

# Self-Calibration Simulation Results

Peter Yoachim (UW), Lynne Jones (UW), Željko Ivezić (UW), Tim Axelrod (LSST)

## ABSTRACT

We describe the planned self-calibration procedure for LSST and compare results of simulated observations to SRD requirements. By combining Opsim pointing histories with a realistic model of the Galactic stellar distribution we generate a mock-catalog of calibrated LSST observations. We simulate the major sources of error expected in LSST observations and measure how sensitive the self-calibration procedure is to them. We demonstrate a new technique for solving the self-calibration problem in parallel using HEALpixels, making it possible to solve an LSST-sized data set on current computer hardware.

## 1. Introduction

LSST is required to deliver photometric calibration with zeropoint variation of at most 1% (root-mean-square scatter of 0.01 mag) across the observed sky. By reducing the photometric error by a factor of two over the current state-of-the-art wide-field optical photometry delivered by SDSS, the error volume in the five-dimensional LSST color-color space will be reduced by more than a factor of 30 when compared to SDSS-like photometry. This improves source classification and the precision of quantities dependent upon color, such as photometric redshifts for galaxies and metallicity determination for stars.

The overall calibration strategy is described in detail in “Level 2 Photometric Calibration for the LSST Survey” by Jones et al., (<git://dev.lsstcorp.org/LSST/sims/doc.git>, hereafter abbreviated “L2PC”). The major advances in the LSST calibration plan include correcting the wavelength dependence of the hardware and atmospheric transmission curves using direct measurements from auxiliary instrumentation and, most importantly, the wealth of repeat observations of  $\sim 10^8$  bright main sequence stars in the LSST survey. These series of repeat observations will enable correction of a wide variety of photometric errors and most importantly, allow LSST to operate in a wide variety of atmospheric conditions. The factor of two reduction in photometric error results from two major differences between LSST and SDSS. First, each source will receive a hundred to two hundred observations (per band) over the ten years of the LSST survey. These series of repeat observations will be used to self-calibrate the photometric system across the sky and for each observation (akin, but not

identical to, the über-calibration procedure used by SDSS (Padmanabhan et al. 2008)), allowing LSST to operate in a wide variety of conditions. Secondly, the wavelength dependence of the hardware and atmospheric transmission response functions will be measured with auxiliary instrumentation on sufficiently fine angular and temporal scales to enable their explicit inclusion in the calibration procedure, rather than resorting to traditional approximations such as linear color terms.

## 2. Simulated LSST Catalog

We combine the following elements to generate a simulated LSST catalog of observed stellar magnitudes.

### Milky Way Model

The stars used as the basis for self-calibration are generated using Mario Jurić’s GALFAST code<sup>1</sup>. This provides a mock Milky Way stellar catalog to act as “ground truth” for our simulations.

Our model Milky Way includes  $\sim 1$  billion stars, with a magnitude range of  $17 < r < 21$ . The model includes several stellar components (thin disk, thick disk, bulge, halo) with magnitude and color distributions based on SDSS. The catalog is dominated by main-sequence and red giant branch stars, but also includes white dwarf stars and blue horizontal branch stars. We do not include O, B, or A stars, however, these stars are concentrated near the galactic plane and are expected to be saturated in LSST exposures. Our catalog is drawn from the object catalog currently used by Imsim. For most of our simulations, we have pre-selected a subset of one million stars that are uniformly distributed across the sky. This pre-selection is done to limit the amount of computational memory the self-calibration requires, and prevents systematic effects from having a steep density gradient of stars across the field of view. The densities of the various stellar populations are shown in Figure 1.

### Telescope and Survey Properties

A simulated pointing history of the telescope is generated by the LSST Operations Simulation. We are using the Opsim 3.61 run as our baseline survey. Opsim simulations include scheduled and unscheduled downtime, main survey visits, deep drilling fields, and a few other observing programs. The Opsim database includes the  $5\sigma$  depth for each observation, calculated using the expected sky background emission in that filter at that altitude/azimuth, lunar distance and phase, as well the seeing and exposure time of each observation. We use

---

<sup>1</sup><http://mwscience.net/trac/wiki/galfast>

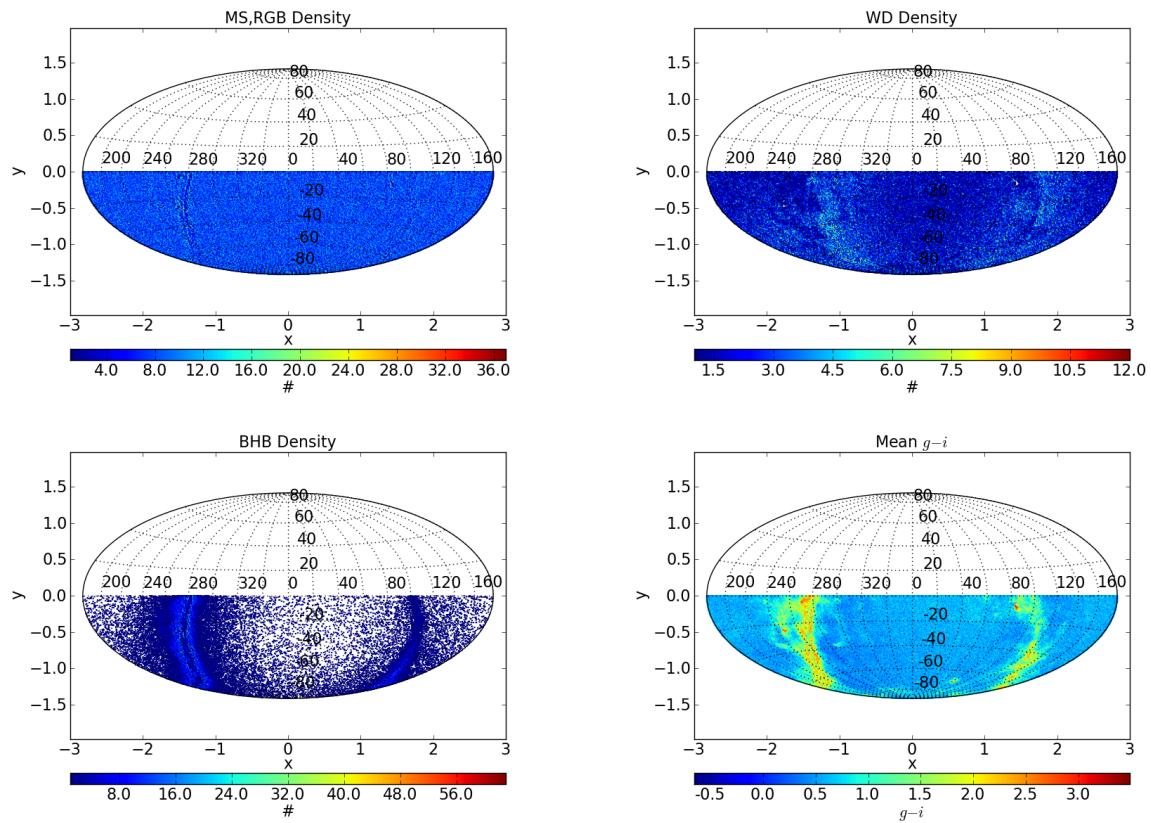


Fig. 1.— Basic properties of the GALFAST stellar catalog used to generate LSST observations. For our sub-sample of  $\sim 1$  million stars, we plot the densities of the different stellar populations along with the average  $g - i$  stellar color.

this  $5\sigma$  value when calculating the signal-to-noise ratio for each observation. We modify the  $5\sigma$  depth across the field of view to model the telescope vignetting. Opsim also includes the extinction from cloud cover. Because Opsim reports cloud extinction in a limited number of bins, we interpolate the Opsim values to fully cover the possible phase space. Opsim typically lists 60-70% of LSST observations as being taken in cloudless conditions, with the remainder of observations being taken through clouds with 0.1-0.8 magnitudes of extinction.

### Atmospheric Extinction

For each observation, we simulate the atmospheric extinction and cloud cover. One of the largest calibration issues for LSST will be dealing with cloud extinction. Unlike most previous surveys, LSST plans to observe in non-photometric conditions, thus we need to be able to perform accurate calibration even when stars are extinguished through cloud cover.

### Instrumentation Effects

We assume there is a systematic limited noise floor for all the photometric observations. We include a radially dependent color term to simulate the effect of a radially varying filter bandpass. We also have a generic color-term to simulate errors/uncertainties in the atmospheric throughput corrections. These are the terms that remain *after* the standard calibrations have been applied to the data.

## 2.1. Resulting Calibration Star Catalog

In building our calibration star catalog, we assume the recorded calibrated magnitudes can be expressed as the true magnitude offset by a number of offset terms that are a function of position on the focal plane, time, magnitude, color, etc.

$$m^{std} = m^{true} + \sum_i \Delta m_i(x, y, t, m, \sigma_5, g - i) \quad (1)$$

We emphasize that these  $\Delta m_i$  terms are what remain *after* the observations have been reduced (i.e., corrected for flat field variations, shifted to a standard bandpass, etc.), but have not had a zeropoint correction applied. We thus pull  $m^{true}$  values from the GALFAST catalog and generate values for  $\Delta m_i(x, y, t, m, \sigma_5, g - i)$ . Descriptions of all our  $\Delta m_i$  terms are listed in Table 1 and described in §2.2.

## 2.2. Simulated Offsets

The self-calibration simulations currently use up to 8 terms to simulate the errors that can be present in LSST standard magnitudes. We emphasize that these are offsets that

remain after the observed magnitudes have been corrected for atmospheric and instrumental effects and transformed to standard bandpass magnitudes. For example, there are currently no offset terms that depend explicitly on the airmass the observation was taken at, as we assume the LSST calibration procedure corrects for any color dependencies leaving only a gray extinction that is adequately modeled with a zeropoint shift. All our applied error terms are listed in Table 1.

The  $\Delta m_{snr}$  term is used to add Gaussian noise based on the signal-to-noise of a given measurement. We use the star’s magnitude along with the Opsim reported 5-sigma visit depth to calculate the value.

$$x = 10^{0.4(m^{std} - \sigma_5)} \quad (2)$$

$$\sigma_{rand}^2 = (0.04 - \gamma)x + \gamma x^2 (\text{mag}^2) \quad (3)$$

$$\sigma_{snr}^2 = \sigma_{sys}^2 + \sigma_{rand}^2 \quad (4)$$

Where  $\sigma_5$  is the  $5\sigma$  depth for a given visit from Opsim and  $\gamma$  is filter dependent and varies between 0.037 for  $u$  to 0.040 for  $z$  and  $y$ . The value for  $\sigma_{sys}$  is set to 3 millimags to simulate a systematic error floor for realistic photometry procedures. The simulation generates  $\Delta m_{snr}$  as a random number drawn from a Gaussian distribution with  $\sigma = \sigma_{snr}$  for each observation.

