Predictive Statistical Cost Estimation Model for Existing Single Family Home Elevation Projects

Arash Taghinezhad[1]*, Carol J. Friedland[1], Robert V. Rohli[2], Brian D. Marx[3], Jeffrey Giering[4] and Isabelina Nahmens[5]

[1]Bert S. Turner Department of Construction Management, College of Engineering, Louisiana State University, Baton Rouge, LA, United states, [2]Department of Oceanography & Coastal Sciences, Louisiana State University, Baton Rouge, LA, United States, [3]Department of Experimental Statistics, Louisiana State University, Baton Rouge, LA, United States, [4]Louisiana Governor's Office of Homeland Security and Emergency Preparedness, Baton Rouge, LA, United States, [5]Department of Mechanical and Industrial Engineering, Louisiana State University, Baton Rouge, LA, United States

One of the most preferred flood mitigation techniques for existing homes is raising the elevation of the lowest floor above the base flood elevation (BFE). Determination of project effectiveness through benefit-cost analysis (BCA) relies on the expected avoided flood loss and the project cost. Conventional construction cost estimates are highly detailed, considering specific details of the project; however, mitigation project decisions must often be made while considering only highly generalized building details. To provide a robust, generalized project cost estimation method, this paper implements data modeling and mining methods such as multiple regression, random forest, generalized additive model (GAM), and model evaluation and selection with cross-validation methods to hindcast elevation costs for existing single-family homes based on average floor area, increase in floor elevation, number of stories, and foundation type. Project cost data for homes elevated in Louisiana, United States, between 2005 and 2015 are used in cost prediction analysis. The statistical modeling results are compared with detailed estimations for several types of home foundations over a range of elevations. The results show substantial agreement between regression predictions and detailed estimates using RSMeans cost data.

Keywords: flood mitigation, Freeboard, cost estimation, regression, random forest, GAM, cross-validation, foundation cost

INTRODUCTION

Elevating the lowest floor of existing homes is widely considered to be the most effective building-scale flood mitigation strategy (Bellomo et al., 1999; FEMA 2010; FEMA 2012; Li and van De Lindt 2012; Bohn 2013), in contrast to acquisition and reconstruction. In spite of the effectiveness of elevation, this construction technique is performed by highly specialized contractors and generalized cost guidance is not widely available. At the project decision stage, benefit-cost analysis (BCA) must demonstrate a positive return on investment (FEMA 2011; Orooji and Friedland 2017). Thus, reasonable cost estimates are needed for comparison with long-term benefits to evaluate the most economically efficient strategies to achieve overall mitigation goals and provide economic justification for specific projects (Renn, 1998; Amoroso and Fennell 2008).

Conventional methods for project cost estimation are unit-cost and unit-area-cost. Unit-cost is project-specific, with exact construction quantities and historical unit-price costs, while unit-area-cost is based on general building attributes such as occupancy, building type, and other building parameters. In the absence of proprietary historical cost data, RSMeans (Waier and Balboni 2018) is commonly used to estimate construction cost. However, RSMeans data do not include all necessary construction activities for elevation projects and prices can vary substantially by contractor (Gair et al., 2011). These and similar shortcomings limit the ability of stakeholders (e.g., federal, state, and local agencies, homeowners) to estimate elevation project cost effectively.

Acknowledging this issue, elevation cost guidance has been developed previously. USACE (1993) reported that for a 0.6 m (2 foot) elevation, elevating wood-frame buildings with existing pile, post, or pier foundations costs $280/m^2 ($26/ft^2), while elevating slab buildings costs $320/m^2 ($30/ft^2) in 1993 dollars. Considering a 140 m^2 (1,500 ft^2) house with 0.6-m (2-ft) elevation, additional costs associated with earthen fill (slab only), landscaping, engineering design, and contract cost bring these values to $380/m^2 ($35/ft^2) for pile, post, or pier foundations and $450/m^2 ($42/ft^2) for slab foundations in 1993 dollars. FEMA (1998) reported that for a 0.6- m (2-ft) elevation, elevating frame buildings with existing basement or crawl-space foundations onto continuous foundation walls or open foundations costs $180/m^2 ($17/ft^2) while elevating frame or masonry slab buildings costs $510/m^2 ($47/ft^2) in 1999 dollars. Newer guidance has moved away from providing elevation costs, as FEMA (2012) indicates that elevation cost relates to the type of construction and existing foundation but does not provide monetary values. In each of these documents, only mean cost values are reported, limiting consideration of the distribution of cost data. Most importantly, the effect of number of stories on elevation project cost is not mentioned in existing guidance. Thus, it is clear that updated cost guidance for existing home elevation projects is needed.

Predictive statistical cost modeling has been used in several construction cost applications (e.g., Herbsman 1986; Adeli and Wu 1998; Wilmot and Mei 2005), although not specific to home elevations. To predict construction cost, Karshenas (1984) used multiple regression, Skitmore and Ng (2003) used regression and cross-validation regression, and Kouskoulas and Koehn (1974) used multiple linear regression and validated the results with two real building case studies. Lowe et al (2006) used multiple linear regression, Jrade and Alkass (2007) developed a set of linear regression models in a computer-based cost estimation program, and Sonmez (2008) used a combination of linear regression and bootstrap techniques for construction cost modeling. Additionally, Shimizu et al. (2014) used switching regression model and generalized additive model (GAM) to predict the housing price, and Liu et al. (2018) used random forest and GAM to predict construction productivity using environmental factors. Specific to natural hazard mitigation, Jafarzadeh et al (2015) applied multiple linear regression to establish construction cost models for seismic retrofit of confined masonry buildings. Although statistical cost prediction models have been used for highways, commercial buildings, residential homes, and seismic retrofits, there are no known studies for existing building elevation cost prediction.

Conventional cost estimation methods are not readily accessible to decision-makers, and existing elevation cost guidance is limited and dated. Therefore, the goal of this paper is to evaluate and improve generalized home elevation construction cost estimation using predictive statistical modeling. This is accomplished by developing a robust, generalized cost estimation method for existing home elevations. Historical home elevation cost data obtained from the Louisiana Governor's Office of Homeland Security and Emergency Preparedness (GOHSEP) are categorized statistically using 10 regression models, a random forest model, and five GAMs with 10-fold cross-validation (CV) RMSE on all tested models. The required assumptions for each model are tested and the model with minimum prediction error is selected. Prediction results are compared with costs from USACE (1993), FEMA (1998), and Gair et al (2011) after modifying and updating them for time and location.

Both the methodology and the findings from the statistical model results are contributions of this research. First, previous statistical cost prediction research has evaluated limited models such as few regressions or GAMs; however, the method proposed in this research evaluates results from three robust statistical techniques, and external prediction accuracy of the selected models are examined. Second, the results themselves offer guidance to predict home elevation costs which enhance the flood mitigation decision-making and BCA (Taghinezhad et al., 2020a). Although the model results are applicable to Louisiana, the methodology itself can be applied for elevation mitigation project cost in other construction markets. Also, if the predicted elevation costs are adjusted for time and location, they may be representative of costs expected for similar buildings in similar construction markets.

BACKGROUND

Elevation project cost varies based on several factors [**Eq. 1**], where C is the cost of the elevation project ($), A is the average floor area (m^2) calculated as the total home area divided by the number of stories, ΔE is the change in first-floor elevation (FFE, m) calculated using **Eq. 2**, S is the number of stories, and F is a categorical variable representing foundation type. The FFE elevation (NAVD88) represents the top of the lowest floor (including basement, crawl-space, or enclosure floor) from elevation certificates, where FFE_0 and FFE_1 represent the FFE before and after elevation, respectively.

$$C = f(A, \Delta E, S, F) \qquad (1)$$
$$\Delta E = FFE_1 - FFE_0 \qquad (2)$$

DATA

Elevation Cost Literature

USACE (1993) calculates total cost of elevation (C_t; **Eq. 3**), where C_e is the cost of elevation; C_l represents the cost of landscaping excluding trees, bushes, and flowers; C_p is the cost of professional

engineering, and P_c is the contract profit percentage. Landscaping cost (C_l) is calculated using **Eqs. 4**, **5**, where A_l represents the landscaping area, C_{ul} represents the unit area landscaping cost, and W_b and L_b are the width and length of the building, respectively.

$$C_t = \left(C_e + C_l + C_p\right) \times \left(1 + P_c\right) \tag{3}$$

$$C_l = A_l \times C_{ul} \tag{4}$$

$$A_l = \left(W_b + 6.1\right) \times \left(L_b + 6.1\right) \left(\text{m}^2\right);$$
$$[A_l = \left(W_b + 20\right) \times \left(L_b + 20\right)] \left(\text{ft}^2\right) \tag{5}$$

According to USACE (1993) the cost elevation values for 0.6 m (2ft) additional elevation are as: The C_e for "wood frame building on piles, posts or piers," "wood frame building on foundation walls" "brick building," and "slab-on-grade building" are $280/\text{m}^2$ ($26/\text{sf}^2$), $205/\text{m}^2$ ($19/\text{sf}^2$), $344/\text{m}^2$ ($32/\text{sf}^2$), and $323/\text{m}^2$ ($30/\text{sf}^2$), respectively. The C_{ub}, C_p, P_c, and earthen fill are $6/\text{m}^2$ ($5/\text{yd}^2$), $7,000, 10\%, $13/\text{m3}$ ($10/\text{yd3}$), respectively. The slab foundation is assumed to be converted to elevated foundations; however, cost values for earthen fill are also provided. Also, it must be noted that values provided in USACE (1993) are assumed to represent 1993 dollars.

