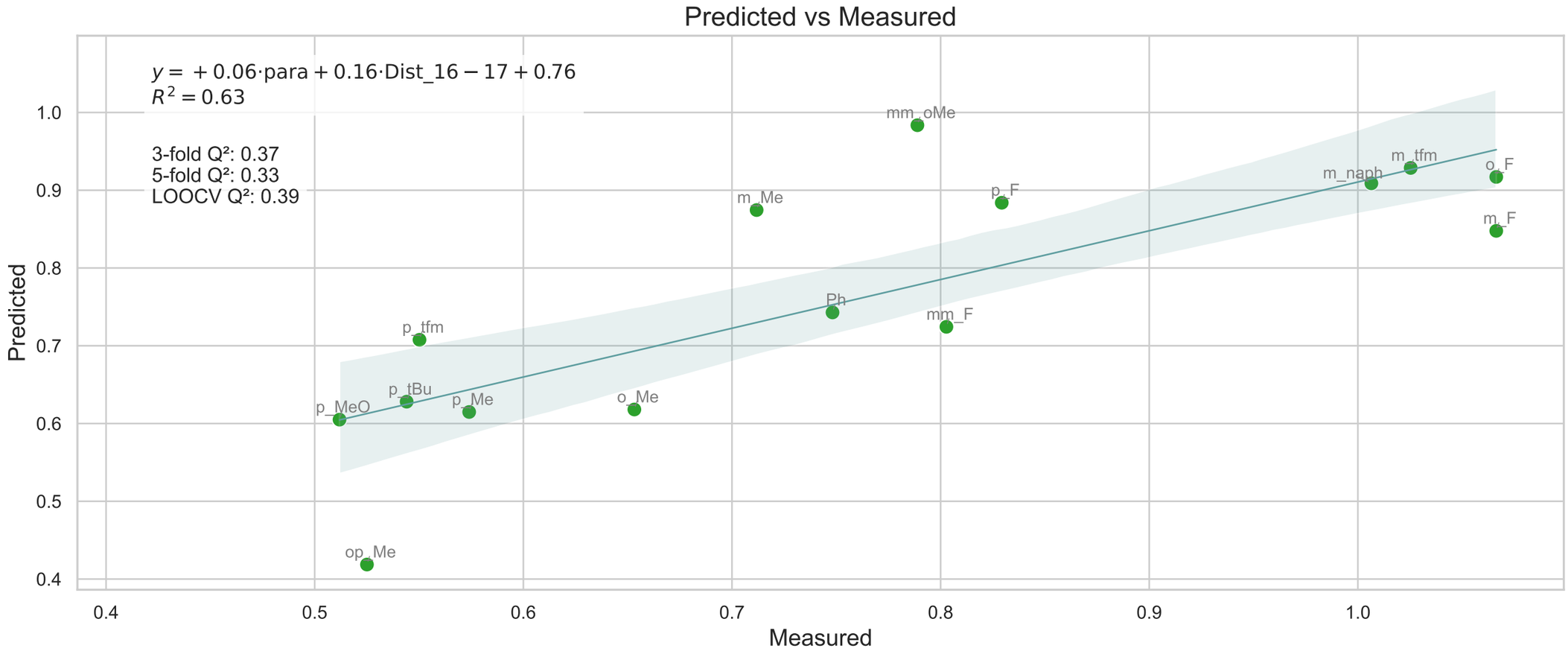

Predicted vs Measured
$y = +0.06 \cdot \text{para} + 0.16 \cdot \text{Dist_16} - 17 + 0.76$
$R^2 = 0.63$
3-fold Q²: 0.37
5-fold Q²: 0.33
LOOCV Q²: 0.39
mm_oMe
m_Me
p_F
m_tfm
m_naph
o_F
m_F
Ph
mm_F
p_tfm
p_tBu
p_MeO
p_Me
o_Me
op_Me
Predicted
Measured
0.4
0.5
0.6
0.7
0.8
0.9
1.0