There are two color terms,  $\Delta m_{color}$  and  $\Delta m_{colordist}$ .  $\Delta m_{color}$  is a color term that simulates errors introduced from imperfect atmospheric corrections (e.g., if the auxiliary telescope incorrectly fits the atmospheric transmission model).  $\Delta m_{colordist}$  simulates the errors introduced by having the filter bandpass shape vary radially on the focal plane.

$$\Delta m_{colordist} = \Delta m_{filter} \left(1 + \frac{\Delta S}{S}\right) \left(1 + \frac{\Delta R}{R}\right) \left(1 + \frac{\Delta(g-i)}{(g-i)}\right) - \Delta m_{filter} \quad (5)$$

where  $\Delta m_{filter}$  is the offset caused by the variable filter throughput,  $S$  is the slope of the offset in mags per  $(g-i)$  per fraction of field of view (see Figure 3 in L2PC),  $R$  is the radial position on the field of view where the star falls. The terms  $\Delta S$ ,  $\Delta R$ , and  $\Delta(g-i)$  represent the difference between the measured and true values of the various quantities. See figures 9 and 10 in L2PC for examples of  $\Delta m_{filter}$  in LSST filters.

We have developed a robust procedure for generating realistic clouds based on previously observed cloud structure functions (Ivezić et al. 2007; Burke et al. 2010). Clouds are particularly difficult for LSST because the short exposure times means clouds have little time to drift across the field of view, resulting in extinction which is more structured than in surveys with longer exposure times like SDSS.

We first calculate an ARMA-based structure function, drawing from realistic distributions of physical scales for clouds and average extinctions from Opsim. The structure

function is then used to generate a cloud image with LSST’s field-of-view and sampled at a resolution similar to the average spacing between stars. An example of a simulated cloud is shown in Figure 2.

We have used FRED runs to calculate the expected illumination correction errors for each filter. Examples of the magnitude of the illumination corrections can be seen in Figure 3. For our simulations, we typically assume the illumination errors will be removed by a map derived from dense star field observations and measurements from the camera calibration optical bench (CCOB). We assume the pipeline reductions will only be able to remove these effects at the 90%-level.

### 3. Self-Calibration Algorithm

The L2PC document describes a number of steps in the complete photometric calibration procedure for LSST. The overall procedure can be summarized as separating the hardware and atmospheric throughput curves, and furthermore, separating the wavelength dependent and independent portions of the transmission curve throughput and shape. The wavelength-independent hardware throughput curve is normalized using a dome screen flat field in a fairly traditional manner. Wavelength-dependent (i.e. color-dependent) corrections to the hardware throughput curve are generated using narrow-band dome screen measurements. Wavelength-dependent corrections to the atmospheric response curve are generated using measurements from the auxiliary telescope. After applying all of these corrections to the extracted counts of objects in each image, we must still correct for wavelength-independent atmospheric transmission response variations due to cloud extinction and set appropriate zeropoints for each observation. This wavelength-independent correction of the atmospheric transmission curve is the primary function of the self-calibration procedure.

In the terms introduced in L2PC, we are starting self-calibration using magnitudes extracted from images which have been corrected to be in a standardized bandpass,  $\phi_b^{std}(\lambda)$ , where

$$\phi_b^{obs}(\lambda, t) = \frac{S^{atm}(\lambda, alt, az, t) S_b^{sys}(\lambda, x, y, t) \lambda^{-1}}{\int_0^\infty S^{atm}(\lambda, alt, az, t) S_b^{sys}(\lambda, x, y, t) \lambda^{-1} d\lambda}. \quad (6)$$

From the many repeat observations (under a variety of conditions) of the many different stars, we will then determine the best-fit zeropoints for each observation.

The self-calibration routine assumes that

$$m_{ij}^{std} = m_i^{best} + Z_j \quad (7)$$

where  $m_{ij}^{std}$  is the observed magnitude of star  $i$  corrected to a standard LSST bandpass in

Table 1. Offsets Generated in Selfcal Sims

$\Delta m$	Arguments	Description
snr	$m, \sigma_5$	Gaussian noise from signal-to-noise ratio + error floor
color	$(g - i)$	color term for a patch (imperfect atmospheric correction)
colordist	$(g - i), x, y$	color term calculated from the variable filter bandpass and placement
cloud_im	$x, y, t$	Offset calculated from a simulated cloud image
temperature	$x, y, t$	temperature variation which effect QE ( $y$ -band only)
expt	$t$	exposure time variation
gain	$x, y, t$	Variation in the gain
illum	$(g - i), x, y$	color-dependent errors from illumination and ghosting

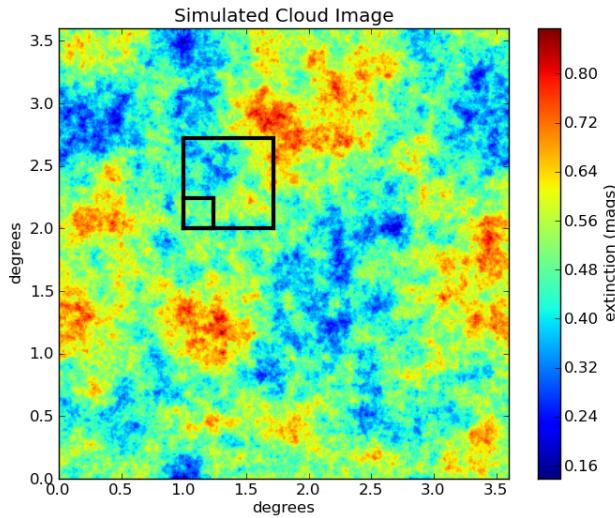


Fig. 2.— Example of a simulated cloud image with an average extinction of 0.5 magnitudes. The large square shows the size of an LSST raft while the small square shows the scale of a single LSST CCD.

patch  $j$ ,  $m_i^{best}$  is the best-fit magnitude for the star, and  $Z_j$  is the zeropoint for patch  $j$ . The patch sizes for  $Z_j$  will be approximately the size of a CCD, containing about 100 calibration stars on average. In most of our simulations, due to limitations on computer memory and processing time, we have simplified the calibration patches to be the size of an LSST raft (3x3 CCDs). The self-calibration procedure uses the multiple observations of stars to find the best-fit values for  $m_i^{best}$  and  $Z_j$  by minimizing

$$\chi^2 = \sum_{ij} \frac{(m_{ij}^{std} - m_i^{best} - Z_j)^2}{(\sigma_{ij}^{std})^2} \quad (8)$$

where  $\sigma_{ij}^{std}$  are the uncertainties in the observed magnitude. Minimizing Equation 8 requires solving for approximately  $10^8 m_i^{best}$  and  $10^8 Z_j$  terms in the basic calibration model. Adding additional terms to Equation 7 to refine the calibration of the hardware normalization or bandpass shape increases the number of parameters that must be fit only slightly, as these terms are assumed constant over large stretches (years) of observations. Of course, not all stars will be observed on all calibration patches, so there will be only about  $10^{10}$  non-zero values of  $(m^{std})_{ij}$  (per band).

Because we include radially dependent terms, we modify Eqn 7 to include an illumination patch correction

$$m_{ij}^{std} = m_i^{best} + Z_j + I_r \quad (9)$$

where  $I_r$  is an illumination correction. Unlike the patch zeropoint, the illumination correction stays the same from exposure to exposure.

We are currently using the LSQR algorithm, which is in the conjugate gradient family of solvers for solving the system  $Ax = b$ . Previous surveys have typically introduced a covariance matrix to perform the self-calibration (Padmanabhan et al. 2008; Schlafly et al. 2012). We have not introduced a covariance matrix and use LSQR because it has the advantage of having well-defined convergence criteria. Using alternative solvers, we were often forced to specify a number of iterations and found ourselves under-converged. LSQR often runs longer than strictly necessary, but stops iterating when it reaches the limits of machine precision.

For smaller simulations where we observe one-million stars at LSST cadence for 2 years, the resulting matrix  $A$  has 130 million non-zero elements and can be solved with modest computational resources.

If flux standard stars are not included in the fit, the self-calibration procedure will include a floating zeropoint (note that there is only one number per filter that converts the magnitudes to a flux-calibrated system).

A key assumption of the self-calibration algorithm is that the majority of stars are non-variable, or variable with amplitudes much smaller than the required calibration precision (0.01 mag). Based on SDSS results, at most 10% of stars will be variable at a level that would make them poor calibration sources. We assume that such variable stars will be easily recognized using their light curves (possibly using an iterative rejection scheme). Another crucial assumption is that for each LSST observation the field of view can be divided into a set of patches that each have a single photometric zeropoint. Presently, the zeropoints for these patches are set independently but it is likely that on photometric nights this assumption is too strong and will be relaxed (resulting in a smaller number of observations to close the system).

#### 4. Solving in Parallel With HEALpixels

The self-calibration problem for LSST is incredibly large. For comparison, the uber-cal of SDSS involved 36 million observations of 12 million unique stars. For LSST in the first two years we will have around 3.2 billion observations of 100 million unique stars. The memory and computation power required to solve such a system demands that we parallelize the problem in some way.

We have developed a technique using the Hierarchical Equal Area isoLatitude Pixelization (HEALpix) tessellation of a sphere. The HEALpix tessellation was originally designed for analyzing all-sky CMB observations. HEALpixels have equal area, and are distributed to make fast calculations of multipole moments and power spectra.

To run self-calibration in parallel, we assign each observation (consisting of a patch ID, star ID, observed magnitude, magnitude uncertainty, and possibly illumination patch ID) to the nearest four HEALpixels on the sky. This divides the observations in such a way that we have regions on the sky that are entire HEALpixels plus an surrounding border of approximately one-half HEALpixel. We have had good results with using 768 (53 square degrees) or 3072 (13 square degree) HEALpixels.

Once the observations have been divided, we run the self-calibration solver on each HEALpix region independently. By solving each HEALpixel in isolation, each result converges to a unique floating zeropoint. To tie the system back to a single floating zeropoint, we construct a matrix based on

$$P_{ij} = P_i + HP_j \quad (10)$$

where the  $P_{ij}$  are the patch zeropoints fitted on each HEALpixel. The equation is solved for the true patch zeropoint  $P_i$  and the HEALpixel floating zeropoint  $HP_j$ , completely analog-

gously to equation 7. If an illumination correction is included, we replace the patch ID with a unique ID for each patch ID and illumination ID combination present. At full density, we expect  $\sim 100$  stars per calibration patch and  $\sim 15$  observations per year. Thus, by using the patches to tie large scale solutions together, we have reduced the computational requirements by three orders of magnitude. For example, the two-year  $r$ -band simulation presented in §6.1 has  $1.3/\text{times}10^{11}$  non-zero matrix elements in the full self-calibration formulation, but the patch solution needs only  $3.2 \times 10^7$  non-zero elements.

This procedure is inefficient in the sense that each patch zeropoint and stellar magnitude is solved for 4 times. However, there is substantial speedup provided by not having to run the solver to convergence over very large spatial scales. Once the final patch zeropoints are solved for, we loop back through the data and apply the zeropoints (and possibly illumination solution) to the stellar observations and calculate a weighted mean for the final best-fit stellar magnitudes. We are currently weighting the patches in Eqn 10 by the number of stars they contain. This should probably be refined, as patches taken in cloudy conditions will be poorly fit even if they contain many stars.

The solutions returned by fitting in parallel are well-matched to solutions which solve the system simultaneously. Figure 5 compares a fit with the traditional global solver with a solution made with HEALpixels and shows the differences are typically less than 1 millimag.

Why does breaking the problem into HEALpixels work? Since LSST has multiple visits over the entire survey area, the problem is very well linked. Figure 6 shows how the fitting with HEALpixels fails if do not use the Opsim Hexdither option. With Hexdither, the position of pointings is offset by a fixed amount each night, resulting in more uniform depth for the survey. With Hexdither off, the self-calibration problem is poorly linked, and we are no longer able to solve regions independently and stitch the solutions back together to find the global solution.

