Pan-STARRS1 as pilot-survey for panoptic time-domain science

Nina Hernitschek[1], Hans-Walter Rix[1], Branimir Sesar[1] and Edward F. Schlafly[2]

[1] Max-Planck-Institut für Astronomie,
Königstuhl 17, 69117 Heidelberg, Germany
email: hernitschek@mpia.de

[2] UC Berkeley, Berkeley, CA 94720-5800

Abstract. For examining possibilities and challenges in doing science with multi-band and non-simultaneous data from upcoming surveys like LSST, the Pan-STARRS1 (PS1) 3π can be used as a pilot survey. This is especially important to explore the possibilities in detection and classification of variable sources within the first years of LSST's 10-year baseline. We had explored the capabilities of PS1 3π for carrying out time-domain science in a variety of applications. We had used structure function fitting as well as period fitting, to search for and classify high-latitude as well as low-latitude variable sources, in particular RR Lyrae, Cepheids and QSOs.

Keywords. methods: data analysis, methods: statistical, catalogs, surveys, (stars: variables:) Cepheids, stars: variables: other, (galaxies:) quasars: general

---

1. Introduction

This contribution deals with the time-domain properties of PS1 3π and why and how PS1 3π can serve as a pilot survey for the upcoming LSST. In the case described here, a *pilot survey* is an existing survey that is fully operationally and scientifically used, but can also be used to gather information for upcoming surveys.

2. Comparison of Pan-STARRS1 3π and LSST

Answering questions related to the important scientific problems of the next decade (such as studying the evolution of our Milky Way and of galaxies in general, discovering the nature of Dark Energy and Dark Matter, and opening up the time domain to discover rare transient and variable objects down to faint magnitudes with a fast cadence) relies strongly on deep, wide-field time-domain imaging of the sky in optical bands. Surveys like Pan-STARRS1 are already preparing for the challenges of the upcoming generation of synoptic sky surveys.†

Pan-STARRS consists of two telescopes, of which the first one, PS1, was used for the Pan-STARRS1 survey. Most of the observing time is dedicated to two surveys: The 3π survey, observing the entire sky north of declination -30°, and the medium-deep survey, a deeper, many-epoch survey of 10 fields, each 7 deg^2 in size (Chambers 2011). At the end of the 3π survey, there are in total $\sim$12 observations in each of the 5 filters available for each part of the observable sky. The PS1 Science Consortium defined 12 key projects, many related to galaxies and large-scale Milky Way structures.

The LSST is one out of the upcoming generation of synoptic sky surveys that will

† Here, this word is adopted from the LSST Science Book, Version 2.0, who use it to refer to "looking at all aspects of something", derived from the Greek word "Synopsis".

Table 1. Comparison of Pan-STARRS1 3π and LSST.

	PS1 3π	**LSST**
sky area	30,000 deg^2	20,000 deg^2 for main survey
sky region	$\delta > -30°$	$\delta < 10°$ for main survey, up to $\delta < 35.5°$
filters	$g_{P1}, r_{P1}, i_{P1}, z_{P1}, y_{P1}$, non-simultaneous	$ugriz y$, five concurrent at a time
single exposure depths (5σ)	$g_{P1} < 22.0$, $r_{P1} < 22.0$, $i_{P1} < 21.9$, $z_{P1} < 21.0$, $y_{P1} < 19.8$ mag	$u < 23.9$, $g < 25.0$, $r < 24.7$, $i < 24.0$, $z < 23.3$, $y < 22.1$ mag
coadded depths (5σ)	$g_{P1} < 23.4$, $r_{P1} < 23.4$, $i_{P1} < 23.2$, $z_{P1} < 22.4$, $y_{P1} < 21.3$ mag	$u < 26.3$, $g < 27.5$, $r < 27.7$, $i < 27.0$, $z < 26.2$, $y < 24.9$ mag
cadence	67 epochs over 5.5 years	1000 visits over 10 years
nightly data volume	1 TB	30 TB
catalog data volume	∼100 TB	30 PB

fulfill the requirements for the foresaid scientific fields, as it will be repeatedly survey the sky with deep short exposures and a fast cadence. Currently under construction, LSST will be sited on Cerro Pachón in the Northern Chilean Andes. The main survey, imaging each region of the sky ∼2000 times over a ten-year survey life time, covers only 20,000 deg^2, as it avoids confusion-affected parts of the Galactic plane.

A summary on various science metrics for PS1 3π and LSST are given in Table 1, based on Schlafly *et al.* (2014) and the LSST Science Book, Version 2.0.

