12NCEE
Salt Lake City
2022

Machine Learning and Statistics-Driven Code-Oriented Shear Strength Equation for RC Structural Walls

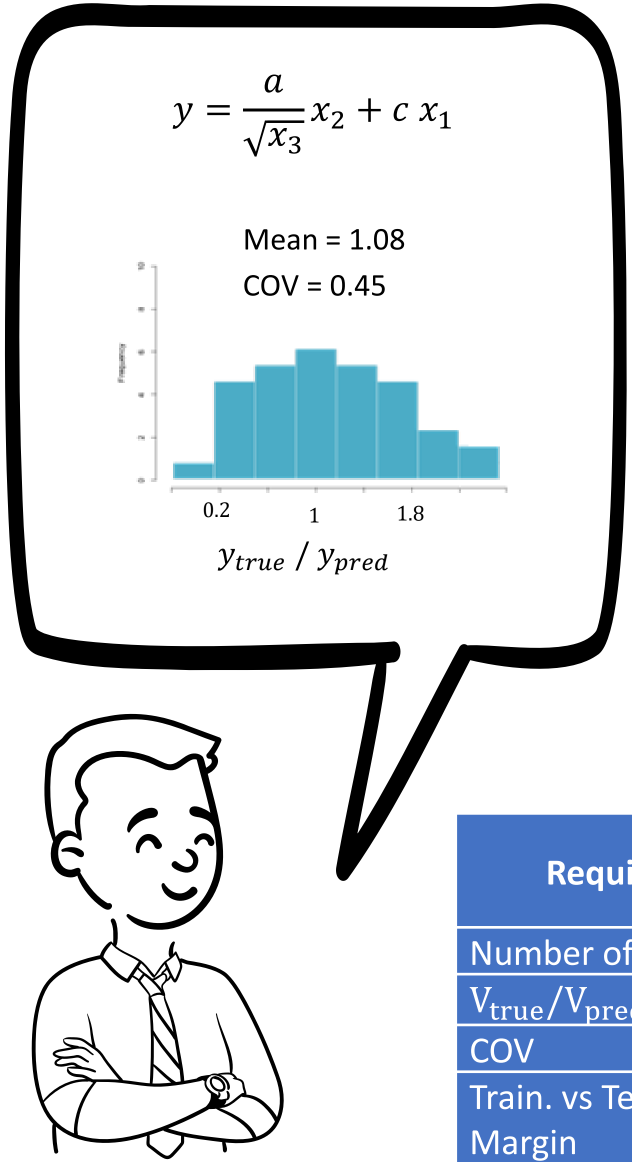
M. Rojas-León[(1)], S. Abdullah[(2)], K. Kolozvari[(3)], J. Wallace[(4)]

[(1)] Ph.D. Candidate, Department of Civil Engineering, University of California Los Angeles, mrojasleon@ucla.edu
[(2)] Postdoctoral Researcher, Department of Civil Engineering, University of California Los Angeles, eng.saman86@gmail.com
[(3)] Associate Professor, Department of Civil Engineering, California State University Fullerton, kkolozvari@fullerton.edu
[(4)] Professor, Department of Civil Engineering, University of California Los Angeles, wallacej@ucla.edu

Motivation & Challenges

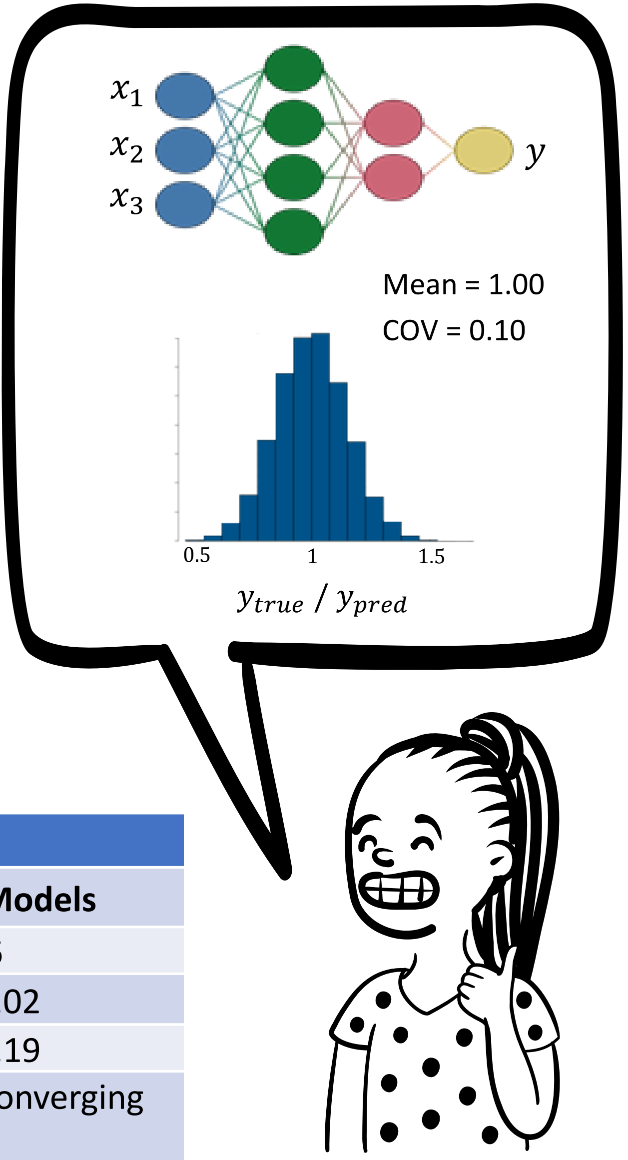
- Equation in ACI is based on limited data and have a large error
- Availability of the UCLA-RC Walls Database
- Significant variance when using the same equation/model to predict values of different databases
- Complexity of ML models and its interpretation

