Comparison of outlier detection techniques

- Federica Bianco, Umaa Rebbapragada, Kelle Cruz, Miguel de Val-Borro, Daniel Giles, Lucianne Walkowicz, Rafael Martinez-Galarza, Matthew Graham, Rishi Paudel

- Used 2500 Kepler light curves to compare outliers found by different algorithms

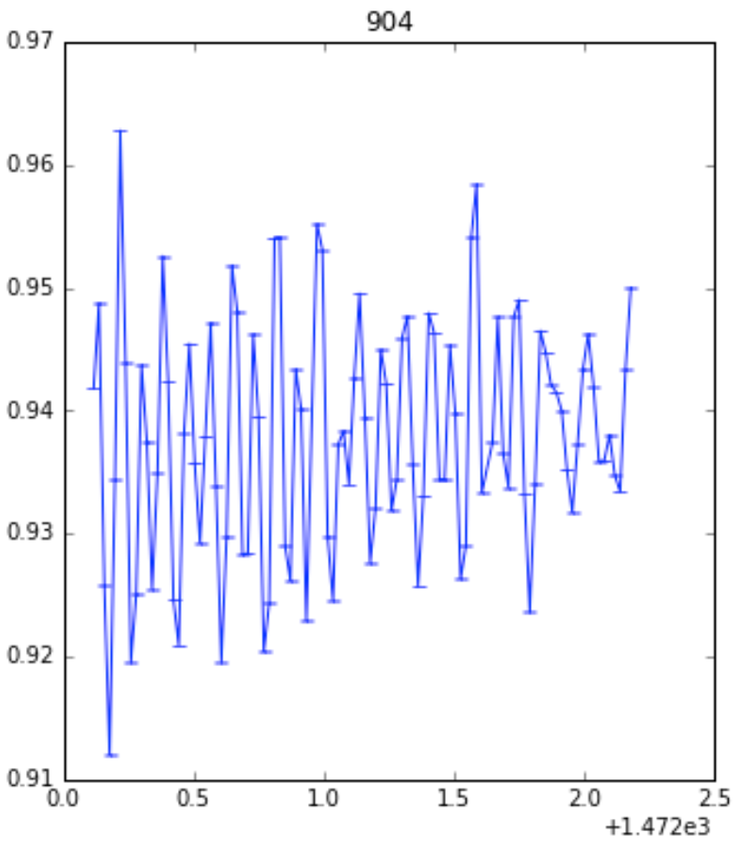
- SVM, Isolation Forest, Clustering, Persistent Homology, Unsupervised Random Forest

