Inferring the distance to Westerlund 1 from Gaia DR2

Mojgan Aghakhanloo1⋆, Jeremiah W. Murphy1, Nathan Smith2, John Parejko3, Mariangelly Díaz-Rodríguez1, Maria R. Drout4,5, Jose H. Groh6, Joseph Guzman1, Keivan G. Stassun7,8.

1 Department of Physics, Florida State University, 77 Chieftan Way, Tallahassee, FL 32306, USA
2 Steward Observatory, University of Arizona, 933 N. Cherry Ave., Tucson, AZ 85721, USA
3 Department of Astronomy, Box 351580, University of Washington, Seattle, WA 98195, USA
4 The Observatories of the Carnegie Institution for Science, 813 Santa Barbara St., Pasadena, CA 91101, USA
5 Hubble and Carnegie-Dunlap Fellow
6 School of Physics, Trinity College Dublin, the University of Dublin, Dublin, Ireland
7 Vanderbilt University, Department of Physics & Astronomy, 6301 Stevenson Center Lane, Nashville, TN 37235, USA
8 Fisk University, Department of Physics, 1000 17th Avenue N., Nashville, TN 37208, USA

25 January 2019

ABSTRACT

Westerlund 1 (Wd1) is potentially the largest star cluster in the Galaxy. That designation critically depends upon the distance to the cluster, yet the cluster is highly obscured, making luminosity-based distance estimates difficult. Using Gaia Data Release 2 (DR2) parallaxes and Bayesian inference, we infer a parallax of $0.31 \pm 0.04$ mas corresponding to a distance of $3.2 \pm 0.4$ kpc. To leverage the combined statistics of all stars in the direction of Wd1, we derive the Bayesian model for a cluster of stars hidden among Galactic field stars; this model includes both the zero-point and astrometric excess noise. We infer the exponential length scale, $L$, for field stars; $L = 0.84_{-0.03}^{+0.02}$ kpc. This is $\sim 1.7$ times smaller than the models used for Bailer-Jones et al. (2018). Previous estimates for the distance to Wd1 ranged from 1.2 to 5.5 kpc, although values around 5 kpc have usually been adopted. The Gaia DR2 parallaxes reduce the uncertainty from a factor of 3 to 15% and rules out the most often quoted value of 5 kpc with 99% confidence. This new distance allows for more accurate mass and age determinations for the stars in Wd1. For example, the previously inferred initial mass at the main sequence turnoff was around 40 $M_\odot$; the new Gaia DR2 distance shifts this down to about 25 $M_\odot$. This has important implications for our understanding of late stages of stellar evolution, including the initial mass of the magnetar and the LBV in Wd1. Similarly, the new distance suggests that the total cluster mass is about three times lower than previously calculated.

Key words: stars: evolution—open clusters and associations: individual: Westerlund 1-methods: Bayesian analysis

1 INTRODUCTION

Massive stars are a central focus of ongoing work in stellar evolution theory. Many gaps exist in our understanding of massive stars due to their rarity, short lifetimes, high fraction of interacting binaries, and imprecise Galactic distances. To better understand the evolutionary path of massive stars, it is helpful to explore associated clusters or OB associations. If one can show that the star is a member of a cluster or association, studying the cluster provides a unique insight into intrinsic properties of cluster members. For example, Westerlund 1 has a Luminous Blue Variable (LBV) (Clark and Negueruela 2004), at least 24 Wolf-Rayet stars (WR) (Fenech et al. 2018; Rosslowe et al. 2015; Clark, J. S. et al. 2005; Groh et al. 2006; Crowther et al. 2006), 6 yellow hypergiants (YHG) (Clark, J. S. et al. 2005), and a magnetar (Muno et al. 2006); knowing the distance to this one cluster will help to constrain the luminosity, mass, and evolution of all of these late phases of stellar evolution.

The massive young star cluster Westerlund 1 (Wd1) was detected by Westerlund (1961) during a survey of the Milky Way. Wd1 is located at RA(2000) = 16h47m04.0s, Dec(2000) = $-45^\circ51'04''.9$ and has large populations of WRs, YHGs, a LBV, and a magnetar (Fenech et al. 2018).
Previous distance estimates mostly based on reddening-distance relationship ranged from 1.2 to 5.5 kpc (Crowther et al. 2006; Clark, J. S. et al. 2005; Piatti, A. E. et al. 1998; Westerlund 1968). Westerlund (1961) first suggested \(d_V = 12.0\ \text{mag.}\) and reported a distance of 1.4 kpc. Later, Westerlund (1968) derived a significantly larger distance of 5 kpc by using VRI photographic photometry with near-infrared photometry of the brightest stars. In contrast, Piatti, A. E. et al. (1998) presented CCD imaging in the V and I bands, and using isochrone fitting, estimated a distance of 1.1 \(\pm 0.4\) kpc. On the other hand, Clark, J. S. et al. (2005) obtained spectra for the brightest members of Wd1, and noted that they are all post main-sequence stars; since the isochrone fitting of Piatti, A. E. et al. (1998) assumed that many of these lie on the MS, Piatti, A. E. et al. (1998) distance estimate was incorrect. Six of the stars are YHGs, and the most luminous YHGs are presumed to have relatively standard luminosity of around \(\log(L/L_{\odot}) \sim 5.7\) (Smith et al. 2004a). Assuming that the YHGs in Wd1 were at the observed upper luminosity limit for cool supergiants, and adopting an extinction of \(A_v = 11.0\), Clark, J. S. et al. (2005) inferred a distance of \(\lesssim 5.5\) kpc. However, they noted that their reddening law is not entirely consistent with Wd1 data. To place a lower limit on the distance, they noted a lack of radio emission from the WR winds; this suggests a minimum distance of \(\sim 2\) kpc. Hence, Clark, J. S. et al. (2005) reported a distance of \(2 < R < 5.5\) kpc. Crowther et al. (2006) inferred a similar distance using near-IR classification of WN and WC stars. More recently, Kothes and Dougherty (2007) derived a distance of 3.9 \(\pm 0.7\) kpc based on the radial velocity of Hi features in the direction of Wd1, and Brandner et al. (2008) derived a distance of 3.55 \(\pm 0.17\) kpc based on MS fitting. However, we caution against adopting this last seemingly accurate distance estimate; it was based upon saturated infrared data. Since the apparent "MS" in the color-magnitude diagram is vertical and featureless between the detection threshold and the saturation limit, it would be difficult to infer much about the cluster. Given all of these difficulties in estimating the distance, it is necessary to use an independent distance estimate.

