The Astrophysical Journal Supplement Series, 259:58 (35pp), 2022 April https://doi.org/10.3847/1538-4365/ac5033

The Impact of Observing Strategy on Cosmological Constraints with LSST

Michelle Lochner[1,2], Dan Scolnic[3], Husni Almoubayyed[4], Timo Anguita[5,6], Humna Awan[7,8], Eric Gawiser[8], Satya Gontcho A Gontcho[9], Melissa L. Graham[10], Philippe Gris[11], Simon Huber[12,13], Saurabh W. Jha[8], R. Lynne Jones[14,15], Alex G. Kim[16], Rachel Mandelbaum[4], Phil Marshall[17], Tanja Petrushevska[18], Nicolas Regnault[19], Christian N. Setzer[20], Sherry H. Suyu[12,13,21], Peter Yoachim[14], Rahul Biswas[20], Tristan Blaineau[22], Isobel Hook[23], Marc Moniez[22], Eric Neilsen[24], Hiranya Peiris[20,25], Daniel Rothchild[26], Christopher Stubbs[27,28]
and
The LSST Dark Energy Science Collaboration

[1] Department of Physics and Astronomy, University of the Western Cape, Bellville, Cape Town, 7535, South Africa
[2] South African Radio Astronomy Observatory (SARAO), The Park, Park Road, Pinelands, Cape Town, 7405, South Africa
[3] Department of Physics, Duke University, 120 Science Drive, Durham, NC 27708, USA
[4] McWilliams Center for Cosmology, Department of Physics, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[5] Departamento de Ciencias Fisicas, Universidad Andres Bello, Fernandez Concha 700, Las Condes, Santiago, Chile
[6] Millennium Institute of Astrophysics, Nuncio Monsenor Sotero Sanz 100, Of. 104, Providencia, Santiago, Chile
[7] Leinweber Center for Theoretical Physics, Department of Physics, University of Michigan, Ann Arbor, MI 48109, USA
[8] Department of Physics and Astronomy, Rutgers, the State University, Piscataway, NJ 08854, USA
[9] Department of Physics & Astronomy, University of Rochester, 500 Wilson Boulevard, Rochester, NY 14627, USA
[10] DiRAC Institute, Department of Astronomy, University of Washington, Box 351580, U.W., Seattle, WA 98195, USA
[11] Laboratoire de Physique de Clermont,IN2P3/CNRS, Université Clermont Auvergne, F-63000 Clermont-Ferrand, France
[12] Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Str. 1, D-85748 Garching, Germany
[13] Physik-Department, Technische Universität München, James-Franck-Straße 1, D-85748 Garching, Germany
[14] Department of Astronomy, University of Washington, Seattle, WA 98195, USA
[15] DiRAC Institute, University of Washington, Seattle, WA 98195, USA
[16] Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720, USA
[17] Kavli Institute for Particle Astrophysics and Cosmology, P.O. Box 20450, MS29, Stanford, CA 94309, USA
[18] Centre for Astrophysics and Cosmology, University of Nova Gorica, Vipavska 11c, 5270 Ajdovščina, Slovenia
[19] Laboratoire de Physique Nucleaire et de Hautes-Energies, Barre 12-22 1er etage, 4 place Jussieu, F-75005 Paris, France
[20] The Oskar Klein Centre for Cosmoparticle Physics, Department of Physics, Stockholm University, Stockholm, SE-106 91, Sweden
[21] Academia Sinica Institute of Astronomy and Astrophysics (ASIAA), 11F of ASMAB, No. 1, Section 4, Roosevelt Road, Taipei 10617, Taiwan
[22] Université Paris-Saclay, CNRS/IN2P3, IJCLab, F-91405 Orsay, France
[23] Department of Physics, Lancaster University, Lancaster, Lancashire LA1 4YB, UK
[24] Fermi National Accelerator Laboratory, P.O. Box 500, Batavia, IL 60510, USA
[25] Department of Physics & Astronomy, University College London, Gower Street, London WC1E 6BT, UK
[26] Department of Electrical Engineering and Computer Sciences, UC Berkeley, 253 Cory Hall, Berkeley, CA 94720-1770, USA
[27] Department of Physics, Harvard University, 17 Oxford St, Cambridge, MA 02138, USA
[28] Department of Astronomy, Harvard University, 60 Garden St, Cambridge, MA 02138, USA
Received 2021 December 15; revised 2022 January 25; accepted 2022 January 26; published 2022 April 6

Abstract

The generation-defining Vera C. Rubin Observatory will make state-of-the-art measurements of both the static and transient universe through its Legacy Survey for Space and Time (LSST). With such capabilities, it is immensely challenging to optimize the LSST observing strategy across the survey's wide range of science drivers. Many aspects of the LSST observing strategy relevant to the LSST Dark Energy Science Collaboration, such as survey footprint definition, single-visit exposure time, and the cadence of repeat visits in different filters, are yet to be finalized. Here, we present metrics used to assess the impact of observing strategy on the cosmological probes considered most sensitive to survey design; these are large-scale structure, weak lensing, type Ia supernovae, kilonovae, and strong lens systems (as well as photometric redshifts, which enable many of these probes). We evaluate these metrics for over 100 different simulated potential survey designs. Our results show that multiple observing strategy decisions can profoundly impact cosmological constraints with LSST; these include adjusting the survey footprint, ensuring repeat nightly visits are taken in different filters, and enforcing regular cadence. We provide public code for our metrics, which makes them readily available for evaluating further modifications to the survey design. We conclude with a set of recommendations and highlight observing strategy factors that require further research.

Unified Astronomy Thesaurus concepts: Cosmology (343); Observational cosmology (1146); Optical telescopes (1174); Sky surveys (1464)

1. Introduction

The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST), with its ability to make rapid, deep observations over a wide sky area, will deliver unprecedented advances in a diverse set of science cases (Abell et al. 2009, hereafter The LSST Science Book).

LSST has an ambitious range of science goals that span the universe: solar system studies, mapping the Milky Way, astrophysical transients, and cosmology; these are all to be achieved with a single 10 yr survey. Around 80% of LSST's observing time will be dedicated to the main or "wide, fast, deep" (WFD) survey, which will cover at least 18,000 deg^2. The remainder of the time will be dedicated to "mini-surveys" (for instance, a dedicated Galactic plane survey), "deep drilling fields" (DDFs), and, potentially, "targets of opportunity."

Because LSST has such broad science goals, the choice of observing strategy is a difficult but critical problem. Important early groundwork was laid in the Community Observing Strategy Evaluation Paper (COSEP; LSST Science Collaboration et al. 2017). To further address this challenge, in 2018, the LSST Project Science Team and the LSST Science Advisory Committee released a call for community white papers proposing observing strategies for the LSST WFD survey, as well as the DDFs and mini-surveys (Ivezić et al. 2019). In response to this call, the DESC Observing Strategy Working Group (hereafter DESC OSWG) performed a detailed investigation of the impact of observing strategy on cosmology with LSST. The DESC OSWG made an initial set of recommendations for both the WFD (Lochner et al. 2018) and DDF (Scolnic et al. 2018c) surveys, with proposals for an observing strategy that will optimize cosmological measurements with LSST. Following this call, many new survey strategies have been simulated to answer the ideas in various white papers submitted; these strategies are discussed in Jones et al. (2021). Furthermore, a Survey Cadence Optimization Committee (SCOC) was formed[29] with the charge of guiding the convergence of survey strategy decisions across the multiple LSST collaborations. The SCOC released a series of top-level survey strategy questions,[30] where answers can be supported using analyses of the simulations in Jones et al. (2021). In this paper, we evaluate a number of simulated LSST observing strategies to support decisions on the survey strategy.

A review of the dark energy analyses planned by the DESC (which is a subset of the fundamental cosmological physics that will be probed by LSST) is given in the DESC Science Requirements Document (The LSST Dark Energy Science Collaboration et al. 2018; hereafter DESC SRD). Each analysis working group (weak lensing, WL, large-scale structure, LSS, galactic clusters, type Ia supernovae, SNe Ia, and strong lensing) within DESC provided a forecast of the constraints on dark energy expected from their probe, given a baseline observing strategy. As a metric, the DESC SRD used the Dark Energy Task Force Figure of Merit (DETF FoM), defined as the reciprocal of the area of the contour enclosing 68% of the credible interval constraining the dark energy parameters, w_0 and w_a,[31] after marginalizing over other parameters (Albrecht et al. 2006). Once statistical constraints were quantified, each group determined the control of systematic uncertainties needed to reach the goals for a Stage IV dark energy mission.

Three ways to increase the likelihood of achieving the goals set out in the DESC SRD are to (a) improve the statistical precision of each probe, (b) reduce each probe's sensitivity to systematic uncertainties, or (c) reduce the total uncertainty when combining multiple probes. Changes in observing strategy have the potential to affect each of these. For instance, observing strategies that yield more uniform coverage across the survey footprint, or strategies with improved cadence, can have a strong impact on both the statistical precision and the systematic control.

DESC encompasses multiple cosmological probes, and it is the ultimate goal of the DESC OSWG to be able to compute the combined DETF FoM using all probes, as well as other combined metrics, for each proposed observing strategy. However, at the level of LSST precision, careful treatment of systematic effects is required, and work is still ongoing to include these in the forecasting analysis tools. In addition, a full cosmological analysis is computationally intensive and not feasible to test on hundreds of simulated LSST observing strategies. Thus, while we include an emulated DETF FoM for certain dark energy probes, we also introduce a suite of metrics that are anticipated to correlate with cosmological constraints but that are faster to run and simpler to interpret. Most of the metrics in this paper are focused on the WFD survey, but we make use of many of the same metrics (particularly for supernova, SN, cosmology) for the DDFs. It should be noted that one of the cosmological probes mentioned, clusters, is not explicitly included in our analysis. This is because it is expected that clusters will have identical requirements to LSS and so should already be accommodated.

While the metrics we have developed focus entirely on the extragalactic part of the WFD survey, there is one cosmological probe that relates to observations near the Galactic plane: the study of dark matter with microlensing. Microlensing is the light magnification due to the transit of a massive compact object (lens) close enough to the line of sight of a background star (Paczynski 1986). The search for the dark matter component of intermediate-mass black holes within the Milky Way through microlensing involves several-year timescale events, which can be efficiently detected only with long timescale surveys such as LSST (Mirhosseini & Moniez 2018). This search will not be sensitive to the details of observing strategy, as long as gaps larger than a few months are avoided. Thus, for this work, we focus only on extragalactic probes.

Although all of the metrics described in this paper are useful for understanding the impact of observing strategy on cosmological measurements with LSST, some are more closely related to the primary cosmology goals as outlined in the DESC SRD than others. One of these is a joint WL and LSS measurement referred to as the 3×2 pt correlation function. It involves the combination of three two-point correlation functions: shear–shear, galaxy–shear, and galaxy–galaxy correlations. This combined probe, which measures structure growth, and SNe, which measures the expansion history of the universe, together have the most constraining power. However, novel probes such as strong lensing and kilonovae (kNe) can be complementary and offer unique tests of cosmology beyond the DETF FoM. Our recommendations and conclusions are generally guided by the priorities outlined in the DESC SRD, but we attempt to quantify performance of observing strategies in terms of the scientific opportunities offered by more novel probes as well.

We highlight here the context of this paper: it is a summary of years of research and development that the DESC has undertaken, guided by the various milestones set up by the

[29] https://www.lsst.org/content/charge-survey-cadence-optimization-committee-scoc

[30] https://docushare.lsst.org/docushare/dsweb/Get/Document-36755

[31] Note that this FoM does not necessarily optimize the constraining power of alternative cosmological models beyond the simple dark energy parameterization considered here.

Rubin Project and outlined at the beginning of this section. It should therefore be noted there are only a few outstanding issues with the current default strategy—but the metrics and considerations that have led us here are worth elaborating on and keeping in mind as we move forward with the optimization of the observing strategy. This work forms part of the focus issue on Rubin observing strategy optimization, with the opening issue presented by Bianco et al. (2022).

We structure the paper as follows: Section 2 outlines the factors that affect LSST observing strategy, the simulator used, and resulting sets of simulations of different observing strategies. We split the metrics descriptions as follows: general static science metrics (Section 3); static science-driven metrics (Section 4), which include WL, LSS, and photometric redshifts; and transient science-driven metrics (Section 5), which include SNe, strong lensing of SNe/quasars, and kNe. We draw together the results of the analysis of our metrics on various simulated observing strategies in Sections 6 and 7. In addition to describing the analysis supporting the proposal for various observing strategy choices, we provide metrics, recommendations, and conclusions in this paper that are meant to be of more general use to future large-scale surveys.

2. LSST Observing Strategy

In this section, we describe the Rubin Observatory, the baseline LSST observing strategy, the software used to generate realistic LSST observing schedules that we make use of in this work, and the metrics framework used to evaluate different strategies.

2.1. LSST Overview

An overview of the Vera C. Rubin Observatory telescope specifications can be found in Ivezić & The LSST Science Collaboration (2013); we summarize here the specifications that impact observing strategy. The Rubin Observatory is under construction in the Southern Hemisphere, at Cerro Pachón in Northern Chile, and will undertake the LSST, a 10 yr survey expected to start in 2023. The system has an 8.4 m (6.7 m effective) diameter primary mirror, a 9.6 deg^2 field of view, and a 3.2 gigapixel camera. The integrated filter exchange system can hold up to five filters at a time, and there are six filters available: *ugrizy*, which cover a wavelength range of 320–1050 nm. Typical 5σ (i.e., the apparent brightness in magnitudes at which a point source is detected at 5σ significance) of 30 s exposures in *ugrizy* are 23.9, 25.0, 24.7, 24.0, 23.3, and 22.1 mag and coadded over the full survey will reach approximately 26.1, 27.4, 27.5, 26.8, 26.1, and 24.9 mag. Several performance specifications influence the survey cadence[32]: filter change time (120 s), closed optics loop delay or slews where altitude changes by more than 9° (36 s), read time (2 s), shutter time (1 s), and median slew time between fields (5 s).[33] The estimated fraction of photometric time is 53% of the 10 yr survey. Standard "visit" exposures are typically 30 s long (referred to as 1×30s). An alternative exposure strategy, to mitigate cosmic-ray and satellite trail artifacts, are two successive short exposures called "snaps" (this is referred to as 2×15 s). The decision between these exposure strategies has not yet been made. Throughout this paper, all simulations use 1×30 s exposures; however, we will return to this point in Section 6 to explicitly examine the impact of using 2×15 s exposures instead.

2.2. LSST Observing Strategy Requirements

The observing strategy of LSST is impacted by several factors, and its optimization is a complex challenge. The LSST Science Requirements Document (Ivezić & The LSST Science Collaboration 2013, hereafter LSST SRD) defines top-level specifications for the survey such that:

1. The sky area uniformly covered by the main survey will not be smaller than 15,000 deg^2, with a design goal of 18,000 deg^2.
2. The sum over all bands of the median number of visits in each band across the sky area will not be smaller than 750, with a design goal of 825.
3. At least 1000 deg^2 of the sky, with a design goal of 2000 deg^2, will have multiple observations separated by nearly uniformly sampled timescales ranging from 40 s to 30 minutes.

There are other additional requirements on point-spread function (PSF) ellipticity correlations and parallax constraints, the former of which will be indirectly analyzed in this paper.

Given that these requirements only constrain a few aspects of observing strategy, many remaining factors can still be optimized to maximize scientific return.

2.3. Baseline Strategy

Here we summarize the baseline observing strategy (Jones et al. 2021), which has evolved significantly over the years. The strategy described here is considered the current nominal observing strategy plan pending further modifications:

1. Visits, which are single exposures in a given filter toward a given pointing, are always 1×30 s long (not 2×15 s). The baseline simulation achieves about 2.2M visits over 10 yr.
2. Pairs of visits in each night are in two filters as follows: $u-g$, $u-r$, $g-r$, $r-i$, $i-z$, $z-y$, or $y-y$. Pairs are scheduled for approximately 22 minutes separation. Almost every visit in g, r, or i has another visit within 50 minutes. These visit pairs assist with asteroid identification.
3. The survey footprint is the standard baseline footprint, with 18,000 deg^2 in the WFD survey spanning declinations from $-62°$ to $+2°$ (excluding the Galactic equator), and additional coverage for the North Ecliptic Spur (NES), the Galactic Plane (GP), and South Celestial Pole (SCP). The baseline footprint includes WFD, NES (*griz* only), GP, and SCP. WFD is ∼82% of the total time.
4. Five DDFs are included,[34] with the fifth field being composed of two pointings covering the Euclid Deep Field–South,[35] devoting 5% of the total survey time to DD fields.

[32] The cadence is defined as the median internight gap over a season.

[33] These times are estimated from the expected performance from the various components of the telescope and camera and are what is used in the scheduling simulators.

[34] Details of the four selected DDFs can be found here: https://www.lsst.org/scientists/survey-design/ddf.

[35] This field has not been officially confirmed as part of the LSST survey. Information on the Euclid field can be found here: https://www.cosmos.esa.int/web/euclid/euclid-survey.

Table 1
Description of FBS Simulation Families Used Including a Descriptive Name of Each Family, the FBS Version, the Number of Simulations, and a Brief Explanation of the Family

Name	FBS Version	No. of Simulations	Description
Baseline	1.4/1.5/1.6	3	Baseline as described, with choice of 2 versus 1 snap, and mixed filters or not
U_pairs	1.4	12	Varies how *u*-band visits are paired with other filters, how many visits occur in the *u* band, and when the *u* band is loaded in or out of filter
Third_obs	1.5	6	Adds third visit to some fields night, where total amount of time dedicated to these visits is 15–120 minutes per night
Wfd_depth	1.5	16	Amount spent on WFD compared to other areas changes from 60% to 99%
Footprints	1.5	12	Changes in WFD footprint, in North/South, Galactic coverage, Large/Small Magellanic Clouds
Bulge	1.5	6	Different strategies for observing the Galactic bulge
Filter_dist	1.5	8	Varying the ratios of time spent in *u*, *g*, *r*, *i*, *z*, *y* filters
Alt_roll_dust	1.5	3	Dust-limited WFD footprint with the alt-scheduler scheduling algorithm, with and without rolling, where a rolling strategy only observes a set fraction of the survey footprint each year
DDFs	1.5	3	Different strategies for the DDFs, ranging from 3% to 5.5% of the total survey time
Goodseeing	1.5	5	Aims to acquire at least one good-seeing visit at each field each year, varies which filters it is needed for
Twilight_neo	1.5	4	Adds a mini-survey during twilight to search for near-Earth objects (NEOs)
Short exposures	1.5	5	Adds short exposures in all filters, from 1–5 s, two to five exposures per year
U60	1.5	1	Swaps 60 s *u*-band visit instead of 30 s
Var_expt	1.5	1	Changes exposure time so that the single image depth is constant
DCR	1.5	6	Adds high airmass observations in different combinations of filters one or two times per year
Even_filters	1.6	4	Bluer filters are observed in moon bright time
Greedy footprint	1.5	1	A greedy survey not run on ecliptic, where a portion of the sky that has the highest reward function is observed two times over a given time span (about 20–40 minutes).
Potential Schedulers	1.6	17	Multiple variations at once for a particular science goal.

5. The standard balance of visits between filters is 6% in *u*, 9% in *g*, 22% in *r*, 22% in *i*, 20% in *z*, and 21% in *y*.
6. Owing to the limitation of five installed filters in the camera filter exchange system, if at the start of the night the moon is 40% illuminated or more (corresponding to an approximately full moon ± 6 nights), the *z*-band filter is installed; otherwise the *u*-band filter is installed.
7. The camera is rotationally dithered nightly between $-80°$ and $80°$. At the beginning of each night, the rotation is randomly selected. The camera is rotated to cancel field rotation during an exposure, then reverts back to the chosen rotation angle for the next exposure.
8. Twilight observations are done in *r*, *i*, *z*, and *y*, and are determined by a "greedy" algorithm, which builds up a solution piece by piece, always choosing the next observation that offers the largest benefit given observing metrics/requirements.
9. Nonrolling cadence: the nominal baseline strategy observes the entire footprint each observing season. A *rolling cadence* would prioritize sections of the footprint at different times (e.g., observing half the footprint for one year and changing to the other half the next) to improve cadence in that section.

2.4. Survey Simulators

The simulations analyzed here are created using the Feature-Based Scheduler (FBS; Naghib et al. 2019), which uses a modified Markov decision process to decide the next observing direction and filter selection, allowing a flexible approach to scheduling, including the ability to compute a detailed reward function throughout a night. FBS is the new default scheduler for the LSST, replacing the LSST Operations Simulator (OpSim Ridgway et al. 2010; Delgado et al. 2014).

We note that, other than the LSST default schedulers (OpSim and FBS), there is an alternate scheduler, AltSched, presented in Rothchild et al. (2019). AltSched is a simple, deterministic scheduler, which ensures that telescope observations take place as close to the meridian as possible, alternating between sky regions north and south of the observatory latitude on alternate nights, while cycling through the filter set and changing filters after observing blocks. We do not include simulations from this scheduler; however, we note that its approach does produce encouraging results.

2.5. Observing Strategy Simulations

Sets of simulations have been periodically released for use by the community. In Table 1, we summarize the families of simulations used, number of simulations in each family, and versions.[36] New versions of simulated strategies are released regularly with improvements to the scheduler, weather simulation, and changes to the baseline strategy. In this paper, we mostly focus on version 1.5 simulations, but select v1.6 and v1.7 simulations are included in certain plots (see Jones et al. 2021, for details of the simulations). Certain simulations are excluded from specific plots because they are unrealistic or differ significantly from the baseline (for instance, with a dramatically different footprint or visit allocation in WFD). It is very important to note that for each version, the baseline changes somewhat. In particular, the default choice of exposure strategy has changed from 2×15 s in older versions, to 1×30 s in v1.5 and v1.6, and back to 2×15 s in v1.7, which has a large impact on overall efficiency and hence metric performance. All figures in this paper are for relative improvements compared to the baseline strategy corresponding to that simulation's version. We note that while the baseline

[36] Simulations can be downloaded at http://astro-lsst-01.astro.washington.edu:8081/.

Table 2
From the Various Simulations that Comprise the Multiple Simulation Families Listed in Table 1, We Choose a Subset of 11 Strategies to Focus on for the Initial Part of This

Simulation Name	Short Name	Longer Description
`baseline_v1.5_10yrs.db`	Baseline	Baseline strategy (described in Section 2.3)
`baseline_2snapsv1.5_10yrs.db`	2 × 15 s exposures	Same as baseline except exposures consist of two 15 s exposures instead of a single 30 s one
`bulges_bs_v1.5_10yrs.db`	Large footprint, Galactic bulge focus	Uses the "big sky" footprint but also includes coverage of the Galactic bulge
`footprint_big_sky_dustv1.5_10yrs.db`	Large footprint, extra-galactic focus	Uses the "big sky" footprint with a dust-extinction cut, completely avoiding the Galactic plane
`footprint_bluer_footprintv1.5_10yrs.db`	Bluer filter distribution	Baseline footprint with more observations in the bluer bands
`footprint_newAv1.5_10yrs.db`	Large footprint, Galactic plane focus	Uses the "big sky" footprint but increases depth in the Galactic plane
`var_expt_v1.5_10yrs.db`	Variable exposure	Baseline strategy but allows exposure time to vary between 20 and 100 s based on observing conditions to try to ensure constant single-visit depth
`wfd_depth_scale0.65_noddf_v1.5_10yrs.db`	65% of visits in WFD	Decreases the number of visits in WFD to 65% of available observing time, placing more emphasis on the mini-surveys
`wfd_depth_scale0.99_noddf_v1.5_10yrs.db`	99% of visits in WFD	Increases the number of visits in WFD to 99% of available observing time meaning there are essentially no mini-surveys or DDFs
`ss_heavy_v1.6_10yrs.db`	Solar system focus	Baseline footprint with ecliptic plane visits added. Some visit pairs are taken in the same filter (as opposed to different filters as is standard for baseline)
`combo_dust_v1.6_10yrs.db`	Combo dust	"Big sky" footprint defined by dust-extinction cut but including Galactic plane coverage.

