Astrophysical validation of Gaia parallaxes

Fredrik Windmark

Lund Observatory
Lund University

2010-EXA39

Degree project of 60 higher education credits (for a degree of Master)
June 2010

Lund Observatory
Box 43
SE-221 00 Lund
Sweden

Acknowledgements

I want to thank my supervisor Lennart Lindegren for his help and for sharing his great knowledge and experience. I also want to thank my co-supervisor David Hobbs for his help with the project in general, and the Java programming in particular. Working with this Masters project has been an incredibly fun and rewarding time.

Big thanks also go to Berry Holl for his help with simulating the Gaia parallax measurements. I have also had countless helpful discussions with my office mate Tobias Albertsson, to whom I owe a lot. Worth mentioning are also my fellow Master students; Nils Håkansson, Hannes Jensen and Carina Lagerholm, who have been really helpful and have had to put up with all my rambling during these two years. My girlfriend Maria Ewerlöf has also been incredibly patient with me, and her support has meant a lot.

Abstract

The Gaia satellite, to be launched in August 2012, will measure highly accurate absolute parallaxes of hundreds of millions of stars. This is done by comparing parallactic displacement of stars in different parts of the sky. The accuracy of this method highly depends on the stability of the so-called basic angle between the two fields of view of the Gaia instrument, and periodic variations could lead to a global zero-point error in the measured parallaxes. Small variations of the basic angle are closely monitored by on-board instruments, but independent verification methods are also needed.

In this project, we use Galactic Cepheid variables as standard candles to compare with the observed parallaxes at a wide range of distances. If there is a parallax zero-point error, the observed parallaxes will not be consistent with a single Period-Luminosity relation. A model is formulated where the complete Galactic Cepheid population is generated and observed in a simulated Gaia mission. Using the observed Cepheids, we then make simultaneous fits to the P-L relation and the parallax zero-point in order to determine whether using Cepheids is a viable zero-point verification method.

Our simulations show that Gaia will observe about 9000 Galactic Cepheids, fifteen times the currently known number. Gaia will alone result in large improvements in the accuracy with which the Galactic P-L relation can be determined. Both constants in the relation can be determined with an accuracy of $\sigma_{a,b} < 0.05$. We show that using Galactic Cepheids, the parallax zero-point can be determined with an accuracy of $\sigma_c = 0.3$ μas, with the largest error contribution coming from the uncertainty with which we can determine the extinction. This is very good, but not enough for the most demanding tasks of Gaia. We conclude that the global verification of the parallax zero-point ultimately will depend on a combination of many different methods.

Sammanfattning

I augusti 2012 kommer Gaia-satelliten, en rymdsond utvecklad av ESA, att skjutas upp för att under fem år observera hundratals miljoner stjärnor. Gaia kommer bland annat att mäta parallaxen, eller avståndet, till alla dessa stjärnor med en noggrannhet som är tusen gånger bättre än dess föregångare Hipparcos. Antalet stjärnor kombinerat med den stora noggrannheten innebär att Gaia kommer att utföra den största och mest noggranna kartläggningen av Vintergatan någonsin. Detta kommer garanterat att leda till ett otal vetenskapliga upptäckter, men för att med säkerhet kunna använda datan är det viktigt att på något sätt verifiera att mätningarna är sanna.

I det här arbetet undersöker vi om denna verifiering skulle kunna ske med hjälp av så kallade cepheider i Vintergatan. Cepheider är jättestjärnor som varierar i storlek och ljusstyrka med en period som beror på hur stor massa de har. Detta innebär att man genom att mäta hur lång en cepheids period är kan bestämma dess avstånd från oss utan att behöva mäta parallaxen. Denna egenskap har gjort cepheiderna till en av de viktigaste metoderna för att kunna mäta avstånd utanför vår egen galax, och en stor del av vår uppfattning om universum beror idag på dem. Det borde även vara möjligt att använda dem för att bekräfta Gaias parallaxmätningar.

För att kunna avgöra hur bra cepheider är för detta ändamål måste vi veta hur många Gaia kommer att observera i Vintergatan. Eftersom vi bara känner till de cepheider som ligger allra närmast solen, och Gaia kommer att kunna se även dem i andra änden av galaxen, så måste vi simulera hur cepheidernas fördelning i Vintergatan kan tänkas se ut. Sedan, eftersom Gaia fortfarande inte blivit uppskjuten, måste vi även simulera Gaias observationer.

Med våra simulationer visar vi att Gaia kommer att observera fler än 9000 cepheider i Vintergatan, vilket kan jämföras med de 600 man känner till idag. Cepheid-metoden kommer att kunna bekräfta Gaias parallaxmätningar med en noggrannhet på ungefär 0.3 mikrobågsekunder. Detta är mycket bra, men inte tillräckligt för de allra mest krävande uppgifter som Gaia är kapabel till att utföra. Verifieringen av Gaias parallaxer kommer troligen inte att kunna ske med hjälp av en enda metod, utan måste nog snarare ske med ett antal olika, där Vintergatans cepheider kommer att spela en viktig roll.

Contents

1 Introduction

The Gaia satellite, due for launch in August 2012, is the successor to the successful Hipparcos mission, an ESA space astrometry mission that was active in the early 90's. The Hipparcos data had a very large impact on the world of astronomy, with its primary catalogue containing approximately 120 000 stars covering the whole sky with a median parallax accuracy of 1.1 milliarcseconds (mas) (Perryman et al. 1997). The advantage of working with large, homogeneously determined data sets is clear. Even today, the Hipparcos mission remains the largest astrometric all-sky survey. This will change with the advent of Gaia, which will result in a catalogue containing roughly a billion objects with parallax accuracies reaching below 10 microarcseconds (μas) (Lindegren & Perryman 1996; Lindegren 2010). With a catalogue ten thousand times larger and a hundred times more accurate than what we currently have, along with simultaneous astrometric, photometric and spectroscopic observations, it is safe to say that Gaia will result in a revolution in the understanding of stellar and Galactic dynamics, formation and evolution (Perryman et al. 2001).

Achieving the desired accuracy requires an exceedingly stable optical instrument for the Gaia satellite, as even extremely small variations in the basic angle could lead to an undesirable global shift in the parallax zero-point. To be able to determine the absolute parallax of an object, Gaia simultaneously observes stars in two regions on the sky that are separated by a large basic angle. The two fields of view cross the same part of the sky with a separation in time of the order of a few hours, and it is the relation between the parallaxes measured in the two fields that lies behind Gaia's ability to do global astrometry and to determine absolute parallaxes. The stability of the basic angle is therefore of great importance to avoid introducing errors in the parallax measurements. As the satellite rotates, however, different parts will be exposed to solar heating. This will inevitably lead to basic angle variations on the scale of hours, creating apparent image shifts that may be indistinguishable from a global offset of all parallaxes (Lindegren 2004). An on-board laser interferometer is therefore used to measure these variations so that they can be included in the instrument calibration model. However, it is still desirable to verify the parallax accuracy, and in particular the parallax zero-point, by independent astrometric means.

One example where the knowledge of the Gaia parallax bias is of great importance is for the use of distance determination to the Large Magellanic Cloud (LMC). This is very important to the extragalactic distance scale, and its distance of around 50 kpc (or 20 μas) is today known with an uncertainty of 5%. It is believed that Gaia will observe 10^7 stars in the LMC with a mean standard error of about 200 μas, which would result in a mean parallax of $200/\sqrt{10^7} \approx 0.06$ μas, corresponding to a relative precision of 0.3%. In reality, the precision with which Gaia can determine the LMC distance is limited by the bias. It is therefore of great interest to be able to verify a Gaia parallax zero-point bias below 0.1 μas.

In this project, we investigate the feasibility of using Cepheid variables to verify the parallax zero-point of Gaia. The Period-Luminosity (P-L) relation has been known for a very long time (Leavitt 1908; Leavitt & Pickering 1912) and its calibration is still of great interest today (Sandage et al. 2004; Fouqué et al. 2007; Ngeow et al. 2009). It shows a clearly defined relation between the period and luminosity of Cepheids, making them excellent standard candles, as the distance to a Cepheid can be determined by measuring the period, apparent magnitude and extinction. Since they are also among the brightest stars, they are visible out to very large distances of tens of Mpc. Today, Cepheids are vital in the cosmic distance ladder, but it is the closest Cepheids that will be of interest to us during this study.

Using the P-L relation, we can with independent measures obtain the distances of all the Galactic Cepheids observed by Gaia, which can be compared to the measured trigonometric parallaxes. With the method described in this thesis, we simultaneously make a calibration of the Cepheid P-L relation and the parallax zero-point. An offset in the P-L relation with ΔM_{V} gives offsets in the computed parallaxes that is proportional to the true parallaxes ($\Delta\pi \propto \Delta M_{\mathrm{V}}\pi$), while the parallax zero-point is independent on parallax. It is therefore possible to separate the two effects, provided that we can use Cepheids on a wide range of distances.

Since Gaia will observe a much larger number of Galactic Cepheids than is currently known today, we will need to model the total Galactic Cepheid population as observed by Gaia. This modelling constitutes a large part of the project, and has resulted in several Cepheid distribution models, as well as models of the Galactic extinction and how the P-L relation may vary over the Galaxy.

In Section 2, we discuss the Gaia mission and the need for a stable basic angle in more detail. This is followed by Section 3, where we go through different possible ways to verify the Gaia parallax measurements. In Section 4, we discuss the Cepheids and their properties in more detail. The work to model the Galactic Cepheid population is described in Section 5 along with the different extinction models used. In Section 6, we explain the Gaia observational model that is used to simulate the Gaia observations, and in Section 7, we describe the statistical investigation of the observed Cepheids in which we determine the parallax zero-point accuracy. In Sections 8 and 9, we describe and discuss the results of the different experiments that have been performed using all the models developed during the project. Worth consulting is also the table of notations attached in Appendix A.

2 The Gaia Mission

Gaia is an astrometry space mission under development by the European Space Agency (ESA). It was originally proposed by Lindegren et al. (1993), with the goal

Table 1: Expected Gaia sky average rms errors of parallax and proper motion of a G0V star. (Lindegren 2010)

V magnitude	6-13	14	15	16	17	18	19	20
parallax [μas]	8	13	21	34	54	89	154	300
proper motion [μas yr^{-1}]	5	7	11	18	29	47	80	158

of making the largest and most precise map of the Milky Way to date, and was chosen as one of the ESA cornerstone missions in 2000. As a successor to the Hipparcos mission, which resulted in a primary catalogue of 118 000 stars, Gaia will make position measurements of roughly a billion stars down to a limiting magnitude of $G = 20$ (corresponding to $V \sim 20$–25) and measure parallaxes with accuracies as good as $\sim$10 μas. A list of expected accuracies in parallax and proper motion depending on the apparent V magnitude of the star, assuming sky average errors of a G0V star, is given in Table 1. The large increase in accuracy and observed number stars can be achieved by employing larger mirrors which collect more than 30 times the light of Hipparcos, as well as benefitting from the rapid development in detector technology allowing the use of much more efficient CCDs. Gaia will also perform spectral and photometric observations of all the objects, as well as derive space velocities for the stars observed in the Galaxy. With all this information, it will be possible to map the six-dimensional phase-space of the Galaxy with unprecedented accuracy and spectral range. The satellite is scheduled for launch in August 2012, and will have an operational lifetime of 5 years. With data processing by the Gaia Data Processing and Analysis Consortium (DPAC) taken into account, the final Gaia catalogue is expected to be released around 2021.

The Gaia satellite will be placed in an orbit around the L2 Lagrangian point, situated about 0.01 AU from Earth in the direction away from the Sun. This will allow continuous observations during the full 5 year mission length. In this orbit, Gaia will rotate slowly, 4 revolutions per day, with a small precession (63 days period) in the spin axis, pointing at a 45 degree angle from the Sun (see Fig. 1). This combined with the rotation around the Sun will in about half a year give full sky-coverage, and will after the full mission length have resulted in between 50 and 250 field transits over each point in the sky. This is known as the Gaia scanning law, and after 5 years, the predicted number of field transits depending on sky position can in equatorial coordinates be seen in Fig. 2.

The satellite is built around two telescopes providing two viewing directions (see Fig. 1). The angle between the two lines of sight is called the basic angle, and is fixed at 106.5 degrees. Both viewing directions are perpendicular to the satellite's spin axis, meaning that they cover the same part of the sky separated by a few hours. The large angle between the Sun and the spin axis (45 degrees) and the large basic-angle between the viewing directions are essential for Gaia's capability

to do global astrometry and to determine absolute parallaxes. With this setup, it is possible to link widely different areas of the sky. By slowly changing the orientation of the rotation axis so that each point on the sky can be observed many times in different directions, we will after the full mission length have complex relationships between all objects in the sky which can then be solved to finally obtain the absolute parallaxes (Hobbs et al. 2008).

2.1 The Gaia basic angle

In order to be able to correctly solve the complex relationship between all objects observed by Gaia, we need the basic angle between the two fields of view to be stable on short timescales during the whole 5 year mission. If the basic angle varies, it can introduce additional errors to the measured parallaxes. There are several reasons for basic angle variations to occur, and of particular interest is the well-known effect caused by solar heating. The parallax measurements rely on image shifts of the stars which depend on geometry with respect to the Sun that varies with the spin of the satellite. However, as the satellite rotates, different parts of it is exposed to the solar radiation and heated, causing the basic angle to vary periodically on the same scale as the satellite spin period (6 hours), which means that there is perfect correlation with the astrometric parameters. This can lead to apparent image shifts indistinguishable from a global offset of all parallaxes (Lindegren 2004). Even with the current heat shield design, the offset this would introduce during parallax measurements would be on the order of a few μas, enough to be a concern during the development of Gaia. This global offset is called the parallax zero-point, which can be written as

$$c = \langle \pi_{\mathrm{G}} - \pi_{\mathrm{true}} \rangle \tag{2.1}$$

where π_{G} is the parallax of an object measured by Gaia, and $\pi_{\mathrm{true}} = 1/r_{\mathrm{true}}$ is the true parallax of the object. The error in the measured parallaxes are generally symmetric around the true parallax, which causes negligible zero-point shifts when taking the mean of a large number of objects. If c is non-zero, it will therefore correspond to a systematic offset in all measured parallaxes.

The periodic variation from solar heating cannot be fully prevented, but with the use of an on-board Basic Angle Monitoring (BAM) device, a laser interferometer, it will be measured and taken into account in the instrument calibration model. The BAM will measure the possible variations in the basic angle to an accuracy of 0.5 μas every 5 minutes, and the information will then be used during the continuous instrument calibration. With this method, the final parallax measurements should end up without a systematic offset on short timescales. However, an independent parallax validation method would be very valuable to verify the calibration done with the help of the BAM.

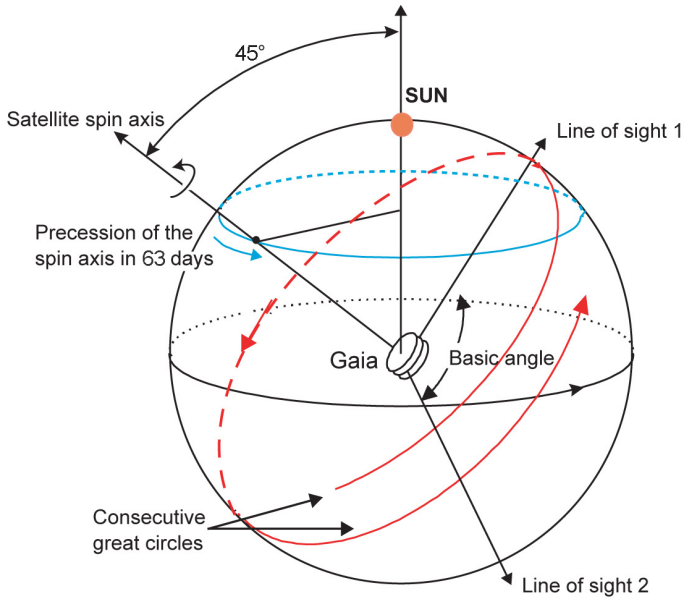


Figure 1: A sketch of the Gaia scanning principle. The spin axis orientation combined with the precession in the spin axis gives full sky coverage in about half a year. Figure credits to Ulrich Bastian.

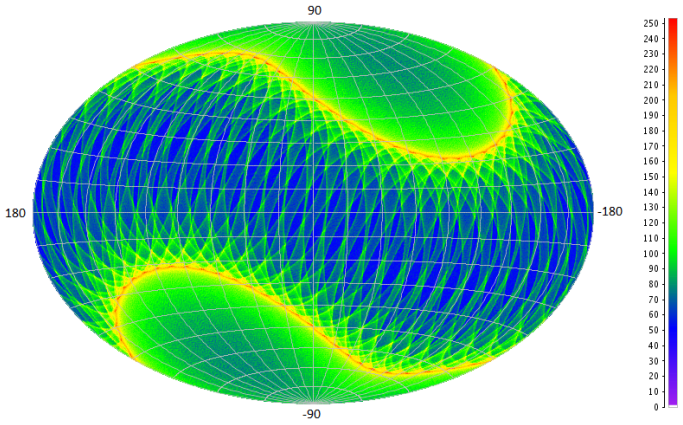


Figure 2: The number of Gaia field transits on the sky after the full 5 year mission length. The map is drawn in equatorial coordinates, with the right ascension, α, on the x-axis and the declination, δ, on the y-axis. The Gaia scanning law is symmetric around the ecliptic, which explains the $\sim$ shaped pattern of increased number of transits. Figure credits to Berry Holl.

3 Methods of parallax validation

In this section, we discuss two different astrophysical methods for the validation of the Gaia parallaxes. Both methods rely on the same fundamental principle; if Gaia can observe a number of objects which have distances that can be determined by other independent means, a comparison can be made between the two distance determinations as in Eq. (2.1). A similar method was used by Arenou et al. (1995) to determine that the zero-point error of the Hipparcos parallaxes should be smaller than 0.1 mas. They used the General Catalogue of Trigonometric Stellar Parallaxes (GCTSP) for their comparison. For Gaia, however, we want to determine the parallax zero-point to a much higher accuracy, as good as < 0.1 μas, and will need to consider alternative parallax references.

Perhaps the most intuitive objects to use for parallax validation are the quasars. Because of their large distances, they will all have parallaxes much below the expected parallax accuracy of Gaia. Cepheid variables are another type of object that might be used. These are widely used as standard candles because of their inherent relation between period and luminosity, providing an observer a simple way to determine their distance. Both of these methods are discussed in more detail below.

3.1 Quasars

Quasars are high redshift galaxies containing an active galactic nucleus (AGN). They are among the most luminous objects in the Universe, with an almost equally strong emission across the whole spectrum, ranging between X-ray and far infrared. The expected number of quasars to be detected by Gaia is large, around 500 000 (Lindegren et al. 2008). Since all quasars are distant enough to have negligible proper motions, they will be used to determine the non-rotating reference frame of Gaia, and might also be used for the verification of Gaia parallaxes.

Quasars do not act as standard candles, but the key property here is instead their large distances. The majority of the Gaia quasars will have redshifts of $z > 0.1$, which gives them parallaxes of $\pi < 3$ nas (nano-arcseconds). With the best Gaia parallax accuracies, roughly three orders of magnitude larger than this, they for all our purposes would have parallaxes $\pi_{\text{true}} = 0$. The quasars observed by Gaia are faint, only 4% are brighter than $V = 18$ and about 75% are fainter than $V = 19$. This would lead to individual parallax measurements with uncertainties not much better than $\sigma_\pi \sim 300$ μas as can be seen in Table 1. However, their large number should bring the accuracy down to slightly better than $300/\sqrt{500000} \approx 0.4$ μas. Using quasars, we would therefore expect to be able to verify the parallax zero-point with an accuracy of roughly 0.4 μas, which is larger than the 0.1 μas aimed for.

The method for identifying quasars is also not without problems. There is currently no certain way of identifying the quasars from the roughly 2000 times larger

sample of foreground stars. This means that the the quasar sample observed by Gaia is very likely to be contaminated by a small fraction of faint foreground stars with photometric properties similar to quasars but with non-zero parallaxes and proper motions. Around 100 000 faint stars in the Galactic halo are expected to have parallaxes and proper motions below 1.5 standard errors of the Gaia measurements. Their mean parallax would be around 100 μas, which means that if 1% of the stars were mistaken as quasars, they would bias the mean parallax of the quasar sample with as much as 0.2 μas. This is not a problem when determining the non-rotating reference frame of Gaia, as the individual scatter of the proper motions of stars are expected to cancel each other out. However, for parallaxes, the errors would consistently work to increase the mean parallax of the sample (Lindegren 2004).

The verification of Gaia parallaxes using quasars therefore mostly relies on being able to create a non-contaminated sample of quasars. How this can be achieved is something that we decided not to focus on during the project, and in the next section, we instead look at the parallax verification method using Cepheids.

3.2 Cepheids

Cepheid variables have been of great interest to the world of astronomy for the last 100 years (Leavitt 1908; Leavitt & Pickering 1912). Henrietta Leavitt observed almost 2000 variables in the Magellanic Clouds, and reported a relation between the period and the luminosity of some of these, now referred to as classical Cepheids. This is now called the Period-Luminosity (P-L) relation, and allows distance determinations by measurements of the Cepheid period. This is of great importance to astronomy, since the Cepheids therefore can be used as standard candles, and have now become an essential part of the cosmic distance ladder. In its simplest form, the P-L relation can be written as

$$\langle M_{\mathrm{V}} \rangle = a \log_{10} P + b \tag{3.1}$$

where $\langle M_{\mathrm{V}} \rangle$ is the magnitude at mean visual luminosity, P is the period in days and $a \approx -3$ and $b \approx -1$ are constants describing the slope and zero-point of the relation. The uncertainties in the a and b constants are very large compared to the uncertainty in the Gaia parallaxes, but with a sufficiently large sample, the uncertainties should become small enough for the P-L relation to prove useful for the verification of parallaxes. The question we need to answer is how large a sample we actually need in order to obtain this accuracy, and how much other factors (such as extinction) can affect the end results.

Currently, we know about 1000 Cepheids in the Large Magellanic Cloud (LMC) and 500 in the Small Magellanic Cloud (SMC). The Magellanic Clouds have been studied in more detail than the Galaxy during the years, largely because of the difficulties associated with extinction in the Galactic disk. The Cepheids in the Clouds have magnitudes around $V \sim 16$, which means that their individual parallaxes can

be determined to a better accuracy than the fainter quasars. The total number of Cepheids in the Magellanic clouds would yield a combined measured parallax uncertainty of $\simeq 1$ μas, which is not accurate enough to be of interest to us. In the Milky Way, we know of about 600 Classical Cepheids, mostly within a few kpc from the Sun, which is also not enough. Today, however, we only see a small fraction of the total number of Cepheids in the Galaxy, and the number of observed Galactic Cepheids is expected to increase by at least an order of magnitude with the advent of Gaia. We also expect many of the Cepheids observed by Gaia to have magnitudes $V < 16$, which will increase the accuracy with which their distances can be determined. This method has a lot of potential, but in order to predict just how accurate it can be, we will need to know a lot more about the Galactic Cepheids.

We conclude that understanding the Cepheid distribution in the Milky Way is a necessity if we want to investigate the use of Cepheids for the parallax verification. The problem is that since our current knowledge of the Galactic Cepheids is so limited, it is no easy matter to estimate the total number of Cepheids in the Galaxy or their distribution. We will therefore need to make a number of assumptions before we can estimate the total number of Cepheids that Gaia will observe, and their distribution of parallax accuracies. A large portion of this project has therefore gone into creating a model for the Galactic distribution of Cepheids. In the end, we want to be able to answer the question whether Gaia will be able to observe enough Cepheids for the verification of Gaia parallaxes. Cepheids and their Galactic distribution are discussed in detail in Sections 4 and 5.

4 Cepheid properties

In this section, we discuss the general properties of Cepheids; what causes them to pulsate, their mass and period ranges, the calibration of the P-L relation and what we currently know about the Galactic Cepheid distribution.

4.1 General properties

When a star of sufficient mass leaves the main sequence, it follows a rather horizontal track on the Hertzsprung-Russel diagram as it heads towards becoming a Red Giant Branch (RGB) star. During this time, it passes through what is commonly called the instability strip. This is a region where many of the variable stars can be found, such as RR Lyrae, Classical Cepheid and W Virginis variables. Stars of these types pulsate due to the different opacities of He II and He III (singly and double ionized helium). In a normal star of spectral type A, F and G, we find neutral He in the stellar photosphere, He II further in as the temperature increases, and He III even deeper inside the star. Stars in the instability strip have their He layers at just the right distances from the center, so that when the star contracts, the density and temperature in the He II layer increases enough to start transforming into He III. This results in a higher opacity, leading to higher absorption rate of the energy flux.

This causes the layer to expand, leading to a decrease in density and temperature which causes He III to again recombine into He II. With the once again lower opacity, the layer starts to contract, starting a new cycle of pulsation. At both end points in the pulsation, the layer recieves a kick inwards or outwards which causes it to overshoot and miss the point of equilibrium. The end result is a regular pulsation of significant amplitude in both radius and luminosity that keeps on going much like a harmonic oscillator. This variable state disappears once the star has reached the RGB branch after a few Myrs.

Early on, the definition of a Cepheid unintentionally included many different types of variable stars in the instability strip. It was not discovered until later that the Cepheids seemed to follow several different period-luminosity relations, resulting in the variable star subclasses that we have today. The population we are interested in during this project is what is called Population I, or Classical Cepheids (in the rest of the paper simply referred to as Cepheids), which have the most clearly defined P-L relation and are the ones most widely used as standard candles. Classical Cepheids are young, metal rich, yellow giants with masses larger than 5 $M_\odot$. Their pulsation periods range from 1 to 100 days and the absolute magnitudes between -1 and -7, with the shortest period stars being the faintest and the longest period stars the brightest. The reason behind this is quite intuitive, as the more massive and bright stars simply have more extended and less dense envelopes, meaning that the timescale of the variability is longer.

With masses above 5 $M_\odot$, the stars capable of becoming Cepheids will spend less than 100 Myrs on the main-sequence before they enter the instability strip. They are believed to have a lifetime of $\tau < 1 - 3$ Myrs before they lose their variability and enter the RGB branch. This means that we can really only expect to find Cepheids in regions with ongoing star formation, and not in old globular clusters or in the Galactic halo. Surveys of the Galactic gas distribution show that most of the hydrogen can be found close to the Galactic plane, and this is where the currently discovered Cepheids have been found so far. This is discussed in more detail in Section 4.3.

4.2 The Cepheid P-L relation

In the just over 100 years since the discovery of the Cepheid P-L relation, it has been calibrated numerous times with greater and greater accuracy. The majority of these calibrations has been done with the Cepheids in the SMC and LMC, as these suffer much less extinction and other problems associated with the Galactic disk. In Table 2a, we give a compilation of LMC P-L relations obtained by various groups during the last 10 years. As can be seen, the constants a and b are quite well determined and the values agree to a good accuracy between the groups. There is a bit of a discrepancy in b, as finding it relies on the ability to determine the distance modulus to the LMC.

Determining the P-L relation using Galactic Cepheids involves more difficulties. For example, the extinction needs to be determined separately for each object as opposed to in the LMC, and we also know of fewer Cepheids in the Galaxy than in the LMC. In Table 2b, we see the P-L relations obtained from different groups during the last 20 years. Here, we see a larger discrepancy because of the problems described. The two relations obtained by Tammann et al. (2003) and Sandage et al. (2004) differ from the others in both a and b. They argue that the reason for this is that their calibration of the zero-point b is only done using Galactic Cepheids, and the other groups assume the same zero-point as in the LMC relation. Something that we can also note is the discrepancy between the P-L relation of Cepheids in the LMC and in the Galaxy. This is also seen for the P-L relation of the SMC, and is believed to be real. The likely reason behind this is the difference in metallicity between the different populations of Cepheids. Why metallicity plays a role is easily understood, as higher metallicity causes the opacity to increase due to the larger amount of possible electron transitions for heavier elements. The exact effect of metallicity on the P-L relation is not known, however, and is something that is currently quite debated. With only a few hundred Cepheids with known metallicities, all of them found within a few kpc from the Sun, it is today difficult to draw any accurate conclusions. Groenewegen (2008) reports a highly uncertain metallicity dependence that is within their error limits, and only a larger Cepheid sample or more accurate metallicity abundances can decrease the uncertainty.

The possible effect of metallicity is something that we need to consider. We intend to fit a single P-L relation to all of the Cepheids observed by Gaia, and that either requires the metallicity effect to be small or that all Cepheids in the Galaxy have the same metallicity. Comparing the intense star formation going on in the Galactic Centre to the much lower star formation in the outskirts of the Galaxy, for example, should logically show quite a wide spread in metallicities. The result should therefore be a relatively large scatter when fitting our single P-L relation to the Galactic population compared to the scatter of the LMC P-L relation. A second alternative is that we can fit all Cepheids to a P-L relation with an added metallicity-dependent term. Our different assumptions regarding this are discussed in more detail later on, and the effect that these assumptions have on the end result is discussed in Section 9.

4.3 Cepheid catalogue data

In order to estimate the number of Galactic Cepheids that Gaia will observe, we need to simulate the observation of the whole population of Cepheids in the Milky Way. Our simulations need to include the distribution of periods, and we need to determine what P-L relation to use in order to create the intrinsic M_V distribution. We also need to know the physical distribution of Cepheids; both their positions projected on the Galactic plane and their vertical distribution around it.

