Modeling of Cement Activity Increase by Dispersed Mineral Additives

L Ilina\(^1\) and I Mukhina\(^1,\ast\)

\(^1\)Novosibirsk State University of Architecture and Civil Engineering (Sibstrin), 113, Leningradskaya str., Novosibirsk, 630008, Russia
E-mail: \ast\) muhina.irina@rambler.ru

Abstract. Obtaining high-strength and ultra-high-strength concrete is a priority research area in construction sciences. Since cement is the main component responsible for the strength of cement conglomerates, increasing its activity is an urgent task. Authors considered the possibility of increasing the compressive strength of hardened cement by introducing dispersed mineral additives (diopside and limestone), which are mining wastes. Multiple regression models were proposed for various periods and modes of cement hardening that describe the dependence of cement strength on the composition and amount of mineral additives. The linear and non-linear regression coefficients were calculated using the least squares method. Statistical analysis of models was fulfilled. Equations’ verification included checking the adequacy of calculated values of cement strength to actually measured. In order to determine the mineral additive composition that would provide the greatest cement strength for corresponding hardening conditions, an optimization (maximization) procedure was performed for equations. In addition, the residuals of regression equations were investigated. The study confirmed that the dispersion of measurement errors was constant, so that the Gauss-Markov conditions were satisfied, and the factors included in the model (the amount of diopside and the amount of limestone) adequately describe the physical process of cement hardening. In order to visualize research results, surface plots and isoline plots for the dependence of cement strength on additives for linear and nonlinear models are presented in the article.

1. Introduction
The priority area for the development in the fields of architecture, urban planning and civil engineering scientific researches is to obtain high-strength and ultra-high-strength concrete. The main strengthening component is cement with the use of aggregates made of strong rocks in concrete. Moreover, no fundamental differences in the composition and structure of cement hydration products in hardened cement, mortar, or concrete were noted [1 - 7].

The use of dispersed mineral additives (mining waste) can significantly increase the strength characteristics of cement conglomerates (all types of mortars and concrete) by increasing the cement activity. In addition, in many cases, the introduction of such additives reduces binders’ expensiveness [8 - 13].

The objective of the current study was a regression model [14-18] that describe the dependence of cement strength on the composition and amount of mineral additives.

2. Mathematical and Methods
In order to assess the optimal amount of complex mineral additives, a regression model was proposed
Y=f(X_1, X_2)+\epsilon, \tag{1}

Where $X_1, X_2$ are the independent variables (factors), meaning the amount of limestone and diopside in the mineral additive, respectively; $Y$ is a dependent variable, in our case it is the cement strength ($Y$ is a random variable); $f(X_1, X_2)$ is unknown function of parameters $X_1, X_2$; $\epsilon$ is random model error. The function $f(X_1, X_2)$ should describe the dependence of cement strength on chemical composition and the amount of introduced mineral additives. The error $\epsilon$ in Eq. (1) is explained by a number of reasons: these are both possible errors in choosing the type of $f(X_1, X_2)$ function and measurement errors of $Y$ value that arise during the experiments. It is assumed that the measurements are equal, that is, all measurement errors have equal variances.

The proportion of complex additives’ in the total cement mass varied from 1 to 11%, while the ratio of limestone: diopside changed as 1: 2, 1: 1, 2: 1; in addition, limestone and diopside were used as separate additives during the experiments. In total, 30 pairs of values ($x_{ij}$) were compiled, which corresponded to 30 strength values $y_i, i$ is the number of observations, $i=1, 2,\ldots,n, n=30$ is the number of observations on the variables. Observation results were presented as cross-section data

\[
\begin{array}{ccc}
 x_{11} & x_{12} & y_1 \\
 x_{21} & x_{22} & y_2 \\
 \vdots & \vdots & \vdots \\
 x_{n1} & x_{n2} & y_n \\
\end{array}
\tag{2}
\]

where the column $x_{ij}$ indicates the portion of limestone in the cement mass, $x_{ij}$ is the portion of diopside, $y_i$ is the strength, obtained in the $i$ observation. To reduce a random error 24 measurements were performed for each pair of values. Thus, all values in matrix (2) were calculated as arithmetic means of 24 measurements ($k$ is the measurement number):

\[
x_{ij}=\frac{\sum_{k=1}^{24} x_{ij}^{(k)}}{24}, \quad y_i=\frac{\sum_{k=1}^{24} y_i^{(k)}}{24}, \quad i=1, n, \quad j=1, 2
\tag{3}
\]

