PROPERTIES OF M31. II. A CEPHEID DISK SAMPLE DERIVED FROM THE FIRST YEAR OF PS1 PANDROMEDA DATA

MIAH KODRIC1,2, ARNO RIFFESER1,2, ULRICH HOPP1,2, STELLA SEITZ1,2, JOHANNES KOPPENHOEFER1,2, RALF BENDER1,2, CLAUS GOESS1,2, JAN SNIGLA1,2, CHIEN-HSIU LEE1,3, CHOW-CHOONG NGEOW3, K. C. CHAMBERS3, E. A. MAGNIER4, P. A. PRICE4, W. S. BURGETT4, K. W. HODAPP4, N. KAISER4, AND R.-P. KUDRITZKI4

1 University Observatory Munich, Scheinerstrasse 1, D-81679 Munich, Germany
2 Max Planck Institute for Extraterrestrial Physics, Giessenbachstrasse, D-85748 Garching, Germany
3 Graduate Institute of Astronomy, National Central University, Jhongli City 32001, Taiwan
4 Institute for Astronomy, University of Hawaii at Manoa, Honolulu, HI 96822, USA

Received 2012 December 5; accepted 2013 January 14; published 2013 March 13

ABSTRACT

We present a sample of Cepheid variable stars toward M31 based on the first year of regular M31 observations of the PS1 survey in the $r_{P1}$ and $i_{P1}$ filters. We describe the selection procedure for Cepheid variable stars from the overall variable source sample and develop an automatic classification scheme using Fourier decomposition and the location of the instability strip. We find 1440 fundamental mode (classical $\delta$) Cep stars, 126 Cepheids in the first overtone mode, and 147 belonging to the Population II types. Two hundred ninety-six Cepheids could not be assigned to one of these classes and three hundred fifty-four Cepheids were found in other surveys. These 2009 Cepheids constitute the largest Cepheid sample in M31 known so far and the full catalog is presented in this paper. We briefly describe the properties of our sample in its spatial distribution throughout the M31 galaxy, in its age properties, and we derive an apparent period–luminosity relation (PLR) in our two bands. The Population I Cepheids nicely follow the dust pattern of the M31 disk, whereas the 147 Type II Cepheids are distributed throughout the halo of M31. We outline the time evolution of the star formation in the major ring found previously and find an age gradient. A comparison of our PLR to previous results indicates a curvature term in the PLR.

Key words: catalogs – distance scale – galaxies: individual (M31, NGC 224) – galaxies: star formation – Local Group – stars: variables: Cepheids

Online-only material: Supplemental data file (tar.gz)

1. INTRODUCTION

Classical $\delta$ Cep variable stars are young massive stars of Population I in their post-main-sequence evolution. Due to their period–luminosity relation (PLR; Leavitt 1908 and Leavitt & Pickering 1912), they are the anchor of the extragalactic distance scale. As such, they are a central topic of many investigations of the period–Wesenheit (Madore 1976, 1982; Opolski 1983) relation (from colors $V$ and $I$) does not depend on metallicity while the red band combinations of the Johnson–Cousins system should not show this dependence (Fiorentino et al. 2007). The discussion of the PLR properties—in individual photometric bands—has become even more complicated as some authors reported a non-universal slope for the PLR (see Sandage & Tammann 2008 and references therein), which has not been confirmed by others and is under debate.

The recent discussion on the calibration of the cosmological distance scale and the subsequent determination of the Hubble constant has somehow shadowed that these relatively young and massive stars can be also used as excellent tracers of the recent evolution of stellar populations as they allow number counts in precisely known time bins.

The LMC has been the historical place for the establishment of the PLR (Leavitt & Pickering 1912) as it is closely and can be well resolved from the ground while already distant enough that the member stars are essentially at the same distance. Thus, the LMC (and Small Magellanic Cloud, SMC) was intensively studied for variable stars and transient phenomena by various projects, also yielding extensive and well-observed samples of Cepheid stars in fundamental, first, and second overtone pulsation as well as Population II pulsators (W Vir, RV Tau, etc.). The most prominent example is probably the OGLE project, which provided a well-established PLR based on $B$, $V$, and $I$ observations of $\sim 3400$ $\delta$ Cep stars in the LMC and SMC (Udalski et al. 1999a, 1999b, 1999c).

The nearest example of a Cepheid population in a typical large spiral galaxy of roughly solar metallicity and outside our own Milky Way galaxy is the one in the Andromeda galaxy M31. The first detections go back to Hubble (1929). Baade &
Swope (1965) summarize the early observational results which already comprised a sample size of ~260 Cepheid variables.

Wide-field CCD imagers time series studies brought M31 again into the focus of several projects aiming to detect and study transients and variables stars (e.g., Macri 2004, the DIRECT project; Firi et al. 2006; Vilardell et al. 2007).

The Firi et al. (2006) sample (based on the WeCAPP microlensing project; Riffeser et al. 2001) was observed over almost eight years with ground-based CCD imaging centered on the M31 bulge. Therefore, the δ Cep stars as typical Population I objects are only a minor contribution (33) to the total sample of variable stars and they are even outnumbered by other types of Cepheids (93).

The Vilardell et al. (2007) sample was obtained from observations covering the northwest quarter of the M31 disk. Spanning almost five years, the ~250 observations in B and V allowed to obtain a sample of 416 Cepheid stars, to study their light curves in detail, and to derive distance estimates to M31. A shortcoming of the DIRECT data is that the 250 individual observations were obtained in only 21 nights. The window function (inherent to the observations) implies a strong bias against finding objects within certain period ranges. This can distort results like, e.g., the period-number distribution of the sample.

Ground-based M31 observations have been supplemented with single-epoch high-resolution HST imaging (PHAT; Johnson et al. 2012), especially in the near-IR (NIR). This allowed improvement on some of the shortcomings of ground-based data for this galaxy and paved the way for a precise distance determination of M31 (Riess et al. 2012).

In this paper, we expand the studies of the Cepheid population in M31 with our wide-field ground-based data taken within the first year of the Pan-STARRS 1 project (Kaiser et al. 2002). Among others, the two major advantages of our data are that, first, they cover the complete M31 disk and, second, they are much more evenly distributed in time using observations taken in up to 102 nights (~80% of the nights have two epochs, giving a total of 183 epochs).

Throughout this paper, we will refer to fundamental mode (FM) Cepheids (classical δ Cep variable stars), first overtone (FO) Cepheids (classical δ Cep variable stars pulsating in the excited mode of the FO), and Type II Cepheids (W Virginis), and use the term Cepheids for all these subgroups. BL Herculis stars are too faint to be detected with our survey and RV Tauri are excluded manually (cf. Section 3), so that our Type II Cepheids only consist of W Virginis stars.

This paper is organized as follows. In Section 2 we present the PAndromeda survey, our observations in 2010 and 2011, and a description of our data reduction, including the determination of the light curves for the resolved sources. Our period determination is described in Section 3. The Cepheids detected in the PAndromeda data set are presented in Section 4. This section also describes the automatic Cepheid detection and classification. An overview of the catalog is shown in Section 5. We discuss our results, such as (among others) the PLR, the spatial distribution of the Cepheids and the spatial age distribution, in Section 6. We conclude the paper in Section 7.

2. OBSERVATIONS AND DATA REDUCTION

2.1. PAndromeda

The 1.8 m Panoramic Survey Telescope and Rapid Response System (Pan-STARRS; Hodapp et al. 2004; Kaiser et al. 2002) on Haleakala, Maui, HI began its regular survey in spring 2010. Pan-STARRS 1 (PS1) uses the Giga Pixel Camera (GPC; Tonry & Onaka 2009) which is currently one of the largest cameras in the world. It consists of 60 orthogonal transfer arrays (OTAs), each of them segmented into an array of 8 x 8 cells. The total number of pixels is 1.4 × 10^9, with a pixel size of 0.258 arcsec. The field of view (FOV) of one OTA is 21’ × 21’. The aims of PS1 are to map 3π of the sky in gP1, rP1, iP1, zP1, and YP1 bands. In addition to the large 3π area, some fields are visited more frequently and have deeper exposure times. One of these so-called medium deep fields targets the Andromeda galaxy (M31). With the ~7 deg^2 FOV of Pan-STARRS, a complete monitoring of the disk and large parts of the halo is possible with one single exposure. The PAndromeda survey observes M31 for 0.5 hr per night (including overhead) during a period of five months per year, equivalent to 2% of the overall PS1 time. We split the integration time into two visits per night (separated by 3 ± 0.5 hr), for the whole season.5 During the first two seasons we focused on the filters rP1 and iP1 (Tonry et al. 2012). Observing M31 two times per night with the rP1 and iP1 filters is a consequence of the PAndromeda survey goal to find microlensing events as described in Lee et al. (2012).

In the first season, PAndromeda monitored M31 from 2010 July 23 to 2011 December 27. In this work we used in addition part of the second season from 2011 July 25 to 2011 August 12. This results in a gap of about half a year in the light curves and gives us a total of 183 epochs.6

2.2. Data Reduction

The details of the data reduction for PAndromeda are described in Lee et al. (2012). In the following section, we summarize the data reduction steps and point out differences to the approach taken in this work.

The GPC1 exposures are de-biased, flat-fielded, and astrometrically calibrated using the Pan-STARRS Image Processing Pipeline (Magnier 2006). During this process the images are remapped to a common grid, the so-called skycells. These warped images have a pixel scale of 0.200 arcsec pixel^{-1}, which is smaller than the natural pixel size of the GPC1 (0.258 arcsec pixel^{-1}). For better data management, we integrated our pipeline (Koppenhoefer et al. 2011) into the AstroWISE data management system (Valenti et al. 2007).

