Report Summary

The comparison chart of variable importance of MARS Model & Random Forest Regression Model are shown and important variables are selected from the comparison. The value chart having the metrics of both the models are shown side by side and the best model is discussed and evaluated based on the chart. All the computations and calculations are provided in the python codes. The plot between predicted and observed data of the MARS Model & Relative Information Plot of the Random Forest Regression are shown. At the end, comments are made on the model free method found from the mutual information and correlation plot with the developed MARS Model & Random Forest Regression Model.

The Variable importance of MARS Model & Random Forest Regression Model

MARS Model			Random Forest Regression Model
Variable	RSS	GCV	Relative Importance (X 10^-3)
pHsoil	0	0	6.35529
SOM	0	0	7.59872
SILT	0.01	0	8.03987
WC15BAR	0	0	7.90923
WC13BAR	0	0	4.75272
CEC	0	0	5.82418
KSAT	0.01	0.01	5.60156
CLAY	0	0	11.64267
SAND	0	0	5.56489
BULKDENSIT	0	0	26.02432
WellDepth	0.03	0.02	29.69777
Aqbott	0	0	16.1885
F	0.02	0.01	34.31689
GWTemp	0	0	15.84
Mg	0	0	59.54016
pHGW	0.01	0	21.51849
K	0.05	0.01	16.98384
SARGW	0.01	0	17.9087
Na	0.08	0	11.40444
SC	0	0	50.14261
SO4	0.05	0.04	38.81802
TDS	0.58	0.81	234.58193
As	0	0	14.81675
HCO3	0	0	16.83386
B	0.02	0.01	31.96667
Ca	0.01	0	152.29443
Cl	0.05	0.02	33.16684
DWT	0.05	0.05	44.76783
PRCP	0	0	25.49979
LSD	0.03	0.02	16.195
AqDepth	0	0	28.20404

Identified Important Variables (MARS Model):

(1) WellDepth (2) F (3) K (4) SO4 (5) TDS (6) CL (7) DWT (8) LSD (9) Ca (10) B (11) KSAT

Identified Important Variables (Random Forest Regression Model): Threshold of Importance is taken as 28 X 10^-3

(1) WellDepth (2) F (3) Mg (4) SC (5) SO4 (6) TDS (7) B (8) Ca (9) Cl (10) DWT (11) AqDepth

* Here, 8 variables are commonly selected as important features by both the models

Model Evaluation Chart

| Metric Parameter	MARS (Train)	MARS (Test)	Random Forest (Train)	Random Forest (Test)
MSE	0.06439921	0.0855760	0.03161246	0.0734863
MAE	0.19433870	0.2063314	0.13909204	0.1928508
COR	0.78262107	0.6642074	0.91534926	0.7028815
Overall (KGE)	0.69257977	0.63332922	0.73461936	0.56047369
COR (KGE)	0.78262107	0.6642074	0.91534926	0.7028815
VAR (KGE)	0.78262107	0.85356119	0.74848406	0.6763047
BIAS (KGE)	1.0	0.98430087	0.99906719	0.98880487

