
46
FIAT
YAMAHA
FASTWEB
agv
MOTUL
YAMAHA
motogp.com
official website

Machine Learning techniques for non-linear Regression and matrix inversion in the Sloan Digital Sky Survey

Michael Way (NASA/Goddard Institute for Space Studies)
Paul Gazis, Jeffrey Scargle (NASA/Ames, Space Sciences Division)
Ashok Srivastava (NASA/Ames Intelligent Systems Division)
Rama Nemani (NASA/Ames Earth Sciences Division)
Les Foster + Students (San Jose State University)

http://astrophysics.arc.nasa.gov/~mway/CESS2009.pdf

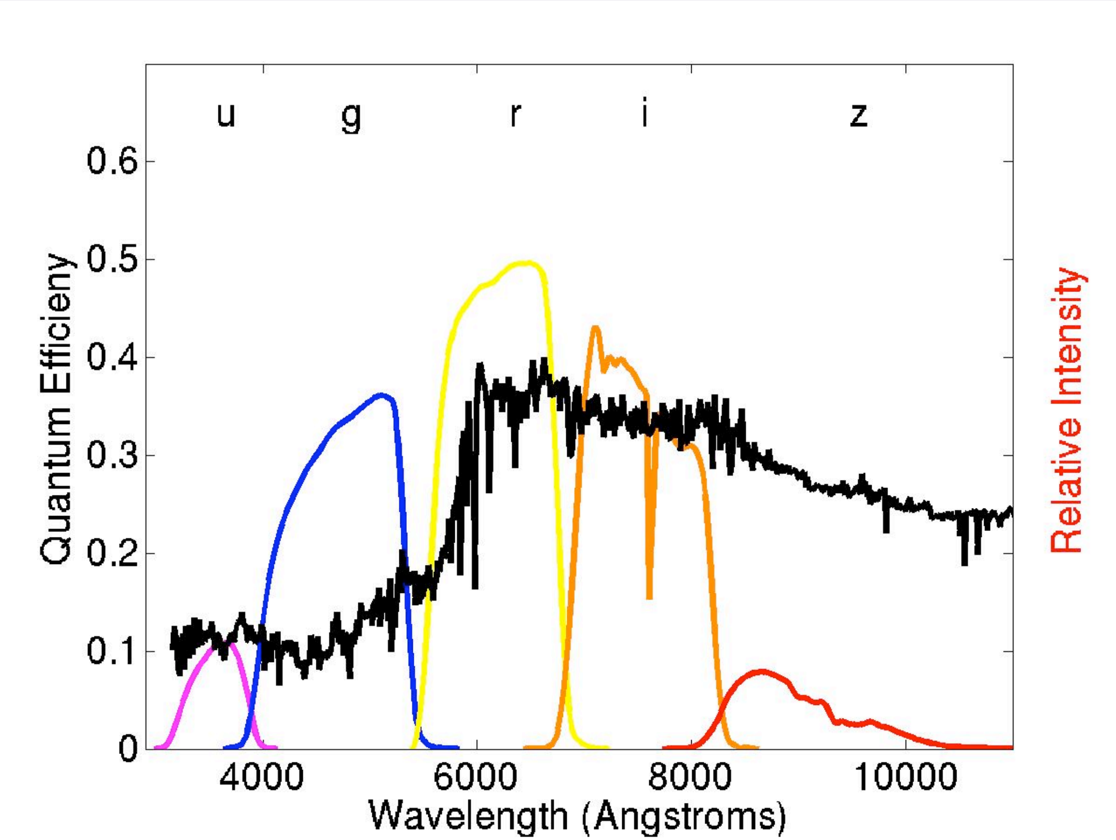
Outline

- What are Photometric Redshifts?
- Some common training set methods
- What is Gaussian Process Regression?
- Do different kinds of Kernels matter?
- How do I invert these huge non-sparse matrices?
- How many galaxies do I need?
- Some results

What are Photometric Redshifts?

Photometric Redshifts: A **rough** estimate of the redshift of a galaxy without having to measure a spectrum.

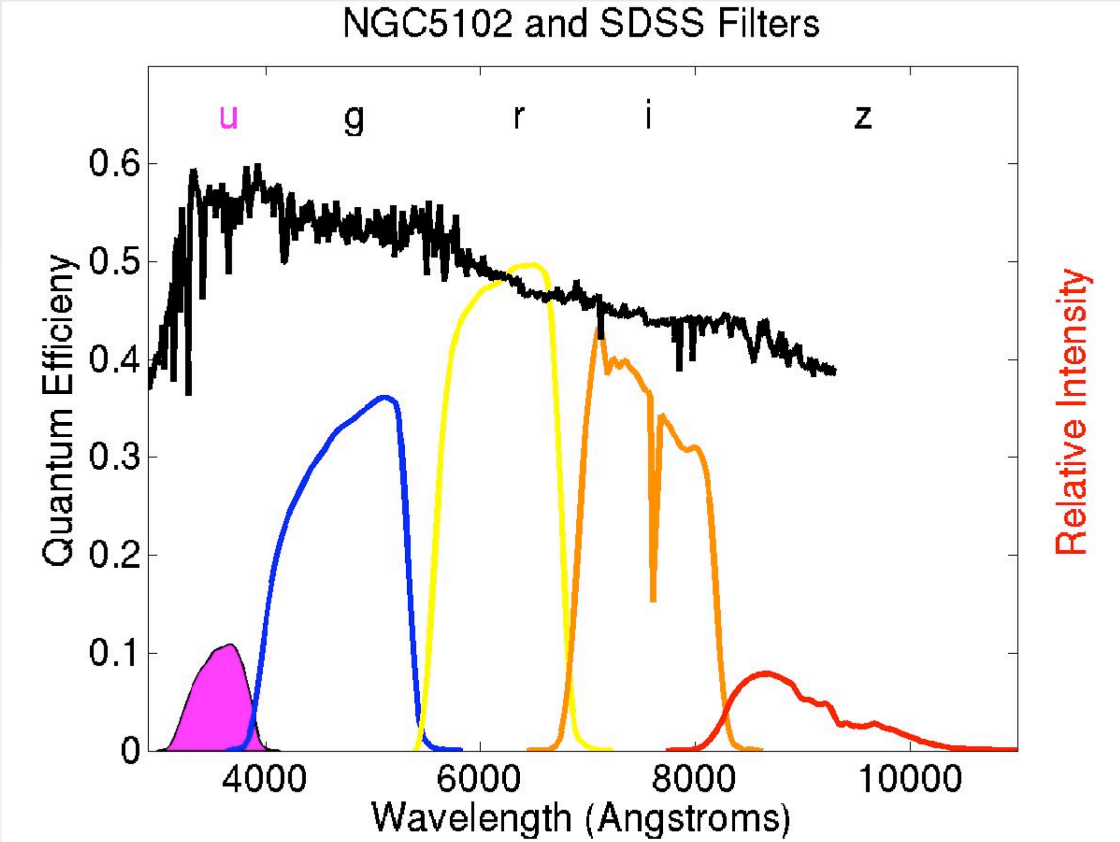
$$\mathbf{Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}} \qquad \mathbf{z_{photo}=z(C,m)}$$



What are Photometric Redshifts?

$Z_{spec} = (\lambda_{measured} - \lambda_{rest})/\lambda_{rest}$ $z_{photo} = z(C,m)$

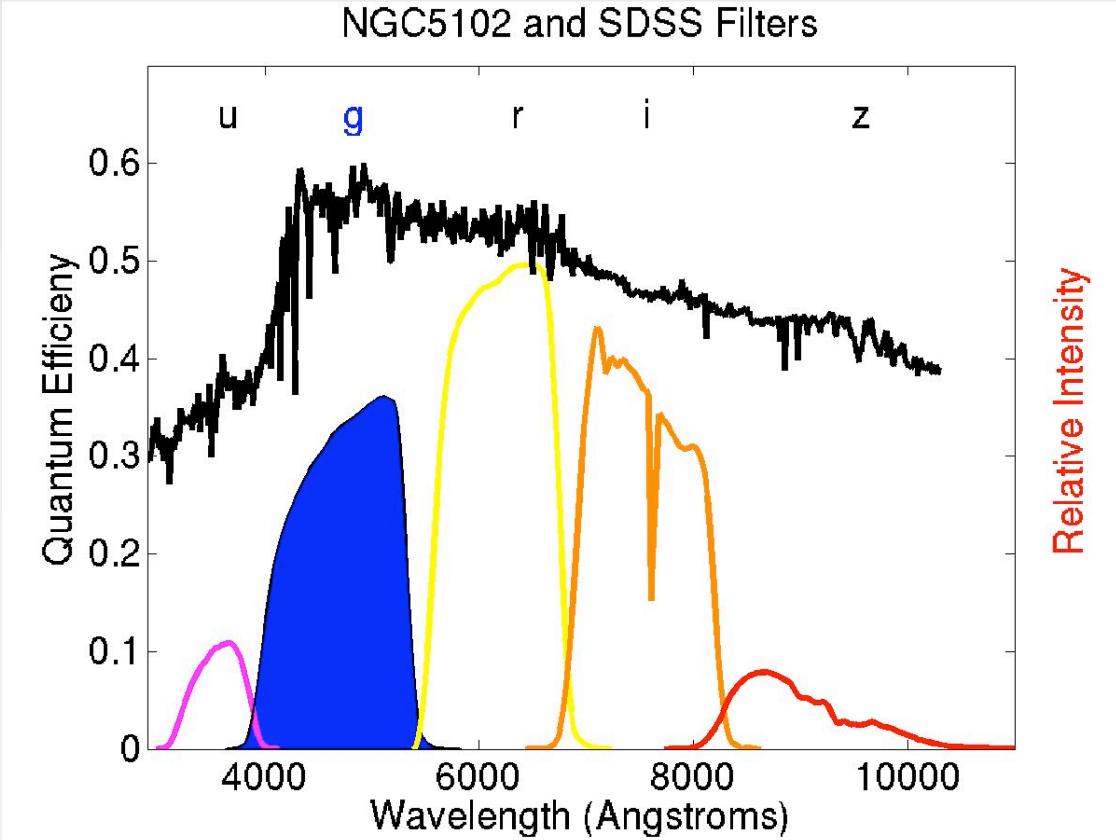
z=0.0



What are Photometric Redshifts?