The main goal of this paper is to infer an independent and geometric distance to Wd1 using Gaia DR2 data. The Gaia DR2 parallax precision for individual stars in Wd1 ranges from 0.05 mas to 1.0 mas. Since Wd1 is probably of order a few kpc in distance, the larger uncertainties prohibit a precise distance estimate for an individual star. However, the combined statistics of all cluster members should easily produce a more precise distance. Section 2 describes the data and method to infer the distance. First, we describe the Gaia DR2 data and possible systematic uncertainties. Then we infer the approximate cluster distance by modeling both cluster and field stars (section 2.2). In Section 2.3 we use a Bayesian inference technique to infer the distance, cluster density, field-star density, and field-star length scale. In Section 3 we present the inferred distance to Wd1 (3.2 \(\pm 0.4\) kpc) and the length scale for the field star distribution. Then, we compare this distance and field-star distribution with previous works (section 3). We also discuss how the revised distance affects the properties of cluster members. In particular, we infer that all luminosities are reduced by \(-0.38\) dex. This also implies a lower main-sequence turnoff mass, and lower initial masses for the magnetar and the LBV that are significantly below 40 \(M_{\odot}\), contrary to previous estimates. Section 3 presents a summary and direction for further investigation.

2 METHOD

In this section, we describe the method to infer the distance to Wd1. The stars in the direction of Wd1 comprise cluster stars as well as Galactic field stars. Therefore, to infer the distance to Wd1, our likelihood model must account for both cluster and field stars. The following sections describe the data and methods required to model both components and infer the distance to Wd1.

2.1 Gaia Data Release 2 (DR2) Data

The source of the data is the Gaia DR2. We collect all Gaia DR2 sources within 10' of the position of Wd1; the position of Wd1 is RA(2000) = \(16^h 47^m 04^s.0\), Dec(2000) = \(-45^\circ 51' 04''9\). For each source, we note the parallax, \(\pi\), the theoretical uncertainty on the parallax, \(\sigma_\pi\), the astrometric excess noise, \(\epsilon\), and the astrometric excess noise significance, \(D\).

Figure 1 presents the positions of all objects within 10' of Wd1. The density of stars is mostly uniform throughout the field of view; however, the density does noticeably increase toward the center. The inner circle marks a region that is 1' from the center of the cluster. The outer annulus extends from 9' to 10'. Objects in the inner circle are mostly associated with the cluster, and the stars in the outer annulus are mostly field stars. This spatial separation between cluster stars and field stars suggests a strategy for constraining the parameters for each population. The inner circle contains both field and cluster stars, but is dominated by cluster stars. Therefore, the inner circle provides a good constraint on the cluster population. The outer annulus likely contains mostly field stars. Therefore, the outer region will constrain the field star distribution.

Figures 2-5 show histograms of parallax, parallax uncertainty, astrometric excess noise, and astrometric excess noise significance for the inner circle and the outer annulus. The average parallax for the inner circle is 0.28 \(\pm 0.04\) mas, and the average parallax for the outer annulus is 0.5 \(\pm 0.02\) mas. The difference in average parallax already indicates that the cluster is farther than the average field star. The distribution of expected parallax uncertainty (Figure 3) is similar between the two regions. The minimum parallax uncertainty in the direction of Wd1 is 0.04 mas, but the vast majority of uncertainties are even larger (up to a few mas), including systematics such as zero-point and excess astrometric noise. The distribution for astrometric excess noise, (Figure 4) shows a considerable difference between the two regions. A higher fraction of objects in the inner circle have

© 2012 RAS, MNRAS
large excess noise. In fact, the significance (Figure 5) is quite high, greater than 2 for most of the stars in the inner region. This dictates that one must include the astrometric excess noise when modeling the likelihood of observations.

2.2 Average Statistics: A First Rough Statistical Inference

Before using Bayesian inference, we employ average statistics to roughly infer the cluster distance.

Figure 6 shows the average parallax as a function of radius from the cluster center. It also shows a clear indication that the cluster has a smaller parallax (larger distance) than the average field star. The blue line represents \( \langle \pi \rangle \) for several annuli. Each annulus is 0.5' in width. To estimate the uncertainty on the average, the vertical error bars show the 68% confidence interval after bootstrapping the average 500 times for each annulus. For radii above \( \sim 3' \), the average parallax seems to be dominated by the field stars. Interior to this radius, \( \langle \pi \rangle \) becomes more influenced by the cluster with decreasing radius.

To roughly infer the cluster parallax, we model the average parallax as follows. For simplicity, first consider the average parallax for the outer ring. We assume that the outer ring is dominated by field stars. Under these assumptions, the average parallax of the field stars is
Even though there is a wide distribution of parallaxes for both the inner and outer ring, there is a clear difference in their average parallax. The inner ring which is dominated by cluster stars has a parallax of $\sim 0.28 \pm 0.04$ mas. The outer ring, which is dominated by field stars along the line of sight has a parallax of $\sim 0.5 \pm 0.02$ mas.

Lindegren et al. (2012) state that any source with $D > 2$ has significant excess noise. Many of the stars in the inner ring have highly significant excess noise. Therefore, one must include the excess noise when inferring the distance.
Figure 6. Observed average parallax, $\langle \varpi \rangle$, for each ring. The average parallax for the outer rings most likely represent the field stars, and the inner rings represent the cluster. Below $3^\prime$, the average parallax transitions from being dominated by the field stars to the cluster stars.

$$\langle \varpi_f \rangle = \int \varpi P_f(\varpi) d\varpi, \quad (1)$$

$P_f$ is the probability density function

$$P_f = \frac{n_f(\varpi)}{n_f} = \frac{n_f(\varpi)}{N_f}, \quad (2)$$

where $A_o$ and $N_f$ are the area and the number of stars in the outer ring respectively. If we assume that each ring $i$ is a combination of the cluster stars and the field stars then the average parallax of each ring is

$$\langle \varpi_i \rangle = \int \varpi \left( \frac{P_{cl,i} N_{cl,i}}{N_{cl,i} + N_f} + \frac{P_f N_f}{N_{cl,i} + N_f} \right) d\varpi. \quad (3)$$

Using the above definitions for the distributions and averages, equation (3) becomes

$$\langle \varpi_i \rangle = \frac{1}{n_f + n_{cl,i}} \left[ < \varpi_{cl,i} > n_{cl,i} + < \varpi_f > n_f \right]. \quad (4)$$

This can be converted to the following expression for the average cluster parallax

$$\langle \varpi_{cl,i} \rangle = \frac{< \varpi_i > - < \varpi_f >}{1 - n_f/n_i}. \quad (5)$$

Where $n_i$ is the total number density of stars in each annulus, $n_f$ is the number density of field stars, and also is the number density of the outer ring, $\langle \varpi_f \rangle$ is the average parallax for the most outer ring, and $\langle \varpi_i \rangle$ is the average parallax for each annulus $i$.

