Top 5 Models (ranked by R^2)

Overview

Rank	R^2 (train)	#Features	#Samples	Formula
1	0.914	4	12	-2-3- + -2-9- + Dist_1-2 + dip_x
2	0.904	4	12	-1-2- + Dist_1-2 + Dist_16-17 + dip_x
3	0.904	1	12	['-1-2-', 'Dist_1-2', 'Dist_16-17', 'dip_x']
4	0.904	1	12	['-1-2-', 'Dist_1-2', 'Dist_16-17', 'dip_x']
5	0.894	4	12	para + -18-20- + Dist_16-17 + L

Rank #1 | R^2=0.914 | Features=4 | Samples=12
-2-3- + -2-9- + Dist_1-2 + dip_x

Coefficients

Estimate	Std. Error	t value	p value
0.7459	0.02136	34.93	4.093e-09
-0.08011	0.0205	-3.908	0.005839
-0.08343	0.02441	-3.419	0.01116
-0.1581	0.02198	-7.195	0.0001783
-0.1314	0.02489	-5.282	0.001146

VIF

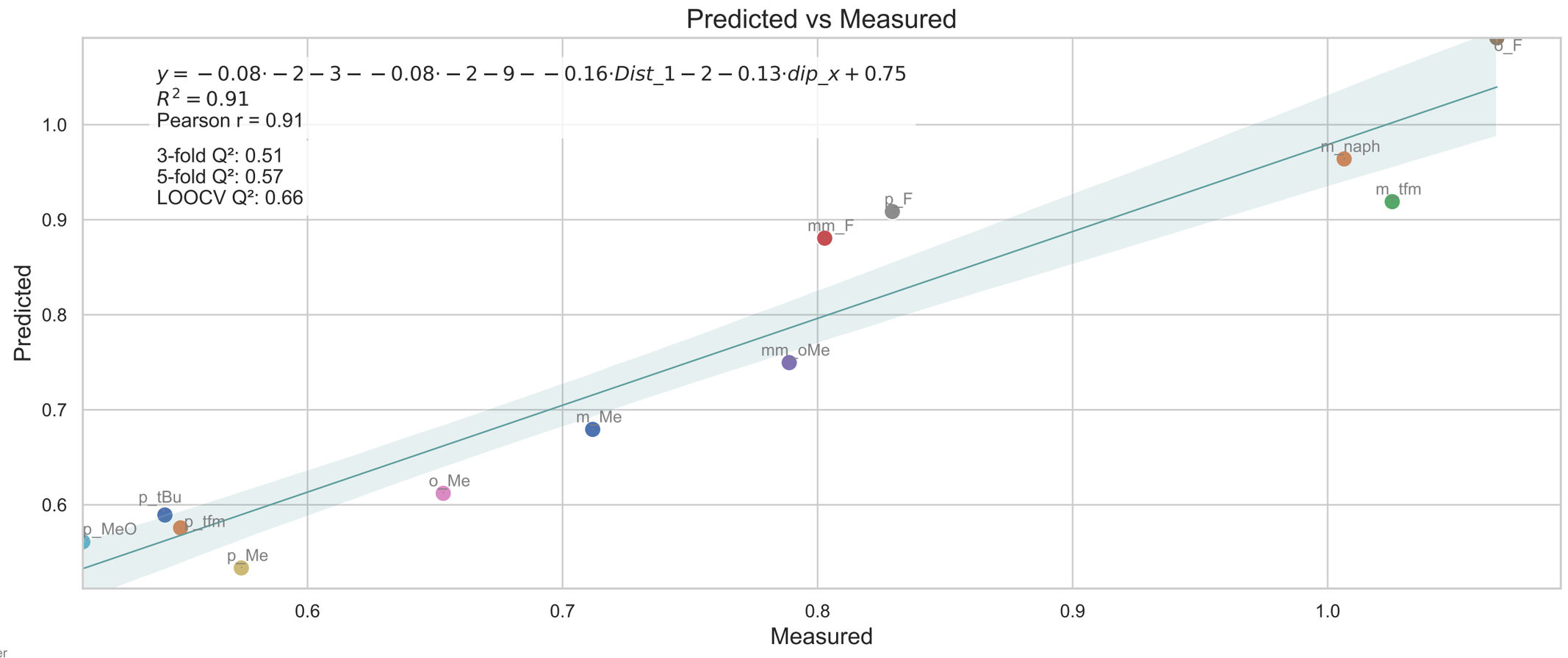
variable	VIF
-2-3-	1.134
-2-9-	1.454
Dist_1-2	1.173
dip_x	1.415