In this work, we use only data from 24 skycells, namely, 039-041, 051-054, 064-068, 076-079, 088-092, and 102-104. As shown in Figure 1, the skycell layout is not optimized in the sense that it does not follow the detector boundaries. We plan to optimize the skycell layout and use an intrinsic pixel scale close to 0.258 arcsec pixel^{-1} in the future. The following data reduction steps are applied to each skycell and both filters rP1 and iP1.

Since we use difference image photometry (Tomany & Crotts 1996), we have to produce a stack of images with the best seeing, the so-called reference image. We photometrically align the 70 best-seeing images, so that the images have the same zero point and the same (non zero) sky background. In the next step, we replace masked areas in the photometrically aligned images with values from another image that has the most similar point-spread function (PSF; Riffeser 2006, p. 134). The

5 Per visit seven exposures of 60 s in the rP1 band and five exposures of 60 s in the iP1 band are taken. The total integration time per night is therefore 1440 s. The remaining time is spent on readout, focusing, and filter change.

6 Due to masking the total number of epochs for a light curve can be less than 183.
A weighted stack of the aligned images is called the reference image.

We use SExtractor to detect sources on the reference image (Bertin & Arnouts 1996). For the data utilized in this work, SExtractor is used in a mode that modifies the detection threshold depending on the background. While this mode is very useful it hampers completeness tests.

For all sources in the reference image, we perform PSF photometry on a background (sky and M31) subtracted reference image. In order to obtain the background in the reference image, we fit a bicubic spline model. To suppress blending effects we iteratively subtract neighboring stars. The fluxes measured in the reference image are added later to the fluxes in the difference images, so as to create the light curves. The flux error in the reference image determined by the PSF photometry does not take into account how well our constant PSF model matches the real sources in the reference frame which can show slight PSF variation over the field. This can result in wrong flux errors. To obtain reasonable errors, we rescale the error of each source with the \( \sqrt{\chi^2} \) of its PSF fit.

To minimize masking and to obtain a better signal-to-noise ratio we stack the images of each visit. The procedure is the same as for the reference images. The visit stacks are photometrically aligned to the reference image, so that the visit stacks and the reference image have the same zero point and sky background. At this point they only deviate in signal to noise and PSF shape.

To create the difference images we first align the PSF of the reference image to match the PSF of the visit stack by calculating a constant convolution kernel by least-squares minimization from all pixels in subareas of 75'' x 75'' (see Alard & Lupton 1998). This PSF aligned reference image is then subtracted from the visit stack. On this difference image we perform PSF photometry to obtain light curves. The PSF is constructed with sub-pixel precision from the convolved reference image using isolated stars. Since we only measure the flux difference we add the flux measured in the reference image to this flux difference to obtain the light curves for the resolved sources. Figure 2 summarizes the PAndromeda data reduction pipeline.

Note that the astrometric precision and photometric stability have been discussed in Lee et al. (2012). For the flux calibration we make use of the fact that the \( r \) and \( i \) bands are rather similar to the Sloan Digital Sky Survey (SDSS; Abazajian et al. 2009) \( r \) SDSS and \( i \) SDSS bands and thus the flux calibration should be insensitive to color terms (Tonry et al. 2012). To calibrate our magnitude to the SDSS magnitude system with stars from SDSS we use

\[
m_{\text{SDSS}} = -2.5 \log\left[F_{\text{ref}}(\text{PSF}_{\text{ref}}) + \Delta F_{\text{diff}}(\text{PSF}_{\text{diff}}) + ZP_{\text{PSF}}\right].
\]

(1)

where \( F_{\text{ref}}(\text{PSF}_{\text{ref}}) \) is the flux of a source determined by PSF photometry in the reference image and \( \Delta F_{\text{diff}}(\text{PSF}_{\text{diff}}) \) is the flux determined by PSF photometry in the difference image. \( ZP_{\text{PSF}} \) accounts for the difference between SDSS magnitudes and our instrumental magnitudes and does also consider the PSF construction.

Our pipeline does not take into account the correlated noise that arises from the change of the pixel scale. To quantify this we constructed light curves of all pixels, normalize each single deviation from the fitted mean by its error. The distribution of these normalized deviations is clearly dominated by the non-varying pixels. Then the correction factor can easily be estimated by comparing this distribution with a standard Gaussian distribution (\( \sigma = 1 \)). This rescaling factor is the same for every skycell and filter.

Throughout this paper, we will use the AB magnitude system.
3. PERIOD DETERMINATION

The periodicity of the light curves is determined with the software **SigSpec** (Reegen 2007). **SigSpec** calculates a quantity called spectral significance (sig) for each peak in the discrete Fourier transform (DFT) amplitude spectrum of a light curve. Similar to the false alarm probability\(^7\) \(\Phi_{FA}(A)\)

\[
sig(A) := - \log[\Phi_{FA}(A)]
\]

is a measure for the probability that a peak arises from noise. This value does not only depend on the amplitude and frequency, but also on the phase and the sampling of the light curve. This fact gives **SigSpec** an advantage over the use of other Fourier-based algorithms such as the Lomb–Scargle algorithm (Lomb 1976; Scargle 1982). While both methods produce comparable results concerning the period, we found that **SigSpec** finds more periodic light curves over a given false alarm probability threshold than the Lomb–Scargle algorithm. This is due to the fact that certain periods have to be excluded in the Lomb–Scargle algorithm, depending on the sampling of the light curve in order to avoid aliasing.

In the following paragraph we outline the parameters with which we run **SigSpec**. As statistical weight we use the inverse variance (\(w_i = \sigma_i^{-2}\)) of the PSF flux measurement and we use \([\text{JD} - 2450000]\) for the epoch. We apply a sig-threshold of 5 (siglimit 5) to determine the periodicity of the light curves. This means that in less than one out of \(10^5\) cases pure noise could produce the measured period. With a lower threshold the number of detections would increase (e.g., by a factor of \(\sim 10\) with a threshold of 3), but most of the detections would be low signal-to-noise detections. With the parameter `ufreq` 1 we define the lowest period to be considered as one day. We exclude smaller periods because they can hardly be detected with a sampling of two visits per night. An upper limit for the period has not been defined, but a period cut is performed later on. The last two parameters (`multisine:lock` and `iterations 1`) force **SigSpec** to consider only the most significant peak\(^8\) in the DFT amplitude spectrum ensuring that noise has no impact on the determined period. We use no variability index since **SigSpec** only determines periods for variable sources.

Inspection by eye of the folded light curves, which **SigSpec** detects, showed the good performance in the period determination of **SigSpec**. One problem, however, is the lack of an upper period limit that permits **SigSpec** to determine periods even if the signal is not periodic.\(^9\) The inspection showed that for large periods some of the folded light curves look like Cepheid light curves. We did not exclude those light curves at this stage, but rather applied a period cut later on. As our data set also contains 10 days of observations of season 2011, the data show a gap of half a year between these data points and the 2010 data points. In some cases, the 2011 data points have a small flux shift from the 2010 data points in the folded light curve. This may have different reasons. First, the shift could be due to flux variations.

\(^7\) The false alarm probability \(\Phi_{FA}(A)\) is the integrated probability density function and describes the probability that an amplitude exceeds a given limit \(A\) as described in Reegen (2007).

\(^8\) The **SigSpec** default value would perform an iterative sequence, the so-called prewhitening sequence. After the most significant peak is identified, a fit to all signals that are identified so far is performed. This fit is then prewhitened (subtracted) from the spectrum and the next iteration starts until the maximum peak is under the threshold.

\(^9\) **SigSpec** can determine a period for almost every signal that is not constant. If the signal is not periodic and not constant the program usually detects the total observation time as the period.

Second, it could be caused by a small deviation from the true period, which can be observed in the folded light curve if the period is small enough. And last the frequency spacing used by **SigSpec** could be too low. Also in the case of RV Tauri variables **SigSpec** detects the wrong period, as it identifies the time span between the primary and secondary minimum as the period. We identified and rejected those light curves manually.\(^10\) From 724894 sources that were detected in the \(r_p\) band and which have light curves in the \(r_p\) and \(i_p\) band, **SigSpec** finds 75,362 periodic variable light curves with sig > 5.

Since **SigSpec** does not determine the period error, we use the bootstrap method to estimate the period uncertainties. A random sampling is drawn from the light curve (one epoch can be drawn multiple times) and the period for that sampling is determined with **SigSpec** (with a sig-threshold of 1). This procedure is performed 1000 times for each light curve, so that the 1σ error can be determined from the resulting distribution. The obtained period errors for the Cepheids of the three-dimensional parameter space classified Cepheid catalog (see Section 4.6) are shown in Figure 3. The classification of the different Cepheid types shown in Figure 3, as well as the meaning of the three-dimensional parameter space classified Cepheid catalog, is described in the next section.

In general, the period errors are small, except for some light curves where the distribution has an additional peak at ~1.5 days due to aliasing. For those light curves the large error is only caused by this additional peak. It is not a broader distribution that causes the larger error and therefore the period error would be small if there is no aliasing. Since we find no aliasing in the final Cepheid light curves and the systematics have a larger effect than the period errors, we disregard the period errors, which means that the period errors shown in the plots do not contribute to any selection criteria or fit.