Because of the large size of such surveys, methods are needed which run without much intervention after a test phase, are reliable, fast, and being able to deal with various kinds of variable sources. Methods being capable of evaluating sparse data in a reliable manner, as presented here, can help especially during the first stages of LSST. LSST will have an observational baseline more than twice as long as PS1 3π. Methods for PS1 3π are developed under the aspect of dealing with such a relatively short baseline in combination with the sparse cadence. For this reason, methods developed for PS1 3π can be helpful especially during the first months and years of LSST.

3. Classifying Variable Sources in Non-Simultaneous Multi-Color All-Sky Surveys

We lay out, develop, test, and apply an approach to characterize variable sources in a survey such as PS1 3 π having multi-band, non-simultaneous observations. As this method – shown here for QSOs and RR Lyrae – is very general, it can be extended to other classes of variable sources. We provide a comprehensive description of this approach and results in Hernitschek *et al.* (2016).

3.1. *Data*

From PS1 3π (e.g. Kaiser *et al.* 2010), we use the preliminary PS1 catalog processing version PV2, where the average number of detections per source is 35 over 3.7 years after rejection of non-photometric data. Nikutta *et al.* (2014) have shown that the color cut $W12 = W1 - W2 > 0.5$ from WISE infrared photometry is able to isolate QSOs, as it is an indicator of its hot dust torus. For sources with ($\sigma_{W1} < 0.3$, $\sigma_{W2} < 0.3$), we use $W12$ as parameter for classification. The SDSS contains a stripe in the South Galactic Cap, Stripe 82 (S82), scanned multiple times for co-addition of the data and to enable discovery of variable objects. S82 contains ∼10^4 spectroscopically confirmed

QSOs (Schneider *et al.* 2007, Schmidt *et al.* 2010) as well as 483 RR Lyrae (Sesar *et al.* 2010). Both are used as training set for classification as well as for testing the method.

3.2. *Methodology*

In this section we describe the steps we take to identify and characterize variable point sources: characterizing their variability, and then attribute classifications.

3.2.1. *Identifying Significantly Varying Sources*

We start with a generic non-parameteric measure to characterize the significance of variability, defining for an object's total number N of observations across all n bands:

$$\hat{\chi}^2 = \frac{\chi^2_{\text{source}} - N_{d.o.f}}{\sqrt{2\,N_{d.o.f}}}, \quad \chi^2_{\text{source}} = \sum_\lambda \sum_{i=1}^N \frac{(m_{\lambda,i} - \langle m_\lambda \rangle)^2}{\sigma^2_{\lambda,i}}. \tag{3.1}$$

The sum over λ is over the PS1 bands g_{P1}, r_{P1}, i_{P1}, z_{P1}, y_{P1}, and $N_{d.o.f} = N - n$.
For non-varying sources, $\hat{\chi}^2$ should form a Gaussian distribution. Varying sources should form a "tail" of higher $\hat{\chi}^2$. The QSOs and RR Lyrae are indeed well separated in the *normalized* distibutions. However, there are only 415 RR Lyrae and 7630 QSO, compared to $\sim 1.85 \times 10^6$ "other" objects in SDSS S82, superimposing the distributions of QSOs and RR Lyrae due to the high number of "other" sources. Therefore, a simple criterion such as $\hat{\chi}^2$ alone is insufficient to identify QSOs or RR Lyrae.

3.2.2. *Non-simultaneous, Multi-band Structure Functions*

The structure function (e.g. Kozłowski *et al.* 2010) of a source describes the mean squared magnitude difference between pairs of observations as a function of the time lag between them. The cadence of PS1 3π data makes it necessary to extend the usual single-band formalism to multi-band. For a source with N observations across all bands, for each observation n, we have magnitude m_n, its error σ_n, its epoch t_n and the corresponding bandpass b_n. The structure function is specified by a multi-band damped random walk (DRW) with a variability timescale τ and a band-depending variability amplitude $\omega(b_n)$ for the band of the observation n in case. The lightcurve variability is described by a $N \times N$ covariance matrix C with elements

$$C_{nn'} = \omega(b_n)\,\omega(b_{n'})\,\exp\left[-\frac{|t_n - t_{n'}|}{\tau}\right] + \sigma_n^2\,\delta_{nn'}. \tag{3.2}$$

To account for wavelength-depending variability amplitudes, we link the $\omega(b_n)$ by a power law with exponent $\alpha = \frac{\log(\omega(b_k)/\omega_r)}{\log(\lambda_k/\lambda_r)}$. The probability of the parameters given the data can be computed, which yields a maximum likelihood approach with the log-likelihood

$$\log \mathcal{L}(\mathbf{m}|\omega_r, \tau, \alpha, \mu) = -0.5 \left(\log|C| + \log|C_\mu| + \chi^2\right), \quad \chi^2 = (m - \overrightarrow{\mu})^{\text{T}} \cdot C^{-1} \cdot (m - \overrightarrow{\mu}) \tag{3.3}$$

and time-series magnitude predictions $\overrightarrow{\mu}$. RR Lyrae and QSOs populate different areas of (ω_r, τ) space: RR Lyrae have typical $\tau \sim 1$ day and QSOs have $\tau \sim 100 - 1000$ days.