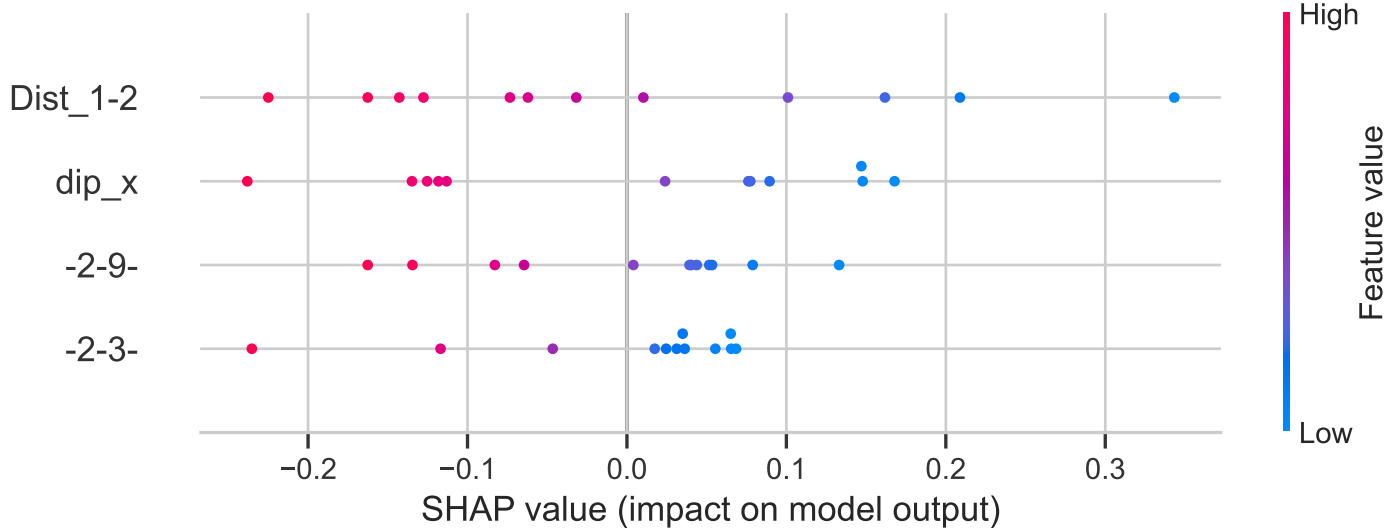
Cross-validation metrics

Q2_3_MAE	Q2_3_RMSD	Q2_5_MAE	Q2_5_RMSD	Q2_LOO	LOO MAE	LOO RMSD
0.511	0.103	0.131	0.57	0.095	0.119	0.657	0.097	0.111

Model Summary

Formula: -2-3- + -2-9- + Dist_1-2 + dip_x
Train R^2: 0.914
Q^2 (3/5/LOO): 0.511 / 0.570 / 0.657


Predicted vs Measured
$y = -0.08 \cdot -2 -3 - -0.08 \cdot -2 -9 - -0.16 \cdot Dist\_1 -2 -0.13 \cdot dip\_x + 0.75$
$R^2 = 0.91$
Pearson r = 0.91
3-fold Q²: 0.51
5-fold Q²: 0.57
LOOCV Q²: 0.66
Predicted
Measured
1.0
0.9
0.8
0.7
0.6
0.6
0.7
0.8
0.9
1.0
o_F
m_naph
m_tfm
p_F
mm_F
mm_oMe
m_Me
o_Me
p_tBu
p_tfm
p_MeO
p_Me