Table 2: Comparison of P-L relations of Cepheids in LMC and the Galaxy written as $M_V = a \log_{10} P + b$. In three of the cases in (a), the zero-point includes the distance modulus to LMC, μ_0, in order to avoid introducing errors. Today, $\mu_0 = 18.54$ is the generally accepted value. For comparison, M_V with $\log P = 0.8$ days has also been calculated for all relations. Note the larger discrepancy between relations in (b) than in (a).

(a) P-L relations of Cepheids in the LMC

Reference	a	b	$M_V(0.8)$
Gieren et al. (1998)	-2.769	-1.294	-3.51
Udalski et al. (1999)	-2.779	$17.066 - \mu_0$	-3.62
Sandage et al. (2004)	-2.702 ± 0.028	-1.491 ± 0.022	-3.65
Fouqué et al. (2007)	-2.734	$17.052 - \mu_0$	-3.60
Ngeow et al. (2009)	-2.769	$17.115 - \mu_0$	-3.56

(b) P-L relations of Cepheids in the Galaxy

Reference	a	b	$M_V(0.8)$
Madore & Freedman (1991)	-2.76	-1.4	-3.61
Feast & Catchpole (1997)	-2.81	-1.43	-3.68
Gieren et al. (1998)	-2.77	-1.29	-3.51
Tammann et al. (2003)	-3.14	-0.83	-3.34
Sandage et al. (2004)	-3.087 ± 0.085	-0.914 ± 0.098	-3.38
Fouqué et al. (2007)	-2.678	-1.275	-3.42
Groenewegen (2008)	-2.60	-1.30	-3.38

We have taken data from two different Galactic Cepheid databases. The first of the two was created by Fernie et al. (1995), and contains 509 Cepheids in total, but only 416 Cepheids with all the important properties known. The second was compiled by Berdnikov et al. (2000), containing 455 Cepheids that could be of use to us. The overlap between the two databases is large, and there would be little gain in trying to merge them. From these two databases, we try to obtain as much information as possible in order to be able to get a good fit of our model to reality. As will be shown later in this section, both these databases are complete out to a distance of about 1 kpc from the Sun, which has implications for how many of the objects that can be used when determining the distribution of various properties. We decided to primarily use the Berdnikov database, as it is the largest and most recently updated of the two.

Distribution in the Galactic plane

In Fig. 3, we show the Berdnikov Cepheids projected on the Galactic plane with the Galactic Center at the origin and the Sun positioned a distance of 8 kpc from it. In this figure, we can identify a possible ridge structure that might indicate a tendency

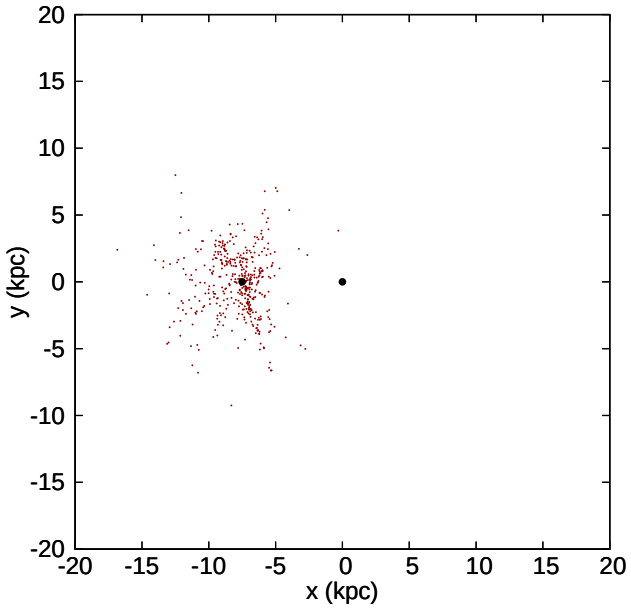


Figure 3: The distribution of Berdnikov Cepheids with the Galaxy seen face on. The Galactic Center is placed at the origin, and the Sun at a distance of 8 kpc from the GC.

for the Galactic Cepheids to be found in the spiral arms. We can also note that in general, we see Cepheids at larger distances away from the Galactic Center than towards it, undoubtedly because of the larger extinction in the direction towards the center.

Distribution vertical to the Galactic plane

Figure 4 shows the observed vertical distribution of Cepheids, with the height around the Galactic plane calculated from $z = r \sin\ b + z_\odot$, where r is the distance from the Sun, b is the galactic latitude and $z_\odot$ is the vertical position of the Sun. At $b = 0$, we therefore look parallel to and slightly above the center of the Galactic disk, resulting in the effect that the majority of the stars appear to be located slightly below it. In the plot, we assume $z_\odot = 0$, and can note a clear offset in the Cepheid distribution. This offset is thought to depend solely on the vertical position of the Sun, and we can therefore use the observed height distribution to determine $z_\odot \approx 25$ pc. Other, more involved investigations of the Sun's position, give slightly lower values as tabulated by Reed (2006). In order to be consistent with these investigations, we adopt $z_\odot = 20$ pc in our Galactic model. In the plot, we can also note that the Cepheids exist solely at low heights above the Galactic plane within a few hundred pc. It is in the Galactic disk that they are formed, and because of their short lifetimes, they do not have time for any major scatterings away from their original place of formation.

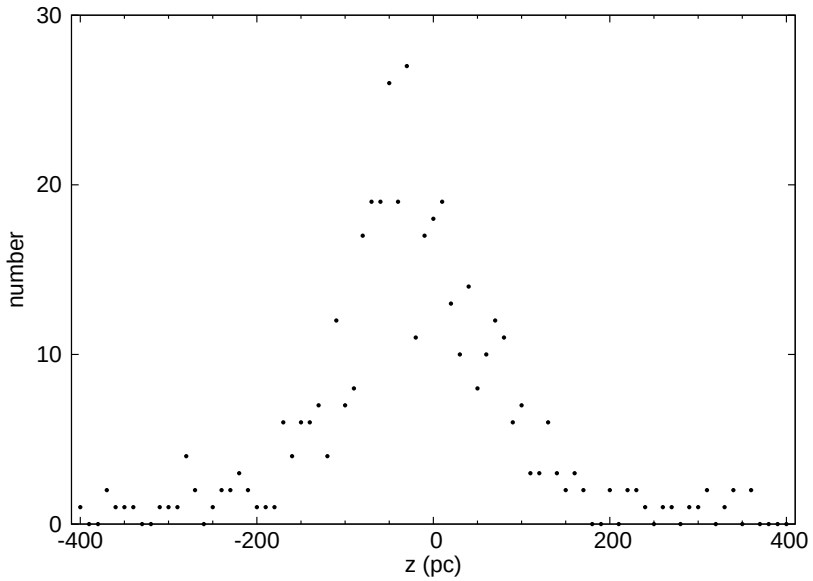


Figure 4: The distribution of Berdnikov Cepheids binned with Δz = 10 pc calculated from $z = r \sin\ b + z_\odot$ with an assumed $z_\odot = 0$. The peak offset indicates that the Sun's real position is at $z \approx 25$ pc.

Period distribution

In Fig. 5, we show the period distribution of the full Berdnikov sample of Cepheids. The distribution shows periods between roughly 1 and 100 days, with the largest peak at $P \sim 5$ days. There also seems to be a second peak at $P \sim 12$ days, possibly hinting at two different populations of Cepheids. However, as can be seen in the P-L relation (see Eq. (3.1)), we expect the longer period Cepheids to be brighter than those with shorter periods. Since this is not a complete sample, but to some extent limited by apparent magnitude, this implies that the sample is biased towards longer periods, as these can be seen at larger distances. In a non-biased sample, we can then expect the long-period Cepheids to be fewer, likely weakening the second peak. In Section 5.2.4, we discuss a method to find the non-biased period distribution.

Completeness of the sample

The number of Galactic Cepheids that have been discovered so far is very small compared to the expected total number, as can be seen in the Berdnikov catalogue given in Fig. 3. This catalogue is a compilation of a number of Cepheid studies, and because of this does not have a strict brightness limit. In order to determine out to which distance the catalogue is complete, we plot the column density of Berdnikov Cepheids, Σ, versus the heliocentric distance projected on the Galactic plane, $r \cos b$. To determine the column density within $r \cos b$, we project the position of

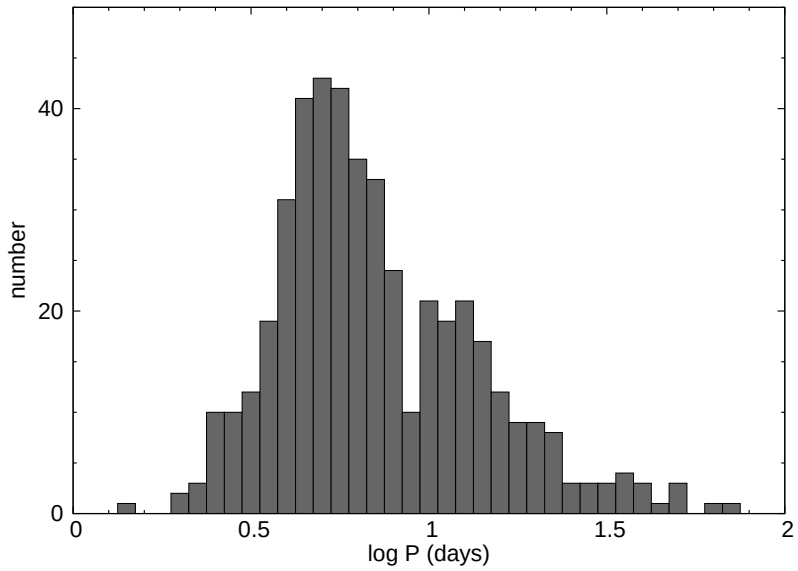


Figure 5: Histogram of the Berdnikov Cepheid periods.

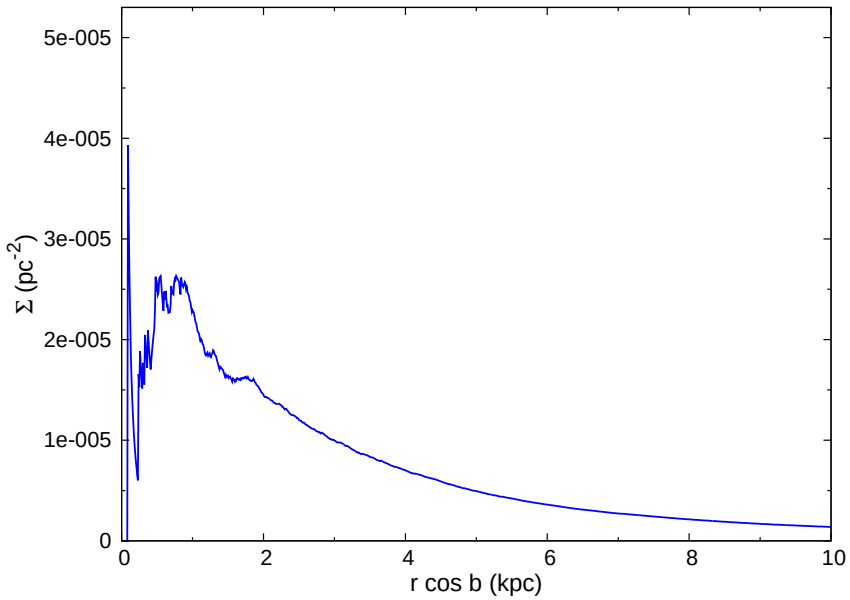


Figure 6: The Berdnikov Cepheid column density plotted versus projected distance $r \cos b$. Note the drop in column density at distances larger than 1 kpc.

each Cepheid on the Galactic plane and calculate the surface density. The result is given in Fig. 6. For small changes in $r\cos b$, we would expect the column density of the currently known Cepheids to be reasonably constant because of the relative flatness of the radial distribution function. For very small distances, the column density varies greatly because of the limited number of data points, but we can see a clear plateau between 0.5 and 1 kpc. After 1 kpc, the curve falls off, which can be attributed to the fact that we do not have a complete sample and do not see all the Cepheids at these larger distances.

Within the completeness limit of 1 kpc, we have 71 Cepheids in the Berdnikov database, corresponding to a column density of $\sim$23 Cepheids kpc^{-2}. A rough estimate of the number of Galactic Cepheids can then be obtained by assuming a radius of the Galactic disk of $R = 15$ kpc and a constant Cepheid density over the disk. This gives a total number of approximately 16 000 Galactic Cepheids. With such an incomplete sample, 71 out of $\sim$20 000, many properties become very difficult to study, and it is clear that we need information from other sources than the Cepheid databases. What these other sources are, and how the Galactic model was implemented, is discussed in detail in the following section.

5 Galaxy modelling

With our Galaxy model, we want to be able to simulate the total population of Galactic Cepheids, with a number of properties for each object. In order to study the P-L relation, each Cepheid needs a period, P, and an absolute visual magnitude, M_V. We also need to determine the intrinsic $V - I$ colour, $(V - I)_0$, in order to determine the magnitude in the Gaia G band. Finally, we also need to determine the position in the Galaxy of each generated Cepheid. The aim is then to determine what each object will look like from the Sun, taking interstellar extinction into account. This will determine whether the Cepheid can be observed by Gaia or not. If it is observable, the full five years of Gaia observations are simulated in order to determine the accuracy of the measured parallaxes.

This section is divided into three parts. In Section 5.1, we discuss the necessary terminology for working with these types of Galactic simulations, and also discuss the general structure of the Galaxy as well as the general assumptions made during the work. In Section 5.2, we discuss the modelling of the Cepheid properties, and in Section 5.3, we discuss the modelling of the three-dimensional Galactic extinction.

5.1 Framework for Galaxy modelling

The Galactic structure is very complicated and not something that is easily modelled. Spiral galaxies like the Milky Way can roughly be divided into four different components: a thin and a thick disk, a halo and a bulge. Our Sun is located in the thin disk, a dense sheet of stars a few hundred pc thick distributed around the

Galactic plane where most of the interstellar gas can be found. The thick disk has a thickness of around 1 kpc, and is much less dense. In the disk, we know that there are a number of spiral arms which have higher density of gas and young stars than the neighbourhood. In the center of the Galaxy, the bulge makes up a spheroidal structure with a radius of roughly 1 kpc. Very little gas and thus little star formation can be found there, and it therefore consists of an older population of stars than the disk. The Galactic halo surrounds the whole galaxy, and contains a low density of stars without any ongoing star formation. Since Cepheids are all young stars (< 100 Myrs old) and the Cepheid stage is relatively shortlived ($\tau < 3$ Myrs), we only expect them to be found near regions with active star formation. Because of this, our modelling will mostly concern the thin disk in the region a few hundred pc from the Galactic plane.

The position of our own Sun in the Galaxy is of great importance for what the generated Cepheids will look like to Gaia and how many that are observable. We assume the commonly used distance of $R_\odot = 8.0$ kpc to the Galactic Center (Reid 1993; Eisenhauer et al. 2003). For the Solar height above the Galactic plane, we choose $z_\odot = 20$ pc, which is taken from the results of several different studies tabulated by Reed (2006).

During the Galactic modelling, we work in two different reference frames. When determining the Cepheid distribution, we work in Galactic x, y and z coordinates, where the Galactic center is at the origin. x and y are here the coordinates projected on the Galactic plane, and z is the vertical distance to the plane. We then assume the position of the Sun to be $(-8, 0, 0.020)$ kpc.

When simulating the Gaia observations, we move over to a heliocentric reference frame. We use the standard galactic coordinate system (l, b, r). The galactic longitude ranges between $0° < b < 360°$ and is measured in the Galactic plane using the direction towards the Galactic center as $l = 0$ at the direction of rotation at $l = 90°$. The North Galactic Pole is found at $b = 90°$, and the South Galactic Pole at $b = -90°$. The Sun's position slightly above the Galactic plane means that the Galactic Center and the majority of the Cepheids are positioned at slightly negative latitudes.

5.2 Cepheid distribution

With the limited knowledge we have of the Galactic Cepheids, it is necessary to make some initial assumptions regarding their distribution. We start by assuming that all Cepheids are distributed axisymmetrically around the Galactic Center, meaning that the number density of Cepheids is a function of only the galactocentric radius and the vertical distance to the Galactic plane, disregarding structures such as spiral arms and individual molecular clouds. This is a common assumption when it comes to Galaxy modelling, but also a somewhat rough one, as we know for a fact that

the Galaxy does not lack features on a smaller scale. In fact, it is quite likely that the Cepheid density increases around the spiral arms, as Cepheids are young stars, which requires continuous star formation, as can be found in the spiral arms. This is also slightly hinted at in Fig. 3. Currently, however, we know relatively little about the properties of the spiral arms. Working with a Cepheid distribution model which included the spiral arms would be quite a lot more complex than the axisymmetric model, and the model would be highly uncertain anyway. Moreover, it is not expected that such a detailed model would radically change for example the total number of Cepheids observed by Gaia. Because of this, we neglect the effect of the higher stellar density in the spiral arms. Similarly, we assume that the Cepheids are symmetrically distributed around the Galactic plane ($z = 0$). The first results from the statistical investigation will tell us just how sensitive we are to changes in the distribution parameters.

We therefore need three different properties for the the physical Cepheid distribution: the radial distribution around the Galactic Center, the vertical distribution around the Galactic disk, and the total number of Cepheids in the Galaxy. We also need a model for the period distribution and a P-L relation to generate the absolute magnitudes, as well as the colour distribution (necessary for the Gaia observational accuracy). Details on these different properties are discussed below.

5.2.1 Radial distribution

Determining the radial distribution of Cepheids in the Galaxy is not trivial. As seen in Fig. 3 and discussed in Section 4.3, the currently known Galactic Cepheids are found at galactocentric radii between about 5 and 10 kpc, but the sample is only complete near the solar neighbourhood, within the range 7 to 9 kpc. This severely limits our ability to study and understand the Cepheid density both close to the Galactic Centre and further out towards the edge of the Galaxy. It is therefore difficult to predict the full radial distribution of Cepheids using only the current Cepheid sample. We need to consider other, indirect ways.

Cepheids are young, massive stars with short lifetimes on the order of tens of Myrs. This means that we will only see Cepheids in regions with ongoing star formation, since all the Cepheids in other regions would have evolved and continued on towards the Red Giant Branch. Star formation can only be initiated in regions containing gas, which implies that they follow the same distribution as the gas in the Galaxy. Fortunately, because of the hydrogen 21 cm line and the low extinction at radio wavelengths, we know a lot more about the distribution of neutral hydrogen than we do about the Cepheids, and there have been several surveys to map the Galactic gas distribution over the years. In general, we expect the density to be highest in the centre and decrease further out. In our first Cepheid distribution model, we use the simple relation for the distribution of gas in the Galaxy as described by Binney & Merrifield (1998) based on infrared observations by DIRBE and COBE by

Freudenreich (1998). They describe the density of the Galactic disk with respect to galactocentric distance as a simple exponential function, written as

$$\rho_{\rm gas}(R) \propto \exp\Big(-\frac{R}{R_0}\Big) \tag{5.1}$$

where $R = \sqrt{x^2 + y^2}$ is the galactocentric distance projected on the Galactic plane and $R_0 = 2.5$ kpc is the radial scale-length. This will from now on be referred to as the disk radial distribution.

With Cepheids being young and massive stars, we would expect them to follow the same distribution as other bright and young objects such as OB stars which form in OB associations in molecular clouds. McKee & Williams (1997) and Williams & McKee (1997) found that their distribution best can be described by an exponential function as in our disk radial distribution, but with a scale-length of $R_0 = 3.5$ kpc. We call this the OB radial distribution.

Using an exponential function to describe the distribution of gas in our Galaxy is an approximation that does not work very well everywhere. We know for example that the bulge contains a much lower density of gas and therefore also a much lower density of young stars. Because of this, there have been more thorough attempts to describe the distribution of gas in our Galaxy. In their construction of a three-dimensional extinction model in the Galaxy, Amôres & Lépine (2005) used data on the Galactic gas to derive a slightly more sophisticated model of the gas distribution. The extinction model itself is discussed and used in Section 5.3, but we also use their work to determine a third model for the radial distribution of Cepheids. The general expression they obtained for both gas phases (neutral and molecular hydrogen, HI and H_2) is

$$\rho_{\rm HI,H_2} = \delta_1 \exp\left(-\frac{R}{\alpha_1} - \Big(\frac{\beta_1}{R}\Big)^2\right) \tag{5.2}$$

where the parameters are $\alpha_1 = 7.0$ and 1.2 kpc, $\beta_1 = 1.9$ and 3.5 kpc and $\delta_1 = 0.7$ and 58 cm^{-3} for HI and H_2 respectively. They also included an additional function to describe the very dense H_2 region in the Galactic center,

$$\rho_{\rm H_2} = \delta_2 \exp\left(-\Big(\frac{R}{\alpha_2}\Big)^2\right) \tag{5.3}$$

where $\alpha_2 = 0.1$ kpc and $\delta_2 = 240$ cm^{-3}. In our third, most extreme Cepheid distribution model, we assume that the Cepheid density is proportional only to the H_2 gas. Molecular hydrogen is thought to exist predominantly in molecular clouds, which are the most important regions of star formation. In this model, since the density function in the Galactic centre is so uncertain, we disregard the Cepheids at galactocentric radii less than 1 kpc, giving us the worst-case scenario for this type of distribution. The result is a doughnut shaped ring with its maximum at a

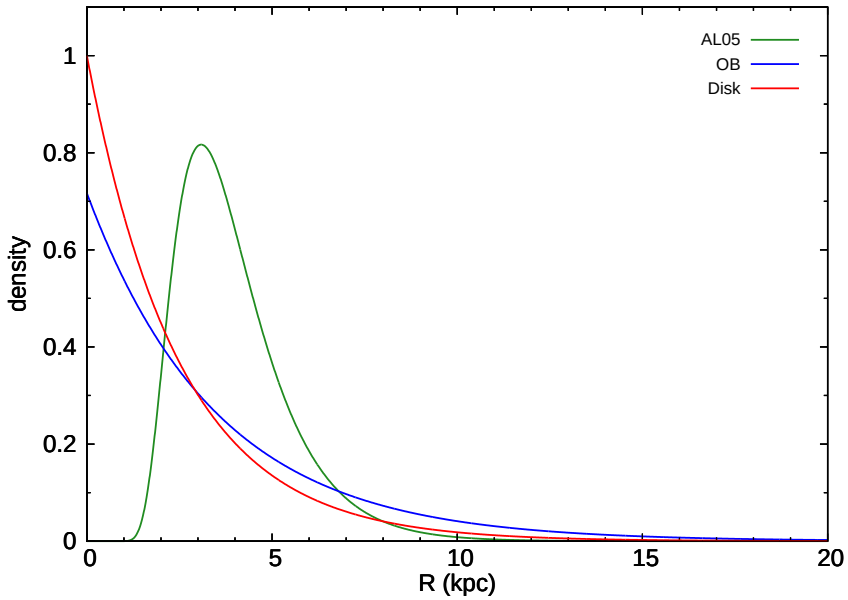


Figure 7: The three radial distribution models used in this work.

galactocentric radius of 3 kpc. We refer to this as the AL05 radial distribution.

The three radial distributions are all given in Fig. 7 for comparison. We can now generate Cepheids that follow any of these distributions. This is done by computing the Cumulative Density Function (CDF) of the distribution. The CDF is very useful when we want to generate discrete data points with a certain distribution. CDF(R) is the fraction of Cepheids contained within the distance R compared to the total Cepheid population, so that $0 \leq \text{CDF}(R) \leq 1$. This means that for every R we want to generate, we draw a random value, U, between 0 and 1, and compute $R = \text{CDF}^{-1}(U)$. We will however first need to normalize the CDF with respect to the area within R. The CDF of our normalized radial distribution functions is

$$\text{CDF}(R) = \frac{\int_0^R \rho_{\text{ceph}}(R) \cdot 2\pi R}{\int_0^\infty \rho_{\text{ceph}}(R) \cdot 2\pi R} \tag{5.4}$$

The above equation then needs to be solved for $R = \text{CDF}^{-1}(U)$. This is however quite difficult to do analytically. In our program, therefore, we rely on pregenerated tabulated values for R and CDF calculated numerically. Because of the shape of the exponential function, we do not need to concern ourselves with finding the outer limit of the Galaxy. As an example, CDF(15 kpc) ≈ 0.98, and CDF(20 kpc) ≈ 0.997. If we determine the CDF out to a maximum distance of 25 kpc, we will miss less than one per mil of the total Cepheid population.

From the radial distance, the galactic coordinates, x and y, are then generated

using θ randomly chosen in the interval $0 \leq \theta < 2\pi$:

$$x = R \sin\theta \tag{5.5}$$

$$y = R \cos\theta \tag{5.6}$$

The generation of the third coordinate, z, is discussed in the next section.

5.2.2 Vertical distribution

The Cepheid distribution normal to the Galactic plane is much more straightforward to study than the radial distribution. In this case, we can reliably use the currently known Cepheid data, since the disk containing the Cepheids is thin enough to allow us to observe all of the Cepheids both above and below us. This is not only true for Cepheids, but for all stars, and the vertical distribution has therefore been studied in more detail than the radial. van der Kruit (1988) gives this general equation to describe the distribution of stars around the Galactic disk

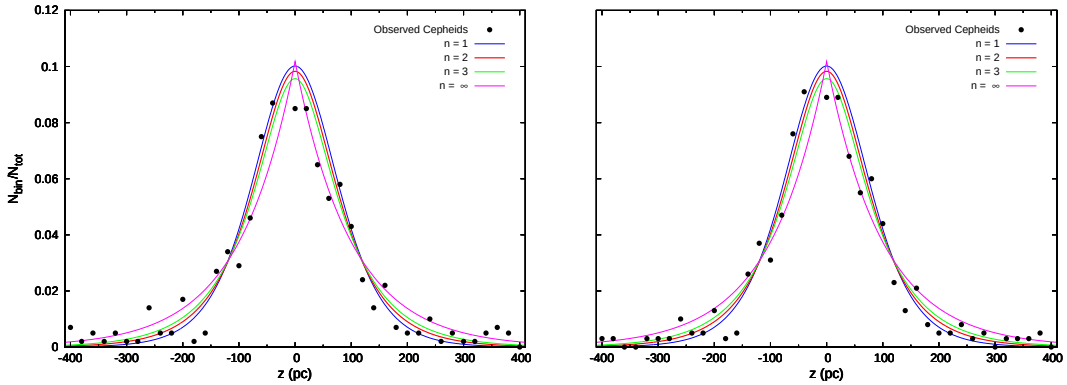
$$\rho(z) = \rho(0)\ \mathrm{sech}^{2/n}\Big(\frac{nz}{2z_0}\Big) \tag{5.7}$$

where $\rho(0)$ is the density in the plane, z_0 is the scale height and

$$\text{sech z} = \frac{1}{\cosh z} = \frac{2}{\mathrm{e}^z + \mathrm{e}^{-z}} \tag{5.8}$$

is the hyperbolic secant. This expression has the property of being smooth around $z = 0$. The n in the expression results in different shapes for the curve: $n = 1$ describes an isothermal distribution, $n \to \infty$ an exponential distribution and the other n values give intermediate states between the two. In Fig. 8a, we compare the commonly used $n = 1$, $n = 2$, $n = 3$ and $n = \infty$ to the observed distribution of Cepheids. The normalized curves have scale-heights between 50 and 100 pc, empirically chosen to fit as well as possible to the vertical distribution described by the currently known Cepheids binned up and represented by the black dots. We can note that almost all Cepheids are found relatively close to the Galactic plane, within a few hundred pc, which is to be expected for young objects that have formed in the disk and have not had time to scatter further out during their short lifetimes.

In the figures, we can see a clear advantage of using a non-exponential distribution, as the exponential gives rise to a very spiky peak at $z = 0$ which we neither observe nor expect physically. In Fig. 8a, however, the exponential actually seems to represent the wings best. This is believed to be due to selection effects. Objects at large z can be expected to be visible at larger distances than those at low z, as the extinction decreases with z. This means that we observe an unproportionally large fraction of large-z objects at distances larger than the completeness limit of 1 kpc. Unfortunately, the complete sample only contains 71 Cepheids, which leads to too large a statistical scatter. To illustrate our point, Fig. 8b only includes the observed



(a) All Cepheids (b) Cepheids within 5 kpc

Figure 8: Histogram of Berdnikov Cepheid distribution around the Galactic plane, with the lines corresponding to different model fits. Different scale-heights have been used for the different n values; $z_{0,1} = 50$ pc, $z_{0,2} = 75$ pc, $z_{0,3} = 85$ pc, $z_{0,\infty} = 100$ pc. Panel (a) contains 455 Berdnikov Cepheids, and (b) contains 387 Cepheids. Note the smaller wings for the limited sample in (b).

Cepheids within 5 kpc (a total of 387 Cepheids). In this sample, we still have a bias towards Cepheids at large z, but we note that the wings have decreased substantially from the sample in Fig. 8a.

In the limited sample, we note that the exponential no longer fits very well. Instead, the low n values are favoured, and $n = 1, 2$ and 3 do not seem to differ much at all. Holmberg et al. (1998) found $n = 3$ to be good fit for the stellar population in the Hipparcos catalogue. For simplicity, however, we use $n = 2$, as this results in nicer expressions for the simulations.