FEMA (1998) simply provides unit costs to elevate existing buildings to continuous foundation walls or open foundations by 0.6 m (2ft) of $510/\text{m}^2$ ($47/\text{ft}^2$) for frame or masonry buildings on slab foundations and $180/\text{m}^2$ ($17/\text{ft}^2$) for frame buildings with basement or crawlspace foundation, assuming 1998 costs.

Gair et al. (2011) evaluated elevation cost for typical 140 m^2 (1,500ft^2) one-story homes in Louisiana using unit-cost estimation and 2011 RSMeans residential cost data for slab and pier and beam foundations, elevated by 0.9 m (3ft), 1.8 m (6ft), and 2.7 m (9ft). However, because standard RSMeans cost data do not cover all construction activities required to elevate homes, Gair et al (2011) obtained unit cost values from a survey of foundation elevation contractors. Gair et al. (2011) divided the elevation process into 12 typical activities for Louisiana: push piling; raise, shore and align; footings; piers; wood stair; sanitary sewer; water; electrical; driveway and sidewalk pavement; platform for air conditioning (AC); remove/replace AC; and insulation below floor framing (per and beam only). Three additional activities are not typical for Louisiana: exterior wall; masonry stair; gas. The average cost/unit area/unit elevation for these three additional activities according to Gair et al. (2011) are 65.6 (1.9), 43.3 (1.2), and 9.0 (0.3), $ \text{m}^{-2}$ m^{-1} ($ \text{ft}^{-2}$ ft^{-1}), respectively.

Cost Adjustment

Cost information from the literature was normalized to represent 2015 dollars using the Engineering News-Record (ENR) average annual building cost index (i.e., average index, AI; (Grogan, 2016), which is commonly used by researchers in the construction industry (e.g., Popescu et al., 2003; Touran and Lopez 2006; Mikhed and Zemčík 2009). AI values have been determined considering nationwide changes (i.e., 20 cities) in labor rates, productivity, material prices, and the competitive condition of the building marketplace. The AI values (Grogan, 2016) are used to calculate project cost in terms of 2015 dollars

[**Eq. 6**], where C_{2015} is cost in 2015, AI_{2015} is the average index of the construction cost in 2015, AI_i is the average index at time i, and C_i represents cost at time i (i.e., either project contract date or year of previous study),. Historical AI values used for 1993, 1998, 2005, and 2015 are 2,996, 3,391, 4,205, and 5,517, respectively.

$$C_{2015} = \frac{AI_{2015}}{AI_i} \times C_i \tag{6}$$

National average project costs (C_{NA}) were adjusted to represent Louisiana costs (C_{LA}) using average location factor, P_l [**Eq. 7**], determined by averaging all Louisiana city RSMeans location factors (RSMeans, 2015). These factors ranged between 77.8 and 87.5%, with an average of 82.6%. Summarized costs are provided along with the results of this paper in **Table 1**

$$C_{LA} = P_l \times C_{NA} \tag{7}$$

Louisiana Elevation Project Data

Data were collected from scanned GOHSEP documents, corresponding to single-family homes elevated after major hurricane and flood events from 15 parishes (counties) in southern Louisiana between 2005 and 2015. Of the 805 total building records evaluated, the 666 with missing or spurious data were discarded from further analysis, thereby leaving 139 projects for statistical analysis. All cost data were adjusted to 2015 dollars, using the contract date as the original cost basis.

Seventy-one percent (71%) of the buildings had elevation certificates, from which elevation data were obtained. For the remaining buildings, FFE was obtained from other related building documents rather than the elevation certificate. The FFE in these documents was assumed to be the top of bottom floor (including basement, crawl-space, or enclosure floor) as specified in the elevation certificates.

Statistical summarization of variables used in the prediction model (**Table 2**) includes mean elevation cost per average floor area per unit ΔE ($825/\text{m}^2/\text{m}$), with a median of $821/\text{m}^2/\text{m}$, standard deviation of $425/\text{m}^2/\text{m}$, and range from $203/\text{m}^2/\text{m}$ to $2,151/\text{m}^2/\text{m}$.

The correlation matrix and boxplot for each variable enhance the understanding of collected data. The correlation matrix (**Table 3**) reveals the dependence between variables before statistical analysis. Cost correlates most strongly with number of stories, followed by ΔE. The elevation project cost boxplot shows many (13 out of 139) outliers above $500,000 (**Figure 1**). Data were weighted toward smaller values, which in turn indicates that the majority of collected data are associated with small and medium-sized homes. However, some outliers appear at the upper tail of the average floor area distribution. The ΔE boxplot shows that 67 out of 139 buildings (48%) were elevated in the range of 1.1 m (3.6 ft.) to 2.7 m (8.9ft). Data for ΔE data are slightly right-skewed but are normally distributed along the available range of elevation data.

Of the 139 elevation projects, 105 buildings are one-story, while 34 buildings are two-story. Four initial foundation types exist in the data: slab (116), crawl-space (2), pier and beam (15),

TABLE 1 | Elevation cost (cost/unit area) comparison between model 5 m and cost guidance, \$/m^2 (\$/ft^2).

ΔE	Slab foundation				Other foundation types			
	USACE	FEMA	Gair et al	Reg. Model 5 m	USACE	FEMA	Gair et al	Reg. Model 5 m
0.9 m (3 ft)	660 (61)	690 (64)	730 (68)	908 (84)	590 (55)	260 (24)	700 (65)	695 (65)
1.8 m (6 ft)	710 (66)	720 (67)	920 (86)	991 (92)	650 (60)	290 (27)	850 (79)	758 (70)
2.7 m (9 ft)	760 (71)	750 (70)	1,080 (99)	1,081 (100)	700 (65)	320 (30)	920 (85)	827 (77)

Note: USACE, FEMA, and Gair et al. costs were adjusted for Louisiana while regression costs were developed for Louisiana; all costs have been economically adjusted to represent 2015 dollars; there is no fill under any of the foundations in these estimates.

TABLE 2 | Statistical mean, median, standard deviation and range for 139 observations.

Variable	Description	Mean	Median	Standard Deviation	Range
C	Elevation cost[a]	\$241,160	\$179,567	\$172,665	[\$57,415:\$896,044]
A	Average floor area, m^2 (ft^2)	169 (1,820)	160 (1,720)	55 (590)	[54:361] [(580:3,890)]
ΔE	Delta elevation, m (ft)	1.9 (6.4)	1.9 (6.2)	0.9 (2.9)	[0.6:3.8] [(2.0:12.3)]
C/A/ΔE	Cost[a]/Unit area/Unit elevation, \$/m^2/m (\$/ft^2/ft)	830 (24)	820 (23)	430 (12)	[200:2,150] [(6:61)]

[a]*All costs have been economically adjusted to represent 2015 dollars.*

TABLE 3 | Correlation matrix for independent variables in sampled elevated homes in Louisiana, 2005–2015.

	C	A	ΔE	S	F
C	1.00*				
A	0.37*	1.00*			
ΔE	0.40*	0.05	1.00*		
S	0.71*	−0.13	0.32*	1.00*	
F	0.23*	0.16	−0.06	0.12	1.00*

Note: Asterisk in cells shows that correlation coefficient differs significantly from zero at p < 0.05.

and piling (6). Since there were only two levels of building stories in the data set, this variable was converted to a categorical variable with levels 0 and 1, representing one- and two-story buildings, respectively. In addition, slab foundations were the most predominant foundation type, with only 23 observations of other foundation types. Thus, the foundation type variable was also converted to a categorical variable, with levels 0 and 1, representing other and slab foundations, respectively.

METHODOLOGY

Multiple Regression

Statistical model prediction depends on the type of regression model and statistical characteristics of the data, including number of variables and the data distribution for each variable (Kim et al., 2004; Sousa et al., 2007; Atici 2011). Determination of the "best" or most appropriate model depends on the model evaluation criteria. In this study, these criteria are defined as: variable

significance, goodness of fit, 10-fold CV RMSE, and adherence to regression assumptions.

Variable Significance

Elevation project cost and average floor area data are non-normal and right-skewed. The elevation change data are slightly right-skewed; such skewness is reasonably expected to translate to the regression surface unless the cost values are transformed in the regression model to satisfy the assumption of normally distributed residuals. Therefore, the dependent cost variable and independent average floor area variable were transformed by a log-transformation, which is supported by other recent studies in construction cost prediction (e.g., Lowe et al., 2006; Jafarzadeh et al., 2015).