\(^10\) RV Tauri light curves are rather easy to distinguish from the Cepheid light curves, since the RV Tauri light curves appear as an overlay of two shifted light curves.

\(^11\) We detect the variable sources in the \(r_p\) band and therefore there is always a light curve in the \(r_p\) band for each variable source. But the same source can be masked in the \(i_p\) band so that there is no light curve in the \(i_p\) band.
4. CEPHEID DETECTION AND CLASSIFICATION

As described in the previous section we narrowed the number of variable sources down to those that have periodic light curves. In this sample of variable sources we want to identify the Cepheids.

The classical approach to find the Cepheids (e.g., Udalski et al. 1999a; Vilardell et al. 2007) is to preselect sources in a certain brightness and color range and to visually inspect them. But the position of the instability strip spans an area depending on brightness and color (Fiorentino et al. 2002 and Marigo et al. 2008, including the updates by Girardi et al. 2010). We choose a different approach than the visual inspection and establish a more robust Cepheid detection. The motivation for this is that the selection based purely on light curves is always biased, in particular if the light curves become more noisy. We define limits in a three-dimensional parameter space which define our Cepheids. In order to define the parameter space, we select a subsample of the periodic light curves and visually select the Cepheids. The manual selection is of course also biased, but the advantage of this method is that the selection of the remaining Cepheids only depends on the defined parameter space. The parameters used are derived from a Fourier decomposition of the light curves. With this method we were also able to identify the Cepheid type. The FM Cepheids and the FO Cepheids can be distinguished in this way as, e.g., shown by Udalski et al. (1999a) and Vilardell et al. (2007). We can additionally also distinguish the Type II Cepheids from the other two mentioned Cepheid types.

Additionally we impose certain requirements on the remaining variable sources detected by SigSpec, so as to exclude bad light curves.

In the following sections, we describe our approach that we outlined here.

4.1. Fourier Decomposition

A Fourier decomposition of the folded light curves can be used to distinguish between FM Cepheids and FO Cepheids (Udalski et al. 1999a; Vilardell et al. 2006). We also use Fourier decomposition to identify Type II Cepheids and to define the three-dimensional parameter space in which we detect the Cepheids.

We fit a Fourier series to the light curve of the form

$$m(t) = a_0 + \sum_{n=-N}^{N} a_n \cos(n\omega t) + \sum_{n=1}^{N} b_n \sin(n\omega t),$$

with \(P\) being the period of the light curve determined by SigSpec, \(\omega = (2\pi / P)\), and \(t \in [0, P]\). Prior to the fit we convert the flux to magnitudes and fold the light curves, so that they show the variation of the magnitude over the phase \(\theta\) times the period\(^{12}\) \(P\). With this fit we determine the minimum and maximum of the light curve and the coefficients \(a_n\). We also calculate the magnitude of the mean flux, needed later for the PLR. This is different to a mean magnitude, which we would get if we take the mean of the light curve. For this reason we transform the Fourier series fit back to flux and calculate the mean flux and then convert this to a mean magnitude.

For an analysis of the Fourier decomposition we transform the Fourier series to the form

$$m(t) = b_0 + \sum_{k=1}^{N} b_k \cos(k\omega t + \phi_k),$$

with \(b_k = \sqrt{a_k^2 + a_{-k}^2}\) and \(\phi_k = \arctan(-a_k/a_{-k})\). This allows us to calculate the amplitude ratio

$$A_{21} = \frac{b_2}{b_1}$$

and the phase difference

$$\varphi_{21} = \varphi_2 - 2\varphi_1$$

for each light curve (Vilardell et al. 2007). The period \(P\) and the coefficients \(A_{21}\) and \(\varphi_{21}\) are used to define the three-dimensional parameter space \((P, A_{21}, \varphi_{21})\) and to identify the Cepheid type.

\(^{12}\) That is why \(t \in [0, P]\) and not \(t \in [0, 1]\), what would be the case for folded light curves that vary over the phase.
We chose \( N = 5 \) for the Fourier series fit. Other degrees of freedom (dof) would result in slightly different amplitudes \( a_k \). In most cases the deviations of the \( a_k \), from other dof, compared to the \( a_k \) of \( N = 5 \), is small, but there are also light curves with significant deviations for higher dof. While \( N = 5 \) is sufficient to fit bumps (cf. left panel of Figure 12) in the light curves, higher dof are more susceptible to “overshoots” and tend to be more influenced by outliers. This is why we do not determine the optimal dof for each light curve as, e.g., described in Deb & Singh (2009). The principal component analysis (PCA) technique demonstrated in Deb & Singh (2009) does not offer any advantage in the form of saving computation time, since we use a linear least-squares fit and not a nonlinear fit for which the PCA is faster.

Additionally, we derive a polynomial fit with two boundary conditions of the form

\[
p(t) = a_0 + \sum_{k=1}^K a_k t^k \quad \text{with} \quad p(0) = p(P) \quad \land \quad p'(0) = p'(P) \quad \land \quad t \in [0, P]
\]

(7)

to the folded light curves (in magnitudes). For this fit a change in the number of dof also changes the \( a_k \), so we chose \( K = 10 \) in order to have the same number of dof as in the Fourier series fit. With this polynomial fit we have an alternative light-curve description to the Fourier series. The best-fitting Fourier and polynomial fits are affected by data gaps in a different way and they produce different “overshoots.” We compare the Fourier and polynomial fits to identify and exclude badly fitted light curves, as we will describe in the next section.

4.2. Wesenheit–Color Cuts

Young stars are located in spiral arms which are also populated by interstellar dust. The dust in the arms shows a clumpy distribution and Cepheids may be either located before, within, or behind this dust.

Although Population II Cepheids are not located in the M31 disk it is not a priori known whether they stay in front of or behind the disk. Therefore, even for the majority of the Population II objects an understanding of the internal reddening is required for each single object.

As already mentioned the preselection of sources in a certain brightness and color range depends on the instability strip. But it also depends on the reddening. By using the Wesenheit magnitude the brightness is independent of reddening and only the color is affected by extinction. For this reason we use the Wesenheit–color plane and select Cepheids according to the location of the instability strip.

The Wesenheit magnitude is defined to be independent of reddening (Madore 1976, 1982; Opolski 1983):

\[
W = r_p - 0.036 (g_p - i_p) = r_{P1,0} - 0.036 (g_{P1,0} - i_{P1,0}).
\]

(8)

With \( r_{P1} = r_{P1,0} + A_{r_p}, \ i_p = i_{P1,0} + A_{i_p} \) and \( A_{r_p} \) and \( A_{i_p} \) being the extinction corrections in the \( r_p \) and \( i_p \) bands we can derive the coefficient

\[
A_{r_p} = \frac{r_{P1,0} - r_{P1,0} - (r_{P1,0} - i_{P1,0})}{r_{P1,0} - 0.036 (g_{P1,0} - i_{P1,0})} = \frac{A_{r_p} - A_{i_p}}{A_{r_p} - A_{i_p}}.
\]

(9)

The galactic extinction coefficients \( A_{r_p}/E(B-V) \) and \( A_{i_p}/E(B-V) \) are (Tonry et al. 2012)

\[
\frac{A_{r_p}}{E(B-V)} = 2.585 - 0.0315(g_{P1,0} - i_{P1,0})
\]

(10)

\[
\frac{A_{i_p}}{E(B-V)} = 1.908 - 0.0152(g_{P1,0} - i_{P1,0}).
\]

Fiorentino et al. (2002) provides the position of the instability strip for Cepheids depending on \( L/L_{\odot} \) and \( T_{\text{eff}} \) (\( L_{\odot} \) is the solar luminosity and \( T_{\text{eff}} \) the effective temperature). We combine this with the isochrones from Marigo et al. (2008) with the corrections by Girardi et al. (2010) with \( Z = 0.019 \), for different \( L/L_{\odot} \) and \( T_{\text{eff}} \), thus allowing us to determine the theoretical location of the Cepheids in the color–magnitude diagram, as shown in the left panel of Figure 4. The right panel of Figure 4 shows the domain defined by this theoretical prediction in the color–Wesenheit diagram (defined by a triangle with the corners \((0.036/21.50), (0.275/16.56), \) and \((0.59/16.56))\), which is a reasonable choice.

A candidate is qualified as a Cepheid variable star if its mean magnitude is located inside this triangle (including its 1σ error). Similar to Figure 4 we also obtained the location of the Cepheids in the \( g_p-i_p \) versus \( g_{P1,0} - i_{P1,0} \) plane; see Figure 5. For \((g_{P1,0} - i_{P1,0}) = 1.0 \) and from Equation (10) we get \((A_{r_p}/E(B-V)) = 2.5535, \) \((A_{i_p}/E(B-V)) = 1.8928, \) and \( R = 3.86. \) Since the \((g_{P1,0} - i_{P1,0}) \) correction is small, the min–max error in \( R \) is \( \Delta R = 0.04, \) which makes \((g_{P1,0} - i_{P1,0}) = 1.0 \) a reasonable choice.