Note. Here we present this list as well as a lookup table for the shorter, simpler names that are used in the figures in Sections 4 and 5. Note that here, "big sky" refers to a larger footprint that extends farther north and south and removes coverage of the Galactic plane (unless otherwise stated).

may change, the conclusions of relative importance compared to the baseline do not. Appendix B captures in detail exactly which simulations are used for which plot and the corresponding baseline simulation. As the number of total simulations is above 100, we choose a subset of the simulations to focus our analysis; these are representative simulations for the families as listed in Table 1 and have the greatest impact on our observing metrics. These are presented in Table 2, which includes a lookup table for the short, simpler names for the simulations used in some figures in this paper. For more details, we refer the reader to the report on Survey Strategy and Cadence Choices for the Vera C. Rubin Observatory LSST.[37]

2.6. Proxy Metrics and Metrics Analysis Framework

As stated in Section 1, the ultimate goal of the DESC OSWG is to compute cosmology figures of merit to evaluate observing strategies. However, this is difficult and computationally intensive, making such an approach impractical for evaluating many simulations. We thus largely focus on proxy metrics, which can be quickly computed on any simulation. We make use of and incorporate our metrics into the metrics analysis framework (MAF; Jones et al. 2014). MAF is a python framework designed to easily evaluate aspects of the simulated strategies. It can compute simple quantities, such as the total coadded depth or number of visits, but it can also be used to construct more complex metrics that can evaluate the expected performance of a simulation for a given science case. Here, we use a combination of independent codes (which are too slow to run as part of MAF) and custom MAF metrics to evaluate the simulated observing strategies. MAF metrics and external metrics created for this paper are linked when described. Unless otherwise specified, each metric is run on a simulation of the full 10 yr survey.

We note that in order to compare metrics directly, they must be transformed to be able to be interpreted as "larger is better" and placed on the same scale. Appendix A includes a table that describes how all metrics are transformed. In all plots, to put the metrics on the same scale, they are normalized using their values for the baseline simulation (which is different for each FBS version) as:

$$x_{\rm normed} = \frac{x - x_{\rm baseline}}{x_{\rm baseline}}, \quad (1)$$

where x is the metric in question, $x_{\rm baseline}$ is the value of that metric for the corresponding baseline simulation, and $x_{\rm normed}$ is the normalized metric.

A final point to note before introducing the metrics is that the focus of this paper is the optimization of the WFD survey; thus, all metrics are evaluated only on the WFD observations of each simulation. However, we include some discussion of DDF optimization in Section 7.2, which is particularly important for SNe and also photometric redshift calibration.

3. General Static Science Metrics

In this section, we introduce metrics relevant to static science topics that will be useful to multiple cosmological probes. This includes metrics related to general WFD characteristics as well as to photometric redshift (z) characteristics from the WFD sample.

3.1. WFD Depth, Uniformity, and Area

Depth–area trade-offs caused by the availability of a fixed amount of observing time are commonly encountered when designing astronomical surveys. One could therefore assume

[37] https://pstn-051.lsst.io/

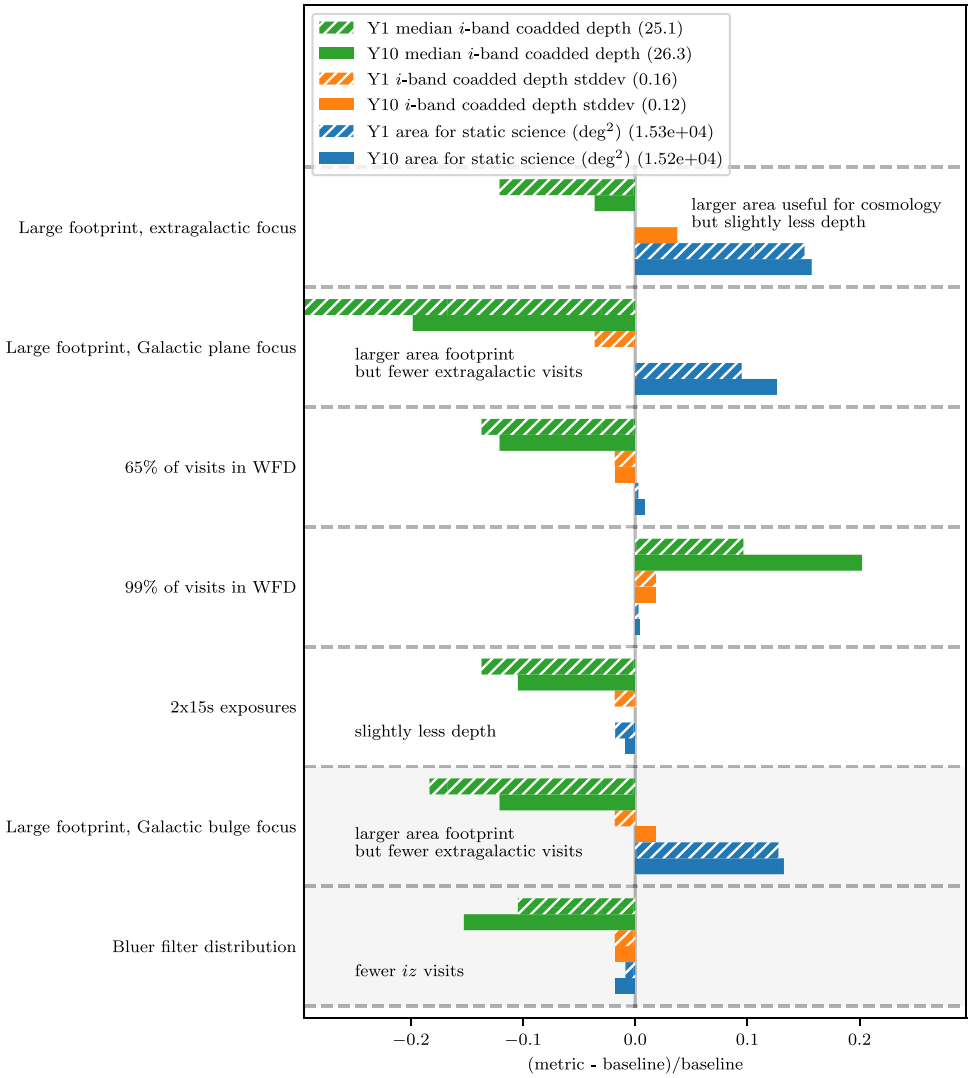

Figure 1. Static science metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. Larger area simulations clearly indicate the usual area vs. depth trade-off, while increasing extragalactic visits improves all metrics. The efficiency lost with 2×15 s exposures is clear, and we note that bluer filter distributions decrease all metrics, which are evaluated on the i and z bands.

that optimizing the LSST survey design for static science requires finding the best-performing location in a two-dimensional space of area versus the number of visits subject to a fixed observing time constraint. However, there are additional complexities that must be considered, including the uniformity of the resulting survey and the proportion of visits assigned to each filter. Those additional complexities motivate the more in-depth studies of observing strategy trade-offs for static science presented in this and subsequent sections.

In this subsection, we introduce three sets of metrics, where each set includes information after years Y1 and Y10 and is also calculated at a few intermediate steps such as Y2, Y4, and Y7 (Y1 refers to the data collected after the first year, Y2 after two years, etc.). Figure 1 shows the results from these metrics. The 5σ depth used below refers specifically to the magnitude of a point source that would be detected with a signal-to-noise ratio (S/N) of 5; see Ivezić et al. (2019).

1. Y1/Y10 area for static science (deg^2): The effective survey area for static science after Y1/Y10; this area is limited by depth and extinction, and requires coverage in all six bands, and is described in more detail below. Note that this area is also referred to as "the extragalactic area" in later discussion.
2. Y1/Y10 median i-band coadded depth: The median i-band 5σ coadded depth in the effective survey area for static science after Y1/Y10.
3. Y1/Y10 i-band coadded depth stddev: The standard deviation of the i-band 5σ coadded depth distribution in the effective survey area for static science after Y1/Y10, quantifying the nonuniformity in depth coverage; smaller values of this metric indicate a more uniform survey.

We follow the LSST Science Book in using the i band to track the brightness limit of galaxies that can be detected in the survey, motivated by the fact that almost all galaxies are brighter in i than in g or r, while the coadded depths are similar for these three filters. We note that this could be misleading when comparing observing strategies that vary the relative time spent observing in the i band versus other bands by a significant factor.

The metrics above are calculated using HEALPix[38] (Górski et al. 2005) maps, with a pixel resolution of 13.′74 (achieved using the HEALPix resolution parameter $N_{\rm side} = 256$). For extragalactic static science using high S/N measurements, we must restrict our analysis to a footprint that provides the deep, high S/N samples needed for our science. To achieve this, we implement an extinction cut and a depth cut, retaining only the area with $E(B-V) < 0.2$ (where $E(B-V)$ is the dust reddening in magnitudes) with limiting i-band coadded 5σ point-source detection depths of 24.65 for Y1 and 25.9 for Y10; the $E(B-V)$ cut ensures that we consider the area with small dust uncertainties[39] while the depth cut ensures that we have high S/N galaxies, with the Y10 cut fixed by the LSST SRD goal of yielding a "gold sample" of galaxies with i-band coadded 5σ (extended source detection depth), $i < 25.3$ after Y10. This is achieved using the MAF Metric object, egFootprintMetric.[40]

3.1.1. Uniformity and Dithering

Survey uniformity, as measured by our i-band coadded depth stddev metric, is critical for all static science probes. Nonuniformity can be introduced by spending more observing time or having better atmospheric conditions in certain parts of the sky, or when a survey is tiled and the overlaps in fields introduce an artificial structure to the survey. The latter effect can be effectively mitigated using dithering: small offsets in the pointing of the telescope when it returns to a field (see, e.g., Awan et al. 2016). Dithering can be translational or rotational, both of which are useful for reducing different types of systematics. Figure 1 shows that the stddev metric varies by less than 5% across the simulations; all of the current observing strategy proposals implement similar dithering strategies, and it appears that none of them stand out in selectively favoring one portion of the sky over another. While most metrics in this paper focus on the performance of the full 10 yr LSST survey, we note that it is important that survey uniformity is achieved at specific release intervals, such as Y1, Y2, Y4, Y7, and Y10, in order to enable periodic analyses of data sets suitable for cosmology. The current baseline achieves this by default, but it will be important to consider if a strategy is chosen whereby only parts of the footprint are observed each season (so-called rolling cadence).

3.1.2. General Conclusions from Static Science

Static science systematics can be reduced by increasing survey uniformity via frequent translational and rotational dithers; the impacts of these can be probe-specific, as discussed below, but overall, the current simulations do not lead to dramatic variations in overall depth and depth uniformity in the extragalactic footprint. We specifically note that Y1 is especially sensitive to the specific cadence, and while the different kinds of cadences/footprints converge for Y3–Y10 area, very few simulations yield close to the desired 18,000 deg^2 WFD area for extragalactic science. We also emphasize the need to check the depth statics at intermediate intervals especially for the case of rolling cadence as that will have immediate impacts on our science during the course of the 10 yr survey.

We also note that spectroscopic observations in the extragalactic part of the survey will be essential to calibrate LSST photometric redshifts. For this purpose, overlap with upcoming spectroscopic surveys is quite critical and is further discussed in Section 6.

3.2. Photometric Redshifts

While photometric redshifts impact multiple probes, including transients such as SNe, they are in turn not affected by time-sensitive aspects of observing strategy. We thus generally include photo-z metrics with the static science metrics. We introduce four metrics for the quality of photo-z determination in the WFD:

1. photo-z standard deviation at high (1.8–2.2) and at low (0.6–1.0) redshift;
2. outlier fraction at high and at low redshift.

The high- and low-redshift bins were chosen to represent two different regimes to be explored by the LSST. A summary of the results of these metrics can be seen in Figure 2.

We evaluate the relative quality of simulated photometric redshift estimates for each simulation by determining the average coadded depth in extragalactic fields, and using those depths to simulate observed apparent magnitudes and photo-z for a mock galaxy catalog using the color-matched nearest neighbors (CMNN) photometric redshift estimator Graham et al. (2018).[41] The CMNN estimator does not produce the "best" or "official" LSST photo-z, but does produce results for which the *relative quality* of the photo-z is directly related to the input photometric quality, and thus is an appropriate photo-z estimator for evaluating the bulk impact on photo-z results due to changes in the photometric depth of a survey. As shown in Graham et al. (2018, 2020) the standard deviation and fraction of outliers for photo-z from the CMNN estimator increase monotonically between our representative low- and high-redshift bins.

First, we determine the 5σ point-source limiting magnitudes of the 10 yr coadded images from the WFD program in sky regions ($\sim 220'$ wide) for all simulations. We consider extragalactic fields as appropriate for cosmological studies if their Galactic dust extinction is $E(B-V) < 0.2$ mag, and if they receive at least five visits per year in all six filters (i.e., the five visits define a minimum coadded depth). The median 10 yr six-filter depths across all such appropriate extragalactic fields for each simulation are then input to the CMNN photo-z estimator. The CMNN estimator uses the depths to synthesize apparent observed magnitudes and errors for a simulated galaxy catalog; then it splits the catalog into test and training sets, and uses the training set to estimate photo-z for the test set. The 5σ depths are the only input; thus, only observing strategies that result in photometric depths that differ significantly from the baseline cadence will result in significantly different photometric redshift results.

We used the same mock galaxy catalog as described in Graham et al. (2018, 2020), which is based on the Millennium

[38] http://healpix.sourceforge.net/

[39] As noted by, e.g., Leike & Enßlin (2019), the behavior of Galactic dust becomes more uncertain as the amount of dust increases, and Schlafly & Finkbeiner (2011) identified $E(B-V) = 0.2$ as a threshold where the dust properties change.

[40] https://github.com/humnaawan/sims_maf_contrib/blob/master/mafContrib/lssmetrics/egFootprintMetric.py

[41] A demonstration of the CMNN photo-z estimator is available on GitHub at https://github.com/dirac-institute/CMNN_Photoz_Estimator.

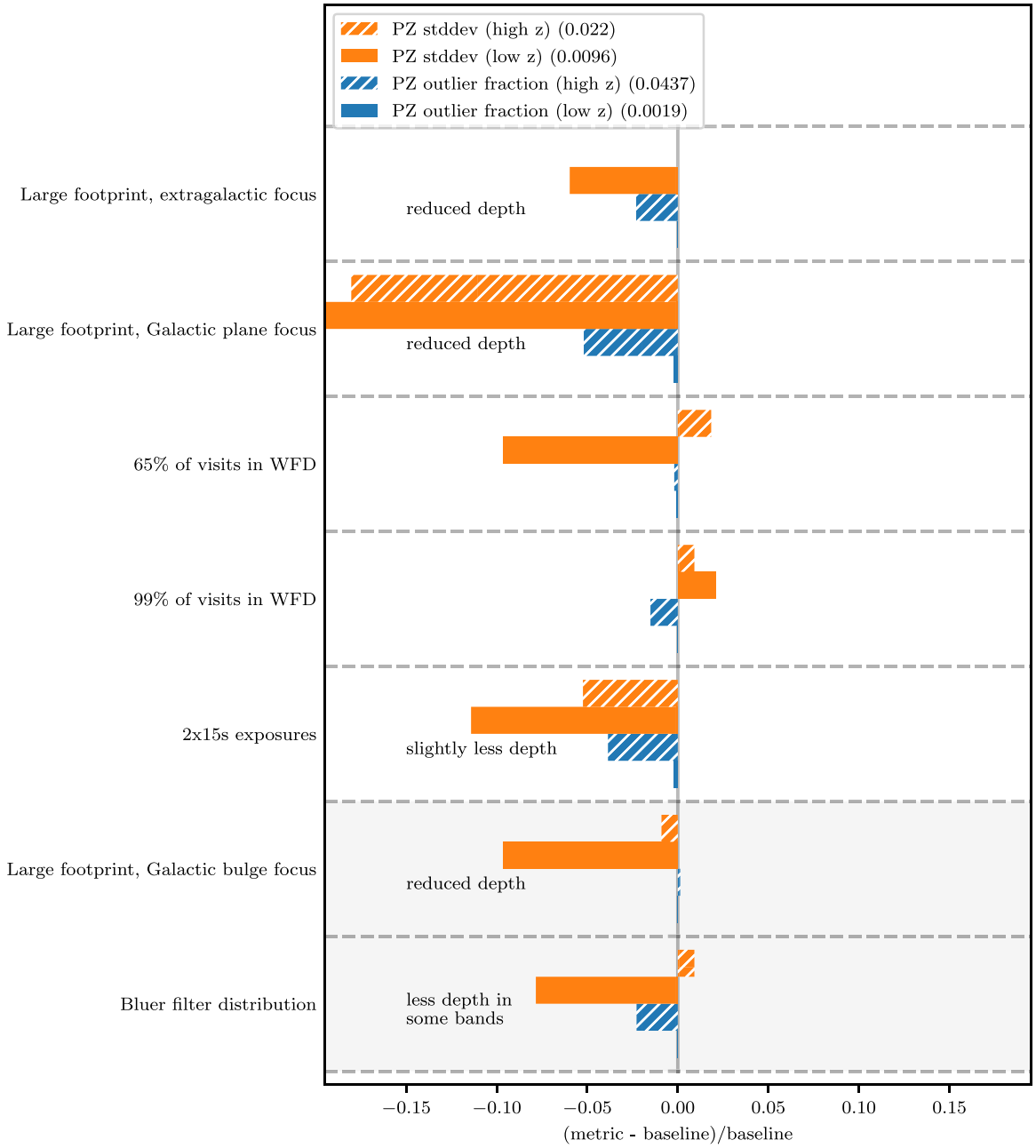

Figure 2. Photometric redshift metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. As described in Section 3.2, the low-z bin is (0.6–1.0) and the high-z bin is (1.8–2.2). The most significant degradation to photo-z metrics comes from reducing the depth in the extragalactic part of the footprint, losing efficiency from 2 × 15 s exposures and redistributing visits to bluer bands.

simulation (Springel et al. 2005) and the galaxy formation models of Gonzalez-Perez et al. (2014), and was fabricated using the light-cone construction techniques described by Merson et al. (2013).[42] To both the test and training sets, we apply cuts on the observed apparent magnitudes of 25.0, 26.0, 26.0, 25.0, 24.8, and 24.0 mag in filters *ugrizy*, and enforce that all galaxies are detected in all six filters. These cuts are all about half a magnitude brighter than the brightest 5σ depth of any given simulation we considered. This cut is applied because it imposes a galaxy detection limit across all simulations that is independent of the depth (i.e., independent of the photometric quality). If such a cut is not imposed, the default setting is for the CMNN estimator to apply a cut equal to the 5σ limiting magnitude. This default setting results in more fainter galaxies being included in the test and training sets for simulations with deeper coadds. Although this default setting is realistic—fainter galaxies are included in real galaxy catalogs made from deeper coadds—due to the fact that fainter galaxies generally have poorer-quality photo-z estimates, this also results in some simulations with deeper coadded depths appearing to produce *worse* photo-z estimates. These bright magnitude cuts ensure an "apples-to-apples" comparison across all simulations.

[42] Documentation for this catalog can be found at http://galaxy-catalogue.dur.ac.uk.

All other CMNN estimator input parameters are left at their default values, except for the number of test (50,000) and training (200,000) set galaxies, and the minimum number of colors, which is set to five (from a default of three) to only include galaxies that are detected in all six filters. The other CMNN parameters[43] impact the final absolute photo-z quality, and so it is important to keep in mind that the results of the CMNN estimator should not be interpreted as absolute predictions for the photo-z quality, but as *relative* predictions for the photo-z quality produced by different observing strategies (i.e., different 10 yr coadded depths). It is important to note that, because the test and training sets are drawn from the same population, they have the same apparent magnitude and redshift distributions. This contrived scenario in which the training set is perfectly representative of the test set does not produce photo-z results with realistic systematic effects or biases. Additionally, we use input parameters for the CMNN estimator that produce photo-zs with a very good absolute quality. The combination of nearly perfectly matched test and training sets, optimized input parameters, and bright magnitude cuts results in very small bias values (where bias is the average of $|z_{\rm true} - z_{\rm phot}|$ over all test galaxies) for our simulations, which is why the photo-z bias is not being used as one of the metrics (described below) for evaluating the simulations. In future photo-z simulations, variations in the test and training sets could be established that correspond to different simulations (e.g., building a deep training set from the DDFs)—but we consider this out of the scope of the present analysis.

We evaluate the photo-z results with two statistics: precision and outlier fraction. To calculate the precision, we first reject catastrophic outliers with $|z_{\rm true} - z_{\rm phot}| > 1.5$ (this is a non-standard definition, chosen for this simulated data set, and used also in Graham et al. 2020). Then we calculate the robust standard deviation in the photo-z error, $\Delta z_{1+z} = (z_{\rm true} - z_{\rm phot})/(1 + z_{\rm phot})$, by using the value of the interquartile range as an FWHM and dividing by 1.349 to convert to standard deviation (by definition, $\sigma = {\rm FWHM}/1.349$). The outlier fraction is the number of galaxies with a photo-z error greater than three times the standard deviation or greater than three times 0.06, whichever is larger (this definition matches that used for photo-z outliers by the LSST SRD). We calculate these two statistics for a low-$z_{\rm phot}$ bin (0.6–1.0) and a high-$z_{\rm phot}$ bin (1.8–2.2), for each of the simulations.

The results are shown in Figure 2. As mentioned above, the results of the CMNN estimator should be considered as *relative* predictions for the photo-z quality. Thus in Figure 2 we show the results as *fractional* changes from the results for the baseline simulation.

3.2.1. General Conclusions from Photometric Redshifts

We find that the photo-z quality is optimized by observing strategies that lead to deeper limiting magnitudes, which is as expected, and vice versa. For Figure 2, most of the selected observing strategies included in that plot lead to shallower limiting magnitudes in the WFD region because they, for example, spend more time in Galactic regions or on mini-surveys, have higher overheads (2×25 s exposures), or increase the WFD area at the expense of depth. The point of Figure 2 is to demonstrate the relative amount of degradation in photo-z quality due to such changes to the observing strategy, but strategies that instead provide deeper limiting magnitudes can improve the standard deviation in the photo-z, as seen in the fourth row of Figure 2 for the strategy in which 99% of the survey time is spent in the WFD region.

As there is a trade-off in any survey between depth and area, and because areal coverage is required by a variety of LSST science goals, we recognize that the observing strategies that optimize photo-z quality might not be optimal for cosmological studies. The ideal approach would be to include photometric redshifts in the science metrics that they impact so they can be jointly optimized. For instance, increasing the effective number density and mean redshift of the WL sample may decrease the photo-z accuracy, leading to weaker cosmological constraints. Such an analysis is, however, beyond the scope of this paper. Different science cases have different photo-z needs, and a single photo-z metric that captures the science impact is very challenging to define. Full end-to-end simulations of scientific results that incorporate photo-z quality would be the correct approach, but are beyond the scope of this work.

4. Weak Lensing and Large-scale Structure

For this section, we discuss two cosmological probes: WL and LSS. Due to the high degree of synergy between these probes, we present our general conclusions jointly for them in Section 4.3.

4.1. Weak Lensing

WL is the deflection of light from distant sources due to the gravitational influence of matter along the line of sight. In practice, the coherent distortions of background galaxy shapes, or "shear" (measured in different redshift ranges), reveal the clustering of matter as a function of time, including both luminous and dark matter. The evolution of matter clustering is affected by the expansion history of the universe, which means that WL is also sensitive to the accelerated expansion rate of the universe caused by dark energy (for a review, see Kilbinger 2015).

We introduce two new metrics associated with WL. A summary of the results from the WL metrics can be found in Figure 3.

1. WL+LSS FoM (Section 4.1.1)—DETF FoM for cosmological WL and LSS measurement; larger numbers correspond to larger statistical constraining power.
2. WL Average visits (Section 4.1.2)—Average number of visits metric in the r, i, and z bands; higher numbers are better for WL shear systematics mitigation.

4.1.1. WL+LSS Figure of Merit

The 3×2 pt (DETF) FoM is the inverse of the area of the 68% confidence interval in the space of the dark energy equation of state parameters w_0 and w_a, and is an indicator of the statistical constraining power of the survey's static probes. In this section, we follow the convention taken by ongoing WL surveys that the canonical analysis is a joint measurement of WL and LSS. Since forecasting the cosmological constraining power for such a measurement is quite resource-intensive, our approach is to carry out the calculation for a limited number of survey strategies, and use that to build an emulator of the $3 \times$

[43] Other CMNN parameters include, e.g., the minimum number of CMNN training-set galaxies, the mode of determining the photo-z from the CMNN subset, and the percent-point function value, which is used to generate the CMNN subset.

