

New Approaches To Gaussian Process Regression In The Sloan Digital Sky Survey

M. J. Way^{1,2}

NASA Goddard Institute for Space Studies, 2880 Broadway, New York, NY 10029, USA

L. Foster

San Jose State University, Department of Mathematics, One Washington Square, San Jose, CA 95192, USA

P. R. Gazis

NASA Ames Research Center, Space Sciences Division, MS 245-6, Moffett Field, CA 94035, USA

A. N. Srivastava

NASA Ames Research Center, Intelligent Systems Division, MS 269-4, Moffett Field, CA 94035, USA

ABSTRACT

Expanding upon the work of Way & Srivastava (2006) we demonstrate how the use of training sets of comparable size continue to make Gaussian Process Regression a competitive and in many ways a superior approach to that of Neural Networks and other least-squares fitting methods. This is possible via new matrix inversion techniques specifically for Gaussian Processes that does not require that the kernel matrix be sparse. This development, combined with a neural-network kernel function appears to give superior results for this problem.

We also demonstrate that there appears to be a minimum number of training set galaxies needed to obtain the optimal fit when using our Gaussian Process Regression rank reduction methods. We also find that morphological information included with many photometric surveys appears, for the most part, to make the photometric redshift evaluation slightly worse rather than better. This would

¹NASA Ames Research Center, Space Sciences Division, MS 245-6, Moffett Field, CA 94035, USA

²Department of Astronomy and Space Physics, Uppsala, Sweden

indicate that morphological information simply adds noise from the Gaussian Process point of view.

In addition we show that cross-match catalog results involving 2MASS and SDSS or GALEX have to be evaluated in the context of the resulting cross-match magnitude and redshift distribution. Otherwise one may be misled into overly optimistic conclusions.

Subject headings: Photometric Redshifts, Sloan Digital Sky Survey, Gaussian Processes, Neural Networks

1. Introduction

General approaches to calculating Photometric Redshifts from Broad Band Photometric data have been discussed elsewhere recently (Way & Srivastava 2006, hereafter Paper I). These involve Template based approaches and what are referred to as Training-Set approaches. In this paper we expand upon the training-set approaches outlined in Paper I using Gaussian Processes (GPs). Previously we were limited to training set sizes of order 1000 because a matrix inversion of order 1000x1000 was required for calculating the GPs. Part of the limitation was due to the amount of single thread accessible RAM on our circa 2005 32bit computers, meaning that one could not invert a matrix larger than about O(1000x1000) in size at one time within Matlab¹, our choice for implementing GPs. Today one can now use commodity based 64bit workstations and invert matrices of O(20000) within Matlab. However, even this is a small fraction of the total potential size of today’s Photometric Redshift training sets. For this reason we have developed new non-sparse rank-reduction matrix inversion techniques that allow one to use over 100,000 training samples. From this work we demonstrate that the new rank-reduction methods only require approximately 30-40,000 samples to get the optimal possible fit from GPs on Sloan Digital Sky Survey (York et al. 2000, SDSS) data.

Since Paper I several new approaches to Galaxy Photometric Redshifts from Broad Band Photometry have come about along with expansion and refinement of previously published methods. Below is a summary of some of these approaches.

Kurtz et al. (2007) have used the Tolman Surface Brightness Test (μ -PhotoZ) using the relation $\mu \approx (1+z)^{-4}$ where μ is the galaxy surface brightness in the SDSS r band via the 50% Petrosian (1976) Radii (petroRad50_r): $\mu = \text{petroMag}_r + 2.5(0.798 + 2\log(\text{petroRad50}_r))$

¹<http://www.mathworks.com>

and the galaxy r-i colors to pick the red galaxies this method is intended for. The Petrosian Radii may add useful information because of the angular diameter distance relation. We also find this to be the case for GPs as discussed in Section 6 below.

Carliles et al. (2008) have used Random Forests (ensembles of Classification and Regression Trees) to estimate photometric redshifts from the SDSS. Like GPs (see Paper I) this method is also supposed to give realistic individual galaxy photometric redshift error estimates and few or no catastrophic failures. Ball et al. (2008) continue their work using Machine Learning methods to derive photometric redshifts for galaxies and quasars using the SDSS and The Galaxy Evolution Explorer (GALEX, Martin et al. 2005)². In particular, they have made good progress in eliminating catastrophic failures in Quasar photo-z estimation while bringing down the root-mean-square values.

Wray & Gunn (2007) have taken a Bayesian approach using the SDSS apparent magnitude colors u-g, g-r, r-i, i-z, surface brightness μ_i in the i band, the Sérsic n-index (Sérsic 1968), and the Absolute Magnitude M_i “corrected” to $z=0.1$. Some of these quantities are only available from The New York University Value Added Catalog (NYC-VAGC) of Blanton et al. (2005) or calculated from the raw photometry directly. Wang et al. (2007) have used Support Vector Machines (also see (Wadadekar 2005)) and Kernel Regression on a SDSS and Two Micron All Sky Survey (2MASS, Skrutskie et al. 2006)³ cross-match list.

D’Abrusco et al. (2007) utilize a Supervised Neural Network using a standard multilayer perception, but operated in a Bayesian framework on two different SDSS datasets. One uses the SDSS Data Release Five Adelman-McCarthy et al. (2007) Luminous Red Galaxy sample, and the other which they term the “General Galaxy Sample” includes all objects classified as “GALAXY”. They then break their sample up into two redshift ranges and after some interpolation fit to the residuals they obtain impressive results, especially for the Luminous Red Galaxy sample (see their Table 4). In a higher redshift study Stabenau et al. (2008) used surface brightness priors to improve their template based scheme for photometric redshifts in the VVDS (Le Fèvre et al. 2004) and GOODS (Giavalisco et al. 2004) surveys.

This certainly does not cover all of the recent work in this field, but is a representative sample to show the intense interest being generated because of near-future large-area multi-band surveys like The Large Synoptic Survey Telescope (LSST, Ivezić et al. 2008)⁴ and PanStarrs (Kaiser et al. 2002).

²<http://www.galex.caltech.edu/>

³<http://www.ipac.caltech.edu/2mass/>

⁴<http://www.lsst.org>

We have used a variety of datasets in our analysis which are discussed in § 2. Discussion of the photometric and spectroscopic quality of the datasets along with other photometric pipeline output properties of interest are found in § 3. The methods used to obtain photometric redshifts are in § 4. How to pick the optimal sample size, matrix rank, and inversion method in § 5. Results are in § 6 and Conclusions in § 6.5.

2. The Sloan Digital Sky Survey, The Two Micron All Sky Survey and The Galaxy Evolution Explorer Datasets

Most of the work herein utilizes the SDSS Main Galaxy Sample (MGS, Strauss et al. 2002) and the Luminous Red Galaxy Sample (LRG, Eisenstein et al. 2001) from the SDSS Data Release Three (DR3, Abazajian et al. 2005) and Data Release Five (DR5, Adelman-McCarthy et al. 2007). We include the DR3 to more easily make the present work comparable to that from Paper I. We utilize the DR5 to maximize the size of our cross-match catalogs.

For comparison with other work we have cross-matched the SDSS datasets with both the 2MASS extended source catalog and GALEX Data Release 4 (GR4)⁵ All Sky Survey photometric attributes. Our method of cross-matching these catalogs have not changed since Paper I except that we now cross-match against the SDSS DR5 instead of the DR3 to increase the size of our catalogs. Many aspects of the SDSS, 2MASS and GALEX surveys relevant to this work were described in Paper I and hence we will not repeat them here. The only new catalog included since Paper I is the SDSS LRG. The SDSS LRG sample is similar to the SDSS MGS except that it explicitly targets the Luminous Red Galaxies. These galaxies have a fairly uniform Spectral Energy Distribution (SED) and a strong 4000Å break which tend to make calculating photometric redshifts easier than for the MGS (e.g. Padmanabhan et al. 2005) since the training-set contains more homogenous Spectral Energy Distributions. Since these galaxies are among the most luminous galaxies in the universe and tend to be found in over dense regions (e.g. Clusters/Groups of Galaxies) they are also good candidates for mapping the largest scales in the universe. See (Eisenstein et al. 2001) for more details.

⁵<http://galex.stsci.edu/GR4>

3. Photometric and redshift quality, morphological indicators and other catalog properties

For SDSS photometric and redshift quality we follow much the same recipe as in Paper I. However, unlike Paper I we refrain from using SDSS photometry of the highest quality (what we referred to as “GREAT”) as we did not see any consistent improvements in our regression fits using this higher quality photometry. We stick with the SDSS photometric “GOOD” flags only: !BRIGHT and !BLENDED and !SATURATED. See Table 2 in Paper I for a description of the flags. We utilize the same photometric quality flags for the GALEX and 2MASS datasets as described in Paper I, Section 3. We incorporate the same SDSS morphological indicators as in our previous work (See Paper I, Section 3.5). The SDSS casjobs⁶ queries used to get the data are the same as those in the Appendix of Paper I except in the case of the Luminous Red Galaxies utilized herein which require `primitarget=TARGET_GALAXY_RED (p.primitarget & 0x00000020 > 0)` instead of `primitarget=TARGET_GALAXY (p.primitarget & 0x00000040 > 0)` for the Main Galaxy Sample.

Tables 1 and 2 contain a comprehensive list of the 6 Data sets used herein.

4. Improved Gaussian Process Methods

In this section we will discuss our investigation of different Gaussian Process transfer functions (Kernels), rank-reduction matrix inversion techniques and results showing that there is the possibility of an upper limit to the number of training-set galaxies needed in the SDSS. However, regarding the latter possibility this particular result should be viewed with caution. Although there have been recent suggestions that one may quantify the maximum number of galaxies required to obtain the optimal fit Bernstein & Huterer (2009) in practice what we see with the GPs could be an artifact of the algorithm itself. One needs to further explore building good “local” models, whereas at present the GPs (and Neural Networks) are global models.

