

Database Contents -- Catalog Simulations

Below, we discuss the characteristics and origin of the tables in the LSST database. For a detailed list of the available tables and columns, see [this page](#).

Star Catalogs

Several database tables exist containing different types of stellar objects (Main Sequence stars, RR Lyrae, etc.). These tables can be accessed through the classes provided in `sims_catUtils/python/lss/sims/catUtils/baseCatalogModels/StarModels.py`. Not all of these catalogs are accurately distributed across the sky. Some of them are just 'proof of concept' tables. Below, we list the available tables of stars, along with the python class which accesses them, and their extent across the sky. Those which cover the 'whole' sky (0-360 degrees in RA; -90 to 35 degrees in Dec) may be taken to be distributed according to some physically accurate model. Those which are not distributed over the whole sky should not be treated as physically accurate.

Main Sequence stars

- database table = `StarMSRGB`
- accessed using the class `MsStarObj` from `StarModels.py`
- distributed over the whole sky

White Dwarfs

- database table = `StarWD`
- accessed using the class `WdStarObj` from `StarModels.py`
- distributed over the whole sky

Eclipsing binaries

- database table = `StarEclipsingBinary`
- accessed using the class `EbStarObj` from `StarModels.py`
- RA ranges from 0.03 degrees to 87.5 degrees; Dec ranges from -17.5 degrees to -2.5 degrees

RR Lyrae

- database table = `StarRRLy`
- accessed using the class `RRLyStarObj` from `StarModels.py`
- distributed over the whole sky

dwarf galaxy stars

- database table = `StarDwarfGalaxy`
- accessed using the class `DwarfGalStarObj` from `StarModels.py`
- RA ranges from 0.64 degrees to 2.2 degrees; Dec ranges from -0.75 degrees to 0.77 degrees

Cepheid variables

- database table = `StarCepheid`
- accessed using the class `CepheidStarObj` from `StarModels.py`
- RA ranges from 72.7 degrees to 90 degrees; Dec ranges from -20 degrees to -0.2 degrees

'easter eggs'

- database table = `AstromEasterEggs`
- accessed using the class `EasterEggStarObj` from `StarModels.py`
- RA ranges from 78 degrees to 321 degrees; Dec ranges from -12.5 degrees to 5.5 degrees

Cosmological and Galaxy Catalogs

The galaxy simulation is based on dark matter haloes from the Millennium simulation [\citep{springel05}](#) (with an assumed standard Λ -CDM cosmology) and a semi-analytic baryon model grafted upon the Millennium results as described in [\citep{springel05}](#) and [\citep{delucia}](#). This semi-analytic model features radiative cooling, star formation, and the dynamics of black holes, supernovae, and AGNs. It includes explicitly following dark matter haloes, even after accretion onto larger systems, in order to follow the dynamics of satellite galaxies for an extended period of time as well as 'radio mode' feedback of AGNs. The model was adjusted to mimic the luminosity, color, and morphology distributions of low redshift galaxies [\citep{delucia}](#). LSST cosmological catalogs were generated from the [\citep{delucia}](#) data by constructing a lightcone, covering redshifts $0 < z < 6$, from 58 500h⁻¹Mpc simulation snapshots. This lightcone covers a 4.5x4.5 degree footprint on the sky and samples halo masses over the range 2.5×10^9 to $10^{12} h^{-1} M_\odot$.

Dynamically tiling this footprint across the sky enables the simulation of the full LSST footprint 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 \citep{bruzual}. The \cite{delucia} 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. These parameters are used in constraining the assignment of SEDs to each disk and bulge component. Fits are undertaken independently for the bulge and disk and include inclination dependent reddening. Morphologies are modeled using two Sersic \cite{Sersic} profiles and a single point source (for the AGN) with bulge-to-disk ratios and disk scale lengths from \cite{delucia}. Half-light radii for bulges are estimated using the empirical absolute-magnitude vs half-light radius relation given by \cite{gonzalez09}. AGNs are derived using the \cite{bongiorno12} luminosity function. The B-band absolute magnitudes are converted to bolometric luminosities using Eqn. 2 in \cite{hopkins07}. Empirical relations derived from the SDSS enable computation of the colors and stellar mass of the AGNs host galaxy from its luminosity. These parameters are used, together with the redshift values from the AGN catalog, to match each AGN to a galaxy in the galaxy catalog. In general, the AGNs match to galaxies having higher stellar masses, approximately $10^9 M_\odot$ to $10^{11} M_\odot$ which is comparable to recent analysis of host galaxies done by \cite{xue11}. The AGN SED is taken from the mean AGN spectrum of \cite{vandenberk}. 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 are "cloned" in magnitude and redshift space until their densities reflect the average observed properties (see \S \ref{sec:galaxycounts}).

