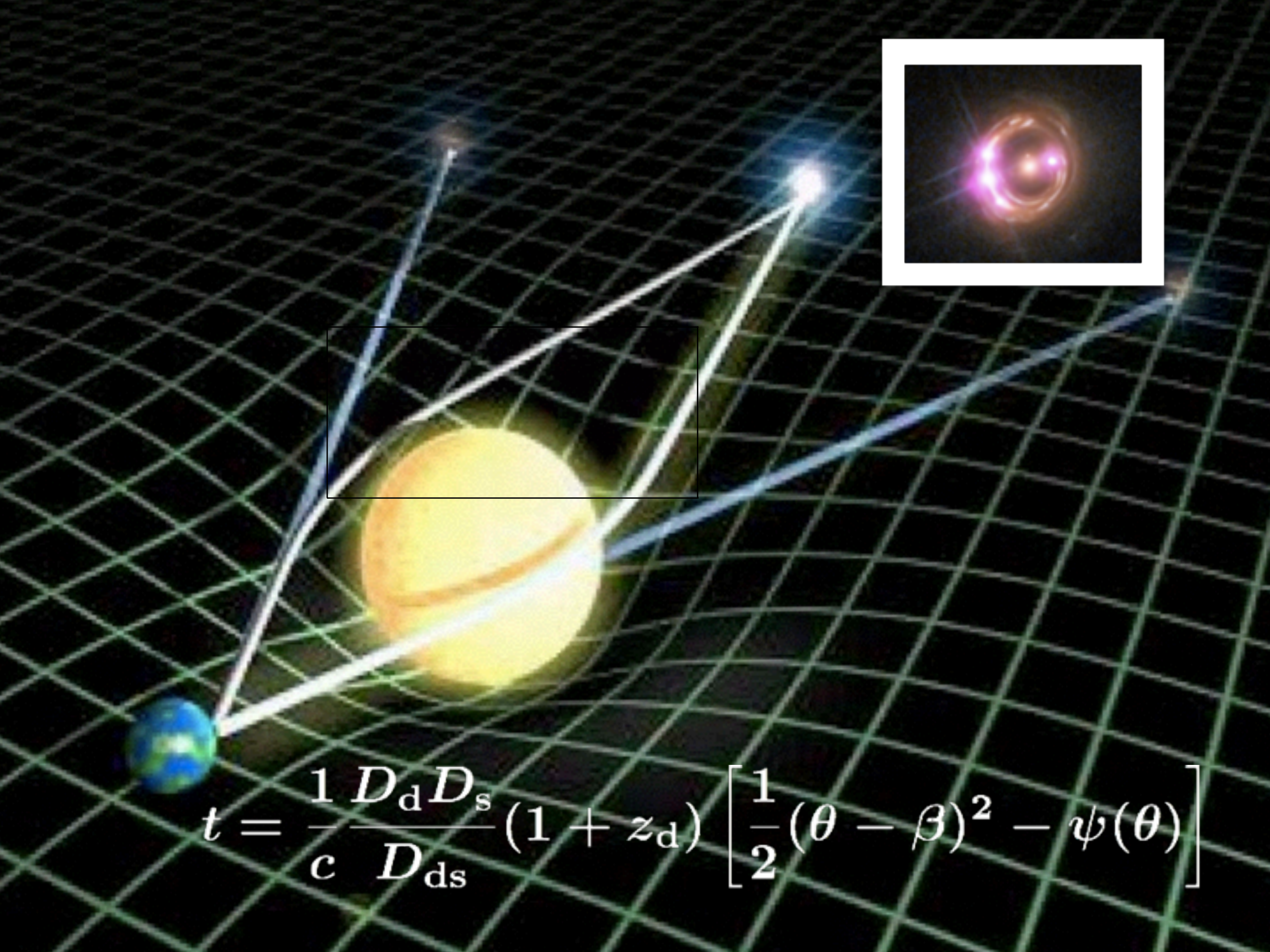
LSST Observing Strategy Metric Analysis Example:

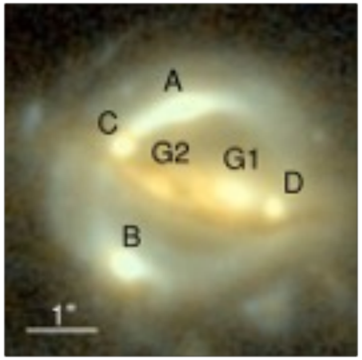
Strong Lens Time Delays

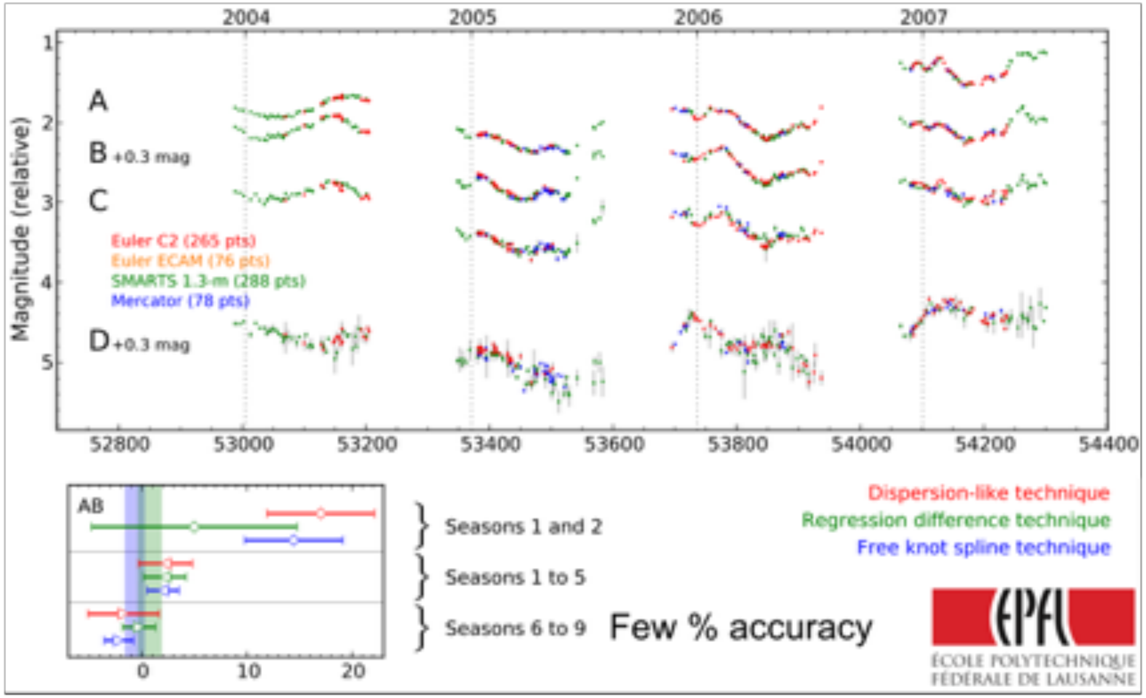
Phil Marshall (SLAC/KIPAC)



$$t = \frac{1}{c}\frac{D_{\mathrm{d}}D_{\mathrm{s}}}{D_{\mathrm{ds}}}(1+z_{\mathrm{d}})\left[\frac{1}{2}(\theta-\beta)^2-\psi(\theta)\right]$$

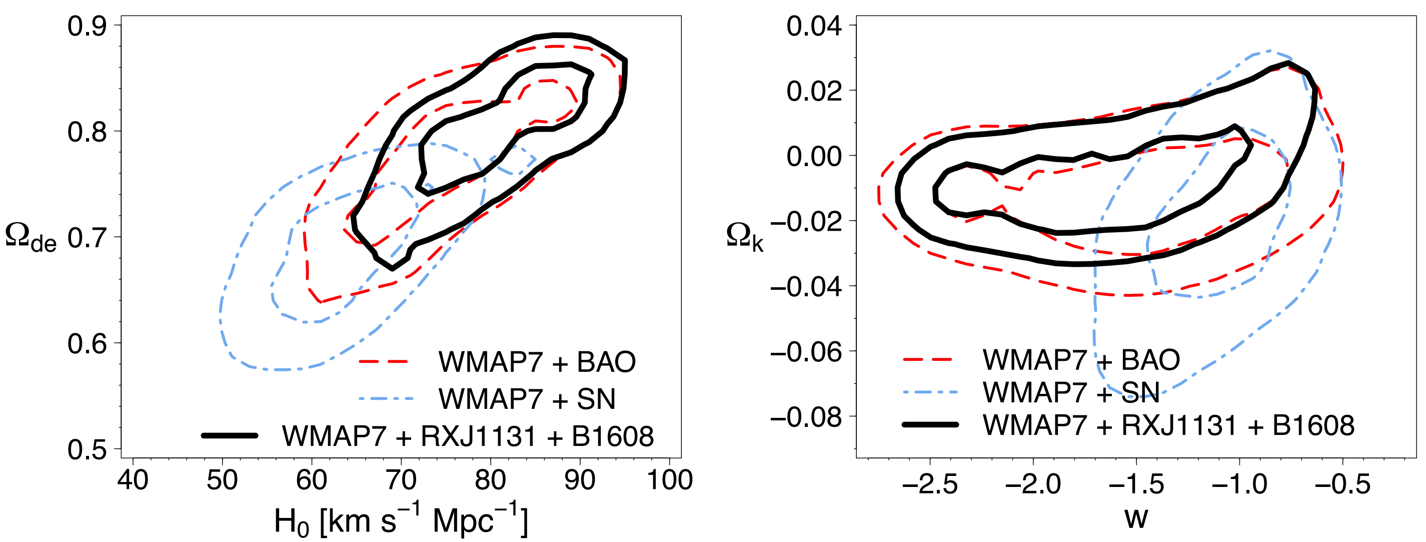
Time Delay Distances





- Strong lens time delay distances are one of the cosmological probes being investigated by the DESC

Time Delay Distances



- LSST will automatically monitor thousands of systems, randomly distributed over the sky
- We need accurate time delays and mass models (from follow-up data) for a sample of 10^{2-3} in order to measure H_0 to sub-% precision and help constrain Dark Energy models