Dist_1-2
dip_x
-2-9-
-2-3-
−0.2
−0.1
0.0
0.1
0.2
0.3
SHAP value (impact on model output)
High
Feature value
Low

Rank #2 | R²=0.904 | Features=4 | Samples=12
-1-2- + Dist_1-2 + Dist_16-17 + dip_x

Coefficients

Estimate	Std. Error	t value	p value
0.7397	0.02247	32.92	6.177e-09
-0.09158	0.02573	-3.559	0.009232
-0.1117	0.02835	-3.941	0.005598
0.1028	0.02764	3.719	0.00747
-0.09002	0.0234	-3.847	0.006318

VIF

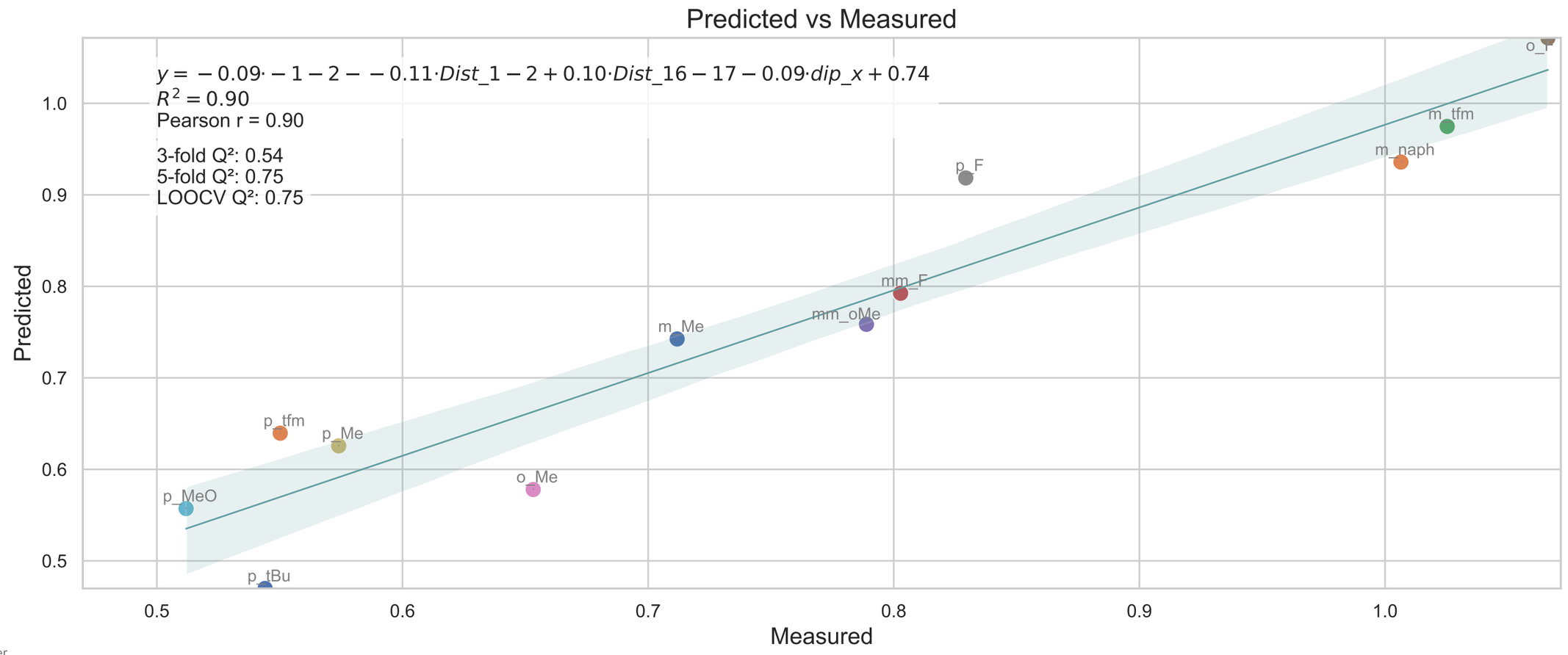
variable	VIF
-1-2-	1.479
Dist_1-2	1.752
Dist_16-17	1.547
dip_x	1.124

