PS1 3π as a pilot survey for panoptic time-domain science

Nina Hernitschek, MPIA Heidelberg

collaborators:
Eddie Schlafly, Branimir Sesar, Hans-Walter Rix, Laura Inno,
Željko Ivezić, David Hogg

LSST@Europe 2, June 20-24, Belgrade

Why a pilot survey?

Does LSST need a pilot survey? Does PS1 3π serve as a pilot survey?

LSST is deep, wide & fast

$r\sim24.5$ (single)
$r\sim27.5$ (coadd.)

20,000 deg^2

15 - 60 min pairs

$\Rightarrow$ doing science with preliminary LSST

Why PS1 3π as pilot survey?

Why a pilot survey?

Does LSST need a pilot survey? Does PS1 3π serve as a pilot survey?

LSST is deep, wide & fast

deep	wide	fast
$r \sim 24.5$ (single) $r \sim 27.5$ (coadd.)	20,000 deg^2	15 - 60 min pairs
$r \sim 21.8$ (single) $r \sim 23.2$ (coadd.)	31,000 deg^2	14 min pairs

$\Rightarrow$ doing science with preliminary LSST

Why PS1 3π as pilot survey?
multiband, sparse time sampling, covers 3/4 of the sky

$\Rightarrow$ testing ground for various modeling approaches

Why a pilot survey?

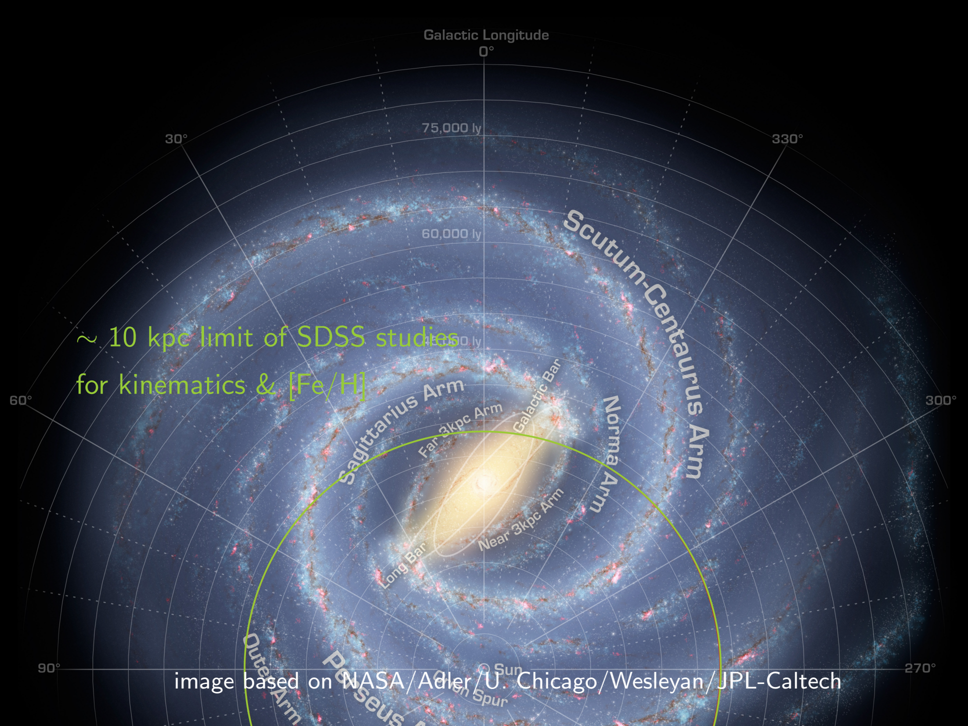
PS1 3π in one sentence: An optical/near-IR survey of 3/4 of the sky in *grizy* to r~23.2 based on ~65 visits over a 3.5-year period.

LSST in one sentence: An optical/near-IR survey of half the sky in *ugrizy* to r~27.5 based on 1000 visits over a 10-year period.

map galactic halo to ~120 kpc/ 400 kpc (~ virial radius of MW)
single-visit depth of $r \sim 21.8$/ $r \sim 24.5$
coadded depth of $r \sim 23.2$ / $r \sim 27.5$
sky coverage of ~31,000/ ~20,000 deg^2
$\delta < 30$ deg / $\delta < 34.5$ deg
65 epochs over 3.5 years/ 1000 over 10 years

grizy nonsimultaneous / *ugrizy* nonsimultaneous

(*u* extremely powerful for separating low-redshift QSOs from hot stars $\Rightarrow$ test variability or color-variability selection with PS1 3π)