Something that cannot be easily studied using any current Cepheid database is the radial dependence of the scale-height. As shown in the previous section, at increasing galactocentric radius, the density of stars decreases. It is also believed that the total matter density in the disk decreases outwards. This leads to a lower gravitational pull in the vertical direction the further out from the Galactic centre we get. If a star receives a kick away from the disk, it therefore ends up at different heights dependent on its galactocentric distance. This means that kicks are more effective further away from the Galactic center, leading to the somewhat counter-intuitive result that the scale-height increases with larger R. The same effect can be observed in the distribution of gas, and was described by Guibert et al. (1978). More recently, Amôres & Lépine (2005), found a scale-height of $z_0 \propto \exp(0.08R)$ pc to work for the distribution of HI and H_2. Comparing this expression with the scale-height in the solar neighbourhood which we know from Fig. 8 to be about 75 pc, we find

$$z_0(R) = 40 \exp(0.08R) \tag{5.9}$$

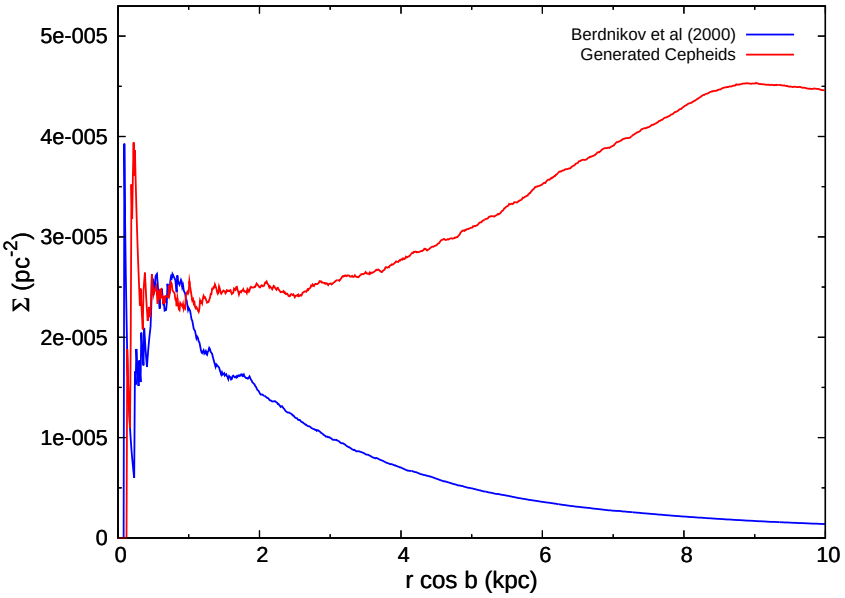


Figure 9: Column density of Cepheids within a distance $r\cos b$. The Berdnikov Cepheids are given by the blue curve, and the red curve represents simulated data using the OB radial distribution model.

where z_0 is given in pc and R in kpc. Finally, we need to find the inverse CDF for Eq. (5.7) in order to generate our Cepheid population. The expression for an hyperbolic secant ($n = 2$) can be solved analytically, yielding

$$z = z_0 \log\Big(\tan(\pi/2 \cdot U)\Big) \tag{5.10}$$

where U has a uniform distribution between 0 and 1. We now have all that we need to give each Cepheid a three-dimensional position in the Galaxy. The next step is to determine the total number of Cepheids in the Milky Way.

5.2.3 Total number of Galactic Cepheids

The total number of Cepheids in the Galaxy can be estimated using the observed Cepheids in the Solar neighbourhood and the radial distribution functions described in Section 5.2.1. To do this, we use the column density plot given in Fig. 6. To determine the total number of Galactic Cepheids for any of the radial distribution functions, we keep on generating Cepheids until the column density of the generated sample fits with the column density from the observed Cepheids within the completeness limit. The column density for the simulated data at the end of this process is shown by the red curve in Fig. 9. The discrepancy between the blue and red curve at larger distances illustrates how incomplete the current Cepheid databases are. The simulations results in $N_{\text{ceph}} = 20000$ Cepheids if the OB radial distribution is assumed. The process is then repeated for the two other radial distributions, yielding

$N_{\rm ceph} = 22000$ for the disk distribution and as many as $N_{\rm ceph} = 35000$ for the AL05 distribution. These numbers are larger than the $N_{\rm ceph} = 15000$ found by the rough approximation by Majaess et al. (2009), who do not take the radial gradient into account. We believe our calculations to be more closer to reality, however, it would also be useful to verify our numbers by some other, independent means, if possible.

From our knowledge of Cepheid masses and lifetimes, we can come to a rough estimate by the use of the Initial Mass Function (IMF). The IMF is a commonly used way to describe the distribution of masses in a stellar population, and is an entirely empirical function determined from large quantities of stars. If N is the number of stars and M is the mass, the IMF is usually assumed to follow a power law,

$$\frac{dN}{dM} \propto M^{-\alpha} \tag{5.11}$$

where α is a constant. One of the most popular IMF functions is the Salpeter IMF, in which $\alpha = 2.35$ (Salpeter 1955). Assuming that all stars in a certain mass interval will at one point become Cepheids, the fraction of Cepheids to the total number of stars is

$$f_{\rm ceph} = \frac{N_{\rm ceph}}{N_{\rm tot}} = \frac{\int_{C_{\rm min}}^{C_{\rm max}} M^{-\alpha} dM}{\int_{T{\rm min}}^{T{\rm max}} M^{-\alpha} dM} = \frac{\left[M^{1-\alpha}\right]_{\rm C_{min}}^{C_{\rm max}}}{\left[M^{1-\alpha}\right]_{\rm T_{min}}^{T_{\rm max}}} \tag{5.12}$$

where $T_{\rm min} = 0.08$ $\rm M_\odot$ and $T_{\rm max} = 120$ $\rm M_\odot$ is the minimum and maximum mass of all stars respectively, and $C_{\rm min}$ and $C_{\rm max}$ are the minimum and maximum mass among the Cepheids. The total number of Cepheids in the Galaxy at one point in time can then be estimated from

$$N_{\rm ceph} = R_{\rm sfr} f_{\rm ceph} \tau_{\rm ceph} \tag{5.13}$$

where $R_{\rm sfr} \sim 3$ $\rm M_\odot$ $\rm yr^{-1}$ is the formation rate, i.e. the average mass of the gas in the Milky Way which every year is converted into stars, and $\tau_{\rm ceph} < 1 - 3$ Myrs is the average Cepheid lifetime.

If we assume $\tau_{\rm ceph} = 1$ Myr and that Cepheid masses range between 5 and 120 $\rm M_\odot$, we get $N_{\rm ceph} \sim 11000$. This number is however very uncertain. Neither the Cepheid lifetime, the Cepheid mass limits nor the star formation rate in the Milky Way are known to very good accuracies, and the IMF α value is also disputed. We easily have uncertainties of a factor two, which makes the IMF method much less trustworthy than catalogue extrapolation. Taking this into account, the value we obtain from the IMF does fit reasonably well with our previous results, and we conclude that $N_{\rm ceph} \simeq 20000$–$30000$ might be a good fit to reality. During our modelling of the Galaxy, we decide to use the three different numbers (20000, 22000 and 35000) obtained from the catalogue extrapolation depending on the assumed radial distribution.

5.2.4 Period distribution

The distribution of Cepheid periods is important to us for two reasons. First of all, because of the inherent relation between period and luminosity, we need a range of periods to fit a P-L relation to the data. The accuracy with which this can be done is affected by the number of short and long period Cepheids, as we need both to determine the P-L relation. We also need it to create the distribution of absolute magnitudes in our Galaxy. From the P-L relation (given in Eq. (3.1)), we can see that longer periods correspond to brighter magnitudes. A period distribution favouring long periods will therefore lead to more Cepheids being visible to Gaia.

Because we are now studying properties related to the magnitude of the Cepheids, we have to be careful to avoid biases when making our period distribution for the simulations. Because long-period Cepheids are brighter, we can expect the currently known Cepheid periods seen in Fig. 5 to be biased towards longer periods. To avoid this, we therefore need to consider only the period sample within the completeness limit, containing only 71 stars. If we look at Fig. 5, we can distinguish two peaks which might be describable by two overlapping Gaussian distributions. When we go down to the smaller sample, the longer period peak decreases slightly, but can still be seen. In order to be able to produce both the shape of the larger short period peak and the long period tail, we need at least two Gaussians. If we introduce $u = \log P$, a log-normal distribution of P is described by the probability density function (pdf)

$$f(u) = \frac{1}{\sigma_{\mathrm{u}}\sqrt{2\pi}} \exp\left(-\frac{(u-\bar{u})^2}{2\sigma_{\mathrm{u}}^2}\right) \tag{5.14}$$

and a mixture of two pdf:s with parameters ($\bar{u}_1$, σ_{u1}) and ($\bar{u}_2$, σ_{u2}) becomes

$$f(u) = \frac{1-F}{\sigma_{\mathrm{u1}}\sqrt{2\pi}} \exp\left(-\frac{(u-\bar{u}_1)^2}{2\sigma_{\mathrm{u1}}^2}\right) + \frac{F}{\sigma_{\mathrm{u2}}\sqrt{2\pi}} \exp\left(-\frac{(u-\bar{u}_2)^2}{2\sigma_{\mathrm{u2}}^2}\right) \tag{5.15}$$

where $1-F$ is the fraction of stars in the pdf with ($\bar{u}_1$, σ_{u1}) and F the fraction with ($\bar{u}_2$, σ_{u2}). In Fig. 10, we plot the CDF of the periods for both the full sample and the limited sample. Fitting the two Gaussians to the CDF numerically, we find the best fit to be

$$\begin{aligned} &\bar{u}_1 = 0.753 \quad (\mathrm{P} = 5.66 \text{ days}), && \sigma_{\mathrm{u1}} = 0.180 \text{ dex} \\ &\bar{u}_2 = 1.477 \quad (\mathrm{P} = 30.0 \text{ days}), && \sigma_{\mathrm{u2}} = 0.201 \text{ dex} \\ &F = 0.039 \end{aligned} \tag{5.16}$$

In Fig. 10, we also include the modelled distribution, and can note that it does indeed seem to fit pretty well to the small sample while still keeping the tail corresponding to the long-period Cepheids. Our fitted model is the main period distribution used in the program, but we also use the CDF of the full sample for reference.

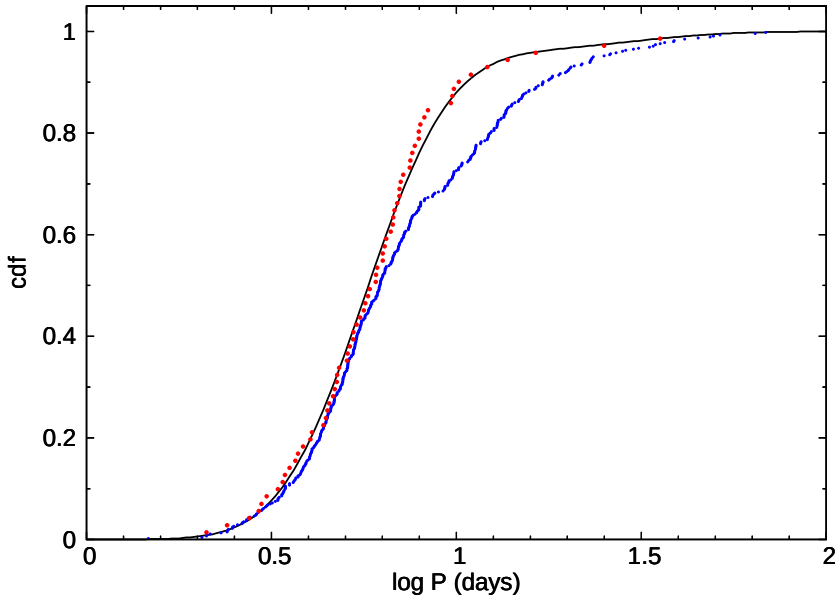


Figure 10: CDF of all observed Cepheids (blue), all Cepheids within the completeness limit (red) and the modeled period distribution (black).

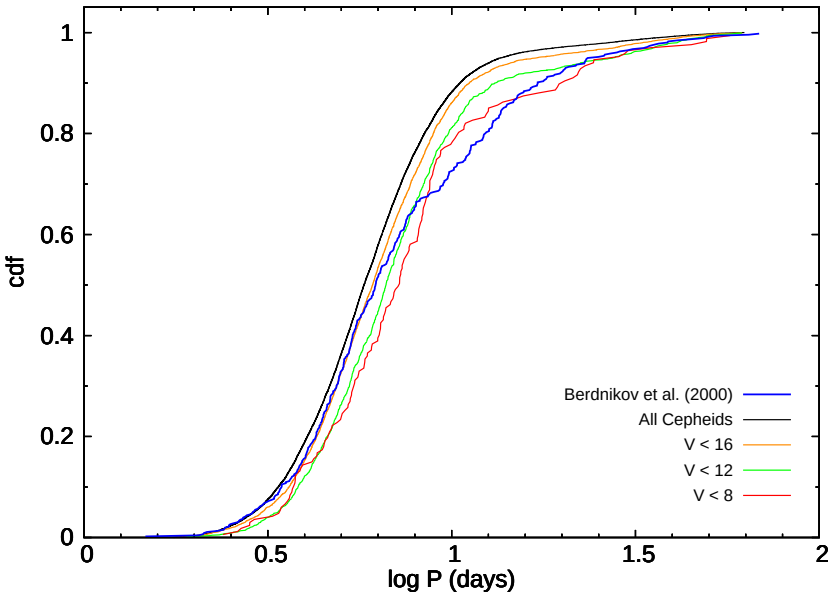


Figure 11: Comparison between the distribution of all Berdnikov periods (blue), the modelled periods (black) and the periods of all simulated Cepheids brighter than $V < 16$ (orange), $V < 12$ (green) and $V < 8$ (red).

To test our modelled distribution, we simulate a Galactic Cepheid population using data set D1 (one of the finished galaxy models described in Table 4), and in Fig. 11 plot the period distribution of all simulated Cepheids brighter than $V = 16$, $V = 12$ and $V = 8$ magnitudes. We also show the period distribution of the model and for all Cepheids in the Berdnikov catalogue. Had the catalogue been obtained from one single survey with a certain limiting magnitude, it would ideally have agreed with the simulated curve with the same magnitude limit. In the figure, we can see that Berdnikov curve seem to fit well for the $V < 16$ sample at shorter periods and with the $V < 12$ sample for longer periods. The problems in fitting a single sample to the Berdnikov catalogue is likely due to the fact that the catalogue is a compilation of several studies, and therefore does not have a strict magnitude limit.

5.2.5 Magnitude distribution

The longer period a Cepheid has, the brighter it is. From the P-L relation, we know the absolute visual magnitude of Cepheids to range roughly between -1 and -7, with the majority of Cepheids at the fainter end. This is very bright compared to most other stars (the solar absolute magnitude being $M_{V,\odot} = 4.83$ for reference). The brightness of Cepheids is one of the properties that makes them interesting for us, since they can be seen at large distances and through large extinctions.

From the P-L relation and a probable period distribution, we can work out the distribution of absolute magnitudes. To do this, any of the P-L relations presented in Table 2b could be used. Which relation we choose to use as our internal P-L relation will affect our results only minimally, as they are both very similar and the parameter fitting described in later sections is quite independent on the exact values for a and b. In our first model, we decided to use the relation from Sandage et al. (2004);

$$M_{\mathrm{V}} = -3.087 \log_{10} P - 0.914 \tag{5.17}$$

This is the primary P-L relation used in our Galaxy modelling, from now on referred to as the STR04 relation.

The different P-L relations found for stellar populations in the Galaxy, SMC and LMC suggest that the P-L relation is not global. It is today believed that the absolute magnitude of a Cepheid depends not only on its period, but also on other properties, with metallicity as the second most important parameter. With the relatively small sample of Cepheids with measured metallicities, the nature and degree of the dependence is still unclear (Sandage & Tammann 2006). In order to take the possible metallicity dependence into account, we consider the P-L relation described by Groenewegen (2008),

$$M_{\mathrm{V}} = -(2.60 \pm 0.09) \log_{10} P - (1.30 \pm 0.10) + (0.27 \pm 0.30)[\mathrm{Fe/H}] \tag{5.18}$$

where [Fe/H] is the metallicity normalized to that of the Sun,

$$[\mathrm{Fe/H}] = \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{star}} - \log_{10}\left(\frac{N_{\mathrm{Fe}}}{N_{\mathrm{H}}}\right)_{\mathrm{sun}} \tag{5.19}$$

where N_{Fe} and N_{H} are the number of iron and hydrogen atoms per unit volume, respectively. We note that the metallicity dependence in Eq. (5.18) is highly uncertain and that a zero dependence lies within the error bars. Groenewegen (2008) points out that their non-metallicity dependent fit to the data would be virtually identical to the metallicity dependent one but without the [Fe/H] term. We will not investigate this question any further, but instead simply study Galaxy models both with and without the last term in Eq. (5.18), and work out how they affect the parallax verification. For reference, we include both relations in our Galaxy model, and refer to them as the metallicity dependent and the plain G08 relation.

With the gas density increasing towards the Galactic center, we expect the star formation rate along with the metallicity to increase towards lower galactocentric radii. This possible metallicity gradient is discussed in Lemasle et al. (2008), who conclude that the metallicity of Cepheids goes as roughly -0.052 dex/kpc, which together with their observed mean metallicity of $\langle[\mathrm{Fe/H}]\rangle = 0$ in the Solar neighbourhood gives

$$[\mathrm{Fe/H}](R) = 0.416 - 0.052R \tag{5.20}$$

assuming $R = 8$ kpc for the Sun. Lemasle et al. (2008) show a scatter in metallicity of Cepheids at equal R which is significantly larger than the uncertainty in their determined metallicity abundances. In order to reproduce this scatter (part of it may be due to observational scatter), we assume the metallicity to be normally distributed around $[\mathrm{Fe/H}](R)$ with a standard deviation of $\sigma_{[\mathrm{Fe/H}]} = 0.1$ dex.

In our Galaxy model, we assume two different metallicity dependences. The first one is the metallicity gradient (Eq. (5.20)) plus scatter, as described above. In the second model, we assume $[\mathrm{Fe/H}](R) = 0$ plus the scatter. Together with Eq. (5.18), we now have several different ways to describe the possible metallicity dependence of Galactic Cepheids.

With Gaia observing not in the V band, but in G (described in Section 6), we also need the intrinsic $V-I$ colour to do the appropriate magnitude transformation. Fortunately, there also exists a Period-Colour (P-C) relation for Cepheids which for the Galactic Cepheids was determined by Tammann et al. (2003) to be

$$(V-I)_0 = (0.256 \pm 0.017)\log_{10} P + (0.497 \pm 0.016) \tag{5.21}$$

where $(V-I)_0$ denotes the intrinsic $V-I$ colour. We now have distribution models for all the necessary Cepheid properties, and are therefore able to simulate an entire population of Galactic Cepheids. We will now need to work out what each of these Cepheids will look like from the Sun. This is discussed in the next section.

5.3 Extinction and apparent magnitudes

When looking at the Cepheids from the Sun, their decrease in brightness will depend on two factors; the distance and the amount of extinction due to the intermediate gas and dust. In order to determine what the sky looks like from the Sun, we need to take both these factors into account. The distance modulus describes the decrease in magnitude in the following way:

$$V - M_V = 5 \log\ r - 5 - A_V \tag{5.22}$$

where V is the apparent visual magnitude of the object, M_V the absolute magnitude, r the distance in parsec and A_V the total interstellar extinction in the V band. In order to use the formula, we need to determine the extinction to an arbitrary three-dimensional point in the Galaxy. In this section, we describe three different models with increasing complexity. These are then evaluated in Section 8. A common property of all three is that they assume smooth extinction distributions, and do not take clumpiness into account. In fact, we know the extinction to be quite clumpy, but the final extinction model described in this section has been shown to fit quite well with reality. Even so, it is only an approximation of reality, but the three-dimensional mapping of the Galactic extinction is simply too limited today to be of any use to us.

In our first, most primitive and naive extinction model, we assume the dust disk to be a flattened cylinder of constant thickness and density. In this disk, within $|z| \leq 100$ pc, we have a constant extinction of $a_V = 1.0$ magnitude/kpc, and outside of the disk, at $|z| > 100$ pc, the extinction per kpc is $a_V = 0$. This uniform extinction model is extremely crude, however, and we need to develop it into something more sophisticated. Assuming a smooth extinction distribution, the extinction to a three-dimensional point in the Galaxy can be written as

$$A_V(l, b, r) = \int_0^r a_{\mathrm{V}}(l, b, r')dr' \tag{5.23}$$

where l is the galactic longitude, b is the latitude, r is the distance and $a_V(l, b, r)$ the extinction per unit length at some arbitrary point in the Galaxy. In reality, as has been discussed before, the density of matter in the disk decreases rapidly as we move away from the Galactic plane. A more accurate extinction model would take this into account, and in our second model, we assume the extinction to decrease exponentially with $|z|$. As we saw in Fig. 9, exponentials have trouble describing the region close to the plane, but do fit the stellar distribution pretty accurately at larger values of $|z|$. We can then write the the extinction per unit length at an arbitrary point as

$$a_V(l, b, r) = a_{\mathrm{V},0} e^{-|z|/h_{\mathrm{a}}} \tag{5.24}$$

where $z = r \sin b + z_\odot$ and h_{a} is the scale-height. The total extinction to a point at a distance r can then be obtained from Eq. (5.23),

$$A_V(l, b, r) = \int_0^r a_{\mathrm{V}}(l, b, r')dr' = a_{\mathrm{V},0} \int_0^r e^{-|z|/h_{\mathrm{a}}} dr' \tag{5.25}$$

This can be evolved analytically with the following result

$$A_V(r) = \begin{cases} \frac{a_{\mathrm{V},0} h_\mathrm{a}}{|\sin b|} \left[-e^{-(r' \sin b + z_\odot)/h_\mathrm{a}} \right]_{r'=0}^{r} & \text{for } z > 0 \\ \frac{a_{\mathrm{V},0} h_\mathrm{a}}{|\sin b|} \left(\left[-e^{-(r' \sin b + z_\odot)/h_\mathrm{a}} \right]_{r'=0}^{r_\mathrm{p}} + \left[e^{(r' \sin b + z_\odot)/h_\mathrm{a}} \right]_{r'=r_\mathrm{p}}^{r} \right) & \text{for } z < 0 \end{cases} \tag{5.26}$$

where $r_\mathrm{p} = z_\odot / |\sin b|$ is the distance to the Galactic plane along the line of sight. If we assume the extinction to be proportional to the density of interstellar matter, we can use a scale-height of $h_\mathrm{a} = 134$ pc as given in Marshall et al. (2006). For the extinction in the plane, we adopt the previously used value of $a_\mathrm{V} = 1.0$ magnitude kpc^{-1}, known to be true within the solar neighbourhood. This exponential extinction model is however also highly simplified, and is for example unable to differentiate between a point towards the Galactic Centre and a point at the same distance in the anti-centre direction. As seen in Section 5.2.1, the density of the disk increases towards the Galactic Centre, and we know that the same thing applies for the extinction.

The third and most realistic extinction model was developed by Amôres & Lépine (2005). They make similar assumptions to ours: that the interstellar dust is well mixed with the gas, and that the extinction therefore is proportional to the gas density. With gas distribution data based on several different Galactic infrared surveys, their model assumes an axisymmetrical distribution around the Galactic center where the extinction at an arbitrary point in the Galaxy depends on the density of neutral and molecular hydrogen and can be written as

$$a_V = \gamma(R) N_\mathrm{HI}(r, z) + 2\gamma(R) N_{\mathrm{H}_2}(r, z) \tag{5.27}$$

where R is the galactocentric radius, z is the vertical distance to the Galactic plane and r is the distance from the Sun, N_HI and N_{H_2} are the column densities of neutral and molecular hydrogen along r and γ is a proportionality constant. They found that a good fit to the gas density in the plane could be obtained by using a radial function given by

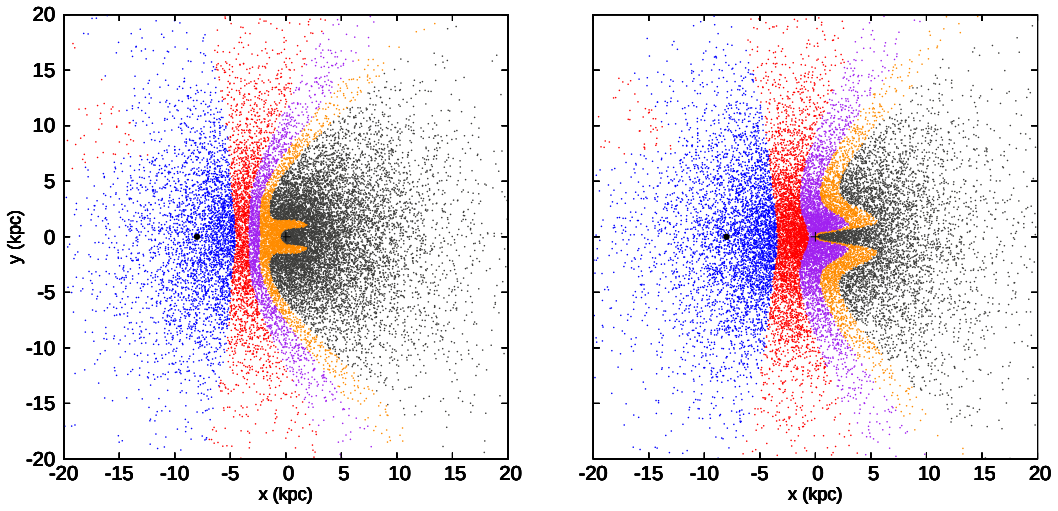
$$n(R) = C \exp\left(-\frac{R}{A} - \left(\frac{B}{R}\right)^2 \right) \tag{5.28}$$

with the parameters $A = 7.0$ and 1.2 kpc, $B = 1.9$ and 3.5 kpc and $C = 0.7$ and 58.0 cm^{-3} for HI and H_2 respectively. By describing the vertical distribution with a Gaussian, the gas density can then be written as

$$N(R, z) = n(R) \exp\left(-\left(\frac{z}{1.2h}\right)^2 \right) \tag{5.29}$$

where h is the scale-height. The scale height increases with the galactocentric radius as

$$h = k \exp(0.08R) \tag{5.30}$$



(a) $z = 0$ (b) $z = 100$ pc

Figure 12: Galactic extinction determined with the AL05 extinction model. In (a), we have determined the extinction from the Sun to points at a height of $z = 0$, and in (b), the points are situated at $z = 100$ pc. The bins are below 5 magnitudes (blue), 5 to 10 magnitudes (red), 10 to 15 magnitudes (purple), 15 to 20 magnitudes (orange) and above 20 magnitudes (black).

where R is given in kpc and $k_{\mathrm{HI}} = 45$ pc and $k_{\mathrm{H_2}} = 81$ pc. The total extinction is then given by Eq. (5.23). Solving this analytically is very complicated, and we therefore have to do this numerically using the numerical solver given by Amôres & Lépine (2005)[1]. This extinction model is hereafter referred to as the AL05 model.

In Fig. 12, we see the Galaxy face on with the Sun situated at $(x, y) = (-8, 0)$ kpc. With the use of the AL05 extinction model, we have calculated the extinction to every data point, and the different colours correspond to different degrees of extinction as described in the legend. In Fig 12a, all points are situated in the disk plane at $z = 0$, and in Fig. 12b, they are above the plane at a height of $z = 100$ pc. Note the asymmetry in extinction in the positive and negative x-direction, and the ridge of higher extinction at $R \approx 1$ kpc. Also note how at even a modest height of 100 pc, the extinction is substantially decreased when compared to the extinction in the disk plane.

We will also need the colour excess in the $V - I$ colour. This is defined as

$$E_{V-I} = (V - I) - (V - I)_0 \tag{5.31}$$

[1]Fixed a typo in the code by changing sign on $z_\odot$.

where $V - I$ is the apparent colour and $(V - I)_0$ the intrinsic colour. We can also define the extinction coefficient in $B - V$ as $R_{B-V} = A_V/E_{B-V} = 3.1$ (Cox 2000). The corresponding extinction coefficient in $V - I$ is not as clearly defined, but Tammann et al. (2003) finds $E_{V-I} = (1.283 \pm 0.011)E_{B-V}$, which means that

$$R_V - I = \frac{A_\mathrm{V}}{E_{V-I}} = \frac{1}{1.283}\frac{A_V}{E_{B-V}} = \frac{1}{1.283}R_{B-V} \tag{5.32}$$

and we can finally write the colour excess as

$$E_{V-I} = 1.283A_\mathrm{V}/R_B - V \tag{5.33}$$

Using Eq. (5.31) and our extinction model for A_V, we can now obtain the apparent V-I colour for every simulated Cepheid.

