Supplementery Information

September 12, 2014

Table 1: Comparison to existing methods.

Model	PKKB	Martel	Average
smlogp	0.859	1.222	1.040
spluslogp	0.938	1.392	1.165
p.crippen	1.196	1.172	1.184
alogps	0.863	1.546	1.205
alogp	1.292	1.281	1.287
mlogp	0.794	2.053	1.423
xlogp3	1.741	1.177	1.459
p.xlogp	1.569	1.393	1.481
milogp	2.479	1.176	1.828
am	2.511	1.187	1.849
p.mlogp	2.028	1.704	1.866
clogp	2.241	1.560	1.900
p.alogp	3.566	4.338	3.952

Table 2: Descriptors used in model training.

nAcid	nAromBond	nH	nC
nN	nS	nF	nCl
nX	ATSc1	ATSc2	ATSc3
ATSc4	ATSc5	ATSm1	ATSm3
ATSm4	ATSm5	nBase	BCUTw.1l
BCUTw.1h	BCUTc.1l	BCUTc.1h	BCUTp.1l
BCUTp.1h	nBondsD	nBondsD2	nBondsM
bpol	C1SP2	C2SP2	C3SP2
C1SP3	C2SP3	C3SP3	C4SP3
SCH.5	SCH.6	SCH.7	VCH.3
VCH.4	VCH.5	VCH.6	VCH.7
SC.3	SC.4	SC.5	SC.6
VC.3	VC.4	VC.5	VC.6
SPC.6	VPC.6	SP.7	VP.7
CrippenLogP	nHBa	nwHBa	nHBint2
nHBint3	nHBint4	nHBint5	nHBint6
nHBint7	nHBint8	nHBint9	nHBint10
nHsOH	nHdNH	nHaaNH	nHCHnX
nHCsats	nHCsatu	nHAvin	nssCH2
nsssCH	ntsC	ndssC	naasC
naaaC	nssssC	ndsN	naaN
nsssN	naasN	ndO	nssO
naaO	nssS	nddssS	SHBd
SHBa	SwHBa	SHBint2	SHBint3
SHBint4	SHBint5	SHBint6	SHBint7
SHBint8	SHBint9	SHsOH	SHssNH
SHdsCH	SHaaCH	SHCsats	SsCH3
SssCH2	SdsCH	SsssCH	SdssC
SaasC	SssssC	SsNH2	SssNH
SsssN	minHBd	minHBa	minwHBa
minHBint2	minHBint3	minHBint4	minHBint5
minHBint6	minHBint7	minHBint8	minHBint9
minHsOH	minHssNH	minHdsCH	minHaaCH
minHCsats	minHCsatu	minHother	minsCH3
minssCH2	mindsCH	minaaCH	minsssCH
mindssC	minaasC	minssssC	minssNH
mindsN	minaaN	minsssN	minsOH
mindO	minssO	minsCl	maxHBa
maxwHBa	maxHBint2	maxHBint3	maxHBint4
maxHBint5	maxHBint6	maxHBint7	maxHBint8
maxHBint9	maxHCsats	maxssCH2	maxsssCH
maxdssC	maxaasC	sumI	hmax
gmax	hmin	gmin	ETA_AlphaP
ETA_dAlpha_B	ETA_Epsilon_1	ETA_Epsilon_2	ETA_Epsilon_4
ETA_Epsilon_5	ETA_dEpsilon_B	ETA_dEpsilon_D	ETA_Psi_1
ETA_dPsi_A	ETA_Shape_P	ETA_Shape_Y	ETA_Shape_X
ETA_BetaP	ETA_BetaP_s	ETA_dBeta	ETA_Beta_ns_d
ETA_BetaP_ns_d	ETA_EtaP	ETA_EtaP_F	ETA_Eta_F_L
ETA_EtaP_F_L	ETA_Eta_B	ETA_EtaP_B	ETA_Eta_B_RC
ETA_EtaP_B_RC	FMF	fragC	nHBAcc
nHBAcc3	nHBAcc_Lipinski	nHBDon_Lipinski	HybRatio
Kier2	Kier3	nAtomLC	nAtomP
nAtomLAC	MDEC.11	MDEC.12	MDEC.13
MDEC.22	MDEC.23	MDEC.33	MDEO.11
MDEO.12	MDEN.22	MDEN.23	MLFER_A
MLFER_BH	MLFER_S	MLFER_E	MLFER_L
PetitjeanNumber	nRing	nFRing	nF9Ring
nF10Ring	nFG12Ring	nTRing	nT5Ring
nT6Ring	nRotB	LipinskiFailures	TopoPSA
VAdjMat	WTPT.2	WTPT.3	WTPT.4
WPATH			