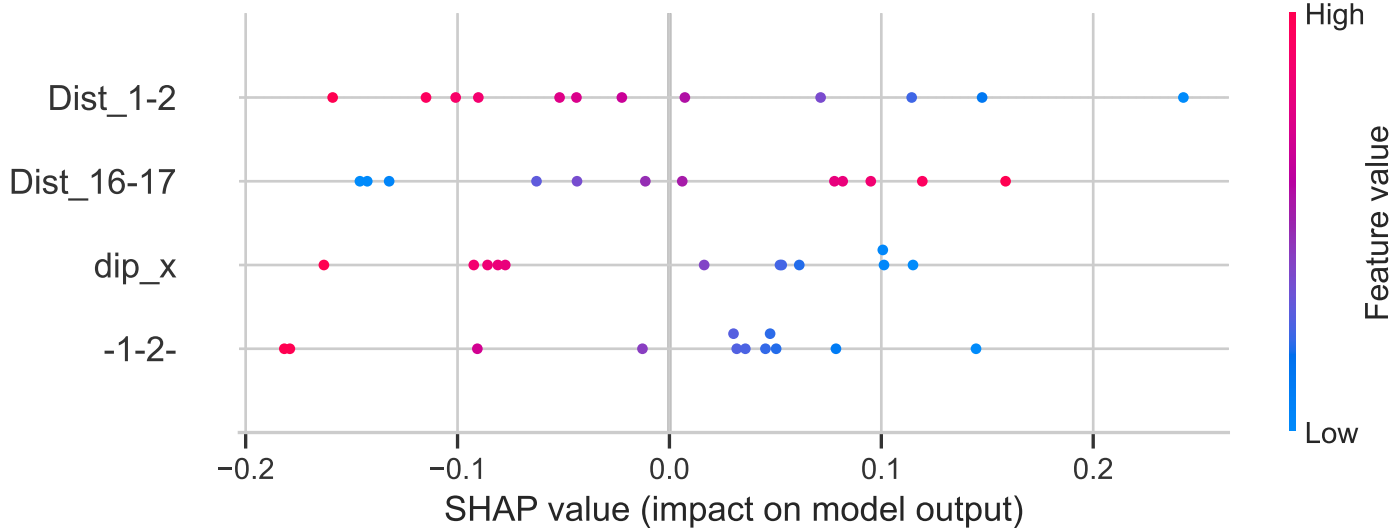
Cross-validation metrics

Q2_3_MAE	Q2_3_RMSD	Q2_5_MAE	Q2_5_RMSD	Q2_LOO_MAE	Q2_LOO_RMSD (LOO)
0.54	0.102	0.124	0.746	0.084	0.095	0.747	0.086	0.096

Model Summary

Formula: -1-2- + Dist_1-2 + Dist_16-17 + dip_x
Train R²: 0.904
Q² (3/5/LOO): 0.540 / 0.746 / 0.747


Predicted vs Measured
$y = -0.09 \cdot -1 - 2 - -0.11 \cdot Dist\_1 - 2 + 0.10 \cdot Dist\_16 - 17 - 0.09 \cdot dip\_x + 0.74$
$R^2 = 0.90$
Pearson r = 0.90
3-fold Q²: 0.54
5-fold Q²: 0.75
LOOCV Q²: 0.75
Predicted
Measured
1.0
0.9
0.8
0.7
0.6
0.5
0.5
0.6
0.7
0.8
0.9
1.0
o_F
m_tfm
m_naph
p_F
mm_F
mm_oMe
m_Me
p_tfm
p_Me
o_Me
p_MeO
p_tBu


