

# Ensembles for predicting structured outputs

## Saturation curves for regression and HMLC

Dragi Kocev, Celine Vens, Jan Struyf, Sašo Džeroski

July 19, 2010

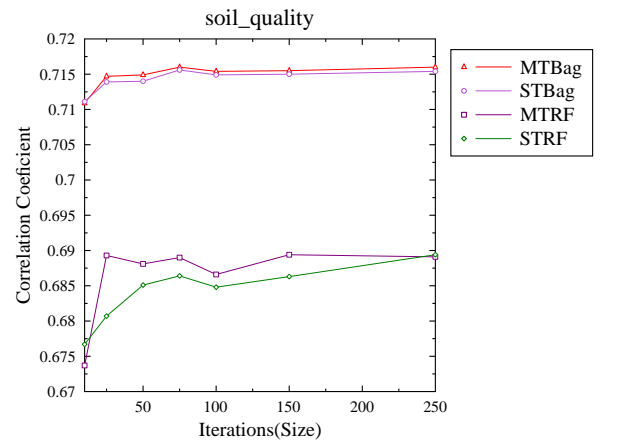
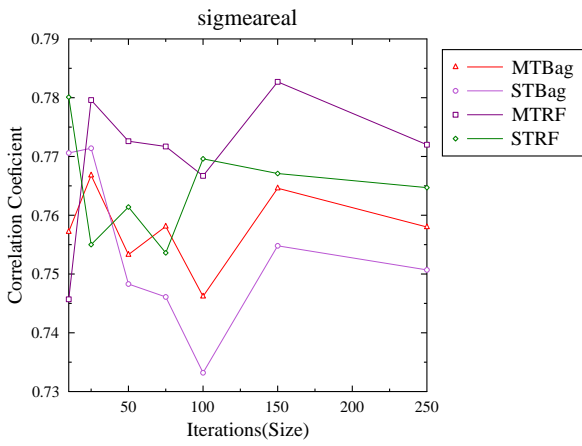
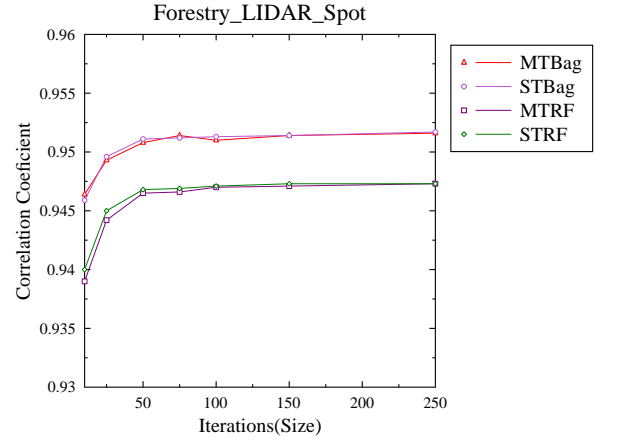
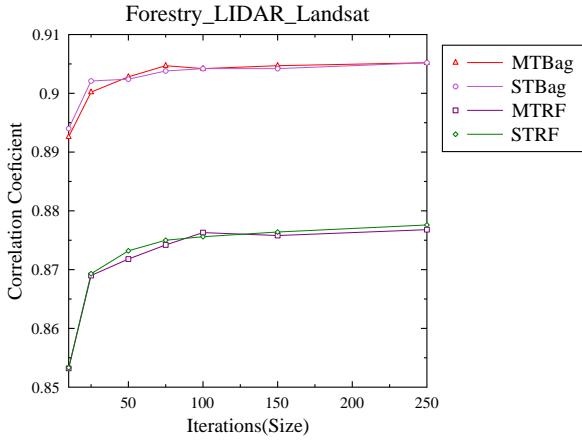