Figure 7 shows this rough inference for the cluster parallax for five inner rings between $0^\prime$ and $2.5^\prime$. Three inner rings suggest that the cluster parallax is anywhere from 0 to 0.3 mas. In the next subsections, we infer the cluster parallax through Bayesian inference.

2.3 Bayesian analysis

To infer the Wd1 parallax, the posterior distribution for both the model parameters, $\theta$ and the nuisance parameters, $\eta$, is the product of the likelihood $\mathcal{L}(D, \eta|\theta)$ and the prior probability ($P(\theta)$):

$$P(\theta, \eta|D) \propto \mathcal{L}(D, \eta|\theta) P(\theta). \quad (6)$$

Using the conditional probability theorem, the likelihood is further expanded as

$$\mathcal{L}(D, \eta|\theta) = P(D|\eta) P(\eta|\theta). \quad (7)$$

Marginalizing over the nuisance parameters gives the posterior distribution for just the model parameters:

$$P(\theta|D) = \int P(\theta, \eta|D) d\eta. \quad (8)$$

To model the likelihood, we consider two sets of stars.
Figure 7. A simple inference of the cluster parallax, $\langle \varpi_{cl} \rangle$. The inner three rings suggest that the cluster parallax is anywhere from 0 to 0.3 mas. Figures 10 and 11 show a thorough Bayesian inference of the cluster parallax.

One set contains both cluster and field stars. The other set only includes field stars, and will give the constraint on the field-star length scale. The full set of observations, $D$, include the parallaxes and parallax uncertainties for the inner circle in Figure 2, $\{ \varpi_j, \sigma_j \}$, and the parallaxes and parallax uncertainties for the outer annulus, $\{ \varpi_k, \sigma_k \}$. $D$ also includes the number of stars in the inner circle, $N_i$, and the number of stars of the outer annulus, $N_o$. The probabilistic graphical model (PGM), shown in Figure 8, shows the interdependency between the observations, the model parameters, and the nuisance parameters.

The set of model parameters, $\theta$, are the cluster parallax, $\varpi_{cl}$ (mas), density of the cluster stars, $n_{cl}$ (number per square arc-minute), density of the field stars, $n_f$ (number per square arc-minute), the field-star length scale, $L$ (kpc). Each set also has two nuisance parameters. The nuisance parameters, $\eta$, are two sets of true parallax, $\varpi$ (mas), and two sets of true zero point parallax, $\varpi_{zp}$ (mas).

The PGM provides a clear map in how to use the conditional probability to further deconstruct the likelihood:

$$
\mathcal{L}(D, \eta | \theta) = \mathcal{L}(N_i | n_{cl}, n_f) \times \mathcal{L}(N_o | n_f)
\times \mathcal{L}_{\text{Outer}}(\{ \varpi_k, \sigma_k \}, \varpi_{zp} | L, \mu_{zp}, \sigma_{zp})
\times \mathcal{L}_{\text{Inner}}(\{ \varpi_j, \sigma_j \}, \varpi_{zp} | \theta, \mu_{zp}, \sigma_{zp}).
$$

$\mu_{zp}$ and $\sigma_{zp}$ are fixed quantities that represent the mean and variance for the zero-point parallax. The likelihood for the outer set of data may be further deconstructed into

$$
\mathcal{L}_{\text{Outer}}(\{ \varpi_k, \sigma_k \}, \varpi_{zp} | L, \mu_{zp}, \sigma_{zp}) = \prod_k P_k(\varpi_k | \varpi_{zp} | L) \times P(\varpi_{zp} | L, \mu_{zp}, \sigma_{zp})
$$

and the likelihood for the inner data is

$$
\mathcal{L}_{\text{Inner}}(\{ \varpi_j, \sigma_j \}, \varpi_{zp} | \theta, \mu_{zp}, \sigma_{zp}) = \prod_j P_j(\varpi_j | \varpi_{zp} | \theta) \times P(\varpi_{zp} | \theta, \mu_{zp}, \sigma_{zp}).
$$

The first two likelihoods in equation (9), $\mathcal{L}(N_i | n_{cl}, n_f)$ and $\mathcal{L}(N_o | n_f)$, represent the number of stars in the inner circle and the outer ring:

$$
\mathcal{L}(N_i | n_{cl}, n_f) = \frac{\lambda_i^{N_i} e^{-\lambda_i}}{N_i!},
$$

and

$$
\mathcal{L}(N_o | n_f) = \frac{\lambda_o^{N_o} e^{-\lambda_o}}{N_o!},
$$

where the expected number of stars in the inner circle is $\lambda_i = (n_f + n_{cl}) \times A_i$ and the expected number of stars in the outer ring is $\lambda_o = n_f \times A_o$.

