An end-to-end simulation framework for the Large Synoptic Survey Telescope

Andrew J. Connolly[1], George Z. Angeli[2], Srinivasan Chandrasekharan[3], Charles F. Claver[2], Kem Cook[2], Zeljko Ivezic[1], R. Lynne Jones[1], K. Simon Krughoff[1], En-Hsin Peng[4], John Peterson[4], Catherine Petry[2], Andrew P. Rasmussen[5], Stephen T. Ridgway[3], Abhijit Saha[3], Glenn Sembroski[4], Jacob vanderPlas[1], Peter Yoachim[1]

[1]Dept. of Astronomy, University of Washington, Seattle, WA, 98195, [2]LSST, Tucson AZ, 85719, [3]NOAO/LSST, 950 N Cherry Ave. Tucson, AZ 85719, [4]Dept. of Physics, Purdue University, Lafayette, IN, [5]SLAC National Accelerator Lab, 2575 Sand Hill Road, Menlo Park, CA 94025

ABSTRACT

The LSST will, over a 10-year period, produce a multi-color, multi-epoch survey of more than 18000 square degrees of the southern sky. It will generate a multi-petabyte archive of images and catalogs of astrophysical sources from which a wide variety of high-precision statistical studies can be undertaken. To accomplish these goals, the LSST project has developed a suite of modeling and simulation tools for use in validating that the design and the as-delivered components of the LSST system will yield data products with the required statistical properties. In this paper we describe the development, and use of the LSST simulation framework, including the generation of simulated catalogs and images for targeted trade studies, simulations of the observing cadence of the LSST, the creation of large-scale simulations that test the procedures for data calibration, and use of end-to-end image simulations to evaluate the performance of the system as a whole.

Keywords: simulations, LSST, cosmology, image, catalogs,

1. INTRODUCTION

The LSST is a large, wide-field and ground-based optical survey that will image half the sky every few nights. Designed to operate from Cerro Pachon in Northern Chile, the LSST comprises an 8.4m (6.7m effective diameter) primary mirror and a 3.2 Gigapixel camera. With a 9.6 degree^2 field-of-view, it will visit each part of its 18000 degree^2 primary survey area ~1000 times over the course of 10 years. Each visit will comprise a 15 second pair of exposures with a single visit depth of ~24.5 magnitudes (AB) (in the six bands u, g, r, i, z, and y). There are four primary science drivers for the LSST project: the characterization of dark energy through the multiple cosmological probes (e.g. gravitational weak lensing, luminosity distances from Type Ia supernovae, and Baryon Acoustic Oscillations); mapping the 3D distribution of stars within our Galaxy; a census of solar system objects within the Solar System; and a detailed study of the transient and variable universe. The requirements associated with these science objectives are described in the LSST Science Requirements Document (SRD)[1]. These requirements include: absolute photometric accuracy of 1%, relative photometric calibration of 0.5%, astrometric uncertainties of 0.01 arcsec (for point sources on scales of 20 arcmin), and residual correlations in the point-spread-function ellipticities of 10^{-7} (on scales >5 arcmin). Throughout pre-construction, construction and commissioning, the LSST must be capable of demonstrating that it can achieve these requirements given its design and as-delivered components, that the system can be calibrated to the required level of fidelity, that the data management software can extract the appropriate astrophysical signals, and that this can be achieved with sufficient efficiency such that the telescope can complete its primary objectives within the ten years of its duration. Realizing these objectives requires that the project be able to characterize the performance of the LSST including the performance of the opto-mechanical systems, the response of the detectors and their electronics, and the capabilities of the analysis software.

The LSST simulation framework is designed to provide such a capability; delivering a virtual prototype LSST against which design decisions, optimizations (including descoping), and trade studies can be evaluated.