In the Gaussian Process method utilized herein one would begin with a training set matrix X of dimensions $n \times d$ (n = number of galaxies, d = number of components such as broad band flux measurements). One would also have a target vector y of dimensions $n \times 1$ (y would contain the known redshift for each galaxy in our case). The testing data is in a $n^* \times d$ matrix X^* with target values in a matrix y^* consisting of $n^* \times 1$ redshifts. We wish to predict the value of y^* given X , y , and X^* . The prediction of y^* requires a covariance

⁶<http://casjobs.sdss.org>

Table 1. Data Sets 1-3

Data Set 1 ^a SDSS-DR3 MGS Training=18045,Testing=20229 ^b	Data Set 2 SDSS-DR5 LRG Training=87002,Testing=9666	Data Set 3 SDSS-DR3 MGS + GALEX-GR4 Training=30036,Testing=3374
g-r-i	g-r-i	g-r-i
u-g-r-i	u-g-r-i	u-g-r-i
g-r-i-z	g-r-i-z	g-r-i-z
u-g-r-i-z	u-g-r-i-z	u-g-r-i-z
-	-	nuv-fuv-u-g-r-i-z
u-g-r-i-z-r50	u-g-r-i-z-r50	nuv-fuv-u-g-r-i-z-r50
u-g-r-i-z-r50-r90	u-g-r-i-z-r50-r90	nuv-fuv-u-g-r-i-z-r50-r90
u-g-r-i-z-r50-r90-ci	u-g-r-i-z-r50-r90-ci	nuv-fuv-u-g-r-i-z-r50-r90-ci
u-g-r-i-z-r50-r90-ci-qr	u-g-r-i-z-r50-r90-ci-qr	nuv-fuv-u-g-r-i-z-r50-r90-ci-qr
u-g-r-i-z-r50-r90-fd	u-g-r-i-z-r50-r90-fd	nuv-fuv-u-g-r-i-z-r50-r90-fd
u-g-r-i-z-r50-r90-fd-qr	u-g-r-i-z-r50-r90-fd-qr	nuv-fuv-u-g-r-i-z-r50-r90-fd-qr

^au-g-r-i-z=5 SDSS magnitudes, r50=Petrosian 50% light radius in r band, r90=Petrosian 90% light radius in r band, CI=Petrosian Inverse Concentration Index, FD=FracDev value, QR=Stokes Q value in the r-band, nuv=GALEX Near UV band, fuv=GALEX Far UV band. See Paper I § 3.6 for more details.

^bThese are the sizes of the testing and training sets used in our analysis

Table 2. Data Sets 4-6

Data Set 4 ^c SDSS-DR5 LRG + GALEX-GR4 Training=4042,Testing=454	Data Set 5 SDSS-DR5 MGS + 2MASS Training=133947,Testing=15050	Data Set 6 SDSS-DR5 LRG + 2MASS Training=39344,Testing=4420
g-r-i	g-r-i	g-r-i
u-g-r-i	u-g-r-i	u-g-r-i
g-r-i-z	g-r-i-z	g-r-i-z
u-g-r-i-z	u-g-r-i-z	u-g-r-i-z
nuv-fuv-u-g-r-i-z	u-g-r-i-z-j	u-g-r-i-z-j
nuv-fuv-u-g-r-i-z-r50	u-g-r-i-z-h	u-g-r-i-z-h
nuv-fuv-u-g-r-i-z-r50-r90	u-g-r-i-z-k	u-g-r-i-z-k
nuv-fuv-u-g-r-i-z-r50-r90-ci	u-g-r-i-z-j-h	u-g-r-i-z-j-h
nuv-fuv-u-g-r-i-z-r50-r90-ci-qr	u-g-r-i-z-h-k	u-g-r-i-z-h-k
nuv-fuv-u-g-r-i-z-r50-r90-fd	u-g-r-i-z-j-h-k	u-g-r-i-z-j-h-k
nuv-fuv-u-g-r-i-z-r50-r90-fd-qr	-	-

^cu-g-r-i-z=5 SDSS magnitudes, r50=Petrosian 50% light radius in r band, r90=Petrosian 90% light radius in r band, CI=Petrosian Inverse Concentration Index, FD=FracDev value, QR=Stokes Q value in the r-band, nuv=GALEX Near UV band, fuv=GALEX Far UV band, j=2MASS j band, h=2MASS h band, k=2MASS k band. See Paper I § 3.6 for more details.

function $k(x, x')$, with x and x' vectors with d components. This covariance function can be used to construct K ($K_{ij} = k(x_i, x_j)$) with x_i, x_j rows of X) and the $n^* \times n$ cross covariance matrix K^* ($K_{ij}^* = k(x_i^*, x_j)$ where x_i^* is the i^{th} row of X^*). Therefore the prediction \hat{y}^* for y^* may be given by the Gaussian Process equation (Rasmussen & Williams 2006, p. 17):

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

λ represents the noise in y and can be used to improve the quality of the model (Rasmussen & Williams 2006). For details about the selection of λ , the covariance function (kernel) k , hyperparameters in the kernel, and Gaussian Process Regression in general see Foster et al. (2009); Rasmussen & Williams (2006). What follows is a summary of Foster et al. (2009). We will use the notation above for the sections that follow.

4.1. Different Kernel choices

In Paper I we relied exclusively on a polynomial kernel, but to investigate the possibility that other kernels might perform better we tried several other common forms.

We test the squared exponential (SE) kernel function given by

$$k_{SE}(r) = \exp\left(-\frac{r^2}{2l^2}\right) \tag{2}$$

where l is the length scale. The length-scale determines the rate at which the kernel function drops to zero away from the origin. This covariance function is infinitely differentiable and hence is very smooth. With such strong smoothness, it is sometimes unrealistic for use in modeling real physical processes. The squared exponential covariance function is also called the radial basis function.

The Matern class covariance function is given by

$$k(r) = \frac{2^{l-v}}{\Gamma(v)} \left(\frac{\sqrt{2vr}}{l}\right)^v K_v\left(\frac{\sqrt{2vr}}{l}\right) \tag{3}$$

where v and l are positive parameters and K_v is a modified Bessel Function. As $v \rightarrow \infty$ this reduces to the squared exponential above. The process becomes very rough for $v = \frac{1}{2}$ and for values of $v \geq \frac{7}{2}$, the function is as rough as noise. The Matern class covariance function is mean square differentiable k times if and only if $v > k$. Hence, the Matern class of covariance functions can be used to model real physical processes and is more realistic than the Squared Exponential (SE) covariance function.

The rational quadratic covariance function is given by

$$k(r) = \left(1 + \frac{r^2}{2\alpha l^2}\right)^{-\alpha} \quad (4)$$

As $\alpha \rightarrow \infty$ this reduces to the SE function. This function is mean square differentiable for every α as opposed to the Matern class covariance function.

The polynomial covariance function is given by

$$k(x, x') = (\sigma_0^2 + x^T \Sigma_p x')^p \quad (5)$$

where Σ_p is a positive semidefinite matrix and p is a positive integer. If $\sigma_0^2 = 0$ the kernel is homogeneous and linear, otherwise it is inhomogeneous. In principle this function may not be suitable for regression problems the variance grows with $|\times|$ for $|\times| > 1$. However there are applications where it has turned out to be effective (Rasmussen & Williams 2006).

The neural network covariance function is given by

$$k_{NN}(x, x') = \frac{2}{\pi} \sin^{-1} \left(\frac{2x^T \Sigma x'}{\sqrt{(1 + 2x^T \Sigma x)(1 + 2x'^T \Sigma x')}} \right) \quad (6)$$

This covariance function is named after Neural networks because the function can be derived from the limiting case of a model of a neural network (Neal 1996)

Two or more covariance functions can be combined to produce a new covariance function. For example sums, products, convolutions, tensor products and other combinations of covariance functions can be used to form new covariance functions. Details are described in Rasmussen & Williams (2006).

For the calculations shown in the rest of the paper we utilized equation 6, the Neural Network kernel, since for our data it outperformed all other kernels.

4.2. Low Rank Approximation Matrix Inversion Techniques

As mentioned in Paper I (section 4.4) to utilize Gaussian Process Regression (GPR) the inversion of the matrix $M = (\lambda^2 I + K)$ in equation 1 is required. This matrix turns out to be an $n \times n$ non-sparse matrix where n =number of training-set galaxies. Paper I mentioned that matrix inversion requires $O(n^3)$ floating point operations. Thus, to accommodate the matrix in memory and to keep the computation feasible, we kept $n \leq 1000$.

This was a severe shortcoming for GPs since they have 1-2 orders of magnitude less training samples to work with than all of the other methods described in Paper I. Nonetheless GPs performed extremely well within this limitation.

Since writing Paper I, we have developed a variety of rank reduction methods, to invert large non-sparse matrices. These will make GPR much more competitive than that shown in Paper I. Foster et al. (2009) outline the rank reduction methods utilized in detail, so we provide a brief summary of their advantages below.

Note that n is the same as that described above, while $m < n$ is the size of the rank reduced matrix. We typically keep $m < 1500$ to keep the numbers of operations to invert the matrices manageable in wall-clock time. Memory usage for the methods below is $O(nm)$.

SR-N: The Subset of Regressors Method. This method has been proposed and utilized in the past (Rasmussen & Williams 2006; Whaba 1990; Poggio & Girosso 1990) and requires nm^2 flops to invert. However, this method is known to have problems with numerical stability. That problem is addressed in the methods below.