The first term in equation (10), $P_k(\varpi_k | \varpi_{zp} | L)$, is
the probability of observing any parallax and parallax uncertainty for the \( k \)th star in the outer ring:

\[
P_k(\varpi_k, \sigma_k | \hat{\varpi}_k, \hat{\varpi}_{sp_k}) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left[-\frac{(\varpi_k - \hat{\varpi}_k - \hat{\varpi}_{sp_k})^2}{2\sigma_k^2}\right].
\]  

(14)

The second term in equation 10 \( P(\hat{\varpi}_k | L) \), represents the field star distribution. If one considers an image populated with stars, then the total number of stars in the image is given by \( N = \text{FOV} \int n r^2 dr \), where FOV is the field of view in square radians, and \( n \) is the number density of stars. If \( n \) is constant, then any random star in the image is drawn from a probability distribution of \( P(r) \propto r^2 \). Given dust in the Galaxy, the sight line will be attenuated by \( \exp(-r/L) \), where \( L \) is an effective optical depth for extinction. Therefore, the distribution of the field stars is

\[
P(r|L) = \frac{1}{2L^3} r^2 \exp(-r/L).
\]  

(15)

After transforming from distance to parallax, the field-star distribution becomes

\[
P(\hat{\varpi}_k | L) = \frac{1}{2L^3} \frac{1}{\hat{\varpi}_k} \exp\left[-1/(\hat{\varpi}_k L)\right].
\]  

(16)

The third term in equation 10 \( P(\hat{\varpi}_{sp_k} | \mu_{sp}, \sigma_{sp}) \), is the distribution of the zero point parallax for the outer ring. Lindegren et al. (2018) show that the zero point is a function of position and has significant variance. While there is likely one zero point for this field, we are uncertain what that value is. Therefore, we model the distribution for this zero point as:

\[
P(\hat{\varpi}_{sp_k} | \mu_{sp}, \sigma_{sp}) = \frac{1}{\sqrt{2\pi}\sigma_{sp}} \exp\left[-\frac{(\hat{\varpi}_{sp_k} - \mu_{sp})^2}{2\sigma_{sp}^2}\right].
\]  

(17)

Together, equations 14,17 represent the likelihood for the outer ring, given in equation 10.

Similarly, the terms for the inner likelihood are as follows. The first term in equation 11 \( P_j(\varpi_j, \sigma_j | \hat{\varpi}_j, \hat{\varpi}_{sp_j}) \), is the probability of observing parallaxes and parallax uncertainties for the inner circle and is the same as equation 14 with a simple exchange of index from \( k \) to \( j \). The second term in equation 11 \( P(\hat{\varpi}_j | \theta) \), is the distribution of stars in the inner ring. Once again, the inner circle is composed of cluster stars and field stars.

\[
P(\hat{\varpi}_j | \theta) = \frac{n_{cl} \times P_{cl}(\varpi_{cl}) + n_j \times P_j(\varpi_j | L)}{n_{cl} + n_j}.
\]  

(18)

For the cluster star distribution, \( P_{cl}(\varpi_{cl}) \), we assume a delta function at \( \varpi_{cl} \); \( \delta(\hat{\varpi}_j - \varpi_{cl}) \). The field star distribution is the same as equation 10 with the simple change of index from \( k \) to \( j \) due to the presence of the same amount of dust towards the cluster. The third term in equation 11 \( P(\hat{\varpi}_{sp_j} | \mu_{sp}, \sigma_{sp}) \), is the distribution of the zero point parallax for the inner circle and is the same as equation 17 with a simple change of index from \( k \) to \( j \). With simple index replacements, equations 14,17 & 18 together form the inner likelihood.

To find the posterior distribution for just the model parameters, \( \theta \), we marginalize over the nuisance parameters, \( \eta \):

\[
P(\theta | D) = \int P(\theta, \eta | D) d\eta
\]  

(19)

In the PGM, Figure 8, the nuisance parameters are the true parallaxes, \( \hat{\varpi}_j \) and \( \hat{\varpi}_k \), and the zero points, \( \hat{\varpi}_{sp_j} \) and \( \hat{\varpi}_{sp_k} \). For both the inner circle and the outer ring, marginalizing over the zero-points involves convolving two Gaussians, equations 14 & 17. The solution is analytic and a Gaussian with \( \hat{\varpi}_{sp} \) being replaced by \( \mu_{sp} \) and the new width is the quadrature sum of \( \sigma_{sp} \) and \( \sigma_{zp} \). Therefore, the likelihood for
outer ring becomes:

\[
\mathcal{L}_{\text{Outer}}(D|\theta) = \int \frac{1}{\sqrt{2\pi(\sigma_{\hat{\omega}}^2 + \sigma_{zp}^2)}} \exp \left[ -\frac{(\hat{\omega}_k - \hat{\omega}_{cl} - \mu_{zp})}{2(\sigma_{\hat{\omega}}^2 + \sigma_{zp}^2)} \right] \times \frac{1}{2L^3} \exp\left[-1/(\hat{\omega}_k L)\right] d\hat{\omega}_k. \tag{20}
\]

The convolution of the Gaussian and the true field distribution is not analytic and requires a numerical solution.

The inner ring likelihood, equation (1), has two terms. The first represents the cluster term and involves the convolution of two Gaussians and delta function. The result is analytic. The second term, which represents the association with the field stars, involves a convolution of two Gaussians and the field-star distribution. Once again, the convolution of the zero-point Gaussian is analytic, but the convolution of the true parallax with a field-star distribution is not. The likelihood for the inner circle is

\[
\mathcal{L}_{\text{Inner}}(D|\theta) = \left( \frac{n_{cl}}{n_{cl} + n_f} \right) \frac{1}{\sqrt{2\pi(\sigma_{\hat{\omega}}^2 + \sigma_{zp}^2)}} \exp \left[ -\frac{(\hat{\omega}_k - \hat{\omega}_{cl} - \mu_{zp})}{2(\sigma_{\hat{\omega}}^2 + \sigma_{zp}^2)} + \int \frac{n_f}{n_{cl} + n_f} \frac{1}{\sqrt{2\pi(\sigma_{\hat{\omega}}^2 + \sigma_{zp}^2)}} \exp\left[-1/(\hat{\omega}_j L)\right] d\hat{\omega}_j \right]. \tag{21}
\]

\(\sigma_{zp}\) is the expected parallax uncertainty, but it does not represent the full uncertainty. For one, Lindegren et al. (2018) found that the empirical uncertainty is 1.081 times larger than the expected value. Second, \(\sigma_{\hat{\omega}}\) is only accurate if the five-parameter model (positions [2], parallax [1], proper motion [2]) is the correct model for the astrometric solution (Lindegren et al. 2012, 2018). If there are other considerations such as motions due to binarity or calibration issues, then there will be excess noise beyond \(\sigma_{\hat{\omega}}\). This excess astrometric noise, \(\epsilon\), represents the true variation of the astrometric observations. Lindegren et al. (2012, 2018). The excess-noise signal, \(D\), indicates the significance of the measured \(\epsilon\). Almost all of the stars in the inner circle have high \(D\). One approach to dealing with the large excess noise is to cull stars with high \(D\). However, such culling could impose a bias on the data. Rather, we propose to include the excess noise in the model. Therefore, we add the excess noise and parallax uncertainty in quadrature to estimate the total uncertainty in the parallax (figure 9).

