EPiC Series in Built Environment
Volume 2, 2021, Pages 28–36
ASC 2021. 57th Annual Associated Schools of Construction International Conference

A Statistical Approach to Predicting Fresh State Properties of Sustainable Concrete

Ruoyu Jin, Ph.D.
London South Bank University
London, UK

Qian Chen, Ph.D. and Alfred Soboyejo, Ph.D.
The Ohio State University
Columbus, OH

Using environmentally friendly concrete materials can help improve the sustainability of the concrete industry. However, the effects of such materials on concrete properties must be fully understood before sustainable concrete can be widely applied. Previous research showed limited applications of statistical methods in analyzing the effects of sustainable concrete materials on fresh concrete properties. This study applies multivariate regression analysis to modeling properties of fresh concrete (i.e., slump, air content, and density) made with multiple sustainable raw materials based on variables in mixture design. Different regression models were tested to explore the best-fit model(s) that can capture and predict how these variables affect fresh concrete properties. The regression analysis showed satisfactory results in predicting air content and density, but not in predicting slump. The regression analysis, as a statistical tool, can provide deep insights into how the selected independent variables affect fresh concrete properties and the degree of the effects.

Key Words: Sustainable concrete, Fresh concrete properties, Regression analysis, Mixture design, Statistical methods.

Introduction

The sustainability movement of the construction industry encourages using building materials with one or more of the following features: containing recycled content, environmentally friendly with reduced greenhouse gas emissions, reserving natural resources, locally available to decrease transportation cost, improved material performance in its life cycle, etc. As the most widely consumed construction material worldwide, concrete has caught increasing attention from researchers who are interested in finding out how concrete sustainability could be improved by replacing conventional concrete ingredients (especially cementitious and aggregate materials) with sustainable materials.

Portland limestone cement (PLC), as an alternative to Portland cement (PC), has been more widely used in European countries to reduce the energy use and emissions associated with cement manufacturing (Livesey, 1991). In the U.S. concrete industry, supplementary cementitious materials (SCMs), such as fly ash (FA), silica fume, and ground-granulated blast-furnace slag, have often been

T. Leathem, A. Perrenoud and W. Collins (eds.), ASC 2021 (EPiC Series in Built Environment, vol. 2), pp. 28–36

used to lower the environmental impacts from the use of Portland cement (Jin et al., 2015; Obla el al., 2012). Alternative aggregates like lightweight aggregate (LWA) and recycled concrete aggregate are also used to produce concrete for suitable applications (Jin et al., 2015). Although these sustainable concrete materials have been individually investigated on how they could affect concrete properties, there are limited studies (e.g., Berry, 2011; Limbachiya et al., 2012) on how they can be jointly applied to further elevate the sustainability of concrete, not to mention wide industry applications.

In the concrete industry, guidelines and standards are usually used for designing concrete mixtures, and the overdesign factor is statistically determined by test data records or calculated based on formulas (when the data is not sufficient) (ACI, 2019; ODoT, 2008; Pardi, M., personal communication, Apr. 16, 2012). Some companies also use their historical data or other methods such as trial batches (Jin et al., 2015). So far neither existing standards nor previously generated test data could cover scenarios where multiple sustainable concrete materials are jointly used to produce sustainable concrete. The investigation of concrete properties in relation to various substitution rates of alternative materials usually requires large sample-size experiments, which are not only time-consuming but also cost-prohibitive. Thus, adopting statistical, mathematical, or other predictive modeling techniques becomes a practical approach to studying how various mixture designs affect concrete properties and improving the production of sustainable concrete (Abounia Omran et al., 2016; Atici, 2011; Jin et al., 2018). Most of previous works on predicting concrete properties focus on hardened concrete properties, e.g., strength. Only a few studies attempted to predict fresh concrete properties like slump (Agrawal & Sharma, 2010; Gomma et al., 2021; Oztas et al., 2006).