Identified Main Problems - Example



Model Complexity – Acceptable Error

Trade Off



Requirements	Model Complexity Level	
	Complex ML Models	**Simplified Models**
Number of parameters	-	~ 3 - 6
V_{true}/V_{pred} mean ratio	0.99 – 1.01	0.98 – 1.02
COV	$\leq$ 0.12	0.16 – 0.19
Train. vs Test. Error Margin	$\pm$20% of the converging error	$\pm$10% of the converging error

Framework to Set Target Errors for Different Model Complexity Levels

Framework to Set Target Errors for Different Model Complexity Levels

6 generic steps of ML + Sensitivity Analysis with Iterative k-Fold CV
+ Elastic Net Models (ENMs) with engineered features
+ Establish desired underfitting levels

Framework to Set Target Errors for Different Model Complexity Levels

(1) Collection and preparation of data

(2) Feature selection

- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm

- Complex ML model (e.g., ANN, RF regression)
- ENMs

(4) Selection hyper-parameters

- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

(5) Model training

(6) Model performance evaluation

(1) Collection and preparation of data

(2) Feature selection

- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm

- Complex ML model (e.g., ANN, RF regression)
- ENMs

(4) Selection hyper-parameters

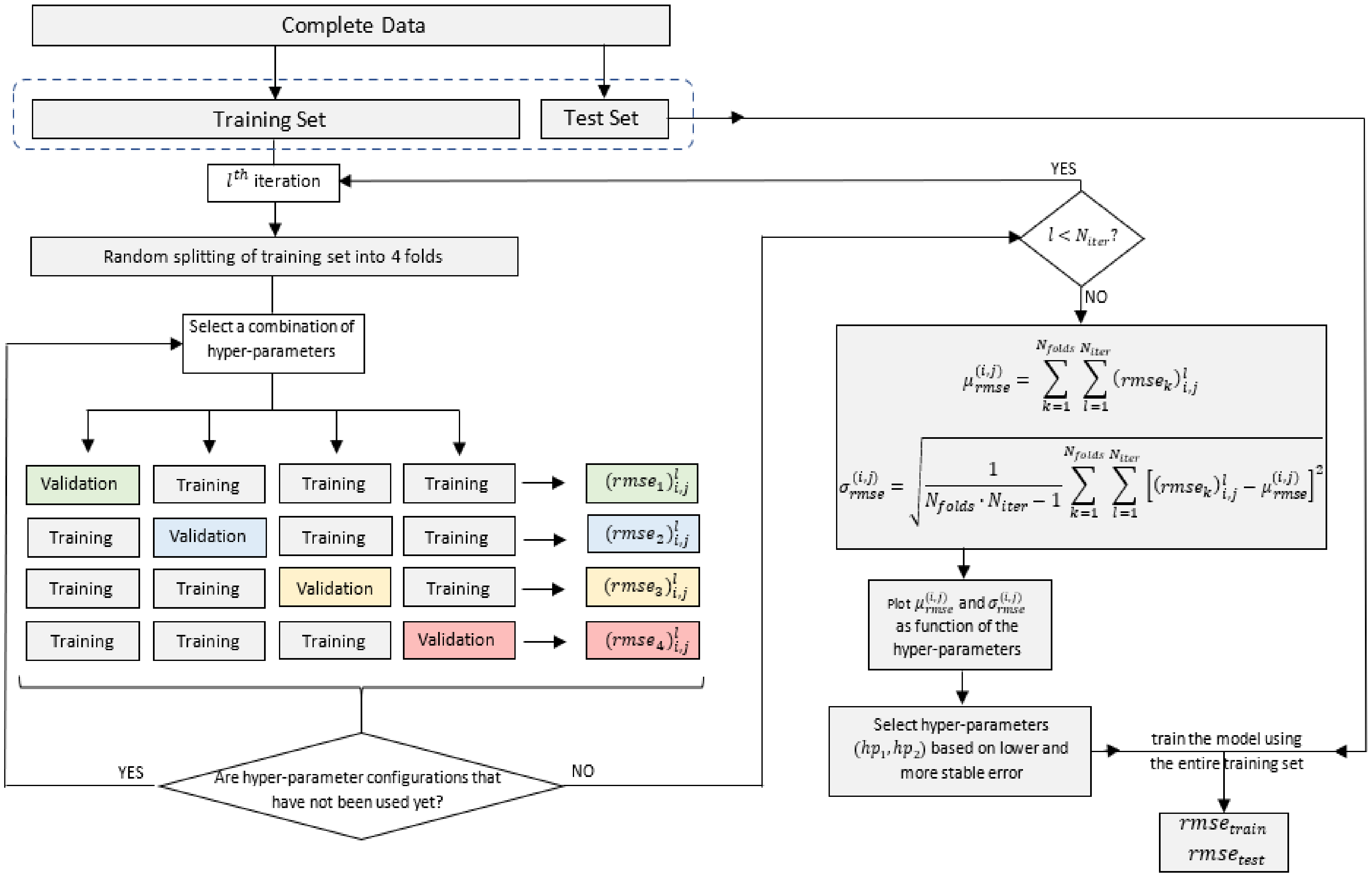
- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