6 The Gaia model

The Gaia satellite will observe objects in the so-called G band, which is the effective transmission curve times the quantum efficiency of the CCD detectors (Jordi et al. 2006). The G band covers the wavelength range from 400 to 1000 nm, with a maximum response at $\sim$ 715 nm. The G magnitude is similar to the V band for blue objects, but much brighter than V for red objects. The limiting magnitude of $G = 20$ therefore corresponds to $V \simeq$ 20–25 depending on the colour of the object. This needs to be taken into account when determining whether a Cepheid is detectable by Gaia or not, and our Gaia observational model includes numerical methods to transform a V magnitude and a $V - I$ colour into the corresponding G magnitude. Lindegren (2010) writes this transformation as

$$G = V - 0.0107 - 0.0879(V - I) - 0.1630(V - I)^2 + 0.0086(V - I)^3 \tag{6.1}$$

The accuracy with which Gaia will determine the parallax of an object depends on its position on the sky and the apparent magnitude. As the satellite rotates around the Sun at the L2 Lagrangian point, it is going to cover the whole sky several times within its five year lifespan. The number of field transits range between 50 and 250 times depending on the position on the sky (see Fig. 2), leading to variations in the statistical accuracy on the order of $\sqrt{250}/\sqrt{50} \sim 2$ between stars with the same brightness. In Table 1, we see the effect of the apparent magnitude using the sky average errors for a G0V type star. We note a variation by a factor 30 between the brightest and the faintest stars on the sky. This is clearly an important factor when determining what the individual Cepheids will look like to Gaia.

The Gaia Data Processing and Analysis Consortium (DPAC) has developed a Java package called GaiaTools which includes means to simulate the Gaia satellite observations. Based on the object's position on the sky, the StandardErrors code in the GaiaTools simulates the number of field transits that Gaia will do on the object over

the complete mission length. In our Gaia model implementation, we first obtain the sky-averaged standard error, σ_π assuming along-scan (AL) estimates for a specific G magnitude (de Bruijne et al. 2010). Gaia is expected to be saturated for $G < 13$, leading to a complex error estimate at these magnitudes. In our implementation, we assume a constant σ_π for all of the saturated objects, leading to a slight overestimate of σ_π for the brightest stars. The sky-averaged error is then scaled depending on the object's position on the sky by simulating the number of transits over that specific point.

For every object, we recieve from GaiaTools a predicted standard error in the measured parallax, σ_π. Assuming the measured errors follow a Gaussian distribution, we simulate the parallax value observed by Gaia as a normal distribution

$$\pi_{\mathrm{G}} \sim N(\pi_{\mathrm{true}}, \sigma_\pi^2) \tag{6.2}$$

where π_{true} is the true value and σ_π^2 the variance. The CDF of the Gaussian function cannot be solved analytically, so in order to obtain the measured parallax, we use the Box-Muller transformation, designed to generate pairs of normally distributed random numbers from a pair of uniformly distributed random numbers. If we let U_1 and U_2 be two uniformly distributed values between 0 and 1, we then get two independent normally distributed values, Z_1 and Z_2, from

$$Z_1 = \sqrt{-2\log U_1}\sin 2\pi U_2 \tag{6.3}$$

$$Z_2 = \sqrt{-2\log U_1}\cos 2\pi U_2 \tag{6.4}$$

and we get the simulated parallax measurement from

$$\pi_{\mathrm{G}} = \mu_\pi + Z\sigma_\pi \tag{6.5}$$

where Z is either of the two generated Box-Muller values.

Gaia will also make multi-epoch photometric observations between 320–1000 nm, which will make it possible for DPAC to provide good measurements of the Cepheid periods without any time-consuming follow-up observations. Over the course of five years, Gaia will make between 50 and 250 observations per object. This is a lot, and means that the periods can be determined with very good accuracies. In our simulations, we assume that σ_P is negligible.

7 Statistical analysis

In the previous sections, we have described different Galactic Cepheid distribution models, different Galactic extinction models and the Gaia observational model used to simulate five years of observations. With all this, we can simulate the number of Cepheids that Gaia will be able to observe, and also the measured parallax and known parallax standard error for every individual Cepheid. In this section, we

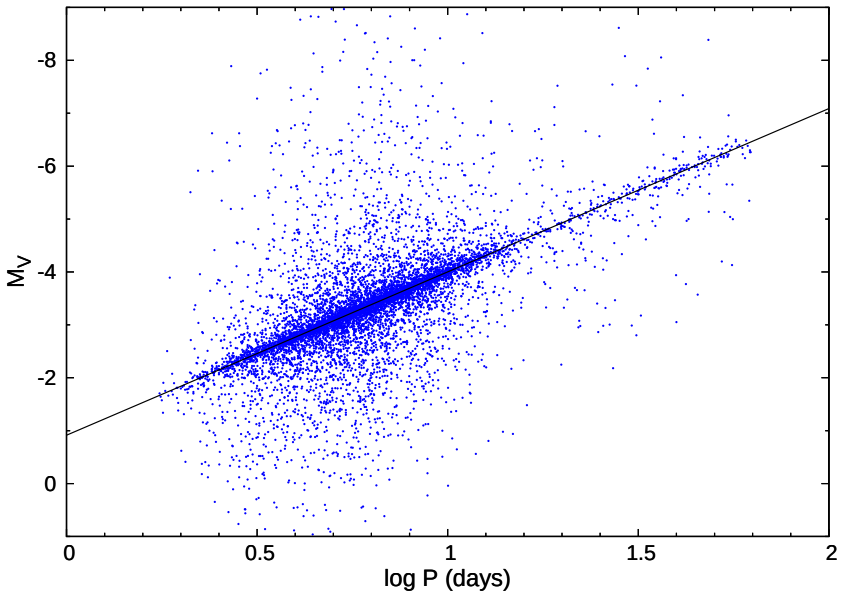


Figure 13: The P-L relation plotted for all observed Cepheids assuming observational parallax errors but perfect knowledge of the extinction. The black line represents the true P-L relation.

describe the statistical tools used to analyze the simulated data and in particular the parallax zero-point. The method simultaneously makes a fit to the P-L relation and the parallax zero-point.

7.1 Parameter fitting

As a first test of our modelled Galaxy, we investigate the relation between the observed Cepheid periods and apparent visual magnitudes in order to recover the P-L relation that was used to generate the Cepheids (as was described in Section 5.2.5). Ideally, we would want our fit to be identical to the P-L relation used, but we will also see some additional scatter due to the uncertainty in the measured parallaxes. We can assume that the uncertainty with which Gaia will determine the period and apparent visual magnitude is negligible. In order to estimate the absolute magnitude for every Cepheid, we use the distance modulus (given in Eq. (5.22)) and solve for the absolute magnitude

$$M_V = V - 5 \log\ r_{\mathrm{G}} + 5 + A_V \tag{7.1}$$

where V is directly observed by Gaia, $r_{\mathrm{G}} = 1/\pi_{\mathrm{G}}$ is determined from the measured parallax and we pick A_V to be the correct one. In this way, we can expect the scatter to be solely due to the error in the parallax measurements. The parallax uncertainty will in some cases lead to negative measured parallaxes. With this simple method, we have to ignore these Cepheids. The P-L relation we obtain with this crude fit is very close to the used P-L relation, and is given in Fig. 13.

Zero-point fitting

We now introduce one more unknown parameter: the parallax zero-point, c. We do not only want to determine the value of c, but also its standard deviation, σ_c. The errors in the parallaxes measured by Gaia are symmetric around the true value, for every individual measurement. This is not the case if we instead convert the parallaxes to distances or absolute magnitudes. This means that in order to avoid introducing errors from non-linear conversion to other units, it is vital that we keep our equations in units of parallax. Working with parallaxes also gives us another advantage versus working with distances, and that has to do with the fact that, due to uncertainties in the measurements, we will end up with a non-negligible fraction of negative parallaxes. Negative observed parallaxes are perfectly physical, but they cannot be converted to distances or magnitudes. When we make our fit to the P-L relation, all measurements are therefore perfectly usable, and the negative parallaxes should indeed be included to avoid biases.

Starting with the distance modulus of Eq. (7.1), we rewrite it and solve for the parallax

$$\pi = 10^{5+0.2(M_V - V + A_V)} \qquad [\mu\text{as}] \tag{7.2}$$

We now use the fact that the absolute magnitude can be determined from the P-L relation and write $M_\text{V} = a \log P + b$ and let the constants a and b be free parameters. Finally, we introduce the parallax zero-point, c, and the observation equation for the parallax can be written as

$$\pi_\text{G} = 10^{5+0.2(a \log P + b - V + A_V)} + c \qquad [\mu\text{as}] \tag{7.3}$$

where we assume that the period, P, and the apparent visual magnitude, V, can be directly obtained from Gaia with negligible uncertainty. In this first version, we assume that the extinction, A_V, can be inferred by some independent means, either from the observed reddening or from the use of some extinction model. We then have three free parameters, a, b and c which can be fitted simultaneously to the complete set of observed Cepheids.

As can be seen in Eq. (7.3), the data are not a linear combination of the three parameters, and we therefore need to use non-linear least-squares to fit the parameters to the data. We choose to solve this using the Newton-Raphson iterative method, where we let $\underline{\theta} = (a, b, c)^T$ be a column vector of parameters. In this method, we assume some initial values $\underline{\theta}^{(0)}$ of the three parameters. $\underline{\theta}^{(k)}$ then stands for the kth parameter approximation, and in each iteration, we add a correction term, $\Delta\underline{\theta}$, meaning that

$$\underline{\theta}^{(k+1)} = \underline{\theta}^{(k)} + \Delta\underline{\theta}^{(k)} \tag{7.4}$$

The correction term is obtained from

$$\Delta\underline{\theta}^{(k)} = \left[(\underline{A}^{(k)})^\text{T} \underline{C}^{-1} \underline{A}^{(k)}\right]^{-1} (\underline{A}^{(k)})^\text{T} \underline{C}^{-1} \left[\Delta\underline{\pi}\right] \tag{7.5}$$

where $\underline{C} = \mathrm{diag}(\sigma_\pi^2)$ is the diagonal covariance matrix of the observed parallaxes, and $\Delta\underline{\pi}$ is the vector of residuals, with elements $\Delta\pi_i = \pi_{G,i}^{(\mathrm{obs})} - \pi_{G,i}(\theta)$. Each element in $\underline{A}^{(k)}$ is given by $A_{ij}^{(k)} = \partial\pi_{G,i}/\partial\theta_j$ and $\pi_{G,i}^{(k)}$ is the parallax obtained from Eq.(7.3) at iteration k.

The correction term is then repeatedly applied, and as long as the initial parameters are not too bad, the correction term will get smaller and smaller and we finally end up with a fit that is sufficiently accurate. The Newton-Raphson solution converges quadratically, which means that we need very few iterations to get an accurate fit. As long as $\pi_G(\theta)$ is not too strongly non-linear, we find that the covariance of the parameters can be written as

$$\underline{\mathrm{Cov}}[\theta] \approx (\underline{A}^{\mathrm{T}}\underline{C}^{-1}\underline{A})^{-1} \tag{7.6}$$

where the resulting parameter covariance matrix in our case will have the form

$$\underline{\mathrm{Cov}}[\theta] = \begin{bmatrix} \sigma_a^2 & \rho_{a,b}\sigma_{\mathrm{a}}\sigma_b & \rho_{a,c}\sigma_a\sigma_c \\ \ldots & \sigma_b^2 & \ldots \\ \ldots & \ldots & \sigma_c^2 \end{bmatrix} \tag{7.7}$$

where σ_θ is the uncertainty in each parameter, and $\rho_{i,j}$ is the correlation factor between two parameters. We can therefore easily determine the formal uncertainties σ_{a}, σ_{b} and σ_{c}.

Extinction-free and metallicity-dependent methods

With the above-mentioned method, we are able to obtain the parallax zero-point, based on the assumption that we can determine the extinction to each individual object to a very good accuracy. Determining the extinction to individual objects is today certainly possible, but would require time-consuming follow-up observations of all the thousands of Cepheids that Gaia observes. As a possible alternative, we introduce a second, extinction-free P-L relation equation, using the fact that for Classical Cepheids we have a Period-Colour relation for which $V-I$ can be written as

$$(V-I)_0 = d\log_{10} P + e \tag{7.8}$$

where $(V-I)_0$ is the intrinsic colour and d and e are constants. The extinction in the V band can be written as

$$A_V = \big((V-I) - (V-I)_0\big)R_{V-I} \tag{7.9}$$

where $V-I$ is the apparent colour and $R_{V-I} = A_{\mathrm{V}}/E_{\mathrm{V-I}}$ the extinction coefficient. Inserting the above equation into Eq. (7.3) gives

$$\pi_G = 10^{5+0.2(a\log P + b - V + ((V-I)-(V-I)_0)R_{V-I})} + c \qquad [\mu\mathrm{as}] \tag{7.10}$$

and inserting the P-C relation then gives

$$\pi_G = 10^{5+0.2(k_1 \log P + k_2 - V + R_{V-I}(V-I))} + c \qquad [\mu\text{as}] \tag{7.11}$$

where $k_1 = a - dR_{V-I}$ and $k_2 = b - eR_{V-I}$. It is necessary to introduce k_1 and k_2, since it would not be possible to solve separately for all four parameter a, b, d and e. Solving this equation with the three free parameters k_1, k_2 and c will therefore not give us any information about the constants in the P-L relation or the P-C relation, but will instead give us a fit to the parallax zero-point that is almost extinction-independent, as we know R to a much better accuracy than A_V.

As a third zero-point fitting method, we assume that the P-L relation is metallicity dependent. This means that we can write the P-L relation as

$$M_V = a \log P + b + m\,[\mathrm{Fe/H}] \tag{7.12}$$

where m is the metallicity dependent constant and [Fe/H] is the metallicity of the Cepheid. The parallax is then given by

$$\pi_G = 10^{5+0.2(a \log P + b + m[\mathrm{Fe/H}] - V + A_V)} + c \qquad [\mu\text{as}] \tag{7.13}$$

We have so far, in all of the three parameter fitting methods, assumed that we know the extinction A_V or extinction coefficient R_{V-I} exactly. In reality, this will never be the case. In order to investigate the extinction dependence of our methods, we therefore need to implement different possible errors in our assumptions.

7.2 Measurement errors

In this section, we discuss the assumption errors that have been used during analysis methods in order to simulate errors in the measurements. Errors are included for extinction, metallicity and $V - I$ colour.

We have used four different extinction error types. During the experiments, the AL05 model was chosen as the default to determine the extinction in the Galaxy. In our first and most crude error type, we simply assume a different model than the one used. Instead of correctly assuming the AL05 model, we assume the exponential model, which lacked the radial dependence. This is likely to lead to major errors in the extinction, and is probably not very realistic.

For our second and third extinction error type, we assume an error related to the real extinction to the Cepheid. In both cases, we assume a Gaussian error σ_A around the true A_V value. The extinction assumed during the parameter fitting is then obtained by use of the Box-Muller transformation (given in Eqs. (6.3)–(6.5)). In Tammann et al. (2003), they show the difference in the colour excess estimated in the Fernie catalogue for 321 Galactic Cepheids and the mean colour excess estimated by nine other groups. The result hints towards a combination of constant σ_A and

a σ_A which is proportional to A_V. In our second error type, we therefore assume a constant $\sigma_A = 0.05$ magnitudes, independent of A_V. In our third error type, we assume $\sigma_A = 0.05 A_V$. This will give us the upper and lower boundaries of the uncertainty in the parameter fitting. The constant error will likely overestimate the error for nearby Cepheids but underestimate it for more distant ones. The scaling error will in turn underestimate the error of nearby Cepheids, and overestimate the distant ones.

Finally, for our extinction-minimized parallax zero-point fit, we do not consider an error in the assumed extinction. Instead, we introduce an error in the extinction coefficient, $R_{\mathrm{V-I}}$. Here, we assume a constant error 5% larger than the real value, from $R_{\mathrm{true}} = 3.1/1.283 = 2.416$ to $R_{\mathrm{bias}} = 1.05 R_{\mathrm{true}} = 2.537$. This will systematically overestimate the colour excess, which should cause a bias and overestimate the Cepheid brightnesses. The 5% bias is likely to be an overestimate of the real error, and how much this actually affects our results will need to be investigated.

During the experiments, we assume an error in the measured metallicity abundance around $[\mathrm{Fe/H}]_{\mathrm{true}}$ of $\sigma_{[Fe/H]} = 0.1$ dex. This is based on Lemasle et al. (2008), where they have formal iron abundance errors of ~ 0.12 dex. For the uncertainty in $V - I$ colour arising from the problems involved in correctly determining the mean colour of a variable object, we assume $\sigma_{V-I} = 0.01$ (private communication with Lennart Lindegren).

8 Results

We are now able to create synthetic Galactic populations of Cepheids and with the help of extinction models determine what they will look like from the Sun. We can simulate the Gaia observations including the parallax measurements and the number of observable Cepheids. From this, we are able to determine the P-L relation constants and the parallax zero-point by means of least-squares estimation. A list of the distribution models, possible analysis assumptions and important parameters that have been discussed in the previous sections is given in Table 3.

To use and analyze the information we have obtained throughout Sections 5–7, we need to do extensive testing with combinations of the different models given in Table 3. For this, we implement all the described models in a Java program called CepheidObsModel. Java is the language of choice here for one major reason, and that is the ability to easily implement the DPAC GaiaTools package, which is written in Java. It is also an object oriented language which makes it natural to divide the code into individual objects such as the Cepheid, Galaxy and Gaia classes. We also know that with a few tens of thousands of Cepheids that will only use a limited number of operations, we will not have any issues with the performance. See Appendix B for a more detailed discussion and description of the program.

Table 3: The properties and their different parameters available in the program.

Property	Models
Cepheid radial distribution	• Disk (Exponential with $R_0 = 2.5$ kpc) • OB (Exponential with $R_0 = 3.5$ kpc) • AL05 (Following H_2)
Cepheid vertical distribution	• Sech distribution with $n = 2$ and $z_0(R)$
Cepheid period distribution	• Modelled following completeness limit sample • Following total Cepheid sample
Cepheid metallicity distribution	• Independent of R. Scatter of $\sigma = 0.1$ dex • Radial gradient with a scatter of $\sigma = 0.1$ dex
Cepheid P-L relation	• STR04 ($M_V = -3.087 \log P - 0.914$) • G08 ($M_V = -2.60 \log P - 1.30$) • G08 met dep ($M_V = -2.60 \log P - 1.30 + 0.27\,[\mathrm{Fe/H}]$)
Extinction model	• Uniform within $\lvert z\rvert = 100$ pc • Exponential around Galactic plane • AL05 following Galactic gas
P-L relation fit	• Least-square fit to a, b and c • Least-square fit to a, b, c and M • Least-square fit to k_1, k_2 and c
Assumed extinction	• True extinction • Exponential extinction model • Gaussian error around A_{true} with $\sigma_A = 0.05$ • Gaussian error around A_{true} with $\sigma_A = 0.05 A_V$ • Biased extinction coefficient with $R_{\mathrm{bias}} = 1.05 R_{\mathrm{true}}$
Including all observed Cepheids	• Yes • Excluding bulge and GC ($R < 1$ kpc)

We will now need to analyze the impact the different distribution models and analysis assumptions have on our resulting parameter fits. Investigating all possible combinations from Table 3 is not possible, however, as there are more than a thousand different combinations. It is obvious that we need to sort out and focus on the most interesting ones. In order to do this, we can divide the different models into two parts. The Cepheid radial, vertical, period and metallicity distributions along with the Cepheid P-L relation and Galactic extinction model constitute what we can call the galactic data. A combination of these corresponds to one specific galaxy, and as such gives rise to one specific data set. This data set can be analyzed in many different ways, depending on the parameter fitting method, the extinction assumed during the fitting and whether the data includes the Cepheids in the bulge and Galactic Centre or not. A combination of these three conditions is what we can call an analysis set. The combination of a specific data and analysis set constitutes

an experiment.

In Table 4, we have compiled a list of the ten most interesting data sets that we have chosen to work with. They are denoted D0 through D9. We decided that the AL05 extinction model was the only one realistic enough for our purposes, and have therefore used it for all ten sets. We also decided that the modelled period distribution using only the sample within the completeness limit would act as the default one, with only one set using the total period sample for reference. The metallicity gradient is included in two sets where we have a metallicity dependent P-L relation.

The ten most interesting analysis sets, denoted A0 through A9, are given in Table 5. Here, we had to rationalize away the removal of the Galactic Center population and limit it to only one set, as it will later be shown that this has a limited effect on the results. Assuming an exponential extinction is only done in one of the sets since this is not a realistic way to determine extinction. In our experiments, we omit the uniform extinction model for the same reason.

Each data set is then investigated with all analysis sets, which gives a total of 100 experiments. They are named after data and analysis set, e.g. D1A3 means using data set D1 and analysis set A3. There are seven or nine parameters of interest depending on the experiment. In all sets, we are interested in the total number of Cepheids observed by Gaia, N_{obs}. We are also interested in the estimated values for the parameters a, b and c (or k_1, k_2 and c for the extinction-free experiments), as well as their formal standard deviations. In some of the experiments, we are also interested in m and σ_m. In order to investigate possible biases, all experiments are realized 100 times each. In these runs, we use the same 100 random generator seeds for each experiment, meaning that each identical data set is analyzed in 10 different ways. From these runs, all individual results as well as the mean result for every parameter of interest have been saved and analyzed.

We divide our discussion of the results into two parts. In the first part, we discuss the results we get and what conclusions we can draw from our modelling of the Galactic Cepheid population. In the second part, we present and discuss the results of of the parameter fitting.

8.1 The simulated galaxies

Our aim to investigate the verification of the Gaia parallaxes has also resulted in a modelling of the Galactic Cepheids at an unpreceded level of detail. In this section, we discuss some of our results from this modelling that can be discussed separately from the main results in Section 8.2. This includes predictions of what the Cepheid population will look like, both in general and as viewed from the Sun in the Gaia survey.

Table 4: The ten data sets investigated in this work. All sets use the AL05 Galactic extinction model.

	Cepheid radial distribution			Cepheid period distribution		Cepheid metallicity gradient		Cepheid P-L relation		
	Exp from H	Exp from OB	Following H_2	Modeled	Total sample	No	Yes	STR04	G08	Met dep G08
Set D0	×			×		×		×		
D1		×		×		×		×		
D2		×		×		×			×	
D3		×		×		×				×
D4		×		×			×			×
D5		×			×	×		×		
D6			×	×		×		×		
D7			×	×		×			×	
D8			×	×		×				×
D9			×	×			×			×

Table 5: The ten different analysis methods investigated in this work.

	P-L relation fit			Assumed extinction					Bulge pop included?	
	a, b, c	a, b, c, m	k_1, k_2, c	True extinction	From exp model	$\sigma_A = 0.05$	$\sigma_A = 0.05A_V$	Wrong R_{V-I}	Yes	No
Set A0	×			×					×	
A1	×				×				×	
A2	×					×			×	
A3	×						×		×	
A4	×						×			×
A5		×		×					×	
A6		×				×			×	
A7		×					×		×	
A8			×	×					×	
A9			×					×	×	

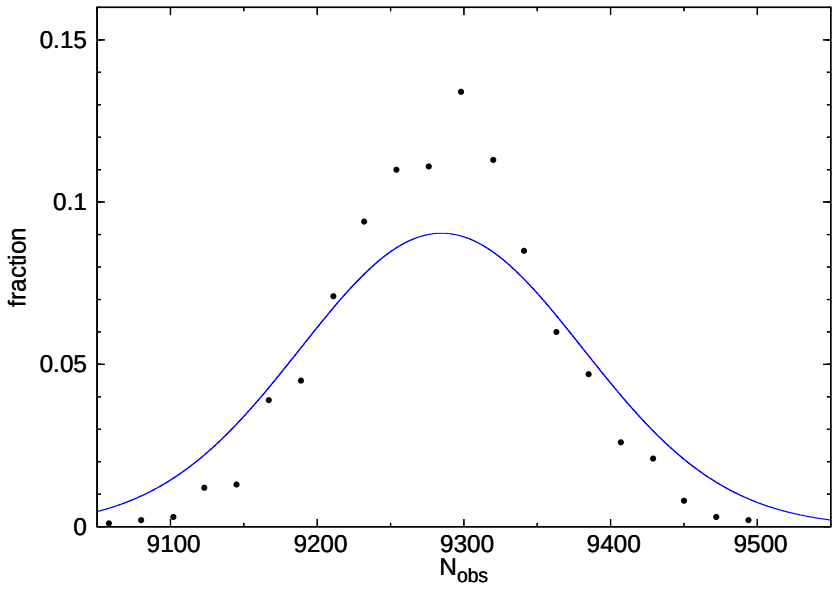


Figure 14: The distribution of observed numbers of Cepheids after 100 runs with data set D1. The blue curve represents a normalized Gaussian distribution with $\mu = \langle N_{\rm obs} \rangle$ and $\sigma = \sqrt{\langle N_{\rm obs} \rangle}$.

Both as a test of our program and in order to study the observed number of Cepheids, we plot in Fig. 14 a histogram of the number of Cepheids observed by Gaia in 100 realizations of data set D1 using CepheidObsModel. In the figure, we also show the normalized Gaussian distribution with a standard deviation $\sigma_N = \sqrt{\langle N_{\rm obs} \rangle}$ as expected from the statistical Poisson noise. We can note that our simulated distributions end up with a standard deviation more narrow than what we expect from the Poisson noise, with $\sigma_{N,\rm obs} = 71$ versus $\sqrt{\langle N_{\rm obs} \rangle} = 96$. We also see a symmetric distribution around the mean number with no unexpected features, and the same goes for the other data sets. This means that the mean observed number describes $N_{\rm obs}$ well, and we can use it in our result tables. After 100 runs, $\langle N_{\rm obs} \rangle = 9283$ for D1 and $\langle N_{\rm obs} \rangle = 9024$ for D0, less than a 3% difference even with the large difference in scale-length between the two exponential distributions. The uncertainty in the radial scale length will therefore not lead to any significant differences. The AL05 radial model, however, deviates from the other two. For D6, we find $\langle N_{\rm obs} \rangle = 14912$ Cepheids, over 60% more than for the two exponential models.

Figure 15 is a histogram of the apparent visual magnitudes for all observed Cepheids in data sets D0, D1 and D6, as well as in the Berdnikov catalogue. The most striking feature of the plot is the increase in numbers across the whole range. Gaia will observe Cepheids more than ten magnitudes fainter than we do today, and the number of Cepheids observed will surprisingly enough increase even at magnitudes as bright as $V = 7$. This leads to a shift in the mean V, which for the Berdnikov catalogue

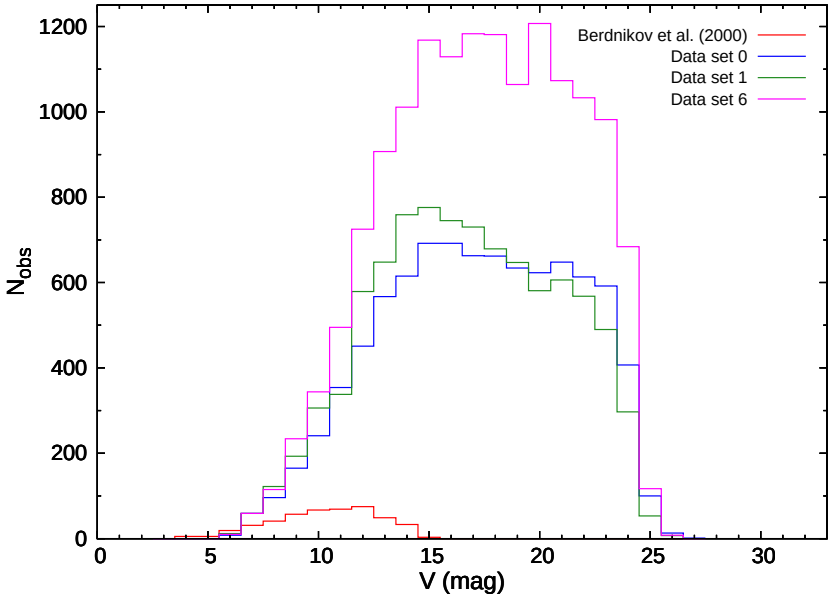


Figure 15: The distribution of V magnitudes for all observed Cepheids in data sets D0 (blue), D1 (green) and D6 (violet). Also included is the V distribution for the Berdnikov catalogue (red).

