CHARACTERIZING OPEN CLUSTERS AND SPECTROSCOPIC ECLIPSING BINARIES WITH

MACHINE LEARNING FRAMEWORKS

By

Karl Oskar Jaehnig

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Astrophysics

August 11, 2023

Nashville, Tennessee

Approved:

Dr. Kelly Holley-Bockelmann, Chair

Dr. Jonathan Bird

Dr. Adrian Price-Whelan

Dr. David Weintraub

This dissertation is dedicated to my mother, Cristina. I finally did it Mamá. I wish you could've seen me blossom.

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

<center>**CHAPTER 1**</center>

<center>**Background**</center>

1.1 Open Clusters

Open clusters (OCs) are single stellar populations that formed together. They were observed in antiquity by their close grouping of positions in the night sky. It is now possible today to group these stars as bonafide clusters based on their relative positions and proper-motions in the sky thanks to astrometric surveys such as Gaia. Currently thousands of OCs have been identified and are widespread throughout the Milky Way disk (Friel, 1995; Kharchenko et al., 2013; Dias et al., 2014; Cantat-Gaudin et al., 2018b; Hunt and Reffert, 2020). Studying OCs offers insights into theories of star formation (Kroupa, 2001) and provides a direct view of the high mass end of the stellar initial mass function. Their lower stellar density allows for clearer observations of stars as compared to the more crowded regions in globular clusters.

Open clusters serve as an important laboratory for astronomers to answer many unresolved questions in the field. OCs are sites of few-body dynamical mechanisms that can produce unique astronomical objects such as Blue Straggler stars (Perets, 2015; Toonen et al., 2020; He and Petrovich, 2017). Taken as a whole, the OC population are key tracers of chemical enrichment throughout the Galactic disk (Spina et al., 2021). The youngest OCs have been used to investigate the distribution and dynamics of the Milky Way spiral arms (Hao et al., 2021).

1.1.1 Identifying Open Cluster Populations

Distinguishing the members of an open cluster from the background of field stars can be difficult and numerous techniques have been developed to tackle the problem. Many techniques rely on finding the population of stars in proper-motion space. In proper-motion space, the population of stars that form an OC will appear as a concentrated grouping of similar proper-motions. This grouping can be identified straightforwardly for those open clusters whose proper-motions are high enough that they are separate and distinct from the galactic field stars (e.g., Melotte 25 (the Hyades) in Figure 1.1, top-panel). However, more embedded OCs, such as Melotte 20 (see Figure 1.1, bottom-panel), present a greater challenge.

With the advent of precise astrometric data from the Gaia survey (see Prusti et al., 2016; Brown et al., 2018), it is now possible to distinguish and computationally extract open clusters even when the kinematics of the cluster and field stars are not significantly different from one another. We briefly review some commonly employed methods here, but see Hunt and Reffert (2020) for a thorough and comprehensive review.

Once OC candidates have been selected and preliminary membership is assigned, it is still necessary

<center>1</center>
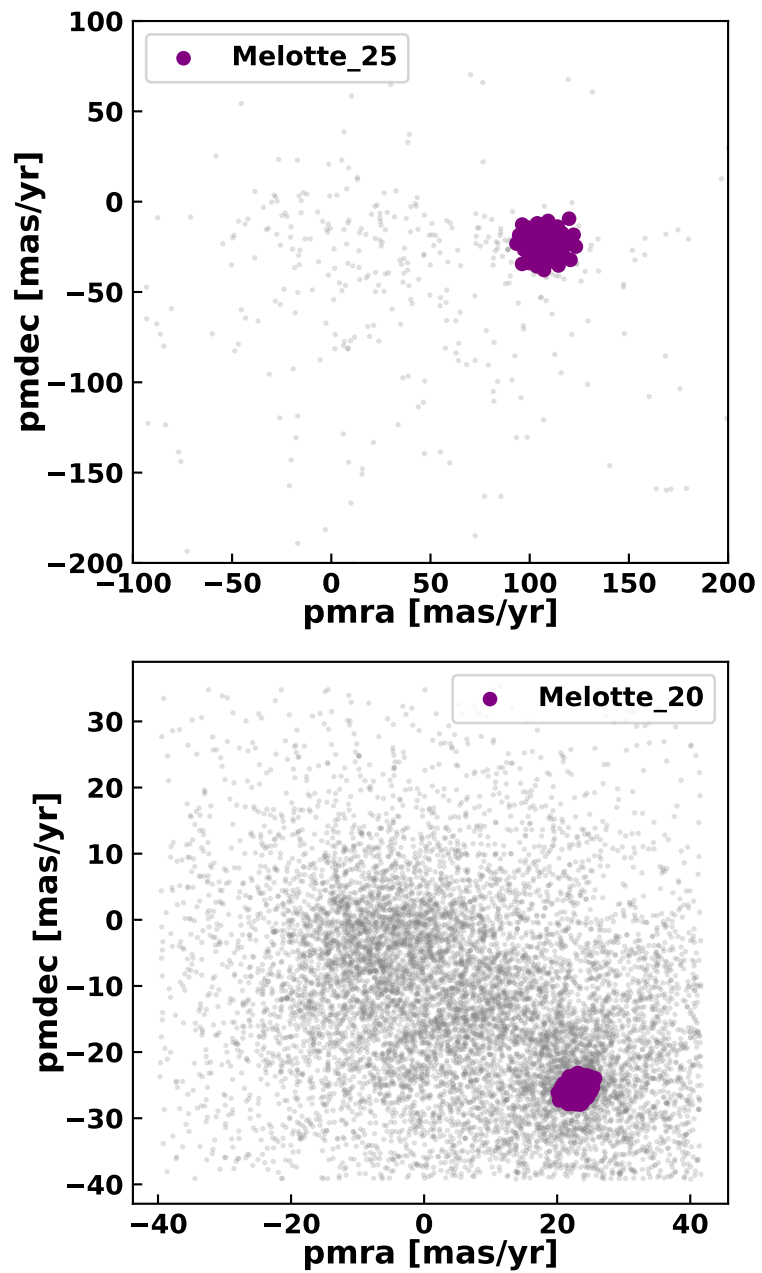
Figure 1.1: Proper-motion vector plot for Melotte-25, aka The Hyades Cluster (top panel, purple dots), and Melotte-20, aka the Alpha Persei Cluster (bottom panel, purple dots). Manually selecting open cluster members using vector point diagrams in proper-motion are easier (top panel) for some clusters than for more embedded clusters (bottom panel).

to ensure that the selected group of stars is a true physical cluster. One of the more reliable confirmation techniques for the overall cluster is to directly inspect the candidates in color-magnitude space. A group of stars with a narrow range of parallax can masquerade as a cluster within a CMD. However, false cluster members with erroneous parallax measurements will appear in unphysical locations within the CMD. Requiring member stars to be clustered in proper-motion as well as distance also limits false-positive cluster identification. As open clusters represent a single stellar population, the color-magnitude plot will exhibit well-known properties: the main sequence, the binary main sequence, and the red giant branch sequence if the cluster is sufficiently old. This technique becomes even more effective when previous values of the star cluster are known, such as metallicity ([Fe/H]), age, and extinction (A_V). A theoretical isochrone can then be generated to compare the observed colors and magnitudes of the stars with what would be expected from stellar evolutionary theory.

Perhaps the hardest aspect of such confirmation methods is that they are often performed manually for each individual OC. This can become a daunting task, considering that the Milky Way Star Cluster (MWSC) catalog Kharchenko et al. (2013) contains over 2000 open clusters. There have been many attempts at automating the detection and identification of open clusters within astrometric data sets in recent years. Cantat-Gaudin et al. (2018b) performed searches of open clusters using the UPMASK (Krone-Martins and Moitinho, 2014) algorithm, based on k-means clustering. Hunt and Reffert (2020) developed a thorough review of numerous clustering algorithms with respect to open clusters.

1.2 Binary Stars

There is a growing consensus that the majority of stars form in binary-star systems (hereafter "binaries") or higher order multiple systems. These systems are thought to have formed within fragmenting giant molecular clouds undergoing collapse due to loss of hydrostatic equilibrium. Binaries are a fundamental component of a plethora of astrophysical problems and are very common throughout the universe. About half of the stars in the Milky Way are found in binaries (Duchêne et al., 2013; Duchene et al., 2018), and the binary demographics and evolution of binaries are key unknowns in many astrophysical contexts (e.g., Price-Whelan et al., 2019; Rix et al., 2019).

However, binaries are also critical benchmark systems: for example, binaries provide one of the few ways to infer stellar distances without the use of stellar evolution models (Soderblom, 2010; Serenelli et al., 2021), and enable an important channel for calibrating models of stellar parameters across a range of stellar masses (Kraus and Hillenbrand, 2009; Stacy et al., 2009) as most binary-star components are assumed to be coeval in the field.

1.2.1 Eclipsing Binary Stars

A crucial class of benchmark binaries are eclipsing binaries (Serenelli et al., 2021) (EBs), in which the binary system orbit is aligned sufficiently to our line of sight for one star to pass in front of the other star. This transit causes a measurable dip in the brightness of the binary system. Most EBs will have two transits visible in their light curve; the primary eclipse (more massive primary star transits in front of the less massive secondary) and a secondary eclipse (less massive star transits in front of more massive primary). The eclipse shapes and timing provide information about both the orbital properties of the system, such as inclination, period, eccentricity, and orientation, and also intrinsic properties of the stellar components, such as the radius ratio and surface brightness ratio (e.g. Andersen, 1991; Torres et al., 2010b; Miller et al., 2020; Cunningham et al., 2020).

Both historically and continuing today, benchmark EB systems are often first identified from time-domain photometric observations and subsequently observed spectroscopically to obtain radial velocity time series of the systems (e.g. Baroch et al., 2022; Hong et al., 2022). With recent all-sky, high-cadence, time-domain photometric surveys such as the Kepler Mission (Prša et al., 2011) and the Transiting Exoplanet Survey Satellite (TESS; Ricker et al., 2014; Stassun et al., 2018), it has become possible to identify large samples of $\sim 10^4$ EB systems across a range of stellar types (Kirk et al., 2016; Prša et al., 2022). Especially useful EB systems are those with small secondary components (e.g., Torres, 2013; Gill et al., 2020; Sebastian et al., 2022) in which a star with well-predicted structure and parameters (e.g., a 1 $M_\odot$ primary star) has a low-mass companion (e.g., a 0.3 $M_\odot$ star) with much more uncertain intrinsic and atmospheric parameters (Kraus et al., 2011; Cassisi and Salaris, 2019).

1.2.2 Spectroscopic Binary Stars

Spectroscopic binaries (SBs), like eclipsing binaries, require some alignment in the orbit with respect to our line of sight to be observed. Spectroscopic binaries can offer different constraints on orbital parameters depending on whether the spectra encodes a detection of one (SB1) or two (SB2) stars. However, due to a degeneracy between the inclination and semi-major axis, SB1s alone can only provide population level inferences of binary demographics (Badenes et al., 2018; Price-Whelan et al., 2020). SB1s, coupled with available ancillary data, are ideal to perform analyses of binary-star field and cluster populations. In addition to this SB2s do not typically sample the entire mass-ratio range due to observed SB2s typically being two near equal mass stars. Observing a binary system both as an EB as well as an SB is the ideal case. It is possible to determine both stellar and orbital parameters to high precision for both stars in a binary only when photometric measurements of EBs are coupled with spectroscopically determined radial velocities of SBs.

1.3 Data

1.3.1 Astrometric data from Gaia DR2

The Gaia satellite (Prusti et al., 2016) has greatly broadened our understanding of stars throughout the Milky Way. Since launching in December 2013, Gaia's repeated measurements of billions of stars across the night sky have produced one of the best and largest astrometric catalogues to date. Now in its third data release DR3 (Gaia Collaboration et al., 2022), there are more than 1.8 billion stars with 5 dimensional (α, δ, ϖ, $\mu_{\alpha*}$, μ_{δ}) astrometric solutions with micro-arcsecond uncertainties, and approaching a faint limit of magnitude 21.5. This catalog has become an indispensable component at numerous scales in the astronomical community from wide binaries to stellar streams throughout the Milky Way.

1.3.2 Time-series flux measurements from the TESS Satellite

The successor to the Kepler spacecraft (Batalha et al., 2010), the Transit Exoplanet Survey Satellite (hereafter, TESS) (Ricker et al., 2014; Sullivan et al., 2015; Stassun et al., 2017, 2019) is now in its 5th year of operation in the second extended mission. The time-domain photometric data from TESS has helped make breakthroughs in the field of exoplanets and in the study of binaries because of its high-cadence, nearly all-sky-coverage observations of over 200,000 stars, covering almost 70% of the night sky. TESS observes targets continuously within a particular sector over a span of 13.5 with a short 2-minute cadence; each sector has at least 27 total days of observing time.

The resulting data contains a large number of high quality time-series light curve flux data for each star in the TESS Input Catalog (Stassun et al., 2017). Having this continuous, short-cadence light curve data enables studying periodic behavior present in the observed stars over many timescales (e.g., stellar flares (Günther and Daylan, 2021), stellar rotation (Doyle et al., 2020), stellar pulsations (Cunha et al., 2019), and stellar/sub-stellar companions (Prša et al., 2022; Gandolfi et al., 2018, respectively). Pre-processed TESS light curves from the TESS-SPOC processing group (Caldwell et al., 2020) provide an easily accessible data set of time-series flux data that can be harnessed to characterize EBs.

1.3.3 Time-series radial velocities from the APOGEE Spectrograph

The Sloan Digital Sky Surveys (SDSS) have been in operation for nearly 20 years. The recently released 17th data release from SDSS (Abdurro'uf et al., 2022) marks the final data release in SDSS-IV (Blanton et al., 2017), and contains a full release of the APOGEE-2 spectroscopic data (spectral resolution $R \sim 22,500$) (Majewski et al., 2017; Beaton et al., 2021; Santana et al., 2021). The APOGEE-2 data set contains over ~ 2.6 million individual spectra for over 700,000 stars. The APOGEE project was originally designed to only perform spectroscopic chemical analysis of stars, which required multiple visits to accumulate a high

enough signal to noise ratio. With multiple spectra and long time baselines, these data can be repurposed to provide time series radial velocity measurements (e.g., Troup et al., 2016; Badenes et al., 2018; Price-Whelan et al., 2018).

1.4 Methods

1.4.1 Distribution-based Clustering with GMM

Gaussian Mixture Models (GMMs) (Pearson, 1893) describe a data set using a number of multivariate Gaussian components. GMM is a 'top-down' method that attempts to represent a data-set as a mixture of multivariate Gaussians. An individual Gaussian, or component, of the model represents the cluster. GMMs have already been used to identify and characterize a number of astronomical objects, including individual open clusters (Gao, 2018), the galaxy red sequence (Hao et al., 2009), and supernovae/host galaxy populations (Holoien et al., 2016).

GMMs allow for mixed memberships, wherein a data point can have a probability of belonging to multiple components. K-means and DBSCAN typically assign each data point only to one component or cluster. GMMs also have fewer hyperparameters to tune than DBSCAN or other similar agglomerative clustering methods. The resulting covariance matrices from a well-fit GMM effectively describe the shape, scaling, and orientation of components within the data. Determining the right number of cluster components to employ for any particular GMM fit is important to draw accurate conclusions. However there are numerous ways to determine the best number of GMM components *a posteriori*, such as the statistical metrics like the Bayesian Information Criterion or methods such as component number cross-validation.

As Cabrera-Cano and Alfaro (1990) cautions in their analysis of ~ 300 stars in NGC-2420, care must be taken when fitting an OC dataset with GMMs. If the field star fraction is too large, the OC may be overwhelmed by field stars along the line of sight. An open cluster may be also be less separable if the distribution of field stars is strongly non-Gaussian. Both of these situations may produce less than optimum results, although a quantitative analysis of the effect of these potential issues on a given cluster was not addressed.

1.4.1.1 Extreme Deconvolution GMM

It is ideal for a GMM to fit the intrinsic, underlying distribution rather than noisy discrete data. To accomplish this the observed measurements must first be deconvolved from their uncertainties Bovy et al. (2011) developed a method to determine the underlying distribution function in the presence of noisy data called 'Extreme Deconvolution'.

Extreme Deconvolution Gaussian Mixture Models (hereafter, XDGMM) combine the two techniques to

ideally fit the true, intrinsic distribution with a Gaussian mixture model. XDGMM has already been employed successfully to characterize members of a Galactic open cluster (Olivares et al., 2019) and to calculate mean cluster positions and proper-motions in the case of an actively disrupting open cluster (Price-Whelan et al., 2019). However, these use cases of XDGMM were for a small number of individual systems.

1.4.2 Bayesian Inference

The most frequently used methods to determine the orbital parameters usually rely on some generative model (such as a two-body keplerian orbit) which takes a vector of parameters which can be called 'θ'. This generative model is incorporated into a likelihood function which computes the similarity of the model synthetic data with the observed data using a statistical metric such as the χ^2 parameter. The likelihood function can be maximized in order to infer the orbital parameters, that coupled with the generative model, best approximate the observed data.

However this process, simply referred to as 'Maximum Likelihood' (hereafter ML), has a number of shortcomings, such as: ML infers the mode of a distribution which is not as robust to skewed distributions as the median of the distribution. ML does not directly generate uncertainties on the parameter modes inferred. ML does not take into account any relevant prior information that may help better infer the desired parameters.

We can correct for these issues by using Bayes Theorem (Bayes, 1763) to perform our inference which is written out as $P(\theta|D) = P(D|\theta) \cdot P(\theta)/P(D)$. Here $P(D|\theta)$ is our likelihood function and $P(\theta)$ is our prior function. $P(D)$ is called the 'Bayesian Evidence' and while it is an important quantity, it is also usually very computationally expensive to calculate and primarily useful in comparing different models. As our framework currently only uses one orbital model, it can be safely ignored for now.

The Bayes theorem equation is fundamentally an integral over the whole of the parameter space we are interested in. Fully evaluating the integral is a computational intractable endeavor. However we can evaluate this equation numerically using Markov Chain Monte Carlo (hereafter, MCMC). MCMC is an iterative way to evaluate $P(\theta|D)$ (known as the posterior distribution) by drawing samples for θ. The probabilities of the values in θ are calculated in the prior. These values are then used in our generative model and compared with the data in the likelihood function.

1.4.3 Hamiltonian Monte Carlo

Most MCMC methods rely on the 'Metropolis-Hastings' algorithm (otherwise known as 'random-walk') (Metropolis et al., 1953) to draw samples to construct the posterior distributions. Most of the samples drawn throughout the MCMC process are correlated and thus most MCMC evaluations rely on running a chain of samples long enough such that sufficient uncorrelated samples are drawn to properly approximate the target

posterior distribution. It is also possible to run N multiple chains of samples with N multiple starting values of θ, known as 'walkers', in order to more effectively explore the parameter space concurrently.

Unfortunately, the efficacy of MCMC relies on the presumption of properly exploring the parameter space from which the distributions of the variable posteriors can be inferred. MCMC algorithms that rely on a random walk to explore a parameter space usually suffer from an incapacity to identify whether effective exploration of the target distribution is being achieved. Characterizing an MCMC run as 'converged' then typically rely on post-MCMC sampling techniques such as calculating the auto-correlation length of samples to determine if the number of samples drawn is enough; this process may need to be repeated several times to properly derive posteriors for very geometrically complicated problems.

Hamiltonian Monte Carlo (Duane et al., 1987) (hereafter HMC) draws from the properties of Hamiltonian physics to exploit the differential geometry of a parameter space, even if its non-physical in nature, in order to efficiently explore that parameter space in building posterior distributions. HMC methods rely on calculating the gradients within a parameter space in order to explore it efficiently (see Neal, 2012).

HMC can still be potentially undermined if the two hyper-parameters, step size (ε) and number of steps (L), are not set adequately. The No U-Turn Sampling (NUTS) algorithm (Hoffman and Gelman, 2014) was developed to allow researchers to bypass the usually time consuming process of performing multiple iterations of sampling runs with varying ε and L values to determine how best to explore a parameter space. The NUTS algorithm actively adjusts ε and L during sampling as well as terminating sampling when the algorithm confirms that the parameter exploration is beginning to double-back onto itself in order to maintain computational efficiency.

<div align="center">

CHAPTER 2

</div>

<div align="center">

Membership lists for 431 open clusters in Gaia DR2 using extreme deconvolution Gaussian Mixture Models

</div>

Abstract [1]

Open clusters are groups of stars that form at the same time, making them an ideal laboratory to test theories of star formation, stellar evolution, and dynamics in the Milky Way disk. However, the utility of an open cluster can be limited by the accuracy and completeness of its known members.

Here, we employ a "top-down" technique, *extreme deconvolution gaussian mixture models* (XDGMM), to extract and evaluate known open clusters from Gaia DR2 by fitting the distribution of stellar parallax and proper motion along a line-of-sight. Extreme deconvolution techniques can recover the intrinsic distribution of astrometric quantities, accounting for the full covariance matrix of the errors; this allows open cluster members to be identified even when presented with relatively uncertain measurement data. To date, open cluster studies have only applied extreme deconvolution to specialized searches for individual systems.

We use XDGMM to characterize the open clusters reported by Ahumada and Lapasset (2007) and are able to recover 420 of the 426 open clusters therein (98.1%). Our membership list contains the overwhelming majority ($> 95\%$) of previously known cluster members. We also identify a new, significant, and relatively faint cluster member population and validate their membership status using Gaia eDR3.

We report the fortuitous discovery of 11 new open cluster candidates within the lines of sight we analyzed. We present our technique, its advantages and challenges, as well as publish our membership lists and updated cluster parameters.

2.1 Introduction

Open clusters (OCs) are single stellar populations that formed together. Thousands of OCs have been identified and are widespread throughout the Milky Way disk (Friel, 1995; Kharchenko et al., 2013; Dias et al., 2014). Studying OCs offers insights into theories of star formation (Kroupa, 2001) and provides a direct view of the high mass end of the stellar initial mass function. OCs are also sites of few-body dynamical mechanisms that can produce unique astronomical objects such as Blue Straggler stars (Perets, 2015; **?**; He and Petrovich, 2017). Taken as a whole, the OC population are key tracers of chemical enrichment throughout the Galactic disk (Spina et al., 2021).

Identifying the members of a particular open cluster can be difficult and numerous techniques have been

---

<div align="center">

9

</div>

developed to tackle the problem. Many techniques rely on finding the population of stars in proper-motion space. In proper-motion space, the population of stars that form an OC will appear as a compact group of similar proper-motions. This grouping can be identified straightforwardly for those open clusters whose proper-motions are high enough that they are separate and distinct from the galactic field stars (e.g., Melotte 25 (the Hyades) in Figure 1.1, top-panel). However, more embedded OCs, such as Melotte 20 (see Figure 1.1, bottom-panel), present a greater challenge.

With the advent of precise astrometric data from the Gaia survey (see Prusti et al., 2016; Brown et al., 2018), it is now possible to distinguish and computationally extract open clusters even when the kinematics of the cluster and field stars are not significantly different from one another. We briefly review some commonly employed methods here, but see **?** for a thorough and comprehensive review.

Cluster and cluster member identification algorithms that use astrometric data leverage the compact nature of a cluster in some combination of position, distance, and proper motion. Most techniques employ what could be called a 'bottom-up' approach in which the star cluster is built up out of smaller components until all the stars within a line of sight are either considered a 'cluster' star or a 'field' star.

Algorithms such as K-means (Utsunomiya et al., 1996), fit a K number of clusters to data by iteratively finding K mean centroids until some convergence threshold is met. Other algorithms, such as DB-SCAN (Daszykowski and Walczak, 2009), assign groupings of stars to clusters using inter-particle distance. One benefit of these algorithms is that they are not confined to proper-motion data, and have been successfully applied to position and multi-dimensional astrometric data (Gao, 2018; Castro-Ginard et al., 2018). Both DBSCAN and K-means have been used to identify stars in known open clusters, but they have also been successfully employed to find new open cluster candidates (**?**Liu and Pang, 2019; Castro-Ginard et al., 2019).

Despite its power and versatility, K-means algorithms necessitate pre-defining the number of clusters that are to be found. This may be difficult to do without explicit prior knowledge of how the data are distributed along the dimensions of interest. Workflows using K-means also typically assume that the data are all distributed spherically, a clear limitation when presented with more realistic cluster shapes (**?**). DBSCAN is more generalized than K-means in that it can find non-symmetric clusters distributed throughout the data, and also doesn't require a preset number of clusters. However, it becomes computationally expensive when considering large data sets in more than 3 dimensions, as all of the pairwise distances between data points must be calculated. The hyperparameters used in DBSCAN to define clusters also needs to be fine-tuned, requiring hyperparameter optimization which can become computationally expensive.

Once OC candidates have been selected and preliminary membership is assigned, it is still necessary to ensure that the selected group of stars is a true physical cluster. One of the more reliable confirmation techniques for the overall cluster is to directly inspect the candidates in color-magnitude space. A group

of stars with a narrow range of parallax can masquerade as a cluster within a CMD. However, false cluster members with erroneous parallax measurements will appear in unphysical locations within the CMD. Requiring member stars to be clustered in proper-motion as well as distance also limits false-positive cluster identification. As open clusters represent a single stellar population, the color-magnitude plot will exhibit well-known properties: the main sequence, the binary main sequence, and the red giant branch sequence if the cluster is sufficiently old. This technique becomes even more effective when previous values of the star cluster are known, such as metallicity ([Fe/H]), age, and extinction (A_V). A theoretical isochrone can then be generated to compare the observed colors and magnitudes of the stars with what would be expected from stellar evolutionary theory.

Perhaps the hardest aspect of such confirmation methods is that they are often performed manually for each individual OC. This becomes a daunting task, considering that the Milky Way Star Cluster (MWSC) catalog Kharchenko et al. (2013) contains over 2000 open clusters.

There have been great strides in automating the process of open cluster membership construction to be able to take a census of the Milky Way population.

Cantat-Gaudin et al. (2018a) recently performed one of the largest OC automated analyses in Gaia DR2, deriving mean cluster parameters (α, δ, $\mu_{\alpha*}$, μ_δ, ϖ, r_{core}, etc) as well as individual membership lists for 1229 OCs throughout the Milky Way. Cantat-Gaudin and Anders (2019) updated these results and grew their catalog to 1481 clusters.

Their method relies on finding all sub-clumps defined by running a K-means algorithm in their chosen sight lines. They then use the minimum spanning tree metric, Λ (Allison et al., 2009), to distinguish between concentrated star cluster sub-clumps and uniformly distributed sub-clumps of the same number that are found in the field. This was done iteratively to construct their final open clusters out of these sub-clumps. Finally, they produce membership probabilities by re-sampling the 3x3 covariance matrices of the stars and re-running their algorithm.

'Top-down' cluster identification algorithms, which aim to characterize the environment that includes the cluster, offer an alternative to the 'bottom-up' approach. Gaussian Mixture Models (GMMs) (**?**) describe a data set using a number of multivariate Gaussian components. GMM is a 'top-down' method that describes a line of sight (or subset thereof) as a mixture of multivariate gaussians. An individual gaussian, or component, of the model represents the cluster. GMMs have already been used to identify and characterize a number of astronomical objects, including individual open clusters (Gao, 2018), the galaxy red sequence (Hao et al., 2009), and supernovae/host galaxy populations (Holoien et al., 2016).

GMMs allow for mixed memberships, wherein a data point can have a probability of belonging to multiple components. K-means and DBSCAN typically assign each data point only to one component or cluster.

GMMs also have fewer hyperparameters to tune than DBSCAN or other similar agglomerative clustering methods. The resulting covariance matrices from a well-fit GMM effectively describe the shape, scaling, and orientation of components within the data.

As Cabrera-Cano and Alfaro (1990) cautions in their analysis of ~ 300 stars in NGC-2420, care must be taken when fitting an OC dataset with GMMs. If the field star fraction is too large, the OC may be overwhelmed by field stars along the line of sight. An open cluster may be also be less separable if the distribution of field stars is strongly non-gaussian. Both of these situations may produce less than optimum results, although a quantitative analysis of the effect of these potential issues on a given cluster was not addressed.

It is ideal for a GMM to fit an intrinsic, underlying distribution rather than noisy discrete data. To accomplish this the observed measurements must first be deconvolved from their uncertainties Bovy et al. (2011) developed a method to determine the underlying distribution function in the presence of noisy data called 'Extreme Deconvolution'.

Extreme Deconvolution Gaussian Mixture Models (hereafter, XDGMM) combine the two techniques to ideally fit the true, intrinsic distribution with a gaussian mixture model. XDGMM has already been employed successfully to characterize members of a Galactic open cluster (**?**) and to calculate mean cluster positions and proper-motions in the case of an actively disrupting open cluster (Price-Whelan et al., 2019). However, these use cases of XDGMM were for a small number of individual systems.

In this paper, we use XDGMM to identify members of known open clusters within Gaia DR2. We construct an automated pipeline to address the shortcomings of GMM cluster fitting – computational cost, the number of gaussian components to fit, and selection of the gaussian component associated with the cluster.

In Section 3.2 we discuss the Gaia DR2 data. In Section 2.3 we describe our method to extract open clusters. We present the mean cluster parameters and the properties of designated cluster members as well as a validation of our results via comparison with the literature in Section 3.4. In Section 2.5, we describe the serendipitous discovery of 11 new candidate open clusters not found in the current literature. Finally, we summarize our findings on open cluster extraction using an automated machine learning method and discuss future work in Section **??**.

2.2 Data

We first assemble a list of open clusters to consider using the cluster catalog from Ahumada and Lapasset (2007), originally created in their search for blue straggler stars. They constructed a catalog of 427 open clusters throughout the Milky Way, using Johnson UBV photometry (Johnson and Knuckles, 1955) and isochrones (Girardi et al., 2000). We remove one known asterism (NGC-1252 Kos et al., 2018; Angelo et al.,

2019; Cantat-Gaudin and Anders, 2019) from consideration. This reduces the total number of open clusters we consider to 426.

The effectiveness of any search for or characterization of open clusters will depend upon the input data. Gaia DR2 contains 5 dimensional astrometric measurements – position (α, δ), proper motion ($\mu_{\alpha*}$, μ_δ), and parallax (ϖ) – and photometry for over 1.3×10^9 sources (Lindegren et al., 2018). The quality of this data set makes it ideal for open cluster searches while its quantity presents a challenge. Attempting a global XDGMM fit over the entire sky is computationally prohibitive and would yield a suboptimal result even if possible (Section 2.3).

For each cluster, we define a target field of potentially associated Gaia DR2 sources. The target field of view is centered on the position of the cluster and has conservatively-wide opening angle to ensure that we do not remove potential cluster members and affect our subsequent analysis. While Ahumada and Lapasset (2007) report both cluster positions and angular sizes, we supplement their measurements with the more recent open cluster catalog efforts of Dias et al. (2018) and Kharchenko et al. (2013). The addition of these catalogs and variation amongst them also captures the uncertainty of these input cluster parameters in our analysis. Our target field center is the median of the positions reported in these cluster catalogs. The opening angle is equal to $1.5\times$ the largest cluster angular diameter reported amongst Ahumada and Lapasset (2007), Dias et al. (2018), and Kharchenko et al. (2013).