We have not yet explored optimizing the amount of overlap needed when solving individual HEALpixels. XXX-expand. The choice to assign each observation to the four closest HEALpixels was done out of convenience rather than any deep mathematical insight.

## 5. Simulation Metrics

Our self-calibration simulations can be used to measure the spatial uniformity of the best-fitting solutions as well as the accuracy of repeat observations of a single object. Example plots showing the spatial uniformity and repeatability can be seen in Figure 7.

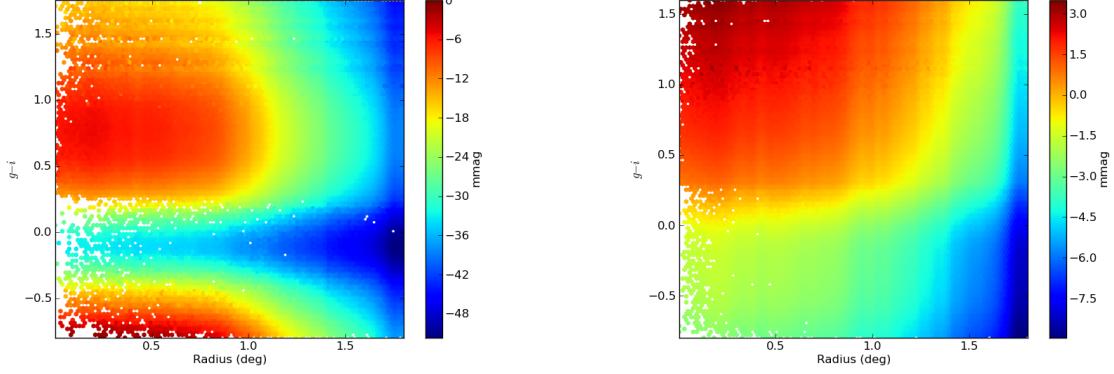


Fig. 3.— The illumination and ghosting error maps. left panel shows the  $u$ -band while the right panel is for  $r$ . For most of the simulations, we assume  $\sim 90\%$  of the illumination errors are calibrated out before passing magnitudes to self-calibration.

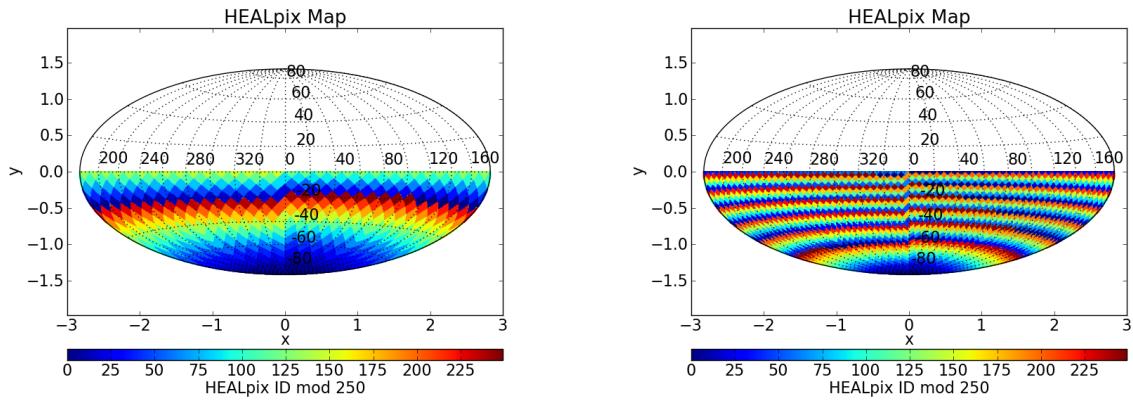


Fig. 4.— HEALpixel maps for 8 and 16 sides, resulting in 400 and 1568 individual pixels respectively.

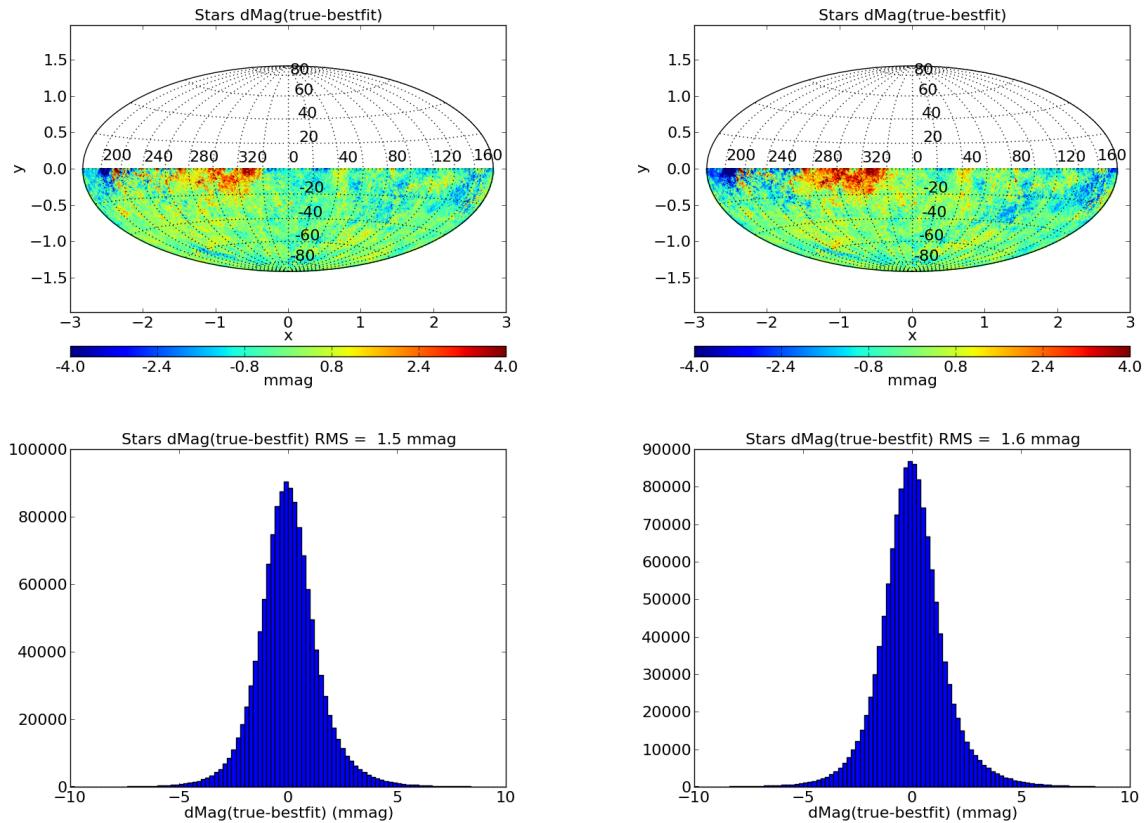


Fig. 5.— Comparison of a simulation solved simultaneously (left) with one solved on individual HEALpixels and combined. Statistically, the differences between the two solutions are at the sub-millimag level, much less than the differences we see between runs with different starting seeds (Figure 15).

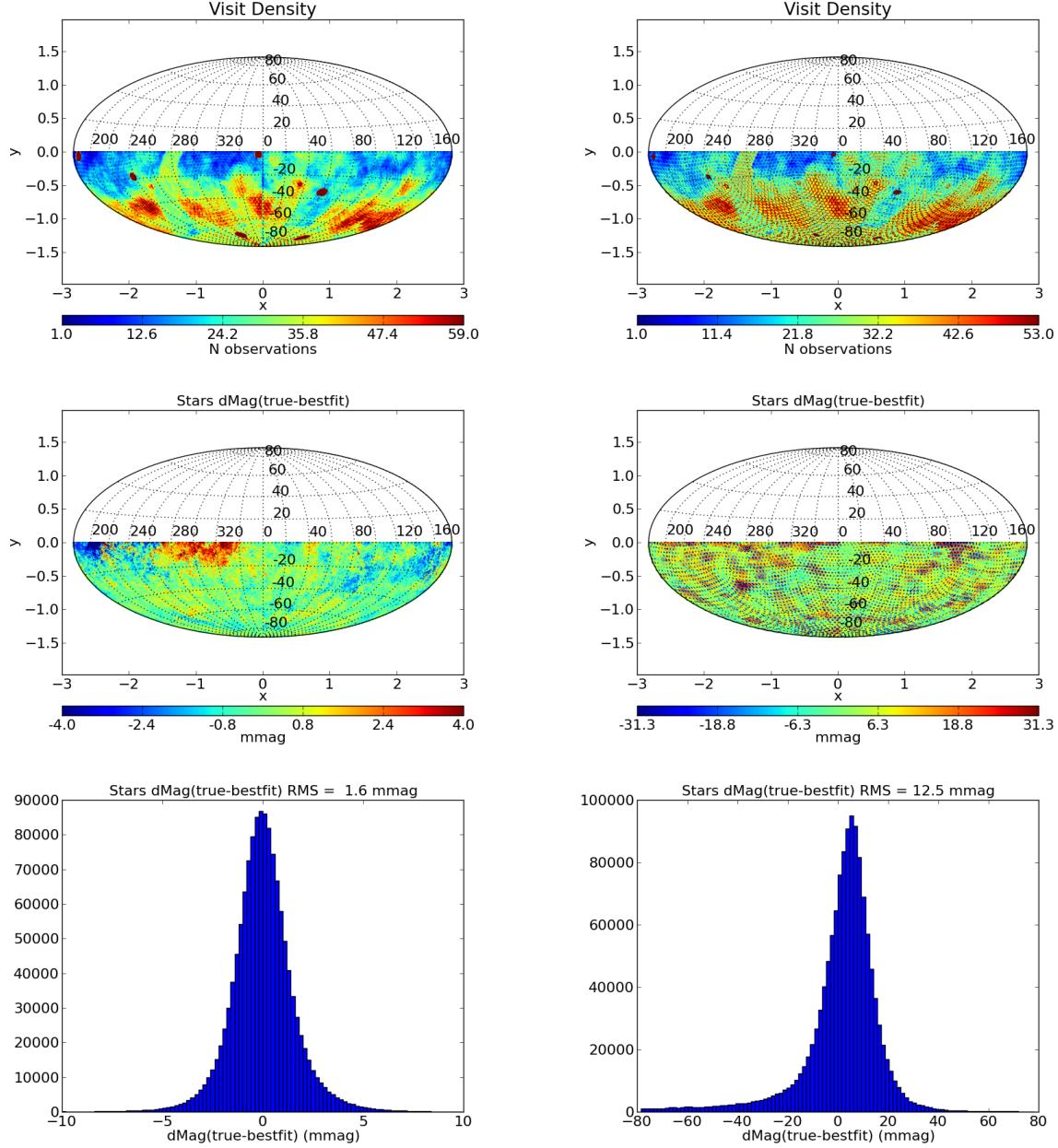


Fig. 6.— Example of when solving the self-calibration problem in parallel fails. On the left are the results from a 1-million star simulation in  $r$ -band. On the right, the same input parameters were used, but the Opsim hex-dither option was turned off, resulting in a survey that has much less spatial coverage uniformity. The top panel shows a map of the number of visits. The bottom two panels show how well stellar magnitudes are recovered by self-calibration. By turning off dithering, the self-calibration problem is poorly-linked and we are not able to solve the system in parallel. We are able to solve the system without Hexdither if we do not use HEALpixels and solve the entire system simultaneously.

