Supplementary Data

1.1 Supplementary Tables

Table S1: Numerical parameters present in the erosion database, along with their range of values and mean value.

Feature	***Min***	***Max***	***Mean***
Particle Size (µm)	8	490	149
Size (mesh)	46	1000	141
Size range (µm)	1-250	100-1000	N/A
Velocity (m/s)	10	554	136
V Pressure (PSI)	20	116	76
V_{air} (m/s)	275	400	342
V_{KinE} (m/s)	241	348	301
Incident angle	3	90	52
Erodent flux (g/(cm^2*s))	1	19	4.9
Total Erodent mass (g)	20	1500	318
Particle feed rate (g/minute)	0.37	361	29
Dust Concentration (mg/ft^3)	25	80	59
Working Distance (mm)	6	102	37
Nozzle Diameter (mm)	0.8	50	6.2
Nozzle length (mm)	36	1600	364
Temperature (°C)	-7	800	76
Publication Year	1960	2023	1995
Target Hardness (GPa)	0.2	8	2.4
Target Density (kg/m^3)	2660	11400	5591
Target Elastic Modulus (GPa)	3	216	139
Target Poisson Ratio	0.27	0.44	0.32
Target Thermal conductivity (W/m*K)	6.8	398	62
Target Yield Strength (MPa)	9	1700	607
Target Tensile Strength (MPa)	15	1800	750
Target Fracture toughness (MPa*m$^{1/2}$)	10	223	79
Target Melting Temp (°C)	328	1704	1338
Erodent Density (kg/m^3)	2230	4000	3023
Erodent Hardness (GPa)	4.5	25.5	16.9
Erodent Melting Temp (°C)	1283	2797	2050
Erodent Young's Modulus (GPa)	63	410	223
Erodent Poisson's Ratio	0.14	0.27	0.18
Erodent Fracture toughness (MPa* m$^{1/2}$)	0.63	4.6	2.4
Erosion Rate (mg/g)	0.001	41.95	1.43

Table S2: Categorical parameters present in the erosion database.

Feature	*Values*
Particle Shape	Angular, Spherical
Carrier Gas	Air, Nitrogen
Test type	Gas Blast, Wind tunnel, Whirling Arm
Target Ductile/Brittle	Ductile, Brittle
Hardness Measurement	Vickers, Rockwell
ASTM G76 Mentioned	Yes, No

Table S3: Missing features in the database, and the number of instances where they are missing.

Feature	***# of missing***
Particle Size (µm)	3
Target Tensile Strength (MPa)	6
Target Hardness (GPa)	9
Target Melting Temp (°C)	9
Target Poisson Ratio	10
Target Yield Strength (MPa)	14
Erosion Rate (mg/g)	18
Target Fracture toughness (MPa*$m^{1/2}$)	24
Erodent Density (kg/m^3)	40
Erodent Hardness (GPa)	40
Erodent Melting Temp (°C)	40
Erodent Young's Modulus (GPa)	40
Erodent Poisson's Ratio	40
Erodent Fracture toughness (MPa*$m^{1/2}$)	40
Velocity (m/s)	41
Incident angle	42
Carrier Gas	85
Particle Shape	106
Nozzle Diameter (mm)	401
Working Distance (mm)	517
Particle Feed rate (g/minute)	729
Total Erodent mass (g)	732
Erosion Rate (mm^3/g)	748
Size range (µm)	772
Size (mesh)	797
Nozzle length (mm)	821
$V_{Pressure}$ (PSI)	887
Dust concentration (mg/ft^3)	908
Erodent flux (g/(cm^2*s))	975
V_{air} (m/s)	987
V_{KinE} (m/s)	987

1.2 Supplementary Data

The optimized hyperparameter values for each model in section 3.2.1 are listed below. These values, as well as the code used to obtain them, are also available on GitHub at https://github.com/stebrown/MLforSPE.

