International Conference on
Advances in Steel Structures
Egypt Steel 19 (ES19)
December 17-19, 2019
Cairo, Egypt

Housing & Building
National Research Center

المؤتمر الدولى
آفـاق التطوير للمنشآت الصلب
17-19 ديسمبر 2019
القاهرة – جمهورية مصر العربية

MACHINE LEARNING APPLICATIONS FOR BLAST PERFORMANCE ASSESSMENT OF COLD FORMED STEEL GIRT SYSTEMS

Islam A. Torky[(i)], Shady Salem[(ii)]

(i) Student, *Civil Engineering Department, The British University in Egypt, Egypt*
(ii) Assistant Professor, *Civil Engineering Department, The British University in Egypt, Egypt*

ABSTRACT

Within the past few decades the awareness against improvised explosive devices has escalated due to the increasing number and magnitudes of terrorist attacks. Moreover, steel cladding has been extensively used within the applications of accelerated blast mitigation construction. In this context, numerous experimental investigations have been conducted to evaluate the blast performance of cold-formed steel girts, yet no design tools are available. This shortage is due to the complexity of proposing a simplified tool that accounts for the multiple input parameters as the load and resistance and their accompanied uncertainties. The aim of this paper is to derive reliable design tools for steel cold-formed girts subjected to blast loading using different machine learning approaches. These tools were developed through machine learning approaches that devise predictive models. This paper lays the foundation for a novel preliminary blast design approach that would be applicable for the assessment of different blast mitigation systems.

KEYWORDS: Cold-formed, blast, machine learning, regression, steel-girts

INTRODUCTION

Blast loads have been considered in different applications for hundreds of years (i.e. peace and war applications). However, the blast effects and their mitigation techniques/measures have been extensively investigated/developed within the past few decades (Krauthammer, 2008). These developments are mainly attributed to the increase of terrorism risk (number and magnitude) and the enhanced awareness of their negative impacts on the different community aspects (economic, social, environmental,…etc.)(Salem, Campidelli, El-Dakhakhni, & Tait, 2018). Moreover, blast wave parameters are highly dependent on the explosive device which constitutes large source of uncertainty, including chemical attacks, vapors, high explosives (Krauthammer, 2008). As such, most of the available literature usually consider the blast wave parameters rather than their sources (Krauthammer, 2008).

On the other hand, cold-formed steel (CFS) structures have been in service for many years and used for domestic and industrial purposes due to their accelerated construction nature (Darcy, 2005). Moreover, Blast–resistant design using CFS has also been the focus of several researches due to their direct exposure to free field blast events (i.e. typically used as cladding). (Lane, 2003; Salim & Townsend, 2004; Woodson & DiPaolo). Steel girt systems, one form of CFS, are considered an effective blast-resistant cladding system compared to high-performance systems such as precast/prestressed concrete facades (Aviram et al., 2012; Godinho et al., 2013). Steel girt systems are secondary framing members that are placed horizontally to provide lateral support for the corrugate sheets to resist surface loads as depicted in Fig.(1).

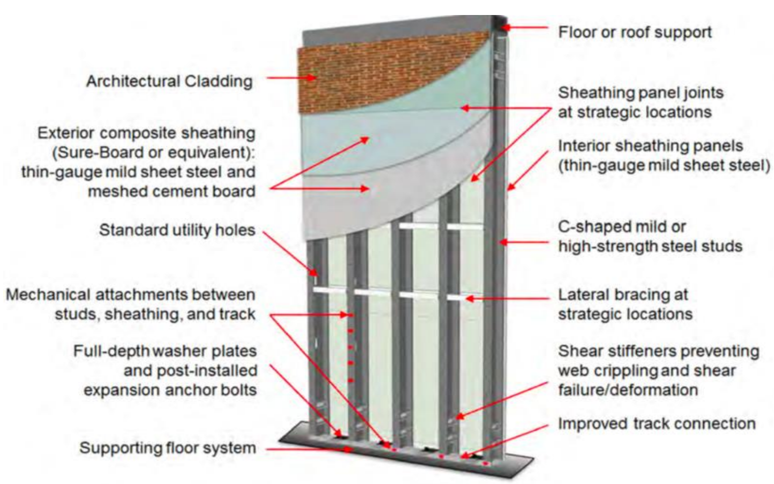

Figure 1. Section view of Steel Girts (adapted from Aviram (2014)).