The spatial uniformity is usually dominated by the visit density of a region of sky, with the expected result that sparsely observed areas of the sky are poorly fit while regions with many observations are smooth. The repeatability is dominated by cloud structure which is not well fit as a single zeropoint. Because cloud structure can generate large offsets, we use a robust-RMS by scaling the inter-quartile range (IQR). How residuals in individual observations departs from Gaussianity is shown in Figure 14.

Since most of our simulations were done with stars uniformly distributed over the sky, the distribution of true minus best-fit stellar magnitude is a good measure of the calibration uniformity. When we make simulations with variable densities, we average the residuals in HEALpixels (see §4).

Measuring how well we can derive stellar colors and tie the system to an external flux scale are discussed in §6.5.

## 6. Results

### 6.1. Full-Sky, Low-Density

Since a full-scale self-calibration simulation is still computationally expensive, we have run a series of simulations using only 1.3 million stars, distributed fairly uniformly across the sky. Since we are using fewer stars, we increase the calibration patch size from one LSST CCD to one LSST raft (3x3 CCDs). With fewer stars, we expect the patch zero-points will not be as well constrained as the full-scale self-calibration. We will also be more sensitive to the extinction structure in clouds with larger calibration patches. Despite these limitations, these simulations still provide useful insight to how well the Opsim pointings tie the sky together. Also, the results from these simulations can be considered worst-case scenarios, as the fits should only improve with more stars and higher resolution patches.

All of these simulations use the same sub-catalog of stars shown in Figure 1. This is one-million bright main-sequence and red giant stars, selected to be distributed fairly uniformly across the sky. There are additionally white dwarf and blue horizontal branch stars added, so we can better detect any color-dependent effects.

In general, areas that have fewer visits have the largest residuals. The repeatability is these simulations is dominated by the cloud cover. The repeatability is particularly bad in regions that have few visits and those visits were in clouds conditions. The visit distribution, uniformity, and repeatability for each filter are shown im Figures 7-12.

XXX-insert table of results compared to SRD requirements.

XXX-different seeds: Figure 15

XXX-the  $y$ -filter is poorly fit because the simulation includes temperature variation, which in turn change the effective quantum efficiency, at the CCD level. Since these simulation have raft-sized patches, the offsets from the temperature variations are not fit well.

## 6.2. Small-Sky, Full-Density

We have simulated observations on a small section of sky at full stellar density with CCD-sized calibration patches. This was two-years of Opsim pointing in  $r$ -band, and included all stars in the range  $17 < r < 21$ , resulting in 60 stars per patch. Results are shown in Figure 16. There does not appear to be substantial improvement over the fits which used fewer stars and larger patches.

## 6.3. Full-Sky, High-Density

XXX-still running. Results show in Figure 17.

## 6.4. Time Evolution

Figure 18 shows the time evolution of the self-calibration solution, using 1 million star simulations in the  $r$ -band. After one-year, the  $r$ -band still has regions that are not well connected to the rest of the survey. However, the solution does meet the SRD requirements for relative photometry repeatability even when regions are not well connected. After 2-years, the self-calibration solution gradually improves with increasing numbers of observations.

XXX 10 yr still running.

## 6.5. Flux Calibration and Colors

By default, the self-calibration procedure does not tie the solution to a physical flux scale. We keep track of which stars are WDs and select a subset as flux standards.

We assume that by the time LSST begins observations, there will be white dwarf standard stars available with  $\sim 1\%$  absolute flux precision. Here, we attempt to flux calibrate the low-density full-sky simulations from §6.1. Using only 10 WD flux standards, the absolute

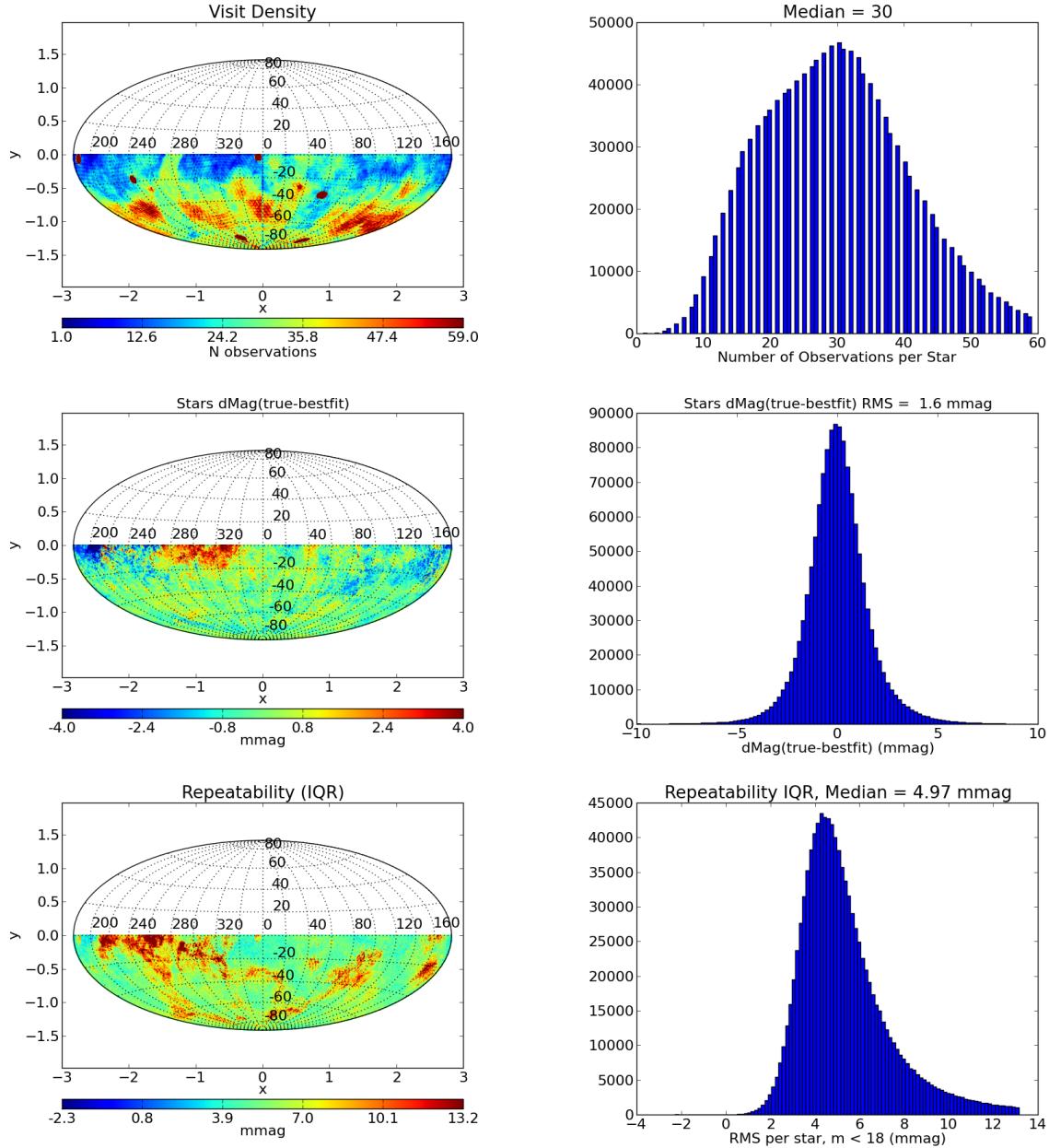


Fig. 7.— Full-Sky, Low-Density simulation for the  $r$ -filter.

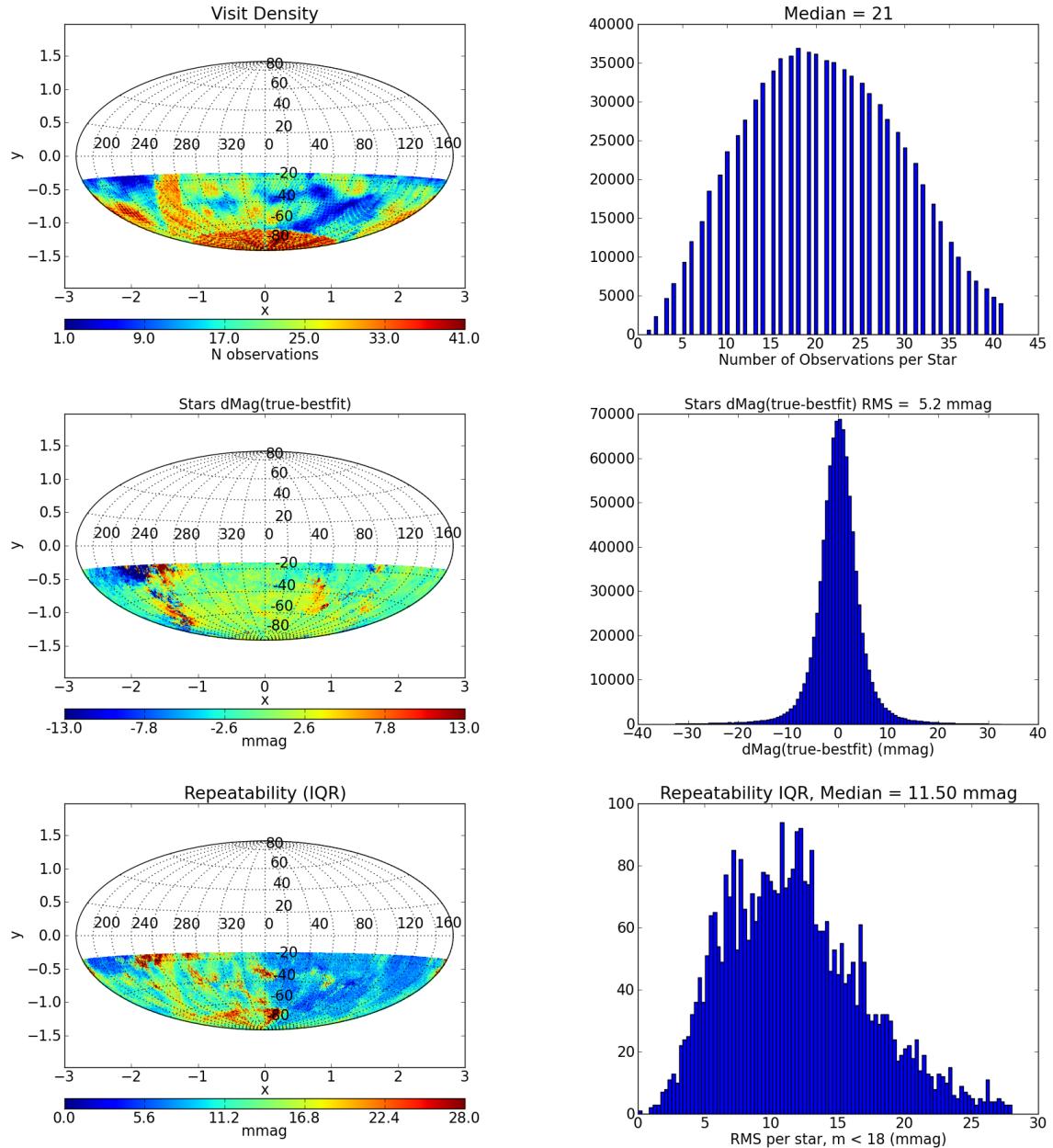


Fig. 8.— **Full-Sky, Low-Density simulation for the  $u$ -filter.** Note this is limited to declinations south of -15 degrees. After 2-years, the Opsim pointings in  $u$  do not cover the entire southern hemisphere.