Ten statistical regression models were tested to find the best predictive model for determination of the estimated cost of elevation ($\widehat{C}$) [**Eqs. 8–17**], where $\widehat{\beta}_0$ is the estimated intercept, $\widehat{\beta}_i$ represents the estimated coefficient of regressor variable i, A is the average floor area (m^2), ΔE is elevation change (m), S represents the categorical number of stories variable, and F represents the categorical foundation type variable.

Model 1.

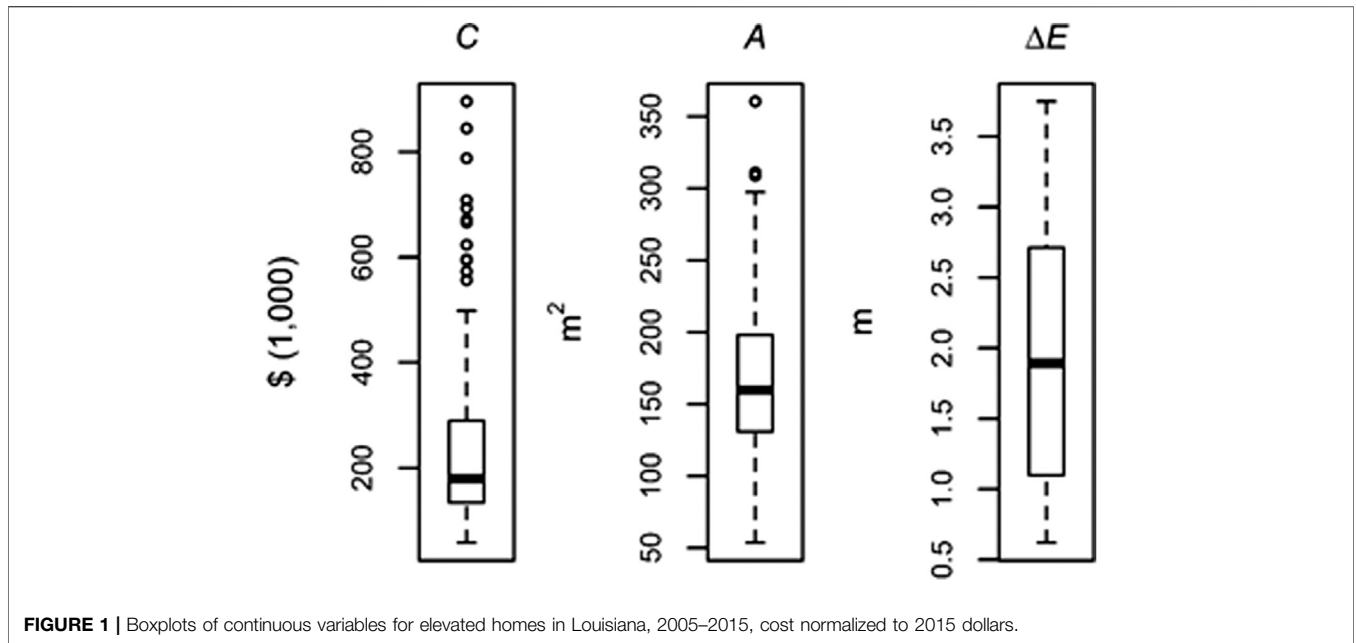$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E \tag{8}$$

Model 2.

$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 S + \widehat{\beta}_4 F \tag{9}$$

Model 3.

$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 \ln(A) + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 S + \widehat{\beta}_4 F \tag{10}$$

Model 4.

$$\ln(\widehat{C}) = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 S + \widehat{\beta}_4 F \tag{11}$$

FIGURE 1 | Boxplots of continuous variables for elevated homes in Louisiana, 2005–2015, cost normalized to 2015 dollars.

Model 5.

$$\ln\left(\widehat{C}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 \ln(A) + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 S + \widehat{\beta}_4 F \qquad (12)$$

Model 6.

$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 (A \times \Delta E) \qquad (13)$$

Model 7.

$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 (A \times \Delta E) + \widehat{\beta}_4 S + \widehat{\beta}_5 F \qquad (14)$$

Model 8.

$$\widehat{C} = \widehat{\beta}_0 + \widehat{\beta}_1 \ln(A) + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 ln(A \times \Delta E) + \widehat{\beta}_4 S + \widehat{\beta}_5 F \qquad (15)$$

Model 9.

$$\ln\left(\widehat{C}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 A + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 (A \times \Delta E) + \widehat{\beta}_4 S + \widehat{\beta}_5 F \qquad (16)$$

Model 10.

$$\ln\left(\widehat{C}\right) = \widehat{\beta}_0 + \widehat{\beta}_1 \ln(A) + \widehat{\beta}_2 \Delta E + \widehat{\beta}_3 ln(A \times \Delta E) + \widehat{\beta}_4 S + \widehat{\beta}_5 F \qquad (17)$$

Model 1 was fit only with continuous variables, and Model 2 expands Model 1 with the addition of both S and F. Model 3 is the same as Model 2, but with logarithmic transformation of the continuous independent variable A, while Model 4 is the same as Model 2 but with logarithmic transformation of the response variable, also known as an exponential model. Model 5, known as a log-semi-log model, is the same as Model 3 with logarithmic transformation of the response variable and A. Models 6 through 10 are the same as the first five models, with the addition of a term representing the interaction between A and ΔE, which is transformed logarithmically in Models 8 and 10. Coefficient estimates, standard errors, and p-values were determined using R (www.r-project.org) for each of the ten models.

Regression Assumptions

For multiple linear regression, three main assumptions were tested: homoscedasticity, multicollinearity, and normality of the residuals. Homoscedasticity was tested through the Breusch-Pagan test (Breusch and Pagan 1979), with multicollinearity tested using the variance inflation factor (VIF). In models that consider interaction, multicollinearity always exists, and the VIF was not evaluated. Normality was tested using the Shapiro-Wilk test (Shapiro and Wilk 1965). Violation of the normality assumption decreases the robustness of regression results when the sample size was not large enough (Lumley et al., 2002). In some cases the violation of regression assumptions can be resolved by nonlinear transformations of regression variables (Montgomery et al., 2015) and by trimming problematic observation outliers (Andersen, 2008).

Before removing model outliers, each problematic observation was evaluated for any distinguishing features, leverage, r-student residual, and Cook's distance. An outlier with a large leverage value is an influential point because it can change the regression results. Cook's distance is another statistical measure that measures the influence of each observation in the model.

The coefficient of determination (R^2) is a statistical parameter that indicates goodness of fit between predicted and observed values; however, to compare the goodness of fit for multiple models that consider non-equal numbers of independent variables, the R^2 can be misleading because the value increases as the number of regressor variables increase. Therefore, to better represent goodness of fit for model comparison, the adjusted R^2 (R^2_{adj}) was calculated.

10-Fold Cross-Validation Root Mean Square Error

The RMSE was used to measure the error rate of prediction models. In order to obtain the RMSE, a prediction model was

constructed on training data and was then used to predict data for the test set. The *RMSE* was obtained by examining the test set data on a training set fitted model [**Eq. 18**], where n is the number of observations for prediction of the test set data, $\hat{Y}_t$ is the predicted value of observation t in the test set data, and Y_t is the actual value of observation t in the test set data.

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} \left(\hat{Y}_t - Y_t\right)^2} \qquad (18)$$

Sometimes RMSE values resulting from only one training and one test set become sensitive to the selection of data for each set. Therefore, obtaining RMSE with K-fold CV ($K > 2$) is preferable (Zhang et al., 2011). Based on the recommendation of Kohavi (1995), this paper uses 10-fold CV for multiple regression to select the best prediction model. In each fold, the prediction error $RMSE_i$ was calculated, and the mean of all prediction errors (E) is the 10-fold CV RMSE for the prediction model (Priddy and Keller, 2005), where $RMSE_i$ is the RMSE for fold i [**Eq. 19**].

$$E = \frac{1}{10} \sum_{i=1}^{10} RMSE_i \qquad (19)$$

Random Forest

Random forest (Breiman, 2001) is a robust data mining model used for both prediction (i.e., regression) and classification. This ensemble method was constructed based on the equal averaging of many random trees in the classification and regression tree (CART) method (Breiman, 2001) to obtain a model with reduced variance. In the random forest, every tree was created by a bootstrap sample from the training data, and the tree grows to a maximum depth without pruning (Breiman, 2001; Cutler et al., 2007). The random forest algorithm selects regressor variables randomly at each node. Additionally, the random forest is useful for ranking regressor variables by their importance in prediction. The "randomForest" package in the R program was used for random forest analysis in this study.

