Prediction of Compressive Strength of Recycled Aggregate Concrete using LASSO

Arif Cekic1* and Jale Tezcan2

1Intertek-PSI, Farmington Hills, USA
2Department of Civil and Environmental Engineering, Southern Illinois University, USA

Submission: March 13, 2018; Published: May 18, 2018

*Corresponding author: Arif Cekic, Intertek-PSI, 37483 Interchange Drive, Farmington Hills, Michigan, USA. Tel: 734-564-3562; Email: arif.cekic@psiusa.com

Abstract

The past few decades have seen a continued interest in the use of industrial waste and byproducts in the production of Portland cement concrete due to increasing landfill costs, concerns on the carbon footprint of the construction industry, and the need to preserve natural resources. While the recent increase in the use of recycled concrete aggregate in the construction industry has made the engineering properties of recycled aggregate concrete a research priority, research efforts have focused mainly on the effect of mix ingredients and concrete maturity on the strength and stiffness properties of concrete. The effect of elevated temperatures has not been investigated. Motivated by the fact that a variety of structures are exposed to elevated temperatures during their service life, this paper investigates the compressive strength of recycled aggregate concrete subjected to elevated temperatures. Using experimental data with 31 test results, the relative importance of four variables, namely, temperature, water-to-cement ratio, recycled aggregate replacement ratio and concrete maturity beyond 28-days, was ranked. Analysis results indicate that the water-to-cement ratio is the strongest predictor of the compressive strength, followed by temperature, replacement ratio, and concrete maturity in that order. The method described in this paper can be applied to a variety of modeling problems in environmental and sustainability research where the relative importance of a set of variables is to be investigated.

Keywords: Recycled aggregate concrete; Compressive strength; Water-to-cement ratio; Replacement ratio; Exposure temperatures; Regression; Lasso

Abbreviations: MSE: Mean Squared Error; DOF: Degrees of Freedom; MSE: Mean Squared Error; LASSO: Least Absolute Shrinkage and Selection Operator; ANN: Artificial Neural Network; SVM: Support Vector Machine; W/C: Water-to-Cement ratio; OLS: Ordinary Least Squares

Introduction

Industrial waste and byproducts have been utilized in the construction industry for decades due to economic and environmental considerations. Increasing landfill costs, concerns on the carbon footprint of the construction industry, and the need to preserve natural resources have motivated a continued interest in the use of industrial waste and byproducts in the production of Portland cement concrete [1]. Fly ash and silica fume [2], waste tire [3-5], waste glass [6,7], and waste plastic [8-10] have been used as a replacement for cement or aggregate in the concrete mix. The use of recycled concrete aggregate as a replacement for aggregate is becoming increasingly common in concrete production [11]. Recycled concrete aggregate, also referred to as recycled crushed concrete, is typically obtained through demolition of Portland cement concrete elements of buildings, roads, and other structures [12]. Typical applications of recycled aggregates include general bulk fill, base or fill in drainage projects, sub base or surface material in pavement and road construction, and new concrete production.

Many researchers [13-17] have focused their attention on engineering properties of Portland cement concrete made with recycled concrete aggregates. The main focus of many of these studies has been the evaluation of the appropriateness of recycled concrete aggregate as described in ASTM C33 [18], the standard specification for concrete aggregates. The mechanical properties of concretes made with recycled aggregates have been observed to be different from those made with natural aggregates. These differences were generally attributed to the mix ingredients and concrete maturity. The effect of elevated temperatures has not been investigated, despite the fact that a variety of non-building structures such as concrete chimneys, pressure furnaces, reactor vessels and coal gasification vessels are exposed to elevated temperatures during their operation, and others may be subjected to high temperatures due to a thermal shock or fire.

This study investigates the effect of four variables, namely temperature, water-to-cement ratio, recycled aggregate replace-
ment ratio expressed as a percentage, and concrete maturity beyond 28 days, on the compressive strength of recycled aggregate concrete. For this purpose, a predictive model has been constructed using a dataset [19] with 31 test results containing recycled aggregates at various water-cement and replacement ratios, exposed to elevated temperatures up to 250 °C. Half of the specimens were tested at 28 days, and the rest at 90 days. The effect of four variables on the compressive strength has been investigated, and their relative importance has been ranked using Least Absolute Shrinkage and Selection Operator (LASSO), which can be described as a penalized regression technique with a built-in capability of variable selection.