While the Wesenheit magnitude is independent of extinction, the color is not. Therefore, we apply an extinction correction to the colors. We use the \( E(B-V) \) map from Montalto et al. (2009) and the foreground extinction of M31 derived by Schlegel et al. (1998) with \( E_{B-V} = 0.062 \) and apply the recalibration factor of 0.86 (Schlafly & Finkbeiner 2011) to determine a mean extinction for each light curve depending on its (projected) position:

\[
A_{r_p} = \frac{E(B-V) + 0.86 \times 0.062}{2} \times \frac{A_{r_p}}{E(B-V)}
\]

(11)
Figure 6. Three-dimensional parameter space for the manually classified Cepheid catalog in the $r_{P1}$ band. The scope defined by the magenta dashed lines defines the parameter space in which most of the FO Cepheids ($P < 6.0 \text{ days}$ and $A_{21} < -0.498289 \log(P) + 0.527744$) or rather most of the Type II Cepheids ($11.95 \text{ days} < P < 53.0 \text{ days}$ and $\phi_{21} > 5 \text{ rad}$) reside, while the FM Cepheids occupy the remaining parameter space. Since there are transitions between the different scopes, we perform a $3\sigma$ clipping in the period–Wesenheit relation (cf. Figure 7) for each Cepheid type. All the clipped Cepheids and all those Cepheids with a large error in the Wesenheit ($\Delta W > 0.9$) are classified as unclassified (UN) Cepheids. Left panel: amplitude ratio $A_{21}$ in the $r_{P1}$ band. Right panel: phase difference $\phi_{21}$ in the $r_{P1}$ band.

\[
A_{ip1} = \frac{E(B-V) + 0.86 \times 0.062}{2} \times \frac{A_{ip1}}{E(B-V)}.
\] (12)

Note that for Cepheids that are outside the Montalto et al. map we only corrected for the foreground extinction.

Obviously we could use an extinction-corrected magnitude and use it instead of the Wesenheit, but since the correction is associated with a large systematic uncertainty it is better to use the Wesenheit, which does not require an extinction correction. The Wesenheit thus has a higher statistical error, but a lower systematic error relative to the extinction-corrected magnitudes.

With the Wesenheit and the extinction-corrected ($r_{P1,0} - i_{P1,0}$) color, we are able to select only those light curves belonging to the instability strip. This Wesenheit–color, applied to the 75,362 periodic light curves detected by SigSpec, excludes 70% of the light curves.

4.3. Manual Classification

In order to find out in which part of the $P$, $A_{21}$, $\phi_{21}$ three-dimensional parameter space the Cepheids are located, we need a “training” set of Cepheids which we obtain from a manual classification, i.e., by visual inspection of the preselected light curves.

The visual inspection of a subsample of Cepheid candidates and admission or rejection of Cepheids based on the light curve can lead to a bias if the Cepheid subsample itself does not represent all Cepheid types. The most unbiased way to select a Cepheid candidate subsample for manual classification is to use the spectral significance ($\text{sig}$). On the one hand the significance threshold has to be chosen low enough to contain a fair amount of Cepheids, so that there are enough Cepheids for the exploration of the three-dimensional parameter space. On the other hand, we do not want the Cepheid candidate subsample to contain so many Cepheids already, that the classification with the parameter space does not detect significantly more Cepheids than the subsample.

For the manual classification, we thus selected all light curves that have $\text{sig} > 25$ in the $r_{P1}$ band. As reference light curves for the manual classification we used the Cepheid light curves from Sterken & Jaschek (2005).

The resulting manual classified subsample contains 1020 Cepheids. This subsample makes up $\sim 50\%$ of the final Cepheid sample.

4.4. Type Classification

As already mentioned in Section 4.1, the Fourier decomposition can be used to distinguish between FM Cepheids and FO Cepheids. This is done by identifying different locations of the Cepheids in the $A_{21} – P$ and $\phi_{21} – P$ planes (e.g., Udalski et al. 1999a; Vilardell et al. 2007). Figure 6 shows these projections for the manually classified Cepheid catalog. When looking at the Cepheids in a three-dimensional plot in $A_{21} – \phi_{21} – P$ space, it becomes clear that there are indeed three different relatively well separated distributions with only small overlap areas.

The FM Cepheids correspond to the “V-shaped” sequence that can be seen in the left panel of Figure 6. In a three-dimensional plot the FM sequence appears spiral shaped and the two branches are shifted slightly from one another. The FO Cepheid sequence is perpendicular to the FM sequence, but there is also an overlap regime between them where FM and FO objects can be hardly distinguished. For objects in or near this overlap area the classification into FM or FO is therefore rather uncertain as a real parameter overlap seems to be present.

We additionally find a third sequence not present in Udalski et al. (1999a) and Vilardell et al. (2007) and corresponding to Type II Cepheids. This sequence is more distinct from the FM Cepheids than the FO Cepheids, nevertheless there is also a transition between the two.

In a first step we define the parameter space where most of the FO Cepheids and the Type II Cepheids are located and assign the according type to the Cepheid. We use

\[
P < 6.0 \text{ days} \quad \text{and} \quad A_{21} < -0.498 \log(P) + 0.528
\]

for the FO parameter space.

For the Type II space we use

\[
11.95 \text{ days} < P < 53.0 \text{ days} \quad \text{and} \quad \phi_{21} > 5 \text{ rad}.
\] (14)

The remaining Cepheids are FM Cepheids. Figure 6 shows the three different sequences. Nevertheless, there are transitions...
are classified as unclassified (UN) Cepheids. Cepheids and all those Cepheids with a large error in the Wesenheit (\(\Delta W > 0.9\)) are classified as unclassified (UN) Cepheids.

Figure 7. Period–Wesenheit relation for the manually classified Cepheid catalog. The period–Wesenheit relation is used to identify those Cepheids that reside between the type classification scopes (cf. Figure 6). This is done by performing 3\(\sigma\) clipping for a linear relation (\(W = a \log(P) + b\), cf. Section 6.4) for each Cepheid type (FM Cepheids: blue symbols and line; FO Cepheids: red symbols and line; Type II Cepheids: green symbols and line). All the clipped Cepheids and all those Cepheids with a large error in the Wesenheit (\(\Delta W > 0.9\)) are classified as unclassified (UN) Cepheids.

between the sequences. There are FM Cepheids in the FO and Type II parameter space and FO Cepheids or rather Type II Cepheids in the FM sequence. In order to account for these ambiguous cases, we identify these Cepheids in a second step by using the period–Wesenheit relation. By using the Wesenheit and not the \(r_{p1}\) band, we get a classification independent of reddening. We perform iterative 3\(\sigma\) clipping for a linear relation (\(W = a \log(P) + b\), cf. Section 6.4) for each Cepheid type and classify all clipped Cepheids and all those Cepheids with a large error in the Wesenheit (\(\Delta W > 0.9\)) as unclassified (UN) Cepheids. Figure 7 shows that this classification works well, but also indicates the problem with this approach. Cepheids that are, e.g., outside the defined Type II parameter space will be classified as UN Cepheids although it is apparent from Figure 7 that they are most likely Type II Cepheids.

4.5. Selection Criteria

With a visual inspection of the light curves only a color cut is needed to identify the Cepheids. Since we want to identify Cepheids using the \(P, A_{21}, \varphi_{21}\) three-dimensional parameter space cut, we have to apply additional selection criteria to exclude bad light curves. Note that the following criteria can be applied in any order. Table 1 summarizes the selection cuts.

The first two criteria (I and II) concern the period of the light curve. As already mentioned in Section 3 we have not applied a period cut yet. Our data cover approximately 150 days in 2010 and a few days only in 2011. While inspecting the light curves by eye, these few days in 2011 can help to confirm that the detected period is valid, even if the period is \(\sim 100\) days. However, since there is a gap of approximately half a year between the 2010 and 2011 data with only few days of observations in 2011, the light curve could be irregular in the time span of the gap and the data points in 2011 could fit to the folded light curve by chance. We decided therefore to select only light curves with periods below 75 days (I), so that the whole period cycle is sampled at least twice in the 2010 data. We note that there could very well be Cepheids with larger periods than that, but we will be able to confirm those only after the release of the full 2011 data. The second criterion (II) is that the period determined in the \(i_{p1}\) filter differs from that in the \(r_{p1}\) filter by less than one percent. This criterion excludes most of the light curves (84%).

We do not require a light curve to contain more than a certain number of epochs. The reason for this is that we use the Fourier series fit and the polynomial fit to account for gaps in the data that might be problematic and produce “overshoots.” Most of our light curves have \(\sim 180\) epochs, but there can also be light curves with as few as 60 epochs.\(^{13}\) The left panel of Figure 8 shows the distribution of the number of epochs for the three-dimensional parameter space classified Cepheid catalog. The right panel of Figure 8 shows that there are only a few Cepheids with a small number of epochs when normalized to the period of the light curve in the three-dimensional parameter space classified Cepheid catalog. A visual inspection shows that these light curves constitute Cepheids. This indicates that we need no criterion requiring a certain number of epochs when applying the three “overshoot” criteria (the classification of the

\(^{13}\) Sixty is the lowest number of epochs for a light curve associated to a Cepheid in the three-dimensional parameter space classified Cepheid catalog.

Figure 8. Epoch distribution for the three-dimensional parameter space classified Cepheid catalog in the \(r_{p1}\) band. Left panel: distribution of the number of epochs. Right panel: distribution of the number of epochs divided by the period.
The criteria III and IV exclude extreme “overshoots” by selecting only those light curves, where the magnitudes $m(t)$ of the fitted light curve are reasonable: $m(t) < 30$ mag and $m(t) > 15$ mag.

Note that criteria I–IV are applied to both $r_{P1}$ and $i_{P1}$ band. All other criteria only concern the $r_{P1}$ band.

