Forecasting the parameters of structures from natural materials taking into account environmental impact

I Ya Lvovich1, Ya E Lvovich2, A P Preobrazhensky1, Yu P Preobrazhensky1, O N Choporov2 and T V Melnikova1

1 Voronezh Institute of High Technologies, 73a, Lenin St., Voronezh, 394043, Russia
2 Voronezh State Technical University, 14, Moscow Prospect, Voronezh, 394026, Russia

E-mail: koronezh@yandex.ru

Abstract. The paper deals with the problems of building structures with certain values of stability parameters based on the corresponding values of material parameters. For modeling, a passive experiment is used based on archival information for building structures. An algorithm using forecast models is built. Its main stages are given, they are indicated in the diagram. The algorithm is based on statistical approaches. The constructed forecast models for stone natural building materials, for binding materials, for forest materials are presented. During the construction of the regression model the authors proceeded from the independence of the experiments, as well as the fact that the nature of various random components remained unchanged in all experiments. The built models may be used to forecast changes in design parameters when the selected characteristics change.

1. Introduction
To date, a large number of building materials have been developed. Material properties may change as a result of external influences. Besides, there is consistency among the structural components that are part of the materials. Then there are opportunities to evaluate the characteristics of materials. The variety of building structures makes it possible to meet a variety of needs under appropriate environmental conditions [1, 2]. The purpose of the work is to study the dependence of the parameters of building structures under various external conditions on the selected material parameters.

2. Development of an algorithm to forecast the parameters of building structures from selected characteristics
When solving control tasks in order to streamline the functioning of construction systems, the nature of control impacts depends on the current state of a studied object, however, if there is a forecast of the ongoing process, the effectiveness of decisions may significantly increase. To solve this problem, it is necessary to build forecast models, on the basis of which it is possible to conduct a simulation experiment in order to select optimal control actions [3].

Active and passive experiment methods based on the regression analysis are most often used to construct the forecast models.
The efficiency of solving problems of modeling, forecasting and optimizing the functioning of building materials increases significantly with the use of computer equipment and specialized software.

The heterogeneous nature of modeling objects requires an individual approach to the control algorithm. It is known that any control system involves two main processes: determining the state of the system and generating the optimal impact for this state. When choosing the control impact tactics, one of the most important stages is the prediction of system parameters based on the forecast models.

To obtain a mathematical description, active and passive experiments based on the regression analysis are most often used. Given the specifics of construction systems, a passive experiment based on archival information is used to model them.

Forecasting changes in construction characteristics is one of the most important estimates, the accuracy of which mainly determines the optimal selection of control impacts.

The algorithm for constructing the forecast models consists of the following stages:

1. A set of indicators is determined \( X_i (i = 1, N) \) based on a survey of experts, which allows fully identifying the state of the modeling object.
2. One or more controlled indicators \( Y_j (j = 1, M) \) are identified.
3. As a result of the dispersion analysis, those indicators \( X_i \) that do not affect the change of any controlled indicator \( Y_j \) are excluded.
4. Information is filtered to select valid measurements.
5. Optimal selection of characteristic space is performed due to elimination of parametric redundancy.
6. The hypothesis of the normal distribution of the values of indicators \( X_i \) is checked.
7. The type of the regression model is selected (linear, incomplete quadratic, quadratic).
8. Estimates of the coefficients of the regression equation are calculated. The least squares method is used. The least squares criterion may be written as follows:

\[
S = \sum_{i=1}^{n} (y_i - y(x))^2 \rightarrow \min .
\]

9. Their significance is checked and insignificant coefficients are excluded from the model.
10. The adequacy of the model is checked. If the model is adequate, this identifies the end of the algorithm. If the model is inadequate, but its complication is possible – refer to item 7, otherwise it is necessary to adjust the original sample (increase its volume, reduce the number of inaccurate measurements). The algorithm for constructing the forecast models is shown in Fig. 1. The construction of multiple regression equations in this case is carried out by the so-called step (multi-step) analysis, during which the model is solved and the selection of factors is completed using statistical and mathematical criteria where the form of connection of each factor with the effective characteristic is specified.

