УДК: 691.3; 519.6

ПАРНАЯ РЕГРЕССИОННАЯ МОДЕЛЬ ПРОГНОЗА ПРОЧНОСТИ БЕТОНА В ПРОИЗВОДСТВЕННЫХ УСЛОВИЯХ

Ключевые слова: производство бетонных смесей, контроль качества, прогнозирование
Key words: concrete mix optimization, filler, industry by-products, slag

Р.О. РЕЗАЕВ, канд. физ.-мат. наук, генеральный директор ООО «Проектирование материалов», г. Москва, доцент ТПУ, г. Томск, Россия;
А.А. ДМИТРИЕВ, главный технолог ООО «Проектирование материалов», г. Москва, Россия

В статье обсуждается возможность прогнозирования свойств бетона на примере его прочности в производственных условиях. Предлагается для этих целей использовать регрессионные модели, простейшей из которых является модель парной регрессии. Используя выборку из данных производственного контроля объемом 890 элементов в работе построена модель регрессии прочности на различные сроки твердения с производственными значениями цементно-водного отношения. Показано, что даже в условиях большой зашумленности контролируемых показателей, регрессионная парная модель обеспечивает удовлетворительный коэффициент детерминации (более 0,5) для прогноза прочности на первые сутки твердения после режима ТВО. На более длительные сроки твердения прогноз по одному только фактору не представляется возможным (коэффициент детерминации менее 0,4), что обуславливает необходимость применения моделей множественной регрессии.

The article discusses the possibility of predicting the properties of concrete using the example of its strength in production conditions. It is proposed to use regression models for these purposes, the simplest of which is the paired regression model. Using a sample of production control data with a volume of 890 elements, a regression model of strength for various hardening periods with production values of the cement-water ratio was built. It is shown that even in conditions of high noise in the controlled indicators, the regression pair model provides a satisfactory coefficient of determination (more than 0.5) for predicting the strength on the first day of hardening after the HPT regime. For longer curing periods, a forecast based on one factor alone is not possible (the coefficient of determination is less than 0.4), which necessitates the use of multiple regression models.

Введение

Контроль качества производства бетонных смесей осуществляется посредством сбора различных показателей технологического процесса и характеристик бетонной смеси с последующим их сравнением с допустимыми значениями из нормативных диапазонов. При этом идеальным сценарием для контроля качества является тот, при котором по собранным показателям можно было бы осуществлять прогноз целевых характеристик/свойств бетона. Математическое моделирование представляет для решения данной проблемы неоценимо полезный инструмент. Используемые в практике производства модели можно условно разделить на два класса: основанные на физико-химических закономерностях, и регрессионные, эмпирические.

Первые имеют большую ценность, так как помогают выявить причинно-следственные связи, которые справедливы независимо от конкретной производственной или лабораторной площадки или сырьевой базы, однако разработать и верифицировать такие модели чрезвычайно сложно.

Второй класс моделей (регрессионные) не выявляет фундаментальной связи между причиной и следствием, а только устанавливает соответствующий количественный факт.

Такие модели разрабатываются на базе некоторой выборки данных, поэтому их применимость ограничена условиями, в которых эта выборка была получена. Как правило, данные условия описываются конкретной производственной площадкой и ее сырьевой базой. Другими словами, регрессионные модели не имеют обобщающей силы, поэтому те модели из этого класса, которые эффективно работают на одном производстве, могут оказаться бесполезными на другом. Естественно, в этом случае возникает вопрос, а зачем вообще нужны такие модели? Ответ заключается в их простоте – для построения регрессионных моделей нужна лишь некоторая историческая выборка данных и простейший редактор табличных данных со встроенными инструментами обработки данных.

Важным преимуществом регрессионных моделей помимо их простоты также является возможность их быстрой корректировки по мере поступления новой информации – это осуществляется посредством перерасчета входящих в регрессионную модель коэффициентов.

Производство бетона/бетонных смесей сопряжено с большим количеством различных факторов, влияющих на целевые свойства выпускаемой продукции. В зависимости от рассматриваемого свойства актуальными могут быть те или иные факторы. В данной работе в качестве целевого свойства рассматривается прочность, точнее ее значения на 1-е, 7-е и 28-е сутки. Общепризнанным фактом является то, что прочность бетона во многом определяется цементно-водным отношением [1]. Однако следует отметить, что данный факт имеет определенную область применимости – стандартные сырьевые материалы тяжелого бетона [2-4]: щебень/гравий, песок, портландцемент, вода и пластифицирующие добавки, а также нормальные условия твердения.