Generalized Additive Model

The GAM is used to identify the relationship between input and output variables in nonlinear models. It relaxes the strictly linear relationship between the response and the regressors, allowing regressors to have a general and flexible relationship to the response, but maintains additive or non-interactive structure (Moore et al., 2011; Shimizu et al., 2014; Larsen, 2015; Taghinezhad et al., 2020b). Although we do not consider it here, GAMs can additionally accommodate non-normal responses with added flexibility through a nonlinear link function (Xiang, 2001; Han et al., 2009; Calabrese and Osmetti, 2015). This study used the "gam" package (Hastie, 2020) in the R program to fit the GAM. The smoothing function of spline fit on continuous variables of A and ΔE is applied to the model. To obtain the optimum fit with the lowest RMSE, the models are varied based on applying the logarithmic

TABLE 4 | Parameter estimate, standard error, and *p*-value for multiple regression models.

Model #	Coefficient	Parameter	Estimate	Std. Error	*p*-value
1	$\hat{\beta}_0$	Intercept	−92,079	48,266	0.058
	$\hat{\beta}_1$	A	1,110	230	<0.001 *
	$\hat{\beta}_2$	ΔE	75,292	14,296	<0.001*
2	$\hat{\beta}_0$	Intercept	−155,123	32,080	<0.001*
	$\hat{\beta}_1$	A	1,397	141	<0.001*
	$\hat{\beta}_2$	ΔE	30,492	9,078	0.001*
	$\hat{\beta}_3$	S	284,495	18,765	<0.001*
	$\hat{\beta}_4$	F	38,510	20,613	0.064
3	$\hat{\beta}_0$	Intercept	−110,7306	114,089	<0.001*
	$\hat{\beta}_1$	ln (A)	233,401	22,712	<0.001*
	$\hat{\beta}_2$	ΔE	34,503	8,898	<0.001*
	$\hat{\beta}_3$	S	282,077	18,420	<0.001*
	$\hat{\beta}_4$	F	33,822	20,366	0.099
4	$\hat{\beta}_0$	Intercept	1.056E+01	9.470E-02	<0.001*
	$\hat{\beta}_1$	A	5.862E-03	4.161E-04	<0.001*
	$\hat{\beta}_2$	ΔE	1.003E-01	2.680E-02	<0.001*
	$\hat{\beta}_3$	S	9.474E-01	5.539E-02	<0.001*
	$\hat{\beta}_4$	F	2.643E-01	6.085E-02	<0.001*
5	$\hat{\beta}_0$	Intercept	6.641	0.340	<0.001*
	$\hat{\beta}_1$	ln (A)	0.964	0.068	<0.001*
	$\hat{\beta}_2$	ΔE	0.118	0.026	<0.001*
	$\hat{\beta}_3$	S	0.935	0.055	<0.001*
	$\hat{\beta}_4$	F	0.247	0.061	<0.001*
6	$\hat{\beta}_0$	Intercept	−52,487	109,178	0.631
	$\hat{\beta}_1$	A	877	621	0.160
	$\hat{\beta}_2$	ΔE	56,872	47,733	0.236
	$\hat{\beta}_3$	(A × ΔE)	108	266	0.686
7	$\hat{\beta}_0$	Intercept	−45,595	69,118	0.511
	$\hat{\beta}_1$	A	791	367	0.033*
	$\hat{\beta}_2$	ΔE	−19,244	29,281	0.512
	$\hat{\beta}_3$	(A × ΔE)	286	160	0.077
	$\hat{\beta}_4$	S	288,621	18,757	<0.001*
	$\hat{\beta}_5$	F	31,394	20,832	0.134
8	$\hat{\beta}_0$	Intercept	−1,123,570	119,220	<0.001*
	$\hat{\beta}_1$	ln (A)	271,491	81,704	0.001*
	$\hat{\beta}_2$	ΔE	55,847	44,864	0.215
	$\hat{\beta}_3$	ln (A × ΔE)	−38,662	79,641	0.628
	$\hat{\beta}_4$	S	282,645	18,510	<0.001*
	$\hat{\beta}_5$	F	32,388	20,637	0.119
9	$\hat{\beta}_0$	Intercept	1.051E+01	2.064E-01	<0.001*
	$\hat{\beta}_1$	A	6.172E-03	1.096E-03	<0.001*
	$\hat{\beta}_2$	ΔE	1.257E-01	8.744E-02	0.153
	$\hat{\beta}_3$	(A × ΔE)	−1.459E-04	4.777E-04	0.760
	$\hat{\beta}_4$	S	9.453E-01	5.601E-02	<0.001*
	$\hat{\beta}_5$	F	2.680E-01	6.221E-02	<0.001*
10	$\hat{\beta}_0$	Intercept	6.638	0.355	<0.001*
	$\hat{\beta}_1$	ln (A)	0.970	0.243	<0.001*
	$\hat{\beta}_2$	ΔE	0.121	0.134	0.368
	$\hat{\beta}_3$	ln (A × ΔE)	−0.005	0.237	0.982
	$\hat{\beta}_4$	S	0.935	0.055	<0.001*
	$\hat{\beta}_5$	F	0.247	0.061	<0.001*

transformation on C and A variables and also changing the degrees of freedom in spline fit smoothing functions (i.e., 4, 2, and 1) because changing degree of freedom tunes the flexibility in the regressors, and is thus explored as a hyperparameter. In GAM Models 11–15, g represents the identity link with normal response, $\hat{s}$ represents the smoothing function of spline fit, and df represents the degree of freedom.

Model #	Homosceda-sticity	Multicollin-earity	Normality	R^2	Adjusted R^2	10-Fold CV RMSE
1	F	p	F	0.28	0.27	133,324
2	F	p	F	0.75	0.74	86,447
3	F	p	F	0.76	0.75	85,436
4	p	p	p	0.82	0.81	70,393
5	F	p	F	0.82	0.81	63,618
6	F	NA	F	0.28	0.27	134,127
7	F	NA	F	0.76	0.75	86,138
8	F	NA	F	0.76	0.75	87,216
9	p	NA	p	0.82	0.81	71,070
10	F	NA	F	0.82	0.82	64,127

Note: p = pass, F = fail, NA = not applicable.

Model 11.

$$g(\widehat{C}) = \widehat{\beta}_0 + \widehat{s}(A, df = 4) + \widehat{s}(\Delta E, df = 4) + \widehat{\beta}_1 S + \widehat{\beta}_2 F \quad (20)$$

Model 12.

$$g\left[ln(\widehat{C})\right] = \widehat{\beta}_0 + \widehat{s}(ln(A), df = 4) + \widehat{s}(\Delta E, df = 4) + \widehat{\beta}_1 S + \widehat{\beta}_2 F \quad (21)$$

Model 13.

$$g\left[ln(\widehat{C})\right] = \widehat{\beta}_0 + \widehat{s}(ln(A), df = 2) + \widehat{s}(\Delta E, df = 2) + \widehat{\beta}_1 S + \widehat{\beta}_2 F \quad (22)$$

Model 14.

$$g\left[ln(\widehat{C})\right] = \widehat{\beta}_0 + \widehat{s}(ln(A), df = 2) + \widehat{\beta}_1 \Delta E + \widehat{\beta}_2 S + \widehat{\beta}_3 F \quad (23)$$

Model 15.

$$g\left[ln(\widehat{C})\right] = \widehat{\beta}_0 + \widehat{\beta}_1 ln(A) + \widehat{s}(\Delta E, df = 2) + \widehat{\beta}_2 S + \widehat{\beta}_3 F \quad (24)$$

Model 11 is the GAM with four degrees of freedom on smoothing functions, Model 12 includes a logarithmic transformation of the response variable and A with inclusion of smoothing function on the continuous variables of A and ΔE. Model 13 is the same as Model 12 but with two degrees of freedom on smoothing functions. Finally, Models 14 and 15 are the same as Model 13 but with smoothing function on only A or ΔE, respectively. It must be noted that the response variable in all the GAMs have identity link function with normal response.

RESULTS

Multiple Regression

The parameter estimate, standard error, and significance p-value of each variable for all ten models are shown in **Table 4**. The results indicate that the p-values of all selected variables in Models 1, 2, 3, and 6 are less than the significance level of 0.05, indicating that all variables in these four models have significant impacts on the dependent cost variable. The standard error shows the variability of each parameter estimate applicable to the regression model. Of these, only Models 4 and 5 show significance of all independent variables with low standard errors.

The criteria for selecting the best among the ten proposed models are the fulfillment of the statistical regression assumptions, p-value significance for all independent variables, adjusted R^2, and minimization of 10-fold CV RMSE. According to **Table 5** the only models passing the main assumptions of multiple linear regression are the exponential models (i.e., Models 4 and 9 with log transformation of dependent variable C).

Although Model 4 appears to be the preferred model for the first three criteria, Model 5 has a lower 10-fold CV RMSE with equal adjusted R^2. However, regression assumptions of normality and homoscedasticity of residuals were not satisfied. In the residual plots of normal Q-Q, scale location, and residuals vs. leverage (**Figure 2**), observations numbered 77, 100, and 101 were detected as problematic observations (2% of total).