The numerical values of the multiple regression equation parameters are usually determined using the least squares method, for which a system of normal equations is constructed and solved. For linear multiple regression \( Y(x_1, x_2, ..., x_m) = b_0 + b_1 x_1 + b_2 x_2 + ... + b_m x_m \), the system of normal equations will be as follows:
The coefficients at \( x_i \) in the multiple linear regression equation show how much on average the effective feature changes with an increase in the corresponding factor by one and with a fixed (constant) value of other factors included in the regression equation.

The value of the aggregate correlation coefficient from the values of the paired coefficients may be determined as follows:

\[
R^2 = 1 - \frac{\sum_{i=0}^{m} r_{0m}^2 + r_{1m}^2 + r_{2m}^2 + \ldots + r_{m0}^2}{\sum_{i=1}^{m} (r_{10} - r_{11})^2 + (r_{20} - r_{21})^2 + \ldots + (r_{m0} - r_{m1})^2}
\]

The \( R^2 \) value, which is the determination coefficient, shows to what extent the variation of the effective feature is caused by the influence of the feature factors included in the considered correlation equation. The value of the aggregate correlation coefficient varies from 0 to 1 and cannot be numerically less than any of the pair correlation coefficients forming it. The closer the aggregate correlation coefficient is to one, the smaller the role of the factors not considered in the model and the more reason to believe that the parameters of the regression model reflect the degree of effectiveness of the factors included into it [4]. In some cases, the scattering of points of the correlation field is so large that it makes no sense to use the regression equation, since the error in the estimation of the analyzed indicator will be large. For the entire set of observed values, the mean quadratic error of the regression equation \( S_e \) is calculated, which is the mean quadratic deviation of the actual values \( y_i \), relative to the values calculated from the regression equation \( \hat{y}_i \), i.e.

\[
S_e = \sqrt{\frac{n}{n-m} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

where \( S_e \) – mean quadratic error of the regression equation;
\( y_i \) – actual values of the effective characteristic obtained from the observation data;
\( \hat{y}_i \) – values of the effective characteristic calculated by the correlation relation equation and obtained by substituting the values of the factor characteristic \( x_i \) into the regression equation;
\( m \) – number of parameters in the regression equation. A standard error is also defined:

\[
E_{std} = \frac{\sum (y_j - \hat{y}_j)}{n-m}
\]
Development of a set of indicators identifying the state of construction materials $X_i, (i = 1, N)$

Selection of controlled parameters $Y_j, (j = 1, M)$

Dispersion analysis to exclude irrelevant indicators $X_i$.

Data filtration

Choice of the optimal feature space

Is the hypothesis on the normal distribution of indicators $X_i, (i = 1, N)$ confirmed?

True

Choice of the model

Linear

Incomplete quadratic

Quadratic

Calculation of the estimates of the regression equation coefficients

Is the model adequate?

True

Model output

False

Correction of the initial set

False

Is it possible to complicate the model?

Figure 1. Algorithm for the forecast models
The smaller the scattering of empirical points around the line, the smaller the mean quadratic error of the equation. Thus, the $S_e$ value serves as an indicator of the significance and utility of the model expressing the relationship between the features [5].

Checking the adequacy of the model is one of the most important procedures for regression analysis, since it is necessary to make sure that the practical use of the obtained model will lead to positive results. When we select a model structure, we want it to be as simple as possible, i.e. it should include as few coefficients as possible. This is the so-called principle of the model efficiency. Reducing the number of coefficients facilitates both the evaluation procedure and the use of the model [6]. After building the model (regression equation), it must be evaluated and analyzed. An important point in evaluating the model is the measurement of bond tightness. The total variance

