Evaluation of v2.0 and v2.1 OpSims for Galactic Science

The purpose of this report is to describe the evaluation of the v2.0 and v2.1 families of Rubin Operational Simulations (opsims) for the purposes of galactic science, particularly in the time-domain.

This topic encompases a range of science and categories of variables in the Milky Way and Local Volume, particularly the Galactic Bulge, Plane, Magellanic Clouds, Star Forming Regions (SFRs) and Open and Globular Clusters.

The aim of this document is to provide high-level feedback to the SCOC on the pros and cons of different opsim experiments for Galactic Science as a whole to inform their discussions. It is intended to complement other reports and publications which explore the impacts of survey strategy on specific categories of variability in greater depth.

Contributors:

R.A. Street, Las Cumbres Observatory, rstreet@lco.global

X. Li
M. Di Criscienzo
R. Bonito
L. Venuti
S. Leccia
I., Musella
W. Clarkson
N. S. Abrams
M. Hundertmark
S. Khakpash
S. Ustamujic
V. Braga
V. Ripepi
L. Girardi

Metrics Relevant to Galactic Science

A number of metrics are highly relevant to this science topic, but in the course of evaluating the opsim families, we recommend focusing on the following set of metrics as particularly representative of this science topic.

- maf.metrics.CountMetric(Nvisits)
- galacticPlaneMetrics.GalPlaneFootprintMetric
- galplaneTimeSamplingMetrics.GalPlaneVisitIntervalsTimescaleMetric
- galplaneTimeSamplingMetrics.GalPlaneSeasonGapsTimescaleMetric
- galacticPlaneMetrics.GalPlaneTimePerFilterMetric
- cadenceMetrics.UniformityMetric
- YoungStellarObjectsMetric.NYoungStarsMetric
- filterPairTGapsMetric.filterPairTGapsMetric
- PulsatingStarRecovery
- PeriodicDetectMetric
- MicrolensingMetric [Ndetect, Npts]

Revised Baseline_v2.0 Nvisits Compared with Earlier Baselines

The inclusion of "the Diamond" (central Galactic Bulge and Plane) and the Magellanic Clouds in the WFD as well as including regions of lower extinction at lower galactic latitude are strongly beneficial for a range of Galactic Science, especially microlensing, and we endorse this change.
The only reservation we have is that this excludes regions in the "wings" of the Plane at higher galactic longitude, reducing the range of different stellar environments surveyed at moderate cadence, and (for microlensing) the lines of sight through different galactic populations. In place of the centrally-concentrated Diamond, **we recommend distributing the same time across a set of pencilbeam fields and the Magellanic Clouds**, which offers a way to address these issues without taking additional time.

Recovery of Periodic Variables

Simulations were conducted to evaluate how well periodic objects (e.g. RR Lyrae) can be recovered from baseline_v2.0 lightcurves. This analysis has shown that after 10 years it is possible to obtain average magnitudes and amplitudes with a precision comparable to the standard of theoretical relations (thousandths of magnitudes on mean magnitudes and hundredths on amplitudes), from lightcurves in at least three of the 6 bandpasses. We note three is the minimum number of bands to build color-color diagrams to discover trends with metallicity.

We were also able to show that it is not necessary to wait 10 years for this type of short period variables to obtain a good recovery of periodicities from light curve: it is possible to recover the periods and amplitudes from the first two years of data, at least for the most sampled filters (rizy). **We emphasize that the early identification of periodic variables is essential** not only for detailed study of the variables themselves, but also in order to distinguish them from transients throughout the rest of the survey.

However, measurement of the average magnitudes and amplitudes in u and g bands would require a different strategy than baseline_v2.0. Simulations from the v2.0 filter-dist family do not cover this gap (for example **bluer_indx0_v2.0_10yrs**). One possible solution could be an early rolling strategy in u and g band to have at the end of the second year of observation accurate mean magnitudes and amplitudes or a dedicated minisurvey but that foresees many more visits than those foreseen by **local_gal_bindx2_v2.0_10yrs.db** of v2.0 in IC1613, the only dwarf in common with our sample.

We have demonstrated that long period variables, treated as static stars, can be observed up to the Local Group outskirts, but only Cepheids are recovered satisfactorily. For Long Period Variables with large amplitudes, the situation is more complicated. For the most distant galaxies, the important part but less luminous part of the light curve falls completely below the detection limits, making it very difficult to recover the lightcurve morphology or period.