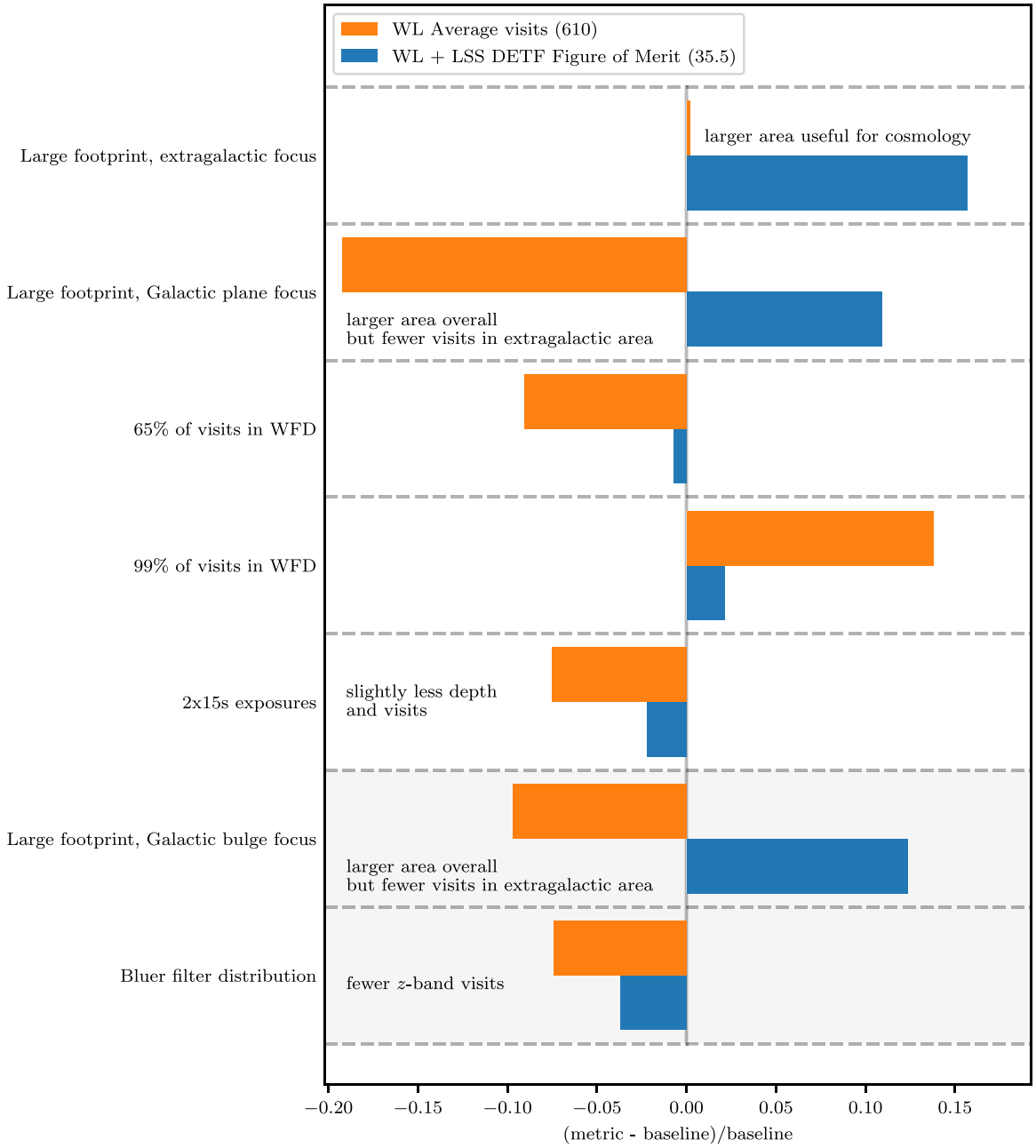

Figure 3. WL metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. Note that the extinction cut $E(B-V)<0.2$ is always applied. As shown, the metrics that describe statistical constraining power ("WL + LSS DETF FoM") and reduction in systematic uncertainties in WL shear ("WL Average visits") do not necessarily correlate, as the former tends to prefer an increased area while the latter tends to prefer an increased number of visits at each point in the WFD survey on average. Thus, observing strategy changes such as increasing the fraction of time in the WFD survey can increase both metrics, while observing strategy changes that increase area at the expense of decreasing the average number of visits across the survey area will lead to opposing changes in the metrics.

2 pt FoM for arbitrary scenarios based on interpolation. This section briefly describes the emulation process for the 3×2 pt FoM; for more detail, see T. Eifler & J. Motka (2022, in preparation).[44]

The emulator is defined in a six-dimensional parameter space, where the dimensions are: effective survey area, survey median depth in the i band, systematic uncertainty in the WL shear calibration, photometric redshift scatter, and the size of priors on photometric redshift bias and scatter. A total of 36 points in this parameter space were selected to build the emulator, following a Latin hypercube design (often used for efficient sampling of high-dimensional parameter spaces in cosmological emulators; e.g., Heitmann et al. 2009; Mead et al. 2021). For each selected point in this parameter space, the galaxy redshift distributions, and the observable quantities for joint WL and LSS measurement, and their covariance matrix are calculated using CosmoLike (Krause & Eifler 2017). The

[44] https://github.com/hsnee/sims_maf/blob/master/python/lsst/sims/maf/metrics/summaryMetrics.py

FoM for this 3×2 pt measurement is then calculated using Markov Chain Monte Carlo constraints on cosmological parameters (marginalizing over key sources of astrophysical systematic uncertainty, such as galaxy bias, intrinsic alignments, and baryonic physics) in a simulated likelihood analysis based on the observables and their covariances. The reason to marginalize over those systematic uncertainties when computing the FoM is that this is the process used in a real WL analysis to propagate the aforementioned systematic uncertainties into uncertainties on cosmological parameters. The emulator was then built from those 36 FoM values in the six-dimensional space using a Gaussian process regression. The accuracy of the emulator was validated by sequentially omitting each point used to build the emulator, rebuilding it with the remaining points, and testing recovery of the emulated FoM compared to the directly computed one. Typical accuracy was better than 10%.

Area and median depth are calculated after making an extinction cut of $E(B-V) < 0.2$, to exclude high extinction areas, along with minimum depth cuts of 24.5 and 25.9 for Y1 and Y10, respectively, to ensure that the survey depth is relatively homogeneous; as well as a cut that guarantees at least one visit in all six bands to ensure photo-z quality. Functionally, the cut on at least one visit in all six bands does very little to the area footprint, because these strategies were designed to ensure there is coverage in all bands throughout the WFD survey. As a result, even though the cut requires at least one visit in the other bands besides i (where there is a strict depth cut), the distribution of magnitudes in all bands is compact and reflects that the points that are sampled across the survey footprint have many observations in each band. In all bands, there is some typical coadded depth with some regions having a depth that is at most 0.5 mag shallower than that typical value. These cuts are consistent with the extragalactic cuts applied in Section 3.1. See the DESC SRD for more details about sample definition and methods of estimating redshift distributions and number counts, though the figures of merit in that document were calculated with slightly different choices of redshift binning and modeling of systematic uncertainty.

Note that the plots in this paper only use the area and depth dimensions of the emulator, with the remaining parameters fixed to their fiducial values from the DESC SRD. We used the emulator with marginalization over default values for the photometric redshift systematics parameters; therefore, the only effects it captures are the varying area and median depth of the different survey strategies. The effects of area and median depth changes are the dominant factors for the emulator, with changes in photometric redshift systematics being subdominant. For example, a 15% change in photometric redshift variance (maximum change for the strategies considered in this work) would cause a $\sim$2% change in the 3×2 pt FoM (T. Eifler & J. Motka 2022, in preparation), while a similar change in area or median depth changes the 3×2 pt FoM by $\sim$10%. Changes due to the priors on the photometric redshift variance and bias have negligible effects for the type of strategies considered (specifically, the considered changes of overlap with spectroscopic surveys used for calibrating photo-z errors), and are likewise not included in this work. As a rule, the 3×2 pt FoM prefers greater median depths and larger survey areas. The 3×2 pt FoM metric described here is an improved version of the associated metric presented in Lochner et al. (2018); while the improved version includes more sources of systematic uncertainty, its trends with survey area and depth are similar to those in Lochner et al. (2018).

4.1.2. WL Average Visits and WL Shear Systematics

The statistical constraining power for WL was covered in the previous subsection. For this reason, the text below summarizes observing strategy considerations related to WL shear systematics in the WFD survey, for which a metric has been introduced in MAF.[45] See Almoubayyed et al. (2020) for further detail.

WL analysis typically involves measuring coherent patterns in galaxy shapes due to WL shear. For this reason, any effects that are not associated with WL but that cause apparent galaxy shape distortions with any spatial coherence must be well understood and controlled to avoid the measurement being systematics-dominated. The LSST provides a new opportunity to control WL systematics using the observing strategy. This opportunity was not as feasible in previous surveys because the LSST will be the first survey to dither at large scales (relative to the field of view) with a very large number of exposures. This large number of exposures means that a source of systematic error with a particular spatial direction in one exposure may contribute with a different direction in other exposures for a given object, thus reducing the amount of systematic error that must be controlled in the image analysis process. Similar studies for systematics associated with differential chromatic refraction and CCD fixed frame distortions were conducted in the COSEP and were found to be minimized for a uniform distribution of parallactic angles and the position angle of the LSST camera over all visits.

Additive shear systematics, such as those induced by errors in modeling the PSF and errors arising from the CCD charge transfer, often have a coherent spatial pattern in single exposures. This type of systematic can potentially be mitigated and averaged down in coadded images, depending on the details of the dithering and observing strategy.

In Almoubayyed et al. (2020), we developed a physically motivated analysis related to the impact of observing strategy WL shear systematics, then used our findings with it to design a simpler proxy metric. Here, we describe the physically motivated analysis in four steps: (a) We select a large number (e.g., 100,000) of random points at which the PSF is to be sampled, distributed uniformly in the WFD area of each survey, with cuts based on the coadded depth and dust extinction as explained in Section 3.1. (b) We create a toy model for the PSF modeling errors as a function of position in each exposure as a radial error in the outer 20% of field of view (and no error in the inner 80%), and for modeling CCD charge transfer errors and the brighter-fatter effect, we use a horizontal (CCD readout direction) error over the stars in the entire field of view. This model is motivated by observed spatial patterns in PSF model errors in ongoing surveys (Jarvis et al. 2016; Bosch et al. 2018). (c) To approximate the effect of coaddition, we average down the modeling errors across exposures via their second moments, since the coaddition process is linear in the image intensity and therefore in the (unweighted) moments. (d) We propagate the systematic errors for the PSF in the coadded image into the bias on the cosmic shear using the ρ-statistics formulation (Rowe 2010; Jarvis et al. 2016).

[45] https://github.com/lsst/sims_maf/blob/master/python/lsst/sims/maf/metrics/weakLensingSystematicsMetric.py

To create a proxy metric that connects more directly with survey parameters and is more practical to run for every simulation, we note that given a chosen dithering pattern (e.g., a random translational dither per visit with random rotational dithering at every filter change), the reduction in systematic errors is directly related to the number of visits that are used in WL analysis. This is due to the fact that more visits leads to a better sampling of a rotationally uniform distribution around the coherent direction associated with the additive shear systematic. We, therefore, use the average number of *r*-, *i*-, and *z*-band visits for a large number of objects—for practical reasons, picked as the centers of cells in a HEALPix grid (Górski et al. 2005). The higher this number, the better a survey strategy performs. Increasing the number of visits need not be done at fixed exposure time, so it is not necessarily the case that increasing the number of visits requires a decrease in survey area; rather, there are a variety of area–depth trade-offs possible for scenarios with increased numbers of visits. Even a decrease to 20 s exposures in some bands can be impactful for this metric. Note that the bands to be used for WL shear estimation have not been decided yet; however, *riz* are likely to dominate due to their higher S/N for WL-selected samples, which is why we choose to focus on them here.

Due to uncertainties in the level of detector effects and other sources of additive shear systematics, and in the performance of instrument signature removal methods (which determine how sensor effects may contaminate the PSF estimated from bright stars), this metric has an arbitrary normalization and can only differentiate between the relative improvement between different strategies. Existence of on-sky LSSTCam data will provide a direct estimate of the level of additive shear systematics that need to be mitigated via observing strategy. Therefore, the impact of this metric can only be directly compared with that of the 3×2 pt FoM defined in Section 4.1.1 once Rubin Observatory commissioning begins. Nonetheless, we can already identify areas of parameter space where it will be interesting to optimize trade-offs between the WL systematics metric and the 3×2 pt FoM. There is a potential trade-off between improving on 3×2 pt FoM and mitigating WL shear systematics, as the 3×2 pt FoM prefers an increase in area, while WL shear systematics are mitigated with a larger number of well-dithered visits; the relative importance of these metrics can be optimized at the time of commissioning. Strategies that increase the usable area for WL and decrease exposure time can lead to improvement in both metrics simultaneously. It is worth noting that the 2×15 s exposures will most likely be combined into a single observation, so this exposure strategy does not actually increase the number of visits and reduces both metrics owing to less efficiency overall (due to the extra readout time between exposures). As shown in Figure 3, increasing the amount of time in the WFD survey, going from 2×15 s to 1×30 s exposures, and using a redder filter distribution leads to improvement in both metrics.

4.2. Large-scale Structure

LSS constrains cosmological parameters via observations of galaxy clustering. LSS is a more localized tracer of the matter distribution, rather than an integral along a line of sight like WL. As a result, the constraining power of LSS is more sensitive to bias, scatter, and catastrophic errors in photometric redshift estimation, as these determine how much the clustering signal is degraded by projection along the line of sight (Chaves-Montero et al. 2018). Artificial modulations in the observed galaxy number density caused by depth variations and observing conditions (e.g., sky brightness, seeing, clouds) provide key systematic errors in measuring galaxy clustering (Awan et al. 2016). Additionally, Galactic dust impacts the brightness and color of each galaxy (e.g., Li et al. 2017), and correcting for these effects becomes more difficult in regions with high levels of Galactic dust reddening.

For the LSS probe, we introduce two new metrics in detail below. A summary of the results of these metrics can be seen in Figure 4.

1. Y1/Y10 $N_{\rm gal}$ at $0.66<z<1.0$ (Section 4.2.1): Estimated number of galaxies at $0.66<z<1.0$ based on the Y1/Y10 *i*-band coadded depth in the effective survey area.
2. LSS systematics FoM for Y1/Y10 (Section 4.2.2): LSS systematics diagnostic FoM for Y1/Y10, comparing the uncertainty added by the Y1/Y10 survey nonuniformity versus that achieved for the baseline strategy using the Y10 gold sample (as defined in Section 3.1).

4.2.1. Galaxy Counts

In order to get cosmological constraints from *n*-point statistics, e.g., two-point correlation functions or the two-point power spectra, LSST will offer an unprecedentedly large and deep galactic sample, allowing us to carry out analyses in a regime where statistical uncertainties will be subdominant to systematic ones. To estimate the number of galaxies, we follow Awan et al. (2016) and propagate the simulated 5σ coadded depth to the number of galaxies using light-cone mock catalogs.

Using the `MAF` object from Awan et al. (2016), we create a new `MAF` object, `depthLimitedNumGalMetric`,[46] that calculates the number of galaxies in the extragalactic footprint (as defined in Section 3.1).

Evaluating the metric on the latest simulations, we find only small variations in the total galaxy counts across the observing strategies. Given that all strategies lead to samples comprising billions of galaxies that easily beat the shot-noise limit, even the 10%–15% variations we see in the galaxy counts are not critical for LSS. The `depthLimitedNumGalMetric` will, however, provide a good sanity check to ensure that no catastrophic changes are introduced, and to test the impacts of more complex strategies like rolling cadence.

4.2.2. Systematics Introduced by the Observing Strategy

Spatial fluctuations in galaxy counts represent LSS and hence are of interest for dark energy science. As discussed in Awan et al. (2016), artificial structure induced by the observing strategy leads to systematic uncertainties for LSS studies. Specifically, while asystematic bias induced in the measured LSS can be corrected, the uncertainty in our knowledge of this bias leads to uncertainties that affect our measurements. In order to quantify the effectiveness of each cadence in minimizing the uncertainties in the artificial structure that is induced by the observing strategy, we update the LSS FoM

[46] https://github.com/humnaawan/sims_maf_contrib/blob/master/mafContrib/lssmetrics/depthLimitedNumGalMetric.py

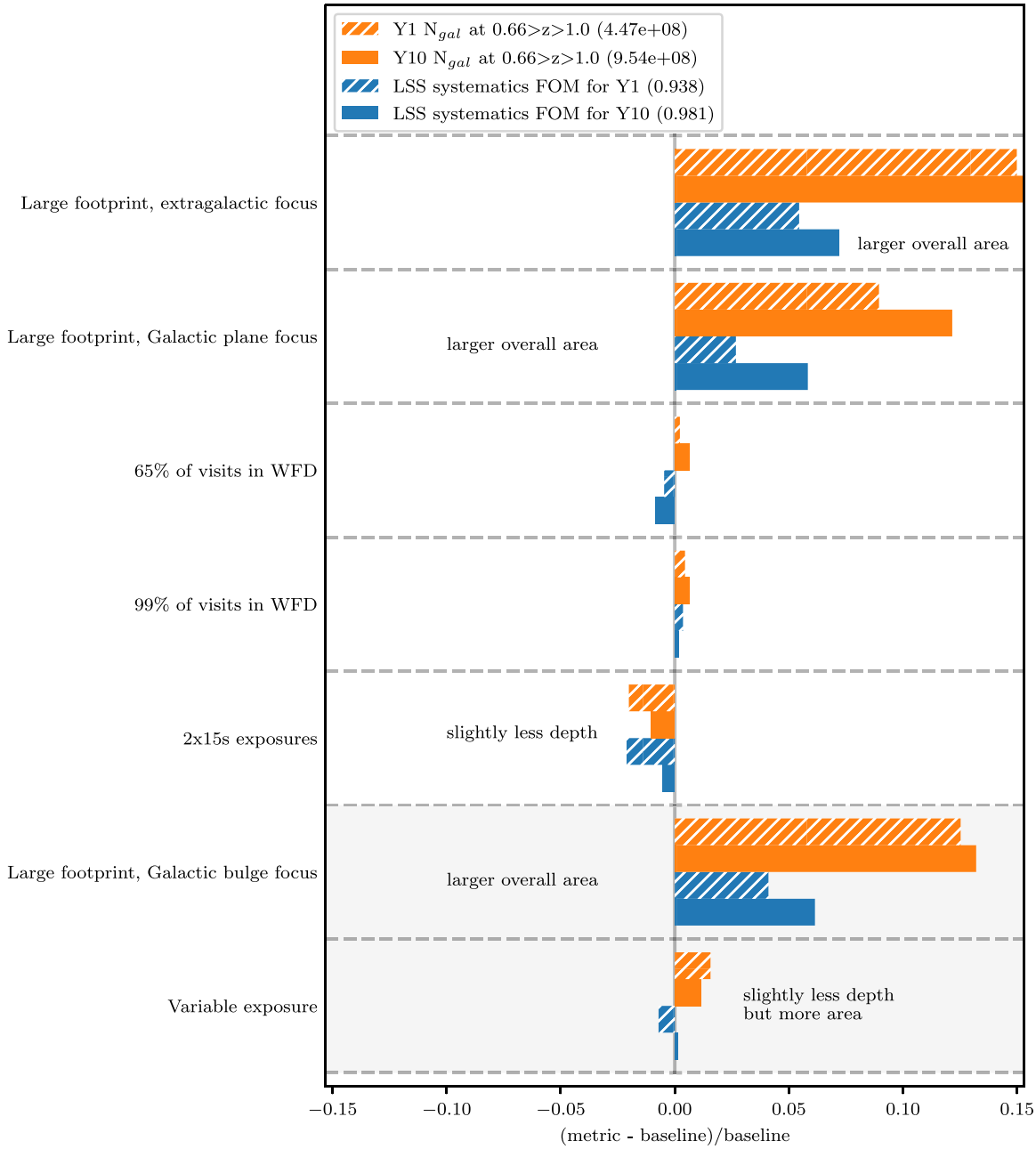

Figure 4. LSS metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. These metrics are largely driven by area, and are improved by strategies with a larger footprint. As noted in the text, all of the strategies generate galactic samples for which shot noise is small, making even 10% changes in galaxy number unimportant for cosmological constraints and placing more emphasis on the LSS systematics FoM and the WL+LSS 3 × 2 pt FoM of Figure 3.

given in Equation (9.4) in the COSEP. Specifically, we have:

$$\text{LSS FoM} \equiv \sqrt{\frac{\sum_\ell \left(\sqrt{\frac{2}{f_{\text{sky,Y10baseline}}(2\ell+1)}} C_{\ell,\text{LSS,Y10baseline}} \right)^2}{\sum_\ell \left[\left(\sqrt{\frac{2}{f_{\text{sky}}(2\ell+1)}} \left\{ C_{\ell,\text{LSS}} + \frac{1}{\bar{\eta}} \right\} \right)^2 + \sigma^2_{C_{\ell,\text{OS}}} \right]}} \tag{2}$$

where, unlike in COSEP, the numerator is fixed to the uncertainty achieved for the baseline strategy using the Y10 gold sample. The denominator contains the uncertainty for each time interval (e.g., Y1, Y10); briefly, f_{sky} is the fraction of the sky at the given time interval used for analysis, while C_ℓ denotes the angular power spectrum, and $\bar{\eta}$ is the surface number density of the galaxies in units of sr^{-1} at the given time interval; we refer the reader to COSEP for further details. The first two terms in the denominator represent the standard sample variance and shot noise, and their combination adds to the final term giving variance caused by the observing strategy (denoted "OS" in Equation (2)); the final term is calculated, as in Awan et al. (2016), as the standard deviation of $C_{\ell,\text{OS}}$ across

the *ugri* bands, to model uncertainties due to detecting galaxy catalogs in different bands. Note that this FoM approaches 1 if the observing strategy and shot-noise contributions are negligible and the statistical power matches the Y10 baseline strategy. It can be greater than 1 by Y10 for an observing strategy that covers more area than the baseline, but in that case, this improvement will duplicate that seen in the static science FoM. Improvements in survey uniformity, however, will affect this LSS FoM and not the static science FoM.

For $f_{\mathrm{fsky,baseline}}$, we use the Y10 sky coverage from the baseline for a given FBS simulation (i.e., we use `baseline_v1.5_10yrs` for v1.5 sims; `baseline_nexp2_v1.6_10yrs` for v1.6 sims that implement the 2×15 s exposure (identified by the `nexp2` tag), and `baseline_nexp1_v1.6_10yrs` for the rest; `baseline_nexp2_v1.7_10yrs` for v1.7 sims); the footprint is the extragalactic footprint (as defined in Section 3.1), designed to achieve the target Y10 gold sample of galaxies.

4.3. General Conclusions from Weak Lensing and Large-scale Structure

For both WL and LSS, statistical constraining power and observational systematics are both impacted by choices in observing strategy. As seen in Figure 3, changes in observing strategy that lead to more visits in the *r*, *i*, or *z* bands are preferred for the WL systematics metric. The 3×2 pt FoM benefits from any of the following: larger survey area at low dust extinction, greater median depth, or improved photometric redshifts. While the latter two both favor depth versus area, within the variations available from simulated surveys, the 3×2 pt FoM shows a greater improvement for simulations that maximize the area at low dust extinction. The LSS metrics in Figure 4 follow the general trend of the 3×2 pt DETF FoM in favoring a larger effective survey area despite the corresponding modest loss of median depth. The LSS FoM metric prefers both increased area and greater survey uniformity; the latter responds favorably to nightly translational dithers (as shown in Awan et al. 2016; COSEP), which have now been implemented as a default in the FBS simulations. On a higher level, we find that the statistical power for combined WL and LSS prefers more area, as do the observational systematics for LSS, while observational systematics for WL prefer more visits. In the end, for the majority of the static science metrics explored in this section, the gain from larger area is greater than that from more visits.

To illustrate the tension specifically using Figure 3, for example, the cases with bluer filter distribution and 65% of visits in the WFD survey both reduce exposure in the extragalactic area (or in bands that are used for WL shear estimation) resulting in worse performance for both the 3×2 pt FoM and the systematics metric; while the opposite is true for the strategy with 99% of visits in the WFD survey. For "large footprint, Galactic bulge focus" and "large footprint, Galactic plane focus," we see a trade-off between the two metrics, due to the fact that these simulations generally increase the area of the survey while decreasing the average number of visits in this area. The FBS simulation "Large footprint, extragalactic focus" is beneficial to the 3×2 pt FoM due to the increase in area without harming the WL systematics metric due to reducing the area that is effectively ignored by the metric.