Most of the available literature addressing the blast behavior of steel girt systems were focused on either experimental (Moen, 2008; Muller, 2002), or numerical (Bondok, 2012) investigations with economical and technical limitations. The experimental investigations are usually characterized with high cost that burdens the repeatability of testing (El-Dakhakhni, Mekky, & Rezaei, 2010). On the other hand, numerical simulations may yield into large computational efforts with sensible accuracy compared to the experimental investigations. For example, Bondok (2014) numerically assessed the blast performance of conventional steel stud wall, however, the developed model had large computational efforts with deviations of 4.1-34.6% from the experimental testing. Moreover, blast loading is characterized with multiple sources of uncertainty, namely, epistemic and aleatory uncertainty (Stewart, Netherton, & Rosowsky, 2006). As such, a new simple reliable model is required to predict the blast performance of steel girt systems to enable probabilistic blast assessment of such elements.

In this context, this paper introduces different statistical models to assess the blast performance of steel girt systems. The statistical models are derived using different machine learning techniques through fitting the experimental results of far field testing for different steel girt systems. These models assess the performance of the steel girts through assessing the deformation (i.e. current North American response limit ASCE, 2011; CSA, 2012). The developed models account for different independent variables; for example, these independent variables are blast intensity, material, and geometrical characteristics. The proposed models (i.e. regression) are mainly selected due to its validity while using limited number of data sets. This paper introduces different multiple regression analysis techniques for far field blast testing of steel girt systems. The presented regressions are using multivariate linear, polynomial, and random forest tree regressions. Finally, the reliability of the presented regression models is assessed using the R score, and root mean square error (RMSE).

METHODOLOGY

MACHINE LEARNING

The aim of using machine learning in this investigation is to formulate an equation or to develop a tool to predict the numerical outcome using multiple inputs (Mohamed, Khan, & Bashier, 2016). Machine learning techniques are usually categorized into supervised and unsupervised techniques.

The difference between supervised and unsupervised learning is that supervised learning utilizes both the input and the output data to formulate a predictive equation while the latter uses the input data only to create a prediction tool (i.e. black box). In the case of this paper supervised learning due to having the output part of the dataset (i.e. steel girt blast performance). These techniques include multivariate linear, polynomial, and random forest regressions. Regression analysis depends on the statistical relationship between the different independent variables and how they will contribute to the dependent variable. Each of these regressors have their own unique way of computing the outcome. These regressors use the input data as training points for the models to predict their output (Witten & Frank, 2005). The following sections elaborate the theory of the used regressors, and their applications.

MULTIVARIATE LINEAR & POLYNOMIAL REGRESSION

Multivariate Linear Regression (MVLR) is a statistical technique uses multiple independent variables to predict a dependent variable. The difference between normal linear regression and MVLR is that MVLR is not limited to a 2D plane. The MVLR is typically applied in n-dimensional space depending on the number of variables (n). The relationship between independent variables, and dependent variables can be developed using linear regression following the general equation shown in Eq. 1 (Neter, Kutner, Nachtsheim, & Wasserman, 1996).

$$Y = \beta_0 + \sum_{j=1}^{n} \beta_j X_j + \varepsilon \qquad \textit{Equation (1)}$$

Y is the dependent variable, β_0 is a constant, β_j is a regression coefficient where j represents the number of variables (j = 1, 2, 3..., n), while ε is the error term which represents the deviation from the true value. In normal linear regression, only a single independent variable X_1 can be considered.

According to Neter et al. (1996) there are five assumptions that should be made for a MVLR. These assumptions include existence, independence, linearity of relationships, homogeneity, and normality. Existence is defined as a specific combination of independent variables X_1, X_2, X_3,.., X_j and the output variable Y is a random variable with a defined average and variance in a specific probability distribution. While independence refers to the Y values which are independent and do not have any relation with each other. Linearity of Relationships means that the average value of Y is a function of the linear combination of X_1, X_2, X_3,.., X_j. Moreover, homogeneity of Variance is having a constant variances in the linear combination of the independent variables (X_1, X_2, X_3,.., X_j). And finally, normality is obtained through distributing the dependent variable by the linear combination of X_1, X_2, X_3,.., X_j.

Polynomial Regression is similar to MVLR with a small difference. This difference is the degree of the equation used in regression. Polynomial Regression uses the same assumptions as the MVLR and shown in Eq. (2) (G. Carmines, A. Stimson, & A. Zeller, 1978). Eq. 2 demonstrates the degrees that can be reached for a polynomial equation (*N*).

$$Y = \beta_0 + \beta_1 X_1 + \ldots \ldots . . \beta_N {X^N}_N + \varepsilon \qquad \textit{Equation (2)}$$

RANDOM FOREST REGRESSION

Random Forest Regression (RFR) is based on the method of decision trees (DT). DT utilizes a decision-making framework based on the information theory, a mathematical model used to store information in data (Wang & Zang, 2011). DT has the capability to predict both continuous (regression) and categorical (classification) data. RFR is considered as a piecewise regression, where the exact regression equation depends on the data point features and how the trees are structured. DT is structured similarly to a real tree in which it has a base called the “root node” and “leaf node” which can be thought of as the actual prediction. DT progresses from one node to another node through “Arcs”. The arcs yield all the information of the previous nodes including the one it originated from as shown in Fig. (2). DT works by starting from the root node and traversing to reach the leaf node. This traversal is done through inquiring about a series of questions regarding the characteristics of the data set. The DT asks questions regarding the inequalities of the nodes value and the actual value and then proceeds to the following node until it reaches the leaf node (i.e. the prediction).