2. AN OVERVIEW OF THE LSST SIMULATIONS AND MODELLING TOOLS

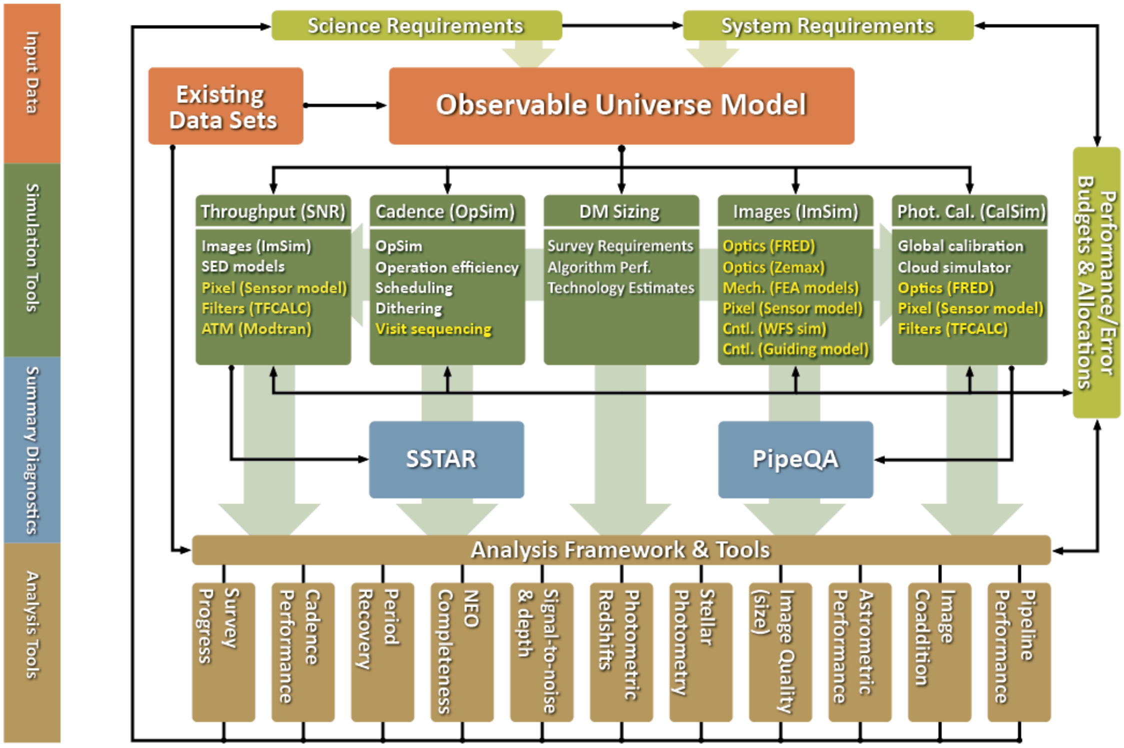

Figure 1. A schematic of the tools available and used within the LSST for modeling and simulations. The orange boxes represent a model of the LSST universe comprising stars, galaxies, and solar system objects. Green boxes depict individual science simulation tools each with different levels of fidelity and complexity. The brown boxes represent the types of investigations that can be accomplished by combining the individual simulation tools. Items highlighted in yellow correspond to engineering tools (e.g. ZEMAX) that are used to validate the fidelity of the science simulations.

Figure 1 illustrates the type of simulations and modeling tools available to the LSST project and the types of applications that these tools have been used for. The LSST Systems Engineering group manages the LSST simulation framework. This framework comprises three main components: a catalog simulator (CatSim) capable of returning catalogs of astrophysical sources (e.g. stars, galaxies, and solar system objects) with properties and noise characteristics that are representative of what the LSST will observe to its coadded depth; an image simulator (Phosim) that is capable of returning images with characteristics consistent with the design of the LSST (i.e. with astrometric, photometric and ellipticity distributions that are appropriate for a large, wide-field long exposure telescope); and an observing strategy simulator (OpSim) that can generate sequences of LSST observations (and their summary statistics) that meet the 10 year cadences and depths required by the survey (while accounting for the expected performance of the telescope and site). In isolation each of these simulation components provides a broad range of capabilities; from the generation of the statistical properties of a year's worth of observations to targeted simulations of stars to evaluate how well the point-spread-function can be interpolated across a sensor. Together, the simulation components enable end-to-end simulations that can trace the properties of the LSST system from the underlying cosmology through to derived data products.