We searched for all Gaia DR2 sources within the 426 target fields of view. We obtained the astrometric [α, δ, $\mu_{\alpha*}$, μ_δ, ϖ] and photometric [G, G_{BP}, G_{RP}] measurements, as well as the relevant uncertainties and correlation coefficients. In addition to the Gaia DR2 data, we collected all previous distance determinations to the clusters in our sample through the Simbad (Wenger et al., 2000) database. These distance measurements will be used in a pre-processing step described in Subsection 2.3.1. We used the AstroQuery python module (Ginsburg et al., 2019) to query both Gaia DR2 and Simbad.

2.3 Methods

Ultimately, we identify a cluster and its members via a XDGMM fit of the density field represented by standardized $\mu_{\alpha*}$, μ_δ, and ϖ measurements and uncertainties of stars in the cluster vicinity. Our procedure's computational cost and its effectiveness correlates with the density contrast between the cluster and field. We address both of these concerns by first restricting each target field to a manageable subsample.

The position information of the subsample of stars associate with each cluster is only used for validation purposes (Section ??). We make this decision based on the work of Hillenbrand and Hartmann (1998), Kuhn et al. (2014), and Kuhn et al. (2017). These works considered clustering of star clusters using their observed on-sky positions. They found that a multivariate normal distribution may not be best fit to these distributions.

We scale the measurements and uncertainties of interest to account for the different variance of each feature before applying the XDGMM fit. Finally, we establish quantitative metrics to select the number of components to include in the XDGMM fit and to select the individual component corresponding to the cluster. The details of our method follow.

2.3.1 Data Preprocessing

The Gaia DR2 data is known to have systematics that, if unaccounted for, can lead to erroneous or biased results (Vasiliev, 2019). We use the recommendations from Lindegren et al. (2018) to account for systematics found within the parallax and proper-motion data. In particular, we subtract the -0.029 mas parallax zero-point from all of our stars and exclude all stars fainter than 18th Gaia G magnitude. Our faint limit reduces the total number of stars included in our analysis. Stars with $m_G > 18$ typically have high fractional astrometric uncertaines in Gaia DR2 making them ill-suited for cluster-finding (e.g. Cantat-Gaudin and Anders, 2019; Hunt and Reffert, 2020). We also correct for the inertial spin found within the proper-motions of bright sources[2] using the recommendations from Lindegren (2019).

After applying the global corrections and limits, we preprocess the input data within each cluster target field independently. The goal is to remove obvious field stars and astrometric outliers from consideration as cluster members. This reduces the dynamic range of parallax and proper motion and, in turn, increases the signal of any over-density associated with the cluster.

We first remove outlier stars in proper motion. Stars with $\mu_{\alpha*}$ or μ_{δ} measurements more than 10σ away from the median of their target are dropped from consideration. Individual high proper-motion stars may skew the feature scaling within a target field. This step is performed for all clusters except for the high proper-motion cluster Melotte-25 (the Hyades cluster).

We remove parallax outliers only if more than 10^4 stars in a target field remain eligible for XDGMM fitting. We found that this numerical threshold strikes a good balance between ensuring a complete cluster membership list given our magnitude limit and maintaining computational efficiency of the XDGMM fitting. If a target sample contains $> 10^4$ stars, we discard the stars least likely to be at the distance of the known cluster.

The open clusters have uncertain literature distances and there is uncertainty measured parallax of each star. For these reasons, we do not simply discard stars using a strict delta parallax cut-off. Rather, we employ a simple probabilistic model to determine the stars least likely to be within the literature parallax range of the cluster.

We assume the parallax of cluster members is distributed normally with a mean μ and standard deviation

---

[2]Gaia DR2 $G < 11$

14

σ given by:

$$\mu = med(\varpi_{lit}), \quad \sigma = \begin{cases} 0.25 \cdot med(\varpi_{lit}), & \text{if } med(\varpi_{lit}) > 1 \text{ mas} \\ 0.25 \text{ mas}, & \text{if } med(\varpi_{lit}) < 1 \text{ mas} \end{cases} \quad (2.1)$$

where $med(\varpi_{lit})$ is the median of the cluster parallaxes reported in the three cluster catalogs sources we examine (Section 3.2). We set σ to depend on $med(\varpi_{lit})$. Clusters with a literature parallax less than 1 mas (distance greater than 1 kpc) have a constant σ_{ϖ} of 0.25 mas. Clusters with a literature parallax greater than 1 mas have their σ_{ϖ} set to a 25% of their median literature parallax. This ensures that the parallax distribution of nearby clusters is not excessively truncated.

We use bootstrap sampling to calculate the probability that each stellar parallax measurement is drawn from the cluster parallax distribution. For each star, we sample a normal distribution with mean $\mu = \varpi_{obs}$ (*gaiadr2.parallax*) and $\sigma = \sigma_{\varpi_{obs}}$ (*gaiadr2.parallax_error*) 1000 times. We then evaluate the PDF of the corresponding cluster, defined in equation 2.1, at the locations of these 1000 samples. The star's probability of parallax overlap is then the mean of the 1000 PDF evaluations.

Only the 10^4 stars with the highest parallax overlap probability within the target field are kept and subject to XDGMM fitting. After this step, all target fields have been reduced to a maximum of 10^4 stars based on the closeness of their parallax and proper motion to the parent cluster. All of the analysis and XDGMM fit procedure described below pertain only to these restricted data.

2.3.2 Scaling the $\mu_{\alpha*}$, μ_{δ}, ϖ data

Feature scaling is necessary to establish an appropriate shape of the parameter space as we are fitting 3-dimensional data with unequal means and variance. We use 'StandardScaler' from the `Scikitlearn` (**?**) python package, which shifts and scales the data in each to dimension to a mean of zero and a variance of one.

We first smooth the data before scaling to ensure we calculate the scaling parameters accurately. The smoothed distribution is a simple resampling of the [$\mu_{\alpha*}$, μ_{δ}, ϖ] distribution according to the uncertainty covariance matrix (equation 2.2). We resample each star $N_{samples}$ times such that StandardScaler uses approximately 5×10^4 coordinate positions (i.e., $N_{samples} \times N_{data} = \sim 5 \times 10^4$, where N_{data} is the number of stars remaining in the field after pre-processing). The shift and scale parameters found for the smooth distribution are then applied to the actual measurements prior to fitting. We scale the measurement uncertainties as well, to preserve the fractional uncertainty of each measurement in the transformed data space.

2.3.3 Extreme Deconvolution Gaussian Mixture Models

Gaussian Mixture Models employ the assumption that any N-dimensional data can be represented as the summation of K components, each of which is an N-dimensional Gaussian. The K Gaussian distributions have K means, and K covariances.

XDGMM takes into full account the full covariance matrix representing the uncertainty on $\mu_{\alpha*}$, μ_δ, and ϖ, and the correlation ρ for each data point. This covariance matrix is assembled as written in equation 2.2.

$$\begin{bmatrix} \sigma_{\mu_{\alpha*}}^2 & \sigma_{\mu_{\alpha*}}\sigma_{\mu_\delta}\rho_{\mu_\alpha\mu_\delta} & \sigma_{\mu_{\alpha*}}\sigma_\varpi\rho_{\varpi\mu_{\alpha*}} \\ \\ \sigma_{\mu_\delta}\sigma_{\mu_{\alpha*}}\rho_{\mu_\alpha\mu_\delta} & \sigma_{\mu_\delta}^2 & \sigma_{\mu_\delta}\sigma_\varpi\rho_{\varpi\mu_\delta} \\ \\ \sigma_\varpi\sigma_{\mu_{\alpha*}}\rho_{\varpi\mu_{\alpha*}} & \sigma_\varpi\sigma_{\mu_\delta}\rho_{\varpi\mu_\delta} & \sigma_\varpi^2 \end{bmatrix} \tag{2.2}$$

We perform an XDGMM fit of the scaled astrometric data and the associated scaled covariance matrices using the XDGMM python wrapper (Holoien et al., 2016). Each cluster target field is fit independently.

2.3.4 The Optimal Number of Mixture Components

We fit each cluster-specific scaled data set a total of nine times, using from 2 to 10 Gaussian components. We do not fit more than 10 components to keep the model fitting computationally efficient. All model fits converge within a tolerance of 1×10^{-8} and use a random seed of 999.

Of the nine XDGMM fits to a cluster, we select the best-fit model according to the Bayesian Information Criterion (BIC Ghosh, Jayanta K. et al., 2006). The BIC is commonly used to select between models when fitting to a data set as it strikes a balance between maximizing the likelihood and penalizing the introduction of new parameters to avoid over-fitting. The BIC is defined as,

$$\text{BIC} = x \cdot \log n - log\hat{L}, \tag{2.3}$$

where x is the number of parameters needed in the XDGMM fit, n is the number of data points, and $\hat{L}$ is the maximized likelihood of the model. The best-fit model has the lowest BIC score.

We find that the optimal number of components in the best-fit model is typically between 5 and 8 and correlates with the number of stars in the target field. Heavily populated target fields ($>$5000 stars) generally required more components in the best-fit model than sparsely populated fields .

2.3.5 Identifying Open Clusters with Information Theory

We use the differential entropy information metric to select the gaussian component that best describes the open cluster. Differential entropy is a measure of how compact a distribution is within a volume. In this case, we can ascertain the compactness of each gaussian component in the scaled data space. For a multivariate gaussian, the differential entropy (h) can be written as,

$$h = \frac{d}{2} + \frac{d}{2}ln(2\pi) + \frac{1}{2}ln(|\Sigma|),$$ (2.4)

where d is the dimension of the parameter space and Σ is the determinant of the covariance matrix for the multivariate gaussian of each component in the best-fit XDGMM (see Ahmed and Gokhale, 1989).

We automatically designate the component with the lowest differential entropy as the cluster component.

2.3.5.1 Individual membership probabilities

Once we have determined the optimal number of components and which component belongs to the open cluster, we employ bootstrap resampling to calculate the open cluster membership probabilities of each star in the scaled target field. We assign each star within a target field to a component using the XDGMM code's built-in 'predict' function. We then generate a new scaled $[\varpi, \mu_{\alpha*}, \mu_{\delta}]$ value for each star from its scaled 3x3 covariance matrix. We recompute component assignments for each star for a total of 100 iterations. The final membership probability of any one star is the number of times that the star was assigned to the open cluster component out of 100.

2.3.5.2 Validation of fit results

We visually inspect the resulting open cluster model results by plotting four different characteristics of the cluster stars with a membership probability above 50%. We plot the proper-motion of the stars in α and δ the G and G_{BP} - G_{RP} color-magnitude diagrams, the positions in l and b, and ϖ as a function of G-band magnitude. We show these four panels for the cluster NGC-6583 in Figure 2.1 as an example.

We define a successful open cluster retrival, in the broadest sense, via visual validation. Statistical tests can be found in Section 3.4. We retrieve a known open cluster when the resulting members (probability >50%) are concentrated in proper-motion vector space, form a relatively well-defined isochronal color-magnitude sequence, are centered within the target field on-sky, and occupy a narrow distribution in parallax as a function of G-band magnitude.
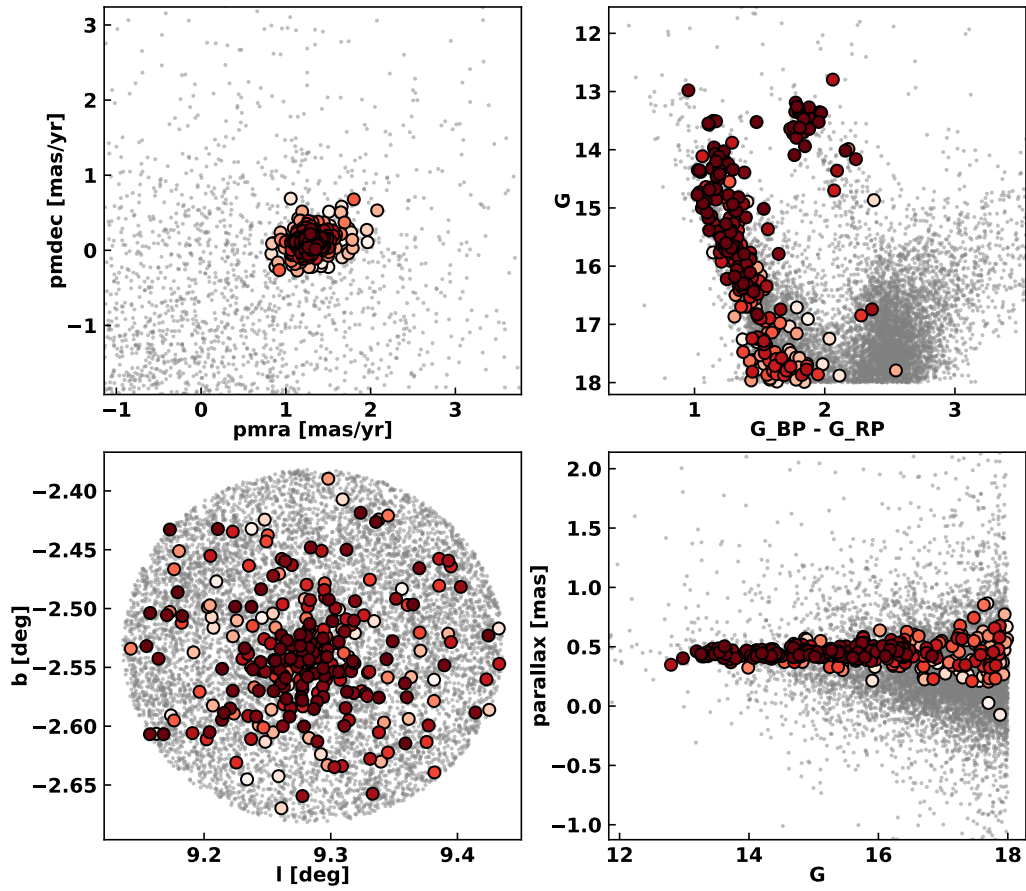
Figure 2.1: An example four panel plot used for visual validation of a fully automatic XDGMM extraction. The open cluster is NGC-6583. In all four panels, the field is given by gray dots, and the cluster stars are given by the red circles, where the lower membership probability stars tend towards white in color. The top left panel is a vector point diagram plot in proper motion. The top right panel is a color-magnitude diagram using Gaia filters. The bottom left panel is a plot of positions in galactic coordinates. The bottom right panel is a plot of the parallax measurements for cluster stars as a function of their Gaia G-band magnitude.

2.4 Results

We successfully recover 420 of the 426 known open clusters targeted. The probable members of these iden-
tified clusters are highly concentrated in proper-motion and parallax space relative to field stars and populate
a narrow isochrone in the color-magnitude diagram (Section 2.3.5.2). We statistically validate our cluster
membership designations and the astrometic location of the recovered clusters in Section 2.4.2.

Our automated pipeline to select the best-fit XDGMM and identify the individual Gaussian component
corresponding to the cluster target required no human intervention in 95% of the investigated target fields. In
these 385 fields, the lowest differential entropy (h) component of the best-fit model represented the targeted
cluster. The remaining 41 fields required manual analysis of XDGMM fits. We found the target cluster was
fit by the second lowest h component of the best-fit model in 29 of the manual analysis fields. In nearly all
($28/29$) of these target fields, the minimum h component was either a spatially adjacent known open cluster
or a new open cluster candidate (Section 2.5). All told, the minimum h component of the best-fit XDGMM
identified a cluster or candidate cluster in 413 of the 426 target fields.

None of the best-fit model components fit describe the known cluster in remaining 12 target fields. We
visually inspected all 54 ($\sum_{i=2}^{10} i$) components within the 9 BIC-scored model fits to each of these target fields
in search of a component that corresponded to the known open cluster. We identified a Gaussian component
that represented the open cluster in 6 of these target fields.

In the remaining 6 target fields, we were unable to recover a cluster candidate component either auto-
matically or manually despite convergence of the XDGMM code. At least 2 of these 6 targeted clusters
likely violate our assumption of the open cluster being a significant over-density in the [$\mu_{\alpha*}$, μ_δ, ϖ] space.
Ruprecht-46 may be an asterism masquerading as a cluster, as was recently argued by Cantat-Gaudin et al.
(2020). Collinder-228 is a well-known cluster in the Carina Nebula, but Feigelson et al. (2011) found that the
cluster is a composition of many sparse groups with no clear central concentration.

Nevertheless, our automated pipeline was successful in a variety of conditions. Within the 385 target
fields with a fully automated and successful target cluster recovery:

1. The target field contains a total number of stars as few as 123 stars and as many as 1×10^4 stars

2. The number of cluster member stars (with a membership probability > 0.5) was as few as 40 stars and
   as many as ~ 3700 stars

3. Cluster members comprised a wide range, $0.64\% - 54.5\%$, of all stars in the target field

This work is, to the best of our understanding, the first to report central parameters using Gaia DR2 for
some of the target clusters. We identify 11 clusters that are not included in the Gaia DR2 based cluster

Name	α	δ	ϖ	σ_ϖ	$\mu_{\alpha*}$	$\sigma_{\mu_{\alpha*}}$	μ_δ	σ_{μ_δ}	$N_{0.85}$	r_{50}	Dm_{16}	Dm	Dm_{84}	R_{GC}	Z_{GC}
	deg	deg	mas	mas	mas/yr	mas/yr	mas/yr	mas/yr		deg	kpc	kpc	kpc	kpc	kpc
Berkeley-42	286.31	1.90	0.09	0.11	-2.84	0.26	-5.95	0.25	236	0.02	10.46	11.041	11.62	6.54	-0.43
Bochum-1	96.33	19.90	0.23	0.06	-0.19	0.08	-0.43	0.07	29	0.13	4.47	4.58	4.69	12.81	0.32
Bochum-14	270.52	-23.70	0.31	0.09	0.25	0.13	-1.06	0.10	107	0.13	3.05	3.08	3.11	5.25	-0.01
Bochum-7	131.10	-45.97	0.18	0.03	-3.09	0.10	3.65	0.15	59	0.19	5.40	5.48	5.56	10.32	-0.16
Collinder-96	97.62	2.84	0.12	0.07	0.13	0.29	0.38	0.36	57	0.11	7.89	8.25	8.61	16.05	-0.43
NGC-1931	82.80	34.24	0.45	0.05	0.37	0.12	-1.91	0.19	27	0.09	2.22	2.26	2.29	10.55	0.04
NGC-2467	118.18	-26.37	0.19	0.03	-2.55	0.08	2.58	0.11	154	0.09	5.09	5.16	5.24	11.59	0.08
NGC-3247	156.71	-57.93	0.32	0.11	-6.49	1.20	3.02	0.60	225	0.05	2.96	2.98	3.00	8.08	0.01
NGC-6514	270.65	-22.90	0.85	0.04	0.33	0.32	-1.68	0.20	113	0.23	1.16	1.17	1.17	7.14	0.02
Trumpler-24	254.01	-40.47	0.84	0.02	0.49	0.62	-1.80	0.54	19	0.49	1.19	1.20	1.20	7.15	0.06
Trumpler-27	264.08	-33.49	0.49	0.01	-0.15	0.08	-1.28	0.06	20	0.20	2.00	2.03	2.06	6.28	-0.01
XDOCC-01	19.90	58.35	1.26	0.06	-1.68	0.11	-0.34	0.06	22	0.17	0.78	0.79	0.80	8.80	-0.03
XDOCC-02	84.05	34.30	0.80	0.05	0.28	0.11	-2.29	0.07	26	0.14	1.21	1.26	1.31	9.57	0.06
XDOCC-03	107.01	-13.05	0.88	0.04	-3.32	0.36	0.55	0.13	25	0.21	1.12	1.14	1.15	8.95	-0.02
XDOCC-04	118.83	-24.17	0.33	0.04	-2.71	0.08	3.06	0.09	60	0.21	2.84	3.02	3.22	10.06	0.14
XDOCC-05	121.75	-29.90	0.60	0.02	-3.66	0.05	3.56	0.03	19	0.12	1.59	1.63	1.68	9.06	0.07
XDOCC-06	130.93	-48.15	0.77	0.04	-6.38	0.11	3.61	0.09	33	0.06	1.27	1.29	1.32	8.47	-0.05
XDOCC-07	159.48	-58.72	2.28	0.04	-14.32	0.23	0.93	0.23	21	0.15	0.43	0.43	0.43	8.01	0.02
XDOCC-08	270.76	-22.73	0.73	0.05	0.67	0.11	-2.57	0.08	51	0.12	1.32	1.34	1.37	6.79	0.01
XDOCC-09	293.60	25.00	1.83	0.04	-0.46	0.10	-8.74	0.37	47	0.66	0.54	0.55	0.55	8.04	0.05
XDOCC-10	305.42	37.29	0.51	0.06	-2.39	0.11	-5.23	0.12	40	0.16	1.88	1.91	1.95	7.87	0.03
XDOCC-11	317.73	47.72	0.47	0.04	-1.01	0.05	-1.74	0.05	21	0.10	1.99	2.06	2.13	8.54	0.02

Table 2.1: Cluster parameters for 11 open clusters confirmed in this work and 11 newly discovered open clusters. All σ values in this table are median absolute deviations (MAD) multiplied by the factor 1.4826 to approximate a standard deviation. The 'dmode$_{16}$' and 'dmode$_{84}$' values are the 16th and 84th percentiles of the distance posterior distribution. 'dmode' is the 50th percentile of the distance posterior distribution. The full table of 431 cluster parameters will be made available electronically.

catalogs of Cantat-Gaudin et al. (2018a), Cantat-Gaudin and Anders (2019), Castro-Ginard et al. (2018), Castro-Ginard et al. (2019) Castro-Ginard et al. (2020), Ferreira et al. (2020), Sim et al. (2019), Liu (2019), or Hunt and Reffert (2020). Cluster parameters for these systems can be found in Table 2.1.

Our cluster member catalog contains 261 662 stars with non-zero membership probability (P_{mem}). In our subsequent analysis, we define members, probable members, and highly probable members of a cluster as those stars with P_{mem} greater than 0.5, 0.85, and 0.95, respectively.

The member catalog of the 420 recovered open clusters includes 171 939 cluster members, 125 235 probable members, and 95 920 highly probable members.

2.4.1 Cluster Astrometric Parameters and Distances

We determine robust central astrometric coordinates [α, δ, ϖ, $\mu_{\alpha*}$, μ_δ] of each cluster using the measurements of probable members (Table 2.1).

To do so, we first remove potential outlier measurements via a sigma-clipping procedure. We calculate

an initial coordinate median and median absolute deviation (MAD), both weighted by inverse measurement uncertainty, of probable members. We estimate σ assuming a normal distribution using $\hat{\sigma} = 1.4826 \times \mathrm{MAD}$ (e.g., Huber, 1980). We discard all outlier measurements not within ± 3 σ of the initial weighted median.

We then recalculate the weighted median and MAD of the non-outlier stars for each of the 5 astrometric parameters. We list these median values and the standard deviation estimator $\hat{\sigma}$ in Table 2.1. We also calculate the r_{50} radius, containing half of the cluster members, using the cluster positions and list these values in 2.1.

Cluster distance, while not a direct astrometric measurement, can be ascertained to a relatively high precision because each cluster member can be assumed to lie at nearly the same distance along the line of sight. Simple parallax inversion is not adequate to measure physical distances using Gaia DR2. This is due to the nature of the Gaia DR2 parallax measurements which may be close to zero or even negative. We obtain individual cluster distances through Bayesian Inference (see Luri et al., 2018, for more details).

The posterior probability of the cluster distance can be expressed as,

$$P\big(r|\{\varpi\},\{\sigma_\varpi\},L\big) \propto P\big(r|L\big)P\big(\{\varpi\}|r,\{\sigma_\varpi\}\big), \tag{2.5}$$

where $P(r|L)$ is the exponential decreasing space density prior (**?**):

$$P(r|L) = \begin{cases} \frac{1}{2L^3}r^2 e^{-r/L}, \text{ if } r > 0 \\ \\ 0, \text{ otherwise} \end{cases} \tag{2.6}$$

and $\mathrm{P}(\{\varpi\}|r,\{\sigma_\varpi\})$ is the likelihood of observing the set of parallax measurements given a true cluster distance r and the set of measurement uncertainties and is given by,

$$P(\{\varpi\}|r,\{\sigma_\varpi\}) = \prod_{i=1} \frac{1}{\sqrt{2\pi\sigma_{\varpi_i}^2}} exp\bigg(-\frac{(\varpi_i - \frac{1}{r})^2}{2\sigma_{\varpi_i}^2} \bigg). \tag{2.7}$$

We set the length scale 'L' within our prior to be 1 kpc and compute the posterior over a distance range of [0.01, 30] kpc.

We sample the posterior distribution and compute the 16th, 50th, and 84th percentile distances for all 420 open clusters and the open cluster candidates (Section 2.5). Table 2.1 contains our distance measurements to the 11 newly confirmed open clusters in Gaia DR2 and our 11 new open cluster candidates. The full table of all 420 open clusters mean astrometric parameters and inferred distances will be made available electronically.

We acknowledge that the range between the 16th and 84th distance percentiles may be artificially narrow. Equation 2.7 implicitly assumes that all cluster members reside at a singular distance rather than occupy a

small, realistic range in distance along the line of sight. The assumption greatly simplifies the calculation and is relatively accurate for distant clusters. However, our treatment may not be the most suitable for very nearby open clusters such as Melotte-25 (the Hyades, $\varpi \sim 21$ mas), and Melotte-22 (the Pleiades, $\varpi \sim 7$ mas).

2.4.2 Verification of our catalog results

To ascertain some measure of the veracity of our results, we first compare our central cluster parameters and membership lists to previously published membership catalogs. In addition, we use the more recent Gaia eDR3 data (Brown et al., 2020) to determine if the cluster members unique to this work are actual cluster stars or field contamination.

2.4.2.1 Comparison with previous results in the literature

We can further scrutinize our results by comparing the mean cluster parameters for our sample with those clusters which have entries within the literature. We plot the empirical cumulative distribution functions(eCDF) of the residuals between each cluster's parameter calculated in this work with the literature parameters in Figure 2.2.

The difference in ϖ, $\mu_{\alpha*}$, and μ_δ was computed for 392 clusters this work has in common with Cantat-Gaudin et al. (2020). We find that the difference in ϖ, $\mu_{\alpha*}$, and μ_δ is quite small, with median values of 0.031 mas, -0.002 mas yr^{-1}, and 0.004 mas yr^{-1}, respectively. The offset of 0.031 mas in the ϖ appears to come from our application of the -0.029 mas zero-point offset (see Section 2.3.1). Removing this zero-point offset from the cluster ϖ values leads to a median difference of only 0.002 mas with the literature values. This level of agreement with a study using a completely different cluster identification algorithm is strong evidence that we successfully recover the central location of our target clusters.

We find good agreement between our inferred cluster distances and those from Cantat-Gaudin and Anders (2019). The median difference in distance is -0.003 kpc. The range between the 84th and 16th percentiles in the distance discrepancy is only 0.036 kpc.

We also find good agreement in r50, the radius containing half the cluster members. The median difference in r50 between the results of Cantat-Gaudin and Anders (2019) and our own is 0.003 deg. We do report much smaller r50 values for about 10% of clusters (Figure 2.2, bottom left, x <-0.10 deg). Upon investigation, we find that these clusters, according to Cantat-Gaudin et al. (2020), extend beyond the aperture radius employed in this work to define the input target field (Section 3.2). Thus, the cluster members recovered in this work may populate as wide a range in position for the clusters with large angular sizes. However, based on the agreement in cluster mean ϖ, $\mu_{\alpha*}$, and μ_δ, this does not appear to affect our ability to recover the bulk of the clusters in question.
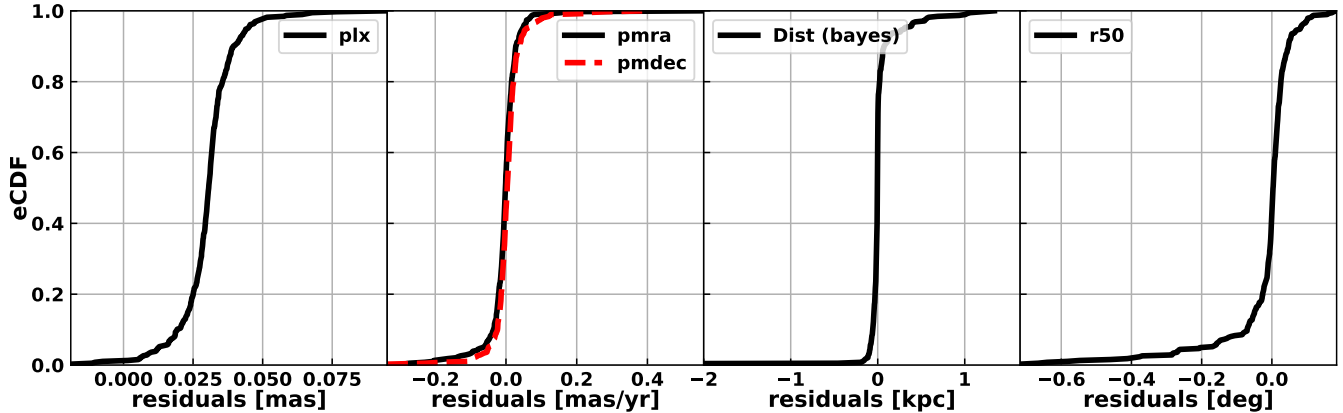
Figure 2.2: (In order from left to right) Comparison empirical cumulative distribution (eCDF) plots of the mean cluster ϖ, $\mu_{\alpha*}$(black solid line) and μ_δ(red dashed line), distance from Bayesian inference (black solid line), and on-sky size containing half of all cluster members (r50). The ϖ, $\mu_{\alpha*}$, μ_δ, distance from Bayesian inference, and r50 eCDF plots are comparing the 386 clusters this work has in common with Cantat-Gaudin and Anders (2019). There is overall excellent agreement between mean open cluster parameters in our catalog with the values in the literature from different methods.

The general agreement in cluster mean parameters suggests that our cluster member lists have significant overlap with those of Cantat-Gaudin et al. (2020). We compare our catalog with the catalog of Cantat-Gaudin et al. (2020) using Gaia DR2 source ids at a membership probability of $> 25\%$, $> 50\%$, and $> 95\%$. We correct the aperture size to match our target field aperture.

At the lowest level of membership probability we find our that our cluster member list contains 98.7% of the members reported by Cantat-Gaudin et al. (2020) at the same membership probability according to their method. This overlap decreases to 96.1% at the next membership level. The overlap decreases to 81.1% at the highest level of membership.

The overlap in cluster members is a strong argument that our cluster member designation is largely accurate, since our method to construct the cluster membership lists differs from that of Cantat-Gaudin et al. (2020). The alternative, that our discrepant methods share the same false positives, appears unlikely.