$$Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest} \qquad z_{photo}=z(C,m)$$

z~0.06 (18000 km/s)

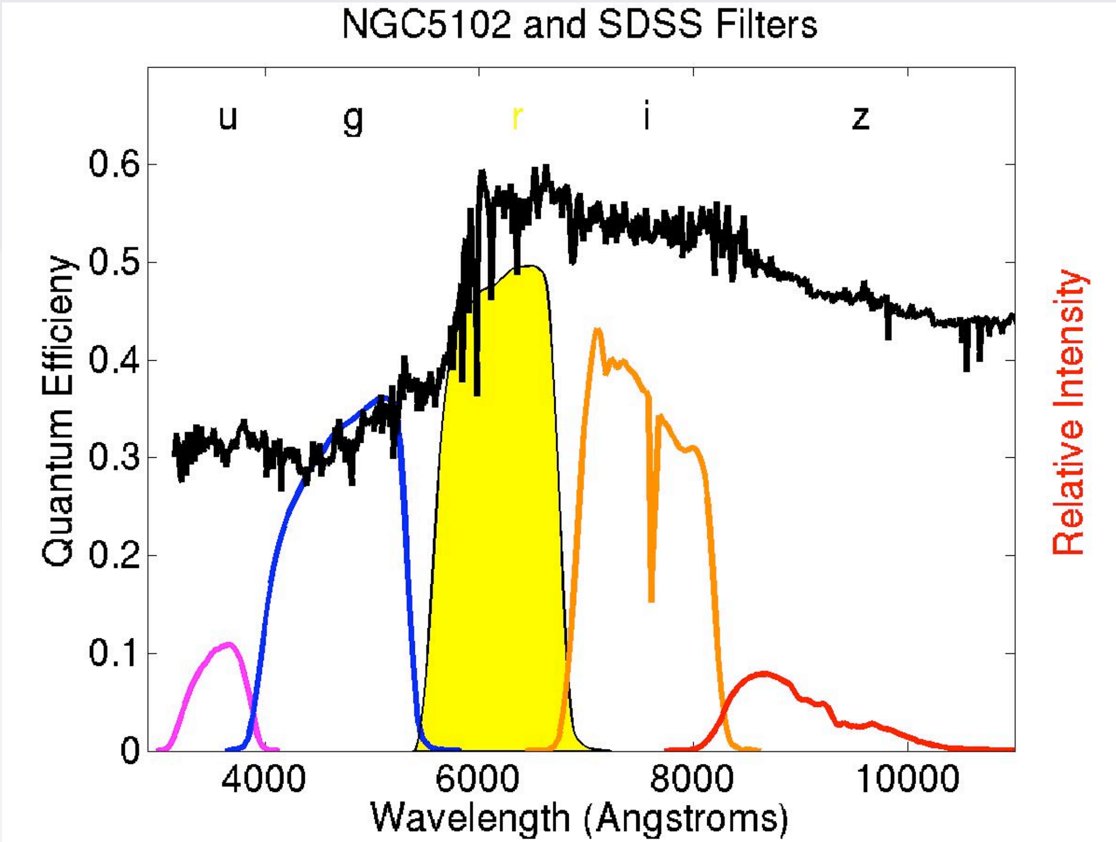


What are Photometric Redshifts?

$Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}$

$z_{photo}=z(C,m)$

z~0.6

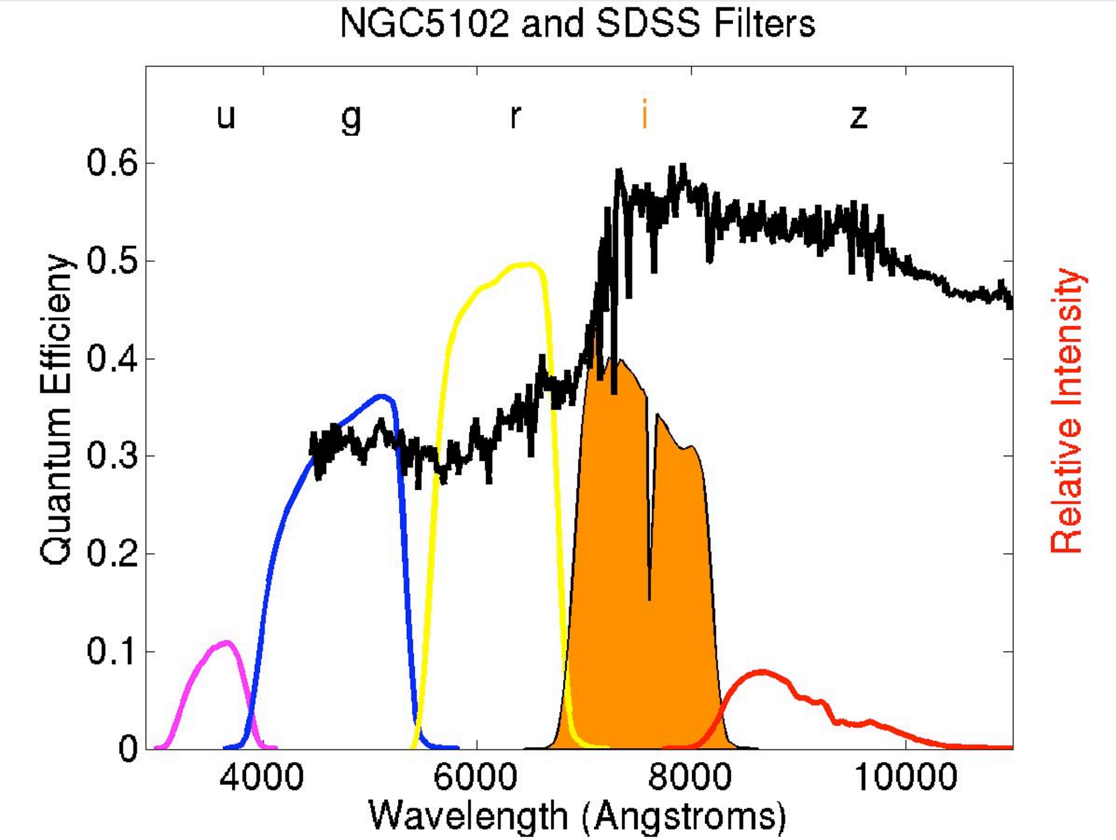


What are Photometric Redshifts?

$\mathbf{Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}}$

$z_{photo}=z(C,m)$

z~0.90

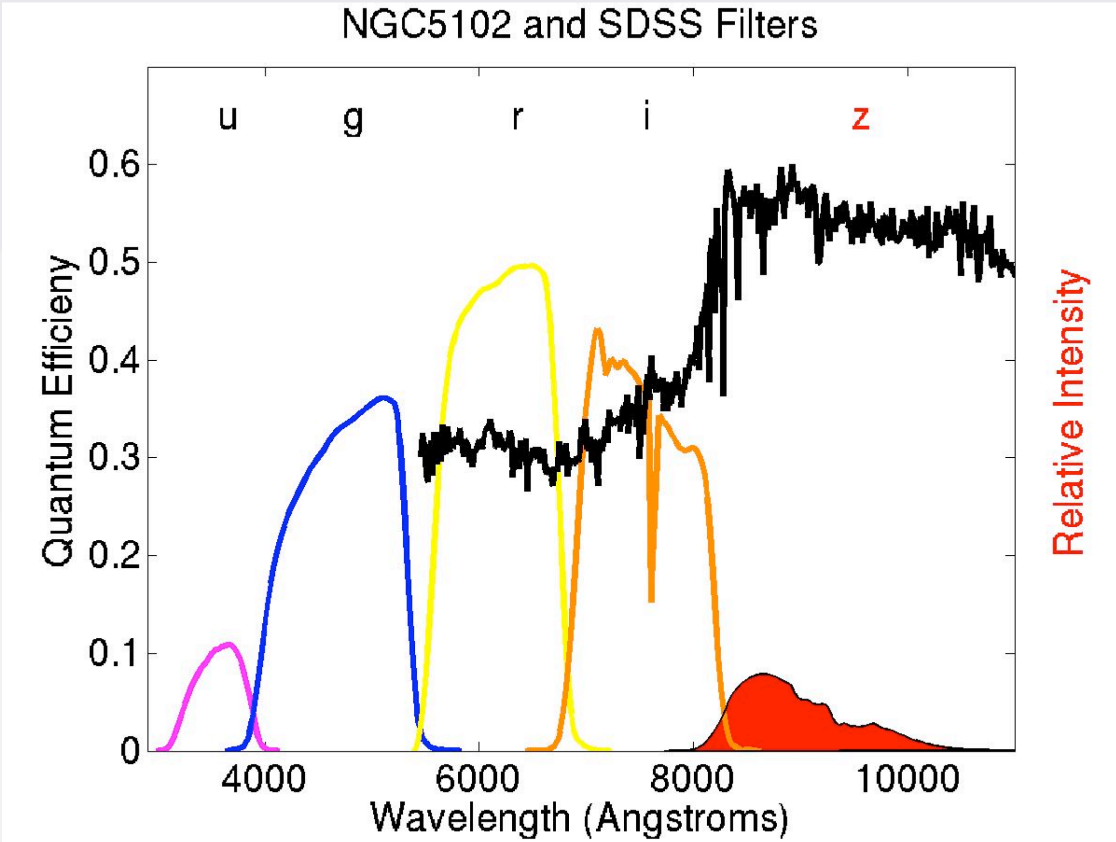


What are Photometric Redshifts?