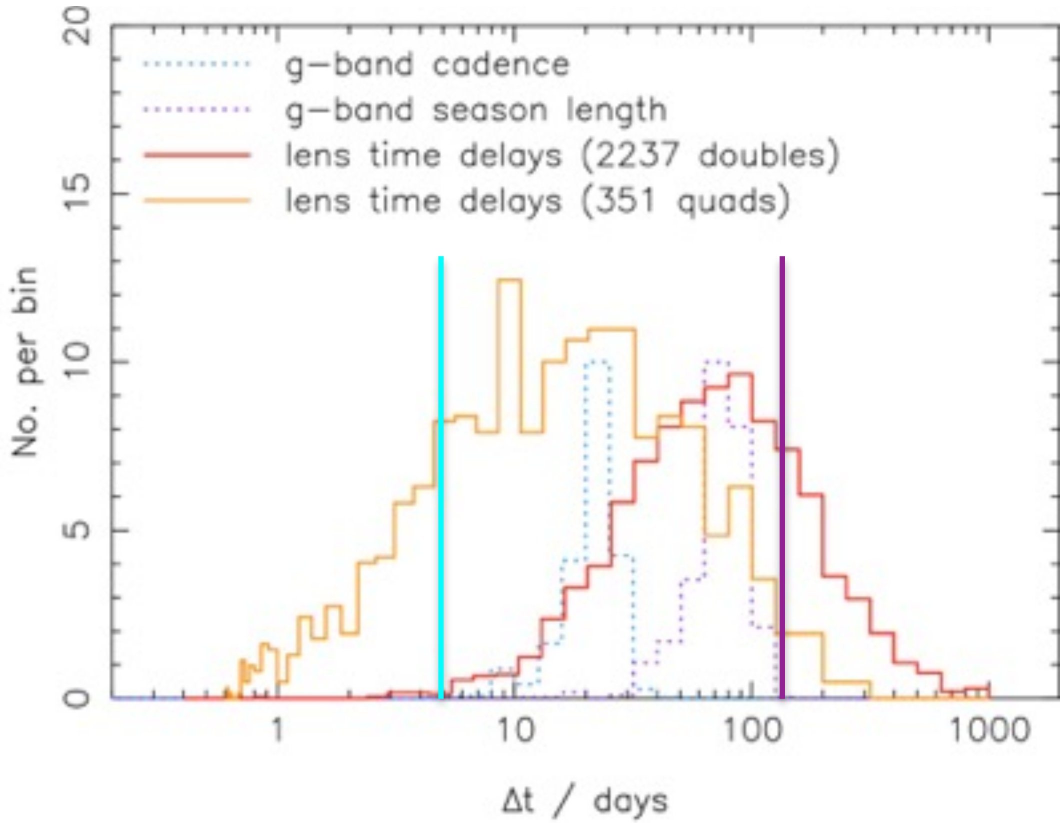
Time Delay Distances

- To measure each time delay distance to 5%, we need the time delay precision to be around 3% per lens
- To measure cosmological parameters with sub-percent accuracy, we need the mean systematic error to be below ~0.2%, and the sample to be larger than 100 lenses

1) How well will we be able to measure lensed quasar time delays with LSST, given the default “universal cadence” observing strategy?

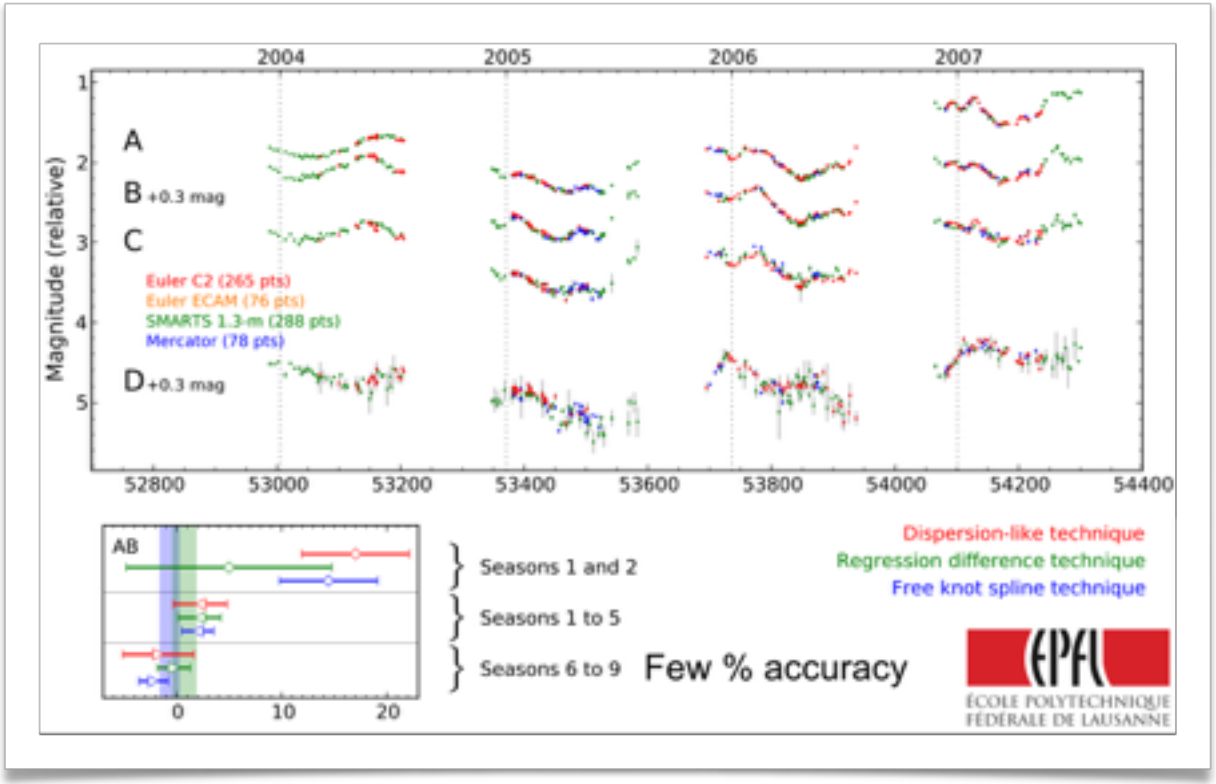
2) Which aspects of the LSST observing strategy would we, the DESC SL working group, change, and how?

