

Method for calculating loads combination on a building using information measures

Elena Chernetsova^{1,*}, Anatoly Shishkin¹

¹Russian State Hydrometeorological University, 195196, Malookhtinsky Ave, 98, St. Petersburg, Russia

Abstract. A method for calculating loads combination on a building is considered using information measures of the connectivity of signals received from sensors of various physical nature, united in a wireless monitoring network. The method includes the definition of the most powerful information measure on the ensemble of process realizations with known a priori load data by the criterion of connectedness of time series. Then, based on the selected information measure, the connectivity of the signals for the ensemble of realizations of the random process of loads to the building from the network formed by the wireless monitoring data bank of time series is calculated. The volume of the data bank sufficient to make the correct decision about the combination of loads on the building with a predetermined error probability is calculated on the basis of a consistent criterion for the ratio of Wald probabilities. This method is easily algorithmized and can be used to develop an automated decision support system.

1 Introduction

Building structures and the foundation of a building or structure must have such strength and stability that during the construction and operation there is no threat of harm to life or health of people, property of individuals or legal entities, state or municipal property, the environment, life and health of animals and plants. Therefore, the analysis of loads on structures is one of the main tasks of construction mechanics.

The problem of combinations of random loads is usually considered as the problem of exceeding a given level of the sum of random processes under the assumption of the variability of the load, which is realized only in time. However, for many loads, the distribution along the surface of a structure or other spatial parameters, such as a random region of impact, is also random, that is, they should be considered as spatio-temporal random processes [1]. To justify such loads, it is recommended to apply probability-statistical methods [2].

When calculating the combination of loads, the values of the constant P_d , temporary long-time P_l and temporary short-time P_t loads [3] are multiplied by the corresponding coupling coefficients. These recommendations are based on engineering experience, which only in some cases was supported by field observations (see, for example, [4]). At the same time, the

* Corresponding author: chernetsova@list.ru

differences in the laws of the changes in loads in time were not taken into account, and the differences in the length of the action of various loads were taken into account to a very small degree.

The ISO standard and Eurocode [5] present an approach to calculating load combinations based on the rules of the Turkstra [6]. In this case, the extreme value of one load is combined with the values of all the others for an arbitrary moment of time. For m temporary loads, there are m combinations of this kind, and the most important of these combinations is the determining one. In this approach, there are as many coupling coefficients as there are cases of loading. The combination of load combinations can be represented as a matrix. In the symmetric case, there are units on the main diagonal of the matrix that can be interpreted as the coefficients of combinations of some load with itself. In the case of an asymmetric matrix, the combination rule recommended in [7] is realized on the basis of theoretical studies of the reliability problem.

In connection with the new national standard of the structural inspection accepted in Russia in 2011 [8], it is necessary to perform the building construction monitoring. This becomes possible with the use of the wireless sensor network (WSN) that allows inspecting such characteristics as vibration, deformation, deviation, etc. with the periodicity convenient for a user. A wireless sensor network consists of spatially distributed autonomous sensor nodes and execution units unified by a radio channel to monitor physical or/and environmental conditions. WSN coverage area can achieve several kilometers due to the messages forwarding from one element to another. The monitoring system includes different types of sensors installed on structural elements of a building with a view to detect both the physical (humidity and temperature) and force (static and dynamic loads) effects on their strength and deformation properties [9].

Time series of effects obtained from different types of sensors can be considered as continuous over the entire time interval of monitoring. They supplement the database created in the framework of Building Information Model development (BIM [10]) and can be used for:

- determining the correlation of random combined loads, since the a priori statement about the independence of loads is not always true;
- accounting for combinations of actual, in some cases rather complex, distributions of random loads, in contrast to the often and not always reasonably used normal law;
- taking into account the specific features of a number of loads that have a complex character of the changes in time associated with production technology and operational features.

The correlation analysis application for loads combination [11] is correct only if the observational data are homogeneous and distributed according to the normal distribution law. But if the relationship between the quantities is determined, it must be remembered that the ratio of independent normally distributed random variables is distributed according to the Cauchy law [12]. The problem of Cauchy distribution is the fact that the sample mean does not evaluate the mathematical expectation. In addition, the sampling elements are strongly ejected into the tails of distributions (taking large values in modulus), as well as the sampling average [13], which makes the use of the Bayesian decision rule difficult [14]. Attempts to find the distribution parameters usually lead to the fact that the same empirical distribution can be described by different hypothetical distributions with the same or better agreement [15].

In addition, the partial correlation coefficient is a measure of the linear relationship between factors. The latter circumstance in the case of the evaluation of pair bonds led to the fact that a large number of dependence measures were heuristically introduced into statistical research, allowing one to take into account the nonlinear character of stochastic dependence between random variables [16].