SR-Q: The Subset of Regressors using a QR factorization. This approach was originated by (Golub & Van Loan 1996, p.239) and is designed to reduce computer arithmetic errors in the SR-N method. This method requires $2nm^2$ flops to invert while storage is $O(nm)$. Therefore it is a little more expensive than SR-N.

SR-V: The V Method. This method is intermediate in in terms of growth of computer arithmetic errors between the normal equations and the SR-Q method, but in general the accuracy is close the SR-Q. This method was first discussed by Seeger et al. (2003) and (Whaba 1990, p.136). This method requires $2nm^2$ flops to invert the matrices.

SR-NP, SR-QP, SR-VP: The use of Pivoting with rank reduction methods. All of the previous methods use the first m columns of K , but one can select any set of columns to construct a low-rank approximation. Selecting these columns is part of the problem to be solved. Our approach is similar to that of Fine & Scheinberg (2001). In the end adding pivoting increases SR-N to $2nm^2$ flops and SR-Q to $3nm^2$ while SR-V stays the same.

5. Comparison: Picking the optimal Sample Size, RANK size, and Matrix Inversion Method

Here we investigate Data Set 1 in detail in order to discern a variety of things including: is there an optimal sample size for a given survey; what is the best Matrix Inversion Method; if using rank-reduction methods what is the optimal RANK size? When discussing the full

matrix inversion we will be limited to a maximum of 20,000 training samples ⁷.

Figures 1 and 2 include the full matrix inversion up to 20,000 training-set samples, which is labeled GPR and is in yellow. The rest of the curves are from the other rank reduction matrix inversion techniques and are labeled as described in the previous section. A few features of note:

1. The SR-N method does not perform well in comparison to any of the other techniques. However, it does invert its matrices much faster than the standard matrix inversion technique.
2. Except for the SR-N method all of the other rank reduction methods outperform the full matrix reduction in the range of 10000-20000 samples.
3. The rank reduction methods with pivoting slightly outperform the non-pivoting methods in term of lower RMS values. However, the pivoting methods take much more time to do the matrix inversions than the non-pivoting methods.
4. More training-set samples give lower RMS values. By around 40,000 samples the curves start to level off regardless of the RANK size.
5. Larger RANK sizes clearly give better performance in terms of lower RMS for a given sample size.

Figure 3 plots RANK versus RMS for a given sample size. Here we plot RANK from 100 to 1000 in increments of 100, but also add RANK=1500 to see if there is a large change in calculated RMS for a much larger RANK. Some important features to note here:

1. As in Figure 1 for large sample sizes the RMS decreases, but again there is not a large difference between Sample sizes of 40,000 and above.
2. For the non-pivoting matrix inversion techniques (not include SR-N) SR-Q and SR-V the RMS increases beyond RANK=800. This shows some instability in the solutions for larger RANK sizes. It is clear that one must stick with the pivoting methods (SR-QP or SR-VP) if one wishes to use a RANK of 800 or larger.

⁷This is because of memory(RAM) limitations. Our 64-bit compute platform is based around a 2 x 2.66 Ghz Dual-Core Intel Xeon with 16GB of 667Mhz DDR2 RAM

3. On average it appears that SR-VP and SR-QP outperform all RANK reduction methods. SR-VP also appears to outperform SR-QP, although the difference is marginal.
4. SR-VP with RANK=800 and Sample size=40000 appear to be optimal choices for our data when looking at Figures 1, 2 and 3 given the accuracy of the result. The timings are much longer for these pivoting methods as shown above, but they outperform all other methods.

6. Results

6.1. SDSS Main Galaxy and Luminous Red Galaxy Results

In Data Set 1, which is based upon the SDSS MGS sample, the trends regarding the choosing of optimal Sample and RANK size should remain consistent for the class of galaxies/SEDs the MGS encompasses. On the other hand Data Set 2 is based upon the SDSS LRG sample where there are far fewer SED types than that found in the SDSS MGS. This clearly makes the job of any regression algorithm different. Compare the plots in Figure 4 and this is immediately apparent. A number of points need to be stressed:

1. Morphological Inputs: The morphological information (p50, p90, CI, FD, QR) may add some information that the regression algorithm can utilize. This includes the Petrosian 50% radii in the r-band (p50), the Petrosian 90% radii in the r-band (p90), the inverse concentration index ($CI=p50/p90$), the FracDev (FD) and Stokes Q in the r-band (QR) parameters as discussed in Paper I. In Data Set 1 Figure 4a only the p50 parameter appears useful. In Data Set 2 Figure 4b the morphological information appears to add noise for the most part making the fits worse than by using only the 5 bandpass filter fluxes.
2. Fewer SEDs: As mentioned in the previous section, around the time sample sizes of 40,000 in Data Set 1 are reached the RMS begins to level off. In Data Set 2 however this is already occurring for most of the inputs in the 10,000–20,000 range. This is clearly the advantage of having less SEDs to worry about. In fact for Data Set 2 it is clear that only 4 of the 5 bandpasses is sufficient for the optimal fit. The u bandpass is superfluous in this data set.

6.2. Cross-Matching GALEX and SDSS Results

Figure 5 shows results from a cross-match of the SDSS and GALEX catalogs which are Data Sets 3 and 4 in Table 1. A list of obvious things to point out include:

1. If one compares Figure 5a on the left to Figure 4a one immediately notices that Figure 5a has lower rms values for the purely SDSS bandpasses (g-r-i, g-r-i-z, u-g-r-i, u-g-r-i-z).
2. The same is even more obviously true for Figure 5b versus Figure 4b.
3. In Figure 5 the additional two GALEX bandpasses (nuv, fuv) appear to lower the rms when added to the SDSS bandpasses.
4. The rms values for the SDSS only bandpasses for Figure 5 over Figure 4 are all the more remarkable when one looks at the training sample sizes used in each. In Figure 4a up to 80,000 samples were used while in Figure 5a only 30,000 were used. In 4b again up to 80,000 are used, while in 5b only 4000 are available and yet the results are incredibly superior in 5b in all bandpass and morphological combinations.
5. The previous observations clearly point to the fact that our underlying SDSS magnitudes and redshifts have changed. If one looks at Figure 7 which compares Data Sets 1 and 3 (Figures 4a and 5a) this hypothesis is confirmed. The differences are even more radical in Figure 8 as one would expect given the differences in Figures 4b and 5b.
6. In Figure 5 the rms values level off more quickly as a function of sample size for all inputs compared to Figure 4.

It is clear that one must be very careful in interpreting greatly improved rms results from SDSS + GALEX cross-match catalogs.

6.3. Cross-Matching 2MASS and SDSS Results

Figure 6 demonstrates our GPR results from a cross-match catalog containing the 2MASS extended source catalog and the SDSS MGS sample and LRG sample (Data Sets 5 and 6). When comparing Figure 6 with 4 the Figure 6 results appear to be surprisingly improved in both cases. However, we would be mistaken again if we thought it was the extra bandpasses that brought all values down. We must take note of a variety of changes between Figures 4 and 6.

1. It is immediately apparent that the rms has suddenly dropped from Data Set 1 to 5 for the SDSS only bandpasses and for Data Set 2 to 6.
2. Figure 6b (Data Set 6) has lower RMS values regardless of input. It also appears to converge to a best fit RMS very quickly in comparison to Data Set 5 on the left in Figure 6.
3. In Figure 6a (Data Set 5) it is clear that adding the 2MASS fluxes improves the RMS fit regardless of which 2MASS filter is combined with the SDSS filters.
4. In Data Set 5 using only the u-g-r-i of the SDSS gives rather excellent results. This would imply that there are very few large spectral features influencing the z band. That would lead one to believe that the redshift for this sample is lower. It would be a correct assumption if one looks at Figures 9 and 10.
5. In Figure 6b (Data Set 6) adding the 2MASS fluxes can improve the RMS fit, but unlike in Figure 6a ((Data Set 5) the circumstances for improvement are different. In fact upon closer inspection one sees using only the u-g-r-i filters of the SDSS-LRG in combination with the 2MASS h and k bands give equivalent best results by the time one reaches 40,000 training-samples. Clearly in this case it is the u band that adds little to the LRG sample which concurs with what we found in Figure 4b (Data Set 2).

6.4. Systematics

For a given Data Set we plot in Figures 11 and 12 that input which yields the lowest rms. The actual rms is also indicated in each plot. There appears to be a systematic shift above the regression line for redshifts less than 0.1 and below the regression line between $0.1 < z < 0.2$ for Data Sets 1, 3 and 5. This effect has been seen or discussed in many papers on this topic (e.g. Collister & Lahav 2004; D’Abrusco et al. 2007; Ball et al. 2008; Wang et al. 2009).

6.5. Conclusion

We have demonstrated that with new non-sparse matrix inversion techniques and a better choice of kernel (or transfer function if you prefer) that Gaussian Process Regression is a competitive way to obtain accurate photometric redshifts for surveys such as the SDSS. However, there are several caveats to be observed when estimating photometric redshifts in

combined catalogs of the SDSS and 2MASS as well as the SDSS and GALEX as discussed in Section 6.

For example, The SDSS + 2MASS cross-match results are astoundingly good, but even when the only bandpasses used are those of the SDSS especially so when focusing on the Luminous Red Galaxy Sample. This is clearly a case where we are sampling a smaller range of redshifts and magnitudes which makes the regression job easier regardless of the algorithm. Clearly one has to be careful when quoting "better" results from a cross-match of any catalog.