$\sim$ 10 kpc limit of SDSS studies

for kinematics & [Fe/H]

image based on NASA/Adler/U. Chicago/Wesleyan/JPL-Caltech

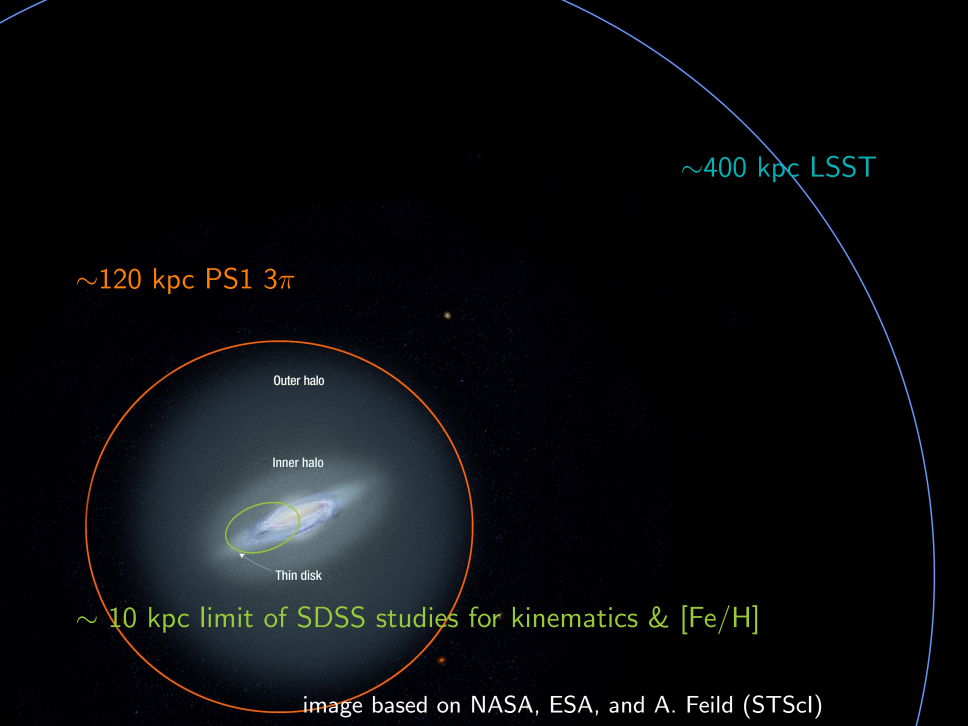


image based on NASA, ESA, and A. Feild (STScI)

Pan-STARRS 1 as a Time Domain Survey

most ambitious panoptic multi-epoch multi-band survey to date:

- solar system objects
- transients
- proper motions (& parallaxes)
- **variable sources**

Pan-STARRS 1 as a Time Domain Survey

most ambitious panoptic multi-epoch multi-band survey to date:

- solar system objects
- transients
- proper motions (& parallaxes)
- **variable sources**

QSOs

- intrinsically interesting, and astrometric reference points
- varying on weeks to years time-scales

Pan-STARRS 1 as a Time Domain Survey

most ambitious panoptic multi-epoch multi-band survey to date:

- solar system objects
- transients
- proper motions (& parallaxes)
- **variable sources**

QSOs

- intrinsically interesting, and astrometric reference points
- varying on weeks to years time-scales

RR Lyrae:

- precision 3D mapping of the (old) Milky Way
- varying on 1/4 day timescales

Pan-STARRS 1 as a Time Domain Survey

most ambitious panoptic multi-epoch multi-band survey to date:

- solar system objects
- transients
- proper motions (& parallaxes)
- **variable sources**

QSOs

- intrinsically interesting, and astrometric reference points
- varying on weeks to years time-scales

RR Lyrae:

- precision 3D mapping of the (old) Milky Way
- varying on 1/4 day timescales

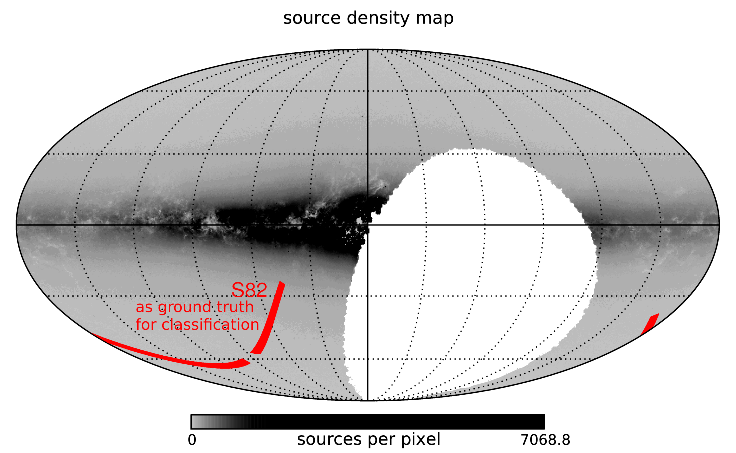
Cepheids:

- map extent of Milky Way's disk
- varying on days time-scales

Pan-STARRS (Panoramic Survey Telescope & Rapid Response System) Kaiser et al. (2010)

panoramic:

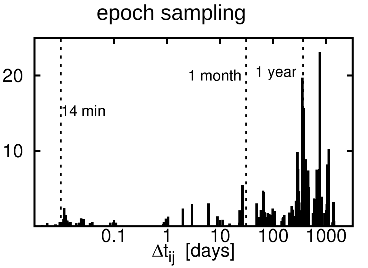
PS1 3π surveyed 3/4 of the sky (DEC > -30 deg) in *grizy*

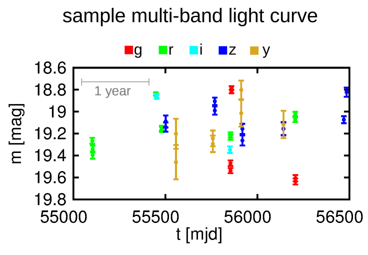


Pan-STARRS (Panoramic Survey Telescope & Rapid Response System) Kaiser et al. (2010)

rapid: time-domain survey

- ~35 (~65) epochs between 2010 and 2014 for PV2 (PV3)
- 5 bands (*grizy*)
- non-simultaneous





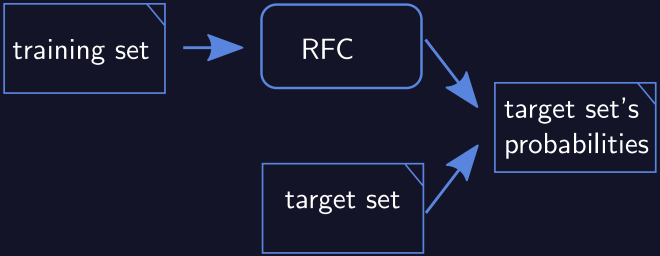
"Astrostatistics" with PS1 3π

"all-sky" time domain surveys looking for variable sources

- outlier detection **(machine learning)**
- variability characterization
- source classification **(machine learning)**
- period fitting

Astrostatistics with PS1 3π

for all machine learning approaches, we use **supervised learning**



training set: set of sources inside/outside set we are looking for

Random Forest Classifier

- provides binary classification
- training and classification can be parallelized
- implemented in Python `sklearn` package

Outlier detection

photometric outliers cause spurious variability

old solution:

- flags (e.g. blending, off chip)
- (psf-aperture) magnitude cuts (morphology)

Outlier detection

photometric outliers cause spurious variability

old solution:

- flags (e.g. blending, off chip)
- (psf-aperture) magnitude cuts (morphology)

machine learning approach (*Sesar et al. 2016 in prep.*):

- training set of non-varying sources (bright K and G stars) $\Rightarrow$ in/outliers
- train classifier on features (e.g. flags, morphology, CCD position, filter)
- feature importance can be used for understanding what causes outliers

$\Rightarrow$ works $3\times$ better:
machine learning approach: 80% recovery rate (i.e. completeness), miss-classification rate of 1
old solution: $\sim$50% recovery rate, miss-classification rate of 3

PS1 as a Time-Domain Survey

PS1 3π interesting because of its size $\Rightarrow$ "all-sky" time domain astronomy

how to characterize variability statistically?

forget about simple single-band models!

PS1 as a Time-Domain Survey

PS1 3π interesting because of its size $\Rightarrow$ "all-sky" time domain astronomy

how to characterize variability statistically?

forget about simple single-band models!
$\Rightarrow$ **solution: model non-simultaneous multi-band variability**

PS1 as a Time-Domain Survey

PS1 3π interesting because of its size $\Rightarrow$ "all-sky" time domain astronomy