The need for science simulations arises because the requirements described in the SRD are a simplification of a complex system that incorporates the physics of the universe, the performance of the subsystems, and our ability to analyze these data under varying conditions. Engineering simulations such as Zemax[2] or FRED[3] have been used to define the optical design of the system. While detailed, these modeling tools do not couple the astrophysical properties of the sky nor the changes in observing conditions to the system performance. They are not designed to scale to the size of simulations of the LSST universe with 20 million sources per focal plane image (to a coadded depth of $i=26.8$). In contrast, science or system simulations provide the ability to take a value specified within the SRD, which incorporates opto-mechanical, atmospheric, electronic, and software components together with the underlying astrophysical distributions of sources and evaluate which systematic uncertainties are most sensitive to individual components (i.e. assuming we can model the simulation components at the appropriate level of fidelity). A science simulation framework can provide an end-to-end implementation of the full flow of photons and information to evaluate how well we can achieve the SRD requirements or a simplification of the flow of information to identify the subcomponents and their contribution to the overall performance.

2.1 The LSST Universe Simulator

The model for the LSST universe is designed to provide a representative view of the night sky above the atmosphere. The galaxy simulation is based on dark matter haloes from the Millennium and a semi-analytic baryon model described in De Lucia et al.[4] The semi-analytic model features radiative cooling, star formation, the dynamics of black holes, supernovae, and AGNs and was adjusted to mimic the luminosity, color, and morphology distributions of low redshift galaxies[4]. LSST cosmological catalogs were generated from the De Lucia et al. data by constructing a lightcone, covering redshifts $0<z<6$ from 58 500 h^{-1}Mpc simulation snapshots. Comparisons between the redshift and number-magnitude distributions of the simulated catalogs with those derived from deep imaging and spectroscopic surveys showed that the De Lucia models under-predict the density of sources at faint magnitudes and high redshifts. To correct for these effects, sources in the original catalogs are "cloned" in magnitude and redshift space until their densities reflect the average observed properties. The final catalog comprises a 4.5x4.5 degree footprint on the sky (sufficient to cover a single LSST field-of-view) and samples halo masses over the range 2.5×10^{9} to 10^{12} $M_{\odot}$.

Dynamically tiling this footprint across the sky enables the simulation of the full LSST survey area while keeping the underlying data volume small (but at the expense of introducing periodicity in the large scale structure). For all sources, a spectral energy distribution (SED), is fit to the galaxy colors using Bruzual and Charlot spectral synthesis models[5]. The De Lucia et al. catalog includes BVRIK magnitudes and dust values for the disk and bulge components of each galaxy as well as radii, redshift, coordinates, stellar age, masses and metallicities. Fits are undertaken independently for the bulge and disk and include inclination dependent reddening. Morphologies are modeled using two Sersic profiles and a single point source (for the AGN). Bulge-to-disk ratios and disk scale lengths are taken from De Lucia et al. Half-light radii for bulges are derived from the absolute-magnitude vs half-light radius relation given by Gonzalez et al[6]. Colors and stellar mass of the AGN host galaxies are estimated from the AGN luminosities. These parameters are used, together with the redshift of the AGN, to assign each AGN to a galaxy in the galaxy catalog.