While the focus of this study is limited to the compressive strength of recycled aggregate concrete, the method used in this study is applicable to a variety of problems in environmental and sustainability research where data-driven approaches are utilized. The LASSO method described in this paper not only facilitates identification of the important factors affecting a quantity of interest, but also allows construction of predictive models with reduced susceptibility to over fitting. Because the risk of over fitting increases as the ratio of the number of predictors to the number of observations increases, selection of a compact set of variables is important in cases where a small number of observations are available, such as this study. Furthermore, identification of a compact set of relevant variables makes the model easier to interpret, which in turn can help corporations and government organizations make deliberate decisions when evaluating and comparing alternative solutions to environmental and sustainability problems.

The remainder of this paper is organized as follows. Section 2 provides a brief overview of the techniques that have been used in predictive modelling of concrete properties and argues that such techniques are not ideal for the current study, which is based on only 31 test results. Section 3 presents a mathematical description of the Least Absolute Shrinkage and Selection Operator (LASSO). Section 4 is devoted to the investigation of the relative importance of the four variables on the compressive strength, and comparison of the output of regression models with increasing levels of complexity. Section 5 presents a discussion of the bias-variance trade-off in the context of model selection. Finally, Section 6 concludes the paper by presenting a brief summary of this study.

Regression Techniques for Predicting Concrete Properties

Many recent research efforts have been focused on the understanding of how different mix ingredients affect various properties of fresh or hardened concrete. A variety of regression models have been constructed in an effort to understand the influence of various mix ingredients on the properties of the concrete, and to predict the property of interest for a specified mix design.

Various recent studies [13-17, 20-26] have used black box approaches such as the Artificial Neural Network (ANN) and Support Vector Machine (SVM), to model concrete properties such as compressive strength and elastic modulus. These approaches have been found to yield satisfactory results, offering better predictability than the Multiple Linear Regression.

Black box approaches are not suitable in this study for two reasons: First, due to the scarcity of experimental data on recycled aggregate concrete exposed to elevated temperatures, this study is based on a dataset of only 31 observations. Black-box approaches learn the input-output relationship from the training data, without a specified parametric model. When a dataset with a large number of observations is available, an accurate model can be obtained using ANNs and SVMs as long as the complexity of the model is controlled. With a small dataset, model over fitting is an important concern. Reserving a portion of the available data as a test set, as commonly done in black box modeling, would further reduce the size of the data to be used in constructing the model. Second, black box approaches produce a model which is difficult, if not impossible to interpret. Our goal in this paper is to rank the importance of the four variables in regards to the compressive strength of recycled aggregate concrete.

In this paper, we use the Least Absolute Shrinkage and Selection Operator (LASSO) to evaluate the importance of the four variables. Being a penalized regression approach, the LASSO reduces the risk of over fitting. The LASSO is suitable when the number of available observations is low relative to the number of predictors. The LASSO model is described in Section 3.

The Least Absolute Shrinkage and Selection Operator

This section describes the mathematical background of the Least Absolute Shrinkage and Selection Operator (LASSO), proposed by Tibshirani [27]. In the discussion below, bold letters are used to denote vectors and matrices.

Consider a dataset containing \( n \) independent observations \((x^{(i)}, y^{(i)})\), \( i = 1, 2, \ldots, n \). Let \( x^{(i)} = [x_{i1}, x_{i2}, \ldots, x_{ip}] \) be the \( i\)th observation and \( y^{(i)} \) is the corresponding scalar output. It is assumed that the input \( x^{(i)} \) and the output \( y^{(i)} \) are related through the linear equation:

\[
y^{(i)} = \beta_0 + \mathbf{x}^{(i)^T} \mathbf{\beta} + \epsilon^{(i)}
\]  

(1)

where \( \mathbf{\beta} \) is the \( p \times 1 \) coefficient vector and \( \epsilon^{(i)} \) is the random error term associated with \( y^{(i)} \). The error term is assumed to have mean zero and variance \( \sigma^2 \). Equation (1) can be written in matrix form as:

\[
y = \beta_0 \mathbf{1}_n + X \mathbf{\beta} + \mathbf{\epsilon}
\]

Where \( Y \) is the \( n \times 1 \) output vector, \( \beta_0 \) is the intercept term, \( \mathbf{1}_n \) is the \( n \times 1 \) vector of ones, \( X \) is the \( n \times p \) input matrix containing \( x^{(i)} \) as its \( i\)th row and \( \mathbf{\epsilon} \) the \( n \times 1 \) vector of independent and identically distributed random errors. Assuming that each column of \( X \) is shifted and scaled to have zero mean and unit variance, the expected value of \( \mathbf{y} \) is equal to the intercept term. Thus, the intercept term is predicted as \( \hat{\beta}_0 = \mathbf{\bar{y}} \), where \( \mathbf{\bar{y}} = \frac{1}{n} \sum_{i=1}^{n} y^{(i)} \). As a
consequence, without loss of generality, it can be assumed that \( \tau = 0 \) effectively removing the intercept term from the model. This results in the model
\[
y = X\beta + e
\]
where \( y \) now measures the deviations from the mean of the observed output.