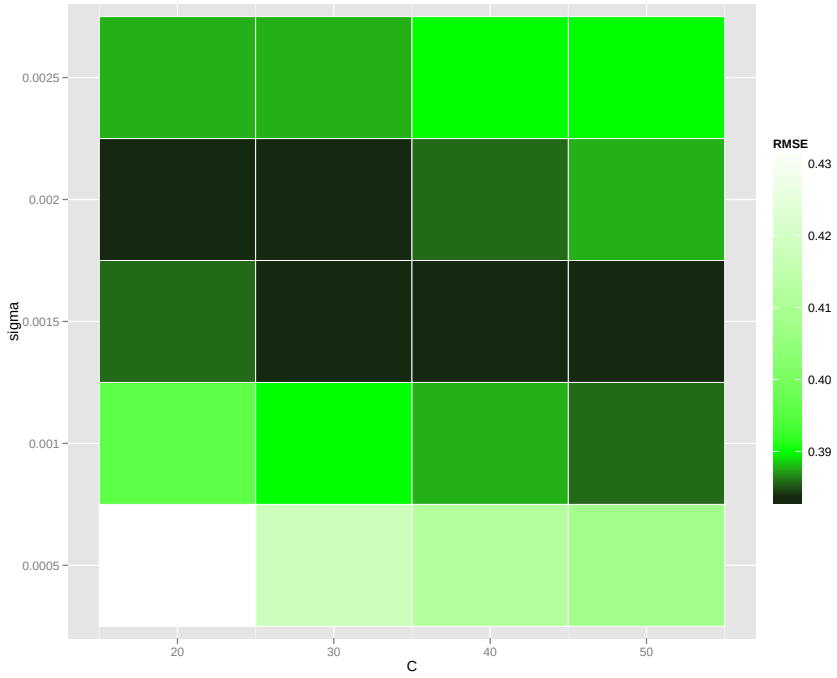

Figure 1: Grid search over a narrow parameter space. σ and Cost hyperparameters are optimised by comparing the model's mean RMSE over the 5 validation sets.