The possibility to use a linear model was studied, since it is preferable for the production technology to use linear equations as the easiest to implement. Correlation coefficients were calculated using cross-section data (2) to determine a linear relationship

Based on the available data sampling with the volume $n=30$ using the least squares method, a linear multiple regression was suggested:

\[
\hat{y}(x_1, x_2)=b_0+b_1 x_1+b_2 x_2.
\tag{4}
\]

Unknown parameters were found from the condition of minimum functional

\[
\sum_{i=1}^{n}(y_i-\hat{y}_i)^2=(y-Xb)^T(y-Xb),
\tag{5}
\]

where $y_i$ is the strength values, observed in experiments; $\hat{y}_i$ is strength values, calculated by the regression equation; $b$ is the vector of unknown coefficients; $X$ is a matrix

\[
X=\begin{bmatrix}
1 & x_{11} & x_{12} \\
1 & x_{21} & x_{22} \\
\vdots & \vdots & \vdots \\
1 & x_{n1} & x_{n2}
\end{bmatrix}, \quad n=30.
\tag{6}
\]

The multicollinearity of the source data may become a problem in the process of multiple regressions constructing. $r_{x_1,x_2}=-0.29$ is the correlation coefficient, calculated by matrix (6). Such a value of the correlation coefficient suggests that multicollinearity is absent in the adopted model [19].

The lack of multicollinearity between the values of limestone and diopside additives indicates the stability of parameter values, calculated by the least squares method to changes in the initial data. The condition number of matrix (6) cond$(X^TX)=112.95$ is small.
A set of five equations was obtained, each of which corresponds to specific hardening mode – under normal conditions for 1, 3, 7, and 28 days and under heat-moisture treatment conditions (respectively, Eq. (7) - (11)): 

\[
\hat{y}_1(x_1, x_2) = 13.59 - 0.297x_1 + 0.047x_2 \quad (7)
\]

\[
\hat{y}_2(x_1, x_2) = 24.82 - 0.495x_1 + 0.079x_2 \quad (8)
\]

\[
\hat{y}_7(x_1, x_2) = 42.33 - 0.744x_1 + 0.288x_2 \quad (9)
\]

\[
\hat{y}_{28}(x_1, x_2) = 75.67 - 1.30x_1 + 0.56x_2 \quad (10)
\]

\[
\hat{y}_{tvo}(x_1, x_2) = 66.31 - 1.65x_1 + 0.68x_2 \quad (11)
\]

where \(x_1\) is the portion of limestone, \(x_2\) is the portion of diopside. Contour plots, constructed using linear multiple regressions (7) - (11) are shown in figure 1. The graphs analysis confirmed that additives increase cement compressive strength. \(x_1\) axis is the limestone portion, \(x_2\) axis is the diopside portion, along the \(y\) axis is the strength of the cement, MPa, hardened for: 1 day (a); 28 days (b).

![Figure 1. Contour plots of cement strength depending on the composition of additives. \(x_1\) axis is the limestone portion, \(x_2\) axis is the diopside portion, along the \(y\) axis is the strength of the cement, MPa, hardened for: 1 day (a); 28 days (b).](image)

In order to check the statistical significance of \(b_0, b_1, b_2\) coefficients, t-statistics were used:

\[
t_j = \frac{b_j}{s_{b_j}} \quad (12)
\]

Here \(s_{b_j}\) is the standard error for \(b_j\) appraisal:

\[
s_{b_j} = \sqrt{\frac{\text{explained variance}}{\text{unexplained variance}}} \quad (13)
\]

where \(s^2\) is an appraisal of variance for a vector \(y_i - \hat{y}_i\). The calculated values \(t_j\) were compared with the critical value \(t(1-\alpha/2,n-m)=2.052\) for \(\alpha=0.05\), \(n=30\), \(m=3\) (\(m\) is the number of unknown coefficients in the regression equation). The comparison showed that the condition was satisfied for all coefficients of Eq. (7) - (11). Therefore, we can conclude that all the coefficients of the proposed equations are statistically significant at \(\alpha=0.05\) significance level.