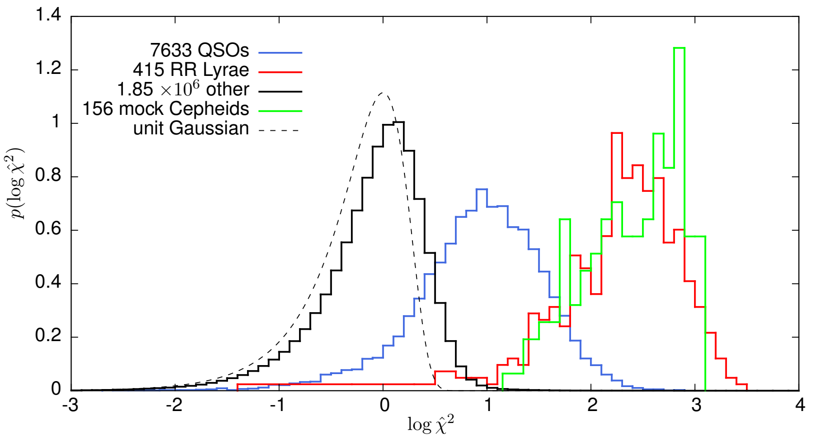
how to characterize variability statistically?

forget about simple single-band models!
$\Rightarrow$ **solution: model non-simultaneous multi-band variability**
$\Rightarrow$ **generic: other variables (min sampling $\lesssim$ time scale $\lesssim$ survey duration)**

Which sources vary at all?

multi-band χ^2 statistics for PS1 photometry, assuming non-varying sources

SDSS S82 (RR Lyrae, QSOs, "other"), mock (Cepheids)



Characterize Light Curves

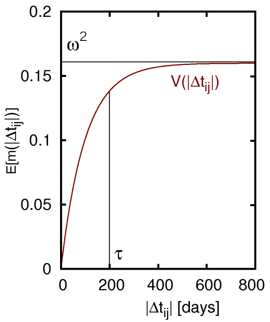
multi-band structure-function variability model $\mathscr{L}(grizy|\omega_\mathbf{r},\tau)$:
how much should you expect a source to vary within Δt?

$$\mathrm{V}(|\Delta t|) \equiv \mathrm{E}[(m(t) - m(t+\Delta t))^2]$$

assume functional form

$$V(\Delta t) \overset{\text{model}}{\equiv} \omega_i(\lambda_i)\omega_j(\lambda_j)\left(1 - \exp\left[-\frac{|\Delta t|}{\tau}\right]\right)$$

with
$\tilde{m}_\lambda(t) = m_\lambda(t) - \bar{m}_\lambda$, $\omega_k(\lambda_k) = \omega_r\left(\frac{\lambda_k}{\lambda_r}\right)^\alpha$

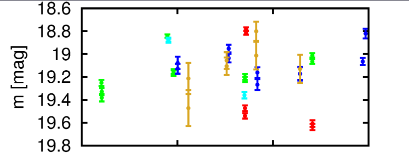


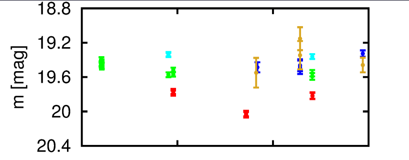
$\Rightarrow$ fit amplitude ω_λ, variability time-scale τ & $\bar{m}_\lambda$
$\Rightarrow$ characteristic variability timescale & amplitude

Characterize Light Curves

multi-band structure-function variability model $\mathscr{L}(grizy|\omega_{\mathbf{r}}, \tau)$:
how much should you expect a source to vary within Δt?

$\Rightarrow$ fit (ω_λ, τ) & $\bar{m}_\lambda$
$\Rightarrow$ characteristic variability timescale & amplitude

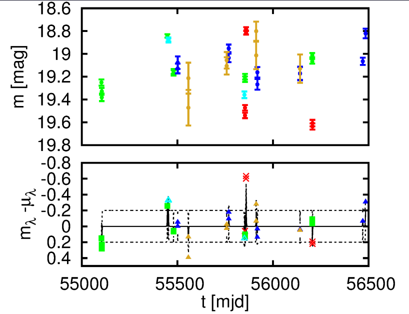




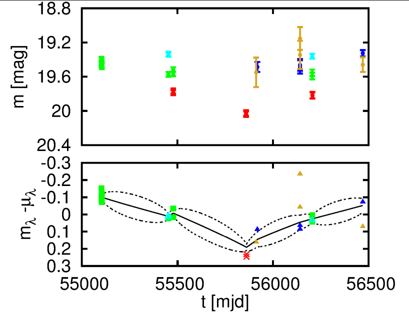
Characterize Light Curves

multi-band structure-function variability model $\mathscr{L}(grizy|\omega_{\mathbf{r}}, \tau)$:
how much should you expect a source to vary within Δt?