$$\sigma_{y}^{2} = \frac{\sum(y - \bar{y})^2}{n} \tag{6}$$

is a measure of how the actual (empirical) values of the feature $y$ vary near their average value. It characterizes the general variation of the effective characteristic $y$ explained by the influence of all the factors on which it depends. The deviations $y - \bar{y}$ are explained by the fact that the combination of the values of factors affecting $y$ in each unit of the population is individual, different. In addition to the total variance, on the basis of the regression equation and deviations $y_i - \bar{y}$, it is possible to calculate the average square of these deviations, i.e. the average square of deviations of theoretical values $y_i$ calculated from the regression equation from their average value (the average value $y_i$ is equal to the total average $\bar{y}$):

$$\sigma_{y_i}^{2} = \frac{\sum(y_i - \bar{y})^2}{n} \tag{7}.$$ 

This is a measure of how the theoretical values of the feature vary near the average value (factor variance). Based on the regression equation and deviations $y - y_i$ it is possible to calculate the average square of these deviations, i.e. the average square of the deviations of the actual values of the effective feature from its theoretical values obtained by substituting the corresponding values of the feature-factor into the regression equation:

$$\sigma_{\varepsilon}^{2} = \frac{\sum(y - y_i)^2}{n} \tag{8}.$$ 

This is a measure of fluctuations in actual values $y$ near the corresponding theoretical values, i.e. near the regression line (residual variance). In mathematical statistics

$$\sigma_{y}^{2} = \sigma_{y_i}^{2} + \sigma_{\varepsilon}^{2}. \tag{9}.$$ 

To measure the tightness of the relationship between $y$ and $x$, it is therefore logical to use the ratio of factor variance to the total variance of the effective feature. This ratio is called the theoretical determination index ($i^2$):

$$i^2 = \frac{\sigma_{y_i}^{2}}{\sigma_{y}^{2}}. \tag{10}.$$ 

The theoretical index of determination shows what part of the total variation of the effective feature-factor $y$ is explained by the feature-factor $x$ included in the corresponding regression equation.
The square root of the determination index, called the correlation index \((i)\), is also used as an indicator of bond tightness.

Regression and correlation indices – parameters of the regression equation, indices or coefficients of determination and correlation – calculated for constraints on the population size, may be distorted by random factors. Therefore, it is necessary to check whether the indicators are characteristic of the complex of conditions in which the studied population is located, whether they are the result of a combination of random circumstances. The significance (materiality) of the regression and correlation indicators is checked using the t-test (Student), the variance F-test (Fisher):

\[
t_0 = b_0 \sqrt{\frac{n-2}{\sigma^2}}, \quad t_i = b_i \sqrt{\frac{n-2}{\sigma^2}} \cdot \frac{\sum (x_i - \bar{x})^2}{n}; \quad F = \frac{\sigma^2_{xy}}{\sigma^2_{x}} \frac{n-m}{m-1},
\]

where \(m\) – number of parameters in the regression equation. The values \(t\) calculated from these formulas are then compared with their critical values at the received significance level and the number of degrees of freedom \(k=n-2\). The critical values \(t\) are found in the Student distribution table. The calculated value \(F\) is compared with the critical (tabular) value for the received significance level and degrees of freedom numbers \(k_1=m-1, k_2=n-m\).

Estimating the significance of regression coefficients using the \(t\) criterion is often used to complete the selection of factors during step analysis. Two procedures are best known that are implemented in application packages: sequential increase and sequential decrease in the group of independent variables [7]. For example, a sequential decrease is that after solving the model and estimating the significance of all regression coefficients, the factor at which the coefficient is insignificant and has the lowest confidence coefficient \(t\) is excluded from the model. After that, the model is solved again, and again the significance of all regression coefficients is estimated. If among them there are again insignificant values, then again the factor with the smallest coefficient \(t\) is excluded. The process of excluding factors continues until a regression equation is obtained, all coefficients in which are significant. Step-by-step regression is used to minimize the number of independent variables included in the studied model [8].

3. Results
In order to predict the efficiency of building materials by characteristics of resistance to loads, models are built that take into account the relationship of analyzed indicators. To use the regression analysis methods, the following prerequisites must be met:

- all experiments should be conducted independently of each other in the sense that the accidents that caused the deviation of the response from the pattern in one experiment did not affect such deviations in other experiments;
- the statistical nature of these random components remained unchanged in all experiments; the main reasons for the fact of inaccuracy of information on construction components include the following: inability to ensure the objectivity of assessments; difficulty or inability to quantify qualitative indicators; data writing errors;
- measures included in the regression equation as independent variables must be unrelated to each other.