We also demonstrate that the addition of SDSS morphological parameters does not systematically improve our regression results. In most cases only adding the Petrosian 50% Radii gives substantially lower rms values. Most of the other morphological parameters appear to simply add noise to our regression fits. For a low redshift survey like those analyzed herein it makes sense that the Petrosian 50% Radii would help given the angular diameter-distance relation. However, in higher redshift surveys this is very unlikely to be the case.

The papers associated with this project and the Code used to generate the results from this paper are available on the NASA Ames Dashlink web site <https://dashlink.arc.nasa.gov/algorithm/stab>

M.J.W thanks Jim Gray, Ani Thakar, Maria SanSebastien, and Alex Szalay for their help in cross-matching the catalogs used herein. Thanks goes to the Astronomy Department at Uppsala University in Sweden for their generous hospitality while part of this work was completed. M.J.W. acknowledges funding received from the NASA Applied Information Systems Research Program. A.N.S. thanks the NASA Aviation Safety Integrated Vehicle Health Management project for support in developing the GP-V method.

The authors acknowledge support from the NASA Ames Research Center Director's Discretionary Fund.

Funding for the SDSS has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy, the Max-Planck-Institute for Astrophysics, New Mexico State University, University of Pittsburgh, Princeton

University, the United States Naval Observatory, and the University of Washington.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

The Galaxy Evolution Explorer (GALEX) is a NASA Small Explorer. The mission was developed in cooperation with the Centre National d'Etudes Spatiales of France and the Korean Ministry of Science and Technology.

This research has made use of NASA's Astrophysics Data System Bibliographic Services.

REFERENCES

- Abazajian, K., et al. 2005, *AJ*, 129, 1755
- Adelman-McCarthy, J.K. et al. 2007, *ApJS*, 172, 634
- Ball, N.M., Brunner, R.J., Myers A.D., Strand, N.E., Alberts, S.L., & Tchong, D. 2008, *ApJ*, 683, 12
- Bernstein, G. & Huterer, D. 2009, arXiv:0909.2782v1
- Blanton, M.R., et al 2005, *AJ*, 129, 2562
- Carliles, S., et al. 2008, *ASPC*, 394, 521
- Collister, A.A. & Lahav, O. 2004, *PASP*, 116, 345
- D'Abrusco, R., Staiano, A., Giuseppe, L., Brescia, M., Paolillo, M. De Filippis, E. & Tagli-
aferri, R. 2007, *ApJ*, 663, 752
- Efron, B., & Tibshirani, R.J. 1993, *An introduction to the bootstrap*. New York, Chapman
& Hall
- Eisenstein et al. 2001, *AJ*, 122, 2267
- Fine, S. & Scheinberg, K. 2001 "Efficient svm training using low-rank kernel representations",
Journal of Machine Learning Research 2, 243264
- Foster, L., Waagen, A., Aijaz, N., Hurley, M., Luis, A., Rinsky, J. Satyavolu, C., Way, M.,
Gazis, P., Srivastava, A. 2009 *Journal of Machine Learning Research*, in press

- Giavalisco, M. et al. 2004 ApJ, 600, L93
- Golub, G.H. & Van Loan, C.F. 1996, “Matrix Computations”. Johns Hopkins University Press, Baltimore, MD, USA third edition, ISBN 0-8018-5413-X, 0-8018-5414-8
- Ivezic, Z., Tyson, J.A., Allsman, R., Andrew, J., Angel, R., et al 2008, arXiv:0805.2366v1
- Kaiser, N. et al. 2002, Proc. Of the SPIE, 4836, 154
- Kurtz, M.J., et al. 2007, astro-ph/0707.0484v1
- Le Fèvre O. et al. 2004 å, 417, 839
- Neal R.M. 1996, *Bayesian learning for neural networks*, Springer-Verlag New York, Inc., Secaucus, NJ, USA
- Martin, D. C., et al. 2005, ApJ, 619, L1
- Padmanabhan, N. et al. 2005, MNRAS, 359, 327
- Petrosian, V. 1976, ApJ, 209, L1
- Poggio, T. and Girosi, F. 1990 Proceedings of IEEE, 78, 1481.
- Rasmussen, C.E. & Williams, C.K.I. 2006 “Gaussian Processes for Machine Learning”, MIT Press, Cambridge Massachusetts
- Seeger, M., Williams, C., & Lawrence, N.D. “Fast forward selection to speed up sparse gaussian process regression”. In C. M. Bishop and B. J. Frey, editors, Proceedings of the Ninth International Workshop on Artificial Intelligence and Statistics, San Francisco, 2003. Morgan Kaufmann.
- Sérsic, J.L. 1968, Atlas de Galaxias Australes (Cordoba: Obs. Astron.)
- Skrutskie, M.F. et al. 2005, AJ, 131, 1163
- Stabenau, H.F., Connolly, A., & Bhuvnesh, J. 2008, MNRAS, 387, 1215
- Strauss, M.A., et al. 2002, AJ, 124, 1810
- Wadadekar, Y. 2005, PASP, 117, 79
- Wang, D., Zhang, Y., Liu, C. & Zhao, Y. 2007, astro-ph/0707.2250v1
- Wang, T., Huang, J. & Gu, Q. 2009, Research in Astron. Astrophys. Vol.9, No.4, p.390

Way, M.J. & Srivastava, A.N. 2006, ApJ, 647, 102

Wahba, G. 1990 “Spline Models for Observational Data”, Society for Industrial and Applied Mathematics, Philadelphia, PA, USA

Wray, J.J. & Gunn, J.E. 2007, astro-ph/0707.3443v1

York, D.G., et al. 2000, AJ, 120, 1579

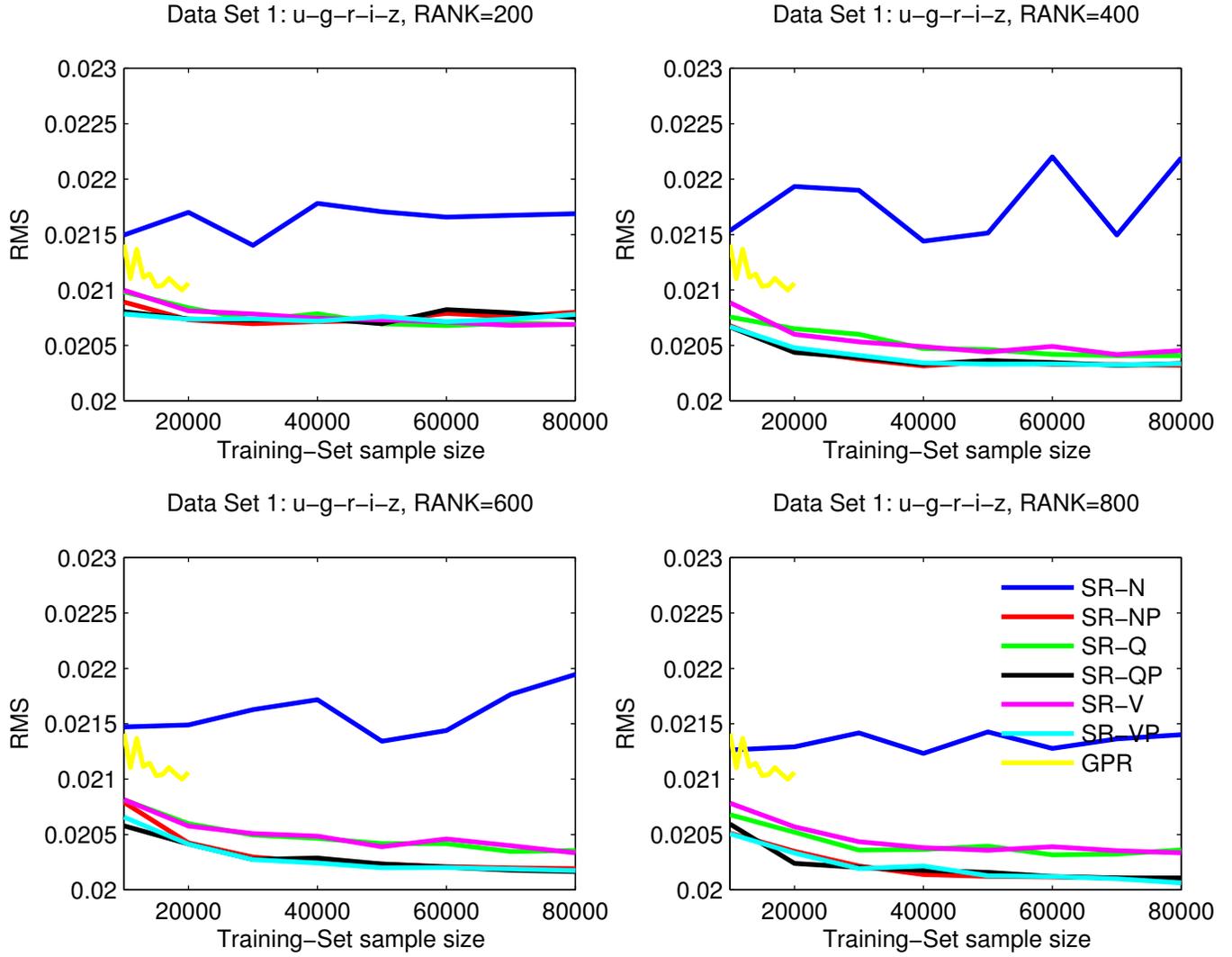


Fig. 1.— From Data Set 1 (see Table 1). Error bars are not plotted for reasons of clarity, however they are of the same order as the scatter in the lines.