Examination of the corresponding buildings for these observations revealed that they are extraordinary projects with an unusual A or E (**Table 6**). For instance, observation #77 has a very low building cost while the building area is large. Therefore, in Model 5m, these three observations were excluded from Model 5, which then satisfied the regression assumptions (**Figure 3**).

Table 7 provides the estimated coefficients, standard errors, and p-values for the Model 5 mm parameters. The p-values are significant for all parameters in the model and the high R^2 and adjusted R^2 values of 0.86 and 0.85, respectively, indicate a good fit between data and model. Additionally, the 10-fold CV RMSE is decreased and changed to 61,542. The results for the Model 5 m reveal no violation of tested assumptions (i.e., the p-value of the Shapiro-Wilk test for the normality assumption is 0.063, the p-value of the Breusch-Pagan test for the homoscedasticity assumption is 0.559, and the VIF results for all regressor variables are less than the threshold of 10 [$VIF_A = 1.06$, $VIF_{\Delta E} = 1.14$, $VIF_S = 1.18$, $VIF_F = 1.04$]).

Random Forest and Generalized Additive Model

The random forest model out-of-bag (OOB) error decreased dramatically with the first 50 trees, after which the test-error becomes nearly constant (**Figure 4**). Therefore, random forest is applied with 800 trees to obtain the best results. The random forest variable importance option indicates that S, A, ΔE, and F are the most important variables in the random forest model, in order.
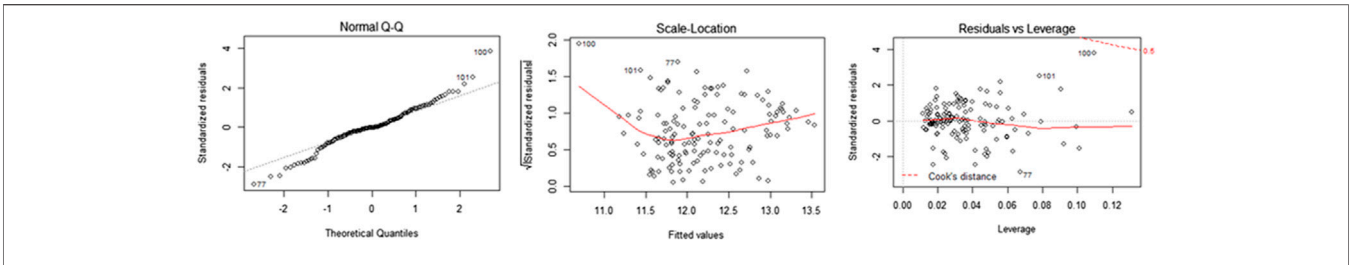
FIGURE 2 | Model 5 residuals plots of normal Q-Q, scale location, and residuals vs. leverage.

TABLE 6 | Outlier observations in model 5 with the description of the issue.

N	C	A	E	S	F	Issue	Leverage	R-student	Cooks D
77	$71,051	206	0.9	0	0	Low cost; big size	0.07	−2.97	0.12
100	$111,767	54	1.9	0	0	Very small size	0.11	4.03	0.36
101	$172,775	99	3.4	0	0	Very high elevation	0.08	2.56	0.11
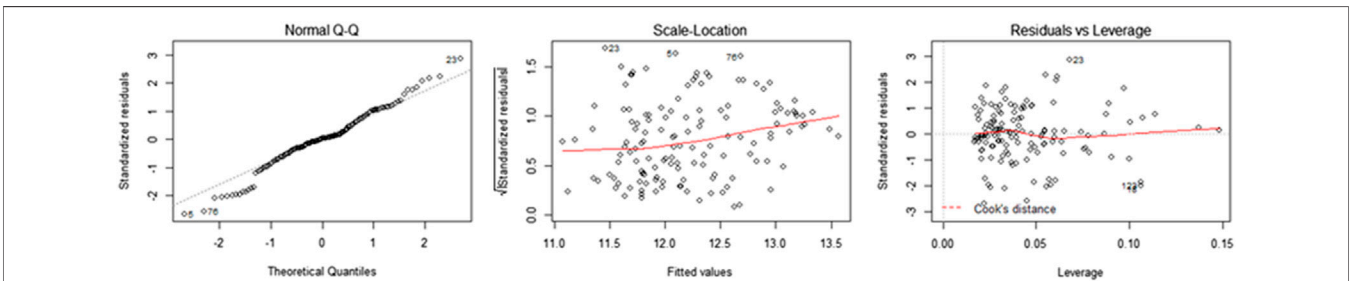
FIGURE 3 | Model 5 m residuals plots of normal Q-Q, scale location, and residuals vs. leverage (Model 5 after deleting observations 77, 100, and 101).

TABLE 7 | Parameter estimate, standard error, and p-value for multiple regression model 5 m.

Coefficient	Parameter	Estimate	Std. Error	p-value
$\hat{\beta}_0$	*Intercept*	6.062 (3.495)	0.319 (0.467)	<0.001 *
$\hat{\beta}_1$	ln(A)	1.080	0.063	<0.001 *
$\hat{\beta}_2$	ΔE	0.096 (0.029)	0.024 (0.007)	<0.001 *
$\hat{\beta}_3$	S	0.969	0.049	<0.001 *
$\hat{\beta}_4$	F	0.268	0.057	<0.001 *

Note: The values in parentheses reflect U.S. units.

The 10-fold CV RMSE for the random forest model is 72,843, which is greater than the best regression model. The RMSEs for five GAMs on Models 11–15 are: 89,728, 68,080, 65,182, 64,641, and 64,200, respectively. The results show that Model 15 with logarithmic transformation of response and A variables and spline smoothing on ΔE variable with two degrees of freedom has the best RMSE among all the other GAMs. The partial residual plots of this model show the nonlinear effect of regressors ln(A) and ΔE (**Figure 5**). We find that ΔE is essentially linear in nature, whereas the ln(A) effect requires mild flexibility.

The 10-fold CV RMSEs in the statistical cost estimation models show that the regression Model 5 m (10-fold CV
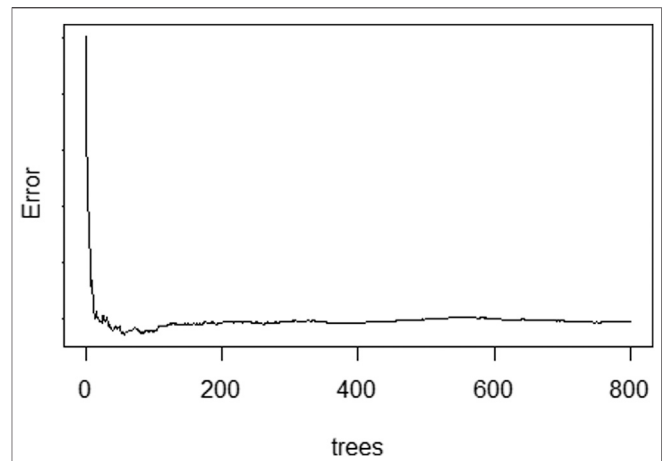

FIGURE 4 | Random forest OOB error based on the number of trees.

RMSE = 61,542) has the best prediction capability. Therefore, this model is selected to use in this research to compare with the elevation costs on the literature. The cost predictions by this model are shown in Appendix Table A1. **Figure 6** shows the
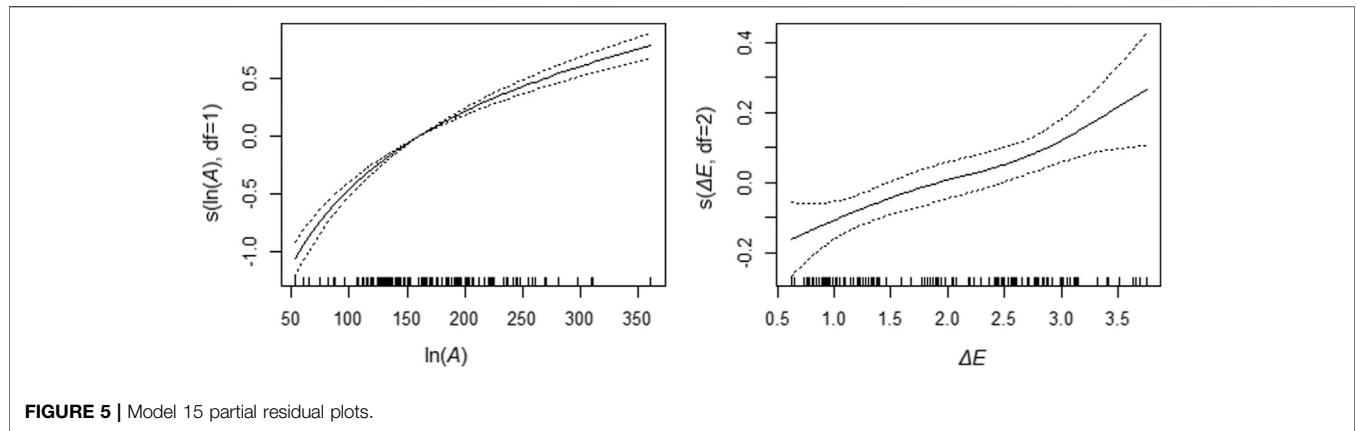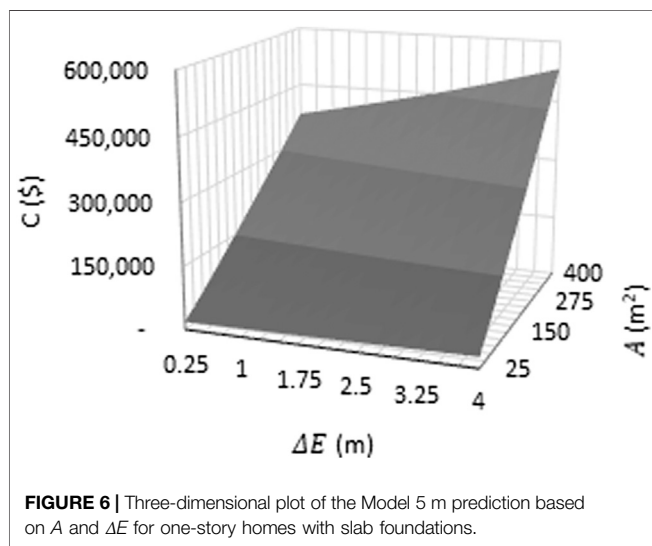
FIGURE 5 | Model 15 partial residual plots.

