Metrics Analysis Framework (MAF) Requirements

1. What is the Metrics Analysis Framework (MAF)?

The Metrics Analysis Framework is intended to provide a framework to evaluate simulated data, both in terms of an individual set of data (such as a single Operations Simulation run or a single Calibration Simulation run) and for comparison of different sets of similar data (such as multiple different Operations Simulations runs or Calibration Simulation runs).

A 'metric' here is simply an evaluation of that data. This evaluation could be performed at many points over the sky or some other division of the input data – this division will be called the metric grid. The evaluation is simply a measurement over that grid, which can then be summarized further and represented by simple numbers (an average over the grid, for example), plots incorporating the full grid, a histogram, or other derived values such as a power spectrum. Neither the summary values or the direct metric values will carry an evaluation of 'good' or 'bad' or 'pass' or 'fail' – these limits, if desired, will be applied as a separate stage of processing, as a user looks at the summary values.

The preliminary purpose of the MAF is to generate metrics for Operations Simulation runs, but the design should be general enough to allow for analysis of other simulation data such as Calibration Simulation runs, as well as real data fitting the same general profile. The major goal is to make it easy for users to write their own metrics, thus we must provide the tools to read data, generate and iterate over a variety of metric grids, calculate a metric at each point in the metric grid, save these metric values, and generate views of that metric data (such as sky maps, histograms, angular power spectra, or simple summary numbers of metrics from a single simulated data set as well as comparisons between simulated data sets), as well as run metric analysis.

2. Requirements for Metrics Analysis Framework

2.1. Standard requirements

We'll follow standard LSST software practices.

1. The code will follow LSST software standards (`https://dev.lsstcorp.org/trac/wiki/CodeStandards`).

2. The software package will contain unit tests (`https://dev.lsstcorp.org/trac/wiki/UnitTestingStd`).

3. Code and documentation will be kept in an LSST git repository.

4. Software will be easily distributable to an LSST software stack user.

2.2. MAF requirements

1. The MAF will calculate metrics on simulated data.

   - Examples of simulated data are Operations Simulation (opsim) runs, which contain a simulated pointing history of LSST over a period of time that can vary from a short period (such as one month, or even one night) to ten years (approximately 2.7M visits), or Calibration simulation (calsim) runs, which contain results of self-calibration simulations, including best-fit stellar magnitudes (up to 50M stars) and patch zero-points (up to 100M zero-points).
   - Examples of metrics are calculations of NEO completeness or coadded $m5$ depth in each filter for opsim data, or $\Delta(m)$ (best-fit vs. true stellar magnitude) for calsim data.

2. The metric set will be easily extensible (i.e. adding a new metric to a simulated data set's evaluation should be simple).

3. Calculation of metrics will be easily configurable (i.e. a single metric can be calculated or a whole set, depending on the user's preferences).

4. The MAF will allow reading simulated data from a flat file or from a database (any database with a python driver).

5. Metrics evaluated on a spatial grid (calculated at a distribution of points in the sky) will be calculable using a variable set of resolutions. This allows quick metric calculation (evaluation on a coarse grid across the whole sky) as well as more precise evaluations (to resolve field overlaps, for example).

6. Where appropriate, metrics evaluated on a spatial grid will be calculable over varying regions of the sky. This allows very high resolution over a small portion of the sky as well as simple parallelization by splitting the sky into regions.

7. Calculation of metrics should be separated from presentation of the results of these metrics to users. This allows calculation of complex metrics (such as NEO completeness), which may take significant time to run, and allows for the possibility of multiple methods to view the metric results.

8. Presentation of metric results to users must allow diving into all metrics from a particular data set. This includes sky plots, histograms, and simple summary statistics, from multiple metrics.

9. Presentation of metric results which have a time dependence (such as opsim coadded $m5$ depths at year 1, 2, 3, etc.) will include the ability to quickly visualize that time dependence (such as with an animated gif or movie).

10. Presentation of metric results to users must also allow comparison of metric results from user-selected multiple data sets (of the same type). This includes comparison of the simple summary statistics, generation of histograms with data from the same metric calculated using multiple simulated data sets, and generation of plots containing the differences in metric values across the sky.

11. When comparing different simulation runs, users will also be able to compare differences in simulations which generated those data; this includes version information of the simulation software and configuration files used to generate the specific simulation data.

12. The presentation of metric results should allow users to determine which simulation data sets meet their particular criteria and determine their preferred simulation data sets (from a group of simulated data sets).

13. The presentation layer of the MAF will allow generation of a static (i.e. PDF or similar) summary report, interactive web-based displays of summaries of the metric data (i.e. a web page with links to all metric summaries - plots and simple numbers - from a particular simulated data set and with the capability to compare user-specified simulated data sets and metrics), as well as providing methods to allow users to interact with metric data from within a python shell.

2.3. Operations Simulation analysis specific requirements

1. Presentation of metric results from a single database must allow comparison to a 'baseline' (typically a 'design' or 'stretch' list of benchmarks). The baseline will be user-definable, but the values of the baseline comparison must be presented with the metric summaries.

2. The framework must read and allow the user to evaluate the opsim input configuration files.

3. The presentation layer of the MAF must be able to create a table of all simple number summaries of metrics, for all simulations, which is possible to sort by a user-defined scheme.

4. The MAF must be able to read all opsim output database tables (including, but not limited to the current 'output_*' table) in order to access sequence and proposal information and other engineering records such as slew time.

5. The metrics calculated by the MAF for opsim runs will include (but are not limited to):

   - Number of visits at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits and for visits split by proposal ID. The

results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and +/- 3σ over the entire region visited.