Dist_1-2
Dist_16-17
dip_x
-1-2-
−0.2
−0.1
0.0
0.1
0.2
SHAP value (impact on model output)
High
Feature value
Low

Rank #5 | R^2=0.894 | Features=4 | Samples=12
para + -18-20- + Dist_16-17 + L

Coefficients

Estimate	Std. Error	t value	p value
0.7248	0.02441	29.69	1.266e-08
0.2215	0.04443	4.986	0.001591
-0.09812	0.03416	-2.872	0.02392
0.1826	0.02856	6.394	0.0003693
-0.09175	0.03269	-2.807	0.02627

VIF

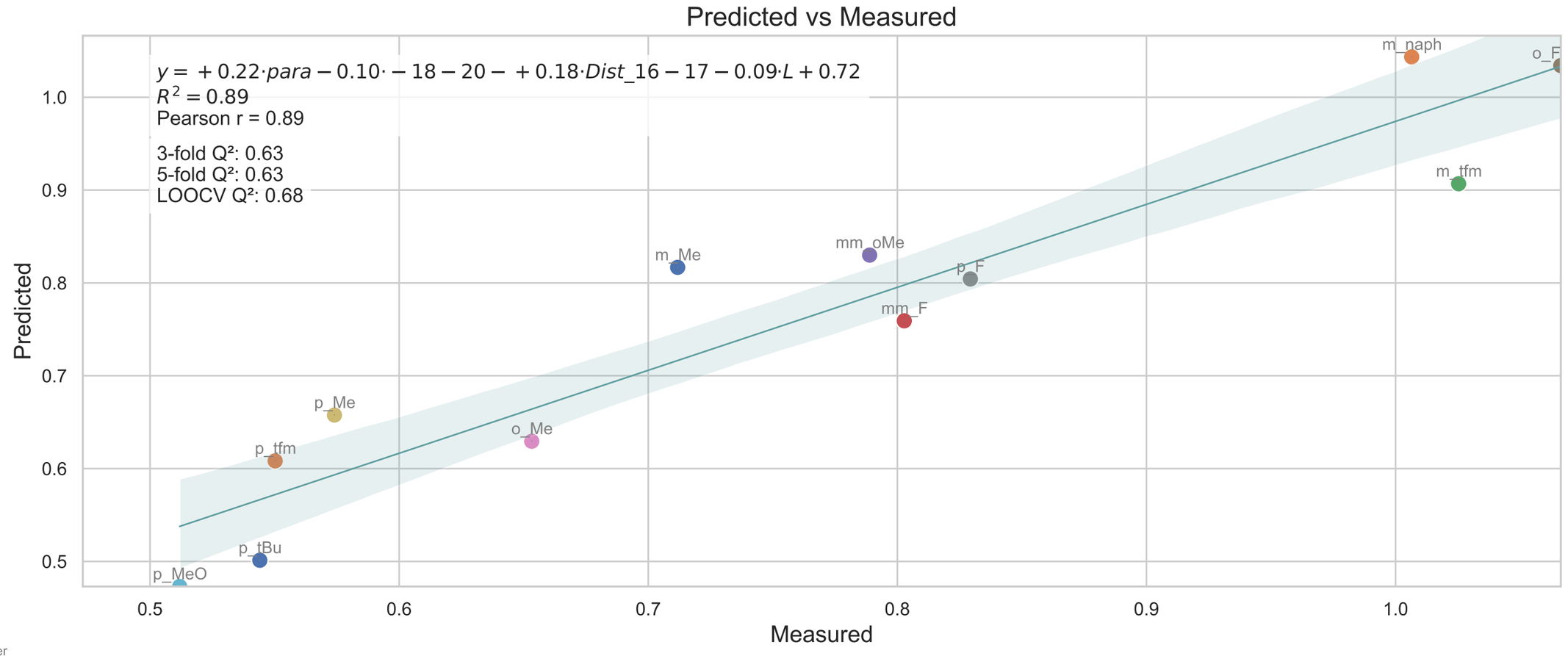
variable	VIF
para	3.369
-18-20-	1.881
Dist_16-17	1.486
L	2.093