In order to also exclude less extreme “overshoots,” we apply criterion V, which selects only light curves with similar Fourier series fit and polynomial fit. We check for similarity by calculating the area between both fits and demand that the area $\int_0^P (m(t) - p(t))^2 dt$ is below a threshold. The median of this area is $\sim 0.2$ mag day and the largest area is $\sim 13$ for the three-dimensional parameter space classified Cepheid catalog. So our threshold of 30 is quite high when compared to those numbers, but the “overshoots” usually produce orders of magnitude larger values than this threshold. The choice of the threshold is arbitrary to some degree, although the threshold of 30 has proven to work well.

The last two selection criteria (VI and VII) concern the noise in the folded light curves. If the maximum and minimum of the Fourier series fit are similar, we cannot exclude that the light curve is constant and the variation is caused by noise. Additionally, light curves affected by blending have smaller amplitudes. Therefore, we select only those light curves with $\text{max}(m(t)) - \text{min}(m(t)) \geq 0.1$ mag. The $\chi^2$ of a fit can be low, although the scatter is large, if the errors in the data points are large enough. Therefore, we introduce a last selection criterion (VII), which concerns the scatter of the light curve. We want to allow for a few points to scatter, which is why we compare the median scatter of the Fourier series fit with the total change in magnitude: $\text{median} \left| m(t) - y(t) \right| \leq 0.2 \left( \text{max}(m(t)) - \text{min}(m(t)) \right)$ ($m(t)$ is the magnitude of the Fourier series fit at time $t$ and $y(t)$ is the measured magnitude at time $t$). The threshold of 0.2 was optimized by a visual inspection.

### 4.6. Three-dimensional Parameter Space Classification

After the selection criteria are applied (cf. Section 4.5), we are left with some objects with well-defined light curves that are not Cepheids. This becomes obvious if we compare their position in the three-dimensional parameter space $(A_{21}, \varphi_{21}, P)$ with that of our manually classified Cepheids. Therefore, we use the manually classified subsample to define a “three-dimensional parameter space” occupied by Cepheids. All variables surviving the selection criteria in Section 4.5 which at the same time reside in the three-dimensional Cepheid subspace are considered Cepheids as well. The complexity of this three-dimensional subspace is so high that there are no two-dimensional projections that could classify Cepheids as effectively.

We span a $10 \times 10 \times 10$ grid$^{14}$ in $(A_{21}, \varphi_{21}, P)$ space with $1 \leq \log(P) \leq 75$, $0 \leq A_{21} \leq 0.75$, and $0 \leq \varphi_{21} \leq 2\pi$. We then identify those boxes around the grid points where the Cepheids of the manually classified subsample reside in. The space defined by those grid points that have Cepheids around them defines the three-dimensional parameter space. The subsample size and the rather large grid span force us to accept grid points that only have one data point around it. While this is not ideal, the alternative would be to accept only grid points that have a number of Cepheids above a certain threshold and this would drastically decrease the number of Cepheids at large periods, Type II Cepheids and FO Cepheids.

After the Wesenheit–color cut and the selection criteria are applied there are 2030 light curves within this parameter space, including also those already identified with the manual classification. As stated in Section 3, these light curves can contain RV Tauri light curves. We inspect those 2030 light curves manually and find 21 RV Tauri light curves, which we exclude from the three-dimensional parameter space classified Cepheid catalog. The final three-dimensional parameter space classified Cepheid catalog contains 2009 Cepheids (including the 1020 manually classified Cepheids).

As can be seen in Figure 9 the transition region between the different Cepheid types becomes more populated, in comparison to the manual classified Cepheid catalog (cf. Figure 6). This makes the classification of the UN Cepheids with the period–Wesenheit relation (Figure 10) even more important. Of the 75362 periodic light curves detected by SigSpec the three-dimensional parameter space cut alone excludes 94%.

It is important to note that while this procedure works well, there are Cepheids that are excluded by this strict three-dimensional parameter space classification cut. This is due to the grid size and the sampling of the parameter space through the manually classified subsample.

## 5. THE CATALOG

The three-dimensional parameter space classified Cepheid catalog contains a total of 2009 Cepheids, consisting of 1440 FM Cepheids, 126 FO Cepheids, 147 Type II Cepheids, and 296 UN Cepheids.

Table 2 contains an excerpt of the three-dimensional parameter space classified Cepheid catalog (the corresponding light curves are shown in Figures 11–17). The full version of Table 2 is available in the .tar.gz package in the online journal. A portion is shown here for guidance regarding its form and content.

We match the position of the Cepheids in our sample with the 416 Cepheids from Vilardell et al. (2007) and found a coincidence for 225 Cepheids (187 FM, 9 FO, 2 Type II, and 27 UN Cepheids) with a matching radius of 1 arcsec. The periods of both samples show a good agreement (Figure 18). The light curve of the Cepheid with the largest difference in the period (5,101 days) is shown in the right panel of Figure 16. The reference image shows no other close source and as can be seen from the right panel of Figure 16 the period determination is reasonable, which suggests a slightly wrong period in the Vilardell sample. The periods for the other five light curves with

---

**Table 1**

Selection Criteria for the Light Curves

| Selection Criterion | Percentage Excluded |
|---------------------|----------------------|
| I $P_{p1}$          | $\leq 75$ days       |
| II $\left| \frac{p_{p2} - p_{p1}}{p_{p1}} \right| < 0.01$ | 84% |
| III $m(t)$          | $< 30$ mag          |
| IV $m(t)$           | $> 15$ mag          |
| V $\int_0^P (m(t) - p(t))^2 dt < 30$ | 31% |
| VI $\text{max}(m(t)) - \text{min}(m(t)) \geq 0.1$ mag | 4% |
| VII $\text{median}[m(t) - y(t)] - \text{max}(m(t)) \leq 0.2$ mag | 42% |

**Notes.** $P$ is the period, $m(t)$ is the magnitude of the Fourier series fit at time $t$, $p(t)$ is the polynomial fit at time $t$, and $y(t)$ is the measured magnitude at time $t$. The last column is the percentage excluded by the criterion (if each criterion is applied individually) from the 75,362 periodic light curves detected by SigSpec.

---

$^{14}$ A grid with more grid points could leave gaps, which we want to avoid.
a relative difference larger than 0.5% we were able to confirm by visual inspection, but in some cases we cannot rule out that there is another close source within the 1 arcsec matching radius and that the difference is caused by that source.

We also matched our Cepheids with the Cepheids of the DIRECT project (Macri 2004) and the WeCAPP project (Fliri et al. 2006) with the matching radius of 1 arcsec. We found matches for 216 Cepheids from 332 DIRECT Cepheids (187 FM, 3 FO, 0 Type II, and 26 UN Cepheids) and 26 matched Cepheids from 126 WeCAPP Cepheids (15 FM, 0 FO, 4 Type II, and 7 UN Cepheids). Taking also the Vilardell et al. (2007) sample into account we matched 354 unique Cepheids.

Figure 19 shows the color–magnitude diagram and the color–Wesenheit diagram for all resolved sources in the \(r_{P1}\) band and the Cepheids of the three-dimensional parameter space classified Cepheid catalog. The left panel of Figure 20 shows the color–Wesenheit diagram for the Cepheids alone. We require that the Cepheids lie within the domain enclosed by the dashed line which results from theoretical predictions of the instability strip (see Section 4.2).

Using a simple color cut for the selection of Cepheids would either yield too many interlopers (wide cut) or cut too many Cepheids (narrow cut) as a simple color cut cannot account for the tilt of the instability strip in the color–magnitude diagrams (cf. right panel of Figure 20). The use of the instability strip allows for a strict sample cut and yields a color distribution dominated by Cepheid stars (cf. left panel of Figure 20).

To describe the overall shape of the Cepheid light curve and to trace the bump progression (also known as Hertzsprung progression; Bono et al. 2000) we introduce the decline/rise factor. This factor describes the ratio of the time from the minimal to the maximal magnitude (not flux!) compared to the time from the maximal to the minimal magnitude. Figure 21 shows that there are different scopes that correspond to the different positions of the bump. Around a period of 10 days two bumps occur on each side of the minimal magnitude and the light curve gets quite symmetrical. This can be verified in our sample light curves in Figure 12.

6. RESULTS

So far, the most complete catalog of Cepheid variable stars in M31 had been presented by Vilardell et al. (2007). They found 416 Cepheids of which they classified 281 as FM and 75 as FO objects. We here present a dramatically larger catalog with 2009 candidates, almost five times the number of objects in the earlier catalog. The factor ∼4.8 difference has to be attributed to the different patrol fields: Vilardell et al. (2007) observed 0.32 deg\(^2\) within northeastern lobe of the disk of M31 (and some contribution from bulge and halo) while our Pan-STARRS 1 data cover the full disk, the complete bulge and a large fraction of the M31 halo (only a small part of the halo has been analyzed for this work).

At the distance of M31 (750–770 kpc), an arcsecond corresponds to ∼3.7 pc. Therefore, ground-based imaging of individual stars within M31 is subject to blending and crowding, especially for young stars like Type I Cepheids which are often located in young clusters or associations. Even at the resolution of the HST cameras, blending is an issue. The impact of blending has been discussed in detail by Vilardell et al. (2007) for M31 and Chavez et al. (2012) for M33. Chavez et al. (2012)—comparing ground-based images to HST images—find
Therefore, it is even more puzzling that the FO-to-FM number completeness for the on average bluer and fainter FO stars. The columns contain (from left to right): identifier, Cepheid type, R.A. (J2000.0), decl. (J2000.0), significance (sig), period (determined from the Fourier series fit), mean magnitude in the rP1 band, mean magnitude in the iP1 band, Fourier coefficient A_{21}, Fourier coefficient ϕ_{21}, Wesenheit index and the decline/rise factor (cf. Figure 21).

