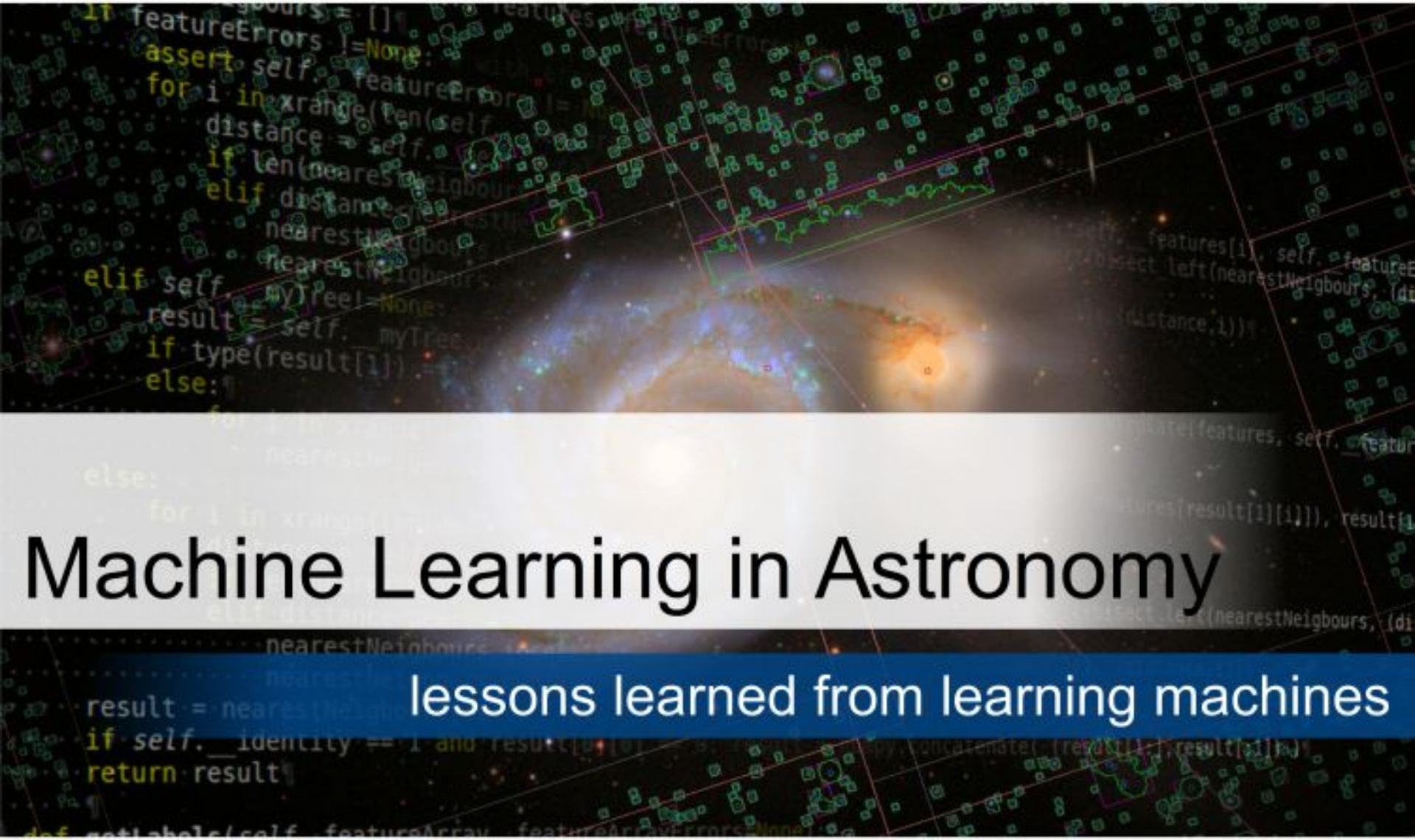
Heidelberg Institute for
Theoretical Studies

HITS



Machine Learning in Astronomy

lessons learned from learning machines

when we talk about
machine learning
in the context of
astronomy

we talk about
analysing

large
and / or
complex **datasets**

Machine Learning, Taxonomy

discriminate between

supervised learning (labeled data)

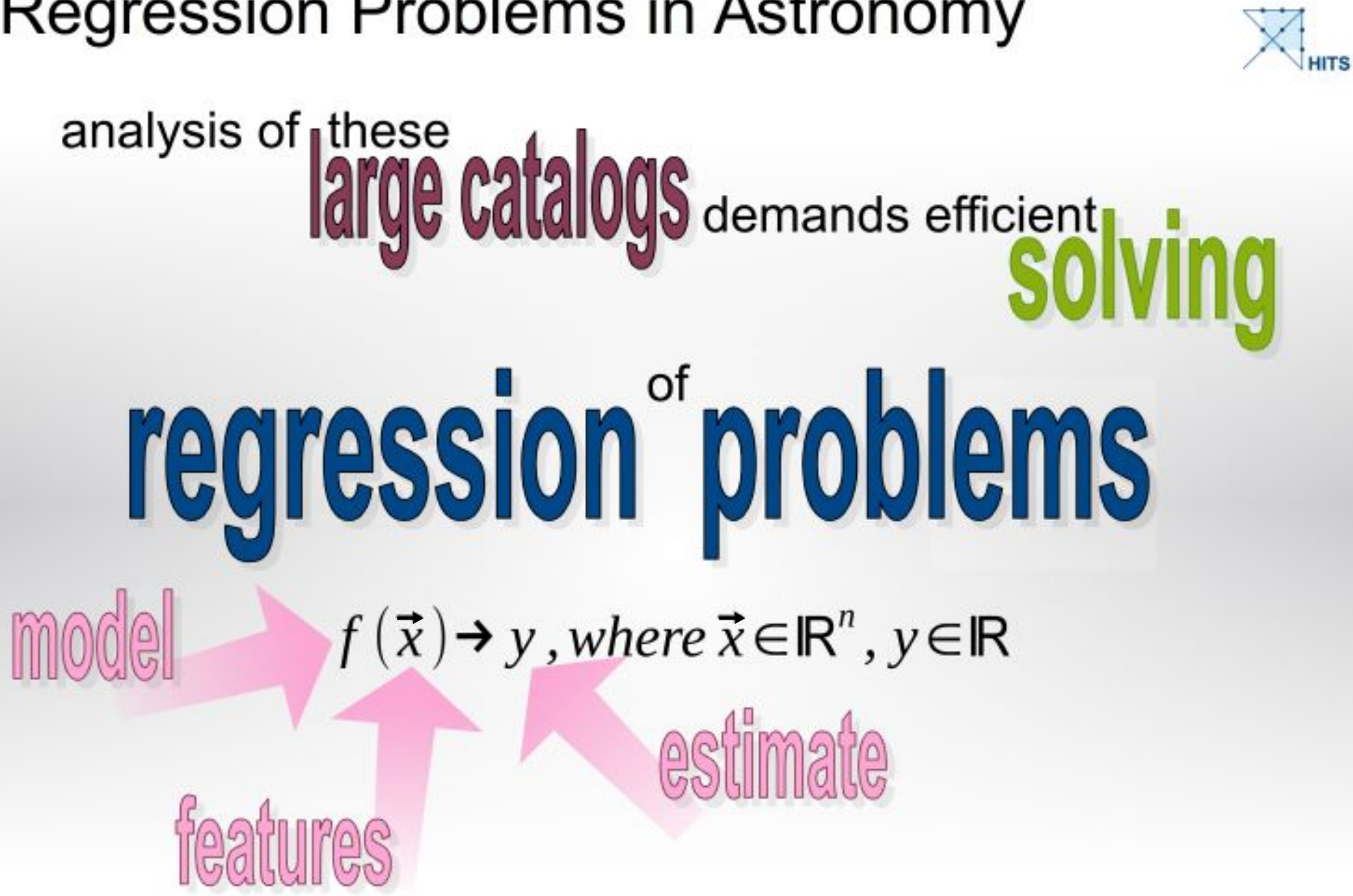
classification, ranking, regression

semi-supervised learning (partially labeled data)

un-supervised learning (unlabeled data)