$$Z_{spec}=(\lambda_{measured}-\lambda_{rest})/\lambda_{rest}$$

$$z_{photo}=z(C,m)$$

z~1.10



CESS 2009

Photo-z methods

1.) Spectral Energy Distribution (SED) Fitting:

- model based approach
- uses redshifts derived from spectra of artificial galaxies (e.g. Bruzual & Charlot)

2.) Training-Set methods:

- **empirical approach**
- **uses *spectroscopic* redshifts from a sub-sample of galaxies with the same band-pass filters**

Photo-z The Empirical Approach

Training Set Methods need a sub-sample of Galaxies:

1. of known spectroscopic redshift with a comparable range of **magnitudes (**u g r i z**)** as found in our Photometric survey objects

2. such that I can train u-g-r-i-z to predict redshift without actually knowing the redshift in a larger sample: u-g-r-i-z <–> redshift

- These will be our "Training Samples"

“Training Set” Methods

Galaxy Photometric Redshift Prediction History

u-g-r-i-z <–> redshift

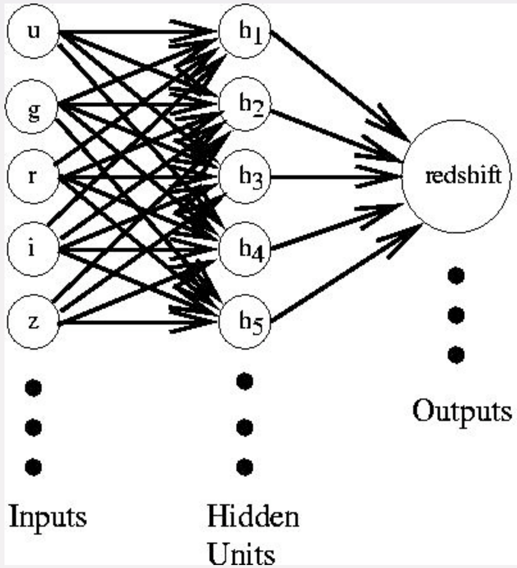
- Linear Regression was first tried in the 1960s
- Quadratic & Cubic Regression (1970s)
- Polynomial Regression (1980s)
- Neural Networks (1990s)
- Kd Trees & Bayesian Classification Approaches (1990s)
- Support Vector Machines & GP Regression (2000s)

Gaussian Process Regression fitting

Gaussian Process Regression ⇔ Kernel Methods

Kernel Methods have replaced Neural Networks in much of the Machine Learning literature

WHY?: given a large # of hidden units => GP (Neal 1996).



$h_n > 100$

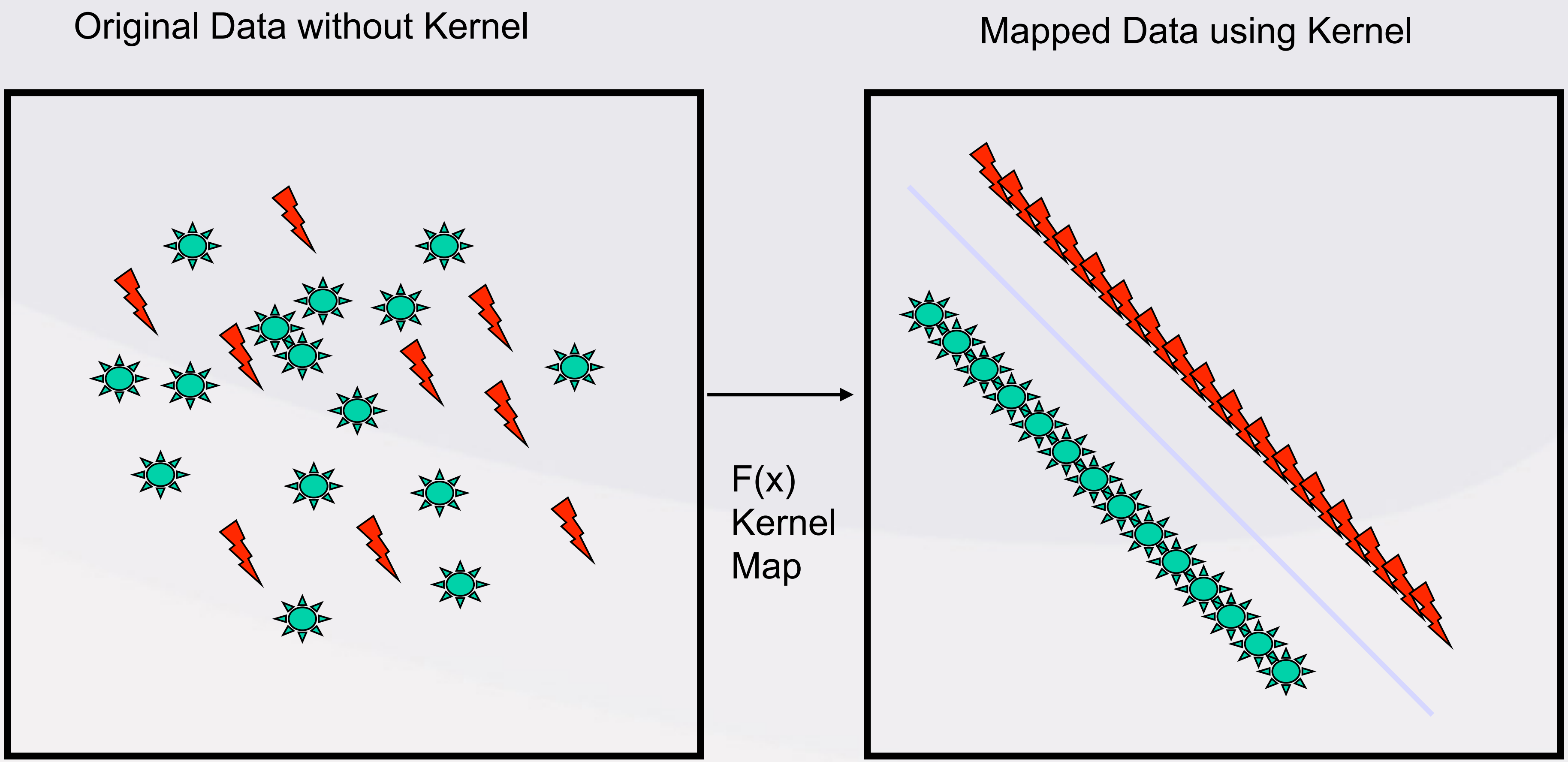
→→→→→

Kernel Methods - Gaussian Process Regression

GP regression builds a linear model in a very high dimensional *parameter space* ("feature space" → Hilbert space).

- One can map the data using a function F(x) [kernel] into this high (or infinite) dimensional *parameter space* where one can perform linear operations.

The value of kernels



Data in original space: highly complex decision boundaries.

Data in high dimensional feature space after mapping through F(x) can yield simple decision boundaries.

GP Regression (Kernels)

GP Advantages:

- Small input data training samples (good for higher redshifts?) yet low errors
- Realistic estimation of individual redshift errors

GP Disadvantages:

- Possibly large CPU time requirements (**Part I**)
  - The Kernel (Covariance Matrix) **can** be large: $K=(\lambda^2 I+XX^T)^2$ if $X=5x180,000$ (our case) then K is a matrix 180,000 x 180,000 and we have:

$$y^* = K^*(\lambda^2 I + K)^{-1} y$$

  - Need to invert this large K matrix - $O(N^3)$ operation
- Kernel Selection is ambiguous? (**Part II**)

Pick a transfer/covariance function (Kernel)

Matern Class Fcn

$$k(r) = \frac{2^{l-v}}{\Gamma(v)}\left(\frac{\sqrt{2vr}}{l}\right)^{v} J_v\left(\frac{\sqrt{2vr}}{l}\right)$$

$v \to \infty$

Radial Basis Fcn

$$k(r) = \exp\left(\frac{r^2}{2l^2}\right)$$

Rational Quadratic

$$k_{RQ}(r) = 1 + \left(\frac{r^2}{2\alpha l^2}\right)^{-\alpha}$$

Polynomial

$$k(x,x') = \left(\sigma_o^2 + x^T \sum_p x'\right)^p$$

Neural Nets

$$k_{NN}(x,x') = \frac{2}{\pi}\sin^{-1}\left(\frac{2x^T\Sigma x'}{\sqrt{(1+2x^T\Sigma x)(1+2x'^T\Sigma x')}}\right)$$

GPR: Part II

That matrix inversion…

With our SDSS (DR3) Main Galaxy spectroscopic sample (180,000 galaxies) the matrix size is 180,000 x 180,000

- Need a SSI supercomputer with a LOT of ram and cpu time?
- One can take a random sample of ~1000 galaxies & invert that while bootstrapping n times from full sample [**Paper I**]
- **How about some *low-rank matrix approximations*?** Remember these are **non-sparse** matrices! [**Paper II**]

Matrix Inversion

We would prefer to try and invert a larger fraction of that matrix. 1000 galaxies isn't much to train on.