It was established in [17] that the information characteristics of the relationship for arbitrary random variables from absolutely continuous distributions have the necessary completeness of properties for revealing both linear and nonlinear dependences and their analysis, therefore they can be used to calculate combinations of loads on buildings. The aim of the work is to develop a method for calculating the loads combination on a building using information measures. To achieve this goal, the following tasks are solved in the work:

- Definition of the most powerful by the criterion of determining the connectivity of time series information measure among the selected measures for analysis.
- Determining the nature of the relationship between the selected information measure and the linear correlation coefficient for Gaussian and non-Gaussian time series.
- Development of a method for determining the size of a database of sensor signals sufficient to determine the combination of loads on a building with a predetermined probability of error of type I or II.

2 Methods

To apply the information method for calculating loads combination on a building, it is necessary to have a time series database that is signals received from heterogeneous sensors for monitoring the building's status, usually via a wireless monitoring network.

Since to calculate load combinations it is necessary to obtain the results of comparison of data having, in general, a different physical nature, then it is necessary to use such measures of time series connectivity in which the units of measurement of the recorded signals are not taken into account. When using information measures of the connectivity of time series, there is no need to know the a priori parameters of the distribution of the initial data. To apply informational connectivity measures, it is only necessary to synthesize from the recorded time series one-dimensional and mutual Probability Density Distributions (PDDs) of compared loads.

To build mutual PDDs, it is also necessary that the two compared time series be comparable by any parameter that is independent of the units of measurement. This parameter can be the envelope of each signal, normalized to its maximum. After constructing such envelopes for each of the compared load signals, two time series are obtained, all of whose energy lies in the positive region, and they differ from each other only in form.

In the process of building a mutual PDD of time series that represent two compared loads on a building, the sample should be presented in the form of a histogram consisting of columns with a certain length of the corresponding intervals. If these intervals are the same, the number of hits of the sample elements at different intervals will be comparable. But there is also an optimal number of class-grouping intervals, in which the stepped envelope of the histogram is closest to the smooth curve of the distribution of the population [18].

Since the optimal number of histogram intervals depends on the type of the distribution law of the experimental data, which is not known in this case, it is recommended to build an equal-interval histogram to select the number of intervals to use the algorithm for selecting an appropriate calculation formula based on the entropic coefficient proposed in [19].

After constructing two-dimensional and two-dimensional PDDs, we define a combination of two loads based on information connectivity measures showing how much the mutual PDD of two time series is removed from the product of their individual PDDs.

As information measures of connectivity of two time series can be, in particular, accepted

1. The information criterion for the Kullback-Leibler distance:

$$\rho_1(x, y) = \int \log \frac{w(x, y)}{w(x)w(y)} w(x, y) \mu(dxdy), \quad (1)$$

where $w(x, y)$ - joint PDD of two random time series x and y ;

$w(x)$, $w(y)$ - one-dimensional (independent) PDDs of two random time series x and y ;

$\mu(dxdy)$ - dominant measure,

which is nothing more than the amount of Shannon information $I(x, y)$ in the parameter y about the message x [20]. When using the logarithm for base 2 in formula (1), the obtained numerical values of the Kullback measure are easily interpreted: if the time series x and y coincide, the Kullback measure is 2 bits (since this corresponds to the resolution of information uncertainty with respect to two random variables).

2. The information criterion for χ^2 :

$$\rho_2(x, y) = \int \frac{(w(x, y) - w(x)w(y))^2}{w(x, y)} \mu(dxdy) \quad (2)$$

3. The information criterion for the Hellinger distance:

$$\rho_3(x, y) = \int (\sqrt{w(x, y)} - \sqrt{w(x)w(y)})^2 \mu(dxdy) \quad (3)$$

4. The information criterion for the square of the distance between probability densities:

$$\rho_4(x, y) = \int (w(x, y) - w(x)w(y))^2 \mu(dxdy) \quad (4)$$

Kullback-Leibler distance plays a special role in information theory and finds a natural application in Bayesian theory. However, neither the Kullback-Leibler distance nor the chi-square distance are symmetric, nor can they be determined for all points of the parametric set in the case when the density carrier depends on the parameter. The Hellinger distance can be used to determine the proximity between measures from one family indexed by different parameters. This distance does not depend on the choice of the dominant measure and is defined for all points of the parametric set [21].

The Kullback-Leibler distance is widely used to compare distributions, but it is considered unsuitable for statistical purposes because it does not have a limiting distribution. In [22], however, it was shown that the Kullback-Leibler distance distribution in the limit is bounded from above by a chi-square. This gives, albeit limited, the possibility of using the Kullback-Leibler distance between distributions as statistics to test the hypothesis that two samples belong to the same distribution and allows us to speak of the statistical significance of the Kullback-Leibler distance.