(This table in its entirety is available in the .tar.gz package available in the electronic edition. A portion is shown here for guidance regarding its form and content.)

Table 2

| Identifier | Type | R.A. | Decl. | sig | P | rP1 | iP1 | A_{21} | ϕ_{21} | W | Decline/Rise |
|------------|------|------|-------|-----|---|-----|-----|-------|-------|---|--------------|
| 103-27590  | FM   | 11.1608 | 42.1585 | 16.40 | 4.398 | 21.34 ± 0.08 | 21.20 ± 0.08 | 0.37 ± 0.03 | 4.37 ± 0.11 | 20.80 ± 0.38 | 2.80 |
| 040-12157  | FM   | 9.8918 | 40.4165 | 16.09 | 6.304 | 21.10 ± 0.05 | 20.92 ± 0.05 | 0.41 ± 0.03 | 4.66 ± 0.10 | 20.40 ± 0.25 | 2.24 |
| 102-03621  | FM   | 11.0967 | 41.9573 | 19.07 | 10.224 | 20.04 ± 0.02 | 19.80 ± 0.05 | 0.19 ± 0.01 | 4.87 ± 0.07 | 19.09 ± 0.18 | 2.14 |
| 103-12114  | FM   | 11.3482 | 42.0301 | 18.53 | 10.370 | 20.73 ± 0.10 | 20.40 ± 0.10 | 0.13 ± 0.02 | 4.32 ± 0.15 | 19.47 ± 0.48 | 1.15 |
| 090-30644  | FM   | 11.0918 | 41.9226 | 18.76 | 66.429 | 18.55 ± 0.02 | 18.20 ± 0.03 | 0.19 ± 0.01 | 4.91 ± 0.04 | 17.21 ± 0.12 | 1.82 |
| 051-15274  | FO   | 9.7241 | 40.7097 | 33.30 | 2.730 | 21.39 ± 0.08 | 21.33 ± 0.08 | 0.08 ± 0.04 | 4.85 ± 0.44 | 21.17 ± 0.39 | 1.14 |
| 090-07109  | FO   | 10.6804 | 41.6543 | 30.35 | 3.663 | 21.23 ± 0.03 | 20.91 ± 0.05 | 0.15 ± 0.04 | 4.35 ± 0.34 | 20.02 ± 0.20 | 1.60 |
| 041-31081  | T2   | 10.5598 | 40.5848 | 29.91 | 21.514 | 21.47 ± 0.03 | 21.23 ± 0.04 | 0.33 ± 0.03 | 5.40 ± 0.09 | 20.36 ± 0.20 | 3.19 |
| 078-12906  | T2   | 10.9476 | 41.4886 | 27.59 | 33.832 | 20.92 ± 0.05 | 20.79 ± 0.07 | 0.44 ± 0.03 | 5.56 ± 0.07 | 20.41 ± 0.33 | 2.01 |
| 103-01636  | UN   | 11.4924 | 41.9421 | 16.31 | 5.251 | 20.39 ± 0.03 | 20.08 ± 0.05 | 0.37 ± 0.03 | 4.14 ± 0.09 | 19.22 ± 0.19 | 3.38 |
| 077-16136  | UN   | 10.5815 | 41.4385 | 28.22 | 26.096 | 20.47 ± 0.21 | 20.08 ± 0.21 | 0.42 ± 0.01 | 4.81 ± 0.04 | 18.98 ± 1.02 | 4.68 |

Notes. The columns contain (from left to right): identifier, Cepheid type, R.A. (J2000.0), decl. (J2000.0), significance (sig), period (determined from the rP1 band light curve), mean magnitude in the rP1 band, mean magnitude in the iP1 band, Fourier coefficient A_{21}, Fourier coefficient ϕ_{21}, Wesenheit index and the decline/rise factor (cf. Figure 21).

Severe effects (≥10% level) for more than 30% of the stars in V and I.

6.1. FO-to-FM Ratio

Interestingly, the relative amount of FO stars as compared to the FM pulsators differs between Vilardell et al. and our findings, namely, 0.27 versus 0.09. Both values have their largest uncertainties in the number of unclassified objects as it is rather uncertain to which of the two pulsational modes these objects belong. Incomplete detection can further add uncertainties. As the Vilardell et al. data were obtained with a slightly larger telescope, and with B, V filter bands, they might have a better completeness for the on average bluer and fainter FO stars. Therefore, it is even more puzzling that the FO-to-FM number ratio based on the David Dunlap Observatory Sample (Fernie et al. 1995) of the Milky Way (~0.06) and of the General Catalog of Variable Stars (Samus et al. 2012) within the Milky Way (~0.13) matches our ratio nicely.

We note that the FO-to-FM ratios as derived from the OGLE surveys of the lower metallicity stars-forming dwarf galaxies LMC, SMC, and IC 1613 yield ratios in the range 0.4 to 0.7 (see, e.g., Udalski et al. 1999b, and references therein, and the OGLE project web pages for the catalogs). Comparing these values to the one of the two spiral galaxies might be indicative for a metallicity trend, despite the rather large scatter of the data in this possible relation. FO Cepheids preferentially populate the blue boundary of the instability strip. As isochrones easily demonstrate (see, e.g., Figure 1 in Becker et al. 1977),

---

Figure 11. Light curves in the rP1 band for a part of the excerpt of the three-dimensional parameter space classified Cepheid catalog. Each top panel shows the unfolded light curve, while each bottom panel shows the variation of the folded light curve over the period times the phase θ. The folded light curve has been continued in both directions for better illustration, only the middle part between the dashed lines shows the real folded light curve. The gray circle data points represent the 2010 season, while the gray triangle data points are within the 2011 season. The Fourier series fit is the black solid line and the black dashed line represents the light curve that is only defined by A_{21}, ϕ_{21}, and the mean magnitude rP1 (cf. Section 4.1). Left panel: light curve of an FM Cepheid with P = 4.398 days. Right panel: light curve of an FM Cepheid with P = 6.304 days.
lower metallicity objects have blue loops in post-main-sequence evolutions which extend far to the blue and cross the complete strip while higher metallicity shrinks the extent more to the red and the strip is no more fully crossed. Therefore, a trend between host metallicity and FO-to-FM ratio is not surprising.

### 6.2. Amplitude Ratio

It is well known that the amplitude of light variations of Cepheids decreases from blue to red wavebands. In several recent investigations of extragalactic samples, the ratio between the amplitudes in two bands has been used to select Cepheid candidates. We do not use the amplitude ratio for selecting good Cepheid candidates, but derive a ratio between the \( r_{P1} \) and \( I_{P1} \) of \( \sim 1.3 \) with a slight dependence on the period (see Figure 22 for our M31 sample). This is in excellent agreement with the high-quality photometry of galactic Cepheids (FM and FO) by Wisniewski & Johnson (1968). From their light curves we find an average ratio in \( R \) and \( I \) of 1.28 \( \pm \) 0.09, and their data points are very well represented by our fit to the M31 sample data (including the slope).

### 6.3. Period Distribution

The FM period distribution of the presented catalog (Figure 23) shows a double-peaked distribution with peaks at \( \sim 0.75 \) and 1.1 in \( \log(P) \) (\( \sim 5.5 \) days and \( \sim 12.5 \) days), respectively. The peaks coincide with the ones found previously by Vilardell et al. (2007). Also, the shape of the distribution agrees quite well with the exception that our sample shows a steeper cutoff to the shortest periods which probably is related to the somewhat shallower PS1 data. Thus, our larger data set fully
supports the more general finding in the Milky Way and M31 that objects of solar metallicity show a complex period distribution as compared to, e.g., the Magellanic Clouds and other low-metal star-forming galaxies (Becker et al. 1977; Vilardell et al. 2007; Alcock et al. 1999).

We note that the period distributions of M 101 and NGC 4258, two other large spirals with more than 800 and 300 detected Cepheids, respectively, show similarly double peaked distribution with a local minimum at \( \sim 10 \) days (Shappee & Stanek 2011; Macri 2004). Thus, this shape of the distribution seems to be a rather common feature of the Cepheid population of large, solar metallicity spiral galaxies. A full discussion of this topic is beyond the scope of this paper.