This study adopted multivariate regression analysis (MRA) to model the relationships between properties (slump, air content, and density) of fresh sustainable concrete made with PLC, FA, and LWA and variables related to concrete mixture proportion. Both linear and nonlinear regression models were investigated to find the best-fit one(s). Multiple statistical methods were used in the regression analysis to comprehensively evaluate the reliability of the tested models. The established models generated from experimental data have great potential for predicting fresh concrete properties given the mixture proportion. The multiple potential models examined in this study expanded the options of quantitative methods from previous research in exploring the best-fit models to capture the relationship between fresh concrete properties and independent variables.

Materials and Methods

PC Type I/II, brown sand (fineness modulus at 2.48), and pea gravel with maximum size at 3/8 inch were conventional concrete materials used in the control group in the experimental tests. Micro Air was chosen as the air-entraining admixture (AEA) to increase air content in concrete batches. PLC, Haydite ® LWA, and FA Class F were selected as promising sustainable concrete materials for experimental study based on the market survey results from Jin et al. (2015). Thirty-six batches of air-entrained concrete were made in the experiment following the guideline of ASTM C31/C31M-06 (ASTM International, 2007a). As illustrated in Figure 1, these batches were divided by PC and PLC concrete with two different water-cement (W/C) ratios (0.40 and 0.65), three different substitution rates of FA Class F (0%, 20%, 30 or 40%) by weight of the cementitious material, and three different replacement rates of LWA to pea gravel by volume (0%, 33%, and 67%). Each concrete batch was tested of its fresh state properties including density following ASTM C138/C138M-01a (ASTM International, 2007b), air content following ASTM C231-04 (ASTM International, 2007c), and slump following ASTM C143/C143M-05a (ASTM International, 2007d). The test results of both PC and PLC batches were combined to create a sufficient data sample for statistical analysis.

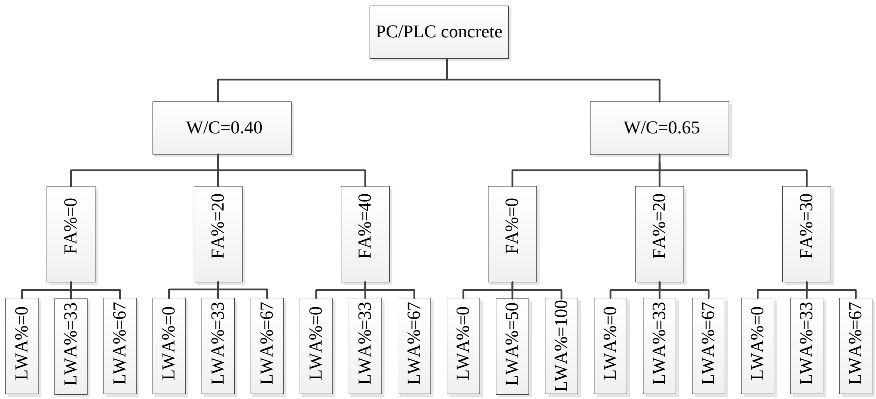


Figure 1. Mixture design involving sustainable concrete materials.

This research explored the potential relationship between selected sustainable concrete properties and predicator variables related to concrete mixture. The response random variable (RRV) included fresh concrete properties, i.e., slump, air content, and density. In total six mixture design related independent predictor variables (IPVs) were defined. Let Y_i denote ith target RRV, and X_{ij} denote the *k*th IPV in totally six IPVs. The target RRVs and the six IPVs are defined as follows:

- Y_i: Slump, air content, or density
- X_{i1}: Fly ash substitution rate in cementitious material (FA%)
- X_{i2}: LWA substitution rate in coarse aggregate (LWA%)
- X_{i3}: W/C ratio
- X_{i4}: Weight ratio of sand to cement (S/C)
- X_{i5}: Volume ratio of sand to coarse aggregate (S/CA)
- X_{i6}: Amount of AEA (fl oz.) per 100 lbs of cement.