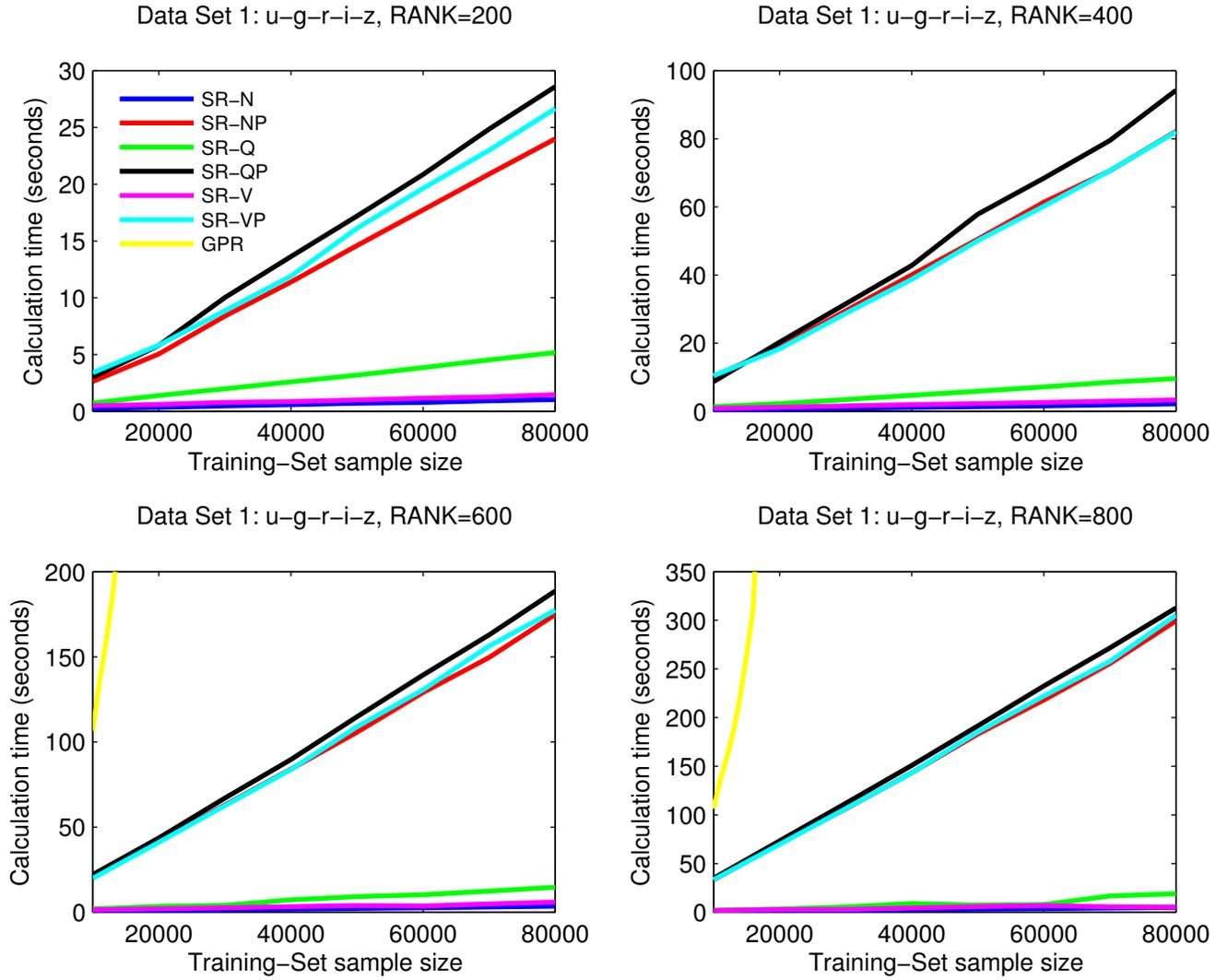


Fig. 2.— From Data Set 1 (see Table 1)

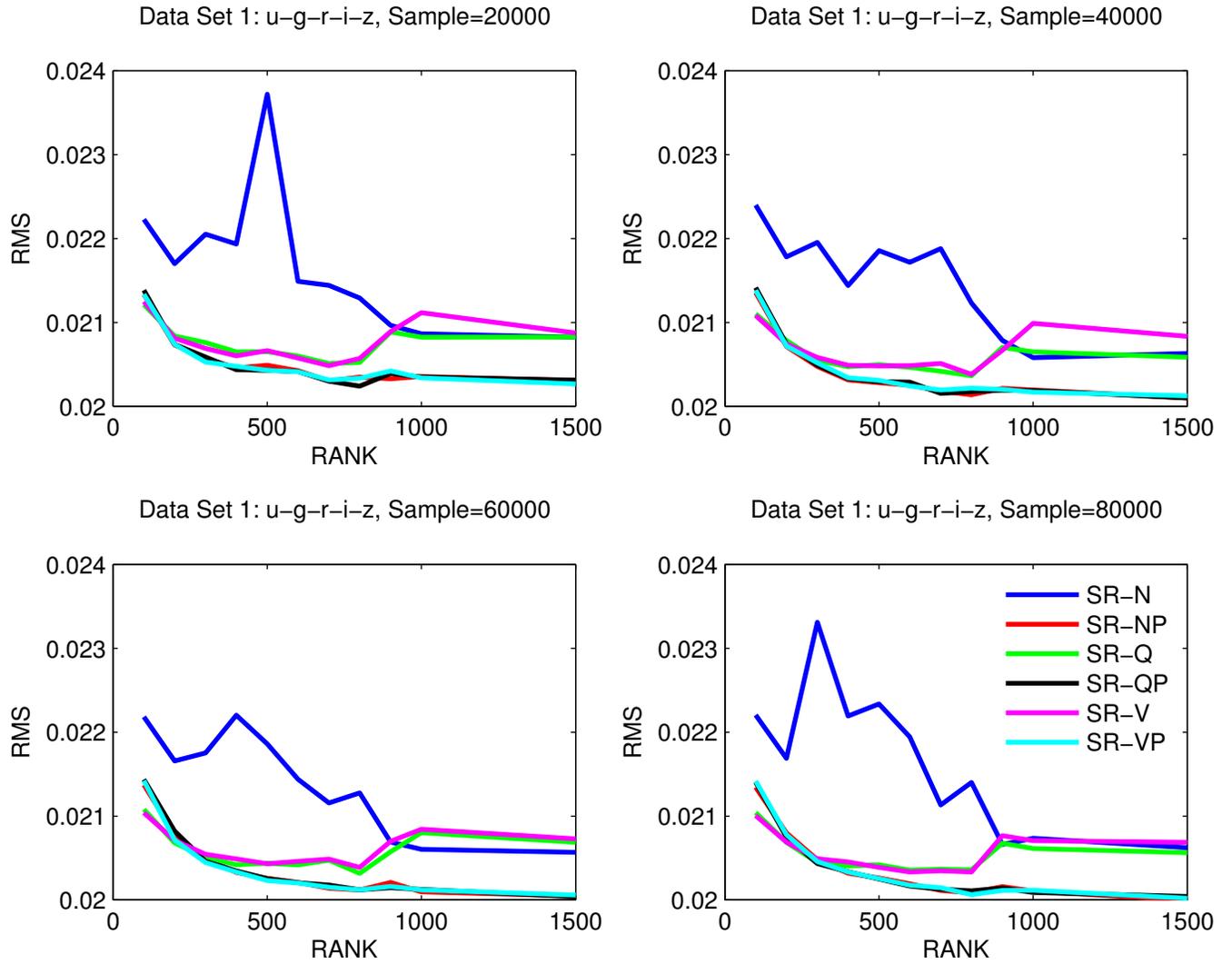


Fig. 3.— From Data Set 1 (see Table 1) Error bars are not plotted for reasons of clarity. They are of the same order as the scatter in the lines.

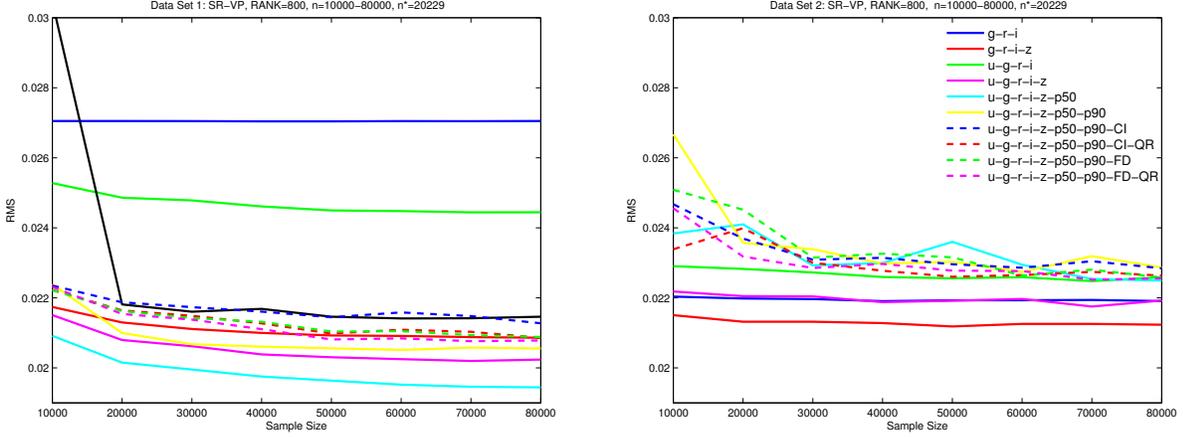


Fig. 4.— From Data Sets 1 and 2 (see Table 1). We utilize the Rank Reduction method termed SR-VP with RANK size of 800. The training-sets (n in the plot, following our earlier notation) range in size from 10000 to 80000 in 10000 increments with 10 bootstraps (Efron & Tibshirani 1993) per run. The testing sample size (n^*) was always 20,229. The mean value of the 10 bootstraps is plotted.