Galaxies are stored in a single table that is 4.5° on a side and comprises 17,428,284 galaxies. The schema for these galaxies is described in [the Database Schema page](#). Through a database stored procedure we provide a virtual replication of this table; tiling it across the full sky (see Figure~\ref{fig:galcoverage}). All queries outside of the footprint of the primary table are transformed, based on the bounding box of the tiles that they intersect, to lie within the primary table. Positions for sources that are returned from the database query are then transformed to their appropriate positions on the sky using the bounding boxes of the input tiles.

\begin{figure}[h]

\centering \includegraphics[width=0.5\textwidth]{validation_figures/basicDemo.png} \caption{The galaxy catalog is replicated (virtually) across the LSST footprint using a series of tiles. These tiles correspond to the footprint of the galaxy table in the database. Queries are transformed using the tile bounding box such that they map to the galaxy table. When positions are returned through this query they are mapped back to the appropriate sky coordinates using the tile bounding boxes.} \label{fig:galcoverage}

\end{figure}

The left panel of Figure \ref{fig:gcounts} shows a comparison of the cumulative galaxy counts in CatSim to a compilation of observations provided by Metcalfe et al. (see \cite{http://ls.st/fus}). This comparison is undertaken in the i -band to minimize the effects of dust extinction which are somewhat uncertain in the Metcalfe compilations. A single transform of $I_{AB} = i_{AB} - 0.6$ has been applied to the Metcalfe magnitudes to take them from the Kron-Cousins photometric system to the SDSS AB photometric system \citep{ellis07}.

The right panel of Figure \ref{fig:gratio} shows the ratio of the cumulative counts taken from the simulations to a polynomial fit to the cumulative counts derived from the Metcalfe data. The error bars are estimated from the published uncertainties on the Metcalfe galaxy number counts. The requirement on galaxy number densities is that they agree within $\pm 10\%$ of the observed counts (to a coadded i -band depth of 26.8). This requirement is set due to the variance in the counts of galaxies at faint magnitudes (due to the small areal coverage of galaxy surveys at these depths). For magnitudes < 25.75 the simulated galaxy catalog meets the LSST requirements. For brighter magnitudes the simulated catalogs over-predict the galaxy counts by up to 25%. This discrepancy is to be expected as the volume sampled by the simulated galaxies (covering $4.5^\circ \times 4.5^\circ$) is small relative to the observations and the cosmic variance in the simulated data will be large. For magnitudes fainter than > 25.75 the galaxy counts fail to meet the number density requirement; deviating by up to 13% from the observed counts. % We have %taken their compilations from: \cite{http://star-www.dur.ac.uk/~nm/pubhtml/counts/ldata.txt} accessed on %06/01/2013. \begin{figure}[h] \centering \includegraphics[width=0.45\textwidth]{validation_figures/Ngals-i.png} \hfil \includegraphics[width=0.45\textwidth]{validation_figures/CumulativeFraction_i.png} \caption{A comparison of the Metcalfe galaxy number counts (symbols) to those derived from the simulated catalog \label{fig:gcounts}. The left panel shows the differential counts and the right panel the ratio of the cumulative counts. Error bars are from derived from those published by the individual surveys. \label{fig:gratio} The vertical dashed line represents the 5σ magnitude limit for galaxies.} \end{figure}

Galactic Structure Catalogs

Stars are represented as point sources and are drawn from the Galfast model of \cite{galfast}. Galfast generates stars according to density laws derived from fitting SDSS data to a model of a thick and thin disk, and a halo \citep{juric}. 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 $\rho(x,y,z,M)$. After this stage, using Fe/H and kinematics models from \cite{ivezic08} and \cite{bond09} (also derived from SDSS data), each star is assigned a metallicity, proper motion, and parallax. Spectral energy distributions are fit to the predicted colors using the models of \cite{kuruczCD} for main sequence stars and giants, \cite{bergeron95} for white dwarfs, and a combination of spectral models and SDSS spectra for M, L, and T dwarfs \citep[e.g.][cushing05,bochanski07,burrows06,pettersen89,kowalski10]. For Galactic reddening, a value of $E(B-V)$ is assigned to each star using the three-dimensional Galactic model of \cite{amores05}. For consistency with extragalactic observations the dust model in the Milky Way is re-normalized to match the \cite{schlegel98} dust maps at a fiducial distance of 100 kpc. Once the extinction and SED are assigned, observed magnitudes are calculated in the SDSS and LSST photometric systems using fiducial system throughput curves. Binary stars are included in the luminosity functions from which the stellar colors are sampled but are assumed to be unresolved and non-variable (except for a selection of eclipsing binaries described later).

Stellar populations included within the current implementation of the model are:

- Main Sequence: F, G, K, M, L, T
- White Dwarf: H and He
- Red Giant Branch
- Blue Horizontal Branch
- RR-Lyrae

- Cepheids

Approximately 10% of the stellar sources are variable at a level detectable by LSST. Variability is modeled for sources within the base catalogs 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). Variables modeled range from 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.

