

Research Article

Linear Regression Estimation Methods for Inferring Standard Values of Snow Load in Small Sample Situations

Xudong Wang ¹ and Jitao Yao ²

¹School of Science, Xi'an University of Architecture and Technology, Xi'an 710055, China

²School of Civil Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

Correspondence should be addressed to Jitao Yao; 32782472@qq.com

Received 21 May 2020; Revised 24 June 2020; Accepted 25 June 2020; Published 27 July 2020

Academic Editor: Zeshui Xu

Copyright © 2020 Xudong Wang and Jitao Yao. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The aim of this paper is to establish a new method for inferring standard values of snow load in small sample situations. Due to the incomplete meteorological data in some areas, it is often necessary to infer the standard values of snow load in the conditions of small samples in engineering, but the point estimation methods of classical statistics adopted till now do not take into account the influences of statistical uncertainty, and the inference results are always aggressive. In order to overcome the above shortcomings, according to the basic principle of optimal linear unbiased estimation and invariant estimation of the minimum type I distribution parameters and the tantile, using the least square method, the linear regression estimation methods for inferring standard values of snow load in small sample situations are proposed, which can take into account two cases such as parameter-free and known coefficient of variation, and the predicted formulas of snow load standard values are given, respectively. Through numerical integration and Monte Carlo numerical simulation, the numerical table of correlation coefficients is established, which is more convenient for the direct application of inferential formulas. According to the results of theoretical analysis and examples, when using the indirect point estimation methods to infer the standard values of snow load in the conditions of small samples, the inference results are always small. The linear regression estimation method is suitable for inferring standard values of snow load in the conditions of small samples, which can give more reasonable results. When using the linear regression estimation to infer standard values of snow load in practical application, even if the coefficient of variation is unknown, it can set the upper limit value of the coefficient of variation according to the experience; meanwhile, according to the parameter-free and known coefficient of variation, the estimation is carried out, respectively, and the smaller value of the two is taken as the final estimate. The method can be extended to the statistical inference of variable load standard values such as wind load and floor load.

1. Introduction

Snow load is one of the main loads of buildings, and the inference of its standard value is the basis for the establishment of structural design and evaluation methods. At present, the inference of the standard values of snow load is generally fitting with the maximum type I distribution for the maximum annual snow pressure [1]; then, using the relation between the representative values and the distribution parameters, an estimate of the distribution parameters is given under a certain guaranteed rate. Among them, the estimation method of distribution parameters includes the moment method and the maximum likelihood method

[2–13]. These are point estimation methods. They are mainly suitable for large samples. They require sufficient meteorological data as statistical samples. It is generally believed that at least 30 years of data are required [14, 15]. However, in engineering practice, it is sometimes necessary to infer the standard values of snow load under the condition that the test data are insufficient, and the actual sample capacity is often very limited. Statistical analysis is done mostly in the case of small sample capacity. This results in a significant reduction in parameter estimation accuracy when the sample capacity is small. In the case of such a small sample, if the current point estimation method is still used for inference, the result of inference is often reduced due to the

influence of statistical uncertainty. A more reasonable choice is to use a small sample inference method [16–19].

For dead loads, the standard for appraisal of reliability of civil buildings has proposed a small sample method for inferring its standard value [20]; however, the snow load usually obeys the maximum type I distribution, which is different from the probability distribution form of the dead load. Therefore, the same method cannot be used to infer the standard values of snow load. A method for estimating the maximum type I distribution parameters and tantile is proposed in the paper [21], which lays a theoretical foundation for the small sample inference of the standard values of snow load without parameter information. Therefore, the inference formula of the standard values of snow load without parameter information can be established. However, in some cases, the variable coefficients of the probability distribution of snow loads are known, or their upper limit values can be set based on experience. Using the additional information, the uncertainty of the statistics in the estimation process can be significantly reduced, and more favorable extrapolation results can be obtained under the same conditions. Therefore, it is necessary to study the inference method of the standard values of snow load when the coefficient of variation is known under the condition of small samples, to provide a better choice for the small sample inference of the standard values of snow load.

Based on the above analysis, the probability distribution model of snow load is established first in this paper because the standard values of snow load usually are expressed as a tantile of the distribution, type I maximum distribution, and type I minimum distribution which belong to the same extreme value distribution families and can be converted with each other; therefore, by using the least square method based on the best linear unbiased estimation and the invariant estimation principle of the current minimum type I distribution parameters, a linear regression estimation method for standard snow load under small sample conditions is proposed [22–31].

The inference formula of standard values of snow load is given in this paper, and no parameter information and the known coefficient of variation are both considered. Through numerical integration and Monte Carlo numerical simulation, a numerical table of correlation coefficient is established to tantile the application of the inference formula, and the conclusions and suggestions are given by comparing the results with the traditional large sample method. The method can be extended to the statistical inference of variable load standard values such as wind load and floor load.

2. Probability Distribution Model of Snow Load

A probability distribution model of snow load is established by using a stationary binomial random process [32].

It is assumed that the design reference period of the building structure is T , and it can be divided into r equal period of time in $[0, T]$; the average time for the snow load to change once is $\tau = T/r$; the probability of action at each time is p ; the probability distribution function appearing on different time segments is $F_Q(x)$, and the random variables x on different time segments are independent of each other. The establishment of the snow load model requires the

identification of the above three key parameters. The Unified Standard for Reliability Design of Building Structures uses a limit state design method based on probability theory, that is, a first-order second-moment method that considers the probability distribution type of basic variables. Since the basic variables are considered as random variables, the random process of the load must be converted into a random variable [33]. If random variables at any point time are used instead of random processes, it will be unsafe, so the maximum load random variables Q_T ($Q_T = \max Q_t, 0 \leq t \leq T$) appearing in the design reference period should be used instead of random processes for statistical analysis. The basic steps to convert the random process to the maximum load within T years are as follows:

Step 1: establish the load probability distribution function at any period time τ :

$$F_i(x) = P[Q(t) \leq x, t \in T] = p \cdot F_i(x) + 1 - p = 1 - p[1 - F_i(x)], \quad (1)$$

where $F_i(x)$ is the load distribution function at i time point.