We find that member stars in our catalog that are not found in Cantat-Gaudin et al. (2020) are fainter than the member stars shared in common between our two catalogs. The median Gaia DR2 G-band magnitude of common member stars with $P_{\mathrm{mem}} > 50\%$ is 15.67. The unique member stars found in this work have a median Gaia DR2 G-band magnitude of 16.77, a difference of 1.1 magnitudes. At $P_{\mathrm{mem}} > 95\%$, the median Gaia G-band of unique member stars identified in this work is still 0.6 magnitudes fainter than the highly probable members common to both catalogs. This could indicate that the XDGMM fit is more sensitive to the fainter stars in each target field, given the full treatment of each star's individual covariance matrix.

2.4.2.2 Hertzsprung-Russell Diagrams of our Open Clusters

Open clusters populate isochrones within the Hertzsprung-Russell Diagram (HRD). The HRD is a well-known figure that relates a star's evolutionary state to its color and absolute magnitude, in lieu of temperature and luminosity measurements, respectively. We can check the reliability of our method of selecting open cluster groups by placing them on the HRD and checking to see the bulk population properties of our member catalog.

We calculate the absolute magnitudes of each cluster star prior to placing it on the HRD. The majority of the open clusters considered in this work reside within the galactic disk of the Milky way, occupying the galactic latitude (b) range [-20,20]. We need to correct for reddening and extinction due to galactic dust, as well as distance, to be able to place the cluster member stars on the HRD accurately.

Dust extinction and distance has already measured for 392 clusters in our recovered sample of 420 (93.3%) by Cantat-Gaudin et al. (2020). In Cantat-Gaudin et al. (2020) a neural network was developed and trained on CMDs of synthetic clusters constructed from 347 real cluster parameters in the literature, the bulk of which came from Bossini et al. (2019).

We use the polynomial fit equations of Babusiaux et al. (2018) to determine the Gaia DR2 passband extinction values, A_G, A_{BP}, A_{RP} for all stars in the 392 clusters with A_V measurements from Cantat-Gaudin et al. (2020). However, some stars have poor photometry such that the Gaia extinction coefficients could not be calculated. We circumvent this by selecting a high quality sample of stars from each cluster. We select the stars within each cluster with a probability > 0.85. We also exclude all stars with *gaiadr2.phot_bp_rp_excess_factor* $>1.5+0.03\cdot(G_{BP} - G_{RP})^2$ as this is indicative of large systematic errors within the G_{BP} and G_{RP} colors (Riello et al., 2018). The median A_G, A_{BP}, A_{RP} values of these high quality cluster stars is then used to de-extinct and de-redden stars all other stars in each cluster. This approach allows us to apply dust corrections to all probable cluster members, even those with more problematic photometry. However, this means we do not account for differential reddening present in very distant or heavily extincted open clusters.

We plot the extinction-corrected absolute G band magnitude as well as the de-reddened G_{BP} - G_{RP} color of cluster members in Figure 2.3. We plot the HRDs of stars with membership probabilities greater than 50% (left panel), greater than 85%(middle panel), and greater than 95% (right panel). We also color code the stars by their parent cluster age as predicted in Cantat-Gaudin et al. (2020).

Our HRD contains various groups at different stages of stellar evolution, such as the main-sequence, main-sequence turn off, and red-giant branch. All of these features are clearly visible even at the 50% membership probability threshold(Figure 2.3, left panel). We also recover 12 white dwarfs as high probability cluster members with this application of XDGMM. 5 are apparent members of the Beehive cluster (NGC-2632), 3

are in the Hyades cluster (Melotte-25) while the Coma star cluster (Melotte-111), the Pleiades (Melotte-22), Graff's cluster (IC-4756), and the Pincushion cluster (NGC-3532) each have 1 white dwarf.

The overwhelming majority of the cluster member stars appear in physically sensible regions of the HRD. While it is true that a group of stars sharing the same parallax will form a diagonal sequence within the HRD, our general and close agreement with the mean cluster parameters of previous studies (Section 2.4.2.1) is good evidence that we do not claim to find a known cluster at an incorrect astrometric position. Instead, the lack of stars at unphysical locations in the diagram, even at $P_{mem} > 50\%$, suggests that our membership classification does not suffer from significant field contamination due to erroneous or uncertain parallax measurements. Stars with intrinsic parallax different from the target cluster would appear above or below the main locus.

The clear age gradient and features in the HRD offer more evidence that our cluster member designations are largely accurate. The main sequence turn-off in older clusters in clearly more red and dim than in younger clusters. If we erroneously labeled all stars near the cluster parallax as cluster members, the correlation between cluster age and turn-off position may have been weakened by field interlopers. In addition, we only find stars along the pre-main-sequence in the youngest clusters. The age labels of each cluster are taken from Cantat-Gaudin et al. (2020), who use an entirely different method of determining cluster membership. The cluster ages shown in Figure 2.3 are completely decoupled from our cluster membership designations, yet the age-dependent features above remain.

Our cluster membership probabilities are calculated independent of any photometry, but we find a correlation between membership probability and position on the HRD relative to the expected locations predicted by cluster isochrones. A small percentage of stars appear between the main sequence and white dwarfs; others are redder than the red giant branch. These outliers are far less prominent in the $P_{mem} > 95\%$ panel. This correlation is clearly apparent for stars just below the main sequence ($G_{BP} - G_{RP} \approx 1$ and $G \approx 7.5$; $G_{BP} - G_{RP} \approx 0$ and $G \approx 2.5$), colder than the main sequence turnoff, and with intermediate colors relative to the red clump and upper main sequence.

2.4.2.3 Membership Testing with Gaia eDR3

Gaia eDR3 (Brown et al., 2020) yields an opportunity to validate our cluster membership lists. The Gaia eDR3 astrometric data is nominally at higher precision than Gaia DR2 and can serve as a check on the quality of our cluster stars memberships. Specifically, we can statistically test for cluster membership by comparing the Gaia DR2 and eDR3 parallax measurements of the same star.

We first assume that the observed parallax does not suffer from systematic uncertainty nor bias and is normally distributed about the true parallax in both Gaia eDR3 and DR2. The observed parallax in either data
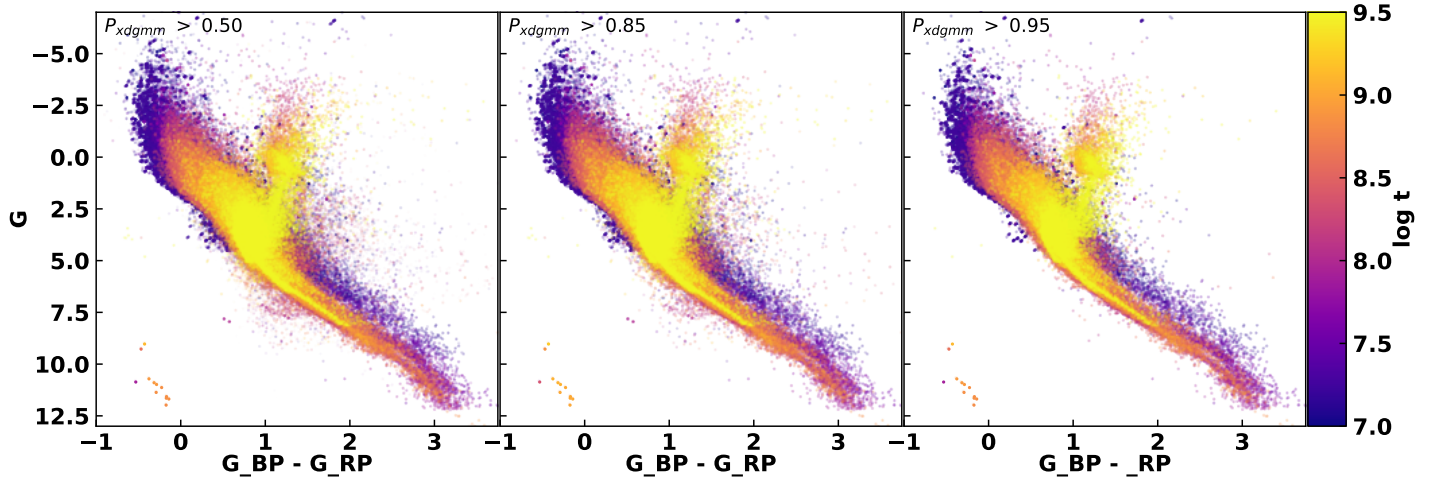
Figure 2.3: Hertzsprung - Russell diagram of all member stars within the 392 open clusters recovered in this work that also have predicted ages, extinctions, and distances in Cantat-Gaudin et al. (2020). All stars plotted have a G_{BP} - G_{RP} excess factor less than $1.5 + 0.03 \cdot (G_{BP} - G_{RP})^2$. Stars with at least a 0.25 membership probability are plotted in the left panel, the 0.50 membership level stars are plotted in the middle panel, and the 0.95 membership level The stars are plotted in the right panel. The three Gaia passbands (G, G_{BP}, G_{RP}) have been corrected for extinction and reddening using the methods described in **?**. We find that our use of XDGMM is able to properly capture multiple phases of open cluster evolution with no apparent biases.

release is then (as in Luri et al., 2018),

$$p(\varpi_j | \varpi_{\text{true}}) \propto \mathcal{N}(\varpi_{\text{true}}, \sigma_{\varpi_j}) \tag{2.8}$$

where the subscript $j \in$ [DR2, eDR3] refers to the specific data release, ϖ_{true} is the true parallax of the star, $\mathcal{N}(\mu, \sigma^2)$ represents a normal distribution with mean μ and variance σ^2, and σ_{ϖ_j} is the uncertainty of the parallax measurement in each data release .

The relationship between the parallax measurements in both Gaia data sets and the parallax (distance) of the cluster will be different for cluster and field stars. The true distance of a cluster corresponds to a singular cluster parallax, $1/d_{\text{clus}} = \varpi_{\text{clus}}$, if we ignore the small intrinsic width of the cluster. The observed parallax of a star relative to the cluster parallax is simply $\delta_{\varpi_j} = \varpi_j - \varpi_{\text{clus}}$, where j again identifies the specific Gaia data release. For a bonafide cluster member, $\varpi_{\text{true}} = \varpi_{clus}$ and thus one expects the eDR3 parallax measurement to be closer to the true cluster parallax; i.e., $|\delta_{\varpi_{\text{eDR3}}}| < |\delta_{\varpi_{\text{DR2}}}|$ if $\sigma_{\varpi_{eDR3}} < \sigma_{\varpi_{DR2}}$. However, $\varpi_{\text{true}} \neq \varpi_{clus}$ for field stars in which case we expect $|\delta_{\varpi_{\text{eDR3}}}| \sim |\delta_{\varpi_{\text{DR2}}}|$.

The *difference* between Gaia eDR3 and DR2 measurements of the observed parallax relative to the cluster parallax, $\delta_{\varpi_{\text{eDR3}}} - \delta_{\varpi_{\text{DR2}}}$ should therefore be different for cluster and field stars. We define this change in relative parallax $\Delta\varpi$ for a single star in terms of the relative Gaia DR2 parallax measurement as,

$$\Delta\varpi = \delta\varpi_{\mathrm{eDR3}} - \delta\varpi_{\mathrm{DR2}}$$

$$= \varpi_{\mathrm{eDR3}} - \varpi_{\mathrm{DR2}}$$

$$= \mathscr{N}(0,1)\sigma_{\varpi_{\mathrm{eDR3}}} - \mathscr{N}(0,1)\sigma_{\varpi_{\mathrm{DR2}}} \qquad (2.9)$$

$$= \mathscr{N}(0,1)\sigma_{\varpi_{\mathrm{eDR3}}} - (\delta\varpi_{\mathrm{DR2}} - \varpi_{\mathrm{true}} + \varpi_{\mathrm{clus}})$$

where the final two equalities make use of $\mathscr{N}(\mu,\sigma^2) \equiv \mathscr{N}(0,1)\sigma + \mu$ and our definition of $\delta\varpi_{\mathrm{DR2}}$.

When the true parallax of a star is different from ϖ_{clus}, as is the case for field stars, Δ_ϖ will have no dependence on $\delta\varpi_{\mathrm{DR2}}$ as $\delta\varpi_{\mathrm{DR2}} \propto \varpi_{\mathrm{true}} - \varpi_{\mathrm{clus}}$. Cluster stars, however, will show a strong correlation between Δ_ϖ and $\delta\varpi_{\mathrm{DR2}}$. Since $\varpi_{\mathrm{true}} = \varpi_{\mathrm{clus}}$ for cluster members, $-\delta\varpi_{\mathrm{DR2}} = -\mathscr{N}(0,1)\sigma_{\varpi_{\mathrm{DR2}}}$ and thus, $\Delta_\varpi \propto -\delta\varpi_{\mathrm{DR2}}$.

Figure 2.4 shows the distribution of $\Delta\varpi$ and $\delta\varpi_{\mathrm{DR2}}$ for the field stars and member stars in NGC-188 line-of-sight. We identified field stars as those with $< 1\%$ membership probability; $\Delta\varpi$ has little or no dependence on $\delta\varpi_{\mathrm{DR2}}$ for these stars (filled contours). Bonafide cluster members (open line contours) show a strong, negative correlation between $\Delta\varpi$ and $\delta\varpi_{\mathrm{DR2}}$. The *bonafide* cluster member designation applies to those stars with cluster membership $> 50\%$ in both our catalog and that of Cantat-Gaudin et al. (2020). Using this definition, we are confident that nothing inherent to our cluster membership analysis is driving the correlation seen in Figure 2.4.

We statistically assess the relationship shown in Figure 2.4 across all target fields using simple linear regression[3]. In each target field, we fit a line to $\Delta\varpi$ vs. $\delta\varpi_{\mathrm{DR2}}$ for three distinction populations – field stars, bonafide cluster members, and unique cluster members. Unique cluster members are probable members ($> 85\%$ membership probability) according to our analysis but have zero cluster membership probability according to Cantat-Gaudin et al. (2020) (see the black points in Figure 2.4 for an example). We attempt to match all stars considered in our XDGMM fits with its Gaia eDR3 counterpart using the crossmatch provided by the Gaia team (*gaiaedr3.dr2_neighbourhood*). To ensure a clean match, we demand that the eDR3 target is within 0.15 mas of the DR2 source on the sky and that the eDR3 and DR2 Gaia magnitudes do not differ by more than 0.1 mag.

The resulting catalog contains $\sim 1.7 \times 10^6$ stars. Along each line of sight, we assume the true cluster parallax is the median observed parallax of the bonafide cluster members. We require at least 20 bonafide members to ensure a robust measurement of ϖ_{clus} and at least 15 members of any other group to perform the fitting procedure. Finally, the intercept of the linear fit must be near zero to guard against significant differences in the observed parallax systematics between the two data releases (see below for details). These

---

[3]as implemented by CURVE_FIT, a part of the SCIPY python package
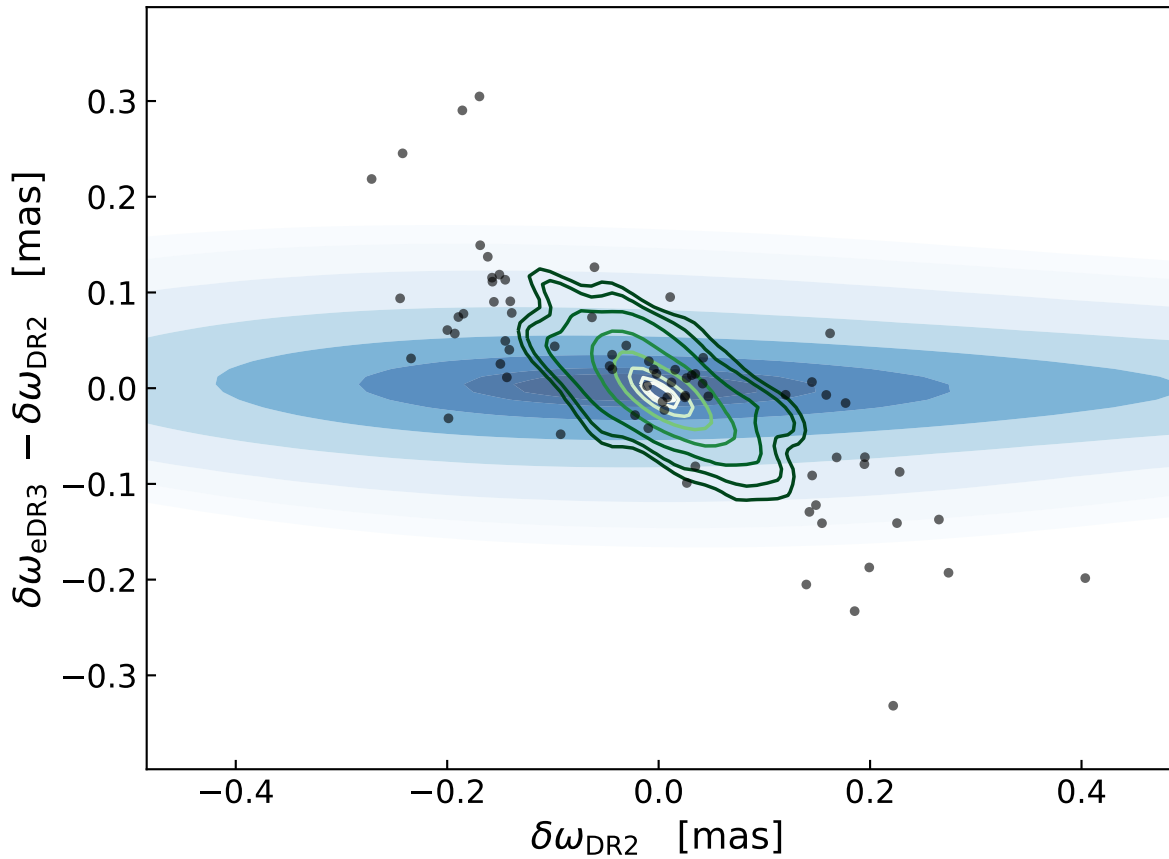
Figure 2.4: The observed parallax relative to the cluster in Gaia DR2 ($\delta\varpi_{DR2}$) compared to the change in this relative parallax between Gaia eDR3 and DR2 for the target field containing the cluster NGC 188. Field stars (filled contours, darker colors represent higher density) show almost no correlation between these quantities. Bonafide cluster members (open contours, lighter colors denote higher density) show a strong negative correlation. Stars identified as cluster members in our analysis but absent from Cantat-Gaudin et al. (2020) are shown as individual points. See text for details.
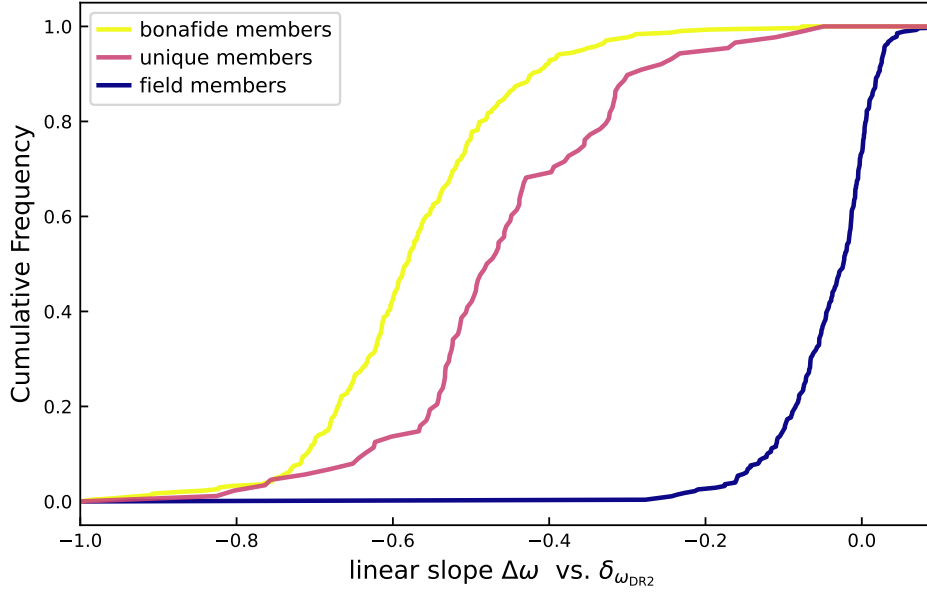
Figure 2.5: The distribution of linear slope fit to the $\Delta\varpi$ vs. $\delta\varpi_{DR2}$ data within three stellar populations along each line of sight. The slope measures the degree of correlation between the two quantities; each line represents the empirical cumulative distribution function of slope within a population. The slope of bonafide cluster members (yellow; median of -0.58) is significantly more negative than that of field stars (blue; median of -0.02). The distribution of slope measured for cluster members uniquely identified in this work (magenta; median of -0.48) is intermediate to the field and bonafide member distributions. The similarity of the bonafide and unique cluster member slope distributions suggests that a majority of the unique cluster members are likely genuine. See text for details.

requirements limit our analysis to 307 sightlines for the bonafide cluster member and field populations and 88 sightlines for the unique member stars. The linear fit is performed across the same domain for all three groups. The fit only applies to stellar measurements within the extent of $\delta\varpi_{DR2}$ for cluster stars, including both bonafide and unique members. Performing the regression across the same domain ensures the most straightforward comparison between cluster and field stars.

We show the distribution of the $\Delta\varpi$ vs. $\delta\varpi_{DR2}$ slope measured from the above regression for field, bonafide cluster member, and unique cluster member stars in Figure 2.5. The slope of the bonafide cluster members (yellow line) is significantly negative with a median slope of -0.58. The bonafide cluster members show a strong negative correlation between the dependent and independent quantities as predicted by equation 2.9. However, the expected slope for a pure cluster member sample is -1. We discuss potential reasons for this deviation below. Field stars (blue line) show almost no correlation between $\Delta\varpi$ and $\delta\varpi_{DR2}$; their median target field slope is -0.02. Given our assumptions, the expected slope value for field stars is zero and near our actual result. The distribution of slopes measured for unique cluster members (pink line) is much

closer to that of bonafide cluster members than that of field stars; the median slope is -0.48.

Figure 2.5 presents strong evidence that the majority of unique cluster members genuinely belong to a cluster. The unique cluster members clearly show a correlation between the observed parallax in Gaia eDR3 and Gaia DR2 that is not seen in the field stars. The most negative slope recorded for field stars is -0.28 for any line of sight. The measured slope of the unique cluster member stars is < -0.28 in over 90% of all lines of sight considered. Further, the similarity of the slope distributions of the bonafide and unique cluster member stars suggests that they are largely drawn from the same distribution of cluster-relative parallax ($\delta\varpi_{\mathrm{DR3}}$, $\delta\varpi_{\mathrm{DR2}}$). However, the distribution of slopes for the unique cluster stars is not as negative as that of the bonafide cluster members. We therefore conservatively conclude between 5% and 45% of the unique cluster stars are possible false positives. We identify 9,118 unique cluster member stars meet the data criteria detailed within this subsection and more than 7,500 probable unique cluster member stars were removed from this analysis. The increase in probable cluster membership due to our unique cluster stars is significant even if the false positive rate is near the high-end of our estimated range.

To simplify the calculation of $\Delta\varpi$ and related quantities, we made several assumptions that are likely incorrect in detail but do not detract from our line of reasoning. The Gaia reported parallax contains some systematic error and is likely biased. In our calculations, we do correct for the global parallax offset of 0.029mas in Gaia DR2 (Lindegren et al., 2018) and 0.017mas in Gaia eDR3 (Lindegren et al., 2020) but then assume no additional systematic errors. A significantly non-zero intercept of the linear fit to $\Delta\varpi$ vs. $\delta\varpi_{\mathrm{DR2}}$ along a sight-line suggest influential and dissimilar bias in the parallax measurements of the two Gaia data releases. To mitigate this where possible, we remove any sightline in which the absolute value of the intercept is greater than one-tenth the median uncertainty in the observed parallax of the population under consideration. All of the reported statistics from Figure 2.5 include this restriction. Finally, clusters do have intrinsic width, often near the scale of $\delta\varpi$ for cluster members. The combination of these effects contributes to the change in the slope distribution for bonafide and unique cluster members relative to the expectation of -1 and likely contributes to the slight negative skew seen in Figure 2.5 for field stars. Still, the slope distributions of the bonafide member and field stars are sufficiently different from that of field stars to support our conclusions above.

2.5 Eleven Previously Unidentified Open Cluster Candidates

We report the serendipitous discovery of 11 previously unpublished open cluster candidates. These candidates were found as components with relatively low differential entropy within the best-fit XDGMM of the 420 target fields with a successful recovery of the central target cluster. When looking for candidates in a given field, we considered all non-target cluster components of the best-fit model that had a differential en-

tropy within a factor of 4 of the target cluster. We find 97 components that meet this differential entropy requirement.

The low differential entropy components must pass several tests before we consider them cluster candidates. First, we check to see that the cluster is spatially compact as well compact in proper-motion.

We check compactness in proper-motion by employing the proper-motion dispersion check developed by Cantat-Gaudin and Anders (2019, section 5.3). We calculate the total dispersion in proper-motion for each component and compare it to their analytical threshold. We examine the on-sky spatial compactness of each candidate using the method of Cantat-Gaudin et al. (2018c, section 2.3). We compare the sum of the minimum spanning tree branch lengths of the candidate cluster member positions to those of a randomly generated position distribution of the same size. We remove 44 candidates that do not meet both these requirements.

We then examine the parallax and sky position distributions of the remaining 53 candidates. We calculate the slope of Δ_ϖ vs. $\delta_{\varpi_{\mathrm{DR2}}}$ using the probable members of each component (see Section 2.4.2.3 for details). Only 3 candidates do not have a slope < -0.4, which is more negative than the field members within any target field (Figure 2.5).

We then determine if any of the 50 candidates were previously reported using Gaia DR2 data. To do so, we checked for overlap between the Gaia DR2 source IDs of our candidate members and the source IDs of membership lists from the literature. We compiled a list of all unique source IDs reported as members of *any* cluster by Cantat-Gaudin et al. (2018c), Cantat-Gaudin and Anders (2019), Castro-Ginard et al. (2018), Castro-Ginard et al. (2019) Castro-Ginard et al. (2020), Ferreira et al. (2020), Sim et al. (2019), Liu (2019), and **?**. One or more of these studies previously identified a cluster with members from 39 of our candidates. Only 11 previously unreported candidate clusters remained.

Finally, we estimate the physical size of each candidate. We calculate a physical r50 value, the radius encompassing 50% of the candidate member stars, using the candidate's bayesian inversion parallax distance (see 2.4.1). The 11 remaining candidates span a physical r50 value ranging from 1.15 parsecs to 11.3 parsecs, placing them in the bulk of typical open cluster sizes (see Cantat-Gaudin and Anders, 2019, Figure 4).

The final 11 newly discovered open cluster candidates all have different parent target fields. The gaussian component associated with the candidate had either the lowest (6 of 11 fields) or second-lowest differential entropy of the best-fit components.

We stress that these 11 systems are open cluster *candidates*. The member stars of these systems occupy a relatively small, compact volume in the space of ϖ, $\mu_{\alpha*}$, and μ_δ. All candidates passed the statistical tests mentioned above and their qualitative characteristics (see below) are consistent with being open clusters. Still, confirming these candidates as bonafide open clusters will require dedicated observational follow-up and

detailed investigation. The 11 clusters are called XDOCC (eXtreme Deconvolution Open Cluster Candidates) and numbered [1-11].

We present the color-magnitude diagrams of the 11 new open cluster candidates in Figure 2.6 and report their central parameters in Table 2.1. We get age, distance, and extinction estimates for the 11 cluster candidates by using the Auriga Neural Net developed by Kounkel et al. (2020). The Auriga neural net has been trained on Gaia DR2 photometry [G, G_{BP}, G_{RP}], 2MASS (Skrutskie et al., 2006) photometry [J, H, K], and Gaia DR2 parallaxes in order to predict cluster age, distance, and interstellar extinction (A_V). We also plot isochrones using the derived parameters from Auriga, assuming solar metallicity in Figure 2.6. We describe the 11 new candidates and their predicted Age, A_V, and distance from the Auriga Neural Net below.
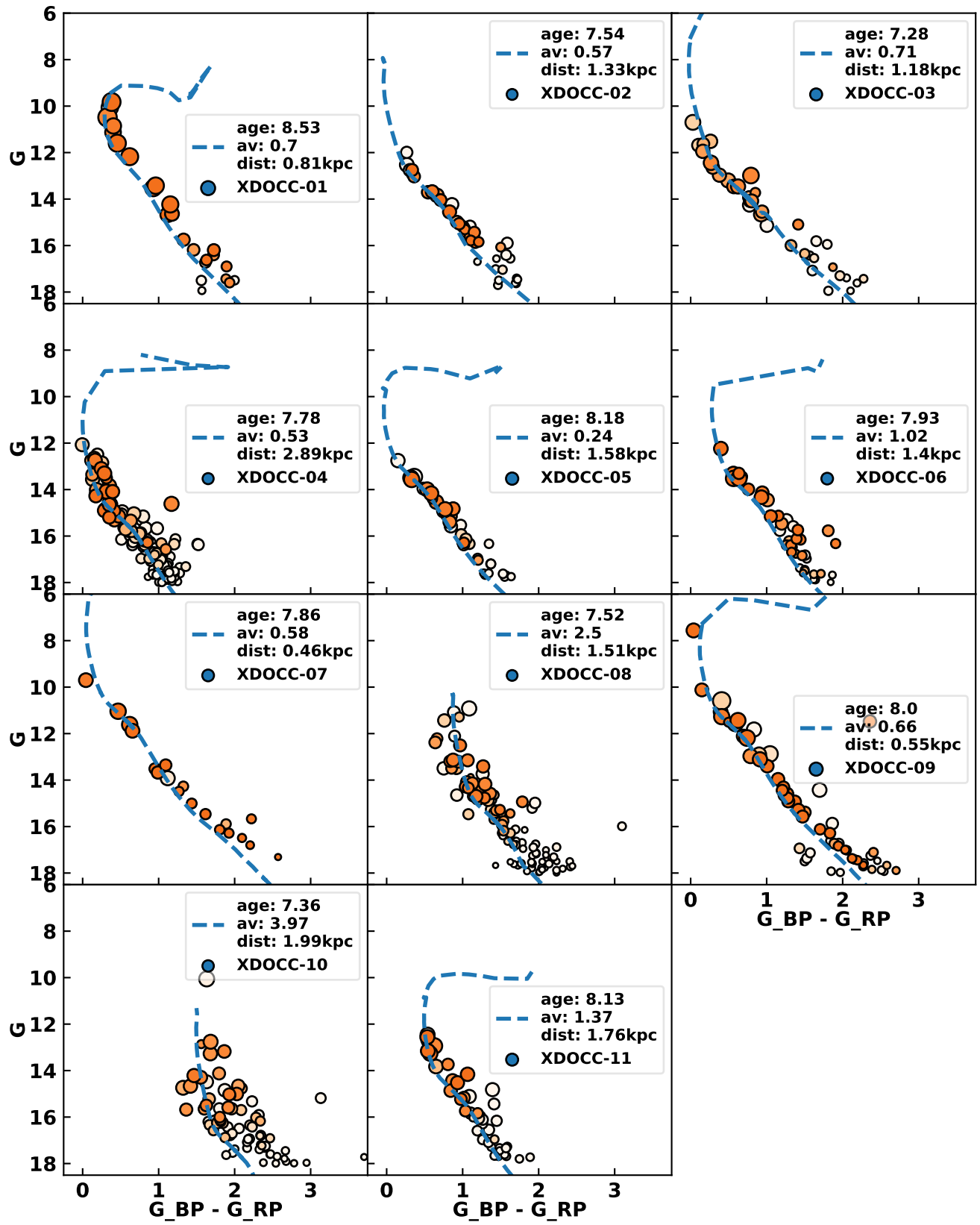
Figure 2.6: Multi plot color-magnitude diagrams of the 11 newly discovered open clusters within this work. Cluster member stars are plotted as orange colored points. The points are colored by their membership probability, where darker colors == higher probability. The points are sized by their G_{BP} - G_{RP} uncertainty, where bigger == lower uncertainty. The isochrone age, distance, and extinction values are predictions from the Auriga Neural Net developed by Kounkel et al. (2020). All isochrones have been set to a solar metallicity.