5. Transient Science

5.1. Supernovae

As of today, the Hubble diagram of SNe Ia contains of the order of $\mathcal{O}(10^3)$ SNe (Betoule et al. 2014; Scolnic et al. 2018b). LSST will discover an unprecedented number of SNe Ia —$\mathcal{O}(10^6)$. The key requirements to turn a significant fraction of these discoveries ($\sim$10%) into distance indicators useful for cosmology are (1) a regular sampling of the SNe Ia light curve in several rest-frame bands, (2) a relative photometric calibration (band-to-band) at the 10^{-3} level, (3) a good understanding of the SNe Ia astrophysical environment, (4) a good estimate of the survey selection function, and (5) a precise measurement of the redshift host galaxy based on LSST photo-*z* estimators. The first point is crucial. It determines how well we can extract the light-curve observables used by current and future standardization techniques (stretch, rest-frame color(s), rise time). It also determines how well photometric identification techniques are going to perform, as live spectroscopic follow-up will only be possible for 10% of SNe Ia.

The average quality of SNe Ia light curves depends primarily on the observing strategy through five key facets: (1) a high observing cadence (2–3 days between visits) delivers well-sampled light curves, which is key to distance determination and photometric identification; (2) a regular cadence allows for minimizing the number of large gaps (>10 days) between visits, which degrades the determination of luminosity distances, and potentially results in rejecting large batches of light curves of poor quality; (3) a filter allocation ensuring the use at least three bands (rest frame) to select high-quality SNe; (4) the season length (the duration a given field is observed between annual Sun constraints) determines the number of SNe Ia with observations before and after peak; due to time dilation, maximizing season length is particularly important in the DDFs; and (5) finally, the integrated S/N over the SNe Ia full light curve determines the contribution of measurement noise to the distance measurement. It is a function of the visit depth and the number of visits in a given band.

All of the studies presented in this section rely on light-curve simulations of SNe Ia. We have used the SALT2 model (Guy et al. 2007, 2010) where an SN Ia is described by five parameters: x_0, the normalization of the spectral energy distribution (SED) sequence; x_1, the stretch; c, the color; T_0, the day of maximum luminosity; and z, the redshift. The time-distribution and the photometric errors of the light-curve points are estimated from observing conditions (cadence, 5σ-depth, season length) given by the scheduler. We consider two types of SNe Ia defined by (x_1, c) parameters to estimate the metrics: (intrinsically) faint SNe, defined by $(x_1, c) = (-2.0, 0.2)$, and medium SNe, defined by $(x_1, c) = (0.0, 0.0)$. (z_{faint}, $N_{z<z_{\mathrm{faint}}}$) gives an assessment of the depth and size of the redshift limited sample (i.e., the sample of SNe usable for cosmology) with the selection function having minimal dependence on understanding the noise properties. (z_{med}, $N_{z<z_{\mathrm{med}}}$) gives an assessment of the depth and size of the sample of SNe Ia with precise distances. We will get higher statistics with the medium sample, but also need a better understanding of noise to determine the selection function.

All of the metrics described below are estimated from a sample of well-measured SNe Ia that passed the following light-curve requirements: visits with S/N > 10 in at least three bands; five visits before and 10 visits after peak, within [−10;

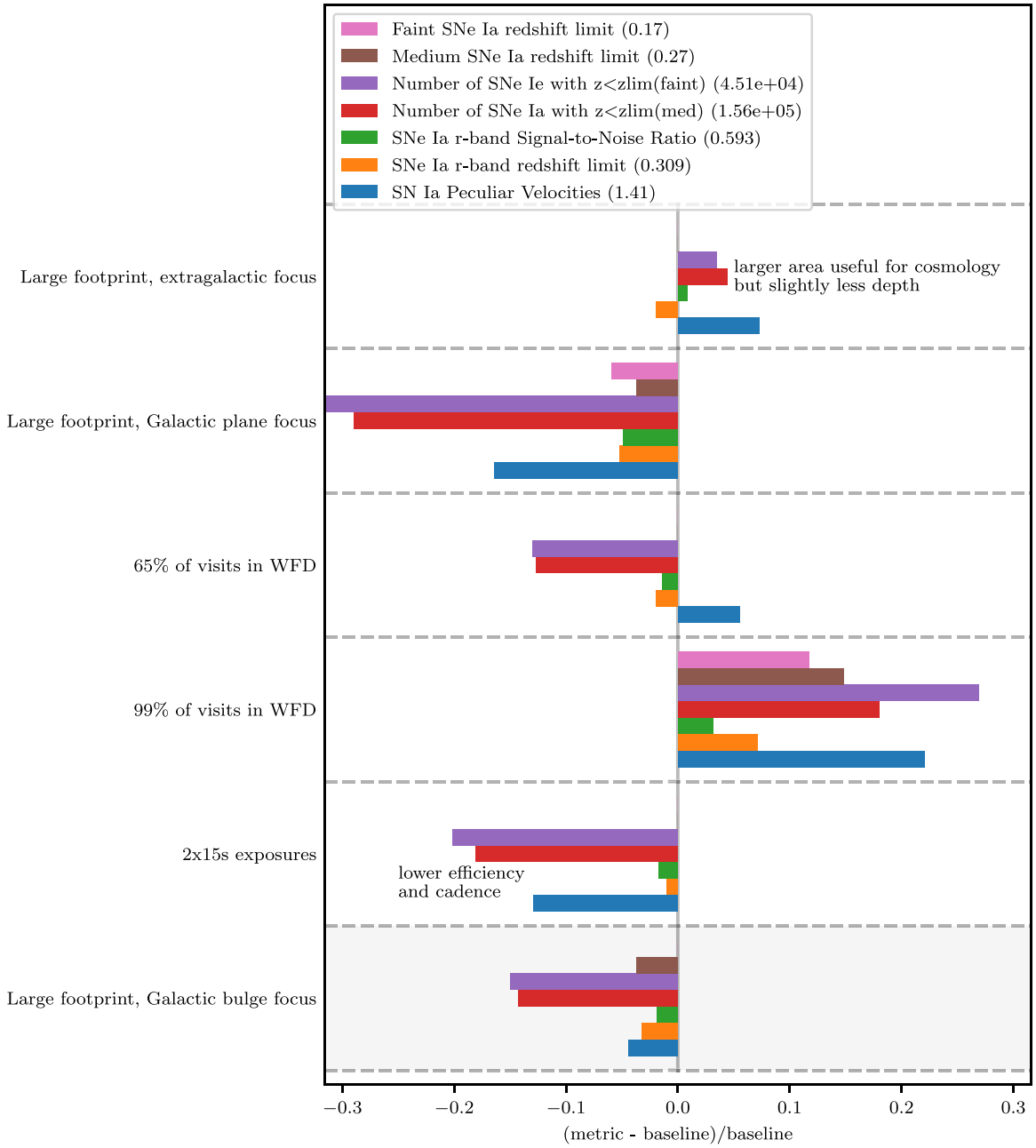

Figure 5. SN metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. As described in Section 5.1.1, $z_{\rm lim}$ corresponds to the redshift beyond which SNe no longer pass light-curve requirements. "med" and "faint" refer to a sample of typical and faint SNe, respectively. SN metrics in general are decreased by any loss in depth (and hence cadence) in the extragalactic part of the WFD footprint.

+30] days (rest frame); $\sigma_C < 0.04$ where σ_C is the SALT2 color uncertainty; all observations satisfying 380 nm $< \bar{\lambda}_{\rm obs}/(1+z) <$ 700 nm.

For the SNe Ia probe, we introduce seven new metrics. A summary of the results from the metrics can be seen in Figure 5.

1. Faint SNe Ia redshift limit—Redshift limit corresponding to a complete SNe Ia sample ($z_{\rm lim}^{\rm faint}$) (Section 5.1.1)
2. Medium SNe Ia redshift limit—Redshift limit corresponding to a complete SNe Ia sample ($z_{\rm lim}^{\rm medium}$) (Section 5.1.1)
3. Number of SNe Ia with $z \leqslant$ zlim(faint)—Number of well-sampled SNe Ia with $z \leqslant z_{\rm lim}^{\rm faint}$ (Section 5.1.1)
4. Number of SNe Ia with $z \leqslant$ zlim(med)—Number of well-sampled SNe Ia with $z \leqslant z_{\rm lim}^{\rm medium}$ (Section 5.1.1)
5. SNe Ia r-band S/N—Fraction of faint SNe Ia with an r-band S/N higher than a reference S/N corresponding to a regular cadence (Section 5.1.2).
6. SNe Ia r-band redshift limit—r-band redshift limit of faint SNe Ia (Section 5.1.2)
7. Peculiar velocities—SNe Ia host galaxy velocities (Section 5.1.3).

5.1.1. Number of Well-measured Type Ia Supernovae/Survey Completeness

We use, as our primary metric, the size and depth of a subset of well-sampled SNe Ia using the redshift limit $z_{\rm lim}$ and the number of well-sampled SNe $N_{z<z_{\rm lim}}$ below this redshift. $z_{\rm lim}$ corresponds to the redshift beyond which SNe no longer pass light-curve requirements.

The WFD footprint is quite large (at least 18,000 deg^2) forbidding the use of full (time-consuming) light-curve simulations. We have opted for a slightly different approach. The celestial sphere is pixellized in HEALPIX superpixels ($N_{\rm side} = 64$, which corresponds to 0.8 deg^2 per pixel). The directions (i.e., (R.A., decl.) positions)/healpixel affected by a Galactic extinction $E(B-V)$ larger than 0.25 are masked (to minimize reddening effects) and not included in our assessment. We consider only the *griz* observations that are the ones that matter to derive SN luminosity distances.

We process observing strategies using a simple model of the LSST focal plane and estimate:

$$z_{\rm lim} = \max\,(z|{\rm LC}(z)\ \text{fulfill requirements}) \tag{3}$$

$$N_{z<z_{\rm lim}} = \delta\Omega_{\rm pix}\int_0^{z_{\rm lim}} \frac{\Delta T_{\rm step}}{1+z}\,\mathcal{R}(z)\,dV(z) \tag{4}$$

where $\delta\Omega_{\rm pix}$ is the solid angle subtended by one pixel; dV is the (differential) comoving volume; $\Delta T_{\rm step}$ is the time interval for SNe simulations (in observer-frame days)—that is, only SNe with a peak luminosity during this time range are simulated; and $\mathcal{R}(z)$ is the SN Ia volumetric rate (Perrett et al. 2012). We also compute the average cadence (in day^{-1}), i.e., the number of g, r, i, or z visits in a fiducial rest-frame interval. The quantities above are determined for each pixel and each night (identified by its Modified Julian Date) and may be used to build full sky maps giving, as a function of the position on the sky (1) the density of SNe, (2) the median maximum redshift (over the observed area), and (3) the median cadence.

This metric is the most precise to assess observing strategies, but also the most intricate to implement. A lot of effort has been put to design algorithms combining speed, reliability, and accuracy. The codebase is accessible through Github[47] in the Metric Analysis Framework.

We note that while we expect the quality cuts described here will ensure accurate classification of SNe Ia and separate them from other classes of transients, we do not yet have a transient metric to ensure this is the case and leave this important step to future work (see Section 7.2 for a detailed discussion).

5.1.2. S/N$_{rate}$ and Redshift Limit

While the metric in Section 5.1.1 is our most accurate metric, we developed two proxy metrics that have also been incorporated in MAF where straightforward, fast, and easy-to-run metrics are preferred. These two metrics are quite simple (they do not require the use of a light-curve fitter) and just need templates of SN light curves as input. They are estimated for each band, thus providing tools for further comparison of observing strategy performance. They are sensitive to two key points of observing strategies: the median cadence, and the internight gap[48] variations. The codebase is accessible through Github.[49]

These two metrics rely on the S/N of the light curves (per band), which may be written as:

$$\mathrm{S}/\mathrm{N}_b = \sqrt{\sum_i \left(\frac{f_i^b}{\sigma_i^b}\right)^2} \tag{5}$$

where b is the band, f^b are the fluxes, and σ^b the flux errors (summation over light-curve points). In the background-dominated regime, flux errors may be expressed as a function of the 5σ limiting flux of each visit $f_{|5}$. We may rewrite Equation (5) by defining δ_b as the number of observations per bin in time ΔT and per band b:

$$f_{|5}\,\langle\delta_b\rangle^{-1/2} = \frac{5\sqrt{\Delta T}\sqrt{\sum_i (f_i^b)^2}}{\mathrm{S}/\mathrm{N}_b}. \tag{6}$$

We describe below two metrics that may be extracted from Equations (5) and (6): the S/N$_{\rm rate}$ and the redshift limit metrics.

S/N$_b$ (Equation (5)) is the result of the combination of observing strategy features (5σ depth and cadence) and SNe parameters. We fix some of these parameters (we considered faint SNe with $z = 0.3$, where the sample is not affected by the Malmquist bias) so as to estimate S/N$_b(t)$ for an SN with $T_0 = t - 10$. We also evaluate $\mathrm{S}/\mathrm{N}_b^{\rm regular}(t)$ using the same SN parameters but opting for median values for 5σ depth and cadence. The S/N$_{\rm rate}$ metric is then defined as the fraction of time (in a season) when the requirement $\mathrm{S}/\mathrm{N}_b(t) \geqslant \mathrm{S}/\mathrm{N}_b^{\rm regular}(t)$ is fulfilled.

The two above-mentioned contributions to S/N$_b$ are clearly visible in Equation (6), with observing conditions on the left side (5σ limiting flux times cadence), and flux (SNe properties) on the right side. We fix some of the SN parameters (faint SNe with $T_0 = 0$), and we use median values of 5σ depth and cadences to estimate redshift values defining the second metric, dubbed as $z_{\rm lim}$. We used Equation (6) with the following S/N (ANDed) requirements: $\mathrm{S}/\mathrm{N}_g > 30$, $\mathrm{S}/\mathrm{N}_r > 40$, $\mathrm{S}/\mathrm{N}_i > 30$, and $\mathrm{S}/\mathrm{N}_z > 20$. Combining these selections is equivalent to requesting $\sigma_c \leqslant 0.04$ and will ensure observation of well-measured SNe Ia.

5.1.3. Peculiar Velocities

The goal of the peculiar velocities metric is to study modified gravity through its effects on the overdensities and velocities of SN Ia host galaxies. Gravitational models are efficiently parameterized by the growth index, γ, which influences the linear growth rate as $f = \Omega_M^\gamma$ where $\Omega_M = \frac{\Omega_{M_0}}{\Omega_{M_0} + (1-\Omega_{M_0})a^3}$ and Ω_{M_0} is the mass density today. The parameter dependence enters through fD, where D is the spatially independent "growth factor" in the linear evolution of density perturbations, and $f \equiv \frac{d\ln D}{d\ln a}$ is the linear growth rate where a is the scale factor (Hui & Greene 2006; Davis et al. 2011). Two surveys with the same fractional precision in fD will have different precisions in γ, with the one at lower redshift providing the

[47] https://github.com/LSST-nonproject/sims_maf_contrib/blob/master/science/Transients/SN_NSN_zlim.ipynb

[48] The internight gap is the number of nights between two subsequent nights of observation.

[49] https://github.com/LSST-nonproject/sims_maf_contrib/blob/master/science/Transients/SNSNR.ipynb

tighter constraint. We thus use the uncertainty in the growth index, σ_γ, as the peculiar velocity metric.

For a parameterized survey and redshift-independent γ, σ_γ is bounded in Kim et al. (2019) by using the Fisher matrix

$$F_{ij} = \frac{\Omega}{8\pi^2}\int_{r_{\rm min}}^{r_{\rm max}}\int_{k_{\rm min}}^{k_{\rm max}}\int_{-1}^{1} r^2k^2\,{\rm Tr} \times\left[C^{-1}\frac{\partial C}{\partial\lambda_i}C^{-1}\frac{\partial C}{\partial\lambda_j}\right]d\mu\,dk\,dr \quad (7)$$

where Ω is the sky coverage of the survey, $r_{\rm max}$ ($r_{\rm min}$) are the comoving distances corresponding to the upper (lower) redshift limits of each redshift bin, and we set $k_{\rm max} = 0.2h$ Mpc^{-1} and $k_{\rm min} = 2\pi/r_{\rm max}$. μ is the cosine of the angle ϕ between the k-vector and the observer's line of sight. The covariance matrix C is defined by:

$$C(k,\,\mu) = \begin{bmatrix} P_{\delta\delta}(k,\,\mu) + \frac{1}{n} & P_{v\delta}(k,\,\mu) \\ P_{v\delta}(k,\,\mu) & P_{vv}(k,\,\mu) + \frac{\sigma^2}{n} \end{bmatrix} \quad (8)$$

and the parameters considered are $\lambda \in \{\gamma,\, bD,\, \Omega_{M_0}\}$.

The SNe Ia host-galaxy radial peculiar velocity power spectrum is $P_{vv} \propto (fD\mu)^2$, the count overdensity power spectrum is $P_{\delta\delta} \propto (bD + fD\mu^2)^2$, the overdensity-velocity cross-correlation is $P_{vg} \propto (bD + fD\mu^2)fD$, where b is the galaxy bias and $\mu \equiv \cos(\hat{k}\cdot\hat{r})$ where $\hat{r}$ is the direction of the line of sight, and n is $\epsilon\phi t$ where ϵ is the sample selection efficiency, ϕ is the observer-frame SN Ia rate, and t is the duration of the survey. While the bD term does contain information on γ, its constraining power is not used here. The variance in γ is $(F^{-1})_{\gamma\gamma}$. Our FoM is the inverse variance, so that a larger value corresponds to a more precise measurement and hence a better survey strategy.

The parameters in Equation (7) that are primarily affected by the survey strategy are the survey solid angle Ω and the number density n of well-measured SNe Ia. The other parameters related to the follow-up strategy of these SN discoveries are the survey depth $r_{\rm max}$ and the intrinsic velocity dispersion σ, which is related to the intrinsic magnitude dispersion of well-measured SNe Ia. The estimate of σ_γ is sensitive to both the sample variance P_{vv} and shot noise $\frac{\sigma^2}{n}$ in the range of proposed surveys, meaning that its accurate determination cannot be taken in either the sample- or shot-noise limit. A follow-up strategy must also be specified for the calculation of σ_γ since Rubin/LSST will not generate all of the information needed for this measurement, e.g., redshift, SN classification. Here, we adopt a maximum survey redshift of $z = 0.2$ and follow-up that gives 0.08 mag magnitude dispersion per SN. The minimum redshift is $z = 0.01$, number densities are based on 65% of the SNe Ia rates of Dilday et al. (2010), and $k_{\rm max} = 0.1h$ Mpc^{-1}, $b = 1.2$, ΛCDM cosmology with $\Omega_{M_0} = 0.3$, and overdensity power spectra for the given cosmology as calculated by CAMB (Lewis & Bridle 2002).

The code used for the calculations are available in Github.[50]

5.1.4. General Conclusions from Supernovae

Collecting a *large* sample of *well-measured* SNe Ia is a prerequisite to measure cosmological parameters with high accuracy. Our analysis (Sections 5.1.1–5.1.3 and Figure 5) has shown that the key parameter to achieve this goal is the effective cadence delivered by the survey. For the WFD survey, a regular cadence in the g, r, and i bands is essential to (1) secure a high-efficiency photometric identification of the detected SNe Ia, and (2) secure precise standardized SNe Ia distances, by optimizing the integrated S/N along the SNe Ia light curves. Gaps of more than ∼10 days in the cadence have a harmful impact on the size and depth of the SNe Ia sample. The cadence of observations is by far the most important parameter for SNe Ia science (see Figure 14), before observing conditions: on the basis of studies conducted, it is preferable to have pointings with suboptimal observing conditions rather than no observation at all.

Three main sources of gaps have been identified: telescope down time (clouds, maintenance), filter allocation, and scanning strategy (i.e., the criteria used to move from one pointing to another). While we are aware that it is difficult to minimize the impact of down time, there is still room for improvement on filter allocation and scanning strategy. Significant efforts have been made to make sure that nightly revisits of the same field are performed in different bands. Relaxing the veto on bluer bands around full moon, or increasing the density of visits (i.e., the number of visits per square degree) during a night of observation (by decreasing the observed area for instance) will help to achieve an optimal cadence for SNe of 2–3 days in the g, r, and i bands.

5.2. Strong Lensing

The Hubble constant H_0 is one of the key parameters to describe the universe. Current observations of the cosmic microwave background assuming a flat ΛCDM cosmology and the standard model of particle physics yield $H_0 = 67.4 \pm 0.5$ km s^{-1} Mpc^{-1} (Planck Collaboration 2020), which is in tension with $H_0 = 73.2 \pm 1.3$ km s^{-1} Mpc^{-1} from the local Cepheid distance ladder (Riess et al. 2021). To probe the $>4\sigma$ tension between the cosmic microwave background and the Cepheid distance ladder further, other independent methods are needed.

One such method is lensing time-delay cosmography, which can determine H_0 in a single step. The basic idea is to measure the time delays between multiple images of a strongly lensed variable source (Refsdal 1964). This time delay, in combination with mass profile reconstruction of the lens and line-of-sight mass structure, yields directly a "time-delay distance" that is inversely proportional to the Hubble constant ($t \propto D_{\Delta t} \propto H_0^{-1}$). Applying this method to six lensed quasar systems and using well-motivated models for the lens mass distributions, the H0LiCOW collaboration (Suyu et al. 2017) together with the COSMOGRAIL collaboration (e.g., Eigenbrod et al. 2005; Tewes et al. 2013; Courbin et al. 2017) measured $H_0 = 73.3^{+1.7}_{-1.8}$ km s^{-1} Mpc^{-1} (Wong et al. 2020) in flat ΛCDM, which is in agreement with the local distance ladder but higher than CMB measurements. Another promising approach goes back to the initial idea of Refsdal (1964) using lensed supernovae (LSNe) instead of quasars for time-delay cosmography (e.g., Grillo et al. 2020; Mörtsell et al. 2020; Suyu et al. 2020). In terms of discovering strong lens systems from the static LSST images for cosmological studies, having g-band observations with comparable seeing as in the r and i bands would facilitate the detection of strong lens systems (Verma et al. 2019).

[50] https://github.com/LSSTDESC/SNPeculiarVelocity/blob/master/doc/src/partials.py

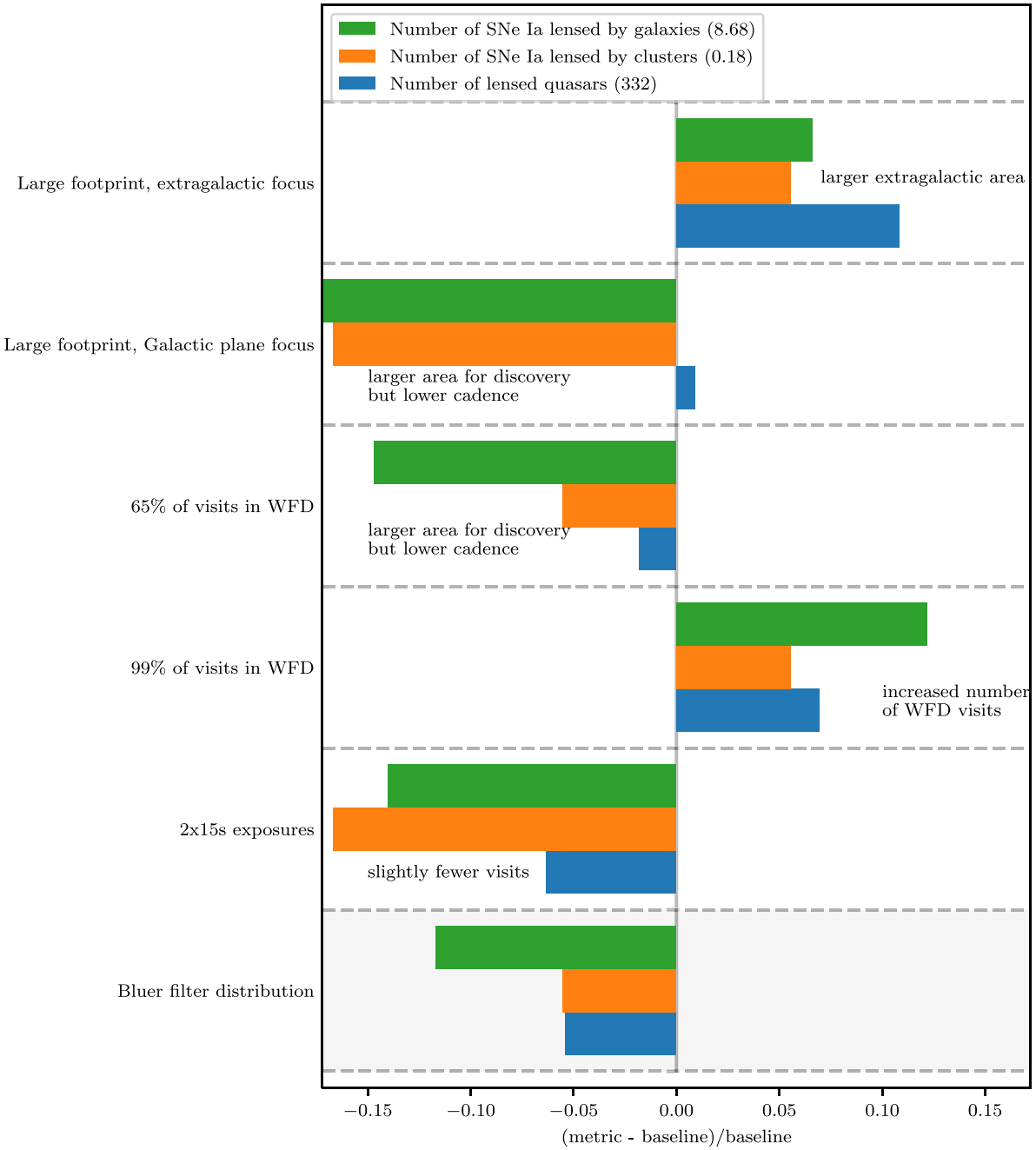

Figure 6. Strong lensing metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. We find that a higher cadence (increased number of visits) is preferred for strong lensing in terms of maximizing the number of lens systems. Further, a larger survey area helps, but only if the cadence is not impacted significantly. The lensed quasar case suffers less from a lower cadence in comparison to SNe because SNe will fade away in contrast to quasars and therefore the time of detection is more important. For the case of 99% of visits in WFD, the slightly larger long cumulative season length also helps for the science case of SNe and quasars lensed by galaxies. Furthermore, given the redshift distribution of the expected cluster-lensed SNe Ia, the yields are also sensitive to the choice of filters, so the bluer filter distribution negatively affects the number of expected cluster-lensed SNe Ia.

