Top 5 Models (ranked by R²)

Overview

Rank	R² (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	0.894	4	12	para + -18-20- + Dist_16-17 + L
4	0.889	4	12	-2-3- + -1-10- + Dist_1-2 + dip_x
5	0.884	4	12	para + -18-20- + Dist_16-17 + dip_x

Rank #1 | R²=0.914 | Features=4 | Samples=12
-2-3- + -2-9- + Dist_1-2 + dip_x

Coefficients

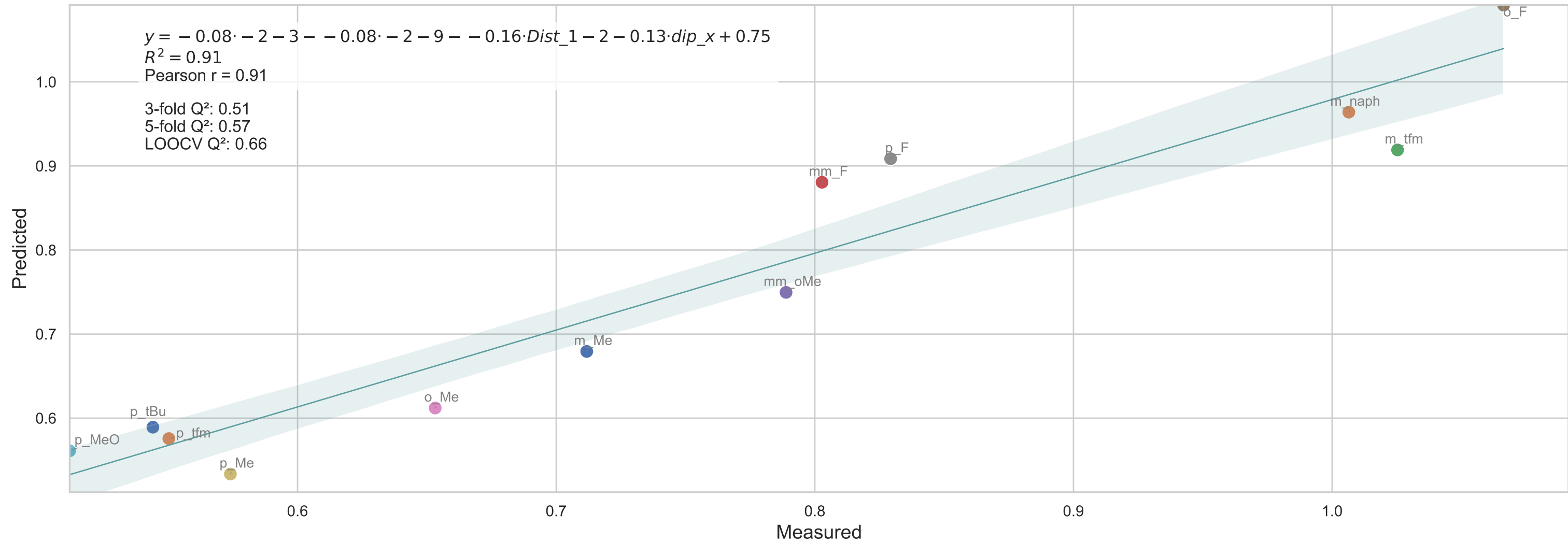| Estimate	Std. Error	t value	p value
0.7459132404050799	0.021357285465540832	34.9254703557052	4.093052341502812e-09
-0.08010638784241192	0.02049810733629733	-3.9079894805976716	0.005838517312024338
-0.08343316793769026	0.024406348230271616	-3.4185027252133797	0.011158089362995227
-0.1581258589048966	0.0219773243266064	-7.194821367008234	0.0001782569125097222
-0.13144529782747694	0.02488652150130307	-5.281786682023617	0.0011458235083039447

VIF

variable	VIF
-2-3-	1.134081095321145
-2-9-	1.4539447166855537
Dist_1-2	1.1725939964567342
dip_x	1.4151101869661404

Cross-validation metrics

Q2_3_F	MAE	RMSD	Q2_5_F	MAE	RMSD	Q2_LOO	MAE	RMSD (LOO)
0.5106	0.1033	0.1309	0.57	0.0951	0.119	0.6566	0.0969	0.1113

Model Summary

Formula: -2-3- + -2-9- + Dist_1-2 + dip_x
Train R²: 0.914
Q² (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 Q^2: 0.51