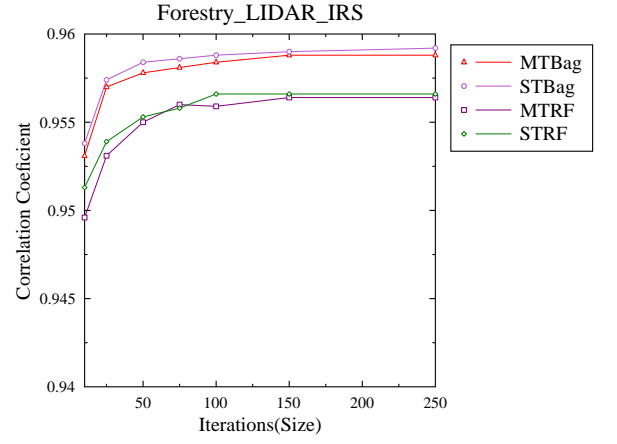
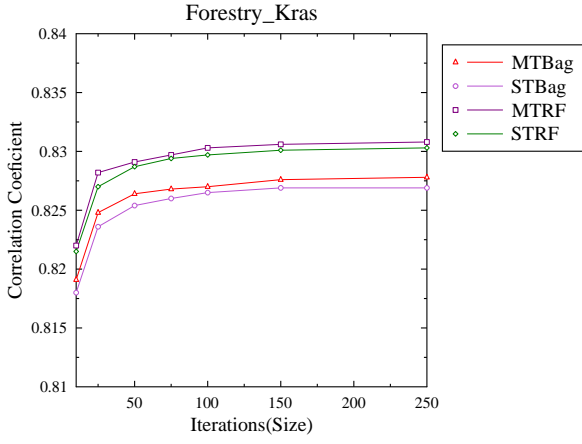
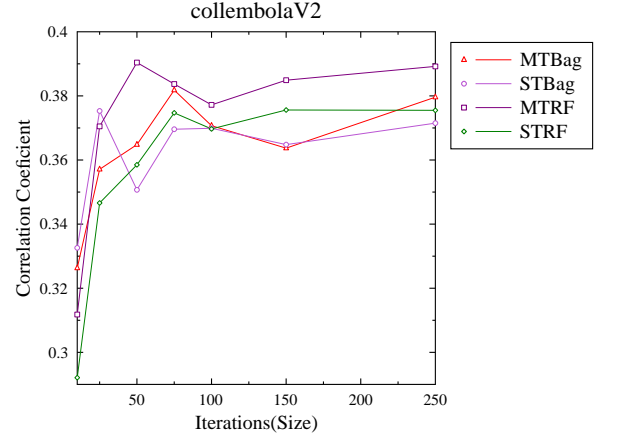
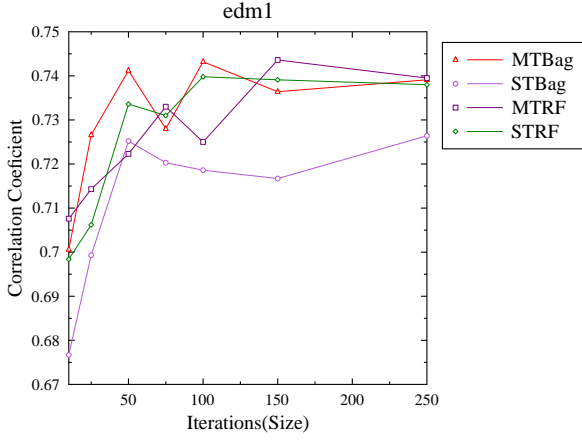
Table 1: Properties of the datasets with multiple targets;  $N$  is number of instances,  $D/C$  number of descriptive attributes (discrete/continous), and  $T$  number of target attributes.