$\Rightarrow$ fit (ω_λ, τ) & $\bar{m}_\lambda$
$\Rightarrow$ characteristic variability timescale & amplitude



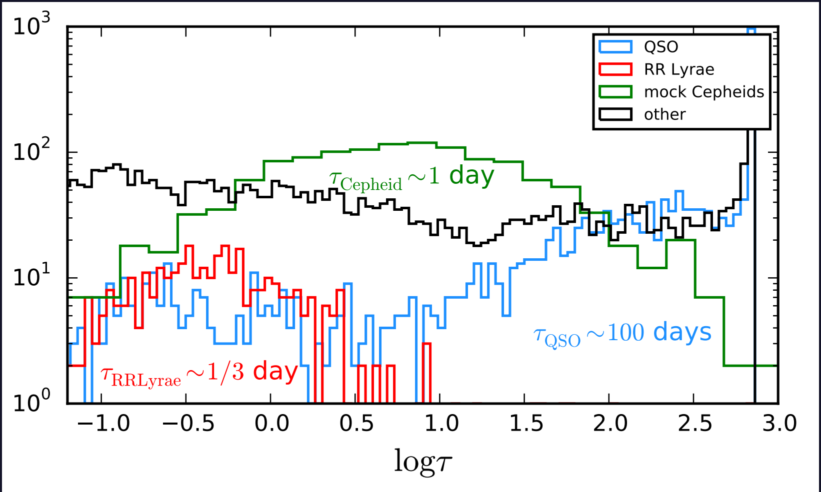
RR Lyrae, ω_r=0.3, τ=1.5 days



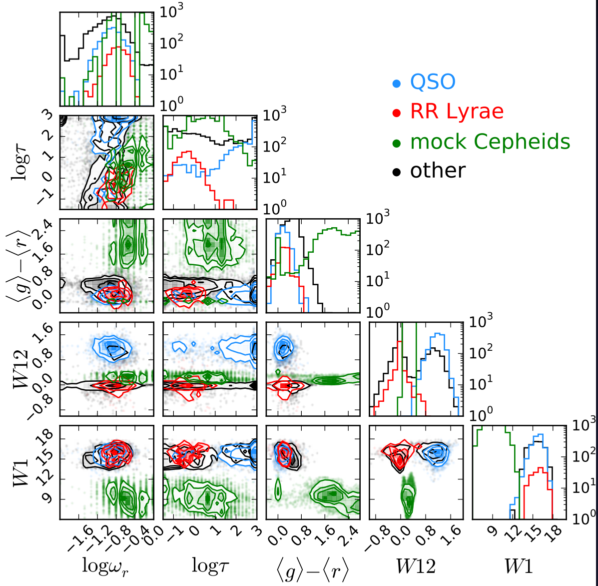
QSO, ω_r=0.13 , τ=560 days

Multi-Band Structure Functions

time-scale variability



Parameter Space



Classifying Variable Objects

How much can variation parameters & mean photometry tell us about classifications?

parameters $\Rightarrow$ classification probabilities

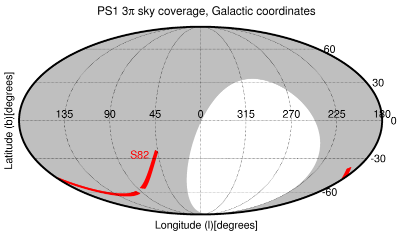
Classifying Variable Objects

How much can variation parameters & mean photometry (also from SDSS and WISE) tell us about classifications?

parameters $\Rightarrow$ classification probabilities

Approach:

- use SDSS Stripe 82 classification in overlapping area as ground truth for RR Lyrae & QSOs
- use mock Cepheid lightcurves at expected reddening
- train RFC, apply to PS1 3π



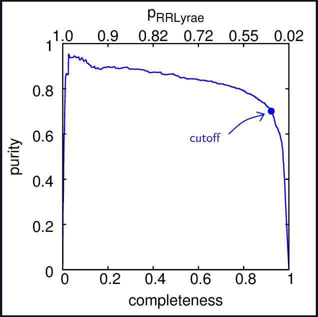
SDSS S82:

- $\sim$ 60 epochs simultaneous *ugriz*
- complete QSO and RR Lyrae classification

Classifying Variable Sources with PS1 3π Data

(I) classify sources using structure function (ω_r, τ), $\hat{\chi}^2$, colors
(II) for periodic variables: get period for likely candidates (template fitting)