The ordinary least squares (OLS) estimate of the \( \beta \) vector, which will be denoted as \( \hat{\beta}_{\text{ols}} \), is obtained by minimizing the sum of the squared residuals. That is,
\[
\hat{\beta}_{\text{ols}} = \arg \min \{ (y - X\beta)'(y - X\beta) \}
\]

Under the Gauss-Markov assumptions, the OLS estimate is unbiased. Further, it has the smallest variance of all linear unbiased estimators. Because the accuracy of an estimator is measured by its mean squared error (MSE), which is equal to the sum of the variance of the estimator and its squared bias, \( \hat{\beta}_{\text{ols}} \) is the best linear unbiased estimator (BLUE) of the true parameter vector \( \beta \). The variance of a biased estimator, however, can be much smaller than that of \( \hat{\beta}_{\text{ols}} \), resulting in a smaller MSE.

Ridge regression, also known as the penalized least squares method, offers a technique for reducing the variance of the estimated regression coefficients. Ridge regression shrinks the coefficients toward zero, making the estimate more stable than the OLS estimate. Ridge regression coefficients are determined by solving the following optimization problem:
\[
\hat{\beta}_{\text{ridge}} = \arg \min \left\{ (y - X\beta)'(y - X\beta) + \lambda \sum_{m} \beta_{m}^{2} \right\}
\]
where \( \lambda > 0 \) is a user-selected parameter that controls the amount of shrinkage. The lasso parameter \( \lambda \) can be selected using cross validation, as described in Section 4.3. Similar to the ridge regression, the LASSO also prefers smaller coefficients, thus resulting in coefficient estimates that are more stable than the OLS method.

Although the LASSO problem given in Equation (3) looks very similar to the ridge regression problem given in Equation (2), the solutions \( \hat{\beta}_{\text{ridge}} \) and \( \hat{\beta}_{\text{lasso}} \) exhibit significant differences. In the process of shrinking the coefficients, the LASSO sets some of the coefficients exactly to zero. This is unlike ridge regression, where the coefficients are shrunk but never set equal to zero. By setting some coefficients equal to zero and retaining the ones strongly affecting the output, the LASSO performs variable selection. Identification of such variables can improve the interpretability of the resulting model, especially when there is a large number of predictors.

LASSO model for predicting the compressive strength

This section presents the application of LASSO described in Section 3 for predicting the compressive strength of concrete. Following a description of the dataset in Section 4.1, we describe the predictor and target variables and construct the LASSO model in Section 4.2. Section 4.3 describes the procedure used in selecting the shrinkage parameter of the model. Section 4.4 presents the results and discusses the relative importance of the predictors.

Data

The dataset used in this paper is taken from Celik [19]. The dataset includes results from 31 compressive strength measurements, where each measurement is obtained as the average of three specimens. The test data is reproduced in Table 1.

### Table 1: Test data reported in Celik [19].

<table>
<thead>
<tr>
<th>Test</th>
<th>W/C</th>
<th>Replacement Ratio (%)</th>
<th>Temperature (°C)</th>
<th>Age (days)</th>
<th>Compressive Strength (MPa)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.35</td>
<td>0</td>
<td>22</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>0.35</td>
<td>0</td>
<td>105</td>
<td>28</td>
<td>33.8</td>
</tr>
<tr>
<td>3</td>
<td>0.35</td>
<td>0</td>
<td>105</td>
<td>28</td>
<td>29.9</td>
</tr>
<tr>
<td>4</td>
<td>0.35</td>
<td>0</td>
<td>175</td>
<td>28</td>
<td>44.8</td>
</tr>
<tr>
<td>5</td>
<td>0.35</td>
<td>0</td>
<td>250</td>
<td>28</td>
<td>52.2</td>
</tr>
<tr>
<td>6</td>
<td>0.35</td>
<td>50</td>
<td>22</td>
<td>28</td>
<td>37.7</td>
</tr>
<tr>
<td>7</td>
<td>0.35</td>
<td>50</td>
<td>105</td>
<td>90</td>
<td>39.3</td>
</tr>
<tr>
<td>8</td>
<td>0.35</td>
<td>100</td>
<td>22</td>
<td>90</td>
<td>40</td>
</tr>
<tr>
<td>9</td>
<td>0.35</td>
<td>100</td>
<td>105</td>
<td>28</td>
<td>37.2</td>
</tr>
<tr>
<td>10</td>
<td>0.35</td>
<td>100</td>
<td>175</td>
<td>90</td>
<td>40.6</td>
</tr>
<tr>
<td>11</td>
<td>0.35</td>
<td>100</td>
<td>250</td>
<td>90</td>
<td>42.7</td>
</tr>
</tbody>
</table>
Input and Target Variables