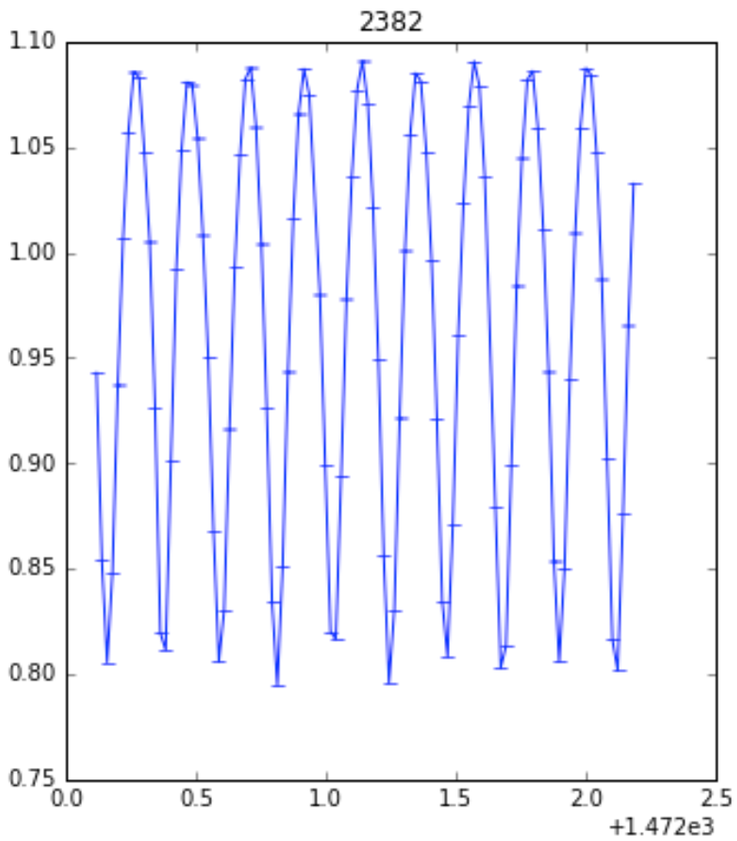
Overlap Between Methods

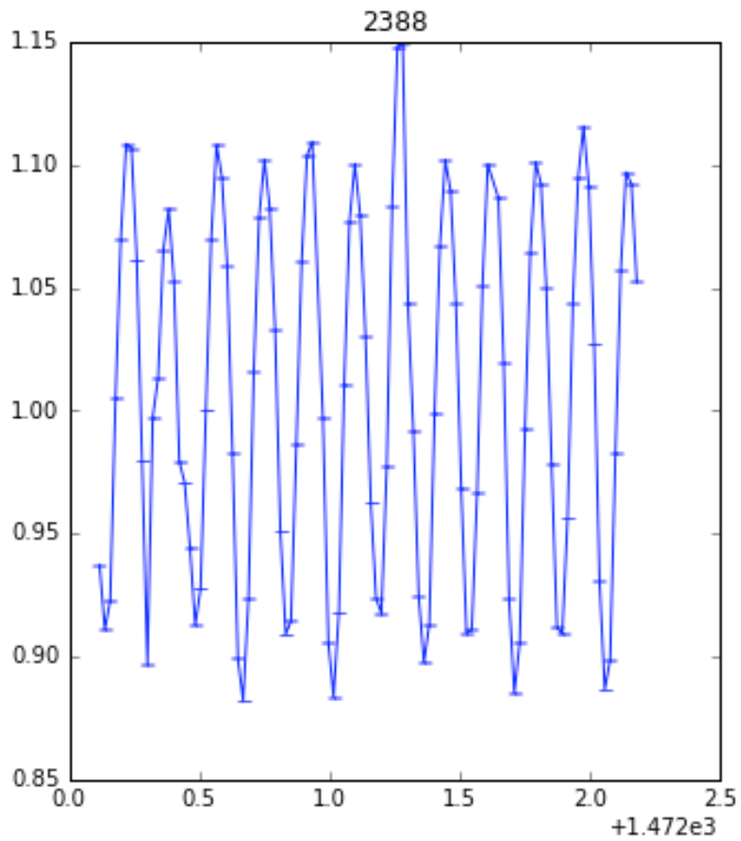
	IsoForest	MinSpanTree	Cluster	OneClassSVM
IsoForest		9	0	17
MinSpanTree			1	12
Cluster				0
OneClassSVM				