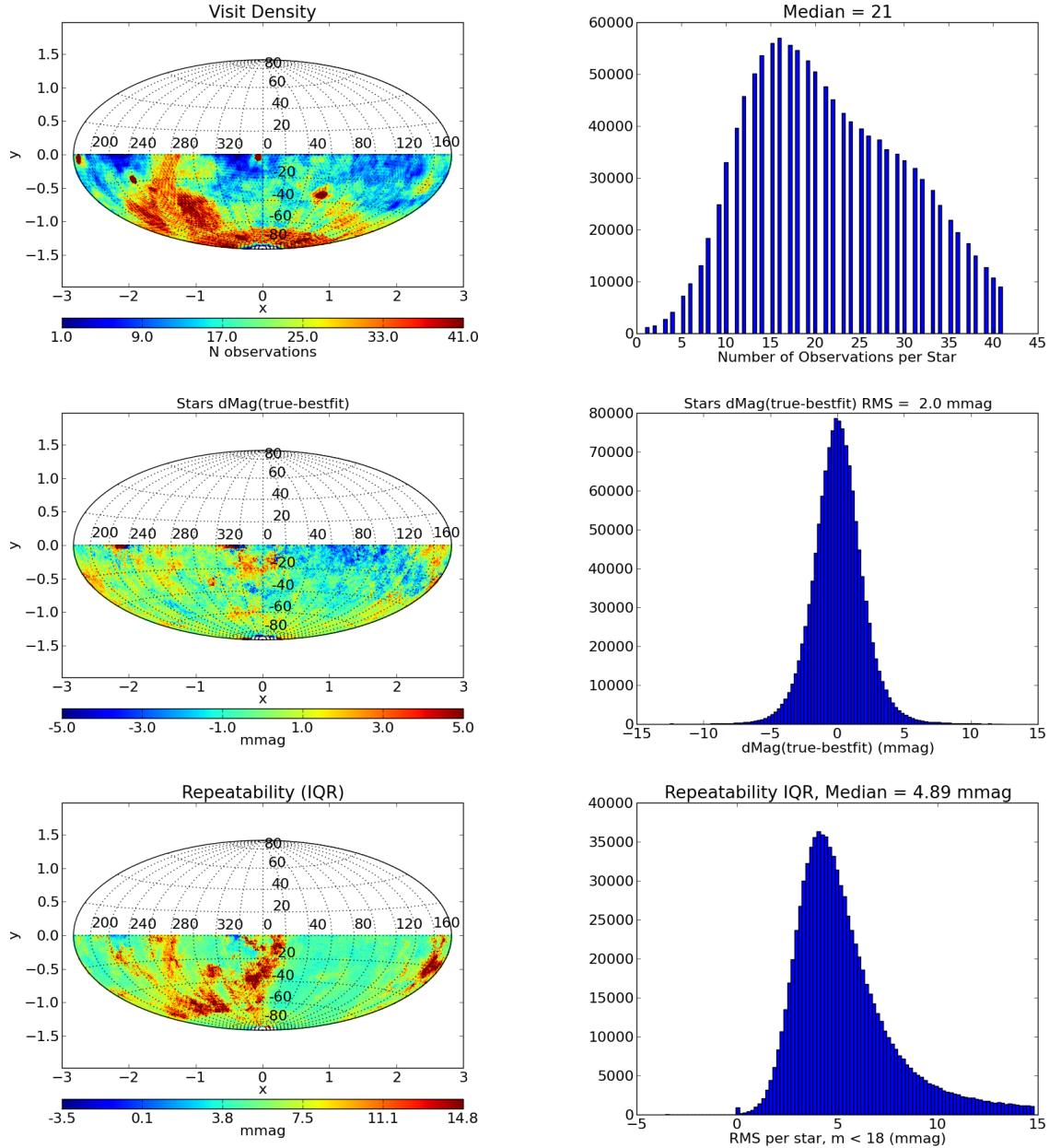


Fig. 9.— Full-Sky, Low-Density simulation for the  $g$ -filter.

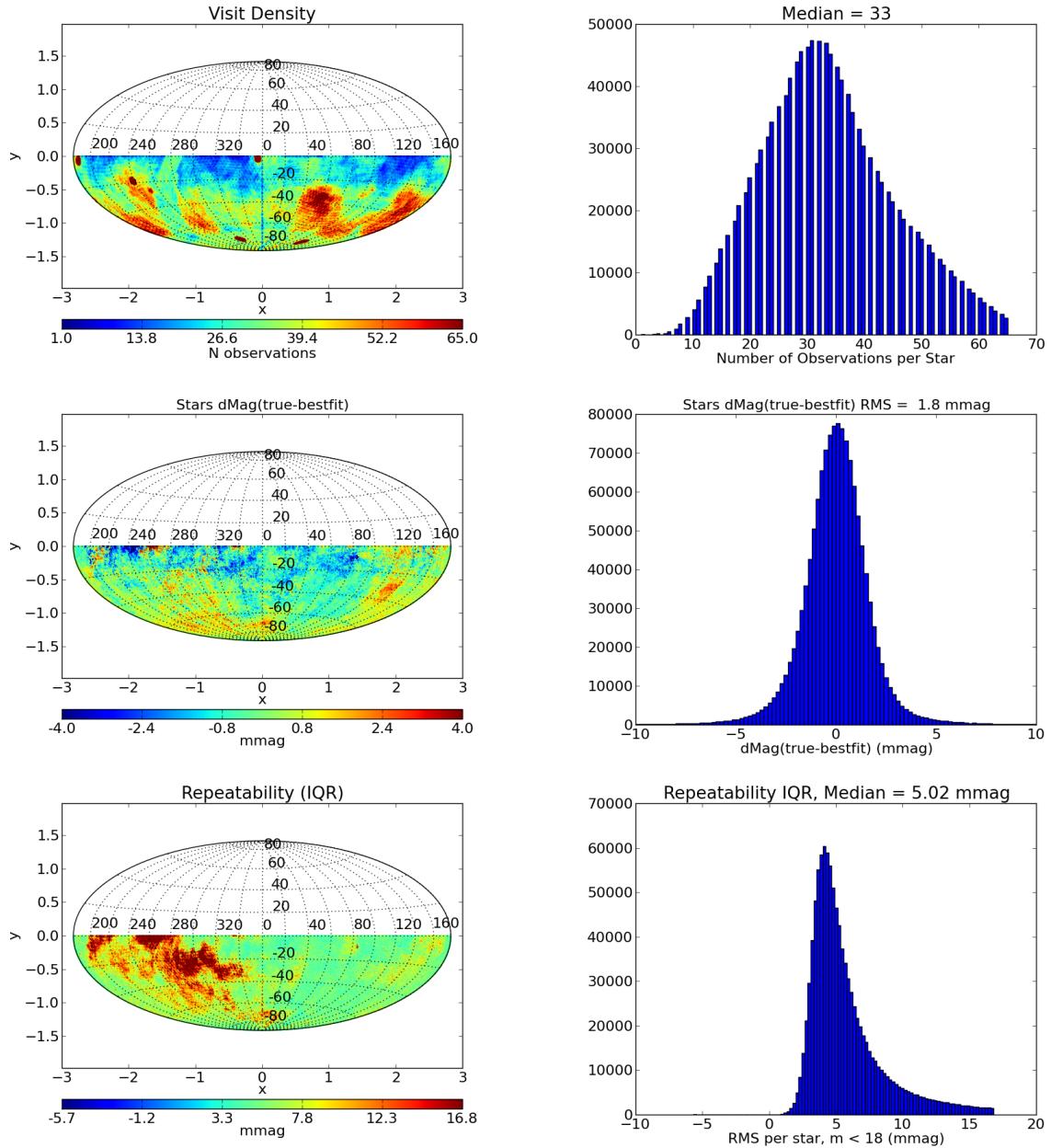


Fig. 10.— Full-Sky, Low-Density simulation for the *i*-filter.

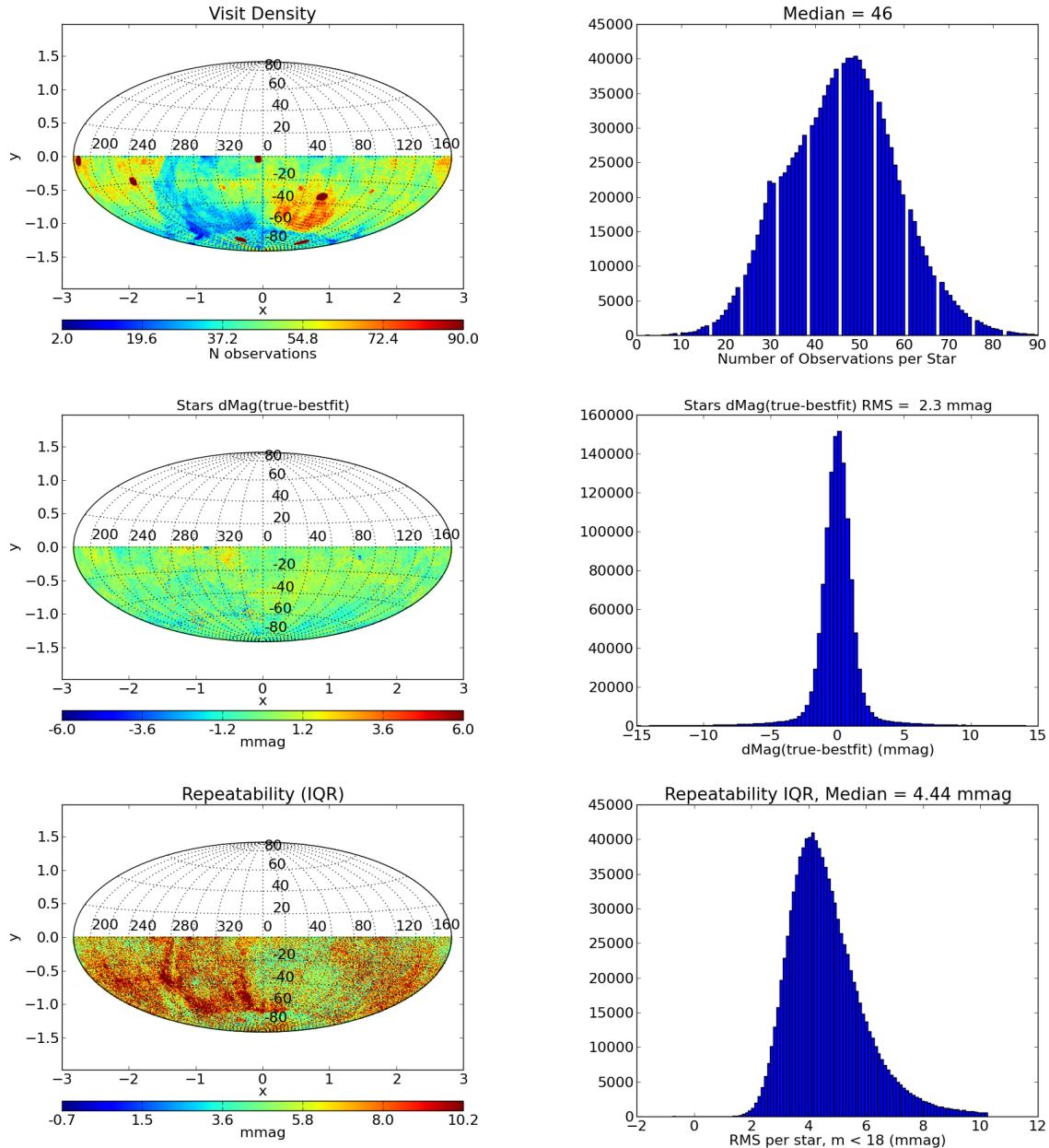


Fig. 11.— Full-Sky, Low-Density simulation for the  $z$ -filter.