The general format of multivariate regression analysis (MRA) is displayed in Eq. (1):

$$Y = f\ (X_1, X_2 \dots, X_k). \tag{1}$$

The potential MRA statistical models introduced in Soboyejo (2009) were applied to describe the relationship between the target RRV and IPVs. Model 1 using the typical multi-linear regression analysis is displayed in Eq. (2) and mixed Models 2 to 7, which were converted into the linear formats, are introduced in Eqs. (3) - (8):

$$Y_i = \alpha\ +\ \sum_{j=1}^{k} \beta_j X_{ij}\ ,\ \ i\ =1, \dots, n, \tag{2}$$

$$\frac{X_{ij}}{Y_i} = \alpha\ \ + \sum_{l=1}^{k} \beta_l X_{il}\ ,\ \ i\ =1, \dots, n,\ \ j\ =1, \dots, k. \qquad \text{(3) to (8)}$$

This study used Minitab to analyze these MRA models. R^2 and residual standard deviation were generated to compare the accuracy of these models in predicting each target RRV. The *F* and *p* values generated from Analysis of Variance (ANOVA) were used to test the significance of the selected regression model in describing the data sample. Regression analysis of all these models was at 95% level of significance. The null hypothesis is that the target RRV cannot be predicted by the selected model involving these IPVs. A *p* value less than 0.05 from ANOVA would reject the null hypothesis and indicate that the selected model fits the data. Residual analysis was also conducted to study the distribution and values of residuals—the differences between predicted RRV and experimental data.

The Durbin-Watson statistic test is based on the null hypothesis that residuals from a least-square regression are not autocorrelated. The Durbin-Watson value ranges from 0 to 4. A value less than 2 indicates positive autocorrelation, a value near 2 indicates non-autocorrelation, and a value falling between 2 and 4 indicates negative autocorrelation. According to Atici (2011), the ideal Durbin-Watson value would be from 1.5 to 2.5. Among the six IPVs, some may have more significant effects on the target RRV than others. The *t*-test of correlation analysis was used to determine the significance regarding the effect of each IPV on RRV. There is a *p* value corresponding to each *t* value for an IPV. At the 95% confidence level, a *p* value lower than 0.05 would indicate that this selected IPV has significant contribution to RRV. In contrast, IPVs with *p* values higher than 0.05 are those without significant contributions.

Results and Discussions

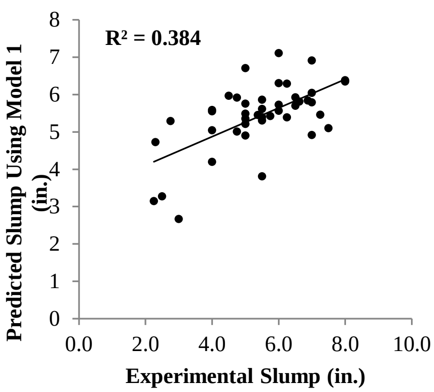
Prediction of Slump

The MRA Model 1 in Eq. (9) did not show a strong correlation between slump and the six IPVs (R^2 = 38.4%), although *p* value at 0.005 indicated a significant correlation. The mixed model in Eq. (10) using AEA/Slump as the RRV displayed the highest R^2 value at 61.2% among the six mixed models.

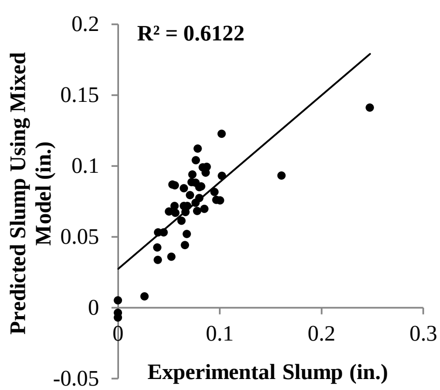
$$Slump\ (in.) = 7.570 + 0.011FA\% + 0.005LWA\% - 13.400\frac{W}{C} + 2.330\frac{S}{C} - 2.070\frac{S}{CA} + 4.450AEA, \tag{9}$$

$$\frac{AEA}{Slump} = -0.0660 - 0.0002FA\% - 0.0001LWA\% + 0.1381\frac{W}{C} - 0.0243\frac{S}{C} + 0.0382\frac{S}{CA} + 0.2299\,AEA. \tag{10}$$