(5) Model training

(6) Model performance evaluation

(7) Set target errors for different model complexity levels

- Algorithm to assess the sensitivity analysis on the hyper parameters



(1) Collection and preparation of data

(2) Feature selection
- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm
- Complex ML model (e.g., ANN, RF regression)
- ENMs

(4) Selection hyper-parameters
- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

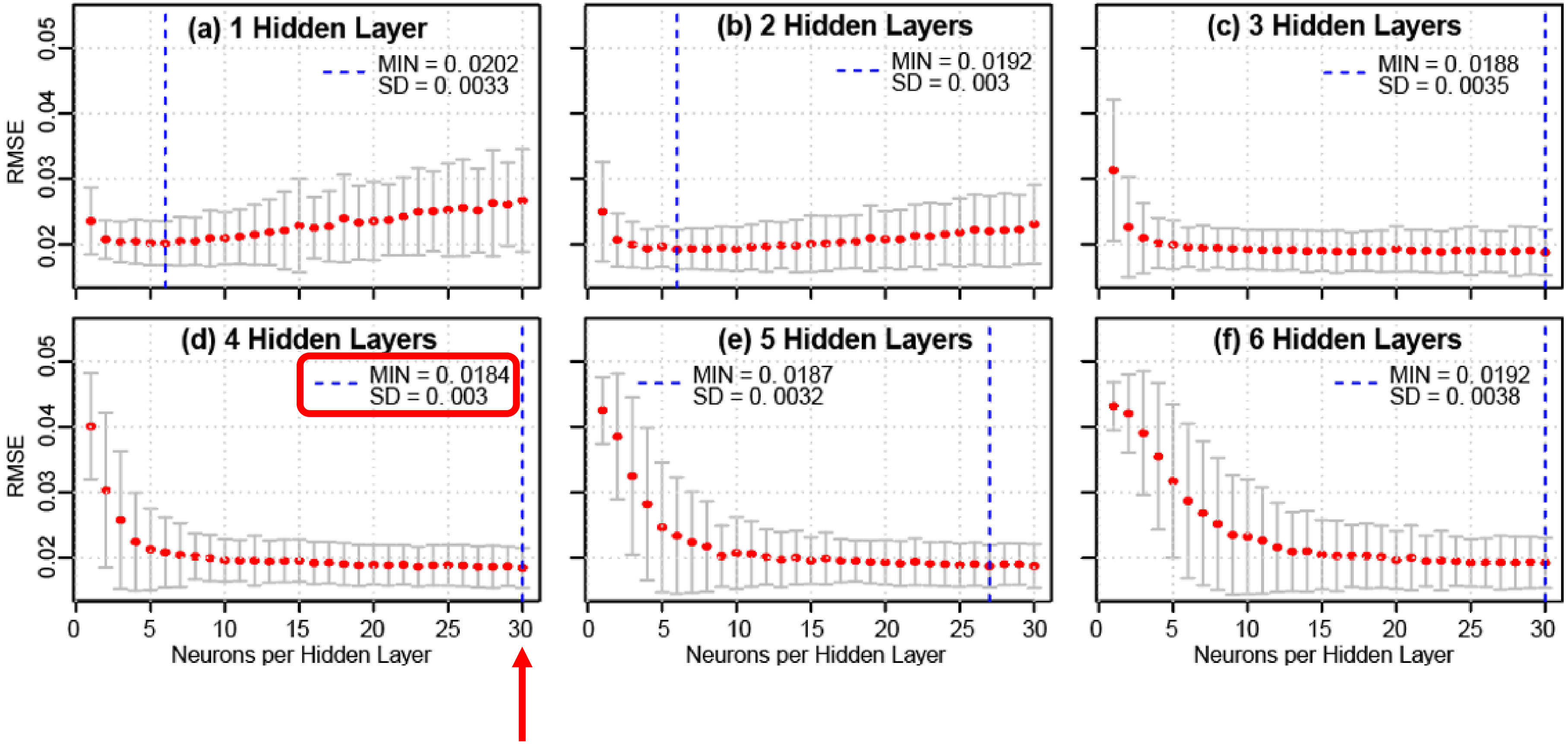
(5) Model training

(6) Model performance evaluation

(7) Set target errors for different model complexity levels

- Artificial Neural Network (ANN)