3-fold CV Metrics

Q2_3_Fold	MAE_3	RMSD_3
0.374	0.131	0.154

5-fold CV Metrics

Q2_5_Fold	MAE_5	RMSD_5
0.326	0.139	0.161

LOOCV Metrics

Q2_LOOCV	MAE_LOOCV	RMSD_LOOCV
0.391	0.135	0.154

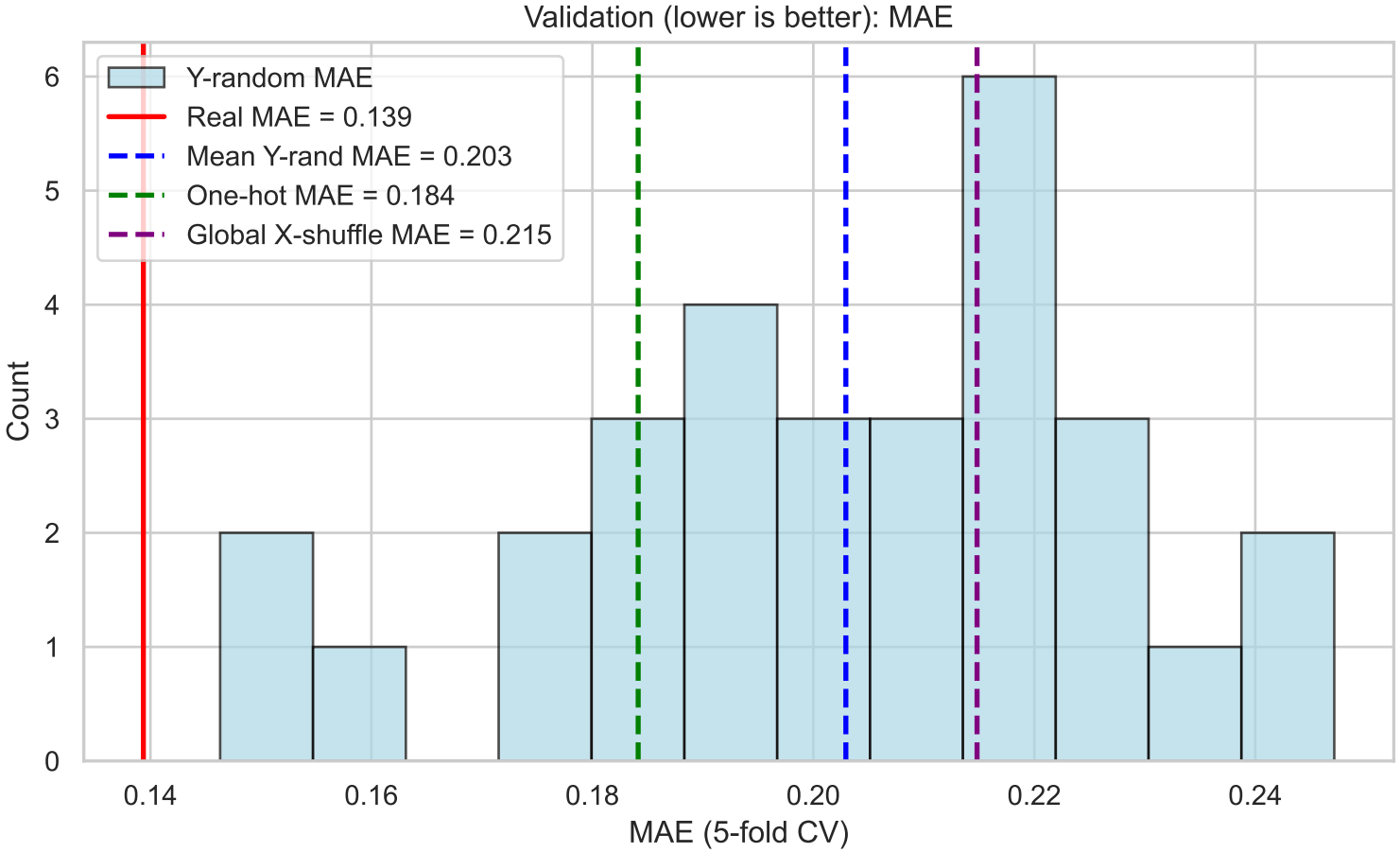
Variance Inflation Factors