In this section, we investigate the prospects of using LSST for measuring time delays of both lensed SNe and lensed quasars. In particular, we focus on the number of lens systems that we would detect for the various observing strategies as our metrics. From the investigation of LSNe by galaxies, we define a metric for the number of LSNe Ia with good time-delay measurement. For lensed quasars, we have additional metrics defining how well we can measure the time-delay distances.

For the strong lensing probe, we introduce three new metrics. A summary of the results from these metrics can be seen in Figure 6.

1. Number of SNe Ia lensed by galaxies—Number of SNe Ia strongly lensed by galaxies with accurate and precise time delays between the multiple SN images (Section 5.2.1);

2. Number of SNe Ia lensed by clusters—Number of strongly lensed SNe Ia in the multiply imaged galaxies behind well-studied galactic clusters (Section 5.2.2);
3. Number of lensed quasars—Number of strongly lensed quasars with accurate and precise time delays between the multiple quasar images (Section 5.2.3).

5.2.1. Number of Supernovae Lensed by Galaxies

For constraining cosmological parameters with SNe lensed by galaxies as well as possible, ideally we would like to maximize the number of accurate and precise time-delay distance measurements. Currently there are only three known lensed SN systems with resolved multiple images, namely SN "Refsdal" (Kelly et al. 2016a, 2016b), iPTF16geu (Goobar et al. 2017), and AT2016jka (Rodney et al. 2021), but LSST will play a key role in detecting many more LSNe (Oguri & Marshall 2010; Goldstein et al. 2018; Wojtak et al. 2019). A measurement of a time-delay distance from a strongly lensed SN system requires (1) the detection of the system, (2) the measurement of time delays between the multiple SN images from their observed light curves, and (3) the lens mass modeling of the system to infer the distance from the time delays. After the LSN fades away, we can get the lens mass modeling from an observation of the multiple images of the SN host galaxy and the lens galaxy, to avoid the bright SN images outshining the lensing galaxy. Therefore, (3) does not depend on LSST's observing strategies; however, it affects both (1) and (2), and the uncertainties in the time delays from (2) enter directly into the uncertainties on the time-delay distances. Therefore, we use as a metric the number of lensed SNe systems that could yield time-delay measurements with precision better than 5% and accuracy better than 1%, in order to achieve H_0 measurement that has better than 1% accuracy from a sample of lensed SNe. We refer to time delays that satisfy these requirements as having "good" delays.

Huber et al. (2019) presented a detailed study about the number of lensed SNe with "good" delay measurement, by simulating realistic mock LSNe Ia for 20 different LSST observing strategies. The results from Huber et al. (2019) showed that using only LSST data for the delay measurement is not ideal and LSST should be rather used as discovery machine for LSNe where the delay measurement should be conducted from follow-up observations with a more rapid cadence than LSST. Furthermore, they find that long cumulative seasonal lengths (sum of each season length over the 10 yr survey, where an "observing season" refers to the duration in a year when a target is observable at night and observed by LSST) and a more frequent sampling are important to increase the number of LSNe Ia with well-measured time delays, but a pure rolling cadence is clearly disfavored, because their shortened cumulative season lengths (only five instead of 10 seasons for two decl. bands) lead to overall a more negative impact on the number of LSNe Ia with delays, compared to the gain from the more rapid sampling frequency.

To evaluate a much larger sample of observing strategies, we have defined a metric based on the investigations of Huber et al. (2019). The number of LSNe Ia with well-measured time delays using LSST and follow-up observations for a given observing strategy can be approximated as

$$N_{\mathrm{LSNeIa,good\ delay}} \approx 45.7 \frac{\Omega_{\mathrm{WFD}}}{20{,}000\ \mathrm{deg}^2} \frac{t_{\mathrm{csl}}}{2.5\ \mathrm{yr}} \cdot \frac{1}{2.15} \exp(-0.37 t_{\mathrm{gap}}). \quad (9)$$

The first part of the metric (separated by a dot from the second part) is the rescaling of the value predicted in OM10 (Oguri & Marshall 2010) by taking into account the survey area of the WFD (Ω_{WFD}) and the mean of the cumulative season length (t_{csl}; summed over all season lengths) for a given observing strategy. The second part contains a fit based on the numbers of LSNe Ia with well-measured time delay, presented in Huber et al. (2019), in comparison with the total number of LSNe Ia, which will be detected (first part of Equation (9)). The fit function depends on the median internight gap between any filter, t_{gap}, and is measured in days, which is an important parameter because we assume a detection of the LSNe Ia in Huber et al. (2019) after the third data point exceeds the 5σ point-source depth in any filter. The internight gap (t_{gap}), survey area Ω_{WFD}, and cumulative season length t_{csl} can be calculated via MAF[51] and a Python script[52] where we only take observations into account that have a 5σ point-source depth greater than 22.7, 24.1, 23.7, 23.1, 22.2, and 21.4 for the filters u, g, r, i, z, and y, respectively. These cuts are motivated by Huber et al. (2019) and are important to restrict visits to the WFD, where the metric from Equation (9) is valid. The results are summarized in Figure 6, where we see that in principle a larger area helps, but only if the cadence is not reduced significantly. Furthermore, we find that less time on the WFD in comparison to the baseline cadence is clearly rejected. In agreement with that, Figure 6 shows that more time on the WFD improves the number of LSNe Ia. In addition, we find that 1×30 s exposures are favored over 2×15 s exposures, and we see that a filter redistribution to bluer bands hurts our science case, which is not surprising given that the median source redshift of the expected LSN Ia sample is around 0.77 (Huber et al. 2021), and therefore LSN Ia are faint in the blue bands (Huber et al. 2022). Our conclusions including results from Huber et al. (2019) are summarized in Section 5.2.4.

5.2.2. Number of Supernovae Lensed by Galaxy Clusters

Here, we focus on prospects of observing LSNe, which are strongly lensed by galaxy clusters that have well-studied lens models. High-z galaxies that appear as multiple images in the cluster field can host SN explosions. The first discovery of this kind was SN Refsdal, which was classified as a core-collapse (CC) explosion (Kelly et al. 2015, 2016b). Several teams predicted the reappearance of SN Refsdal almost a year later, which allowed us to test their lens models (e.g., Grillo et al. 2016; Kelly et al. 2016b). By measuring the time delays of SN Refsdal and having a high-quality strong lensing model of the galaxy cluster, it was shown that it is possible to measure H_0 with 6% total uncertainty (Grillo et al. 2018, 2020). Dedicated ground-based searches for lensed SNe behind galaxy clusters have been performed using near-infrared instruments at the Very Large Telescope (Goobar et al. 2009; Stanishev et al. 2009; Petrushevska et al. 2016, 2018a). Most notably, they reported the discovery of one of the most distant CC SNe ever found, at redshift $z = 1.703$ with a lensing magnification factor

[51] https://me.lsst.eu/gris/
[52] https://github.com/shuber891/LSST-metric-for-LSNe-Ia/

Table 3
The Galaxy Clusters Considered in This Work

Cluster	$N_{\rm systems}$	$N_{\rm images}$	$z_{\rm min} - z_{\rm max}$
A1689	18	51	1.15–3.4
A370	21	67	0.73–5.75
A2744	12	40	1.03–3.98
AS1063	14	42	1.03–3.71
MACS J0416.1-2403	23	68	1.01–3.87
Total	88	268	

Note. The number of unique galaxies behind the cluster is given in Column 2, and the number of their multiple images of these galaxies is given in Column 3. The redshift range of these galaxies is given in Column 4.

of 4.3 ± 0.3 (Amanullah et al. 2011). Furthermore, thanks to the power of the lensing cluster, it was possible to estimate the volumetric CC SN rates for $0.4 \leqslant z < 2.9$, and compare them with the predictions from cosmic star formation history (Petrushevska et al. 2016). Knowing the absolute brightness of SNe Ia permits the estimation of the absolute magnification of SNe Ia, therefore breaking the so-called mass-sheet degeneracy of gravitational lenses (Holz 2001). Thus, LSNe Ia could be used to put constraints on the lensing potential, if the background cosmology is assumed to be known (see, e.g., Nordin et al. 2014; Patel et al. 2014; Rodney et al. 2015).

As a metric, we use the expected number of LSNe in the selected galaxy cluster fields in 10 yr of LSST. For the details regarding the methods in this section, we refer to Petrushevska (2020) based on the work in Petrushevska et al. (2016, 2018a, 2018b). Here, we present a short summary. We consider the six Hubble Frontier Fields clusters (Lotz et al. 2017) and A1689, given in Table 3. These clusters have been extensively studied, and given the good-quality data, well-constrained magnification maps and time delays can be obtained from the lensing models (Petrushevska et al. 2016, 2018a, 2018b). We consider the multiply imaged galaxies in the cluster fields that have a spectroscopic redshift. Given the redshift range of the multiply imaged galaxies considered here (see Table 3), the most important bands are i, z, and y (see Figure 2 in Petrushevska 2020). The observability of an SN in the multiply imaged galaxies is sensitive to the redshift, star formation rate, and magnification of the galaxy, but also on the observing strategy parameters such as depth, separation between the two consecutive observations, and the filter. The expected number of LSNe Ia in the five cluster fields is relatively low, mainly for two reasons. First, we have only considered 268 images of the background galaxies in the five cluster fields, which are listed in Table 3. Second, given the redshift range of $0.73 < z < 5.75$ (see Table 3), the ground-based Rubin Observatory filter set is not optimal for detecting SNe in these galaxies. However, thanks to the magnification from the galaxy clusters, LSST is sensitive to detecting LSNe Is to very high redshifts ($0.73 < z < 1.95$). We note that the resulting expected number of LSNe in the selected galaxy cluster fields is a lower limit, since we have only considered few clusters and the multiply imaged galaxies with spectroscopic redshift. Beyond the clusters that we have considered here, LSST will observe $\sim$70 galaxy clusters with Einstein radii larger than $\theta_{\rm E} > 20''$ that have $\sim$1000 multiply imaged background galaxies (The LSST Science Book). As the expectations of strongly lensed SNe in cluster fields depend on several factors, including the star formation rate and the stellar mass of the host galaxy, it is not straightforward to make a reliable prediction, and we leave the expectation in all clusters visible to LSST for a future study.

The conclusions of all of the metrics presented in the Section 5.2 are presented in Section 5.2.4. The conclusions for the galaxy-lensed SNe and cluster-lensed SNe are similar in general (see Figure 6). Given the aforementioned dependence of the observing strategy parameters, what drives the difference in the yields is mostly the number of observations, as the depth and the mean gap between the observations of the considered clusters remain roughly the same, for the observing strategies plotted in Figure 6. Furthermore, in order to optimize the sensitivity to high-redshift SNe with multiple images in galactic cluster fields, deeper images in the reddest bands (i, z, and y) are preferred for the cluster-lensed SNe Ia. This can be obtained by coadding images from visits closely separated in time. As mentioned in the previous section, the LSST will serve to detect the LSNe, but additional follow-up by other photometric and spectroscopic instruments will be needed to securely measure the time delays.

5.2.3. Number of "Golden" Lensed Quasars

The goal of this section[53] is to evaluate the precision we can achieve in measuring time delays in strongly lensed active galactic nuclei (AGNs), and as such, the precision on the measurement of the Hubble constant from all systems with measured time delays.

Anticipating that the time-delay accuracy would depend on night-to-night cadence, season length, and campaign length (number of survey years), we carried out a large-scale simulation and measurement program that coarsely sampled these schedule properties. In Liao et al. (2015), we simulated five different light-curve data sets, each containing 1000 lenses, and presented them to the strong lensing community in a "time-delay challenge" (TDC). These five challenge "runs" differed by their schedule properties. Entries to the challenge consisted of samples of measured time delays, the quality of which the challenge team then measured via three primary diagnostic metrics: time-delay accuracy ($|A|$), time-delay precision (P), and usable sample fraction (f). The accuracy of a sample was defined to be the mean fractional offset between the estimated and true time delays within the sample. The precision of a sample was defined to be the mean reported fractional uncertainty within the sample.

Focusing on the best challenge submissions made by the community, we derived a simple power-law model for the variation of each of the time-delay accuracies, time-delay precisions, and usable sample fractions, with the schedule properties cadence (cad), season length (sea), and campaign length (camp). They are given by the following equations:

$$|A|_{\rm model} \approx 0.06\% \left(\frac{\rm cad}{3\ \rm days}\right)^{0.0} \left(\frac{\rm sea}{4\ \rm months}\right)^{-1.0} \left(\frac{\rm camp}{5\ \rm yr}\right)^{-1.1} \quad (10)$$

$$P_{\rm model} \approx 4.0\% \left(\frac{\rm cad}{3\ \rm days}\right)^{0.7} \left(\frac{\rm sea}{4\ \rm months}\right)^{-0.3} \left(\frac{\rm camp}{5\ \rm yr}\right)^{-0.6} \quad (11)$$

[53] Summarized and updated version of the Cosmology chapter of LSST Science Collaboration et al. (2017).

$$f_{\rm model} \approx 30\% \left(\frac{\rm cad}{3\ \rm days}\right)^{-0.4} \left(\frac{\rm sea}{4\ \rm months}\right)^{0.8} \left(\frac{\rm camp}{5\ \rm yr}\right)^{-0.2}. \quad (12)$$

All three of these diagnostic metrics would, in an ideal world, be optimized: this could be achieved by decreasing the night-to-night cadence (to better sample the light curves), extending the observing season length (to maximize the chances of capturing a strong variation and its echo), and extending the campaign length (to increase the number of effective time-delay measurements).

The accuracy and precision in time-delay measurements (assuming identical lens "modeling uncertainty") are roughly proportional to the statistical uncertainty on the Hubble constant. Our analysis thus consists of selecting only the sky survey area that allows time-delay measurements with accuracies of $<1\%$ and precision $<5\%$. This high accuracy and precision area can be used to define a "gold sample" of lenses. The TDC usable fraction averaged over this area gives us the approximate size of this sample: we simply rescale the 400 lenses predicted by Liao et al. (2015) by this fraction over the 30% found in TDC. Note that naturally there is a strong dependence with footprint: assuming the single-visit depth is deep enough to detect the typical lensed quasar image, the number of lenses will scale linearly with the survey area. The uncertainty on the Hubble constant will then finally scale as one over the square root of the number of lenses. While these numbers are approximate, the ratios between different observing and analysis strategies provide a useful indication of relative merit.

Our calculations are performed using the full 10 yr of LSST operations with observations in all bands contributing equally (i.e., monochromatic intrinsic variability). However, it is important to note that there is a rather large caveat: even when AGN variability can show almost a negligible difference between bands close in wavelength, the difference can be important between the bluest and reddest LSST bands. As such, the numerical values have to be considered as an optimistic lower limit, but, as mentioned before, the ratios between observing strategies are an indication of relative merit.

5.2.4. General Conclusions from Strong Lensing

For our science case, a long cumulative season length and improved sampling frequency (cadence) at a sufficient depth are important. The cumulative season length provided by the baseline cadence is sufficient, but rolling cadences are disfavored as pointed out by Huber et al. (2019), because of the reduced cumulative season length. In terms of the cadence, Huber et al. (2019) showed that revisits of the same field within a single night should be done in different filters. Further improvements could be achieved by doing single snaps instead of two snaps, as shown in Figure 6. A filter redistribution to bluer bands is clearly rejected for all strong lensing science cases, which we have presented here. We also note that increasing the overall area would naturally yield an improvement in the number of lensed SNe and quasars, but only if the cadence is not influenced much, because fewer WFD visits would be detrimental for the number of lensed SNe with accurate and precise time delays, given that SNe fade away and therefore an earlier detection plays a more important role in contrast to the stochastically varying quasars.

5.3. Kilonovae

Within the next decade, kNe will mature as a probe of cosmological physics. Gravitational-wave (GW) observatories will begin to run nearly continuously and, with expected upgrades, become ever more sensitive to the mergers of binary neutron stars (BNSs) that produce kNe. Detecting the electromagnetic (EM) counterpart to these sources, i.e., the kNe, will enable improved constraints on the accelerating expansion of the universe, via measurements of the Hubble constant (Holz & Hughes 2005; Nissanke et al. 2013), complementary to other probes of cosmic expansion (Nissanke et al. 2010; Chen et al. 2018; Mortlock et al. 2019). This is predicated on the detection of EM and GWs from the same BNS mergers. The LSST will be able to detect kNe at distances beyond the range that future GW observatories operating in the 2020s will be sensitive (Scolnic et al. 2018a; Setzer et al. 2019).

Combined with the large area and rapid cadence of LSST's WFD survey, this offers the opportunity for the LSST to detect and identify kNe that will not be detected with any other instrument. However, GW detectors do not need to point in a given direction to make a measurement of a signal, and they are sensitive to signals from the entire sky. In principle, if GW detectors are operating coincidentally to a detection of a kN by the LSST, the signal of the BNS merger will be in the GW data, but possibly below the significance threshold used to claim a merger detection. The kN detection by the LSST can then be used as prior information to reverse-trigger a search through GW data for an accompanying merger signal (Kelley et al. 2013). With this information, the population of sources with both EM and GW signals, i.e., standard sirens, used for studies of fundamental physics can be increased, and studies can be made of the underlying BNS population. This may be critical for studying the selection effects of GW-detection of standard sirens.

5.3.1. Serendipitous Kilonova Detections

For the kNe probe, we introduce three new metrics. A summary of the results from this section can be seen in Figure 7.

1. GW170817-like kNe Counts 10 yr—This metric represents the number of GW170817/AT2017gfo-like (Abbott et al. 2017) kNe that are detected according to a set of criteria over the full 10 yr survey.
2. kNe MAF Mean Counts—This metric represents the MAF implementation, which evaluates the number of GW170817/AT2017gfo-like kNe that are detected on average per region of the sky assuming a kN is always "going-off" at a fixed redshift of 0.075.
3. kN Population Counts 10 yr—This metric represents the number of kNe drawn from a population that are detected according to a set of criteria over the full 10 yr survey.

To classify a kNe as being detected, we use the criteria from Scolnic et al. (2018a) and used again by Setzer et al. (2019). These criteria are as follows:

1. Two LSST alerts separated by $\geqslant$30 minutes.
2. Observations in at least two filters with $S/N \geqslant 5$.
3. Observations with $S/N \geqslant 5$ separated by maximum of 25 days (i.e., no large gaps).

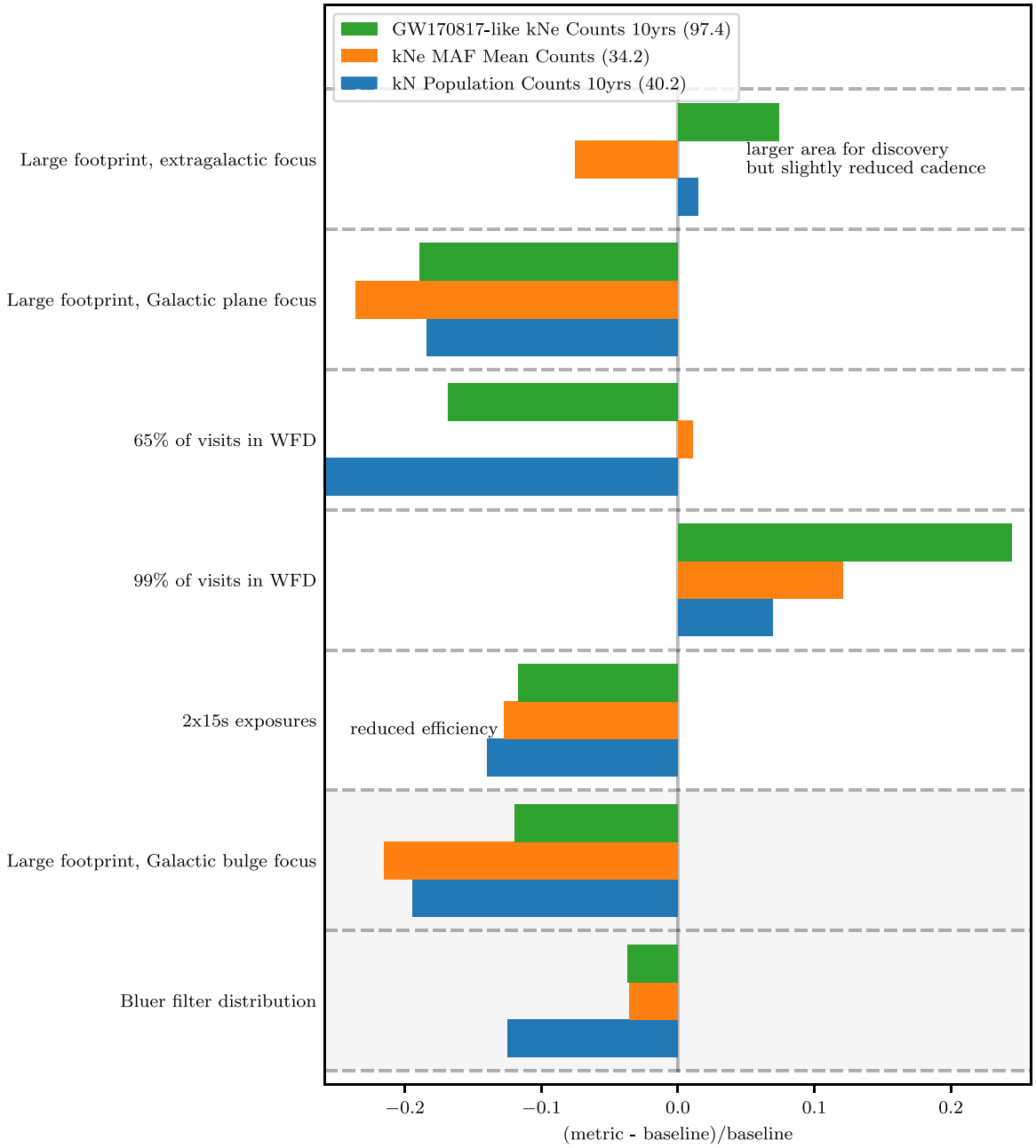

Figure 7. KN metrics as a function of selected observing strategies. Table 2 contains the exact simulation names corresponding to the short names used here. Metrics are transformed using the equations in Table 5 and are taken relative to their values at baseline in order to be directly comparable, with larger values always being better. Metric values at baseline are indicated in parentheses. Select annotations are added to highlight factors driving metric behavior. The larger area footprints tend to lower metric performance, indicating that good cadence is more important than area for detection of well-sampled KNe.