CHAPTER 3

An Open, Flexible Framework for Binary Orbital Parameter Inference: Modeling Spectroscopic Eclipsing Binaries with APOGEE and TESS data

3.1 Introduction

Binary-star systems (hereafter "binaries") are a fundamental component of a plethora of astrophysical problems and are very common throughout the universe. About half of the stars in the Milky Way are found in binaries (Duchene et al., 2018), and the binary demographics and evolution of binaries are key unknowns in many astrophysical contexts (e.g., Price-Whelan et al., 2019; Rix et al., 2019). However, binaries are also critical benchmark systems: for example, binaries provide one of the few ways to infer stellar distances without the use of stellar evolution models (Soderblom, 2010; Serenelli et al., 2021), and enable an important channel for calibrating models of stellar parameters across a range of stellar masses (Kraus and Hillenbrand, 2009; Stacy et al., 2009) as most binary-star components are assumed to be coeval in the field.

A crucial class of benchmark binaries are eclipsing binaries (Serenelli et al., 2021) (EBs), in which the eclipse shapes and timing provide information about both the orbital properties of the system, such as inclination, period, eccentricity, and orientation, and also intrinsic properties of the stellar components, such as the radius ratio and surface brightness ratio (e.g. Andersen, 1991; Torres et al., 2010b; Miller et al., 2020; Cunningham et al., 2020). Especially useful EB systems are those with small secondary components (e.g., Torres, 2013; Gill et al., 2020; Sebastian et al., 2022) in which a star with well-predicted structure and parameters (e.g., a 1 $M_\odot$ primary star) has a low-mass companion (e.g., a 0.3 $M_\odot$ star) with much more uncertain intrinsic and atmospheric parameters (Kraus et al., 2011; Cassisi and Salaris, 2019). Spectroscopic binaries, on the other hand, can offer different constraints on orbital parameters depending on whether the spectra encodes a detection of one (SB1) or two (SB2) stars. However, due to a degeneracy between the inclination and semi-major axis, SBs alone can only provide population level inferences of binary demographics (Badenes et al., 2018; Price-Whelan et al., 2020). SB1s, coupled with available ancillary data, are ideal to perform analyses of binary-star field and cluster populations.

Both historically and continuing today, benchmark EB systems are often first identified from time-domain photometric observations and subsequently observed spectroscopically to obtain radial velocity time series of the systems (e.g. Baroch et al., 2022; Hong et al., 2022). With recent all-sky, high-cadence, time-domain photometric surveys such as the Kepler Mission (Prša et al., 2011) and the Transiting Exoplanet Survey Satellite (TESS; Ricker et al., 2014; Stassun et al., 2018), it has become possible to identify large samples of $\sim$

10^4 EB systems across a range of stellar types (Kirk et al., 2016; Prša et al., 2022). These systems then require spectroscopic follow-up to obtain radial velocity and spectroscopic stellar parameter constraints to determine accurate intrinsic parameters (masses, radii, etc.), but follow-up programs at this scale are challenging given the required time-domain observations needed. Luckily, large spectroscopic surveys, in terms of the number of targets, obtain time-resolved spectroscopic data as a part of their survey strategy or targeting programs, such as the APO Galactic Evolution Experiment (APOGEE; Majewski et al., 2017).

In the case of APOGEE, the time coverage is typically sparse and over long baselines (months to a decade; see, e.g., Figure 2 in Price-Whelan et al. 2020). Still, APOGEE observed over 50,000 observed sources have more than 5 spectroscopic epochs. In this work, we are ultimately interested in generating a framework that can reliably infer orbital parameters for these EBs with sparse or small number of radial velocity measurements. Radial velocities are difficult and expensive to measure with high enough cadences and over time baselines *en masse*. Mining the population of EBs with sparse time-series radial velocity data for orbital parameters will enlarge the overall sample of binary parameters from which population demographics can be better constrained.

Here we develop a framework for the generalized joint modeling of SEBs with this goal in mind; we use a smaller subset of SEBs as a testbed to demonstrate the method, characterize its reliability and precision, and understand its limitations.

Multiple packages already exist and are currently in use by the community to perform joint fitting to time-series radial velocities and light curve data, such as PlanetPack (Baluev, 2013; Baluev et al., 2020), Allesfitter (Günther and Daylan, 2021), Exofast (Eastman et al., 2013, 2019), Exoplanet (Foreman-Mackey et al., 2021), Juliet (Espinoza et al., 2019), and Ellc (Maxted, 2016). However, a majority of these packages are typically well-suited to fit to double-lined spectroscopic binaries (SB2s) or light curves with multiple transiting exoplanets. These modeling packages employ Bayesian inference to obtain orbital solutions and intrinsic parameters of the component orbiting bodies based on the observed data provided. The utility of these packages is that they enable constructing or specifying a likelihood function for the observed data simultaneously given parameters that pertain to both the intrinsic stellar parameters of the binary components and the orbital architecture of a system. This likelihood function can then be combined with prior probability distribution functions (PDFs) to construct a posterior PDF for the parameters.

Modern modeling frameworks usually perform Markov Chain Monte Carlo (hereafter MCMC) sampling to draw samples from the posterior PDF, which can then be used to compute statistics of the posterior or marginal posteriors for parameters of interest. Different packages offer different methods to perform MCMC sampling (such as ensemble sampling methods, nested sampling, or Hamiltonian Monte Carlo methods), each with advantages and disadvantages for generating posterior samples. For example, ensemble sampling

methods typically have trouble generating posterior samples when the shape of the posterior PDF is highly-correlated or irregular, and generally do not scale well to large numbers of parameters. Nested sampling handles multimodality and complex posterior PDF shapes better, but also scales poorly with the number of parameters. Hamiltonian Monte Carlo (hereafter HMC) methods can generally handle more irregular posterior PDF shapes and scale well with the number of parameters. HMC performs inference by using the gradients of the posterior PDF with respect to the parameters to explore it more efficiently. However, calculating the gradients during sampling can be computationally expensive.

The existing EB modeling frameworks currently implement one or more of these sampling methods, which we briely discuss. Allesfitter (Günther and Daylan, 2021), for example, uses ensemble or nested sampling methods through the emcee (Foreman-Mackey et al., 2013) or dynesty packages (Speagle, 2020), respectively. Juliet (Espinoza et al., 2019) enables modeling time-series SB1 data with light curve data using nested sampling. Exoplanet (Foreman-Mackey et al., 2021) can fit both flux and radial velocity time-series data and is able to use HMC sampling methods because it is built upon the PyMC3 (Salvatier et al., 2016) model building framework which is optimized for fast gradient calculations. This infrastructure is built with performing fast calculations of gradients in probabilistic models in mind, correcting a computational bottleneck when performing HMC sampling.

One of the main issues with constructing a flexible, generalized Bayesian probabilistic framework is the necessity to be able to handle many different and highly-irregular posterior parameter spaces. The data we want to model, light curve flux measurements and time-series radial velocity measurements, are both influenced by the orbital parameters of their host binary system, as well as the orientation in space with respect to us. Jointly modeling both the radial velocity data with the light curve allows us to infer the binaries orbital parameters and orientation completely. However, the posterior PDF can also be irregularly-shaped due to choices of parametrization or complex degeneracies between parameters, causing many samplers to under-perform or fail.

The implementation of HMC used by Exoplanet includes the No U-Turn Sampler (hereafter NUTS, see Hoffman and Gelman, 2014). NUTS is a modification of an HMC sampler that tunes the model's hyperparameters as samples are drawn to efficiently explore the posterior parameter space. Tuning hyperparameters otherwise typically involves costly re-runs of models with different hyperparameter values.

We build our framework using Exoplanet as we focus on flexibility to be able to properly infer orbital parameters for a varied combination of SEBs with only SB1 data and automation to tackle a large enough sample to infer binary population demographics. The NUTS algorithm to perform flexible HMC inference without costly external hyperparameter tuning meets both design requirements of our proposed framework.

We demonstrate the capabilities of this framework with a sample of spectroscopic eclipsing binaries with

light curve data from the TESS mission and single-line radial velocity data from the APOGEE surveys. Currently, there has not yet been an investigation into the capabilities of jointly fitting SEBs with TESS and APOGEE data en masse. We hope to showcase both the exquisite nature of the high quality data within TESS and APOGEE, as well as the robustness of our framework as a tool to the astronomical community.

3.2 Data

The successor to the Kepler spacecraft (Batalha et al., 2010), the Transit Exoplanet Survey Satellite (hereafter, TESS) (Ricker et al., 2014; Sullivan et al., 2015; Stassun et al., 2017, 2019) is now in its 5th year of operation in the second extended mission. The time-domain photometric data from TESS has helped make breakthroughs in the field of exoplanets and in the study of binaries because of its high-cadence, nearly all-sky-coverage observations of over 200,000 stars, covering almost 70% of the night sky. TESS observes targets continuously within a particular sector over a span of 13.5 with a short 2-minute cadence; each sector has at least 27 total days of observing time.

The resulting data contains a large number of high quality time-series light curve flux data for each star in the TESS Input Catalog (Stassun et al., 2017). Having this continuous, short-cadence light curve data enables studying periodic behavior present in the observed stars over many timescales (e.g., stellar flares (Günther and Daylan, 2021), stellar rotation (Doyle et al., 2020), stellar pulsations (Cunha et al., 2019), and stellar/sub-stellar companions (Prša et al., 2022; Gandolfi et al., 2018, respectively).

The Sloan Digital Sky Surveys (SDSS) have been in operation for nearly 20 years. The recently released 17th data release from SDSS (Abdurro'uf et al., 2022) marks the final data release in SDSS-IV (Blanton et al., 2017), and contains a full release of the APOGEE-2 spectroscopic data (spectral resolution $R \sim 22,500$) (Majewski et al., 2017; Beaton et al., 2021; Santana et al., 2021). The APOGEE-2 data set contains over ~ 2.6 million individual spectra for over $700,000$ stars. The APOGEE project was originally designed to only perform spectroscopic chemical analysis of stars, which required multiple visits to accumulate a high enough signal to noise ratio. With multiple spectra and long time baselines, these data can be repurposed to provide time series radial velocity measurements (e.g., Troup et al., 2016; Badenes et al., 2018; Price-Whelan et al., 2018).

For our initial list of binaries we sought to have well-sampled radial velocity data. We opted to use the candidate binaries identified by modeling the time-resolved "epoch" radial velocity data in APOGEE data release 17 (Abdurro'uf et al., 2022), as released in the *The Joker* value-added catalog. This sample was identified using the custom Monte Carlo sampler *The Joker* (Price-Whelan et al., 2017), which was designed to use radial velocity measurements from single lined binaries (SB1s) to fit a six parameter binary orbital model and infer posterior distributions with the capacity to handle multi-modal solutions. The pipeline used

to process and run *The Joker* on the APOGEE DR17 data follows the procedure used in Price-Whelan et al. (2020). We will demonstrate with our framework that the inclusion of photometric data into the orbital model will break some of the degeneracies present when only modeling with radial velocities.

Further quality cuts were implemented on this catalog of binaries from APOGEE DR17 to produce a "gold sample" that can be reliably employed to consider orbital properties with stellar parameters. In general we require that each system have clean APOGEE data, reliable APOGEE stellar parameters, well-behaved posteriors from *The Joker*, and measurements determined by the STARHORSE project. We restate those quality cuts below; in particular, we require that the APOGEE binary-star systems have:

1. unimodal (in orbital period) posterior samples produced from *The Joker*,

2. close cross-matches with *Gaia* and 2MASS within ($2''$),

3. analyzed by the STARHORSE project (Queiroz et al., 2018),

4. no APOGEE fiber contaminants ($< 2''$ with $\Delta G > -5$), using *Gaia* EDR3 (Gaia Collaboration and Brown, 2020) source positions and photometry,

5. no bright neighbors ($< 10''$ with $\Delta G > -2.5$), again using *Gaia* EDR3 photometry,

6. reliable stellar parameters from the APOGEE pipeline,[1]

7. small inferred excess radial-velocity variance ($s_{MAP} < 0.5\,\mathrm{km\,s^{-1}}$),

8. more than 5 APOGEE visits ($N_{visits} > 5$).

The resulting Gold sample contains 2,666 binary systems. We truncate the gold sample further by requiring that all binary systems have a spectroscopic orbital period less than 30 days according to the maximum *a posteriori* period sample generated from *The Joker* for each system. We set this period limit to 30 days to limit our initial sample to low eccentricity binaries, due to tidal circularization of tight binaries. We reduce our binary sample to 1,741 binary systems with this period cut.

We choose to only consider pre-processed TESS light curves from the TESS-SPOC processing group (Caldwell et al., 2020). This is not a fundamental step in our framework, nor is it necessary for the proper employment of our framework, but is a choice in data to reduce the amount of pre-processing of TESS full frame images to produce the light curves used in this work. We build our framework to be able to handle any time-series data with flux measurements, time-stamps, and flux uncertainties.

---

[1] $-0.5 < \log g < 5.5$
$3500\,\mathrm{K} < T_{\mathrm{eff}} < 10\,000\,\mathrm{K}$
$-2.5 < [\mathrm{M/H}] < 0.5$

We employ the LightKurve (Lightkurve Collaboration et al., 2018) Python module to search and download all available TESS short cadence (2 min) light curves across all sectors for the 1,741 spectroscopic binary systems identified from *The Joker* and APOGEE data. We find 104 binary systems with TESS light curve data. We find that there are 57 binary systems in our sample with noticeable or conspicuous transits visible in the light curve after a rudimentary BLS period determination using LightKurve. We visual inspect the resulting light curves for both a primary and secondary transit to limit our sample to non-grazing orbital orientations. Our final sample has 12 binaries which met this requirement.

3.3 Methods

We construct a generalized framework that will be used to build different SEB orbital models with parameters constrained using both the radial velocity measurements from APOGEE and the light curve flux measurements from TESS. To that end, we design our framework to perform a series of functions such that flexibility is maximized while only minimal external intervention is required once the framework is initiated for a target binary system.

The likelihood function in our framework comes from the internal orbital model Exoplanet assuming a Keplerian orbit. The orbital model is constructed with the primary component mass and radius, the mass ratio, the surface brightness ratio, the radius ratio, along with the eccentricity, argument of periastron, impact parameter, and period. The secondary component mass and ratio are depend on the primary component parameters and ratios mentioned above. The light curve model is constrained by these parameters as well as limb darkening parameters for the primary and secondary component.

We list the full suite of parameters that go into our framework along with short descriptions in Table 3.1. We initialize the framework with the period and transit epoch times derived in Section 3.2, where we perform a manual inspection to decide to utilize either the TESS period/transit epoch or the APOGEE *The Joker*-derived period/transit epoch. Future versions of our framework may rely on calculating the signal detection efficiency (SDE, Alcock and Parker, 2019) of phase folded data to automatically select best period/transit epoch combination.

We calculate initial estimates for some of the parameters we use to initialize our priors using the Isochrones python package. We infer initial individual component stellar parameters by sampling the MIST (Dotter, 2016) evolutionary isochrones built into the Isochrones package. These posterior PDFs derived from Isochrones help inform the prior PDFs used in our orbital modeling framework for the log-primary mass $\log M_1$, the log-primary radius $\log R_1$, the log-radius ratio $\log k$, the log-mass ratio $\log q$, and the log-surface brightness ratio $\log s$.

With these estimates we initialize our priors, as well as our transit model likelihood function, within
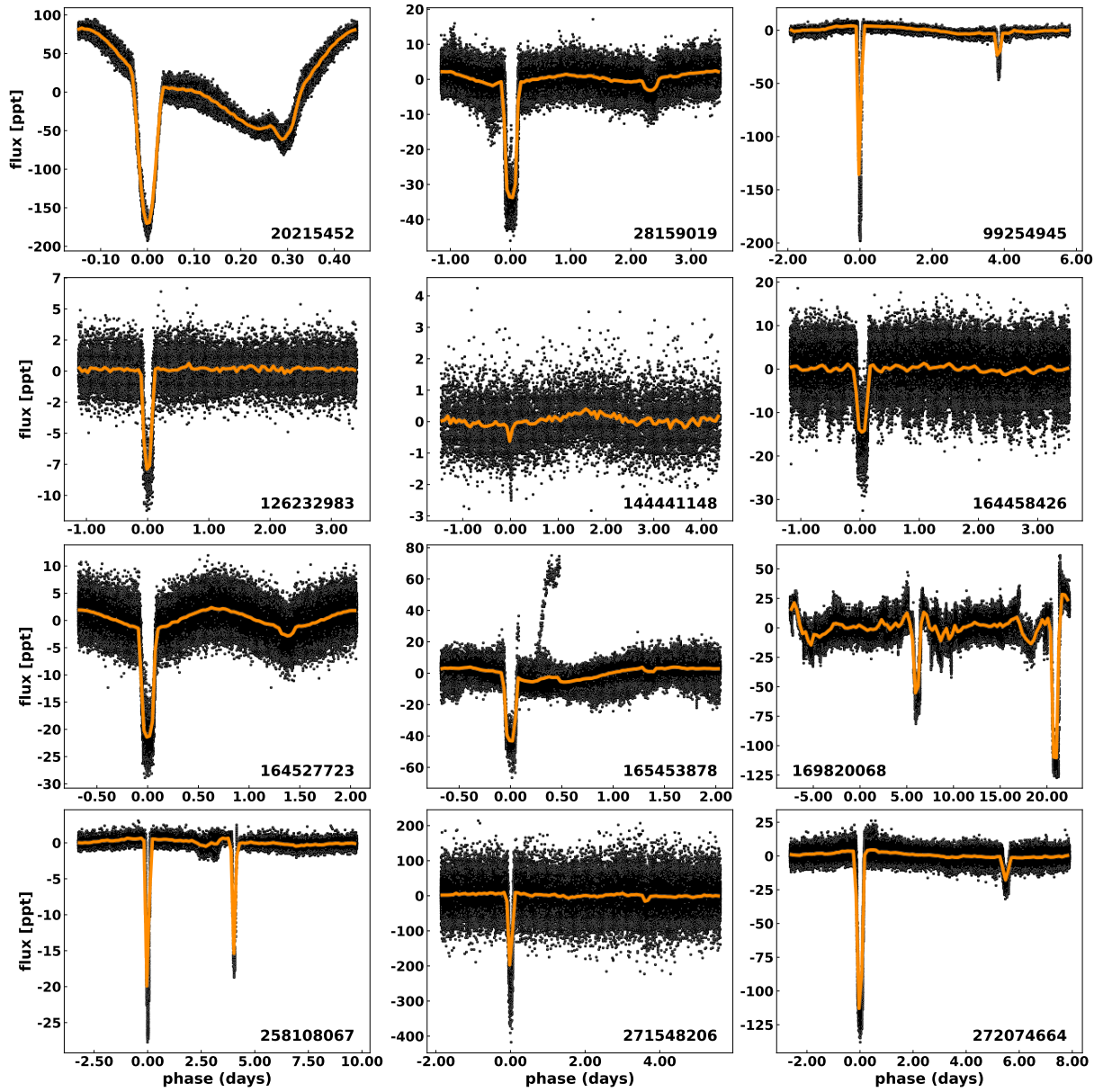
Figure 3.1: Each panel shows a phase-folded TESS light curve for the 12 spectroscopic eclipsing binary systems in our sample. All light curves have been phase-folded with a Box Least Squares-derived period and eclipse epoch from the TESS light curves, except for TIC 169820068, which is phase-folded using the maximum *a posteriori* period from *The Joker* using APOGEE radial velocities. The flux values for all systems have been normalized to parts per thousand, ppt, and plotted in black, and phases are shown in time units. A rolling median flux of the light curve is over-plotted as the solid (orange) line. The light curves selected for this sample have a wide range of transit geometries, making them suitable as an initial test sample for our framework.

a PyMC3 model within the Exoplanet package. We sample using HMC to construct the posterior SEB orbital parameters, allowing effective Bayesian sampling within ill-behaved and high dimensional parameter spaces. Posterior distributions which are multi-modal or have irregularly shaped geometries are an obstacle for any Monte Carlo sampler. Spectroscopic eclipsing binaries, in particular, can have high dimensional posterior parameter spaces. These issues stem from the combination of intrinsic binary dynamics as well as observational orientations from our line of sight. There are also numerous degeneracies between these two groups of parameters, further complicating sampling the posteriors effectively. HMC sampling, by design, is able to sample posteriors such as these and is the best choice to build our generalized framework.

3.3.1 Hamiltonian Monte Carlo

The most frequently used methods to determine the orbital parameters usually rely on some generative model (such as a two-body keplerian orbit) which takes a vector of parameters which can be called 'θ'. This generative model is incorporated into a likelihood function which computes the similarity of the model synthetic data with the observed data using a statistical metric such as the χ^2 parameter. The likelihood function can be maximized in order to infer the orbital parameters, that coupled with the generative model, best approximate the observed data.

However this process, simply referred to as 'Maximum Likelihood' (hereafter ML), has a number of shortcomings, such as: ML infers the mode of a distribution which is not as robust to skewed distributions as the median of the distribution. ML does not directly generate uncertainties on the parameter modes inferred. ML does not take into account any relevant prior information that may help better infer the desired parameters.

We can correct for these issues by using Bayes Theorem (Bayes, 1763) to perform our inference which is written out as $P(\theta|(D) = P(D|\theta) \cdot P(\theta)/P(D)$. Here $P(D|\theta)$ is our likelihood function and $P(\theta)$ is our prior function. $P(D)$ is called the 'Bayesian Evidence' and while it is an important quantity, it is also usually very computationally expensive to calculate and primarily useful in comparing different models. As our framework currently only uses one orbital model, it can be safely ignored for now.

The Bayes theorem equation is fundamentally an integral over the whole of the parameter space we are interested in. Fully evaluating the integral is a computational intractable endeavor. However we can evaluate this equation numerically using Markov Chain Monte Carlo (hereafter, MCMC). MCMC is an iterative way to evaluate $P(\theta|D)$ (known as the posterior distribution) by drawing samples for θ. The probabilities of the values in θ are calculated in the prior. These values are then used in our generative model and compared with the data in the likelihood function.

Most MCMC methods rely on the 'Metropolis-Hastings' algorithm (otherwise known as 'random-walk') (Metropolis et al., 1953) to draw samples to construct the posterior distributions. Most of the samples drawn

throughout the MCMC process are correlated and thus most MCMC evaluations rely on running a chain of samples long enough such that sufficient uncorrelated samples are drawn to properly approximate the target posterior distribution. It is also possible to run N multiple chains of samples with N multiple starting values of θ, known as 'walkers', in order to more effectively explore the parameter space concurrently.

Unfortunately, the efficacy of MCMC relies on the presumption of properly exploring the parameter space from which the distributions of the variable posteriors can be inferred. MCMC algorithms that rely on a random walk to explore a parameter space usually suffer from an incapacity to identify whether effective exploration of the target distribution is being achieved. Characterizing an MCMC run as 'converged' then typically rely on post-MCMC sampling techniques such as calculating the auto-correlation length of samples to determine if the number of samples drawn is enough; this process may need to be repeated several times to properly derive posteriors for very geometrically complicated problems.

Hamiltonian Monte Carlo (Duane et al., 1987) (hereafter HMC) draws from the properties of Hamiltonian physics to exploit the differential geometry of a parameter space, even if its non-physical in nature, in order to efficiently explore that parameter space in building posterior distributions. HMC methods rely on calculating the gradients within a parameter space in order to explore it efficiently (see Neal, 2012).

HMC can still be potentially undermined if the two hyper-parameters, step size (ε) and number of steps (L), are not set adequately. The No U-Turn Sampling (NUTS) algorithm (Hoffman and Gelman, 2014) was developed to allow researchers to bypass the usually time consuming process of performing multiple iterations of sampling runs with varying ε and L values to determine how best to explore a parameter space. The NUTS algorithm actively adjusts ε and L during sampling as well as terminating sampling when the algorithm confirms that the parameter exploration is beginning to double-back onto itself in order to maintain computational efficiency.

Posterior distributions which are multi-modal or have irregularly shaped geometries are an obstacle for any Monte Carlo sampler. Spectroscopic eclipsing binaries, in particular, can have high dimensional posterior parameter spaces. These issues stem from the combination of intrinsic binary dynamics as well as observational orientations from our line of sight. There are also numerous degeneracies between these two groups of parameters, further complicating sampling the posteriors effectively. HMC sampling is able to sample posteriors such as these and is the best choice to build a generalized framework.

3.3.2 Assumptions in our framework

Jointly modeling a binary with its light curve flux measurements as well as radial velocity measurements makes it possible to infer all of the binary orbital parameters individually. We make certain assumptions to develop a clear structure to the solution for the problem. The assumptions we make with respect to the

APOGEE radial velocity data are similar to the assumptions made in Price-Whelan et al. (2017), only now we use the more recent DR17 APOGEE data from Abdurro'uf et al. (2022).

Our assumptions about the radial velocity data summarily concern the time dependent gravitational pull from a single unseen companion acting on the observed star within a detached two-component binary. The assumptions on the light curve data similarly concern a completely-detached, eclipsing, two-component binary that does not have significant stellar activity contributing to the flux variability. For both the individual flux measurements and the individual radial velocity measurements we assume that the noise is Gaussian with variances set by the reported measurement uncertainties.

In detail, our assumptions about the APOGEE radial velocity data are:

1. The radial velocity data represent unbiased measurements of only the more luminous star in a binary system.

2. The time dependence of the radial velocity measurements originates only from movement of the observed star caused by the gravitational pull of a (spectroscopically-unseen) secondary star (i.e. there are no internal processes in the star which may produce a significant RV signal).

3. The noise model for the radial velocity measurements is Gaussian.

Our assumptions about the TESS light curves are:

1. The primary (secondary) transits in the light curve are the result of a primary star being eclipsed by a secondary star (and vice-versa) as they orbit around their common center of mass.

2. Limb-darkening effects from either star can be adequately described by a two-parameter quadratic function (Espinoza and Jordán, 2015, 2016).

3. The out-of-eclipse flux of each binary's light curve can be effectively modeled by a Gaussian process employing a stochastically-damped harmonic oscillator (SHO) kernel.

4. The individual light curve flux measurements can be viewed as samples from a Gaussian distribution with variances given by the reported flux uncertainties.

3.3.3 Phase parameter estimates from TESS light curves

We must first convert all of the measurements (radial velocity and flux) to the same epoch, as well as sort the data points within each set of measurements to be chronologically in order, and zeroed out such that both sets of observation times are centered with respect to either the APOGEE epoch time or TESS epoch time. We

already have two estimates of these parameters: The *The Joker* results in Price-Whelan et al. (2020) (`MAP_P` and `MAP_T0`) and the TESS BLS-derived parameters (see Section 3.2).

We plot all of the phased radial velocity and flux data using both the period, and epoch time (t_0) estimates from both the APOGEE MAP parameters from the *The Joker* results for that system, as well as the BLS-derived period and epoch time from the TESS light curve. We visually inspect the quality of the phase curves folded and centered by each different transit epoch time estimate. We find in 11/12 systems that the TESS light curve estimates for period and t_0 are more successful at folding both sets of data (TESS and APOGEE) into tightly defined phase curves.

3.3.4 Informed Priors with Isochrones

We set the the priors of log primary mass ($\log M_1$), log primary radius ($\log R_1$), log mass ratio ($\log q$), log surface brightness ratio ($\log s$), and log radius ratio ($\log k$) using values derived from the Isochrones package (Morton, 2015). Isochrones is a python package that wraps data from the MESA Isochrones & Stellar Tracks (MIST, also see Choi et al., 2016; Dotter, 2016) with additional functions to perform stellar parameter fitting to single stars as well as binary stars.

We use the Isochrones 'Binary Star Model' to infer binary component stellar parameters from the available photometry of our 12 SEB systems. `BinaryStarModel` performs nested sampling using the Multi-Nest (Feroz et al., 2009) package which is wrapped into the PyMultiNest (Buchner et al., 2014) package. Nested sampling performed by MultiNest allows for quick, efficient sampling of the parameter space possibly multi-modal posterior distributions. The Binary Star Model fits the available photometry under the assumption that each SEB is an unresolved binary in each the pass-bands.

We query and download the available 2MASS J, H, K magnitudes (Skrutskie et al., 2006), the *Gaia* eDR3 G, G_{BP}, and G_{RP} magnitudes (Riello et al., 2020), as well as the TESS magnitude (Sullivan et al., 2015) data for each binary system in our sample. The calculated TESS magnitude is retrieved from the header data of the LightKurve data for each TIC system.

The Binary Star Model has support for custom prior distributions for distance, age, [Fe/H], mass, AV, and a special parameter used internally by Isochrones called the Equivalent Evolutionary Phase (EEP). The EEP is used to approximate stellar evolution along the non-domain, curvilinear main-sequence as well as the later stellar evolutionary stages of a star. We set the distance prior to be a Normal distribution centered on the photo-geometric distance (`r_med_photogeo`) derived by Bailer-Jones et al. (2020) using *Gaia* eDR3 photometry and astrometry. The spread of this distribution is set to $3\times$ the average of the two inferred quantiles of the photo-geometric distance (`r_lo_photogeo` and `r_hi_photogeo`)

We use the available APOGEE measurements of [Fe/H], $T_{\rm eff}$, and $\log g$ to truncate the eventual parameter

space that Isochrones nested sampler explores. Truncating the parameter space helps the nested sampling reach convergence faster. The posterior distributions generated from this sampling will be used as initial values for the priors we ultimately use in our PyMC3 framework to model SEBs.