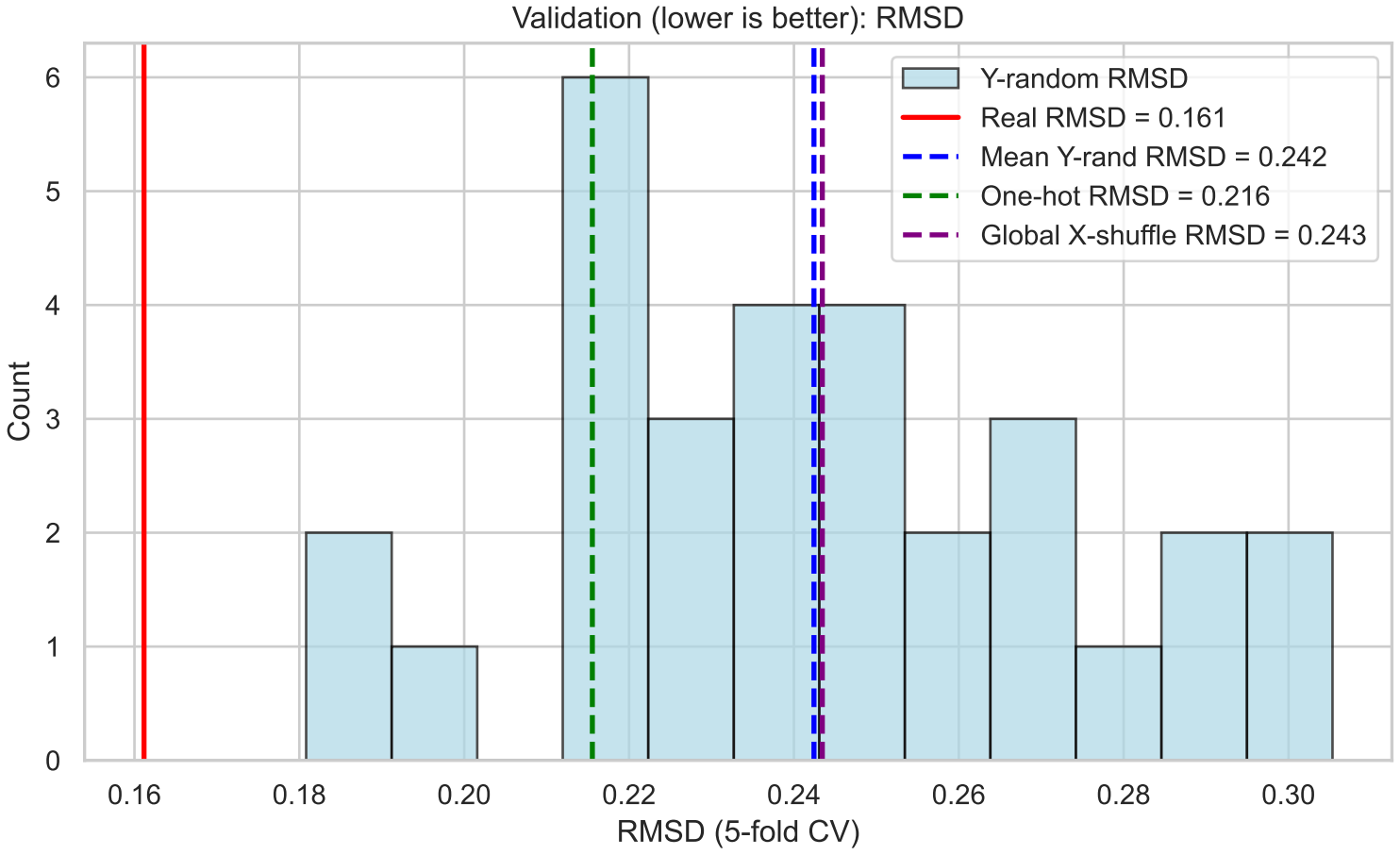
variable	VIF
para	1.04
Dist_16-17	1.04

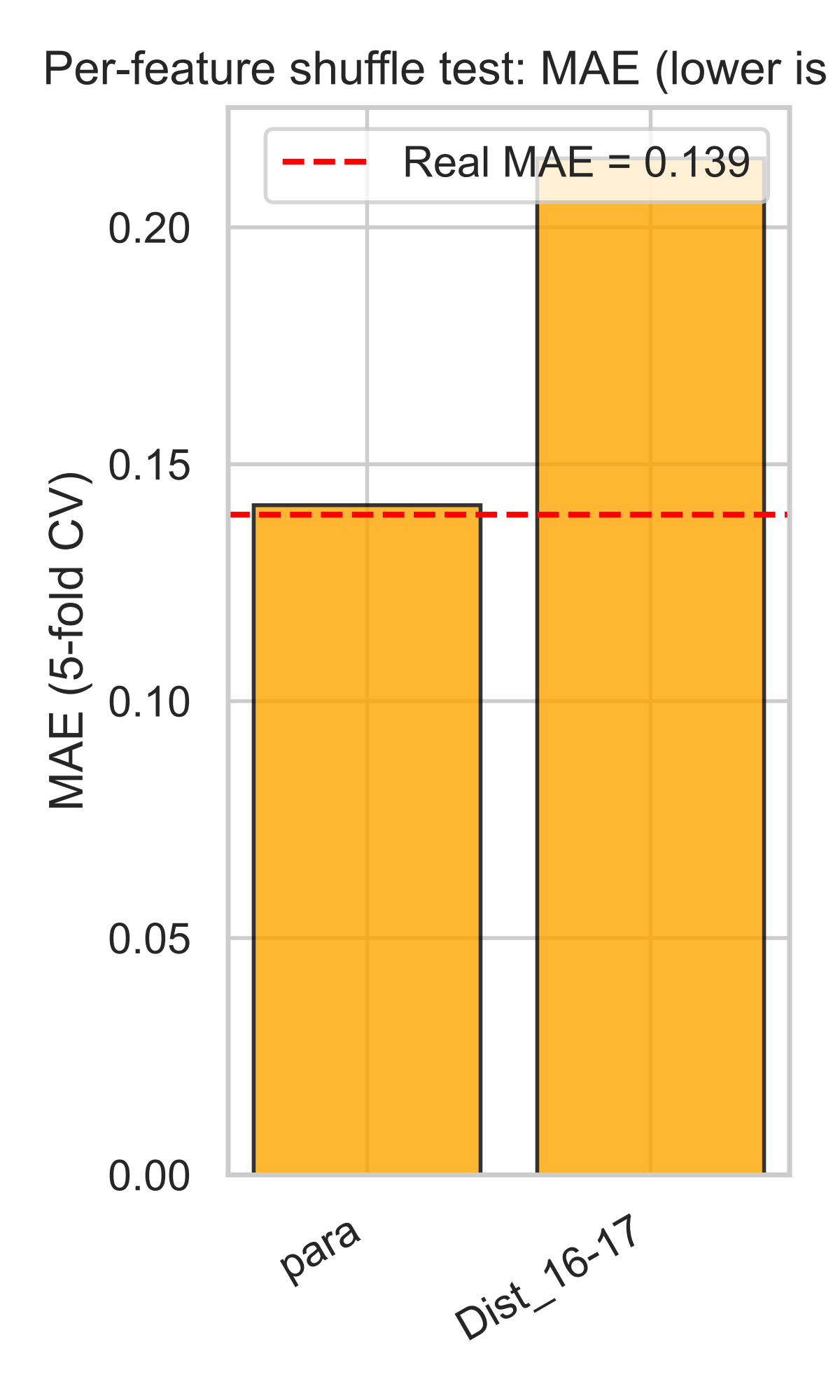
Coefficient Estimates