9 LCs common between IsoForest, MinSpanTree and OneClassSVM (1 of 3)

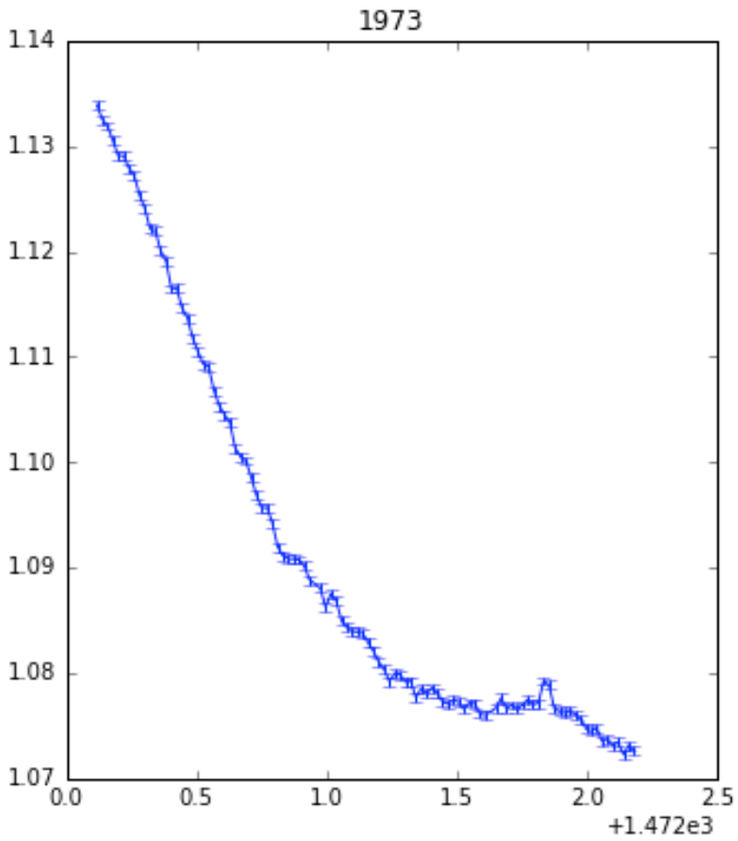
in no particular order

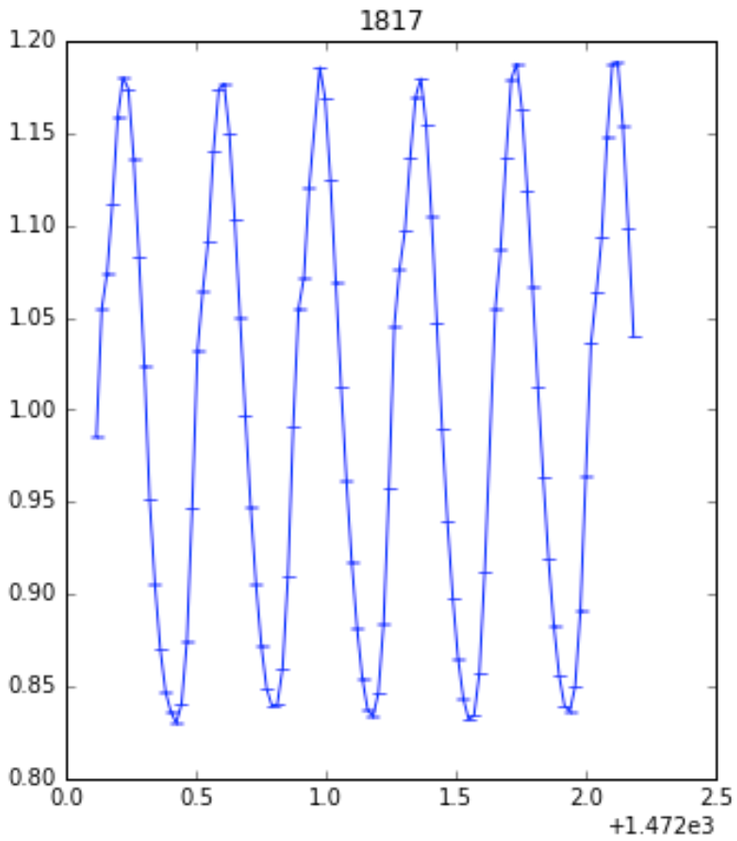


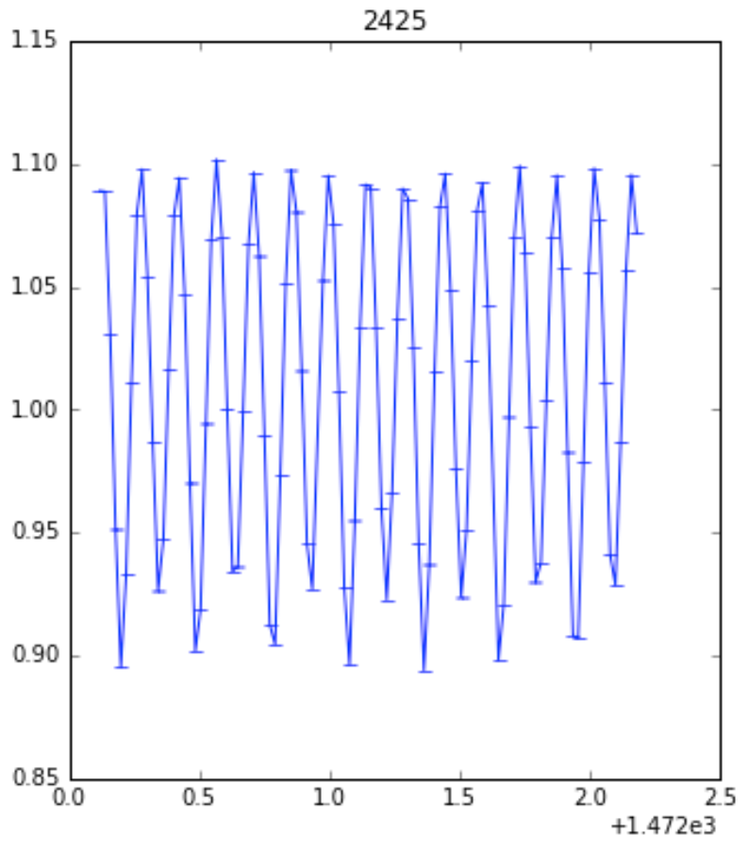




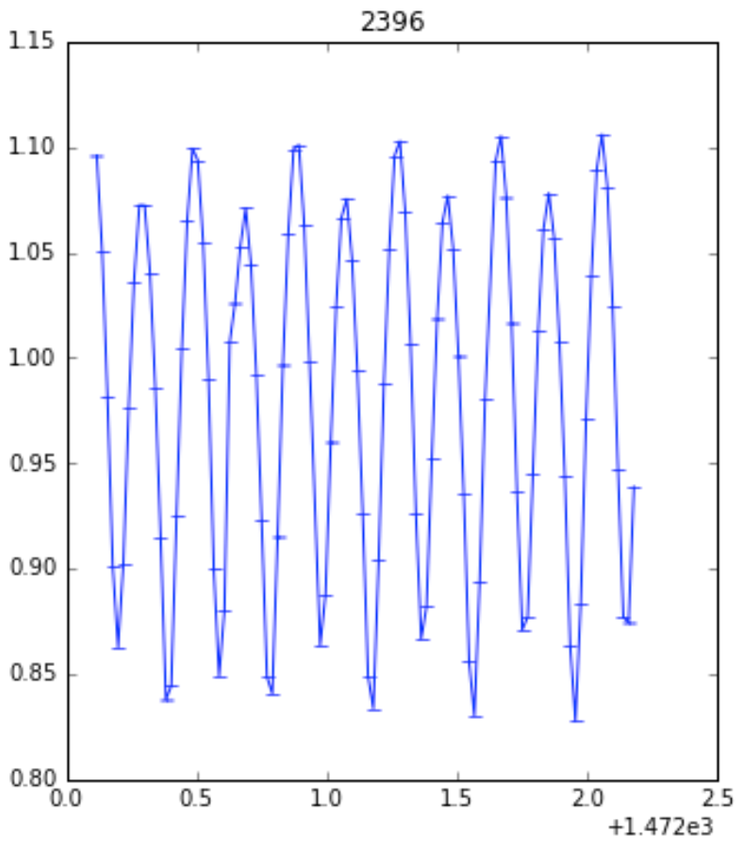
9 LCs common between IsoForest, MinSpanTree and OneClassSVM (2 of 3)