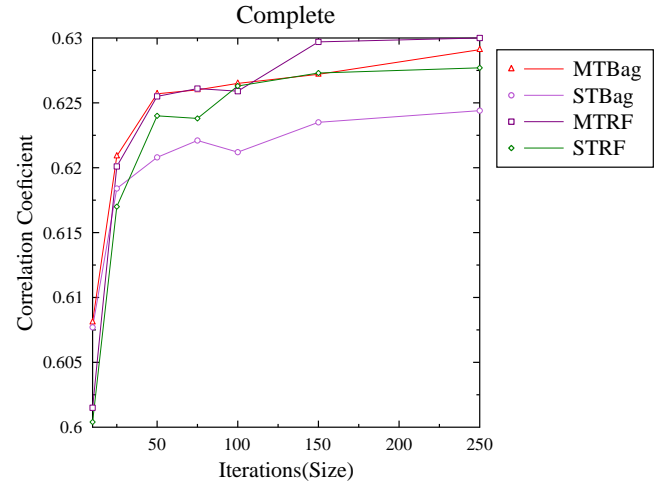
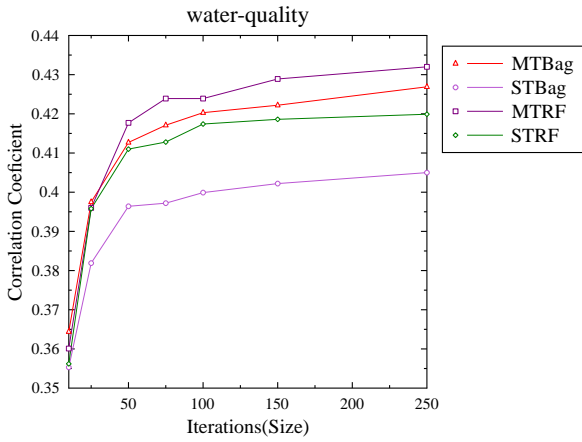
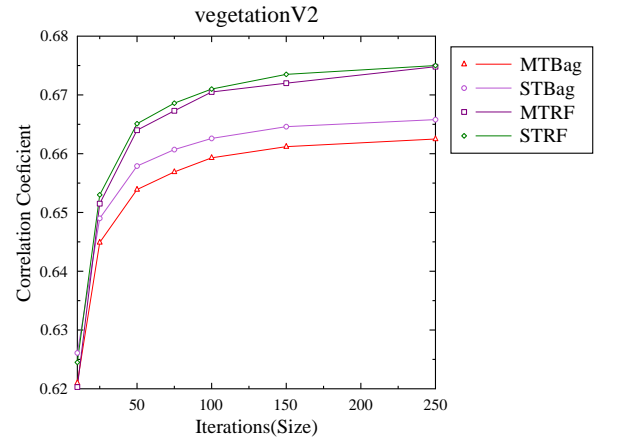
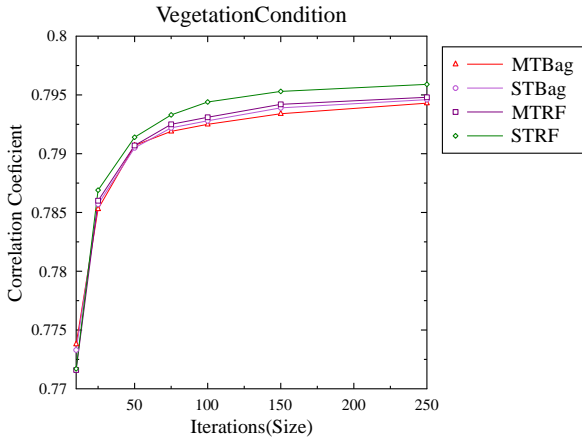
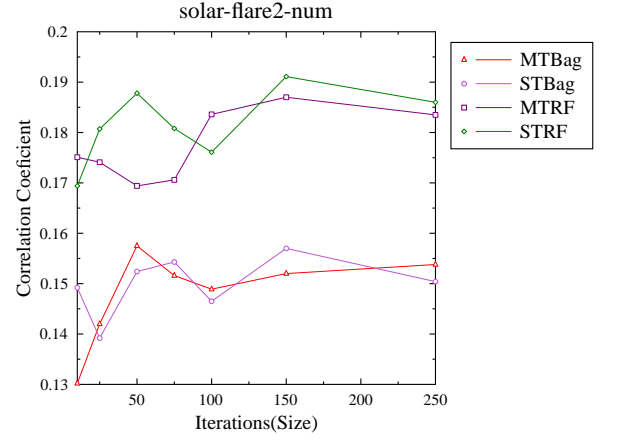
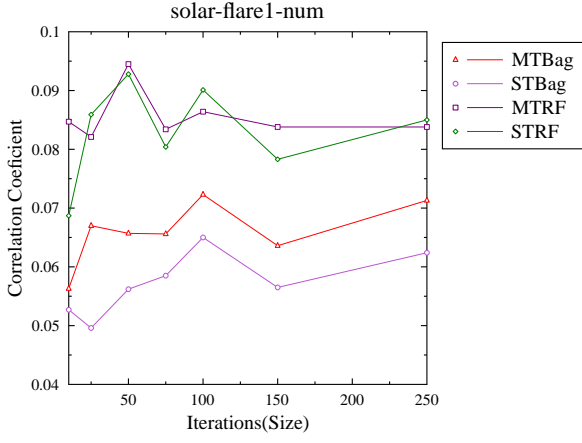
	$N$	$D/C$	$T$
Collembola	393	8/39	3
EDM - 1	154	0/16	2
Forestry-Kras	60607	0/160	11
Forestry-Slivnica-LandSat	6218	0/150	2
Forestry-Slivnica-IRS	2731	0/29	2
Forestry-Slivnica-SPOT	2731	0/49	2