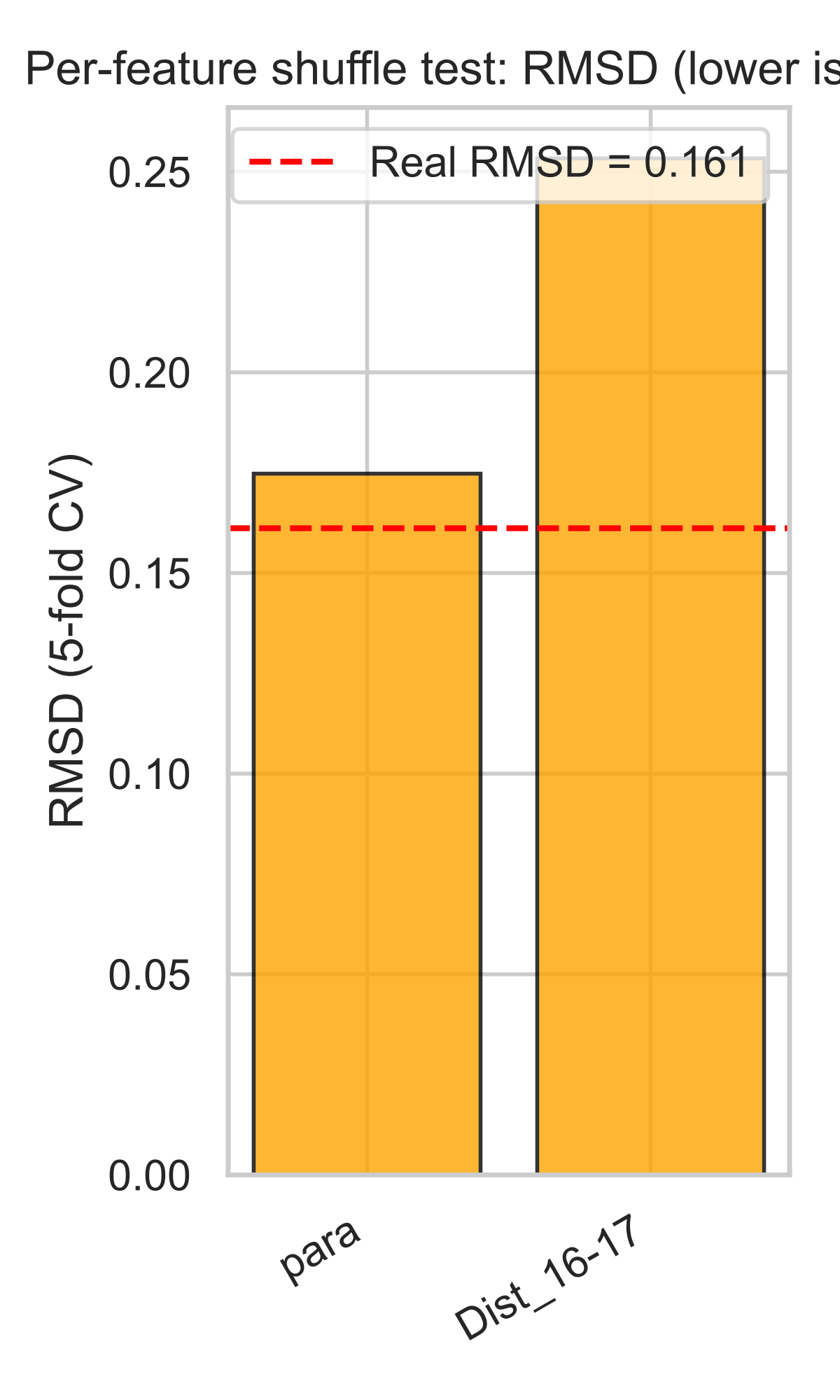
Estimate	Std. Error	t value	p value
0.76	0.035	21.886	0.0
0.055	0.035	1.559	0.145
0.157	0.035	4.433	0.001

