A Plan for Photo-z Calibrations

Abstract: This memo describes two complementary general solutions to the challenge of calibrating photometric redshifts. One relies purely on spectral information while the other relies on angular cross-correlations between spectroscopic and photo-z samples. They complement each other: the more spectroscopic redshifts are obtained, the better the cross-correlation constraints. Priors and angular autocorrelations offer additional improvements. Simulations suggest that these approaches can reach the precision in photo-z bias error and variance required for the next generation deep wide photometric surveys for cosmology.

For a variety of cosmological probes it is necessary to develop a plan for calibrating photometric redshifts over the range in redshift $0<z<3$ for galaxies of all types to R~26 AB mag. Traditionally there are two basic approaches to this spectroscopic calibration problem: (1) empirically calibrate a relation using a large training set, and (2) get a better understanding of the spectral energy distributions for galaxies out to z=3. The difficulty in both approaches (and hybrids) is in getting a fair sample of galaxies at z>1 from which to test or train the relations. Spectroscopic surveys often have color-dependent efficiencies which then bias the distribution of galaxies with redshift. Having a properly sampled redshift survey (i.e. optimal sampling in color space, optical and IR) would solve many of these issues. If one could take all of the existing high-*z* redshift samples that have optical through mid-IR photometry and compare their color distributions with the underlying galaxy populations it might be possible to do a small number of targeted redshift surveys to fill in the gaps (in type, redshift and color). One possibility would be to generate a series of selected fields distributed across the sky with hundreds of thousands of galaxies to R~26 AB calibrated via selected deep spectroscopy. These fields would then be used as a training set for photometric redshifts from a multicolor survey; e.g. the 6-band photometric redshifts for LSST.

This is a challenge for any such deep galaxy photometric survey; here we shall use parameters of the LSST survey for specificity. Aside from the issue of the number of filters, it is very unlikely that the accuracy of photo-z calibration or the results of the kind of analysis we attempt here will influence the design of any survey hardware. This photo-*z* calibration plan uses existing facilities and could be accomplished before 2010.

Surveys like LSST generally use only a limited number of bands in order to cover a wide area deeply. For an LSST-like survey covering *ugrizy*, it is estimated that a faint sample of all types of galaxies would have a statistical photo-*z* error of d*z* ~ 0.05 (1+*z*) per galaxy. Figure 1 shows Monte Carlo simulations at S/N=30 for accuracy of photo-z's from 5 or 6 band systems. Note the non-Gaussian errors.

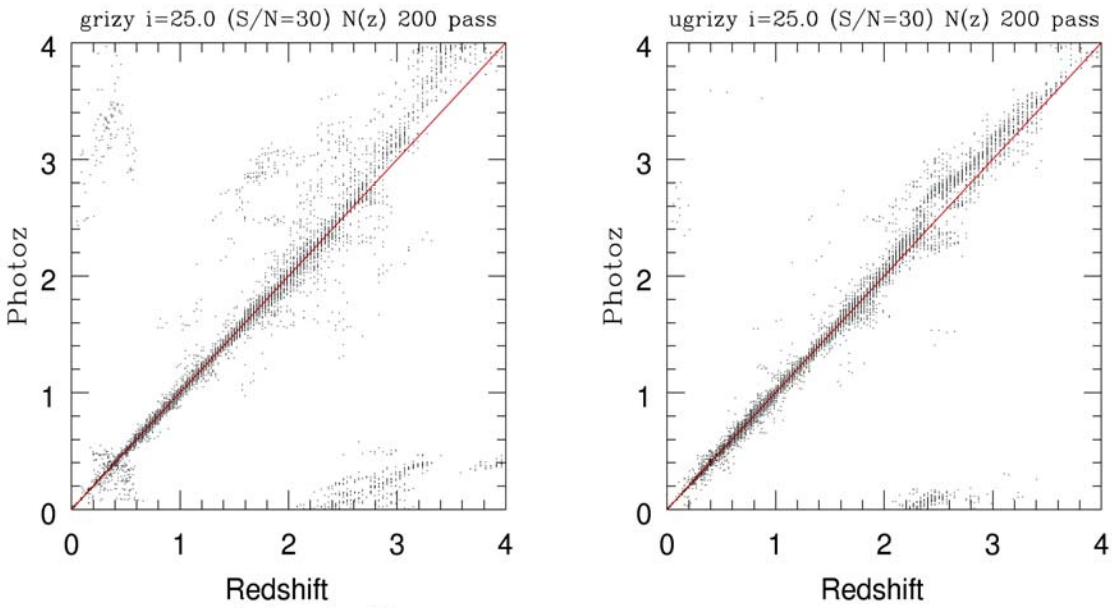

Fig. 1. Monte Carlo simulations of the LSST survey to i=25 AB mag using a 5-band and a 6-band system. No luminosity or surface brightness priors have been used. The task is to calibrate the bias and variance errors to high precision for the sample between 0 < z < 3.