clustering, dimension reduction

Regression Problems in Astronomy

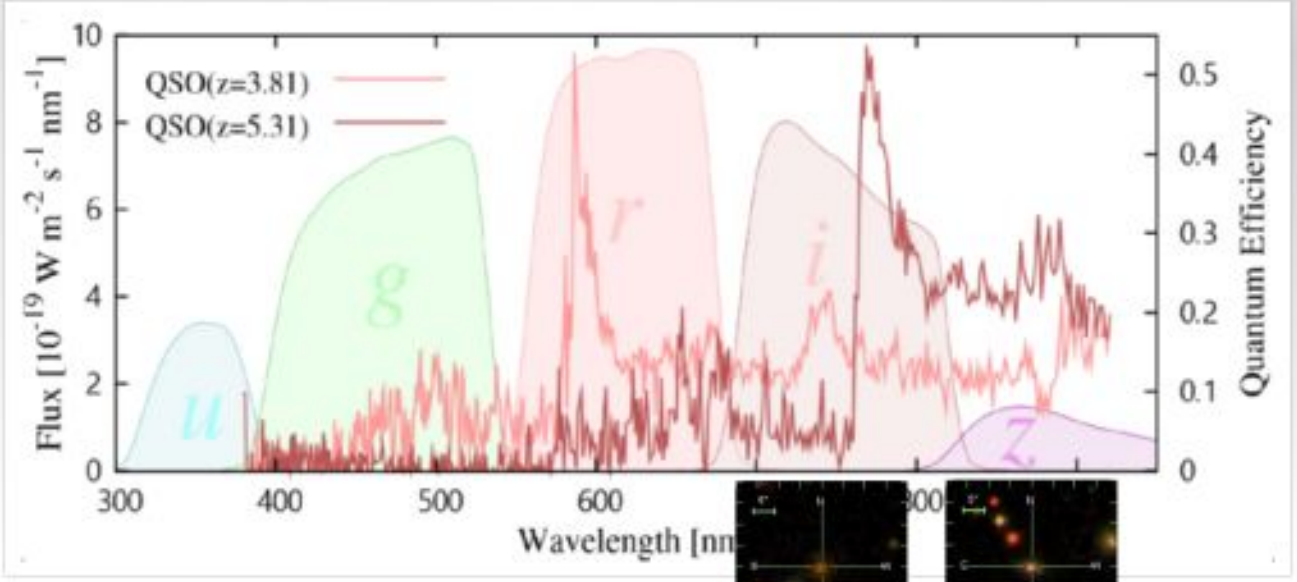


Redshift Estimation

we want to **determine** important **properties** of objects

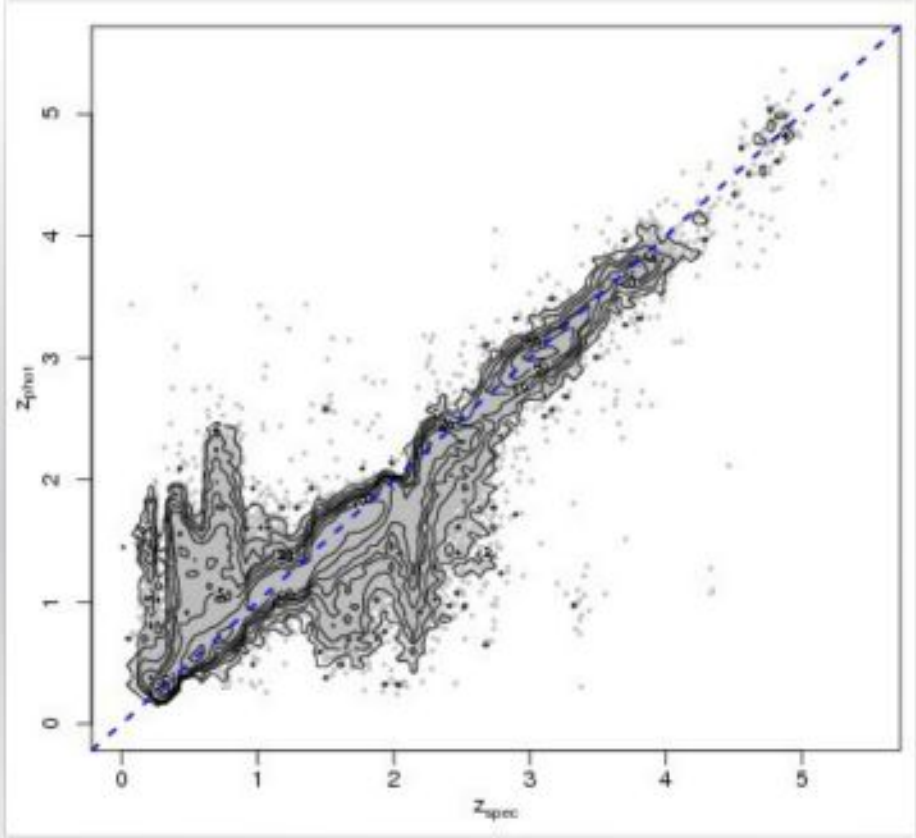
like **redshift**

$$1 + z = \frac{\lambda}{\lambda_0}$$



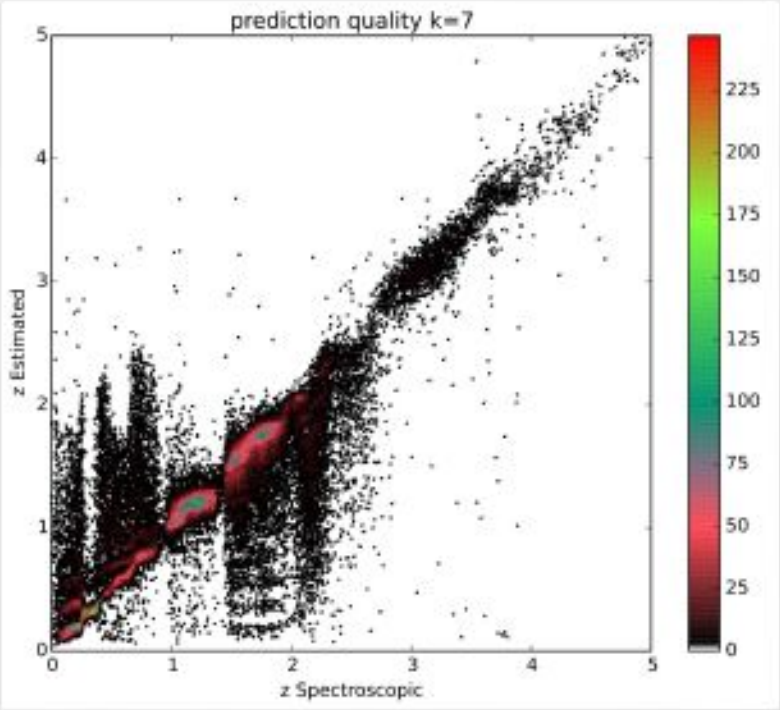
Redshift Estimation

Existing Models



Laurino et al. 2011

$$RMSE(\Delta z_{norm}) = 0.19$$
$$MAD(\Delta z_{norm}) = 0.041$$



Polsterer et al. 2012

$$RMSE(\Delta z_{norm}) = 0.23$$
$$MAD(\Delta z_{norm}) = 0.048$$

Feature Selection

not all of the **features** are useful …

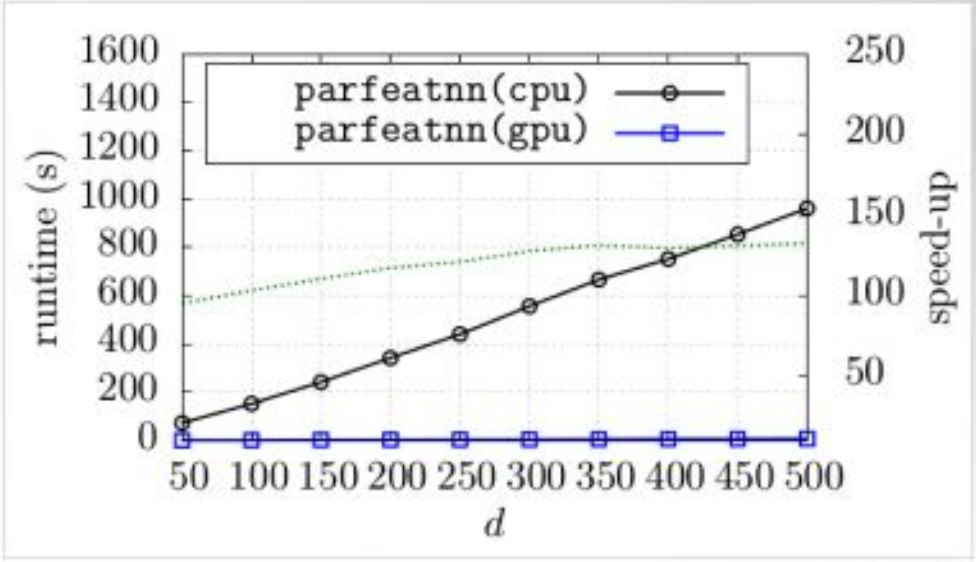
… test different feature **combinations**

$$\frac{n!}{(n-r)!\,r!},\ with\ n=55, r=4$$

$$\rightarrow 341{,}055\ combinations$$

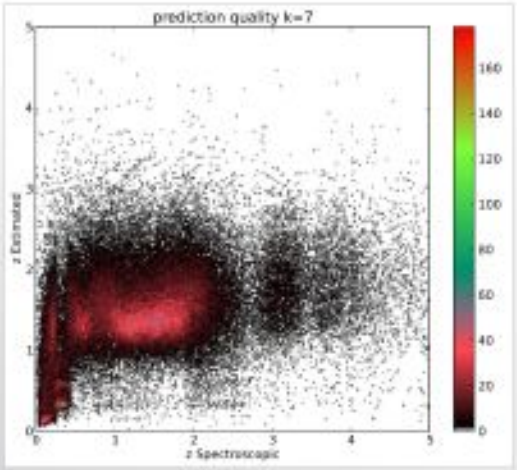
evaluating **1** model = 100 sec.