The predicted RRVs using Eqs. (9) and (10) were compared with the experimental data in Figure 2. The plots of residuals based on Eq. (9) are presented in Figure 3. The normal probability plot and histogram of residuals showed a satisfactory trend of normal distribution. The residuals over the fitted value, i.e., the predicted RRV value using Eq. (9), and observation order also displayed ideal constant band. One potential cause of the relative lower R^2 value in the regression could be due to the combination of PC and PLC concrete batches, since the different cement types could affect the concrete workability. However, when divided by sub-samples of PC and PLC concrete, the same MRA models received R^2 values of 68.9% and 41.5%, respectively. The low correlation indicated that factors other than the mixture proportion might affect fresh concrete slump.



(a)



(b)

Figure 2. Comparison between predicted slump and experimental data: (a) RRV as slump from Model 1 and (b) RRV as AEA/Slump from the mixed model.

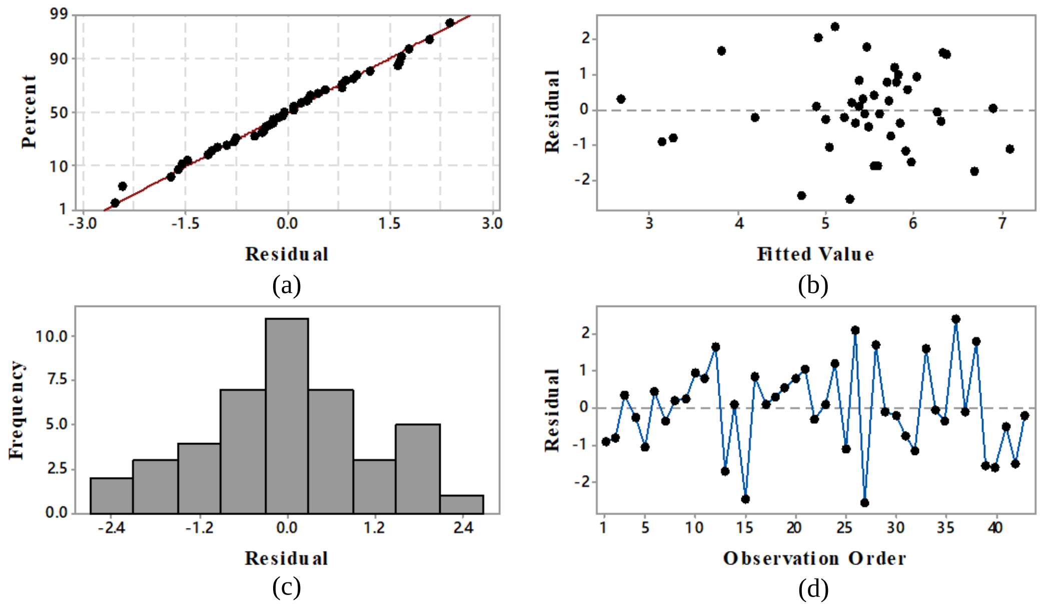


Figure 3. Residual plot analysis for Eq. (9): (a) normal probability plot; (b) residual vs. fitted plot; (c) histogram; and (d) residual vs. observation order.

The Durbin-Watson values at 2.54 and 2.64 revealed that the residuals from the predictive values generated by both models were in negative correlation. Three factors—W/C, S/C, and amount of AEA—were found to be significantly related to slump. This study revealed that MRA achieved lower accuracy in predicting slump than some other predictive techniques, e.g., neural networks achieving R^2 above 0.9 in Agrawal & Sharma (2010).