After choosing the information measure of connectivity to determine the coefficient of combination of pairs of loads and calculating the numerical value of the chosen measure, it is necessary to determine the number of realizations of the random process, sufficient for averaging over the ensemble of realizations.

At this stage of the computation, we can apply the Hodges-Lehmann estimate, which was found to differ little from Wilcoxon statistics that is widespread for solving similar problems, but reduces the number of computational costs required [23].

Estimation of the shift of the Hodges-Lehmann distributions. is the median of all possible pairs of differences of elements of one and the other group:

$$r_{ij} = med(x_{ui} - x_{vj}); u = 1, \dots, n_i; v = 1, \dots, n_j. \tag{5}$$

where x_{ui} and x_{vj} - time series i and j parameters.

The positive property of this median is that as far as the first group is "bigger" than the second, so the second is "less" than the first, that is: $r_{ij} = -r_{ji}$.

An essential shortcoming of the Hodges-Lehmann median is nontransitivity. If x is greater than y by a , and s is greater than y by b , then it is desirable that s is greater than x by $a + b$. The median does not possess this property [24].

However, since in our case there are only two alternatives, then the Hodges-Lehmann statistics are legitimate.

As a tool, we use the Successive Criterion for the ratio of Wald Probabilities (S.C.W.P.) [25], based on successive ranks, by making the measurements of the compared samples so

that they alternate: $x_1, y_1, x_2, y_2, \dots, x_n, y_n$. We denote the combined measurements at

the k -th step by the vector $V(k) = [v_1, v_2, \dots, v_k]$, where $v_1 = x_1, v_2 = y_1$ etc. Let

$S(k) = [S_1, S_2, \dots, S_k]$ is a vector of consecutive ranks for $V(k)$, and

$$\lambda_k = \frac{P_k(S(k)/H_1)}{P_k(S(k)/H_0)} \tag{6}$$

is a successive ratio of probabilities at the k -th step of the process. If the hypothesis H_0 is true, then for an arbitrary vector S from $S(k)$ we have $P_k(S(k)/H_0) = 1/k!$ and, consequently, it is possible to calculate $P_k(S(k)/H_1)$, taking into account that each obtained vector S corresponds in a one-to-one manner to a certain order of combined measurements x_i and y_i .

Thus, it suffices to calculate

$$P(v_1 \leq v_2 \leq \dots \leq v_k / H_1) = \int_{-\infty < t_1 \leq t_2 \leq \dots \leq t_k < \infty} \dots \int \prod_{i=1}^k df_i(P(t_i)), \tag{7}$$

where $f_i(P(t_i)) = P(t_i)$, when v_i is x and $f_i(P(t_i)) = f(P(t_i))$, when v_i is y .

In the case of Lehmann's alternatives, we have two hypotheses.

Hypothesis about the connection of two time series corresponding to two loads on the building $H_0: G=P(X)$ against the hypothesis of their independence

$$H_1: G = f(P(X)) = P^r(X); r > 0.$$

In this case, the successive ratio of probabilities at the k -th step

$$\lambda_k = \frac{P_k(S(k)/H_1)}{P_k(S(k)/H_0)} = \frac{k!r^{k/2}}{\prod_{i=1}^k (\sum_{j=1}^i A_j)} \text{ for even values } k \tag{8}$$

and

$$\lambda_k = \frac{P_k(S(k)/H_1)}{P_k(S(k)/H_0)} = \frac{k!r^{(k-1)/2}}{\prod_{i=1}^k (\sum_{j=1}^i A_j)} \text{ for odd values } k \tag{9}$$

$$A_j = \begin{cases} 1; & \text{if } v_j \text{ is } x, \\ r; & \text{if } v_j \text{ is } y. \end{cases} \tag{10}$$

where

Thus, a non-parametric procedure for determining the amount of realization of a random process in a data bank necessary to establish a connection between two building loads using the successive Wald probability ratio criterion is reduced to the following steps:

Step 1: obtaining the successive rank $(k + 1)$ -th sample of the classification parameter.

Step 2: Form a vector $A(k + 1)$ from $A(k)$ and S_{k+1} ...

Step 3. Calculate the successive ratio of probabilities by formula (8) or (9) and compare it with the stopping boundaries.

We denote by e_{ij} - probability of accepting a hypothesis H_i , while in fact the hypothesis H_j ; $i, j = 0, 1$ is true. Then the stopping boundaries (thresholds) in S.C.W.P. of Wald can be approximated by the formulas

$$A = \frac{1 - e_{01}}{e_{10}}, \quad B = \frac{e_{01}}{1 - e_{10}} \tag{11}$$

and the relation between the average number of measurements $E(k)$ and the parameter r is approximately described by the dependence

$$E_r(k) \cong \frac{\log[(1 - e_{10})/e_{01}]}{\log \frac{1}{2}(r^{-1/2} + r^{1/2})} \tag{12}$$

To solve the problem under consideration, we take $e_{10} = e_{01} = 0.1$.