Step 2: establish the probability distribution function Q_T :

$$\begin{aligned} F_T(x) &= P[Q_T \leq x] = P\left[\max_{0 \leq t \leq T} Q(t) \leq x, t \in T\right] \\ &= \prod_{j=1}^r P[Q(t_j) \leq x, t_j \in t] \\ &= \prod_{j=1}^r \{1 - p[1 - F_i(x)]\} = \{1 - p[1 - F_i(x)]\}^r. \end{aligned} \quad (2)$$

Assuming that the average number m is the snow load which occurs in T years, then

$$m = pr. \quad (3)$$

Obviously, when $p = 1$ and $m = r$, then,

$$F_T(x) = [F_i(x)]^m. \quad (4)$$

When $p < 1$, using the approximate relationship,

$$e^{-x} \approx 1 - x. \quad (5)$$

If $p[1 - F_i(x)]$ is sufficiently small in (2), then

$$F_T(x) \approx \left\{e^{-p[1 - F_i(x)]}\right\}^r = \left\{e^{-[1 - F_i(x)]}\right\}^{pr} \approx \{1 - [1 - F_i(x)]\}^{pr}. \quad (6)$$

Hence,

$$F_T(x) \approx [F_i(x)]^m. \quad (7)$$

It can be seen from the above equation that the probability distribution function $F_T(x)$ of the maximum load Q_T in the design reference period is equal to

the m -power of the probability distribution function $F_i(x)$ of the load at any time, and $F_i(x)$ can be obtained according to statistics [34].

Step 3: the probability distribution of snow load is fitting with the maximum type I distribution. The distribution function and probability density functions are, respectively, as follows:

$$F(x) = e^{-e^{-((x-\mu)/\alpha)}}, \quad (8)$$

$$f_X(x) = \frac{1}{\alpha} e^{-((x-\mu)/\alpha)} \exp\{-e^{-((x-\mu)/\alpha)}\},$$

where α and μ are the scale parameters and position parameters of the distribution.

3. Linear Regression Estimation Method for Standard Values of Snow Load

3.1. In the Condition of Unknown Parameter Information. The value at any time point of snow load obeys the maximum type I distribution, and the probability density function is

$$f_X(x) = \frac{1}{\alpha} e^{-((x-\mu)/\alpha)} \exp\{-e^{-((x-\mu)/\alpha)}\}, \quad (9)$$

where μ and α are distributed parameters, $-\infty < \mu < \infty$ and $0 < \alpha < \infty$. The standard values of snow load usually are expressed as down tantes with p calibration of the random variable X , and they can be written as x_k ; it meets that

$$P\{X \leq x_k\} = \exp\{-e^{-((x_k-\mu)/\alpha)}\} = p, \quad (10)$$

$$x_k = \mu - \alpha \ln(-\ln p) = \mu + k\alpha,$$

where p is a guaranteed rate of the standard value x_k .

It is assumed that n samples of X are arranged from a small to large order: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, and the test values are $x_{(1)}, x_{(2)}, \dots, x_{(n)}$, respectively.

Let

$$Y' = -X. \quad (11)$$

Then, Y' obeys the minimum type I distribution with two parameters $-\mu$ and α , and the order statistic and up tantes with p calibration are

$$Y'_{(j)} = -X_{(n-j+1)}, \quad j = 1, 2, \dots, n, \quad (12)$$

$$y'_p = -\mu - k\alpha = -x_k.$$

For obtaining x_k which is the characteristic value of variable actions, we usually select the upper limit estimated value as the inferring results; when the load effect is favorable for the structure, the lower limit estimated value should be selected, but it has been rarely found in the most unfavorable combinations of the action effect [35]. We can use the upper and lower limit estimated values of y'_p to infer the upper and lower limit estimated values of x_k because the estimate values

$$-\tilde{\mu} = \sum_{j=1}^n D_I(n, n, j) [-x_{(n-j+1)}], \quad (13)$$

$$\tilde{\alpha} = \sum_{j=1}^n C_I(n, n, j) [-x_{(n-j+1)}], \quad (14)$$

$$y_{p,U}' = -\tilde{\mu} - v_{p,1-C}\tilde{\alpha}, \quad (15)$$

$$y_{p,L}' = -\tilde{\mu} - v_{p,C}\tilde{\alpha}. \quad (16)$$

And, for the random variables $-\tilde{\mu}$ and $\tilde{\alpha}$, we have,

$$P\left\{\frac{-\tilde{\mu} - y_p}{\tilde{\alpha}} \leq v_{p,C}\right\} = P\{y_p \geq -(\tilde{\mu} + v_{p,C}\tilde{\alpha})\} = P\{-y_p \leq \tilde{\mu} + v_{p,C}\tilde{\alpha}\} = P\{x_k \leq \tilde{\mu} + v_{p,C}\tilde{\alpha}\}, \quad (17)$$

$$P\left\{\frac{-\tilde{\mu} - y_p}{\tilde{\alpha}} \geq v_{p,1-C}\right\} = P\{y_p \leq -(\tilde{\mu} + v_{p,1-C}\tilde{\alpha})\} = P\{-y_p \geq \tilde{\mu} + v_{p,1-C}\tilde{\alpha}\} = P\{x_k \geq \tilde{\mu} + v_{p,1-C}\tilde{\alpha}\}.$$