Cadence and season length



Most lens time delays fall between a few days and a few months: we expect the night-to-night cadence and the season length to be important

Campaign length



Accuracy increases as the number of seasons builds up: campaign length is also important

Time Delay Challenge

Goals:
- Assess performance of time delay estimation algorithms on LSST-like data (cf STEP in WL community)
- Assess the impact of universal cadence strategy on time delay estimation, and possibly recommend changes

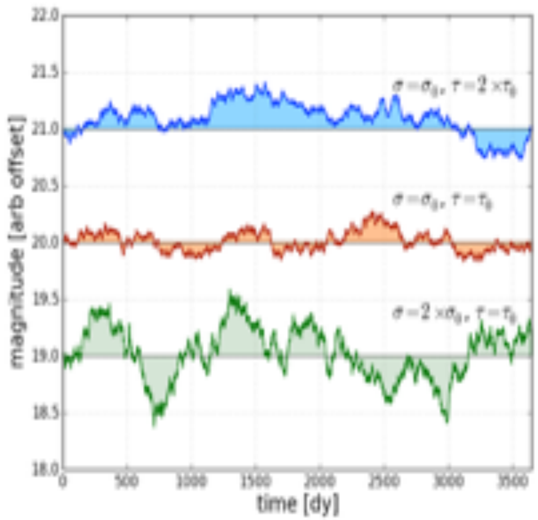
Plan:
- "Evil Team" generated a large set of simulated lightcurves spanning expectations for Stage II-IV
- Challenged wider community "Good Teams" to infer time delays blindly, and submit results

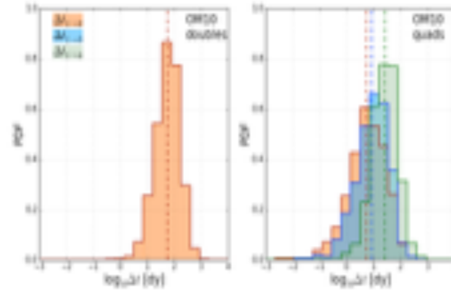
Evil Team for TDC1:
Kai Liao, Greg Dobler, Tommaso Treu, Chris Fassnacht, Nick Rumbaugh, Phil Marshall

TDC timeline

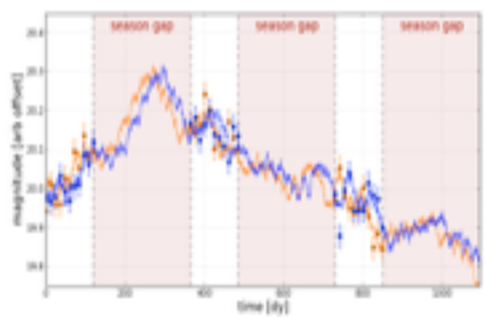
- Paper describing challenge: arXiv:1310.4830
- "TDC0" - small "training set," to get Good Teams started: released October 2013, soft deadline December 2013
- "TDC1" - large "test set," for primary analysis: released December 2013, hard deadline for submissions July 1 2014
- Robotic TDC0 feedback ongoing until then
- Evil Team analysed entries, wrote paper with Good Teams, submitted August 2014. ~1 year, not including sims R&D