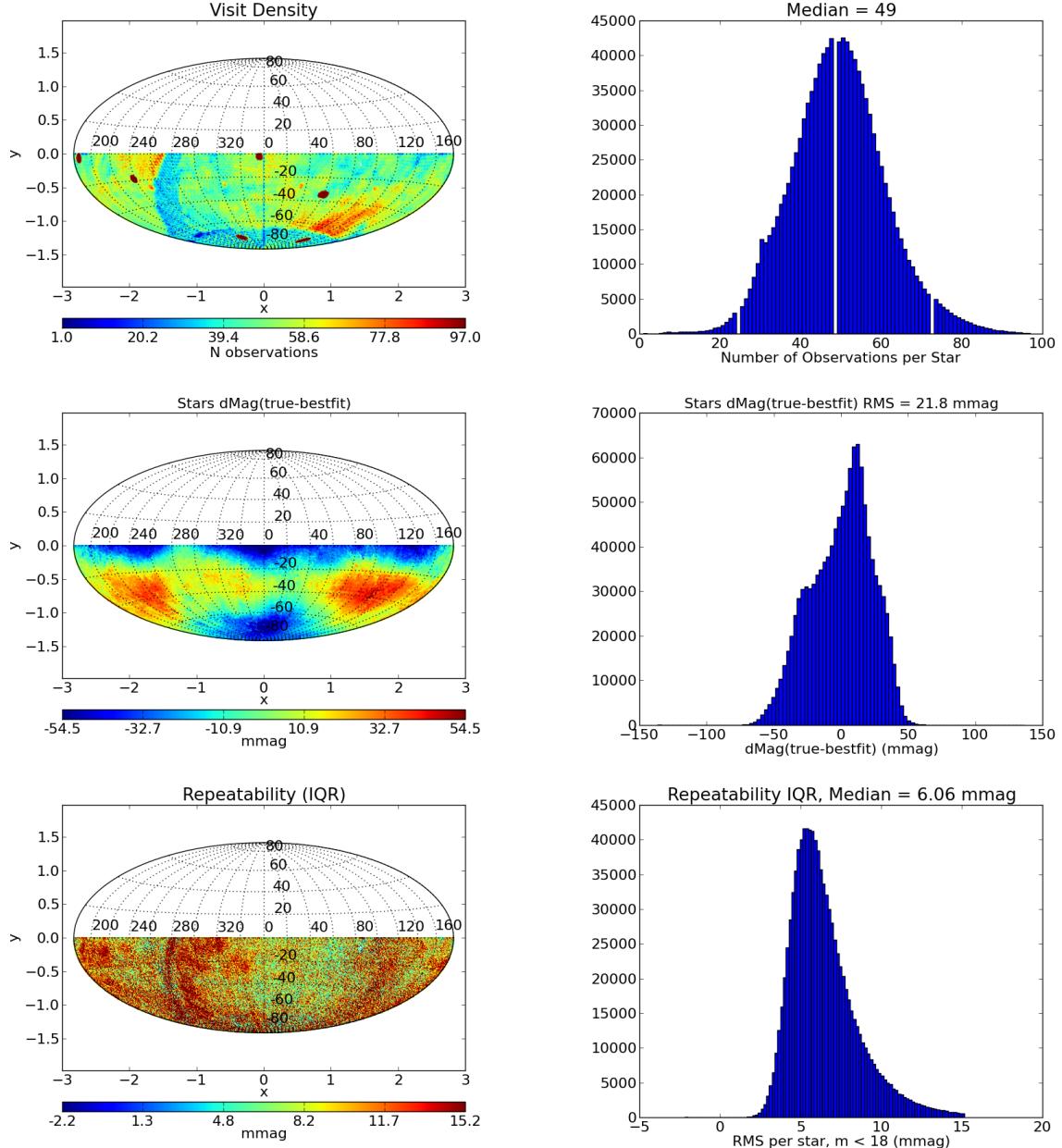


Fig. 12.— **Full-Sky, Low-Density simulation for the  $y$ -filter.** Since the  $y$ -filter simulations include temperature variation, our raft-size calibration patches do a poor job fitting the data. Figure 13 shows a  $y$ -band simulation with a higher density of stars and CCD-size patches.

Fig. 13.— **Partial-Sky, High-Density simulation for the  $y$ -band.**

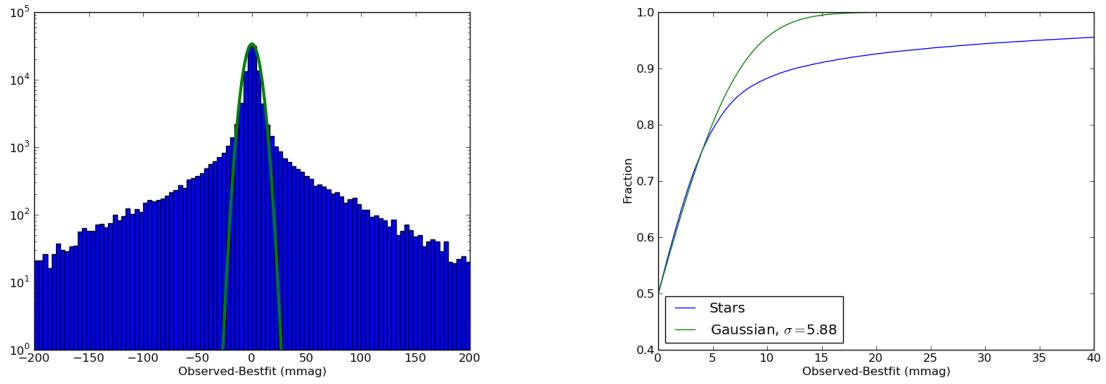


Fig. 14.— The residual distribution for individual observations is very non-Gaussian. This is an example of all the observations in a single  $r$ -band HEALpixel (1 million star simulation). The left panel shows the histogram of residuals of individual observations while the right shows the cumulative distribution. Both panels show a Gaussian for comparison. This shows for any individual star, about 60% of the observations follow a Gaussian distribution with  $\sim 5$  mmag RMS, while the other 40% of the observations are in a much broader tail. This well matches the Opsim input which has about 60% of observations taken in photometric conditions. The distribution should improve as we reduce the calibration patch size down to a single CCD.

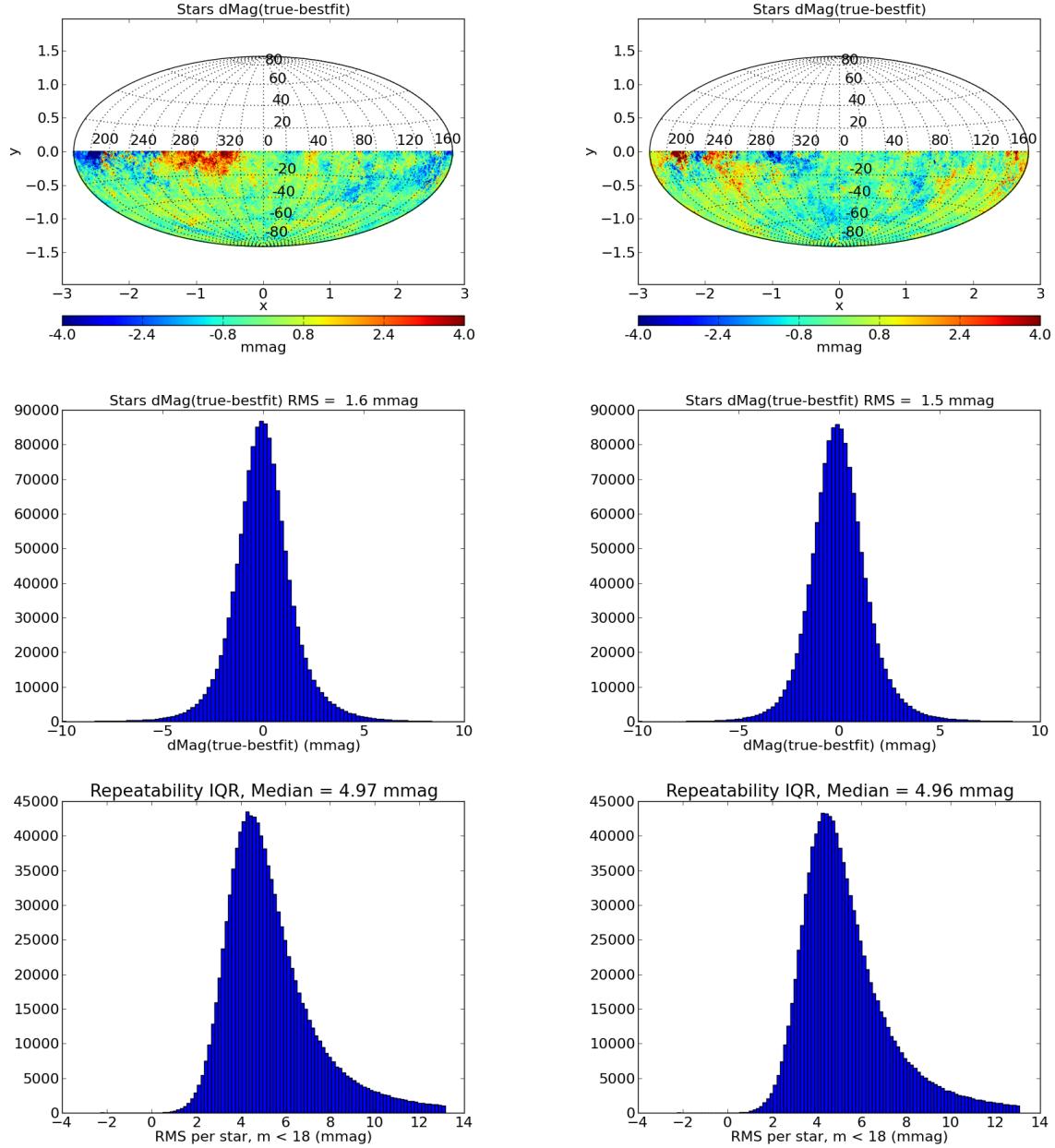


Fig. 15.— Comparing two  $r$ -band simulations started with different random number generator seeds. The residual histograms look very similar, but the sky maps show large changes in the structure.