Thus, the upper and lower limit estimated values of x_k , which are the characteristic values of the variable actions, are

$$x_{k,U} = -y_{p,L}' = \tilde{\mu} + v_{p,C}\tilde{\alpha} = \sum_{j=1}^n D_I(n, n, j)x_{(n-j+1)} + v_{p,C} \sum_{j=1}^n [-C_I(n, n, j)x_{(n-j+1)}], \quad (18)$$

$$x_{k,L} = -y_{p,U}' = \tilde{\mu} + v_{p,1-C}\tilde{\alpha} = \sum_{j=1}^n D_I(n, n, j)x_{(n-j+1)} + v_{p,1-C} \sum_{j=1}^n [-C_I(n, n, j)x_{(n-j+1)}], \quad (19)$$

where $D_I(n, n, j)$ and $C_I(n, n, j)$ can be directly determined from the numerical table. Because $v_{p,C}$ and $v_{p,1-C}$ are present, numerical tables only give the numerical values when $p = 0.90, 0.95, \text{ and } 0.99$, so the values of $v_{p,C}$ and $v_{p,1-C}$ usually cannot be directly determined from the numerical table [36]. Therefore, we propose a new approximate method which uses the present numerical table to infer the values of $v_{p,C}$ and $v_{p,1-C}$ in this paper.

We use p_0 to denote the guarantee rate when the present numerical tables are adopted, and down tantes with p_0 calibration of the random variable X are expressed as x_{k0} , so we have

$$x_{k0} = \mu - \alpha \ln(-\ln p_0),$$

$$x_k = \mu + \frac{\ln(-\ln p)}{\ln(-\ln p_0)} (x_{k0} - \mu). \quad (20)$$

It is approximate let that

$$x_{k0,U} - \mu \approx x_{k0,U} - \tilde{\mu} = v_{p_0,C} \tilde{\alpha},$$

$$x_{k0,L} - \mu \approx x_{k0,L} - \tilde{\mu} = v_{p_0,1-C} \tilde{\alpha}. \quad (21)$$

Because of the corresponding relationships between $x_{k,U}$ and $x_{k0,U}$, $x_{k,L}$, and $x_{k0,L}$, we can obtain that

$$v_{p,C} = \frac{\ln(-\ln p)}{\ln(-\ln p_0)} v_{p_0,C},$$

$$v_{p,1-C} = \frac{\ln(-\ln p)}{\ln(-\ln p_0)} v_{p_0,1-C}. \quad (22)$$

In practical applications, we should select p_0 which is approximate to p . It is proved by calculation that when $n = 5 - 25$, $p \geq 0.90$, and $C = 0.6 - 0.95$, the margins of errors of $v_{p,C}$ and $v_{p,1-C}$ are -0.017 to 0.017 and -0.031 to 0.036 , respectively. This method is more convenient and accurate than the interpolation method.

3.2. In the Condition of Known Coefficient of Variation

3.2.1. Linear Unbiased Estimation of Distribution Parameters. It is assumed that the value at any time point of snow load is X and the coefficient of variation δ_X of X is known, and the distribution parameters are

$$\alpha = \frac{\sqrt{6}}{\pi} \sigma_X,$$

$$\mu = \mu_X - C_E \alpha = \mu_X \left(1 - \frac{C_E \sqrt{6}}{\pi} \delta_X \right), \quad (23)$$

where μ_X , σ_X , and δ_X are the mean, standard deviation, and coefficient of variation of X , respectively, and C_E is the Euler constant. It is assumed that the samples of X are arranged from the small to large order: $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, that is, $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$. At this time, the probability density function of $X_{(i)}$ and the joint probability density function of $X_{(i)}$ and $X_{(j)}$ ($1 \leq i < j \leq n$) are

$$f_{X_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} F_X(x)^{i-1}$$

$$[1 - F_X(x)]^{n-i} f_X(x), \quad i = 1, 2, \dots, n,$$

$$f_{X_{(i)}, X_{(j)}}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} F_X(x)^{i-1}$$

$$[F_X(y) - F_X(x)]^{j-i-1} [1 - F_X(y)]^{n-j}$$

$$f_X(x) f_X(y), \quad 1 \leq i < j \leq n, x \leq y. \quad (24)$$

Let

$$Z = \frac{X - \mu}{\alpha}. \quad (25)$$

Then, Z obeys the standard maximum type I distribution, and its order statistic is

$$Z_{(i)} = \frac{X_{(i)} - \mu}{\alpha}, \quad i = 1, 2, \dots, n, \quad (26)$$

the probability density function of $Z_{(i)}$ and the joint probability density function of $Z_{(i)}$ and $Z_{(j)}$ ($1 \leq i < j \leq n$) are

$$f_{Z_{(i)}}(x) = \frac{n!}{(i-1)!(n-i)!} (e^{-e^{-x}})^{i-1}$$

$$(1 - e^{-e^{-x}})^{n-i} e^{-x} e^{-e^{-x}}, \quad i = 1, 2, \dots, n,$$

$$f_{Z_{(i)}, Z_{(j)}}(x, y) = \frac{n!}{(i-1)!(j-i-1)!(n-j)!} (e^{-e^{-x}})^{i-1}$$

$$(e^{-e^{-y}} - e^{-e^{-x}})^{j-i-1} (1 - e^{-e^{-y}})^{n-j} e^{-x} e^{-e^{-x}}$$

$$e^{-y} e^{-e^{-y}}, \quad 1 \leq i < j \leq n, x \leq y. \quad (27)$$

The mean, the variance of $Z_{(i)}$, and the covariance of the $Z_{(i)}$ and $Z_{(j)}$ are

$$\mu_i = \int_{-\infty}^{\infty} x f_{Z_{(i)}}(x) dx = \frac{n!}{(i-1)!(n-i)!} \int_0^1$$