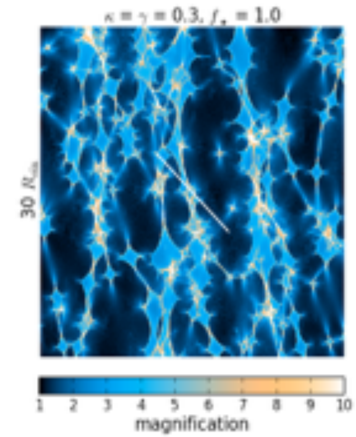
TDC1 ingredients

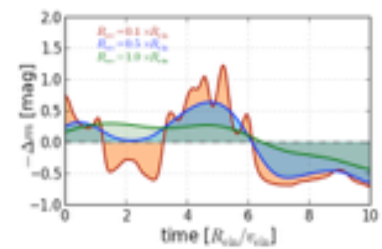




Dobler et al 2013





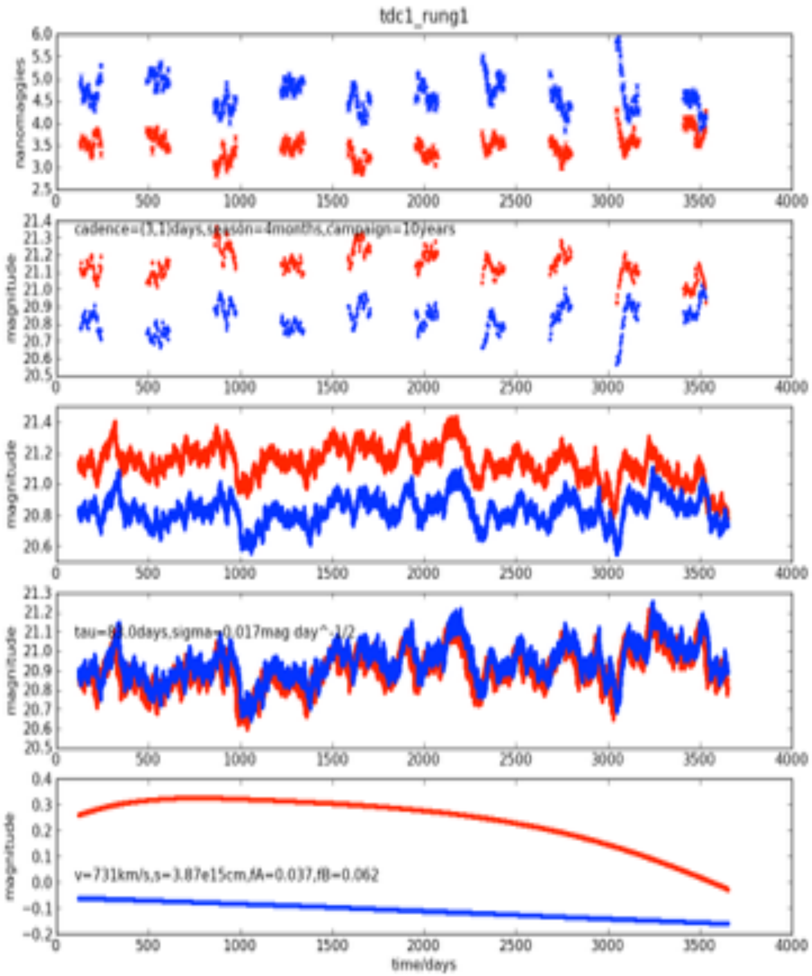


- Lightcurve catalog data only (single filter)
- Realistic AGN variability, microlensing, lenses
- “OpSim-ish” observations (depth, cadence, etc)

TDC1: example lightcurves

10 years, 3 day cadence

5 years, 3 days



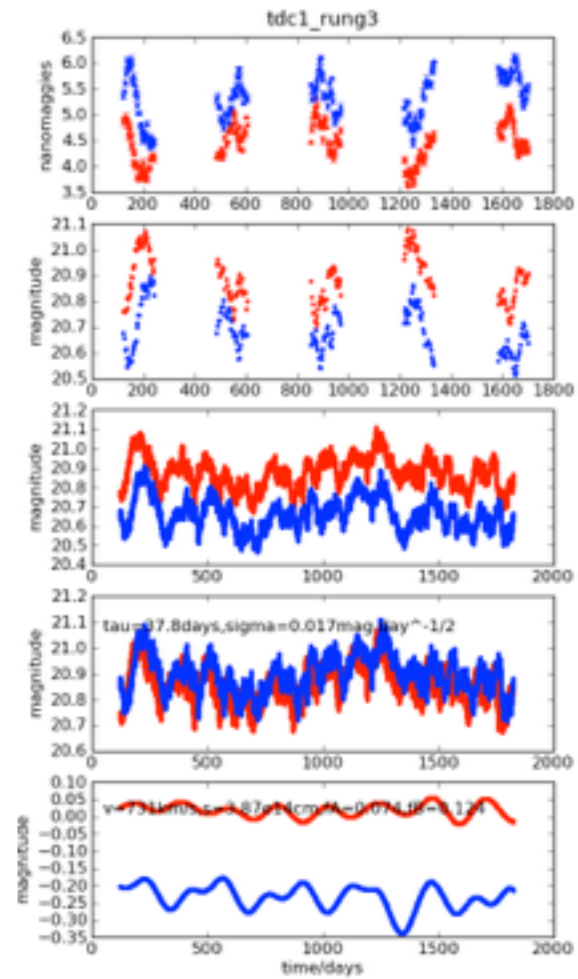
Mock data

- without noise

- fully sampled

- no lensing

- microlensing



Multiple TDC metrics

$$A = \frac{1}{fN} \sum_i \frac{\tilde{\Delta t}_i - \Delta t_i}{|\Delta t_i|}$$

$$P = \frac{1}{fN} \sum_i \left(\frac{\sigma_i}{|\Delta t_i|} \right)$$

$$\chi^2 = \frac{1}{fN} \sum_i \left(\frac{\tilde{\Delta t}_i - \Delta t_i}{\sigma_i} \right)^2$$

1. $f > 0.3$
2. $0.5 < \chi^2 < 2$ (1.09f)
3. $P < 15\%$ (3%)
4. $A < 15\%$ (0.2%)

- 4 metrics (chisq, A, P, f) used to define TDC0 pass
- No single metric, no leaderboard, no cosmology
- LSST/TDC1 requirements are stricter (in parentheses)

TDC1: challenge "rungs"

Rung	Cbar	Cerr	Season	Campaign	Nepochs	maglim	N_lens
0	3.0	1.0	8.0	5	400	24	1000
1	3.0	1.0	4.0	10	400	24	1000
2	3.0	0.0	4.0	5	200	24	1000
3	3.0	1.0	4.0	5	200	24	1000
4	6.0	1.0	4.0	10	200	24	1000

- Rungs enable simple A-B testing of LSST observing scenarios, *assuming that all filters can be used in joint analysis*
- Double (SNe) Cadence, all filters: cadence = 3 +/- 1 days
- Worst-case Universal Cadence, all filters: cadence = 6 +/- 1 days
- Double, and more uniform cadences might be possible with customization of observing strategy, as suggested by SNe group, who are happy to trade campaign length for cadence.