\[
\sigma = \sqrt{(1.081\sigma_{\hat{\omega}})^2 + \epsilon^2}. \tag{22}
\]

\(\mu_{zp}\) and \(\sigma_{zp}\) are the mean and variance for the zero point parallax. Lindegren et al. (2018) used quasars to infer the zero point. They found an average of \(-0.029\) mas. They also found that the zero point is a function of color, magnitude and position. Therefore, one needs to solve the zero point for each sample considered. The two most effective means to calculate zero-point are to either use background quasars or to use independent distance measurements. WD1 is in the Galactic plane, so there are no background quasars, and the current distance estimates for WD1 are too uncertain to use to constrain the zero point. Therefore, we will use previous analyses to estimate the zero-point distribution. Riess et al. (2018) inferred a zero point of \(-0.046 \pm 0.013\) mas for the Cepheid sample. Zinn et al. (2018) compared the distances inferred from astroseismology to infer a zero point of \(0.0528 \pm 0.0024\) mas. Stassun and Torres (2018) also reported the zero point of \(-0.082 \pm 0.033\) mas from eclipsing binaries. The mean of the above four investigations is \(\mu_{zp} = 0.05\) mas, and the spatial variation from Lindegren et al. (2018) for \(\sigma_{zp} = 0.03\).

We choose uniform priors, \(P(\theta)\), for all model parameters in equation (7).

### 2.4 Numerical Solution for the Posterior Distribution

To find the posterior distribution, we use a Monte Carlo Markov Chain package (MCMC), emcee (Foreman-Mackey et al. 2013, Goodman and Weare 2010). For each step in the chain, emcee evaluates the posterior by calculating the likelihood for the inner circle (equation 21) and the likelihood for the outer ring (equation 20). Both likelihoods require the convolution of a Gaussian with the true field distribution. Evaluating these integrals is time intensive. Instead of calculating the integrals at every step in the chain, we create look-up tables for each integral.

Each object has its own look-up table evaluated at a grid of points in \(L\). For each trial of \(L\) in the MCMC, we find the convolution by first order interpolation in the look-up table. To construct each look-up table, we use trapezoidal numerical integration, which requires bounds of integration. Formally, the bounds extend from \(\hat{\omega} = -\infty\) to \(\hat{\omega} = \infty\), but that is not practical for trapezoidal numerical integration. Fortunately, the integrands in the likelihoods have a peak and fall off quickly on either side of

Figure 9. Histogram of total uncertainties, \(\sigma^2 = \sigma_{\hat{\omega}}^2 + \epsilon^2\), where \(\sigma_{\hat{\omega}}^2\) is the expected parallax uncertainty and \(\epsilon\) is the excess noise. Technically, the excess noise in the astrometric fit could be distributed among any of five astrometric parameters. We take the most conservative approach by adding the excess noise to parallax uncertainty.
this peak. To ensure that the numerical integration adequately samples this peak, we set the bounds of integration to be centered on the peak and have a width that extends just outside the peak. To roughly estimate the position of the peak and the extent of the bounds, we approximate the integrand as the convolution of two Gaussians. The mode and width of the first Gaussian is straightforward, \( \mu_1 = \bar{x} \) and \( \sigma_1 = \sigma_{\bar{x}} \). The mode of the field star distribution is \( \mu_2 = 0.25/L \), and the width is \( \sigma_2 = 0.5/L \). In the two Gaussian approximation, the mode of the integrand is roughly at \( \mu = p \times \mu_1 + q \times \mu_2 \) and the width is roughly \( \sigma = \sqrt{(p \times \sigma_1^2 + q \times \sigma_2^2) + ((p \times \mu_1^2 + q \times \mu_2^2) - \mu^2)} \) where weighting factors are \( p = \frac{4 \times (L \times \sigma)^2}{1 + 4 \times (L \times \sigma)^2} \) and \( q = \frac{1}{1 + 4 \times (L \times \sigma)^2} \).

Therefore, we integrate from \( \bar{x} - \sigma \) to \( \bar{x} + \sigma \) roughly at \( \hat{x} = \bar{x} + \sigma \).

For each MCMC run, we use 100 walkers, 2000 steps each, and we burn 1000 of those. For the results presented in section 3, the acceptance fraction is in the \( \alpha = 0.5 \) range.

### 3 PARALLAX AND DISTANCE TO WESTERLUND 1

Figure 10 shows the posterior distribution for \( \pi_{cl}, n_{cl}, n_{cl} \), and \( L \). The two regions used to constrain these parameters are an inner circle centered on the position of Wd1 and with a radius of 1', and an outer annulus from 9' to 10'. The values in the top right corner show the highest 68% density interval (HDI) for all parameters. The parallax of the cluster is \( \pi_{cl} = 0.31 \pm 0.04 \) mas, which corresponds to a distance of 3.2 ± 0.4 kpc, density of the cluster is \( n_{cl} = 97.62 \pm 4.85 \) stars per square arc-minute, of field stars is \( n_f = 35.95 \pm 0.79 \) stars per square arc-minute, and the field-star length scale is \( L = 0.84 - 0.03 \) kpc. The posterior shows a single-peak marginalized probability distribution for most parameters. The multimodal behaviour in the length scale distribution is due to the discreetness of the look-up table (section 2.4).

Presumably, the cluster density is a function of radius. Since we model the cluster density with one average density and not a radial profile, there is a potential for bias to affect the inference. As long as the width of each annulus is small compared to the change in density, then approximating the density in each ring with an average annulus should work well. To test this hypothesis, we infer the full posterior distribution for several inner annuli. For each inference, we use the outer annulus from 9' to 10' to constrain the field-star parameters. Figure 11 shows the inferred parallaxes for each annulus. The rings extend from 0' to 0.3', 0.3' to 0.6', 0.6' to 0.9', and 0.9' to 1.2'. The HDI parallaxes are 0.26, 0.37, 0.23, and 0.36 mas. All are consistent with our main result of \( 0.31 \pm 0.04 \) mas from using an inner circle with radius 1' (Figure 10). Therefore, we conclude that modeling the average cluster density rather than a radial density profile is sufficient for the chosen inner region sizes.