While it is desirable to reduce the scatter in photo-*z*'s, it is far more important to know both the bias and the dispersion to good precision. For extraction of the dark energy parameters w and w_a, simulations of 20,000 sq.deg weak lensing analysis show that one must reach a mean bias precision of dz = 0.002 (1+$\underline{z}$) in each redshift bin and a measure of the mean uncertainty in the dispersion to $\Delta\sigma_z$ = 0.003(1+z) in order not to degrade the precision more than 2x the statistical limit for the dark energy parameter w. Bear in mind that the statistical noise in w in such a survey comes from the shot noise of a sample of several billion galaxies, so it is very small. For the parameter w_a the corresponding goals become (0.005, 0.006). For baryon acoustic oscillations, the most important error is $\Delta\sigma_z$ = 0.003(1+z) for the same condition. In fact, we need to characterize the full distribution function since it is non-Gaussian.

We therefore take 0.002(1+z) as a goal for the mean bias uncertainty in each z-bin and 0.003(1+z) as a goal for dispersion calibration precision. We need sufficient numbers of galaxies per redshift bin to beat down the statistical errors to at least the level of the systematic errors. For a *ugrizy* photo-z error of 0.05(1+z) per galaxy, this implies >600 galaxies per redshift bin for a 1σ statistical limit at the required 0.002(1+z) level. Of course we would like the precision in any one redshift bin to be dominated by systematic errors rather than statistical noise, so that the system can be calibrated. This then leads to over 6,000 galaxies per bin for a precision dominated not by statistics but by accuracy of calibration. Given the non-Gaussian nature of the errors (including catastrophic outliers) it is likely that more would be needed; roughly 100,000 galaxies total if the sample were split up into ten redshift bins. For reference, to R~26 mag AB there are a million galaxies of all types in three square degrees; the field of view of LSST is ten square degrees.

We then need several calibration fields each ~2 square degrees and containing ~100,000 galaxies whose mean redshift we trust to a precision of 0.002 (1+z) per z-bin in total. 100,000 spectroscopic redshifts at these magnitudes is out of the question with current facilities. An efficient route to the required calibration is to undertake a two-phase procedure in which we first assemble a deep "super" photometric (12 or more bands over an extended wavelength range from the far UV to the IR) photo-z training set of 200,000 galaxies, each with a photo-z precision of 5% or better. We then need to calibrate this transfer photometric training set. A model for the evolution of the SED vs. galaxy type could be used to guide how we select galaxies for the spectroscopic calibration of the transfer photometric standard fields in different redshift regimes by optimally sampling. The numbers are a question of sampling the type distribution of galaxies. Random sampling requires about 5,000 galaxies per bin (so we sample sufficient numbers of galaxies to include extreme line emitters), but if we sample based on rest frame color (in an iterative sense where we use the photo-z to initially estimate the type) we can gain about a factor of 10 in efficiency by down-weighting the sampling of typical galaxies. An optimistic assessment (assuming the photometric calibration of these fields was accurate to 1% or better) would result in a need for 500-1000 spectra per redshift bin.

Estimates of the rms photo-z error of a 12-band UV-IR photo-z deep survey (with S/N > 10 in each band) are dz ~0.03(1+z) at 1-sigma. Figure 2 shows the result of a Monte Carlo simulation (including SED template noise) of a 12 band system to S/N=10 at 26 AB mag, log sampled in wavelength from 0.15 to 2 microns, where we sample evenly in redshift and galaxy type to reproduce the distribution of galaxies within the super-calibration data set. Comparing to Figure 1, the error distribution is much better than that from a 0.3-1 micron system because we can detect the Lyman and Balmer breaks simultaneously and remove the degeneracies that occur when only one break is observable.