FIGURE 6 | Three-dimensional plot of the Model 5 m prediction based on A and ΔE for one-story homes with slab foundations.

predicted project cost calculated using the Model 5 m based on A and ΔE for homes with one-story and slab foundation. The other choices of S and F have exactly the same surface, but shifted vertically. The additive structure, and that perhaps GAMs, although having similar structure (see partial residual plots), are overfitting the smooth relationship and thus mildly suffers with external prediction.Comparison With Cost Literature

In this section, the regression Model 5 m predictions are compared with the USACE (1993), FEMA (1998), and Gair et al (2011) estimates previously described. As a fair basis for comparison, all estimates are adapted to 2015 dollars using **Eq. 6** and Louisiana location using **Eq. 7**. In both Gair et al. (2011) and USACE (1993), the general contractor's charge for overhead and profit is considered to be 10% of the estimated final costs according to the recommendations by these two guidelines. Additionally, Gair et al (2011) estimates include a 5.9% charge for insurance and a 20% contingency factor due to the uncertainty and any unpredicted issue that may happen during the construction work. According to instructions for USACE

(1993) estimates, the professional engineering design and landscaping costs must be added to original represented costs in USACE (1993) for elevation.

Table 1 shows the elevation cost based on USACE (1993), FEMA (1998), and Gair et al (2011) cost guidance and regression prediction for one-story buildings in six specific case studies. In all examined case studies, elevation of buildings with existing slab foundations is more expensive than elevation of buildings with other foundation types.

Figure 7 demonstrates graphically the difference between the predicted elevation cost using regression models and cost guidance estimates. The results indicate that USACE (1993) and FEMA (1998) estimates are lower than those in Gair et al. (2011) and regression approaches employed here.

DISCUSSION

The statistical prediction model is based on the generalization from real and completed elevation projects; therefore, it gives a more realistic estimation with actual cost varieties in the market. Additionally, because a wide range of buildings with different conditions was used in the statistical prediction model, it is able to predict cost based on simple achievable building attributes. The elevation cost comparison in **Table 1** and **Figure 7** shows that elevating other foundation types is considerably less expensive than elevating slab foundations. Also, for slab foundation elevation, USACE and FEMA guidance underpredict Louisiana elevation costs; for other foundations, FEMA continues to underpredict, but USACE is closer to Louisiana costs.

The partial plot of the selected GAM model shows that cost has a nonlinear relationship with building average floor area. Therefore, the previous cost guidance (USACE, 1993; FEMA, 1998; Gair et al., 2011) that estimates elevation cost only with a single building size, and then generalizes the cost based on that case study, biases results in buildings with different average floor area. Furthermore, the random forest model shows that the number of stories is the most important variable in prediction of elevation project cost, but this variable is not included in current elevation cost guidance.
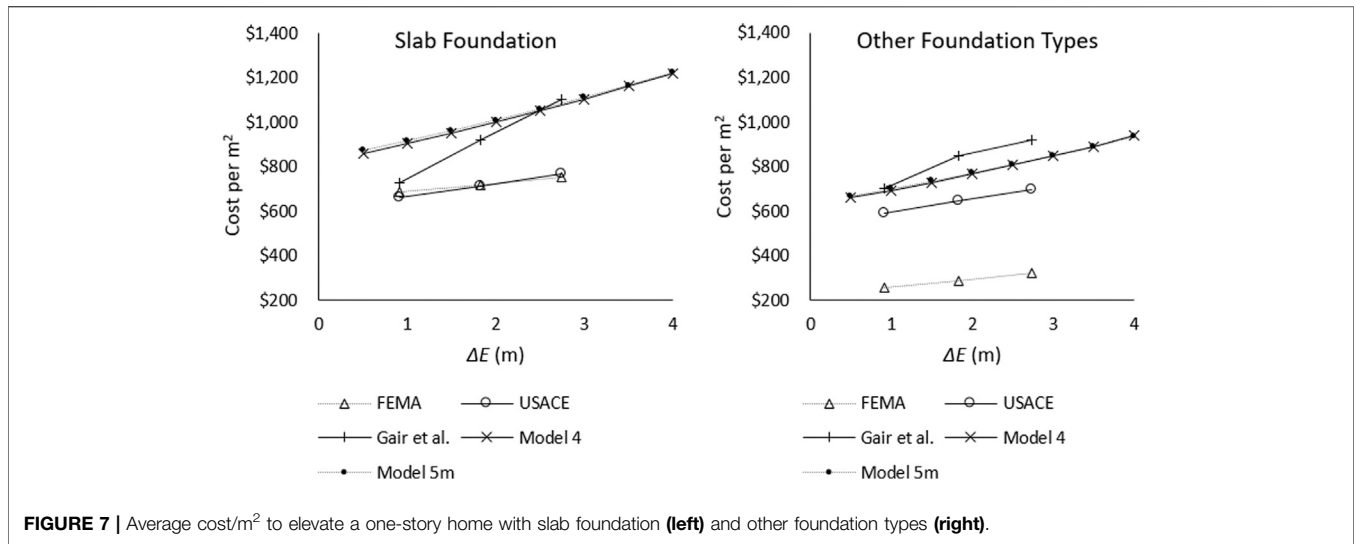
FIGURE 7 | Average cost/m² to elevate a one-story home with slab foundation **(left)** and other foundation types **(right)**.

However, none of the three above-mentioned guidelines have evaluated the effect of important variables such as the building average floor area and number of stories. The USACE (1993) and FEMA (1998) estimates are lower than the newer estimates by Gair et al (2011) and statistical prediction models. The differences may come from changing the construction techniques and equipment over time, and the inherent error in cost adjustment over time. This result suggests that the USACE (1993) and FEMA (1998) guidelines do not have advantages over the newer estimates by Gair et al. (2011) and the statistical prediction models described here. The Gair et al. (2011) study is more conservative than other cost guidance because it considers the 25% contingency factor for any unpredictable construction activities.

Among the tested regression models, Model 5 has the best external prediction ability, with all significant coefficient variables, higher adjusted R^2, and lower 10-fold CV RMSE. But unlike Model 4, which satisfies all regression assumptions, the normality and homoscedasticity assumptions may be violated based on the p-values of these tests, which fall below the significance level of 0.05. Therefore, this study suggests using the modified Model 5 (i.e., Model 5 m) with trimmed outliers, because it passes all regression assumptions. However, the trimmed otliers did not considerably change the trendline of Model 5 as the plots of Models 5 and 5 m are nearly identical (**Figure 6**). The random forest and GAM prediction accuracy are inferior to that of regression Models 5 and 5 m. Accordingly, the regression Model 5 m has a better prediction ability for C among all the models and is selected for use in this study. Also, the regression models are preferable to random forest and GAM in ease of interperation and prediction of the results because the equation and estimated coeffcents can be used easily to estimate the dependent variable without using sophisticated computer programs.

The cost as calculated in statistical predictions can change based on variables that do not exist in the current guidelines. However, regression Model 5 m shows a substantial agreement between its predictions and the guidelines. For instance, there is a difference of between 0.1 and 24.4% in the Model 5 m estimates vs. Gair et al. (2011) case studies. Therefore, the results suggest that project cost prediction with regression Model 5 m enhances future BCA for flood-mitigated properties.

CONCLUSION AND SUMMARY

To provide a series of building elevation project cost case studies based on cost guidance, this study adjusted the costs in the available guidance to represent those in year 2015 for a Louisiana location. According to the cost guidance results for single-family homes with three levels of elevation and three disparate cost analyzing methods, the occupancy phase elevation cost with USACE estimation is between \$590/m² (\$55/ft²) and \$760/m² (\$71/ft²), with FEMA estimation falling between \$260/m² (\$24/ft²) and \$750/m² (\$70/ft²), and the Gair et al. (2011) method suggesting between \$700/m² (\$65/ft²) and \$1,100/m² (\$99/ft²).

