a Prediction performances in train

b Prediction performances in test

* Proportion of full data sampled for training connected by label (green curve)
distance_stat:
Kolmogorov-Smirnov statistics with testing the distance between the training data and the full data