- 32bit computers are limited to 2GB of RAM. O(1000 x 1000)

- Reasonably priced 64bit desktops can be purchase with upwards of 16GB of RAM O(20,000 x 20,000)

Matrix Inversion

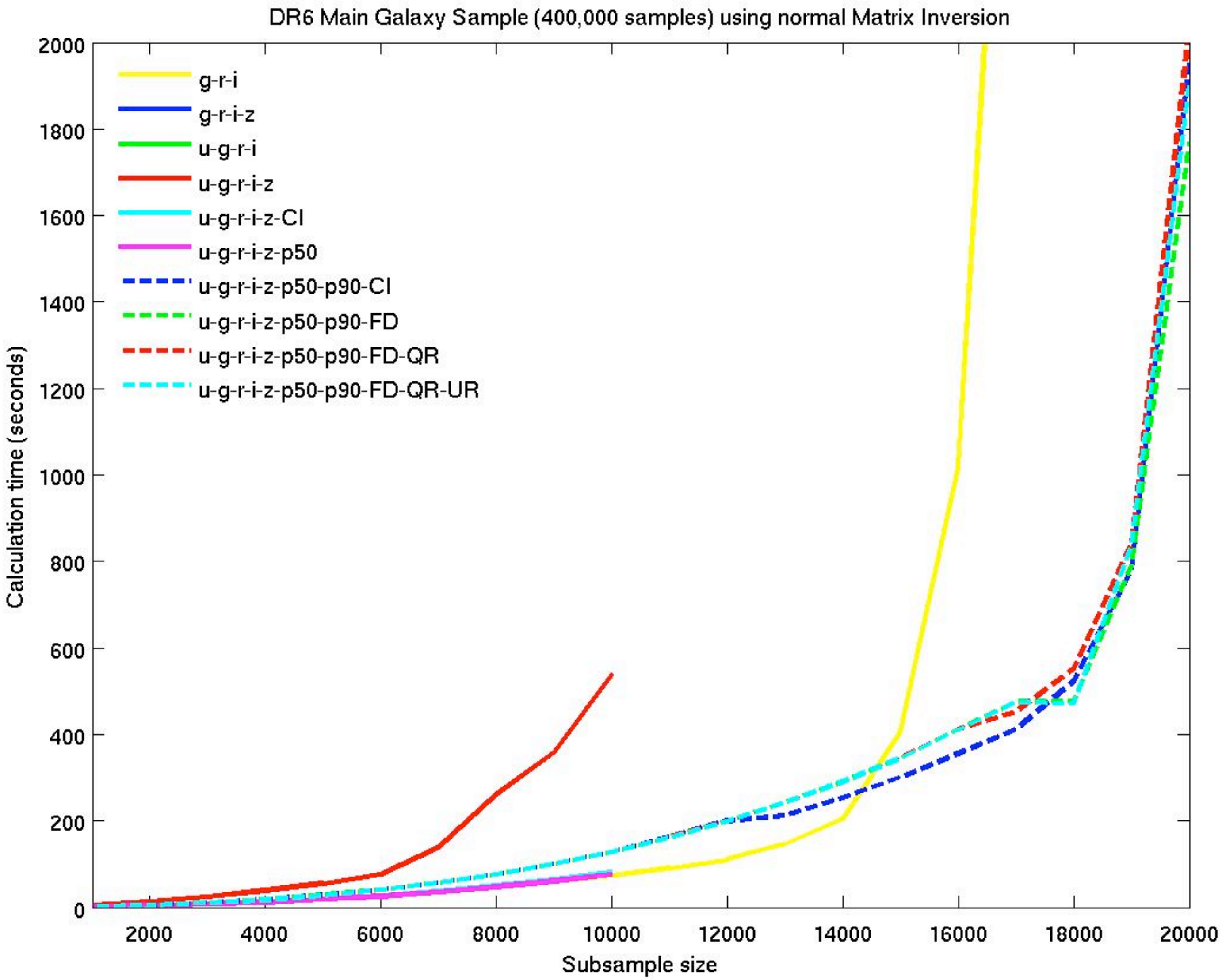
Low-rank approximation works well?

- Cholesky Decomposition, Subset of Regressors, Projected Process Approx, etc…
- With Les Foster + Students (SJSU) we've written a paper (JMLR) to invert these non-sparse matrices with sizes of 80000 x 80000 and above
  - To be published in a few months (referee process completed)

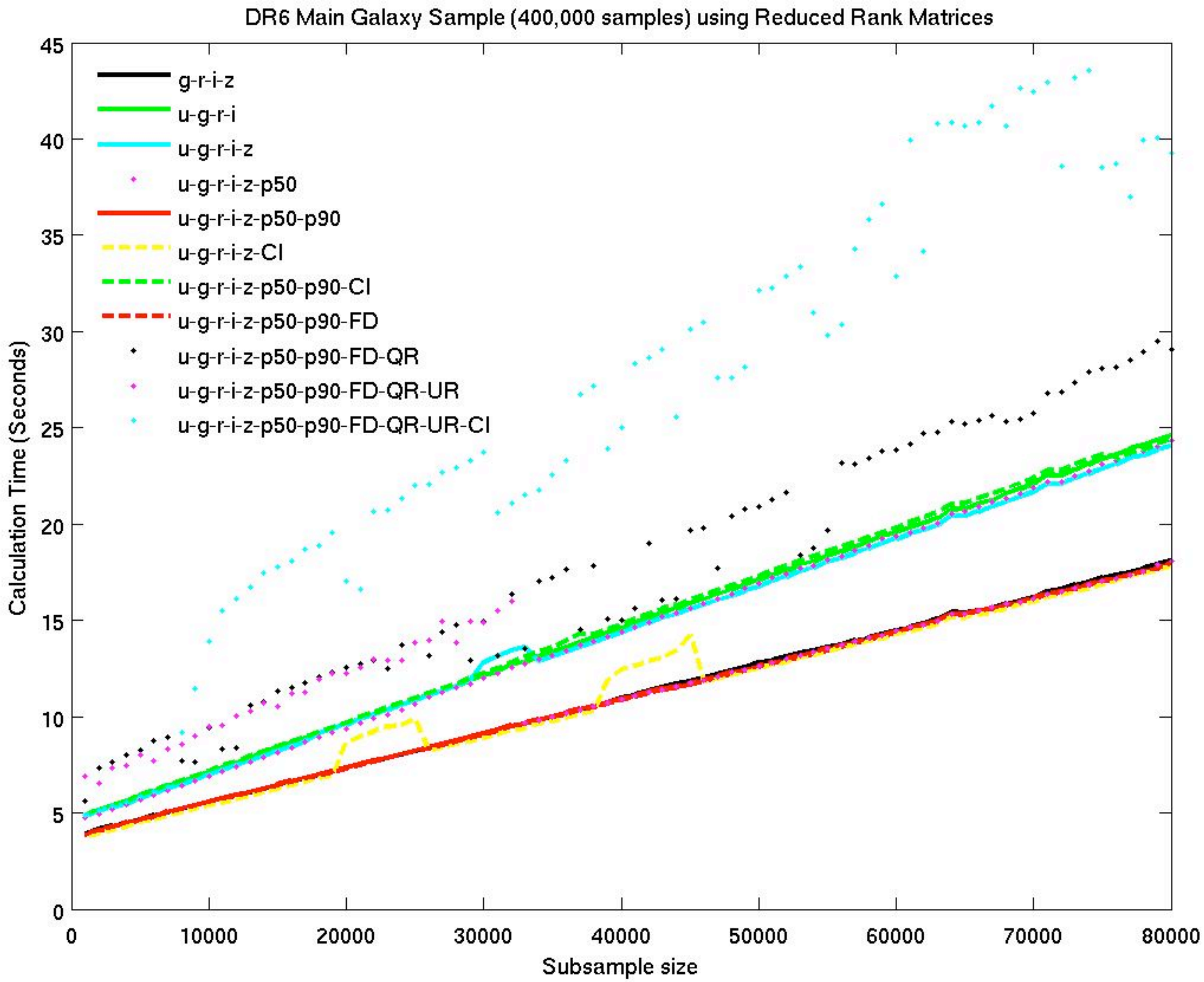
Time to Invert!

- TIME & MEMORY: These rank-reduction methods are much much faster than brute force inversion and use much less RAM
- We also discovered we only need ~30,000 points to gain enough training samples to minimize our rms error in the SDSS dataset