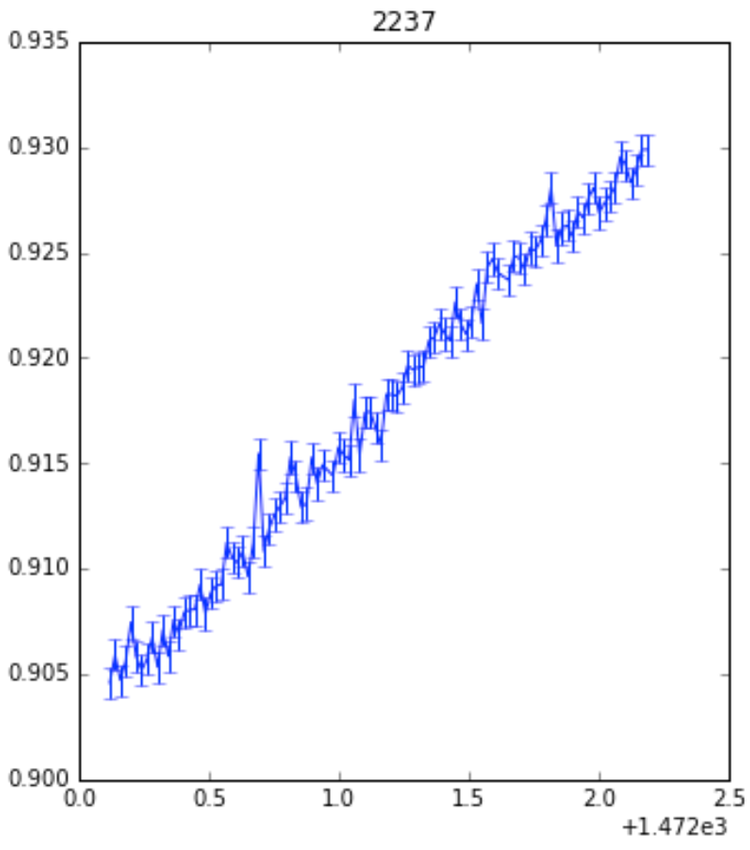


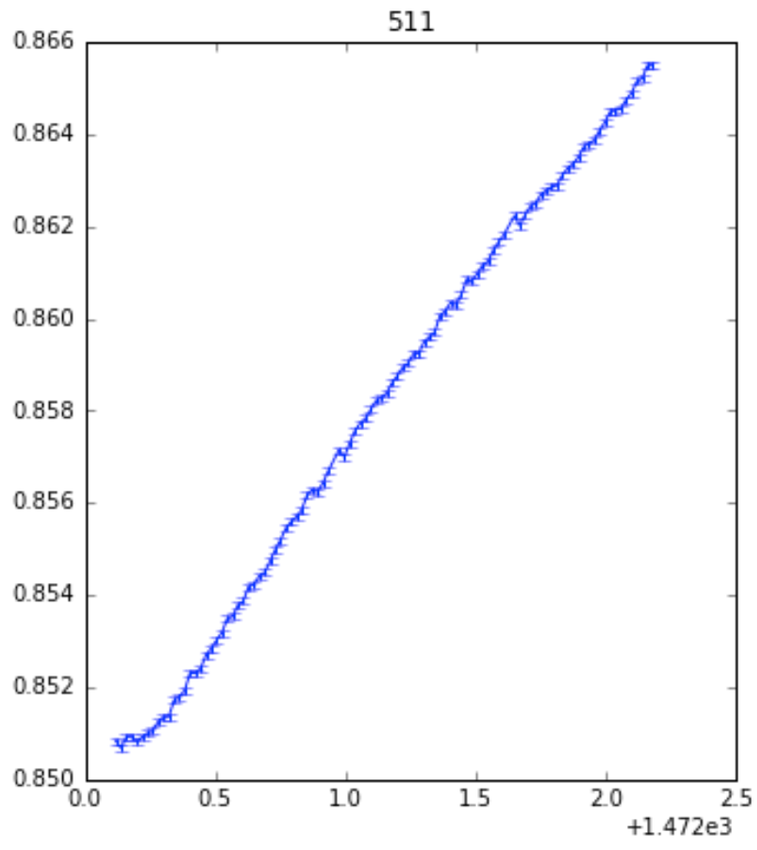




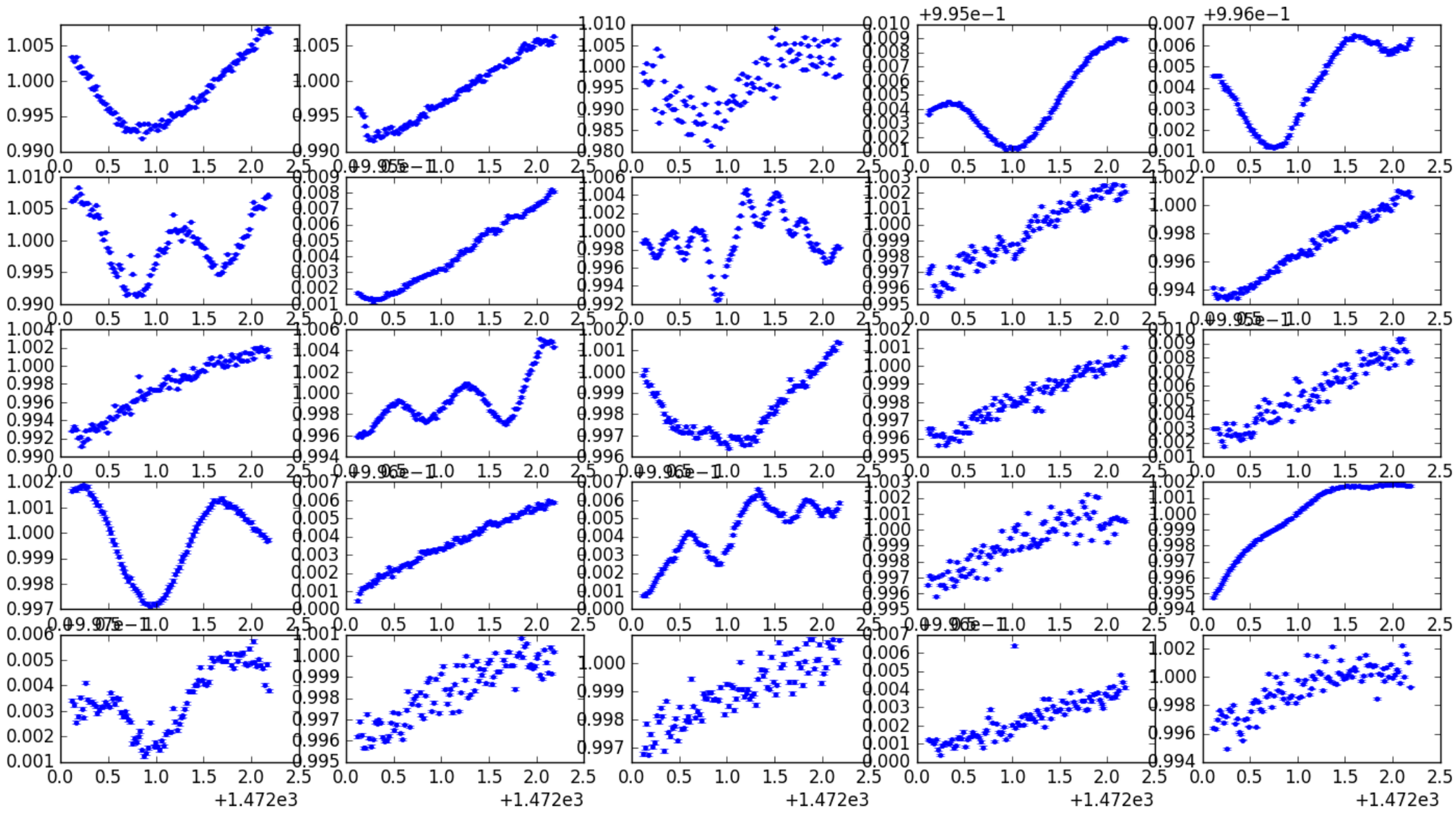
9 LCs common between IsoForest, MinSpanTree and OneClassSVM (3 of 3)