Prediction of Air Content

Data of air content was only available for batches without LWA. The generated Model 1 and one mixed model are shown in Eqs. (11) and (12):

$$Air\ content\ (\%) = 4.35 - 0.02FA\% - 7.03\frac{W}{C} + 3.35\frac{S}{C} - 4.68\frac{\mathrm{S}}{\mathrm{CA}} + 9.05\mathrm{AEA}, \quad (11)$$

$$\frac{AEA}{Air\ content}(\%) = 0.0021 + 0.0005FA\% - 0.0462\frac{W}{C} - 0.0279\frac{S}{C} + 0.0640\frac{S}{CA} + 0.1750AEA. \quad (12)$$

The RRVs using Eqs. (11) and (12) were compared with the experimental data in Figure 4. The plots of residuals of the mixed model generated from Minitab are presented in Figure 5. Like the residual analysis in slump test, all the four graphs in Figure 5 showed a satisfactory trend of normal distribution. The mixed model was found to be superior to Model 1 considering the increased R^2 (0.9466 vs. 0.915) and *F* values. The Durbin-Watson value from the mixed model (1.85) was also closer to 2, showing satisfactory residual correlation. The coefficient analysis found that AEA has the highest effect on air content. Besides AEA, other influencing factors with *p* values lower than 0.05 include FA% and S/C in the mixed model. Therefore, removing other remaining IPVs would not significantly reduce the correlation between RRV and concrete mixture related IPVs nor increase the residual standard deviation. It was also noticed that keeping only these critical IPVs would increase *F* values in ANOVA and achieve more satisfactory Durbin-Watson values. However, further removing these critical values would cause significant reductions in correlation. For example, with AEA as the sole IPV would only explain around 67% of variation for RRV and the residuals would be biased with the low Durbin-Watson value at 1.080. It was suggested that dividing the test sample into PC and PLC batches would further increase the correlation using the same model considering that the cement type might also have some effects on air content.

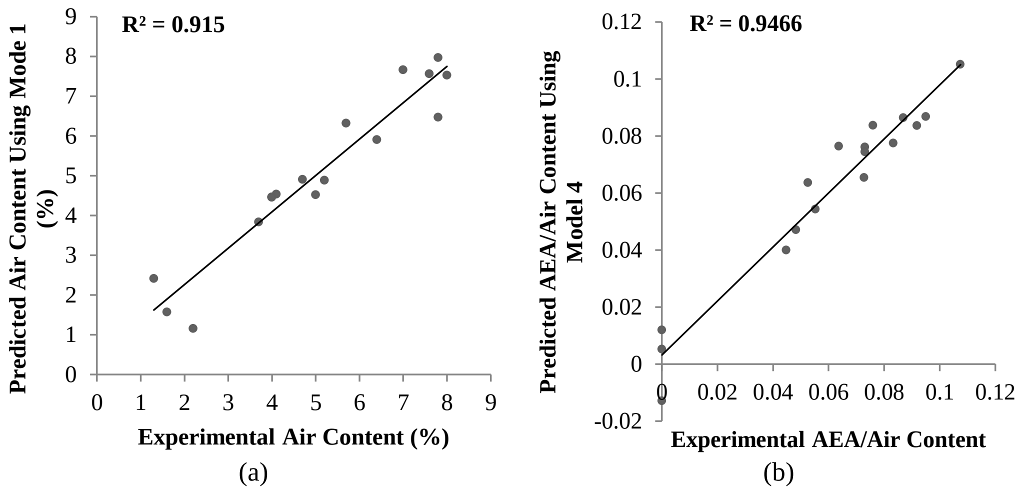


Figure 4. Comparison between predicted RRVs and experiment data: (a) RRV as air content from MRA and (b) RRV as AEA/Air content from the mixed model.