3.2.3. *Classification using a Random Forest Classifier*

For classifying objects based on the previously calculated parameters, we use a Random Forest Classifier (RFC). Using a training set with observed object parameter values as well as class labels, it gives the classification probability of a target set's object being of a certain class. Table 2 summarizes the parameter set for the RFC. For details, we refer to Hernitschek *et al.* (2016).

Table 2. Parameter set for the Random Forest Classifier.

$\omega_{r,\mathrm{grid}}, \tau_{\mathrm{grid}}$	best fit structure function parameter on log-spaced grid
$\hat{\chi}^2$	normalized χ^2 statistic, see Equ. (3.1)
$(g-r)_{\mathrm{P1}}, (r-i)_{\mathrm{P1}}, (i-z)_{\mathrm{P1}}, (z-y)_{\mathrm{P1}}$	colors from dereddened PS1 mean magnitudes
$\langle r \rangle_{\mathrm{P1,deredd}}$	dereddened PS1 mean r_{P1} mag., for calculation of p_{QSO}
$W12$	$W1-W2$, helps with QSO identification
$i_{\mathrm{P1}} - W1$	RR Lyrae gravity indicator

3.2.4. *Limitations of the Method*

Our method is subject to some limitations. At low latitudes, as these regions include very large numbers of disk stars as well as distance depending reddening is present, the amount of contaminants is much larger than in S82. As the S82 QSO sample is only complete down to $i_{\mathrm{P1}} \lesssim 21.25$, fainter training set objects are labeled as non-QSOs, and our classifier learns to discard these objects. A final concern is that the classification of sources may be relatively uncertain, if they lack specific colors.

3.2.5. *Results*

We applied this variability characterization and subsequent Random Forest classification to 3.88×10^8 PS1 3π sources. In the following, all "purity" and "completeness" refer to the case having the full parameter set from Table 2 available and making sure the sources fulfilling the criterion of $15 < \langle g_{\mathrm{P1}} \rangle, \langle r_{\mathrm{P1}} \rangle, \langle i_{\mathrm{P1}} \rangle < 20$.

The assumption of an isotropic distribution of ~ 20 QSOs per deg^2 within $15 < \mathrm{mag} < 21.5$ (Schneider *et al.* 2007, Schmidt *et al.* 2010) allows us to test the large scale homogeneity of our classification. We have selected 3.99×10^5 likely QSO candidates at a level of purity of 82%, completeness of 75%. The candidate selection is homogeneous and similar to S82 to a high degree away from the Galactic plane, until dust extinction and disk star contamination become severe at low latitudes. At low latitudes, the number density of sources per deg^2 is much by a factor of ~ 5 higher than around the Galactic north pole.

We find a total of 1.5×10^5 highly likely RR Lyrae candidates. In the Galactic halo, at Galactic latitudes of $|b| > 20°$, we have 6.0×10^4 likely candidates, some extending to distances as large as ~ 140 kpc. We can make use of this to verify our candidate selection by comparing to known halo substructure (see Fig. 18 in Hernitschek *et al.* 2016). Using the RR Lyrae candidates in Draco dSph, we show that we get distance precisions of 6% at a distance of ~ 80 kpc. A projection of our candidate sample into the orbital plane of the Sagittarius stream reveals the stream morphology.

References

Chambers, K. 2011, *AAS Meeting Abstracts*, 218, 113.01
Hernitschek, N., Rix, H.-W., Bovy, J., *et al.* 2016, *ApJ*, 817, 1, 73
Kaiser, N., Burgett, W., Chambers, K., *et al.* 2010, *Proc. SPIE*, 77330E
Kozłowski, S., Kochanek, Ch. S., Udalski, A., *et al.* 2010, *ApJ*, 708, 2, 927
LSST Science Book, Version 2.0 2009, *arXiv:0912.0201*
Nikutta, R., Hunt-Walker, N., Nenkova, M., *et al.* 2014, *MNRAS*, 442, 4, 3361
Schlafly, E. F., Green, G., Finkbeiner, D. F., *et al.* 2014, *ApJ*, 789, 1, 15
Schmidt, K. B., Marshall, P. J., Rix, H.-W., *et al.* 2000, *ApJ*, 714, 2, 1194
Schneider, D. P., Hall, P. B., Richards, G. T., *et al.* 2007, *ApJ*, 134, 1, 102
Sesar, B., Ivezić, Ž., & Grammer, S. H. 2000, *ApJ*, 708, 1, 717