To find an appropriate statistical prediction model, ten regression models along with one random forest model and five GAMs were studied for cost modeling. The correlation matrix prior to regression analysis shows the existence of correlation between cost and all independent variables. However, according to the random forest variable importance function, elevation cost is most strongly affected by the number of stories — an attribute that has been neglected in previous elevation cost guidance — and change in elevation.

The regression 10-fold CV RMSE results suggest that a log-semi-log model without an interaction term and with trimmed outliers (i.e., Model 5 m) has the lowest RMSE among the tested regression models. In addition, this model makes all independent variables significant with no violation of statistical assumptions and high goodness of fit with R^2 of 0.85. Therefore, the results suggest that regression models can be used successfully in project cost prediction for elevation projects to address the cost issue in BCA and to overcome barriers in existing cost guidance methods.

The regression study shows that for projects undertaken in Louisiana with adjusted costs to 2015 dollars, the elevation costs for slab foundations are \$908/m² (\$84/ft²) to elevate 3 ft, \$991/m² (\$92/ft²) to elevate 6ft, and \$1,081/m² (\$100/ft²) to elevate 9ft. The elevation costs for other foundation types are \$695/m² (\$65/ft²) to elevate 3ft, \$758/m² (\$70/ft²) to elevate 6ft, and \$827/m² (\$77/ft²) to elevate 9ft.

In recent decades new data collection technologies make data more available for analysis in machine learning prediction models.

The results suggest that statistical data prediction models in this study can be used successfully in cost estimation for construction projects, especially for estimation of project costs in natural hazard mitigation projects. However, the statistical modeling of cost in this study suggests that proper model selection is important for improving model prediction. For instance, the RMSE in regression modeling can be improved substantially by choosing proper independent variables and transformation on regression variables specifically when the variables are not distributed normally. The random forest error is decreased by selection of the proper number of trees and the RMSE in GAM analysis can be improved by transformation of variables, applying the smoothing functions on proper variables, and changing the degrees of freedom for smoothing functions.

In future studies, the same methodology can be used for prediction of elevation cost for new buildings during the construction phase. Such information would be useful for adjusting economically the elevation mitigation benefits for new buildings and comparing that estimate with elevation cost in the occupancy phase. Additionally, by knowing the additional cost of elevation in new construction, builders could offer the choice of freeboard (elevation higher than BFE) to the owners as an option for construction in floodprone areas. Also in future studies, the mitigation cost can be predicted by statistical methods for other types of mitigation projects, such as hurricane and tornado wind mitigations.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the Louisiana Governor's Office of Homeland Security and Emergency Preparedness. Restrictions apply to the availability of these data, which were used under license for this study.

AUTHOR CONTRIBUTIONS

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Adeli, H., and Wu, M. (1998). Regularization Neural Network for Construction Cost Estimation. *J. Construction Eng. Manag.* 124 (1), 18–24. doi:10.1061/(asce)0733-9364(1998)124:1(18)

Amoroso, S. D., and Fennell, J. P. (2008). A Rational Benefit/cost Approach to Evaluating Structural Mitigation for Wind Damage: Learning "the Hard Way" and Looking Forward. *Proc. Structures Congress*, 1–10. doi:10.1061/41016(314)249

Andersen, R. (2008). *Modern Methods for Robust Regression*. Newbury Park, CA, United States: Sage. doi:10.4135/9781412985109

Atici, U. (2011). Prediction of the Strength of mineral Admixture concrete Using Multivariable Regression Analysis and an Artificial Neural Network. *Expert Syst. Appl.* 38 (8), 9609–9618. doi:10.1016/j.eswa.2011.01.156

Bellomo, D., Pajak, M. J., and Sparks, J. (1999). Coastal Flood Hazards and the National Flood Insurance Program. *J. Coastal Res.*, 21–26.

Bohn, F. (2013). "Design Flood Elevations beyond Code Requirements and Current Best Practices,". Master's thesis(Baton Rouge, LA: . Louisiana State University).

Breiman, L. (2001). Random Forests. *Machine Learn.* 45 (1), 5–32. doi:10.1023/a:1010933404324

Breusch, T. S., and Pagan, A. R. (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation. *Econometrica* 47, 1287–1294. doi:10.2307/1911963

Calabrese, R., and Osmetti, S. A. (2015). Improving Forecast of Binary Rare Events Data: A GAM-Based Approach. *J. Forecast.* 34 (3), 230–239. doi:10.1002/for.2335

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., Hess, K. T., Gibson, J., et al. (2007). Random Forests for Classification in Ecology. *Ecology* 88 (11), 2783–2792. doi:10.1890/07-0539.1

FEMA (2012). "Engineering Principles and Practices of Retrofitting Floodprone Residential Structures," in *Department of Homeland Security* (Washington, DC: Federal Emergency Management Agency).

FEMA (2010). "Home Builder's Guide to Coastal Construction," in *Department of Homeland Security* (Washington, D.C: Federal Emergency Management Agency).

FEMA (1998). "Homeowner's Guide to Retrofitting; Six Ways to Protect Your House from Flooding," in *Department of Homeland Security* (Washington, D.C: Federal Emergency Management Agency).

FEMA (2011). "Supplement to the Benefit-Cost Analysis Reference Guide," in *Department of Homeland Security* (Washington, DC: Federal Emergency Management Agency).

Gair, R., Balboni, B., Phelan, M., Charest, A., Waier, P., and Mossman, M. J. (2011). *Final Report–Louisiana House Raising–Slab on Grade and Pier and Beam Construction*. RS Means: Reed Construction Data.

Grogan, T. (2016). How to Use ENR's Indexes. Available at: http://enr.construction.com/economics/historical_indices/(Accessed August 10, 2017). doi:10.4324/9781315255712

Han, S.-R., Guikema, S. D., and Quiring, S. M. (2009). Improving the Predictive Accuracy of Hurricane Power Outage Forecasts Using Generalized Additive Models. *Risk Anal.* 29 (10), 1443–1453. doi:10.1111/j.1539-6924.2009.01280.x

Hastie, T. (2020). Generalized Additive Models: Package 'gam'. Available at: https://cran.r-project.org/web/packages/gam/gam.pdf (Accessed December 1, 2020).

Herbsman, Z. (1986). *Model for Forecasting Highway Construction Cost*. Gainesville, FL, United States: Transportation Research Record (1056).

Jafarzadeh, R., Ingham, J., Walsh, K., Hassani, N., and Ghodrati Amiri, G. (2015). Using Statistical Regression Analysis to Establish Construction Cost Models for

Seismic Retrofit of Confined Masonry Buildings. *J. Construction Eng. Manag.* 141 (5), 04014098. doi:10.1061/(asce)co.1943-7862.0000968

Jrade, A., and Alkass, S. (2007). Computer-integrated System for Estimating the Costs of Building Projects. *J. Archit. Eng.* 13 (4), 205–223. doi:10.1061/(asce)1076-0431(2007)13:4(205)

Karshenas, S. (1984). Predesign Cost Estimating Method for Multistory Buildings. *J. Construction Eng. Manag.* 110 (1), 79–86. doi:10.1061/(asce)0733-9364(1984)110:1(79)

Kim, G.-H., An, S.-H., and Kang, K.-I. (2004). Comparison of Construction Cost Estimating Models Based on Regression Analysis, Neural Networks, and Case-Based Reasoning. *Building Environ.* 39 (10), 1235–1242. doi:10.1016/j.buildenv.2004.02.013

Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. Proceedings of the 14th international joint conference on Artificial intelligence, Montreal, Quebec, Canada Vol. 2, 1137–1145.

Kouskoulas, V., and Koehn, E. (1974). Predesign Cost Estimating Function for Buildings. *J. Construct. Div.* 100 (4), 589–604. doi:10.1061/JCCEAZ.0000461

Larsen, K. (2015). "GAM: The Predictive Modeling Silver Bullet," in *Multithreaded* (San Francisco, CA, United States: Stitch Fix), 30.