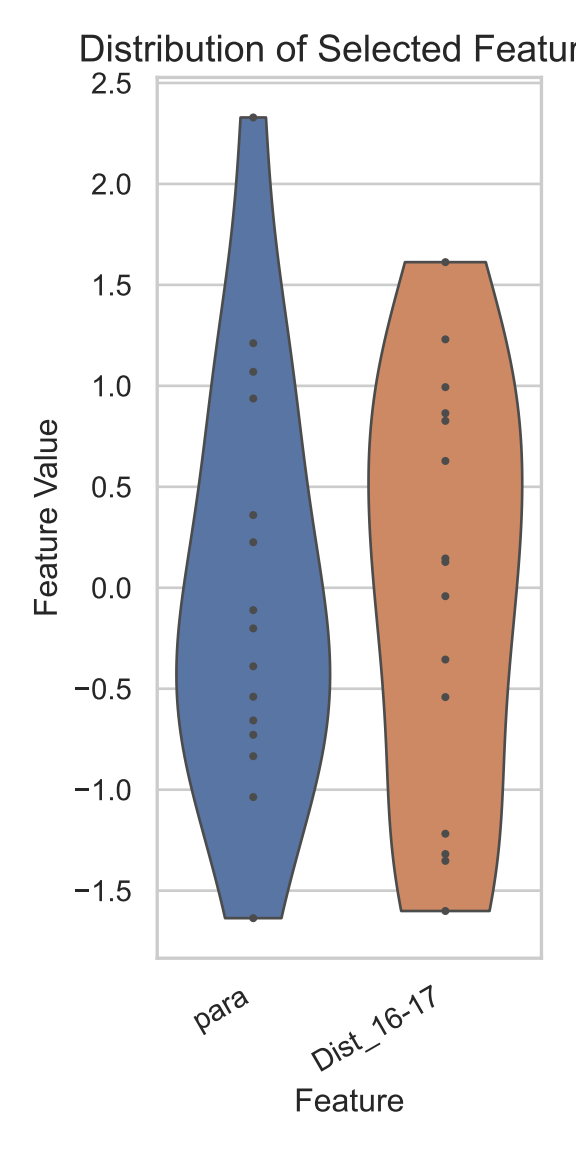
Left-out Predictions (none)

