than 30000 samples of Shanghai are acquired from the open-access data on real estate websites. The lifetime maximum values of the live load are computed according to the survey results.

2. PROBABILISTIC LIVE LOAD MODEL

As shown in Figure 1, the live load remains constant until a load change occurs. In addition, the live load amplitudes at different intervals are assumed to be independent with each other.

The number of the independent, identically distributed live load amplitudes are designated n. The cumulative distribution function (CDF) of the lifetime maximum load, $F_{Lm}(l)$, is expressed as (Harris et al. 1981):

$$F_{Lm}(l)=\sum_{k=1}^{\infty}\left[F_L(l)\right]^k P(n=k) \tag{1}$$

where $F_L(l)$ is the CDF of the arbitrary point-in-time live load L. $P(n=k)$ represents the probability that there are k independent live load amplitudes during the lifetime T.

In this study, L is assumed to follow the Gamma distribution (JCSS 2001). For a considered area A, the mean of L is provided by (JCSS 2001):

$$E(L)=m \tag{2}$$

The variance of L is expressed as the following (JCSS 2001):

$$D(L)=\sigma_V{}^2+\sigma_U{}^2\frac{A_0}{A}\kappa \tag{3}$$

in which κ is determined by the influence surface function over A and is taken as 2.0 for simplicity. A_0 is a reference area. When $A<A_0$, A_0/A equals to 1 (JCSS 2001). Moreover, the above three parameters m,σ_V and σ_U are derived from statistical results.

According to the Poisson process assumption (Peir 1972), $P(n=k)$ is calculated by:

$$P(n=k)=\exp(-\lambda T)\frac{(\lambda T)^{k-1}}{(k-1)!} \tag{4}$$

λ is the mean change rate of the live load, which is typically estimated by the inverse of the mean change interval.

Substituting Eq. (4) into Eq. (1), after some derivations the CDF of the lifetime maximum load is provided by (Harris et al. 1981):

$$F_{Lm}(l)=F_L(l)\exp\left\{-\lambda T\left[1-F_L(l)\right]\right\} \tag{5}$$

The probability density function (PDF) of the lifetime maximum load is:

$$f_{Lm}(l)=f_L(l)\left[1+\lambda T F_L(l)\right]\exp\left\{-\lambda T\left[1-F_L(l)\right]\right\} \tag{6}$$

3. NEW SURVEY METHOD

Based on the theoretical analyses in Section 2, change interval samples are essential for estimating the mean change rate λ. With the development of Big Data, unlimited electronic records of the real estate transaction are concentrated on the public websites, based on which a new survey method is proposed.

3.1. *Basic ideas*

The data concentration and survey implementation processes are illustrated in Figure 2. For residential buildings, the background of the live load change is the change of homeowners. A typical process of homeowner changes in a residence is illustrated in Figure 2. Under the new survey method, the live load change interval, denoted as τ, is decomposed into two periods:

$$\tau=\tau_1+\tau_2 \tag{7}$$

in which τ_1 is the occupied period and τ_2 is the sales period. The occupied period is the length between the residence being occupied and being listed for sale by a homeowner. The sales period is the length between the residence being listed for sale and finally being sold by a homeowner. It is noted that τ_1 and τ_2 are both random variables and are assumed to be independent with each other.

Through observations on Figure 2, the live load change interval is exactly the sum of τ_1 and τ_2.

3.2. *Data concentration process*

Owing to the guiding roles for the real estate marketplace, the data of τ_1 and τ_2 are extensively collected by real estate agencies.