We perform a rudimentary grid search with Isochrones to set informed limits on the sampling priors Isochrones uses in the Binary Star Model. We use the Isochrones MISTEvolutionTrackGrid interpolator to bootstrap sample across a grid of isochrones spanning the Isochrones EEP range after the pre-main sequence phase [202-1700] and masses spanning [0.1 $M_\odot$–50 $M_\odot$, $M_\Delta = 0.025\ M_\odot$]. We set the metallicity [Fe/H] value to the APOGEE measured [Fe/H]. The T_{eff} and $\log g$ APOGEE measurements are then used to perform acceptance/rejection of the grid search generated isochrones. We keep the generated isochrones which have a T_{eff} and $\log g$ value within the APOGEE measured $T_{\text{eff}} \pm 500$ K and $\log g \pm 0.5$ dex, respectively. The bounds for the uniform priors Isochrones uses for mass, age, and EEP are set to the min/max ranges calculated from the accepted isochrones.

We run the Isochrones nested sampler with the aforementioned bounds and run until convergence. We take the primary component mass ($\log M_1$) to be the greater of the two masses derived with Isochrones. We calculate the surface brightness of each component according to Graczyk et al. (2017, Equations 2 & 3) to get the surface brightness ratio, $\log s$. We construct $\log q$ and $\log k$ as the ratio of the secondary component value over the primary component value for mass and radius, respectively. We sigma clip the $\log M_1$, $\log R_1$, $\log q$, $\log k$, and $\log s$ distributions to within 5 σ. We use the median absolute deviation (MAD) of each distribution multiplied by a factor of 1.4286 to perform sigma clipping. The MAD parameter is more robust to outliers than the standard deviation. The scale parameter of 1.4286 is used so the MAD more closely approximates the standard deviation the of normal distribution (see Leys et al., 2013). The μ, σ values from the sigma clipped $\log M_1$, $\log R_1$, $\log q$, $\log k$, and $\log s$ distributions are used to generate the values for the priors we use in our SEB orbital model (see Section 3.3.5).

3.3.5 The Exoplanet Binary Orbital Model

Our priors will be used to construct the orbital model within the Exoplanet python module (Foreman-Mackey et al., 2021), along with relevant noise terms. We compute the likelihood of the parameters of this orbital model using both the APOGEE-measured radial velocities and the TESS light curve flux measurements. The $\log M_1$, $\log R_1$, $\log q$, $\log k$, and $\log s$ priors have μ values that are set using the μ values derived from Isochrones (see Section 3.3.4). We let the variance of these five priors vary within the Exoplanet model by setting all five variances with two inverse Gamma distribution priors. We set one inverse gamma prior for the $\log M_1$ and $\log R_1$ variances, and another inverse gamma prior for the three SEB ratio ($\log k$, $\log q$, and $\log s$) variances. The inverse gamma priors are initialized with σ values from Isochrones priors multiplied by a

large factor of 256, which we set arbitrarily, to account for possible biases within the Isochrones framework arising from its use of the MIST stellar model grid fitting under an assumption of blended photometry.

We employ a Gaussian process using the Celerite2 package (Foreman-Mackey et al., 2017; Foreman-Mackey, 2018) to model the out-of-transit light curve flux of each eclipsing system. We use the SHO kernel in our Gaussian process. The two parameters that dictate the behavior of our Gaussian process are ρ_{GP}, the un-damped period of the SHO, and σ_{GP}, the standard deviation of the process. We also include a σ_{LC} prior to model the jitter present in the TESS light curve flux measurements. We only have one noise parameter for the radial velocity measurements, $\log \sigma_{RV}$, to model the jitter present in the APOGEE radial velocity measurements. We set $\log \sigma_{RV}$ with a log-normal prior, bounded at an upper value of $10 \mathrm{km\,s^{-1}}$.

We employ a unit disk function built into Exoplanet to sample ($e \cos \omega$, $e \sin \omega$). We then calculate the eccentricity e, and argument of periastron ω as $(e \cos \omega)^2 + (e \sin \omega)^2$, and $\arctan \frac{e \sin \omega}{e \cos \omega}$, respectively. We provide an estimate of $e \cos \omega$ to our model by performing a two stage BLS fit to the TESS light curve. We use the built-in transit masking capabilities of the LightKurve module to mask out the primary transit of each binary's TESS data and refit a second BLS to estimate transit parameters for the secondary eclipse.

We derive an estimate for $e \cos \omega$ by rearranging Equation 33 from Winn (2010) using the transit epoch time $T_{0,1}$ for the primary eclipse and the transit epoch time $T_{0,2}$ for the secondary eclipse. The $e \cos \omega$ value we derive and $e \sin \omega$, which we set to 0.0, are used to initialize the Exoplanet unit-disk function from which we derive values of eccentricity e and argument of periastron ω.

The impact parameter b is modeled internally using the Exoplanet impact parameter distribution function, which is conditioned on the radius ratio k (not to be confused with the radial velocity semi-amplitude K). We use quadratic limb darkening parameters (Kipping, 2013) for both the primary and the secondary light curve transits in our model.

We parameterize the orbital period P of the binary as a function of the epoch time of first transit t_0, and the last epoch of transit t_N, such that $P = \frac{(t_N - t_0)}{N_{transits}}$. This parameterization of the period removes a degeneracy present between period and t_0. This degeneracy arises from the long time baselines in the TESS observations and the shortness of the period of these binaries ($\lesssim 50$ days) which introduces a numerical sensitivity in the calculation of t_N. We model both the mean light curve flux and mean radial velocity present in each set of observations with normal distribution priors with a μ of zero and a σ of 10 ppt and $50 \mathrm{km\,s^{-1}}$, respectively.

3.3.6 Calculating the Maximum a posteriori values to initiate sampling

We use PyMC3 to calculate the Maximum *a posteriori* (hereafter, MAP) estimates of our orbital parameters. The MAP estimates are useful to initiate sampling within a parameter space which bolsters the HMC sampler's ability to construct reliable posterior distributions moreso than random initializations.
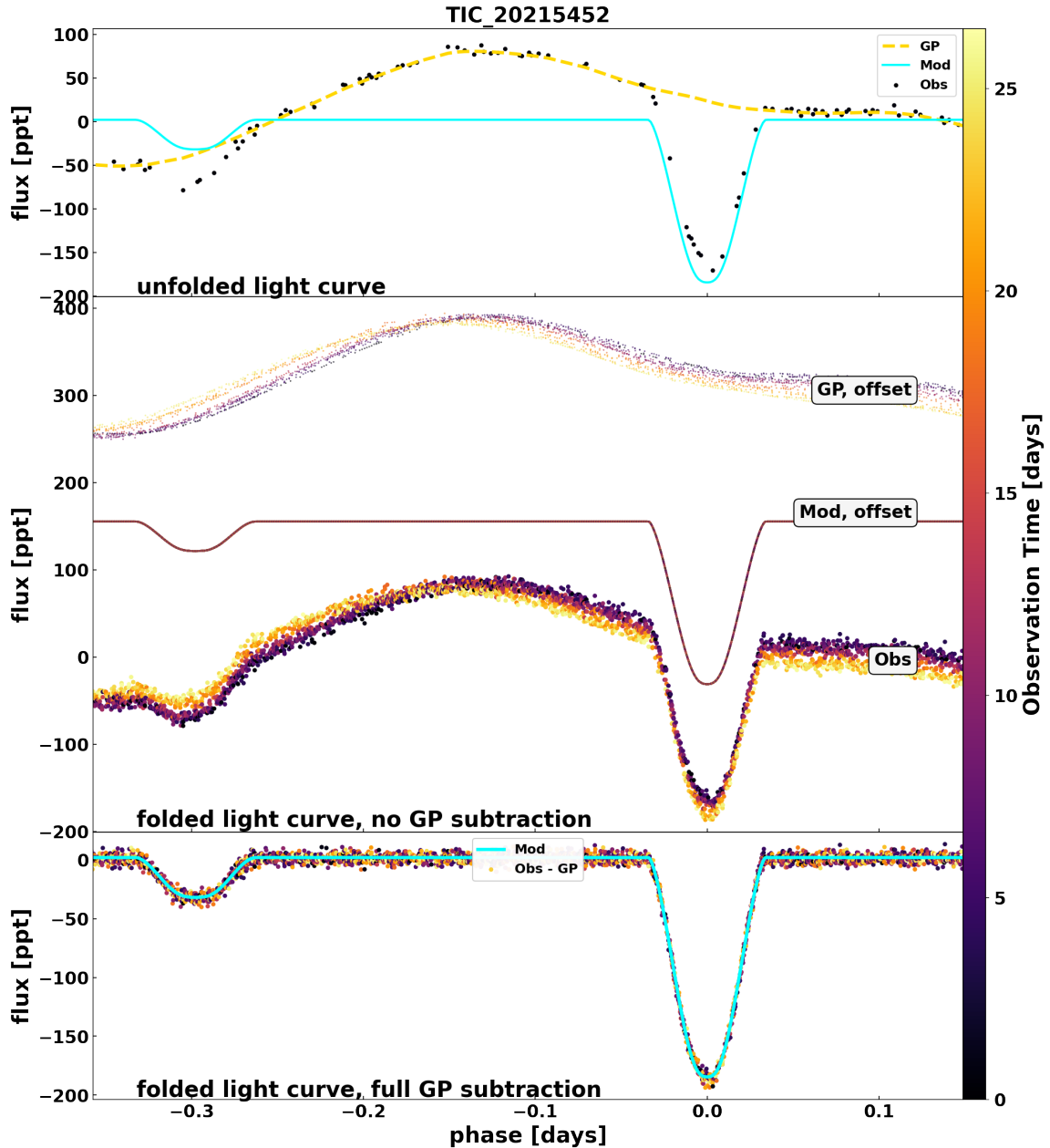
Figure 3.2: A visualization of the light curve of TIC 20215452 to demonstrate the usage of a Gaussian process to remove the background flux from a light curve. **Top**: A small section of the unfolded light curve of this system shown as colored markers, with the unfolded Gaussian process plotted as the dashed (gold) line, and the unfolded eclipse model over-plotted as the solid (cyan) line. **Middle**: The complete folded phase curve (markers, colored by absolute time), the folded Gaussian process model (group of dotted lines), and the folded eclipse model (colored solid line). The Gaussian process and the light curve data points are colored by absolute time in all panels. **Bottom**: The folded light curve (colored markers) with the fitted Gaussian process subtracted from the TESS data points. The light curve model is over plotted in cyan. Using a Gaussian process to model the background light curve flux allows for generalized fitting of multiple types of stellar variability.

Table 3.1: Priors used in the joint RV–LC orbit model (this work)

Parameter Name	Distribution	Hyper-parameters	Description
$\sigma_{\log m_1}, \sigma_{\log r_1}$	Inverse Gamma	$(7, 0.25)$	SD of M_1, R_1 priors
$\sigma_{\log k}, \sigma_{\log q}, \sigma_{\log s}$	Inverse Gamma	$(6, 0.5)$	SD of $\log k$, $\log q$, $\log s$ priors
$\log M_1$	Normal	$(\mu_{\log m_1}, \sigma_{\log m_1})$	Mass of primary
$\log R_1$	Normal	$(\mu_{\log r_1}, \sigma_{\log r_1})$	Radius of primary
$\log q$	Normal	$(\mu_{\log q}, \sigma_{\log q})$	Ratio of component masses
$\log s$	Normal	$(\mu_{\log s}, \sigma_{\log s})$	Ratio of component surface brightness
$\log k$	Normal	$(\mu_{\log k}, \sigma_{\log k})$	Ratio of binary component radii
b	Uniform	$[0, 1+k]$	Impact parameter
(e, ω^a)	Unit-Disk	$(\sqrt{e \cos \omega_{est}}, 0.0)$	Unit disk function
t_0	Normal	$(t_0, 3 \text{ d})$	Time of the first transit
t_N	Normal	$(t_N, 3 \text{ d})$	Time of the Nth transit
u_1	–	$[u_{1,0}, u_{1,1}]^b$	Limb darkening coefficients of primary
u_2	–	$[u_{2,0}, u_{2,1}]^b$	Limb darkening coefficients of secondary
μ_{RV}	Normal	$(0, 50 \text{ km s}^{-1})$	Mean background radial velocity
μ_{LC}	Normal	$(0, 10 \text{ ppt })$	Mean background light curve flux
σ_{GP}	Inverse Gamma	†	SD of Gaussian process
ρ_{GP}	Inverse Gamma	†	Un-damped period of Gaussian process
σ_{LC}	Inverse Gamma	†	Light curve flux 'jitter' term
$\log \sigma_{RV}$	Truncated Normal	$(\log(\widetilde{\sigma_{RV}}), 10 \text{ km s}^{-1})$	Radial velocity jitter term

[a] - argument of periastron

[b] - Uninformative prior for quadratic limb-darkening parameters as implemented by Kipping (2013)

† - We utilize the inverse gamma distribution parameter estimation algorithms outlined by Michael Betancourt

We perform MAP estimation steps on individual parameters in a sequential order before estimating all 19 parameters simultaneously. Sequential MAP estimates increase the numerical stability of the final MAP estimate of all 19 parameters. The order of the sequence is allowed to vary between different SEB systems to increase the flexibility of the framework.

The first MAP estimations that take place occur on radius ratio '$\log k$', which greatly influences the depth of the two transit dips. The impact parameter 'b' is next optimized as it influences the transit duration in our orbital model. The unit disk function 'ecs' is then optimized as the eccentricity calculated from it potentially influences the time between the primary and secondary transit as well as also affect the shape of the model fit to the radial velocity data.

The MAP parameter estimates for $[(t_0, t_N), (u_1, u_2), \log M_1, \log R_1, \log s, \log q, \text{mean-rv}, \text{mean-lc}]$ are ordered to maximize the gradient of the likelihood. This order is calculated in a brute force manner, looping through the set of parameters until the MAP step resulting in the maximum positive change in the log-likelihood is found. The parameter considered in this MAP step is then removed from the set and the loop is restarted. This loop continues until all parameters in the set are optimized.

The MAP estimates of the σ_{GP}, σ_{LC}, and ρ_{GP} parameters are performed simultaneously in a single step. This allows the transit model has been fit to the data as much as possible before attempting to fit the Gaussian

process to the out-of-transit background flux. Performing these MAP estimates before the transit model is fully fit to the light curve data may lead to the Gaussian process over-fitting. The $\log \sigma_{RV}$ parameter is also optimized after the completion of the brute force loop to avoid over-estimating the jitter in the radial velocity background.

The two inverse gamma priors which set the variances for $\log M_1$, $\log R_1$, $\log k$, $\log q$, and $\log s$ are not optimized until the final MAP step. The very broad initial variance parameters on the five orbital parameter priors remain throughout the previous steps to maximize flexibility in estimating the MAP values for all parameters. All 19 orbital model parameters have their MAP estimations performed simultaneously in the final step.

We iteratively perform sigma clipping to remove outliers from the SEB light curve. We calculate the residuals of the light curve model with the TESS data and mask light curve data points with $\|residuals\| > 5\sigma$. MAP values are recomputed after each sigma clipping of the light curve for a maximum of 10 iterations or until no light curve data points are masked as outliers.

3.3.7   Sub-sampling the light curve using the transit model

We use the MAP values computed after sigma clipping is completed to sub sample the out-of-transit flux to reduce computational time of sampling. The out-of-transit flux in a light curve is not simply background. It contains important information about the stars such as stellar variability, surface pulsations, and rotation signatures. Our framework models the background flux with a Gaussian process and treats this flux variability as "nuisance" or additional, correlated noise. We acknowledge that there is information to be distilled from the background flux and may integrate that functionality into our framework at a later time.

We perform targeted sub-sampling of the light curve background flux for our sample of SEB systems. This step is performed to (a) reduce computational cost of evaluating the likelihood across the entire light curve observation for a binary system with every sampling step, and (b) reduce density of background flux to increase transit visibility, as in the case for TIC 144441148 (Figure 3.1, row 2, column 2).

We mask out the primary and secondary transits along the entire light curve using the transit model with the MAP values estimated in the previous sub section. We sub-sample the out-of-transit light curve flux data with a random uniform draw down to a specified size. We find that sub-sampling the light curves to $\sim$5000 data points provides the best balance of computational efficiency and sampling reliability for all 12 SEB systems.

3.3.8 Generating posterior samples with the Exoplanet orbital model

We perform sampling using the priors and likelihood function in the Exoplanet model with 6 sampling chains. Each chain is set to run for 5000 tuning steps and 1000 draw steps. We consider each model to have converged upon the completion of sampling when the rank normalized split-$\hat{r}$ values for all of the model sampling parameters are below 1.01 (see Vehtari et al., 2019). The rank normalized split-$\hat{r}$ value improves upon the split-$\hat{r}$ () diagnostic's ability to detect convergence between multiple sampling chains.

3.4 Results

We successfully infer orbital and stellar parameters for all 12 of our spectroscopic eclipsing binary systems in the short-period APOGEE–TESS cross-match sample. We plot both the radial velocity phase curves and the GP-subtracted, transit-phased light curves for all 12 fitted systems using the median orbital model parameters in Figure 3.3. We also plot the 68th and 95th confidence intervals of the orbital model in both the phase curves in Figure 3.3. We plot an orrery of the 12 SEBs with a top-down view of their orbits from median of our model posteriors in Figure 3.4 to visualize the different binary component configurations modeled with our framework.

We also list a summary of important binary parameters in Table 3.2. The super-scripts (sub-scripts) listed with each column in 3.2 are the upper (lower) bound uncertainty calculated from the Bayesian Credible Interval (hereafter, BCI) for each parameter. The BCI for a parameter can be interpreted as the interval of the marginal posterior PDF within which the true parameter value falls with a specified probability (see Dahn, 1978). The BCI probability is we set to 0.68.

The median mass of the primary components in our SEB sample is $1.06\text{M}_\odot$ with a 16th and 84th percentile of $0.75\text{M}_\odot$ and $1.31\text{M}_\odot$, respectively. The median mass of the secondary components is $0.32\text{M}_\odot$ with a 16th and 84th percentile of $0.17\text{M}_\odot$ and $0.49\text{M}_\odot$, respectively. There are 10 SEB systems which have a median eccentricity $\lesssim 0.05$ but BCIs largely consistent with zero. SEB system (TIC 285108067) is the only non-circular orbit in our sample, with a median eccentricity of 0.35.

Our framework infers tight posterior distributions for the component masses in each SEB. The primary masses in our sample of 12 SEBs have a median uncertainty of 3.81%, with a 16th and 84th percentile of 3.68% and 4.63%, respectively. The secondary masses have a median uncertainty of 2.54%, with a 16th and 84th percentile of 2.20% and 3.80%, respectively.

The median values of the orbital solution period largely agree with the orbital periods derived from the light curve BLS periodigram analysis. The median orbital period of our SEB sample is 5.25 days with a 16th and 84th percentile of 2.73 days and 11.07 days, respectively. We briefly describe the inferred parameters and orbital structure of the 12 systems in our sample below, ordered by their TIC ID.

TIC-ID	M_1	R_1	M_2	R_2	P	a	ecc	q
	$M_\odot$	$R_\odot$	$M_\odot$	$R_\odot$	$days$	$R_\odot$	$-$	$-$
20215452	$0.69^{+0.04}_{-0.04}$	$0.79^{+0.02}_{-0.02}$	$0.33^{+0.02}_{-0.01}$	$0.35^{+0.01}_{-0.01}$	$0.59576^{+8e-07}_{-7e-07}$	$3.0^{+0.05}_{-0.04}$	$0.0112^{+0.0063}_{-0.0112}$	$0.48^{+0.01}_{-0.02}$
28159019	$1.23^{+0.05}_{-0.05}$	$1.99^{+0.04}_{-0.04}$	$0.34^{+0.01}_{-0.01}$	$0.35^{+0.01}_{-0.01}$	$4.61877^{+3.6e-06}_{-4.1e-06}$	$13.57^{+0.18}_{-0.17}$	$0.0018^{+0.0009}_{-0.0018}$	$0.28^{+0.0}_{-0.01}$
99254945	$0.77^{+0.03}_{-0.03}$	$0.71^{+0.01}_{-0.01}$	$0.33^{+0.01}_{-0.01}$	$0.32^{+0.0}_{-0.0}$	$7.70784^{+2.83e-05}_{-2.58e-05}$	$16.97^{+0.19}_{-0.19}$	$0.0052^{+0.0005}_{-0.0004}$	$0.43^{+0.01}_{-0.01}$
126232983	$0.89^{+0.03}_{-0.03}$	$1.65^{+0.09}_{-0.08}$	$0.14^{+0.0}_{-0.0}$	$0.14^{+0.01}_{-0.01}$	$4.53432^{+3.91e-05}_{-4.35e-05}$	$11.66^{+0.14}_{-0.14}$	$0.003^{+0.0014}_{-0.003}$	$0.16^{+0.0}_{-0.0}$
144441148	$0.82^{+0.03}_{-0.04}$	$1.76^{+0.12}_{-0.09}$	$0.14^{+0.0}_{-0.0}$	$0.71^{+0.08}_{-0.08}$	$5.81856^{+5.07e-05}_{-4.03e-05}$	$13.41^{+0.14}_{-0.18}$	$0.0169^{+0.0037}_{-0.0021}$	$0.17^{+0.0}_{-0.0}$
164458426	$1.54^{+0.06}_{-0.06}$	$1.93^{+0.05}_{-0.05}$	$0.2^{+0.01}_{-0.0}$	$0.21^{+0.01}_{-0.01}$	$4.69331^{+1.35e-05}_{-1.29e-05}$	$14.2^{+0.19}_{-0.15}$	$0.0022^{+0.0011}_{-0.0021}$	$0.13^{+0.0}_{-0.0}$
164527723	$1.15^{+0.04}_{-0.04}$	$1.85^{+0.05}_{-0.04}$	$0.18^{+0.01}_{-0.01}$	$0.26^{+0.01}_{-0.01}$	$2.73987^{+3e-06}_{-2.9e-06}$	$9.06^{+0.12}_{-0.1}$	$0.0097^{+0.0109}_{-0.0097}$	$0.16^{+0.01}_{-0.01}$
165453878	$1.07^{+0.05}_{-0.04}$	$1.14^{+0.02}_{-0.02}$	$0.2^{+0.0}_{-0.01}$	$0.22^{+0.0}_{-0.0}$	$2.7076^{+5.3e-06}_{-5.3e-06}$	$8.84^{+0.14}_{-0.11}$	$0.0019^{+0.0017}_{-0.0019}$	$0.19^{+0.0}_{-0.0}$
169820068	$1.2^{+0.05}_{-0.05}$	$4.29^{+0.08}_{-0.09}$	$1.04^{+0.02}_{-0.02}$	$1.13^{+0.03}_{-0.03}$	$29.74447^{+0.0001755}_{-0.0001856}$	$52.85^{+0.58}_{-0.52}$	$0.01^{+0.0012}_{-0.0011}$	$0.87^{+0.01}_{-0.02}$
258108067	$0.69^{+0.03}_{-0.02}$	$1.62^{+0.02}_{-0.02}$	$0.44^{+0.01}_{-0.01}$	$0.5^{+0.01}_{-0.01}$	$12.96863^{+0.0001022}_{-0.0001597}$	$24.14^{+0.24}_{-0.24}$	$0.3477^{+0.001}_{-0.0013}$	$0.63^{+0.01}_{-0.01}$
271548206	$1.05^{+0.04}_{-0.04}$	$0.93^{+0.02}_{-0.02}$	$0.36^{+0.01}_{-0.01}$	$0.42^{+0.01}_{-0.01}$	$7.44835^{+1.41e-05}_{-1.21e-05}$	$17.99^{+0.19}_{-0.21}$	$0.0322^{+0.0022}_{-0.0019}$	$0.35^{+0.01}_{-0.0}$
272074664	$1.61^{+0.1}_{-0.07}$	$1.9^{+0.04}_{-0.03}$	$0.7^{+0.02}_{-0.02}$	$0.63^{+0.01}_{-0.01}$	$10.47626^{+5.7e-06}_{-5.8e-06}$	$26.63^{+0.45}_{-0.34}$	$0.0546^{+0.0005}_{-0.0005}$	$0.44^{+0.01}_{-0.01}$

Table 3.2: Relevant posterior distribution values for orbital parameters of the 12 SEBs considered in this work, ordered in ascending TIC number. The uncertainties for all parameters listed are calculated from the Highest Density Interval (HDI) for each parameter posterior distribution at a 68% probability (The HDI can be considered as the minimum width Bayesian Credible Interval)

TIC 20215452 This system has the shortest orbital period of the systems in our sample, with a period of about $P \approx 0.60$ days. The median of the eccentricity posterior distribution is about 0.01. The computed semi-major axis of the system is 2.99 $R_\odot$. The primary component has a mass of $M_1 \approx 0.69$ $M_\odot$. The mass ratio q^2 of this system is about $q \approx 0.48$.

TIC 28159019 This system has a significantly curved primary eclipse, while its secondary transit displays a flatter eclipse. This system has the lowest inferred eccentricity in our sample with a posterior distribution mode of 0.002 and a credible interval on the eccentricity posterior that is consistent with zero. The computed semi major-axis of the system is 13.57 $R_\odot$. The orbital period of this system is $P \approx 4.62$ days. The primary component mass is $M_1 \approx 1.23 M_\odot$ and the system's mass ratio is $q \approx 0.28$.

TIC 99254945 This binary system has a computed orbital period of about $P \approx 7.71$ days. The semi-major axis of this system has been calculated to be 16.97 $R_\odot$. The median of the eccentricity posterior distribution is found to be 0.005. The primary mass is found to be $M_1 \approx 0.77$ $M_\odot$. The mass ratio for this binary system is $q \approx 0.43$. This system has the deepest observed primary transit (-18.52%) in our sample.

TIC 126232983 This binary system has a very shallow and inconspicuous secondary transit. The orbital period of this system is about $P \approx 4.53$ days. The median of the eccentricity posterior distribution is about 0.003 but the credible interval is consistent with zero. The semi-major axis of this system has

---

[2]We define q to be (M_2/M_1)

been computed to be 11.66 $R_\odot$. The primary component mass is $M_1 \approx 0.89\ M_\odot$. The mass ratio is $q \approx 0.161$.

TIC 144441148 This binary system is the only binary system in our sample which has a non-existent secondary transit in the light curve. The orbital period for this system was computed to be $P \approx 5.82$ days. The semi-major axis was computed to be 13.41 $R_\odot$. The median of the eccentricity posterior distribution was found to be 0.02. The primary component mass was found to be $M_1 \approx 0.82\ M_\odot$. The mass ratio is $q \approx 0.17$. Our framework infers a secondary component mass of $M_1 \approx 0.14\ M_\odot$ but a radius of $R_2 \approx 0.71\ M_\odot$. We take this radius inference as un-reliable as it is non-physical, as seen on the mass-radius plot in Figure 3.5 (black triangle). We discuss this un-reliable inference further in Section 3.5.

TIC 164458426 This binary system has some interesting periodic signals in the background flux that prohibit a visual observation of the secondary transit without subtraction of the background flux by a Gaussian process. The orbital period was computed to be $P \approx 4.69$ days. The median of the eccentricity posterior distribution was found to be 0.002 but the credible interval is consistent with zero. The semi-major axis was computed to be 13.42 $R_\odot$. The primary component mass was found to be $M_1 \approx 1.54\ M_\odot$. The mass ratio is $q \approx 0.13$.

TIC 164527723 This binary system has a computed orbital period of $P \approx 2.71$ days. The semi-major axis was computed to be 9.06 $R_\odot$. The median of the eccentricity posterior distribution was found to be 0.010 but the credible interval is consistent with zero. The primary component mass was found to be $M_1 \approx 1.10\ M_\odot$ and the mass ratio is $q \approx 0.16$.

TIC 165453878 This binary system has a computed orbital period of $P \approx 2.71$ days. The median of the eccentricity posterior distribution is found to be 0.002 but with a credible interval consistent with zero. The semi-major axis was computed to be 8.84$R_\odot$. The primary mass component was found to be $M_1 \approx 1.07\ M_\odot$ and the mass ratio was found to be $q \approx 0.19$.

TIC 169820068 This binary is the only system in our sample where the APOGEE-derived period and epoch transit time served to fold the RV phase curve and flux time-series phase curve better than the TESS light curve derived values. The orbital period was found to be $P \approx 29.74$ days, making it the longest period binary in our sample. The semi-major axis is computed to be about 52.85 $R_\odot$. The median of the eccentricity posterior distribution was found to be 0.01. The primary component mass was found to be $M_1 \approx 1.20\ M_\odot$. The secondary component mass was found to be $M_2 \approx 1.04\ M_\odot$. The mass ratio was found to be $q \approx 0.87$ This is also the only system in our sample for which the final orbital

model produces a larger secondary transit than the primary transit. The primary component radius is $R_1 \approx 4.29\ R_\odot$, which is much larger than main-sequence mass-radius relations. These peculiarities raise the implication that the primary component star is an evolved star.

TIC 258108067 This binary system has the highest median eccentricity in our sample at 0.35. This system necessitated the relaxation of automatically setting a ceiling on $e \cos \omega$ of 0.01 in the initialization of the orbital model. The orbital period was computed to be $P \approx 12.97$ days. The semi-major axis is computed to be $24.14\ R_\odot$. The primary component mass and mass ratio were found to be $M_1 \approx 0.69\ M_\odot$ and $q \approx 0.63$, respectively.

TIC 271548206 This binary system has very broad background flux in the light curve data, making the secondary transit more than difficult to visual identify. The orbital period was computed to be $P \approx 7.45$ days. The semi-major axis was computed to be $17.99\ R_\odot$. The median of the eccentricity posterior distribution was found to be 0.03. The primary component mass was found to be $M_1 \approx 1.05\ M_\odot$. The mass ratio was found to be $q \approx 0.35$.

TIC 272074664 This binary system has a computed orbital period of $P \approx 10.48$ days. The semi major was computed to be $26.63\ R_\odot$. The median of the eccentricity posterior distribution was found to be 0.05. The primary component mass was found to be $M_1 \approx 1.60\ M_\odot$, which is the largest in our sample. The mass ratio was found to be $q \approx 0.44$.

3.5 Discussion

We demonstrated here the flexibility of our framework to derive complete orbital parameter posterior distributions for a sample of 12 different SEBs with a varying number of radial velocity measurements and different light curve transit geometries. There are 11 binary systems in our sample did not have model framework settings that required further manual fine-tuning before sampling took place. Only one system, TIC 285108067, required manual intervention in relaxing the circular eccentricity setting. Our framework only requires a TESS TIC ID in order to be initialized, allowing a list of EB systems to be modeled simultaneously.

The final orbital models constructed from our inferred orbital parameters and plotted over the TESS + APOGEE data in Figure 3.3 highlight the varied configurations of SEBs that our framework is able to model. Figure 3.3 also illustrates the power of jointly considering the TESS + APOGEE data. In the case of a very noisy out-of-transit flux and a very shallow secondary transit, such as TIC 144441148, the well sampled radial velocity measurements likely bolsters the framework's ability to constrain the orbital parameters.