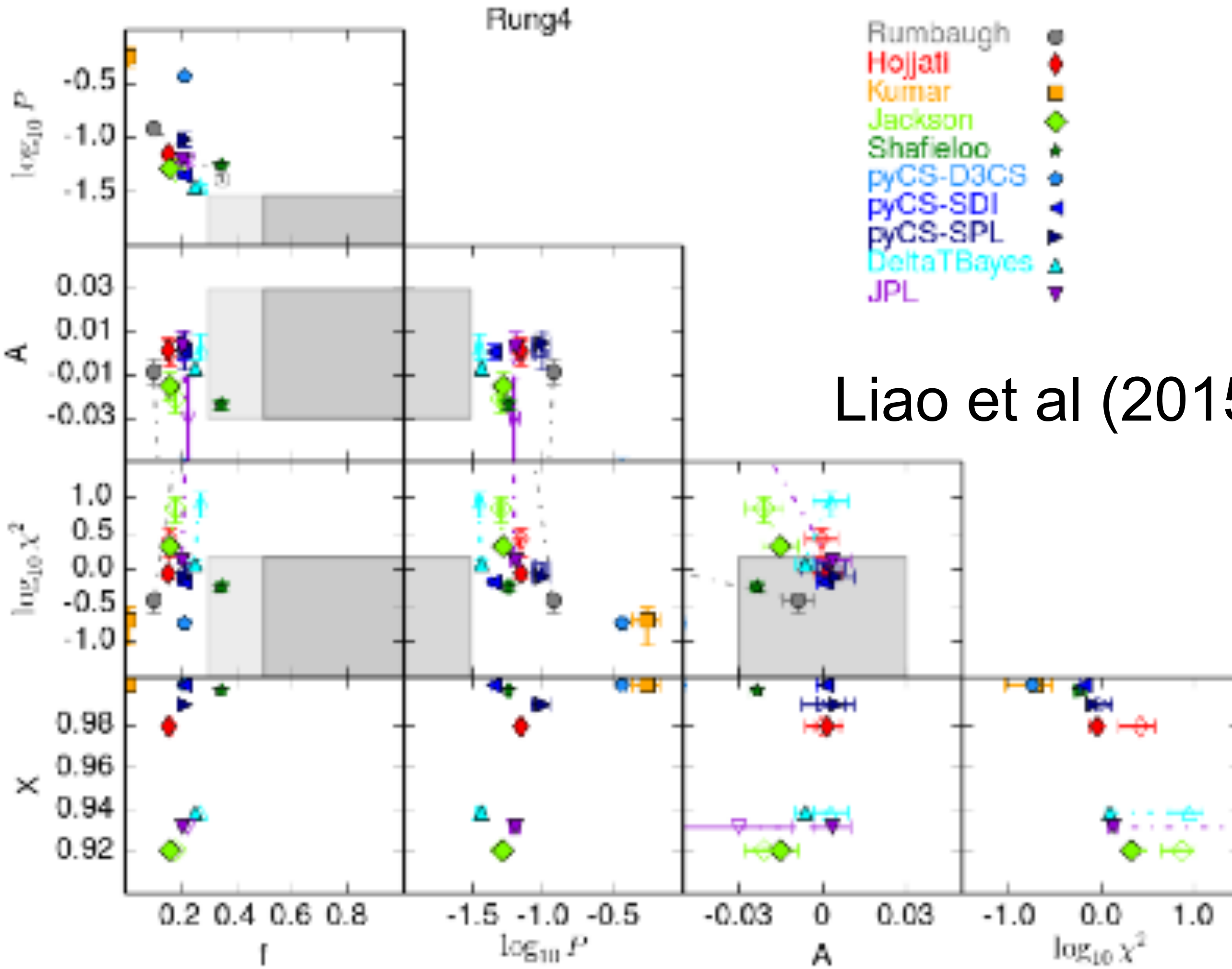
TDC1: results

7 Good Teams plus Evil Team, 10 methods (with variations)

Outliers a problem for some methods

Required high accuracy is achievable, even at universal cadence.

Fraction of useable systems is ~20%



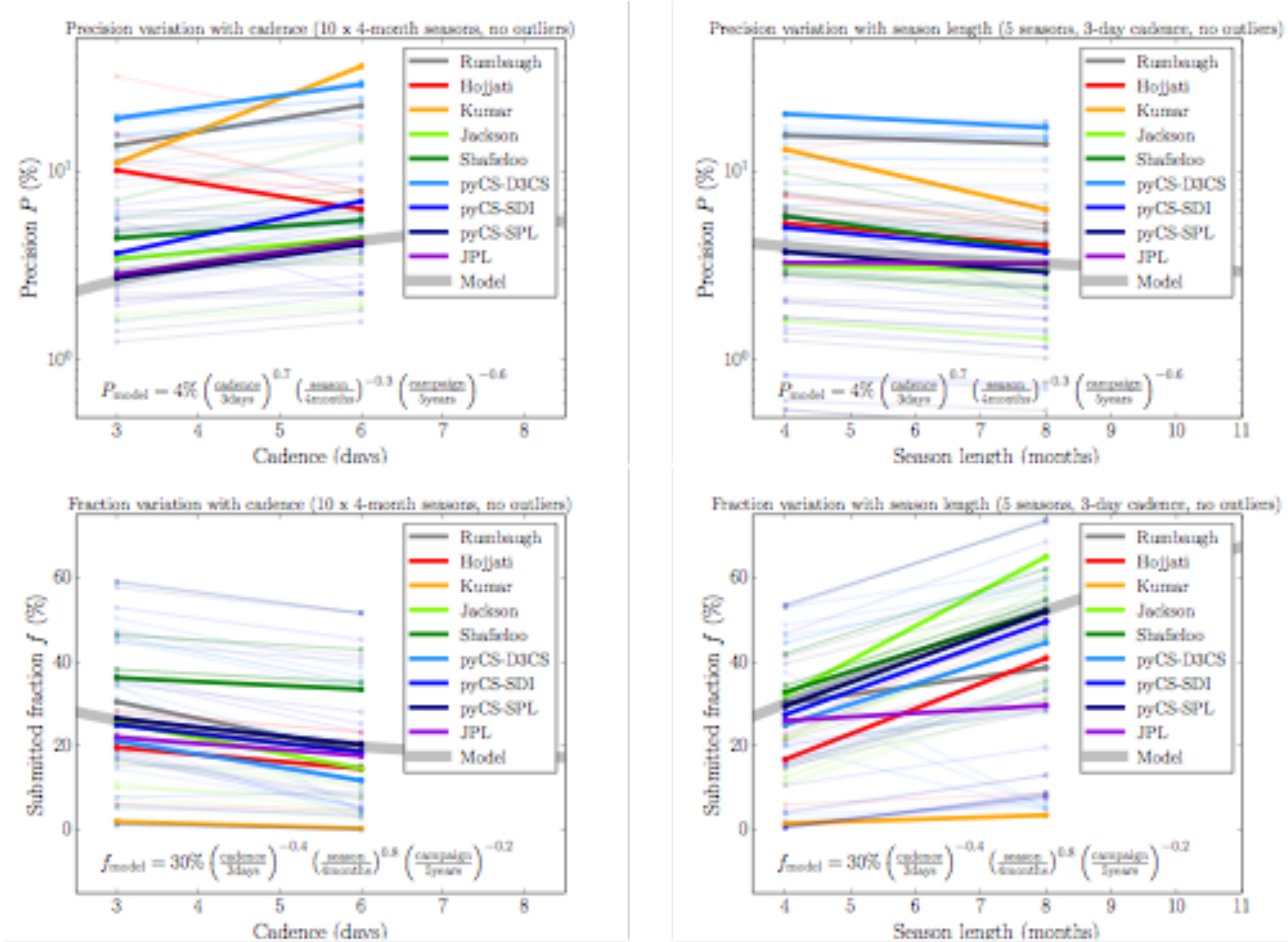
Liao et al (2015)