(Sensitivity Analysis → 180 ANN were trained)



- RF Regression (same)

(Sensitivity Analysis → 120 RF Regressions were trained)

- ENMs (same)

(Sensitivity Analysis → (10 models) x (4 complexity levels) = 40 ENMs were trained)

(1) Collection and preparation of data

(2) Feature selection
- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm
- Complex ML model (e.g., ANN, RF regression)
- ENMs

(4) Selection hyper-parameters
- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

(5) Model training

(6) Model performance evaluation

(7) Set target errors for different model complexity levels

- Selected ENMs

Optimum LASSO models

Model	$RMSE_{train}$	$RMSE_{test}$	$RMSE_{conv}$	Acceptable?
6	0.0117	0.0137	0.0127	YES
5	0.0114	0.0167	0.0141	YES
7	0.0112	0.0177	0.0145	NO
3	0.0131	0.0181	0.0156	YES
8	0.0120	0.0206	0.0163	NO
4	0.0148	0.0186	0.0167	YES
2	0.0134	0.0202	0.0168	YES
1	0.0197	0.0216	0.0207	YES
9	0.0122	0.0631	0.0377	NO
10	0.0122	0.3469	0.1796	NO

1-SD away LASSO models

Model	$RMSE_{train}$	$RMSE_{test}$	$RMSE_{conv}$	Acceptable?
9	0.0137	0.0147	0.0142	YES
6	0.0140	0.0175	0.0157	YES
10	0.0139	0.0185	0.0162	NO
3	0.0137	0.0189	0.0163	NO
5	0.0146	0.0181	0.0163	YES
4	0.0144	0.0183	0.0164	NO
2	0.0138	0.0199	0.0168	NO
7	0.0140	0.0213	0.0176	NO
8	0.0145	0.0244	0.0194	NO
1	0.0203	0.0219	0.0211	YES

6-feature LASSO models

Model	$RMSE_{train}$	$RMSE_{test}$	$RMSE_{conv}$	Acceptable?
3	0.0153	0.0194	0.0173	NO
9	0.0152	0.0201	0.0177	NO
4	0.0164	0.0190	0.0177	YES
8	0.0147	0.0209	0.0178	NO
10	0.0164	0.0198	0.0181	YES
6	0.0203	0.0212	0.0207	YES
1	0.0203	0.0219	0.0211	YES
5	0.0200	0.0222	0.0211	YES
7	0.0220	0.0233	0.0226	YES
2	0.0213	0.0240	0.0227	YES

3-feature LASSO models

Model	$RMSE_{train}$	$RMSE_{test}$	$RMSE_{conv}$	Acceptable?
10	0.0192	0.0218	0.0205	YES
8	0.0184	0.0256	0.0220	NO
4	0.0218	0.0237	0.0227	YES
9	0.0208	0.0291	0.0250	NO
2	0.0245	0.0292	0.0268	YES
3	0.0249	0.0288	0.0269	YES
1	0.0273	0.0289	0.0281	YES
5	0.0264	0.0310	0.0287	YES
6	0.0279	0.0320	0.0299	YES
7	0.0276	0.0360	0.0318	NO

(1) Collection and preparation of data

(2) Feature selection
- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm
- Complex ML model (e.g., ANN, RF regression)
- ENMs

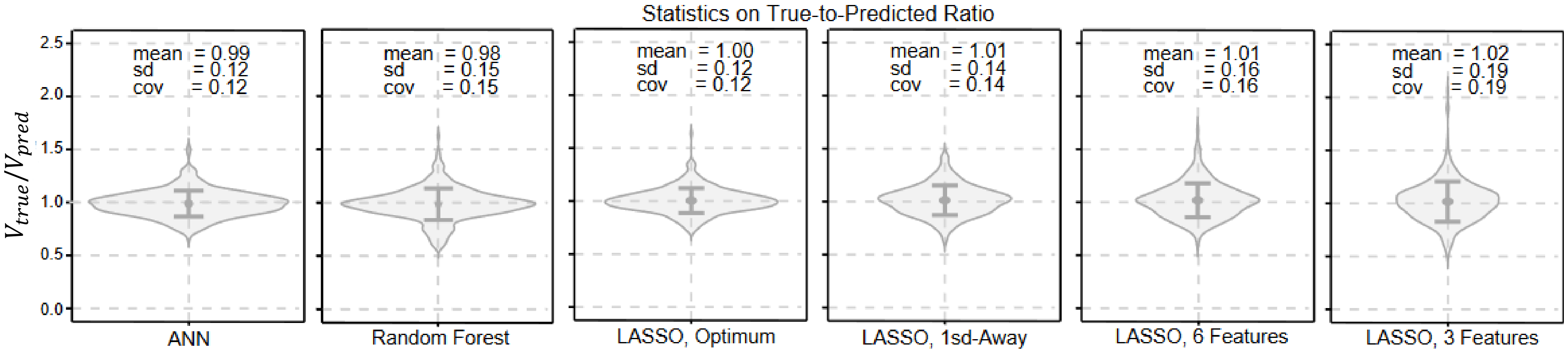
(4) Selection hyper-parameters
- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