Examination of Eq. (7) - (11) statistical significance was checked using the Fisher criterion. The following value was calculated:

\[
F = \frac{\text{explained variance}}{\text{unexplained variance}} \quad (14)
\]
where the “explained variance” is \( \frac{1}{n-m} \sum_{i=1}^{n} (\bar{y} - \hat{y}_i)^2 \), the “unexplained variance” is \( \frac{1}{n-m} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \), \( n=30, m=3 \), \( \bar{y} \) - an average value of the strength vector. \( F \)-statistic (14) follows the F-distribution with degrees of freedom \( (m-1) \) and \( (n-m) \) under the null hypothesis. It was compared to a critical value \( F_{1-\alpha, m-1, n-m} \), where \( F_{1-\alpha, m-1, n-m} \) is an \( F \)-distribution quantile with the \( (1-\alpha) \) level and degrees of freedom \( (m-1) \) and \( (n-m) \). The calculations showed that at the significance level \( \alpha=0.05 \) all the equations of linear multiple, regression (7) - (11) are statistically significant.

In order to determine how well the regression predictions approximate the real data points the multiple coefficient of determination was calculated:

\[
R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}. \tag{15}
\]

\( R^2 \) is a statistical measure that gives some information about the goodness of fit of a model. It is known that adding of variables increases the "complexity" of the model. This leads to the alternative approach of looking at the adjusted coefficient of determination:

\[
\hat{R}^2 = 1 - \frac{n-1}{n-m} (1 - R^2). \tag{16}
\]

The explanation of this statistic is almost the same as \( R^2 \) but it penalizes the statistic as extra variables are included in the model. Table S1 (Supplementary Information) shows the calculated values of coefficient of determination \( R^2 \) and the adjusted coefficient of determination \( \hat{R}^2 \) for Eq. (7) - (11). Table S1 shows the low level of the linear model’s accuracy, so the next stage of the research was the construction of a nonlinear multiple model [18].

**Table 1.** Determination coefficient and the adjusted determination coefficient for Eq. (7) - (11).

| Mode of cement hardening           | \( R^2 \) | \( \hat{R}^2 \) |
|-----------------------------------|----------|----------------|
| 1 day, normal conditions         | 0.484    | 0.446          |
| 3 days, normal conditions        | 0.447    | 0.406          |
| 7 days, normal conditions        | 0.395    | 0.351          |
| 28 days, normal conditions       | 0.575    | 0.543          |
| heat and humidity treatment      | 0.693    | 0.671          |
The nonlinear model was constructed as a parabolic multiple regression:

$$\hat{y}(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2. \quad (17)$$

A feature of this equation is the pair interaction coefficient, which allows one to take into account the mutual influence of two additives.

unknown coefficients were calculated in order to determine the function (17), based on the least squares method. A set of five equations was obtained. Each equation corresponded to certain hardening mode – under normal conditions for 1, 3, 7, and 28 days and under conditions of heat-moisture treatment (respectively, Eq. (18) - (22)):

$$\hat{y}_1(x_1, x_2) = 12.56 - 0.057 x_1 + 0.823 x_2 - 0.01 x_1^2 - 0.07 x_2^2 - 0.097 x_1 x_2. \quad (18)$$

$$\hat{y}_3(x_1, x_2) = 21.72 + 1.616 x_1 + 0.981 x_2 - 0.16 x_1^2 - 0.072 x_2^2 - 0.296 x_1 x_2 \quad (19)$$

$$\hat{y}_7(x_1, x_2) = 37.295 + 1.311 x_1 + 3.199 x_2 - 0.158 x_1^2 - 0.254 x_2^2 - 0.41 x_1 x_2 \quad (20)$$

$$\hat{y}_{tvo}(x_1, x_2) = 62.5 - 0.29 x_1 + 3.07 x_2 - 0.1 x_1^2 - 0.21 x_2^2 - 0.34 x_1 x_2 \quad (21)$$

$$\hat{y}_{28}(x_1, x_2) = 68.33 + 2.031 x_1 + 4.468 x_2 - 0.268 x_1^2 - 0.332 x_2^2 - 0.61 x_1 x_2 \quad (22)$$

Figure 2 shows the surface plots and contour plots, constructed with the use of nonlinear multiple regression Eq. (18), (20)-(22). The area of greatest strength is highlighted in red. The plots clearly demonstrate ways to increase the strength of cement through the additives introduction.