### 4 DISCUSSION

The results in Section 3 have significant implications for both the distance to Wd1 and the distribution of field stars. The inferred length scale for the field stars, \( L \), is significantly different from the value used in the Bailer-Jones et al. (2018) catalog. The inferred \( L \) is 0.84 ± 0.02 kpc, while the model of Bailer-Jones et al. (2018) gives \( L = 1.38 \) kpc. The Bailer-Jones length scale is \( \sim 1.7 \) times larger than the inferred value from the outer ring. The discrepancy is not that surprising given that the Bailer-Jones et al. (2018) estimate for \( L \) was derived from a model of the Galaxy in the pre-Gaia era. Considering that Bailer-Jones et al. (2018) length scale was predicted before the era of accurate Gaia parallaxes, it is encouraging that the Bailer-Jones et al. (2018) model is only off by a little less than a factor of two. However, in this analysis this discrepancy significantly impacted our inference for the cluster parallax of Wd1. With an incorrect field-star distribution, the inferred cluster parallax gets skewed toward the average field-star parallax. Those inferences failed to produce any parallax other than the average parallax of the field stars. For the purpose of estimating priors for individual stars the Bailer-Jones et al. (2018) modeled length scale is most likely sufficient. However, if one is trying to reduce the uncertainty through the statistics of large samples of stars we find it important to model the field-star distribution for our sample.

The inferred distance to Wd1 is 3.2 ± 0.4 kpc, which represents the highest precision distance estimate for Wd1 to date. Historical estimates to Wd1 range from 1.2 to 5.5 kpc. (Crowther et al. 2006; Clark, J. S. et al. 2005; Piatti, A. E. et al. 1998; Westerlund 1968). Recently, Clark, J. S. et al. (2005) estimated that the distance to Wd1 ranged from 2 to 5.5 kpc. To infer this distance Clark, J. S. et al. (2005) noted that most of the bright stars associated with Wd1 are evolved stars and are not on the MS. Notable among these bright stars are yellow hypergiants (YHG). Noting their locations in the HR diagram from Smith et al. (2001a), Clark, J. S. et al. (2005) assumed that the YHGs may have a narrow range of luminosity or at least an upper limit to their luminosities. Given this presumed YHG luminosity range, and using the average observed brightness of YHGs in Wd1 combined with an adopted extinction and reddening law, Clark, J. S. et al. (2005) inferred a distance to Wd1 of 5.5 kpc. They noted, however, that this should be regarded as an upper limit, since the reddening law has trouble reproducing the color variations. Since there is no radio emission detected from WR winds, they also suggested a lower limit on the distance of 2 kpc. However, these constraints on the distance and reddening depend sensitively on assumptions about how the strengths of winds for various classes of evolved stars depend on luminosity, which is still often debatable (Smith 2014). Given the new Gaia DR2 distance, one may derive more accurate extinction and distance values for Wd1.

The bounds given by Clark, J. S. et al. (2005) corresponds to a factor of ~2.8 in distance; in contrast, the precision in the Gaia DR2 inferred distance is 15%. One can use the new distance to infer the fundamental parameters of the cluster such as luminosity, mass, and age via isochrone fitting. In this manuscript, we do not perform isochrone fitting. Instead, we estimate these fundamental parameters using two techniques. First, we scale previous estimates using the new distance, and we infer the luminosity, mass, and age of two bright stars in Wd1. The estimates of two bright stars provides a good proxy for the whole cluster.

The 15% precision in distance will lead to \( \sigma_L/L \approx 30\% \) precision in luminosity. For stars below about 20 \( M_\odot \), the
Figure 10. Posterior distribution for the cluster parallax. We report the mode and the highest density 68% confidence interval. The parallax of the cluster is $\varpi_{cl} = 0.31^{+0.04}_{-0.04}$ mas, which corresponds to distance of $R \sim 3$ kpc, the number density of inner region with 1’ radius is $n_{cl} = 97.62^{+4.85}_{-6.75}$/arcmin$^2$, the number density of field stars is $n_f = 35.95^{+0.74}_{-0.79}$/arcmin$^2$, and field star length scale is $L = 0.84^{+0.02}_{-0.03}$ kpc.

The new Gaia DR2 distance also provides strong constraints on the luminosity, mass, and age of cluster members. Wd1 hosts a diverse population of evolved massive stars such as WR stars, red and blue supergiants, YHG, an LBV, and a magnetar. Previous studies have inferred a turn-off mass of around 40 $M_{\odot}$ and cluster age of 4-5 Myr with a presumed distance of around 5 kpc \cite{Lim2013, Koumpia2012, Negueruela2010, Clark2005, Crowther2006, Ritchie2009}. These estimates were based on modeling the luminosity and temperatures of YHGs, RSGs, and WR stars. By association, this would imply that the magnetar progenitor had an initial mass of >40 $M_{\odot}$ \cite{Koumpia2012, Ritchie2009, Muno2006}.

Without reevaluating the bolometric and extinction corrections for each star, the new distance of 3.2$^{+0.4}_{-0.4}$ kpc reduces all luminosities by -0.38 dex as compared to a distance of 5 kpc. Using single-star stellar evolution models...
Figure 11. Bayesian inferred cluster parallax for each ring. Top and bottom dashed lines represent the highest density 68% confidence interval. All rings are consistent with the inference from all stars within 1′.