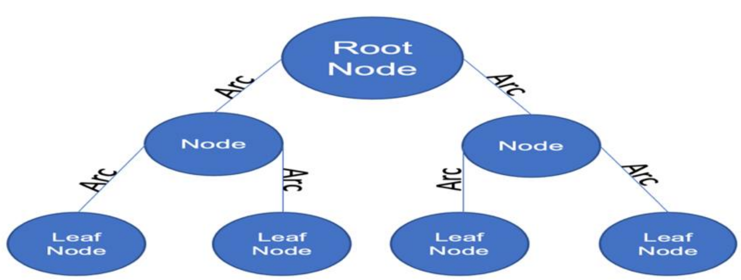

Figure 2. Illustration of a Decision Tree.

To avoid any overfitting of a DT, several trees can be used simultaneously together to form several predictions, which is known as the RFR technique. After creating a prediction from each single tree, the values are averaged to mitigate the performance of any faulty single tree.

MODEL RELIABILITY

To determine the accuracy of the outcome provided by the regressors, a scale should be used to compare the accuracy of the different models. In the presented paper, the coefficient of determination (i.e. R^2) score along with the Root Mean Square Error (RMSE) were used. R^2 is the proportion of variance of the dependent variable by the independent variable and can be calculated using Eq. (3) (Cameron & Windmeijer, 1995). While, the RMSE is the standard deviation of the residuals (prediction errors) as illustrated in Eq. (4) (Chai & Draxler, 2014). Residuals of the RMSE indicates for how far the predictions from the regression line data points (HAYES, 2019).

$$R^2 = 1 - \frac{Explained\,Variation}{Total\,Variation}$$ *Equation (3)*

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(Predicted_i - Actual_i)^2}{N}}$$ *Equation (4)*

DATABASE

The blast response of steel girt systems is influenced by multiple parameters such as the material properties, cladding thickness, girt spacing, boundary conditions, and the applied blast wave. Typically, the positive peak pressure (P_{max}) and positive specific impulse (I) are the most influential factors on the response of structural members (Salem, Campidelli, El-Dakhakhni, & Tait, 2019; Shin, Whittaker, & Cormie, 2015). The positive peak pressure is the sudden rise of the ambient pressure (P_o) due to detonation, while the specific impulse is the integration of the pressures over the loading duration (t_d) as shown in Fig. (3). As such, including all the aforementioned parameters in training the regression models may yield in a complicated model influenced by each input variable.

