

Comparing XGBoost Models

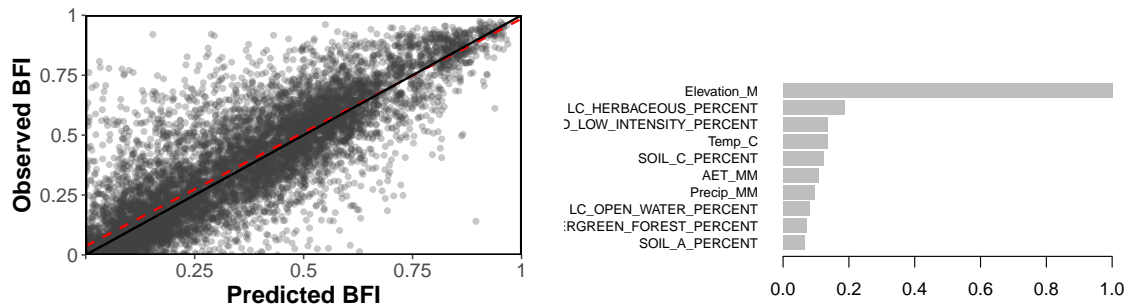
Caelum Mroczek

Full Statewide Model

This model uses all 45 features from the original data set.

Hyper-parameters

nrounds	max_depth	eta	gamma	colsample_bytree	min_child_weight	subsample
650	7	0.05	0.1	0.6	10	1



- (a) Observed BFI vs. Predicted BFI. Solid black line indicates the 1-to-1 line, red dashed line is regressed to the data.
- (a) Top 10 most important features for the statewide model

Statewide Model Results

R2	MSE	RMSE	MAE	Nash.Sutcliffe	pbias
0.7599832	0.0168738	0.1298992	0.0875466	0.7517887	-5.8

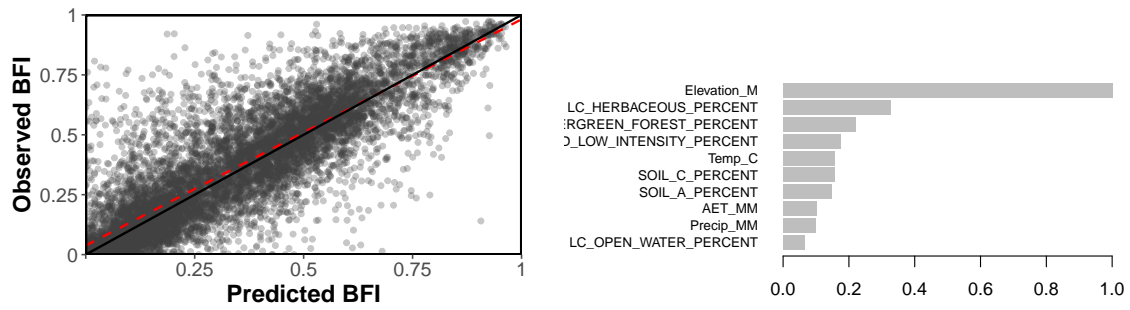
Statewide Model - Goodness of Fit Stats

Trimmed Statewide Model

This model has undergone feature selection uses only the top 10 features from the **Full Statewide Model**.

Hyper-parameters

nrounds	max_depth	eta	gamma	colsample_bytree	min_child_weight	subsample
700	7	0.05	0.1	0.8	10	1



(a) Observed BFI vs. Predicted BFI. Solid black line indicates the 1-to-1 line, red dashed line is regressed to the data. (a) Top 10 most important features for the trimmed model

Trimmed Model Results

R2	MSE	RMSE	MAE	Nash.Sutcliffe	pbias
0.761536	0.0167591	0.1294568	0.0872035	0.7534768	-5.5

Trimmed Model - Goodness of Fit Stats

Comparing Results

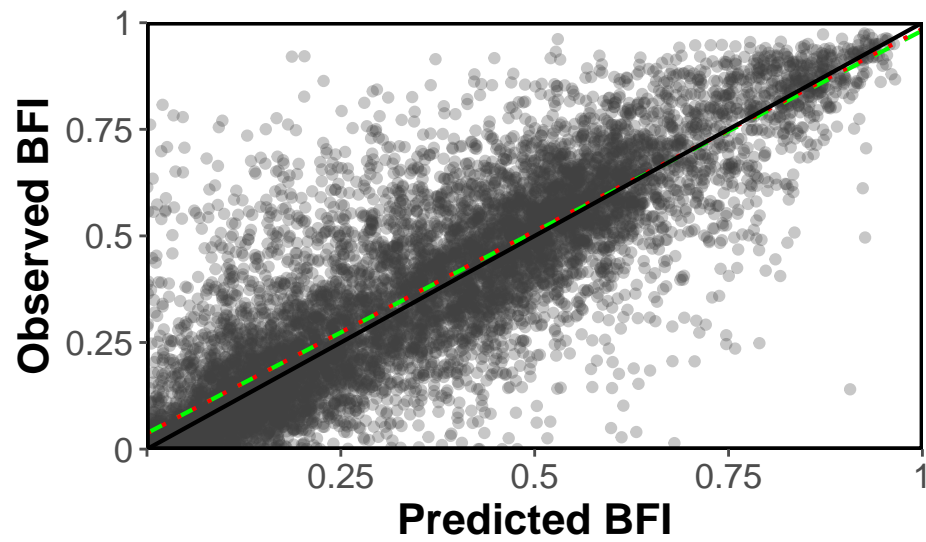


Figure 5: Observed BFI vs. Predicted BFI. Solid black line indicates the 1-to-1 line, red dotted line is regressed to the statewide model data, green dashed line is regressed to the trimmed model data.

Trimmed Model Results