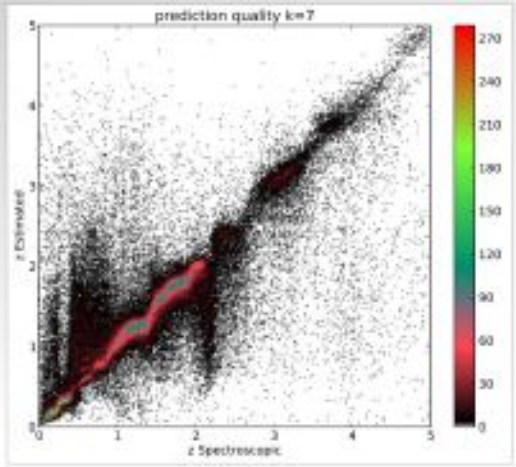
- improve data set
- do it in parallel



Gieseke et al. 2014

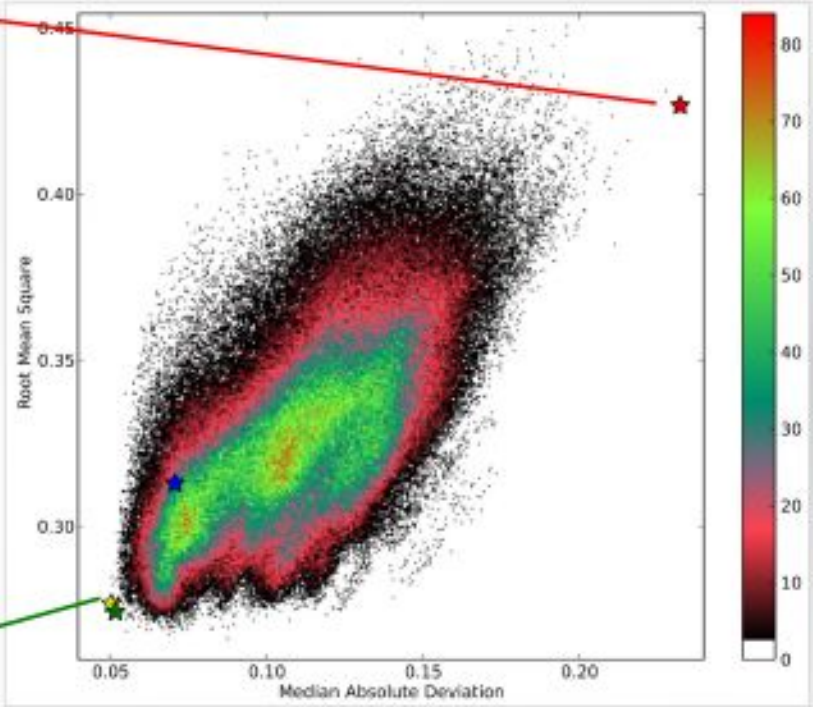
Complete Test





$i_{psf}, z_{psf}, i_{model} - z_{model}, i_{psf} - z_{psf}$

$u_{model} - g_{model}, g_{psf} - r_{model}, z_{psf} - r_{model}, i_{psf} - z_{model}$



Polsterer et al. 2013

Improvement

can we be even better ?

psf-
model-
petrosian-
} magnitudes in (*u, g, r, i, z*)
+ errors
extinction in (*u, g, r, i, z*)

585 features

best 10 out of 585

1,197,308,441,345,108,200,000

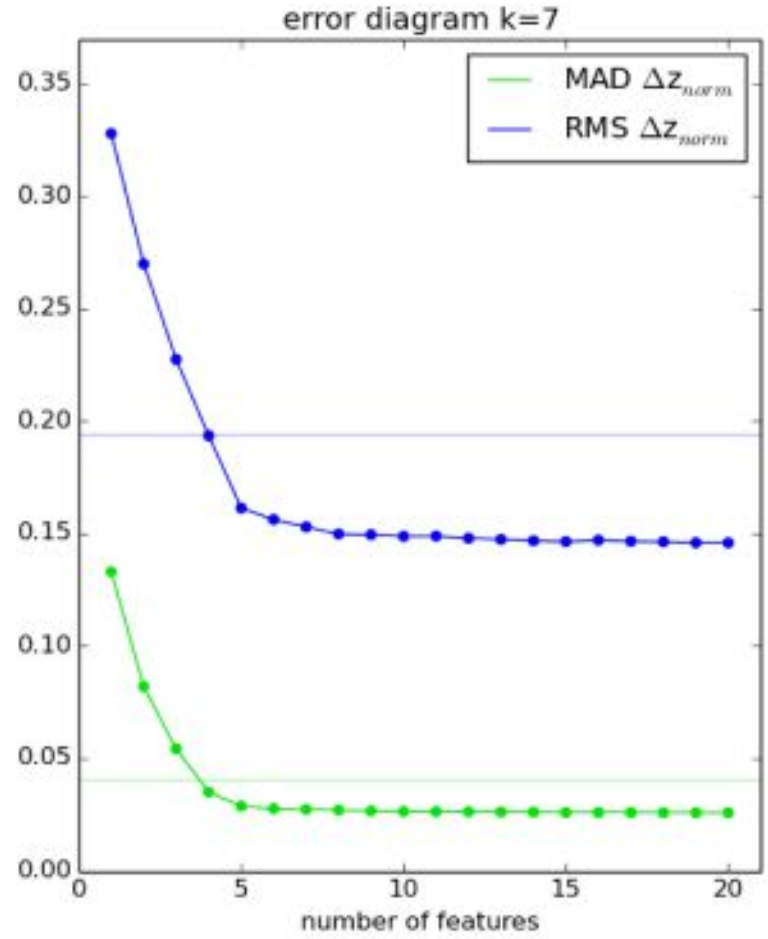
$1.2 * 10^{21}$ combinations!

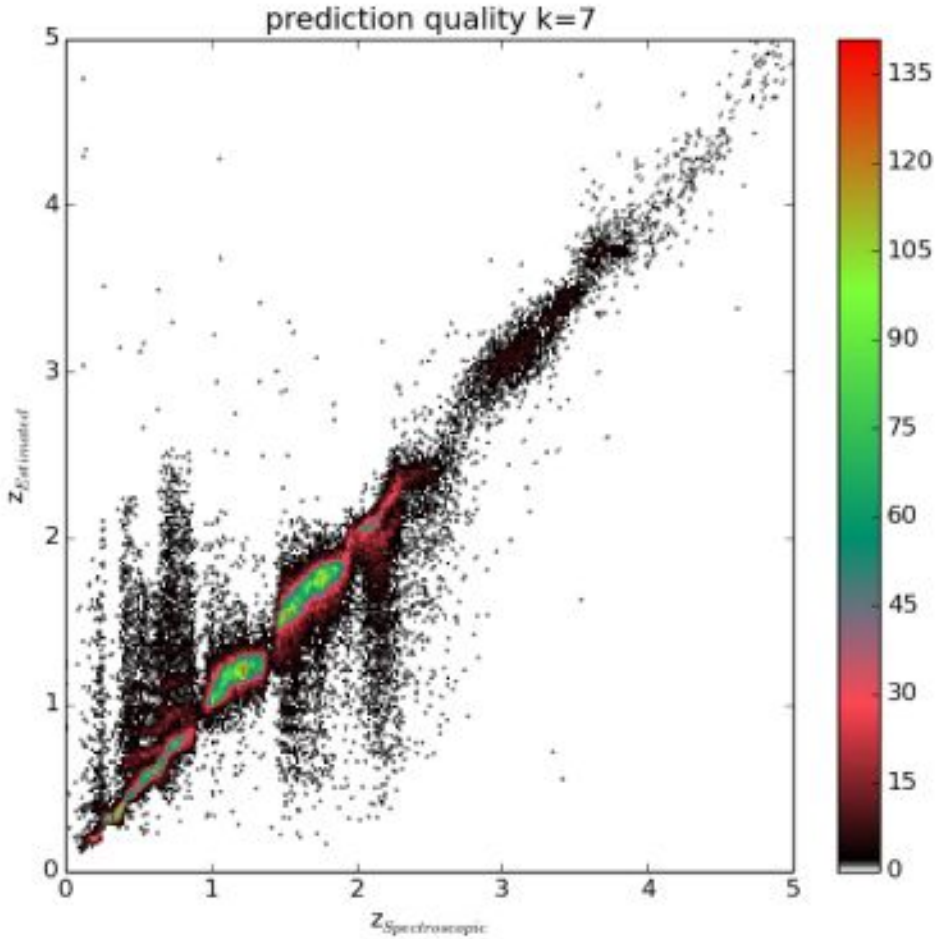
1.2 Sextillions

we need a better strategy!