(Brott et al. 2011), we now infer the mass, age, and corresponding main sequence turn-off mass for two of the brightest stars in Wd1, the LBV W243 and a YHG4. The inferred $\log(L/L_{\odot})$ for W243 is $5.4 \pm 0.1$, and for YHG4 is $5.6 \pm 0.1$. The spectral type of W243 is B2I (to A2I) (Clark and Negueruela 2004, Westerlund 1987), and for YHG 4 is F2Ia+. This corresponds to temperatures of 9.17 kK (to 17.58 kK) and 7.2 kK, respectively. For these temperatures, using single-star stellar evolution models (Brott et al. 2011), the masses are $28.7^{+3.1}_{-3.7} M_{\odot}$ for W243 and $35^{+5.8}_{-5.7} M_{\odot}$ for YHG4, the corresponding ages are $5.56^{+1.34}_{-0.35}$ and $5.0^{+0.49}_{-0.68}$ Myr, respectively. While we did not fit isochrones, these ages would correspond to isochrones with main-sequence-turn-off masses of $25.9^{+2.2}_{-2.5} M_{\odot}$ (W243) and $32.1^{+0.4}_{-3.7} M_{\odot}$ (YHG4).

If we assume that LBV W243 is a representative of the cluster, then the age of the cluster is $5.56^{+1.34}_{-0.35}$ Myr, the turn-off mass is $25.9^{+2.2}_{-2.5} M_{\odot}$ (down from $40 M_{\odot}$), and the mass of the most evolved stars is $35^{+5.8}_{-5.7} M_{\odot}$. Figure 12 shows that the new inferred luminosity brings the LBV W243 to the lower edge of the S Doradus instability strip (Smith et al. 2004a).

However, Figure 12 also clearly shows that there are RSGs in Wd1 with implied initial masses below $20 M_{\odot}$, well below the presumed turnoff mass even at the nearer distance, and implied ages of around 10 Myr. This may suggest either uncertain bolometric corrections, a range of ages in Wd1, or may point to the influence of binary evolution on the evolved star population.

Most of the prior distances for Wd1 relied on measuring an apparent magnitude, assigning an absolute magnitude based upon the stellar type, and calculating the distance...
Figure 12. The HR diagram for evolved stars in Westerlund 1, including the LBV, W243. The open circles show the luminosities when $d = 5$ kpc, and the filled circles show the luminosities with the new Gaia DR2 distance of 3.21±0.4 kpc. All filled circles have the same uncertainty as W243. The orange symbols represent YHGs, the red symbols represent RSGs, the purple symbols represent WNs, and the green circles represent WCs (Crowther et al. 2006; Fenech et al. 2018). The gray boxes show the locations of the temperature dependent S Doradus instability strip (Wolf 1989) and the constant temperature strip of LBVs in outburst, as in Smith et al. (2004b). The new Gaia DR2 distance brings the LBV W243 to the lower edge of the S Doradus instability strip. YHGs and RSGs (Mengel and Tacconi-Garman 2007; Clark, J. S. et al. 2005) have a wide range of zero-age-main-sequence masses which could be due to errors in bolometric correction (Davies and Beasor 2018), variations in reddening, or could be due to binaries. The single-star model tracks (blue) are from (Brott et al. 2011). The evolutionary tracks do not reproduce the WR phases. The LBV W243 has an inferred mass of $28.7^{+3.1}_{-3.7}M_\odot$, an age of $5.56^{+1.34}_{-0.35}$ Myr. The brightest YHG has an inferred initial mass of around $35^{+3.8}_{-5.7}M_\odot$, an age of $5.0^{+0.49}_{-0.68}$ Myr.
would make Wd1’s initial mass comparable to or less than the mass of the Arches cluster in the Galactic center, although Wd1 would be significantly older (Harfst et al. 2010; Stolte et al. 2002; Kim et al. 2000).

5 CONCLUSION

We use Gaia DR2 parallax measurements and Bayesian inference to estimate the distance to the Westerlund 1 (Wd1) massive star cluster, as well as the distribution of field stars along the line of sight. We model both cluster stars and Galactic field stars, and we find that the cluster parallax is 0.31 ± 0.04 mas, which corresponds to a distance of 3.2 ± 0.4 kpc. We also infer that length scale, \( L \), for field stars to be 0.84 ± 0.02 kpc, which is ~1.7 times smaller than the models used by Bailey-Jones et al. (2018). The new distance represents the highest precision, 15%, to Wd1 to date. Much of this precision is limited by the systematics such as zero point and astrometric excess noise. Both are included in the Bayesian model. However, the models are rough and conservative, and require improvement in the future. For example, we model the full uncertainty as a combination of the expected parallax uncertainty and the excess noise. In addition, rather than using one zero-point value, we consider a distribution of zero points due to the observed variation of the zero-point parallax. Both models are rough estimates for these systematics. To further improve the parallax precision we either require better models for the systematics, or better calibration in subsequent data releases.

Wd1 has been discussed as potentially one of the most massive young star clusters in the Galaxy, but revising the distance to this one cluster reduces its total mass and increases its age, and may have profound consequences for stellar evolution theory. An improved distance can significantly narrow the precision on luminosity, mass, and age of the cluster, which provides constraints on the post-main-sequence evolution of cluster members. Based on the new Gaia distance, we infer turn-off mass of around 25 \( M_\odot \), which implies that the progenitor mass of the magnetar CXO J164710.2455216, and LBV W243 is a little bit above 25 \( M_\odot \).

ACKNOWLEDGEMENTS

Support for MA and JWM was provided by the National Science Foundation under Grant No. 1313036. This project was developed in part at the 2018 Gaia Sprint, hosted by the eScience and DIRAC Institutes at the University of Washington, Seattle. Support for NS was provided by NSF awards AST-1312221 and AST-1515559, and by the National Aeronautics and Space Administration (NASA) through HST grant AR-14136 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555.

REFERENCES

C. A. L. Bailey-Jones, J. Rybicki, M. Fousneau, G. Mantelet, and R. Andrei. Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2. 2015. 1508.8588. 2018.

W. Brandner, J. S. Clark, A. Stolte, R. Waters, I. Negueruela, and S. P. Goodwin. Intermediate to low-mass stellar content of Westerlund 1. 2008. 10.1051/0004-6361:20075779.

I. Brott, S. E. de Mink, M. Cantat, N. Langer, A. de Koter, C. J. Evans, I. Hunter, C. Trundle, and J. S. Vink. Rotating massive main-sequence stars: Grids of evolutionary models and initial conditions. 2011. 10.1051/0004-6361/201016113.

J. S. Clark and I. Negueruela. A newly identified Luminous Blue Variable in the galactic starburst cluster AS TroBJ, Westerlund 1/AS TroBJ, 2000. 10.1051/0004-6361:20002413.