Когда речь идет о прочности на 1-е сутки, как правило, подразумевается режим тепло-влажностной обработки, который, естественно, отличается от нормальных условий твердения. Во многих практических случаях взаимосвязь прочности с цементно-водным (Ц/В) отношением положена в основу подбора состава бетона – соотношение формализовано в виде линейной зависимости, коэффициенты которой определяются, исходя из серии экспериментальных замесов. Ориентируясь на данную зависимость, возникает закономерный вопрос, насколько эффективно (в количественном смысле) можно рассматривать Ц/В как технологический параметр, подлежащий контролю, с целью прогнозирования прочности бетона на заданный период твердения. Подчеркнем, что речь идет о производственных условиях, где в силу присутствия гораздо большего количества разнообразных факторов по сравнению с лабораторными, может оказаться, что вариативность и нестабильность свойств сырьевых материалов ослабит предиктивный потенциал фактора Ц/В для прочности. В таких случаях постановка проблемы прогнозирования влечет за собой выявление релевантных факторов производства, от которых зависят свойства изделий. Например, вполне реальна ситуация, когда для контроля прочности после режима ТВО необходимо наряду с Ц/В использовать также контроль доли крупной фракции заполнителя, формирующего каркас материала и играющего критическую роль в прочности на ранних стадиях. Для того чтобы эффективно решать обозначенные задачи, необходимо собирать разнообразные производственные данные. На первый взгляд, может показаться, что это является дополнительным обременением для инженерно-технического персонала производства, однако в действительности большинство необходимых для прогноза данных фактически собирается на производстве в соответствии с текущими нормативными требованиями. В этом контексте большое значение имеет информация о расходах материалов на единицу объема бетонной смеси, показателях плотности смеси, осадки конуса и т.д. Все эти факторы можно использовать для построения регрессионной модели и выявить, на какой минимально необходимый набор параметров нужно обращать внимание с целью прогнозирования прочности (или какого-либо другого свойства). Следует оговориться, что итоговая прочность бетона в конструкции определяется не только расходом сырьевых материалов (рецептурой), но также существенный вклад вносит уход за твердеющим бетоном. В этой связи контроль качества и прогноз свойств бетона позволят разделить ответственность между производителем бетонной смеси и подрядчиками, эксплуатирующими бетонную смесь. В случае производства готовых изделий из бетона (ЖБИ) вопрос о разделении ответственности не возникает, тем не менее, прогноз и управлением браком также и в этом случае представляют интерес.

В данной работе на примере одной из производственных площадок в РФ показано и обосновано, что одного только фактора Ц/В недостаточно для надежного прогнозирования прочности бетона на 1-е, 7-е и 28-е сутки. На основании регрессионной модели, связывающей значения прочностей R_7 и $R_{28,}$ оценен порядок точности построения регрессионных моделей, что дает ориентир для их оценки. В работе используется парная регрессионная модель – с целью демонстрации ее возможностей. Модель множественной регрессии, как более мощный вариант построения прогнозов с учетом большого количества факторов, будет рассмотрена в продолжении настоящей работы.

Модель парной регрессии

В этом разделе максимально детально представлена модель парной регрессии с целью ознакомления с процедурой обработки данных специалистами, не обладающими специальными знаниями в этой области. Хотя излагаемый материал и общеизвестный, его изложение во многих источниках [5] зачастую ограничивается формальными математическими выкладками без иллюстративных примеров. Рассмотрим задачу – имеется набор данных (таб.1) по цементно-водному отношению Ц/В = [250, 275, 300, ..., 500]/175 и соответствующие значения прочности R_{28} = [13,8; 22,6; 27,4; 30,2; 32,1; 44,1; 37,7; 46,6; 48,8; 54,8; 61,5] МПа. В соответствии со схемой модели будем предполагать, что прочность прямопропорциональна Ц/В:

$$R_{расч} = a + b \times \frac{Ц}{В}. \qquad (1)$$

Это соотношение представляет ничто иное, как формулу Боломея-Скрамтаева [6]. В уравнении (1) неизвестными величинами являются коэффициенты a,b, которые определяются из экспериментальных данных методом наименьших квадратов. Экспериментальные данные никогда не будут строго описываться прямой линией типа (1) – всегда будет присутствовать случайное отклонение ε:

$$\varepsilon = R_{эксп} - R_{расч}, \qquad (2)$$

Величину ε называют остатками регрессии. При осуществлении поиска значений коэффициентов a,b по экспериментальным данным естественной идеей кажется использование критерия минимальности отклонений ε расчетных значений от экспериментальных. При этом нужно иметь в виду, что ε может быть как положительным, так и отрицательным, следовательно, критерий минимальности нужно ставить на абсолютные значения отклонений, т.е. взятые без учета знака. Однако для получения «хороших» статистических свойств оценок коэффициентов a,b обычно используют не критерий минимизации абсолютных значений отклонений, а критерий минимальности квадратов отклонений [5]:

$$\varepsilon_1^2 + \varepsilon_2^2 + \cdots + \varepsilon_n^2 \to \min_{a,b}, \qquad (3)$$

где n – количество экспериментальных значений. Поиск значений коэффициентов a,b, удовлетворяющих критерию (3) и называется методом наименьших квадратов (МНК, ordinary least squares – OLS). Чтобы решить задачу минимизации (3), сперва выражаются отклонения в соответствии с уравнением (2) и (1):

$$\sum_{i=1}^{n}\varepsilon_i^2=\sum_{i=1}^{n}\left(R_i-a-b\times\left(Ц/В\right)_i\right)^2\rightarrow\min_{a,b}, \quad (4)$$

где R_i соответствует экспериментальным данным. Далее к уравнению (4) применяются стандартные математические методы поиска экстремума (минимума/максимума) – дифференцирование по параметрам a,b и приравнивание к нулю полученных выражений. Далее можно в явном виде выразить соотношения для расчета искомых коэффициентов:

$$a=R_{эксп}-\frac{cov\left(\frac{Ц}{В},R_{эксп}\right)}{var\left(\frac{Ц}{В}\right)}\times\frac{Ц}{В},$$

$$b=\frac{cov\left(\frac{Ц}{В},R_{эксп}\right)}{var\left(\frac{Ц}{В}\right)}, \quad (5)$$

$$\text{где } cov(x,y)=\frac{1}{n}\sum_{i=1}^{n}(x_i-x)(y_i-y),$$

$$x=\frac{x_1+x_2+\cdots+x_n}{n};\ var(x)=\frac{1}{n}\sum_{i=1}^{n}(x_i-x)^2$$

(для y расчетные формулы для среднего и дисперсии аналогичные). Проанализировав входящие в (5) величины, можно заметить, что в них входят только величины $Ц/В,(Ц/В)^2$ и $(Ц/В)\times R_{эксп}$, что позволяет эти формулы легко записывать в табличный редактор Excel. В качестве примера в табл. 1 рассчитаны значения коэффициентов a,b по вышеприведенным данным (отметим, что истинные значения a,b равны -25 и 30, соответственно).

Помимо нахождения коэффициентов a,b необходимо также оценивать, насколько найденная зависимость хорошо описывает наши экспериментальные данные. Например, на рис. 1 приведен график регрессии с коэффициентами из таб. 1. При этом одинаковые коэффициенты получаются как для набора данных, представленных на панели слева, так и для набора данных на панели справа. Т.е. одна и та же зависимость описывает различные наборы данных. При этом очевидно, что для панели слева модель описывает данные более качественно, чем на панели справа. Таким образом, зная только численные значения a,b оценить качество модели нельзя.

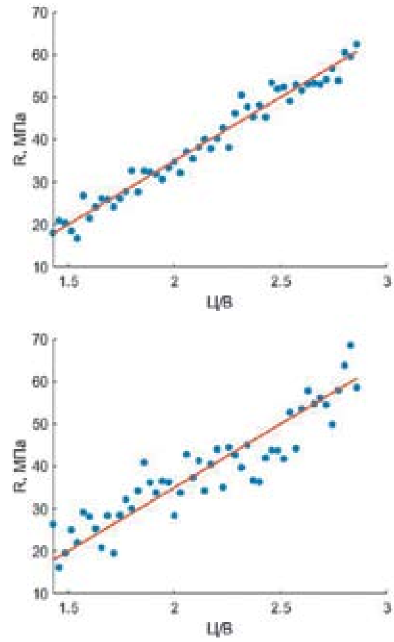

Рис. 1. Графическое представление регрессионной модели для различных наборов данных. Модель описывает данные на ерхней панели лучше, чем на нижней