5-fold Q^2: 0.57

LOOCV Q^2: 0.66

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

Coefficients

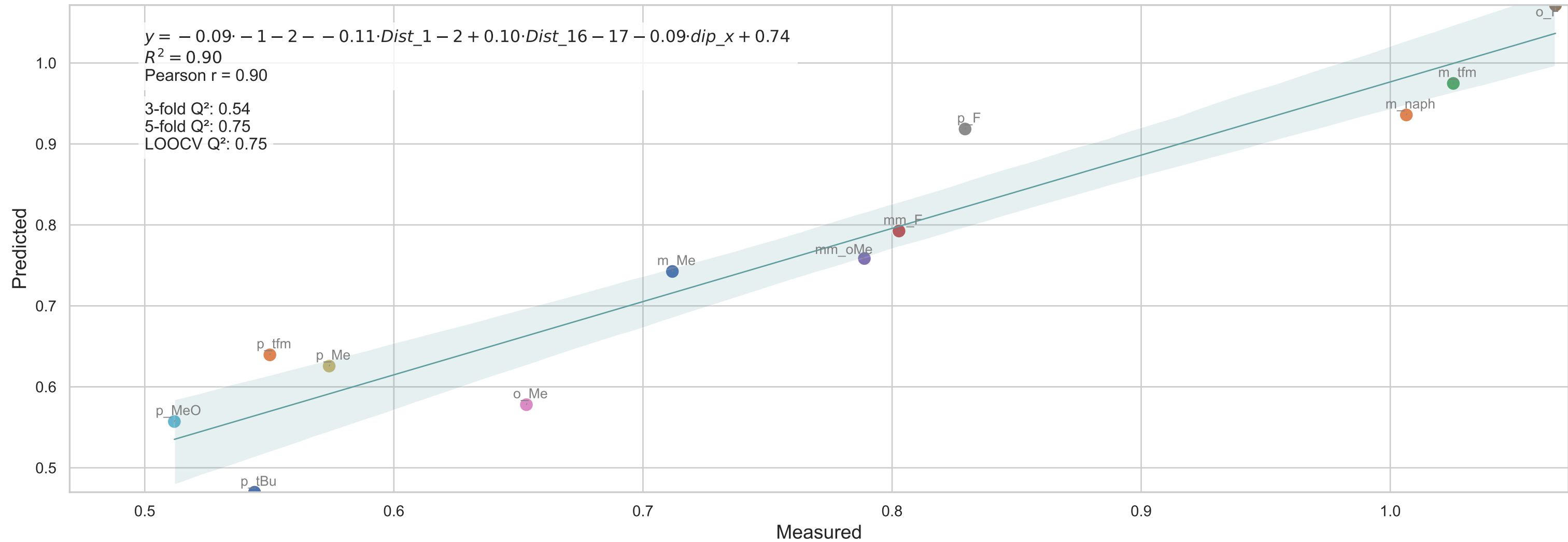| Estimate	Std. Error	t value	p value
0.7397193039115785	0.02246943465650091	32.921135543459755	6.176641420552187e-09
-0.09157604442283115	0.025730994280836916	-3.5589780722555346	0.009232059630158718
-0.11172462939367887	0.028351217476326896	-3.9407348021991755	0.005598068190179495
0.10277341168527394	0.027637759592761862	3.718586933225571	0.007469944943916174
-0.0900245614467871	0.023402089972242048	-3.846859898135938	0.006318097195845418

VIF

variable	VIF
-1-2-	1.4792247451910958
Dist_1-2	1.7519844169925736
Dist_16-17	1.547144657483777
dip_x	1.1235077202530293

Cross-validation metrics

Q2_3_F	MAE	RMSD	Q2_5_F	MAE	RMSD	Q2_LOO	MAE	RMSD (LOO)
0.539	0.102	0.124	0.746	0.083	0.095	0.747	0.085	0.0955

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

Rank #3 | R²=0.894 | Features=4 | Samples=12
para + -18-20- + Dist_16-17 + L

Coefficients

Estimate	Std. Error	t value	p value
0.7247616160944855	0.024408315370127723	29.693225652987497	1.2661580894857138e-08
0.22150852111574018	0.04442859176882552	4.985720057667176	0.0015906808763765934
-0.09811676515921307	0.03416095209281431	-2.872190590374433	0.023916995604722047
0.18260144803481265	0.028557473978240103	6.394173664447675	0.0003693240945257248
-0.09175488728997185	0.03269214257018799	-2.806634257543073	0.026272623904523007