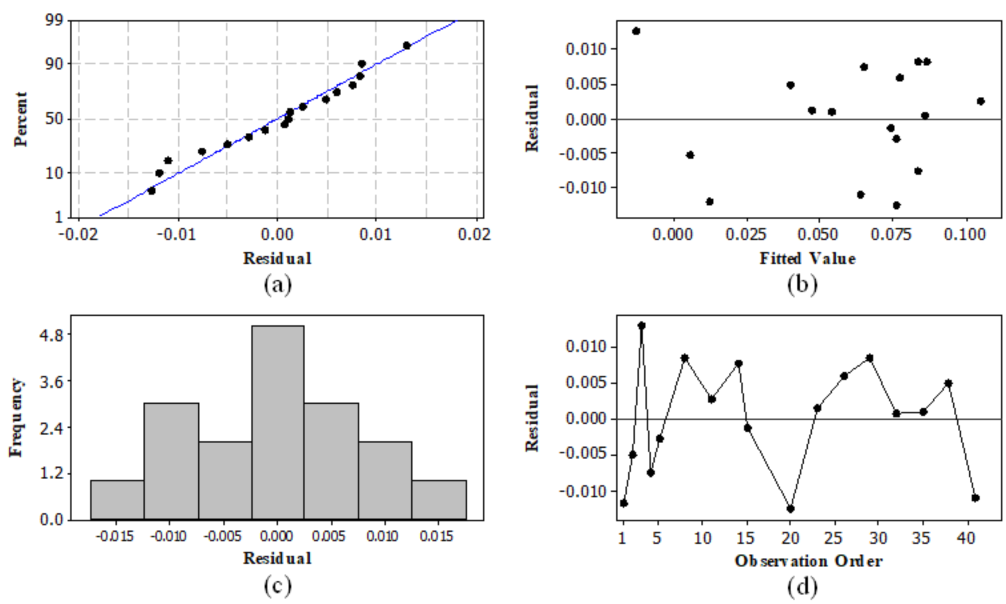


Figure 5. Residual plot analysis for Eq. (12): (a) normal probability plot; (b) residual vs. fitted plot; (c) histogram; and (d) residual vs. observation order.

Prediction of Density

The MRA was applied to predict density (lbs/cf) in relation to six IPVs in mixture design. Eq. (13) based on Model 1 generated the excellent simulation results considering the low residual standard deviation at 1.00 lb/cf, high R^2 value at 99.1%, high F values at ANOVA, and ideal Durbin-Watson value at 1.908, which was close to 2. The predicted density using Eq. (13) and experimental data were compared in Figure 6. Satisfactory normal probability plot and residual over fitted value were observed in the residual analysis as displayed in Figure 7.

$$Density\left(\frac{lbs}{cf}\right) = 151 + 0.03FA\% - 0.260LWA\% + 0.28\frac{W}{C} - 2.94\frac{S}{C} + 0.56\frac{S}{CA} - 5.31AEA. \quad (13)$$

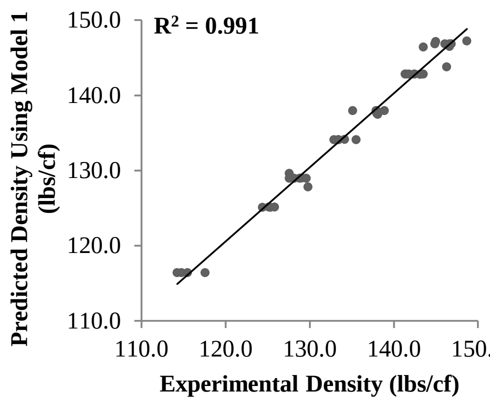


Figure 6. Comparison between predicted density based on Model 1 and experiment data.