6.4. PL Relation

The Cepheids of the three-dimensional parameter space classified Cepheid catalog are used to determine the period–apparent-magnitude relations (PL) for the different Cepheid types (Figure 24). To remove systematic outliers that, e.g., result from blending or the wrong extinction correction, we additionally perform iterative 3\( \sigma \) clipping (Figure 25). But in contrast to the 3\( \sigma \) clipping that is used in Section 4.6 for the period–Wesenheit relation, we use the dispersion of the data points relative to the best-fit relation (which we calculate for each iteration step) as an error for each Cepheid magnitude. Otherwise the small errors would result in the elimination of most of the Cepheids, instead of just clipping the systematic outliers. In comparison, the use of the dispersion is not necessary for the period–Wesenheit clipping since the errors in the Wesenheit index are large. The slopes, zero points, the numerical fit errors of these values and the dispersion of the different

---

15 Note that as described in Section 4.2 we only correct for the foreground extinction if there is no color excess given in the Montalto et al. (2009) map.
relations are given in Table 3 for a linear fit of the form

\[ m = a \log(P) + b. \]  

(15)

For comparison we also added the period–Wesenheit relation from Figure 10 and the relations for the manually classified Cepheid catalog. The value of the slope \( a \) and the values for the intersection \( b \) in the usually applied linear approximation of the PLR have been a long-standing matter of debate (see Section 1 and the literature cited therein). Recently, Freedman & Madore (2011) derived slope values from basic physical principles taking into account the surface temperature and effective surface area. These slope values depend on wavelength (expressed as observing band dependence). The predictions of Freedman & Madore (2011) agree rather well with empirical values of Fouqué et al. (2007) (Milky Way Cepheids) and of Ngeow & Kanbur (2007) (LMC) including the wavelength dependence. Bono et al. (2010) obtained slope values from pulsation models, but show an underlying relation with a weak additional curvature term (see their Figure 1). This quadratic behavior is well approximated by a two-slope linear approach with different slope for periods shorter and longer than 10 days (Ngeow et al. 2005, 2008) and are in good agreement with the broken slope proposed by Sandage et al. (2009). Bono et al. (2010) provide their slope predictions as a function of metallicity and waveband for short, long, and overall period range. Di Criscienzo et al. (2012) obtain slope values similar to Bono et al. (2010) from pulsation models for different period ranges in the SDSS filters. The obtained slopes for \( Z = 0.02 \) differ from those obtained by Bono et al. (2010).

The slope values we derive are surprisingly shallow. In Table 3, we also provide slopes for a subsample restricted to
Cepheids with periods longer than 10 days. The slopes of the full sample fits in Table 3 and those for the subsample with \( P > 10 \) days yield basically identical values.

As can be seen in Figure 26, the shallow slope of the subsample with \( P > 10 \) days agrees with the prediction of Bono et al. (2010) for long periods. This is a warning for applying local PLRs to distant galaxies Cepheid samples which have even shallower (absolute) magnitude limits and are even more dominated by long-period Cepheids: the slope values used by Riess et al. (2012) for M31 and by Riess et al. (2011) for nine more distant hosts perfectly agree with our findings and with the predictions of Bono et al. (2010) for long periods. Thus, the effective period range available for a galaxy impacts the derived apparent distance values if the single slope solution for the PL is applied.

In a next step, we want to check if our sample is dominated by long-period Cepheids. If our sample would be dominated by the long-period Cepheids, this would explain why the slopes of our full sample do not agree with the slopes of Bono et al. (2010) for all Cepheids. Therefore, we want to compare the slopes of the short-period (\( P \leq 10 \) days) Cepheids to those of the long-period Cepheids. Simple linear fits to the short-period samples provide almost similar slopes to those of the long-period samples, but with a different zero points. This means that the fits for the different period ranges are not continuous at 10 days. Hence, we introduce a different procedure for the fit of the broken slope proposal. We fit two linear slopes (\( a_1 \) and \( a_2 \)) with a common

---

Table 3

| Relation | Catalog | Type | Slope (mag/log(d)) | Zero Point (mag) | Dispersion (mag) |
|----------|---------|------|--------------------|------------------|------------------|
| PL (rP band) | 3d | FM | -2.39370 ± 0.00110 | 22.66600 ± 0.00380 | 0.42500 |
| PLC (rP band) | 3d | FM | -2.35440 ± 0.00930 | 22.73900 ± 0.03390 | 0.35100 |
| PLP (rP band, P > 10 d) | 3d | FM | -2.39060 ± 0.00160 | 22.66420 ± 0.00980 | 0.55550 |
| PL (rP band) | Manual | FM | -2.39800 ± 0.00130 | 22.63410 ± 0.00490 | 0.40700 |
| PL (rP band) | 3d | FO | -2.56090 ± 0.00370 | 22.23350 ± 0.01800 | 0.35040 |
| PLC (rP band) | 3d | FO | -2.34510 ± 0.02650 | 22.18960 ± 0.12520 | 0.29360 |
| PL (rP band) | Manual | FO | -2.46520 ± 0.00680 | 22.14830 ± 0.02920 | 0.23840 |
| PL (rP band) | 3d | T2 | -2.68030 ± 0.00380 | 25.03480 ± 0.03850 | 0.38690 |
| PLC (rP band) | 3d | T2 | -2.27150 ± 0.02850 | 24.51130 ± 0.28290 | 0.34340 |
| PL (rP band) | Manual | T2 | -2.40420 ± 0.00720 | 24.61830 ± 0.06830 | 0.30380 |
| PL (iP band) | 3d | FM | -2.50490 ± 0.00130 | 22.61890 ± 0.00470 | 0.35710 |
| PLC (iP band) | 3d | FM | -2.48490 ± 0.00830 | 22.66520 ± 0.03000 | 0.31090 |
| PLP (iP band, P > 10 d) | 3d | FM | -2.43540 ± 0.00190 | 22.52990 ± 0.01240 | 0.43030 |
| PL (iP band) | Manual | FM | -2.46910 ± 0.00160 | 22.55290 ± 0.00610 | 0.33780 |
| PL (iP band) | 3d | FO | -2.84470 ± 0.00490 | 23.31730 ± 0.02380 | 0.31750 |
| PLC (iP band) | 3d | FO | -2.64490 ± 0.02290 | 22.22470 ± 0.10760 | 0.25230 |
| PL (iP band) | Manual | FO | -2.94140 ± 0.00840 | 22.36100 ± 0.04000 | 0.24290 |
| PL (iP band) | 3d | T2 | -2.63990 ± 0.00460 | 24.81500 ± 0.04680 | 0.34530 |
| PLC (iP band) | 3d | T2 | -2.57460 ± 0.02710 | 24.72140 ± 0.26840 | 0.32680 |
| PL (iP band) | Manual | T2 | -2.68790 ± 0.00830 | 24.83940 ± 0.07920 | 0.30570 |
| PLW | 3d | FM | -3.00260 ± 0.00620 | 22.57140 ± 0.02180 | 0.44290 |
| PLW | Manual | FM | -3.00380 ± 0.00760 | 22.57660 ± 0.02830 | 0.40260 |
| PLW | 3d | FO | -3.57730 ± 0.02210 | 22.43020 ± 0.10730 | 0.41600 |
| PLW | Manual | FO | -3.59930 ± 0.03830 | 22.50680 ± 0.17810 | 0.39570 |
| PLW | 3d | T2 | -3.12580 ± 0.02140 | 24.98560 ± 0.21670 | 0.51350 |
| PLW | Manual | T2 | -2.89640 ± 0.03900 | 24.60340 ± 0.36910 | 0.49050 |

Notes. For the error bars in the 3\(\sigma\) clipping (PLC) we use the dispersion of the relation which we calculate for each iteration step. Additionally, a fit to a subsample with a period restriction (PLP) for FM Cepheids is provided.

Figure 18. Relative difference of our periods with the 225 (187 FM (circles), 9 FO (upwards pointing triangles), 2 Type II (downward pointing triangles), and 27 UN Cepheids (squares)) matched Cepheids from Vilardell et al. (2007). The light curve of the Cepheid with the largest difference in the period is shown in the right panel of Figure 16. The cascaded form of the relative difference is due to the precision of the compared values.
Figure 19. Color–Wesenheit diagram and color–magnitude diagram for all the 724,894 point sources that were detected in the $r_p$ band. Cepheid variable stars are color coded. Left panel: color–magnitude diagram. Right panel: color–Wesenheit diagram.

Figure 20. Left panel: color–Wesenheit diagram with the domain (black dashed lines) we require the data points to be in, consistent with their margin of error (cf. Section 4.2). Right panel: histogram of the color distribution.

suspension point ($y_{10}$) at 10 days ($x_{10} = 10$ days):

$$y = a_1 \log \left( \frac{x}{x_{10}} \right) + y_{10} \quad x \leq x_{10}$$

(16)

$$y = a_2 \log \left( \frac{x}{x_{10}} \right) + y_{10} \quad x > x_{10}.$$  

(17)

For the FM Cepheids in the $r_p$ band this results in

$$y_{10} = 20.26755 \pm 0.000004$$

$$a_1 = -2.41618 \pm 0.00007$$

$$a_2 = -2.37669 \pm 0.00004$$

with a dispersion of 0.36345 for the short periods and a dispersion of 0.55620 for the long periods. This fit is shown in Figure 27. As can be seen also this fit with a common suspension point does produce similar slopes for the short- and long-period Cepheids.

Cepheids stay the same, but the magnitude is given by a Gaussian distribution around the slope $-2.8$. The width of the Gaussian distribution is chosen to be the dispersion of the common suspension fit of the short-period Cepheids. The subsample of the long-period Cepheids stays the same. We generate 100,000 Gaussian distributions and perform the common suspension point fit. The resulting distribution of the short-period slopes is shown in Figure 28. As can be seen the shallow slope of the short-period subsample cannot be explained by the dispersion of the PLR.

The most likely explanation for the shallow slope at short periods is that we seem to select the brighter Cepheids at short periods due to the Malmquist bias and that most short-period Cepheids are too faint to be detected, so that we cannot determine the slope for the short periods correctly. To confirm this we would need to perform completeness tests that are beyond the scope of this work.

Nevertheless, the good agreement of our empirical values for the long-period Cepheids with the predictions of Bono et al. (2010) supports the finding of Sandage et al. (2009) of a different slope for long- and short-period Cepheids, depending on the observational band, and the more general prediction of a curvature term in the PLR.
1.0 10.0 100.0

Figure 21. Decline/rise factor for the three-dimensional parameter space classified Cepheid catalog. This factor describes the ratio of the time from the minimal to the maximal magnitude (not flux!) compared to the time from the maximal to the minimal magnitude and it is a good indicator of the overall shape of the light curve. We observe different scopes that correspond to the position of the bump. The 14 Cepheids framed by a black circle are shown in Figures 11–17.