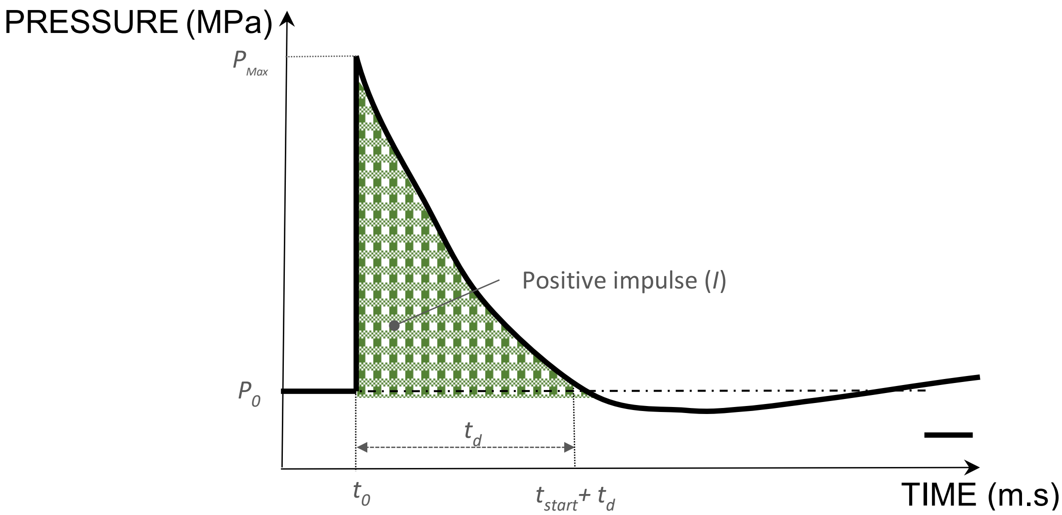

Figure 3. Idealized blast wave

Yet, all the input parameters are still influencing the expected outcomes. Consequently, the influencing parameters are lumped into two groups, namely, resistance and loading groups. The resistance group is an indication for the resistance function of the steel girt system derived using different input parameters as the material properties, cladding thicknesses, girt spacing, and boundary conditions. The resistance group is presented in the form of a resistance function developed by the Unified Facilities Criteria ((UFC), 2008). The UFC proposed an idealized bilinear function for steel systems presented by the initial stiffness "K" and ultimate capacity "R_u" as shown in Fig. (4).

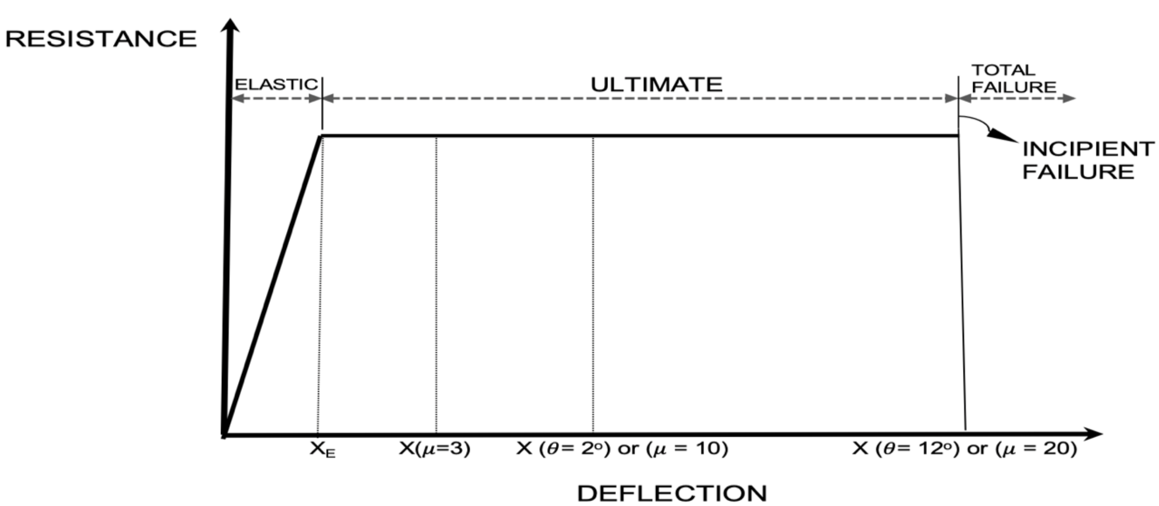

Figure 4. Idealized steel girt systems resistance function (adapted from USDOD2014)

On the other side, the loading is presented in the form of P_{max} and I. In the meanwhile, the output parameter of the steel girt systems is considered as the chord rotation (θ) as per the current North American Standards (ASCE, 2011; CSA, 2012). The developed models are based on the experimental results of 21 steel girt system exposed to far-field blast event (Oswald, 2005) while their input/output variables are summarized in table 1.

Table 1. Used dataset for training

P **kPa (psi)**	***I*** **kPa - ms (psi - ms)**	***Ru*** **kPa (psi)**	***K*** **kPa/m (psi/in)**	θ
41(5.9)	1241(180)	4(0.62)	57(0.21)	13.5
17(2.5)	517(75)	4(0.62)	57(0.21)	12.7
10(1.5)	179(26)	4(0.62)	57(0.21)	6.0
38(5.5)	655(95)	7(1.04)	244(0.9)	14.0
17(2.5)	290(42)	7(1.04)	244(0.9)	9.5
32(4.7)	476(69)	7(1.04)	244(0.9)	16.7
41(5.9)	552(80)	7(1.04)	244(0.9)	12.9
54(7.8)	579(84)	7(1.04)	244(0.9)	18.4
29(4.2)	379(55)	7(1.04)	244(0.9)	15.2
21(3.1)	296(43)	7(1.04)	244(0.9)	8.3
29(4.2)	496(72)	7(1.04)	244(0.9)	12.9
57(8.2)	945(137)	27(3.87)	1482(5.46)	8.9
34(5)	689(100)	17(2.53)	980(3.61)	6.7
10(1.4)	172(25)	10(1.38)	138(0.51)	1.3
14(2.1)	221(32)	10(1.38)	138(0.51)	2.1
17(2.5)	303(44)	10(1.38)	138(0.51)	2.3
17(2.4)	290(42)	10(1.38)	138(0.51)	2.3
241(35)	1462(212)	15(2.18)	632(2.33)	6.7
234(34)	1510(219)	21(3)	1194(4.4)	8.8
234(34)	1510(219)	10(1.5)	600(2.21)	7.5
228(33)	1379(200)	6(0.87)	204(0.75)	5.4

To ensure the efficiency of the used data; mutual variable correlation diagrams are plotted to visualize the distribution of the parameters (input and output) as shown in Fig. (5). Fig.