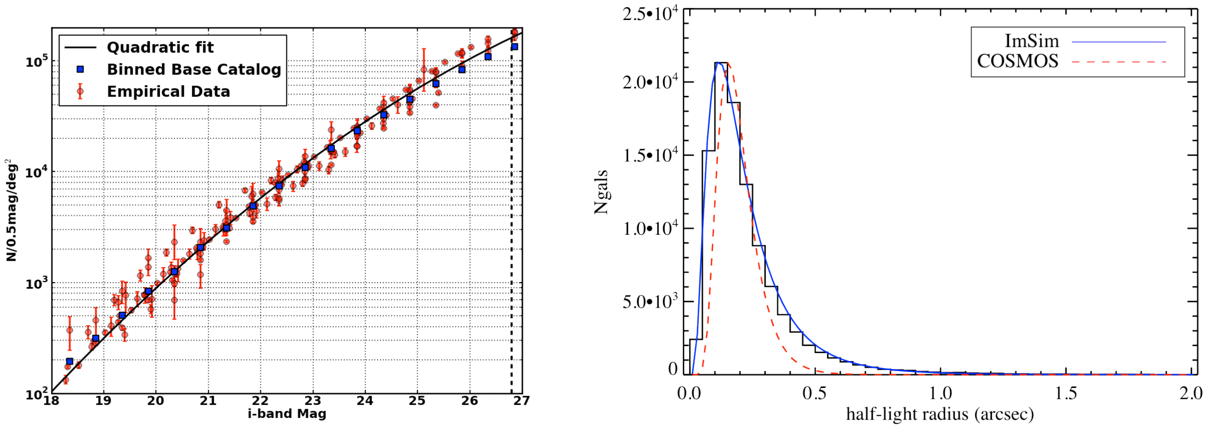

Figure 2 Models of the observable universe include the clustering properties of galaxies, distributions of stars that follow the Galactic structure, and solar system sources. These are designed to reproduce the key observable properties that drive the survey. For example, the left panel compares the simulated density of galaxies (blue points) with observed galaxy number counts (red points) and the right panel compares the predicted size distributions of galaxies (blue line) with those observed from the COSMOS HST survey.

Stars are represented as point sources and are drawn from the Galfast model of Juric et al.[7]. Galfast generates stars according to density laws derived from fitting SDSS data to a model of a thick and thin disk, and a halo. Using an input luminosity function measured from SDSS for each class of star (e.g. main sequence, white dwarf, blue horizontal branch, etc.), Galfast samples stars in space and magnitude from a 4-dimensional probability density function. After this stage, using Fe/H and kinematics models from Bond et al. (2010)[8], each star is assigned a metallicity, proper motion, and parallax. Spectral energy distributions are fit to the predicted colors using the models of Kurucz[9] for main sequence stars and giants, Bergeron et al.[10] for white dwarfs, and a combination of spectral models and SDSS spectra for M, L, and T dwarfs. For Galactic reddening, a value of E(B-V) is assigned to each star using the three-dimensional Galactic model from Amores and Lepine[11]. For consistency with extragalactic observations, this reddening model is re-normalized to match the values in the Schlegel et al.[12] dust maps at a fiducial distance of 100 kpc.

Approximately 10% of the stellar sources are variable at a level detectable by LSST. Variability is modeled by defining a light curve, its amplitude, a period, and a phase. For queries that contain time constraints, the magnitude of the source is adjusted based on the properties of the light curve (the current implementation only allows for monochromatic variations in the fluxes). Current variability models include cataclysmic variables, flaring M-dwarfs, and micro-lensing events. For transient sources, the period of the light curve is set to >10 years such that the sources will not repeat within the period of the LSST observations.

The Solar System model is a realization of the Grav et al.[13] model. All major groups of Solar System bodies are represented including: main belt asteroids, near-Earth objects, trojans of the major planets, trans-Neptunian objects, and comets. There are approximately 11 million objects in the Solar System catalog with the vast majority (about 9 million) being main belt asteroids. Populations are complete down to apparent magnitudes of V=24.5 and each object is assigned a carbonaceous or stony composition spectrum. The choice of a C or S type spectra for an object is based upon a simple relation to the size of its orbit that approximately matches SDSS asteroid observations. The location of the Earth at the time of a particular observation is incorporated through the orbital ephemeris software OOrb.