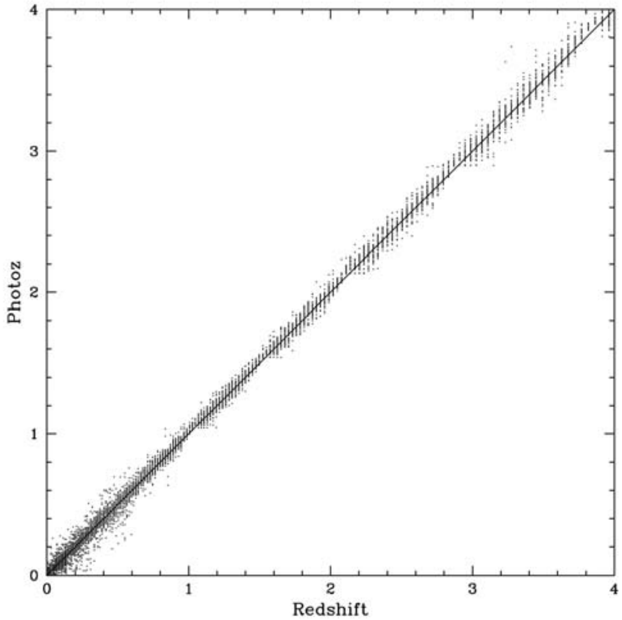

Fig. 2. A super-photo-z transfer calibration system with 12 log spaced bands from 0.15 to 2 microns. While there is smaller scatter, the more significant change is in the distribution of errors. This makes spectroscopic calibration more reliable. This Monte Carlo is without application of any priors. The rms error per galaxy is ~0.03 for most redshift bins.

The task of spectroscopic calibration of this super-photo-z transfer system is less challenging, leading to a 1-sigma requirement of 200 deep spectra per z-bin over the z=0-3 range to beat down the random errors to 0.002(1+z). However, again, for the systematic part of the error to dominate this should be increased to 2000. This leads to a spectroscopic calibration campaign of ~20,000 galaxies to R~26 mag if we want 10 z-bins. Extrapolation beyond the photometric limit of the spectroscopic sample, e.g. to 26.5, should be possible since some of the faint galaxies will be line emitters, so long as line emission does not correlate strongly with SED. In addition, if the ensemble spectral properties of the galaxies do not evolve rapidly with redshift, not as many spectra are required in all redshift ranges. Such an observing campaign is achievable with current facilities if one includes IR spectroscopy. The COSMOS, VVDS, GOODS, and DEEP2 fields are examples where this work has begun, albeit only to I or R~24 AB. Deeper ground-based, HST, and Spitzer photometry than currently available will be required for this campaign, in addition to deeper spectroscopy. Reaching the required depths at the shortest and longest wavelengths required will be the most difficult; e.g. it would take ~100 nights to cover the super-photo-z fields with existing near-IR spectroscopic instruments from the ground.

If the underlying type-color-magnitude-z distribution is fully understood, selection effects in any observation can be simulated and these selection effects can be used to improve the precision; e.g. the errors from z-bin to z-bin could be regularized by a model of SED evolution. How many z bins are needed to train the model depends on the evolution of the SEDs with redshift; i.e., if all galaxies have the same spectral properties as a low redshift sample then we would need to only study the SED distribution locally. This is clearly not the case, but when sampling at z>1 it becomes a question of how rapidly the metalicity and dust content of galaxies evolves with lookback time. In principle we could alternate between higher density sampling of some redshift intervals and lower density sampling of others depending on the type-SED evolution at that redshift.

What about the effects of “cosmic” or sample variance among these 2-3 deep pencil-beam calibration fields? Certainly cosmic variance affects the n(z). However, we are after the type-dependent errors in a photo-z system, so if we can adequately sample the full range of color-z-type space in these small training fields then to first order we have the calibration we need. Cosmic variance will affect the amount of SED evolution in a given field, and if that model is used too heavily to derive likelihoods then the $n(z)$ in the calibration fields can influence the photo-z calibration residuals. This is the prime motivation, aside from having a calibration field available at any time of the year, for having more than one. For $n(z,\theta)$ generally it is sufficient to sample 200 Mpc on a side, and the noise just decreases with the number of 200 Mpc intervals. Of course this is not really relevant to this color-z calibration unless there is a strong density-type dependence. If we assume that morphology-density relations extend to 10 h^{-1} Mpc (based on SDSS environment studies), then each 2$\times$2 degree patch (2 deg = 80 h^{-1} Mpc comoving at redshift z=1) should contain 520 spheres of this size within a Δz=0.2 interval, so there is not much of a residual sample variance effect. One could also sample based on projected density as well as color.