4. A minimum of one observation of the same location within 20 days before the first $S/N \geqslant 5$ observation.
5. A minimum of one observation of the same location within 20 days after the last $S/N \geqslant 5$ observation.

Several of these requirements were implemented to reject potential contaminants such as asteroids, AGNs, and superluminous SNe (Scolnic et al. 2018b; Setzer et al. 2019). However, they are insufficient to perfectly separate kNe from other transient classes. While work is ongoing to implement a transient classification metric (as discussed in Section 7.2) and we anticipate such a metric to correlate with the same quantities that impact our kNe metric, we acknowledge that this is an important step that must be studied in future work.

As was concluded by Setzer et al. (2019), survey strategies that increase the cadence of observing a region of the sky in multiple filters, e.g., obtaining the nightly pair of observations in different filters, provide greater numbers of detected kNe. We direct the reader to Setzer et al. (2019) for a full discussion of the kN models and methodology used for this similar analysis. Since that work, the default setting for new observing strategies has been to implement the nightly pairs of observations in different filters. In this analysis of additional survey strategies, we find several more features that improve the numbers of detected kNe relative to the baseline strategy.

The first feature-change that benefits our science is when the number of visits in the WFD survey is increased. Whether this

is achieved through the use of 1×30 s exposures instead of 2×15 s to increase survey efficiency or directly increasing the time-allocation of the WFD, more extragalactic observations increases the number of detected kNe. This increase in the number of observations, given a fixed survey length, effectively increases the cadence of observations for any field.

This is expected, as kNe are fast-evolving transients and their detection will be sensitive to the cadence of observations. Rolling observing strategies, which most directly increase the cadence of observations, have the potential to improve the number of kN-detections over the baseline observing strategy. However, kNe are not only fast-evolving, but they are also rare, and thus our metric is sensitive to the amount of cosmological volume, effectively sky area, that is observed. Rolling-style observing strategies fundamentally decrease the observed sky area active within an observing season by separating the sky into bands of decl. While we advocate for further exploration in this direction, it is possible that any improvement that might be expected from the increased cadence will be negated by the substantial decrease in actively observed sky area.

Conversely, increasing the total survey area with a fixed number of observations will decrease the cadence of observations. The larger extragalactic footprint simulation only marginally improves the population counts metric due to the corresponding decrease in cadence. However, reducing visits in the extragalactic part of the footprint is highly detrimental, as can be seen in Galactic plane and Galactic bulge focused simulations. We naturally see a significant improvement in all metrics if the number of visits to the WFD is artificially increased (99% of visits in WFD). We conclude increasing the footprint area of the survey is beneficial for increasing the number of detected kNe only if the number of observations per field is not significantly decreased. Furthermore, a larger extragalactic sky area can be achieved by moving the observed sky area to regions of low dust extinction. We find, additionally, that the proposed survey strategies that avoid the Milky Way by a dust or galactic-latitude cut are quite beneficial for detecting greater numbers of kNe.

Lastly, we note that while kNe are expected to be quite red transients, we do not find a redistribution of observations into redder filters to improve detections. Of the proposed distributions of observations by filter, the baseline filter distribution performs best for both the population model of kNe and the model based on GW170817 (see Section 6.4 for an analysis of filter distribution). This can be understood by considering the LSST bandpass efficiency, which is highest for the *g* band and decreases for the redder bands (LSST SRD). The best filter distribution with which to detect kNe is hence dependent on both the underlying SED of kNe and the bandpass efficiencies of LSST. It appears that the current baseline filter allocation is close to optimal, at least for the chosen population model.

While the full detection analysis is in general computationally prohibitive for hundreds of proposed survey strategies, a simplified version of this is implemented into the MAF that is used by the LSST Project and is publicly available. The MAF implementation,[54] labeled above descriptively as *kN MAF Mean Counts*, considers only GW170817-like kNe. Like other MAF transient metrics, it considers that these transients go off at a single, user-specified, cosmological redshift and occur one after another at each point on the sky. From this, observations are made of the light curves according to the chosen survey strategy, and these are checked for whether they pass certain detection criteria. Of the criteria listed above, we are limited to implementing only criteria two and three. From this we obtain a number of detected kNe; however, as this is not a proper cosmological transient distribution based on comoving volumetric rates, this metric should not be used to forecast detected counts of kNe per survey, as was done in the full analysis. In this case, we instead compute the mean number of detections, i.e., the number of detected kNe averaged over the number of fields in the analyzed survey strategy, as our metric for comparison. The use of these numbers should only be in comparison to other survey strategies. Given the limitations in the MAF implementation, the MAF metric does not exactly emulate all metric results we see from the full analysis. However, the number of outliers is small, and the overall trends are reproduced.

[54] https://github.com/LSST-nonproject/sims_maf_contrib/blob/master/mafContrib/GW170817DetMetric.py

5.3.2. General Conclusions from Kilonovae

We conclude the most important observing strategy feature for improving the number of kNe detections is still obtaining a pair of observations in different filters within a single night. Second to this, increasing the number of the observations in the extragalactic WFD survey is very beneficial. Lastly, we find observing strategies that increase the observed sky area in low dust-extinction regions, and do not substantially decrease the cadence of observations to achieve this larger area, are also preferable.

6. Discussion

Combining the insights gained from our cosmology-related metrics is not a trivial task. In an ideal world, a full DETF FoM could be calculated for each simulation allowing for the objective determination of the "optimal" observing strategy for cosmology. However, as we expect systematic effects to play a significant role in final cosmological constraints with LSST, we do not currently have a realistic enough FoM to make final decisions based on it alone. We make use of the 3×2 pt FoM described in Section 4.1.1 to summarize the impact of observing strategy on the main static science probes, but also consider separately WL systematics and transient metrics. These simpler, more interpretable metrics can assist in gaining deeper insight and making general recommendations for observing strategy. We expect metrics such as the total number of well-measured SNe and survey uniformity to correlate strongly with cosmological parameter constraining power once systematic effects are taken into account. In this section, we use select metrics to investigate and draw insights into various aspects of observing strategy.

6.1. Footprint, Area, and Depth

In general, there is a trade-off between depth and area in any survey. Figure 8 shows several key metrics as a function of effective survey area (the survey area meeting the cuts described in Section 3.1). The first thing to note is that the 3×2 pt FoM has a simple linear relationship with area, and the larger the survey, the better the 3×2 pt FoM. Photometric redshifts however tend to prioritize depth over area. There is also a trade-off between WL shear systematics, which improve as more visits are taken (essentially ending up with more depth) and the 3×2 pt FoM, which prefers a larger area. To fully

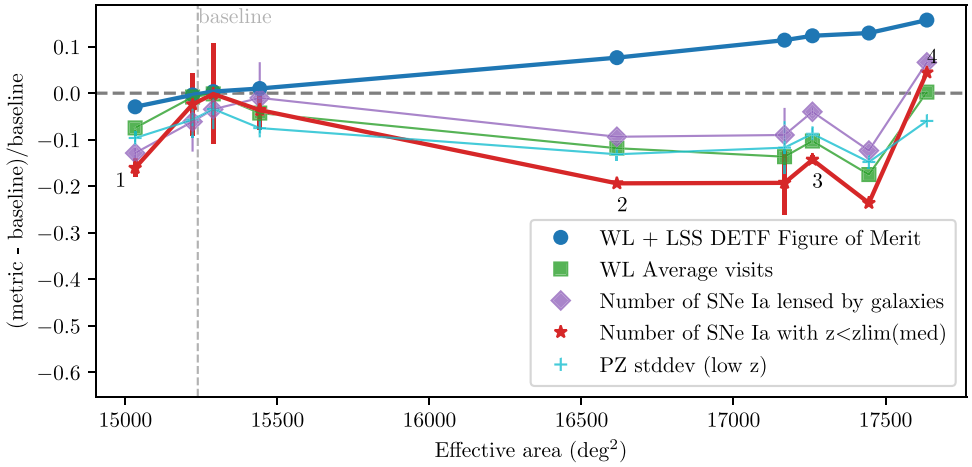

Figure 8. Selected metrics, relative to their values at baseline, as a function of effective area (i.e., the area that meets the cuts described in Section 3.1) of different observing strategies. To improve readability, points that are nearby in the x-axis are binned with only the mean and error bar plotted for that bin. It can be seen that the 3×2 pt FoM metric simply prefers more area while the situation is more complex for time-dependent metrics. The photo-z metric tends to decrease with any reduction of depth or changes in filter distribution. We have highlighted specific simulations with numbered annotations. The simulation at 1 reduces the overall area available for cosmology by reducing visits in the crucial redder bands. Simulations 2 and 3 have larger area footprints, but degrade transient metrics and photometric redshifts due to a reduced number of visits in the extragalactic part of the WFD. Simulation 4 reverses this by completely prioritizing visits in areas of low dust extinction, resulting in both large area and a large number of visits and the best performance for all cosmology metrics. List of annotations: (1) `footprint_bluer_footprintv1.5_10yrs`, (2) `footprint_newBv1.5_10yrs`, (3) `bulges_bs_v1.5_10yrs`, and (4) `footprint_big_sky_dustv1.5_10yrs`.

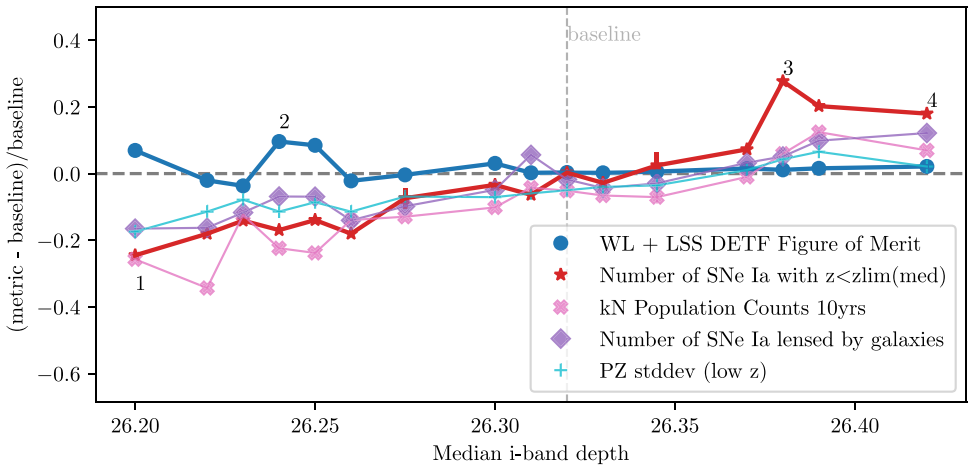

Figure 9. Selected metrics, relative to their values at baseline, as a function of i-band median coadded depth of different observing strategies. To improve readability, points that are nearby in the x-axis are binned with only the mean and error bar plotted for that bin. Here we note that the 3×2 pt FoM metric is generally indifferent to depth, implying that larger area is more important (assuming the changes in depth remain small). Photometric redshifts improve with more depth and the transient science probes, particularly SNe and kNe, are strongly affected since generally greater depth corresponds to increased cadence. We have highlighted specific simulations with numbered annotations. Simulation 1 has the poorest performance because it removes visits from the extragalactic part of the footprint. Simulation 2 is a large area footprint that improves the static science metrics, but still reduces overall cadence and depth. Simulations 3 and 4 artificially perform well for the transient metrics because they unrealistically ignore all mini-surveys and put all visits into WFD. List of annotations: (1) `footprint_newAv1.5_10yrs`, (2) `bulges_cadence_i_heavy_v1.5_10yrs`, (3) `wfd_depth_scale0.99_v1.5_10yrs`, and (4) `wfd_depth_scale0.99_noddf_v1.5_10yrs`.

quantify this trade-off, the WL systematics would need to be included in the full DETF FoM pipeline. Each of the transient metrics has a more complicated relationship with area since they are impacted by other observing strategy choices such as cadence and filter distribution.

Of particular interest are the observing strategies with the largest area. It can be seen that some strategies yield a large area (around 16,500 deg^2) but poor performance for the transient metrics and the WL systematics metric. This is because these strategies have a large footprint but prioritize visits in the Galactic bulge and plane, reducing cadence in the extragalactic area. The largest area simulation, which gives simultaneous high performance for the 3×2 pt FoM and the number of SNe, is the footprint_big_sky_dust simulation that uses an extinction-based cut to define the WFD footprint. This allows both increased area and depth for extragalactic science. The number of visits in WFD and the other surveys is the same as baseline; however, the number of visits in the Galactic plane is dramatically reduced.

Figure 9 shows a subset of metrics as a function of median coadded 5σ i-band depth. Here we see the lack of sensitivity of the 3×2 pt FoM, as long as sufficient depth is achieved. On the other hand, two transient metrics, number of SNe and number of kNe, improve by as much as 27% for a deeper survey. This is really a consequence of an improved number of visits and thus increased cadence. However, most of the simulations that are significantly deeper than the baseline are not actually realistic, as they place artificially large amounts of

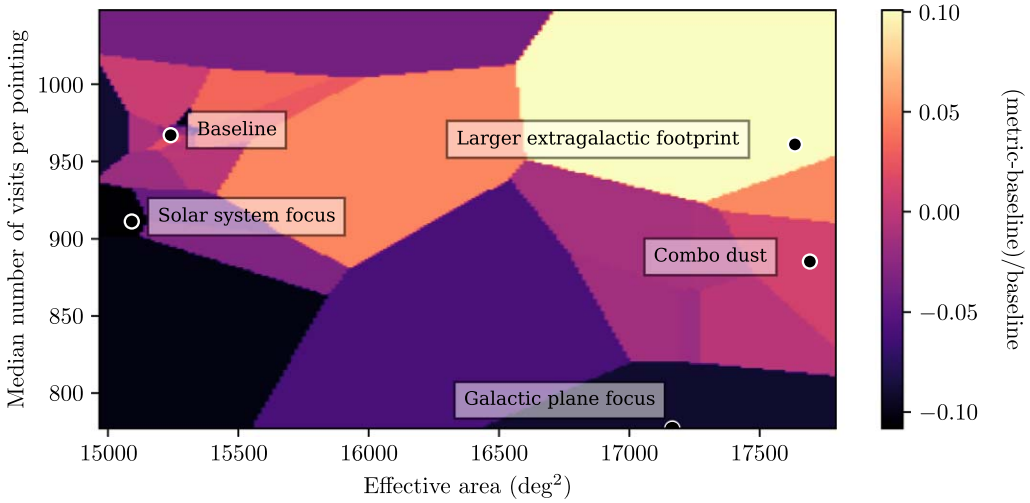

Figure 10. A simple attempt to combine multiple metrics to produce a combined cosmology metric. We show the combined metric (using the usual (metric-baseline)/baseline)) as a function of effective area (i.e., the area that meets the cuts defined in Section 3.1) and median number of visits (which correlates with depth for most simulations). Although the produced map is complex, the trade-off between depth and area can be seen. Several specific strategies are highlighted, the simulation names of which can be found in Table 2. Unlike for the rest of the figures, which focus on simulations that more or less vary one aspect of observing strategy at a time, we have highlighted here strategies from FBS v1.6, which are proposed observing strategies that aim to satisfy multiple science goals. "Solar system focus" and "Galactic plane focus" both take observations away from WFD to prioritize other science cases. "Combo dust" is a proposed large area footprint that has more area but less depth than "Baseline," thus producing a similar performance. "Larger extragalactic footprint" is a larger area footprint that is defined by dust extinction and gives improved cosmological constraints due to the area gained by avoiding extinction. We caution the reader against assuming the performance in this metric would correspond to the exact percentage improvement/degradation in cosmological constraints. Only a full DETF FoM including systematics can indicate the exact numerical impact of observing strategy choice.

survey time into WFD. While these experiments are useful to understand the behavior of the metric, they do not represent viable observing strategies. However, it is entirely possible to choose a footprint or strategy that reduces the overall depth of the survey, to the detriment of the transient probes and photometric redshifts.

Within the constraints of the observing strategy requirements from the LSST SRD, it appears to be quite challenging to achieve a depth much greater than the current baseline. The ideal approach from the standpoint of cosmology is to maintain a WFD footprint of 18,000 deg^2 but prioritize regions with low dust extinction, allowing for increased area while maintaining current depth.

6.2. A Simple Combined Metric to Analyze the Depth–Area Trade-off

To better understand the trade-off between area and depth, we combine metrics in a very simple way based on the DESC SRD. Because of the complexity involved in combining the systematic related metrics such as the WL systematics and photometric redshift metrics, we here only include the 3×2 pt FoM and the number of well-measured SNe metric, in a 50–50 ratio after normalizing each metric to represent the ratio in the amount of information provided by a given simulation versus the baseline. We note that both of these metrics incorporate some (but not all) systematic effects. While we caution the reader that this is a simple approximation of the much more complex full DETF FoM that includes all systematics, it can be helpful in gaining insight. The general trade-off between number of visits (and hence depth) and area can be seen and, while it is difficult to improve over the current baseline, many simulations are much worse. Our preferred footprint, indicated by "Larger extragalactic footprint" that implements a dust-extinction cut, has the best overall performance for cosmology by increasing area without reducing depth. One proposed simulation, indicated as "Combo dust" also has a larger area footprint defined by dust extinction, but includes a significant number of visits in the Galactic plane. This simulation is roughly equivalent to baseline for this combined metric, as can be seen in Figure 10, because what is gained in area is lost in depth.

It is important to note that key probes such as strong lensing and kNe are not included in the combined metric because their constraining power is not comparable to the other probes, and they would be unfairly downweighted. Although they do not contribute as significantly to the overall DETF FoM, these probes are still essential for other aspects of cosmology (unraveling the H_0 tension, for example). We thus consider them separately from this combined metric.

Both kNe and strong lensing suffer if the area is increased, so our recommendation is that any increases in area should maintain the depth in WFD as far as possible by redefining the footprint, using rolling cadence (discussed in Section 7.2), and reoptimizing the number of visits and filter distribution in the Galactic plane and other mini-surveys for those specific science cases.

6.3. Overlap with Other Surveys

Data from other surveys is key to enhancing the cosmological analysis with LSST data. Planned surveys using multi-object spectroscopy including the Dark Energy Spectroscopic Instrument (DESI; DESI Collaboration et al. 2016) and the 4 m Multi-Object Spectroscopic Telescope (4MOST; de Jong et al. 2019) will provide spectroscopic data for millions of objects. This can be used for photometric redshift calibration and training, cross-correlation to improve WL, and host galaxy redshift identification for transients (Mandelbaum et al. 2019). The Time-Domain Extragalactic Survey (TiDES; Swann et al. 2019) on 4MOST will provide live spectra for up to 30,000 transients, including SNe. The Euclid telescope (Laureijs et al. 2011) will survey

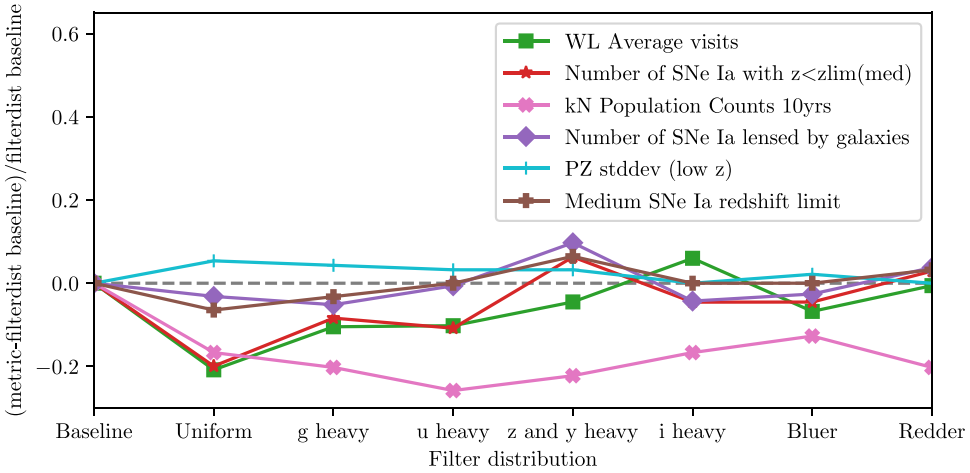

Figure 11. Selected metrics, relative to their values at baseline, as a function of varying choices of filter distributions. Note that these simulations use a different footprint, so all metrics are measured against the baseline `filter_dist` simulation, and not the standard baseline. We do not show the 3 × 2 pt FoM or LSS metrics because they do not vary significantly here. We find no compelling reason to vary the baseline filter distribution: different distributions improve different probes but at the cost of others.

Table 4
Area Overlap for the 10 yr LSST Survey (Extragalactic Footprint) with Other Surveys for Two Types of Footprint: Baseline and a Larger Area Footprint with a Boundary Defined by a Dust Extinction Cut

Simulation Short Name	4MOST Overlap (deg^2)	DESI Overlap (deg^2)	Euclid Overlap (deg^2)
Baseline	14421.22	4178.63	8201.79
Larger extragalactic footprint	14925.47	5951.48	9332.37

Note. See Table 2 for the precise simulation names. The larger footprint provides nearly 2000 deg^2 additional overlap with DESI and over 1000 deg^2 with 4MOST and Euclid. Even greater gains can be made by extending the footprint farther north.

over 15,000 deg^2, obtaining photometric and spectroscopic data in the infrared bands, complementary to LSST and enabling better photometric redshift calibration and training, improvements to WL, and improved de-blending of galaxies (Capak et al. 2019).

Table 4 shows how the overlap with DESI, 4MOST, and Euclid can be improved with a larger area footprint. Even better coverage, particularly for DESI, can be achieved by extending the footprint farther north, as described in Section 7.

Finally, we also note that external data in the DDFs will be crucial for photometric redshift training and SN spectroscopic follow-up (both live and later host galaxy redshift determination). It is particularly important and challenging to align DDF observations temporally with telescopes such as Euclid, meaning that the optimal strategy will likely be to prioritize certain DDFs in certain years. We leave it to future work to determine the optimal DDF strategy for cosmological measurements with LSST.

6.4. Filter Distribution

While the LSST SRD enforces a minimum number of visits in each filter, there is still room to increase or decrease the number of visits (and hence the cadence and depth) in any particular band. This choice of filter distribution has a complex impact on our cosmology metrics, as seen in Figure 11. Here we see some small tensions between probes: SNe prefer a redder distribution while any change from baseline is detrimental to the number of kNe. The choice of filter distribution is especially complex for transients, where the SED of the transient, the underlying efficiency of the LSST bandpasses, and the cadence in each band all play a role. For example, kNe are likely intrinsically redder transients and yet our metric does not improve with a redder filter distribution. This could be because the LSST *g*-band single-visit depth is much deeper than those of the redder bands, meaning the optimum for kNe detection is somewhere between the redder and the bluer distribution and ends up being close to baseline. The baseline filter allocation was heavily influenced by photometric redshifts, so any radical departures from it could prove detrimental. Our photometric redshift metric actually improves slightly for some of the filter distribution simulations, but not significantly. From our results, it is difficult to find a compelling reason to change the distribution from baseline. While the overall filter distribution may already be acceptable for cosmology, *when* those filters are used can still have significant impact on transient metrics, as discussed in Section 6.6.

6.5. Visit Pairs and Exposure Time

Cadence remains a critical choice when designing an observing strategy for any survey that aims to detect transients. One of the most important changes to observing strategy simulations has been to ensure visit pairs are taken in different filters. Figure 12 shows the dramatic decrease in the number of SNe detected if this is not enforced. Although taking pairs in the same filter is more efficient, it severely degrades the overall cadence of the survey. Taking visit pairs in different filters has been the default in simulations since the 2018 white paper call.

Another important decision to be made is whether the visits should be made in a single exposure (1 × 30 s) or separated into two exposures (2 × 15 s), which could help reject cosmic rays. While this decision will only be made during commissioning, Figure 13 shows how beneficial the 1 × 30 s exposure would be to all major cosmology metrics.

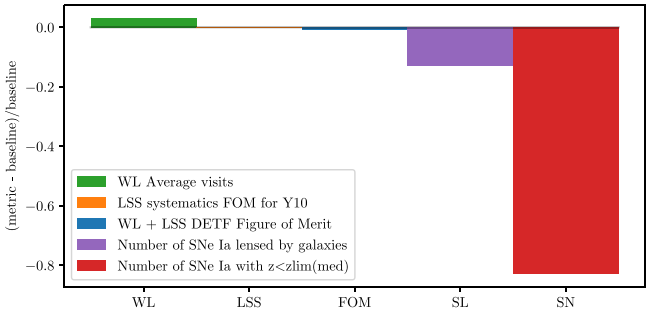

Figure 12. Selected metrics, relative to their values at baseline, for a simulation that takes visit pairs in the same filter instead of changing filters between visits (`baseline_samefilt_v1.5_10yrs`). This clearly shows that while taking visit pairs in the same filter does not impact static science much, it dramatically degrades the transient metrics.