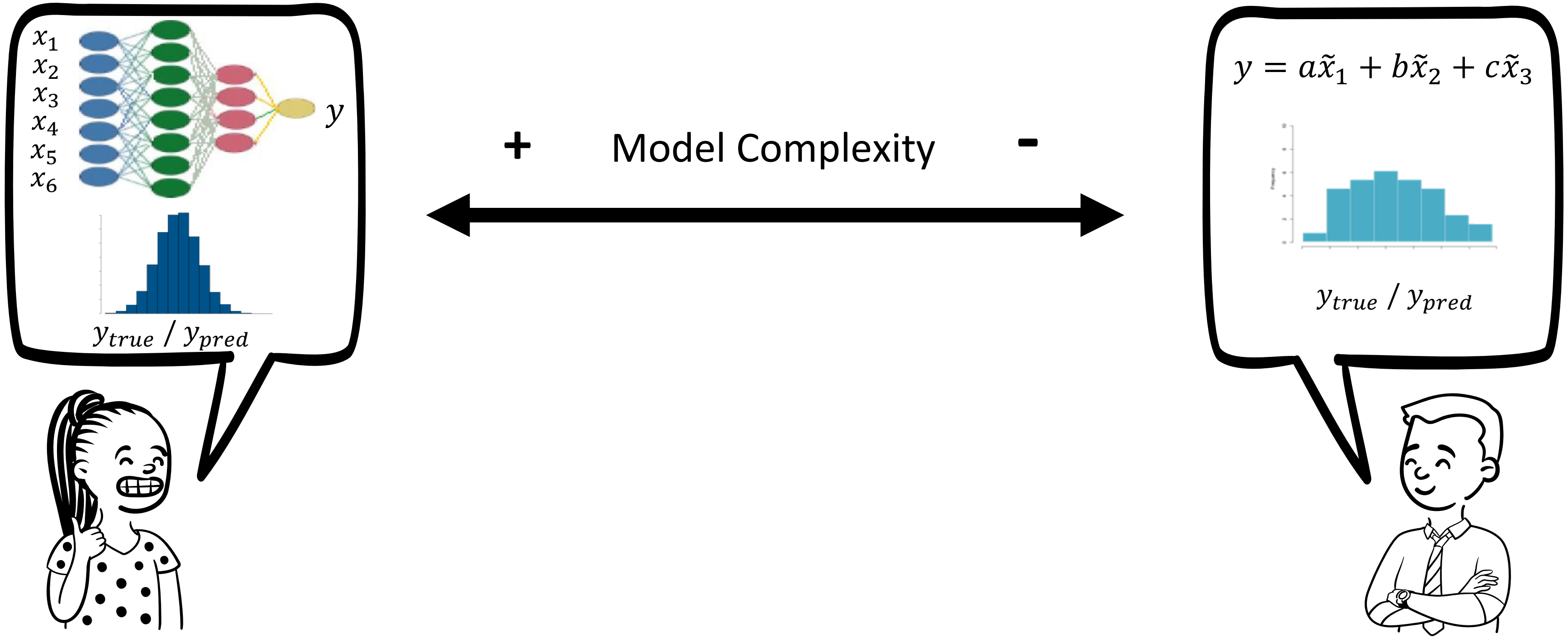
(5) Model training

(6) Model performance evaluation

(7) Set target errors for different model complexity levels

Target Errors for Different Model Complexity Levels





(1) Collection and preparation of data

(2) Feature selection

- Mechanics + Literature Review
- Normalized/scaled features + Mechanics → Engineered Features

(3) Choice of ML algorithm

- Complex ML model (e.g., ANN, RF regression)
- ENMs

(4) Selection hyper-parameters

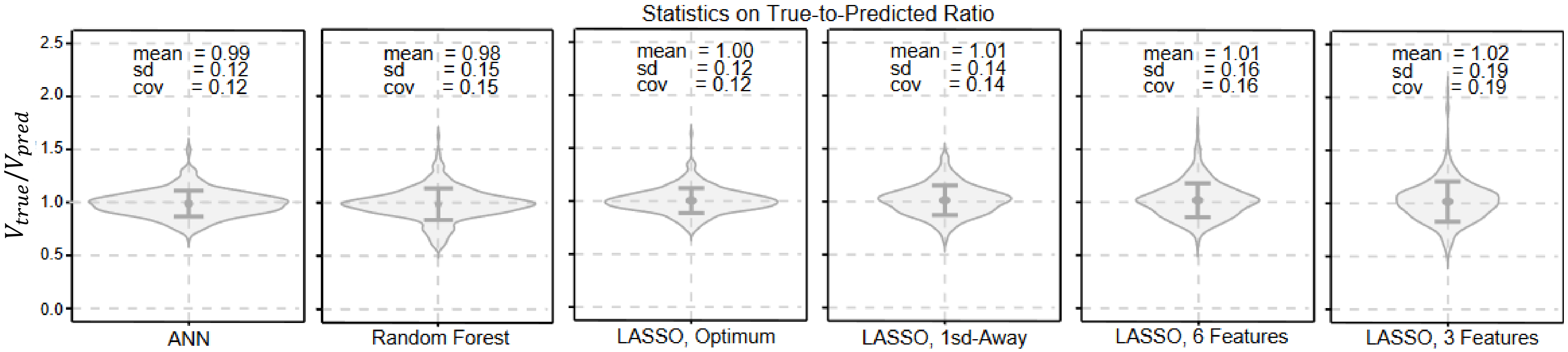
- Sensitivity analysis with iterative k-Fold CV
- Define underfitting levels to be used for the ENMs training

