

Bias in the Cepheid PL relation

P. Lanoix¹, G. Paturel and R. Garnier

CRAL (UMR 5574 CNRS) - Observatoire de Lyon

F69230 Saint-Genis Laval, FRANCE

ABSTRACT

We show that the Cepheid PL relation is affected by a Malmquist type bias, the so-called *population incompleteness bias*. Its calculated slope appears shallower than the true one because of the cutoff in apparent magnitude resulting from the instrumental limiting magnitude. Furthermore, the use of the PL relation, even with the correct slope, leads to underestimation of distances which is not negligible. We confirm this finding by studying simulated PL relations, and we show that this bias may be as large as 0.2 or 0.3 magnitudes on distance modulus. We also test the efficiency of a cutoff in $\log P$ and show that it is a good way to minimize this bias. However, a correction of this is difficult as long as the completeness of the sample is not perfectly well established.

Subject headings: Cepheids — distance scale

1. Introduction

The Cepheid Period-Luminosity relation (hereafter PL relation) is one of the major tools used for computing extragalactic distances from data like those provided by the *Hubble Space Telescope* (HST).

However, Teerikorpi (1984) has pointed out for a long time that such a linear relationship is prone to be affected by biases. For extragalactic Cepheids a bias is expected, identical to that encountered in a cluster of galaxies, the so-called *population incompleteness bias* (Teerikorpi 1987, Bottinelli et al. 1987).

Sandage (1988) has already noticed that truncating a complete sample of Cepheids in LMC leads to too shallow a slope of the PL relation. We intend to show that HST observations, although much deeper in apparent magnitude, are affected by this kind of bias. This is because a sample of Cepheids in a given external galaxy (thus, at the same distance) is complete only at the bright

¹Université Claude Bernard Lyon I F69622 Villeurbanne, FRANCE

end of the Cepheid luminosity function (i.e. only for large $\log P$ values), since observations are limited to a given apparent limiting magnitude.

After introducing the data and the method used (section 2), we investigate how such a bias affects the slope, a , (section 3), by studying NGC 4536 as an example. We then highlight the existence of the bias in this way. In section 4 we analyze the effect on the zero-point b , using the correct slope of the PL-relation. Section 5 presents a simulation that confirms previous results and gives an estimate of the bias, as well as its reduction by applying a cutoff in $\log P$.

It is important to address this question because its effect is poorly studied at the present time, and, yet, is larger than generally admitted.

2. Data and distance moduli

We use a sample of 750 Cepheids which have both V and I photometry available from literature. These Cepheids are located in 23 galaxies, 14 of them have Cepheids observed with HST whereas the other 9 galaxies have only ground based observations. We checked all light curves in order to detect overtone pulsators (low amplitude, $\log P < 1$ and symmetrical curves) and to remove them from our sample since they are subject to a different PL relation (we didn't try to correct their period).

In order to compute the distance moduli of these galaxies, we choose the absolute calibration of the PL relation from Gieren et al. (1998). Their PL calibration is based on infrared Barnes-Evans surface brightness technique (note that it is insensitive to both Cepheid metallicity and reddening). We then avoid comparing PL relations of different galaxies with different metallicities and inaccurate reddenings. The PL relations are :

$$M_V = -2.769 \log P - 1.294 \quad (1)$$

$$M_I = -3.041 \log P - 1.726 \quad (2)$$

We can compute easily the V and I apparent distance moduli for each Cepheid :

$$\mu_V = V - M_V \quad (3)$$

$$\mu_I = I - M_I \quad (4)$$

The supposed true distance moduli are then :

$$\mu_0 = \mu_V - R(\mu_V - \mu_I) \quad (5)$$

where

$$R = \frac{A_V}{A_V - A_I} = 2.446 \quad (6)$$

according to Cardelli et al. (1989), A_V and A_I being respectively the V and I extinction coefficients. The supposed true distance modulus of a galaxy is then assumed to be the mean of the individual distance moduli of its Cepheids.

3. Effect on the PL relation slope

The V-band slope is very well defined from ground-based observations of LMC Cepheids. Gieren et al. (1998) found $a = -2.769 \pm 0.073$. Tanvir (1998) found $a = -2.774 \pm 0.083$ and Madore & Freedman (1991) found $a = -2.76 \pm 0.11$. We will adopt :

$$a = -2.77 \pm 0.09 \quad (7)$$

The slope of the PL relation cannot be determined from distant galaxies. We almost always obtain apparent slopes (a') significantly shallower than the LMC one (assumed to be the true one). Yet, the slope of the PL relation does not vary with metallicity. This is predicted by models (Chiosi et al. 1993) and is widely acknowledged even by people who claim that metallicity effects are strong (Beaulieu et al. 1997). Therefore, we cannot argue about metallicity effects.

