DESC input to SCOC: June 10, 2022

The DESC needs more time to conclusively evaluate the v2.0/v2.1 strategies. We are extremely grateful for the metric bundles provided; this has significantly accelerated our analysis process. In validating some of the metrics, however, we have found that some of them in MAF are not behaving as expected. This requires a thorough investigation before we can fully comment and we still have a need for more time for experts to look at the metric bundle outputs.

At this stage, we can share these preliminary thoughts:

1. In general, the v2.0/v2.1 simulations have a number of benefits for DESC science, particularly in specifying an extragalactic footprint (defined by low Milky-Way reddening) where observing strategy can be optimized independently of other regions, and the exploration of rolling cadences for DESC transient science.
2. DESC static science metrics in MAF appear to be consistent with previous versions. The increased extragalactic footprint going from v1 to v2 improves DESC static science constraining power by ~10%; this is excellent.
3. DESC transient metrics in MAF are behaving in ways that do not agree with our findings for v1.5, and we need more time to investigate this. More details are below.
4. The triplet strategies, which add a 3rd visit to some/all fields observed each night, are largely detrimental to DESC static science because they seem to come with a decrease in the usable extragalactic area (for reasons we do not yet understand). A small subset of those variants are tolerable. We will elaborate on this tolerance in the near future. The impact to DESC transient science is likely to also be negative (as we generally prefer more epochs to sample light curves over better intra-night data that can help early classification).
5. The strategies that give more priority to the Galactic plane are, not surprisingly, detrimental to DESC static science, as they reduce time spent in low extinction regions. Modest enhancements to the Galactic plane may be tolerable and we will elaborate on this tolerance in the near future.
6. We are not yet able to comment on rolling cadence because of the issue with DESC transient science metrics. DESC static science, measured via the Y10 3x2pt FOM metric, is largely insensitive to rolling, but there are some complications (e.g., uniformity of intermediate data releases) that are important to DESC static science.

We are grateful for the SCOC's efforts to push forward on the optimization of observing strategy. The interaction of DESC science with Rubin Observing Strategy is complex and multi-faceted, and as a large collaboration that plans to analyze multiple cosmological probes with different and at times competing observing strategy needs, we may not be as nimble in being able to deliver considered responses to SCOC questions on very short timescales. We thus appreciate the opportunity to engage with the SCOC via continuing, two-way communication. Having more advance notice of what questions the SCOC is considering (and on what dates), would help us plan ahead and provide better, more robust feedback.

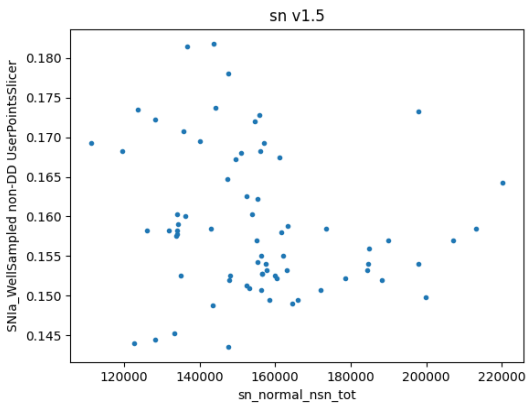
Validation of DESC MAF Metrics: June 10, 2022

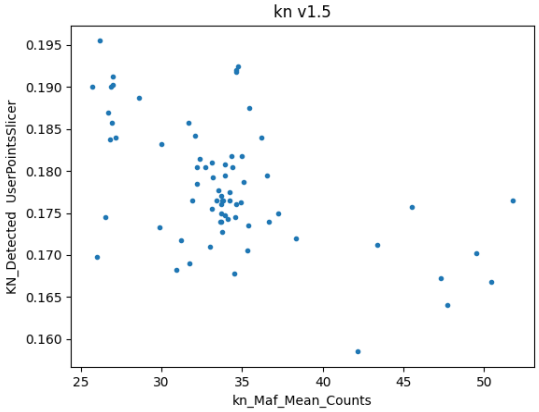
Executive Summary

We were not able to validate DESC transient-science (supernovae and kilonovae) MAF metrics in v2.0 simulations against previously published results from v1.5 simulations. We advise the SCOC not to rely on these transient science metrics yet to investigate implications for the observing strategy. DESC needs more time to diagnose and correct this issue.

Analysis of v1.5 metrics

The only set of simulations for which we have the original DESC metrics for are the v1.5 simulations (used in Lochner et al. 2022). The MAF metrics bundle is incredibly useful in allowing a far more rapid analysis of new simulations. Unfortunately, the MAF metrics bundle run for the v1.5 simulations did not yet include several key metrics contributed by the community. They did include the DESC static science metrics metrics, which match our original metrics extremely well, but the transient metrics were not included. Thus we could only compare against previously developed transient metrics.





In the figures above, we compared the original DESC metrics for SN and KN (x-axis) against their MAF counterparts for v1.5 (y-axis). These MAF metrics were *not* the same implementation as the original ones (e.g. the SNIa_WellSampled metric is based on the PLAsTiCC data). However they are supposed to be measuring similar things (well-sampled SNe ready for cosmology and detection of GW170817-like KNe). **We do not yet know whether this undesired behavior is due to a bug or unintended assumption.**

Analysis of the v2.0 Metrics

Without the original metrics run on the same set of simulations, we cannot validate the v2.0 metrics. Discussions with metric contributors has gone on behind the scenes, but our understanding from the OpSim team is that the SN MAF metric does NOT match the original DESC SN metric, for unknown reasons. The OpSim team does make changes to the code contributed by metrics-writers. While these changes are no doubt quite sensible (such as including dust extinction cuts), it means the metric needs to be fully validated before it can be assumed to accurately reproduce the trends found in the original metrics. Any significant deviations between MAF metrics and original need to be understood.

Conclusions

MAF metrics need to be validated against originally contributed metrics before drawing conclusions about observing strategy from them. The simplest way to do this is to create a unit test for key metrics where any changes to the MAF implementations are rerun on a subset of past simulations, for which the original metrics have been run. Any differences should be understood and either the MAF or original metrics updated until they match.