2.2 The Throughput Simulator

To manage the broad range of use cases for the LSST project (from all-sky catalogs used in modeling the LSST calibration pipeline, to time domain data used to characterize variability as a function of signal-to-noise and temporal sampling, to sequences of images of gravitational lenses from which to measure cosmological time delays) the tools available in the LSST simulation framework trade fidelity for computational complexity. An example of this is the Throughput Simulator, which is included within the CatSim package. The Throughput Simulator takes as input spectral energy distributions, together with the transmissivity and reflectivity of the optical surfaces (e.g. the mirrors and lenses), the transmission of the atmosphere (as a function of airmass), and the integration time and readout noise of the sensor. From this the photometric properties of a source can be calculated. An example of the use of the

Throughput Simulator is shown in Figure 3 where the dashed lines give the throughput of the LSST u, g, r, i, z, and y filters, the solid lines are the reflectivity and transmission of the optical surfaces, and the resulting effective throughputs of the LSST (as a function of filter) are shown at the bottom of the plot.

The throughput simulator is integrated within CatSim through a Python interface and includes methods for integrating an SED over multiple bandpasses, for calculating fluxes and colors, for estimating the signal-to-noise as a function of readout noise or exposure time, and for predicting the expected number of photons within an exposure.

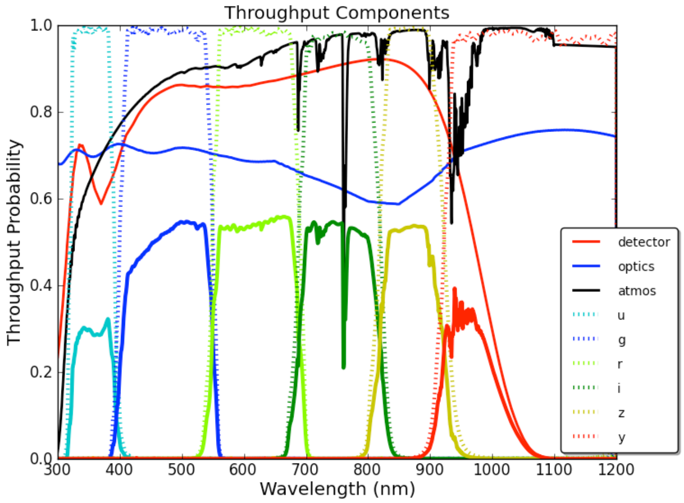

Figure 3 Modeling of the throughput of the LSST (including the atmosphere, optics, sensor efficiencies and noise properties) is undertaken using a python-based tool that can estimate the 5σ depth of the survey as a function of integration time, observing conditions and source spectra energy distribution. The effective throughput as a function of filter is shown as solid lines at the bottom of the plot.

2.3 The Operations Simulator

The Operations Simulator (OpSim) is designed to simulate sequences of visits (i.e. pairs of 15 second exposures) and is described in detail in Ridgway et al[15]. To accomplish this, the simulator takes as an input a description of the configuration of the LSST system including its dynamical and mechanical properties (e.g. the slew time, dome acceleration, readout time for the camera, filter exchange time etc), a simplified model for the weather (based on 10 years of historical records from the Cerro Tololo Inter-American Observatory), a model for the sky brightness[16], and a series of observing proposals that characterize the cadence required by the LSST science cases (e.g. how often must a point on the sky be observed in order to characterize a supernovae lightcurve).

In the current implementation of OpSim, to create the simulated exposures, the 18,000 square degrees of the primary LSST survey area are divided into 2293 overlapping fields (each 9.6 sq degrees). OpSim simulates the performance of the LSST scheduler by continuously calculating the priority of a field in terms of when an observation needs to occur. In considering the next field to observe, this scheduler computes the sky brightness and weather for the observations, the visibility of the field, and the priority of that particular observation given a science proposal or objective (including the history of the previous observations for that proposal and field). These science and visibility ranks are combined with the cost of the observation in terms of slew time (from a dynamical model of the observatory) and the highest priority target (or observation) is selected. The current implementation of the scheduling algorithm takes a greedy approach when calculating ranks by searching through all potential fields before selecting the appropriate next observations.