Для оценки качества модели представим экспериментальное значение, как сумму расчетного и некоторой случайной добавки:

$$R_{эксп}=R_{расч}+\varepsilon. \quad (6)$$

Далее рассчитаем дисперсию экспериментального значения в соответствии с этой формулой:

ТАБЛИЦА 1. ПРИМЕР РАСЧЕТА КОЭФФИЦИЕНТОВ ПАРНОЙ РЕГРЕССИОННОЙ МОДЕЛИ

$Ц/В$	1,4286	1,5714	1,7143	1,8571	2,0	2,1429	2,2857	2,4286	2,5714	2,7143	2,8571
$R_{эксп}$	13,8	22,6	27,4	30,2	32,1	44,1	37,7	46,6	48,8	54,8	61,5
$Ц/В$	2,1429										
$R_{эксп}$	38,2										
$cov((Ц/В), R_{эксп})$	6,11										
$var(Ц/В)$	0,204										
a	-26										
b	30										

$$var\left(R_{эксп}\right) = var\left(R_{расч} + \varepsilon\right) =$$
$$= var\left(R_{расч}\right) + var\left(\varepsilon\right) + 2cov\left(R_{расч}, \varepsilon\right) =$$
$$= var\left(R_{расч}\right) + var\left(\varepsilon\right), \qquad (7)$$

где учтено, что $cov(R_{расч}, \varepsilon)=0$ в силу свойств метода наименьших квадратов. Расписав входящие в уравнение (7) величины с использованием определения дисперсии, получим выражение:

$$\frac{1}{n}\sum_{i=1}^{n}\left(R_{эксп} - R_{эксп}\right)^2 =$$
$$= \frac{1}{n}\sum_{i=1}^{n}\left(R_{расч} - R_{расч}\right)^2 + \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_i - \varepsilon\right)^2,$$

которое после сокращения на $1/n$ примет вид:

$$\sum_{i=1}^{n}\left(R_{эксп} - R\right)^2 = \sum_{i=1}^{n}\varepsilon_i^2 + \sum_{i=1}^{n}\left(R_{расч} - R\right)^2, \quad (8)$$

где учтено $R_{эксп}=R_{расч}=R$ и $\varepsilon=0$.

Интерпретация выражения (8) состоит в следующем: общая сумма квадратов отклонений экспериментальных значений от среднего равна аналогичной сумме отклонений расчетных от того же среднего и сумме квадратов случайных отклонений. Следовательно, чем лучше модель объясняет экспериментальные данные, тем меньше будет сумма квадратов случайных отклонений. Иными словами, отношение $\sum\left(R_{расч} - R\right)^2 / \sum\left(R_{эксп} - R\right)^2$ будет ближе к 1 для более качественных моделей. Данную величину принято называть коэффициентом детерминации – обозначается как R^2 (не путать с прочностью!). Свойства метода наименьших квадратов позволяют не только оценить коэффициенты a,b, но также и получить явный вид оценки их неопределенности:

$$var\left(a\right) = \frac{1}{n}\frac{S^2\sum\left(Ц/В\right)^2}{\sum\left(\left(Ц/В\right)_i - Ц/В\right)^2},$$
$$var\left(b\right) = \frac{S^2}{\sum\left(\left(Ц/В\right)_i - Ц/В\right)^2},$$
$$S^2 = \frac{1}{n-2}\sum_{i=1}^{n}\varepsilon_i^2.$$