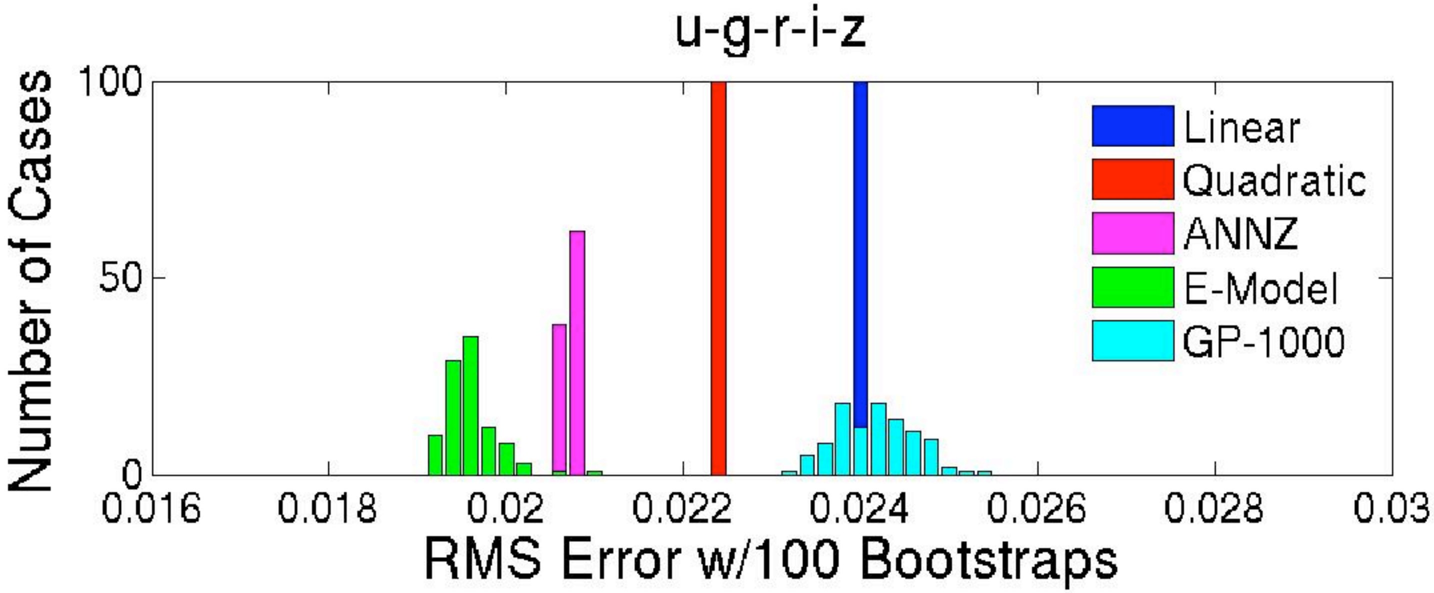
Normal Matrix Inversion Timing Results

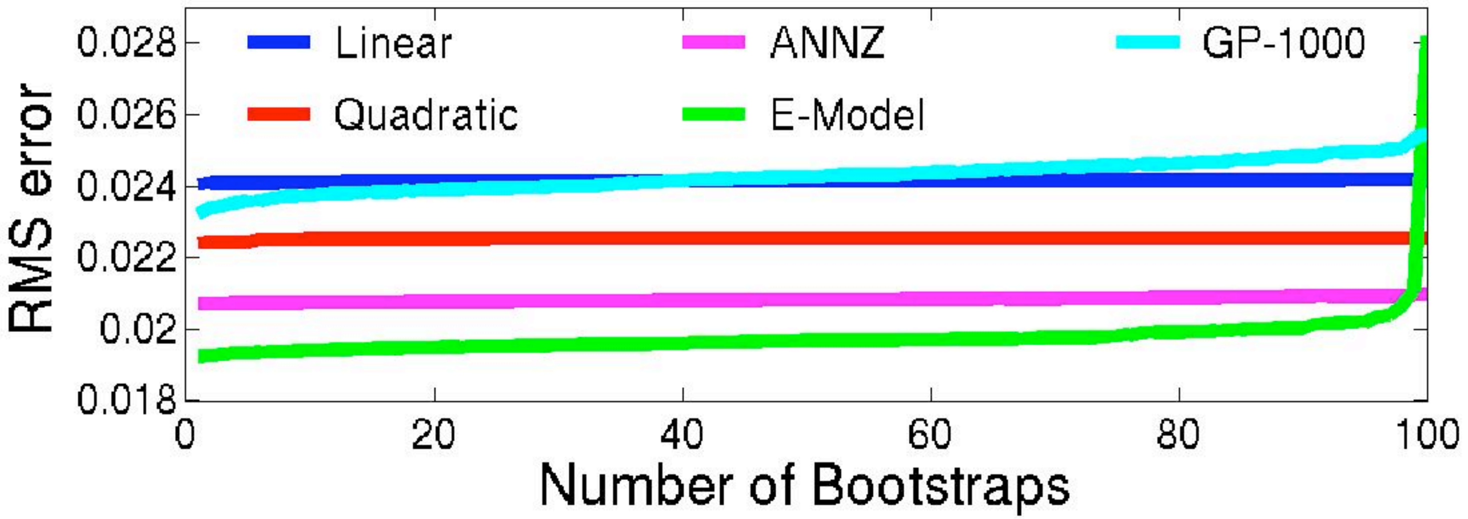


Reduced Rank Timing Results



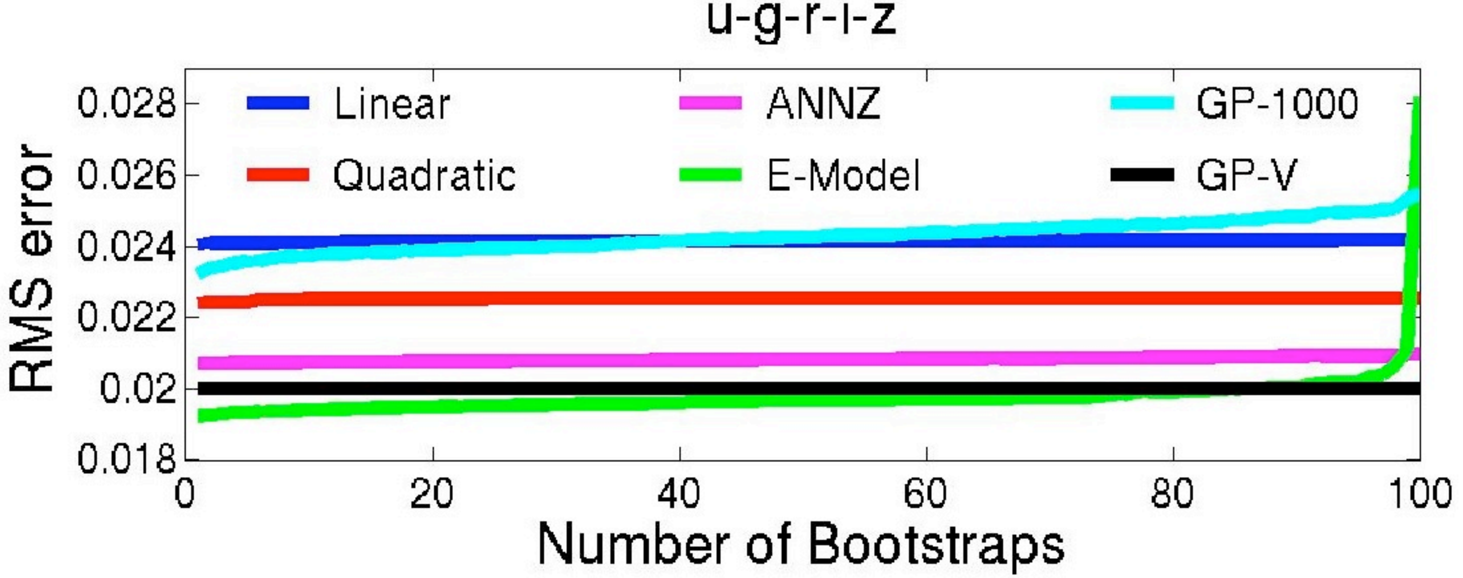
Paper I Results: Comparing Methods

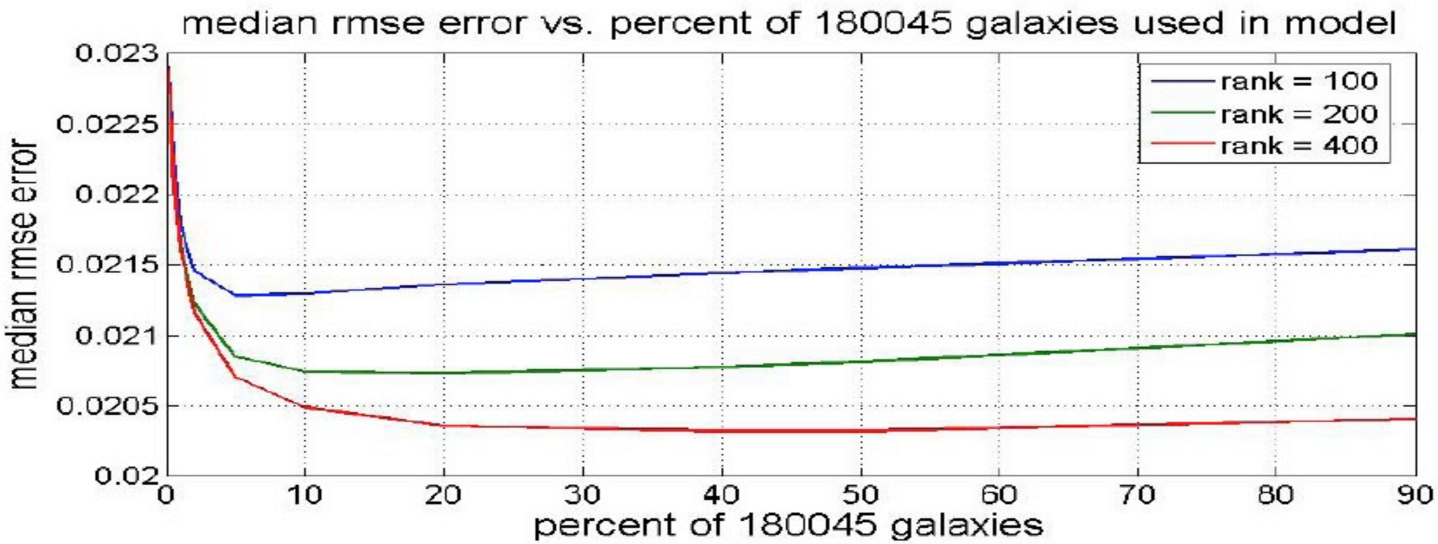




Latest Results: Comparing Methods

➔ GP-V: Rank=1000 for 36000





🡐 Beyond 20% (~36000) Rank 400 is fairly flat

Conclusions

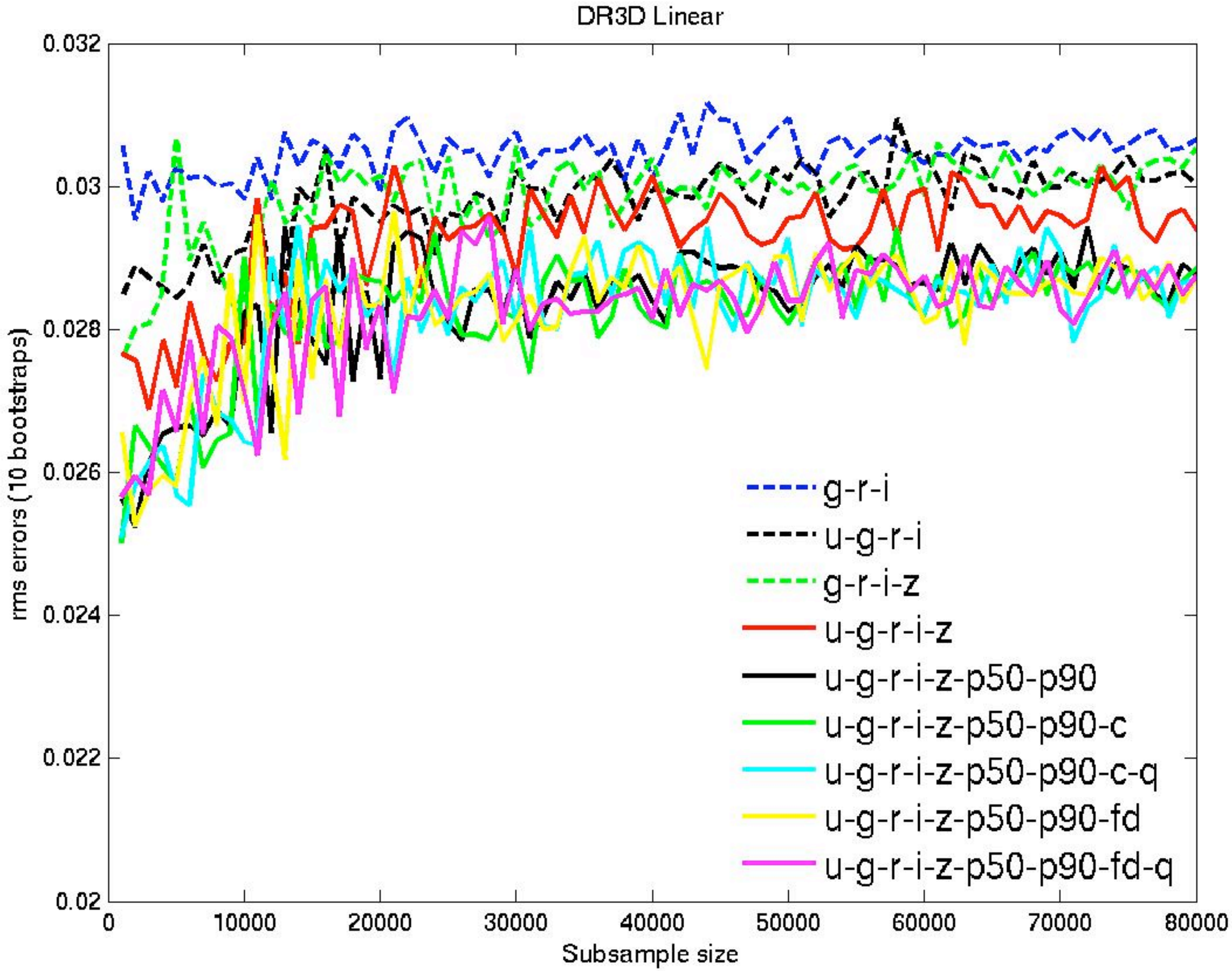
- Gaussian Process Regression is superior to most other regression methods for Photo-z
- GPR is feasible for large datasets with modern matrix inversion techniques
- **Interdisciplinary**: This was made possible by the cooperation of Earth Scientists, Astronomers, Machine Learning people and Mathematicians in 4 different groups.