Sigma real	817	0/4	2
Soil quality 1	1944	0/142	3
Solar-flare 1	323	10/0	3
Solar-flare 2	1066	10/0	3
Vegetation Clustering	29679	0/65	11
Vegetation Condition	16967	1/39	7
Water quality	1060	0/16	14

Table 2: Properties of the datasets with hierarchically organized multiple labels;  $N_{train}$  is number of instances in the training dataset,  $N_{test}$  is number of instances in the testing dataset,  $D/C$  is number of descriptive attributes (discrete/continous),  $HierarchySize$  is number of classes in the hierarchy, and  $\overline{Label}$  is average number of labels per example.

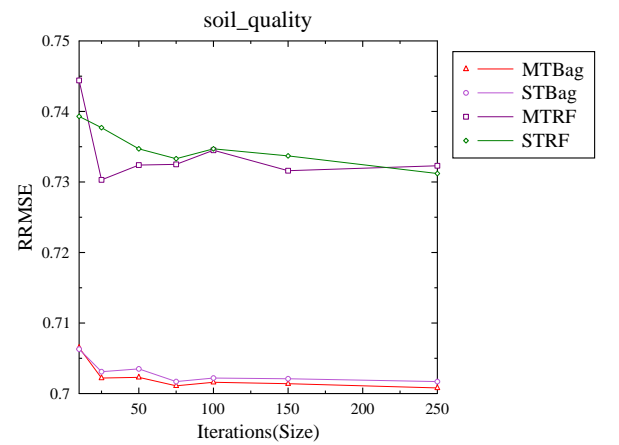
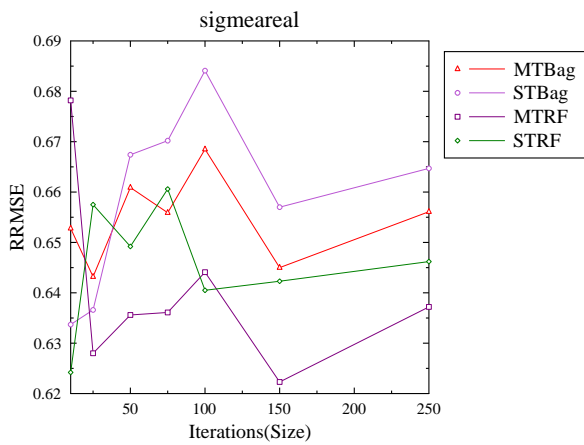
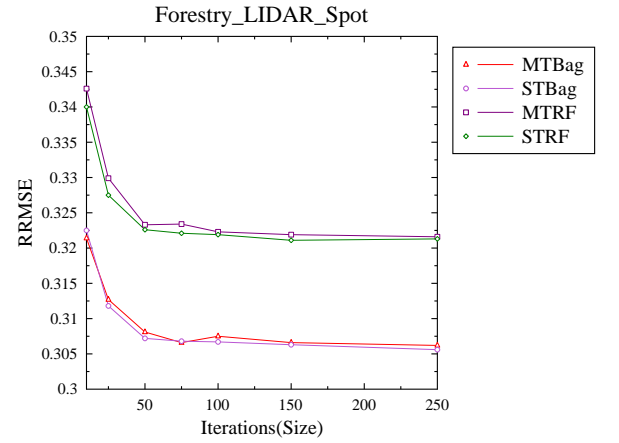
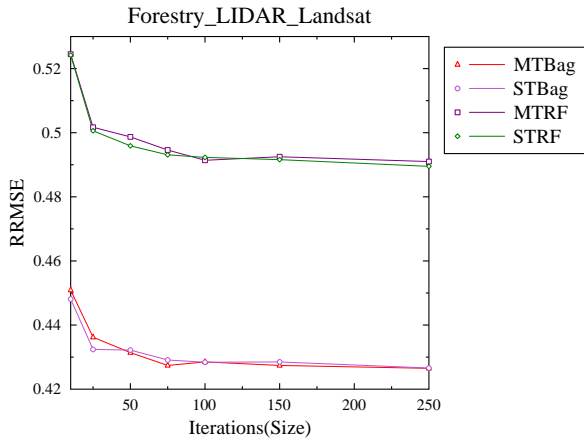
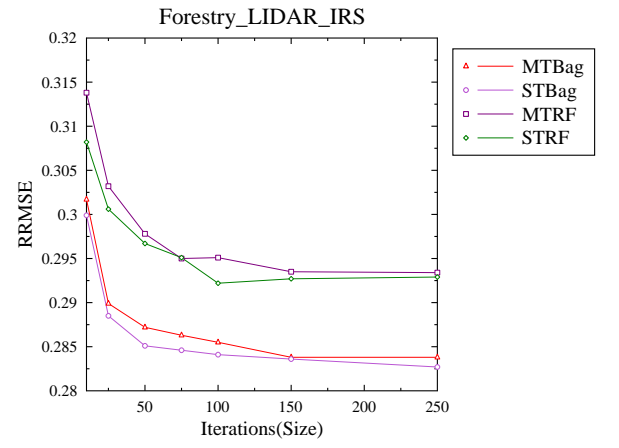
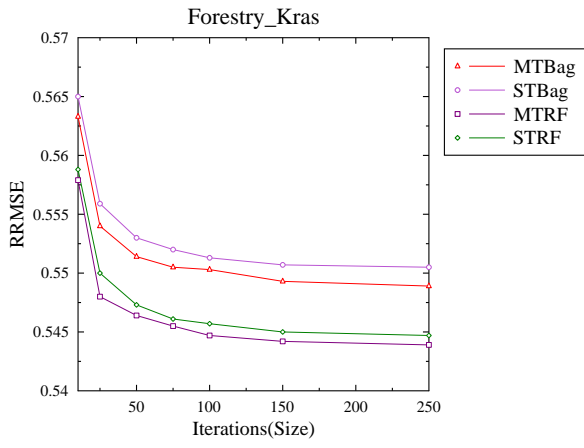
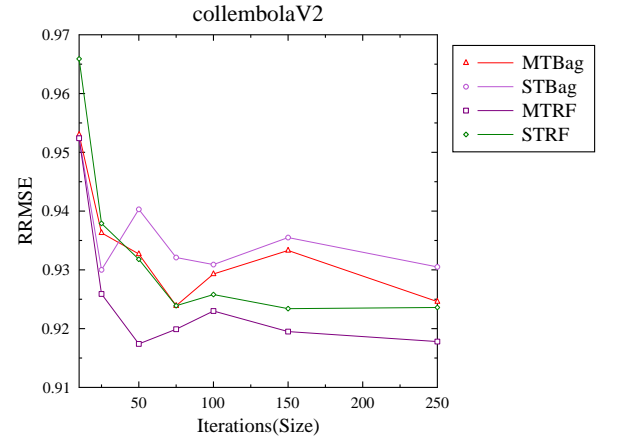