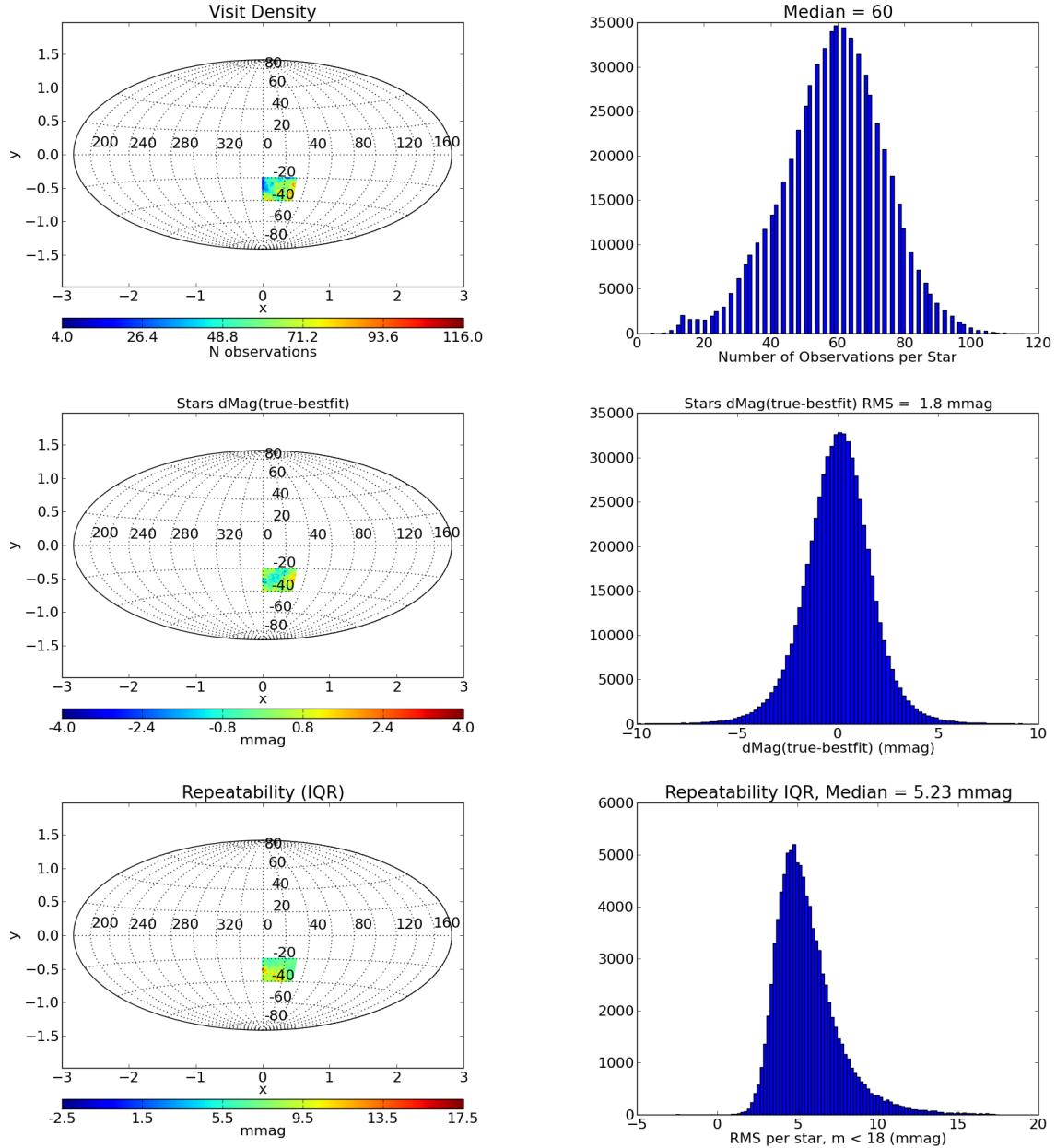


Fig. 16.— Small-Sky, Full-Density simulation for the  $r$ -filter.

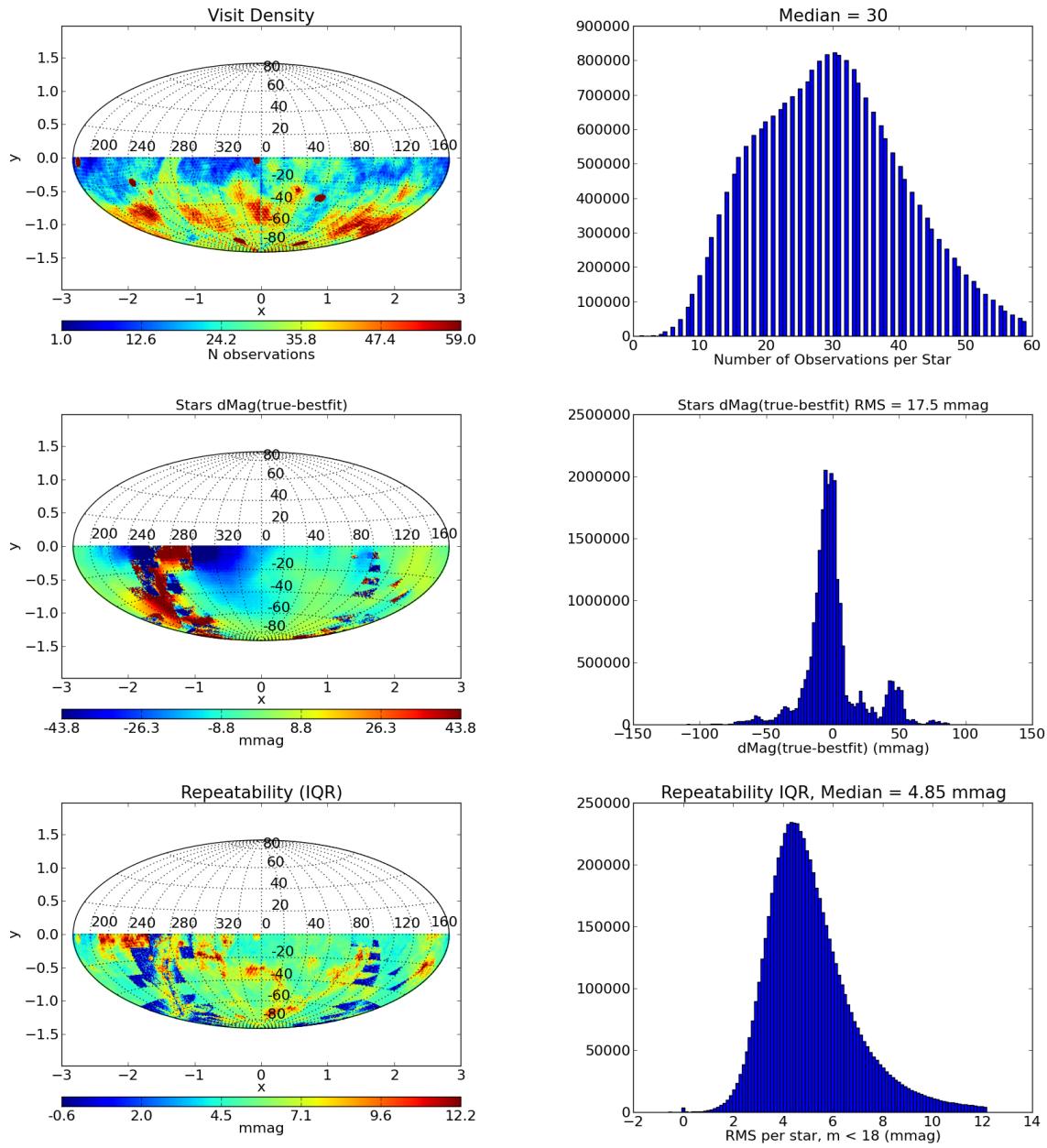


Fig. 17.— Full-Sky, High-Density simulation for the  $r$ -filter. XXX-updated plots coming, these suffer from strong stellar density gradients.

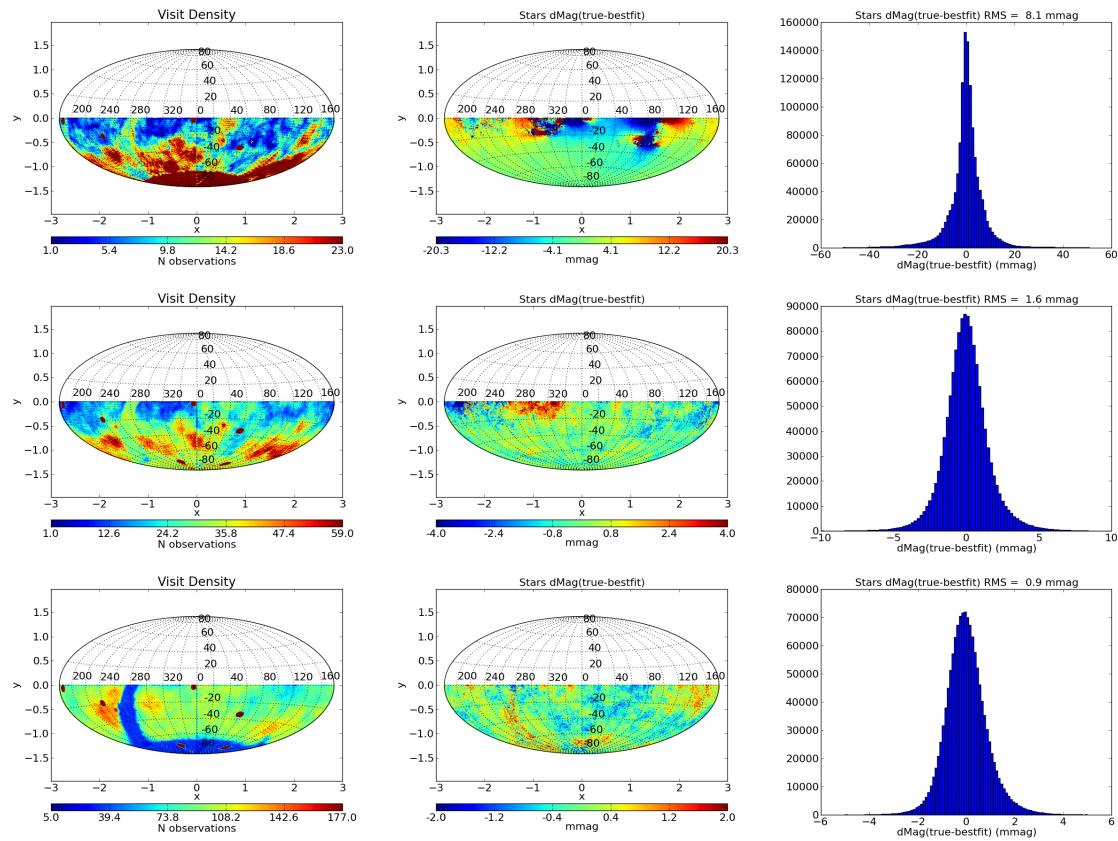


Fig. 18.— The evolution of the  $r$ -band self-calibration solution. From top to bottom, 1, 2, 5, and 10 years of  $r$ -band observations. xxx-10 year still running.

flux level in each filter can be set with an RMS of 3-5 mmag. If we have 30 standards, that drops to 2-3 mmag. Figure 19 shows the residuals of the colors after calibrating with 30 randomly chosen WD stars.

XXX-insert table comparing to SRD requirements

## 6.6. Magnitude Recovery

As a test of how accurately we recover patch zeropoints, we randomly selected 20% of the stars from the  $r$ -band full-sky low-density simulation and re-computed the self-calibration solution without including those stars. We then used the best-fit patch zeropoints to calculate the stellar magnitudes of the withheld stars. Figure 20 shows the residuals are nearly identical for stars included in the fit as those that were not, confirming that our zeropoint fits can be used for objects not included in the fit without introducing additional noise.

## 6.7. What Self-Calibration Can't Do

It is very easy to introduce degeneracies when adding additional terms to the self-calibration solver. For example, we were tempted to solve the illumination correction in separate color bins. Unfortunately, this makes the solution degenerate and each color bin is free to diverge on its own floating zeropoint.