(5) presents the relationship between the variables and their frequencies. For example, the bar chart in the top left corner represent the frequency of the peak pressure within the used database, while the rest of the row represents the variability of the considered I, R_u, K, and θ.

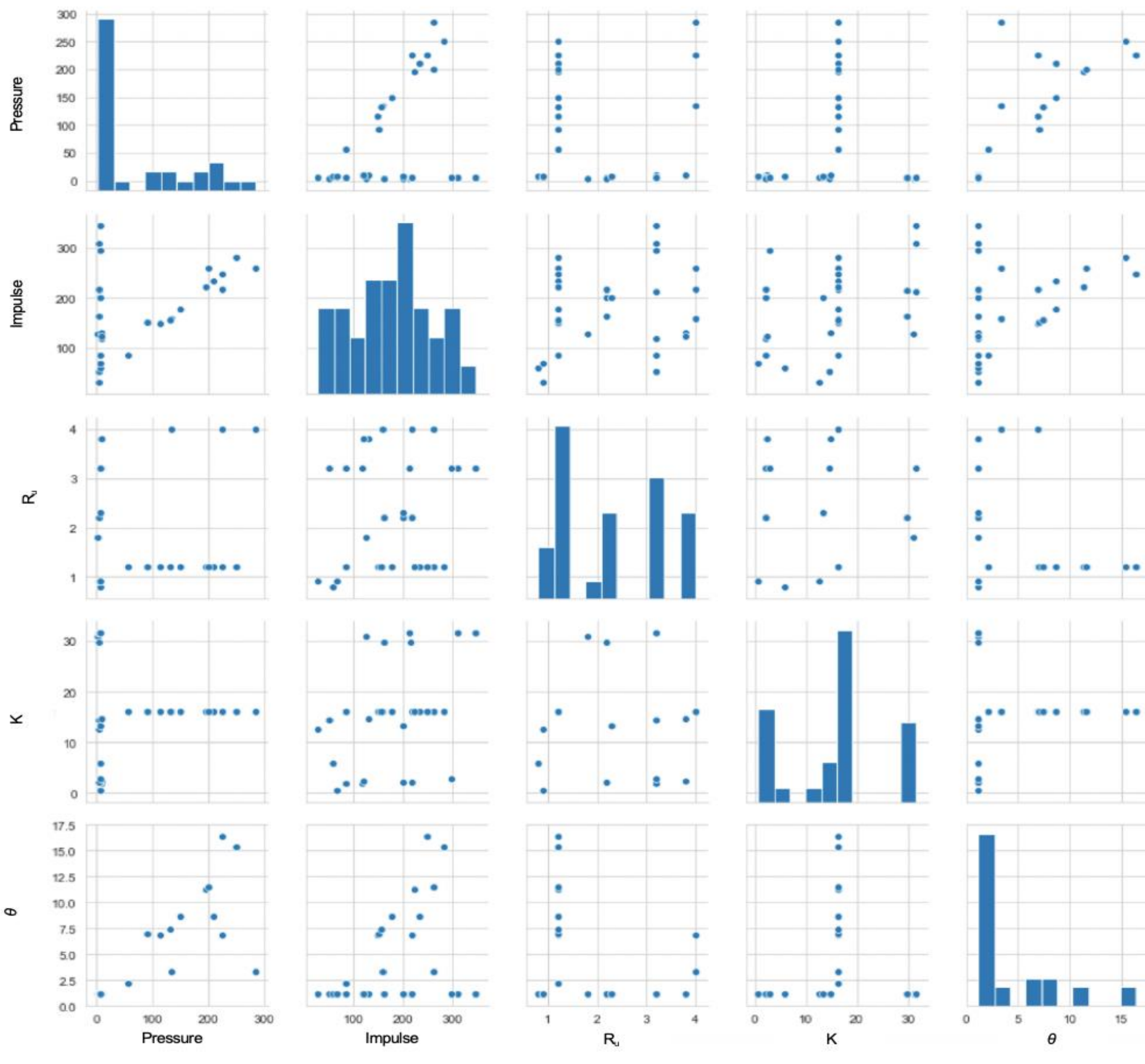

Figure 5. Relationship between variables in the used dataset.