Vary_gp Family of OpSims

The GalPlaneFootprintMetric %ofPriority and the VisitIntervalsMetric were examined for the different timescale categories of variable. The results clearly show the benefits of awarding more time to the high galactic longitude regions of the Galactic Plane, particularly for gpfrac weightings >0.5, with strong increases in the metric for the combined galactic science region and pencilbeams. Increasing the sampling of regions outside the central Plane increases the total population of variables of many types, including microlensing events, Young Stellar Objects and X-ray Binaries, which boosts the corresponding metric values when summed over the survey region for longer (τ_{var}>55days) timescale categories of variability.

However, the two shortest categories of variable timescale (t_{var}<25days) show that ***virtually no Galactic Plane region is adequately sampled to characterize transient targets on those timescales*. This significantly curtails the detected population of microlensing events, and other transient phenomena, including X-ray Binary/Cataclysmic Variables outbursts.**

There is a sweet-spot to be identified in terms of cadence and survey footprint. Observing a large area with too low a cadence will mean that transient events can't be properly identified in time for characterization or follow-up. But reducing the area surveyed in order to boost the cadence will eventually also reduce the total number of transient events discovered. We recommend evaluating the MicrolensingMetric with two timescales (30d and 200d) in order to identify the optimum balance for transient events [see separate report by Abrams et al].

However, we recommend caution in prioritizing simple numbers of microlensing events detected alone, as the distribution of microlensing events is heavily concentrated towards the Bulge. This neglects the high scientific importance of microlensing events towards the Magellanic Clouds and other populations in the Milky Way. It also neglects other time-variable populations with different spatial distributions, such as pulsating stars (including the RR Lyrae that are essential to map the Milky Way, one of LSST's key science objectives), Cataclysmic Variables and X-ray binaries.

We recommend that the GalPlaneFootprintMetric and GalPlaneVisitIntervalsTimescaleMetric values be explored for the pencilbeam fields, Bulge and Magellanic Clouds when reviewing the vary_gp family of opsims. These metrics show a relatively large range of values in the pencilbeam regions, reflecting increasingly good coverage of these fields with increasing gpfraction.

Plane Priority OpSims

Unsurprisingly, the footprint metrics show a marked improvement for the combined survey region when the threshold for HEALpixel selection is lowered and a larger region is included; this is also particularly important for coverage of SFRs and clusters.
There is a notable distinction at priority level=0.9 between the maps with and without the pencilbeams included separately. If a priority selection level of 0.6 or lower is used then the majority of the pencil beams regions are sampled at a cadence consistent with the rest of the survey region - ***regardless of whether they are distinctly included in the map or not.***

If time available for the Galactic Plane survey is limited, we recommend including the pencil beams in the priority map and setting the priority threshold to 0.9 or lower.

If time permits, then ***we recommend selecting a survey region using a priority threshold of at least 0.4***. The metric indicates that this is a sweet spot whereby the highest priority regions are included and the cadence is maximized. If a lower threshold is used, the metric values begin to decline again, indicating that a smaller region can be monitored at the desired cadence.

In terms of observing known MW star clusters to be used as calibrators for the LSST stellar population work, the changes between the different baselines are minimal: Baseline_v2.0 and baseline_v2.1 observe 70 less open clusters in the Galactic Plane, and essentially the same numbers of globular clusters. Nevertheless, there are plenty of open and globular clusters left in all baseline plans (~2000 and 145, respectively), with a good coverage of the age and metallicity space. There are small improvements in the maximum photometric depth reachable in the latest baselines (up to 0.3 mag in the r band), mostly for clusters in less crowded areas of the Galactic Plane.

Rolling Cadences

The survey footprint metrics show essentially no change across all of the rolling cadence realizations, consistent with a very similar footprint being used in all cases. The rolling cadence was applied to the WFD region in all cases, and included the Galactic Bulge in a few cases. One simulation, rolling_all_sky_ns2, performed the rolling cadence across the whole sky.

In reviewing the rolling cadence simulations, we found it most valuable to review the mean value of the VisitIntervalMetric (VIM) for each year of the survey (e.g. Figure 1), for different variability timescales and regions of interest.

As expected, the rolling_ns2 and roll_early family of simulations show no advantage over baseline_v2.0, since the rolling cadence is not applied to any region of interest to galactic science.