$$[-\ln(-\ln u)] u^{i-1} (1-u)^{n-i} du, \quad i = 1, 2, \dots, n, \quad (28)$$

$$v_{ii} = \int_{-\infty}^{\infty} x^2 f_{Z_{(i)}}(x) dx - \mu_i^2 = \frac{n!}{(i-1)!(n-i)!} \int_0^1$$

$$[-\ln(-\ln u)]^2 u^{i-1} (1-u)^{n-i} du - \mu_i^2, \quad i = 1, 2, \dots, n, \quad (29)$$

$$v_{ij} = \int_{-\infty}^{\infty} \int_x^{\infty} x y f_{Z_{(i)}, Z_{(j)}}(x, y) dy dx - \mu_i \mu_j$$

$$= \frac{n!}{(i-1)!(j-i-1)!(n-j)!} \int_0^1 \int_u^1$$

$$[-\ln(-\ln u)] [-\ln(-\ln v)] u^{i-1}$$

$$(v-u)^{j-i-1} (1-v)^{n-j} dv du - \mu_i \mu_j, \quad i = 1, 2, \dots, n. \quad (30)$$

Then, the mean of $X_{(i)}$ is

$$\mu_{X_{(i)}} = \mu + \alpha \mu_i = \mu (1 + c \mu_i), \quad i = 1, 2, \dots, n, \quad (31)$$

$$c = \frac{(\sqrt{6}/\pi) \delta_X}{1 - (C_E \sqrt{6}/\pi) \delta_X}.$$

The covariance matrix and its inverse matrix of $Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}$ are, respectively, denoted as

$$V = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nm} \end{pmatrix}, \quad (32)$$

$$V^{-1} = \begin{pmatrix} v^{11} & v^{12} & \cdots & v^{1n} \\ v^{21} & v^{22} & \cdots & v^{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v^{n1} & v^{n2} & \cdots & v^{nm} \end{pmatrix}.$$

According to the least square method of parameter estimation [36], take the weighted average sum as

$$Q = \sum_{i=1}^n \sum_{j=1}^n \left[(X_{(i)} - \mu_{X_{(i)}}) v^{ij} (X_{(j)} - \mu_{X_{(j)}}) \right]. \quad (33)$$

When the coefficient of variation δ_X is known, let

$$\frac{dQ}{d\mu} = 0. \quad (34)$$

When the coefficient of variation δ_X is known, the least square estimate of the unknown parameter μ is

$$\mu^* = \sum_{j=1}^n D(n, n, j, \delta_X) X_{(j)}, \quad (35)$$

$$D(n, n, j, \delta_X) = \frac{\sum_{i=1}^n [(1 + c\mu_i) v^{ij}]}{\sum_{i=1}^n \sum_{j=1}^n [(1 + c\mu_i) v^{ij} (1 + c\mu_j)]}, \quad j = 1, 2, \dots, n. \quad (36)$$

Since

$$X_{(j)} = \mu + \alpha Z_{(j)} = \mu(1 + cZ_{(j)}), \quad j = 1, 2, \dots, n. \quad (37)$$

Hence, the mean of μ^* is

$$E(\mu^*) = \sum_{j=1}^n \frac{\sum_{i=1}^n (1 + c\mu_i) v^{ij}}{\sum_{i=1}^n \sum_{j=1}^n [(1 + c\mu_i) v^{ij} (1 + c\mu_j)]} \mu(1 + c\mu_j) = \mu, \quad (38)$$

that is, the linear unbiased estimator of μ [37]. If $x_{(1)}, x_{(2)}, \dots, x_{(n)}$ are the test values of $X_{(1)}, X_{(2)}, \dots, X_{(n)}$, then the linear unbiased estimate is

$$\mu^* = \sum_{j=1}^n D(n, n, j, \delta_X) x_{(j)}. \quad (39)$$

Compared to the corresponding coefficient in the current linear unbiased estimate $D(n, n, j)$ [37], the influence of the coefficient of variation δ_X is considered in the coefficient $D(n, n, j, \delta_X)$. Since it is difficult to obtain the analytical expressions of equations (28)–(30), the values of v_{ij} , μ_i , and v_{ii} must be determined by numerical integration, the value of $D(n, n, j, \delta_X)$ can be determined from (36), the values when $n = 10$ are listed in Table 1, and others are omitted.

3.2.2. *Interval Estimation of Tantile.* For the standard values of snow load, a relatively large tantile is usually selected

TABLE 1: Numerical of $D(n, n, j, \delta)$.

n	j	δ						
		0.2	0.3	0.4	0.5	0.6	0.7	0.8
10	1	0.2379	0.1713	0.0952	0.0173	-0.0529	-0.1076	-0.1434
10	2	0.1670	0.1454	0.1173	0.0852	0.0528	0.0239	0.0009
10	3	0.1355	0.1268	0.1131	0.0952	0.0749	0.0547	0.0369
10	4	0.1127	0.1118	0.1067	0.0974	0.0847	0.0702	0.0559
10	5	0.0944	0.0985	0.0992	0.0959	0.0885	0.0783	0.0667
10	6	0.0786	0.0862	0.0910	0.0920	0.0887	0.0817	0.0725
10	7	0.0644	0.0744	0.0822	0.0863	0.0861	0.0818	0.0745
10	8	0.0513	0.0628	0.0725	0.0790	0.0811	0.0790	0.0736
10	9	0.0386	0.0506	0.0614	0.0693	0.0733	0.0730	0.0691
10	10	0.0263	0.0380	0.0490	0.0576	0.0627	0.0638	0.0615

under the condition of a small sample, and the upper limit value in the interval estimate should be used as its estimate value.