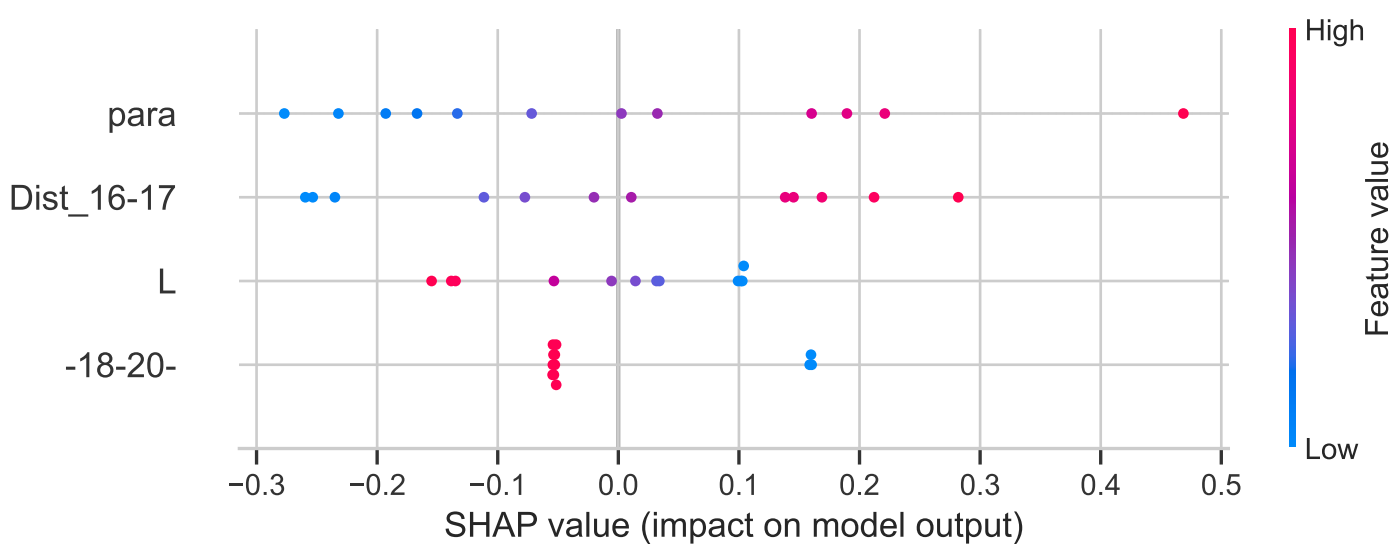
Cross-validation metrics

Q2_3_MAE	R(MSD	Q2_5_MAE	R(MSD	Q2_LOO	MAE	RMSD (LOO)
0.629	0.103	0.63	0.101	0.682	0.097	0.107

Model Summary

Formula: para + -18-20- + Dist_16-17 + L
Train R^2: 0.894
Q^2 (3/5/LOO): 0.629 / 0.630 / 0.682


Predicted vs Measured
$y = +0.22 \cdot para - 0.10 \cdot -18 - 20 - +0.18 \cdot Dist\_16 - 17 - 0.09 \cdot L + 0.72$
$R^2 = 0.89$
Pearson r = 0.89
3-fold Q²: 0.63
5-fold Q²: 0.63
LOOCV Q²: 0.68
m_naph
o_F
m_tfm
mm_oMe
m_Me
p_F
mm_F
p_Me
o_Me
p_tfm
p_tBu
p_MeO
Predicted
Measured
0.5
0.6
0.7
0.8
0.9
1.0



Ranked features SHAP