The simulations where rolling cadence was applied to the Galactic Bulge are clearly highlighted. OpSim rolling_bulge_6 indicates that this strategy can reach almost the ideal cadence for the scientifically valuable shortest timescale (2 day) variability category for 1-2yrs during the survey, at the expense of achieving ~40-60% of the desired cadence in the remaining years. **If the years of high cadence in the Galactic Bulge were to coincide with the Roman Mission's survey of that region, then the complementary datasets would be very valuable for a wide range of science, particularly microlensing planet characterization.** This would offset the otherwise negative impacts of lower cadence in the other years, provided Rubin continues to regularly observe the region. Regular observations in all years are necessary to ensure that transient events continue to be discovered, even if the cadence is so low that characterization depends on additional follow-up. However, this would depend on close coordination between Roman and Rubin to ensure that the high-cadence years coincided with the Roman Bulge survey.

The simulation of rolling cadence across the whole sky is also noteworthy, as it is shown to achieve ~60% of the desired sampling for the shortest timescale variable categories in the high-cadence years. In “off” years, this drops to ~50%. The effect of this is that variables with timescales ≳55days would be detectable across all regions of interest to galactic science, but shorter timescale transients would require substantial follow-up observations in order to characterize them. An alternative realisation of this strategy would be to explore rolling cadence across the high-priority regions of the combined galactic science region (priority threshold>0.4) instead of just the Galactic Bulge.

Reviewing the time spent in each filter for the galactic science regions of interest suggests that **further work to rebalance the time allocated in each filter is strongly recommended.** While some regions received at least the desired ratio of observations in one or two filters (e.g. the Galactic Bulge was well sampled in g and z in almost all opsims), almost all regions were below the desired ratio in several other filters. The Magellanic Clouds appear to receive a high number of observations in u and y, but extremely low (<10%) number of observations in g, r, i, z, which is detrimental for the detection of microlensing.

Intranight cadence (presto_ and long_gaps_ families)

In terms of survey footprint, there was some reduction in the area covered at adequate cadence for τ_{var}⩾55days in the Magellanic Clouds during if the gap between exposures is short, i.e. <3hrs. There was a similar reduction in cadence across Open and Globular Clusters, though the impact was less marked. Otherwise the results were very similar to the baseline_v2.0 strategy, and shorter timescales of variability are not well sampled.

We examined the cadence achieved in all years (including all filters) compared with variables of different characteristic timescales and found relatively little variation between the simulations in this category.

Evaluating the balance of time allocated to different filters showed a dramatic variation in the results for different regions of science interest in different filters (illustrated in Figure 2). For example, sufficient data was obtained in g-band for the Galactic Bulge in all simulations, but the results in this filter for the Magellanic Clouds, Galactic Plane, X-ray Binaries and resolved stellar populations were very poor. Conversely, the results in z-band were much better, with most regions receiving at least 60% of the desired data.

In r-band, there was marked preference for the “non-mixed” filter pairings among the presto_gap strategies, since these focus on the {g,r,i,iz} filterset recommended for most galactic science. The achievable cadence in r for the “mixed” filterset {g,r,i,z,y} was considerably lower in all science regions (<20% of desired), and is therefore disfavored. However, this was reversed in the i-band, where improved cadence was found with the mixed filterset in most presto_gaps implementations.

We note that all of this family of opsims appeared to apply the same filter strategy to the whole sky. **We recommend that different filtersets be prioritized with independent cadences separately for the WFD and galactic science regions of interest respectively. This, combined with the selection of minisurvey sky area based on priority, can be**

used to optimize the time spent in the Galactic Plane, while maintaining support for the WFD and other science cases.

As a general comment, pairs of exposures in different filters are very useful for distinguishing variability classes in real time, and this is particularly important for brokers to be able to identify transients such as microlensing and Cataclysmic variable outbursts. We note that the characterization of microlensing will depend on our ability to constrain the source magnification as a function of time in at least two colors during the event. Regular multi-color observations are also important for the characterization of periodic variables, but the metrics to evaluate this are computationally intensive and are still being calculated for the many different opsims.

Triple exposure sets are likely to be particularly valuable in the Galactic Bulge field if coordinated with the Roman survey of the same field, since this will focus on the shorter timescale microlensing than the wide-area Rubin survey of the whole Galactic Plane and Magellanic Clouds.