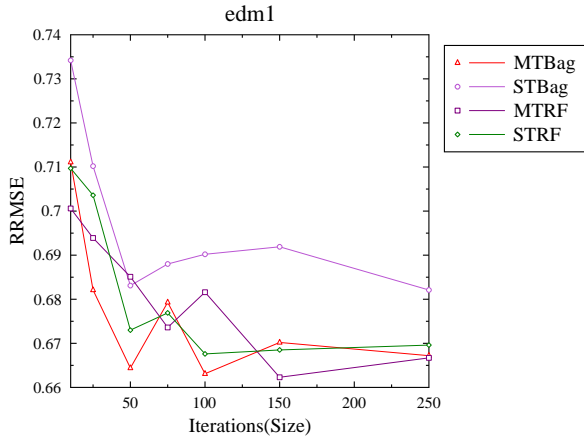
Domain	$N_{train}$	$N_{test}$	$D/C$	$HierarchySize$	$\overline{Label}$
ImageCLEF2007(Directional)	10000	1006	0/80	XX	1.0
ImageCLEF2007(Anatomical)	10000	1006	0/80	XX	1.0
Diatoms	2065	1054	0/371	XX	1.0
Enron	988	660	0/1001	XX	YY
Reuters	3000	3000	0/47236	XX	YY
WIPO	1352	358	0/74435	XX	YY
Expression (GO)	2485	1288	0/551	XX	YY
SCOP (GO)	6507	3336	0/2003	XX	YY