Our framework is also capable of modeling SEBs with sparsely observed secondary transits which is the case for the systems TIC 169820068 and TIC 258108067 (see Figure 3.8 in the Appendix). TIC 258108057,
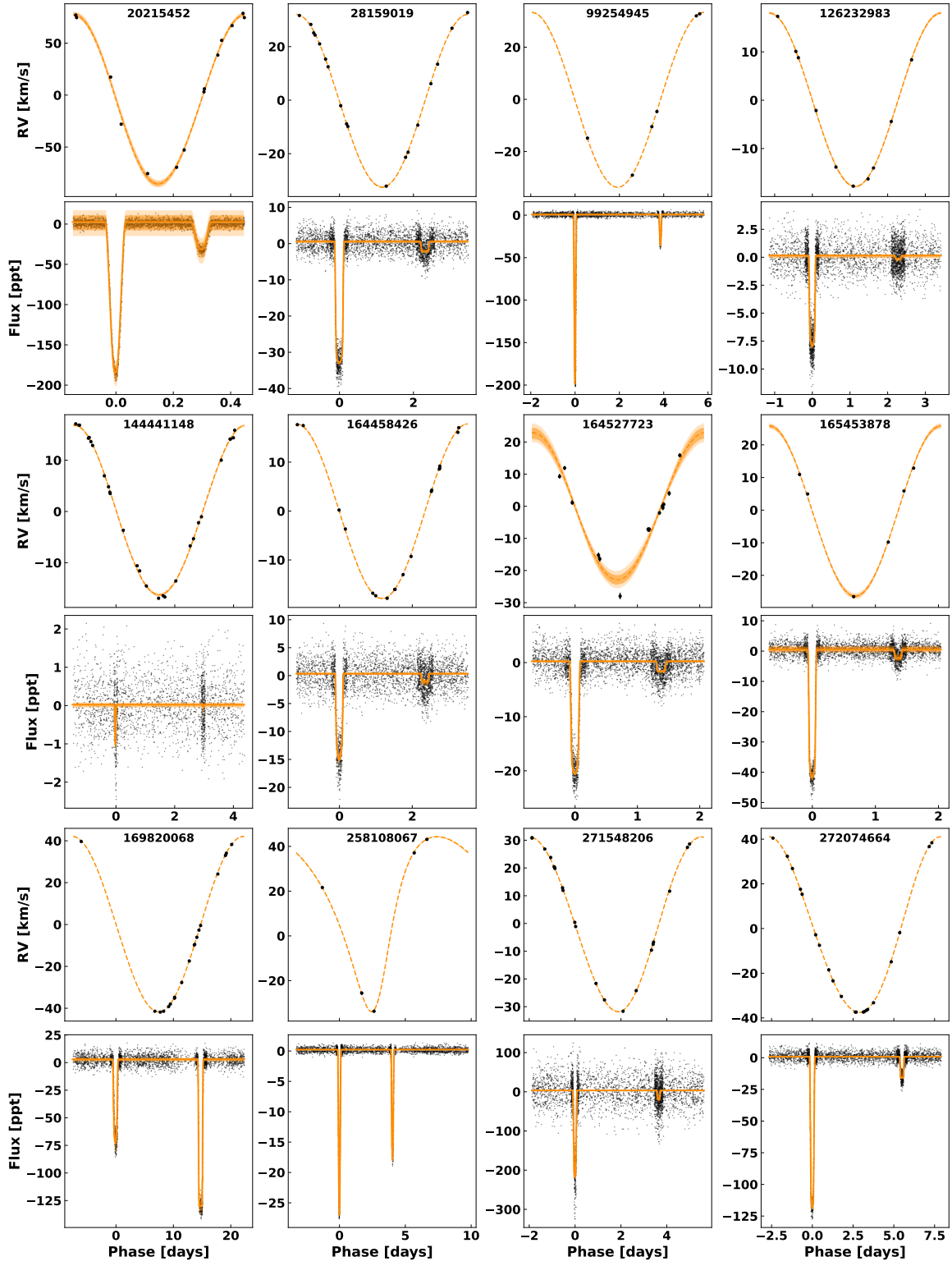
Figure 3.3: Plotted phased radial velocity - light curve for the 12 modeled TESS–APOGEE SEBs in this work. Even rows are the radial velocity phase curves for the 12 systems. Odd rows are the transit light curves for the 12 systems. The data points in each system are plotted as black dots. The final orbital model in both the light curve and radial velocity phase curves is plotted in orange. The 68% and 95% confidence intervals plotted as transparent orange fill. Our framework is able to reliably infer orbital parameter posterior distributions for SEBs with varied data.

54
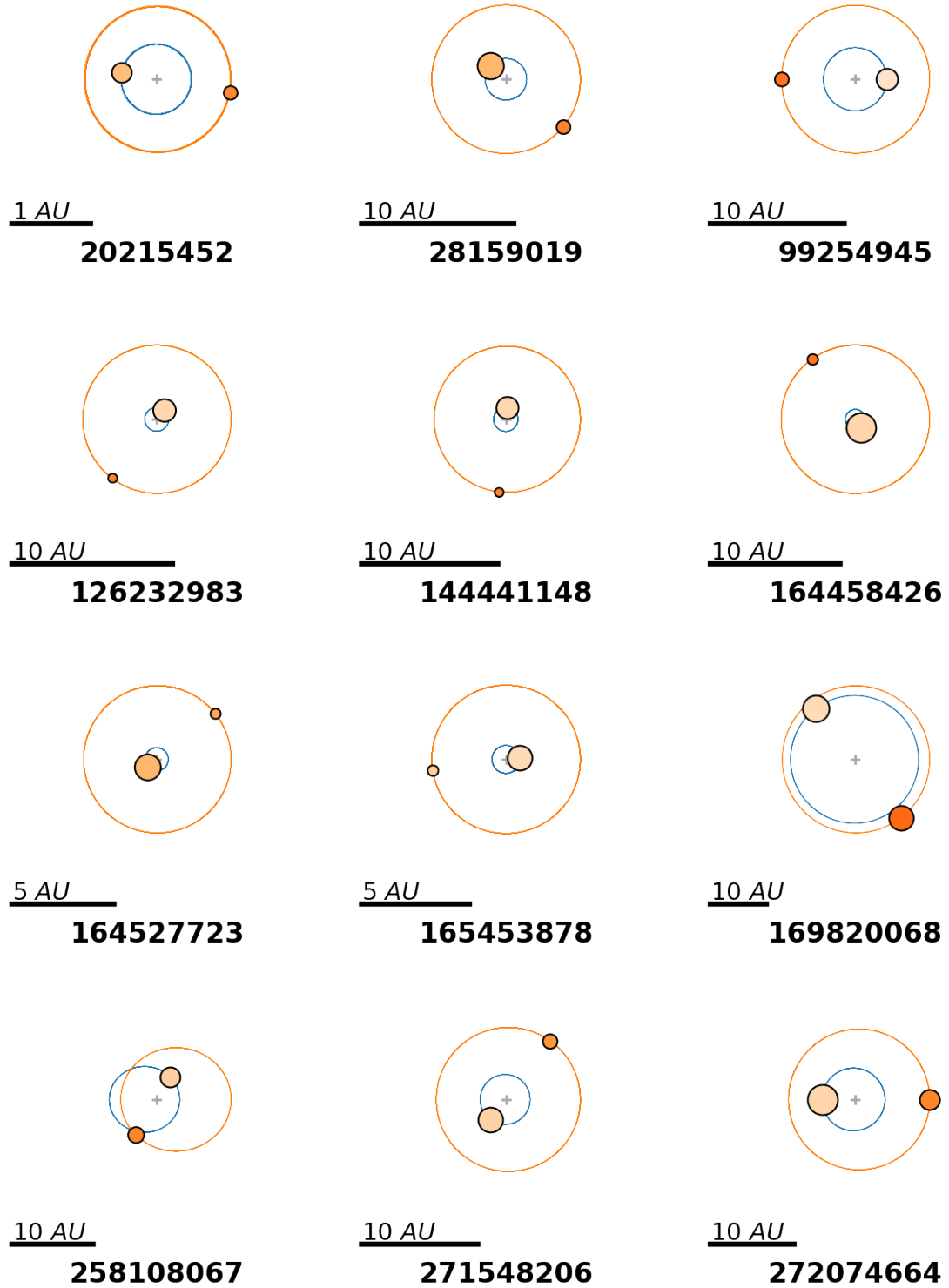
Figure 3.4: Face-on plot of all 12 TIC SEBs considered in this work, ordered by increasing TIC-ID. The primary component is plotted on the blue orbit, the secondary component is plotted on the orange orbit. The relative symbol sizes of all binary components are to scale with each other. The relative sizes of plotted component symbols are not to scale to the plotted orbits. Each plot has a small scale indicator of either 1 AU, 5 AU, or 10 AU. Each SEB has it's TIC-ID listed below it's orbit.

in particular, has a number of characteristics that highlight the flexibility of our framework: it has only 5 radial velocity measurements, and has a sparsely observed secondary transit in the light curve data, and is an eccentric binary.

We have plotted the mass-radius contours created from the posterior distributions of the 12 SEBs we modeled in this work in Figure 3.5. We have plotted the masses and radii of the 95 eclipsing binaries which were characterized in Torres et al. (2010a) in Figure 3.5 as well. The Torres et al. (2010a) binaries trace out a mass-radius main sequence that most of the median mass and radii estimates for our sample fall within as well.

We have marked out the minimum mass and radius measured in the Torres et al. (2010a) EB sample in Figure 3.5 as gray dash lines. The inferred primary masses and radii in our sample fall within this range of masses and radii. There are 4 secondary components which have masses and radii that fall below these ranges. These lower masses and radii highlight the higher overall sensitivity that is achievable by jointly modeling both the TESS + APOGEE data with a Bayesian probabilistic framework.

3.5.1 Spectroscopic Eclipsing Binaries populations as benchmarks

Recently there has been increased work into measuring the parameter distributions of binary populations using Bayesian hierarchical inference. Bayesian hierarchical inference is performed similarly to 'standard' Bayesian inference with a prior and a likelihood informing a posterior. The 'hierarchical' aspect refers to the prior being modifiable during sampling with a 'hyper-prior' informing on what shape the prior distribution might have. The prior distribution shape is then inferred as sampling takes place.

Bayesian hierarchical inference has been used to explore the a number of different characteristic distributions important to the understanding of binaries. Price-Whelan et al. (2020) measured the long period binaries eccentricity distributions with results from *The Joker*. Lyttle et al. (2021) measured the distributions of stellar helium abundance (Y) using Kepler dwarfs and sub-giants in a Bayesian hierarchical model. Hwang et al. (2022) used Bayesian hierarchical inference to measure the eccentricity distribution of wide binaries using astrometric data from *Gaia* DR3.

Our framework generates full posterior distributions for all binary orbital parameters rather than just a single value for the peak (or median) and the range of an appropriate interval. The complete posterior distributions of these orbital parameters can be used to perform more robust Bayesian hierarchical inference. The shape of the posterior distributions of each population member in a Bayesian hierarchical inference can better inform the shape of the hyper-prior of interest.

Our framework is able to infer full orbital parameter estimates for SEBs with single-component radial velocities. Single-lined SEBs are the more commonly observed than double-lined SEBs. This arises from the
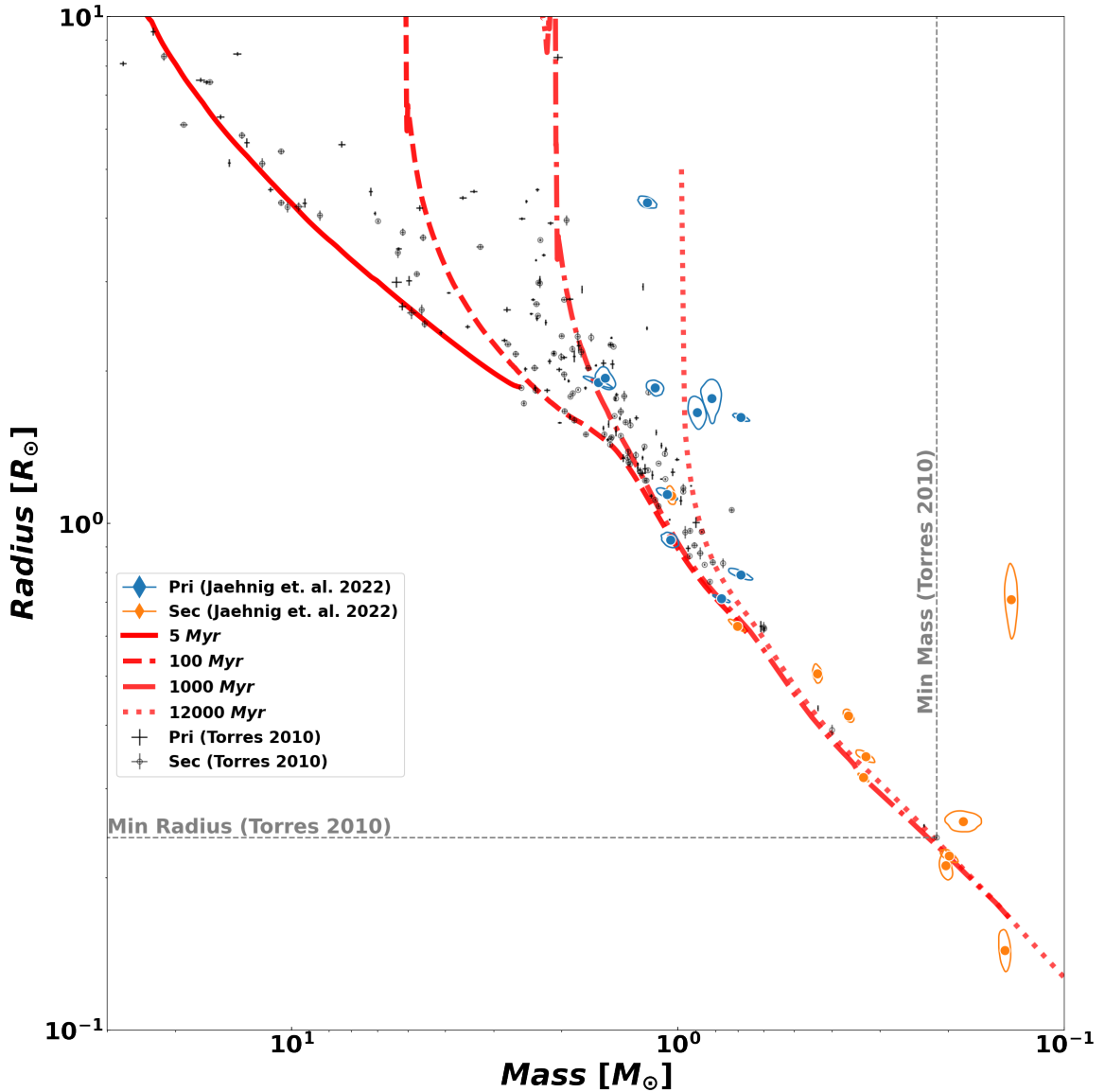
Figure 3.5: Plot of the masses and radii for primary (blue diamonds) and secondary (orange diamonds) components in the 12 SEBs considered in this work. The 1σ contours (blue and orange lines) for each binary component's mass and radius posterior distributions are plotted to highlight the covariance between mass and radius. We plot the mass-radius measurements from Torres et al. (2010a) for the primary (filled black circles) and secondary (empty black circles) components. The vertical/horizontal gray dashed line demarcate the minimum determined mass/radius from the Torres et al. (2010a) catalog, respectively. We plot mass-radius values of a solar metallicity MIST isochrone at 5 Myr (solid red line), 100 Myr (dash red line), 1000 Myr (dashed dot red line), and 12000 Myr (dotted red line).

57

lower probability of observing shifts from both stars in the spectra unless directed long-term spectroscopic observations are carried out. Double-lined SEBs are also not as frequently observed as single-lined SEBs since single-lined SEBs occur more frequently across the range of mass ratio 'q'. Double-line SEBs are usually only observed when the masses of both components are near-equal. The more frequently occurring, and more easily observed single-lined SEBs can be used as population level calibrators for stellar evolution models.

Our framework currently only focuses on modeling single-lined SEBs, and in future work we will expand our framework to also model double-lined SEBs. The addition of a secondary set of radial velocity measurements will result in tighter posterior distributions of the inferred parameters. This will only increase our framework's capabilities for flexible SEB parameter inference and its potential for binary population parameter inference.

3.5.2 Comparison with results from *Gaia* DR3

The recent release of the *Gaia* DR3 (Prusti et al., 2016; Vallenari et al., 2022) there has been a massive quantity of eclipsing, astrometric, and spectroscopic binary data released taken by the *Gaia* spacecraft and processed by the *Gaia* collaboration. We some compare our orbital parameters and stellar parameters for the binary systems which were bright enough in the *Gaia* passbands to have available data in Figure 3.6. In the left most panel of Figure 3.6 we have plotted the derived orbital period for our binary systems on the x-axis and the *Gaia* DR3 periods on the y-axis. The residuals of the two period measurements lie in the range between $\pm 2 \times 10^{-3}$ days, indicating excellent agreement between the two.

In the middle panel of Figure 3.6 we have plotted the orbital eccentricity for our binary systems (x-axis) as well as the *Gaia* DR3 eccentricities (y-axis). It appears that our framework infers more circularized eccentricities than the *Gaia* DR3 measurements. This is true for all the plotted binary systems except TIC 281508067 which has the highest eccentricity in our sample; both the *Gaia* DR3 eccentricity as well as our framework inferred eccentricity agree very well. Error-bars are plotted, though smaller than plot symbols for our measurements of eccentricity.

In the right most panel of Figure 3.6 we have plotted the inferred surface gravity, $\log g$, calculated from our framework (x-axis) for the primary stars in our sample against their $\log g$ (column=LOGG) measurements from APOGEE (Majewski et al., 2017) DR17 (y-axis). The spectrographic $\log g$ APOGEE measurements show good agreement within errors with the Bayesian framework derived $\log g$ values derived in this work.

3.5.3 Possible radius inflation in secondary components

There is a developing consensus that there are observed deviations in measured radii of low-mass stars ($<1\ M_\odot$) from the radii predicted by stellar evolutionary models. The mechanism that generates this observed 'radius inflation' is still not entirely clear but it's believed to be originating from strong magnetic fields interrupting the convective process, large star spots affecting the pressure and temperature of a star, or a combination of both. Radius inflation can cause a single star to move vertically on the Hertzsprung-Russell diagram because of changes in luminosity and be mistakenly identified as belonging to the binary main-sequence, thereby biasing photometric studies of binaries, especially in open clusters. Moreover, radius inflation can directly impact exoplanet characterizations as precision of the transit parameters from a light curve are greatly driven by the precision of the stellar radius.

There are some secondary components near the lower end of the mass range in Figure 3.5 that lie above the expected main sequence that is outlined by the MIST isochrones also plotted (red lines). We speculate that these secondary stars might be undergoing radius inflation and perform a rudimentary analysis to explore that possibility.

We generate several thousand theoretical radii for secondaries with component masses $<1\ M_\odot$ using the MIST isochrones in the Isochrones package. These isochrones span a uniform grid in age, consider only the main-sequence phases of stellar evolution ($202 >$ EEP > 654), and use the APOGEE derived [Fe/H] for each system. We calculate the median percent deviation of our framework's inferred radii from the theoretical MIST radii we generate.

We find five secondary components in our sample may be undergoing radius inflation. The secondary components in TIC 20215452 and TIC 28159019 have radii that are inflated on the order of $\sim$7% and $\sim$6%, respectively. The other three secondary components of TIC 285108067, TIC 271548206, and TIC 164527723 have larger inflated radii on the order of $\sim$20%, $\sim$16%, and $\sim$17%, respectively. This analysis is largely speculative and non-exhaustive but it does illustrate the versatility of this framework in inferring component radii at the low-mass end.

3.5.4 Non-physical stellar parameter inferences

Our framework is able to infer orbital parameter solutions for detached SEB systems that are largely agnostic of stellar evolutionary models. However this also opens up the framework to possibly inferring non-physical parameters in some solutions. This is the case for SEB system TIC 144441148. The median inferred mass and radius for the secondary component are $M_2 \approx 0.14\ M_\odot$ and $R_2 \approx 0.71\ R_\odot$, respectively. These stellar parameters are non-physical for a star, given its position (black triangle) on the mass-radius plot in Figure 3.5. The mass-radius plot can be considered an analog for the Hertzsprung-Russel diagram, and as such, there
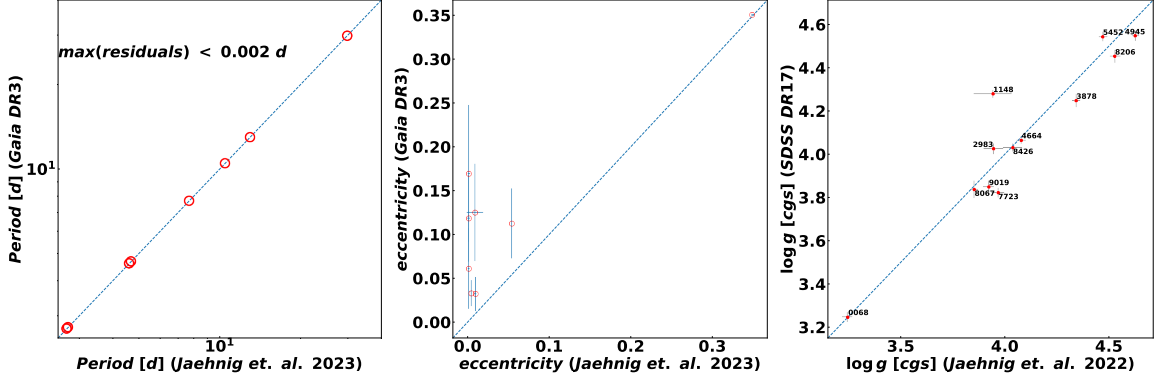
Figure 3.6: Plot of three separate orbital parameters as well as stellar parameters one to one plots comparing the inferred values for our binary systems and the values derived by the *Gaia* collaboration from their recently released *Gaia* DR3 (Holl et al., 2022) as well as APOGEE(Majewski et al., 2017). Panel A is comparing the orbital period of the binary system, and each system for which there is a period measurement is labeled by the last four digits of its `TIC` id. Panel B is a similarly labeled plot comparing the orbital eccentricity for our binary systems. Panel C is a stellar parameter comparison plot of the inferred surface gravity, $\log g$, of the primary stars in our sample with the $\log g$ values calculated by APOGEE.

should not be stars in that area of the mass-radius space.

We consider the secondary component mass to be reliable since the component masses are derived from orbital dynamics constrained by the radial velocity measurements. TIC 144441148 is one of the SEB systems in our sample with over 20 APOGEE radial velocity measurements. The model radial velocity phase curve agrees very well with the radial velocity measurements in Figure 3.3[3].

The source of this non-physical solution is likely a result of the light curve data for this system, which constrains the primary component radius and secondary component radius via the radius ratio prior $\log k$. The light curve for this SEB does not appear to have secondary transit present, which might indicate a grazing eclipsing binary. Our framework infers a large impact parameter of $b \approx 1.38$ which further suggests a grazing eclipse for near circular orbits.

There are a number of different sources for an eclipsing binary to have a grazing orbit. A third orbiting body may induce perturbations in the orbit of the binary, affecting the transit depths. Stellar pulsations may cause changes in the stellar radii and affect the inference of $\log k$. Stellar pulsations can be found by considering the out-of-transit flux in a light curve. Our framework only considers the out-of-transit background flux as a nuisance parameter in its current implementation by design. Future implementations of this framework would have to address the particularities of grazing eclipsing binaries.

---

[3]column 1, row 3

3.5.5 Covariances within the posteriors

We find most of the posterior distributions of the inferred orbital parameters for these systems to be well-behaved. However we do visually identify a covariance between $\log M_1$, $\log q$, and $\log R_1$. This covariance appears to be present in all posterior distributions, even in systems where the posterior distributions are well behaved (i.e. a defined unimodal shape with little to no covariance between other parameters).

We plot an example of 'well behaved' posterior distributions in the form of a corner plot for TIC 272074664 in Figure 3.7. In the corner plot of $\log R_1$ and $\log M_1$ (Figure 3.7, row 2, column 1) we can see the skewed covariance present between the these two orbital parameters.

This skewed contour shape is present after transforming to the $\log M_1$ and $\log R_1$ posterior values to M_p and R_p, respectively. Figure 3.5 shows the still skewed shape of the transformed $\log M_1$ and $\log R_1$ posteriors. This skewed shape is indicated that the $\log M_1$ and $\log R_1$ priors are influenced by their relation in defining the primary star's density, which is an important value attainable through light curve analysis. The shape of the secondary components mass-radius contours are also skewed despite no sampling taking place over any secondary component parameters.

The skewed shape of the secondary component's mass and radius stems from being calculated from multiplying (or adding in log space where we sample) the mass ratio q by the primary component mass. The skewed shape of the sampled $\log q$ posteriors broadcast into the calculated secondary component mass. Additionally, the secondary component radius is calculated from multiplying the radius ratio k by the primary component ratio, reproducing a similarly skewed shape.

3.6 Conclusion

In this work we have introduced a novel framework to infer orbital parameters for 12 spectroscopic eclipsing binaries (SEBs). We use available photometric TESS data and the spectroscopic APOGEE data together to infer the parameters of a two-component binary orbital model within a Bayesian Probabilistic framework.

The TESS light curves are downloaded using the LightKurve package and a high resolution box-least squares (BLS) fit is performed to infer an initial orbital period and time of transit. The TESS light phase curve and the APOGEE radial velocity phase curve are centered and folded by this initial orbital period and time of transit for 11 out of the 12 binary systems. We find the two similar parameters derived using only radial velocities within the *The Joker* catalog (Price-Whelan et al., 2020) resulted in a better folded phase curve for TIC 169820068.

This framework uses Hamiltonian Monte Carlo (HMC) to perform sampling of posteriors. We select HMC to perform sampling within our framework because of its reliability to properly explore a parameter space, as well as scaling well to large number of dimensions. Our framework is constructed to be able to

perform all necessary operations to model an binary system without further manual modification because of the built in capabilities of Exoplanet that we construct our framework upon.

We run our framework around a small sample of 12 spectroscopically eclipsing binaries. We use the Isochrones package to derive informed values for the parameters that we use to initialize our priors for each binary orbital model. We recover reliable SEB model parameter distributions for 11 of our binary systems using the same framework with almost no manual intervention necessary. Only one of our systems has a partially reliable model parameter inference, due to a non-physical secopndary radius. We will continue this work and scaling our framework to fit the larger sample of binaries within both the APOGEE and TESS surveys in order to construct a benchmark catalog of SEBs with fully modeled orbital parameters.
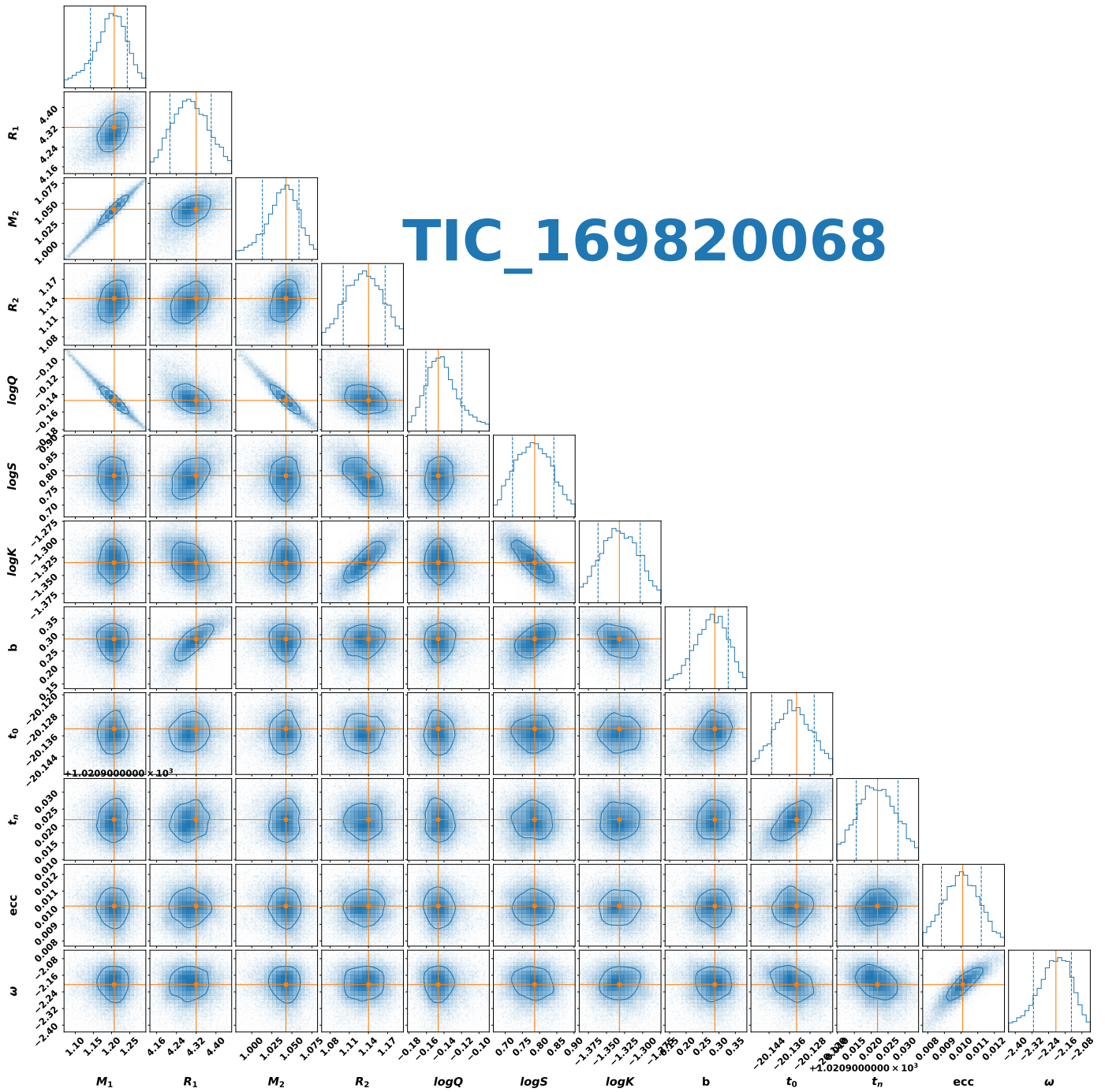
Figure 3.7: Corner plot of posterior values of $(\log M_1, \log R_1, \log q, \log s, b, t_0, t_N, ecc, \omega)$ for binary system TIC 196820068. The 2-dimensional contour on each corner is the equivalent 1-σ. The orange lines on each contour plot are the final MAP values for each variable after the 2nd round of MAP optimizations (see Section 3.3.6). The left and right vertical dashed blue lines on each 1-dimensional histogram plots of each variable are the 16th and 84th quantiles, respectively.
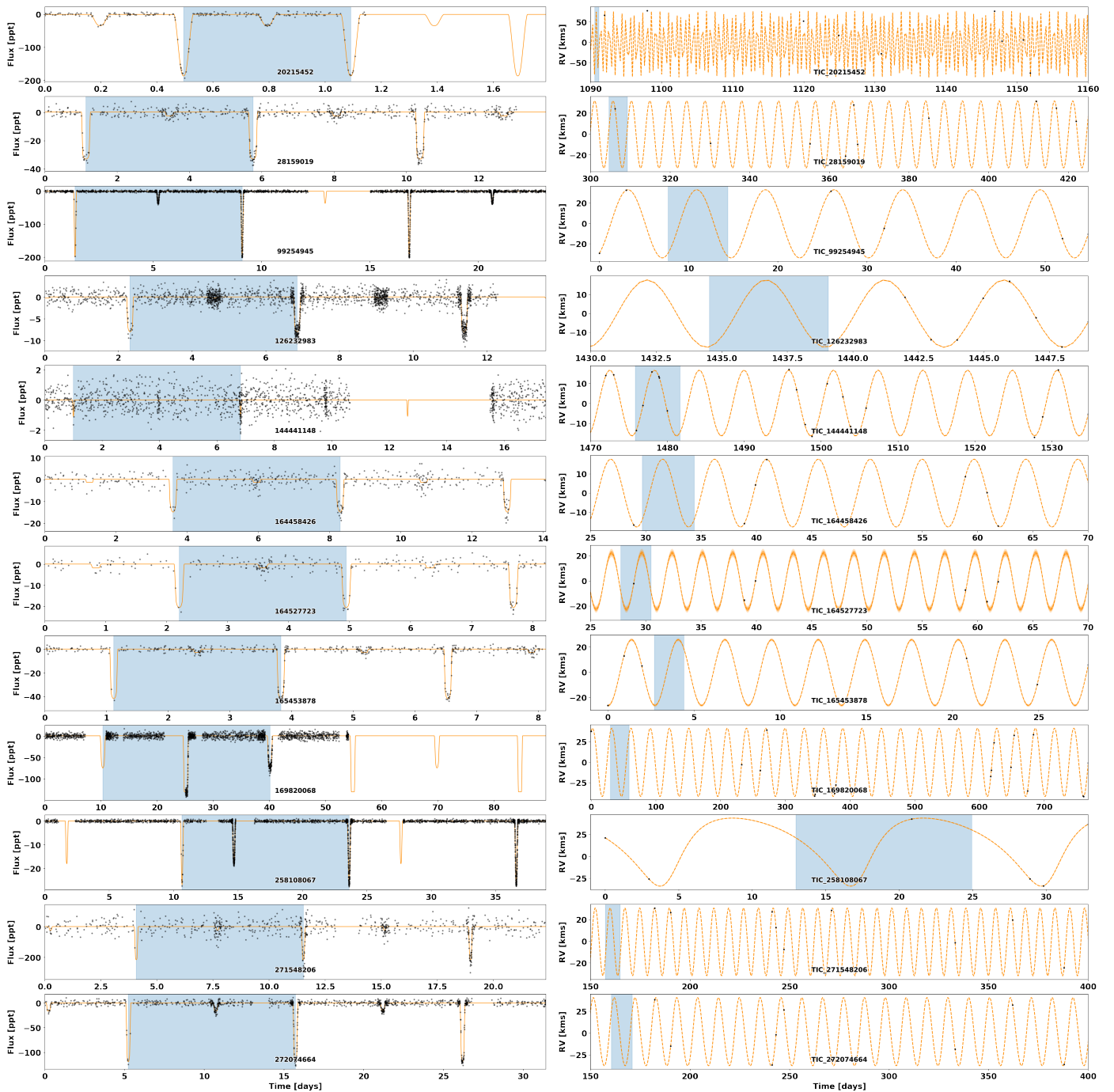
Figure 3.8: Side by side plots of all light curves (left panel) and all radial velocity curves (right panel). Both the radial velocity data and the light curve data have been plotted chronologically. The orbital model constructed with the best parameters are used to produce both the model transit light curve (left panel, orange line) and the radial velocity curve (right panel, orange line). Light curve data and radial velocity data are plotted in both panels as black dots. Each system also has its full orbital period is plotted as blue shading. One full orbital period is plotted for each plot starting from either the first transit or the start of radial velocity observations.