Bluer Balance

The simulation with an emphasis on g-exposures shows enhanced cadence for Star Forming Regions. The simulation with enhanced u and g coverage shows a detrimental decrease in cadence for almost all regions.

Long u

The simulation with the same number of u visits but a 50s exposure time showed enhanced results for X-ray binaries, the Galactic Bulge and pencilbeam fields in the u-band. **However, this came at the expense of observations in the other filters** which are more important for most galactic science.

Vary Exposure Time/Shave

This opsim family indicates that a slightly reduced area of the desired footprint covered at a cadence sufficient to detect variability timescales of 55d or longer, impacting primarily science from the Galactic Plane, open clusters and star forming regions.

Interestingly the VisitIntervalMetric plotted on an annual basis shows significant improvement (1.0 (max) vs. 0.8) over baseline_v2.0 for longer (τ_{var}>55d) variability timescales. This may be due to a higher number of observations reaching the required signal-to-noise thresholds, thereby improving the cadence achieved.

The distribution of visits in different filters show some interesting variations in this family.
In the g-band, the time spent per filter generally improves with longer exposure times for all regions of interest. The same is true in i and r, though for star forming regions there seems to be a sweet spot peaking with exposures around 28s. The trends are reversed for the z- and u-filters, with shorter exposures favored.

Vary North Ecliptic Spur

By including more visits to fields at high galactic longitude, more Star Forming Regions are covered at higher cadence. The nesfrac⩾0.75 show a significant improvement for this science case. However, it appears to come at the expense of coverage in the Galactic Plane, reducing the cadence in the combined region of interest by ~10%. The high nesfrac simulations also offer a more ideal balance between time spent in different filters, but only for the Star Forming Regions, and here they are below ideal in the u, y bands.

No Repeat

The imposition of an additional basis function to avoid repeated visits in a single night seems to slightly reduce the number of visits to the Galactic Plane relative to baseline_v2.1, evidenced by a drop in the area of receiving the cadence required in each time category. This is detrimental to many of our science goals.

However, higher rpw weights do show a strong improvement in the SeasonGaps metric specifically for the Galactic Bulge region, which is beneficial for long-term variable classes in that region. More information on what the rpw value represents is required to interpret this result further. Similar, though much weaker, trends are seen in other regions, such as the Galactic Plane.

Good Seeing

Our footprint metrics produce a similarly small reduction in the region of interest receiving adequate cadence as seen for the no-repeat family of opsims, relative to baseline_v2.1.
Overall, we emphasize the importance of obtaining good seeing images in all bandpasses for the particularly crowded regions in the Galactic Plane and Magellanic Clouds and note that this is is best represented in our metrics by the PulsatingStarMetric, which has not yet been integrated into the MAF, so our analysis to date is not sensitive to this criterion.

Microsurveys

- **Virgo Cluster:** This region lies outside of the desired regions for galactic science, so our metrics are insensitive to it. It does not seem to have a significant impact on the cadence achieved within our desired regions.
- **Carina:** This simulation significantly enhances the coverage of desired Star Forming Regions.
- **SMC movie:** Our metrics are designed to examine median cadence over extended periods of time and are insensitive to the two nights of intensive observations of this relatively small region. However, we note the scientific benefit of these observations for periodic phenomena.
- **Roman:** We intent to apply specific metrics to this simulation which are in development.
- **Local_gal:** These simulations produce a slight increase in coverage for Globular Clusters.
- **Too_rate:** No significant impact is observed

- **North Stripe:** Our metrics show an appreciable improvement over baseline for this opsim, most likely due to a larger number of visits to fields in the Plane at high galactic longitude.
- **Short exposure:** Our metrics show little change from baseline, suggesting that allocation of time to short exposures does not sacrifice science that requires exposures at the program exposure time. An allocation of short exposures extends the science that can be done with Rubin, and would greatly aid calibration (e.g. Gizis et al. 2018 LSST Cadence White Paper, Clarkson et al. 2021 Cadence Note).
- **Multi_short:** We see reductions in the cadence in almost all of our desired survey regions,except the Galactic Bulge.
- **Twilight NEO:** These simulations achieve a slightly lower cadence (relative to baseline_v2.0) in all years for all regions of interest, such as Star Forming Regions and the Galactic Plane, which dominates the corresponding cadence drop in the combined region. This was somewhat surprising as it was understood that twilight observations were not routinely included in the baseline.