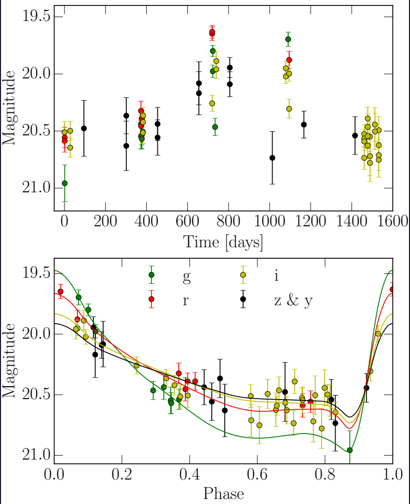
How well does this classification work?
We "know" the answer in SDSS S82.



completeness:
\# selected true RR Lyrae / \# true RR Lyrae

purity:
\# selected true RR Lyrae / \# all selected sources

(II) Period Fitting of RR Lyrae Candidates

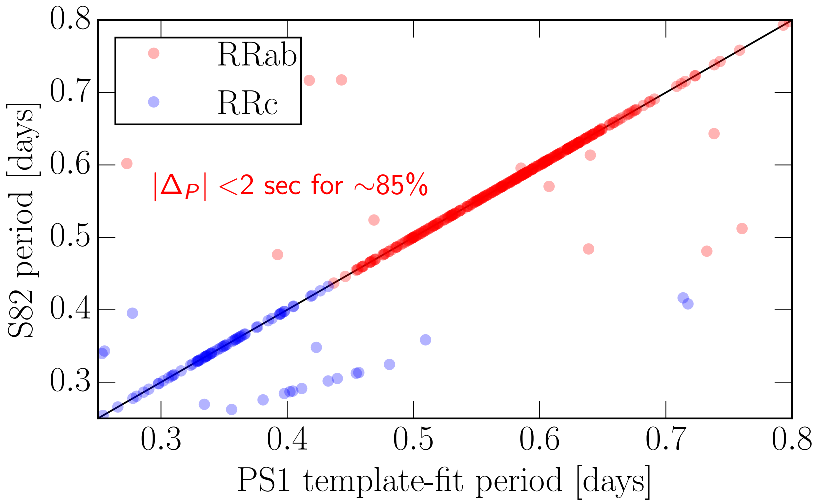


period fitting, using light curve templates from S82 (Sesar et al.)

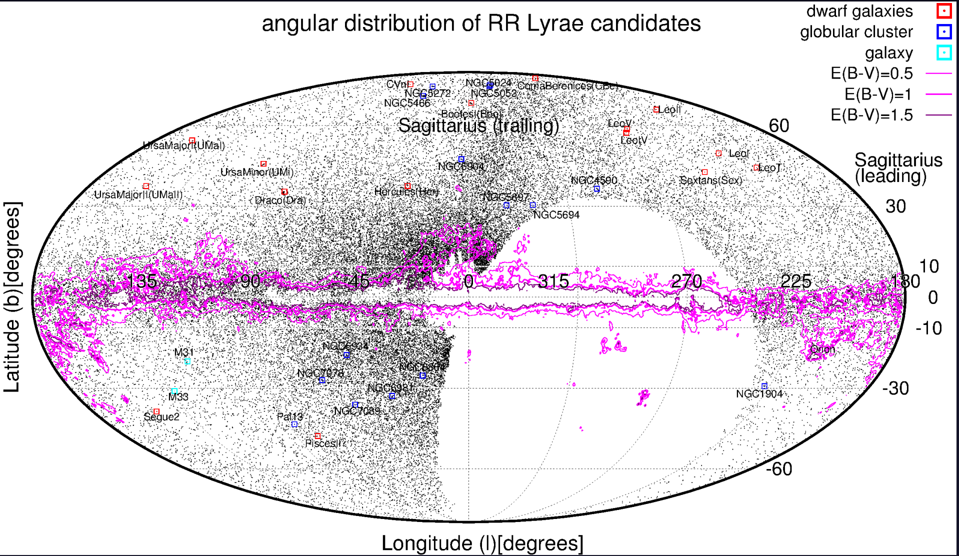
fit period of likely candidates from (I)

- Differential Evolution algorithm to find the optimal period, phase offset, *r*-band magnitude, and multi-band template
- recover periods to 1 sec. & measure distance within 5%
- period fitting improves from 77% purity, 75% completeness to 90% purity, 70% completeness in S82