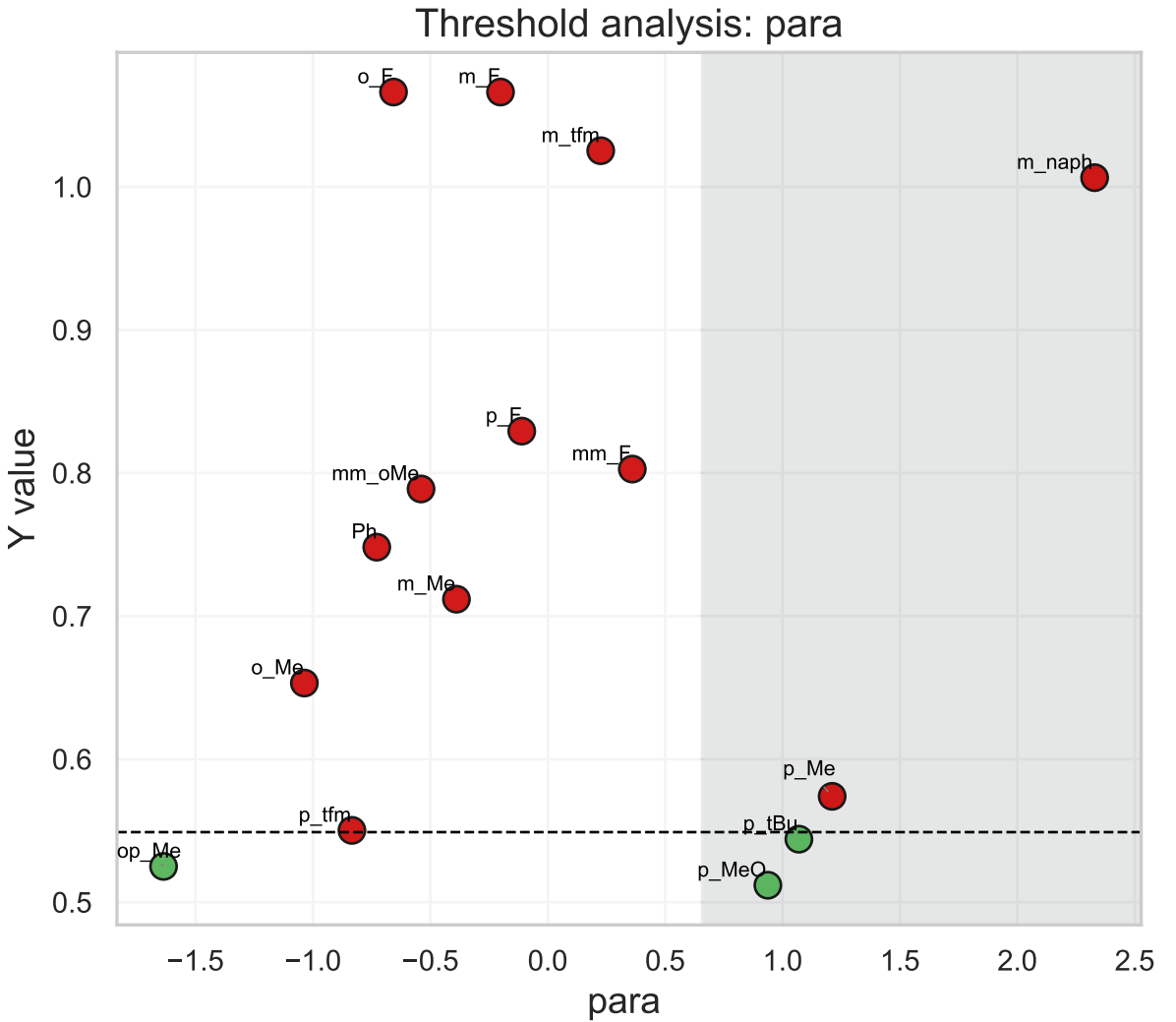

Validation (lower is better): MAE
Y-random MAE
Real MAE = 0.139
Mean Y-rand MAE = 0.203
One-hot MAE = 0.184
Global X-shuffle MAE = 0.215
Count
0
1
2
3
4
5
6
0.14
0.16
0.18
0.20
0.22
0.24
MAE (5-fold CV)