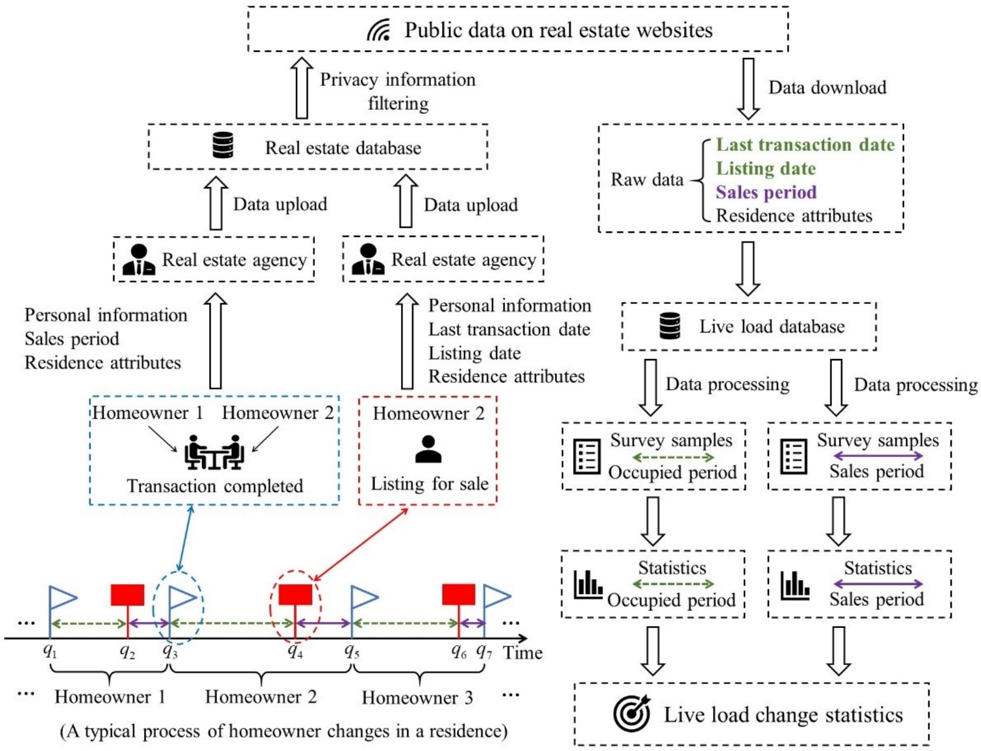

Figure 2 The transaction data concentration and survey implementation process.
(Icons in the dotted frames are provided by the Microsoft Office 2019)

In Figure 2, the data acquisition procedure related to homeowner 2 is detailed for illustration. At q_3 when the transaction between homeowner 1 and homeowner 2 is completed, the data including the personal information, sales period $(q_3 - q_2)$ and residence attributes (such as the residence area, floor location et al.) are recorded by the agency member involving in the transaction coordination.

At q_4, the residence is listed for sale and homeowner 2 is required to provide the agency with the personal information, the last transaction date q_3, the listing date q_4 and residence attributes.

Subsequently, the collected data are organized into the required format and uploaded to the real estate database. The database is typically operated and maintained by large real estate agencies. Then after the filtering of privacy information, the remainders are publicly presented on the real estate website.

3.3. *Survey implementation process*

For live load investigators, four types of raw data require to be obtained from real estate websites and stored in the live load database, as shown in Figure 2. The samples of τ_2 are directly obtained online. Moreover, based on each pair of the last transaction date and the listing date (For example, q_3 and q_4), a sample of τ_1 is acquired. Furthermore, the residence attributes are stored in the database for future study of their impacts on the live load change.

Limited by the data organization of websites, it is infeasible to obtain the samples of τ_1 and τ_2 in pairs. Residences displayed on real estate websites are twofold: on sale and sold, from which the samples of τ_1 and τ_2 are respectively obtained.

As shown in Figure 2, the raw data are directly collected online, thereby no walk-in investigation (such as face-to-face interviews about moving experience) is required. Moreover, aided by the computer technology, the well organized online data is easy to be automatically and rapidly obtained.

4. APPLICATIONS

The new method is applied to an actual survey on the live load change of Shanghai, with more than 30000 samples acquired. Raw data are obtained from the open-access data on Lianjia website (available from: https://sh.lianjia.com/).

With different properties in geographical location, local history, economic development and so forth, 9 districts in Shanghai are involved in the survey. For each district, the sample sizes of τ_1 and τ_2 are listed in Table 2.

Table 2 The sample sizes of the 9 districts in Shanghai

District name	Baoshan	Hongkou	Jingan	Minhang	Putuo
τ_1	2003	476	1035	2157	1021
τ_2	2993	2986	2988	2984	2987
District name	Xuhui	Yangpu	Fengxian	Pudong	(Total)
τ_1	1059	948	525	1977	(11201)
τ_2	2987	2974	2994	2982	(26875)

Based on the Poisson process assumption, the mean of the load change interval is required. Obviously, there exists:

$$E(\tau) = E(\tau_1) + E(\tau_2) \tag{8}$$

The means of τ_1 and τ_2 of the 9 districts are shown in Figure 3 and Figure 4 respectively.