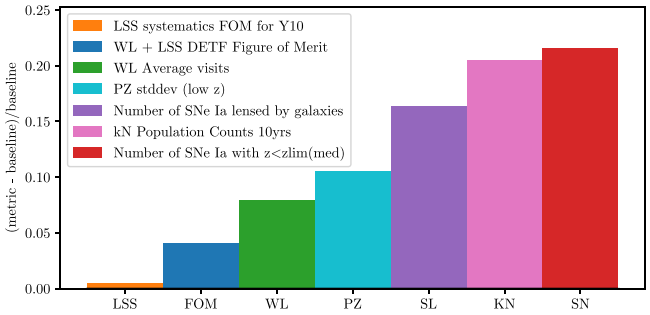

Figure 13. Selected metrics, relative to their values at baseline, for a simulation that uses a single 30 s exposure (`baseline_nexp1_v1.7_10yrs.db`), as compared to a simulation that uses two 15 s exposures (`baseline_nexp2_v1.7_10yrs.db`). Note that the "baseline" used here to normalize the metrics, `baseline_nexp2_v1.7_10yrs.db`, differs from that of other plots. It is clear that the efficiency gained through a single exposure improves all metrics, some dramatically.

6.6. Cadence

Figure 14 shows how different metrics behave as a function of the median internight gap. While the median internight gap is a good measure of overall cadence, it gives an incomplete picture as the tails of the distribution of internight gap can be quite broad. It is still useful however to quantify the average cadence of a simulation. The behavior of the metrics correlates fairly well with the usual area and depth trade-off: larger areas tend to produce worse cadence overall. An interesting point is that while there are only a few simulations that produce better cadence than baseline, there are many that can dramatically reduce the cadence, which has a serious impact on the transient metrics and metrics that depend on depth.

Figure 15 takes the cadence analysis a step further by investigating the impact of internight gap per filter on the number of SNe detected. For fair comparison, only simulations with a similar number of visits in WFD are included. The average cadence varies significantly between the bands, mostly due to the nonuniform filter distribution but also due to decisions of what filters to use in different moon conditions. For SN observations, the redder bands seem to be more important than the bluer bands. We find similar results for strongly lensed SNe and kNe, although kNe have less dependence on the z band.

Figure 16 shows selected metrics as a function of cumulative season length (how long in years a field is observed for on average). We find that as long as the season length remains comparable to the baseline, most metrics are fairly insensitive to changes in season length.

6.7. Weather

Figure 17 shows observing strategies that simulated different weather scenarios, from realistic to impossibly optimistic. Naturally, metrics sensitive to cadence suffer immensely when weather conditions restrict observing time. Although these optimistic simulations are not realistic, they do indicate that much could be gained in transient science by ensuring fields missed by bad weather are observed as soon as possible to avoid large light-curve gaps.

7. Conclusions

Because of its effect on a wide variety of science cases, optimization of LSST observing strategy is critical and has become an active area of research that has already resulted in significant improvements to the scheduler and baseline strategy. In this paper, we have introduced metrics used within the LSST DESC to investigate the impact of observing strategy choices on cosmological measurements with LSST. While we continue to work toward our goal of a full DETF FoM combining all probes, which includes systematics, the metrics introduced here are still valuable in gaining insights about observing strategy.

7.1. LSST DESC Recommendations for Observing Strategy

Here we summarize our conclusions drawn from the metrics outlined in this paper, combined with those from the observing strategy white paper (Lochner et al. 2018):

1. *Footprint and area:* The nominal area for the LSST WFD survey is 18,000 deg^2, but its area and definition can be changed. Lochner et al. (2018) and Olsen et al. (2018) both proposed to shift the WFD footprint away from regions of high Galactic dust extinction to better benefit extragalactic science and rather have a dedicated Galactic plane survey that does not have the same requirements as an extragalactic survey. Figure 8 shows that all metrics decrease with increased effective area, aside from the 3×2 pt FoM, but that one simulation, `footprint_big_sky_dust`, improves all metrics. This simulation proposes a large area footprint but uses a dust-extinction cut to define the boundary. However, it should be noted that this footprint would degrade Galactic science, and so, while it should be considered ideal for cosmology, some compromise must be made to ensure the integrity of other science goals of LSST. Area is not the only consideration of course; there is a natural trade-off between depth and area, with depth (and correspondingly good cadence) being critical for mitigating systematic effects and for transient probes. Figure 10 shows a simple attempt to illustrate this trade-off. In Lochner et al. (2018), we recommended an 18,000 deg^2 footprint, which uses an $E(B-V) < 0.2$ extinction cut, with decl. limits of $-70 <$ decl. < 12.5, which is similar to `footprint_big_sky_dust`. This footprint obtains a large area suitable for extragalactic science and, crucially, extends the footprint north to increase overlap with other surveys like DESI and Euclid. Our work in this paper shows that such a footprint would ensure cosmology goals with

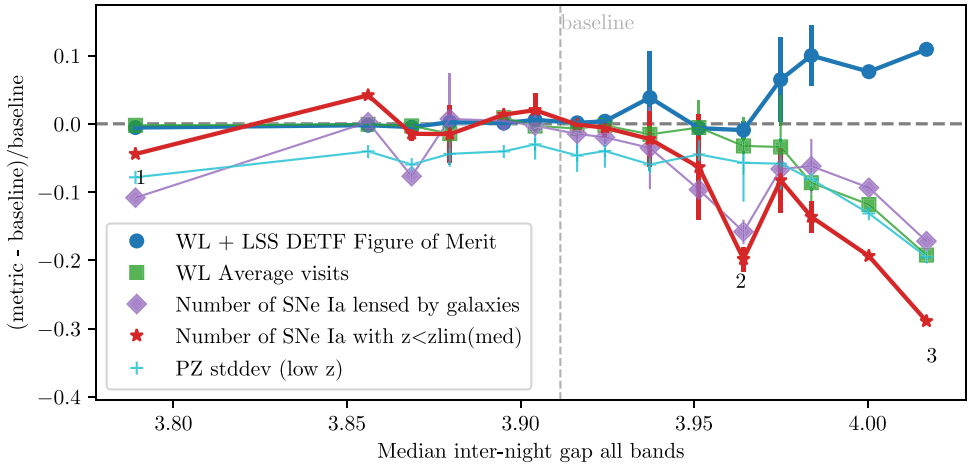

Figure 14. Selected metrics, relative to their values at baseline, as a function of the median internight gap of different observing strategies. To improve readability, points that are nearby in the *x*-axis are binned with only the mean and error bar plotted for that bin. The SN metric prefers shorter gaps between observations, but the situation is complicated by many other factors. Generally it is important that the cadence is at least no worse than the baseline. We have highlighted specific simulations with numbered annotations. Simulation 1 includes visits at high airmass, which is generally avoided, indicating that high airmass observations can improve cadence without too much degradation of metrics. Simulation 2 is when 2 × 15 s exposures are used instead of 1 × 30 s. Simulation 3 is a large area footprint but removes visits from the extragalactic area reducing cadence. List of annotations: 1-`dcr_nham2_ugr_v1.5_10yrs`, 2-`baseline_2snaps_v1.5_10yrs`, 3-`footprint_newAv1.5_10yrs`.

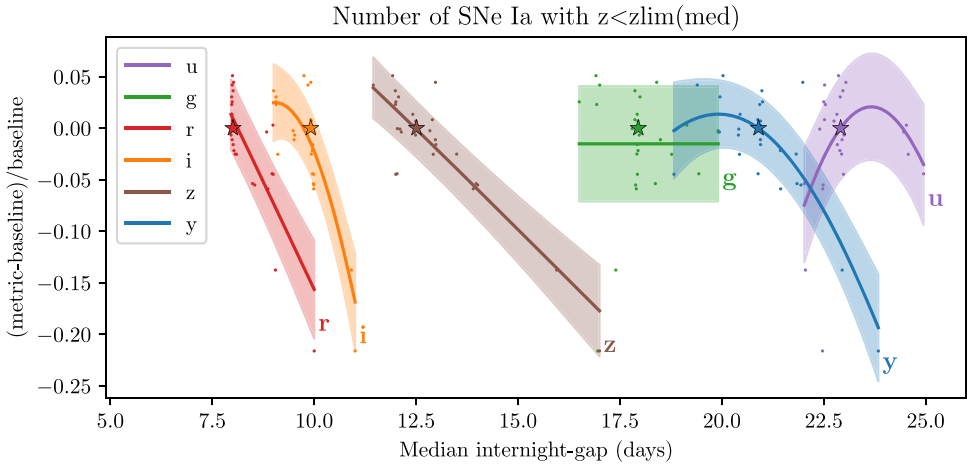

Figure 15. The relative improvement over baseline for the number of SNe as a function median internight gap in each filter. The baseline cadence is represented by star markers. In general, the number of SNe improve as the cadence improves, but note the relatively small dependence on the *g* and *u* bands, indicating the higher importance of good cadence in the redder bands. A Gaussian process has been fitted to the set of points in each filter to help guide the eye; the mean is plotted with a solid line, and the standard deviation indicated with a colored envelope. Also note that we have restricted this plot to simulations with similar numbers of visits in the WFD.

LSST are met. Of course, dedicating 18,000 deg^2 to the extragalactic area means visits must be moved from the WFD to the Galactic plane to support Galactic science goals, potentially reducing the cadence and impacting the transient probes and some systematics. This could be mitigated with a change in strategy such as implementing rolling cadence (see Section 7.2).

2. *Survey uniformity:* Uniformity of depth across the WFD is a strict requirement for cosmology to avoid systematic effects being introduced by nonuniform measurements. We have a metric to measure uniformity (described in Section 3.1.1), which finds that most strategies do not deviate significantly from baseline in terms of uniformity, indicating the SRD requirements are being met. However, it is worth pointing out that there is some flexibility at what points in the survey the observations need to be uniform. Lochner et al. (2018) proposed to ensure observations be uniform by the data releases at the end of years 1, 3, 6, and 10 to enable periodic cosmological analyses; however, these are not yet finalized and can be changed. The important point is that any strategy should allow regular "checkpoints" where uniformity is achieved and analysis-ready data can be released. This should naturally be straightforward with most strategies, but we note that rolling cadence (further discussed in Section 7.2) could affect when these uniform "checkpoints" are possible.
3. *Dithering:* The process of adding random rotational and translational offsets when repeating an observation in a field, known as dithering, is extremely important for cosmology systematics. Uniform coverage of 180° of camera rotation angle in every field is important for reduction of camera-based PSF systematics. All simulations used here implement translational and rotational

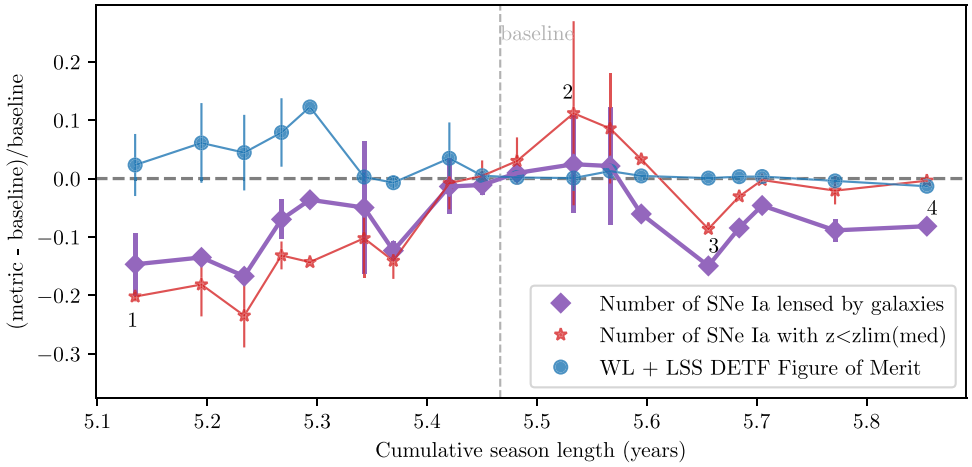

Figure 16. Selected metrics, relative to their values at baseline, as a function of the cumulative season length (the total amount of time a field is observed) of different observing strategies. To improve readability, points that are nearby in the *x*-axis are binned with only the mean and error bar plotted for that bin. Shorter season lengths degrade strong lensing and SNe performance, but as long as the season length does not fall much below its value at baseline, these metrics do not seem to be strongly affected. We have highlighted specific simulations with numbered annotations. Simulation 1 is a larger area footprint, which removes visits from the extragalactic area. Simulation 2 artificially improves cadence (and hence SN performance) by removing the mini-surveys. Simulation 3 replaces some visits with short exposures that cannot be used for SN observations, thus reducing signal-to-noise. And Simulation 4 takes some observations at high airmass, which is usually avoided. This indicates that taking some number of high airmass observations is an effective way to increase season length without degrading metric performance significantly. List of annotations: (1) `footprint_newBv1.5_10yrs`, (2) `wfd_depth_scale0.99_v1.5_10yrs`, (3) `short_exp_5ns_5expt_v1.5_10yrs`, and (4) `dcr_nham1_ugri_v1.5_10yrs`

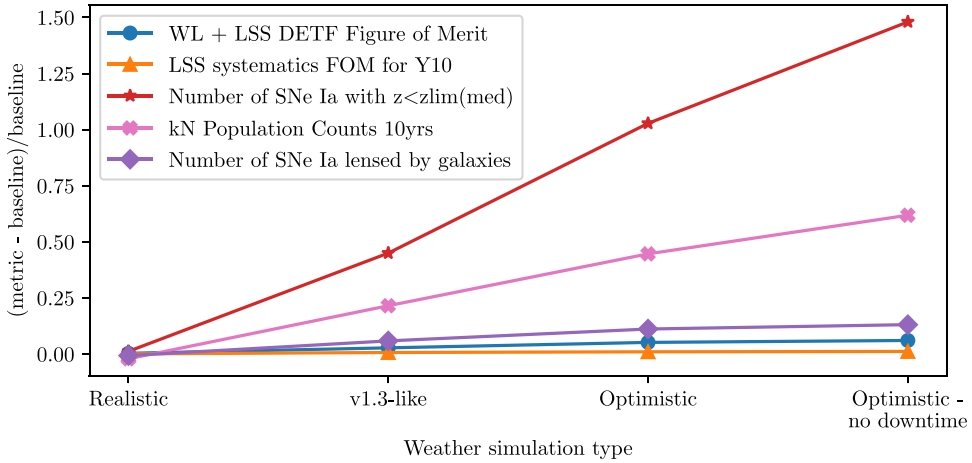

Figure 17. Selected metrics, relative to their values at baseline, as a function of different weather simulations, for an older FBS version, v1.3. The simulations make use of real weather data and differ on what cloud coverage would be required before observations are paused: 30% cloud cover (realistic), 70% cloud cover (v1.3-like, in other words similar to previous simulations), any cloud cover (optimistic, where the dome would never close due to bad weather), and no down time (optimistic, no down time, where the telescope would always observe no matter what). While bad weather only has a small impact on static science metrics by slightly reducing the median depth, it has an enormous impact on transient science metrics by as much as 150%. This highlights the importance of including realistic weather in simulations to make accurate predictions of scientific returns.

dithers, which help maintain survey uniformity. While some investigation remains, we consider the current baseline dithering strategy to be largely effective. We note, however, that it is important that any translational dithers used in the DDFs should be as small as possible so that the depth of the fields is not compromised.

4. *Overlap with other surveys:* As discussed in Section 6.3, overlap with other surveys such as DESI, 4MOST: TiDES, and Euclid is critical for a number of cosmological probes. Spectroscopic data and infrared photometry can provide better calibration and training of photometric redshifts, improved WL constraints, and more accurate de-blending of galaxies. Transient probes, in particular SNe, will rely heavily on spectroscopy for host galaxy redshift identification and live spectroscopic follow-up to provide training sets for photometric classifiers. Improving overlap with other surveys further supports our recommendation for a larger footprint, particularly one that extends farther north to improve overlap with DESI and Euclid.
5. *Exposure time:* The LSST SRD describes a visit to a field as a pair of 15 s exposures (2 × 15 s). The current plan is to combine the two 15 s snaps into a single 30 s exposure after basic processing (e.g., flat-fielding, cosmic-ray removal) but before any PSF modeling is performed. The primary goal of the 2 × 15 s exposure strategy is to robustly reject cosmic rays, but Lochner et al. (2018) reported a potential gain of 7% efficiency, as well as improved image quality, by switching to a single exposure (1 × 30 s) and using standard cosmic-ray

rejection techniques (e.g., Aihara et al. 2018). Our results show that a single exposure can improve our metrics by as much as 20% (see Figure 13). The only advantage of short exposures is that they may mitigate saturation effects for very low-redshift ($z < 0.05$) SNe. However, this advantage does not outweigh the severe degradation to the larger SNe sample. While we strongly advocate for single 30 s exposures, we acknowledge that the final decision about exposure times will only be taken after commissioning tests prove that cosmic rays and satellite trails can be rejected accurately.

6. *Repeated visits in a night:* A requirement in the SRD is that any given LSST field must be visited at least twice within a short period of time (15–40 minutes) to ensure accurate asteroid detection and assist with orbit characterization. There is a clear trade-off between the number of *intra-night* visits and *inter-night* visits, which can heavily impact cosmology with transient objects. Hence, Lochner et al. (2018) proposed that repeat visit pairs be in different filters to improve transient characterization. This has been adopted for all of the simulations used in this paper and remains one of our strongest recommendations, as it significantly improves the transient metrics (Figure 12).
7. *Cadence:* The return time with which each field in the WFD is observed is of critical importance to many science cases and is one of the most difficult factors to optimize. While the LSST SRD states that every field must be observed on average 825 times over the survey, it does not specify how those visits should be distributed, and while each field is visited every 3 days *on average*, there is a large spread in this distribution. Figure 14 shows the impact of the median internight gap on various metrics. Significant gains can be made by changing how the filters are allocated and ensuring the *griz* filters maintain regular cadence, whatever the moon conditions. We find that the redder filters, *riz*, are particularly important for SN characterization and need to have high cadence (see Figure 15). The *u* band, while important for photometric redshifts, does not heavily impact our transient probes and thus does not require high cadence. Cadence and filter allocation are expected to also be important for transient classification, which, while not included as a metric, is discussed in Section 7.2.

7.2. Future Work

While significant gains have been made in understanding the impact of observing strategy on cosmological constraints with LSST, there remain some unanswered questions and active areas of research:

1. *Rolling cadence:* This is a proposed strategy that can increase the cadence in a certain area of sky by prioritizing it for a period of time at the expense of other areas. This would then be repeated for the initially deprioritized areas, allowing a pattern of rolling cadence. While rolling cadence is very promising, it has proved difficult to simulate. The LSST SRD proper-motion requirements mean that the survey must have uniform coverage at its start and end, meaning a rolling cadence strategy must start and stop smoothly. It also has to roll for full observing seasons, constraining when a roll can start. Finally, rolling cadence strategies have many parameters to optimize, including how large an area to roll at a time, for how long and whether to roll completely, and allowing the remaining part of the survey to still be observed but at a lower priority. All of these complexities and constraints have made it difficult to find a rolling strategy that works well for transient science without impacting other science cases. Although it is difficult, we still consider it a promising option to obtain a large sky area without reducing cadence, and are continuing to research this approach. We note that while rolling cadence may negatively affect strong lensing, the potential gains to the number of well-measured SNe are significant enough to warrant further exploration.
2. *Transient classification:* All of our transient metrics have made the explicit assumption that transients detected with LSST can be perfectly classified, which is obviously not true. Not only does photometric transient classification impact the ability to constrain cosmology with SNe, kNe, and strongly lensed SNe, it is also in turn affected by the quality of the photometric redshift estimation. Additionally, transient classification will likely be highly sensitive to observing strategy, making a transient classification metric critical. This metric is a challenge to develop as it would be affected by many aspects of observing strategy, would need to encompass a broad array of transient classes, and could be somewhat dependent on the choice of classifier used. Because photometric transient classification for large surveys is still a young research field, not enough is known to build a reliable emulator to use as a fast metric. Thus, to evaluate a single observing strategy, sophisticated and expensive simulations must be run. An important step in this direction was taken by the PLAsTiCC challenge (The PLAsTiCC team et al. 2018), which was a community challenge that released a set of LSST-like simulations incorporating a wide variety of transients and variables. Work is ongoing to understand the effect of observing strategy on existing transient classification pipelines (e.g., Lochner et al. 2016; Möller & de Boissière 2019) using the PLAsTiCC simulations themselves (Alves et al. 2022), and also by running new simulations based on some of the observing strategies discussed in this paper.
3. *A third visit:* There is a proposal to add an additional third visit some time after the visit pair, repeating one of the filters used in the original pair. The motivation is to improve early classification of transients, which will affect all transient science and be of particular importance to spectroscopic follow-up. However, a third observation in the same band will reduce overall cadence and hence the number of well-measured SNe, kNe, and strongly lensed transients. As stated in the previous point, we do not yet have a reliable classification metric to evaluate this approach. Classification performance must be carefully weighed against overall cosmological constraints for a balance to be found. Considerations must also be made about the spectroscopic follow-up strategy to build training sets for classifiers. Thus, the question of whether or not to add a third visit is actually only a single component of an important and complex analysis that is ongoing.

4. *DDFs:* Cosmology with LSST relies heavily on the DDFs for both photometric redshift training and particularly to provide a deep sample of high-redshift SNe. While this paper focused on the WFD, many of the metrics developed for the SN probe in Section 5.1.1 can be applied to evaluate different DDF strategies. The optimal outcome for the DDFs is that they produce a well-measured SN Ia sample at significantly higher redshifts than the WFD. The cadence for a DDF survey will be on the scale of every ∼2 days rather than every 10–15 days, which results in higher-quality light curves than WFD. However, to improve the redshift range of SNe detected in the DDFs, we advocate here for two strategies: (1) rolling deep fields such that in a given year, one to two fields can go significantly deeper than the other fields, and (2) a significantly skewed filter allocation toward the redder bands.
5. *Satellite Trails:* The rapid increase of low-Earth orbit satellites (McDowell 2020) presents a fundamental challenge for ground-based telescopes such as LSST. Tyson et al. (2020) studied various mitigation techniques for limiting the impact on the LSST observing program, including changes to the observing strategy. Future observing-strategy work should continue on this initial study so as to ensure that satellites have a minimal effect on LSST science.

The metrics developed in this paper have been invaluable in gaining insights about observing strategy and lay a useful foundation for ongoing work on this challenging optimization problem.

T.A. acknowledges support from Proyecto Fondecyt N 1190335 and the ANID-Millennium Science Initiative Program —ICN12_009 awarded to the Millennium Institute of Astrophysics MAS. H.Aw. also acknowledges support by the Rutgers Discovery Informatics Institute (RDI2) Fellowship of Excellence in Computational and Data Science, Rutgers University & Bevier Dissertation Completion Fellowship, Leinweber Postdoctoral Research Fellowship, and DOE grant DE-SC009193. H.Aw. also thanks the LSSTC Data Science Fellowship Program, which is funded by LSSTC, NSF Cybertraining Grant #1829740, the Brinson Foundation, and the Moore Foundation, as participation in the program has benefited this work. This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics Cosmic Frontier Research program under Award Nos. DE-SC0011636 and DE-SC0010008, which supported E.G. and H.Aw. S.G.G. acknowledges support from the University of Rochester and the DOE Office of High Energy Physics under grant No. DE-SC0008475. M.L.G. acknowledges support from the DIRAC Institute in the Department of Astronomy at the University of Washington. The DIRAC Institute is supported through generous gifts from the Charles and Lisa Simonyi Fund for Arts and Sciences, and the Washington Research Foundation.

S.H. and S.H.S. thank the European Research Council (ERC) for support under the European Union's Horizon 2020 research and innovation program (LENSNOVA: grant agreement No. 771776).