Multiple Linear Regression

N/A

Decision Tree

{'max_depth': 33, 'min_samples_split': 10}

K-Nearest Neighbors

{'K': 2}

Random Forest

{'max_depth': 18, 'n_estimators': 7}

Neural Network

{'hidden_layer_sizes': (300, 150, 75), 'learning_rate_init': 0.001625, 'max_iter': 200}

Support Vector Regression

{'svr__C': 10, 'svr__epsilon': 0.05, 'svr__gamma': 0.1, 'svr__kernel': 'poly'}

XGBoost

{'learning_rate': 0.13, 'n_estimators': 300}

1.3 Supplementary Figures

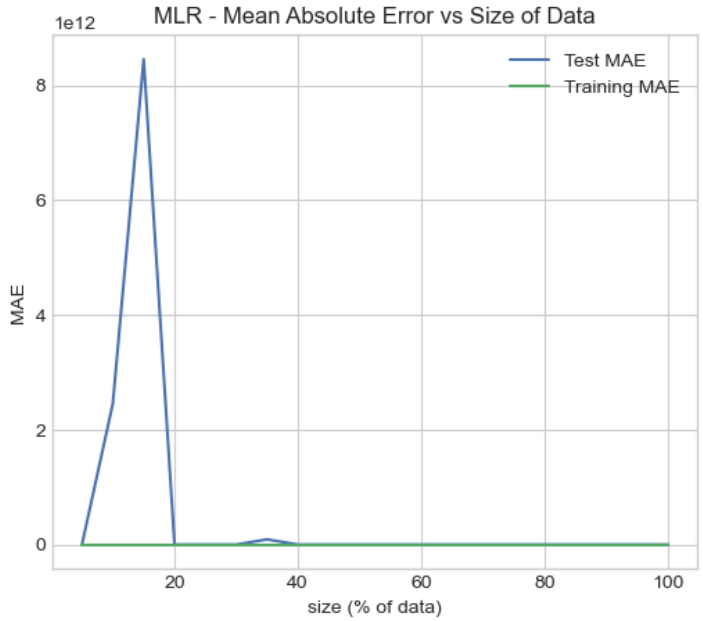

Figure S1: Effect of increasing amount of training data on training and test MAE for multiple linear regression.

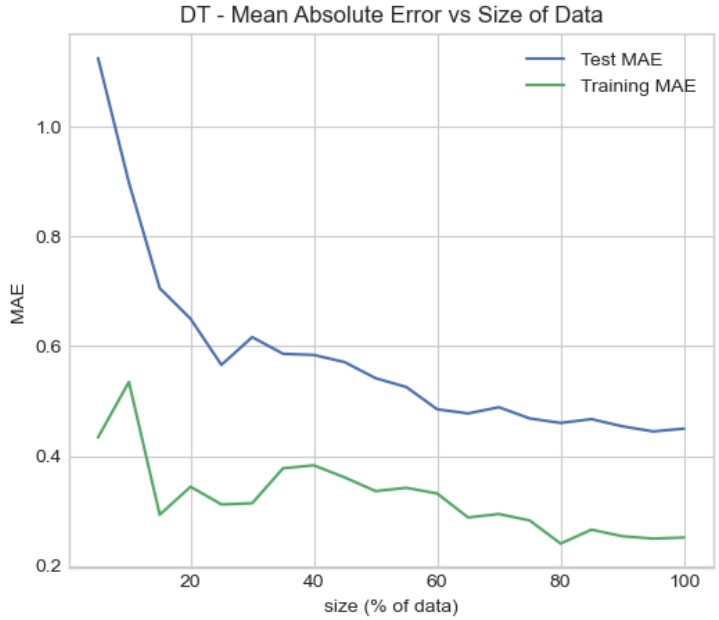

Figure S2: Effect of increasing amount of training data on training and test MAE for decision trees.