Let

$$x_p = \mu + k\alpha = \mu(1 + ck), \quad (40)$$

$$k = -\ln(-\ln p),$$

where p is the guaranteed probability when $\{X \leq x_p\}$, and the value is relatively large. When the coefficient of variation δ_X is known, it can be known from (35) and (37) that

$$\frac{\mu}{\mu^*} = \frac{1}{\sum_{j=1}^n D(n, n, j, \delta_X) (1 + cZ_{(j)})}. \quad (41)$$

Let $U = \mu/\mu^*$, then U is a statistic that is independent of the distribution parameters μ and α . Based on the probability distribution of U , the estimate available of μ by the upper limit in the interval estimate is

$$\bar{\mu} = u(n, n, \delta_X, C) \mu^*, \quad (42)$$

where $u(n, n, \delta_X, C)$ is the down tantile of U , C is the confidence degree, and the value is relatively large. The value of x_p estimated from the upper limit is

$$\bar{x}_p = \bar{\mu}(1 + ck) = u(n, n, \delta_X, C) \mu^* (1 + ck). \quad (43)$$

It is almost impossible to determine $u(n, n, \delta_X, C)$ by the analytical method, so we need to use Monte Carlo numerical simulation to determine the value. The random numbers that obey the standard maximum type I distribution which are first generated in each simulation, a set of sample values $Z_{(1)}, Z_{(2)}, \dots, Z_{(n)}$, can be obtained after sorting, and the sample value of U is obtained from equation (41). When the number of simulations is sufficient, any tantile $u(n, n, \delta_X, C)$ can be obtained by statistics. Here, the number of simulations is 50,000. Table 2 lists the partial numerical tables when $n = 10, 15, 20$.

4. How to Choose the Confidence Degree in Engineering Practice

The confidence degree C is a representative of the trust level for the inferring results, and it has a direct impact on the inferring results; the higher the confidence degree, the higher the upper limit estimated value and the lower the lower limit

TABLE 2: Numerical of $u(n, n, \delta, C)$.

n	C	δ						
		0.2	0.3	0.4	0.5	0.6	0.7	0.8
10	0.50	1.0030	1.0046	1.0066	1.0095	1.0130	1.0164	1.0183
10	0.75	1.0417	1.0678	1.0960	1.1238	1.1518	1.1760	1.1962
10	0.90	1.0770	1.1266	1.1818	1.2409	1.2976	1.3490	1.3923
15	0.50	1.0020	1.0032	1.0047	1.0061	1.0086	1.0107	1.0123
15	0.75	1.0336	1.0537	1.0754	1.0974	1.1186	1.1365	1.1511
15	0.90	1.0626	1.1023	1.1466	1.1908	1.2332	1.2698	1.3001
20	0.50	1.0013	1.0023	1.0035	1.0047	1.0061	1.0069	1.0079
20	0.75	1.0284	1.0456	1.0640	1.0827	1.1002	1.1148	1.1266
20	0.90	1.0537	1.0870	1.1229	1.1595	1.1942	1.2233	1.2483

estimated value, and vice versa. At present, there is no suggestion about how to choose the confidence degree for inferring representative values and design values of variable actions in the literature.

Generally speaking, the larger the variability of the random variable X , the larger the value range of the distributed parameter and tantes in the inference; moreover, the changes of C have great influences on the upper and lower limit estimated values when the confidence degree C gets larger and larger; that is, along with the C becoming larger, the upper limit estimated value is increased relatively quickly and the lower limit estimated value is reduced relatively quickly. When the variability of the random variable X is larger, and if we select the higher confidence degree in this case, the inferring results are too conservative, so we should select the relatively lower confidence degree. On the contrary, we should select the relatively higher confidence degree when the variability of the random variable X is smaller; it can avoid too aggressive inferring results.

The National Standard of the People's Republic of China (standard for appraiser of reliability of civil buildings) (GB50292-1999) throws out some suggestion about how to select the confidence degree to infer the standard value of the material strength: for steel, select $C = 0.90$; for concrete, select $C = 0.75$; for masonry, select $C = 0.60$; and it is suggested that $C = 0.95$ to infer the standard value of permanent action. The variation coefficients of material strength of the permanent action, steel, concrete, and masonry are in the order as 0.07, 0.06–0.10, 0.16–0.23, and 0.20–0.24 [20]. It can be seen that the value of the confidence degree in the National Standard of the People's Republic of China conforms to the above rules. The variation coefficient of snow load is more variable [38]; according to the above rules, we should select the relatively lower confidence degree, but the value should not be less than 0.5; otherwise, the lower limit estimated value will be higher than the upper limit estimated value. When the confidence degree $C = 0.5$ and all other conditions are equal, we contrast the inferring results of the moment method and the liner regression estimation method. It shows that the inferring result of the liner regression estimation method is slightly higher than the moment method, which is close to the result without considering the influence of statistical uncertainty. When the confidence degree C is not great, the changes of C have no great influence on the inferring results [39], so we can select the slightly higher confidence degree, which can take into account the influences of statistical uncertainty more fully and avoid the

too aggressive inferring results. In this article, we suggest to select the confidence degree $C = 0.75$ to infer the standard values of snow load. We will illustrate the correctness by an example in Section 5.

5. Examples

In this section, we use the established linear regression estimation method to estimate the standard values of snow load. The data used in this article are the actual snowfall data from 1989 to 2008 in the Shuyang county [40], and the specific values are shown in Table 3.

To compare and analyze the differences in the extrapolation results under different sample sizes, the above sample values are divided into three groups: 2008–1999 (10 sample data), 2008–1994 (15 sample data), and 2008–1989 (20 sample data) [41–45]. The results are calculated according to different inference methods.

We select the guarantee rate $p = 1 - 1/50 = 0.98$, and the confidence degree C in the interval estimate is 0.75.