The End

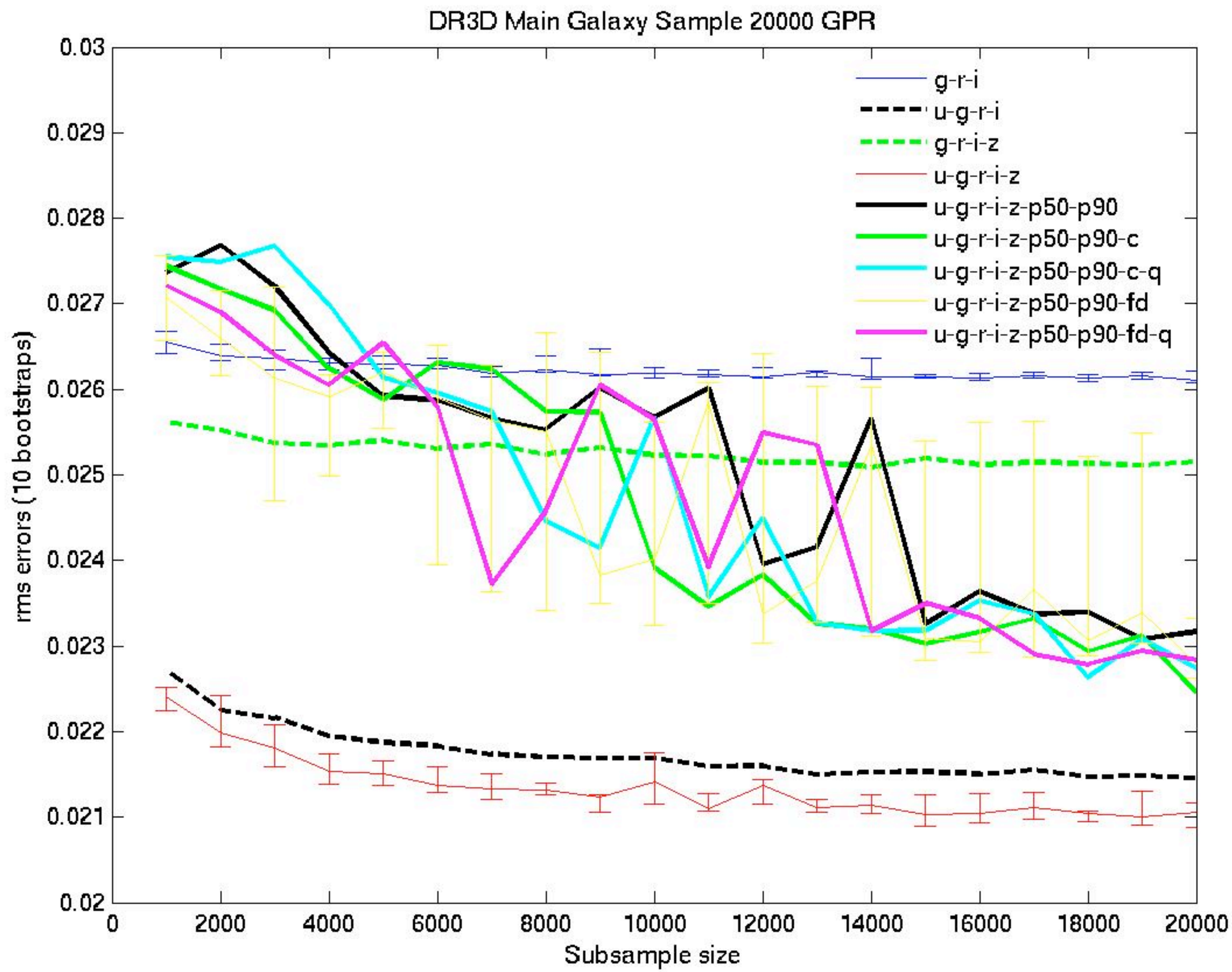
Thank You

http://www.giss.nasa.gov/cess2009/presentations/way.pdf

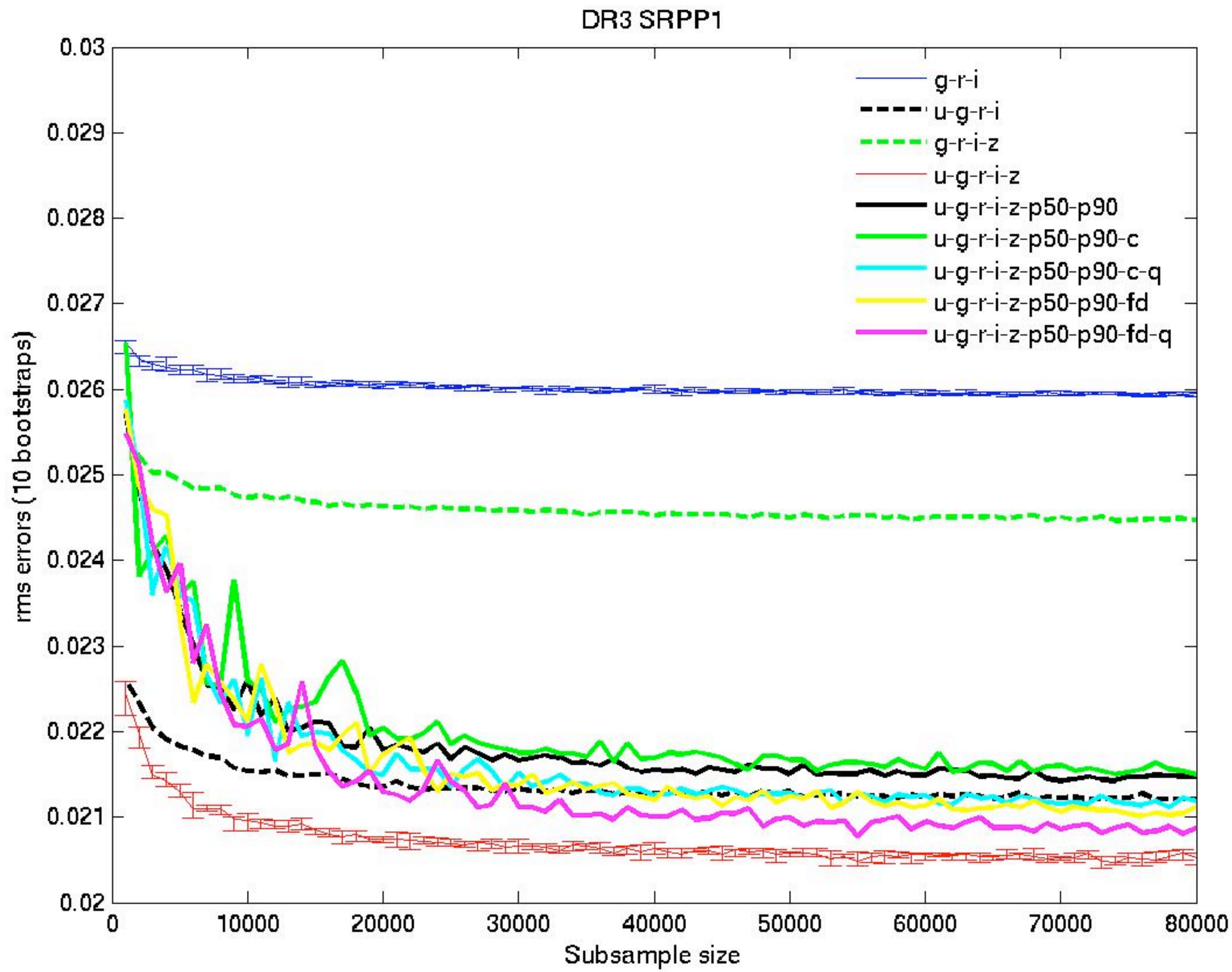
Main Galaxy Sample 80000 Linear