64

CHAPTER 4

Discussion

This work encompasses machine learning frameworks developed and applied to perform: Detection and identification of open cluster (OC) members from astrometric data. Inference of orbital parameters of single-lined spectroscopic eclipsing binaries (SEB1s) using different, sparse time-series data.

In Chapter 2 the Gaia astrometric measurements of α, δ, and ϖ, were used with their accompanying measurement uncertainties to perform unsupervised clustering using a variant of Gaussian mixture models, extreme deconvolution Gaussian mixture models (XDGMM) to find open cluster groups and determine their individual stellar members. The high precision of Gaia DR2 astrometry can be fully employed in XDG-MMs with the full 3x3 covariance matricies for all stars to increase sensitivity to lower mass, fainter cluster members, which typically have higher measurement uncertainties. This XDGMM framework used two other metrics that were used to increase flexibility in the framework to automatically fit hundreds of lines of sight. The Bayesian Information Criterion served to automatically select the best number of Gaussian components to fit in any line of sight. The Differential Entropy was used to quantify which Gaussian component had the smallest volumetric 'spread' in the parameter space under the assumption that the open cluster would be the most concentrated against the more uniform field stars.

These metrics were successfully employed in majority of open cluster classifications in this work, but improvements in future works can greatly improve their efficacy. The Bayesian Information Criterion marginalizes the XDGMM fit by the number of components used but this takes into account the entire data set, including field and open cluster stars. This led to a small number of instances in the framework where more components were used than was necessary. This also revealed that some of the open clusters could be better modeled by fitting more than just one Gaussian component. Indeed, this has started appearing in the literature, with Tarricq et al. (2022) demonstrating it is possible to fit a 'core' and a 'halo' Gaussian mixture component to open clusters to better characterize sub-structure. This highlights that a better metric in the future will better measure how well the model is performing solely centering the open cluster and not the entire line of sight.

Similarly the use of differential entropy as a cluster component detection method leaves room for improvement in future works. For example, the differential entropy could not be reliably used in a blind search for open clusters. Its usage in this work relied on the assumption that in the open cluster component has the lowest differential entropy compared to the surrounding distribution of field stars. The few times this assumption was broken in this work was when the line of sight data set contained another, previously identified, open

cluster, or an undiscovered open cluster candidate.

There have been numerous ventures into a more encompassing method to identify true open clusters. This comes as many previously designated open clusters in pre-Gaia catalogs have been found to be asterisms once they were reanalyzed with higher precision Gaia astrometry (Cantat-Gaudin and Anders, 2019). Neural networks have been presented as a worthwhile method to identify open clusters. Neural networks are capable of learning non-linear relationships within data, and can correctly learn to identify open cluster features present the color-magnitude distribution. Cantat-Gaudin et al. (2020) used a dense neural network to predict open cluster parameters (age, extinction A$_v$, distance) based on a training set composed of identified open clusters and simulated open clusters. Hunt and Reffert (2023) employed a Bayesian convolutional neural network to perform True/False positive validations of open clusters identified in Gaia DR3 data, and compare the neural networks efficacy to human identifiers.

In Chapter 3 a flexible Bayesian framework was constructed using the Exoplanet python package to perform inference of single-lined Spectroscopic Eclipsing binary (SEB1s) orbital parameters using time-series flux data from TESS and sparse time-series radial velocities from APOGEE. This framework performed Bayesian inference using Hamiltonian Monte Carlo (HMC) sampling. HMC differs from other Markov Chain Monte Carlo methods by using gradient information of the parameter space to guide its sampling. This implementation also uses the "No U-Turn sampler" (NUTS) in order to both tune HMC hyperparameters as well as terminating sampling when the sampler has sufficiently explored the parameter space.

Future work on this framework would consist of expanding its flexibility as well as bolstering its ability to handle sparse data sets. Currently as little as 5 radial velocity measurements prove sufficient to infer reliable orbital parameters. However it may be possible to reduce that threshold to 3 radial velocity measurements given a more well parameterized orbital model. Removing the framework's reliance on a near circular eccentricity will also expand it's capabilities. TIC 258108067 was an eccentric binary which needed the circular eccentricity assumption manually relaxed for a successful fit to take place. This demonstrates that the framework *can* handle non-circular eccentricities, but a more nuanced treatment will be necessary to take advantage of the all of identified binaries in the *The Joker* catalog.

Further improvements to this framework can take place in the initial determination of period, epoch of first transit, and transit duration times from the TESS flux light curves. The box least squares (BLS) is very ubiquitous in constructing an accurate periodigram of a light curve. However there have been recent attempts at constructing a model that will better fit the transit flux data without the need for a high resolution frequency grid to search through. Hippke et al. (2019) developed 'Transit Least Squares' as a more optimized algorithm that also takes into account relevant physics present in a transit such as limb darkening. A better fitting model to the light curve will also produce more accurate estimates of transit duration lengths, improving light curve

masking in the framework discussed in Chapter 3.

As more and more high precision measurements are produced by surveys such as the Gaia satellite, JWST, and the upcoming Vera Rubin observatory, it is clear that automated methods of machine learning will be essential to perform analyses on such large and growing data sets to reach science goals. It is also clear that as the field pushes the limits of observations to the most distant and faintest objects that measurement uncertainties be addressed via Bayesian methods to ensure both uncertainties in measurements and models are considered. Bayesian methods of inference also advantageous when trying to infer parameters from sparse data sets. This is crucial in order to perform large uniform analyses of stellar populations such as SEB1s which have are more numerous in the literature than SEB2s. Automated Bayesian machine learning inference methods will continue to improve our understanding of star clusters and binary stars so that a fully realized star formation theory can be achieved.

References

Abdurro'uf, Accetta, K., Aerts, C., Silva Aguirre, V., Ahumada, R., Ajgaonkar, N., Filiz Ak, N., Alam, S., Allende Prieto, C., Almeida, A., Anders, F., Anderson, S. F., Andrews, B. H., Anguiano, B., Aquino-Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Ata, M., Aubert, M., Avila-Reese, V., Badenes, C., Barbá, R. H., Barger, K., Barrera-Ballesteros, J. K., Beaton, R. L., Beers, T. C., Belfiore, F., Bender, C. F., Bernardi, M., Bershady, M. A., Beutler, F., Bidin, C. M., Bird, J. C., Bizyaev, D., Blanc, G. A., Blanton, M. R., Boardman, N. F., Bolton, A. S., Boquien, M., Borissova, J., Bovy, J., Brandt, W. N., Brown, J., Brownstein, J. R., Brusa, M., Buchner, J., Bundy, K., Burchett, J. N., Bureau, M., Burgasser, A., Cabang, T. K., Campbell, S., Cappellari, M., Carlberg, J. K., Wanderley, F. C., Carrera, R., Cash, J., Chen, Y.-P., Chen, W.-H., Cherinka, B., Chiappini, C., Choi, P. D., Chojnowski, S. D., Chung, H., Clerc, N., Cohen, R. E., Comerford, J. M., Comparat, J., da Costa, L., Covey, K., Crane, J. D., Cruz-Gonzalez, I., Culhane, C., Cunha, K., Dai, Y. S., Damke, G., Darling, J., Davidson Jr., J. W., Davies, R., Dawson, K., De Lee, N., Diamond-Stanic, A. M., Cano-Díaz, M., Sánchez, H. D., Donor, J., Duckworth, C., Dwelly, T., Eisenstein, D. J., Elsworth, Y. P., Emsellem, E., Eracleous, M., Escoffier, S., Fan, X., Farr, E., Feng, S., Fernández-Trincado, J. G., Feuillet, D., Filipp, A., Fillingham, S. P., Frinchaboy, P. M., Fromenteau, S., Galbany, L., García, R. A., García-Hernández, D. A., Ge, J., Geisler, D., Gelfand, J., Géron, T., Gibson, B. J., Goddy, J., Godoy-Rivera, D., Grabowski, K., Green, P. J., Greener, M., Grier, C. J., Griffith, E., Guo, H., Guy, J., Hadjara, M., Harding, P., Hasselquist, S., Hayes, C. R., Hearty, F., Hernández, J., Hill, L., Hogg, D. W., Holtzman, J. A., Horta, D., Hsieh, B.-C., Hsu, C.-H., Hsu, Y.-H., Huber, D., Huertas-Company, M., Hutchinson, B., Hwang, H. S., Ibarra-Medel, H. J., Chitham, J. I., Ilha, G. S., Imig, J., Jaekle, W., Jayasinghe, T., Ji, X., Johnson, J. A., Jones, A., Jönsson, H., Katkov, I., Khalatyan, D. A., Kinemuchi, K., Kisku, S., Knapen, J. H., Kneib, J.-P., Kollmeier, J. A., Kong, M., Kounkel, M., Kreckel, K., Krishnarao, D., Lacerna, I., Lane, R. R., Langgin, R., Lavender, R., Law, D. R., Lazarz, D., Leung, H. W., Leung, H.-H., Lewis, H. M., Li, C., Li, R., Lian, J., Liang, F.-H., Lin, L., Lin, Y.-T., Lin, S., Lintott, C., Long, D., Longa-Peña, P., López-Cobá, C., Lu, S., Lundgren, B. F., Luo, Y., Mackereth, J. T., de la Macorra, A., Mahadevan, S., Majewski, S. R., Manchado, A., Mandeville, T., Maraston, C., Margalef-Bentabol, B., Masseron, T., Masters, K. L., Mathur, S., McDermid, R. M., Mckay, M., Merloni, A., Merrifield, M., Meszaros, S., Miglio, A., Di Mille, F., Minniti, D., Minsley, R., Monachesi, A., Moon, J., Mosser, B., Mulchaey, J., Muna, D., Muñoz, R. R., Myers, A. D., Myers, N., Nadathur, S., Nair, P., Nandra, K., Neumann, J., Newman, J. A., Nidever, D. L., Nikakhtar, F., Nitschelm, C., O'Connell, J. E., Garma-Oehmichen, L., Luan Souza de Oliveira, G., Olney, R., Oravetz, D., Ortigoza-Urdaneta, M., Osorio, Y., Otter, J., Pace, Z. J., Padilla, N., Pan, K., Pan, H.-A., Parikh, T., Parker, J., Peirani, S., Peña Ramírez, K., Penny, S., Percival, W. J., Perez-Fournon, I., Pinsonneault, M., Poidevin, F., Poovelil, V. J., Price-Whelan, A. M., Bárbara de Andrade Queiroz, A., Raddick, M. J., Ray, A., Rembold, S. B., Riddle, N., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Rodríguez-Puebla, A., Roman-Lopes, A., Román-Zúñiga, C., Rose, B., Ross, A. J., Rossi, G., Rubin, K. H. R., Salvato, M., Sánchez, S. F., Sánchez-Gallego, J. R., Sanderson, R., Santana Rojas, F. A., Sarceno, E., Sarmiento, R., Sayres, C., Sazonova, E., Schaefer, A. L., Schiavon, R., Schlegel, D. J., Schneider, D. P., Schultheis, M., Schwope, A., Serenelli, A., Serna, J., Shao, Z., Shapiro, G., Sharma, A., Shen, Y., Shetrone, M., Shu, Y., Simon, J. D., Skrutskie, M. F., Smethurst, R., Smith, V., Sobeck, J., Spoo, T., Sprague, D., Stark, D. V., Stassun, K. G., Steinmetz, M., Stello, D., Stone-Martinez, A., Storchi-Bergmann, T., Stringfellow, G. S., Stutz, A., Su, Y.-C., Taghizadeh-Popp, M., Talbot, M. S., Tayar, J., Telles, E., Teske, J., Thakar, A., Theissen, C., Tkachenko, A., Thomas, D., Tojeiro, R., Hernandez Toledo, H., Troup, N. W., Trump, J. R., Trussler, J., Turner, J., Tuttle, S., Unda-Sanzana, E., Vázquez-Mata, J. A., Valentini, M., Valenzuela, O., Vargas-González, J., Vargas-Magaña, M., Alfaro, P. V., Villanova, S., Vincenzo, F., Wake, D., Warfield, J. T., Washington, J. D., Weaver, B. A., Weijmans, A.-M., Weinberg, D. H., Weiss, A., Westfall, K. B., Wild, V., Wilde, M. C., Wilson, J. C., Wilson, R. F., Wilson, M., Wolf, J., Wood-Vasey, W. M., Yan, R., Zamora, O., Zasowski, G., Zhang, K., Zhao, C., Zheng, Z., Zheng, Z., and Zhu, K. (2022). The Seventeenth Data Release of the Sloan Digital Sky Surveys: Complete Release of MaNGA, MaStar, and APOGEE-2 Data. *The Astrophysical Journal Supplement Series*, 259(2):35.

Ahmed, N. A. and Gokhale, D. V. (1989). Entropy Expressions and Their Estimators for Multivariate Distri-

butions. *IEEE Transactions on Information Theory*, 35(3):688–692.

Ahumada, J. A. and Lapasset, E. (2007). New catalogue of blue stragglers in open clusters. *Astronomy and Astrophysics*, 463(2):789–797.

Alcock, H. L. and Parker, R. J. (2019). On the mass segregation of cores and stars. 9(September):1–9.

Allison, R. J., Goodwin, S. P., Parker, R. J., Portegies Zwart, S. F., De Grijs, R., and Kouwenhoven, M. B. (2009). Using the minimum spanning tree to trace mass segregation. *Monthly Notices of the Royal Astronomical Society*, 395(3):1449–1454.

Andersen, J. (1991). Accurate masses and radii of normal stars. *The Astronomy and Astrophysics Review*, 3(2):91–126.

Angelo, M. S., Santos, J. F. C., Corradi, W. J. B., and Maia, F. F. S. (2019). Investigating dynamical properties of evolved Galactic open clusters. *Astronomy & Astrophysics*, 624:A8.

Babusiaux, C., Van Leeuwen, F., Barstow, M. A., Jordi, C., Vallenari, A., Bossini, D., Bressan, A., Cantat-Gaudin, T., Van Leeuwen, M., Brown, A. G., Prusti, T., De Bruijne, J. H., Bailer-Jones, C. A., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D., Lattanzi, M. G., Bakker, J., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Holl, B., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Panuzzo, P., Portell, J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P., Busso, G., Carry, B., Cellino, A., Clementini, G., Clotet, M., Creevey, O., Davidson, M., De Ridder, J., Delchambre, L., Dell'Oro, A., Ducourant, C., Fernández-Hernández, J., Fouesneau, M., Frémat, Y., Galluccio, L., García-Torres, M., González-Núñez, J., González-Vidal, J. J., Gosset, E., Guy, L. P., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Hutton, A., Jasniewicz, G., Jean-Antoine-Piccolo, A., Jordan, S., Korn, A. J., Krone-Martins, A., Lanzafame, A. C., Lebzelter, T., Löffler, W., Manteiga, M., Marrese, P. M., Martín-Fleitas, J. M., Moitinho, A., Mora, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J. M., Recio-Blanco, A., Richards, P. J., Rimoldini, L., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Süveges, M., Torra, J., Van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Altavilla, G., Álvarez, M. A., Alvarez, R., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Arcay, B., Astraatmadja, T. L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Balm, P., Barache, C., Barata, C., Barbato, D., Barblan, F., Barklem, P. S., Barrado, D., Barros, M., Bartholomé Muñoz, S., Bassilana, J. L., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Bienaymé, O., Blanco-Cuaresma, S., Boch, T., Boeche, C., Bombrun, A., Borrachero, R., Bouquillon, S., Bourda, G., Bragaglia, A., Bramante, L., Breddels, M. A., Brouillet, N., Brüsemeister, T., Brugaletta, E., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cannizzaro, G., Carballo, R., Carlucci, T., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Charlot, P., Chemin, L., Chiavassa, A., Cocozza, G., Costigan, G., Cowell, S., Crifo, F., Crosta, M., Crowley, C., Cuypersy, J., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., De Laverny, P., De Luise, F., De March, R., De Martino, D., De Souza, R., De Torres, A., Debosscher, J., Del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Diakite, S., Diener, C., Distefano, E., Dolding, C., Drazinos, P., Durán, J., Edvardsson, B., Enke, H., Eriksson, K., Esquej, P., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Federici, L., Fedorets, G., Fernique, P., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fraile, E., Fraser, M., Frézouls, B., Gai, M., Galleti, S., Garabato, D., García-Sedano, F., Garofalo, A., Garralda, N., Gavel, A., Gavras, P., Gerssen, J., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Glass, F., Gomes, M., Granvik, M., Gueguen, A., Guerrier, A., Guiraud, J., Gutiérrez-Sánchez, R., Haigron, R., Hatzidimitriou, D., Hauser, M., Haywood, M., Heiter, U., Helmi, A., Heu, J., Hilger, T., Hobbs, D., Hofmann, W., Holland, G., Huckle, H. E., Hypki, A., Icardi, V., Janßen, K., De Fombelle, G. J., Jonker, P. G., Juhász, L., Julbe, F., Karampelas, A., Kewley, A., Klar, J., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, M., Kontizas, E., Koposov, S. E., Kordopatis, G., Kostrzewa-Rutkowska, Z., Koubsky, P., Lambert, S., Lanza,

A. F., Lasne, Y., Lavigne, J. B., Le Fustec, Y., Le Poncin-Lafitte, C., Lebreton, Y., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lenhardt, H., Leroux, F., Liao, S., Licata, E., Lindstrøm, H. E., Lister, T. A., Livanou, E., Lobel, A., López, M., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marinoni, S., Marschalkó, G., Marshall, D. J., Martino, M., Marton, G., Mary, N., Massari, D., Matijević, G., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Nelemans, G., Nicastro, L., Noval, L., O'Mullane, W., Ordénovic, C., Ordóñez-Blanco, D., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Panahi, A., Pawlak, M., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Poujoulet, E., Přsa, A., Pulone, L., Racero, E., Ragaini, S., Rambaux, N., Ramos-Lerate, M., Regibo, S., Reylé, C., Riclet, F., Ripepi, V., Riva, A., Rivard, A., Rixon, G., Roegiers, T., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sanna, N., Santana-Ros, T., Sarasso, M., Savietto, H., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Shih, I. C., Siltala, L., Silva, A. F., Smart, R. L., Smith, K. W., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Teyssandier, P., Thuillot, W., Titarenko, A., Torra Clotet, F., Turon, C., Ulla, A., Utrilla, E., Uzzi, S., Vaillant, M., Valentini, G., Valette, V., Van Elteren, A., Van Hemelryck, E., Vaschetto, M., Vecchiato, A., Veljanoski, J., Viala, Y., Vicente, D., Vogt, S., Von Essen, C., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Wertz, O., Wevers, T., Wyrzykowski, Yoldas, A., Žerjal, M., Ziaeepour, H., Zorec, J., Zschocke, S., Zucker, S., Zurbach, C., and Zwitter, T. (2018). Observational hertzsprung-russell diagrams. *Astronomy and Astrophysics*, 616:1–29.

Badenes, C., Mazzola, C., Thompson, T. A., Covey, K., Freeman, P. E., Walker, M. G., Moe, M., Troup, N., Nidever, D., Allende Prieto, C., Andrews, B., Barbá, R. H., Beers, T. C., Bovy, J., Carlberg, J. K., De Lee, N., Johnson, J., Lewis, H., Majewski, S. R., Pinsonneault, M., Sobeck, J., Stassun, K. G., Stringfellow, G. S., and Zasowski, G. (2018). Stellar Multiplicity Meets Stellar Evolution and Metallicity: The APOGEE View. , 854(2):147.

Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Demleitner, M., and Andrae, R. (2020). Estimating distances from parallaxes. V: Geometric and photogeometric distances to 1.47 billion stars in Gaia Early Data Release 3. (December).

Baluev, R. V. (2013). PlanetPack: A radial-velocity time-series analysis tool facilitating exoplanets detection, characterization, and dynamical simulations. *Astronomy and Computing*, 2:18–26.

Baluev, R. V., Shaidulin, V. S., and Veselova, A. V. (2020). *High-velocity moving groups in the solar neighborhood in gaia DR2*, volume 70.

Baroch, D., Giménez, A., Morales, J. C., Ribas, I., Herrero, E., Perdelwitz, V., Jordi, C., Granzer, T., and Allende Prieto, C. (2022). Absolute dimensions and apsidal motion of the eclipsing binaries V889 Aquilae and V402 Lacertae. *Astronomy & Astrophysics*, 665:A13.

Batalha, N. M., Borucki, W. J., Koch, D. G., Bryson, S. T., Haas, M. R., Brown, T. M., Caldwell, D. A., Hall, J. R., Gilliland, R. L., Latham, D. W., Meibom, S., and Monet, D. G. (2010). Selection, prioritization, and characteristics of kepler target stars. *Astrophysical Journal Letters*, 713(2 PART 2).

Bayes, T. (1763). LII. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. *Philosophical Transactions of the Royal Society of London*, 53:370–418.

Beaton, R. L., Oelkers, R. J., Hayes, C. R., Covey, K. R., Chojnowski, S. D., De Lee, N., Sobeck, J. S., Majewski, S. R., Cohen, R. E., Fernández-Trincado, J., Longa-Peña, P., O'Connell, J. E., Santana, F. A., Stringfellow, G. S., Zasowski, G., Aerts, C., Anguiano, B., Bender, C., Cañas, C. I., Cunha, K., Donor, J., Fleming, S. W., Frinchaboy, P. M., Feuillet, D., Harding, P., Hasselquist, S., Holtzman, J. A., Johnson, J. A., Kollmeier, J. A., Kounkel, M., Mahadevan, S., Price-Whelan, A. M., Rojas-Arriagada, A., Román-Zúñiga, C., Schlafly, E. F., Schultheis, M., Shetrone, M., Simon, J. D., Stassun, K. G., Stutz, A. M., Tayar,

J., Teske, J., Tkachenko, A., Troup, N., Albareti, F. D., Bizyaev, D., Bovy, J., Burgasser, A. J., Comparat, J., Downes, J. J., Geisler, D., Inno, L., Manchado, A., Ness, M. K., Pinsonneault, M. H., Prada, F., Roman-Lopes, A., Simonian, G. V. A., Smith, V. V., Yan, R., and Zamora, O. (2021). Final Targeting Strategy for the Sloan Digital Sky Survey IV Apache Point Observatory Galactic Evolution Experiment 2 North Survey. , 162(6):302.

Blanton, M. R., Bershady, M. A., Abolfathi, B., Albareti, F. D., Allende Prieto, C., Almeida, A., Alonso-García, J., Anders, F., Anderson, S. F., Andrews, B., Aquino-Ortíz, E., Aragón-Salamanca, A., Argudo-Fernández, M., Armengaud, E., Aubourg, E., Avila-Reese, V., Badenes, C., Bailey, S., Barger, K. A., Barrera-Ballesteros, J., Bartosz, C., Bates, D., Baumgarten, F., Bautista, J., Beaton, R., Beers, T. C., Belfiore, F., Bender, C. F., Berlind, A. A., Bernardi, M., Beutler, F., Bird, J. C., Bizyaev, D., Blanc, G. A., Blomqvist, M., Bolton, A. S., Boquien, M., Borissova, J., van den Bosch, R., Bovy, J., Brandt, W. N., Brinkmann, J., Brownstein, J. R., Bundy, K., Burgasser, A. J., Burtin, E., Busca, N. G., Cappellari, M., Delgado Carigi, M. L., Carlberg, J. K., Carnero Rosell, A., Carrera, R., Chanover, N. J., Cherinka, B., Cheung, E., Gómez Maqueo Chew, Y., Chiappini, C., Choi, P. D., Chojnowski, D., Chuang, C.-H., Chung, H., Cirolini, R. F., Clerc, N., Cohen, R. E., Comparat, J., da Costa, L., Cousinou, M.-C., Covey, K., Crane, J. D., Croft, R. A. C., Cruz-Gonzalez, I., Garrido Cuadra, D., Cunha, K., Damke, G. J., Darling, J., Davies, R., Dawson, K., de la Macorra, A., Dell'Agli, F., De Lee, N., Delubac, T., Di Mille, F., Diamond-Stanic, A., Cano-Díaz, M., Donor, J., Downes, J. J., Drory, N., du Mas des Bourboux, H., Duckworth, C. J., Dwelly, T., Dyer, J., Ebelke, G., Eigenbrot, A. D., Eisenstein, D. J., Emsellem, E., Eracleous, M., Escoffier, S., Evans, M. L., Fan, X., Fernández-Alvar, E., Fernandez-Trincado, J. G., Feuillet, D. K., Finoguenov, A., Fleming, S. W., Font-Ribera, A., Fredrickson, A., Freischlad, G., Frinchaboy, P. M., Fuentes, C. E., Galbany, L., Garcia-Dias, R., García-Hernández, D. A., Gaulme, P., Geisler, D., Gelfand, J. D., Gil-Marín, H., Gillespie, B. A., Goddard, D., Gonzalez-Perez, V., Grabowski, K., Green, P. J., Grier, C. J., Gunn, J. E., Guo, H., Guy, J., Hagen, A., Hahn, C., Hall, M., Harding, P., Hasselquist, S., Hawley, S. L., Hearty, F., Gonzalez Hernández, J. I., Ho, S., Hogg, D. W., Holley-Bockelmann, K., Holtzman, J. A., Holzer, P. H., Huehnerhoff, J., Hutchinson, T. A., Hwang, H. S., Ibarra-Medel, H. J., da Silva Ilha, G., Ivans, I. I., Ivory, K., Jackson, K., Jensen, T. W., Johnson, J. A., Jones, A., Jönsson, H., Jullo, E., Kamble, V., Kinemuchi, K., Kirkby, D., Kitaura, F.-S., Klaene, M., Knapp, G. R., Kneib, J.-P., Kollmeier, J. A., Lacerna, I., Lane, R. R., Lang, D., Law, D. R., Lazarz, D., Lee, Y., Le Goff, J.-M., Liang, F.-H., Li, C., Li, H., Lian, J., Lima, M., Lin, L., Lin, Y.-T., Bertran de Lis, S., Liu, C., de Icaza Lizaola, M. A. C., Long, D., Lucatello, S., Lundgren, B., MacDonald, N. K., Deconto Machado, A., MacLeod, C. L., Mahadevan, S., Geimba Maia, M. A., Maiolino, R., Majewski, S. R., Malanushenko, E., Malanushenko, V., Manchado, A., Mao, S., Maraston, C., Marques-Chaves, R., Masseron, T., Masters, K. L., McBride, C. K., McDermid, R. M., McGrath, B., McGreer, I. D., Medina Peña, N., Melendez, M., Merloni, A., Merrifield, M. R., Meszaros, S., Meza, A., Minchev, I., Minniti, D., Miyaji, T., More, S., Mulchaey, J., Müller-Sánchez, F., Muna, D., Munoz, R. R., Myers, A. D., Nair, P., Nandra, K., Correa do Nascimento, J., Negrete, A., Ness, M., Newman, J. A., Nichol, R. C., Nidever, D. L., Nitschelm, C., Ntelis, P., O'Connell, J. E., Oelkers, R. J., Oravetz, A., Oravetz, D., Pace, Z., Padilla, N., Palanque-Delabrouille, N., Alonso Palicio, P., Pan, K., Parejko, J. K., Parikh, T., Pâris, I., Park, C., Patten, A. Y., Peirani, S., Pellejero-Ibanez, M., Penny, S., Percival, W. J., Perez-Fournon, I., Petitjean, P., Pieri, M. M., Pinsonneault, M., Pisani, A., Poleski, R., Prada, F., Prakash, A., Queiroz, A. B. d. A., Raddick, M. J., Raichoor, A., Barboza Rembold, S., Richstein, H., Riffel, R. A., Riffel, R., Rix, H.-W., Robin, A. C., Rockosi, C. M., Rodríguez-Torres, S., Roman-Lopes, A., Román-Zúñiga, C., Rosado, M., Ross, A. J., Rossi, G., Ruan, J., Ruggeri, R., Rykoff, E. S., Salazar-Albornoz, S., Salvato, M., Sánchez, A. G., Aguado, D. S., Sánchez-Gallego, J. R., Santana, F. A., Santiago, B. X., Sayres, C., Schiavon, R. P., da Silva Schimoia, J., Schlafly, E. F., Schlegel, D. J., Schneider, D. P., Schultheis, M., Schuster, W. J., Schwope, A., Seo, H.-J., Shao, Z., Shen, S., Shetrone, M., Shull, M., Simon, J. D., Skinner, D., Skrutskie, M. F., Slosar, A., Smith, V. V., Sobeck, J. S., Sobreira, F., Somers, G., Souto, D., Stark, D. V., Stassun, K., Stauffer, F., Steinmetz, M., Storchi-Bergmann, T., Streblyanska, A., Stringfellow, G. S., Suárez, G., Sun, J., Suzuki, N., Szigeti, L., Taghizadeh-Popp, M., Tang, B., Tao, C., Tayar, J., Tembe, M., Teske, J., Thakar, A. R., Thomas, D., Thompson, B. A., Tinker, J. L., Tissera, P., Tojeiro, R., Hernandez Toledo, H., de la Torre, S., Tremonti, C., Troup, N. W., Valenzuela, O., Martinez Valpuesta, I., Vargas-González, J., Vargas-Magaña, M., Vazquez, J. A., Villanova, S., Vivek, M., Vogt, N., Wake, D., Walterbos, R., Wang, Y., Weaver, B. A., Weijmans, A.-M., Weinberg, D. H.,

Westfall, K. B., Whelan, D. G., Wild, V., Wilson, J., Wood-Vasey, W. M., Wylezalek, D., Xiao, T., Yan, R., Yang, M., Ybarra, J. E., Yèche, C., Zakamska, N., Zamora, O., Zarrouk, P., Zasowski, G., Zhang, K., Zhao, G.-B., Zheng, Z., Zheng, Z., Zhou, X., Zhou, Z.-M., Zhu, G. B., Zoccali, M., and Zou, H. (2017). Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. , 154(1):28.