In our case, the experiment refers to the design of the next data on the construction material, so the first two prerequisites of the regression analysis are carried out based on data collection technology. To achieve independence, risk factors need to be pre-selected, i.e. there is a need to exclude a number of parameters, and choose those that carry the minimum information. Based on the developed method, the following groups of interrelated indicators were formed: A) for stone natural building materials: \(X_1\) – true density, \(X_2\) – porosity, \(X_3\) – thermal conductivity, \(X_4\) – heat capacity, \(X_5\) – elasticity, \(X_6\) – residual deformation, B) for binding materials: \(X_1\) – true density, \(X_2\) – porosity, \(X_3\) – thermal conductivity, \(X_4\) – heat capacity, \(X_5\) – capillary suction, \(X_6\) – linear temperature expansion, C) for forest materials: \(X_1\) –
true density, $X_2$ – porosity, $X_3$ – thermal conductivity, $X_4$ – heat capacity, $X_5$ – humidity deformation, $X_6$ – hygroscopicity.

The calculations resulted in the following models:

$$Y_1 = 0.686298 - 0.00595939X_1 - 0.00534681X_2 + 0.0301896X_3 - 0.0511363X_4 + 0.00551003X_5 + 0.00499884X_6,$$

$$Y_2 = 0.647462 + 0.00439514X_1 + 0.0213804X_2 + 0.056029X_3 - 0.114879X_4 - 0.0314114X_5 + 0.108863X_6 + 0.105519X_7,$$

$$Y_3 = -0.056528 + 0.0402237X_1 + 0.0188449X_2 - 0.029715X_4 - 0.100122X_5 + 0.0979803X_6 + 0.029715X_7,$$

where $Y_1$ – stability of structures in conditions corresponding to latitude at the level of St. Petersburg; $Y_2$ – stability of structures in conditions of latitude at the level of Moscow; $Y_3$ – stability of structures in latitude conditions at the level of Rostov.

Based on the built models, it is possible to predict changes in design parameters when the selected corresponding characteristics change.

4. Conclusion

Thus, the analysis of the degree of influence of different characteristics on the stability of building structures in different conditions, which corresponded to different latitudes of construction facilities, was carried out. Based on the built models, it is possible to predict changes in the parameters of building materials, when the selected various characteristics will be changes.

References

[1] Anigbogu N A and Anunike E B 2014 Standard of Materials Specifications, their Implementation and Enforement on Building Construction Projects in Nigeria ATBU Journal of Environmental Technology 7(1) 33–44

[2] Oladapo A A 2007 An Investigation into the use of ICT in the Nigerian Construction Industry Journal of Information Technology in Construction 12 20–33

[3] Sorokin S O 2018 Optimization modeling of the functioning of the system of homogeneous objects in a multidimensional digital environment Modeling, optimization and information technologies 6(3(22)) 153–164

[4] Orlova D E 2018 Stability of solutions in ensuring the functioning of organizational and technical systems Modeling, optimization and information technologies 6(1(20)) 325–336

[5] Neittaanmäki P, Repin S and Tuovinen T (Ed.) 2016 Mathematical Modeling and Optimization of Complex Structures; Series: Computational Methods in Applied Sciences (Switzerland: Springer International Publishing AG)

[6] Rios, L M and Sahinidis N V 2013 Derivative-free optimization: a review of algorithms and comparison of software implementations Journal of Global Optimization 54 1247–1293

[7] Lvovich I Ya, Lvovich Ya E, Preobrazhenskiy A P, Preobrazhenskiy Yu P and Choporov O N 2019 Modelling of information systems with increased efficiency with application of optimization-expert evaluation Journal of Physics p 33079

[8] Bernard F and Gaffet E 2001 Mechanical alloying in SHS research Int. J. Self-Propag. High-Temp. Synth. 10(2) 109–132