By using the moment method, we can obtain the estimate of the standard values of snow load through calculation as follows:

$$\begin{aligned}\alpha &= \frac{\sqrt{6}}{\pi} s = 0.042, \\ \mu &= \bar{x} - C_E \alpha = 0.121, \\ x_p &= \mu + k\alpha = 0.285.\end{aligned}\quad (44)$$

According to the maximum likelihood estimation method, by introducing the likelihood function, we can derive the parameter estimation formula as follows:

$$\begin{aligned}\mu &= -\alpha \ln \left(\frac{1}{n} \sum_{i=1}^n e^{-(x_i/\alpha)} \right), \\ \alpha &= \bar{x} - \frac{\sum_{i=1}^n (x_i e^{-(x_i/\alpha)})}{\sum_{i=1}^n e^{-(x_i/\alpha)}}.\end{aligned}\quad (45)$$

We can obtain the estimated values of the parameters, respectively, by the iterative method, and finally obtain the standard values of snow load through calculation as follows:

$$\begin{aligned}\hat{\alpha} &= 0.041, \\ \hat{\mu} &= 0.122, \\ \hat{x}_p &= 0.282.\end{aligned}\quad (46)$$

By using the linear regression estimation method of the standard value of snow load in the condition of unknown parameter information (equations (13), (14), and (18)), we can obtain the estimate through calculation as follows:

$$\begin{aligned}\tilde{\alpha} &= -\sum_{j=1}^n C_I(n, n, j) x_{(n-j+1)} = 0.042, \\ \tilde{\mu} &= \sum_{j=1}^n D_I(n, n, j) x_{(n-j+1)} = 0.119, \\ \tilde{x}_p &= \tilde{\mu} + v_{p,C} \tilde{\alpha} = 0.339.\end{aligned}\quad (47)$$

TABLE 3: Statistics of the maximum snow depth and pressure every year in the Shuyang county.

Year	Snow depth (m)	Snow pressure (kN·m ⁻²)
2008	0.18	0.265
2007	0.05	0.074
2006	0.11	0.162
2005	0.09	0.132
2004	0.09	0.132
2003	0.1	0.147
2002	0.08	0.118
2001	0.10	0.147
2000	0.06	0.088
1999	0.13	0.191
1998	0.11	0.162
1997	0.08	0.118
1996	0.08	0.118
1995	0.06	0.088
1994	0.13	0.191
1993	0.14	0.206
1992	0.08	0.118
1991	0.21	0.309
1990	0.11	0.162
1989	0.11	0.162

TABLE 4: Numerical table and calculation process.

j	$x_{(10-j+1)}$	$D_I(10, 10, j)$	$C_I(10, 10, j)$	$x D_I(10, 10, j)$	$-x C_I(10, 10, j)$
1	0.265	0.0273	-0.0727	0.0072	-0.0192
2	0.191	0.0400	-0.0780	0.0076	-0.0149
3	0.162	0.0525	-0.0772	0.0085	-0.0125
4	0.147	0.0654	-0.0719	0.0096	-0.0106
5	0.147	0.0793	-0.0617	0.0117	-0.0091
6	0.132	0.0946	-0.0454	0.0125	-0.0060
7	0.132	0.1124	-0.0207	0.0149	-0.0027
8	0.118	0.1342	0.0179	0.0158	0.0021
9	0.088	0.1642	0.0851	0.0145	0.0075
10	0.074	0.2300	0.3246	0.0169	0.0239
Sum		1	0	0.119	0.042

The values of $D_I(10, 10, j)$ and $C_I(10, 10, j)$ are listed in Table 4, and $v_{0.99,0.75} = 6.23$.

In the condition of the known coefficient of variation, to facilitate comparison under the same conditions, we select $\delta_X = 0.4$, by using the moment method, then

$$\mu' = \bar{x} \left(1 - \frac{C_E \sqrt{6}}{\pi} \delta_X \right) = 0.120, \tag{48}$$

$$x'_p = \mu' (1 + ck) = 0.297.$$

By using the maximum likelihood estimation method, we can obtain the estimated values of the parameters, respectively, by the iterative method, and finally obtain the standard values of snow load:

$$\tilde{x}_p = 0.286. \tag{49}$$

By using the linear regression estimation method that the coefficient of variation is known (equations (39) and (42)), then

$$\mu^* = \sum_{j=1}^n D(n, n, j, \delta_X) x_{(j)} = 0.120, \tag{50}$$

$$\tilde{x}_p = u(n, n, \delta_X, C) \mu^* (1 + ck) = 0.327,$$

where $u(10, 10, 0.4, 0.75) = 1.096$ and the value of $D(10, 10, j, 0.4)$ is shown in Table 1; Table 5 shows the calculation process.

To facilitate comparative analysis, Table 6 lists the statistical extrapolation results of different estimation methods and different sample sizes in two cases where the coefficient of variation is known and the parameter information is unknown.

A comparative analysis of the calculations presented in Table 6 shows that

- (1) Regardless of whether the coefficient of variation is unknown or known, the results estimated by the maximum likelihood method are the smallest, followed by the moment method, because these two

TABLE 5: Numerical table and calculation process.

j	x_j	$D(10, 10, j, 0.4)$	$x D(10, 10, j, 0.4)$
1	0.074	0.0952	0.013
2	0.088	0.1173	0.012
3	0.118	0.1131	0.012
4	0.132	0.1067	0.012
5	0.132	0.0992	0.013
6	0.147	0.091	0.013
7	0.147	0.0822	0.014
8	0.162	0.0725	0.013
9	0.191	0.0614	0.010
10	0.265	0.0490	0.007
Sum			0.120

TABLE 6: Comparison of inference results of different inference methods.