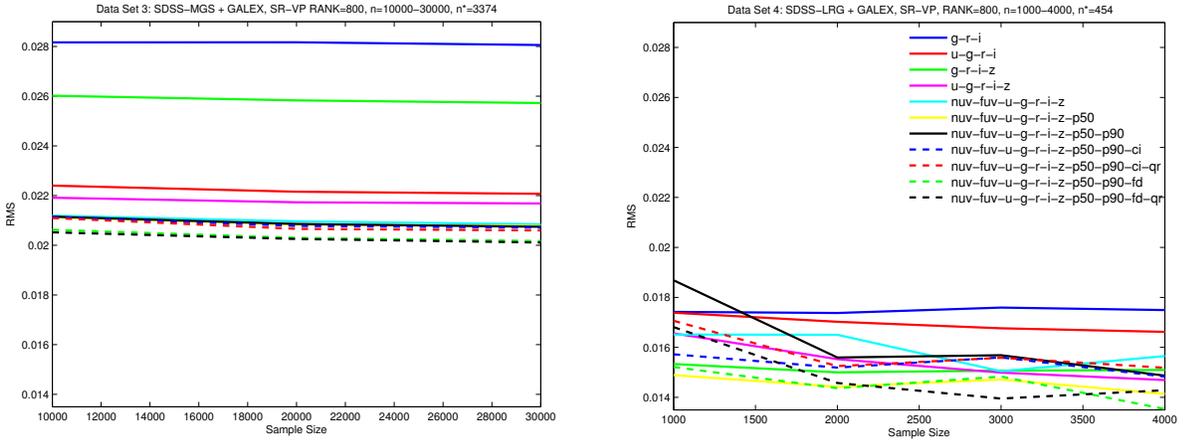


Fig. 5.— From Data Sets 3 and 4 (see Table 1). We utilize the Rank Reduction method termed SR-VP with RANK size of 800. On the left we use training-sets (n in the plot, following our earlier notation) ranging in size from 10000 to 30000 in 10000 increments with 10 bootstraps per run. The testing sample size (n^*) was 3374. The mean value of the 10 bootstraps is plotted. Error bars from the bootstrap resampling are approximately triple the size of the line width. On the right we use similar notation, but we have smaller training (1000-4000) and testing (454) sets. Note that the yellow marker for $\text{nuv-fuv-u-g-r-i-z-p50}$ is hidden under the magenta and black lines in the left plot.

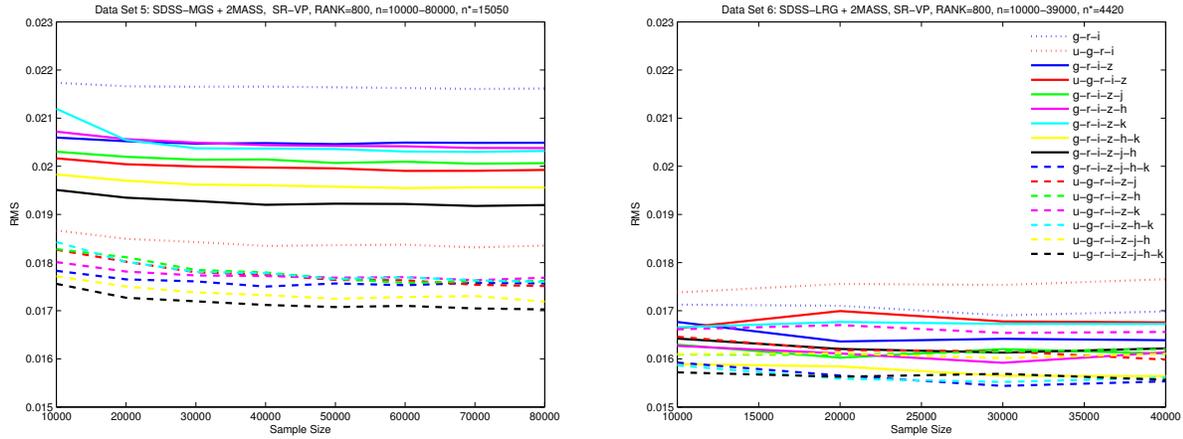


Fig. 6.— From Data Sets 5 and 6 (see Table 1). We utilize the Rank Reduction method termed SR-VP with RANK size of 800. For Data Set 5 the training-sets (denoted as n) range in size from 10000 to 80000 in 10000 increments with 10 bootstraps per run and a testing-set (n^*) size of 15050. On the right Data Set 6 training-sets range from 10000 to 40000 in increments of 10000 with 10 bootstraps per run and a testing-set size of 4420.

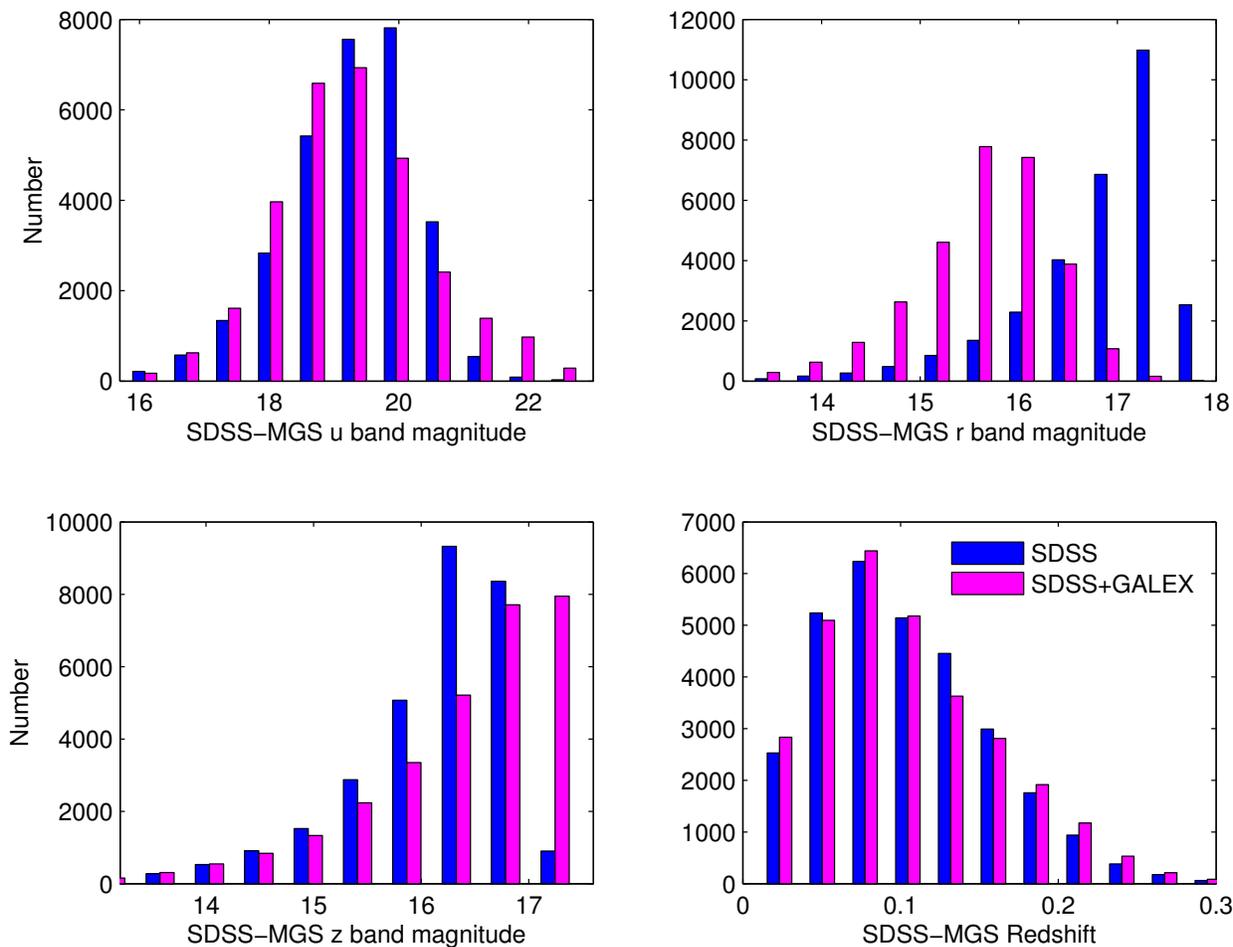


Fig. 7.— Overlapping histograms for Data Sets 1 and 3 (see Table 1) from 3 of the 5 SDSS magnitudes (u,g,z). Data Set 1 are in blue, Data Set 2 in green. Of course the SDSS+GALEX cross-match catalogs (Data Set 3) are smaller, so the SDSS only data (Data Set 1) was randomly resampled to be the same size as the cross-match catalog so that trends in the plots are directly comparable.