To complete such a spectroscopic calibration program there would be several steps along the way:

Step 1) A group of experts analyzes a combination of archived data in regions of the sky where a combination of deep spectroscopy and both *HST* /ACS and *Spitzer* near-IR imaging are available; e.g. the DEEP2, GOODS, and COSMOS fields. In combination with existing ground-based imaging, these datasets give photometry in effectively ~10 broadband filters calibrated with spectroscopy. The combination will greatly refine evolution and selection function models. This step is already happening.

Step 2) These deep fields would be further enhanced via deep imaging in new bands, covering the far UV to IR. The superphotometric bands/facilities should include *GALEX* (FUV + NUV for isolated galaxies), *HST* ACS (*I,g*), Subaru (deeper *BVgrizy*), CFHT (*u* + *JHK*), and possibly *Spitzer* (3 shortest wavelength IR bands for isolated galaxies). This deep super-photo-z system then requires spectroscopic calibration.

Step 3) The first two steps are likely to reveal important holes in our understanding of the models. This knowledge will be used to inform a dedicated effort using of order 50 nights of VLT time and also of order 50 nights of Keck and/or Gemini time to undertake the necessary spectroscopy to constrain the models well. This would produce about 20,000 spectroscopic calibrations out to z=3 for the super photo-z training set. It is possible that fewer calibrators could be used. The key would be to understand the selection function. For example, DEIMOS can go to $R = 26$ in about 40 hours. Beyond $z = 1.4$ one would probably want to use a blue sensitive spectrograph, or a near-IR spectrograph for emission line galaxies around $z = 2$. For z >~ 2.5 or so one would be back to DEIMOS for the Ly-a line. If a calibration precision of $0.001(1+z)$ were required, four times this investment would be needed. We would select an initial sample and then iterate the next selection based on how well we do. Spectroscopic facilities which could be used include Keck/DEIMOS, VLT/VIMOS, Gemini/GMOS, Magellan/COSMOS, and Subaru IRSPEC.

Step 4) Once these steps are complete, we would have some understanding of the remaining systematics leading to confidence in the redshifts of ~20,000 galaxies in these selected fields in order to calibrate the local super-photo-z system. The ~1 million galaxies with super-photo-*z* in each of these training fields then would be used to calibrate the 6-band photo-z's for programs like LSST. The precision of bias calibration and the variance will vary with redshift, being worst in the “optical desert.” However, these error estimates will be known both as a function of type and redshift, and that can be used for more precise extraction of cosmological parameters from WL and BAO analyses. Furthermore, galaxies in portions of color space which are poorly constrained (or fail to ever yield redshifts) could be excluded from WL or BAO analyses with minimal impact.

Cross-correlation techniques for photo-z calibration

We have begun to explore an alternative method that can get around any incompleteness issues in determining redshift distributions. Even if we are not successful in obtaining redshifts for all of the galaxies selected for spectroscopy (and past experience suggests we will not be), it is clear we will have a large set of faint galaxies (~20,000) with spectroscopic redshifts at the end of this campaign. This spectroscopic sample (along with other, existing redshift samples not considered here) can then be used to determine the actual redshift distribution for any set of galaxies selected photometrically via **cross-correlation** methods. In particular, the two-point angular cross-correlation function (on the sky) between galaxies with spectroscopic redshifts in some narrow bin about redshift z and some set of galaxies from the photo-z dataset, $w_{sp}(\theta)$, will be proportional to an integral of the number of photometric galaxies at that z, $n_p(z)$ times the two-point real-space cross-correlation, $\xi_{sp}(r)$, over the redshift dimension. The angular cross-correlation can be measured with great precision in uniform, well-calibrated photometry, as there will be many photometric galaxies near each spectroscopic galaxy on the sky.

This measurement alone is not sufficient to determine $n_p(z)$, but we have additional information: the redshift survey will measure the autocorrelation function of spectroscopic galaxies, $\xi_{ss}(r)$, while the total surface density and angular autocorrelation functions of the photometric sample are directly measurable from the photometric data. These quantities provide the integrals of

$n_p(z)$ and of $n_p(z)^2 \times \xi_{pp}(r)$, respectively. For simple biasing schemes, we can then iteratively determine $\xi_{sp}(r) = (\xi_{ss}(r) \times \xi_{pp}(r))^{1/2}$ and $n_p(z)$. For more complicated biasing, additional modeling is required, but this should not be an insurmountable difficulty.