Parameter information	Inference method	Inference results (kN·m ⁻²)			Relative error (%)		
		$n = 10$	$n = 15$	$n = 20$	$n = 10$	$n = 15$	$n = 20$
Unknown information	Moment	0.285	0.268	0.304	18.9	17.9	16.8
	Likelihood	0.282	0.265	0.295	20.2	19.2	20.3
	Regression	0.339	0.316	0.355			
Known coefficient of variation	Moment	0.297	0.289	0.314	10.1	9.0	7.0
	Likelihood	0.286	0.268	0.304	12.9	14.9	9.6
	Regression	0.327	0.315	0.336			

classical statistical methods do not take into account the influences of statistical uncertainty, and the inferred results are always on the aggressive side.

- (2) Regardless of whether the coefficient of variation is unknown or known, with the gradual increase in the sample capacity, the relative errors of the results of the linear regression method and the classical statistical methods (the moment method and the maximum likelihood method) are gradually reduced, which is due to the gradual reduction of the influence of statistical uncertainty when the sample capacity increases.
- (3) According to the linear regression method, at different sample sizes, the estimated results when the coefficient of variation is known are smaller, which is due to the significant reduction in the uncertainty of the statistic U in the estimation process.

Due to space limitation, this paper only compares the extrapolation values of different inference methods in three groups of sample sizes. By calculating and comparing the extrapolation of the standard values of snow loads in other sample sizes, the same results can be obtained.

6. Conclusions

- (1) When we use the current statistical inference method to infer the standard value of snow load in the conditions of small samples, the inferring results are always on the aggressive side because it does not take into account the influences of statistical uncertainty.

- (2) The linear regression estimation method presented in this paper can reduce the influence of statistical uncertainty when the sample is small, and it is applicable to the statistical inference of the standard values of snow load in small sample conditions; considering the two cases where no parameter information and the coefficient of variation are known, more reasonable inference results can be given.
- (3) We suggest to select the confidence degree $C = 0.75$ to infer the standard value of snow load.
- (4) In the practical application, even if the coefficient of variation is unknown, the upper limit of the coefficient of variation can be set based on experience, the estimation can be made according to the no parameter information, the coefficients of variation are known, and the smaller value of the two is taken as the final estimate. The research in this paper can provide a theoretical basis for adjusting snow load.

Data Availability

The authors solemnly inform that the key data in the calculation process have been listed in the data list of the article, and other nonkey detailed data can be obtained by contacting the corresponding author if necessary.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This project was supported by the National Natural Science Foundation of China (nos. 50678143 and 51278401) and the Education Department of Shaanxi (no. 17JK0440).