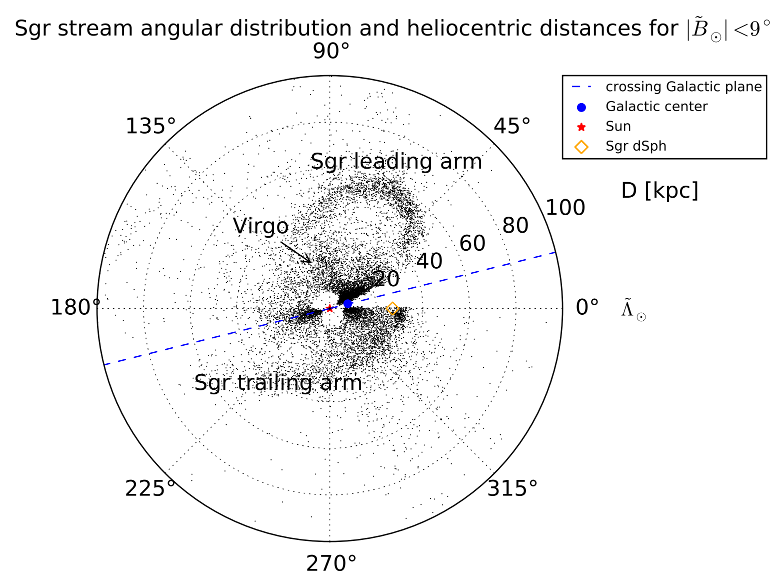
(II) Period Fitting of RR Lyrae Candidates



RR Lyrae Candidates



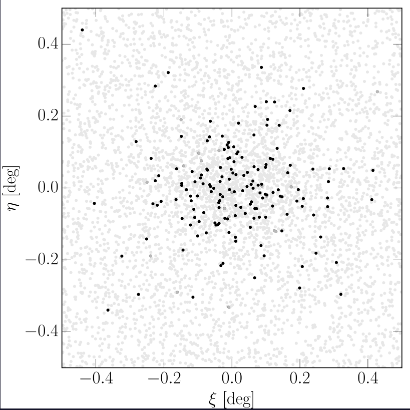
Sagittarius stream: an example

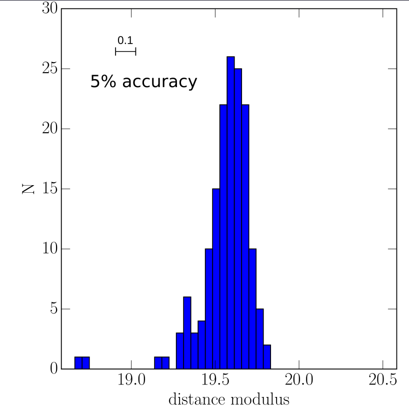


Draco dSph galaxy: an example

RR Lyrae candidates selected using classification by color & structure function parameters, passing period fitting ($\sim$70% completeness, 90% purity)

grey: all PS1 stars with $18 < r < 21.5$



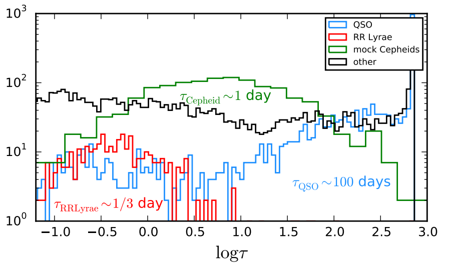


QSOs

QSO selection by colors & variability

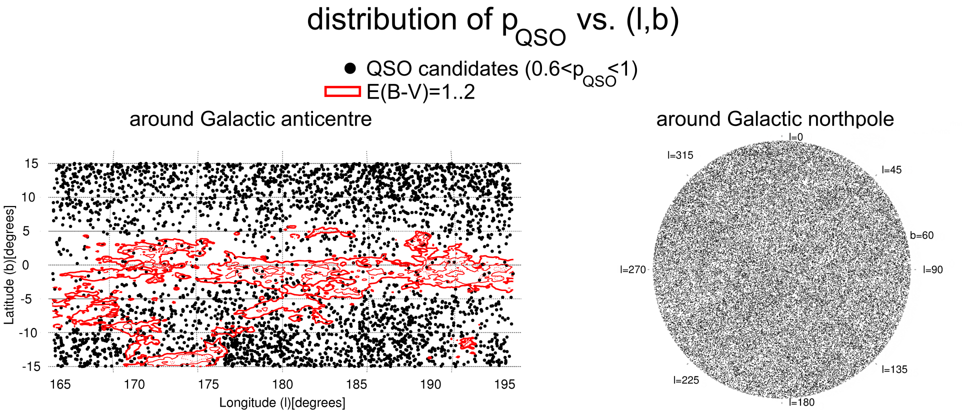
LSST has *u* band (unlike PS1 3π), but will benefit from PS1 3π:

- LSST is (like PS1 3π) nonsimultaneous $\Rightarrow$ effect on colors
- low luminosity QSOs: color contaminated by host galaxy
  $\Rightarrow$ variability selection



$\Rightarrow$ variability approach for non-simultaneous lightcurves

QSO Candidates



- QSO: in S82, 85% purity, 85% completeness for $p_{\mathrm{QSO}} > 0.5$
- plausible number densities, plausible area distributions:
- $\sim$ const. area density for QSO candidates ($\sim 20/\mathrm{deg}^2$)

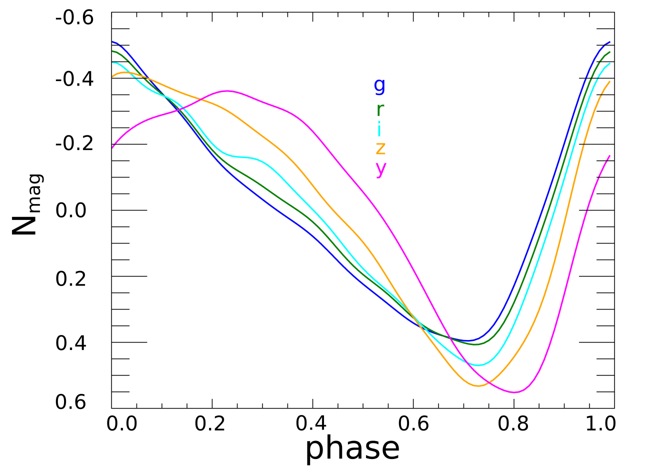
Cepheids

(I) classify sources using structure function (ω_r, τ), $\hat{\chi}^2$, colors from PS1 3π and WISE
(II) likely candidates: get period (template fitting, by Laura Inno)

known MW Cepheids too bright for PS1 3π $\Rightarrow$ synthetic light curves as training set

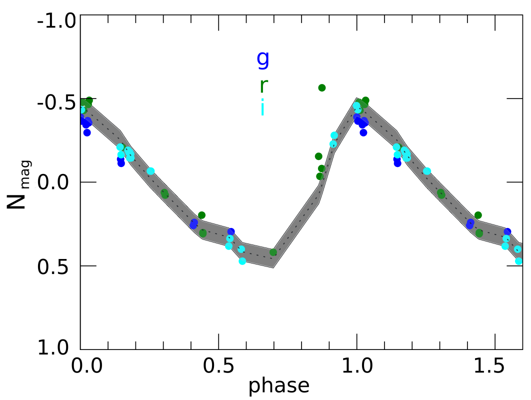
(II) Period Fitting of Cepheid Candidates

example: *grizy* templates for 5 days < P ⩽7 days



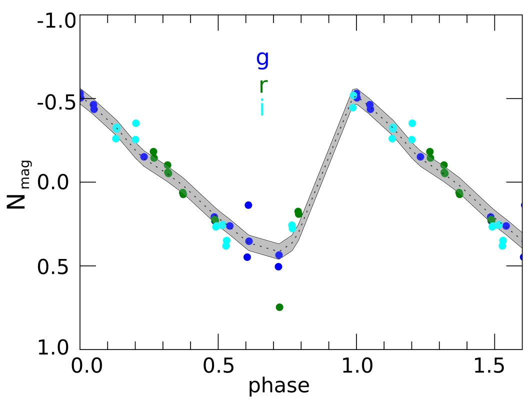
(II) Period Fitting of Cepheid Candidates

two example fits for PS1 3π Cepheids:
a distant Galactic Cepheid (D=12.1 kpc, E(B-V)=1.43):



(II) Period Fitting of Cepheid Candidates

two example fits for PS1 3π Cepheids:
... and a more severely reddened Cepheid (D=6.1 kpc, E(B-V)=2.5):



PanSTARRS1 as a time domain survey: successes & lessons learned

Multi-band structure functions work well for (initial) variable classification [Hernitschek+2016]:

- $\sim$1.000.000 QSOs
- 150.000 RR Lyrae candidates
- at $\sim$75 - 90% purity and >70% completeness
- should also work well for obscured Cepheids throughout the disk

With template light curves, 85% of RRL periods correct to 2 secs, despite extremely sparse sampling [Sesar, Hernitschek et al. 2016]

Take home message

with the right math, even sparse light curves can lead to (surprisingly) good variable classification