For all of these sources, the generation of magnitudes and colors, and the application of time dependent astrometric corrections (e.g. precession, parallax, proper motion) are calculated using Python subclasses of the InstanceCatalog object. \subsubsection{Variable Sources} The framework is able to support several types of variability: periodic, stochastic, and repeating. The variability models used in the database include: \pagebreak \begin{itemize} \item M-dwarf flares -- full sky \item AGN/QSOs -- full sky \item RRly -- full sky \item Cepheids -- exemplar individuals \item Eclipsing binaries -- exemplar individuals \item Am CVn -- exemplar individuals \item Micro lensing -- exemplar individuals \end{itemize} Each type of variability is described by either a parametric model or an interpolated lookup table (see \S\ref{sec:determine} for a description of these models). To date only mono-chromatic variability has been implemented. % (see Figure \ref{fig:lcs} for example lightcurves). %Variable sources are implemented through the InstanceCatalog API %\citep{XXX}. This API takes the name of the variability model and the %parameters associated with that model (both of which are stored in the %database) and modifies the brightness of a source based on the time of %observation.

We consider five representative fields at varying Galactic latitudes and at a Galactic longitude of $l=90^\circ$. We compare the number counts of main sequence stars as a function of i -band limiting magnitude for the Galfast model \citep{juric} (using the composite dust model of \cite{amores05} normalized to \cite{schlegel98}) to the \cite{besancon} model (using their standard dust model). Figure \ref{fig:scounts_90} shows the cumulative number counts as a function of magnitude for the Besancon (dashed) and Galfast (solid) models for five values of Galactic latitude. We also show SDSS counts for two latitudes ($b=30^\circ$, squares, and $b=70^\circ$, circles). For high galactic latitudes ($|b| > 30^\circ$), the agreement with the SDSS counts is best with the Galfast models (with a 13% difference between Galfast and SDSS at the limit of the SDSS data). For low Galactic latitudes there is insufficient observational data to constrain the Galfast of Besancon models. In Figure \ref{fig:sratio_90} we show the ratio of Besancon counts to Galfast counts for the five test fields. The dashed lines are the $\pm 30\%$ limits for the low latitude sizing model constraints and the dash-dot lines are the $\pm 20\%$ limits for the high latitude limits requirements. For all Galactic latitudes $b < -30^\circ$ the Besancon and Galfast models disagree at $>20\%$ requirement for stellar densities. For low Galactic latitudes ($b = -10^\circ$) the Besancon and Galfast models are in good agreement (i.e. within the 20% requirement on stellar number densities).

Solar System Catalogs

The Solar System model is a realization of the \cite{grav11} 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 $V=24.5$. Each object is assigned a carbonaceous or stony composition spectrum derived from extending the reflectance spectra from \cite{demeo} by linear extrapolation from 4500 Å to 3000 Å and then multiplying by a Kurucz solar spectrum. The choice of a C or S type spectra for an object is assigned based upon a simple relation to the size of its orbit that approximately matches SDSS asteroid observations. Each object's brightness during a specific observation is calculated from its location, phase, H_V and g values. H_V is the object's absolute magnitude and corresponds to the brightness if it were observed at 1 AU from the Sun and at zero phase angle. The H_V distribution is modeled independently for each source population (NEO, TNO, main belt, etc.) as described in \S 3 of \cite{grav11}. The g value relates the change in brightness of an object with the change in phase and is set at 0.15 for all objects across all bands, which is a typical value for asteroid phase curves. A more accurate modeling of the asteroid phase curves would require more realistic rotation and composition models which may be included in future work. The location of the Earth at the time of a particular observation is incorporated through the orbital ephemeris software oorb (\cite{granvik}; \tt http://ls.st/rd4) that calculates a V band apparent magnitude which is then used with the object's assigned C or S type SED to derive the corresponding LSST band observations.

Solar System sources are the most complicated table in the database. The typical way to characterize a Solar System object is to store its 6 orbital elements and propagate the orbit of the source to the time of the observation. Propagation of all orbits would require a numerical integration over 11 million sources (for each query). To accomplish for all sources within the Solar System table for each query would be computationally prohibitive (requiring 222,000 s to propagate one year into the future). We, therefore, pre-cache the positions of asteroids within the database and interpolate their positions based on the time of the observation. Ephemerides are calculated for all Solar System sources within the database for a ten year period. The time between ephemerides is variable and depends on the asteroid population (i.e. it is set by the velocity of the asteroid and the complexity of its orbital track). For main belt asteroids the positions are stored every two days together with the Chebyshev polynomial coefficients required to interpolate between these positions. Figure-\ref{fig:asteroid} shows that, using a cubic interpolation, asteroid positions are returned with an rms accuracy of <1 mas (sufficient to meet requirement {\it Catalogs: Requirements 5}). These cached positions are indexed using a HTM to speed the spatial lookup. This results in a query for 20,000 asteroids (i.e. larger than a full focal plane) requiring 30 ms to complete. \begin{figure}[h] \centering \includegraphics[width=0.65\textwidth]{validation_figures/ErrorHistogramsLinear.pdf} \caption{The distribution of maximum errors introduced to the orbital positions of asteroids due to the adopted interpolation scheme. These distributions are given as a function of asteroid populations. All populations meet the requirement that the interpolation be accurate to an rms of 1 mas.} \label{fig:asteroid} \end{figure}

[Return to the main catalog simulations documentation page](#)