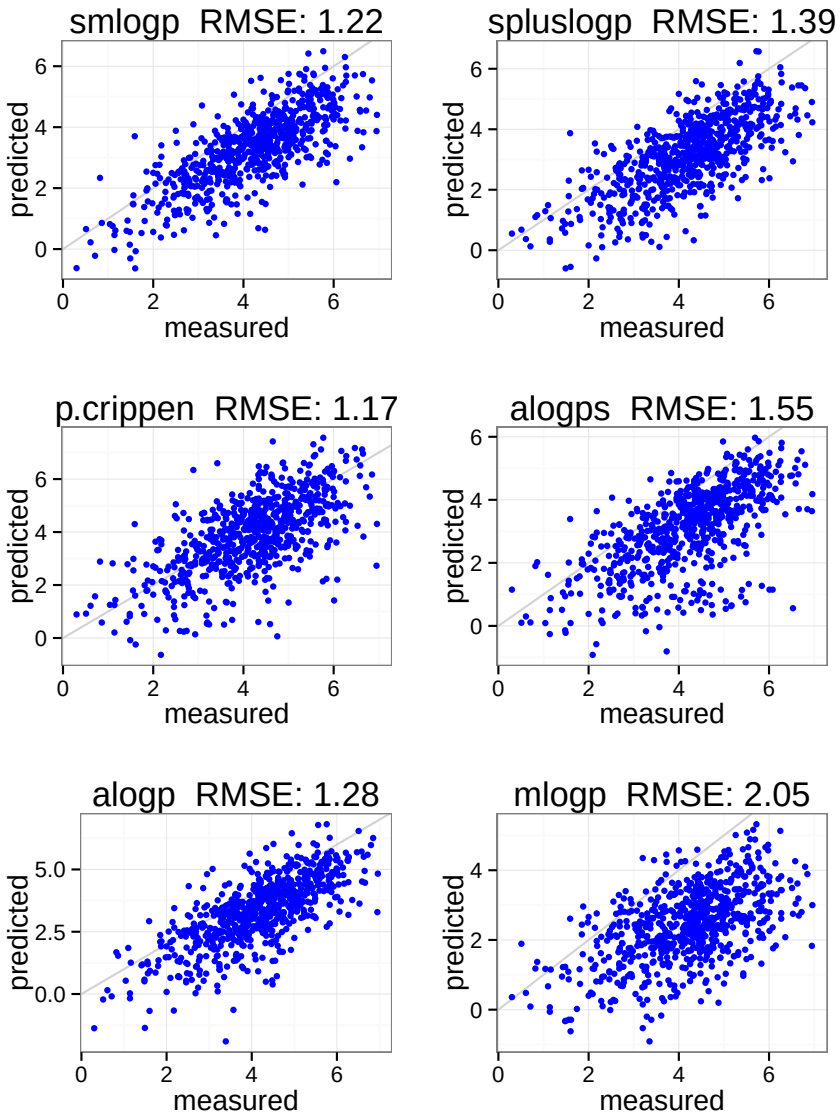

Figure 2: Correlation plots of the top (averaged over both external sets) 6 methods on the Martel set.