Assume we have a sample of N_s galaxies with spectroscopic redshifts distributed as $z^2 e^{-z/z0}$, with median z=0.75, and we are attempting to measure the redshift distribution of some sample of galaxies selected photometrically (e.g. to fall in a given photo-z bin) with surface density in galaxies per square arcminute Σ and a true redshift distribution which is a Gaussian centered at z=1, for example, with sigma σ_z. Monte Carlo tests indicate that we can then measure both the mean redshift and σ_z for the photometric sample with an uncertainty of 4.7×10^{-3} $(\sigma_z/0.1)^{1.6}$ $(N_s/20{,}000)^{-0.5}(\Sigma/10)^{-0.5}$, comparable to the required tolerances for LSST at z=1; see Figure 3. At higher redshifts, it is more likely that targeted samples of a few thousand galaxies per unit z will be obtained, as described in the previous section. If spectra are obtained for dN_s/dz galaxies per unit redshift, the errors will be 7.7×10^{-3} $(\sigma_z/0.1)^{1.5}$ $(dN_s/dz\ /\ 5000)^{-0.5}$ $(\Sigma/10)^{-0.5}$, meeting tolerances at z=2-3. Detecting non-Gaussianities (tails, etc.) in the photo-z distributions should also be straightforward in this method. The assumed numbers of spectroscopic galaxies are fairly modest (worse than the combination of existing samples at z~1, a factor of a few increase on current samples at z~3).

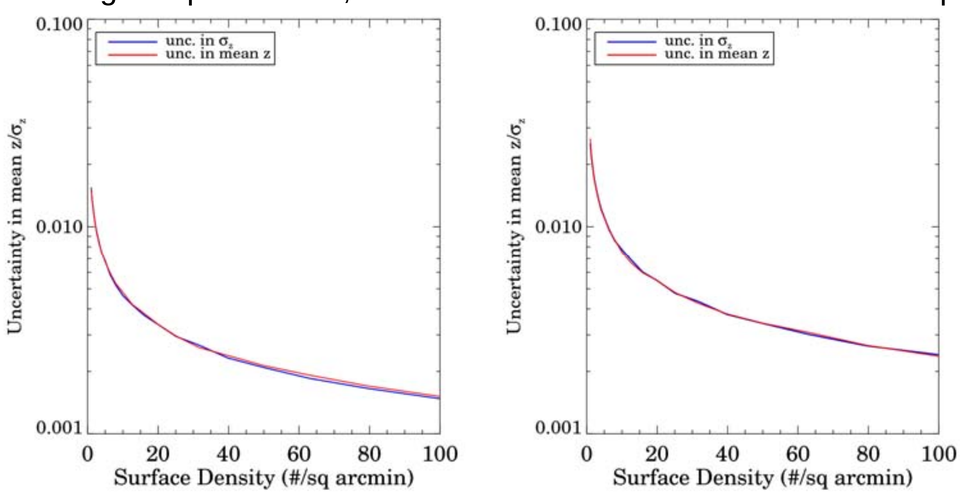

Figure 3. Results from Monte Carlo simulations of uncertainties in cross-correlation measurements of redshift distributions. Plotted are the RMS errors in the recovery of the mean and sigma of a photometric sample distributed as a Gaussian in z with σ_z=0.1, as a function of the surface density of that sample on the sky. On the left, we assume a fiducial spectroscopic survey of 20,000 galaxies with median *z*=0.75 and a photo-*z* bin with central *z*=1 and surface density as shown. On the right, we assume a spectroscopic sample of 5000 galaxies per unit *z*, more appropriate for targeted high-redshift samples (e.g. at *z*~3). The uncertainties in both cases scale as (surface density)$^{-0.5}$. Existing spectroscopic samples are already sufficient to reach the required LSST photo-*z* calibration tolerances at *z*~1, but larger sets of redshifts than currently available at *z*~0.5 and *z*>1.5 are required.

We have tested the impact of several potential pitfalls in this method. The first of these is *bias evolution*: if the bias (or clustering) of the photometric sample evolves with z, that will be degenerate with dN_{phot}/dz, as we have constraints only on the combination of number density and clustering at a given z (from the cross-correlation), and the integrated number and clustering of the photometric sample (from the surface density and angular autocorrelation for the photometric sample). For a Gaussian redshift distribution of galaxies in the photometric bin and a constant bias evolution per unit z of db/dz, if we assume there is no bias evolution but there in fact is, we will make an error in the mean z of the photometric sample of $(db/dz) \times \sigma_z^2$. For realistic samples, db/dz<~0.3; then we could make an error of 0.003 in the mean z if σ_z=0.1 and if we have no knowledge of db/dz, just below LSST tolerances at z=1. This is a highly pessimistic scenario, though, as e.g. by comparing cross-correlation results from adjoining photo-z bins we can determine db/dz directly.