The output of these ranks and the sequence of observations generated by OpSim are stored in a relational database. An example of the how these data are used within the LSST project is shown in Figure 4. This Hammer-Aitoff projection of the sky shows, for each of the six LSST filters, the number of observations as a function of position on the sky. The color reflects the number of visits after 10 years of observations. From these data, the ability of the LSST to meet the requirements described in the SRD can be characterized as a function of the system properties (e.g. how many visits can be delivered in 10 years as a function of readout time, or filter exchange time), and the sensitivity of the LSST's science requirements to any as-delivered components can be calculated.

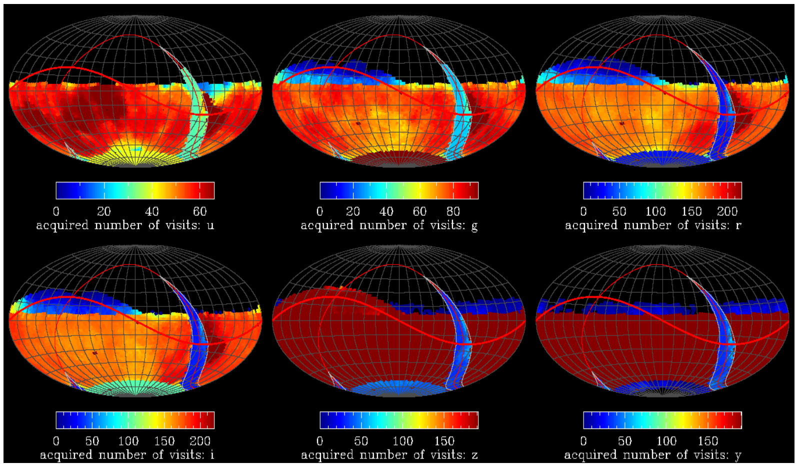

Figure 4. An example output from the Operations Cadence Simulator (OpSim). The panels show the distribution of the number of observations (visits) as a function of the survey footprint for the u, g, r, i, z, and y passbands (after 10 years of the survey). The maximum value of the color scale in each panel is the requested number of visits for that filter.

2.4 The Photon Simulator

The final component in the LSST simulation framework is the image generator (Phosim)[17]. Images are generated by drawing photons from the spectral energy distribution of each source (scaled to the appropriate flux density based on the apparent magnitude of a source and accounting for the spatial distribution of light for extended sources). Each photon is ray-traced through the atmosphere, telescope and camera to generate a CCD image. The atmosphere is modeled using a Taylor frozen screen approximation. Density fluctuations within these screens are generated from a von Karman - Kolmogorov spectrum with an outer scale of ~10m. All screens move during an exposure with velocities derived from NOAA measurements of the wind velocities above the LSST site in Chile.

Once through the atmosphere, photons are reflected and refracted by the optical surfaces within the telescope and camera. The mirrors and lenses are simulated using geometric optics techniques in a fast ray-tracing algorithm and all optical surfaces include perturbations based on design tolerances. Fast techniques for finding intercepts on the aspheric surface and altering the trajectory of a photon by reflection or wavelength-dependent refraction have been implemented to optimize the efficiency of the simulated images.

Ray-tracing of the photons continues into the silicon of the detector. Conversion probability and refraction (a function of wavelength and temperature) and charge diffusion within the silicon are modeled for all photons. Photons are pixelated and readout, simulating the effects of blooming, charge saturation, charge transfer inefficiency, gain and offsets, hot pixels and columns, and QE variations. A cutout of the simulation of a single 15 exposure is shown in Figure 5.