S. Mengel and E. L. Tacconi-Garman. Medium revolution 2.5 instruction set architecture of the massive Galactic open cluster Westerlund 1. 2010. 10.1051/0004-6361/201013772.

Paul A. Crowther, L. J. Hadfield, J. S. Clark, I. Negueruela, and W. D. Vacca. A census of the Wolf-Rayet content in Westerlund 1 from near-infrared imaging and spectroscopy. 2009. 10.1111/j.1365-2966.2009.14925.x.

A. Damiani, L. A. Almeida, R. D. Blum, D. S. C. Damiani, N. Navarrete, M. S. Rubinho, and M. Teodoro. Downward limit in the parameter 0.44.8 m and the 8620Å line towards the stellar cluster Westerlund 1. Monthly Notices of the Royal Astronomical Society, 463(3):2653–2666. 2016. 10.1093/mnras/stw2122.

Ben Davies and Emma R. Beaulieu. The initial masses of the red supergiant progenitors to Type II supernovae. 2004. 10.1046/j.1365-8711.2004.08367.x.

ACKNOWLEDGEMENTS

Support for MA and JWM was provided by the National Science Foundation under Grant No. 1313036. This project was developed in part at the 2018 Gaia Sprint, hosted by the eScience and DIRAC Institutes at the University of Washington, Seattle. Support for NS was provided by NSF awards AST-1312221 and AST-1515559, and by the National Aeronautics and Space Administration (NASA) through HST grant AR-14136 from the Space Telescope Science Institute, which is operated by AURA, Inc., under NASA contract NAS5-26555.

REFERENCES

C. A. L. Bailey-Jones, J. Rybicki, M. Fousneau, G. Mantelet, and R. Andrei. Estimating Distance from Parallaxes. IV. Distances to 1.33 Billion Stars in Gaia Data Release 2. 2015. 1508.8588. 2018.

W. Brandner, J. S. Clark, A. Stolte, R. Waters, I. Negueruela, and S. P. Goodwin. Intermediate to low-mass stellar content of Westerlund 1. 2008. 10.1051/0004-6361:20075779.
A neutron star with a massive progenitor in Westerlund 1. The Astrophysical Journal Letters, 636(1):L41, 2006. URL http://stacks.iop.org/1538-4357/636/i=1/a=L41.

I. Negueruela, J. S. Clark, and B. W. Ritchie. The population of OB supergiants in the starburst cluster Westerlund 1. aap, 516:A78, June 2010. doi: 10.1051/0004-6361/201014032.

Ignacio Negueruela, J. Simon Clark, Lucy J. Hadfield, and Paul A. Crowther. Westerlund 1 as a Template for Massive Star Evolution. In F. Bresolin, P. A. Crowther, and J. Puls, editors, Massive Stars as Cosmic Engines, volume 250, pages 301–306, June 2008. doi: 10.1017/S1743921308020620.

Piatti, A. E., Bica, E., and Clariá, J. J. Fundamental parameters of the highly reddened young open clusters Westerlund 1 and 2 ***. Astron. Astrophys. Suppl. Ser., 127(3):423–432, 1998. doi: 10.1051/aas:1998111. URL https://doi.org/10.1051/aas:1998111.

A. G. Riess, S. Casertano, W. Yuan, L. Macri, B. Bucciarelli, M. G. Lattanzi, J. W. MacKenty, J. B. Bowers, W. Zheng, A. V. Filippenko, C. Huang, and R. I. Anderson. Milky Way Cepheid Standards for Measuring Cosmic Distances and Application to Gaia DR2: Implications for the Hubble Constant. apj, 861:126, July 2018. doi: 10.3847/1538-4357/aac82e.

B. W. Ritchie, J. S. Clark, I. Negueruela, and F. Najarro. Spectroscopic monitoring of the luminous blue variable Westerlund1-243 from 2002 to 2009. aap, 507:1597–1611, December 2009. doi: 10.1051/0004-6361/200912986.

B. W. Ritchie, J. S. Clark, I. Negueruela, and N. Langer. A VLT/FLAMES survey for massive binaries in Westerlund 1. I. Dynamical constraints on magnetar progenitor masses from the eclipsing binary W13. aap, 520: A48, September 2010. doi: 10.1051/0004-6361/201014834.

C. K. Rosslowe, P. A. Crowther, J. S. Clark, and I. Negueruela. Physical properties of the WR stars in Westerlund 1. In Wolf-Rayet Stars: Proceedings of an International Workshop held in Potsdam, pages 105–108, January 2015.

N. Smith. Mass Loss: Its Effect on the Evolution and Fate of High-Mass Stars. aara, 52:487–528, August 2014. doi: 10.1146/annurev-astro-081913-040025.

N. Smith, J. S. Vink, and A. de Koter. The Missing Luminous Blue Variables and the Bistability Jump. apj, 615:475–484, November 2004a. doi: 10.1086/424030.

Nathan Smith, Jorick S. Vink, and Alex de Koter. The Missing Luminous Blue Variables and the Bistability Jump. apj, 615:475–484, November 2004b. doi: 10.1086/424030.

K. G. Stassun and G. Torres. Evidence for a Systematic Offset of -80 μas in the Gaia DR2 Parallaxes. apj, 862:61, July 2018. doi: 10.3847/1538-4357/aacc9c.

A. Stolte, E. K. Grebel, W. Brandner, and D. F. Figer. The mass function of the Arches cluster from Gemini adaptive optics data. aap, 394:459–478, November 2002. doi: 10.1051/0004-6361:20021118.

B. Westerlund. A Heavily Reddened Cluster in ARA. aapl, 73:51, February 1961. doi: 10.1086/127618.

B. E. Westerlund. On the Extended Infrared Source in ARA. aapl, 154: L67, November 1968. doi: 10.1086/180270.

B. E. Westerlund. Photometry and spectroscopy of stars in the region of a highly reddened cluster in ARA. aaps, 70:311–324, September 1987.

B. Wolf. Empirical amplitude-luminosity relation of S Doradus variables and extragalactic distances. aap, 217:87–91, June 1989.

J. C. Zinn, M. H. Pinsonneault, D. Huber, and D. Stello. Confirmation of the zero-point offset in Gaia Data Release 2 parallaxes using asteroseismology and APOGEE spectroscopy in the Kepler field. ArXiv e-prints, May 2018.