The examining of statistical significance of coefficients $b_j$, $j=0, 1, ..., 5$ in Eq. (18) - (22) was checked using formula (12). The calculated values $t_j$ were compared with the critical value $t(1-a/2, n-m) = 2.064$ for $a=0.05$, $n=30$, $m=6$. The comparison showed that $|t_j| > t(1-a/2, n-m)$ condition was satisfied for all coefficients in Eq. (18) - (22). Therefore, the hypothesis is accepted at the significance level of 0.05, that all the coefficients are significant and non-zero.

| Mode of cement hardening | $R^2$ | $\tilde{R}^2$ |
|--------------------------|-------|-------------|
| 1 day, normal conditions | 0.658 | 0.587       |
| 3 days, normal conditions| 0.607 | 0.526       |
| 7 days, normal conditions| 0.681 | 0.615       |
| 28 days, normal conditions| 0.838 | 0.805       |
| heat and humidity treatment| 0.756 | 0.705       |

Examining the significance Eq. (18) - (22) according to Fisher’s test showed that Eq. (18) - (22) of nonlinear multiple regressions are statistically significant at $\alpha = 0.05$ significance level.

Table 2 shows the calculated values of $R^2$ and $\tilde{R}^2$ for Eq. (18) - (22). The calculation data indicate a greater accuracy of the nonlinear equation in comparison with the linear model.

3. Results

Figure 2 shows how the amount of limestone and diopside affects the strength of cement and how the strength varies depending on the hardening time. It is possible to choose a ratio of additives and their total amount, which allows to obtain the greatest strength for a given hardening mode by changing the composition of the additives.

The optimal amount of additives and their ratio can be determined by solving the optimization problem (maximization) of regression Eq. (18) - (22). The conditions are non-negativity $x_1$ and $x_2$, as well as a restriction on the total amount of added additives $x_1 + x_2 \leq 11\%$. For example, maximization
of functional (22) corresponding to the mode of heat-moisture treatment gave the following results: 

\[ x_1 = 1 \quad x_2 = 6.54, \text{ corresponding strength } y_{\text{tv}} = 71.1 \text{ MPa}. \]

Calculations were performed for different hardening times and different mixture compositions in order to verify the model. For example, 1 day hardening, the ratio of limestone: diopside is 1: 1, 1% in total, the strength according to the equation is 12.9 MPa, and 13.0 MPa is from the experiment (difference less than 1%). 7 days hardening, the ratio of limestone: diopside is 1: 2. 5% in total, the strength according to the equation is 44.56 MPa and 44.3 MPa is from the experiment (difference less than 1%). 28 days hardening, the ratio of limestone: diopside is 2: 1. 7% in total, the strength according to the equation is 73.94 MPa, 70.8 MPa is from the experiment (1.8% difference).

In order to verify the adequacy of the model, the following tasks were set: to evaluate the validity of assumptions about equally accurate measurements (Gauss-Markov conditions) and using two factors (diopside amount and limestone amount) in the regression equation and is it sufficient to use only two factors (the amount of diopside and the amount of limestone) in regression.

The residuals for nonlinear models (18) - (22) were calculated solving these problems (a residual is the difference between an observed value and the fitted value, provided by a model) [23]:

\[ e_i = y_i - \hat{y}_i, \quad i = 1, 2, \ldots, n \]  

a quantity, having a t-distribution with \( k = n - m \) degrees of freedom:

\[ d_i = \frac{e_i}{\sqrt{s^2 / \sum_{i=1}^{n} e_i^2}}, \]

Where \( s^2 = \frac{1}{n-m} \sum_{i=1}^{n} e_i^2 \) – variance estimation for the model error in the OLS, \( p_{ii} \) – diagonal matrix elements \( P = \bar{X} (\bar{X}^T \bar{X})^{-1} \bar{X}^T \)

\[ \bar{X}^T = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \end{bmatrix}. \]  

The calculated values \( d_i \) were compared with a critical value \( t(\gamma, n-m) = 2.064 \) for confidence probability \( \gamma = 0.95 \), \( n = 30 \), \( m = 6 \). If the condition \( |d_i| < t(\gamma, k) \) is satisfied for all \( d_i \) values, we can talk about equally accurate measurements and the goodness of regression model.

Figure 3 shows a general graph of the residues for the hardening mode under normal conditions for 28 days and a graph of the corresponding statistical \( d_i \) values. Evidently, that all \( d_i \) values fell into the interval \( (-t(\gamma, k), t(\gamma, k)) \). Therefore, we can conclude that the dispersion of measurement errors is constant and the factors included in the model adequately describe the rate of cement hardening.
4. Conclusion
When complex mineral additives containing several components are introduced into the composition of cement materials, they can influence each other, enhancing or weakening the action of individual components. Thus, the authors studied the influence of individual mineral additives (diopside and limestone) and their complex on increasing the cement strength and established factors affecting this process.