Li, Y., and van De Lindt, J. W. (2012). Loss-based Formulation for Multiple Hazards with Application to Residential Buildings. *Eng. Structures* 38, 123–133. doi:10.1016/j.engstruct.2012.01.006

Liu, X., Song, Y., Yi, W., Wang, X., and Zhu, J. (2018). Comparing the Random Forest with the Generalized Additive Model to Evaluate the Impacts of Outdoor Ambient Environmental Factors on Scaffolding Construction Productivity. *J. Construction Eng. Manag.* 144 (6), 04018037. doi:10.1061/(asce)co.1943-7862.0001495

Lowe, D. J., Emsley, M. W., and Harding, A. (2006). Predicting Construction Cost Using Multiple Regression Techniques. *J. Constr. Eng. Manage.* 132 (7), 750–758. doi:10.1061/(asce)0733-9364(2006)132:7(750)

Lumley, T., Diehr, P., Emerson, S., and Chen, L. (2002). The Importance of the Normality assumption in Large Public Health Data Sets. *Annu. Rev. Public Health* 23 (1), 151–169. doi:10.1146/annurev.publhealth.23.100901.140546

Mikhed, V., and Zemčík, P. (2009). Do house Prices Reflect Fundamentals? Aggregate and Panel Data Evidence. *J. Housing Econ.* 18 (2), 140–149. doi:10.1016/j.jhe.2009.03.001

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2015). *Introduction to Linear Regression Analysis*. United States: John Wiley & Sons.

Moore, L., Hanley, J., Turgeon, A., and Lavoie, A. (2011). A Comparison of Generalized Additive Models to Other Common Modeling Strategies for Continuous Covariates: Implication for Risk Adjustment. *J. Biomet Biostat* 2 (1), 109. doi:10.4172/2155-6180.1000109

Orooji, F., and Friedland, C. J. (2017). Cost-benefit Framework to Generate Wind hazard Mitigation Recommendations for Homeowners. *J. Architectural Eng.* 23 (4), 04017019. doi:10.1061/(asce)ae.1943-5568.0000269

Popescu, C. M., Phaobunjong, K., and Ovararin, N. (2003). *Estimating Building Costs*. Oxfordshire, United Kingdom: CRC Press.

Priddy, K. L., and Keller, P. E. (2005). *Artificial Neural Networks: An Introduction*. Bellingham, Washington, United States: SPIE Press. doi:10.1117/3.633187

Renn, O. (1998). Three Decades of Risk Research: Accomplishments and New Challenges. *J. Risk Res.* 1 (1), 49–71. doi:10.1080/136698798377321

RSMeans (2015). *Building Construction Cost Data* (73rd annual edition ed.), Gordian Group, Rockland, MA, United States: The Gordian Group, Inc..

Shapiro, S. S., and Wilk, M. B. (1965). An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52 (3/4), 591–611. doi:10.1093/biomet/52.3-4.591

Shimizu, C., Karato, K., and Nishimura, K. (2014). Nonlinearity of Housing price Structure. *Int. J. Housing Markets Anal.* 7 (4), 459–488. doi:10.1108/ijhma-10-2013-0055

Skitmore, R. M., and Ng, S. T. (2003). Forecast Models for Actual Construction Time and Cost. *Building Environ.* 38 (8), 1075–1083. doi:10.1016/S0360-1323(03)00067-2

Sonmez, R. (2008). Parametric Range Estimating of Building Costs Using Regression Models and Bootstrap. *J. Constr. Eng. Manage.* 134 (12), 1011–1016. doi:10.1061/(asce)0733-9364(2008)134:12(1011)

Sousa, S., Martins, F., Alvimferraz, M., and Pereira, M. (2007). Multiple Linear Regression and Artificial Neural Networks Based on Principal Components to Predict Ozone Concentrations. *Environ. Model. Softw.* 22 (1), 97–103. doi:10.1016/j.envsoft.2005.12.002

Taghinezhad, A., Friedland, C. J., and Rohli, R. V. (2020a). *Benefit-cost Analysis of Flood-Mitigated Residential Buildings in Louisiana*. Oxfordshire, United Kingdom: Taylor & Francis, 1–18.

Taghinezhad, A., Friedland, C. J., Rohli, R. V., and Marx, B. D. (2020b). An Imputation of First-Floor Elevation Data for the Avoided Loss Analysis of Flood-Mitigated Single-Family Homes in Louisiana, United States. *Front. Built Environ.* 6 (138). doi:10.3389/fbuil.2020.00138

Touran, A., and Lopez, R. (2006). Modeling Cost Escalation in Large Infrastructure Projects. *J. Constr. Eng. Manage.* 132 (8), 853–860. doi:10.1061/(asce)0733-9364(2006)132:8(853)

USACE (1993). *Flood Proofing: How to Evaluate Your Options*. Washington DC, United States: US Army Corps of Engineers.

Waier, P. R., and Balboni, B. (2018). *Building Construction Cost Data*. Rockland, MA, United States: RS Means Company.

Wilmot, C. G., and Mei, B. (2005). Neural Network Modeling of Highway Construction Costs. *J. Constr. Eng. Manage.* 131 (7), 765–771. doi:10.1061/(asce)0733-9364(2005)131:7(765)

Xiang, D. (2001). Fitting Generalized Additive Models with the GAM Procedure. *Proc. SUGI Proceedings, Citeseer*, 256–226.

Zhang, F., Lai, T., Rajaratnam, B., and Zhang, N. R. (2011). *Cross-Validation and Regression Analysis in High-Dimensional Sparse Linear Models*. Stanford, CA, United States: Stanford University. doi:10.1109/icip.2011.6116121

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

APPENDIX A. ELEVATION PROJECT COST ESTIMATES

Table A1 | Elevation project cost estimates by the selected regression model (Model 5 m).

	One-story; slab foundation							
	ΔE (m)							
A (m²)	0.5	1	1.5	2	2.5	3	3.5	4
50	$40,255	$42,234	$44,311	$46,490	$48,776	$51,174	$53,690	$56,330
100	$85,100	$89,285	$93,675	$98,281	$103,113	$108,183	$113,503	$119,084
150	$131,859	$138,342	$145,145	$152,282	$159,769	$167,625	$175,867	$184,515
200	$179,905	$188,751	$198,032	$207,769	$217,985	$228,704	$239,949	$251,748
250	$228,931	$240,188	$251,998	$264,389	$277,389	$291,029	$305,339	$320,353
300	$278,754	$292,461	$306,841	$321,929	$337,758	$354,366	$371,790	$390,071
350	$329,248	$345,438	$362,423	$380,244	$398,941	$418,557	$439,137	$460,730
400	$380,325	$399,026	$418,646	$439,231	$460,829	$483,488	$507,261	$532,204

	Two-story; slab foundation							
	ΔE (m)							
A (m²)	0.5	1	1.5	2	2.5	3	3.5	4
50	$106,084	$111,300	$116,773	$122,515	$128,539	$134,859	$141,490	$148,447
100	*$224,265*	$235,292	$246,862	$259,000	$271,735	$285,097	$299,115	$313,823
150	$347,488	$364,574	$382,501	$401,309	$421,041	$441,744	$463,465	$486,254
200	$474,104	$497,416	$521,875	$547,536	$574,458	$602,705	$632,340	$663,433
250	$603,305	$632,970	$664,093	$696,747	$731,007	$766,951	$804,662	$844,228
300	$734,603	$770,724	$808,621	$848,381	$890,096	$933,863	$979,782	$1,027,958
350	$867,671	$910,335	$955,097	$1,002,059	$1,051,331	$1,103,026	$1,157,262	$1,214,166
400	$1,002,274	$1,051,556	$1,103,262	$1,157,510	$1,214,426	$1,274,140	$1,336,790	$1,402,521

	One-story; other foundations							
	ΔE (m)							
A (m²)	0.5	1	1.5	2	2.5	3	3.5	4
50	$30,791	$32,305	$33,894	$35,560	$37,309	$39,143	$41,068	$43,087
100	$65,094	68$,294	$71,653	$75,176	$78,872	$82,750	$86,819	$91,088
150	$100,860	$105,819	$111,022	$116,481	$122,209	$128,218	$134,522	$141,137
200	$137,611	$144,377	$151,476	$158,924	$166,739	$174,937	$183,539	$192,564
250	$175,111	$183,722	$192,755	$202,233	$212,177	$222,610	$233,556	$245,040
300	$213,221	$223,705	$234,705	$246,246	$258,354	$271,057	$284,385	$298,369
350	$251,845	$264,228	$277,220	$290,851	$305,153	$320,157	$335,900	$352,416
400	$290,914	$305,218	$320,226	$335,972	$352,491	$369,824	$388,008	$407,087

	Two-story; other foundations							
	ΔE (m)							
A (m²)	0.5	1	1.5	2	2.5	3	3.5	4
50	$81,144	$85,134	$89,320	$93,712	$98,320	$103,155	$108,227	$113,548
100	$171,542	$179,977	$188,826	$198,111	$207,852	$218,073	$228,795	$240,046
150	$265,796	$278,866	$292,578	$306,964	$322,058	$337,894	$354,508	$371,939
200	$362,646	$380,478	$399,186	$418,814	$439,408	$461,014	$483,682	$507,465
250	$461,473	$484,163	$507,970	$532,947	$559,153	$586,647	$615,492	$645,757
300	$561,903	$589,532	$618,520	$648,933	$680,842	$714,319	$749,443	$786,293
350	$663,688	$696,322	$730,561	$766,483	$804,171	$843,713	$885,199	$928,725
400	$766,647	$804,344	$843,894	$885,388	$928,924	$974,599	$1,022,521	$1,072,799