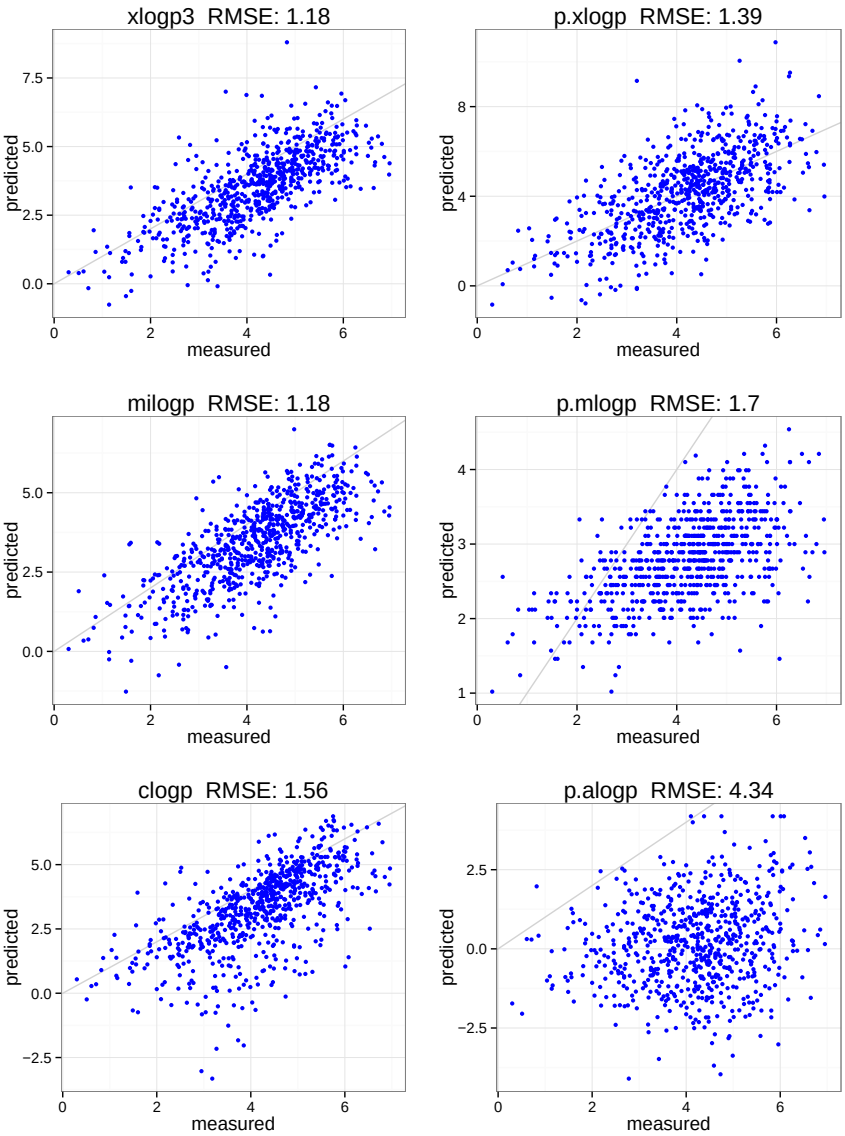

Figure 3: Correlation plots of the bottom (averaged over both external sets) 6 methods on the Martel set.