(5) Model training

(6) Model performance evaluation

(7) Set target errors for different model complexity levels

Target Errors for Different Model Complexity Levels



Requirements	Model Complexity Level	
	Complex ML Models	Simplified Models
Number of parameters	-	~ 3 - 6
V_{true}/V_{pred} mean ratio	0.99 – 1.01	0.98 – 1.02
COV	$\leq$ 0.12	0.16 – 0.19
Train. vs Test. Error Margin	$\pm$20% of the converging error	$\pm$10% of the converging error

Methodology to Get an Equation with a Code-Oriented Format

(Preliminary Results)

Methodology to Get an Equation with a Code-Oriented Format

(1) Identification of relevant parameters

(2) Differentiate parameters between:

- Materiality-related parameters (V_i)
- Other parameters $(\gamma_{j,i})$

(3) Re-arrange parameters ("$V_c + V_s$" format):

$$V_n = \beta_0 V_c + \sum_{i}^{N_i} \beta_i \left(\prod_{j}^{N_j} \gamma_{j,i} \right) V_i$$

(4) Re-write equation in its normalized version

(5) Train model

(6) Drop less significant variable until reaching target error

(7) Verify performance

(1) Identification of relevant parameters

(2) Differentiate parameters between:
- Materiality-related parameters (V_i)
- Other parameters $(\gamma_{j,i})$

(3) Re-arrange parameters into this equation format:

$$V_n = \beta_0 V_c + \sum_{i}^{N_i} \beta_i \left(\prod_{j}^{N_j} \gamma_{j,i} \right) V_i$$

(4) Re-write the equation in its normalized version

(5) Use iterative k-fold analysis to find regression coefficients and p-values

(6) Drop less significant parameter and go back to step (5) until reaching the target error

(7) Verify performance of the proposed equation

$$y = \beta_0 + \beta_1 \left(\frac{M_u}{V_u l_w}\right)^{a_c} \left(\frac{h_w}{l_w}\right)^{b_c} \left(1 + \frac{P_u}{A_g f_c'}\right)^{c_c}$$

$$+ \beta_2 \left(\frac{M_u}{V_u l_w}\right)^{a_{be}} \left(\frac{h_w}{l_w}\right)^{b_{be}} \left(1 + \frac{P_u}{A_g f_c'}\right)^{c_{be}} \frac{\rho_{be} f_{ybe} A_{be}}{A_g f_c'}$$

$$+ \beta_3 \left(\frac{M_u}{V_u l_w}\right)^{a_{wh}} \left(\frac{h_w}{l_w}\right)^{b_{wh}} \left(1 + \frac{P_u}{A_g f_c'}\right)^{c_{wh}} \frac{\rho_{wh} f_{ywh} A_{cv}}{A_g f_c'}$$

$$+ \beta_4 \left(\frac{M_u}{V_u l_w}\right)^{a_{wv}} \left(\frac{h_w}{l_w}\right)^{b_{wv}} \left(1 + \frac{P_u}{A_g f_c'}\right)^{c_{wv}} \frac{\rho_{wv} f_{ywv} A_{cv}}{A_g f_c'}$$

$$V_n = \alpha_c A_g f_c' + \alpha_s \left(A_{sb} f_y + \rho_{wh} f_{ywh} A_{cv}\right)$$

(1) Identification of relevant parameters

(2) Differentiate parameters between:
- Materiality-related parameters (V_i)
- Other parameters ($\gamma_{j,i}$)

(3) Re-arrange parameters into this equation format:

$$V_n = \beta_0 V_c + \sum_{i}^{N_i} \beta_i \left(\prod_{j}^{N_j} \gamma_{j,i} \right) V_i$$