Квадратный корень из дисперсии называется стандартной ошибкой и используется для построения доверительного интервала. Как, используя модель регрессии, проводить тестирование гипотез на значимость фактора (в текущих примерах обсуждаемым фактором является – Ц/В)? Допустим, что требуется ответить на вопрос: «Верно ли, что фактор Ц/В не влияет на прочность бетона?», что эквивалентно вопросу – «Равен ли в регрессии $R_{расч} = a + b \times \left(Ц/В\right) + \varepsilon$ коэффициент b нулю?». Сами по себе найденные количественные значения a,b, построенные по наблюдаемым данным на основе случайной выборки, не дадут ответа на этот вопрос, поскольку также являются случайными величинами. Исходя из этого, необходимо определять, достаточно ли сильно b отличается от нуля для того, чтобы можно было с уверенностью утверждать, что и истинное значение коэффициента b также не равно нулю. Для этого по оценкам b находят значения тестовой статистики $b/\sqrt{var\left(b\right)}$ и выбирают уровень значимость α – вероятность отклонить тестируемую гипотезу при условии, что она верна. Чаще всего используется $\alpha=0,05$. Из таблиц распределения Стьюдента [5] находят критическое значение тестовой статистики t_{n-2}^{α} для выбранного уровня значимости и числа степени свободы (n–2). Далее, если $\left|b/\sqrt{var\left(b\right)}\right| > t_{n-2}^{\alpha}$, то b достаточно велик по абсолютной величине и следует отвергнуть гипотезу, что b=0. В противном случае, исследуемый фактор признают статистически незначимым при уровне значимость α. Наконец, доверительный интервал для коэффициента b имеет следующий вид:

$$\left(b - \sqrt{var\left(b\right)} \times t_{n-2}^{\alpha}, b + \sqrt{var\left(b\right)} \times t_{n-2}^{\alpha}\right),$$

который содержит истинное значение b с вероятностью $1-\alpha$.

Построение и анализ регрессионной парной модели на реальных данных

Представленная в предыдущем разделе модель регрессии в данном разделе применена к реальным производственным данным, – накопленным за год наблюдения за прочностью изделий из железобетона на одной из производственных площадок. Весь объем выборки (по всем используемым на производстве составам) состоит из 890 значений прочности на 1-е, 7-е и 28-е сутки твердения (R_1, R_7, R_{28}) и соответствующих значений Ц/В. Значения R_1, R_7, R_{28} соответственно находятся в диапазонах 11-45, 20-57 и 21-62 МПа, т.е. покрывают практически весь диапазон используемых

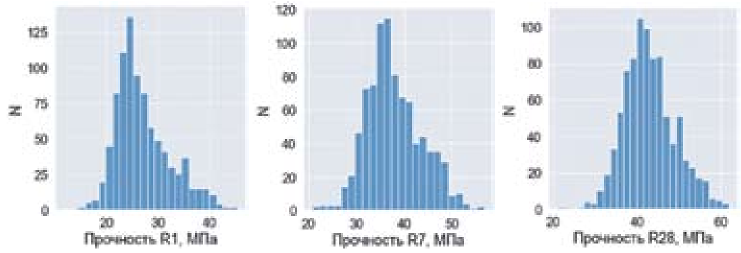

Рис. 2. Распределение производственных значений прочности на различные сроки твердения. Выборка взята по всем составам.

в производственной практике большинства производств бетона. Диапазон изменения значений Ц/В – от 2,25 до 3,5, что также покрывает значимое в практике распределение значений. Таким образом, исходные условия для построения регрессионной парной модели прочности от Ц/В вполне удовлетворительные. На рис. 2 представлены частотные диаграммы разбросов значений прочности.

Из общефизических рассуждений ожидается, что наилучшая корреляция будет для данных между значениями прочности на 28-е и 7-е сутки твердения. Более слабой будет корреляция между значениями на 28-е и 1-е, а также на 7-е и 1-е сутки твердения. Следуя схеме, изложенной в предыдущем разделе, была построена регрессионная модель (см. рис. 3) между этими данными и рассчитан коэффициент детерминации:

$$R_{28} = 9,6971 + 0,88 \times R_7 \left(R^2 = 0,680\right),$$
$$R_{28} = 19,3842 + 0,8732 \times R_1 \left(R^2 = 0,564\right),$$
$$R_7 = 14,0485 + 0,8793 \times R_1 \left(R^2 = 0,650\right).$$

Результаты построения регрессионной модели между прочностями демонстрируют неплохую корреляцию. Можно ожидать, что уровень коэффициента детерминации ~ 0,7 соответствует некоторому эталонному значению.