Feature Selection Strategy

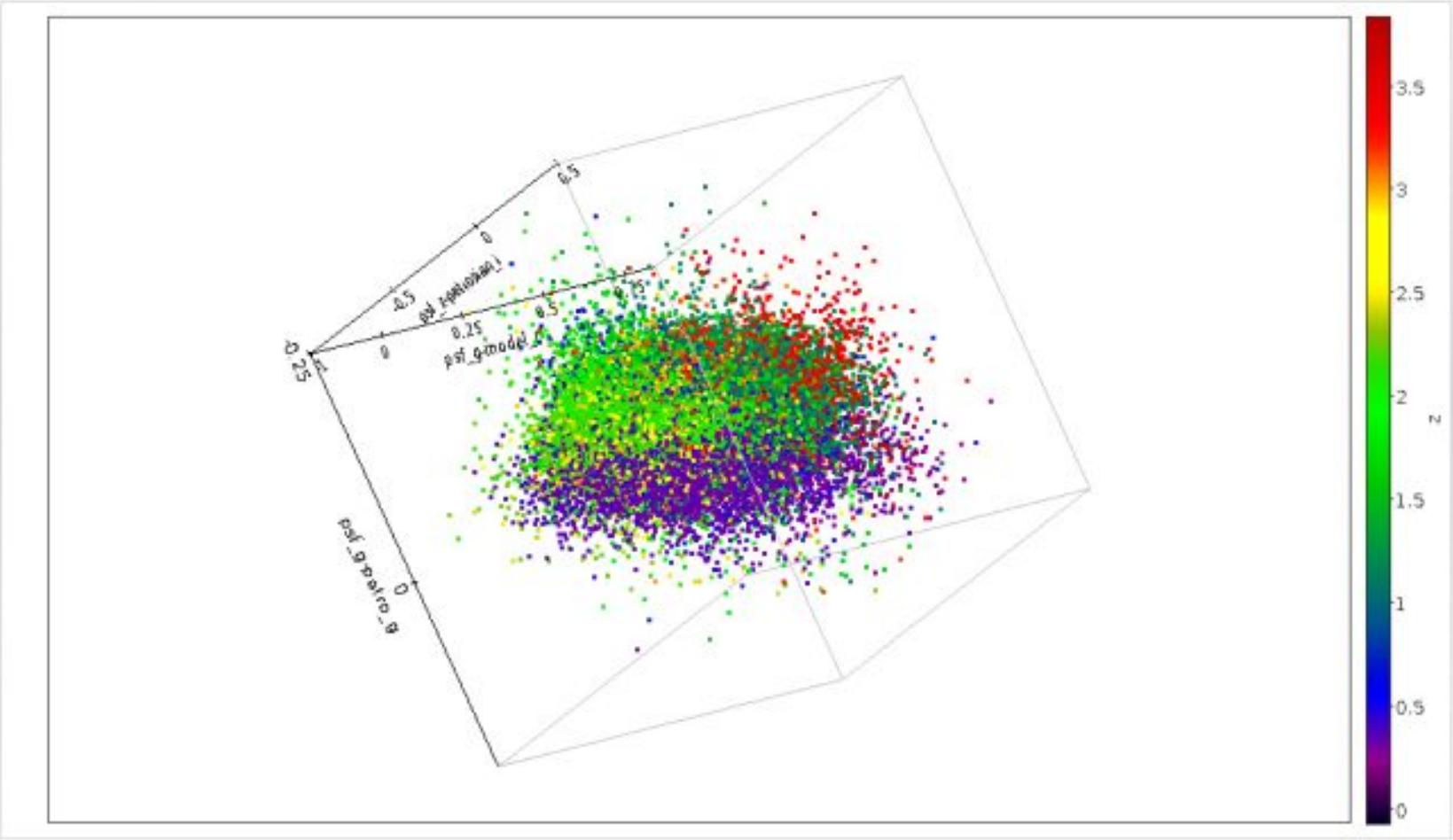
apply greedy forward selection



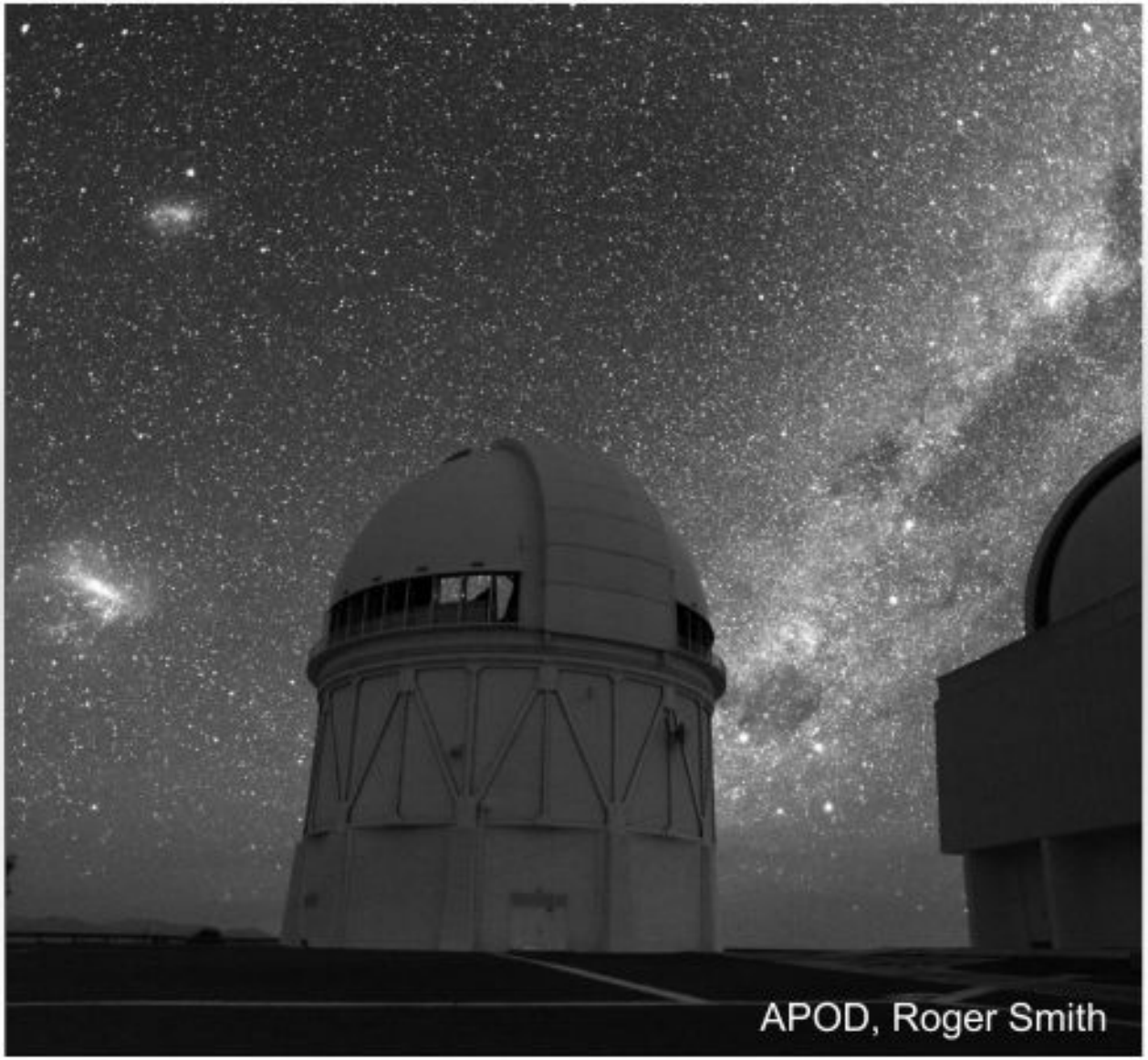


Forward Selection

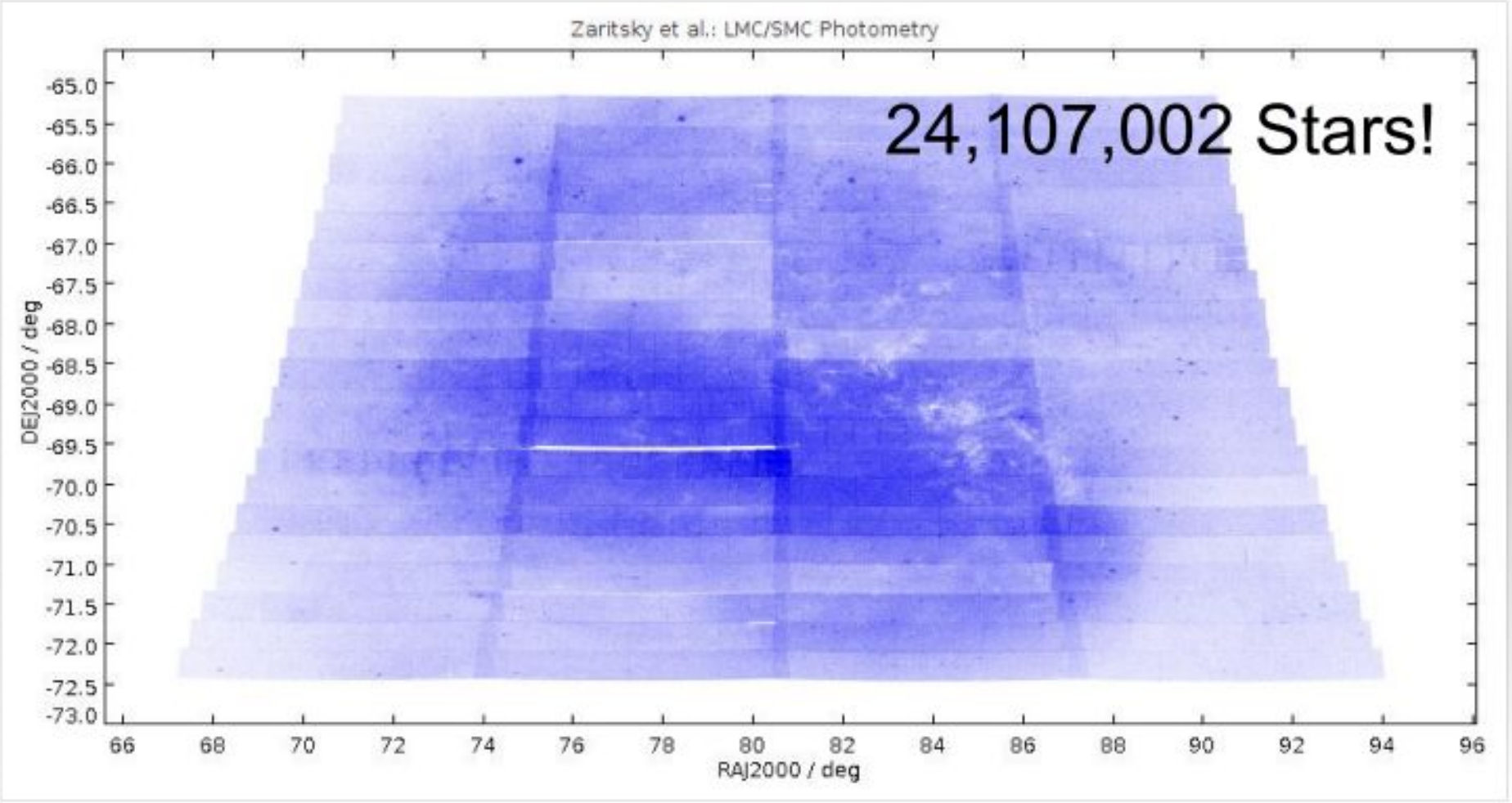
resulting features:



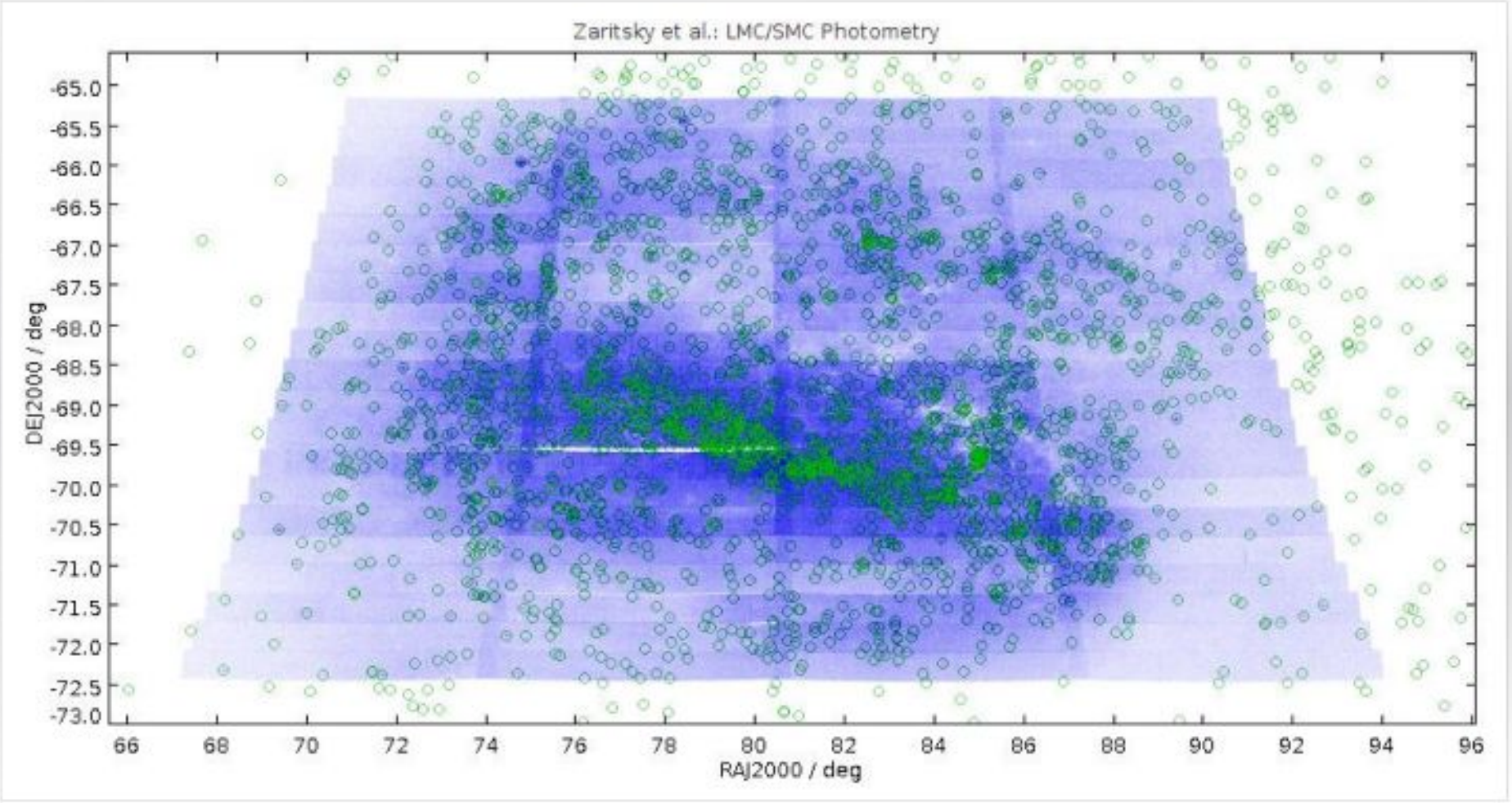
Dimension Reduction



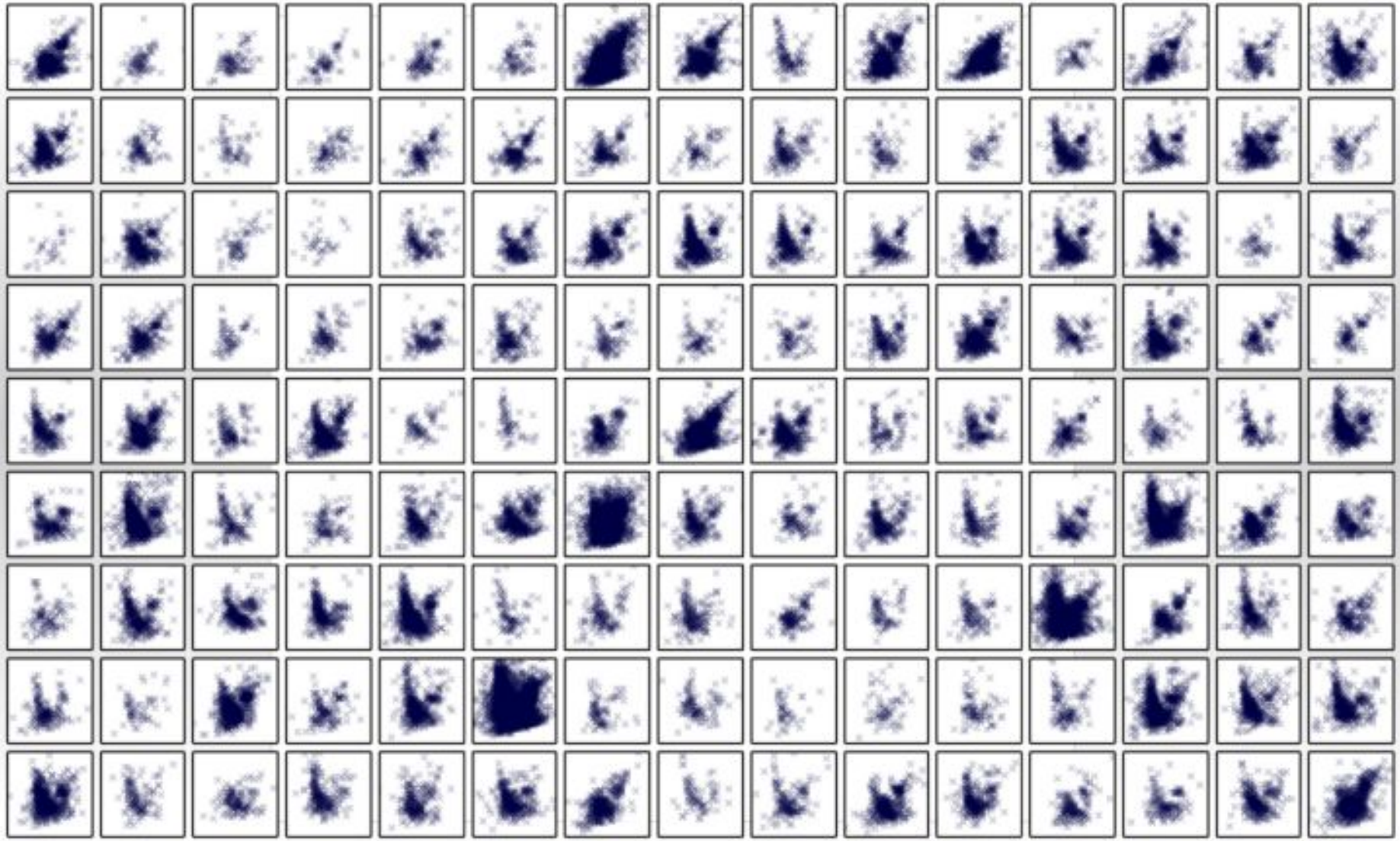
Star Formation History



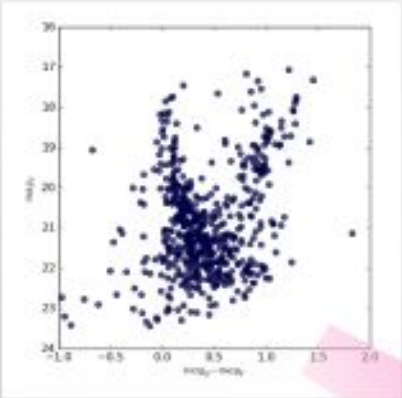
Star Formation History

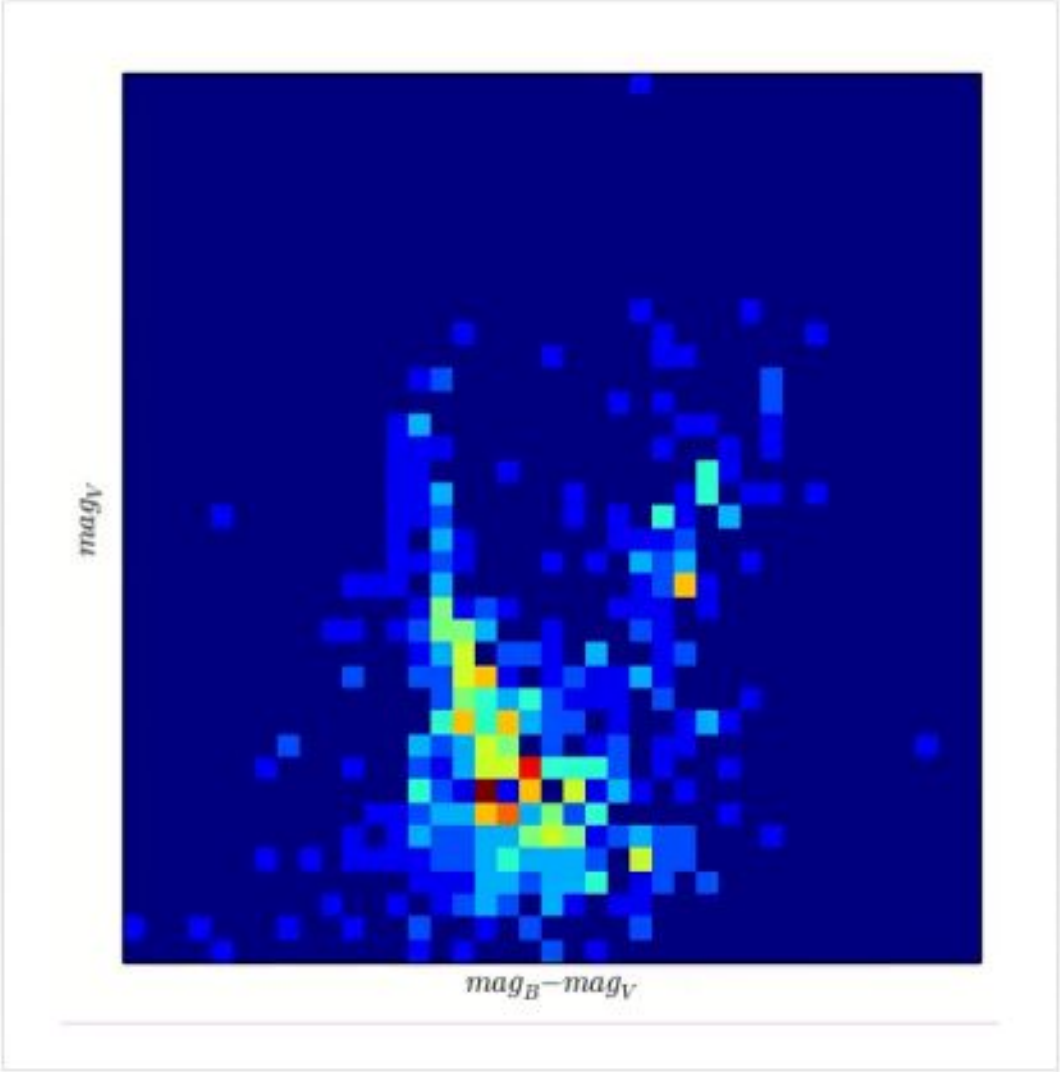


Analysis of Stellar Cluster

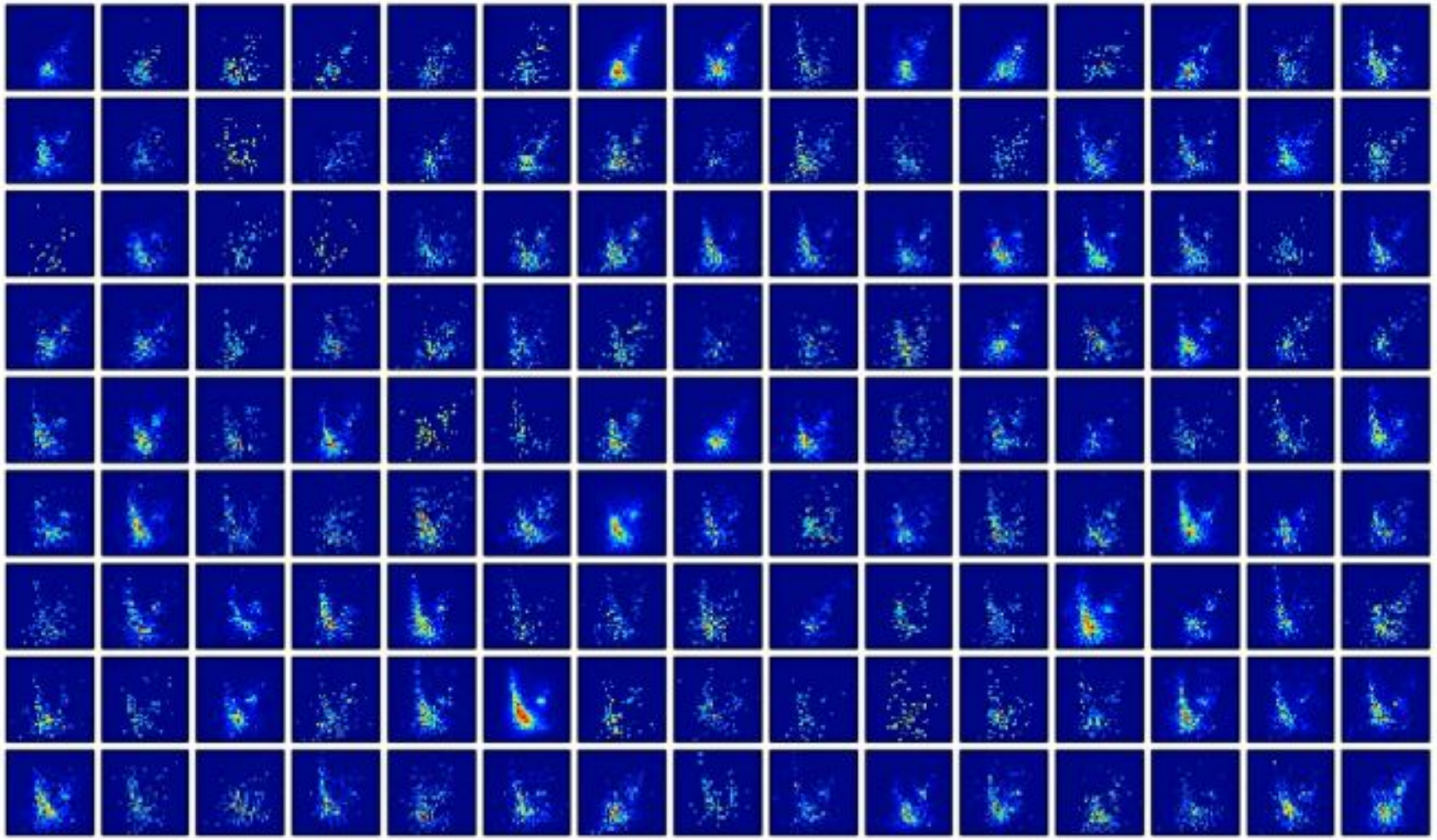


Analysis of Stellar Cluster