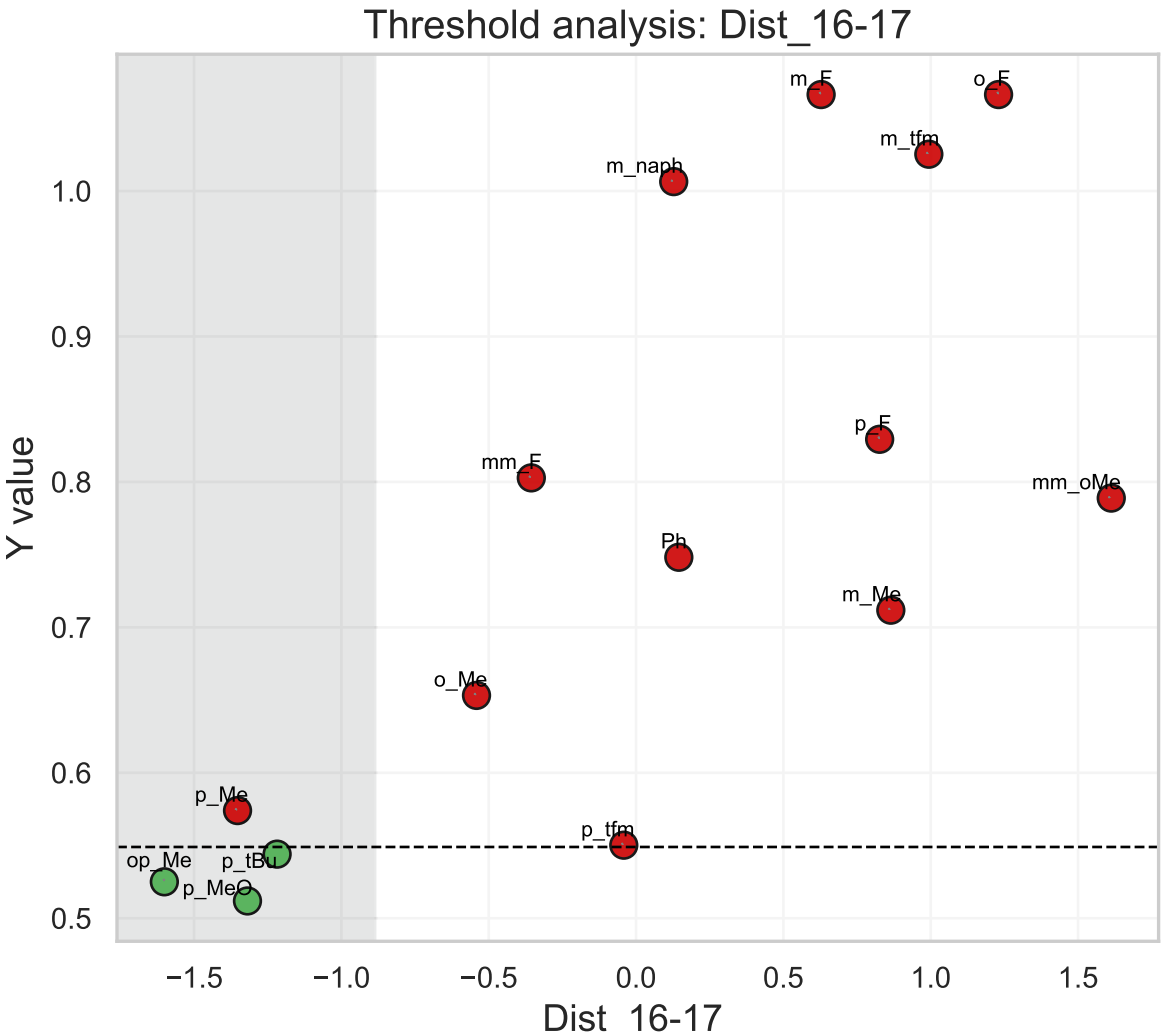

Validation (lower is better): RMSD
Y-random RMSD
Real RMSD = 0.161
Mean Y-rand RMSD = 0.242
One-hot RMSD = 0.216
Global X-shuffle RMSD = 0.243
Count
0
1
2
3
4
5
6
0.16
0.18
0.20
0.22
0.24
0.26
0.28
0.30
RMSD (5-fold CV)


Per-feature shuffle test: MAE (lower is
Real MAE = 0.139
MAE (5-fold CV)
0.20
0.15
0.10
0.05
0.00
para
Dist_16-17


Per-feature shuffle test: RMSD (lower is
Real RMSD = 0.161
RMSD (5-fold CV)
0.25
0.20
0.15
0.10
0.05
0.00
para
Dist_16-17


Distribution of Selected Featur
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
Feature Value
para
Dist_16-17
Feature


Threshold analysis: para
o_F
m_F
m_tfm
m_naph
p_F
mm_F
mm_oMe
Ph
m_Me
o_Me
p_Me
p_tfm
p_tBu
op_Me
p_MeO
1.0
0.9
0.8
0.7
0.6
0.5
Y value
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
2.5
para


Threshold analysis: Dist_16-17
m_F
o_F
m_tfm
m_naph
p_F
mm_F
mm_oMe
Ph
m_Me
o_Me
p_Me
p_tfm
op_Me
p_tBu
p_MeO
Y value
1.0
0.9
0.8
0.7
0.6
0.5
−1.5
−1.0
−0.5
0.0
0.5
1.0
1.5
Dist_16-17