Statistical analysis and verification of a non-linear multiple model confirmed the adequacy of the model: all regression coefficients are significant at the accepted level; all equations are significant; Gauss-Markov conditions are met; model specification is consistent with experimental data. It is possible to choose a composition of mineral additives that would provide the greatest strength of the cement matrix, solving the optimization problem with an equation corresponding to a certain mode of cement hardening.

References
[1] Akhverdieva T A, Dzhafarov R 2019 Construction materials, 3 pp 73-76 doi:10.31659/0585-430X-2019-768-3-73-76
[2] Vernigorova V N, Sadenko S M 2017 Construction Materials, 1-2, pp 86-89 doi:10.31659/0585-430X-2017-745-1-2-86-89
[3] Bazhenov Yu M, V.R. Falikman V R, B.I. Bulgakov B I 2012 Vestnik MGSU, 12, pp 125-133 doi:10.22227/1997-0935.2012.12.125-133
[4] Ahtyamov V F, Hafizova E N 2018 The influence of non-metallic waste on the properties of heavy concrete, The Russian automobile and highway industry, J. 2(15), pp 261–268 doi:10.26518/2071-7296-2018-2-261-268
[5] Berra M, F. Carassiti F, T. Mangialardi T, A.E. Paolini A E 2012 Construction and Building Materials, 35, pp 666–675 doi:10.1016/j.conbuildmat.2012.04.132
[6] Nadeem S A, Memon T Y Lo 2013 Construction and Building Materials, 38, pp 338-347
[7] V.V. Belov V V, Yu.Yu. Kuryatnikov Yu Yu 2018 Gas concrete of non-autoclave soliding based on reactive powder dry mixtures, Building Sciences, pp 63-70 doi:10.22337/9785432302663-63-70 (2018)
[8] Ilina L V, Mukhina I N, Teplov A A 2016 Dry building mixture with complex dispersed mineral additives, AIP Publishing. Advanced Materials in Technology and Construction, 1698 070001. doi:10.1088/1757-899X/451/1/070001
[9] Ilina L V, Tulyaganov A K 2018 Effect of dispersed mineral additive on properties of mortar mixture and solution, Materials Science and Engineering, 012031 doi: 10.1088/1757-899X/451/1/012031
[10] Stefanidou M, Papayianni I 2012 Composites, 43(6), pp 2706-2710 doi:10.1016/j.compositesb.2011.12.015
[11] Berdov G I, Ilina L V, Mukhina I N 2015 Aerated concrete with mineral dispersed reinforcing additives. Materials Science and Engineering, 71, 012017 doi:10.1088/1757-899X/71/1/012017
[12] Ilina L V, Mukhina I N 2016 Estimation of the Applicability for the Filler Produced by Recycling of Concrete and Reinforced Concrete Used in Heavy Concrete. Procedia Engineering, 150, pp. 1525 – 1530. doi: 10.1016/j.proeng.2016.07.102
[13] Ilina L, Mukhina I, Khakimullina S 2017 Increasing the mechanical strength of fine–grained concrete by introducing finely dispersed mineral additives. AIP Conference Proceedings 1800 020009 doi: 10.1063/1.4973025
[14] Yan X 2009 Linear Regression Analysis: Theory and Computing, World , 1-2, ISBN 9789812834119
[15] Berk R A 2007 Regression Analysis: A Constructive Critique, Criminal Justice Review. 32(3) 301-302. doi:10.1177/0734016807304871.
[16] Freedman D A 2009 Statistical models: Theory and Practice, Cambridge University Press
[17] Bühlmann P 2011. Statistics for high-dimensional data: methods, theory and applications, Springer
[18] Aldrich J 2005 Statistical Science, 20(4), pp 401-417 doi:10.1214/08834230500000331.
[19] Reducing Structural Multicollinearity
[20] Timm N H 2007 Applied Multivariate Analysis, Springer Science & Business Media p 107
[21] Warne R T 2011 Beyond multiple regression: Using commonality analysis to better understand R2 results, Gifted Child Quarterly, 55(4) pp 313-318 doi:10.1177/0016986211422217
[22] Good P I, Hardin J W 2009 Common errors in statistics (and how to avoid them) (3rd ed.). Hoboken, New Jersey Wiley, pp 211