Sequence (FunCat)	2455	1264	2/4448	XX	YY
Yeast (GO)	2310	1155	5588/342	XX	YY

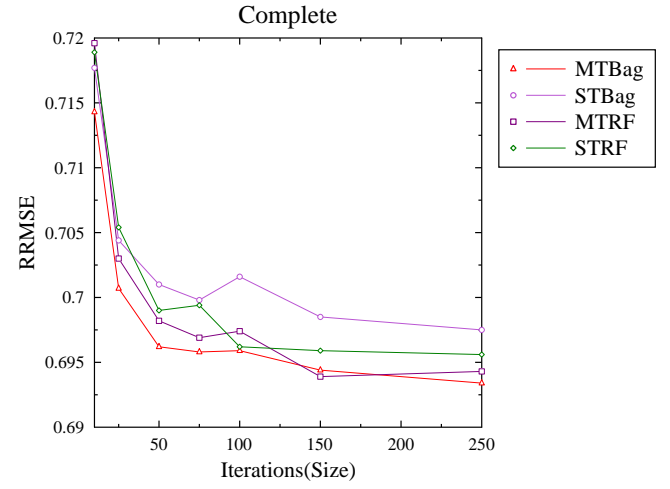
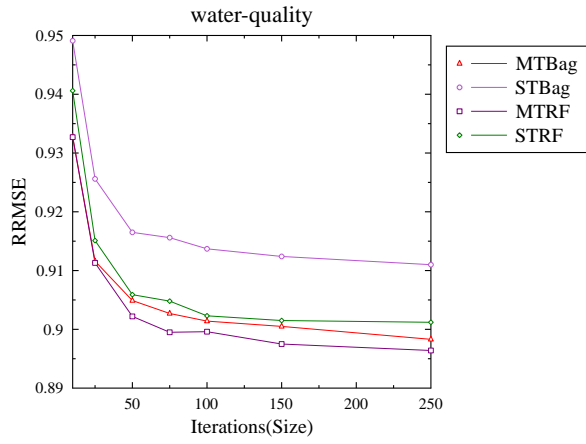
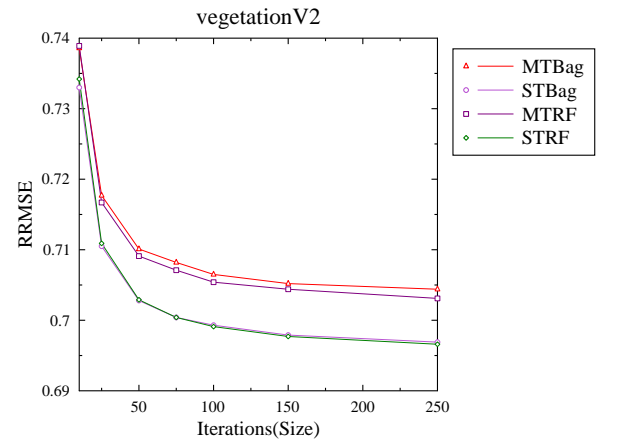
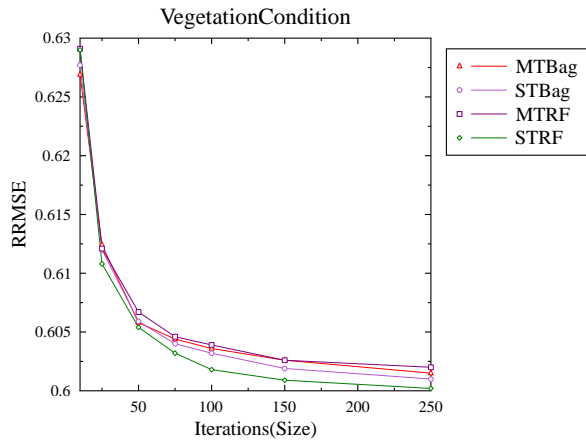
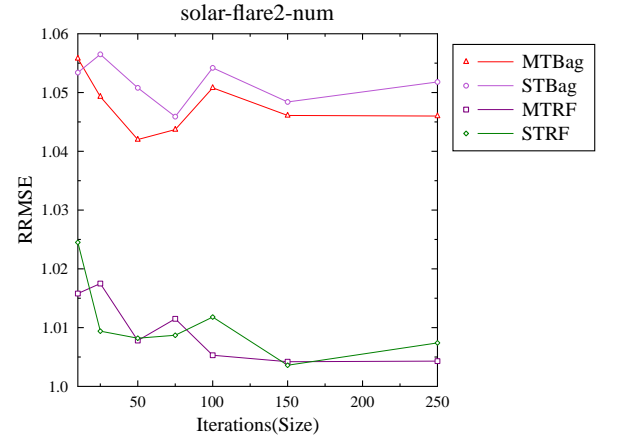
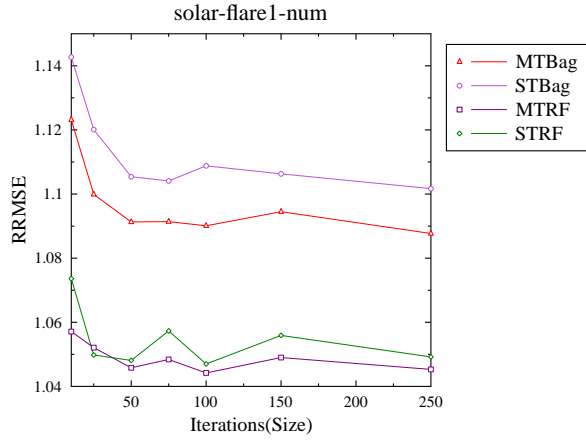
# 1 Multi-Target Regression - Correlation Coefficient





## 2 Multi-Target Regression - RRMSE





3 HMLC -  $\overline{AUPRC}$ 