Carina microsurvey:

The ideal case among the OpSims is the carina OpSim v1.7, as discussed in Bonito & Venuti et al. 2021 Cadence Note. In fact, this OpSim has been developed to meet the requirements of collecting one visit every 30 minutes in each of the filters g, r, and i (and possibly also u-band) in one night of 10 hours and for 7 consecutive days. DDF and baseline OpSim (all versions available from 1.6 to 2.1) do not allow us to collect the proper number of points at the proper cadence to retrieve the shape of the light curves and discriminate the physical process at work in young stellar objects showing short-term variability.
It is worth noting that the v2.0 of the carina OpSim appears to be less populated with respect to the v1.7 discussed in Bonito & Venuti net al. 2021. This can be seen both across the projected coverage along the entire survey duration, and during a specific observing week and night, as illustrated below in Fig. 3, where it emerges that the simulated number of visits has approximately been halved compared to the earlier OpSim v1.7.
The ideal number of visits for our science case, which had informed the development of the Carina OpSim, has been established by comparing different patterns of short-term YSO variability (as reconstructed in detail from space-based observatories) after applying a filter with varying simulated number of visits and cadence, to assess whether the defining variability features would be retained in each case for each simulated filter. This work is described in the Bonito & Venuti et al. Cadence Note. A first version of the YSO variability metric developed for this science case was developed and distributed as a notebook before the deadline for the Cadence Note submission in April 2021. We are currently working on fine-tuning this YSO variability metric further, in collaboration with the MAF team.
It is worth noting that as this project has been defined as a micro-survey, a final decision regarding its implementation is expected only after the observing strategy for the main survey is finalized (MAF team private communication).

Deep Drilling Fields

Reducing the number of visits to the DDFs to 3% causes a small but unsurprising increase in the cadence in galactic science regions of interest, with a corresponding decrease with higher DDF fraction.

Since none of the DDFs lie within the galactic science regions of interest, none of the DDF families of opsims have any impact on our footprint or cadence metrics.
Curiously, some of the DDF “accordion” opsims produce strong variations in the frequency of r-band monitoring in the Galactic Bulge specifically though it was unclear why this should be the case. This is undesirable, and underscores the importance of implementing different cadences per filter for different survey regions.

Pencilbeam Fields

Two simulations were compared that implement alternative sets of pencilbeam fields to cover regions of high interest in the Galactic Plane. The pencil_fs1 simulation added 20 single-pointing pencilbeams, distributed across the Galactic Plane, to the baseline_v2.0 footprint, whereas pencil_fs2 added an alternative set of 4 larger pencilbeams. Both sets of pencilbeams include the same number of square degrees of sky in total. The goal of the comparison was to evaluate whether there was any impact on survey efficiency if these pencilbeams were in contiguous regions or separate pointings.

Two metrics were used to evaluate survey efficiency: the number of visits realised to the science regions of interest, and the OpenShutterFractionMetric.

We found that larger, consolidated pencilbeams result in a 7% increase to these regions overall, with little change to the open shutter fraction in pencilbeam fields. Most of the other regions of interest show small (~2%) increases or decreases in their number of visits, depending on whether they overlapped the pencilbeams. However, coverage of the Bonito Star Forming Regions dropped considerably - receiving 39% fewer visits - when larger pencilbeams were used.

In general, the changes in the open shutter fraction were small for all regions, but we note that the standard deviation of the shutter fraction within each science region of interest changed by an order of magnitude for some regions. Within the pencilbeam regions, the stddev dropped from 0.16 to 0.02 between the smaller and larger sets respectively.

Overall, we conclude that survey efficiency can be maximized by selecting larger contiguous regions for galactic science, thereby improving survey cadence and coverage.
The community will revise the galactic science priority maps accordingly.

Outstanding Questions

- Evaluation of Detection of Periodic Phenomena
  In general, Rubin’s ability to accurately measure period from lightcurve data should be evaluated.

- **The PulsatingStarMetric remains to be fully integrated with the MAF.** We recommend focusing on the Delta_Period_abs metric value as a measure of how well periods can be recovered from the lightcurve sampling, together with the deltamag and deltaamp values in each filter. However, we recommend that it be restricted to the first two years' worth of survey data, as it is widely beneficial to identify periodic variables sooner rather than later to the limiting magnitude of LSST. Not only does this enable science with the variables themselves, but it also enables these objects to be more accurately distinguished from transient objects throughout the survey.
- Results from the **PeriodicDetectMetric** appears to be available for a few opsims only, which makes comparison difficult. Attempts to run this metric independently for multiple opsims have crashed.