The predictor set consists of four input variables: the water-to-cement ratio by weight (W/C), replacement ratio of the recycled aggregate, expressed as a weight percentage (RR), temperature in °C (T), and age in days (A). The target variable is the compressive strength of concrete in MPa ($f_c$), obtained as the average of three specimens. Thus the output of the ith observation is $y^{(i)} = f_c^{(i)}$ and the corresponding input vector is $x^{(i)} = [W/C^{(i)}, RR^{(i)}, T^{(i)}, A^{(i)}]$

Estimation of the LASSO parameter

The LASSO has a user-defined parameter $\lambda$, which controls the degree of shrinking of the regression coefficients. Referring to Equation (3), note that $\lambda = 0$ corresponds to the unpenalized, Ordinary Least Squares method, while $\lambda = \infty$ corresponds to all coefficients in $\hat{\beta}_{\text{lasso}}$ being equal to zero.

We select the LASSO parameter $\lambda$ using five-fold cross-validation [28]. In 5-fold cross-validation, the dataset is randomly partitioned into 5 subsets of approximately equal size. The model for $\lambda^*$, a candidate value for $\lambda$ is constructed using four of the subsets, and the prediction error is computed over the remaining subset, called the test set. The process is repeated 5 times, each time using one of the subsets as the test set, and average prediction error over the 5 repetitions is considered to be cross-validation error associated with $\lambda^*$. Repeating this process over a grid of $\lambda$ values, a plot of the average prediction errors and the standard errors is generated. The value $\lambda$ can be selected by examining a plot of the cross validation errors versus $\lambda$. The value minimizing the cross validation error, hereafter referred to as $\lambda_{\text{min}}$ can be selected. However, as observed by Breiman, Friedman [29], a model constructed using $\lambda_{\text{min}}$ is still susceptible to overfitting. Breiman, Friedman [29], define the one standard error rule (1SE rule) as using the simplest model whose error is no more than one standard error above the minimum. In the LASSO formulation, the 1SE rule suggest using the largest $\lambda$ value for which the cross-validation error is at most one standard error above the minimum, hereafter referred to as $\lambda_{\text{opt}}$, instead of $\lambda_{\text{min}}$.

Results

To select the LASSO parameter, we used a grid of 100 $\lambda$ values logarithmically spaced between 0.01 and 10. Figure 1 shows the cross validated mean-square-error versus the $\lambda$ value. The vertical broken lines indicate the $\lambda$ values at which a coefficient is shrunk to zero, eliminating the corresponding variable from the model. The $\lambda$ values at which a coefficient is shrunk to zero are listed in Table 2, along with the remaining coefficients. The column titled $\lambda = 0$ corresponds to unpenalized, Ordinary Least Squares regression. This column was provided as a reference to check the amount of shrinkage of the coefficients. Up to $\lambda = 0.7565$, all four variables are retained. At $\lambda = 0.7565$, the coefficient of the Age (A) is shrunk to zero. At $\lambda = 0.8698$, the coefficient of re-
placement ratio (RR) is shrunk to zero. The temperature variable (T) is retained until $\lambda = 2.4771$ after which it is also discarded, leaving the water-to-cement ratio (W/C) as the only predictor.

### Table 2: Coefficients of water to cement ratio (W/C), replacement ratio (RR), temperature (T) and age (A) for values of $\lambda$ at which a variable is eliminated.