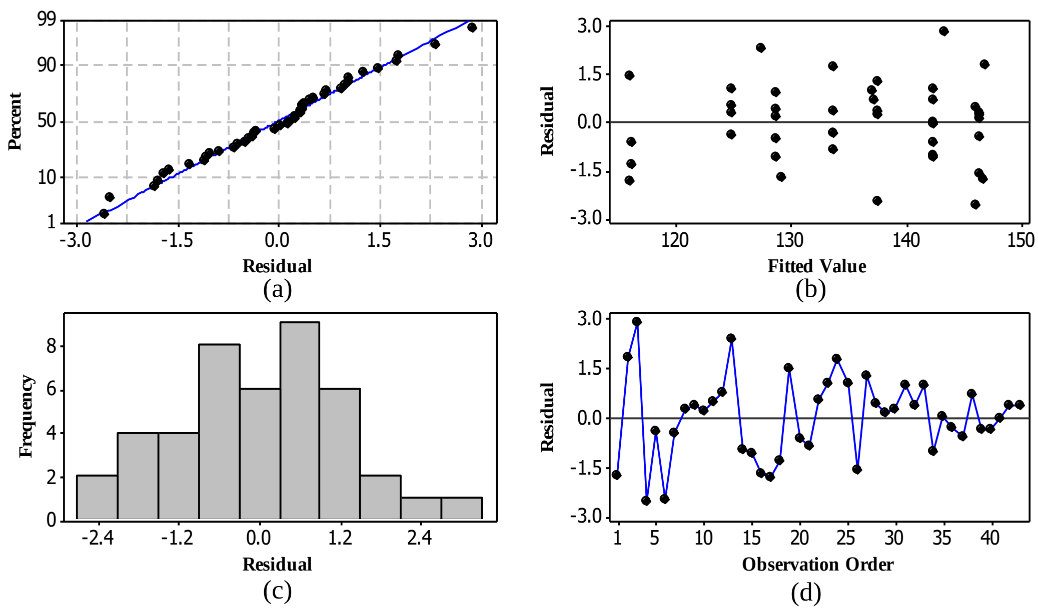


Figure 7. Residual plot analysis for Eq. (13): (a) normal probability plot; (b) residual vs. fitted plot; (c) histogram; and (d) residual vs. observation order.

After removing the two insignificant IPVs (i.e., W/C and S/CA) in the initial MRA, this research found that the remaining four IPVs performed equivalently in predicting density with even slightly lower residual standard deviation, higher *F* value, and comparable Durbin-Watson value. Further reducing IPVs until only LWA% and S/C were kept would still perform comparably in the prediction of density. When only LWA% (the most significant IPV in affecting concrete density) was adopted in the single linear regression analysis, relatively high R^2 (94.5%) was achieved. This indicated that LWA% was the major predictive factor for density accounting for 94.5% change of concrete density. However, the low Durbin-Watson value (0.69) showed relatively strong positive correlation of residuals. Therefore, at least one more IPV would be necessary in addition to LWA% to predict the concrete density. Eq. (14) describes density in relation to two IPVs (LWA% and S/C):

$$Density\ \left(\frac{lbs}{cf}\right) = 151 - 0.26LWA\% - 3.01\frac{S}{C}. \tag{14}$$

Conclusion

This study applied multivariate regression analysis (MRA) in modeling the relationships between fresh state properties of sustainable concrete and a comprehensive list of independent predictor variables (IPVs) related to mixture proportion. The six IPVs included in this study covered all ingredients in concrete mixture design, from conventional sand, gravel, Portland cement to sustainable concrete materials (i.e., PLC, FA, and LWA), plus AEA and water usage. As the response random variables, the simulated fresh concrete properties included slump, air content, and density. The regression analysis showed satisfactory results in describing air content and density of sustainable concrete in relation to IPVs except for slump simulation. The MRA and mixed model results indicated that although AEA was the most critical factor to concrete air content, other mixture design factors, including fly ash content (FA%) and sand to cement ratio (S/C), also played important roles. The MRA showed satisfactory prediction of density with only two critical IPVs, i.e., LWA% and S/C.