- Cadence of color measurements.
    - The FilterPairGaps metric calculates the time gaps between successive observations in different colors, which is valuable. To complement this, we are interested to measure the time gap between pairs of observations in different colors. The latter quantity is important to evaluate how well LSST will track color changes that are important for distinguishing different variable types.
- Roman/Rubin Metrics.
    - We are in the process of developing metrics to evaluate the complementarity of Roman and Rubin survey strategies.
- Deep Drilling Fields
    - Analysis is in progress for these opsim families

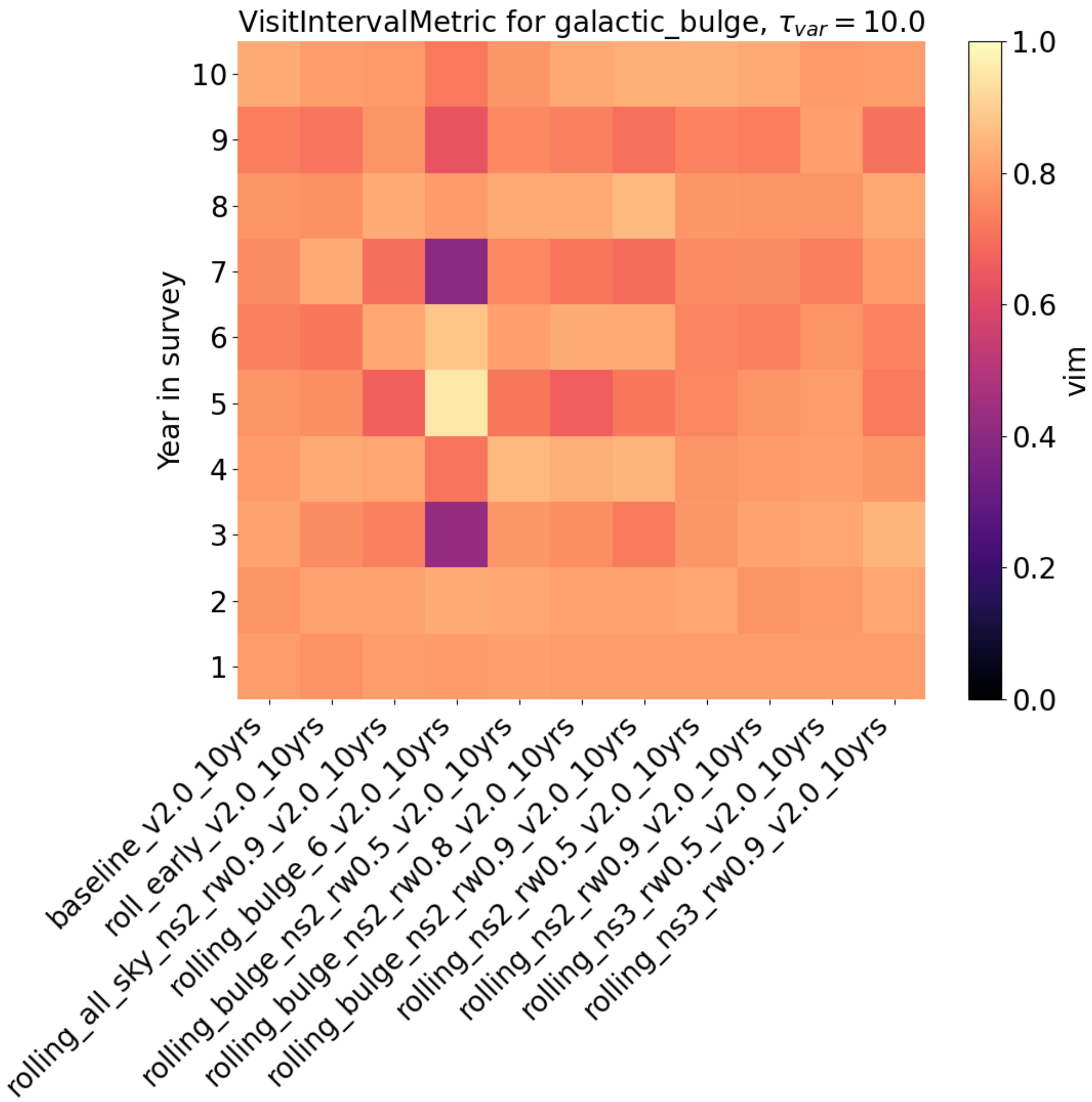

Figure 1: The VisitIntervalMetric evaluated for each year of LSST, presented as a fraction of the ideal cadence needed to characterize a variable with a timescale of <10 days. The median value of the metric is calculated across the region of interest, in this case the Galactic Bulge.