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics Cosmic Frontier Research program under Award Nos. DE-SC0011636, which supported S.W.J. The work of A.G.K. is supported by the U.S. Department of Energy, Office of Science, Office of High Energy Physics, under contract No. DE-AC02-05CH11231. M.L. acknowledges support from the South African Radio Astronomy Observatory and the National Research Foundation (NRF) toward this research. Opinions expressed and conclusions arrived at, are those of the authors and are not necessarily to be attributed to the NRF. R.M. and H.A. are supported by the Department of Energy grant DE-SC0010118. This document was prepared by DESC using the resources of the Fermi National Accelerator Laboratory (Fermilab), a U.S. Department of Energy, Office of Science, HEP User Facility. Fermilab is managed by Fermi Research Alliance, LLC (FRA), acting under Contract No. DE-AC02-07CH11359. C.S., R.B., and H.V.P.'s work was partially supported by the research environment grant "Gravitational Radiation and Electromagnetic Astrophysical Transients (GREAT)" funded by the Swedish Research Council (VR) under Dnr 2016-06012 and the research project grant "Understanding the Dynamic Universe" funded by the Knut and Alice Wallenberg Foundation under Dnr KAW 2018.0067. T.P. acknowledges the financial support from the Slovenian Research Agency (grants I0-0033, P1-0031, J1-8136, and Z1-1853). S.H.S. thanks the Max Planck Society for support through the Max Planck Research Group.

The DESC acknowledges ongoing support from the Institut National de Physique Nucléaire et de Physique des Particules in France; the Science & Technology Facilities Council in the United Kingdom; and the Department of Energy, the National Science Foundation, and the LSST Corporation in the United States. DESC uses resources of the IN2P3 Computing Center (CC-IN2P3–Lyon/Villeurbanne—France) funded by the Centre National de la Recherche Scientifique; the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; STFC DiRAC HPC Facilities, funded by UK BIS National E-infrastructure capital grants; and the UK particle physics grid, supported by the GridPP Collaboration. This work was performed in part under DOE Contract DE-AC02-76SF00515. This paper has undergone internal review in the LSST Dark Energy Science Collaboration. The authors thank the internal reviewers, who were Ariel Goobar, Joshua Meyers, and Samuel Schmidt. The authors also thank Jean-Charles Cuillande and Christopher Frohmaier for providing the footprint maps for Euclid and 4MOST, respectively.

This work was supported by the Preparing for Astrophysics with LSST Program, funded by the Heising Simons Foundation through grant 2021-2975, and administered by Las Cumbres Observatory.

Author Contributions

H.Al.: Developed WL systematics metric, contributed MAF metric for 3×2 pt FoM. Analysis and interpretation for WL metric results. Writing and discussion for Sections 4 and 6. T.A.: Helped develop time-delay lensed quasar metric, run metrics, analysis, and discussion of Section 5.2. H.Aw.: Provided technical assistance with simulations. Developed MAF metrics for the extragalactic footprint and the number of galaxies. Provided numbers for basic static metrics (e.g., depth statistics) and those for LSS. Helped with discussions, and writing Sections 3 and 4. R.B.: Contributed to software and

metric development. T.B.: Contributed to long timescale microlensing efficiency understanding. E.G.: Conceptualization, discussion, and writing in Sections 3 and 4. Edited Sections 1–4 and 7. S.G.G.: Section 5.1.3, metric for SNe PV. M.L.G.: Section 3.2, photo-z, data set creation, writing. P.G.: Section 5.1: discussion, oversight, and writing. Developed a set of SNe metrics implemented in MAF. Processed all observing strategies. Developed real-time display (https://me.lsst.eu/gris/OS_Videos/) to visualize the progress of observing strategies on a daily basis to study cadence/gap effects. Implementation of the lensed SNe Ia by galaxies metric in MAF and processing of observing strategies (Section 5.2.1). I. H.: Section 5.1, discussion. Writing (review and editing) of Sections 1, 2, and 5. S.H.: Section 5.2.1, metric, Section 5.2, validation, analysis, writing, and editing. S.W.J.: Conceptualization and development of observing strategies related to SN Ia (Section 5.1) and rolling cadences. Editing of Sections 1 and 2. R.L.J.: Created simulations, technical assistance with simulations, and metric interpretation. Worked with various groups to develop metrics and analysis. A.G.K.: Section 5.1.3, writing/metric-creation. M.L.: Conceptualization, software, validation, formal analysis, investigation, data curation, writing—original draft, writing—review and editing, visualization, and supervision. R.M.: Discussion and writing in Sections 3 and 4. P.M.: Section 5.2 (designed and implemented metrics, wrote text). M.M.: Microlensing discussion and writing in introduction. E.H.N.: Improved seeing input to opsim simulations (Sections 2.5, 6.7), general discussions, and recommendations. H.P.: Section 5.3: conceptualization, validation, writing—review, supervision, and funding acquisition. T. P.: Metric development and writing of Section 5.2.2. N.R.: Section 5.1, discussion and writing. Developed a set of metrics to evaluate the expected quality on the SN light curves, on the DDF and WFD surveys. Built a pipeline to evaluate the pre-white-paper, white-paper, and fbs1.3–fbs1.7 cadences. D.R.: Developed deterministic scheduler (alt sched) as a conceptual alternative to OpSim for comparison purposes. D.S.: Conceptualization, writing—original draft, writing—review and editing, visualization, and supervision. C.N.S.: Software, metrics, validation, analysis, and writing of Section 5.3. C.S.: Developed deterministic scheduler (alt sched) as a conceptual alternative to OpSim for comparison purposes. S.H.S.: Coordinated strong lensing group; writing, validation, analysis, and editing of Section 5.2. P.Y.: Created survey simulations. Section 2 writing. Helped with development of MAF metrics for Sections 3, 5.

Appendix A
Metric Descriptions and Transformations

In order to directly compare metrics that have different interpretations, we apply transforms to several metrics listed in Table 5. We also list the value of each metric at the baseline observing strategy and provide a quick link to the corresponding section in the paper.

Table 5
Metrics Used in This Paper

Metric Description	Transform	Value at Baseline	Section in Paper
Y1 median i-band coadded depth	$1/(10^{-\frac{x}{2.5}})^2$	25	3.1
Y10 median i-band coadded depth	$1/(10^{-\frac{x}{2.5}})^2$	26	3.1
Y1 i-band coadded depth stddev	$1/(10^{-\frac{x}{2.5}})^2$	0.16	3.1
Y10 i-band coadded depth stddev	$1/(10^{-\frac{x}{2.5}})^2$	0.12	3.1
Y1 area for static science (deg^2)	x	1.5e+04	3.1
Y10 area for static science (deg^2)	x	1.5e+04	3.1
PZ outlier fraction (low z)	$1-2x$	0.0019	3.2
PZ stddev (low z)	$1/x^2$	0.0096	3.2
PZ outlier fraction (high z)	$1-2x$	0.044	3.2
PZ stddev (high z)	$1/x^2$	0.022	3.2
WL + LSS DETF FoM	x	36	4.1.1
WL average visits	x	6.1e+02	4.1
LSS systematics FOM for Y1	x	0.94	4.2.2
LSS systematics FOM for Y10	x	0.98	4.2.2
Y1 $N_{\rm gal}$ at $0.66 < z < 1.0$	x	4.5e+08	4.2.1
Y10 $N_{\rm gal}$ at $0.66 < z < 1.0$	x	9.5e+08	4.2.1
Number of SNe Ie with $z <$ zlim(faint)	x	4.5e+04	5.1.1
Faint SNe Ia redshift limit	x	0.17	5.1.1
Number of SNe Ia with $z <$ zlim(med)	x	1.6e+05	5.1.1
SNe Ia r-band S/N	x	0.59	5.1.2
SNe Ia r-band redshift limit	x	0.31	5.1.2
Medium SNe Ia redshift limit	x	0.27	5.1.1
SN Ia peculiar velocities	x	1.4	5.1.3
Number of SNe Ia lensed by galaxies	x	8.7	5.2.1
Number of lensed quasars	x	3.3e+02	5.2.3
Number of SNe Ia lensed by clusters	x	0.18	5.2.2
GW170817-like kNe counts 10 yr	x	97	5.3.1
kNe MAF mean counts	x	34	5.3.1
kN population counts 10 yr	x	40	5.3.1

Note. The first column describes the metric, and the second indicates how the metric (x) is transformed such that it can be directly compared with other metrics and interpreted as "larger is better." The third column indicates the original metric value at baseline, and the last column directs the reader to the section where the metric is described.

Appendix B
List of Observing Strategy Simulations

All simulations used in this paper, as well as those that have been explicitly excluded, are given in Table 6. We have focused on the v1.5 FBS simulations but indicate where some have been excluded, usually due to bugs or a variation in the simulation that artificially changes a metric and is thus not directly comparable to the baseline. We highlight which figures each simulation is used in, noting particularly that Figure 15 is restricted to simulations with similar total numbers of visits. We do not indicate the list of simulations used in the figures in Sections 4 and 5 since it is explicitly indicated there.

Table 6
List of All Simulations Used in This Paper, Which Baseline Simulation It Is Compared Against, and Which Figures the Simulation Was Used in (Aside from Early Figures in the Paper, Where the Simulation is Explicitly Indicated)

Simulation Name	Comparison Baseline	Included in Which Figures
weather_0.3_v1.4_10yrs	weather_0.3_v1.4_10yrs	17
weather_0.7_v1.4_10yrs	weather_0.3_v1.4_10yrs	17
weather_1.2_ndt_v1.4_10yrs	weather_0.3_v1.4_10yrs	17
weather_1.2_v1.4_10yrs	weather_0.3_v1.4_10yrs	17
agnddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
alt_dust_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
alt_roll_mod2_dust_sdf_0.20_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
baseline_2snaps_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
baseline_samefilt_v1.5_10yrs	baseline_v1.5_10yrs	12
baseline_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 12, 13, 14, 15, 16, 17
bulges_bs_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
bulges_bulge_wfd_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
bulges_cadence_bs_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
bulges_cadence_bulge_wfd_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
bulges_cadence_i_heavy_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
bulges_i_heavy_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
dcr_nham1_ug_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
dcr_nham1_ugr_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
dcr_nham1_ugri_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
dcr_nham2_ug_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
dcr_nham2_ugr_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
dcr_nham2_ugri_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
descddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
filterdist_indx1_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx2_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx3_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx4_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx5_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx6_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx7_v1.5_10yrs	baseline_v1.5_10yrs	11
filterdist_indx8_v1.5_10yrs	baseline_v1.5_10yrs	11
footprint_add_mag_cloudsv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
footprint_big_sky_dustv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
footprint_big_sky_nouiyv1.5_10yrs	baseline_v1.5_10yrs	Excluded
footprint_big_skyv1.5_10yrs	baseline_v1.5_10yrs	Excluded
footprint_big_wfdv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
footprint_bluer_footprintv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
footprint_gp_smoothv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
footprint_newAv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
footprint_newBv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
footprint_no_gp_northv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
footprint_standard_goalsv1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
footprint_stuck_rollingv1.5_10yrs	baseline_v1.5_10yrs	Excluded
goodseeing_gi_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
goodseeing_gri_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
goodseeing_griz_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
goodseeing_gz_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
goodseeing_i_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
greedy_footprint_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
roll_mod2_dust_sdf_0.20_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
rolling_mod2_sdf_0.10_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
rolling_mod2_sdf_0.20_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
rolling_mod3_sdf_0.10_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
rolling_mod3_sdf_0.20_v1.5_10yrs	baseline_v1.5_10yrs	Excluded

Table 6
(Continued)

Simulation Name	Comparison Baseline	Included in Which Figures
rolling_mod6_sdf_0.10_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
rolling_mod6_sdf_0.20_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
short_exp_2ns_1expt_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
short_exp_2ns_5expt_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
short_exp_5ns_1expt_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
short_exp_5ns_5expt_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
spiders_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt120v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt15v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt30v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt45v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt60v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
third_obs_pt90v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
twilight_neo_mod1_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
twilight_neo_mod2_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
twilight_neo_mod3_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
twilight_neo_mod4_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
u60_v1.5_10yrs	baseline_v1.5_10yrs	Excluded
var_expt_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 16
wfd_depth_scale0.65_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.65_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.70_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.70_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.75_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.75_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.80_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 15, 16
wfd_depth_scale0.80_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.85_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 15, 16
wfd_depth_scale0.85_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 15, 16
wfd_depth_scale0.90_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.90_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 10, 14, 15, 16
wfd_depth_scale0.95_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.95_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.99_noddf_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
wfd_depth_scale0.99_v1.5_10yrs	baseline_v1.5_10yrs	8, 9, 16
barebones_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
baseline_nexp1_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
combo_dust_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
ddf_heavy_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
dm_heavy_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
even_filters_alt_g_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
even_filters_altv1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
even_filters_g_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
even_filtersv1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
mw_heavy_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
rolling_exgal_mod2_dust_sdf_0.80_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
rolling_fpo_2nslice0.8_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_2nslice0.9_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_2nslice1.0_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_3nslice0.8_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_3nslice0.9_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_3nslice1.0_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_6nslice0.8_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_6nslice0.9_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
rolling_fpo_6nslice1.0_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	Excluded
ss_heavy_v1.6_10yrs	baseline_nexp1_v1.6_10yrs	10
baseline_nexp1_v1.7_10yrs	baseline_nexp2_v1.7_10yrs	13
baseline_nexp2_v1.7_10yrs	baseline_nexp2_v1.7_10yrs	13

ORCID iDs

Michelle Lochner https://orcid.org/0000-0003-2221-8281
Husni Almoubayyed https://orcid.org/0000-0001-5717-9310
Timo Anguita https://orcid.org/0000-0003-0930-5815
Humna Awan https://orcid.org/0000-0003-2296-7717
Eric Gawiser https://orcid.org/0000-0003-1530-8713
Melissa L. Graham https://orcid.org/0000-0002-9154-3136
Saurabh W. Jha https://orcid.org/0000-0001-8738-6011
Rachel Mandelbaum https://orcid.org/0000-0003-2271-1527
Tanja Petrushevska https://orcid.org/0000-0003-4743-1679
Christian N. Setzer https://orcid.org/0000-0002-7439-2735
Sherry H. Suyu https://orcid.org/0000-0001-5568-6052
Peter Yoachim https://orcid.org/0000-0003-2874-6464
Rahul Biswas https://orcid.org/0000-0002-5741-7195
Isobel Hook https://orcid.org/0000-0002-2960-978X
Eric Neilsen https://orcid.org/0000-0002-7357-0317
Hiranya Peiris https://orcid.org/0000-0002-2519-584X
Christopher Stubbs https://orcid.org/0000-0003-0347-1724

References

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, ApJL, 848, L12
Abell, P. A., Allison, J., Anderson, S. F., et al. 2009, arXiv:0912.0201
Aihara, H., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S8
Albrecht, A., Bernstein, G., Cahn, R., et al. 2006, arXiv:astro-ph/0609591
Almoubayyed, H., Mandelbaum, R., Awan, H., et al. 2020, MNRAS, 499, 1140
Alves, C. S., Peiris, H. V., Lochner, M., et al. 2022, ApJS, 258, 23
Amanullah, R., Goobar, A., Clément, B., et al. 2011, ApJL, 742, L7
Awan, H., Gawiser, E., Kurczynski, P., et al. 2016, ApJ, 829, 50
Betoule, M., Kessler, R., Guy, J., et al. 2014, A&A, 568, A22
Bianco, F. B., Željko, I., Jones, R. L., et al. 2022, ApJS, 258, 1
Bosch, J., Armstrong, R., Bickerton, S., et al. 2018, PASJ, 70, S5
Capak, P., Cuillandre, J.-C., Bernardeau, F., et al. 2019, arXiv:1904.10439
Chaves-Montero, J., Angulo, R. E., & Hernández-Monteagudo, C. 2018, MNRAS, 477, 3892
Chen, H.-Y., Fishbach, M., & Holz, D. E. 2018, Natur, 562, 545
Courbin, F., Bonvin, V., Buckley-Geer, E., et al. 2017, arXiv:1706.09424
Davis, T. M., Hui, L., Frieman, J. A., et al. 2011, ApJ, 741, 67
de Jong, R. S., Agertz, O., Berbel, A. A., et al. 2019, Msngr, 175, 3
Delgado, F., Saha, A., Chandrasekharan, S., et al. 2014, Proc. SPIE, 9150, 915015
DESI Collaboration, Aghamousa, A., Aguilar, J., et al. 2016, arXiv:1611.00036
Dilday, B., Smith, M., Bassett, B., et al. 2010, ApJ, 713, 1026
Eigenbrod, A., Courbin, F., Vuissoz, C., et al. 2005, A&A, 436, 25
Goldstein, D. A., Nugent, P. E., Kasen, D. N., & Collett, T. E. 2018, ApJ, 855, 22
Gonzalez-Perez, V., Lacey, C. G., Baugh, C. M., et al. 2014, MNRAS, 439, 264
Goobar, A., Paech, K., Stanishev, V., et al. 2009, A&A, 507, 71
Goobar, A., Amanullah, R., Kulkarni, S. R., et al. 2017, Sci, 356, 291
Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759
Graham, M. L., Connolly, A. J., Ivezić, Ž., et al. 2018, AJ, 155, 1
Graham, M. L., Connolly, A. J., Wang, W., et al. 2020, AJ, 159, 258
Grillo, C., Karman, W., Suyu, S. H., et al. 2016, ApJ, 822, 78
Grillo, C., Rosati, P., Suyu, S. H., et al. 2020, ApJ, 898, 87
Grillo, C., Rosati, P., Suyu, S. H., et al. 2018, ApJ, 860, 94
Guy, J., Astier, P., Baumont, S., et al. 2007, A&A, 466, 11
Guy, J., Sullivan, M., Conley, A., et al. 2010, A&A, 523, A7
Heitmann, K., Higdon, D., White, M., et al. 2009, ApJ, 705, 156
Holz, D. E. 2001, ApJL, 556, L71
Holz, D. E., & Hughes, S. A. 2005, ApJ, 629, 15
Huber, S., Suyu, S. H., Ghoshdastidar, D., et al. 2022, A&A, 658, A157
Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2019, arXiv:1903.00510
Huber, S., Suyu, S. H., Noebauer, U. M., et al. 2021, A&A, 646, A110
Hui, L., & Greene, P. B. 2006, PhRvD, 73, 123526
Ivezić, Ž., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Ivezić, Ž. & The LSST Science Collaboration 2013, The LSST System Science Requirements Document, LPM-17, http://ls.st/LPM-17
Jarvis, M., Sheldon, E., Zuntz, J., et al. 2016, MNRAS, 460, 2245
Jones, L., Yoachim, P., & Ivezic, Z. 2021, lsst-pst/pstn-051: v1.2.1, Zenodo, doi:10.5281/zenodo.4598873
Jones, R. L., Yoachim, P., Chandrasekharan, S., et al. 2014, Proc. SPIE, 9149, 91490B
Kelley, L. Z., Mandel, I., & Ramirez-Ruiz, E. 2013, PhRvD, 87, 123004
Kelly, P. L., Brammer, G., Selsing, J., et al. 2016a, ApJ, 831, 205
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2015, Sci, 347, 1123
Kelly, P. L., Rodney, S. A., Treu, T., et al. 2016b, ApJL, 819, L8
Kilbinger, M. 2015, RPPh, 78, 086901
Kim, A. G., Aldering, G., Antilogus, P., et al. 2019, BAAS, 51, 140
Krause, E., & Eifler, T. 2017, MNRAS, 470, 2100
Laureijs, R., Amiaux, J., Arduini, S., et al. 2011, arXiv:1110.3193
Leike, R. H., & Enßlin, T. A. 2019, A&A, 631, A32
Lewis, A., & Bridle, S. 2002, PhRvD, 66, 103511
Li, L., Shen, S., Hou, J., et al. 2017, AJ, 153, 88
Liao, K., Treu, T., Marshall, P., et al. 2015, ApJ, 800, 11
Lochner, M., McEwen, J. D., Peiris, H. V., Lahav, O., & Winter, M. K. 2016, ApJS, 225, 31
Lochner, M., Scolnic, D. M., Awan, H., et al. 2018, arXiv:1812.00515
Lotz, J. M., Koekemoer, A., Coe, D., et al. 2017, ApJ, 837, 97
LSST Science Collaboration, Marshall, P., Anguita, T., et al. 2017, arXiv:, 1708.04058
Mandelbaum, R., Blazek, J., Chisari, N. E., et al. 2019, BAAS, 51, 363
McDowell, J. C. 2020, ApJL, 892, L36
Mead, A. J., Brieden, S., Tröster, T., & Heymans, C. 2021, MNRAS, 502, 1401
Merson, A. I., Baugh, C. M., Helly, J. C., et al. 2013, MNRAS, 429, 556
Mirhosseini, A., & Moniez, M. 2018, A&A, 618, L4
Möller, A., & de Boissière, T. 2019, MNRAS, 491, 4277
Mortlock, D. J., Feeney, S. M., Peiris, H. V., Williamson, A. R., & Nissanke, S. M. 2019, PhRvD, 100, 103523
Mörtsell, E., Johansson, J., Dhawan, S., et al. 2020, MNRAS, 496, 3270
Naghib, E., Yoachim, P., Vanderbei, R. J., Connolly, A. J., & Jones, R. L. 2019, AJ, 157, 151
Nissanke, S., Holz, D. E., Dalal, N., et al. 2013, arXiv:1307.2638
Nissanke, S., Holz, D. E., Hughes, S. A., Dalal, N., & Sievers, J. L. 2010, ApJ, 725, 496
Nordin, J., Rubin, D., Richard, J., et al. 2014, MNRAS, 440, 2742
Oguri, M., & Marshall, P. J. 2010, MNRAS, 405, 2579
Olsen, K., Di Criscienzo, M., Jones, R. L., et al. 2018, arXiv:1812.02204
Paczynski, B. 1986, ApJ, 304, 1
Patel, B., McCully, C., Jha, S. W., et al. 2014, ApJ, 786, 9
Perrett, K., Sullivan, M., Conley, A., et al. 2012, AJ, 144, 59
Petrushevska, T. 2020, Symmetry, 12, 1966
Petrushevska, T., Amanullah, R., Goobar, A., et al. 2016, A&A, 594, A54
Petrushevska, T., Goobar, A., Lagattuta, D. J., et al. 2018a, A&A, 614, A103
Petrushevska, T., Okamura, T., Kawamata, R., et al. 2018b, ARep, 62, 917
Planck Collaboration 2020, A&A, 641, A6
Refsdal, S. 1964, MNRAS, 128, 307
Ridgway, S., Cook, K., Miller, M., et al. 2010, Proc. SPIE, 7737, 77370Z
Riess, A. G., Casertano, S., Yuan, W., et al. 2021, ApJL, 908, L6
Rodney, S. A., Brammer, G. B., Pierel, J. D. R., et al. 2021, NatAs, 5, 1118
Rodney, S. A., Patel, B., Scolnic, D., et al. 2015, ApJ, 811, 70
Rothchild, D., Stubbs, C., & Yoachim, P. 2019, PASP, 131, 115002
Rowe, B. 2010, MNRAS, 404, 350
Schlafly, E. F., & Finkbeiner, D. P. 2011, ApJ, 737, 103
Scolnic, D., Kessler, R., Brout, D., et al. 2018a, ApJL, 852, L3
Scolnic, D. M., Jones, D. O., Rest, A., et al. 2018b, ApJ, 859, 101
Scolnic, D. M., Lochner, M., Gris, P., et al. 2018c, arXiv:1812.00516
Setzer, C. N., Biswas, R., Peiris, H. V., et al. 2019, MNRAS, 499, 4260
Springel, V., White, S. D. M., Jenkins, A., et al. 2005, Natur, 435, 629
Stanishev, V., Goobar, A., Paech, K., et al. 2009, A&A, 507, 61
Suyu, S. H., Huber, S., Cañameras, R., et al. 2020, A&A, 644, A162
Suyu, S. H., Bonvin, V., Courbin, F., et al. 2017, MNRAS, 468, 2590
Swann, E., Sullivan, M., Carrick, J., et al. 2019, Msngr, 175, 58
Tewes, M., Courbin, F., Meylan, G., et al. 2013, A&A, 556, A22
The LSST Dark Energy Science Collaboration, Mandelbaum, R., Eifler, T., et al. 2018, arXiv:1809.01669
The PLAsTiCC team, Allam, T. J., Bahmanyar, A., et al. 2018, arXiv:1810.00001
Tyson, J. A., Ivezić, Ž., Bradshaw, A., et al. 2020, AJ, 160, 226
Verma, A., Collett, T., Smith, G. P. & Strong Lensing Science Collaboration, DESC Strong Lensing Science Working Group 2019, arXiv:1902.05141
Wojtak, R., Hjorth, J., & Gall, C. 2019, MNRAS, 487, 3342
Wong, K. C., Suyu, S. H., Chen, G. C. F., et al. 2020, MNRAS, 498, 1420