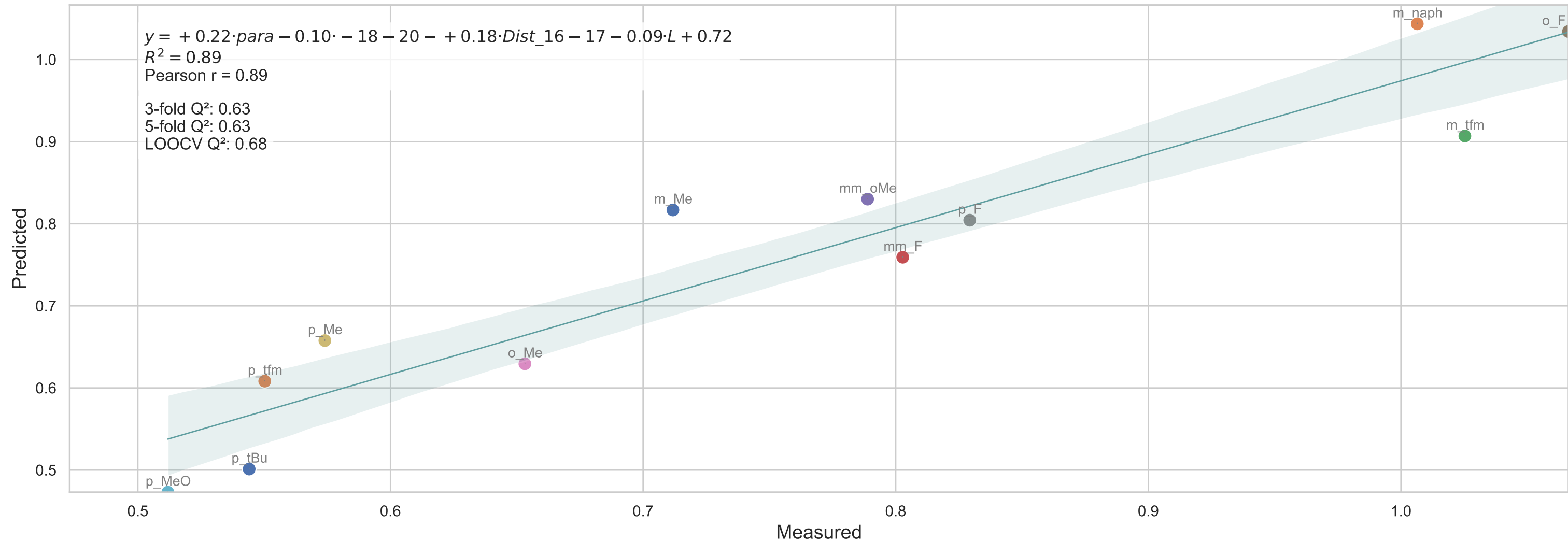## VIF

Cross-validation metrics

Model Summary

Formula: para + -18-20- + Dist_16-17 + L
Train R²: 0.894
Q² (3/5/LOO): 0.629 / 0.630 / 0.682

variable	VIF
para	3.3686779150934734
-18-20-	1.8810277986076578
Dist_16-17	1.4864233399461537
L	2.092523381411499

Q2_3_F	MAE	RMSD	Q2_5_F	MAE	RMSD	Q2_LOO	MAE	RMSD (LOO)
0.6294	0.1029	0.1149	0.6303	0.1013	0.1148	0.6819	0.0967	0.1071

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

Coefficients

Estimate	Std. Error	t value	p value
0.7361918962708925	0.024267006925775866	30.33715276558998	1.0907192926623566e-08
-0.09183336865256804	0.0237633944900076	-3.864488665160111	0.006175532521599081
0.08040410838017992	0.029557733141243724	2.7202393362157777	0.029755306986628982
-0.12330290487862906	0.023262018983041775	-5.300610620622312	0.0011226238551600431
-0.14232634776002306	0.031091746002667352	-4.57762480588298	0.002550935472346527

VIF

variable	VIF
-2-3-	1.1745370538839237
-1-10-	1.518718117423049
Dist_1-2	1.0123015368899815
dip_x	1.702100762697295

Cross-validation metrics

Q2_3_F	MAE	RMSD	Q2_5_F	MAE	RMSD	Q2_LOO	MAE	RMSD (LOO)
0.6005	0.1043	0.1199	0.4215	0.1187	0.1376	0.5982	0.106	0.1204

Model Summary

Formula: -2-3- + -1-10- + Dist_1-2 + dip_x
Train R²: 0.889
Q² (3/5/LOO): 0.600 / 0.421 / 0.598

Predicted vs Measured

$y = -0.09 \cdot -2 - 3 - + 0.08 \cdot -1 - 10 - -0.12 \cdot Dist\_1 - 2 - 0.14 \cdot dip\_x + 0.74$