The low value of a' can actually be traced to the population incompleteness bias resulting from apparent limiting magnitudes ($V_{lim} \approx 26.5$ for HST measurements and $V_{lim} \approx 21.5$ for ground-based observations). Figure 1 and 2 show respectively the V and I PL relations for one of the best measured galaxies, NGC 4536, and illustrate how the bias works. For instrumental reasons, apparent magnitudes are limited to a given limit V_{lim} . Thus, the distribution in the plot V against $\log P$ (or similarly M_V against $\log P$ for a given distance modulus) is distorted. If we force a linear regression, the line will tend to pass by the point having the largest inertia (note A in fig. 1 and 2). This leads to a shallower slope. Assuming a Gaussian distribution of residuals (as a first approximation of top-hat distribution), the model given by Paturel et al. (1997) predicts a slope a' :

$$a' = \frac{a}{1 + \sigma / (V_{lim} - \mu - M_{Max})} \quad (8)$$

where a is the unbiased slope, σ is the scatter of the V band PL relation in magnitudes ($\sigma \approx 0.3$ mag.), μ is the adopted distance modulus and M_{Max} the brightest end of absolute V magnitudes ($M_{Max} \approx -7$). Figure 3 gives the predicted variation of a' with the distance modulus for both ground-based and HST observations. On the same figure we give the observed slopes obtained by a direct regression for Cepheids in 23 galaxies (ground-based and HST observations are represented with different symbols). The agreement fully supports the interpretation of the bias and proves in this way that PL relations in external galaxies are definitely affected by incompleteness bias. It must be noted that the slope dramatically increases with distance modulus, preventing us from using an observed slope without caution.

The inverse regression line is a way which should lead to the correct result (Kelson et al. 1996). Unfortunately, it is subject to the opposite bias (bias in $\log P$) because very long period Cepheids are difficult to detect. Furthermore, the inverse slope varies with the scatter of the sample prohibiting the direct comparison with calibrating PL relation whose scatter is smaller. This situation is identical to that encountered for the inverse Tully-Fisher relation (see Teerikorpi 1990).

4. Effect on the zero-point

We would like to stress the most important consequence of the bias. Even with the right slope, the determination of the distance modulus will be underestimated when data are limited in apparent magnitude. This fact can easily be understood from the distribution of the points in figures 1 and 2. Because of the cutoff in magnitude there are more points above the mean PL line for short periods. Of course, since the residuals on μ_V and μ_I are correlated, the effect of incompleteness is reduced in the second part ($\mu_V - \mu_I$) of the equation (5). However, the effect on the first part (μ_V) of this equation remains affected by incompleteness, and this term is moreover one order larger than the second one ($R(\mu_V - \mu_I)/\mu_V < 5\%$).

In order to highlight this effect, the mean distance modulus is calculated for NGC 4536 by progressively removing the shortest period Cepheids. Figure 4 shows the variation of the mean distance modulus with the lower $\log P$ cutoff ($\log P_l$). The first point (on the left) represents the mean distance modulus of NGC 4536 when using its 73 Cepheids *vs.* the lower $\log P$ of these 73 Cepheids. The next point represents the mean distance modulus *vs.* the lower $\log P$ of the 72 Cepheids having the larger periods. The same applies to 71 Cepheids until 1 Cepheid. The effect on the mean distance modulus as a function of period cutoff is then directly readable on this figure. Presently, the detection of a plateau in the growth curve of mean distance modulus seems the most efficient way to get unbiased distance moduli from Cepheids PL relation. The beginning of the plateau indicates the cutoff in $\log P$ which should be applied ($\log P_l = 1.25$ in this case). In this example, the biased modulus equals to 30.50 while the unbiased modulus is 30.76. This effect is therefore, far from being negligible. The rule of thumb for calculating unbiased distance moduli is simply to use only $\log P$ values larger than the given limit $\log P_l$. This limit can also be calculated from figure 1:

$$\log P_l = \frac{V_{lim} - \mu - b - 2\sigma}{a} \quad (9)$$

For NGC 4536, using the PL relation given in equation (1), and the same parameters as in relation (8), $V_{lim} = 26.5$ and $\sigma = 0.3$, we obtain $\log P_l = 1.19$ for $\mu = 30.50$. From the growth curve of distance modulus (fig. 4) it appears that this limit is quite acceptable. The definition of V_{lim} is namely crucial and difficult.