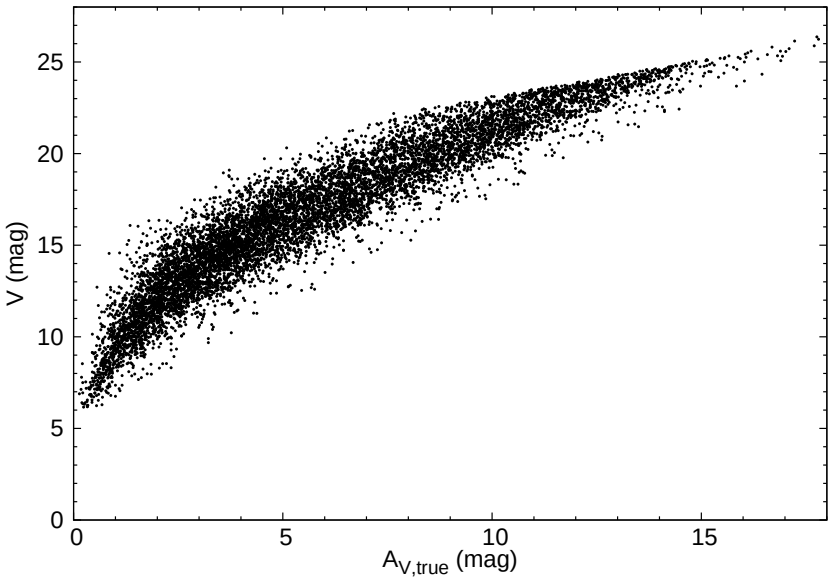
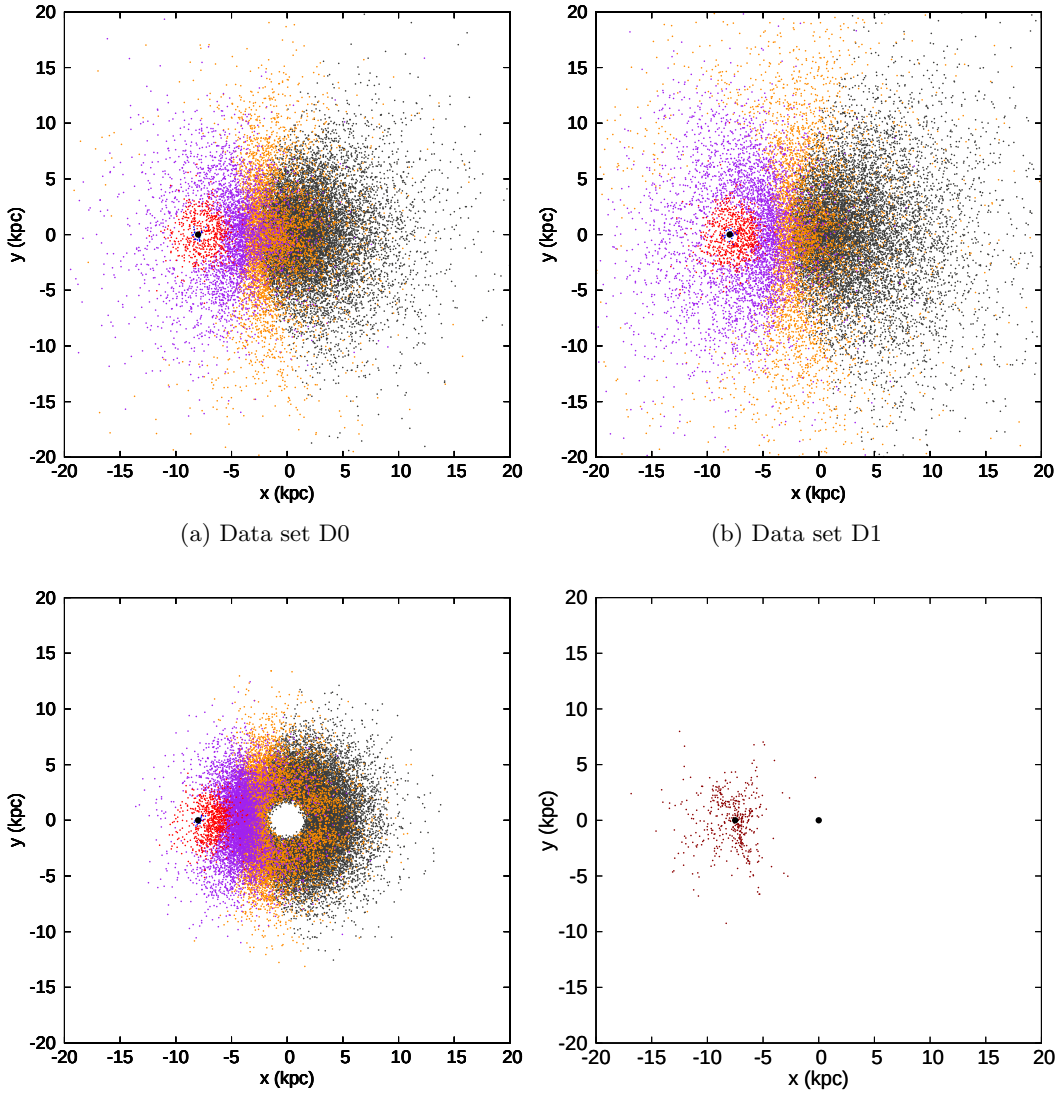


Figure 16: V magnitude versus A_V plotted for all observed Cepheids in data set D1.

is $\langle V \rangle = 10.3$ and for D0, D1 and D6 is 17.2, 16.7 and 17.4, respectively. Overall, the distribution in V seems quite independent on what data set is used, although the numbers are different. It is interesting to note that the Gaia G band allows observation of stars as faint as $V = 25$. This happens because the G band is designed to be more sensitive towards redder colours than V. As can be seen in Fig. 16, all Cepheids above $V = 20$ have suffered extinctions between 5 and as much as 18 magnitudes. From Eqs. (5.31) and (5.33), we can see that an extinction of $A_V = 10$ will lead to a very large colour excess of $E_{V-I} = 4.1$. Where the mean intrinsic colour of a Cepheid is $(V - I)_0 = 0.7$, its observed colour will be $V - I = 4.8$. This is a significant reddening which will cause a notable difference between the G and V bands. This effect leads to a significant increase in the observed Cepheid number by observing in G. In D1, there are 2290 observed Cepheids with $V > 20$ and $G < 20$.

Figure 17 shows the Galaxy face on with all its Cepheids divided into different apparent G magnitude bins. Figures 17a–17c show the results from the runs of data sets D0, D1 and D6. The differing factor between these three data sets is the radial distribution. As we also saw in the distribution of observed numbers and magnitudes, the two exponential distributions yield very similar results, with a slightly less compact distribution in D1 than D0. The AL05 model given in Fig. 17c results in larger differences. The doughnut shape arises because we exclude the Galactic Centre population. Overall, the Cepheid density is much higher and centrally focused here than in the other models. With this compact structure, we can see more Cepheids than in the previous two models, since the denser distribution towards the Galactic Centre means that we have more Cepheids that suffer little extinction. When comparing these three plots with the Berdnikov Cepheids plotted in Fig. 17d, it is quite impossible to rule out any of the models. Had we today known of more Cepheids towards the edge of the Galaxy, however, it would have been much easier to determine a more accurate radial distribution model. Overall, these figures illustrates very nicely the huge boost in knowledge of the Galaxy that Gaia will bring to us. Whatever the true Cepheid distribution is, Gaia will be a huge help when it comes to determining both radial distribution and the total number of Cepheids, as we will go from seeing only a few per cent of the Galactic Cepheids to almost half the population.

In Fig. 17, we note that even though the extinction towards the Galactic Centre is believed to be very large, several hundred Cepheids are still visible near and even behind the Galactic Centre in all the data sets. In order to study this further, we plot in Fig. 18 the distance from the Sun projected on the plane versus the vertical position for all Cepheids within $|l| < 5°$. We can clearly see the effect of a rapidly increasing extinction. At distances larger than 5 kpc, there are no visible Cepheids in the plane. We can also note the decrease in extinction with increasing $|z|$. All Cepheids with projected positions near the centre are positioned at $|z| > 100$ pc with the trend continuing at larger distances from the Sun. In fact, according to our simulations, Galactic Cepheids will be visible to Gaia even if they are positioned



(a) Data set D0

(b) Data set D1

(c) Data set D6

(d) Berdnikov catalogue

Figure 17: The Galactic Cepheids as observed from the Sun. The Galaxy is seen face on, with the different colours corresponding to different apparent G magnitude bins ($G < 0$ blue, $5 < G < 10$ red, $10 < G < 15$ violet, $15 < G < 20$ orange, $G > 20$ black). The Sun is positioned at $(-8, 0)$ kpc, and the Galactic center at (0,0). Panels (a), (b) and (c) correspond to different data sets (see Table 4), and (d) shows the Berdnikov Cepheids for reference.

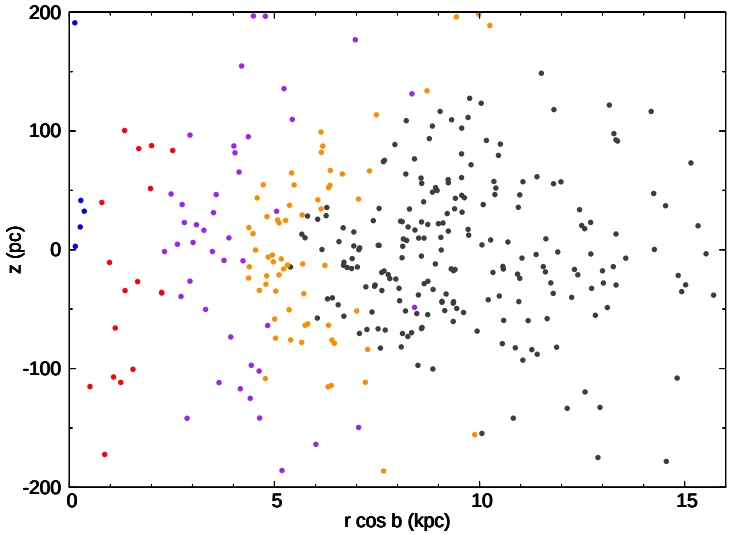


Figure 18: The simulated inner Galaxy ($|l| < 5°$) as observed using data set D1, with projected distance to the Sun plotted versus height above the Galactic plane. The different colours correspond to different apparent G magnitude bins ($G < 0$ blue, $5 < G < 10$ red, $10 < G < 15$ violet, $15 < G < 20$ orange, $G > 20$ black).

at the other side of the galaxy with heliocentric distances of up to 30 kpc. These stars are all found at large $|z|$ between 1 and 2 kpc, where the extinction is relatively small. Finding any Cepheids at all at these heights requires large scale-heights. In the solar neighbourhood, the scale-height is around 75 pc, which results in a very small fraction of stars with $|z| > 300$ pc. From Eq. (5.9), we see that a scale-height of 300 pc is possible at a galactocentric distance of 25 kpc, which would allow Cepheids at $|z| > 1$ kpc.

Figure 19 shows a heliocentric view of the Galaxy in galactic coordinates. For this figure, we use data set D1, divided into different G magnitude bins. We can first of all note that all Cepheids are found at very low latitudes, with only a few objects at $|b| > 5°$ and the absolute majority at $|b| < 1°$. A closer inspection shows that even though the physical distribution is more or less symmetric around $b = 0$, the distribution in the different magnitude bins is not. There are more black dots ($G > 20$) at negative latitudes than positive, and the number of orange dots ($15 < G < 20$) is visible larger on positive latitudes. This is expected, and comes from the Sun's position at $z_\odot = 20$ pc: light from Cepheids at negative latitudes will need to pass through the highest extinction region at $z = 0$, resulting in generally fainter objects.

8.2 Results of the parameter fitting

The results from all 100 experiments are given in Tables 8–10 found in Appendix C. The experiments have been divided into three groups, with each table corresponding

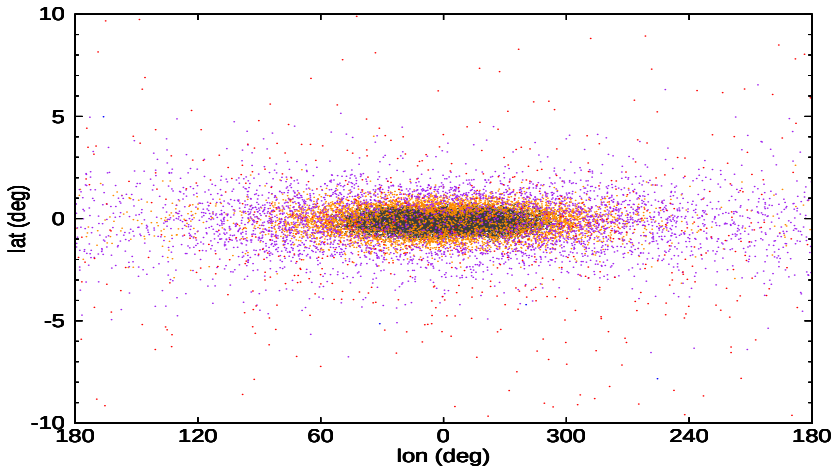


Figure 19: The Galactic Cepheids of data set D1 plotted in galactic coordinates. The different colours correspond to different apparent G magnitude bins ($G < 0$ blue, $5 < G < 10$ red, $10 < G < 15$ violet, $15 < G < 20$ orange, $G > 20$ black).

to one intrinsic P-L relation. Each experiment is realized 100 times, and we list the mean number of observed Cepheids as well as the mean value, $\langle\theta\rangle$ and formal uncertainty, σ_θ for each fitted parameter. The difference between the true value and $\langle\theta\rangle$ shows the bias. In the tables, we also give the real sample standard deviation for the c parameter (denoted $\sigma_{c,\mathrm{s}}$). This parameter does not depend on the possible zero-point bias, and can therefore be compared with the formal uncertainty σ_c. Finally, we also include the trueness of fit, which we define as the fraction of runs with a fitted parallax zero-point $|c| < 0.5$ μas. This limit is chosen so that the fits obtained in experiments with correctly assumed extinction have a trueness of 1 (for example experiments D0A0 and D1A0). This means that the trueness statistic measures the quality of the assumed analyis model, taking both bias and uncertainty into account.

In order to determine the trueness of each experiment, we need to differentiate between uncertainties and biases. Some effects give rise to statistical uncertainties, symmetric scatter around the mean parallax zero-point of all experiments, $\langle c\rangle$. Others give rise to biases which cause $\langle c\rangle$ to shift from the true value of 0. These biases can occur for example when we have a systematic under- or overestimation of the extinction. In these experiments, we have such a large number of models and intertwined effects in each experiment that the outcome is very difficult to predict. In order to study both uncertainties and biases, we have plotted the distribution of the fitted c values for all 100 runs of each experiment. These plots, sorted after data set, have all been included in Appendix C. In all these plots, we have also included the expected formal distribution from $c \sim N(0, \sigma_c^2)$ as the blue curve and the sample mean value $\langle c\rangle$ as the red vertical line.

We divide the rest of this section into three parts. In the first, we discuss in more

detail the results from a selected four typical experiments. In the second part, we show the effects of some of the other parameters (such as an intrinsic metallicity dependent P-L relation) and in the third, we discuss the effects of limiting the Cepheid sample by extinction or magnitude.

8.2.1 Typical experiments

For the detailed study, we select data set D1, and experiments D1A0–D1A3 in particular. We let each of these experiments run 1000 times and study the distribution of the fitted parameters. In Fig. 20, we plot the distribution of fitted a values, and in Figs. 21 and 22, we do the same for b and c. In all these plots, we include the expected formal distribution and the sample mean value as described earlier. In general, the formal error σ_θ comes from the uncertainty in the measured parallax of each Cepheid, but we also have uncertainties from the assumed extinction, $A_{V,\mathrm{obs}}$, as well as metallicity, $[\mathrm{Fe/H}]_{\mathrm{obs}}$ and colour, $(V-I)_{\mathrm{obs}}$. In experiments D1A0–D1A3, however, only the extinction uncertainty is of importance during the parameter fitting. The discrepancy between the histogram and the blue line therefore arises only from the uncertainty in the assumed extinction.

First of all, it can be noted that for experiment D1A0, where we assume the correct extinction during the analysis stage, we have a distribution of a, b and c virtually identical to the formal error distribution. This shows that σ_θ correctly describes the uncertainty in parallax and that there is no bias in this case. With an uncertainty arising only from σ_π, A0 represents the best-case scenario for zero-point determination using Gaia. D1A0 results in very well-determined P-L relation parameters, with $\sigma_a = 0.0026$ and $\sigma_b = 0.0021$. The parallax zero-point uncertainty is $\sigma_c = 0.159$ μas. The implications of this are discussed later on.

In experiment D1A1, we use the same data set but instead assume an extinction according to the exponential extinction model during the analysis stage. We know that the exponential extinction model differs too much from the AL05 model to be realistic, and this is indeed what we see in all three parameter fits. The exponential model is calibrated to the solar neighbourhood, but without the radial dependence, it systematically underestimates the extinction towards the Galactic Centre, and overestimates it away from the Galactic Centre. The result is both a huge bias and uncertainty for the parallax zero-point, with $\langle c \rangle = 223$ μas and $\sigma_c = 37$ μas. We also note that the parameters of the P-L relation (a and b) are almost undefined (the slope a even has the wrong sign in several realizations of the experiment). Estimating the extinction to the Gaia Cepheids using an extinction model (and specially a crude one such as this) is not a realistic method and will certainly not be useful for the verification of Gaia parallaxes. The exponential extinction model is of no interest to us, and will not be further considered in the discussions.

Experiment D1A2 assumes a Gaussian extinction error with a constant uncertainty

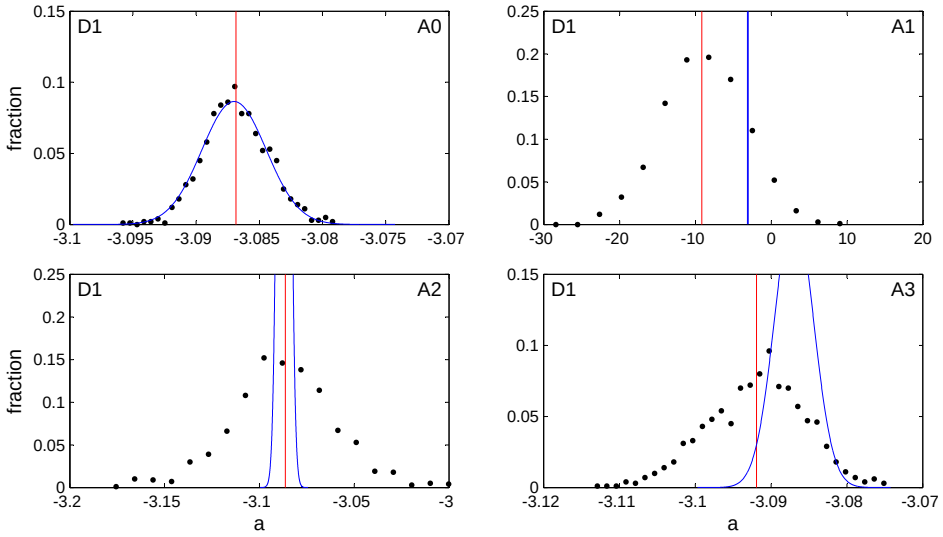


Figure 20: Distribution of a values (slope of P-L relation) obtained for four different experiments with 1000 realizations each (black dots). Also shown is the expected formal distribution (blue curve) and the observed mean value (red line).

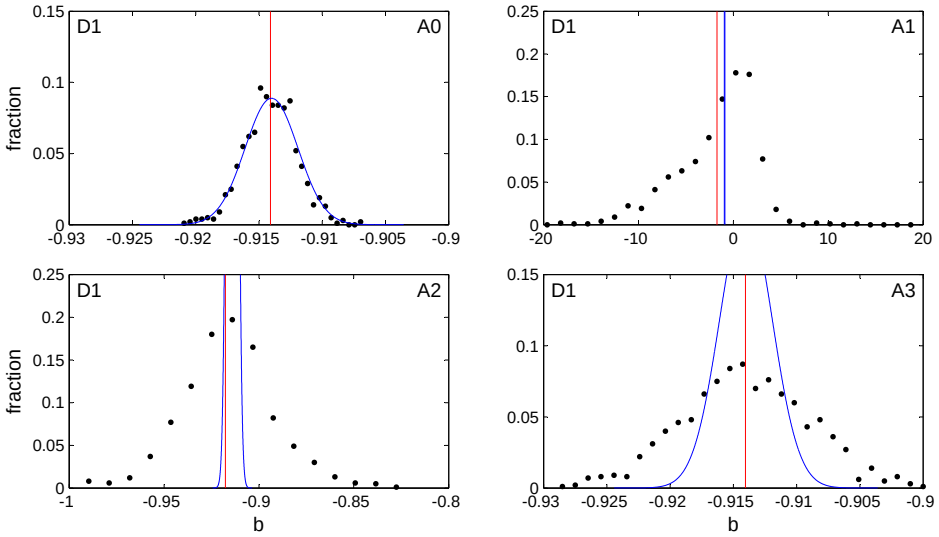


Figure 21: Distribution of b values (zero-point of P-L relation) obtained for four different experiments with 1000 realizations each (black dots). Also shown is the expected formal distribution (blue curve) and the observed mean value (red line).

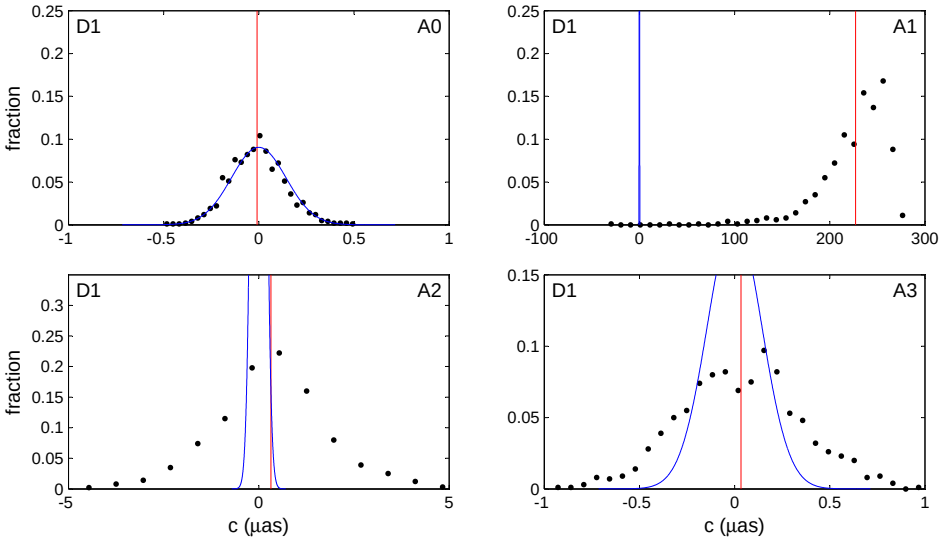


Figure 22: Distribution of c values (parallax zero-point) obtained for four different experiments with 1000 realizations each (black dots). Also shown is the expected formal distribution (blue curve) and the observed mean value (red line).

$\sigma_A = 0.05$ mag during the analysis. This assumption leads to slightly too pessimistic errors for low-extinction Cepheids (typically bright and nearby) but optimistic errors for the high-extinction Cepheids (typically faint and distant). In D1A3, we instead have an extinction uncertainty which scales with the total extinction as $\sigma_A = 0.05 A_V$. This assumption leads to optimisic errors for low-extinction Cepheids, and pessimistic errors for high extinction ones. A2 assumes larger errors than A3 for all Cepheids with $A_V < 1$, but smaller for $A_V > 1$. As an example, a Cepheid with an extinction of $A_V = 10$ will have a ten times larger error in A3 than in A2. If we look at the extinction distribution of the Cepheids observed with Gaia, we find only some hundred Cepheids with extinctions less than 1 magnitude compared to the roughly 9000 observed Cepheids in total. Surprisingly enough, A3 yields a scatter in the fitted c several times smaller than A2. The reason for this requires some investigation, and is discussed in Section 8.2.3.

We can note that both of the realistic extinction errors lead to large improvements in the accuracy with which the P-L relation parameters can be determined. Currently, the Galactic P-L relation is not as well-determined as the P-L relation of the LMC. There are large discrepancies between different attempts to calibrate the Galactic P-L relation (see Table 2b), both due to the small current number of Galactic Cepheids and the problems involved in estimating the extinction. Our simulations show that using Gaia, a and b are much less sensitive to errors in the extinction than c. We estimate that Gaia will be able to determine both constants with an accuracy of $\sigma_{a,b} = 0.01$–0.05.

8.2.2 Other experiments

The inability to measure the extinction accurately enough is a limiting factor in the normal least-squares fitting method. In the extinction-free method used in analysis methods A8 and A9, the accuracy of fit depends on the parallax measurements, the small uncertainty with which we measure $V - I$ and the possible error in R_{V-I}. Looking at data set D1, however, we can see that the accuracy has not improved for set A8. From experiments D1A2 and D1A3 with $\sigma_{c,\mathrm{s}} = 0.305$ and 1.45 μas, the sample uncertainty is $\sigma_{c,\mathrm{s}} = 0.721$ μas for D1A8. This is surprisingly large, with the reason related to the same effect that caused the errors for method A2 (compare Eqs.(7.3) and (7.11)). In set A9, we note a significant sensitivity towards biasing. The 5% bias in R_{V-I} means that we consistently overestimate the colour excess of the observed stars, leading to a very large parallax zero-point bias with $\langle c \rangle = -11.17$ μas. The extinction-free method is however more promising than it looks from these results, as we will show in the following section when limiting the sample.

Data sets D3 and D4 are two of the four data sets in which we have an inherent metallicity dependence in the P-L relation. In set D4, we have a metallicity gradient in the Galaxy, whereas we in set D3 have not. In A0 to A4, we try to fit a non-metallicity dependent solution to these Cepheids, which has a very small effect on the uncertainties of a and b, but causes σ_c to increase to ~ 1 μas even with perfect knowledge of the extinction. When the metallicity is measured and accounted for during the parameter fitting, $\sigma_{c,\mathrm{s}}$ improves by a factor of two for analysis sets A5 and A7. Because of the measurement errors in metallicity, however, we still have a relatively large uncertainty of $\sigma_{c,\mathrm{s}} = 0.63$ for experiment D4A5. If the Cepheid metallicity dependence is true, it will severely limit the capability of using Cepheids for the verification of the Gaia parallaxes unless the metallicity measurements are improved.

In experiment D1A4, we have excluded the inner kpc in the Galaxy. As we can see in Table 8, this removes approximately 200 Cepheids from the sample, and leads to minimal changes in the parameter fits. By comparing the results from data sets D1 and D5, we can also study the effect of using our modelled period distribution versus the observed one. By not taking into account the bias towards brighter Cepheids, the expected number of observed Cepheids would have been somewhat larger, with 9491 Cepheids for D5 instead of 9283 for the more realistic D1 set. This, along with the larger number of bright stars in the sample, would have lead to slightly too optimistic predictions for the zero-point accuracy.

8.2.3 Limiting the sample

In the previous sections, we have noted surprisingly large errors when we assumed a constant extinction error (as in analysis set A2), as well as in our extinction-free method (as in analysis set A8). This has large implications for the quality of our method, and it is therefore important for us to understand the reason behind it and

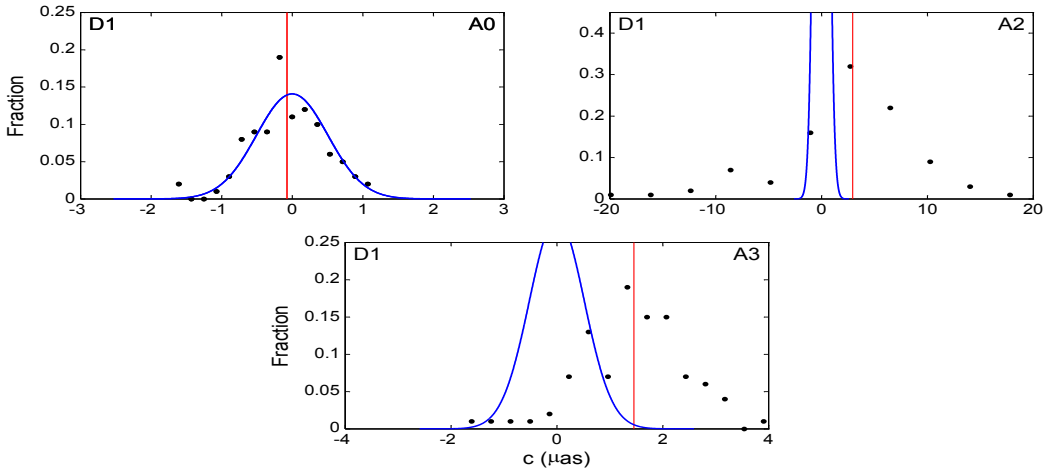


Figure 23: Distribution of c values (parallax zero-point) obtained using Cepheids with $A_{V,\mathrm{obs}} < 1$ after 100 realizations each of three different analysis methods (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

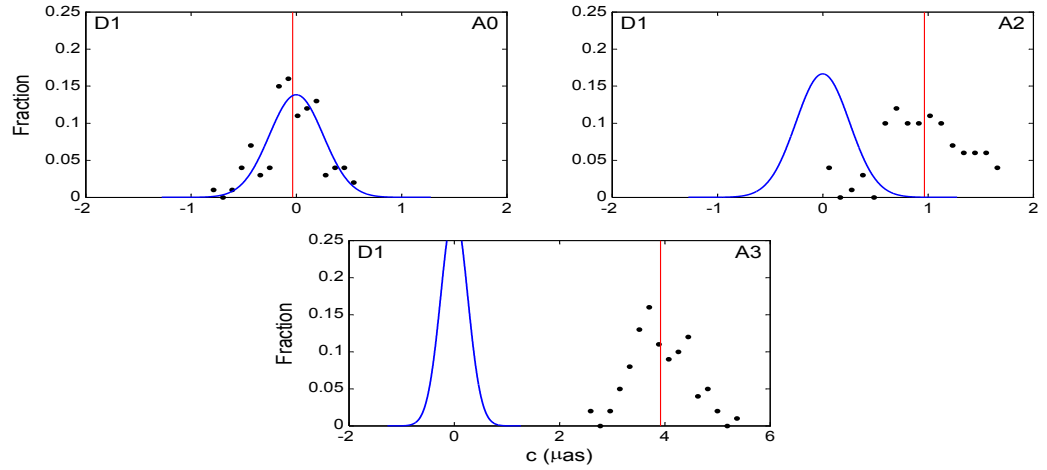


Figure 24: Distribution of c values (parallax zero-point) obtained using Cepheids with $A_{V,\mathrm{obs}} > 1$ after 100 realizations each of three different analysis methods (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

if possible improve the methods.