(4) Re-write the equation in its normalized version

(5) Use iterative k-fold to find regression coefficients and p-values

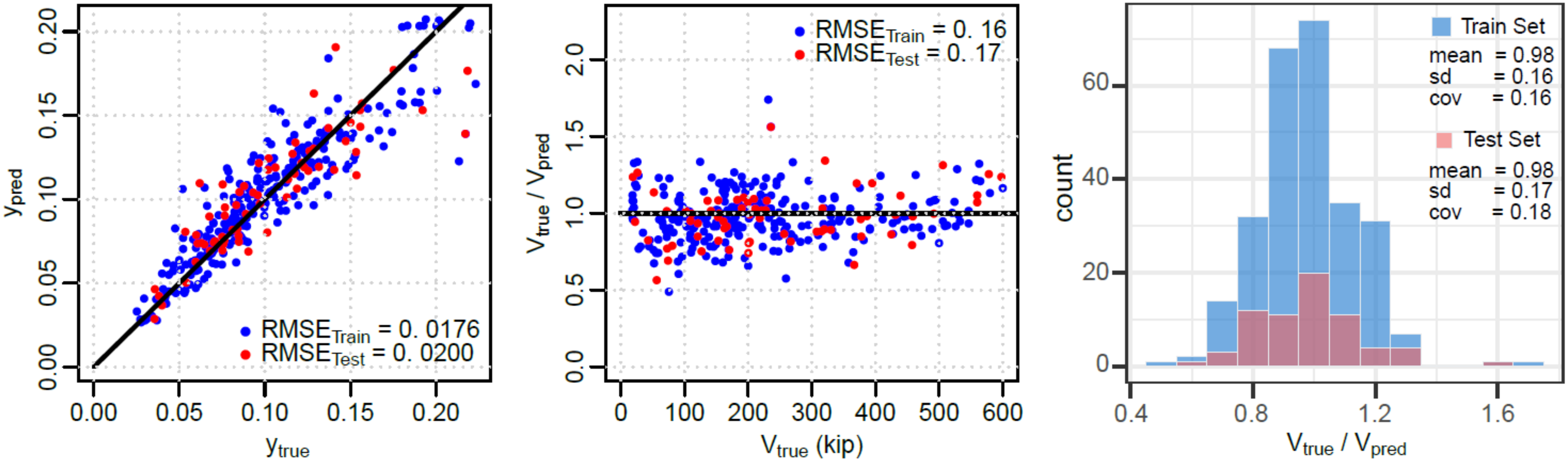
(6) Drop less significant parameter and go back to step (5) until reaching the target error

(7) Verify performance of the proposed equation

$$V_n = \alpha_c A_g f_c' + \alpha_s \left(A_{sb} f_y + \rho_{wh} f_{ywh} A_{cv} \right)$$

$$\alpha_c = \frac{1}{100} \left(14 \frac{\left(1 + \frac{P_u}{A_g f_c'}\right)^2}{\left(\frac{M_u}{V_u l_w}\right)^2} \right) \qquad \alpha_s = \frac{1}{3\left(\frac{M_u}{V_u l_w}\right)^{1/4}}$$

Proposed Equation



(1) Identification of relevant parameters

(2) Differentiate parameters between:
- Materiality-related parameters (V_i)
- Other parameters ($\gamma_{j,i}$)

(3) Re-arrange parameters into this equation format:

$$V_n = \beta_0 V_c + \sum_{i}^{N_i} \beta_i \left(\prod_{j}^{N_j} \gamma_{j,i} \right) V_i$$

(4) Re-write the equation in its normalized version

(5) Use iterative k-fold to find regression coefficients and p-values

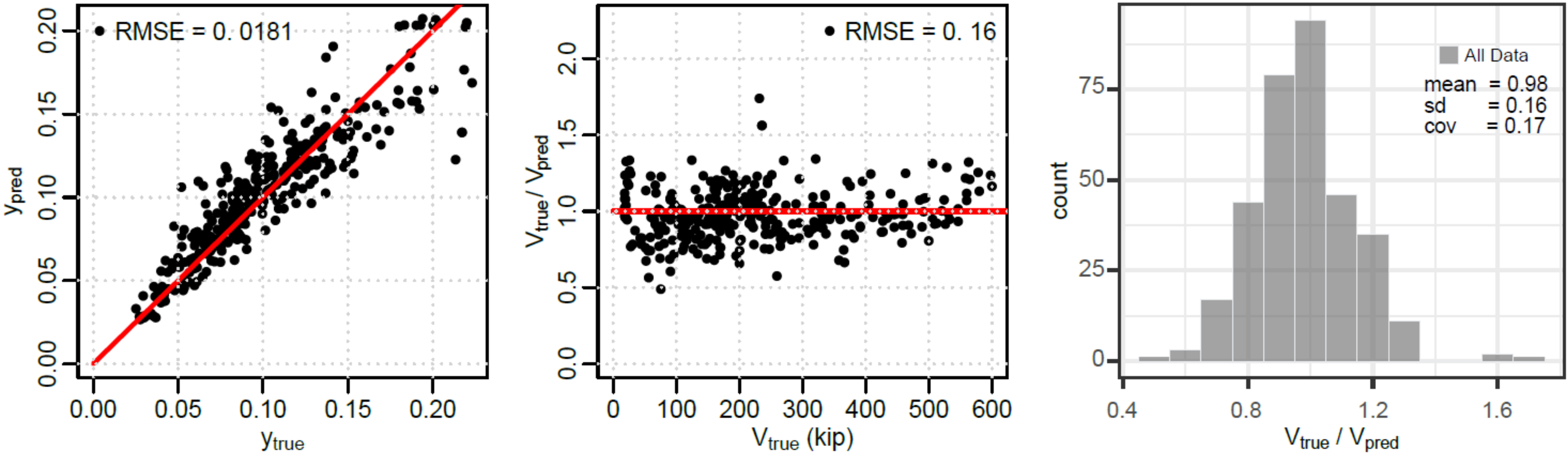
(6) Drop less significant parameter and go back to step (5) until reaching the target error