- 'Joint completeness' - calculation of amount of area with more than Nfraction of Nvisits in each filter, where Nvisits is the number of visits set by the 'baseline' value for each filter, and Nfraction ranges from 0 to 100% in steps of 10%. The visits to be considered for this metric can be restricted to only visits tagged as part of the WFD proposal, or visits evaluated over a spatial grid matching the WFD footprint and meeting requirements in seeing and single visit $m5$ values.

- Median, minimum and maximum single visit $m5$ depth at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits and for visits split by proposal ID. The results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and $+/3\sigma$ over the entire region visited.

- Single visit $m5$ depth per filter, evaluated for all visits (regardless of RA/Dec position). This will be calculated for all visits, as well as only visits tagged for the WFD proposal or visits within the WFD footprint and meeting requirements in seeing and single visit $m5$ depth. The results will be presented as a histogram as a function of number of visits and simple summary numbers of the median, mean + rms over all visits.

- Coadded $m5$ depth at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits and for visits restricted to those meeting requirements in seeing. The results will be presented as a sky map, a histogram as function of area, and simple summary numbers of median, mean + rms, and +/- 3σ over the sky.

- 'Filter map' - a plot of visit day vs visit hour (relative to local midnight), color-coded by visit filter. This plot will be produced for each year of the survey (separately) as well as for the entire survey.

- Median, minimum and maximum sky brightness at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits. The results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and $+/3\sigma$ over the entire region visited.

- Sky brightness per filter, evaluated for all visits (regardless of RA/Dec position). This will be calculated for all visits and as well as only for visits tagged with the WFD proposal ID. The results will be presented as a histogram as a function of number of visits and simple summary numbers of the median, mean + rms over all visits.

- Median, minimum and maximum seeing at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits. The results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and $+/3\sigma$ over the entire region visited.

- Seeing per filter, evaluated for all visits (regardless of RA/Dec position). This will be calculated for all visits and as well as only for visits tagged with the WFD proposal ID. The results will be presented as a histogram as a function of number of visits and simple summary numbers of the median, mean + rms over all visits.

- Median, minimum and maximum airmass at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits. The results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and +/3σ over the entire region visited.

- Airmass per filter, evaluated for all visits (regardless of RA/Dec position). This will be calculated for all visits and as well as only for visits tagged with the WFD proposal ID. The results will be presented as a histogram as a function of number of visits and simple summary numbers of the median, mean + rms over all visits.

- Median, minimum and maximum normalized airmass (airmass divided by the minimum possible airmass for that RA/Dec) at a particular RA/Dec, per filter, evaluated over a spatial grid across the sky. This will be calculated for all visits. The results will be presented as a sky map, a histogram as a function of area, and simple summary numbers of the median, mean + rms, and +/3σ over the entire region visited.

- Normalized airmass per filter, evaluated for all visits (regardless of RA/Dec position). This will be calculated for all visits and as well as only for visits tagged with the WFD proposal ID. The results will be presented as a histogram as a function of number of visits and simple summary numbers of the median, mean + rms over all visits.

- 'Visit pairs' - the maximum number of nights (over the whole survey lifetime) within 30 nights that have pairs of visits (in the same filter and in any of g, r, i, z filters) with inter-visit time between 15 and 90 minutes at a particular RA/Dec, evaluated over a spatial grid across the sky. The results will be presented as a sky map, a histogram as a function of area and simple number summaries of the median, mean + rms and +/- 3σ values across the sky.

- Time between visits, evaluated for all visits. The results will be presented as a histogram and simple number summary of median, mean + rms over all visits.

- Slew distance between visits, evaluated for all visits. The results will be presented as a histogram and simple number summary of median, mean + rms over all visits.

The list here is intended to capture SSTAR functionality. More metrics are listed in the design document, and the framework shall be user-extensible.

3. Discussion

Generally speaking, all use cases for the MAF involve reading simulated data or ingesting it from a database, calculating some value or values using that data (calculating a 'metric'), saving

the resulting values to a file or database, and then presenting these results and/or comparisons of these results against the same metric calculated on a similar data set to the user. Almost always, a particular simulated data set will be used to generate many metrics, although some of these metrics may use different subsets of the simulation data.

When presenting the metric results to the user, we will generally need plots showing the entire metric results (e.g. a map of the coadded $m5$ depths in a particular filter across the sky), histograms to simplify the metric results slightly (e.g. a histogram of the coadded $m5$ depth as a function of area), and simple summary numbers (e.g. the median value of the coadded $m5$ depth across the sky). We may also additional evaluations of the metric values, such as calculating a power spectrum of the metric value (e.g. the power spectrum of the coadded $m5$ values to show spikes which could affect weak lensing).

When presenting metric results from multiple simulations, we will generally need all of the above representations of the metric, but highlighting the differences between multiple simulations. The examples above can be extended in illustration: median coadded $m5$ values for a series of OpSim runs could be presented in sorted order, a single plot showing histograms of the coadded $m5$ values as a function of area for the series of OpSim runs could be generated, and then as the user narrows down to comparison of two different OpSim runs, a sky plot of the difference in coadded $m5$ values between those two runs could be generated.

The need for plots, such as sky maps, combined with the requirement to compare multiple simulation data sets, means the full metric result (which could be thought of as intermediate data, before the graphical and summary representations are generated) from each simulation must be stored. Thus, each metric use case will result in *(a)* a file containing all of the metric values at each point of evaluation, and *(b)* a set of summary outputs (summary number(s), plots).