RESULTS & DISCUSSION

SENSITIVITY

After the data was extracted, regression models using RFR, MVLR, and Polynomial regression were created through the scikit-learn library (Pedregosa, et al., 2011). The developed models, specially RFR and polynomial models, had several parameters controlling the accuracy of the model. For example, the accuracy of the RFR is related to the number of trees used to predict the outcome. It is worth nothing that there is no limit to the number of trees that can be used to create a RFR model. Subsequently, a sensitivity analysis is performed to optimize the number of trees. To assess the sensitivity, the R^2 values were used as an indication for the model stability. Fig. (6) shows the model sensitivity to the number of trees used. Moreover, Fig. (6) demonstrates that the optimum number of trees would be around 20 to 25 trees. This optimum decision is based on the fact that the R^2 did not significantly enhance after using more than 25 DT, however, increasing the DT than this limit would increase the processing time. On the contrary, less than 20 may yield into unreliable model. As such, the RFR model used in this study is based on with 21 trees as the sample size.

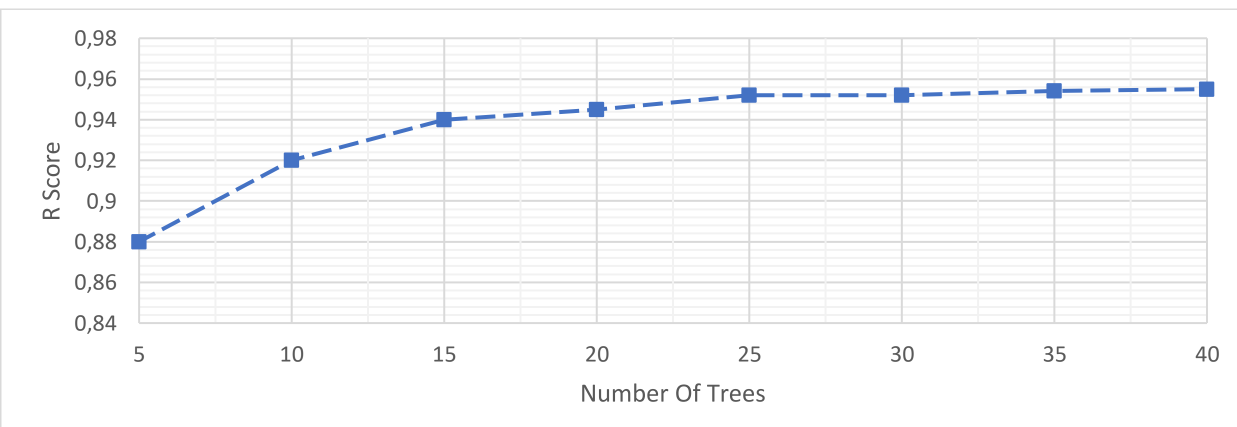

Figure 6. Sensitivity of the RFR against number of trees.

Similarly, Polynomial Regression provided good results when increasing the degree, however, it is prone to overfitting. The model was tested at different degrees against the R^2 as shown in Fig. (7). In Fig. (7), it is evident that the model started to overfit once it reached to the 3rd-degree (R^2=1.0). It is also evident that the 1st-degree equation offers the same value as the MVLR which is a verification to the model. Therefore a 2nd-degree equation is used to avoid any overfitting.

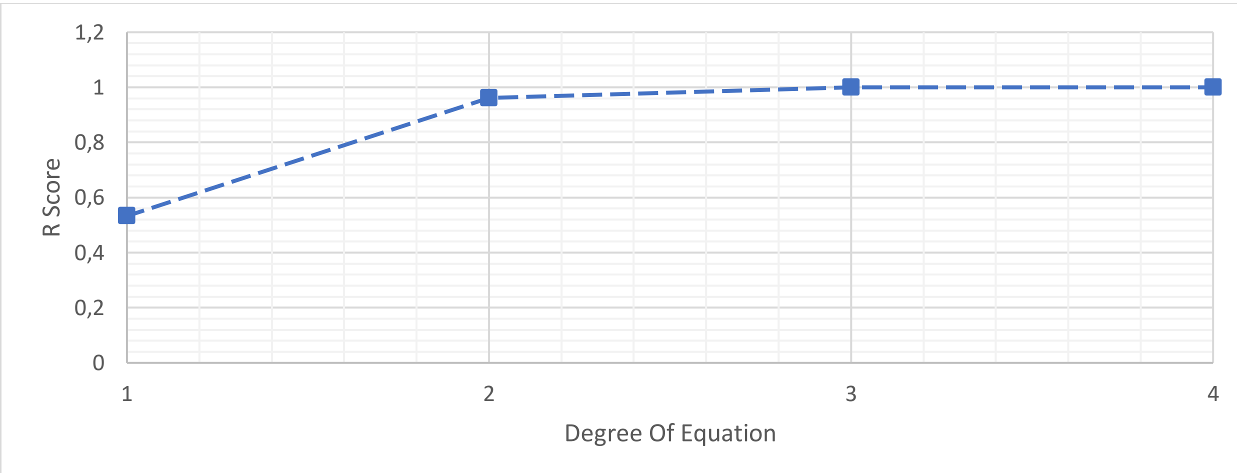

Figure 7. Sensitivity of the polynomial degree.