The regression analysis in this study provided a quantitative tool to predict fresh concrete properties (including air content and density) of sustainable concrete purely based on mixture proportion variables. The statistical tool has advantages of being easy-to-use and low-cost, not requiring huge datasets, and achieving high degree of accuracy. Although the best-fit models identified in this research still need further test data to validate and improve accuracy by assigning adjustment factors, the statistical tool could serve as a potential alternative method in assisting concrete mixture design and estimating and controlling fresh properties of sustainable concrete in the construction field.

References

Abounia Omran, B., Chen, Q., & Jin, R. (2016). Comparison of data mining techniques for predicting compressive strength of environmentally friendly concrete. *Journal of Computing in Civil Engineering*, *30*(6), http://ascelibrary.org/doi/abs/10.1061/(ASCE)CP.1943-5487.0000596.

Agrawal, V. & Sharma, A. (2010). Prediction of slump in concrete using artificial neural networks. *International Journal of Civil, Environmental, Structural, Construction and Architectural Engineering*, *4*(9), 279-286.

American Concrete Institute (ACI). (2019). *ACI CODE-318-19: Building code requirements for structural concrete and commentary*. Farmington Hills, MI: ACI.

ASTM International. (2007a). ASTM C31/C31M-06: Standard practice for making and curing concrete test specimens in the field. 3rd Ed. West Conshohocken, PA.

ASTM International. (2007b). ASTM C138/C138M-01a: Standard test method for density (unit weight), yield, and air content (gravimetric) of concrete. 3rd Ed. West Conshohocken, PA.

ASTM International. (2007c). ASTM C231-04: Standard test method for air content of freshly mixed concrete by the pressure method. 3rd Ed. West Conshohocken, PA.

ASTM International. (2007d). ASTM C143/C143M-05a: Standard test method for slump of hydraulic-cement concrete. 3rd Ed. West Conshohocken, PA.

Atici, U. (2011). Prediction of the strength of mineral admixture concrete using multivariable regression analysis and an artificial neural network. *Expert Systems with Applications*, *38*(8), 9609-9618.

Berry, M., Stephens, J., & Cross, D. (2011). Performance of 100% fly ash concrete with recycled glass aggregate. *ACI Materials Journal*, 108(4), 378-384.

Gomaa, E., Han, T., ElGawady, M., Huang, J., & Kumar, A. (2021). Machine learning to predict properties of fresh and hardened alkali-activated concrete. *Cement and Concrete Composites*, *115*(January), 103863.

Jin, R, Chen, Q., & Soboyejo, A. (2015). Survey of the current status of sustainable concrete production in the U.S. *Resources, Conservation and Recycling*, 105, 148-159.

Jin, R., Chen, Q., & Soboyejo, A. (2018). Non-linear and mixed regression models for predicting sustainable concrete strength. *Construction and Building Materials*, 170, 142-152.

Limbachiya, M., Meddah, M. S., & Ouchagour, Y. (2012). Performance of Portland/silica fume cement concrete produced with recycled concrete aggregate. *ACI Materials Journal, 109*(1), 91-100.

Livesey, P. (1991). Strength characteristics of Portland-limestone cements. *Construction Building Materials*, 5(3), 147-150.

Obla, K. H., Lobo, C. L., & Kim, H. (2012). The 2012 NRMCA supplementary cementitious materials use survey. *NRMCA Concrete InFocus Magazine*, Fall 2012 Edition.

Oztas, A., Pala, M., Ozbay, E., Kanca, E., Caglar, N., & Bhatti, M. A. (2006). Predicting the compressive strength and slump of high strength concrete using neural network. *Construction Building Materials, 20*(9), 769–775.

Pardi, M. PE, LEED GA, Central/SE Ohio Concrete, Ohio Concrete Construction Association (OCCA). Personal communication, Apr. 16, 2012.

Soboyejo, A. (2009). *Probabilistic methods in engineering, environmental and biosystems engineering*. The Ohio State University, Columbus, OH. Chapters 7 and 8, 25-41.

State of Ohio Department of Transportation (ODoT). (2008). Construction and material specifications. Item 499.03, pp. 263-279.