TDC1: cadence etc

Cadence drives precision.

Season length also matters for the useable fraction (rate).

Was accuracy prioritized by Good Teams? It depends much less on cadence and season length.

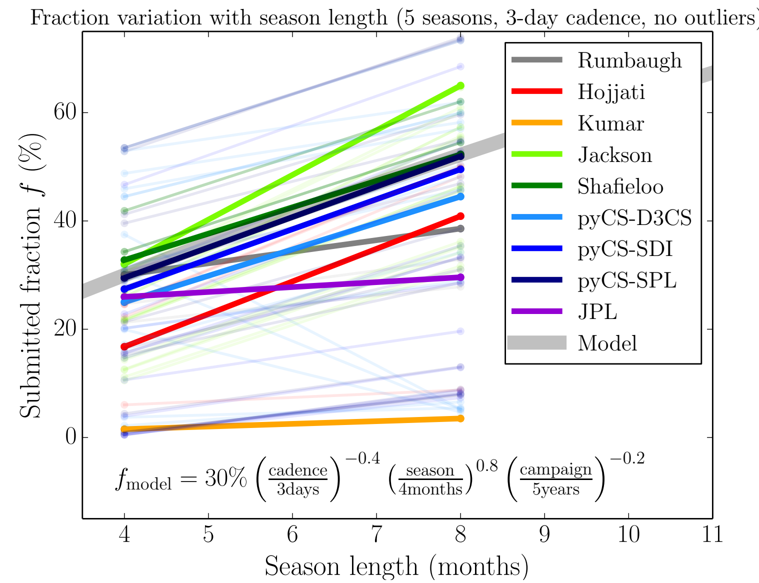


TDC1: results

- Analysis of 1000s of light curves in manageable time (6 months) was demonstrated by 7 Good Teams
- Mean accuracy (or bias) of 0.06% was achieved, independent of cadence, in LSST-like campaigns of just five 4-month seasons. Many methods yielded sub-percent accuracy.
- Low cadence and season length cause low precision and useable fraction. Mean precision predicted for the best method in a 5-day cadence, 10-season campaign is 3.8%. The useable fraction in this case is 20%, which corresponds to 400 lenses.

TDC model metrics

The TDC model approximately captures the dependence of our metrics on observing strategy:



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

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

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

MAF Analysis

mafContrib/tdcMetric.py

```
# =====================================================================
# Phil Marshall (drphilmarshall) <pjm@slac.stanford.edu>
# Lynne Jones (rhiannonlynne) <ljones@astro.washington.edu>
# =====================================================================

from lsst.sims.maf.metrics import BaseMetric
from .campaignLengthMetric import CampaignLengthMetric
from .seasonLengthMetric import SeasonLengthMetric
from .meanNightSeparationMetric import MeanNightSeparationMetric
```

- The TDC “cadence”, “season length” and “campaign length” are *metrics* in the MAF sense
- They *summarize* the LSST observations of a particular point on the sky, and our time delay model metrics depend on them