In the analysis methods A2 and A3, we can identify a break point in the extinction at $A_V = 1$, where the two methods have equal errors. We therefore divide the Cepheid sample in two and analyze $A_V < 1$, where we expect A3 to be favoured, and $A_V > 1$, where A2 should be favoured. The selection is made based on the measured extinction, $A_{V,\mathrm{obs}}$, and the results are given in Figs. 23 and 24. In these plots, we note two major features. First of all, we see that A3 is favoured for $A_{V,\mathrm{obs}} < 1$. However, contrary to our prediction, the two methods actually result in equally large uncertainties in c for $A_{V,\mathrm{obs}} > 1$. Secondly, we can note a significant bias in c for both methods and samples. For $A_{V,\mathrm{obs}} < 1$, the bias is roughly 2 μas for both methods, but for $A_{V,\mathrm{obs}} > 1$, it has decreased to 1 μas for A2 and increased to 4 μas for A3.

The bias might arise from the way we divide the Cepheids by means of the assumed extinction, which we know has a Gaussian distribution with a certain σ_A around the true A_V. Depending on the distribution of the true extinction values, the number of Cepheids that scatter into and out of the selected sample could be different, leading to a systematic difference between the true and assumed extinction in the sample. This is similar to the Lutz-Kelker effect known to cause bias effects during the selection of stars by a lower parallax limit (Lutz & Kelker 1973). To test this, we have used the true extinction to divide up our sample in Figs. 25 and 26. These plots show an improvement in the case with $A_{V,\mathrm{true}} < 1$, but only slight improvements for $A_{V,\mathrm{true}} > 1$. There are obviously additional biasing effects in action here.

Returning to our first point, where we noted in Fig. 24 the surprisingly good fit for method A3 with $A_V > 1$. With the very large errors associated for large extinctions with this method, this implies that the large extinction stars in general are much less important than the low extinction Cepheids. In Fig. 27, we plot the measured parallax uncertainty versus A_V for all observed Cepheids in data set D1. We note that half of all observed stars have extinctions less than 5 magnitudes, and that the mean parallax uncertainty in this sample is $\sigma_\pi = 10$ μas, a factor of 4 less than the mean parallax uncertainty of the total sample. For extinctions less than 1 magnitude, the mean uncertainty is even lower, with $\sigma_\pi = 6$ μas. During our parameter fitting, the weighting of each Cepheid depends on its σ_π. With the correlation between extinction and parallax accuracy, it means that the low-extinction Cepheids in general are more important than the high-extinction Cepheids for our parameter fitting. In order to test this, we give all Cepheids the same weight by artifically assigning all Cepheids the same parallax uncertainty of $\sigma_\pi = 10$ μas. The results confirm our theory, and the distribution of c fits finally become what we originally expected, with A2 being the favoured analysis method for $A_V > 1$.

Interestingly enough, we can with the low-extinction Cepheids removed (see Figs. 24 and 26) note a huge improvement in the sample standard deviation $\sigma_{c,\mathrm{s}}$ for analysis set A2, and the same is also true for set A8. This is not only due to the large weight-

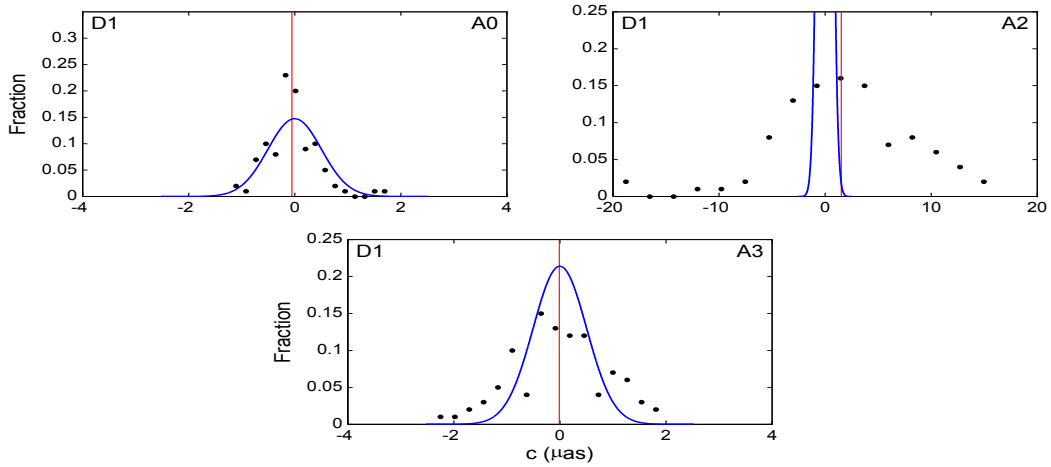


Figure 25: Distribution of c values (parallax zero-point) obtained using Cepheids with $A_{V,\mathrm{true}} < 1$ after 100 realizations each of three different analysis methods (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

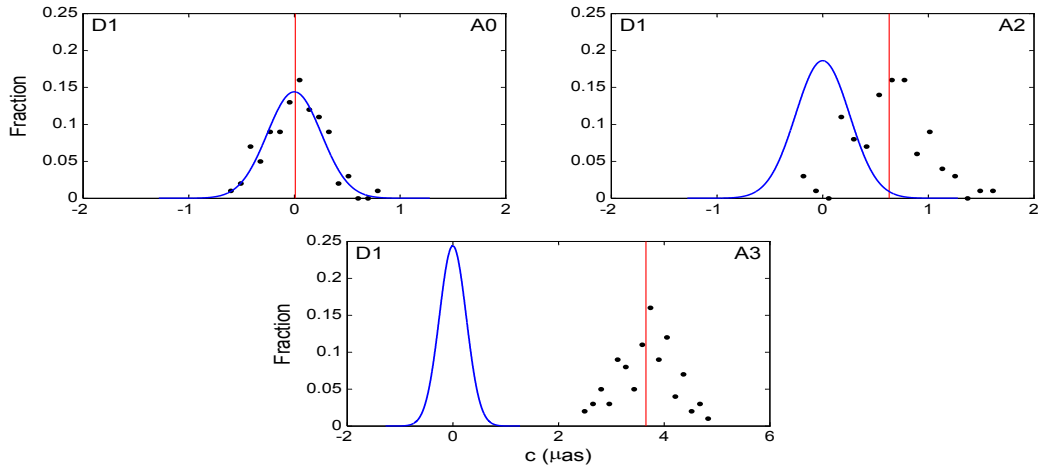


Figure 26: Distribution of c values (parallax zero-point) obtained using Cepheids with $A_{V,\mathrm{true}} > 1$ after 100 realizations each of three different analysis methods (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

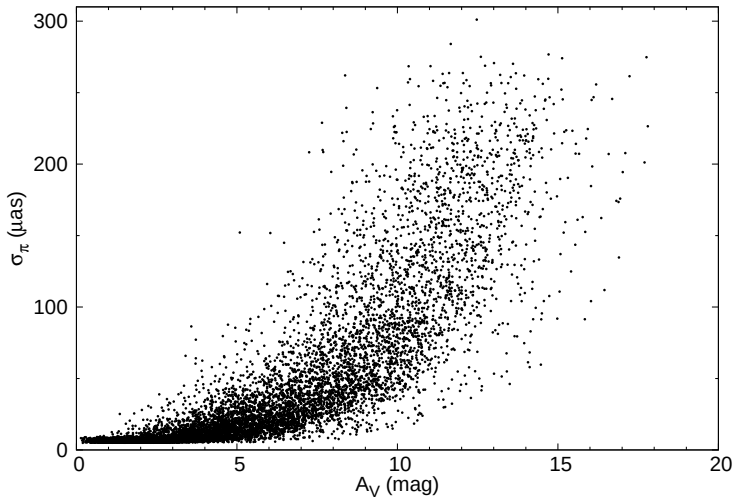


Figure 27: Uncertainty in parallax versus the true extinction for all observed stars in data set D1.

ing for the low σ_c Cepheids, which we realize from the drastic improvements by only removing the 10 or 20 closest Cepheids from the full sample. In reality, the Gaia satellite will not be able to observe stars brighter than $G = 6$, due to saturation of the CCDs. This limit would exclude only the brightest (and also closest) 30 stars, but has not been included in the simulations as it caused complex biasing effects that should be avoidable given more time to adjust the least-squares fitting method. In Table 6, we show a comparison between the sample standard deviation of the full sample and for $G > 6$. Here, we see the same tendency as we saw for the sample with $A_V > 1$. For $G > 6$, the constant and the scalable extinction errors (sets A2 and A3) show a much smaller discrepancy. We can here see the capability of the Cepheid verification method to match the method using quasars with $\sigma_{c,\mathrm{s}} \approx 0.4$ μas for the normal fitting method. The extinction-free method (sets A8 and A9) show similar improvements as method A2 and results in $\sigma_{c,\mathrm{s}} \approx 0.3$ μas, making the extinction-free method the most promising as expected.

It is obvious that the closest Cepheids are incredibly important in our current least-squares fitting method. This means that errors in the measurements of these stars have disproportionally large effects on the end fitting results, and it is also the source of unwanted biasing in c. Due to time limitations, we have not been able to develop and test any alternative weighting methods. With the improvements seen for the sample with $G > 6$, however, it is clear that such a method would be worth pursuing.

Even if the large extinction Cepheids seem to be weighted very low because of their large parallax uncertainties, we investigate whether method A3 can be improved by limiting the sample to some realistic extinction limit. Roughly half of the observed

Table 6: Resulting sample standard deviations using data set D1 and analysis sets A0–A9 for all observed Cepheids and for $G > 6$. The magnitude limited sample excludes only 30 Cepheids.

	Complete $\sigma_{\mathrm{c,s}}$ (μas)	$G > 6$ $\sigma_{\mathrm{c,s}}$ (μas)
A0	0.159	0.177
A2	1.45	0.459
A3	0.305	0.384
A4	0.309	0.370
A5	0.160	0.208
A6	1.37	0.464
A7	0.312	0.385
A8	0.721	0.276
A9	0.808	0.332

sample is included within $A_V < 5$, so this seems like a reasonable limit. The largest extinction will in this case lead to uncertainties of $\sigma_A = 0.25$ magnitudes. The sample standard deviation has for method A3 decreased from $\sigma_{c,\mathrm{s}} = 0.305$ μas for the full sample to $\sigma_{c,\mathrm{s}} = 0.301$ μas, a small change that can be explained by statistical scatter. In order to determine if the zero-point uncertainty is also dependent on the uncertainty in the extinction limit, we redo the same experiment but for $A_{\mathrm{true}} < 5$. This gives us a marginal improvement to $\sigma_{c\mathrm{s}} = 0.290$ μas.

9 Discussion and conclusions

We have formulated several models of the Galactic Cepheid population together with models of the Galactic extinction and Gaia observations. These have been implemented in CepheidObsModel, a Java program capable of simulating Gaia observations of the Galactic Cepheids. We have used this code to investigate whether the P-L relation can be used for verification of the Gaia parallax zero-point. This has been done by making a simultaneous fit to the P-L relation and the parallax zero-point.

Our results indicate that the determination of the extinction is the biggest issue for the Cepheid method. With perfect knowledge of the extinction, the parallax zero-point can in principle be determined with an accuracy of $\sigma_c = 0.16$ μas. Even in this perfect scenario, the uncertainty is slightly larger than the < 0.1 μas needed in the most demanding tasks of Gaia. With random extinction errors included, the zero-point uncertainty rises to $\sigma_c = 0.3$–1.4 μas depending on the type of error. By removing the brightest Cepheids, the uncertainty for the same two error types changes to $\sigma_c = 0.38$–0.46 μas. We have investigated whether the method can be improved by limiting the sample to a maximum extinction, but have concluded

that this improves the results only slightly because of the importance of the closest Cepheids that all have low extinction. Moreover, selecting the sample based on the measured extinction could very well introduce additional biases.

A promising parameter fitting method is the extinction-free method, where the accuracy depends on the ability to determine R_{V-I} and accurately measure the $V-I$ colour. R_{V-I} is a property much better known than extinction to individual stars, and a similar method is commonly used during P-L calibrations of Galactic Cepheids. It leads to zero-point accuracies on the order of $\sigma_{c,\mathrm{s}} = 0.3$ μas, which is not much larger than in the case with perfect knowledge of extinction. We have however also shown that small errors in R_{V-I} can lead to large biases in c.

The majority of the experiments have used all Cepheids below $G < 20$ and have, due to time limitations, not taken Gaias brightest limit of $G > 6$ into account. This limit would only exclude about 30 Cepheids, but will because of complex effects result in bias effects for some scenarios with measurement errors in extinction, colour or metallicity. Solving this issue will require additional work, but we believe that this can lead to improvements in the method making it slightly more accurate than using quasars for the zero-point verification.

As a side effect to our main goal, we have created several possible Cepheid distribution models. We have shown that the advent of Gaia will lead to an increase in the number of observed Cepheids with a factor of fifteen, to roughly 9000 Cepheids. We have also shown that by the use of Gaia alone it will be possible to very accurately determine the P-L relation. The P-L relation constants have been shown to be much less sensitive to errors in the assumed extinction than the parallax zero-point.

The program written in this work allows for other Galactic studies than ours. Due to the modular structure in our Java program, it is relatively simple to implement other objects than Cepheids in the code. One example of this that was originally planned to be investigated during this thesis was the method of parallax zero-point verification using quasars. It would be possible to create a uniform distribution of quasars over the sky with a reasonable magnitude distribution, and then study how the Galactic extinction affects the number of observed quasars. The quasars could then be observed using our Gaia model in order to make a more detailed study of the magnitude distribution and the effect that it has on the verification possibilities of quasars.

In conclusion, we have shown that zero-point verification of the Gaia parallaxes using Galactic Cepheids might not reach the required accuracy of the most demanding tasks. It is however still a promising method with potentially the same order of accuracy as using quasars. The global verification of the parallax zero-point will ultimately depend on the combination of many different methods.

References

Amôres, E. B. & Lépine, J. R. D. 2005, AJ, 130, 659

Arenou, F., Lindegren, L., Froeschle, M., et al. 1995, A&A, 304, 52

Berdnikov, L. N., Dambis, A. K., & Vozyakova, O. V. 2000, A&AS, 143, 211

Binney, J. & Merrifield, M. 1998, Galactic astronomy (Princeton University Press)

Cox, A. N. 2000, Allen's astrophysical quantities, ed. Cox, A. N.

de Bruijne, J., Siddiqui, H., Lammers, U., et al. 2010, in IAU Symposium, ed. S. A. Klioner, P. K. Seidelmann, & M. H. Soffel, Vol. 261, 331–333

Eisenhauer, F., Schödel, R., Genzel, R., et al. 2003, ApJ, 597, L121

Feast, M. W. & Catchpole, R. M. 1997, MNRAS, 286, L1

Fernie, J. D., Evans, N. R., Beattie, B., & Seager, S. 1995, IBVS, 4148, 1

Fouqué, P., Arriagada, P., Storm, J., et al. 2007, A&A, 476, 73

Freudenreich, H. T. 1998, ApJ, 492, 495

Gieren, W. P., Fouque, P., & Gomez, M. 1998, ApJ, 496, 17

Groenewegen, M. A. T. 2008, A&A, 488, 25

Guibert, J., Lequeux, J., & Viallefond, F. 1978, A&A, 68, 1

Hobbs, D., Lindegren, L., Holl, B., Lammers, U., & O'Mullane, W. 2008, in IAU Symposium, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, Vol. 248, 268–269

Holmberg, J., Lindegren, L., & Flynn, C. 1998, Highlights of Astronomy, 11, 570

Jordi, C., Høg, E., Brown, A. G. A., et al. 2006, MNRAS, 367, 290

Leavitt, H. S. 1908, Annals of Harvard College Observatory, 60, 87

Leavitt, H. S. & Pickering, E. C. 1912, Harvard College Observatory Circular, 173, 1

Lemasle, B., François, P., Piersimoni, A., et al. 2008, A&A, 490, 613

Lindegren, L. 2004, GAIA-LL-057, Gaia Livelink technical note (restricted access)

Lindegren, L. 2010, in IAU Symposium, ed. S. A. Klioner, P. K. Seidelmann, & M. H. Soffel, Vol. 261, 296–305

Lindegren, L., Babusiaux, C., Bailer-Jones, C., et al. 2008, in IAU Symposium, ed. W. J. Jin, I. Platais, & M. A. C. Perryman, Vol. 248, 217–223

Lindegren, L. & Perryman, M. A. C. 1996, A&AS, 116, 579

Lindegren, L., Perryman, M. A. C., Bastian, U., et al. 1993, Gaia - Proposal for an ESA Cornerstone Mission

Lutz, T. E. & Kelker, D. H. 1973, PASP, 85, 573

Madore, B. F. & Freedman, W. L. 1991, PASP, 103, 933

Majaess, D. J., Turner, D. G., & Lane, D. J. 2009, MNRAS, 398, 263

Marshall, D. J., Robin, A. C., Reylé, C., Schultheis, M., & Picaud, S. 2006, A&A, 453, 635

McKee, C. F. & Williams, J. P. 1997, ApJ, 476, 144

Ngeow, C., Kanbur, S. M., Neilson, H. R., Nanthakumar, A., & Buonaccorsi, J. 2009, ApJ, 693, 691

Perryman, M. A. C., de Boer, K. S., Gilmore, G., et al. 2001, A&A, 369, 339

Perryman, M. A. C., Lindegren, L., Kovalevsky, J., et al. 1997, A&A, 323, L49

Reed, B. C. 2006, JRASC, 100, 146

Reid, M. J. 1993, ARA&A, 31, 345

Salpeter, E. E. 1955, ApJ, 121, 161

Sandage, A. & Tammann, G. A. 2006, ARA&A, 44, 93

Sandage, A., Tammann, G. A., & Reindl, B. 2004, A&A, 424, 43

Tammann, G. A., Sandage, A., & Reindl, B. 2003, A&A, 404, 423

Udalski, A., Szymanski, M., Kubiak, M., et al. 1999, Acta Astronomica, 49, 201

van der Kruit, P. C. 1988, A&A, 192, 117

Williams, J. P. & McKee, C. F. 1997, ApJ, 476, 166

A Table of notations

Table 7: A list of the various notations used in this work.

Notation	Description
a_{V}	extinction per kpc in the V band
A_{V}	total extinction in the V band
A, B, C	constants to describe the extinction distribution
a	slope of the P-L relation
b	zero-point of the P-L relation / galactic latitude
c	parallax zero-point
CDF	the cumulative density function
d	slope of the P-C relation
e	zero-point of the P-C relation
$E_{\mathrm{V-I}}$	colour excess in $V-I$
G	apparent magnitude in the Gaia G band
h	the extinction scale-height
k_1, k_2	fitting parameters for the extinction-free metod
l	galactic longitude
m	metallicity dependence in the P-L relation
M_{V}	absolute visual magnitude
N	number of Cepheids
P	period of the Cepheid
r	distance from the Sun
R	galactocentric radius
$R_{\odot}$	distance from Sun to the Galactic Center
$R_{\mathrm{V-I}}$	extinction coefficient in $V-I$ colour
V	apparent visual magnitude
$(V-I)_0$	intrinsic $V-I$ colour
$V-I$	apparent $V-I$ colour
x, y	galactocentric coordinates
z	vertical distance to the Galactic plane
z_0	the Cepheid scale-height
$z_{\odot}$	Solar height above the Galactic plane
α, β, δ	constants describing the Cepheid radial distribution
π	parallax
π_{G}	the parallax measured by Gaia
σ	the uncertainty or scatter of a parameter
$\sigma_{c,\mathrm{s}}$	sample standard deviation in c
Σ	the column density within a distance $r\cos b$
τ	lifetime of a Cepheid

B The CepheidObsModel program

CepheidObsModel is the program written in Java (using Eclipse) in order to use and tie together all the theoretical models obtained in Sections 2–7. It can roughly be divided into four parts; generating the Galactic Cepheid population from the distribution functions discussed in Section 5; determining what the Cepheids look like from the Sun as discussed in Section 5.3; simulating the observations of the Cepheids using the Gaia observational model from Section 6; making the statistical investigation of the observable sample using the method in Section 7. These four parts are discussed in more detail in Sections B.1-B.4.

In Fig. 28, we see a UML (Unified Modelling Language) class diagram showing a sketch of the main Java classes included in the program. We can here note the modular structure, where we can easily create a new class representing for example a new type of observable object such as quasars and add them to the observations. It would also be possible to include observational models of other Galactic survey missions than Gaia if necessary. Displayed in the class boxes is a selection of the methods available publically from outside the class. We can for example see that for every individual Cepheid object we obtain its Galactic x and y coordinates from the getGalX() and getGalY() methods.

In Fig. 29, we include a more detailed class diagram, generated automatically by the eUML2 Eclipse add-on. This diagrams also includes the utility classes responsible for things such as the reading input data, printing to external data files, making the binary search through the data and plotting the Java generated plots. This information is then wrapped up in the CepheidObsModel main class, which accesses and initiates the necessary classes. CepheidObsModel is initiated by either the SingleModelWrapper or MultiModelWrapper classes, capable of making one or several realizations of the galaxy simulations. We can in the figure also see the relation between the classes; which class initiates which class and so on.

Finally, we see a so-called sequence diagram in Fig. 30, displaying the order of which the methods are called and initiated, starting with creating the Galaxy and ending with doing the statistics on the observed Cepheids. The boxes do not directly correspond to individual classes or methods given in Figs. 28 or 29, but rather the principle idea behind the program design. Below the names in the boxes to the right, we see the new Cepheid properties determined by that particular step. The different steps are discussed in more detail below, but this can be referred to for clarity.

B.1 The Galaxy

The Galaxy class is responsible for initiating the creation of the Galactic Cepheid population. The Galaxy class is initiated with the total number of Galactic Cepheids to generate and what P-L relation and radial and period distribution model to use.

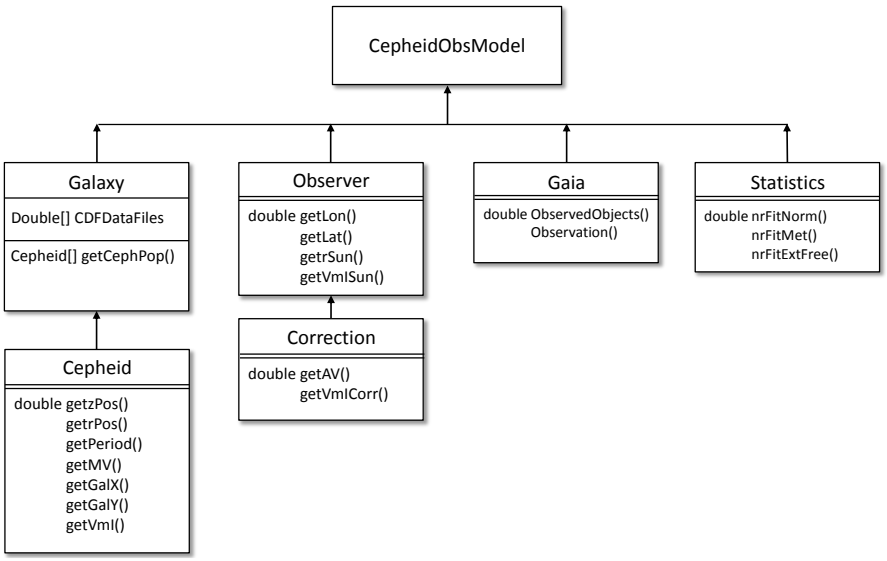


Figure 28: UML class diagram on the principle program design. Each box represents a class, where name and a selection of private and public methods are displayed.

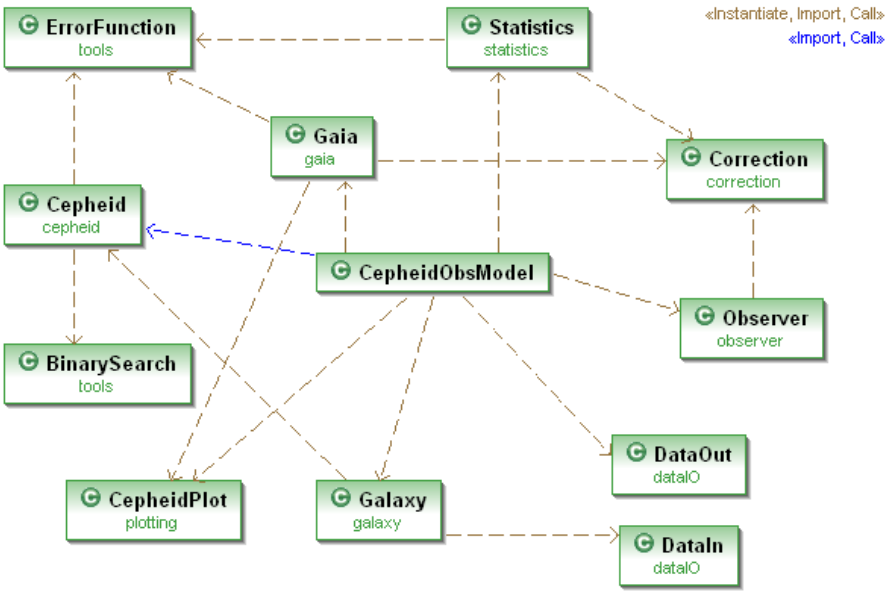


Figure 29: UML class diagram autogenerated from the code by eUML2. The blue line between the main and the Cepheid class comes from the fact that they do not interact directly, since the Galaxy is responsible for creating the individual Cepheids.

It then reads off one of the external pregenerated CDF tables for the radial and period distribution and passes this information down to each Cepheid it creates. The different distributions are all discussed in Section 5.

Each Cepheid is represented as one Java object. In each object, we randomly generate the galactocentric distance, R, after the given radial distribution, and from this the Galactic x and y coordinates. We also generate the vertical distance to the plane, z, the period P from the given period distribution, and from the the absolute magnitude M_{V} after the given P-L relation. From the period, we also give each Cepheid a certain $V - I$ colour.

The information about each individual Cepheid is then passed down to the CepheidObsModel main class.

B.2 The Observer

In the Observer class, we move the reference frame from the Galactic Center to the reference frame of the observer which looks at the Cepheids. In this work, we put the observer at the position of the Sun, 8 kpc from the Galactic Center, but it would also be possible to see what the sky would look like from an arbitrary point, for example somewhere in the bulge, the Galactic halo or the Andromeda galaxy.

Here, we determine what the Cepheids look like from our point of view, and determine the galactic coordinates l, b and r, for every Cepheid, discussed in Section 5.1. The Observer also does not observe the absolute magnitude or colour of an object, but instead sees the apparent magnitude and colour as determined by the Correction class.

The Correction class determines the faintening of the light caused by both distance and extinction from dust, which in turn scatter blue light more than red and therefore causes the reddening of the $V - I$ colour as discussed in Section 5.3. The class can determine the extinction between any two points in the Galaxy as determined by the extinction model. In the Amôres & Lépine (2005) model, the extinction is described by such a complicated model that the integration between the two points can not be done easily. Instead, this is done numerically as described in Amôres & Lépine (2005), by linear interpolation along the line of sight in steps of a certain number of pc. Different step lengths were tested, and we found 50 pc to give the best balance between accuracy and speed.

B.3 Gaia

The Gaia class simulates the full five-year mission observations at the point of the observer, and includes the information obtained from Section 6. This class relies on the GaiaTools and AgisLab libraries developed by the Gaia Data Processing and

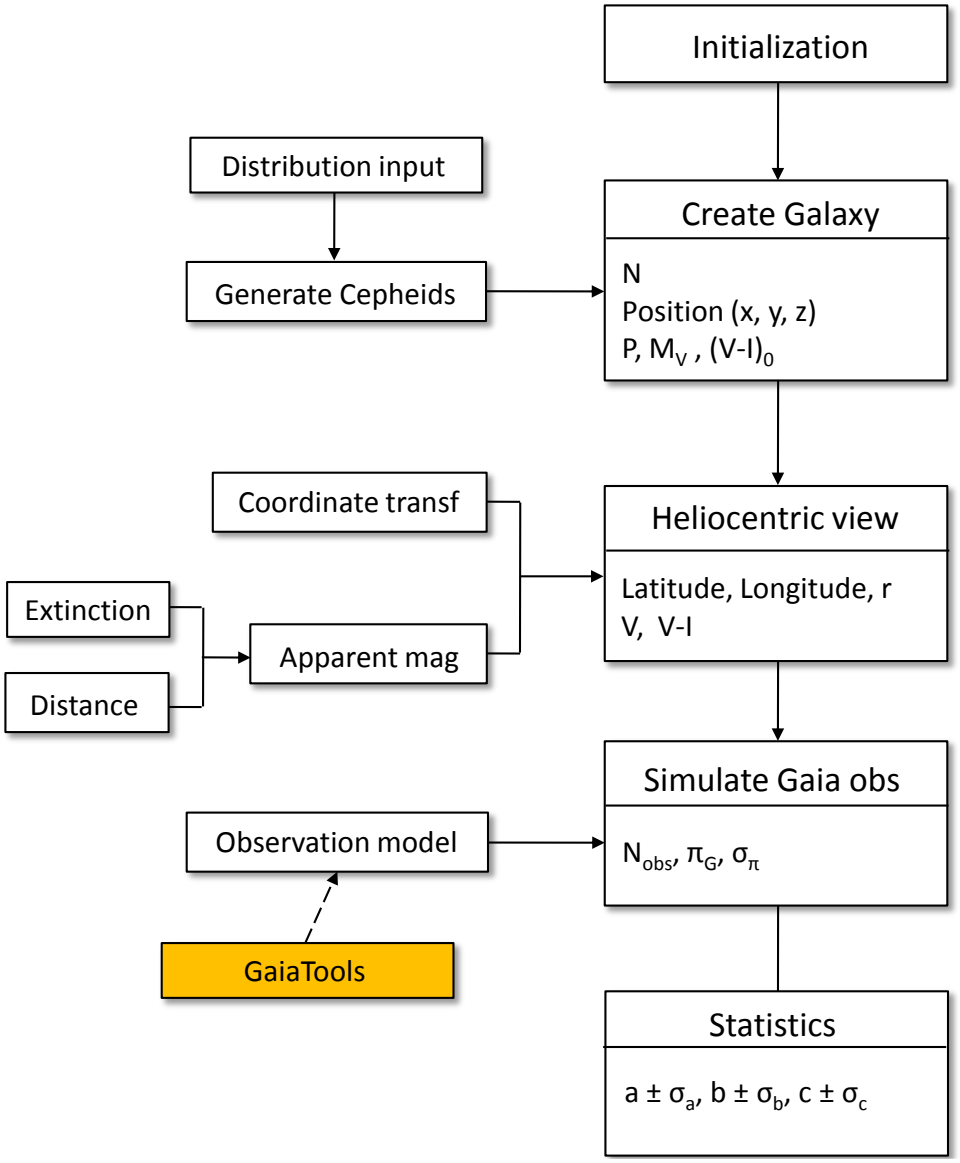


Figure 30: Flow diagram of the principle program design, showing a sketch of how the data flows from the initialization to the end step.