Figure 22. Amplitude ratio for the three-dimensional parameter space classified Cepheid catalog. We observe a slight slope ($A_{R1}/A_{P1} = a + b \log(P)$) with $a = 1.309 \pm 0.006, b = -0.0478 \pm 0.0014$ and a dispersion of 0.1680.

6.5. Spatial Cepheid Distribution

We now investigate the location of the Cepheids in the M31 plane. Figure 29 shows the position of the Cepheids plotted over the $E(B-V)$ map of Montalto et al. (2009). The FM and FO Cepheids are concentrated toward the disk and clearly show a ring like structure (left panel in Figure 29), while the older Type II Cepheids trace the halo of M31 (right panel in Figure 29). Figure 30 shows a contour plot of the distribution of the Cepheids. While we can see the same behavior of the FM and FO Cepheids as in Figure 29, we observe that the Type II distribution is limited by the area of the skycells that were analyzed. For the next data release, we will increase this area so as to better trace the halo. We do not discuss the spatial distribution of Cepheids around NGC 206 but refer the reader to Magnier et al. (1997).

6.6. Spatial Age Distribution

The spatial distribution of FM and FO Cepheids can be used to obtain a spatial age distribution for M31. We use the period–age relation for FM and FO Cepheids given by Bono et al. (2005) to determine the age of each Cepheid. The spatial age distribution shown in Figure 31 indicates that star formation in the last $\sim 100$ Myr was concentrated in a ring which is in good agreement with Davidge et al. (2012). Figure 31 also hints at a correlation between the Cepheid age and the distance to the center of M31.

In order to check for this correlation we deproject the spatial distribution. We use 75° for the inclination and 37° for the tilt angle and the radius and offset of the 10 kpc ring given by Gordon et al. (2006). The deprojected age map is shown in Figure 32. We found the splitting of the ring as previously described in Gordon et al. (2006). Gordon et al. (2006) attribute the star formation of the ring and its splitting to a passage of M32 through the M31 disk.

In a next step, we analyze the age of the Cepheids as a function of distance to the 10 kpc ring. For this analysis we exclude all Cepheids in the splitted part of the ring (150 $\leq \varphi \leq 260$, cf. right panel in Figure 32). Figure 33 shows the median Cepheid age as a function of the distance to the 10 kpc ring in bins of 0.5 deg. The first and the last two bins contain less than 10 Cepheids and can therefore be neglected.

The errors in each bin have been determined with the bootstrap method (cf. Section 3) and are rather small compared to the dispersion in the bins we consider. The upper left corner of the upper panel of Figure 33 contains no Cepheids. In this region (the center of M31) the signal-to-noise ratio is low, so that this is likely the reason that we hardly detect faint Cepheids (i.e., Cepheids with small periods; the study of the detection efficiency for Cepheids will be subject of a further work), and thus we hardly detect old Cepheids in this region.

In the seventh bin, we detect old (faint) Cepheids but no young
Figure 24. Period–apparent-magnitude relations for the three-dimensional parameter space classified Cepheid catalog. The according fits shown as solid lines are given in Table 3. Left panel: period–apparent-magnitude relation in the $r_{P1}$ band. Right panel: period–apparent-magnitude relation in the $i_{P1}$ band.

Figure 25. Period–apparent-magnitude relations with iterative $3\sigma$ clipping. For the error bars in the $3\sigma$ clipping we use the dispersion of the relation which we calculate for each iteration step. The according fits shown as solid lines are given in Table 3. Left panel: period–apparent-magnitude relation in the $r_{P1}$ band. Right panel: period–apparent-magnitude relation in the $i_{P1}$ band.

Figure 26. PLR slope values as function of wavelength. Our full sample and subsample with $P > 10$ days yield basically identical slopes (cyan). They agree with the theoretical predictions of Bono et al. (2010) for long periods (brown).

Figure 27. Common suspension point fit for FM Cepheids in the $r_{P1}$ band (cf. Equations (16) and (17)). The resulting slopes for the short-period and long-period subsample are very similar.
Cepheids, although they should be more easily detectable than old Cepheids. Therefore, the lack of young Cepheids in the outer region of the ring is no selection bias. This leaves us with the conclusion that the age gradient is real. To measure the strength of the age gradient we perform a fit to the median age with

\[ y = a + bx \]

\[ a = 55.4941 \pm 0.7692 \text{ (Myr)} \]

\[ b = 33.8517 \pm 0.6100 \text{ (Myr deg}^{-1} \text{)} \]

with \( (\Delta \chi^2 / \text{dof}) = 0.36 \). We therefore conclude that the median age of the stellar population decreases by \( \sim 34 \text{ Myr/deg inward.} \) A possible interpretation for this is that the star formation is related to the interaction with M32 moving inward.

7. CONCLUSION AND OUTLOOK

We present a large sample of Cepheids in M31 in the \( r_{P1} \) and \( i_{P1} \) filters. We develop an automatic Cepheid detection and classification scheme based on the Fourier parameters \( P \), \( A_21 \), and \( \phi_21 \) and the location of the instability strip in the Wesenheit–color plane. This makes the Cepheid detection and classification less biased than traditional methods and facilitates Cepheid detection in large surveys.

We find 1440 FM Cepheids, 126 Cepheids in the FO mode, 147 belonging to the Population II types, and 296 Cepheids that could not be classified. Three hundred fifty-four of the 2009 Cepheids could be found in other surveys (Vilardell et al. 2007; Fliri et al. 2006; and the DIRECT project, Macri 2004) with a
The spatial age distribution for the three-dimensional parameter space classified Cepheid catalog for different Cepheid ages, plotted over the $E(B - V)$ map of Montalto et al. (2009). To calculate the age we use the period–age relations from Bono et al. (2005). The spatial age distribution shows that the star formation in last $\sim$100 Myr was concentrated in a ring, which is in good agreement with Davidge et al. (2012). Blue points: FM and FO Cepheids with $t \geq 70$ Myr; green upward pointing triangle: FM and FO Cepheids with $70 \text{ Myr} > t \geq 40$ Myr; red downward pointing triangle: $t < 40$ Myr. Left panel: R.A.–decl. distribution. Right panel: R.A.–decl. distribution with the median age for each bin.

![Spatial age distribution for the three-dimensional parameter space classified Cepheid catalog for different Cepheid ages, plotted over the $E(B - V)$ map of Montalto et al. (2009). To calculate the age we use the period–age relations from Bono et al. (2005). The spatial age distribution shows that the star formation in last $\sim$100 Myr was concentrated in a ring, which is in good agreement with Davidge et al. (2012). Blue points: FM and FO Cepheids with $t \geq 70$ Myr; green upward pointing triangle: FM and FO Cepheids with $70 \text{ Myr} > t \geq 40$ Myr; red downward pointing triangle: $t < 40$ Myr. Left panel: R.A.–decl. distribution. Right panel: R.A.–decl. distribution with the median age for each bin.](image)

Figure 31.

The spatial distribution of FM and FO Cepheids follows the 10 kpc ring in M31, while the Type II Cepheids are distributed throughout the halo of M31. The spatial age distribution shows a positive age gradient with increasing distance from the 10 kpc ring, which implies that the star formation in the 10 kpc ring moved inward.

The analysis of our PLRs indicates that the PLR is slightly curved. This has ramifications on the determination of extragalactic distances, since most of the commonly used calibrations of the PLR are dominated by the short-period Cepheids and the typical Cepheids that are used for the extragalactic distance determination are long-period Cepheids.

The next data release will cover more area of the M31 halo, thus we will be able to trace the halo better. With another year of observations we will be able to extend the sample to larger periods. Overlapping skycells will improve the calibration and enable completeness tests.

The complete catalog of $r$ and $i$ light curves for the 2009 Cepheids along with cross-identification tables can be found in the .tar.gz package in the online journal.

We acknowledge the careful reading and helpful comments by the anonymous referee. The Pan-STARRS1 Surveys (PS1) have been made possible through contributions of the Institute for Astronomy, the University of Hawaii, the Pan-STARRS Project Office, the Max-Planck Society and its participating institutes, the Max Planck Institute for Astronomy, Heidelberg, and the Max Planck Institute for Extraterrestrial Physics, Garching, The Johns Hopkins University, Durham University, the University of Edinburgh, Queen's University Belfast, the Harvard-Smithsonian Center for Astrophysics, the Las Cumbres Observatory Global Telescope Network Incorporated, the
Figure 33. Age distribution as a function of distance to the 10 kpc ring. The definition of the symbols and colors is the same as in Figure 32. The split part of the ring (150 \( \leq \varphi \leq 260\), cf. right panel in Figure 32) was excluded. In the top panel the age distribution for the FM and FO Cepheids can be seen. The bottom panel shows the median age for bins with width of 0.5 deg. The first and the last two bins contain less than 10 Cepheids and thus can be neglected. The errors of the medians have been determined with the bootstrap method (cf. Section 3). The dispersion of each bin is shown in black. We observe an age gradient and the fit to the median age (\( y = a + bx \)) is shown as a solid magenta line. The age gradient suggests that the star formation related to the interaction to M32 moved inward. No Cepheids can be found in the top left corner of the upper panel (the center of M31). We hardly detect old (faint) Cepheids in this region, most likely due to the low signal-to-noise ratio.