$R^2 = 0.89$

Pearson r = 0.89

3-fold Q²: 0.60
5-fold Q²: 0.42
LOOCV Q²: 0.60

Rank #5 | R²=0.884 | Features=4 | Samples=12
para + -18-20- + Dist_16-17 + dip_x
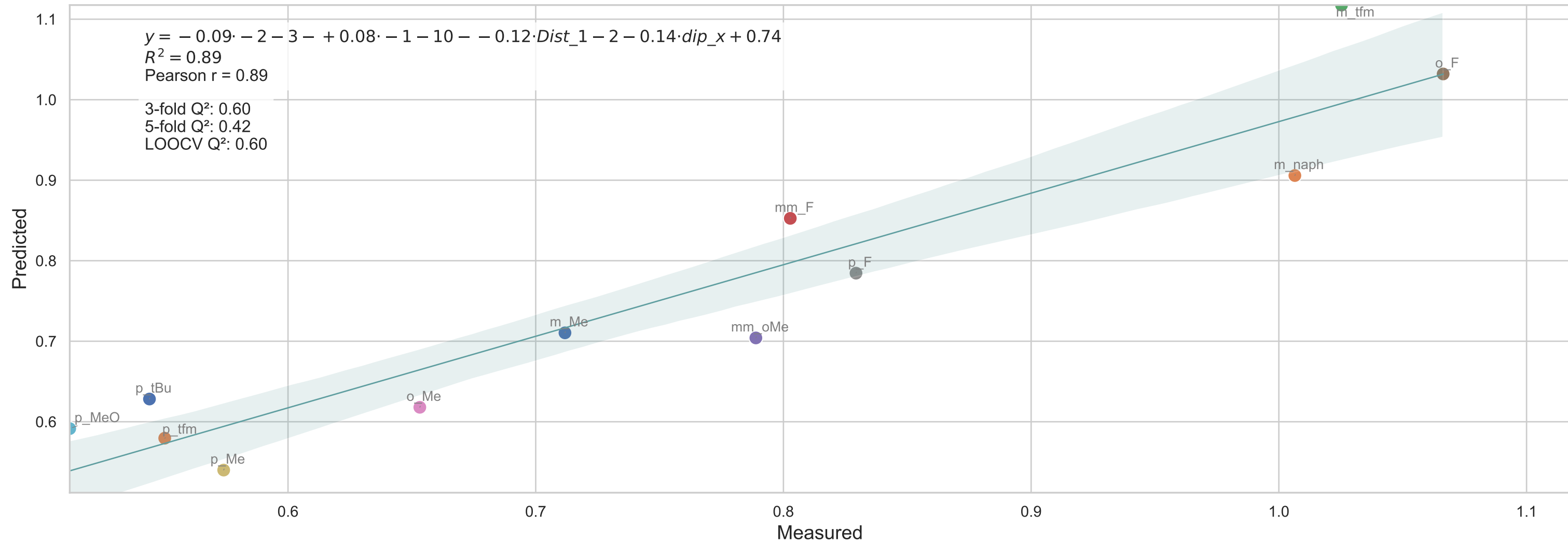
Coefficients

Estimate	Std. Error	t value	p value
0.7274312153498056	0.025614323196283706	28.39939239368019	1.7255781470382203e-08
0.13553896601341767	0.039347691479679696	3.444648489312619	0.01076911875634301
-0.14030862172717193	0.0418591716168093	-3.351920649830263	0.012218604804743238
0.18515172426649532	0.029762560191344448	6.220960934682668	0.0004362675296871732
-0.089645218884703	0.034976414068091766	-2.563019145135419	0.037386583911884985

VIF

variable	VIF
para	2.4099101185631384
-18-20-	2.575991802774773
Dist_16-17	1.4725547350956167
dip_x	2.0597900491063457

Cross-validation metrics

Q2_3_F	MAE	RMSD	Q2_5_F	MAE	RMSD	Q2_LOO	MAE	RMSD (LOO)
0.3845	0.118	0.141	0.5367	0.1128	0.1267	0.6444	0.1079	0.1132

Model Summary

Formula: para + -18-20- + Dist_16-17 + dip_x
Train R²: 0.884
Q² (3/5/LOO): 0.385 / 0.537 / 0.644

Predicted vs Measured

$y = +0.14 \cdot para - 0.14 \cdot -18 - 20 - +0.19 \cdot Dist\_16 - 17 - 0.09 \cdot dip\_x + 0.73$

$R^2 = 0.88$

Pearson r = 0.88

3-fold Q^2: 0.38
5-fold Q^2: 0.54
LOOCV Q^2: 0.64