MAF Analysis

mafContrib/tdcMetric.py

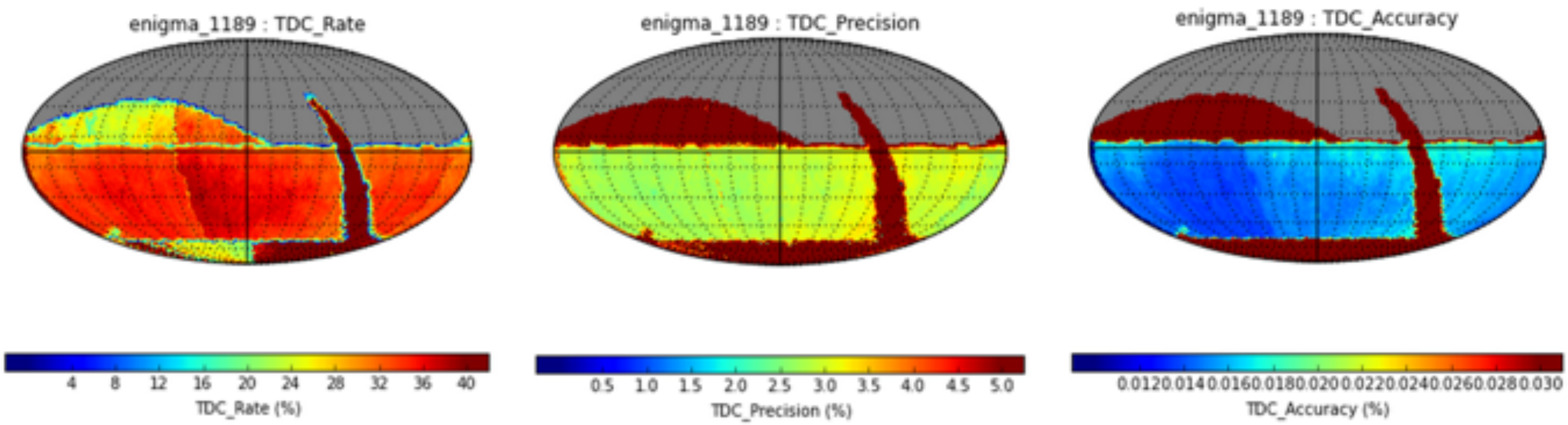
```
class TdcMetric(BaseMetric):

    def run(self, dataSlice, slicePoint=None):
        # Calculate accuracy from combined individual metrics.
        camp = self.campaignLength.run(dataSlice)
        sea = self.seasonLength.run(dataSlice)
        cad = self.meanNightSeparation.run(dataSlice)
        if sea * cad * camp == 0:
            accuracy = self.badval
            precision = self.badval
            rate = 0.0
        else:
            accuracy = 0.06 * (self.seaNorm / sea) * (self.campNorm / camp)**(1.1)
            precision = 4.0 * (cad/self.cadNorm)**(0.7) * (self.seaNorm/sea)**(0.3) * (self.campNorm/camp)**(0.6)
            rate = 30. * (self.cadNorm/cad)**(0.4) * (sea/self.seaNorm)**(0.8) * (self.campNorm/camp)**(0.2)
        return {'accuracy':accuracy, 'precision':precision, 'rate':rate}
```

- TDC model accuracy, precision and rate are components of the “complex” MAF metric “TdcMetric”
- Now we can look at, for example, how each of these metrics varies across the sky

MAF Analysis

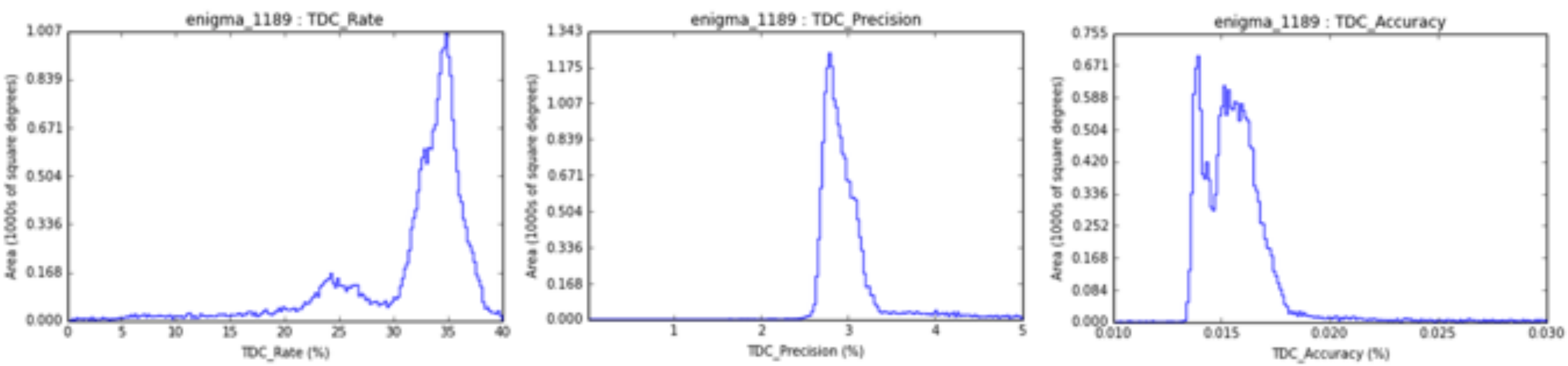
mafContrib/tdcMetric.py



- Main survey area is remarkably uniform in all metrics after 10 years; other survey areas likely not useable

MAF Analysis

mafContrib/tdcMetric.py



- Precision of 3% at accuracy less than 0.02% accuracy is very promising for time delay lens cosmology
- Rates of 30% may correspond to 600 useable lenses - but integral has to be done more carefully
- “Opsim-ish” assumptions in TDC were pessimistic?

Final Thoughts

- Caveat: *optimal combination of all filters* has been assumed so far. TDC2 planned to test this, but more conservative metric plots could be made. Only use *gri*?

- Approach involves some heavy *summarizing:* is it necessary to relax this, and simulate mock surveys? To cope with "rolling cadence"?

- Where might we be after 3 years? 5 years? Early science?

- Time delay accuracy, precision and rate are still intermediate metrics: the number N of useable lenses would be good, and a *Dark Energy figure of merit W* even better.

- Strategy: go back to TDC1 results, and look at N and W for each method, as a function of observing strategy. Extend model, and then upgrade TdcMetric.py