RESULTS:

For each regression model the R score, and RMSE was calculated to assess their reliability. The results of the R score, and RMSE are summarized in Table 2. From Table 2, it is apparent that MVLR has the worst values. MVLR yielding the lowest value R score ($R^2 = 0.53$), which indicates that the data is not fitting in linear manner(Ang & Tang, 2007). This low R value is due to the fact that MVLR is derived using the first-degree equation, however, the data fit differently. Both RFR and the 2nd Degree Polynomial Regression provided satisfactory R score, and RMSE scores. However, it is more reliable to use the RFR model due to not being affected by overfitting compared to the 2nd degree polynomial regression. Moreover, the reliability of the models was tested against the experimental for the 21 data set as shown in Fig. (8).

Table 2. R score (R^2), and RMSE values for the proposed models.

Model	R Score	RMSE
MVLR	0.53	1.87
RFR (21 Trees)	0.95	1.01
Polynomial Regression (2nd Degree)	0.94	1.07

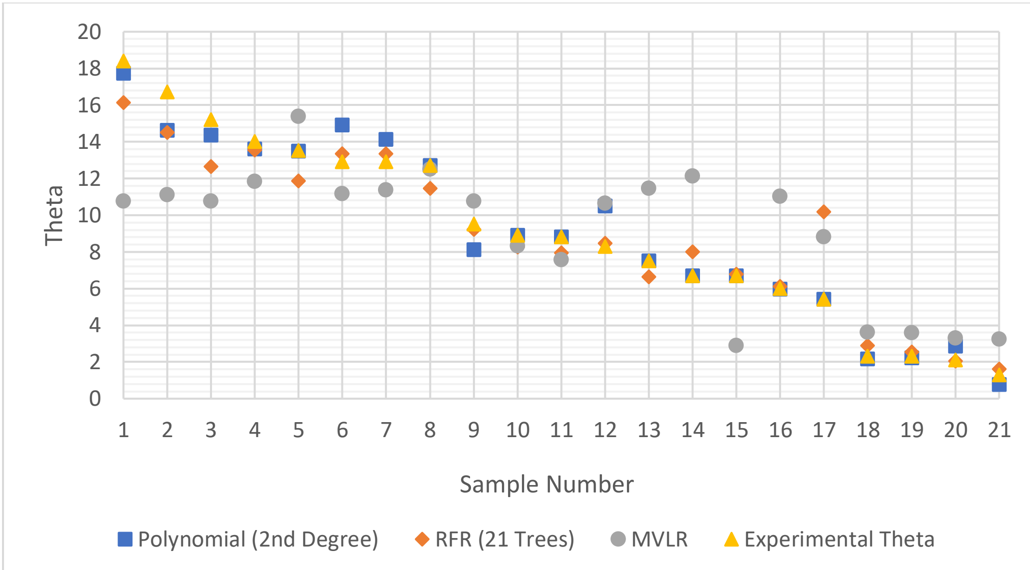

Figure 8. Reliability of the regression methods.

CONCLUSIONS:

This paper developed regression models for blast assessment of steel girt systems using machine learning techniques, namely, MVLR, polynomial, and RFR regressions. The developed models were based on the results of real explosive testing database. The reliability of the models was assessed using R^2 and RMSE. The results of the reliability assessment showed that the RFR was the most reliable model with highest accuracy and no overfitting. In the meanwhile, the polynomial regression offered a higher level of accuracy, however, it starts to overfit at the 3rd degree. The 2nd degree polynomial regression may be accurate in terms of R^2, however, there is no evidence that it did not already start overfitting.

The presented models constitute cheap computational models that can be used in further probabilistic blast performance investigations for steel girt systems. On the other side, the used regression models had the capability of using limited data, as main advantage of using machine learning, expanding the test database may yield into more reliable models. Additionally, different predictive models and reliability assessment indices may yield into reliable models as well.