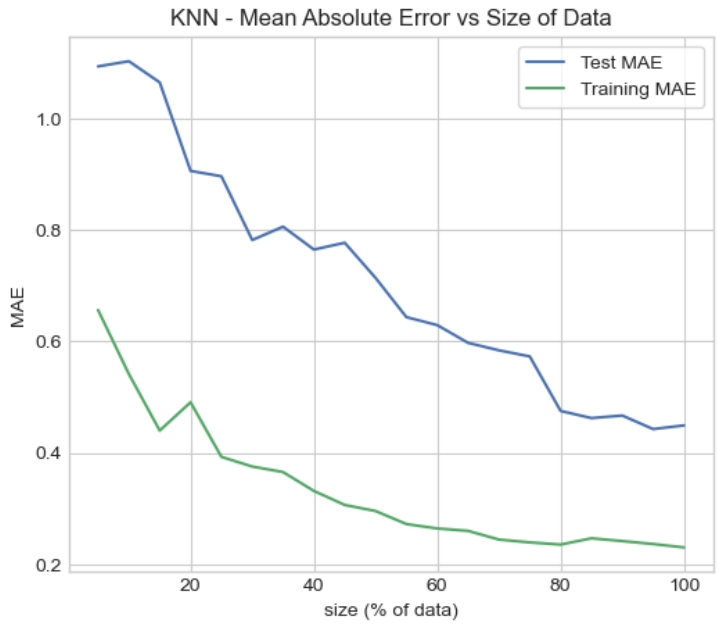

Figure S3: Effect of increasing amount of training data on training and test MAE for K-nearest neighbours.

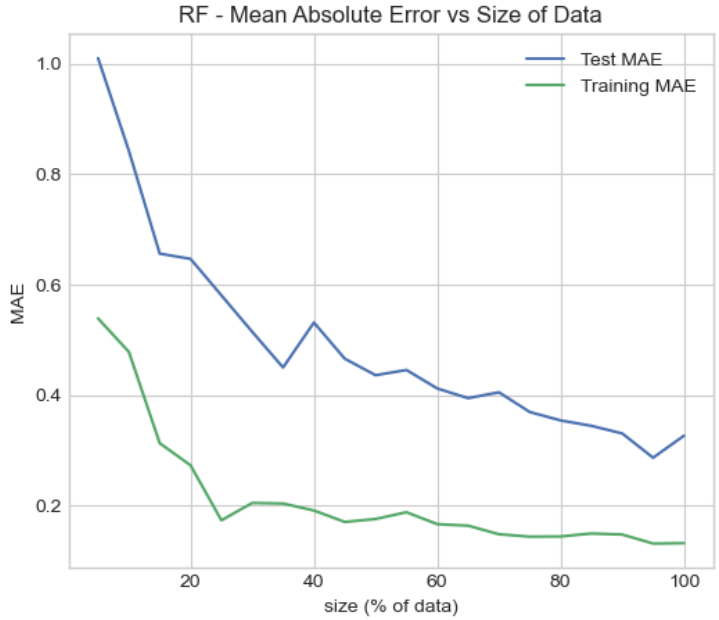

Figure S4: Effect of increasing amount of training data on training and test MAE for random forest.

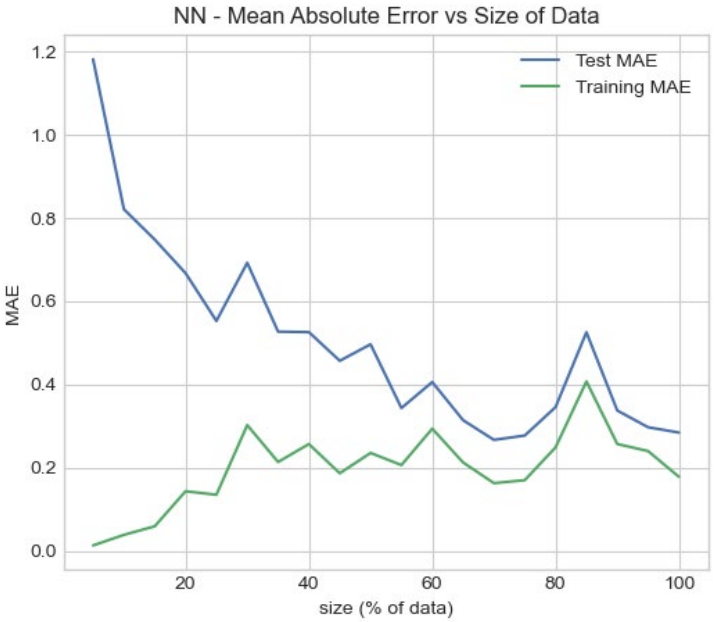

Figure S5: Effect of increasing amount of training data on training and test MAE for neural network.

Figure S6: Full partial dependence plots for optimized XGBoost model, indicating the influence of each feature.