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>W/C</th>
<th>RR</th>
<th>T</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>-6.338</td>
<td>-0.9569</td>
<td>2.4735</td>
<td>0.7098</td>
</tr>
<tr>
<td>0.01</td>
<td>-6.328</td>
<td>-0.9456</td>
<td>2.4635</td>
<td>-0.7</td>
</tr>
<tr>
<td>0.7565</td>
<td>-5.582</td>
<td>-0.0994</td>
<td>1.7097</td>
<td>0</td>
</tr>
<tr>
<td>0.8698</td>
<td>-5.4631</td>
<td>0</td>
<td>1.5855</td>
<td>0</td>
</tr>
<tr>
<td>2.4771</td>
<td>-3.7629</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Figure 1, which shows the coefficients of the four predictors versus the LASSO parameter $\lambda$.

These results suggest that the four variables affecting the compressive strength of recycled aggregate concrete can be ranked in the order of decreasing importance as follows: (1) Water-to-cement ratio, (2) Temperature, (3) Replacement ratio, and (4) Age. Note that in this study, the age variable only takes two values, 28 days and 90 days. Thus, the ranking of Age variable as the least important implies that little strength is gained after 28 days, as expected.

The LASSO model with the two selected variables (water-to-cement ratio and temperature) was used to obtain the predictions for the compressive strength. The predicted strength values versus the measured values are shown in Figure 3. The coefficient of multiple determination ($R^2$) was computed as 0.74, indicating that the LASSO model is able to explain 74% of the observed variance in the compressive strength.

### Discussion

In this section, we provide a comparison of the OLS and LASSO approaches in terms of (1) their ability to explain the observed variance in the compressive strength, quantified by the coefficient of multiple determination, and (2) how well they can generalize to unseen data, based on theoretical understanding of the two methods.

To compare OLS and LASSO in terms of their ability to explain the observed variance in the compressive strength, we define four model types represented by the following parametric forms:

- **Type 1:** $y = b_0 + \sum_{i=1}^{\text{DOF}_i} b_i x_i$
- **Type 2:** $y = b_0 + \sum_{j=1}^{\text{DOF}_j} b_j x_j + \sum_{j=1}^{\text{DOF}_j} \sum_{i=1}^{\text{DOF}_i} b_{ij} x_i x_j$
- **Type 3:** $y = b_0 + \sum_{i=1}^{\text{DOF}_i} b_i x_i + \sum_{i=1}^{\text{DOF}_i} c_i x_i^2$
In the equations above, the parameters $b_0, b_1, b_2$, and $c_i$ represent the constant term, the coefficient of the linear term $x_i$, the coefficient representing the interaction between $x_i$ and $x_j$, and the coefficient of the squared term $x_i^2$, respectively. Note that all four model types defined above are linear in parameters even though three of the four models are nonlinear in variables. Because of the linearity in parameters, linear regression applies to all four model types.

For each model type described above, we perform regression using one, two, three, and four degrees of freedom (DOF), overall producing 16 OLS models and 16 LASSO models. The predictors corresponding to different degrees of freedom are shown in Table 3.

**Table 3:** Predictors corresponding to different degrees of freedom (DOF). The predictors are water to cement ratio (W/C), temperature (T), replacement ratio (RR), and age (A).

<table>
<thead>
<tr>
<th>DOF</th>
<th>W/C</th>
<th>T</th>
<th>RR</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>yes</td>
<td>no</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>2</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>3</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>4</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
<td>yes</td>
</tr>
</tbody>
</table>

Table 4 & 5 show the coefficients of multiple determination ($R^2$) obtained using OLS and LASSO approaches. Note that as the OLS approach requires a preselected set of variables, variable selection has been performed using LASSO, regardless of the choice of the regression approach used in the final model.

**Table 4:** Coefficient of multiple determination ($R^2$) obtained using Ordinary Least Squares (OLS) regression.

<table>
<thead>
<tr>
<th>DOF</th>
<th>Type=1</th>
<th>Type=2</th>
<th>Type=3</th>
<th>Type=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.7082</td>
<td>0.7082</td>
<td>0.7094</td>
<td>0.7094</td>
</tr>
<tr>
<td>2</td>
<td>0.8206</td>
<td>0.8345</td>
<td>0.8617</td>
<td>0.8759</td>
</tr>
<tr>
<td>3</td>
<td>0.8349</td>
<td>0.8872</td>
<td>0.8769</td>
<td>0.9311</td>
</tr>
<tr>
<td>4</td>
<td>0.8347</td>
<td>0.9057</td>
<td>0.8957</td>
<td>0.966</td>
</tr>
</tbody>
</table>