Figure 21 illustrates how passing data with large color terms can result in poor fits. This was a  $u$ -band simulation where we did not remove the illumination and ghosting errors before running the solver. Normally, these errors are assumed to be removed at the 90%-level. Note that since the stellar sample is dominated by main-sequence stars, extreme blue and red stars are more poorly fit when there are color terms. This is especially concerning since it is blue white dwarf stars that are considered the best flux calibration standards.

## 7. Conclusions

A summary of lessons we have learned through running many self-calibration simulations.

- Dithering is important. If LSST observations are not dithered, the system is not linked well enough to solve the problem in parallel.

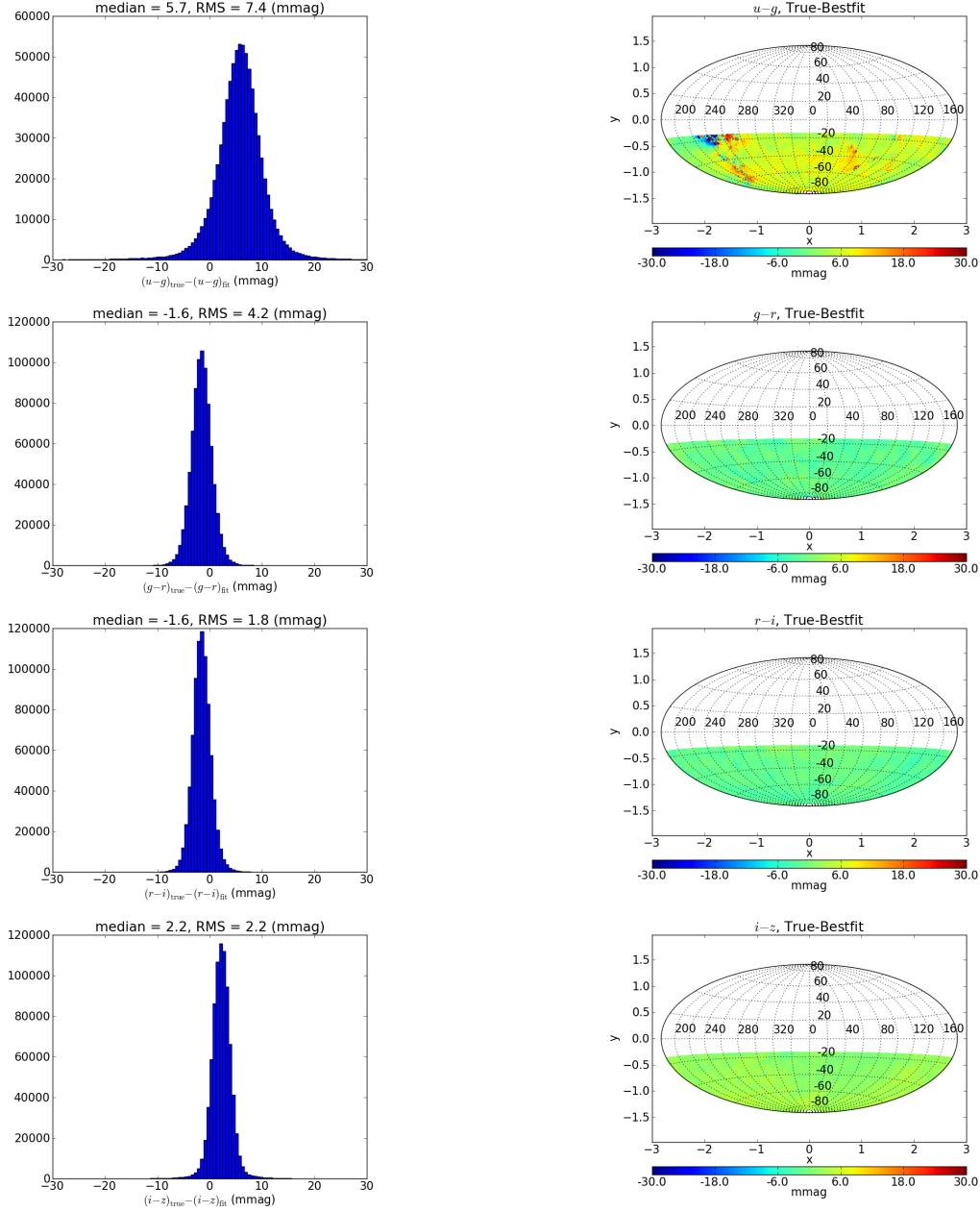


Fig. 19.— Color residuals after using 30 WD flux standards to move each band to a flux-calibrated system.

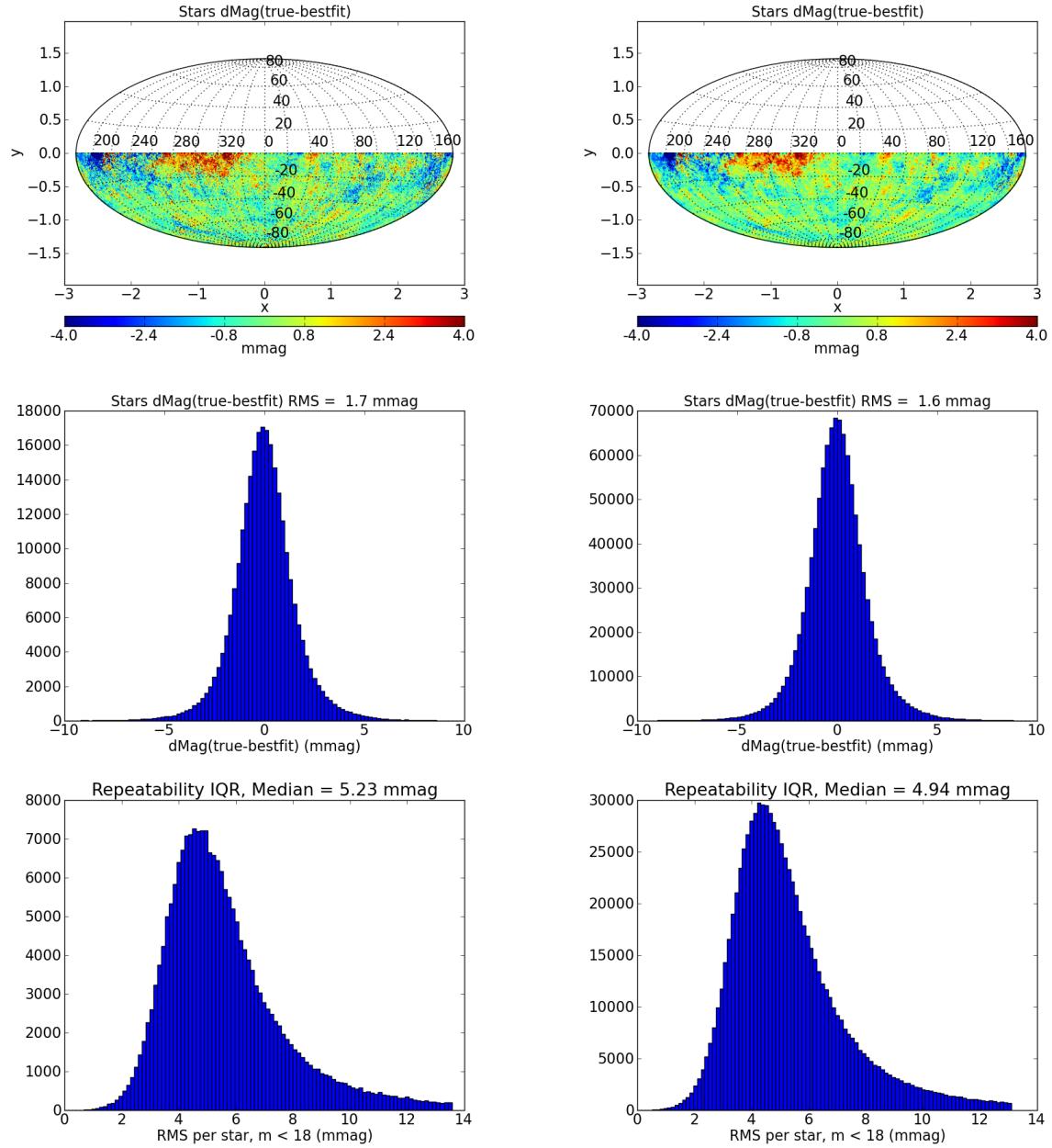


Fig. 20.— Recovering stellar magnitudes that were not included in the self-calibration fit. The panels on the left show the residuals of stars that were not included in the fit, while the panel on the right shows the residuals for stars that were.

- Even with a small number of stars and large calibration patches, self-calibration converges to meet SRD requirements for uniformity and repeatability.
- $u$  observations are very tough to calibrate. This is due in large part to Opsim focusing observations on the poles and leaving regions near the celestial equator disconnected from the rest of the survey
- The ability to meet the SRD uniformity requirement is very sensitive to color terms present in the data passed to the solver.
- The ability to meet the repeatability requirement depends strongly on the cloud cover and structure.
- We have demonstrated the ability to run self-calibration in parallel on LSST-sized data sets.
- With a sufficient number of flux standards, the self-calibration solutions can be moved to a flux-calibrated system, meeting the SRD color requirements.

## 8. Future Refinements

- We need to inspect the returned uncertainties on the fit parameters to see if they match the actual residuals
- Variable patch sizes. We can easily identify observations that were taken in cloudy conditions. For these observations, if there are an adequate number of stars, we could consider decreasing the patch size to remove additional cloud extinction structure.
- Stars in the galactic plane have no additional errors added due to crowding. In the future, it would be good to include the effects of crowded field photometry.
- A more complete simulation of the residuals left after fitting for the atmosphere.

## REFERENCES

Burke, D. L. et al. 2010, ApJ, 720, 811

Ivezic, Ž. et al. 2007, AJ, 134, 973

Padmanabhan, N. et al. 2008, ApJ, 674, 1217

Schlafly, E. F. et al. 2012, ApJ, 756, 158

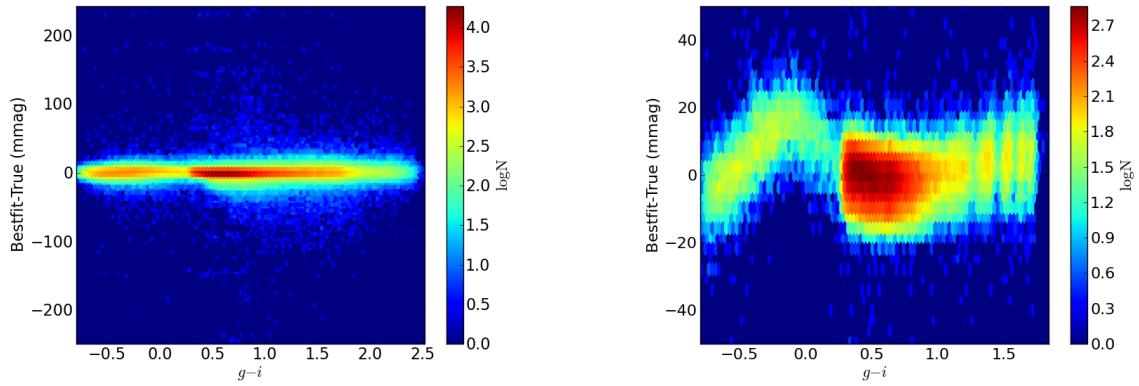


Fig. 21.— The left panel shows the standard  $u$ -band simulation, where the color-dependent illumination errors have been mostly removed before solving. On the right, the illumination errors are uncorrected before solving.