A second potential issue would be spatial variation in the effective photometric *zero points* (due to seeing, calibration issues, etc.). Note that random photometric errors or absolute zero point uncertainties do not affect this method; it will empirically calibrate dN/dz for whatever falls in a given photo-z bin, regardless of whether a galaxy is put in that bin due to errors. However, variation in zero points over the sky could cause more galaxies to be placed in the photo-z bin in some parts of the sky than others artificially, causing a false correlation signal. Even for a conservative scenario (N_{spec}=100,000, $\Sigma=10$, only 3 independent patches of sky surveyed, and RMS zero point errors 0.05), this source of error is tiny, increasing the errors in mean z and σ_z by less a fraction of a percent; we note that the LSST specification is 0.005 mag RMS zero point variation.

We note that these simulations all ignore “cosmic” or sample variance. Its impact could be minimized by sparsely sampling over large areas of the sky, or can be corrected to first order by scaling the observed deviations from smoothness in the dN/dz of the spectroscopic sample by the measured bias to predict variations in the dN/dz of the photometric sample. We also ignore cross-correlations induced by gravitational lensing, as we can iteratively remove the small lensing induced signal as we build up dN/dz (e.g., once we have a first approximation to dN/dz for a photo-z bin, we can predict how much cross-correlation with galaxies in that zbin we should have from the observed lensing and remove it).

In principle, this method would work even if we obtain spectra of only the brightest objects at all redshifts (there are many z=2 galaxies with R<24, or z=3 galaxies with R<25) – we do not need an excessively deep sample, only a well-defined one. However, to maximize the avenues for and crosschecks of photo-z calibration, it would still be best to obtain spectra to a limit approaching the LSST photometric depth. Ideally, one might separately analyze two sets of spectroscopic/photo-z angular cross-correlations: one inside the deep calibration fields (where one understands better each galaxy), and another in which the main wide field LSST survey photo-z sample is correlated in angle with a sparsely sampled spectroscopic sample of galaxies. If the main wide field photo-*z* sample has been calibrated via the deep calibration training set, this additional angular correlation would provide a cross check which could probe in a new way the robustness of this technique (i.e., the results should be consistent vs. type, redshift, and magnitude cuts).

Additional constraints

Two improvements are imminent which will significantly improve photo-z calibration: (1) the efficient use of priors, and (2) the use of the angular-redshift galaxy autocorrelation function $w(\theta,\Delta z)$ for self-calibration.

While it would be dangerous to heavily weight galaxy priors (surface brightness, luminosity function, or size vs. *z*) for the training set, it is effective to use such priors for the main survey photo-*z* to reduce the impact of photometric outliers. Tests on existing survey data with both photo-*z* and spectroscopy show a marked decrease in catastrophic photo-*z* errors when priors are included. For example, the decrease in catastrophic errors shown in Figure 1 when including the *u* filter is even more dramatic when using a luminosity function prior.

The angular-redshift galaxy correlation function $w(\theta,\Delta z)$, which can be determined using data purely from a photo-*z* survey, can also be very helpful. Catastrophic photo-*z* errors induce cross-correlations between the photo-*z* bins that can be more than two orders of magnitude larger than the intrinsic cross-correlation. In principle, measurement of the angular cross-correlation function contains more than enough information to reconstruct the true redshift distribution (i.e. without catastrophic photo-*z* errors). Preliminary Fisher matrix forecasts indicate that the distribution of catastrophic photo-z errors can be determined to within a few percent of the number density of galaxies within each redshift bin, assuming a fiducial model for an LSST survey. To be effective, this autocorrelation method requires large sky coverage to reduce the sample variance and a prior on the galaxy bias, which is otherwise completely degenerate with the galaxy number density on large scales (where scale-dependence of the bias can be ignored).

Andy Connolly
Jeffrey Newman
Tony Tyson
March 20, 2006

With input from Sandy Faber, Lloyd Knox, David Hogg, David Koo, Vera Margoniner, Michael Schneider, David Wittman, and Hu Zhan