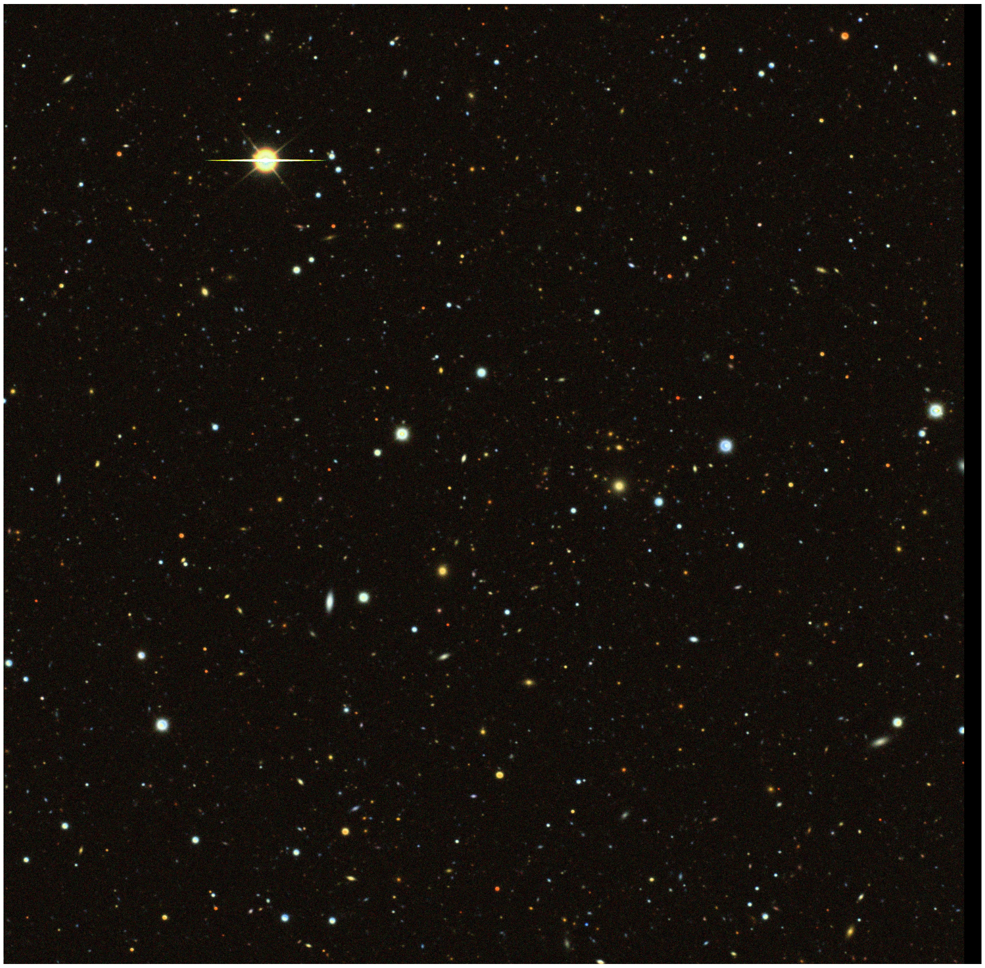

Figure 6. The photon simulator generates images that are representative of the LSST. This color image is generated for a single CCD from simulated 15-second exposures in the LSST g, r, and i passbands.

3. AN END-TO-END SIMULATION FRAMEWORK

As can be seen from the previous sections the design of a framework to simulate the data expected from the LSST requires flexibility and scalability. This is accomplished by dividing the simulation workload into three separate components: a component that stores a model of the universe, a system for querying the underlying model of the universe using simulations of sequences of LSST observations, and a framework for the generation of images via the ray-tracing of photons from individual sources. The LSST universe catalog is stored as a SQL database (using Microsoft SQLServer). Data are accessed through a Python interface that uses SQLalchemy to provide a database agnostic view of the sources. For LSST pointing, sources can be queried as a function of position and time with the returned data accounting for any change in brightness due to variability. For large-scale runs, the catalog is queried using sequences of observations derived from OpSim. Magnitudes and source counts can be calculated using the atmospheric and filter response functions appropriate for the airmass of the observation and after applying corrections for source variability. The resulting catalogs (instance catalogs) can be formatted for use in a science application (e.g. measuring the proper motions of high velocity stars) or fed to the image simulator.

The query framework is written in Python and takes an object-centric view of the data. For each object type (galaxy, main sequence star, strong lens, etc.), a class is defined that knows how to query for, format, and transform objects of that type. Overall this approach provides a simple and extensible framework for expressing different astrophysical sources and incorporating new catalogs within an existing framework (i.e. different components of the universe model are logically distinct and so can be separated into independent tables in a database or in different databases all together).