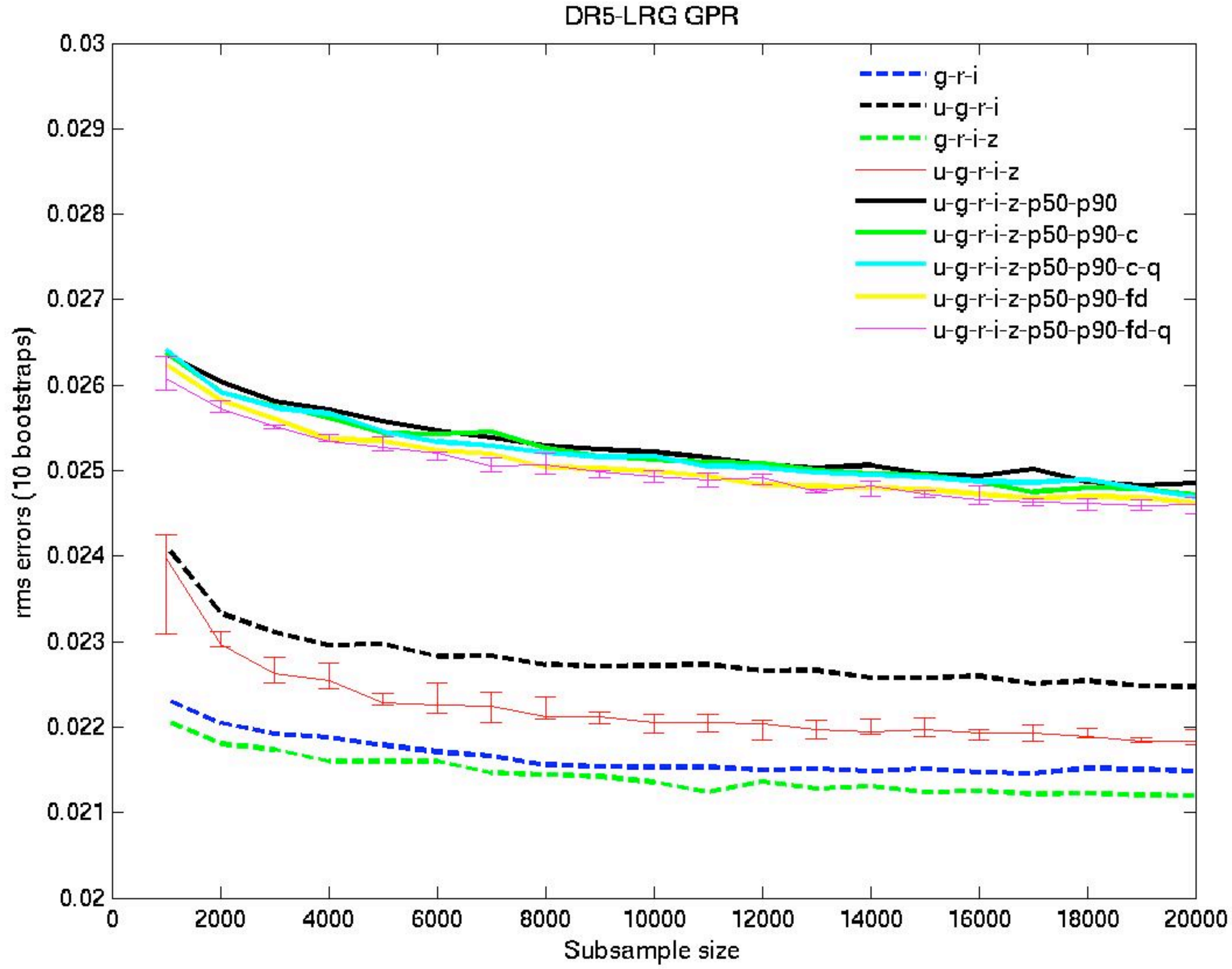
Main Galaxy Sample 20000 GPR



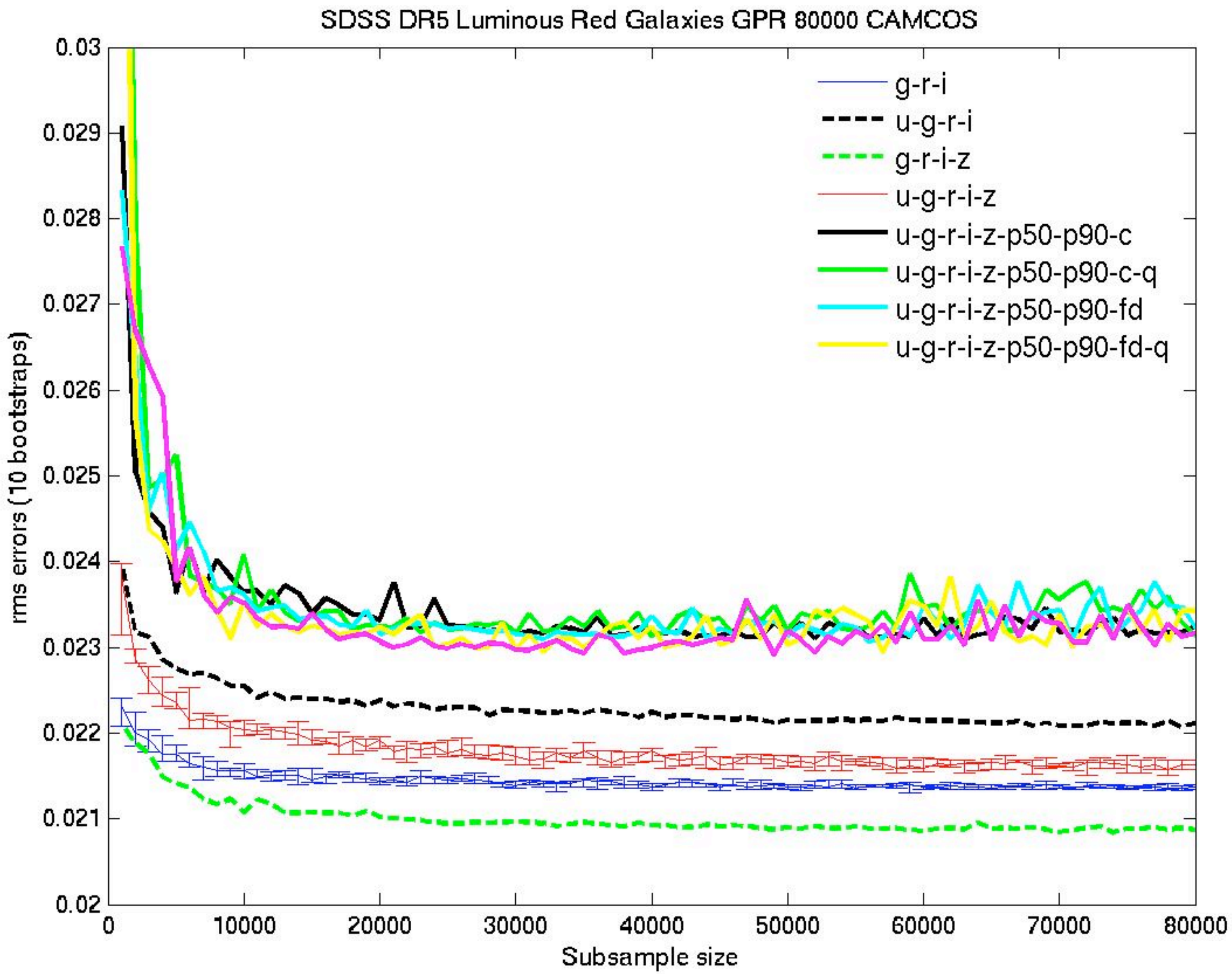
Main Galaxy Sample 80000 GPR



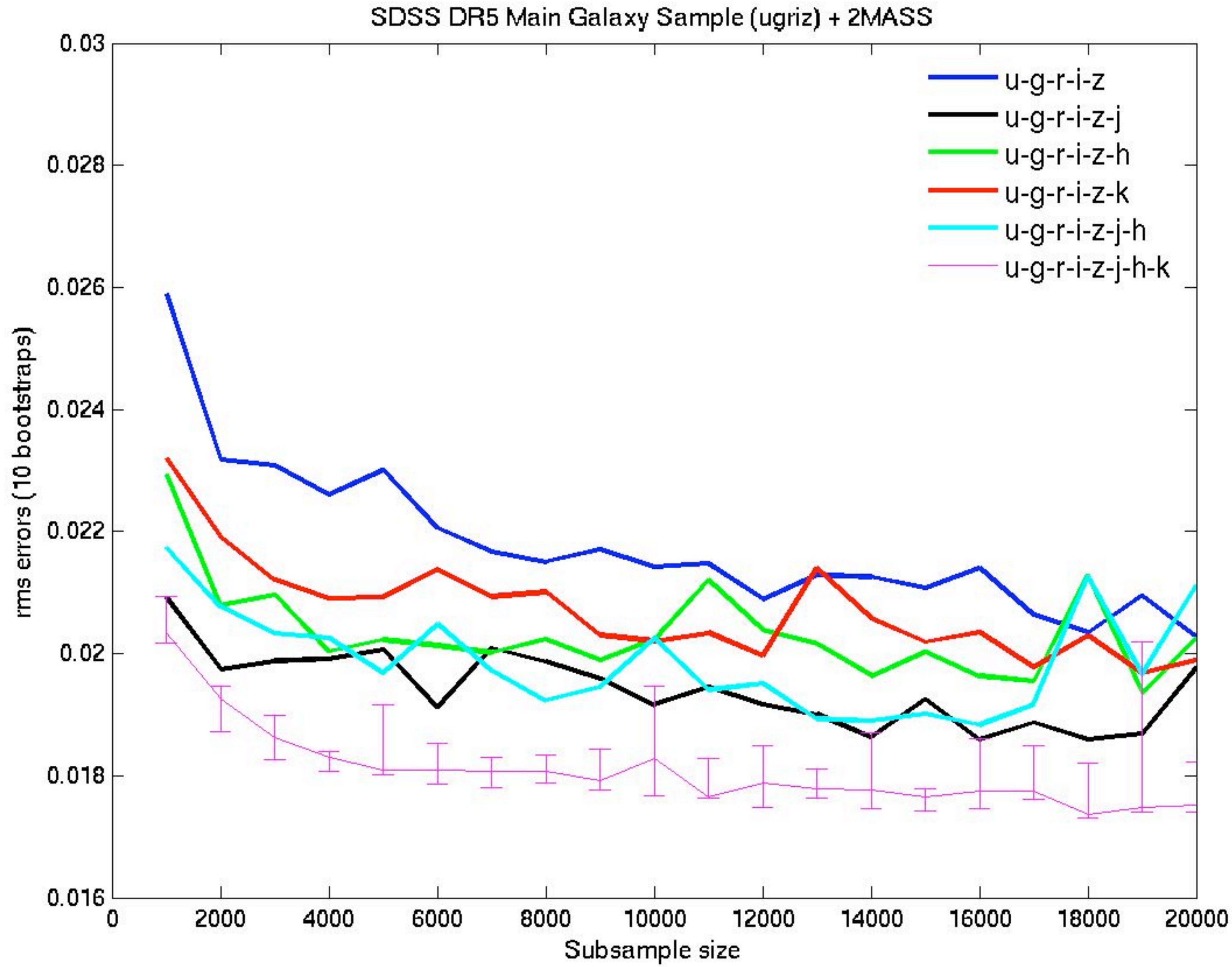