**Table 5:** Coefficient of multiple determination ($R^2$) obtained using LASSO regression.

<table>
<thead>
<tr>
<th>DOF</th>
<th>Type=1</th>
<th>Type=2</th>
<th>Type=3</th>
<th>Type=4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.6167</td>
<td>0.598</td>
<td>0.6167</td>
<td>0.5755</td>
</tr>
<tr>
<td>2</td>
<td>0.785</td>
<td>0.7303</td>
<td>0.8111</td>
<td>0.815</td>
</tr>
<tr>
<td>3</td>
<td>0.8173</td>
<td>0.7378</td>
<td>0.8111</td>
<td>0.8545</td>
</tr>
<tr>
<td>4</td>
<td>0.7456</td>
<td>0.8285</td>
<td>0.7303</td>
<td>0.8815</td>
</tr>
</tbody>
</table>

Examination of the $R^2$ values presented in Tables 4 & 5 leads to the following observations. First, as expected, the $R^2$ values from the OLS approach (Table 4) increases as the number of predictors used in the model (DOF) increases. Second, the OLS approach produce higher $R^2$ values as the complexity of the model’s parametric form increases. In this paper, the Type specification is intended to represent this complexity, although the complexities of Type 2 (constant, linear and interaction terms) and Type 3 (constant, linear, and squared terms) are not directly comparable. This fact is reflected in Table 4, which shows an increase in $R^2$ values between Types 2 and 3 for one and two degrees of freedom, and a decrease for three and four degrees of freedom.

The changes in the $R^2$ values with increasing DOF and Type values is less predictable in the LASSO models, as Table 5 indicates. While the small size of the dataset and the lack of knowledge regarding the true parametric form of the relationship between the compressive strength and the four predictors make it difficult to interpret the $R^2$ values as a reliable indicator of the model’s accuracy, the lack of a strong trend in $R^2$ values from the LASSO approach shows its lessened susceptibility to over fitting, compared to the OLS approach. Over fitting is an issue strongly linked to variance of a model, affecting the generalizability of the model to observations not included in the dataset.

In the context of bias-variance trade-off, the differences between the OLS and LASSO approaches can be summarized as follows. It is well known that the regression estimates from the OLS approach are unbiased. The LASSO approach introduces bias into the solution, which can lead to a considerable reduction in variance compared to the OLS solution. The reduction in variance, in turn, leads to a reduction in the Mean Squared Error (MSE), despite the increase in bias. As the MSE, which takes into account both the bias and the variance, is the most relevant quantity in model selection, the LASSO offers better generalization to the unseen data, compared to the OLS approach.

It should be emphasized that the bias-variance discussion presented above is based on the well-established theoretical understanding of the two approaches. In practical applications, bias and variance of a model cannot be computed since the true conditional distribution of the output given the input is not known. Further, bias-variance decomposition, which relies on averaging over ensembles of datasets cannot be performed with a single observed data set.

**Conclusion**

In this paper, the compressive strength of recycled aggregate concrete subjected to elevated temperatures has been investigated through LASSO and OLS regression, using a dataset of 31 observations, containing four predictors: temperature, water-to-cement ratio, recycled aggregate replacement ratio and concrete maturity beyond 28-days. Using the Least Absolute Shrinkage and Selection Operator, the four variables were ranked in the order of decreasing relative importance as water-to-cement ratio, temperature, replacement ratio, and concrete maturity beyond 28-days. Regression models with four different parametric forms were constructed using one, two, three, and four best predictors, and the models’ ability to explain the observed variance in the compressive strength were compared. The advantage of the LASSO approach in terms of the bias-variance trade-off was discussed in the context of model selection.
References


This work is licensed under Creative Commons Attribution 4.0 License
DOI: 10.19080/CERJ.2018.05.55565

Your next submission with Juniper Publishers will reach you the below assets

- Quality Editorial service
- Swift Peer Review
- Reprints availability
- E-prints Service
- Manuscript Podcast for convenient understanding
- Global attainment for your research
- Manuscript accessibility in different formats (Pdf, E-pub, Full Text, Audio)
- Unceasing customer service

Track the below URL for one-step submission
https://juniperpublishers.com/online-submission.php

How to cite this article: Arif C, Jale T. Prediction of Compressive Strength of Recycled Aggregate Concrete using LASSO. Civil Eng Res J. 2018; 5(1): 555654. DOI: 10.19080/CERJ.2018.05.55565.