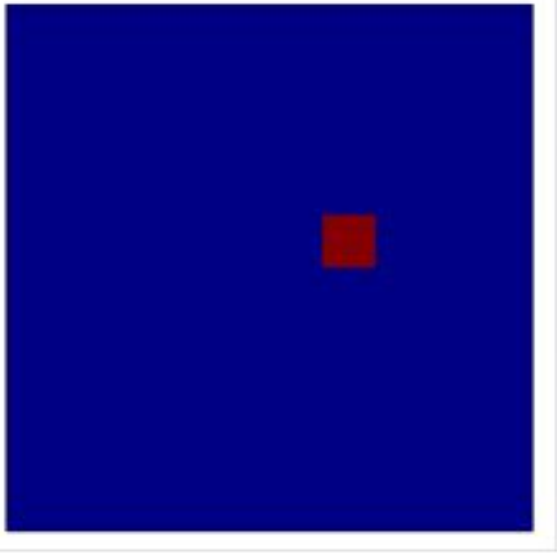
Analysis of Stellar Cluster

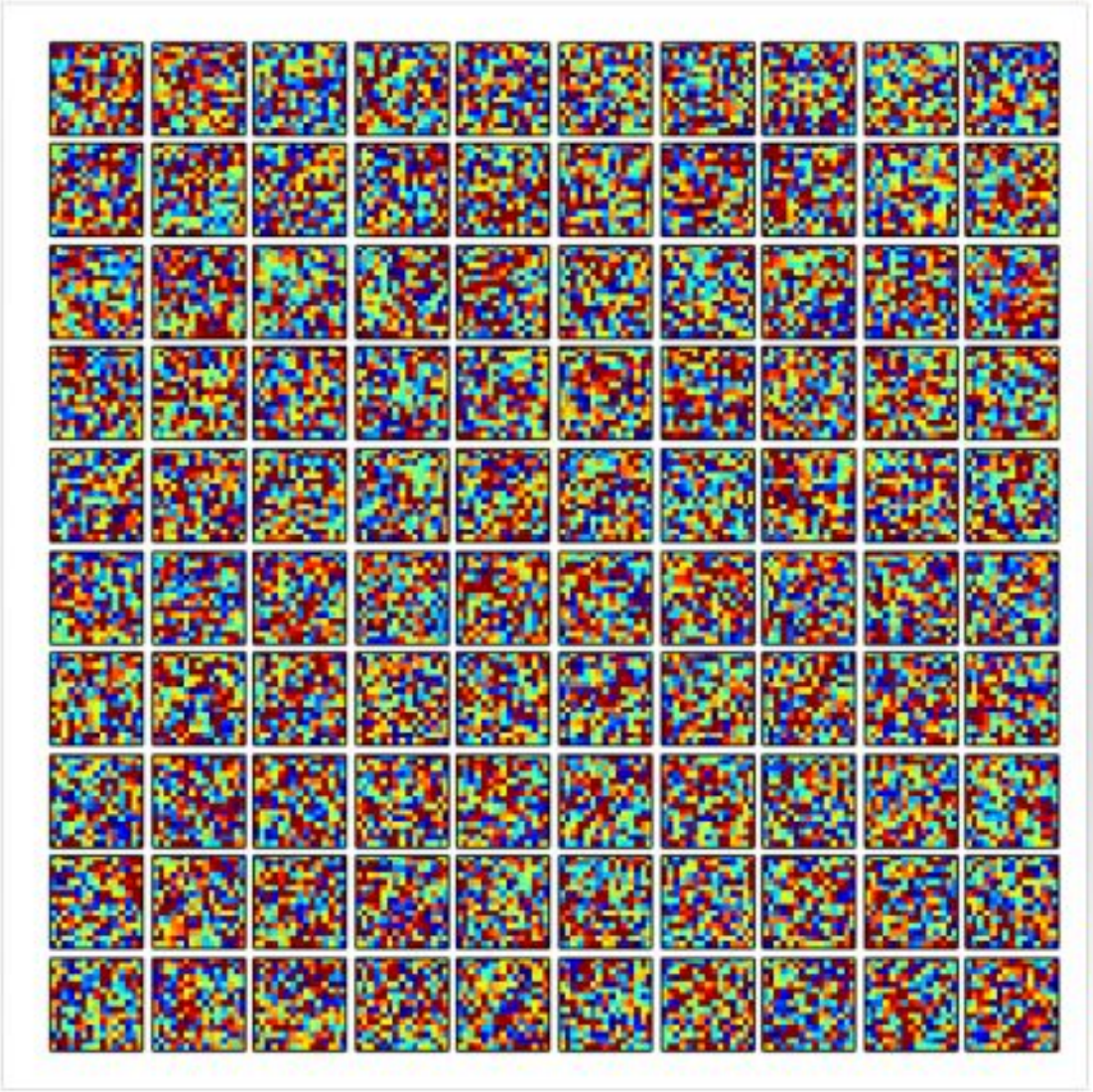


Dimension Reduction / SOM

$$\Delta(A,B)=\sqrt{\sum_{x=1}^{D_x}\sum_{y=1}^{D_y}(A_{xy}-B_{xy})^2}$$

$$\Delta(A,B)=\sqrt{\sum_{x=1}^{D_x}\sum_{y=1}^{D_y}\frac{\left(\frac{A_{xy}}{N_A}-\frac{B_{xy}}{N_B}\right)^2}{\frac{A_{xy}}{N_A^2}}}$$