Luminous Red Galaxies 20000 GPR



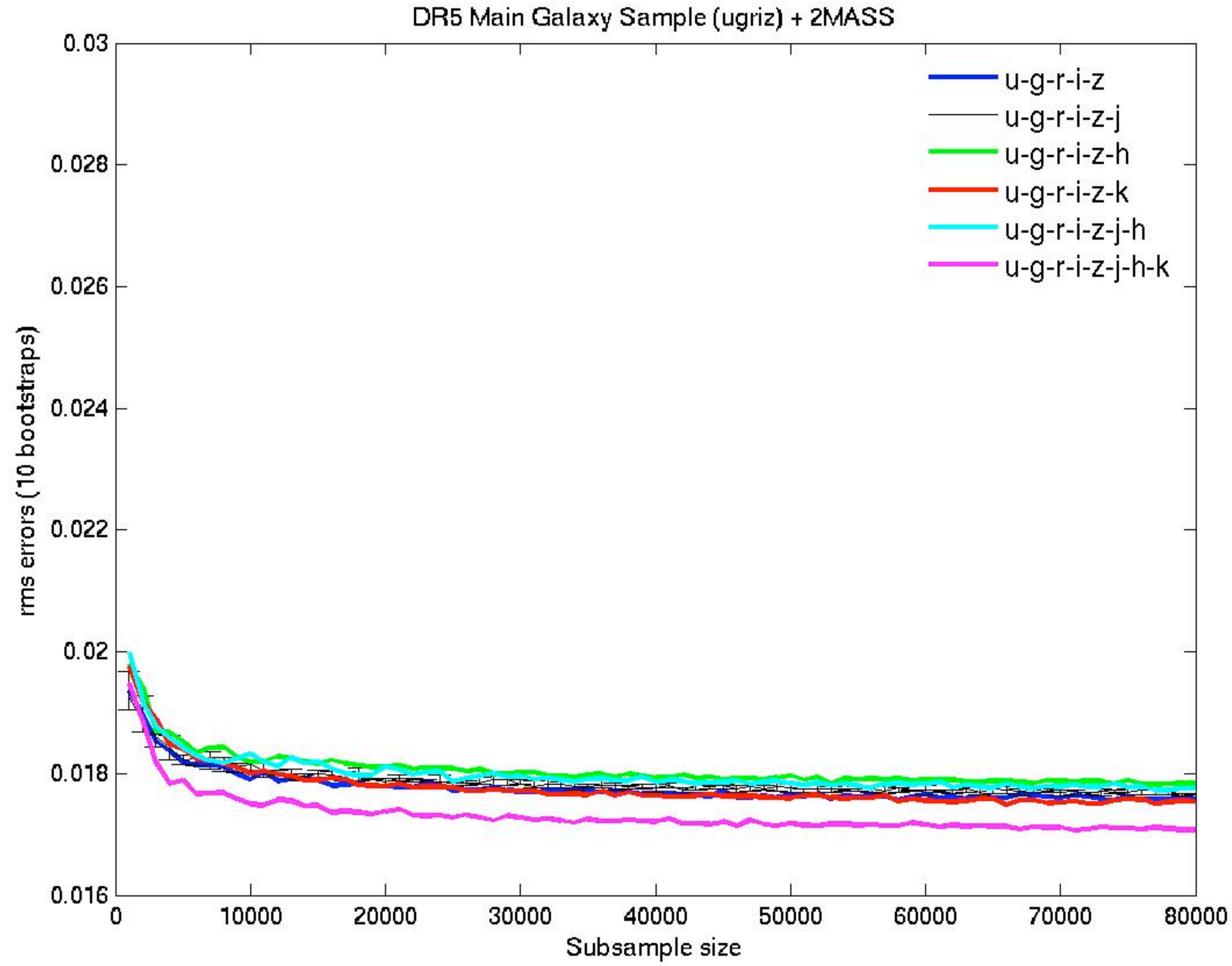
LRG 80000 GPR



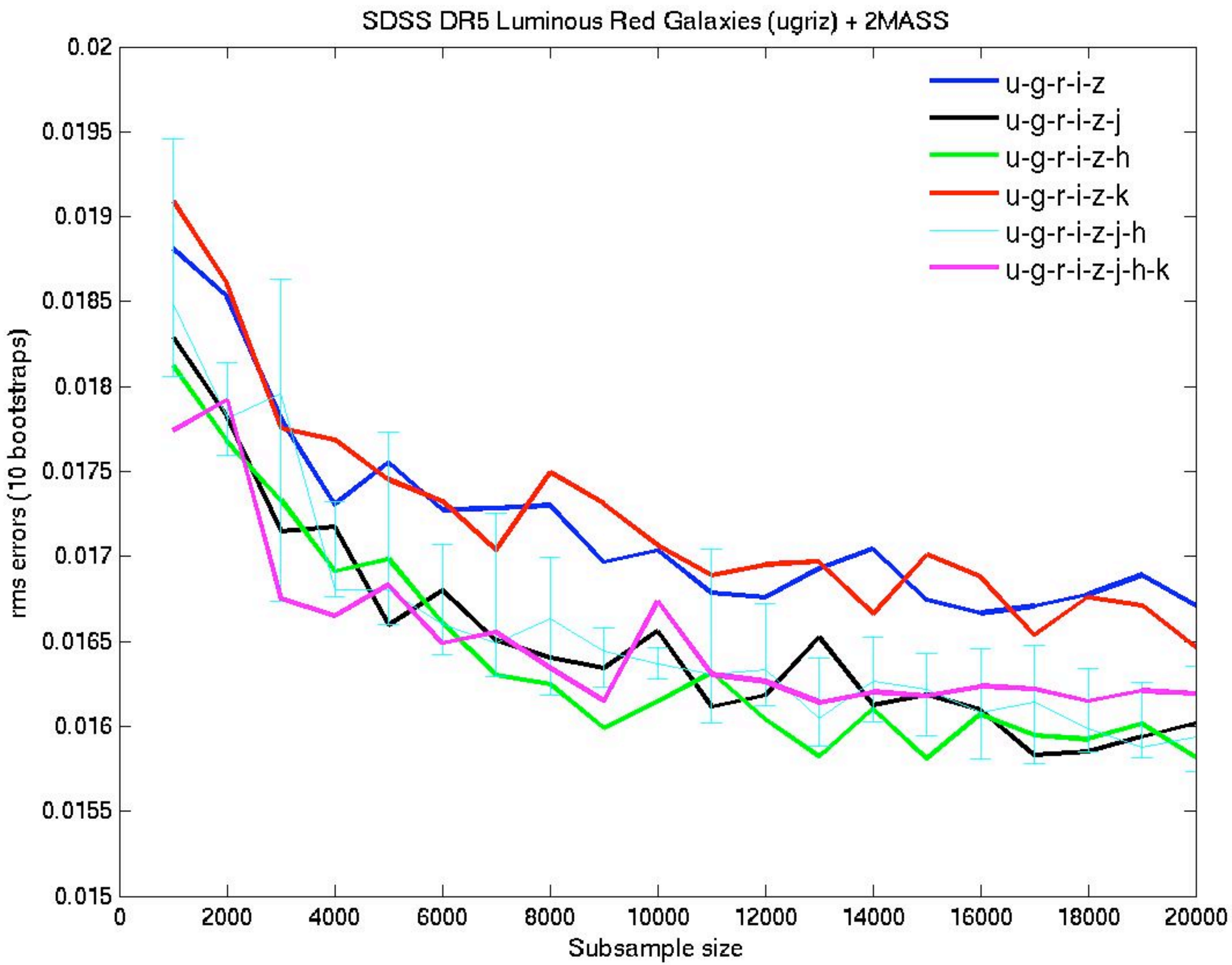
MGS + 2MASS 20000



MGS + 2MASS 80000



LRG+ 2MASS 20000



LRG+ 2MASS 80000