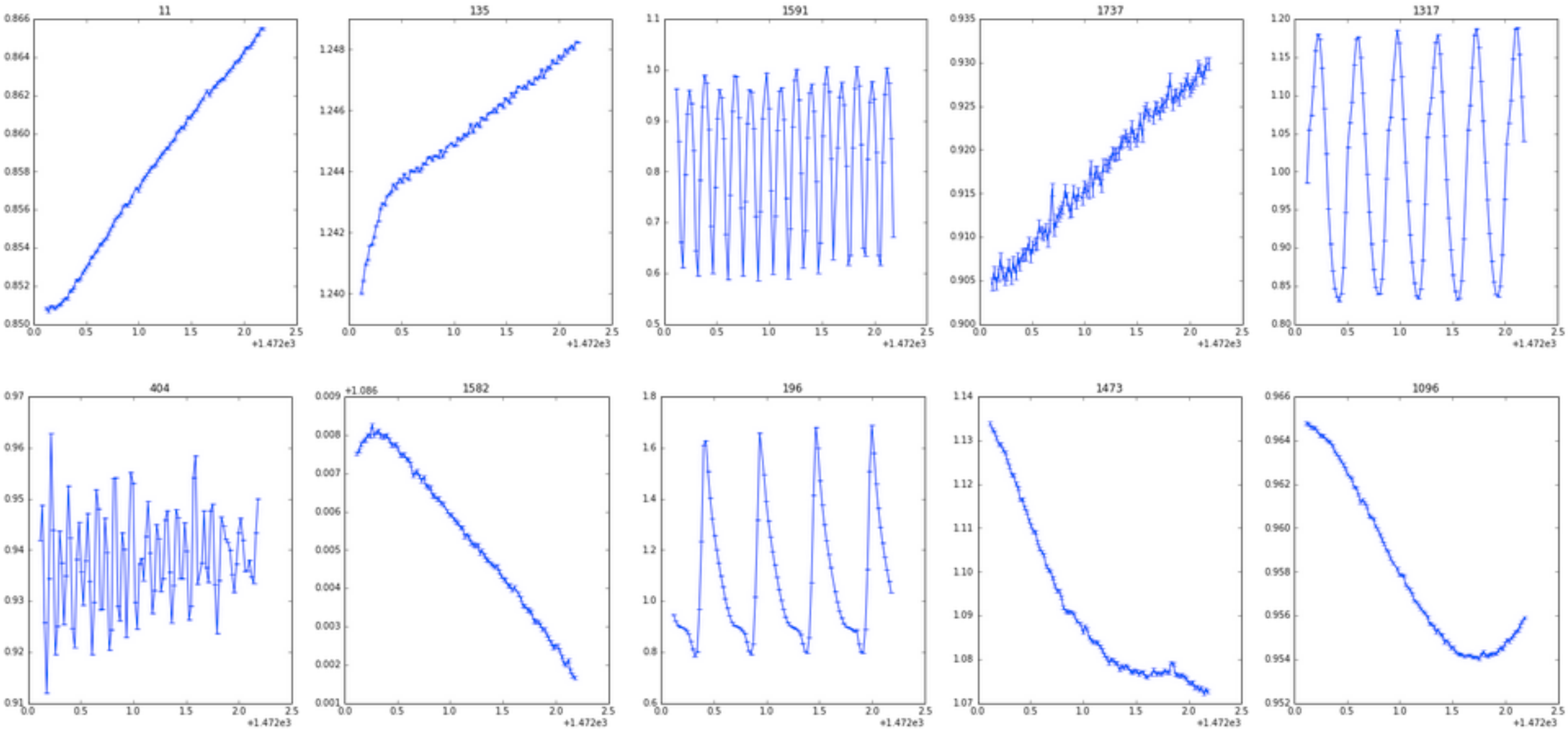




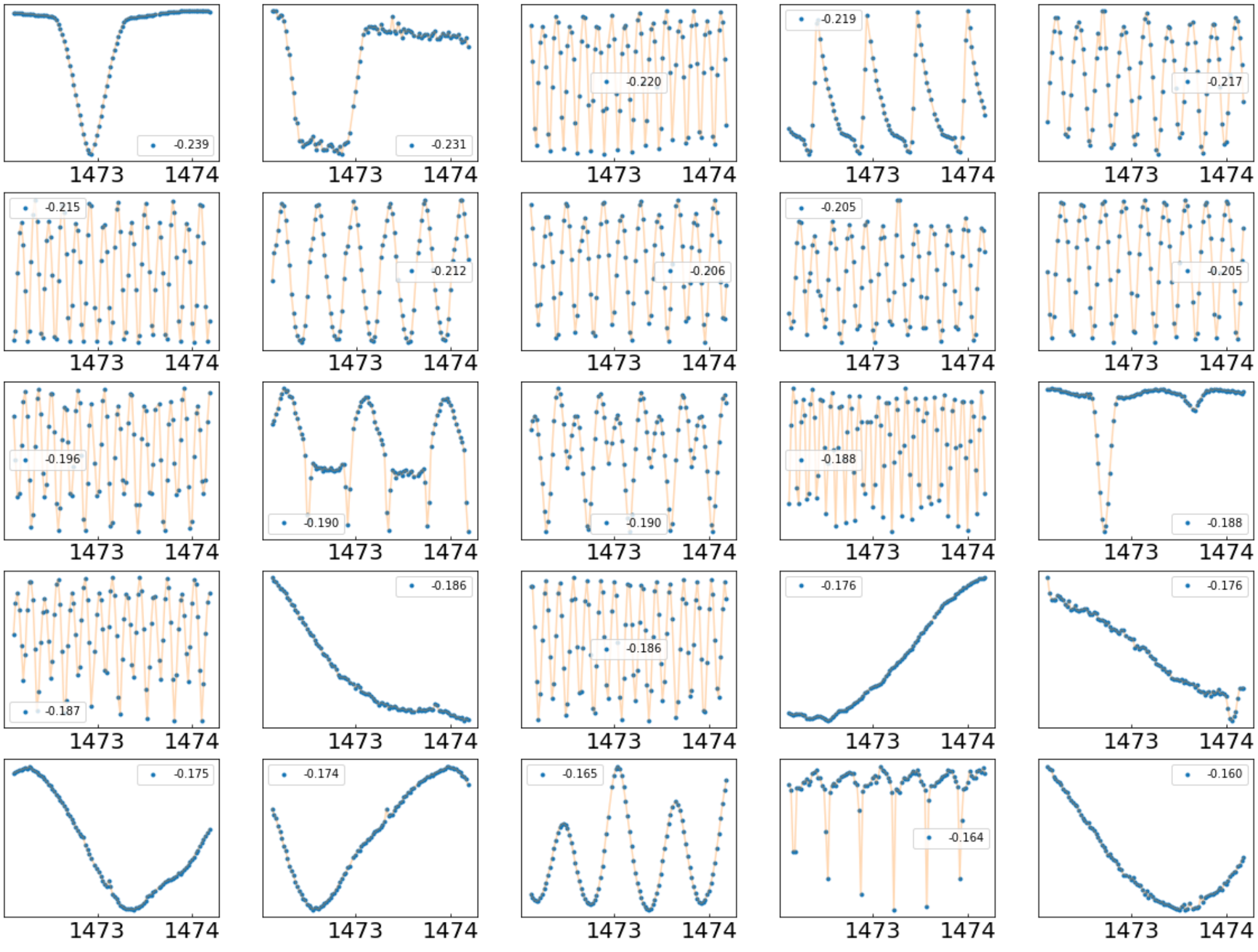
SVM

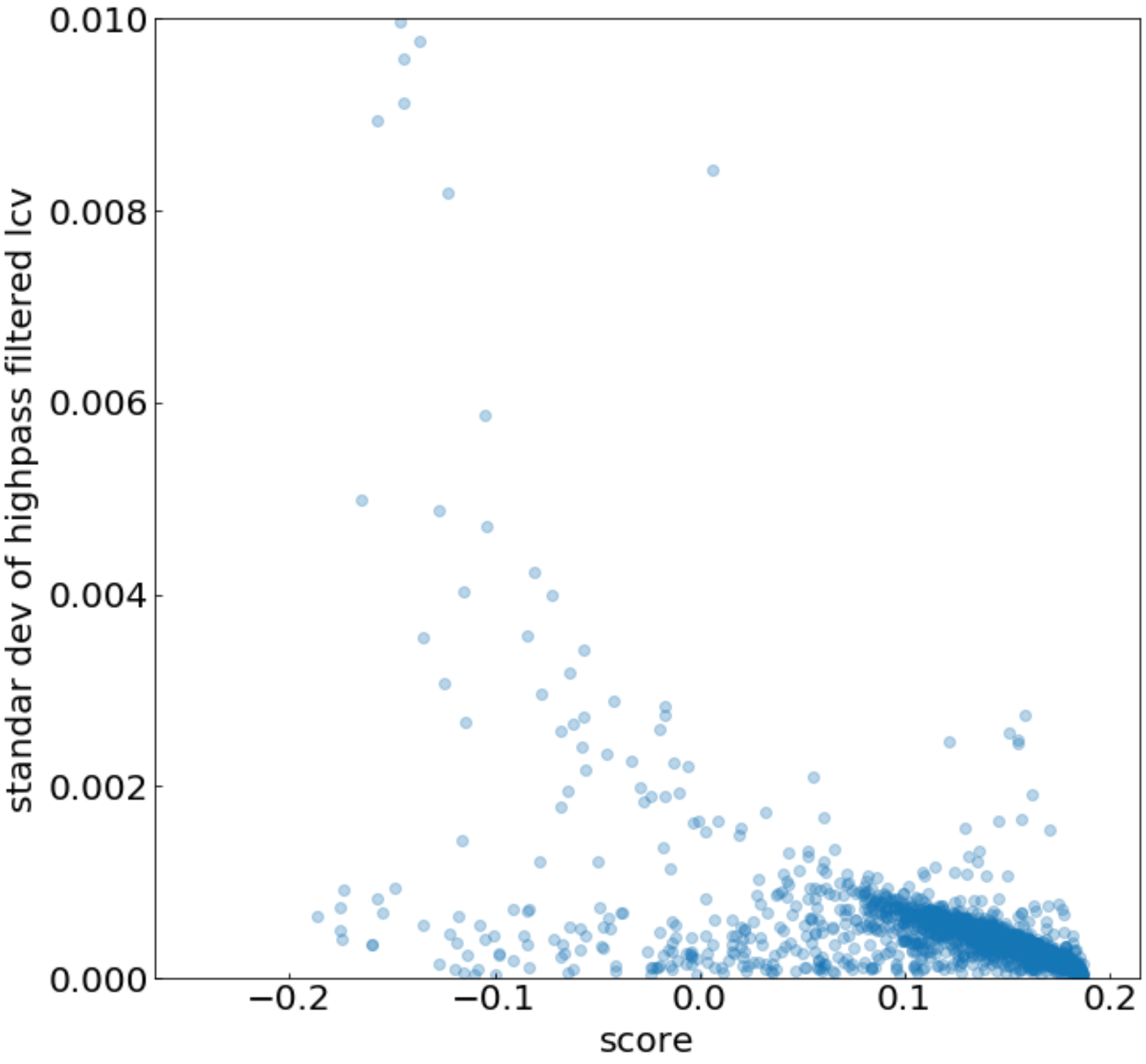