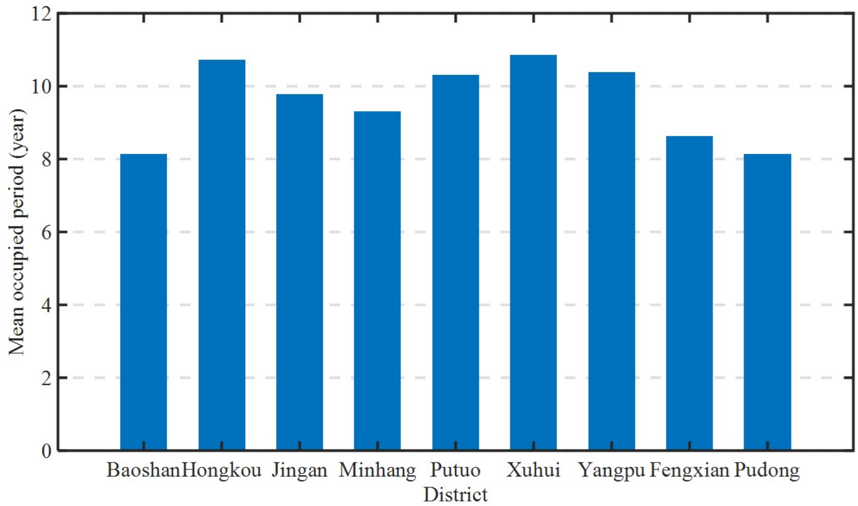

Figure 3 The mean of τ_1 of the 9 districts in Shanghai.

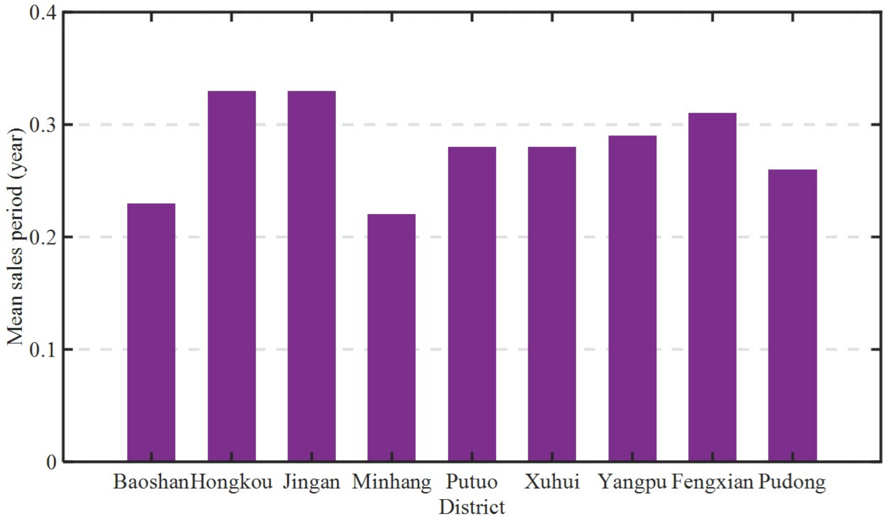

Figure 4 The mean of τ_2 of the 9 districts in Shanghai.

Comparing Figure 3 and Figure 4, the means of τ_2 are one order magnitude smaller than those of τ_1. Based on the engineering background, a sample of τ_1, covering most of the length when a homeowner lives in the residence, is typically in years. However, a sample of τ_2, representing the time cost of waiting for residence buyers and transaction preparation, is generally in weeks. Therefore, it is reasonably assumed that:

$$\tau = \tau_1 \tag{9}$$

which further simplifies the survey process of the live load change.

Considering the statistical results of τ_1, the ratio between the minimum and the maximum of the mean is 75%, suggesting that the live load change statistics do not vary significantly within the same city. More than the data support, the determinants of real estate transactions are considerably similar within the same city. Even more significantly, excessively detailed region related statistics are not practical for engineering applications, thereby furture surveys can be implemented on the city level.

The mean change interval of Shanghai is 9.29 years. For residential buildings, m, σ_V and σ_U are taken as 0.3, 0.15 and 0.3 $\mathrm{KN/m^2}$ respectively (JCSS 2001). A_0 equals to 20 m^2 (JCSS 2001) and T is specified as 50 years. Subsequently, PDFs of the lifetime maximum live loads of some representative areas are derived and shown in Figure 5.