References

- [1] GB50009-2012, *Load Code for the Design of Building Structure*, China Architecture & Building Press, Beijing, China, 2012.
- [2] J. Aitchison and S. D. Silvey, "Maximum-likelihood estimation of parameters subject to restraints," *The Annals of Mathematical Statistics*, vol. 29, no. 3, pp. 813–828, 1958.
- [3] D. D. Dorfman and E. Alf, "Maximum-likelihood estimation of parameters of signal-detection theory and determination of confidence intervals-rating-method data," *Journal of Mathematical Psychology*, vol. 6, no. 3, pp. 487–496, 1969.
- [4] S. I. Aihara, "Regularized maximum likelihood estimate for an infinite-dimensional parameter in stochastic parabolic systems," *SIAM Journal on Control and Optimization*, vol. 30, no. 4, pp. 745–764, 1992.
- [5] S. S. Li, "Estimation method for parameters of extreme value I distribution," *Journal of Fuzhou University*, vol. 16, no. 1, pp. 79–84, 1998.
- [6] Z. D. Duan and D. C. Zhou, "A comparative study on parameter estimate method for extremal value distribution," *Journal of Harbin Institute of Technology*, vol. 36, no. 12, pp. 1605–1609, 2004.
- [7] G. Aneiros-Pérez and P. Vieu, "Semi-functional partial linear regression," *Statistics & Probability Letters*, vol. 76, no. 11, pp. 1102–1110, 2006.
- [8] P. Hall and J. L. Horowitz, "Methodology and convergence rates for functional linear regression," *The Annals of Statistics*, vol. 35, no. 1, pp. 70–91, 2007.
- [9] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2010.
- [10] F. Cribari-Neto and W. B. da Silva, "A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model," *AStA Advances in Statistical Analysis*, vol. 95, no. 2, pp. 129–146, 2011.
- [11] Y. Zhao, X. Meng, and H. Yang, "Jackknife empirical likelihood inference for the mean absolute deviation," *Computational Statistics & Data Analysis*, vol. 91, pp. 92–101, 2015.
- [12] G. C. Manjunath Patel, P. Krishna, and M. B. Parappagoudar, "Squeeze casting process modeling by a conventional statistical regression analysis approach," *Applied Mathematical Modeling*, vol. 40, no. 15–16, pp. 6869–6888, 2016.
- [13] T. Sun, L. Cheng, and H. Jiang, "Greedy method for robust linear regression," *Neurocomputing*, vol. 243, pp. 125–132, 2017.
- [14] M. O'Rourke, A. DeGaetano, and J. Tokarczyk, "Analytical simulation of snow drift loading," *Journal of Structural Engineering*, vol. 131, no. 4, pp. 660–667, 2005.
- [15] Y. N. Zhang, Y. Zhang, Y. Q. Wang, and Y. J. Shi, "Calculation and analysis of basic snow pressure in Liaoning province," *Journal of South China University of Technology (Natural Science Edition)*, vol. 38, no. 9, pp. 108–112, 2010.
- [16] S. Basu, "Improved small sample inference procedures for epidemiological parameters under cross-sectional sampling," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 50, no. 3, pp. 309–319, 2001.
- [17] S. Paul and X. Zhang, "Small sample GEE estimation of regression parameters for longitudinal data," *Statistics in Medicine*, vol. 33, no. 22, pp. 3869–3881, 2014.
- [18] T. P. Talafuse and E. A. Pohl, "Small sample discrete reliability growth modeling using a grey systems model," *Grey Systems: Theory and Application*, vol. 8, no. 3, pp. 246–271, 2018.
- [19] S.-K. Fan, C.-H. Jen, and J.-X. Lee, "Profile monitoring for autocorrelated reflow processes with small samples," *Processes*, vol. 7, no. 2, p. 104, 2019.
- [20] GB50292-1999, *Standard for Appraiser of Reliability of Civil Buildings*, China Architecture & Building Press, Beijing, China, 1999.
- [21] J. T. Yao and X. D. Wang, "Minor sample estimation of the parameters and fractile of largest extreme value distribution type I," *Statistics and Decision*, vol. 19, pp. 11–14, 2014.
- [22] S. Emilio and B. Sen, "Nonparametric least squares estimation of a multivariate convex regression function," *The Annals of Statistics*, vol. 39, no. 3, pp. 1633–1657, 2011.
- [23] S. Aihara, "Consistency property of extended least-squares parameter estimation for stochastic diffusion equation," *Systems & Control Letters*, vol. 34, no. 5, pp. 249–256, 1998.
- [24] D. Kumar, S. F. Ali, and A. Arockiarajan, "Structural and aerodynamics studies on various wing configurations for morphing," *IFAC-PapersOnLine*, vol. 51, no. 1, pp. 498–503, 2018.
- [25] B. Keshtegar and P. Hao, "A hybrid self-adjusted mean value method for reliability-based design optimization using sufficient descent condition," *Applied Mathematical Modelling*, vol. 41, pp. 257–270, 2017.
- [26] Y.-G. Zhao and T. Ono, "Moment methods for structural reliability," *Structural Safety*, vol. 23, no. 1, pp. 47–75, 2001.
- [27] X. Shi, A. P. Teixeira, J. Zhang, and C. Guedes Soares, "Structural reliability analysis based on probabilistic response modelling using the maximum entropy method," *Engineering Structures*, vol. 70, pp. 106–116, 2014.
- [28] X. Zhang and M. D. Pandey, "Structural reliability analysis based on the concepts of entropy, fractional moment and dimensional reduction method," *Structural Safety*, vol. 43, pp. 28–40, 2013.
- [29] M. Koliou and A. Filiatrault, "Development of wood and steel diaphragm hysteretic connector database for performance-based earthquake engineering," *Bulletin of Earthquake Engineering*, vol. 15, no. 10, pp. 4319–4347, 2017.
- [30] Z. Xi, C. Hu, and B. D. Youn, "A comparative study of probability estimation methods for reliability analysis," *Structural and Multidisciplinary Optimization*, vol. 45, no. 1, pp. 33–52, 2012.
- [31] S. S. Mao, *Statistics Handbook*, Science Press, Beijing, China, 2003.
- [32] SJ/T11099-1996, *Tables for Best Linear Unbiased Estimate (BLUE) (Extreme-Value Distribution, Weibull Distribution)*, China Architecture & Building Press, Beijing, China, 1996.
- [33] GB50068-2001, *Unified Standard for Reliability Design of Building Structures*, China Architecture & Building Press, Beijing, China, 2001.
- [34] GB50153-2008, *Unified Standard for Reliability Design of Engineering Structures*, China Architecture & Building Press, Beijing, China, 2009.
- [35] S. S. Dai and H. L. Fei, *Reliability Test and Statistical Analysis (First Book)*, National Defence Industry Press, Beijing, China, 1983.
- [36] Research Department of Machinery Industry Standard Fourth, *Table for Reliability Test*, National Defence Industry Press, Beijing, China, 1979.

- [37] GB12282 1-90, *Tables for Life Testing Tables for Best Linear Unbiased Estimate (BLUE) (Extreme-Value Distribution, Weibull Distribution)*, Publishing House of Electronics Industry, Beijing, China, 1996.
- [38] Z. Chang and X. M. Xie, "Research and suggestions on wind and snow live load value of buildings," *Building Science*, vol. 27, no. 1, pp. 83–85, 2011.
- [39] J. T. Yao and X. D. Wang, "Linear regression estimation of representative values of variable actions," *Journal of Building Structures*, vol. 35, no. 10, pp. 98–103, 2014.
- [40] C. J. Kang, W. F. Du, X. Yang, and Z. Zhu, "Calculation and analysis of snow loads based on the maximum likelihood method," *Journal of Henan University (Natural Science)*, vol. 46, no. 2, pp. 220–225, 2016.
- [41] A. Q. Baig, M. Naeem, and W. Gao, "Revan and hyper-Revan indices of octahedral and icosahedral networks," *Applied Mathematics and Nonlinear Sciences*, vol. 3, no. 1, pp. 33–40, 2018.
- [42] M. Dewasurendra and K. Vajravelu, "On the method of inverse mapping for solutions of coupled systems of nonlinear differential equations arising in nanofluid flow, heat and mass transfer," *Applied Mathematics and Nonlinear Sciences*, vol. 3, no. 1, pp. 1–14, 2018.
- [43] Y. Qin, Y. Luo, Y. Zhao, and J. Zhang, "Research on relationship between tourism income and economic growth based on meta-analysis," *Applied Mathematics and Nonlinear Sciences*, vol. 3, no. 1, pp. 105–114, 2018.
- [44] J. F. Gómez-Aguilar and A. Atangana, "Time-fractional variable-order telegraph equation involving operators with Mittag-Leffler kernel," *Journal of Electromagnetic Waves and Applications*, vol. 33, no. 2, pp. 165–177, 2019.
- [45] D. P. Ahokpossi, A. Atangana, and P. D. Vermeulen, "Hydro-geochemical characterizations of a platinum group element groundwater system in Africa," *Journal of African Earth Sciences*, vol. 138, pp. 348–366, 2018.