Analysis Consortium (DPAC). Specifically, it uses the StandardErrors program written by Berry Holl.

The class determines the G magnitude of each object from the V magnitude and $V - I$ colour. Strictly, by our methodology, this is something that would be done in the Observer class, but we here use G magnitude converter by GaiaTools that we would prefer to have in the Gaia class. We then determine which Cepheids that will be observed by inserting the $G = 20$ limiting magnitude. All objects fainter than this are ignored from here on. Depending on the position and brightness of an object, the Gaia class then uses the by GaiaTools simulated number of field transits for the object and calculates the objects estimated parallax standard error. It then assumes that the measurement errors are Gaussian and symmetric around the expected parallax (equal to the true parallax which the satellite does not know about). Using the Box-Muller approximation and the known parallax standard error, it then simulates the parallax measurement.

B.4 Statistics

Finally, we have the Statistics class, which is based on the statistics methods in Section 7. In this class, we use the parallax measurements as determined by the Gaia class, along with the observed apparent V magnitude and period to determine the parallax zero-point value and standard deviation as well as the P-L relation constants. We also need an assumed A_V value, which is generated depending on the error type specified in CepheidObsModel. The same thing is done for the observed $V - I$ colour and measured metallicity. The parameters are fitted by means of least-squares fitting with the Newton-Raphson method to the free variables in Eqs. (7.3), (7.11) or (7.13).

Table 8: Results from all experiments run with the CepheidObsModel using an intrinsic Cepheid P-L relation following STR04. Each value given is the mean value of 100 runs of each experiment.

Data	Anlys	N_{obs}	$a \pm \sigma_a$	$b \pm \sigma_b$	$m \pm \sigma_m$	$k_1 \pm \sigma_{k_1}$	$k_2 \pm \sigma_{k_2}$	$c \pm \sigma_c$	$\sigma_{c,s}$	true
True			-3.087	-0.914	0	-3.7056	-2.1149	0		
D0	A0	9019	-3.08697 ± 0.00277	-0.91407 ± 0.00227	–	–	–	-0.009 ± 0.160	0.170	1.00
D0	A1	9019	-4.31491 ± 0.00374	-1.07401 ± 0.00296	–	–	–	62.198 ± 0.160	35.493	0.00
D0	A2	9019	-3.08548 ± 0.00277	-0.91615 ± 0.00227	–	–	–	0.037 ± 0.160	1.714	0.28
D0	A3	9019	-3.09386 ± 0.00277	-0.91352 ± 0.00227	–	–	–	0.089 ± 0.160	0.364	0.80
D0	A4	8778	-3.09435 ± 0.00277	-0.91294 ± 0.00227	–	–	–	0.082 ± 0.160	0.360	0.86
D1	A0	9293	-3.08660 ± 0.00249	-0.91430 ± 0.00207	–	–	–	-0.004 ± 0.143	0.159	1.00
D1	A1	9293	-8.70878 ± 0.01163	-1.78811 ± 0.00821	–	–	–	224.913 ± 0.118	30.087	0.00
D1	A2	9293	-3.09104 ± 0.00250	-0.91611 ± 0.00208	–	–	–	0.611 ± 0.143	1.446	0.25
D1	A3	9293	-3.09191 ± 0.00250	-0.91405 ± 0.00207	–	–	–	0.084 ± 0.143	0.305	0.91
D1	A4	9171	-3.09113 ± 0.00250	-0.91496 ± 0.00207	–	–	–	0.062 ± 0.143	0.309	0.87
D5	A0	9500	-3.08716 ± 0.00196	-0.91381 ± 0.00181	–	–	–	-0.012 ± 0.138	0.131	1.00
D5	A1	9500	-9.11314 ± 0.01092	-1.50056 ± 0.00784	–	–	–	222.850 ± 0.114	25.455	0.00
D5	A2	9500	-3.08552 ± 0.00196	-0.91672 ± 0.00181	–	–	–	0.105 ± 0.138	1.553	0.27
D5	A3	9500	-3.09175 ± 0.00196	-0.91415 ± 0.00181	–	–	–	0.026 ± 0.138	0.298	0.90
D5	A4	9366	-3.09113 ± 0.00196	-0.91467 ± 0.00181	–	–	–	0.048 ± 0.138	0.272	0.95
D6	A0	14907	-3.08713 ± 0.00245	-0.91388 ± 0.00205	–	–	–	0.007 ± 0.147	0.156	1.00
D6	A1	14907	-3.14415 ± 0.00282	-1.30243 ± 0.00233	–	–	–	48.352 ± 0.137	10.818	0.00
D6	A2	14907	-3.08710 ± 0.00246	-0.91763 ± 0.00205	–	–	–	0.412 ± 0.146	1.819	0.28
D6	A3	14907	-3.09844 ± 0.00246	-0.91490 ± 0.00205	–	–	–	0.580 ± 0.147	0.465	0.31
D6	A4	14907	-3.09844 ± 0.00246	-0.91490 ± 0.00205	–	–	–	0.580 ± 0.147	0.465	0.31
D0	A5	9035	-3.08711 ± 0.00271	-0.91388 ± 0.00223	-0.00021 ± 0.00423	–	–	-0.002 ± 0.161	0.167	1.00
D0	A6	9035	-3.08925 ± 0.00270	-0.91518 ± 0.00223	-0.00660 ± 0.00432	–	–	0.321 ± 0.161	1.874	0.26
D0	A7	9035	-3.09294 ± 0.00270	-0.91451 ± 0.00223	-0.00018 ± 0.00432	–	–	0.140 ± 0.161	0.424	0.72
D1	A5	9293	-3.08662 ± 0.00252	-0.91427 ± 0.00209	0.00023 ± 0.00398	–	–	-0.007 ± 0.144	0.160	1.00
D1	A6	9293	-3.09083 ± 0.00252	-0.91583 ± 0.00210	0.00266 ± 0.00403	–	–	0.555 ± 0.144	1.367	0.24
D1	A7	9293	-3.09202 ± 0.00252	-0.91406 ± 0.00210	0.00012 ± 0.00403	–	–	0.097 ± 0.144	0.312	0.87
D5	A5	9500	-3.08715 ± 0.00197	-0.91383 ± 0.00183	0.00039 ± 0.00395	–	–	-0.009 ± 0.139	0.131	1.00
D5	A6	9500	-3.08512 ± 0.00198	-0.91692 ± 0.00183	-0.00559 ± 0.00400	–	–	0.086 ± 0.139	1.480	0.31
D5	A7	9500	-3.09173 ± 0.00198	-0.91430 ± 0.00183	-0.00000 ± 0.00400	–	–	0.041 ± 0.139	0.297	0.92
D6	A5	14907	-3.08713 ± 0.00247	-0.91389 ± 0.00206	0.00053 ± 0.00395	–	–	0.008 ± 0.147	0.160	1.00
D6	A6	14907	-3.08632 ± 0.00247	-0.91814 ± 0.00206	0.00222 ± 0.00389	–	–	0.395 ± 0.148	1.715	0.27
D6	A7	14907	-3.09852 ± 0.00247	-0.91513 ± 0.00207	0.00150 ± 0.00389	–	–	0.618 ± 0.148	0.455	0.29
D0	A8	9035	–	–	–	-3.70656 ± 0.00268	-2.11498 ± 0.00221	0.109 ± 0.160	0.961	0.47
D0	A9	9035	–	–	–	-3.74928 ± 0.00261	-2.16618 ± 0.00216	-14.681 ± 0.163	1.076	0.00
D1	A8	9293	–	–	–	-3.70720 ± 0.00250	-2.11551 ± 0.00207	0.235 ± 0.143	0.721	0.51
D1	A9	9293	–	–	–	-3.74618 ± 0.00244	-2.17271 ± 0.00203	-12.583 ± 0.146	0.808	0.00
D5	A8	9500	–	–	–	-3.70485 ± 0.00196	-2.11553 ± 0.00181	-0.011 ± 0.138	0.747	0.55
D5	A9	9500	–	–	–	-3.74395 ± 0.00191	-2.17222 ± 0.00177	-12.961 ± 0.141	0.883	0.00
D6	A8	14907	–	–	–	-3.70554 ± 0.00246	-2.11593 ± 0.00205	0.127 ± 0.146	0.883	0.56
D6	A9	14907	–	–	–	-3.75639 ± 0.00236	-2.15561 ± 0.00198	-18.721 ± 0.152	1.002	0.00

Table 9: Results from all experiments run with the CepheidObsModel using an intrinsic Cepheid P-L relation following G08. Each value given is the mean value of 100 runs of each experiment.

Data	Anlys	N_{obs}	$a \pm \sigma_a$	$b \pm \sigma_b$	$m \pm \sigma_m$	$k_1 \pm \sigma_{k_1}$	$k_2 \pm \sigma_{k_2}$	$c \pm \sigma_c$	$\sigma_{c,s}$	true
True			-2.60	-1.30	0	-3.3286	-2.5009	0		
D2	A0	9297	-2.60010 ± 0.00249	-1.29997 ± 0.00208	–	–	–	-0.005 ± 0.143	0.141	1.00
D2	A1	9297	-7.64722 ± 0.01162	-2.64054 ± 0.00834	–	–	–	225.615 ± 0.118	30.196	0.00
D2	A2	9297	-2.60302 ± 0.00250	-1.29996 ± 0.00208	–	–	–	0.217 ± 0.143	1.397	0.32
D2	A3	9297	-2.60346 ± 0.00250	-1.30131 ± 0.00208	–	–	–	0.057 ± 0.143	0.312	0.91
D2	A4	9176	-2.60341 ± 0.00250	-1.30142 ± 0.00208	–	–	–	0.086 ± 0.143	0.342	0.81
D7	A0	14922	-2.60005 ± 0.00246	-1.30013 ± 0.00205	–	–	–	0.025 ± 0.146	0.145	1.00
D7	A1	14922	-2.64907 ± 0.00281	-1.69333 ± 0.00233	–	–	–	47.836 ± 0.138	10.826	0.00
D7	A2	14922	-2.59607 ± 0.00246	-1.30562 ± 0.00205	–	–	–	0.262 ± 0.147	2.129	0.31
D7	A3	14922	-2.60941 ± 0.00246	-1.30240 ± 0.00206	–	–	–	0.565 ± 0.147	0.466	0.42
D7	A4	14922	-2.60941 ± 0.00246	-1.30240 ± 0.00206	–	–	–	0.565 ± 0.147	0.466	0.42
D2	A5	9297	-2.60015 ± 0.00252	-1.29992 ± 0.00211	0.00008 ± 0.00398	–	–	-0.007 ± 0.144	0.146	1.00
D2	A6	9297	-2.60182 ± 0.00252	-1.30098 ± 0.00210	-0.00405 ± 0.00399	–	–	0.225 ± 0.144	1.266	0.34
D2	A7	9297	-2.60352 ± 0.00252	-1.30142 ± 0.00210	0.00126 ± 0.00399	–	–	0.077 ± 0.144	0.318	0.89
D7	A5	14922	-2.60002 ± 0.00247	-1.30016 ± 0.00207	-0.00007 ± 0.00390	–	–	0.027 ± 0.148	0.148	1.00
D7	A6	14922	-2.59601 ± 0.00247	-1.30540 ± 0.00207	0.00271 ± 0.00387	–	–	0.227 ± 0.147	2.071	0.30
D7	A7	14922	-2.60942 ± 0.00247	-1.30264 ± 0.00207	-0.00183 ± 0.00388	–	–	0.596 ± 0.147	0.456	0.38
D2	A8	9297	–	–	–	-3.22008 ± 0.00249	-2.50022 ± 0.00208	0.046 ± 0.143	0.687	0.68
D2	A9	9297	–	–	–	-3.25768 ± 0.00244	-2.55847 ± 0.00203	-12.766 ± 0.146	0.805	0.00
D7	A8	14922	–	–	–	-3.21650 ± 0.00246	-2.50305 ± 0.00205	0.064 ± 0.147	1.032	0.49
D7	A9	14922	–	–	–	-3.26423 ± 0.00236	-2.54528 ± 0.00198	-18.733 ± 0.152	1.199	0.00

Table 10: Results from all experiments run with the CepheidObsModel using an intrinsic Cepheid P-L relation following the metallicity dependent G08. Each value given is the mean value of 100 runs of each experiment.

Data	Anlys	N_{obs}	$a \pm \sigma_a$	$b \pm \sigma_b$	$m \pm \sigma_m$	$k_1 \pm \sigma_{k_1}$	$k_2 \pm \sigma_{k_2}$	$c \pm \sigma_c$	$\sigma_{c,s}$	true
True			-2.60	-1.30	0.27	-3.3286	-2.5009	0		
D3	A0	9298	-2.59769 ± 0.00249	-1.30180 ± 0.00208	–	–	–	-0.037 ± 0.143	0.925	0.52
D3	A1	9298	-7.63864 ± 0.01161	-2.63984 ± 0.00833	–	–	–	225.482 ± 0.118	30.258	0.00
D3	A2	9298	-2.59280 ± 0.00249	-1.30597 ± 0.00207	–	–	–	-0.067 ± 0.143	1.744	0.23
D3	A3	9298	-2.60137 ± 0.00250	-1.30276 ± 0.00208	–	–	–	-0.010 ± 0.143	1.017	0.47
D3	A4	9176	-2.60118 ± 0.00250	-1.30301 ± 0.00208	–	–	–	0.016 ± 0.143	0.979	0.49
D4	A0	9273	-2.59701 ± 0.00249	-1.29914 ± 0.00207	–	–	–	-0.704 ± 0.142	0.941	0.26
D4	A1	9273	-8.31110 ± 0.01516	-3.13464 ± 0.01075	–	–	–	239.593 ± 0.115	23.004	0.00
D4	A2	9273	-2.59572 ± 0.00249	-1.30022 ± 0.00207	–	–	–	-0.762 ± 0.142	1.872	0.26
D4	A3	9273	-2.60028 ± 0.00250	-1.30035 ± 0.00207	–	–	–	-0.654 ± 0.142	0.991	0.29
D4	A4	9155	-2.59941 ± 0.00249	-1.30084 ± 0.00207	–	–	–	-0.691 ± 0.142	0.993	0.29
D8	A0	14922	-2.60048 ± 0.00246	-1.30028 ± 0.00205	–	–	–	0.056 ± 0.146	0.842	0.55
D8	A1	14922	-2.64947 ± 0.00281	-1.69333 ± 0.00233	–	–	–	47.848 ± 0.138	10.914	0.00
D8	A2	14922	-2.59543 ± 0.00246	-1.30742 ± 0.00205	–	–	–	0.400 ± 0.146	1.902	0.25
D8	A3	14922	-2.60972 ± 0.00246	-1.30263 ± 0.00206	–	–	–	0.614 ± 0.147	0.989	0.28
D8	A4	14922	-2.60972 ± 0.00246	-1.30263 ± 0.00206	–	–	–	0.614 ± 0.147	0.989	0.28
D9	A0	14836	-2.59881 ± 0.00248	-1.30244 ± 0.00207	–	–	–	3.229 ± 0.146	0.806	0.01
D9	A1	14836	-2.65301 ± 0.00283	-1.69541 ± 0.00234	–	–	–	49.315 ± 0.138	11.033	0.00
D9	A2	14836	-2.59985 ± 0.00248	-1.30589 ± 0.00207	–	–	–	3.721 ± 0.146	1.851	0.02
D9	A3	14836	-2.60800 ± 0.00248	-1.30435 ± 0.00207	–	–	–	3.728 ± 0.146	0.937	0.01
D9	A4	14836	-2.60800 ± 0.00248	-1.30435 ± 0.00207	–	–	–	3.728 ± 0.146	0.937	0.01
D3	A5	9298	-2.59737 ± 0.00251	-1.30150 ± 0.00210	0.13684 ± 0.00397	–	–	-0.088 ± 0.144	0.583	0.60
D3	A6	9298	-2.59483 ± 0.00252	-1.30405 ± 0.00210	0.13685 ± 0.00403	–	–	-0.085 ± 0.144	1.534	0.29
D3	A7	9298	-2.60264 ± 0.00252	-1.30155 ± 0.00210	0.13604 ± 0.00403	–	–	-0.017 ± 0.144	0.726	0.56
D4	A5	9273	-2.59832 ± 0.00251	-1.29941 ± 0.00209	0.16382 ± 0.00356	–	–	-0.345 ± 0.143	0.630	0.63
D4	A6	9273	-2.59403 ± 0.00251	-1.30295 ± 0.00209	0.16428 ± 0.00356	–	–	-0.393 ± 0.143	1.606	0.27
D4	A7	9273	-2.59908 ± 0.00251	-1.30264 ± 0.00209	0.15555 ± 0.00356	–	–	-0.302 ± 0.143	0.678	0.60
D8	A5	14922	-2.59901 ± 0.00247	-1.30125 ± 0.00207	0.13544 ± 0.00391	–	–	0.047 ± 0.147	0.620	0.64
D8	A6	14922	-2.59618 ± 0.00247	-1.30682 ± 0.00207	0.13404 ± 0.00389	–	–	0.413 ± 0.147	1.811	0.16
D8	A7	14922	-2.61027 ± 0.00248	-1.30227 ± 0.00207	0.13546 ± 0.00389	–	–	0.638 ± 0.147	0.864	0.38
D9	A5	14836	-2.59955 ± 0.00248	-1.30146 ± 0.00208	0.15488 ± 0.00361	–	–	1.442 ± 0.153	0.612	0.06
D9	A6	14836	-2.60019 ± 0.00248	-1.30421 ± 0.00207	0.14812 ± 0.00360	–	–	1.872 ± 0.154	1.705	0.12
D9	A7	14836	-2.60856 ± 0.00248	-1.30306 ± 0.00208	0.11707 ± 0.00360	–	–	2.305 ± 0.154	0.813	0.00
D3	A8	9298	–	–	–	-3.21384 ± 0.00249	-2.50410 ± 0.00207	-0.110 ± 0.143	1.177	0.37
D3	A9	9298	–	–	–	-3.25147 ± 0.00244	-2.56230 ± 0.00203	-12.934 ± 0.146	1.296	0.00
D4	A8	9273	–	–	–	-3.21488 ± 0.00249	-2.49996 ± 0.00207	-0.789 ± 0.142	1.271	0.27
D4	A9	9273	–	–	–	-3.25189 ± 0.00244	-2.55811 ± 0.00203	-13.582 ± 0.145	1.371	0.00
D8	A8	14922	–	–	–	-3.21646 ± 0.00246	-2.50398 ± 0.00205	0.146 ± 0.146	1.168	0.31
D8	A9	14922	–	–	–	-3.26425 ± 0.00236	-2.54622 ± 0.00198	-18.637 ± 0.152	1.333	0.00
D9	A8	14836	–	–	–	-3.21776 ± 0.00248	-2.50433 ± 0.00207	3.391 ± 0.146	1.148	0.01
D9	A9	14836	–	–	–	-3.26441 ± 0.00238	-2.54716 ± 0.00200	-15.059 ± 0.151	1.266	0.00

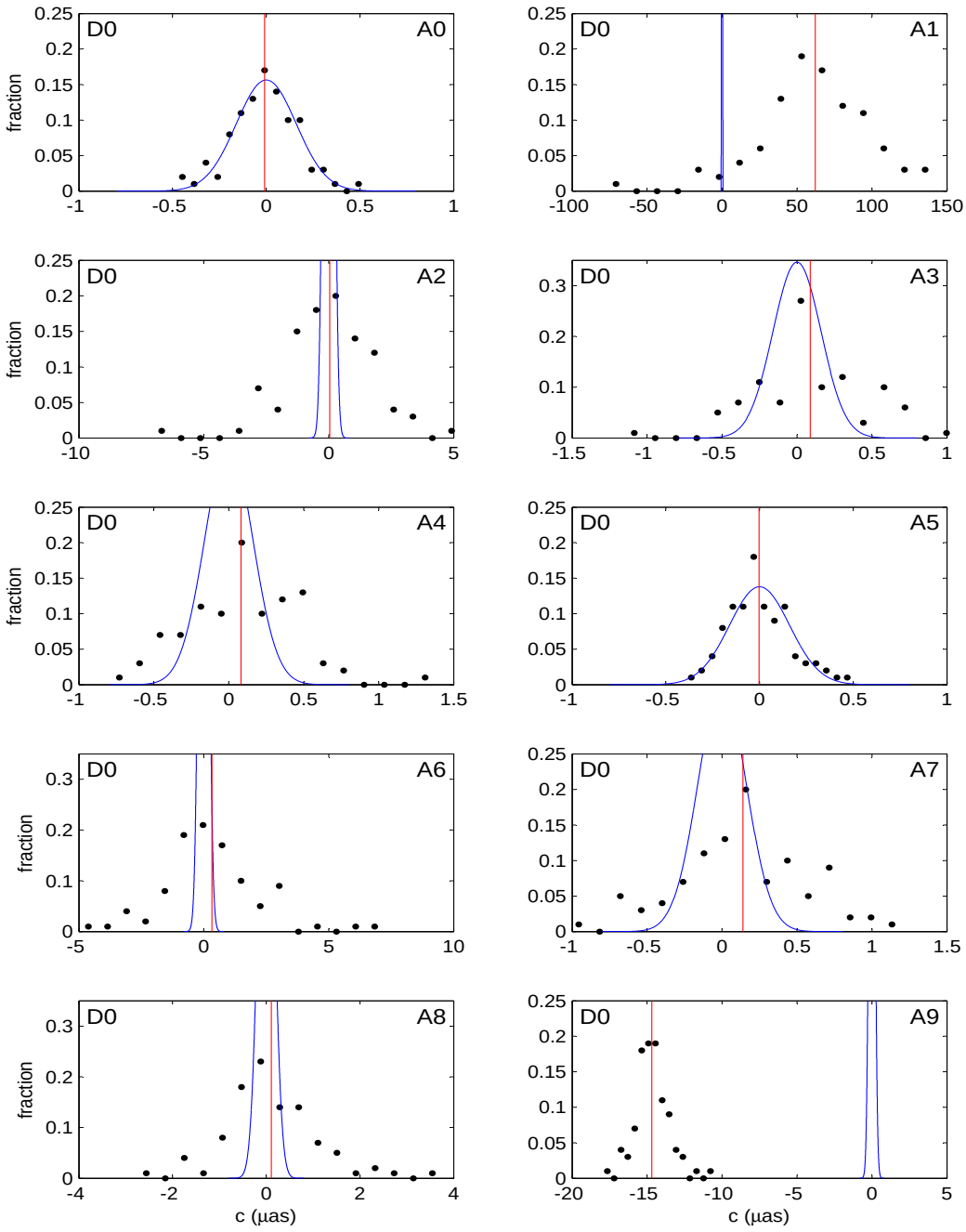


Figure 31: Distribution of fitted c values for each experiment using data set D0 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

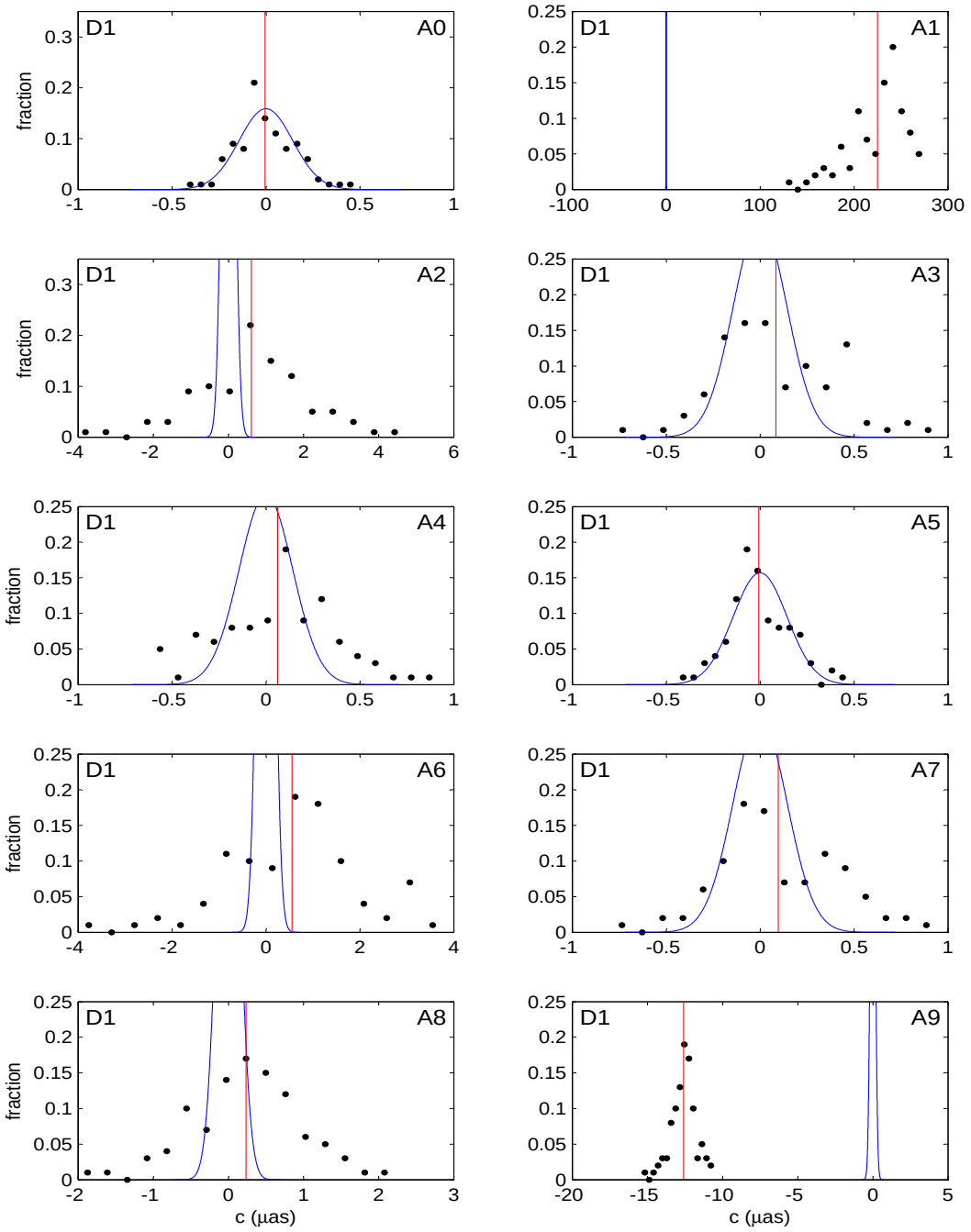


Figure 32: Distribution of fitted c values for each experiment using data set D1 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