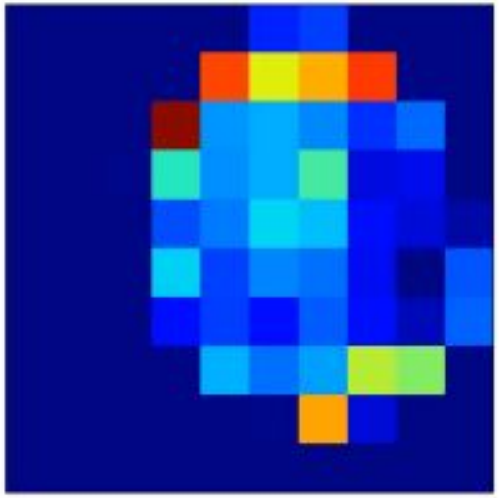


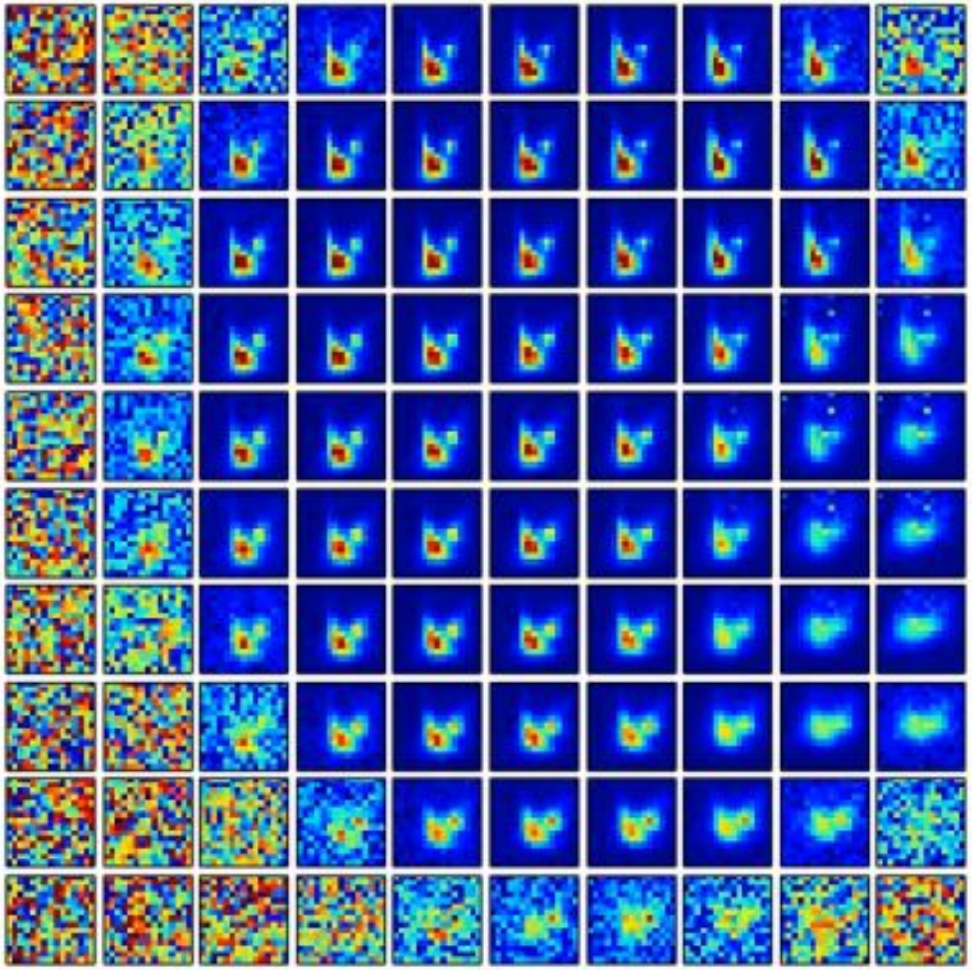


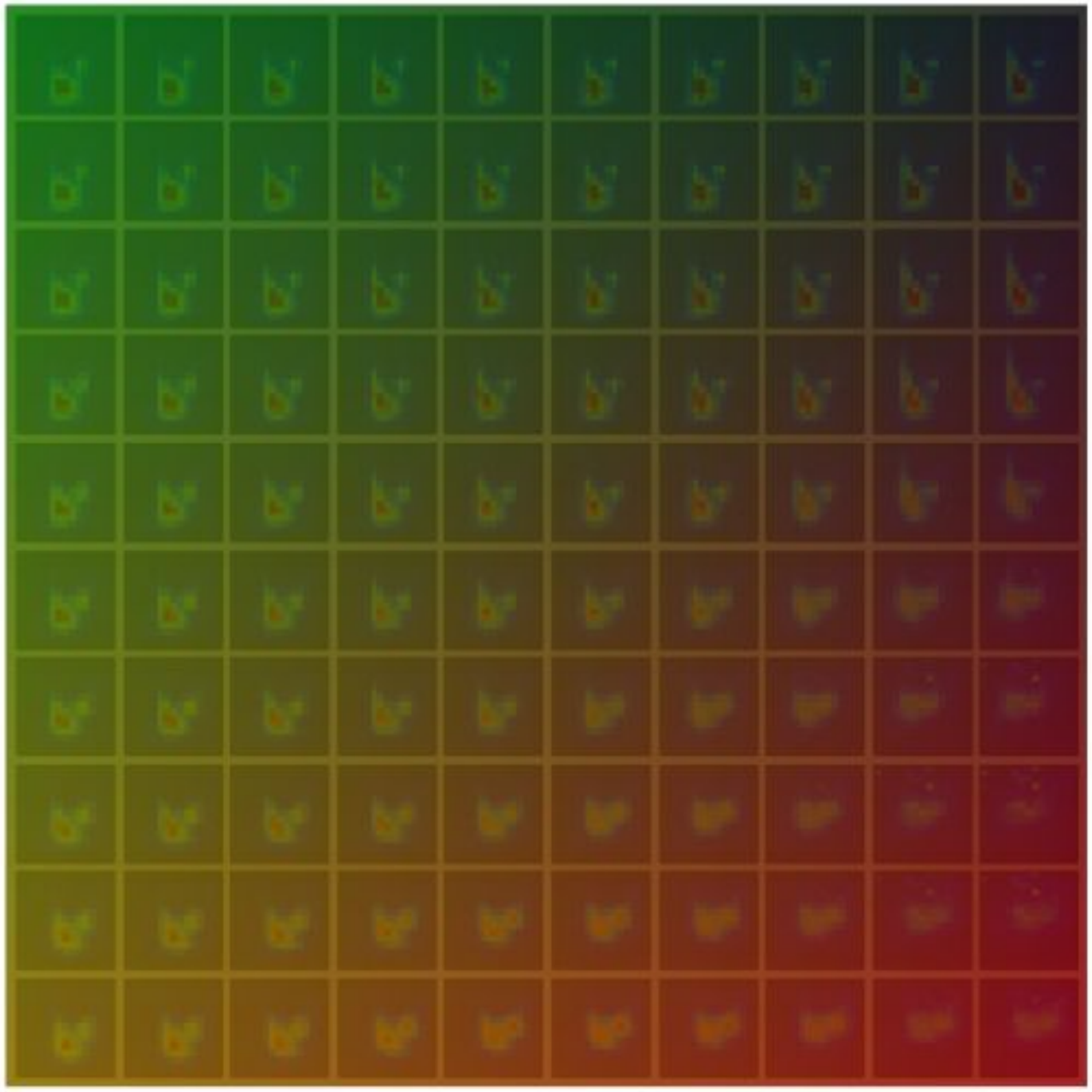
Dimension Reduction / SOM

$$\Delta(A,B)=\sqrt{\sum_{x=1}^{D_x}\sum_{y=1}^{D_y}(A_{xy}-B_{xy})^2}$$

$$\Delta(A,B)=\sqrt{\sum_{x=1}^{D_x}\sum_{y=1}^{D_y}\frac{\left(\frac{A_{xy}}{N_A}-\frac{B_{xy}}{N_B}\right)^2}{\frac{A_{xy}}{N_A^2}}}$$

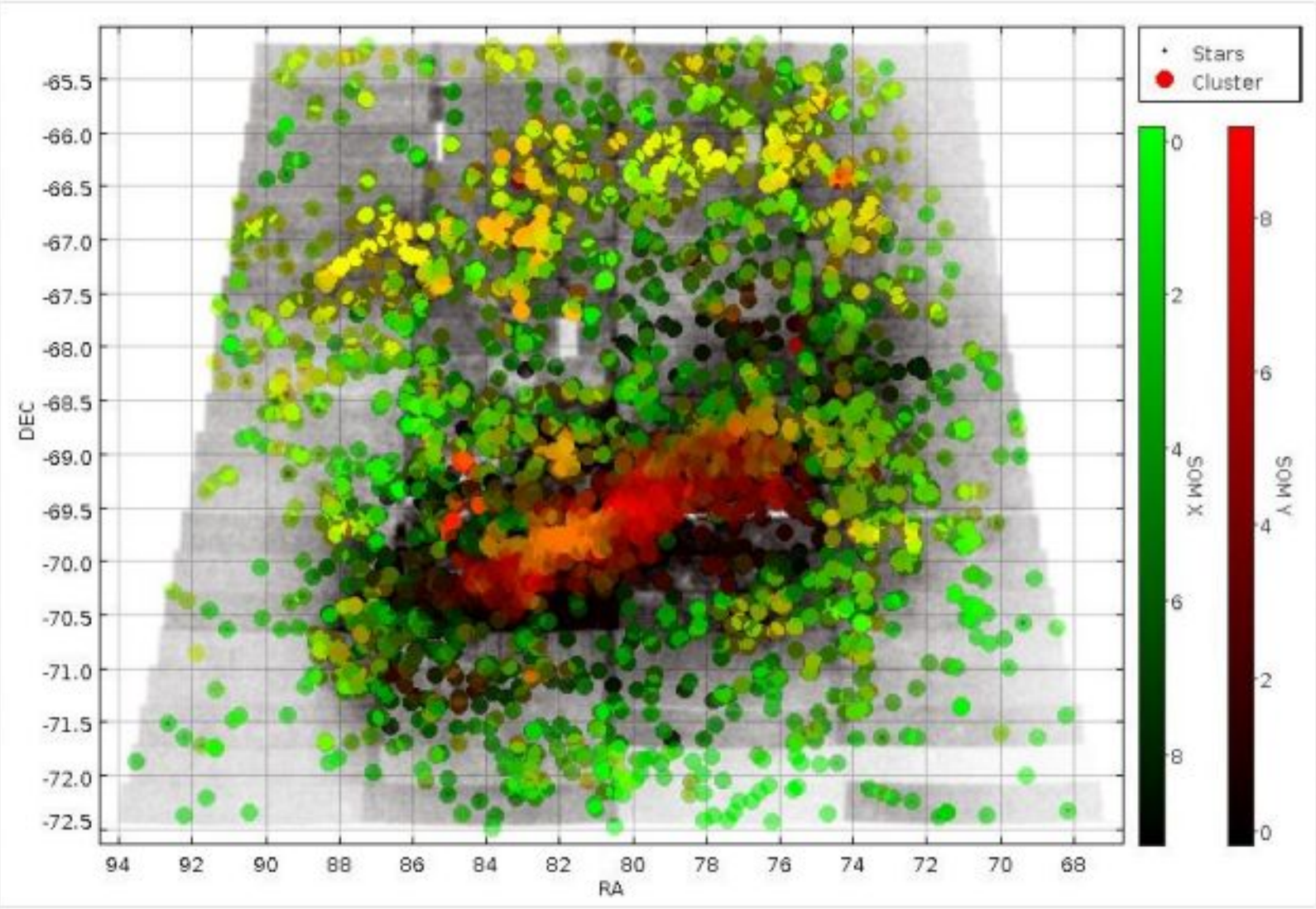






what is it good for?

Results


Stars
Cluster
DEC
RA
SOM X
SOM Y
-65.5
-66.0
-66.5
-67.0
-67.5
-68.0
-68.5
-69.0
-69.5
-70.0
-70.5
-71.0
-71.5
-72.0
-72.5
94
92
90
88
86
84
82
80
78
76
74
72
70
68
0
2
4
6
8

Conclusion

machine learning is not a magic tool that solves all your problems!

80% of all the work is pre-/post- processing


Thank you for your attention !
@AstroInformatix