Далее была построена регрессионная модель для выявления связи между прочностями R_1, R_7, R_{28} и фактором Ц/В. Получены следующие соотношения (см. рис. 4):

$$R_1 = 1,6872 + 9,7469 \times Ц / В \left(R^2 = 0,524\right),$$
$$R_7 = 14,654 + 8,9091 \times Ц / В \left(R^2 = 0,368\right),$$
$$R_{28} = 20,6476 + 8,5922 \times Ц / В \left(R^2 = 0,301\right).$$

В отличие от корреляции значений прочности друг с другом, регрессионные модели для связи прочности с фактором Ц/В показывают не очень высокие значения коэффициента детерминации. Успешной можно считать корреляцию между значениями прочности на первые сутки твердения (после режима ТВО) – коэффициент детерминации равен 0,524. При таком коэффициенте детерминации, в принципе, можно использовать полученную модель с целью оценки прогнозного значения прочности. Однако для прогноза значений прочности R_7 и R_{28} полученные регрессионные модели будут неудовлетворительные в силу низкого коэффициента детерминации. Интерпретация данного факта состоит в том, что необходимо помимо Ц/В включать и другие факторы в рассмотрение и строить более сложную модель множественной регрессии или привлекать модели машинного обучения [7].

Заключение

Сбор информации об операционных значениях как параметров технологического процесса, так и различных производственных факторов, обеспечивает перспективы

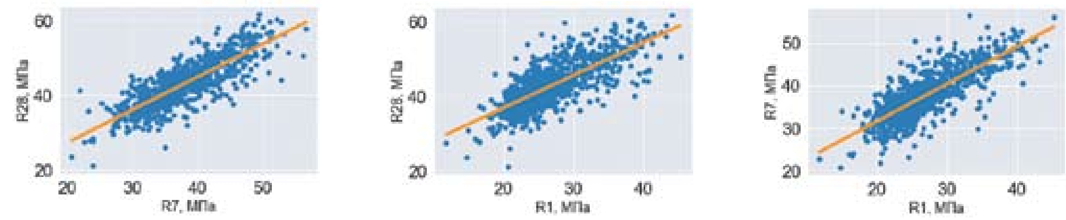

Рис. 3. Модели парной регрессии между значениями прочности на различные периоды твердения.

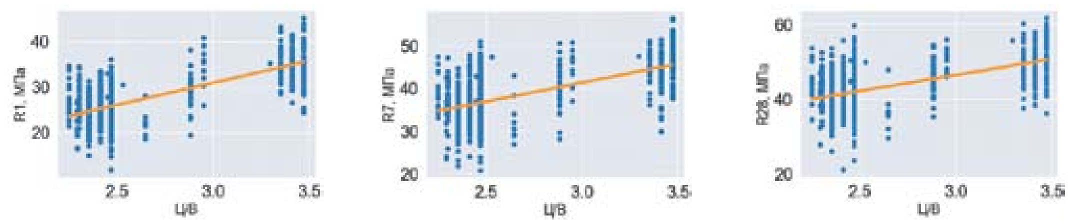

Рис. 4. Модели парной регрессии между значениями прочности и фактором Ц/В.

прогнозирования свойств бетона, в частности, его прочности на различные сроки твердения. Инструментом такого прогноза могут стать различные регрессионные модели, простейшим вариантом которых является парная модель, связывающая какой-либо один интересующий фактор и целевое свойство. Как показано в работе, несмотря на большое количество разнообразных факторов, для прочности на первые сутки твердения можно построить удовлетворительную регрессионную модель (с коэффициентом детерминации выше 0,5), ограничиваясь лишь одним фактором цементно-водного отношения. Ожидается, что прогнозной потенциал моделей множественной регрессии будет значительно выше по сравнению с моделями парной регрессии.

Библиографический список:

1. *Невилль А.М., Свойства бетона. – М.: Стройиздат, 1972 – 344 с.*
2. *ГОСТ 25192-2012. Бетоны. Классификация и общие технические требования.*
3. *ГОСТ 7473-2010. Смеси бетонные. Технические условия.*
4. *ГОСТ 26633-2015. Бетоны тяжелые и мелкозернистые. Технические условия.*
5. *Дрейпер Н., Смит Г., Прикладной регрессионный анализ. Множественная регрессия. – Изд-во Диалектика, 2017. – 912 с.*
6. *Баженов Ю.М., Технология бетона. – М.: Изд-во АВС, 2003. – 499 с.*
7. *Р.О. Резаев, А.А. Дмитриев, Д.В. Чернявский, Применение вероятностных подходов для построения моделей «состав-свойство». – Бетон и железобетон, 4-5, 2022, с. 25-37.*