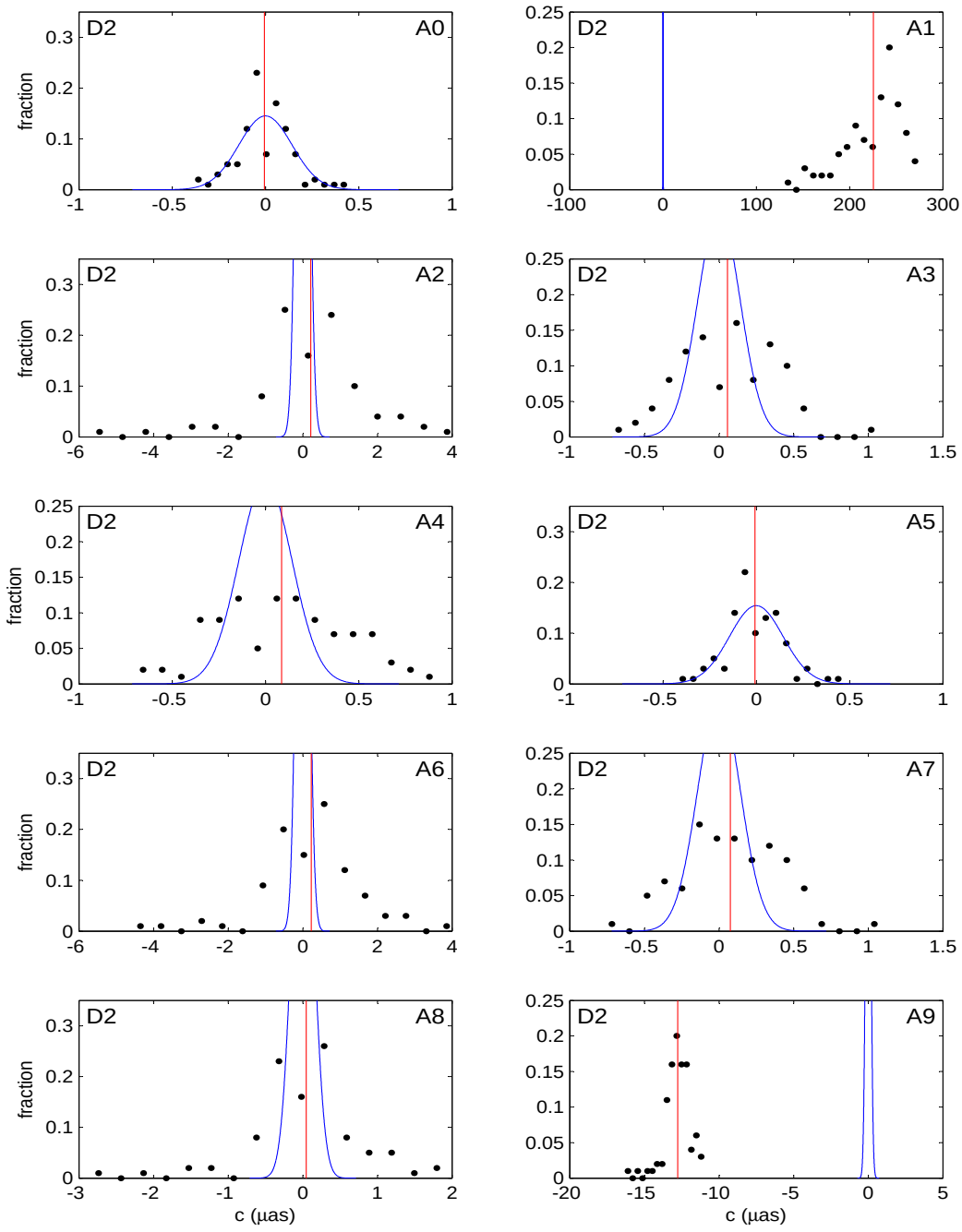


Figure 33: Distribution of fitted c values for each experiment using data set D2 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

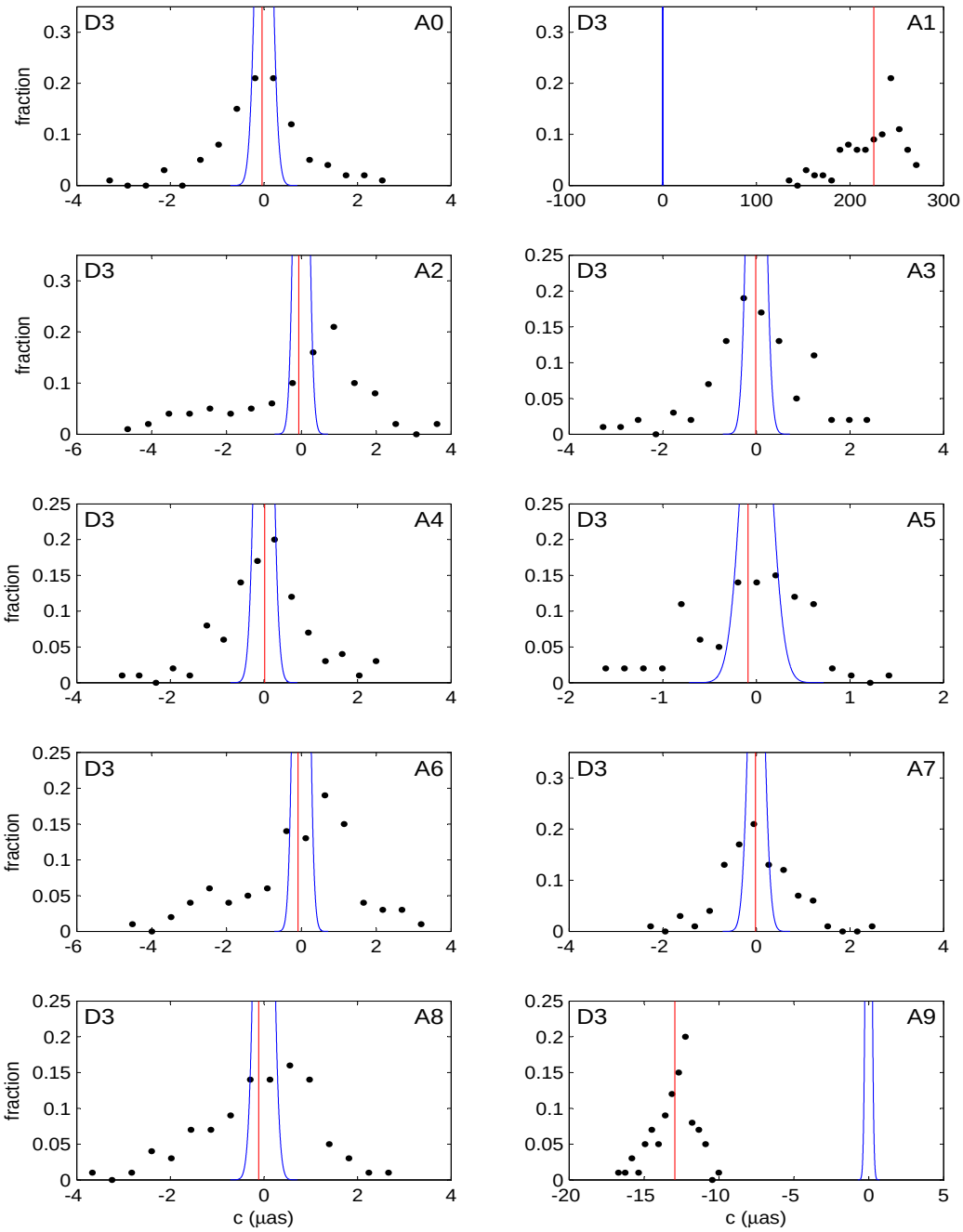


Figure 34: Distribution of fitted c values for each experiment using data set D3 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

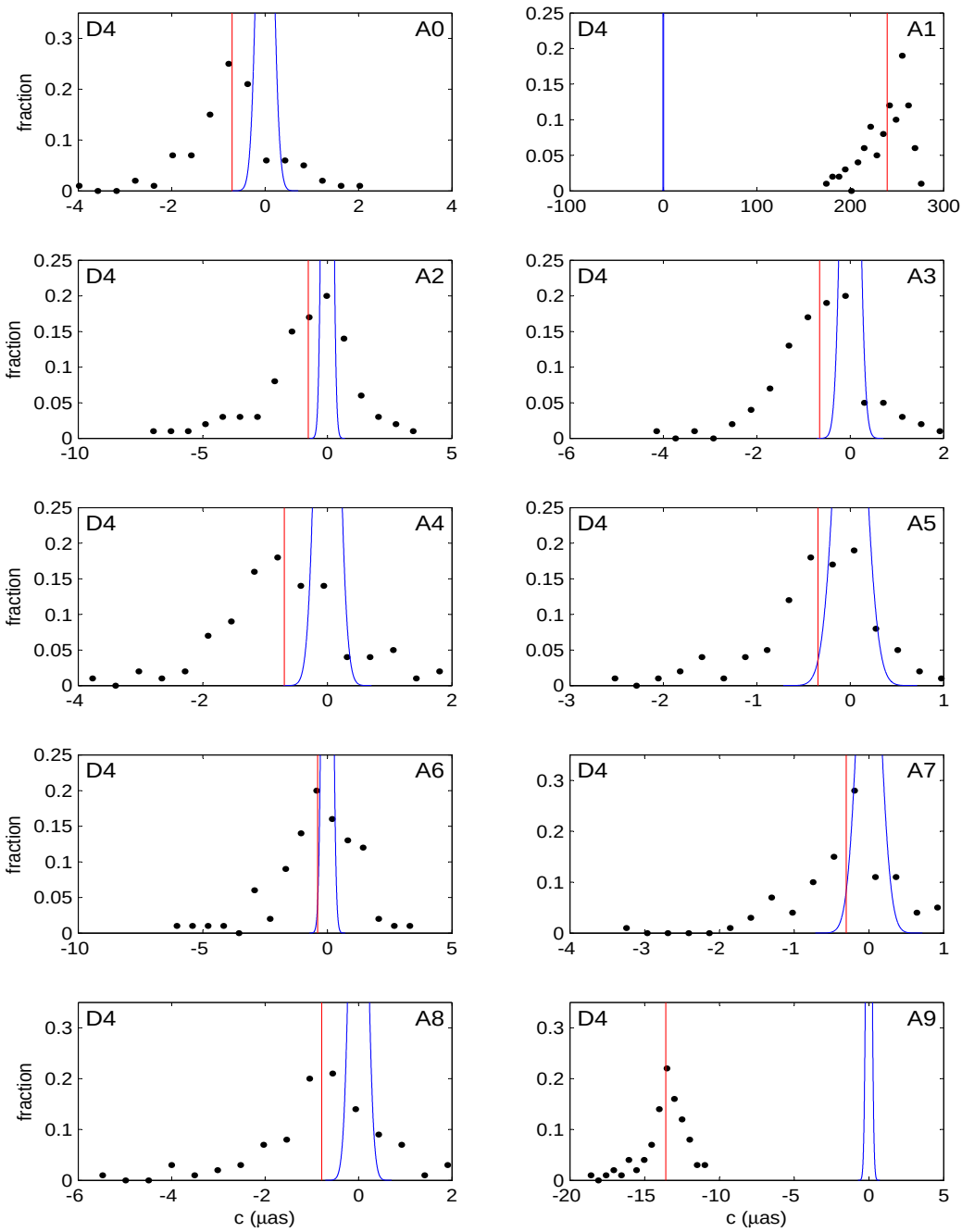


Figure 35: Distribution of fitted c values for each experiment using data set D4 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

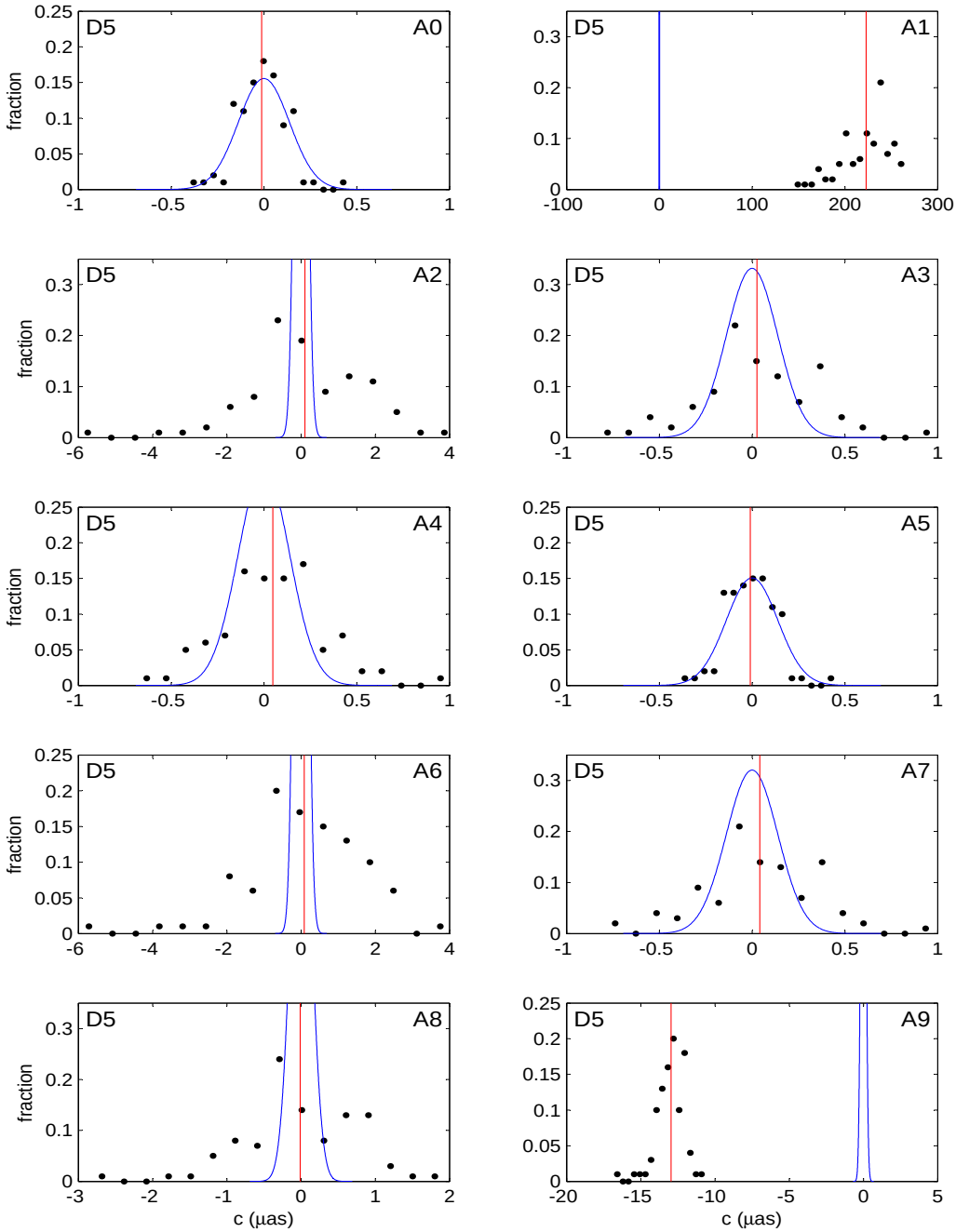


Figure 36: Distribution of fitted c values for each experiment using data set D5 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

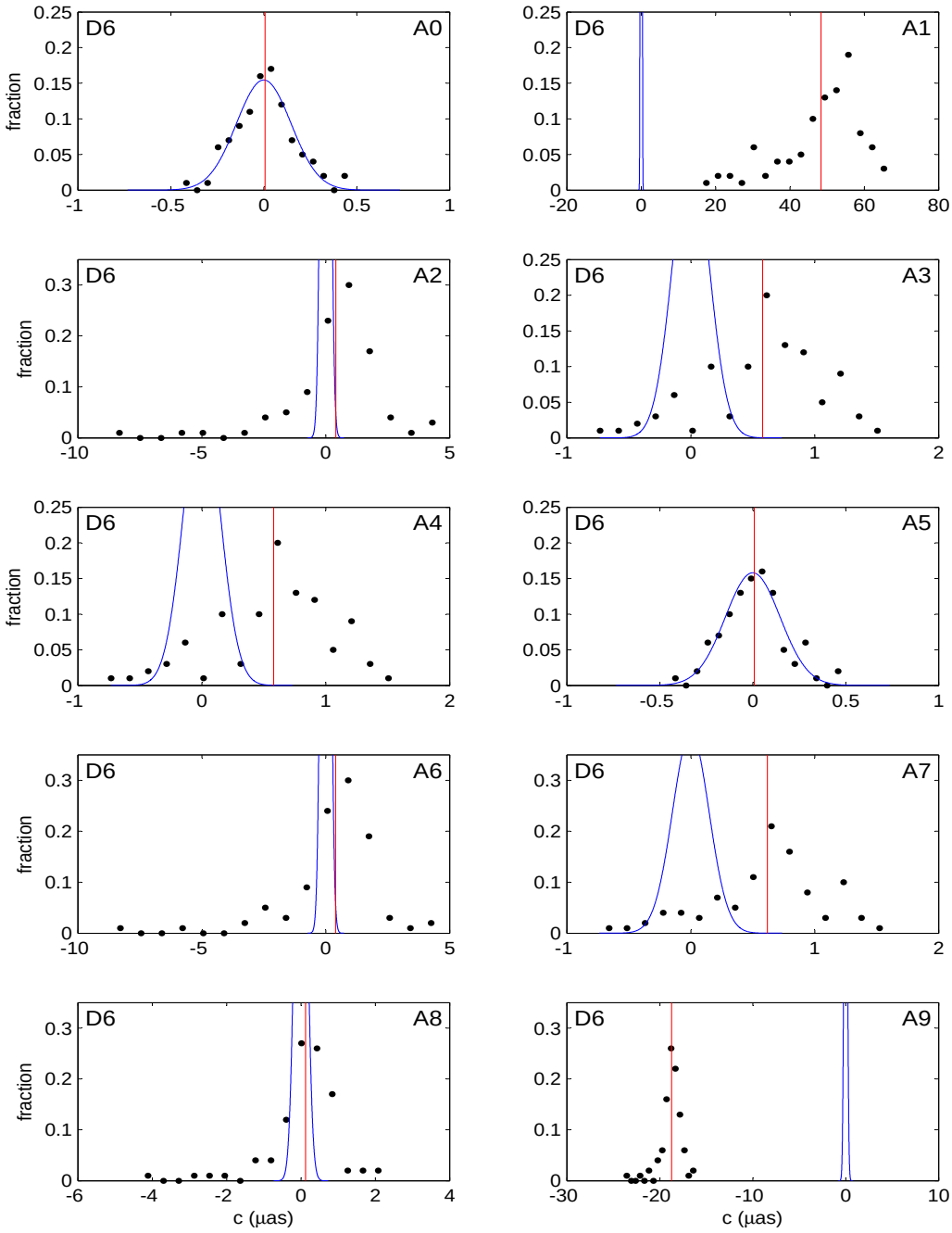


Figure 37: Distribution of fitted c values for each experiment using data set D6 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

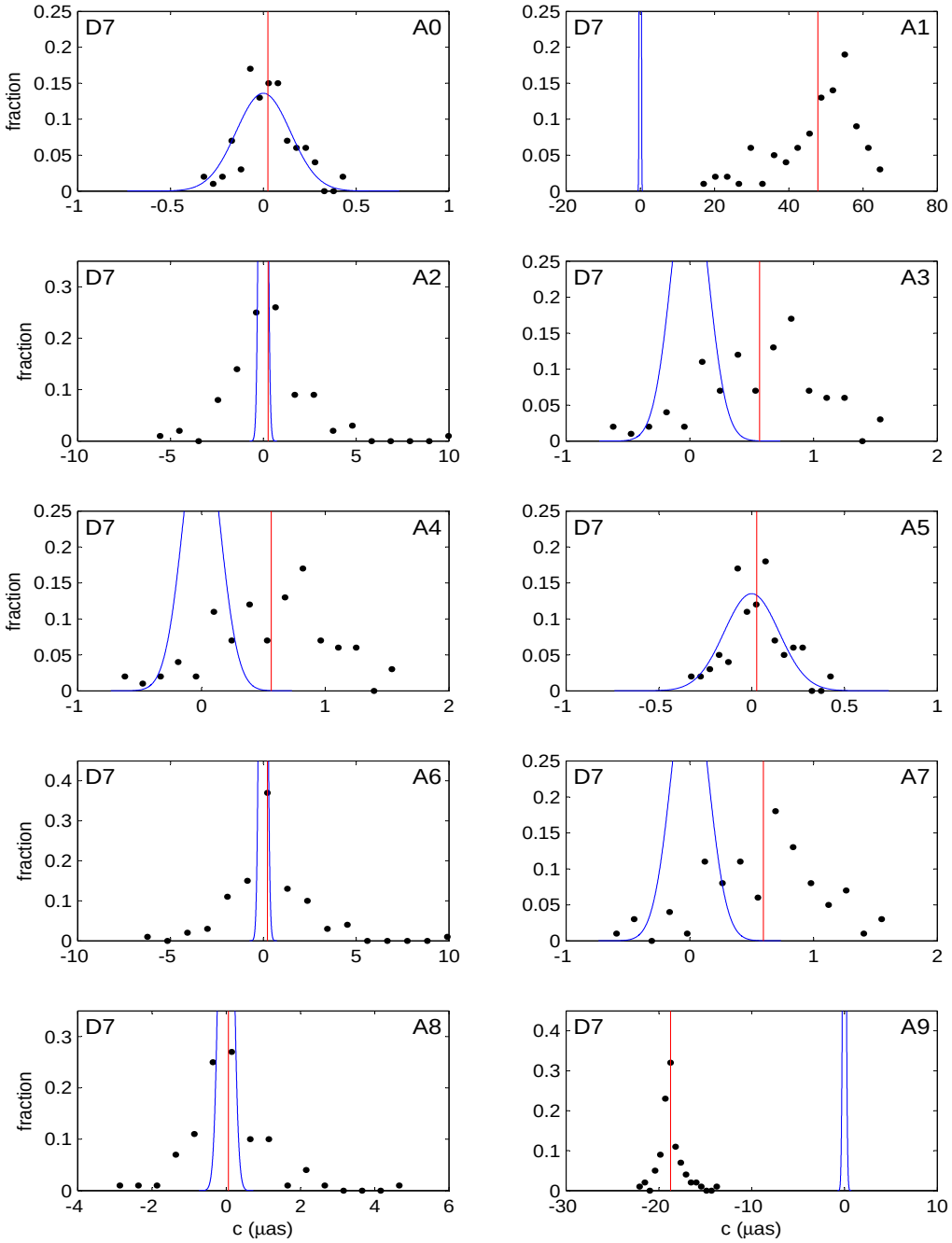


Figure 38: Distribution of fitted c values for each experiment using data set D7 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

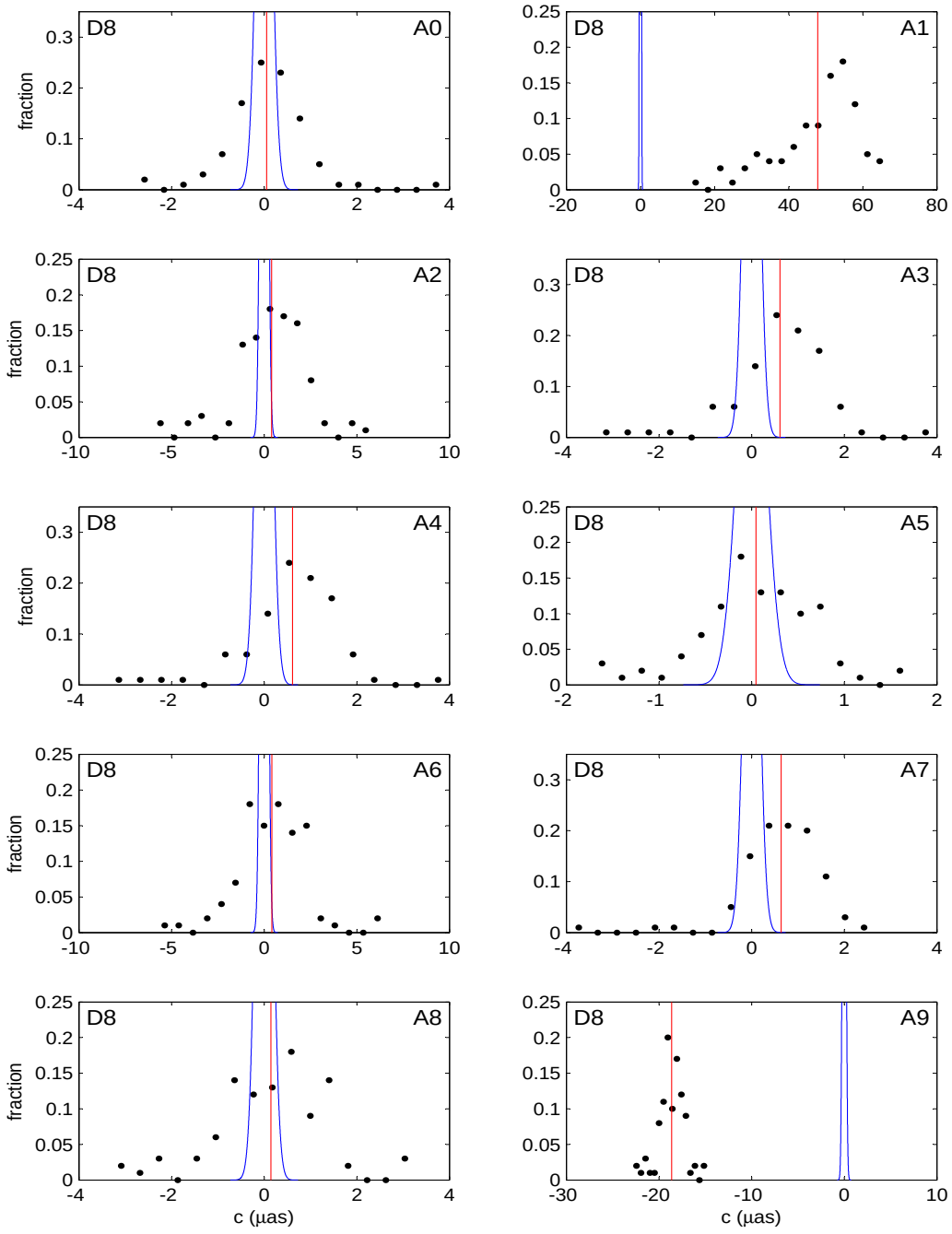


Figure 39: Distribution of fitted c values for each experiment using data set D8 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).

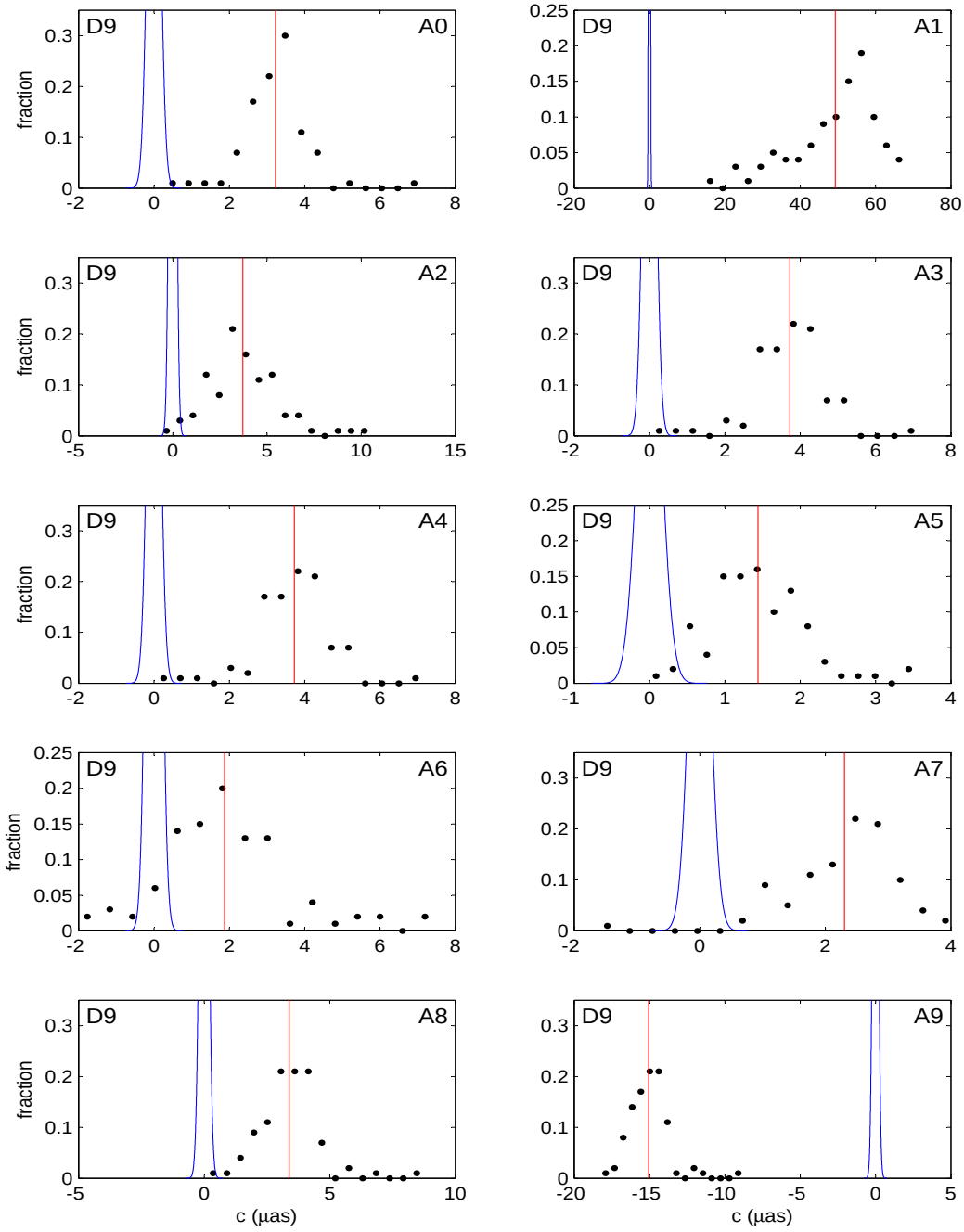


Figure 40: Distribution of fitted c values for each experiment using data set D9 after 100 realizations (black dots). Also included is the expected formal distribution (blue curve) and the sample mean value (red line).