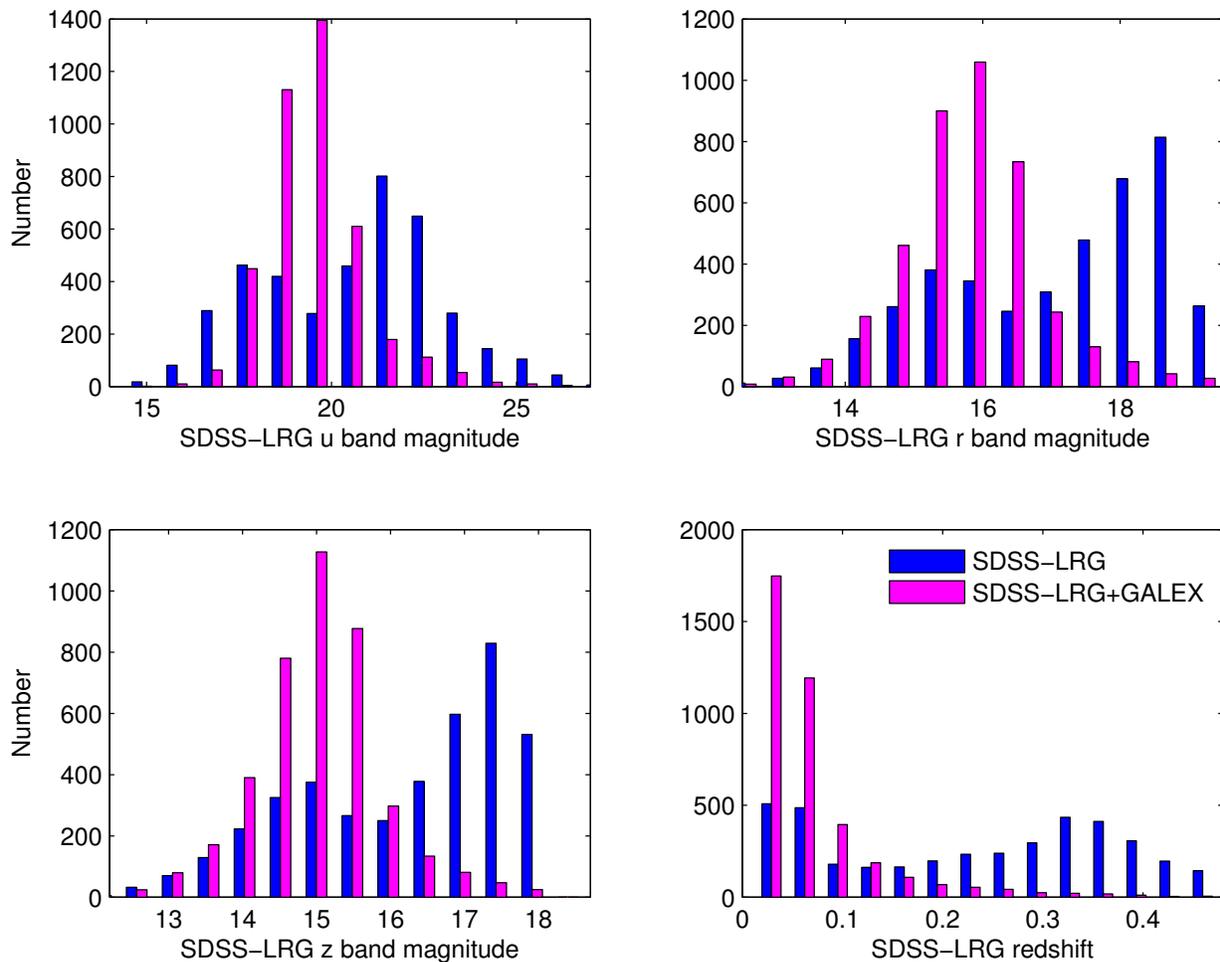


Fig. 8.— Overlapping histograms for Data Sets 2 and 4 (see Table 1) from 3 of the 5 SDSS magnitudes (u,g,z). Data Set 1 are in blue, Data Set 2 in green. Of course the SDSS+GALEX cross-match catalogs (Data Set 3) are smaller, so the SDSS only data (Data Set 1) was randomly resampled to be the same size as the cross-match catalog so that trends in the plots are directly comparable.

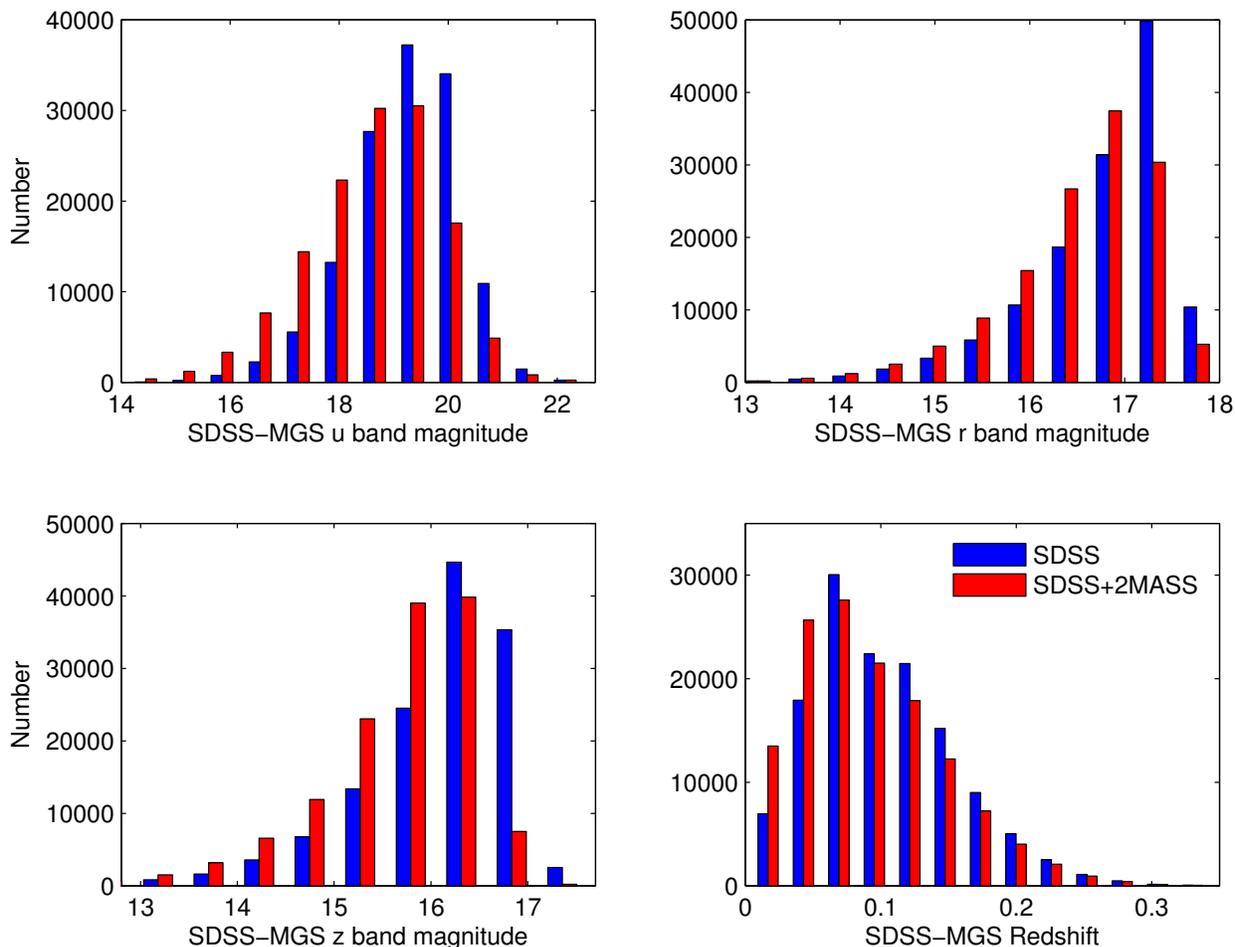


Fig. 9.— Overlapping histograms for Data Sets 1 and 5 (see Table 1) from 3 of the 5 SDSS magnitudes (u,g,z). Data Set 1 are in blue, Data Set 2 in red. Of course the SDSS+2MASS cross-match catalogs (Data Set 5) are smaller, so the SDSS only data (Data Set 1) was randomly resampled to be the same size as the cross-match catalog so that trends in the plots are directly comparable.