5. Simulation

5.1. Construction of simulated PL relations

In order to confirm that our interpretation of the incompleteness bias is the right one, we construct simulated PL relations in the following way : for each Cepheid of a simulated galaxy, we first choose a true distance modulus (μ_{True}) and we determine randomly according to a Gaussian shape the following values :

- $\log P$; ($\langle \log P \rangle = 1.0, \sigma = 0.35$)

- $E_{B-V} = (B - V) - (B - V)_0$; ($\langle E_{B-V} \rangle = 0.08, \sigma = 0.04$)

Using equations (1) and (2) we compute the V and I absolute magnitudes of the Cepheid¹ and add to them a random dispersion Δ , intrinsic dispersion of PL relations with $\langle \Delta \rangle = 0$ and $\sigma_\Delta = 0.3$. We then compute and add V and I absorptions according to $A_V = 3.3E_{B-V}$ and $A_I = 1.95E_{B-V}$. We add random error measurements, ϵ , calculated as a linear function of the distance modulus, $\epsilon = 0.2 - (32 - \mu_{True})/45$, with no correlation between V and I, and obtain :

$$V = \mu_{True} - 2.769 \log P - 1.294 + \Delta + A_V + \epsilon_V \quad (10)$$

$$I = \mu_{True} - 3.041 \log P - 1.726 + \Delta + A_I + \epsilon_I \quad (11)$$

Finally, we compute the probability for the Cepheid to be detected both in the V and I bands. We draw a random parameter $t \in [0, 1]$ and compute the quantity :

$$t_0 = \frac{1}{1 + e^{\alpha(V - V_{lim})}} \quad (12)$$

Whenever $t \leq t_0$, the Cepheid may be observed by the HST and we keep it in our simulated galaxy. In the other case it will be rejected. We assume $\alpha = 3$ and $V_{lim} = 26.5$ for HST measurements. We also assume that a typical galaxy has 250 Cepheids, so that about 80 may be detected for $\mu_{True} \approx 30.5$

5.2. Results

We first simulate a galaxy similar to NGC 4536 by locating it at a true distance modulus $\mu_{True} = 30.76$. Figures 5 and 6 show respectively the kind of absolute V and I PL relations we obtain. This simulated galaxy has for instance 77 observable Cepheids and is comparable in every respect to NGC 4536. We can reduce these data to compute its distance modulus (μ_0) according to the procedure described in section 2.

We find :

$$\mu_0 = 30.59 \quad (13)$$

And then the bias is :

$$\mu_0 - \mu_{True} = -0.17 \quad (14)$$

At the very least, when using HST data, authors involved in the HST Key Project make a cut at $\log P = 1$. When we choose such a lower limit in our simulation, the bias induced is worth -0.12 mag. Figure 7 shows the variation of its computed distance modulus as a function of $\log P_l$. The behaviour is the same as the one observed for NGC 4536 (see fig. 4) and, as such, it confirms our interpretation. We plot the plateau on figure 7, which is obtained with a cutoff at $\log P_l = 1.24$

¹note that the color distribution $(V - I)_0 = 0.272 \log P + 0.432$ is underlying

(then $\langle \log P \rangle = 1.45$).

If we compute the theoretical cutoff according to equation (9), we obtain $\log P_l = 1.23$ in close agreement with the real value. This formula (9) may be very useful whenever the identification of the plateau is difficult.

Moreover, in order to make sure that the result does not depend on this particular random draw, we prepare a random set of 200 galaxies and we place them at increasing distance moduli.

We first test the variation of the slope of the PL relation : figure 8 shows the variation of the slope as a function of μ_{True} . This figure is directly comparable to figure 3 and fully supports our interpretation.

Of greater significance is our evaluation of the bias on the mean distance moduli $\langle \mu_0 \rangle$ from the same set of galaxies as a function of μ_{True} when the correct slope is used. One can see that the bias is indeed far from negligible (fig. 9) . It reaches :

$$\langle \mu_0 \rangle - \mu_{True} = 0.20 \pm 0.01 (\sigma = 0.20) \quad (15)$$

for $\mu_{True} = 32.2$ and $V_{lim} = 26.5$.

We also test the efficiency of the bias correction by applying a $\log P$ cutoff. We use our model (see eq. [9]) and compute again $\langle \mu_0 \rangle - \mu_{True}$ (open dots in fig. 9). The importance of the bias is strongly reduced in this manner although it still exists. Clearly, the bias depends critically on the limiting V_{lim} . Our simulation is made with a constant $V_{lim} = 26.5$, while, e.g. in the Key Project conducted with the HST, distant galaxies have observations taken with longer exposure times leading to fainter V_{lim} (for galaxies in Virgo and Fornax clusters, the limiting magnitude may be between 26.5 and 27 mag). This will infer different biases depending on the actual limiting magnitude.