(7) Verify performance of the proposed equation

$$V_n = \alpha_c A_g f_c' + \alpha_s \left(A_{sb} f_y + \rho_{wh} f_{ywh} A_{cv} \right)$$

$$\alpha_c = \frac{1}{100} \left(14 \frac{\left(1 + \frac{P_u}{A_g f_c'}\right)^2}{\left(\frac{M_u}{V_u l_w}\right)^2} \right) \qquad \alpha_s = \frac{1}{3\left(\frac{M_u}{V_u l_w}\right)^{1/4}}$$

Prpoposed Equation Applied on Complete Dataset



(1) Identification of relevant parameters

(2) Differentiate parameters between:
- Materiality-related parameters (V_i)
- Other parameters ($\gamma_{j,i}$)

(3) Re-arrange parameters into this equation format:

$$V_n = \beta_0 V_c + \sum_i^{N_i} \beta_i \left(\prod_j^{N_j} \gamma_{j,i} \right) V_i$$

(4) Re-write the equation in its normalized version

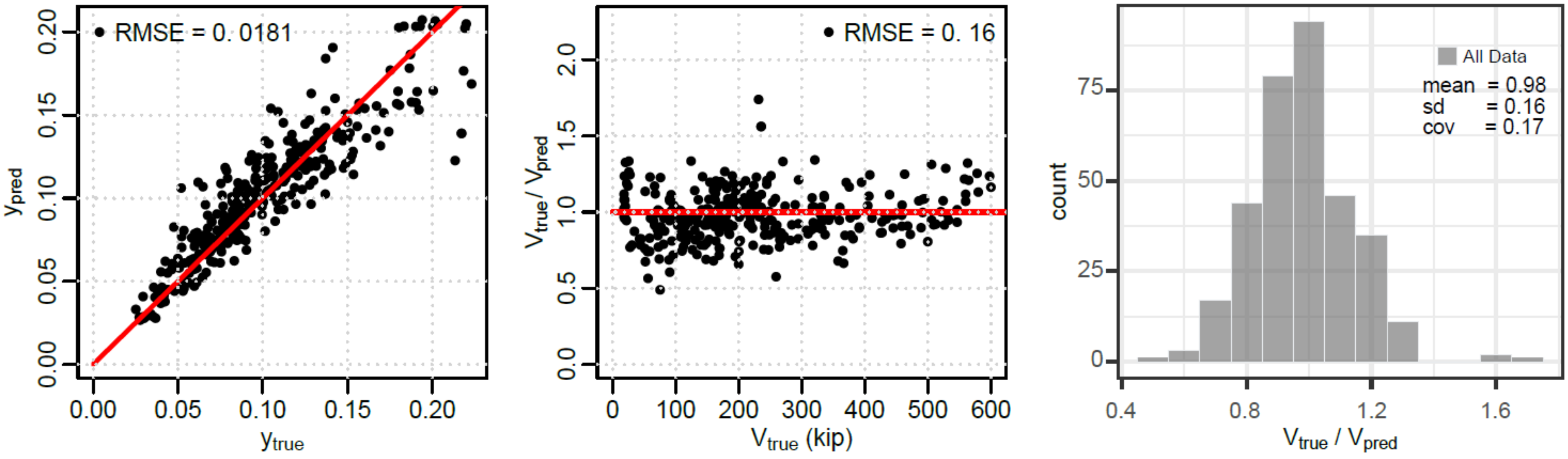
(5) Use iterative k-fold to find regression coefficients and p-values

(6) Drop less significant parameter and go back to step (5) until reaching the target error

(7) Verify performance of the proposed equation

Requirements	Simple Models
Number of parameters	~ 3 - 6
V_{true}/V_{pred} mean ratio	0.98 – 1.02
COV	0.16 – 0.19
Train. vs Test. Error Margin	$\pm 10\%$ of the converging error

Prpoposed Equation Applied on Complete Dataset



Relevant points to investigate:

(1) Does the equation make sense?

(Contribution from each term, parameter ranges, etc.)

(2) How to use the equation on unsymmetrical cross-section shaped walls?

(3) Modification of wall shear strength upper limit?

(4) Is the model complexity level acceptable?

(5) How to use the equation in the design?

(6) Consequences of the equation in the design

Upcoming articles with more details on the results for both:

(1) Target errors for different model complexity levels

(2) Proposed RC wall shear strength equation

Matías Rojas León

https://mrojasl.cl

mrojasleon@g.ucla.edu