REFERENCES:

ASCE. (2011). Blast Protection of Buildings. In American Society of Civil Engineers (Ed.), *ASCE 59-11*. https://doi.org/10.1061/9780784411889

Aviram , A., Mayes, R.L, and Hamburger, R.O. (2012). “enhanced Blast – Resistance of an Innovative High-Strength Steel Stud Wall System”. *In Proceedings of the 2012 Structural Engineers Association of California (SEAOC) Convention*, Santa Fe, New Mexico.

Bondok, J. (2012). *Numerical modeling of static nonlinear behavior of cold-formed steel stud walls.* Columbia, MO.: Dept. of Civil Engineering, Univ. of Missouri,.

Cameron, A., & Windmeijer, F. (1995). *An R-squared measure of goodness of fit for some common nonlinear regression models.*

Chai, T., & Draxler, R. (2014). *Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature.* Geoscientific Model Development.

CSA. (2012). *CSA S850-12 Design and assessment of buildings subjected to blast loads*. Mississauga, ON, Canada: Canadian Standards Association.

Darcy, G. (2005). *Structural Behaviour of an Innovative Cold-Formed Steel Building System.*

El-Dakhakhni, W. W., Mekky, W. F., & Rezaei, S. H. C. (2010). Validity of SDOF Models for Analyzing Two-Way Reinforced Concrete Panels under Blast Loading. *Journal of Performance of Constructed Facilities*, *24*(4), 311–325. https://doi.org/10.1061/(ASCE)CF.1943-5509.0000090

G. Carmines, E., A. Stimson, J., & A. Zeller, J. (1978). *Sociological Methods & Research.*

Godinho, J.m Gallant, S., Quiter, L., Aviram, A., Mayes, R.L., and Hamburger, R. O. (2013). “ Design and Detailing of Metal Stud Wall Systems in Response to Air-Blast Loading Effects”. In Porceedings of the 2013 Structures Congress, Pittsburgh, Pennsylvania.

HAYES, A. (2019, May 8). *Investopedia*. Retrieved from https://www.investopedia.com/terms/r/r-squared.asp

Krauthammer, T. (2008). *Modern Protective Structures.*

Lane, J. W. (2003). *Modeling and design of explosion-resistant steel stud wall systems.* Columbia, Columbia, MO.: Dept. Of Civil Engineering, Univ. Of Missouri.

Moen, C. (2008). *Evaluation of conventional steel stud walls for blast design.* Baltimore: Johns Hopkins Univ.

Mohamed, M., Khan, M., & Bashier, E. (2016). *Machine Learning: Algorithms and Applications .* CRC Press.

Muller, P. (2002). *Static response evaluation of cold-formed steel stud walls.* University of Missouri: Static response evaluation of cold-formed steel stud walls.

Neter, J., Kutner, M., Nachtsheim, C., & Wasserman, W. (1996). *Applied Linear Statistical Models, Fourth ed. IRWIN.*

Oswald, C. J. (2005). *Component explosive damage assessment workbook (CEDAW). Methodology Manual V1. 0.* (Baker Engineering and Risk Consultants, ed.). Washington, DC.

Pedregosa, F., Varouquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., . . . Cournapeau, D. (2011). *Scikit-learn: Machine Learning in Python.* Journal of Machine Learning Research 12.

Salem, S., Campidelli, M., El-Dakhakhni, W. W., & Tait, M. J. (2019). Resilience-Based Design for Blast Risk Mitigation: Learning from Natural Disasters. In *Resilient Structures and Infrastructure* (pp. 177–194). https://doi.org/10.1007/978-981-13-7446-3_7

Salim, H., & Townsend, P. (2004). *Explosion-resistant steel stud wall system.* Explosion-resistant steel stud wall system.

Shin, J., Whittaker, A. S., & Cormie, D. (2015). Incident and Normally Reflected Overpressure and Impulse for Detonations of Spherical High Explosives in Free Air. *Journal of Structural Engineering*, *141*(12), 04015057. https://doi.org/10.1061/(ASCE)ST.1943-541X.0001305

Stewart, M. G., Netherton, M. D., & Rosowsky, D. V. (2006). Terrorism Risks and Blast Damage to Built Infrastructure. *Natural Hazards Review*, *7*(3), 114–122. https://doi.org/10.1061/(ASCE)1527–6988(2006)7:3(114)

Unified Facilities Criteria (UFC). (2008). *Structures to resist the effects of accidental explosions*.

Wang, L., & Zang, X. (2011). *Implementation of a scalable decision forest model based on information theory.*

Witten, I., & Frank, E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques, Second Edition Morgan Kaufman.*

Woodson, S., & DiPaolo, B. (n.d.). *An overview of research at ERDC on steel stud exterior wall systems subjected to severe blast loading.* Proc., ASCE Struct. Congr. and Exposition: Structural Engineering and Public Safety.