From a practical standpoint, one way of taking this bias into account might be to cut systematically the studied sample in $\log P$ according to equation (9). For instance, the effects on the distance moduli of NGC 4321 and NGC 2541 are respectively 0.09 and 0.05 magnitudes. Another approach, more statistically oriented, could be to use values given in figure 9 and to add the value of the bias we compute to the distance modulus obtained from the complete sample.

6. Conclusion

In conclusion, we draw the attention to the difficulties involved in using the PL relation. The slope cannot be determined without considering the population incompleteness bias. Furthermore, it is most important to note that the use of the relation, even with the correct slope, should at least be limited to long period Cepheids. The correction for bias effect is difficult as long as the completeness of the sample is not perfectly well established. We consider, however, that this incompleteness bias should be systematically taken into account when deriving extragalactic distances from Cepheid PL relation, for instance according to the equation (9), even though it may be of little importance in some cases.

We would like to thank Joseph Necker Lanoix for his useful corrections, as well as the anonymous referee who helped us to improve the paper.

REFERENCES

- Beaulieu, J. P., Sasselov, D., Renault, C. et al. 1997 A&A, 318, L47
- Bottinelli, L., Fouqué, P., Gouguenheim, L., Paturel, G., & Teerikorpi, P. 1987, A&A, 181, 1
- Cardelli, J. A., Clayton, G. C., & Mathis, J. S. 1989, ApJ, 345, 245
- Chiosi, C., Wood, P. R., & Capitanio, N. 1993, ApJS, 86, 541
- Gieren W., Fouqué, P., & Gómez M. 1998, ApJ, 496, 17
- Kelson, D. D. et al. 1996, ApJ, 463, 26
- Madore, B. F., & Freedman, W. L. 1991, PASP, 103, 933
- Paturel, G., Lanoix, P., Garnier, R. et al. 1997, in Proceedings of the ESA Symposium "Hipparcos Venice '97" (eds. Perryman, M. A. C., & Bernacca, P. L. ESA SP-402)
- Sandage, A. 1988, PASP, 100,935
- Tanvir, N. 1998, Proceeding of the STScI symposium on the "Extragalactic Distance Scale"
- Teerikorpi, P. 1984, A&A, 141, 407
- Teerikorpi, P. 1987, A&A, 173, 39
- Teerikorpi, P. 1990, A&A, 234, 1

Fig. 1.— See next figure.

Fig. 2.— NGC 4536 V and I band absolute PL relations. The absolute magnitudes are simply $M_V = V - \langle \mu_V \rangle$ and $M_I = I - \langle \mu_I \rangle$, $\langle \mu_V \rangle$ and $\langle \mu_I \rangle$ being respectively the mean V and I apparent distance modulus of the considered galaxy. The solid line corresponds to the adopted PL relation (eq. [1] and [2]). The dotted lines give the corresponding $\pm 2\sigma$ limits. The dashed line gives the relation calculated from a direct regression. The horizontal line gives the approximate cutoff in absolute magnitude resulting from the instrumental limiting magnitude. The linear regression (dashed line) forced onto the distorted distribution tends to pass through point A, leading to too shallow a slope.

Fig. 3.— Apparent slope a' of the PL relation as a function of the distance modulus. The trend is typical of a statistical bias. Open stars represent ground-based observations whereas filled stars correspond to HST observations. The solid curves give the predicted variation in both cases from our model (see eq. [8]).

Fig. 4.— Variation of the mean distance modulus *vs.* $\log P_l$ for NGC 4536. The mean distance modulus depends on the adopted cutoff in $\log P$ (see text).

Fig. 5.— See next Figure.

Fig. 6.— Simulation of V and I band absolute PL relations. The simulated galaxy is located at $\mu_{True} = 30.76$. The solid line corresponds to the adopted PL relation (see eq. [2] and [1]). The dotted lines give the corresponding $\pm 2\sigma$ limits. The dashed line gives the relation calculated from a direct regression.

Fig. 7.— Variation of the mean distance modulus *vs.* $\log P_l$ for a typical simulated galaxy located at $\mu_{True} = 30.76$. The mean distance modulus depends on the adopted cutoff in $\log P$ (see text).

Fig. 8.— Simulation of the bias affecting the slope of the PL relation as a function of the distance modulus. The plotted slope is the mean value of a set of 200 simulated galaxies having increasing distance moduli. The horizontal dotted line indicates the unbiased value. The curve represents our model given by equation (8).

Fig. 9.— Simulation of the bias we made while computing the distance modulus of a galaxy, because of incompleteness. Filled dots represent the mean error on $\mu_0 - \mu_{True}$ as a function of μ_{True} , for a set of 200 galaxies, when keeping all observed Cepheids in the calculation. Open dots represent the same quantity when applying a cutoff in $\log P$ according to equation (9). One can see that this procedure significantly reduces the error on the distance modulus.

