Model Evaluation Comments

Here, the Random Forest Regression Model performs better than MARS Model in all aspects of the metrics provided.
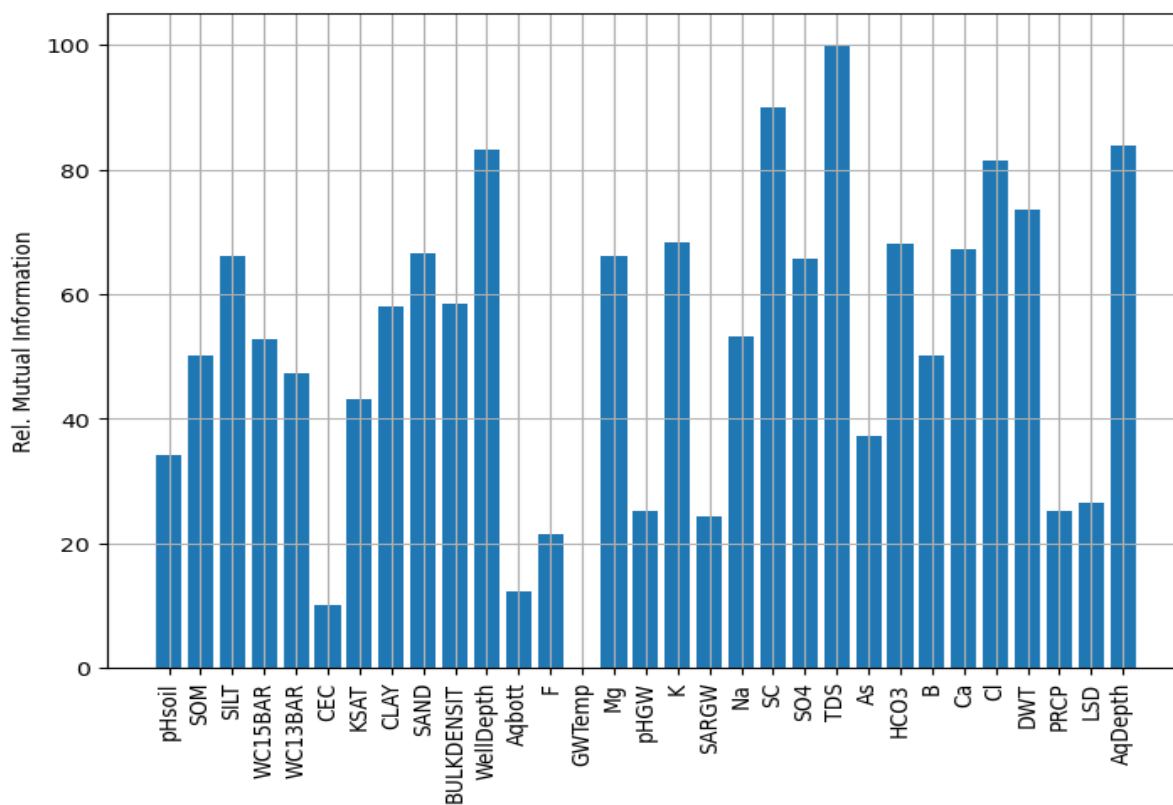
MARS Model Plot
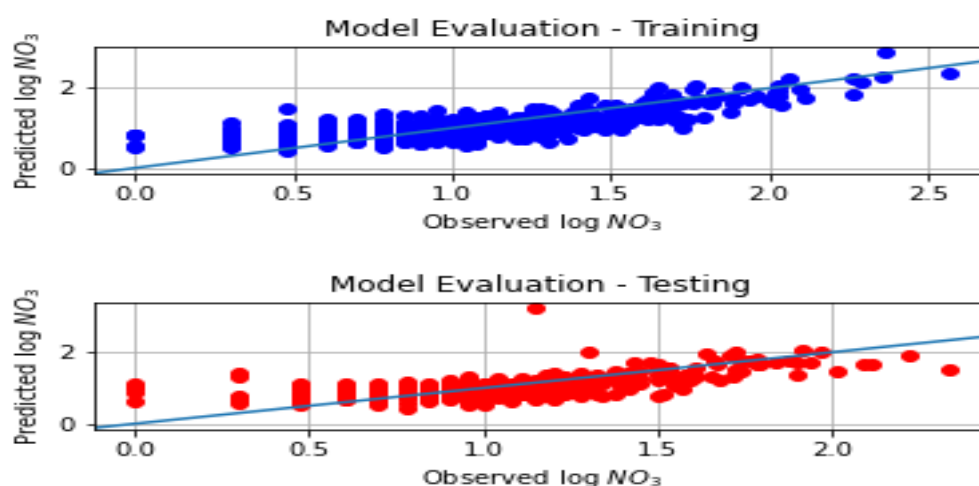
Random Forest Regression Relative Importance Plot
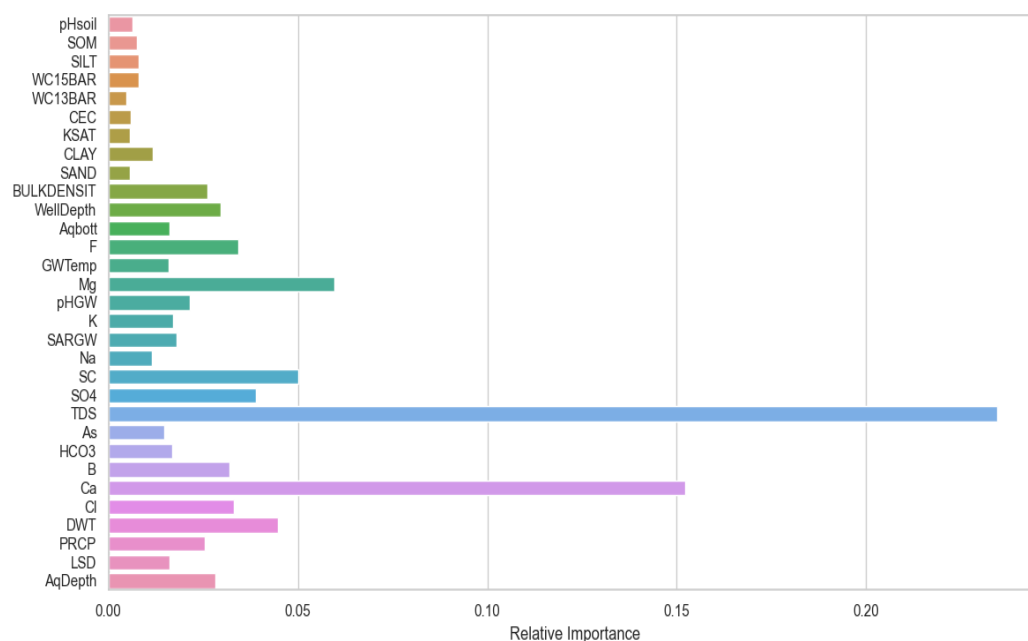
Model Free Method

Relative Mutual Information

Selected Variables from both MARS Model & Random Forest Regression:

(1) WellDepth (2) F (3) SO4 (4) TDS (5) B (6) Ca (7) Cl (8) DWT

From the model free method above it can be seen that all the variables selected in the model dependent method shows a high value. Even though there are many other variables such as SAND & AqDepth which shows high relative importance in model free method are found to be having no such significance in the model dependent method. The correlation plot from the model independent method could also be further shown.