Figure 2: Heat maps of the percentage of different science regions of interest to receive 100% of their desired cadence in (top) r-band and (bottom) i-band.

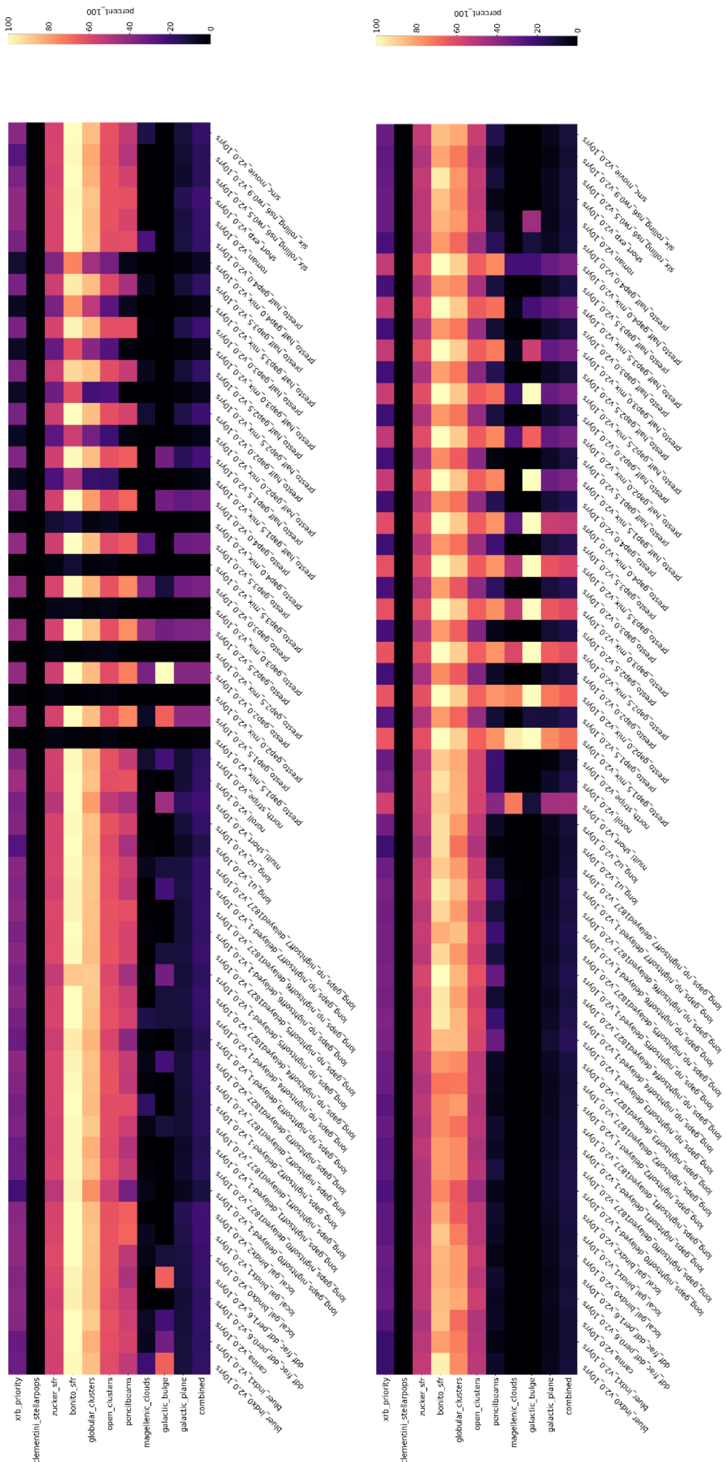

Figure 3: Simulated number of visits for the carina OpSim v1.7 (upper panels) and v2.0 (lower panels) across different coverage times: the entire survey duration (left column), during a specific observing week (middle column) and for a specific night (right column).