4. CONCLUSIONS

The volume of data that will be collected by surveys such as the LSST will result in an era where much of our science will be limited not by the shot noise in the data but by how well we can characterize and correct for systematics. To address this, the LSST has undertaken a program to develop an extensive suite of simulation tools that can model not just the engineering properties of the LSST design but also their impact on the science. This system provides a range of modeling capabilities that trade fidelity for computational cost (from simple parametric models to full ray tracing of images). We have shown here how these tools are deployed and used within the LSST project. All components of the LSST software tools are open source and publically accessible through http://stash.lsstcorp.org.

REFERENCES

[1] Z. Ivezic., et al., "Large Synoptic Survey Telescope (LSST) Science Requirements Document", http://www.lsst.org/files/docs/SRD.pdf (2011)
[2] Radiant Zemax. 2013, "Getting Started Using ZEMAX", http://www.radiantzemax.com/downloads/Getting_Started_With_Zemax_version2.1.pdf.
[3] Photon Engineering. "FRED version 12.31 Technical Description", http://www.photonengr.com/downloads/FREDSpecs.pdf. (2013)
[4] De Lucia, G., Springel, V., White, S.D.M., Croton, D., and Kauffmann, G., "The Formation History of Elliptical Galaxies", Monthly Notices of the Royal Astronomical Society, 366, 499-509 (2006)
[5] Bruzual, G. and Charlot, S., "Stellar population synthesis at the resolution of 2003", Monthly Notices of the Royal Astronomical Society, 344, 1000-1028 (2003)
[6] González, J.E., Lacey, C. G., Baugh, C. M., Frenk, C. S., and Benson, A. J., "Testing model predictions of the cold dark matter cosmology for the sizes, colours, morphologies and luminosities of galaxies with the SDSS", Monthly Notices of the Royal Astronomical Society, 397, 1254-1274 (2009)
[7] Juric, M., et al., "The Milky Way Tomography with SDSS. I. Stellar Number Density Distribution", Astrophysical Journal, 673, 864-914 (2008)
[8] Bond et al, "The Milky Way Tomography with SDSS. III. Stellar Kinematics", ApJ, 716, 1, (2010)
[9] Kurucz, R.L., "CD-ROM No.13", Cambridge, Mass.,Smithsonian Astrophysical Observatory (1993)
[10] Bergeron, P., Wesemael, F., and Beauchamp, A., Wesemael, F., "Photometric Calibration of Hydrogen- and Helium-Rich White Dwarf Models", Publications of the Astronomical Society of the Pacific, 107, 1047-1054 (1995)
[11] Amôres, E.B., and Lépine, J.R.D., "Models for Interstellar Extinction in Galaxy", Astronomical Journal, 130, 659-673 (2005)
[12] Schlegel, D.J., Finkbeiner, D.P., and Davis, M., "Maps of Dust Infrared Emission for Use in Estimation of Reddening and Cosmic Microwave Background Radiation Foregrounds", Astrophysical Journal, 500, 525-553 (1998)
[13] Grav, T., Jedicke, R., Denneau, L., Holman, M. J., and Spahr, T., "The Pan-STARRS Synthetic Solar System Model and its Applications", BAAS, 211, 4721 (2007)
[14] Granvik, M., Virtanen, J., Oszkiewicz, D., and Muinonen, K., "OpenOrb: Open-Source Asteroid Orbit Computation Software Including Statistical Ranging", Meteoritics and Planetary Science, 44, 1853-1861 (2009)
[15] Ridgway, S., et al, "Simulation of autonomous observing with a ground-based telescope: the LSST experience", Proc. SPIE 7737, Observatory Operations: Strategies, Processes, and Systems III, 77370Z (2010)
[16] Krisciunas, K., & Schaefer, B. E., "A model of the brightness of moonlight", PASP, 103, 1033 (1991)
[17] Peterson, J.P., et al., "Simulation of Astronomical Images from Optical Survey Telescopes using a Comprehensive Photon Monte Carlo Approach", in preparation (2014)