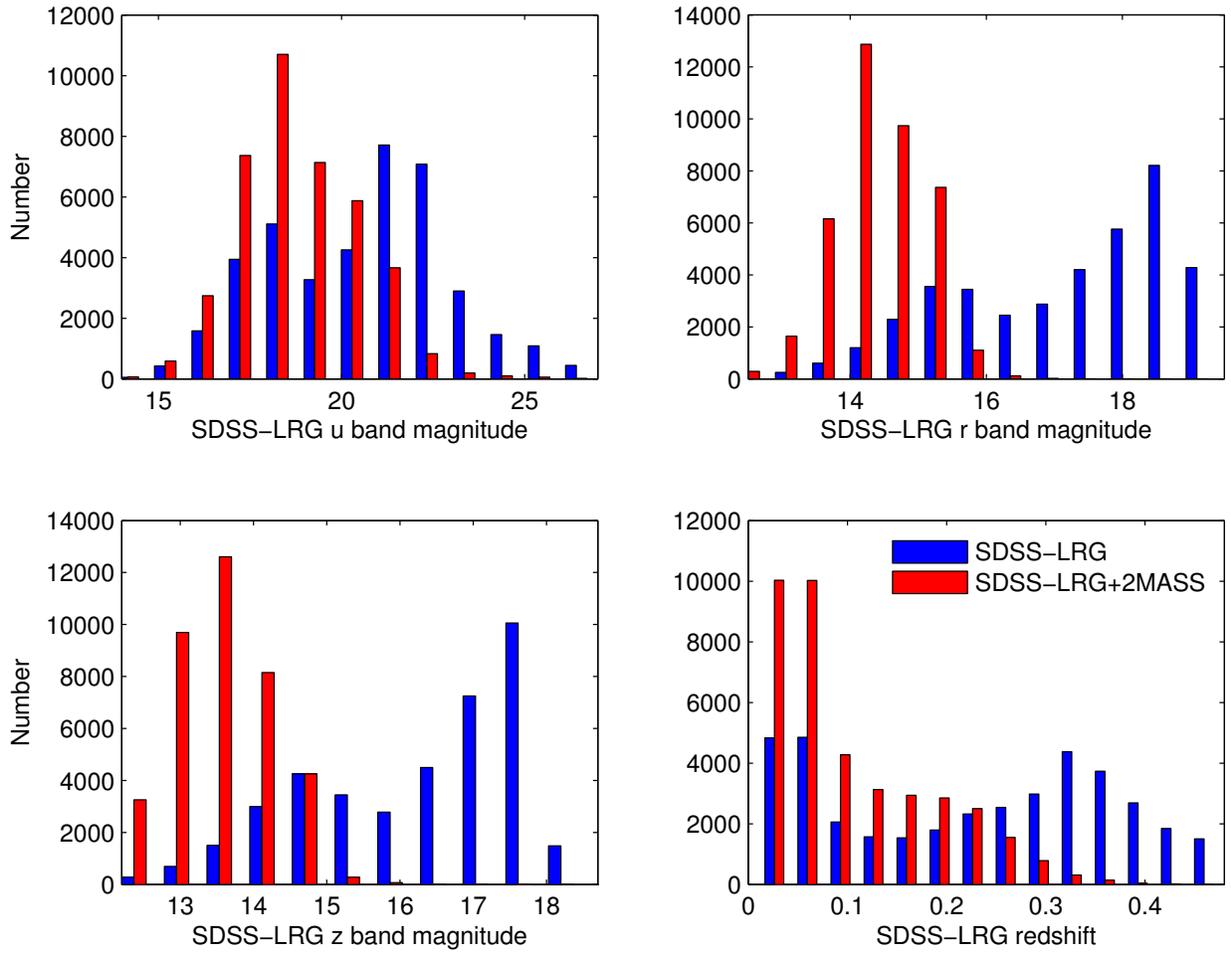


Fig. 10.— Same as Figure 9 except we use Data Sets 2 and 6

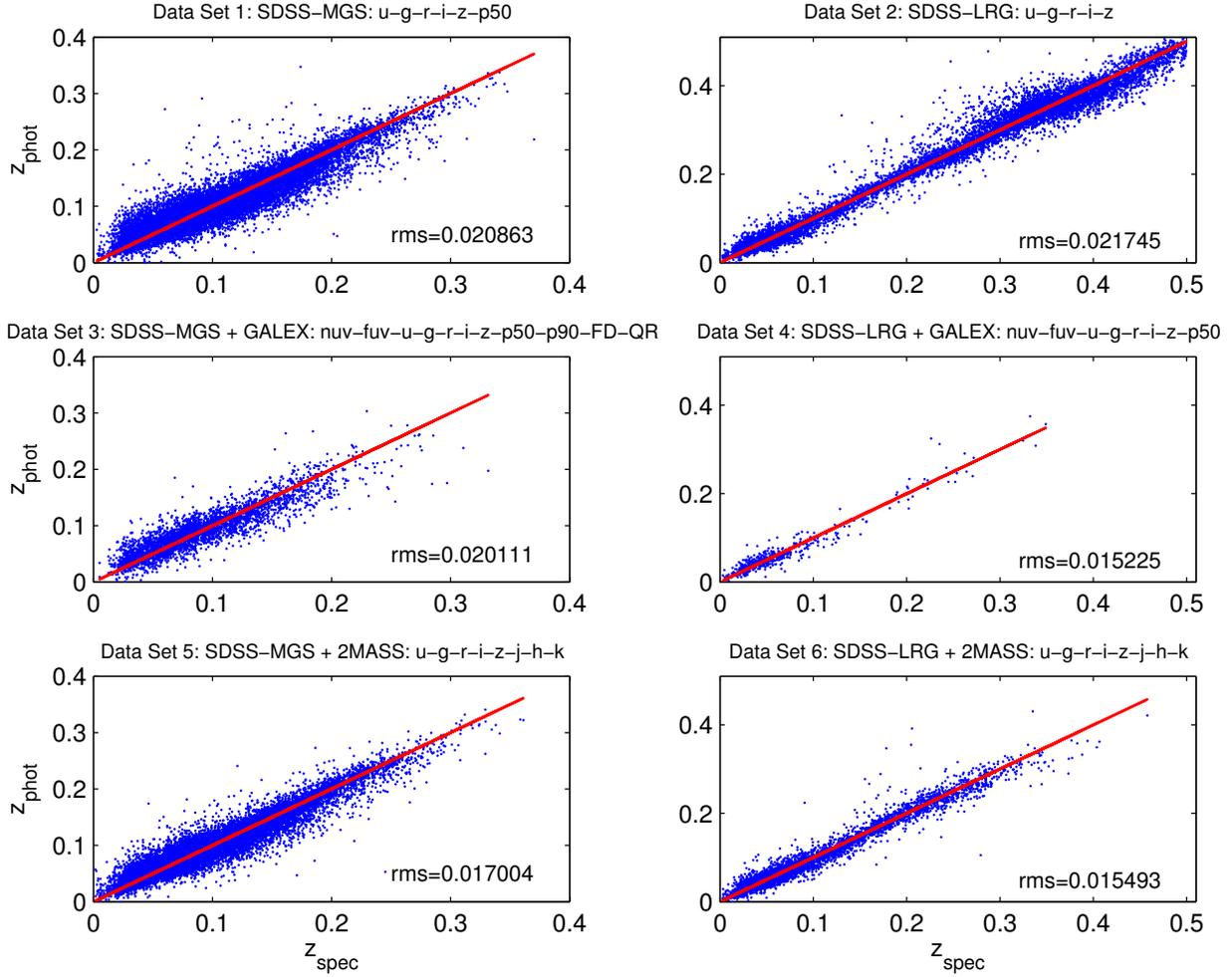


Fig. 11.— Spectroscopic redshift plotted again predicted photometric redshift for the best performing input from each of the Data Sets in Table 1.

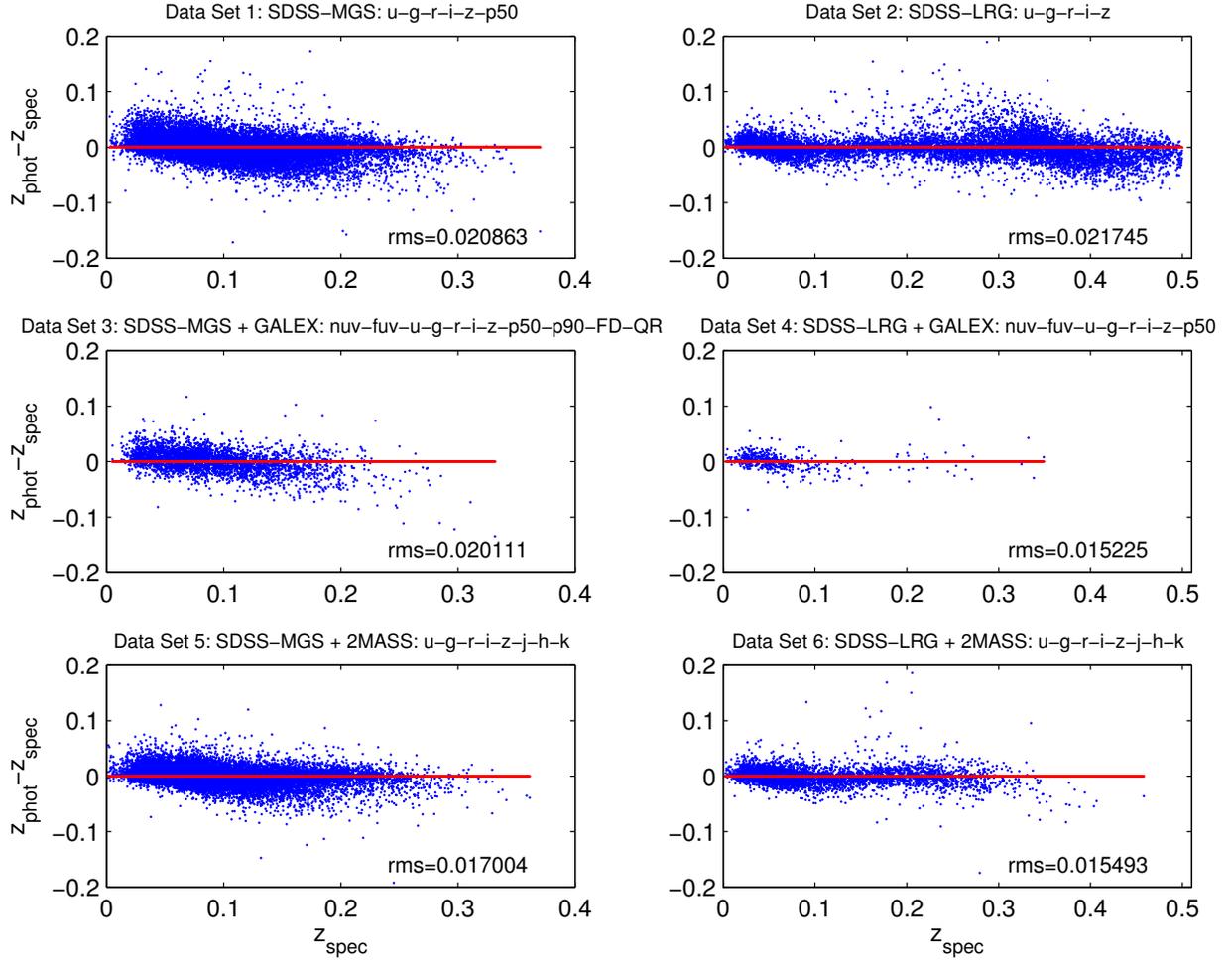


Fig. 12.— Residuals as a function of spectroscopic redshift for the best performing input from each of the Data Sets in Table 1.