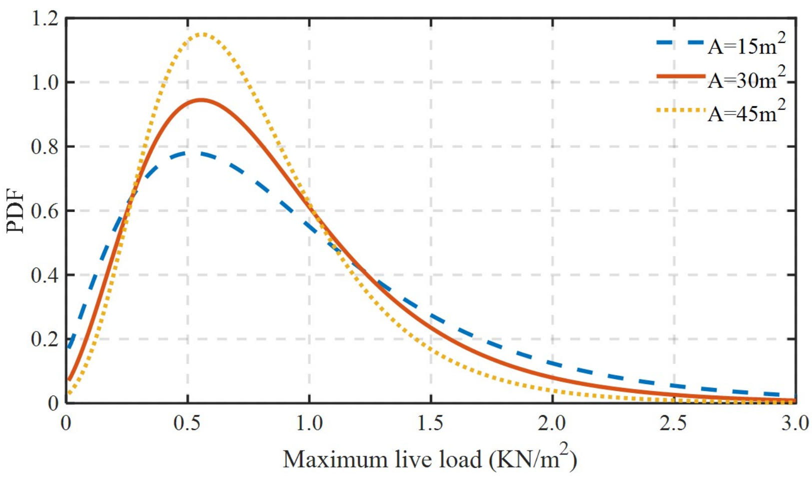

Figure 5 PDFs of the lifetime maximum live loads of representative areas.

5. CONCLUSIONS

(1) A new online survey method on the live load change is proposed to overcome the problems of traditional ones. The new method decomposes the load change interval into the occupied period and sales period, the samples of both two can be easily acquired from real estate websites.

(2) Using the new method, a survey on the live load change is implemented in Shanghai, with a total of more than 30000 samples being collected. The present sample size is nearly two orders of magnitude larger than previous surveys on residential buildings. Even more significantly, the survey was conducted merely by one investigator in his office demonstrating the higher efficiency and lower cost of the proposed method.

(3) The mean change interval is 9.29 years in Shanghai. The influence of the sales period is negligible, thereby the survey process can be further simplified. Based on the survey results, related surveys in the future can be carried out on the city level.

ACKNOWLEDGEMENTS

The supports of the key project of National Natural Science Foundation of China (No. 51538010) and key project of Guangdong Big Data Center Joint Foundation (No. U1711264) are gratefully appreciated.

REFERENCES

Andam, K.A. 1986. Floor live loads for office buildings. *Building and Environment*, 21(3/4): 211-219.

Ge, S.J., Chen, H., Sun, Z.S. and Li, J.B. 2008. Survey and statistics of floor live load of residential building in Central plains region. *Building Structure*, 38(7):88, 125-128(in Chinese).

Harris, M.E., Corotis, R.B. and Bova, C.J. 1981. Area-dependent processes for structural live loads. *Journal of the Structural division (ASCE)*, 107(ST5):857-872.

Joint Committee on structural safety. 2001. Probabilistic model code: Part 2: load models.

Kumar, S. 2002. Live loads in office buildings: point-in-time load intensity. *Building and Environment*, 37:79-89.

Kumar S. 2002. Live loads in office buildings: lifetime maximum load. *Building and Environment*, 37:91-99.

Karman, T. 1965. Investigations of the useful loads of floor structure of dwellings translated from the scientific reports of the institute of construction science, No. 51, Budapest, Hungary.

Karman, T. 1969. Statistical investigations on live loads on floors. International council for building research, studies and documentation, Madrid, Spain.

Li, J.H., Lin, Z.M., Li, M.S., Ma, K.Z., Shao, Z.M., Chen, J.F. et al. 1990. Design of probabilistic limit state of building structures. Beijing: China Architecture & Building Press (in Chinese).

Mitchell, G.R. and Woodgate, R.W. 1971. Floor loadings in office buildings – the results of a survey. CP3/71, Building research station, Garston, U.K.

Niu, J.G. and Niu, D.T. 2006. Investigation and statistical analysis on the floor loads of residential buildings. *Journal of Xi'an University of Architecture & Technology (Natural Science Edition)*, 38(2):214-220 (in Chinese).

Peir, J.C. 1972. A stochastic live load model for buildings (PH.D. thesis). America: Massachusetts institute of technology, MA.

Ruiz, S.E. and Soriano, A. 1997. Design live loads for office buildings in Mexico and the United States. *Journal of Structural Engineering*, 123(6): 816-822.