Then the stopping boundaries (thresholds) in S.C.W.P. of Wald will take the values $A=0.9$ and $B=0.11$.

3 Results and Discussion

Power of information connectivity criteria (1-4) can be obtained on a set of time series with known a priori information on connectivity or unconnectivity of sensor signals.

Calculation of information measures is quite easily algorithmized. In the Matlab package, a program was developed that made it possible to calculate the Kullback measure for Gaussian and non-Gaussian signals and to establish the relationship between the calculated measure and the linear correlation coefficient. At large sizes of signal samples, there is a tendency to normalize the processes, so we can observe an approximate coincidence of the value of the Kulbak measure.

The developed method for determining the size of the data bank of sensor signals received from the building's wireless monitoring network uses the successive Wald probability ratio test, and allows to determine the number of random load implementations sufficient to determine the combination of loads on the building with a predetermined probability of type I or II error.

References

1. YU. V. Krasnoshchokov, *Probabilistic basis for the calculation of building structures* (SibADI, Omsk, 2016)
2. I. N. Priadko, V. P. Mushchanov, E. Bartolo, N. I. Vatin, I. N. Rudnieva, *Journal of Civil Engineering*, **5** (2016)
3. *SP 20.13330.2016 Loads and effects: updated version of SNiP 2.01.07-85* (Standartinform, Moscow, 2017)
4. *Operational reliability of metal structures and structures of industrial buildings in extreme conditions of the North* (Fizmatlit, Moscow, 2012)
5. Eurocode: basis of structural design: EN 1990:2001 (Brussels, CEN, 2002)
6. R. E. Melchers, A. T. Beck, *Structural reliability: analysis and prediction* (Wiley & Sons, New York, 2018)
7. B. Ellingwood, D. Rosowsky, *Journal of Structural Engineering*, **122**, 11 (1996)
8. *GOST 31937-2011 Buildings and structures. Rules for inspection and monitoring of technical condition* (Standartinform, Moscow, 2014)
9. E. Barbin, A. Koleda, T. Nesterenko, S. Vtorushin, *Journal of Physics: Conference Series*, **671**, 012003 (2016)
10. M. Najjar, K. Figueiredo, M. Palumbo, A. Haddad, *Journal of Building Engineering*, **14** (2017)
11. D. M. Rotshteyn, *Probabilistic methods in calculations of building structures* (TIU, Tyumen, 2017)
12. V. M. Galkin, L. N. Yerofeyeva, S. V. Leshcheva, *Proceedings of Nizhny Novgorod State Technical University*, **2**, 104 (2014)
13. V. Pisarenko, M. Rodkin, *Mathematical geosciences*, **43**, 4 (2011)
14. S. Huang, W. Zuo, M. Sohn, *Building Simulation*, **11**, 1 (2018)
15. L. A. Valitova, *Relationship between the process of generating data and the resulting distribution of the socio-economic indicator* (MSU, Moscow, 2016)
16. V. A. Simakhin, *Bulletin of Tyumen State University - Physico-Mathematical Modeling - Oil, Gas, Power Engineering*, **7** (2011)
17. S. N. Moiseyev, *Avtometriya*, **5** (2001)
18. V. V. Kalmykov, F. I. Antonyuk, N. V. Zenkin, *Southern-Siberian Scientific Herald*, **3**, 7 (2014)
19. A. N. Tynnyka, *Technology and Design in Electronic Equipment*, **3** (2017)
20. V. YU. Terebizh, *Uspekhi Fizicheskikh Nauk*, **165**, 2 (1995)
21. A. Ye. Shemyakin, *Applied Econometrics*, **4**, 28 (2012)
22. A. P. Motrenko, V. V. Strizhov, *Informatics and Its Applications*, **8**, 2 (2014)
23. A. A. Royenko, V. V. Lukin, A. A. Zelenskiy, *Radioelektroni i Kompjuterny Sistemy*, **2**, 10 (2005).
24. N. S. Kremer, *Mathematical statistics* (Yurayt Publishing House, Moscow, 2017)
25. A. D. Shishkin, E. A. Chernetsova, *Control Systems and Information Technology*, **2**, 52 (2013)
26. A. Bogoviz, S. Lobova, Y. Ragulina, A. Alekseev, S. Litvinova, *Mediaobrazovanie-Media Education*, **4**, 83-91 (2017)