Isolation Forest

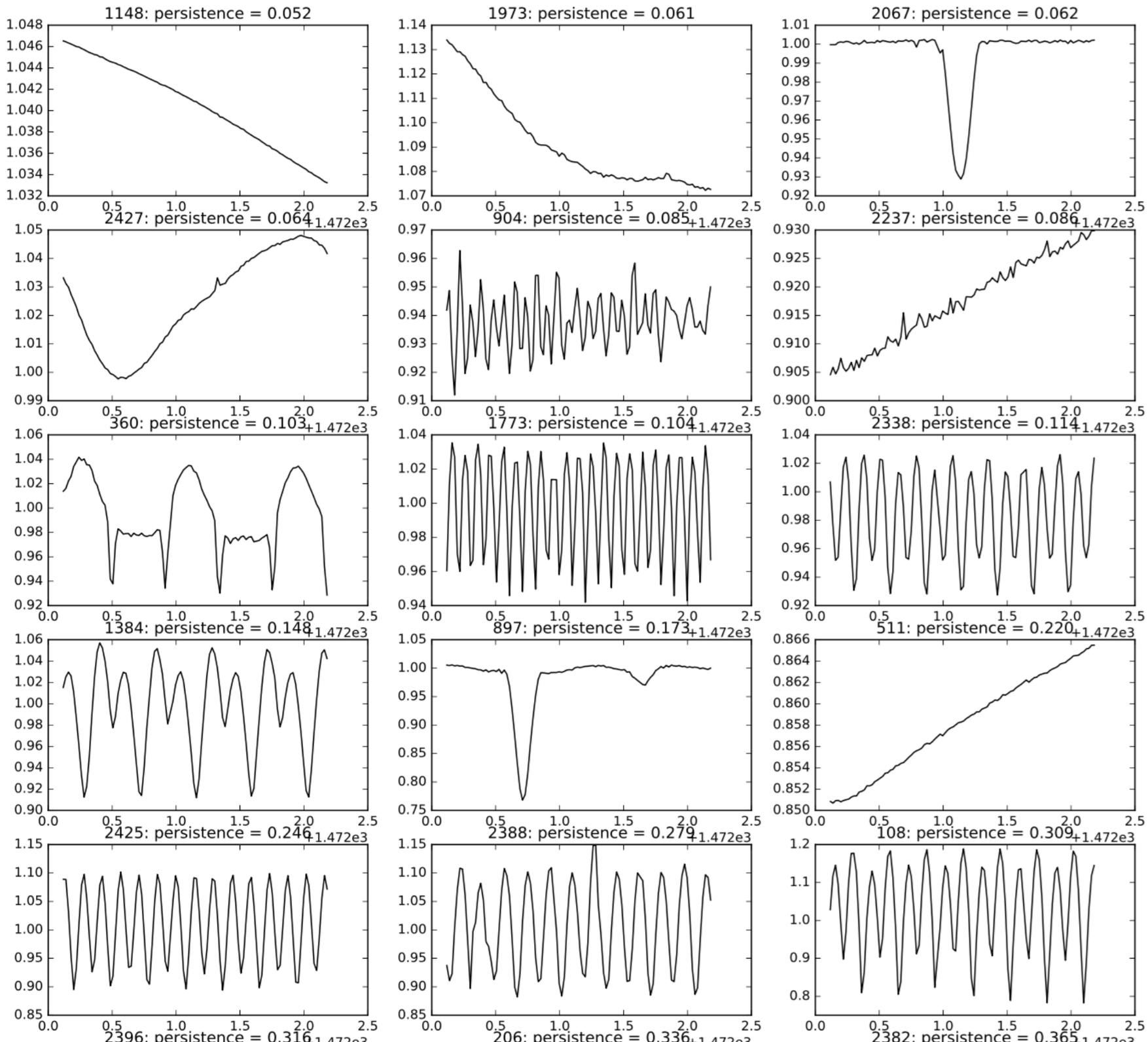


outliers (top 25)

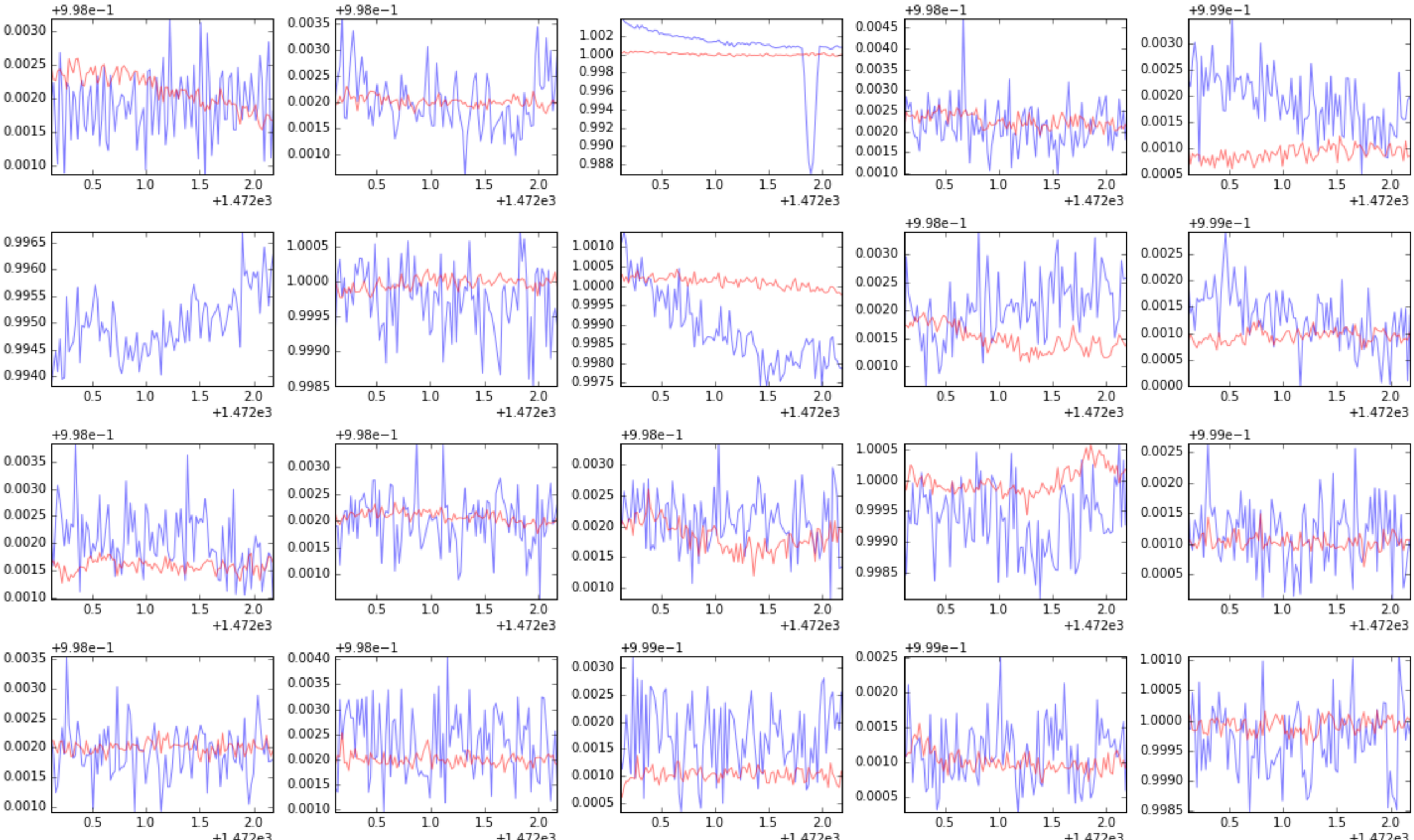



0.010
0.008
0.006
0.004
0.002
0.000
standar dev of highpass filtered lcv
−0.2
−0.1
0.0
0.1
0.2
score

MinSpanTree



Unsupervised Random Forest



Clustering

(lightcurve snippets)

Clustering

(full lightcurves)