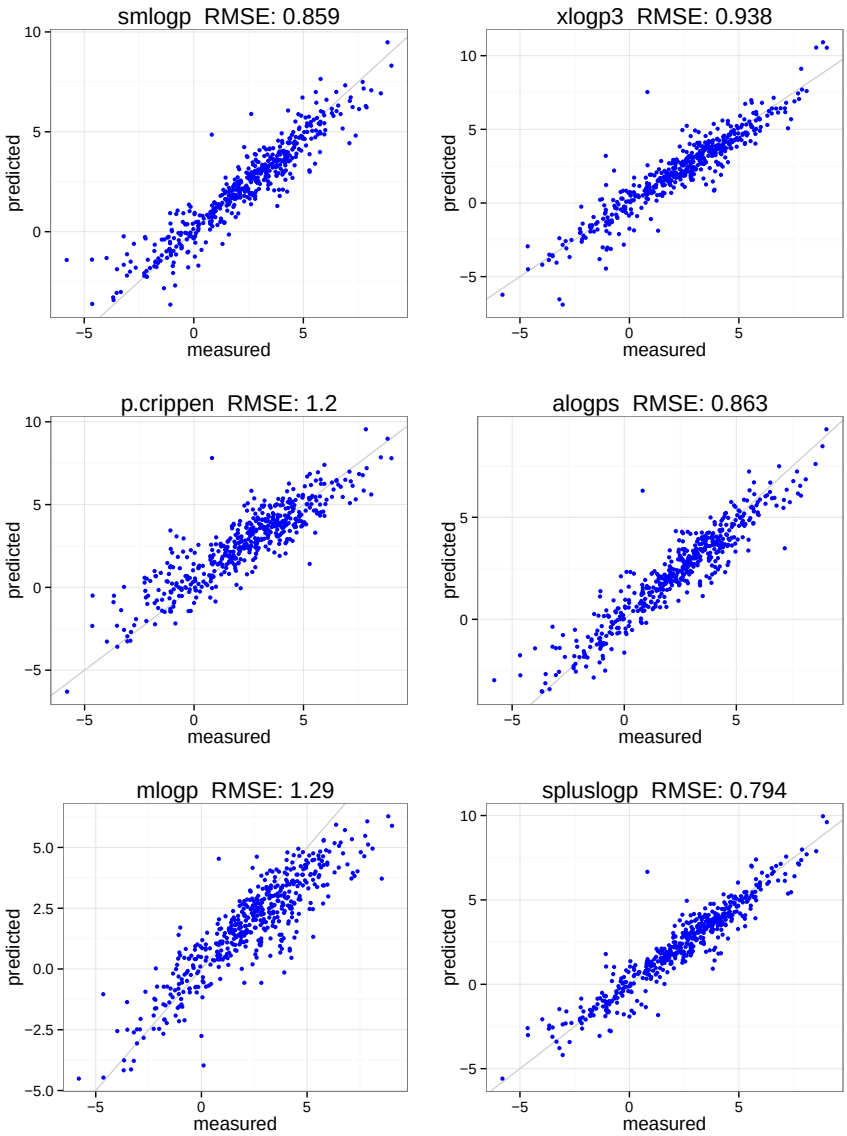

Figure 4: Correlation plots of the top (averaged over both external sets) 6 methods on the PKKB set.

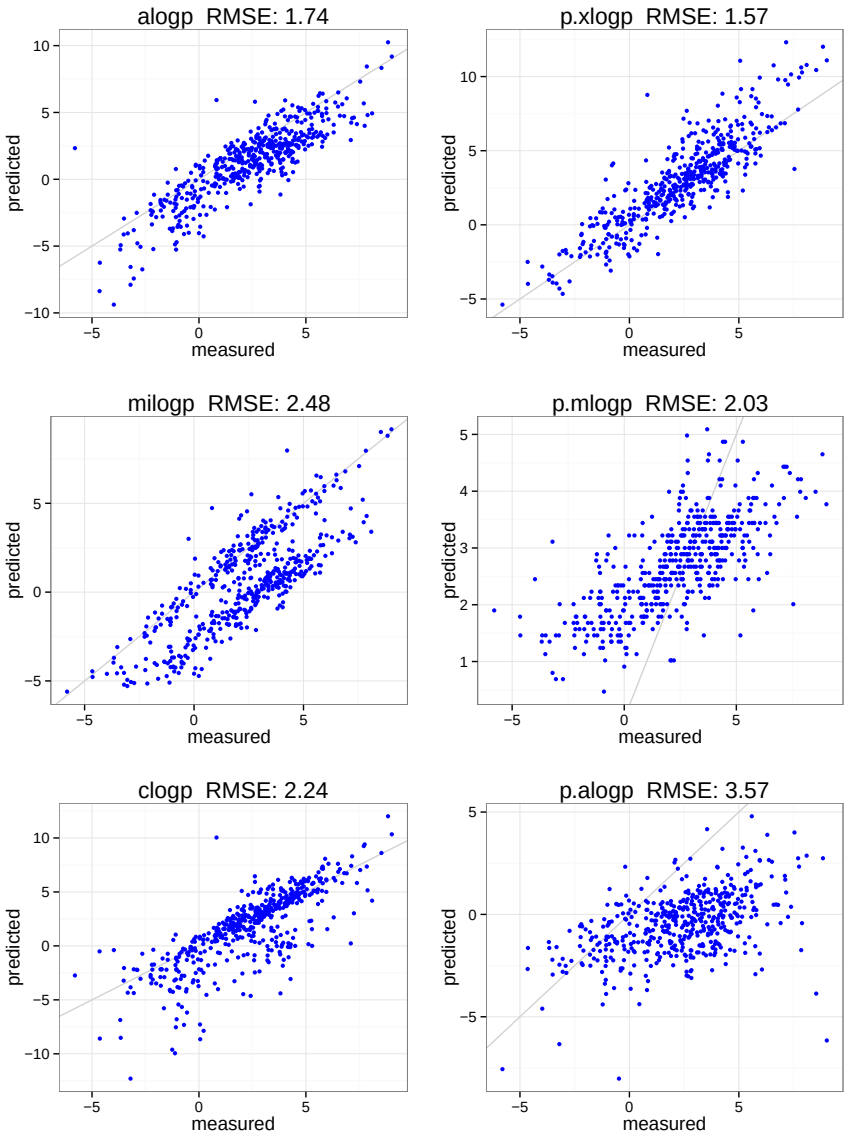

Figure 5: Correlation plots of the bottom (averaged over both external sets) 6 methods on the PKKB set.