Bossini, D., Vallenari, A., Bragaglia, A., Cantat-Gaudin, T., Sordo, R., Balaguer-Núñez, L., Jordi, C., Moitinho, A., Soubiran, C., Casamiquela, L., Carrera, R., and Heiter, U. (2019). Age determination for 269 Gaia DR2 open clusters. *Astronomy & Astrophysics*, 623(2018):A108.

Bovy, J., Hogg, D. W., and Roweis, S. T. (2011). Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *Annals of Applied Statistics*, 5(2 B):1657–1677.

Brown, A. G., Vallenari, A., Prusti, T., and de Bruijne, J. H. (2020). Gaia Early Data Release 3. Summary of the contents and survey properties. *Astronomy & Astrophysics*, 649:A1.

Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Babusiaux, C., Bailer-Jones, C. A. L., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Siddiqui, H. I., Soubiran, C., van Leeuwen, F., Walton, N. A., Arenou, F., Bastian, U., Cropper, M., Drimmel, R., Katz, D., Lattanzi, M. G., Bakker, J., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Holl, B., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Panuzzo, P., Portell, J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Teyssier, D., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P., Busso, G., Carry, B., Cellino, A., Clementini, G., Clotet, M., Creevey, O., Davidson, M., De Ridder, J., Delchambre, L., Dell'Oro, A., Ducourant, C., Fernández-Hernández, J., Fouesneau, M., Frémat, Y., Galluccio, L., García-Torres, M., González-Núñez, J., González-Vidal, J. J., Gosset, E., Guy, L. P., Halbwachs, J.-L., Hambly, N. C., Harrison, D. L., Hernández, J., Hestroffer, D., Hodgkin, S. T., Hutton, A., Jasniewicz, G., Jean-Antoine-Piccolo, A., Jordan, S., Korn, A. J., Krone-Martins, A., Lanzafame, A. C., Lebzelter, T., Löffler, W., Manteiga, M., Marrese, P. M., Martín-Fleitas, J. M., Moitinho, A., Mora, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J.-M., Recio-Blanco, A., Richards, P. J., Rimoldini, L., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Sozzetti, A., Süveges, M., Torra, J., van Reeven, W., Abbas, U., Abreu Aramburu, A., Accart, S., Aerts, C., Altavilla, G., Álvarez, M. A., Alvarez, R., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Arcay, B., Astraatmadja, T. L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Balm, P., Barache, C., Barata, C., Barbato, D., Barblan, F., Barklem, P. S., Barrado, D., Barros, M., Barstow, M. A., Bartholomé Muñoz, S., Bassilana, J.-L., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Bienaymé, O., Blanco-Cuaresma, S., Boch, T., Boeche, C., Bombrun, A., Borrachero, R., Bossini, D., Bouquillon, S., Bourda, G., Bragaglia, A., Bramante, L., Breddels, M. A., Bressan, A., Brouillet, N., Brüsemeister, T., Brugaletta, E., Bucciarelli, B., Burlacu, A., Busonero, D., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cannizzaro, G., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Charlot, P., Chemin, L., Chiavassa, A., Cocozza, G., Costigan, G., Cowell, S., Crifo, F., Crosta, M., Crowley, C., Cuypers†, J., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., de Laverny, P., De Luise, F., De March, R., de Martino, D., de Souza, R., de Torres, A., Debosscher, J., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., Di Matteo, P., Diakite, S., Diener, C., Distefano, E., Dolding, C., Drazinos, P., Durán, J., Edvardsson, B., Enke, H., Eriksson, K., Esquej, P., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Federici, L., Fedorets, G., Fernique, P., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fraile, E., Fraser, M., Frézouls, B., Gai, M., Galleti, S., Garabato, D., García-Sedano, F., Garofalo, A., Garralda, N., Gavel, A., Gavras, P., Gerssen, J., Geyer, R., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Glass, F., Gomes, M., Granvik, M., Gueguen, A., Guerrier, A., Guiraud, J., Gutiérrez-Sánchez, R., Haigron, R., Hatzidimitriou, D., Hauser, M., Haywood, M., Heiter, U., Helmi, A., Heu, J., Hilger, T., Hobbs, D., Hofmann, W., Holland, G., Huckle, H. E., Hypki, A., Icardi, V., Janßen, K., Jevardat de Fombelle, G., Jonker, P. G., Juhász, L., Julbe, F., Karampelas, A., Kewley, A., Klar, J., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, M., Kontizas, E., Koposov, S. E., Kordopatis, G., Kostrzewa-Rutkowska, Z., Koubsky, P., Lambert, S., Lanza, A. F., Lasne, Y., Lavigne, J.-B., Le Fustec, Y., Le Poncin-Lafitte, C., Lebreton, Y., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lenhardt, H., Leroux,

F., Liao, S., Licata, E., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., López, M., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marinoni, S., Marschalkó, G., Marshall, D. J., Martino, M., Marton, G., Mary, N., Massari, D., Matijevič, G., Mazeh, T., McMillan, P. J., Messina, S., Michalik, D., Millar, N. R., Molina, D., Molinaro, R., Molnár, L., Montegriffo, P., Mor, R., Morbidelli, R., Morel, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Nelemans, G., Nicastro, L., Noval, L., O'Mullane, W., Ordénovic, C., Ordóñez-Blanco, D., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Panahi, A., Pawlak, M., Piersimoni, A. M., Pineau, F.-X., Plachy, E., Plum, G., Poggio, E., Poujoulet, E., Prša, A., Pulone, L., Racero, E., Ragaini, S., Rambaux, N., Ramos-Lerate, M., Regibo, S., Reylé, C., Riclet, F., Ripepi, V., Riva, A., Rivard, A., Rixon, G., Roegiers, T., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sanna, N., Santana-Ros, T., Sarasso, M., Savietto, H., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Shih, I.-C., Siltala, L., Silva, A. F., Smart, R. L., Smith, K. W., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Teyssandier, P., Thuillot, W., Titarenko, A., Torra Clotet, F., Turon, C., Ulla, A., Utrilla, E., Uzzi, S., Vaillant, M., Valentini, G., Valette, V., van Elteren, A., Van Hemelryck, E., van Leeuwen, M., Vaschetto, M., Vecchiato, A., Veljanoski, J., Viala, Y., Vicente, D., Vogt, S., von Essen, C., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Wertz, O., Wevers, T., Wyrzykowski, , Yoldas, A., Žerjal, M., Ziaeepour, H., Zorec, J., Zschocke, S., Zucker, S., Zurbach, C., and Zwitter, T. (2018). Gaia Data Release 2. *Astronomy & Astrophysics*, 616:A1.

Buchner, J., Georgakakis, A., Nandra, K., Hsu, L., Rangel, C., Brightman, M., Merloni, A., Salvato, M., Donley, J., and Kocevski, D. (2014). X-ray spectral modelling of the AGN obscuring region in the CDFS: Bayesian model selection and catalogue. *Astronomy and Astrophysics*, 564:1–25.

Cabrera-Cano, J. and Alfaro, E. J. (1990). A Non-parametric Approach to the Membership Problem In Open Clusters. *Astronomy and Astrophysics*, 235:94–102.

Caldwell, D. A., Tenenbaum, P., Twicken, J. D., Jenkins, J. M., Ting, E., Smith, J. C., Hedges, C. L., Fausnaugh, M. M., Rose, M., and Burke, C. J. (2020). TESS Science Processing Operations Center FFI Target List Products. *Publications of the Astronomical Society of the Pacific*, 124(919):1000–1014.

Cantat-Gaudin, T. and Anders, F. (2019). Clusters and mirages: cataloguing stellar aggregates in the Milky Way. (2010):1–24.

Cantat-Gaudin, T., Anders, F., Castro-Ginard, A., Jordi, C., Romero-Gomez, M., Soubiran, C., Casamiquela, L., Tarricq, Y., Moitinho, A., Vallenari, A., Bragaglia, A., Krone-Martins, A., and Kounkel, M. (2020). Painting a portrait of the Galactic disc with its stellar clusters. (1925).

Cantat-Gaudin, T., Jordi, C., Vallenari, A., Bragaglia, A., Balaguer-Núñez, L., Soubiran, C., Bossini, D., Moitinho, A., Castro-Ginard, A., Krone-Martins, A., Casamiquela, L., Sordo, R., and Carrera, R. (2018a). A Gaia DR2 view of the open cluster population in the Milky Way. *Astronomy & Astrophysics*, 618:A93.

Cantat-Gaudin, T., Jordi, C., Wright, N. J., Armstrong, J. J., Vallenari, A., Balaguer-Núñez, L., Ramos, P., Bossini, D., Padoan, P., Pelkonen, V. M., and Mapelli, M. (2018b). Expanding associations in the Vela-Puppis region: 3D structure and kinematics of the young population. (2017).

Cantat-Gaudin, T., Vallenari, A., Sordo, R., Pensabene, F., Krone-Martins, A., Moitinho, A., Jordi, C., Casamiquela, L., Balaguer-Núnez, L., Soubiran, C., and Brouillet, N. (2018c). Characterising open clusters in the solar neighbourhood with the Tycho-Gaia Astrometric Solution. *Astronomy & Astrophysics*, 615:A49.

Cassisi, S. and Salaris, M. (2019). Effective temperature - radius relationship of M dwarfs. , 626:A32.

Castro-Ginard, A., Jordi, C., Luri, X., Cantat-Gaudin, T., and Balaguer-Núñez, L. (2019). Hunting for open clusters in \textit{Gaia} DR2: the Galactic anticentre. pages 1–8.

Castro-Ginard, A., Jordi, C., Luri, X., Cid-Fuentes, J. , Casamiquela, L., Anders, F., Cantat-Gaudin, T., Monguió, M., Balaguer-Núñez, L., Solà, S., and Badia, R. M. (2020). Hunting for open clusters in \textit{Gaia} DR2: 582 new OCs in the Galactic disc. (2019):1–11.

Castro-Ginard, A., Jordi, C., Luri, X., Julbe, F., Morvan, M., Balaguer-Núñez, L., and Cantat-Gaudin, T. (2018). A new method for unveiling Open Clusters in Gaia: new nearby Open Clusters confirmed by DR2. (2014):1–19.

Choi, J., Dotter, A., Conroy, C., Cantiello, M., Paxton, B., and Johnson, B. D. (2016). MESA Isochrones and Stellar Tracks (MIST). I: Solar-Scaled Models. *The Astrophysical Journal*, 823(2):102.

Cunha, M. S., Antoci, V., Holdsworth, D. L., Kurtz, D. W., Balona, L. A., Bognár, Z., Bowman, D. M., Guo, Z., Kołaczek-Szymański, P. A., Lares-Martiz, M., Paunzen, E., Skarka, M., Smalley, B., Sódor, Kochukhov, O., Pepper, J., Richey-Yowell, T., Ricker, G. R., Seager, S., Buzasi, D. L., Fox-Machado, L., Hasanzadeh, A., Niemczura, E., Quitral-Manosalva, P., Monteiro, M. J., Stateva, I., De Cat, P., García Hernández, A., Ghasemi, H., Handler, G., Hey, D., Matthews, J. M., Nemec, J. M., Pascual-Granado, J., Safari, H., Suárez, J. C., Szabó, R., Tkachenko, A., and Weiss, W. W. (2019). Rotation and pulsation in Ap stars: First light results from TESS sectors 1 and 2. *Monthly Notices of the Royal Astronomical Society*, 487(3):3523–3549.

Cunningham, J.-M. C., Feliz, D. L., Dixon, D. M., Pepper, J., Stassun, K. G., Siverd, R. J., Zhou, G., Bayliss, D., Tan, T.-G., Cargile, P., James, D., Kuhn, R. B., and Kounkel, M. (2020). A KELT–TESS Eclipsing Binary in a Young Triple System Associated with the Local "Stellar String" Theia 301. *The Astronomical Journal*, 160(4):187.

Dahn, B. I. (1978). Foundations of Probability theory, statistical inference, and statistical theories of science. *Studia Logica*, 37(2):213–219.

Daszykowski, M. and Walczak, B. (2009). Density-Based Clustering Methods. In *Comprehensive Chemometrics*, volume 2, pages 635–654.

Dias, W. S., Monteiro, H., and Assafin, M. (2018). Update of membership and mean proper motion of open clusters from UCAC5 catalogue. *Monthly Notices of the Royal Astronomical Society*, 478(4):5184–5189.

Dias, W. S., Monteiro, H., Caetano, T. C., Lépine, J. R. D., Assafin, M., and Oliveira, A. F. (2014). Proper motions of the optically visible open clusters based on the UCAC4 catalog. *Astronomy & Astrophysics*, 564:A79.

Dotter, A. (2016). MESA Isochrones and Stellar Tracks (MIST) 0: Methods for the construction of stellar isochrones. *The Astrophysical Journal Supplement Series*, 222(1):8.

Doyle, L., Ramsay, G., and Doyle, J. G. (2020). Superflares and variability in solar-type stars with TESS in the Southern hemisphere. *Monthly Notices of the Royal Astronomical Society*, 494(3):3596–3610.

Duane, S., Kennedy, A., Pendleton, B. J., and Roweth, D. (1987). Hybrid Monte Carlo. *Physics Letters B*, 195(2):216–222.

Duchêne, G., Bouvier, J., Moraux, E., Bouy, H., Konopacky, Q., and Ghez, A. M. (2013). Substellar multiplicity in the Hyades cluster. *Astronomy & Astrophysics*, 555:A137.

Duchene, G., Lacour, S., Moraux, E., Goodwin, S., and Bouvier, J. (2018). Is stellar multiplicity universal? Tight stellar binaries in the Orion Nebula Cluster. 12(March):1–12.

Eastman, J., Gaudi, B. S., and Agol, E. (2013). EXOFAST: A Fast Exoplanetary Fitting Suite in IDL. *Publications of the Astronomical Society of the Pacific*, 125(923):83–112.

Eastman, J. D., Rodriguez, J. E., Agol, E., Stassun, K. G., Beatty, T. G., Vanderburg, A., Gaudi, B. S., Collins, K. A., and Luger, R. (2019). EXOFASTv2: A public, generalized, publication-quality exoplanet modeling code.

Espinoza, N. and Jordán, A. (2015). Limb darkening and exoplanets: testing stellar model atmospheres and identifying biases in transit parameters. *Monthly Notices of the Royal Astronomical Society*, 450(2):1879–1899.

Espinoza, N. and Jordán, A. (2016). Limb-darkening and exoplanets II: Choosing the Best Law for Optimal Retrieval of Transit Parameters. *Monthly Notices of the Royal Astronomical Society*, 457(4):3573–3581.

Espinoza, N., Kossakowski, D., and Brahm, R. (2019). juliet: a versatile modelling tool for transiting and non-transiting exoplanetary systems. *Monthly Notices of the Royal Astronomical Society*, 490(2):2262–2283.

Feigelson, E. D., Getman, K. V., Townsley, L. K., Broos, P. S., Povich, M. S., Garmire, G. P., King, R. R., Montmerle, T., Preibisch, T., Smith, N., Stassun, K. G., Wang, J., Wolk, S., and Zinnecker, H. (2011). X-ray star clusters in the Carina Complex. *Astrophysical Journal, Supplement Series*, 194(1).

Feroz, F., Hobson, M. P., and Bridges, M. (2009). MultiNest: An efficient and robust Bayesian inference tool for cosmology and particle physics. *Monthly Notices of the Royal Astronomical Society*, 398(4):1601–1614.

Ferreira, F. A., Corradi, W. J. B., Maia, F. F. S., Angelo, M. S., and Santos, J. F. C. (2020). Discovery and astrophysical properties of Galactic open clusters in dense stellar fields using Gaia DR2. *Monthly Notices of the Royal Astronomical Society*, 496(2):2021–2038.

Foreman-Mackey, D. (2018). Scalable backpropagation for Gaussian Processes using celerite. *Research Notes of the AAS*, 2(1):31.

Foreman-Mackey, D., Agol, E., Ambikasaran, S., and Angus, R. (2017). Fast and scalable Gaussian process modeling with applications to astronomical time series. *The Astronomical Journal*, 154(6):220.

Foreman-Mackey, D., Hogg, D. W., Lang, D., and Goodman, J. (2013). emcee : The MCMC Hammer. *Publications of the Astronomical Society of the Pacific*, 125(925):306–312.

Foreman-Mackey, D., Luger, R., Agol, E., Barclay, T., Bouma, L., Brandt, T., Czekala, I., David, T., Dong, J., Gilbert, E., Gordon, T., Hedges, C., Hey, D., Morris, B., Price-Whelan, A., and Savel, A. (2021). exoplanet: Gradient-based probabilistic inference for exoplanet data other astronomical time series. *Journal of Open Source Software*, 6(62):3285.

Friel, E. (1995). The Old Open Clusters of the Milky Way. *Annual Review of Astronomy and Astrophysics*, 33(1):381–414.

Gaia Collaboration, Klioner, S. A., Lindegren, L., Mignard, F., Hernández, J., Ramos-Lerate, M., Bastian, U., Biermann, M., Bombrun, A., de Torres, A., Gerlach, E., Geyer, R., Hilger, T., Hobbs, D., Lammers, U. L., McMillan, P. J., Steidelmüller, H., Teyssier, D., Raiteri, C. M., Bartolomé, S., Bernet, M., Castañeda, J., Clotet, M., Davidson, M., Fabricius, C., Torres, N. G., González-Vidal, J. J., Portell, J., Rowell, N., Torra, F., Torra, J., Brown, A. G. A., Vallenari, A., Prusti, T., de Bruijne, J. H. J., Arenou, F., Babusiaux, C., Creevey, O. L., Ducourant, C., Evans, D. W., Eyer, L., Guerra, R., Hutton, A., Jordi, C., Luri, X., Panem, C., Pourbaix, D., Randich, S., Sartoretti, P., Soubiran, C., Tanga, P., Walton, N. A., Bailer-Jones, C. A. L., Drimmel, R., Jansen, F., Katz, D., Lattanzi, M. G., van Leeuwen, F., Bakker, J., Cacciari, C., De Angeli, F., Fouesneau, M., Frémat, Y., Galluccio, L., Guerrier, A., Heiter, U., Masana, E., Messineo, R., Mowlavi, N., Nicolas, C., Nienartowicz, K., Pailler, F., Panuzzo, P., Riclet, F., Roux, W., Seabroke, G. M., Sordo, R., Thévenin, F., Gracia-Abril, G., Altmann, M., Andrae, R., Audard, M., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Burgess, P. W., Busonero, D., Busso, G., Cánovas, H., Carry, B., Cellino, A., Cheek, N., Clementini, G., Damerdji, Y., de Teodoro, P., Campos, M. N., Delchambre, L., Dell'Oro, A., Esquej, P., Fernández-Hernández, J., Fraile, E., Garabato, D., García-Lario, P., Gosset, E., Haigron, R., Halbwachs, J. L., Hambly, N. C., Harrison, D. L., Hestroffer, D., Hodgkin, S. T., Holl, B., Janßen, K., de Fombelle, G. J., Jordan, S., Krone-Martins, A., Lanzafame, A. C., Löffler, W., Marchal, O., Marrese, P. M., Moitinho, A., Muinonen, K., Osborne, P., Pancino, E., Pauwels, T., Recio-Blanco, A., Reylé, C., Riello, M., Rimoldini, L., Roegiers, T., Rybizki, J., Sarro, L. M., Siopis, C., Smith, M.,

Sozzetti, A., Utrilla, E., van Leeuwen, M., Abbas, U., Ábrahám, P., Aramburu, A. A., Aerts, C., Aguado, J. J., Ajaj, M., Aldea-Montero, F., Altavilla, G., Álvarez, M. A., Alves, J., Anderson, R. I., Varela, E. A., Antoja, T., Baines, D., Baker, S. G., Balaguer-Núñez, L., Balbinot, E., Balog, Z., Barache, C., Barbato, D., Barros, M., Barstow, M. A., Bassilana, J. L., Bauchet, N., Becciani, U., Bellazzini, M., Berihuete, A., Bertone, S., Bianchi, L., Binnenfeld, A., Blanco-Cuaresma, S., Boch, T., Bossini, D., Bouquillon, S., Bragaglia, A., Bramante, L., Breedt, E., Bressan, A., Brouillet, N., Brugaletta, E., Bucciarelli, B., Burlacu, A., Butkevich, A. G., Buzzi, R., Caffau, E., Cancelliere, R., Cantat-Gaudin, T., Carballo, R., Carlucci, T., Carnerero, M. I., Carrasco, J. M., Casamiquela, L., Castellani, M., Castro-Ginard, A., Chaoul, L., Charlot, P., Chemin, L., Chiaramida, V., Chiavassa, A., Chornay, N., Comoretto, G., Contursi, G., Cooper, W. J., Cornez, T., Cowell, S., Crifo, F., Cropper, M., Crosta, M., Crowley, C., Dafonte, C., Dapergolas, A., David, P., de Laverny, P., De Luise, F., De March, R., De Ridder, J., de Souza, R., del Peloso, E. F., del Pozo, E., Delbo, M., Delgado, A., Delisle, J. B., Demouchy, C., Dharmawardena, T. E., Diakite, S., Diener, C., Distefano, E., Dolding, C., Enke, H., Fabre, C., Fabrizio, M., Faigler, S., Fedorets, G., Fernique, P., Fienga, A., Figueras, F., Fournier, Y., Fouron, C., Fragkoudi, F., Gai, M., Garcia-Gutierrez, A., Garcia-Reinaldos, M., García-Torres, M., Garofalo, A., Gavel, A., Gavras, P., Giacobbe, P., Gilmore, G., Girona, S., Giuffrida, G., Gomel, R., Gomez, A., González-Núñez, J., González-Santamaría, I., Granvik, M., Guillout, P., Guiraud, J., Gutiérrez-Sánchez, R., Guy, L. P., Hatzidimitriou, D., Hauser, M., Haywood, M., Helmer, A., Helmi, A., Sarmiento, M. H., Hidalgo, S. L., Hładczuk, N., Holland, G., Huckle, H. E., Jardine, K., Jasniewicz, G., Piccolo, A. J.-A., Jiménez-Arranz, , Campillo, J. J., Julbe, F., Karbevska, L., Kervella, P., Khanna, S., Kordopatis, G., Korn, A. J., Kóspál, , Kostrzewa-Rutkowska, Z., Kruszyńska, K., Kun, M., Laizeau, P., Lambert, S., Lanza, A. F., Lasne, Y., Campion, J. F. L., Lebreton, Y., Lebzelter, T., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Liao, S., Licata, E. L., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., Lorca, A., Loup, C., Pardo, P. M., Romeo, A. M., Managau, S., Mann, R. G., Manteiga, M., Marchant, J. M., Marconi, M., Marcos, J., Santos, M. M. S. M., Pina, D. M., Marinoni, S., Marocco, F., Marshall, D. J., Polo, L. M., Martín-Fleitas, J. M., Marton, G., Mary, N., Masip, A., Massari, D., Mastrobuono-Battisti, A., Mazeh, T., Messina, S., Michalik, D., Millar, N. R., Mints, A., Molina, D., Molinaro, R., Molnár, L., Monari, G., Monguió, M., Montegriffo, P., Montero, A., Mor, R., Mora, A., Morbidelli, R., Morel, T., Morris, D., Muraveva, T., Murphy, C. P., Musella, I., Nagy, Z., Noval, L., Ocaña, F., Ogden, A., Ordenovic, C., Osinde, J. O., Pagani, C., Pagano, I., Palaversa, L., Palicio, P. A., Pallas-Quintela, L., Panahi, A., Payne-Wardenaar, S., Esteller, X. P., Penttilä, A., Pichon, B., Piersimoni, A. M., Pineau, F. X., Plachy, E., Plum, G., Poggio, E., Prša, A., Pulone, L., Racero, E., Ragaini, S., Rainer, M., Rambaux, N., Ramos, P., Fiorentin, P. R., Regibo, S., Richards, P. J., Diaz, C. R., Ripepi, V., Riva, A., Rix, H. W., Rixon, G., Robichon, N., Robin, A. C., Robin, C., Roelens, M., Rogues, H. R. O., Rohrbasser, L., Romero-Gómez, M., Royer, F., Mieres, D. R., Rybicki, K. A., Sadowski, G., Núñez, A. S., Sellés, A. S., Sahlmann, J., Salguero, E., Samaras, N., Gimenez, V. S., Sanna, N., Santoveña, R., Sarasso, M., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Ségransan, D., Semeux, D., Shahaf, S., Siddiqui, H. I., Siebert, A., Siltala, L., Silvelo, A., Slezak, E., Slezak, I., Smart, R. L., Snaith, O. N., Solano, E., Solitro, F., Souami, D., Souchay, J., Spagna, A., Spina, L., Spoto, F., Steele, I. A., Stephenson, C. A., Süveges, M., Surdej, J., Szabados, L., Szegedi-Elek, E., Taris, F., Taylor, M. B., Teixeira, R., Tolomei, L., Tonello, N., Elipe, G. T., Trabucchi, M., Tsounis, A. T., Turon, C., Ulla, A., Unger, N., Vaillant, M. V., van Dillen, E., van Reeven, W., Vanel, O., Vecchiato, A., Viala, Y., Vicente, D., Voutsinas, S., Weiler, M., Wevers, T., Wyrzykowski, , Yoldas, A., Yvard, P., Zhao, H., Zorec, J., Zucker, S., and Zwitter, T. (2022). Gaia Early Data Release 3: The celestial reference frame (Gaia-CRF3). pages 1–30.

Gaia Collaboration, G. and Brown, A. G. (2020). Gaia early data release 3 (gaia edr3). 3(December):1–21.

Gandolfi, D., Barragán, O., Livingston, J. H., Fridlund, M., Justesen, A. B., Redfield, S., Fossati, L., Mathur, S., Grziwa, S., Cabrera, J., García, R. A., Persson, C. M., Van Eylen, V., Hatzes, A. P., Hidalgo, D., Albrecht, S., Bugnet, L., Cochran, W. D., Csizmadia, S., Deeg, H., Eigmüller, P., Endl, M., Erikson, A., Esposito, M., Guenther, E., Korth, J., Luque, R., Montañes Rodríguez, P., Nespral, D., Nowak, G., Pätzold, M., and Prieto-Arranz, J. (2018). TESS's first planet. *Astronomy & Astrophysics*, 619:L10.

Gao, X. (2018). Memberships of the Open Cluster NGC 6405 Based on a Combined Method: Gaussian Mixture Model and Random Forest. *The Astronomical Journal*, 156(3):121.

Ghosh, Jayanta K., Delampady, M., and Samanta, T. (2006). *An Introduction to Bayesian Analysis: Theory and Methods*, volume 103. Springer.

Gill, S., Cooke, B. F., Bayliss, D., Nielsen, L. D., Lendl, M., Wheatley, P. J., Anderson, D. R., Moyano, M., Bryant, E. M., Acton, J. S., Belardi, C., Bouchy, F., Burleigh, M. R., Casewell, S. L., Chaushev, A., Goad, M. R., Jackman, J. A. G., Jenkins, J. S., McCormac, J., Günther, M. N., Osborn, H. P., Pollacco, D., Raynard, L., Smith, A. M. S., Tilbrook, R. H., Turner, O., Udry, S., Vines, J. I., Watson, C. A., and West, R. G. (2020). A long-period (P = 61.8 d) M5V dwarf eclipsing a Sun-like star from TESS and NGTS. , 495(3):2713–2719.

Ginsburg, A., Sipőcz, B. M., Brasseur, C. E., Cowperthwaite, P. S., Craig, M. W., Deil, C., Guillochon, J., Guzman, G., Liedtke, S., Lim, P. L., Lockhart, K. E., Mommert, M., Morris, B. M., Norman, H., Parikh, M., Persson, M. V., Robitaille, T. P., Segovia, J. C., Singer, L. P., Tollerud, E. J., de Val-Borro, M., Valtchanov, I., and Woillez, J. (2019). Astroquery: An astronomical web-querying package in python. *arXiv*, 98:1–7.

Girardi, L., Bressan, A., Bertelli, G., and Chiosi, C. (2000). Evolutionary tracks and isochrones for low- and intermediate-mass stars: From 0.15 to 7 M, and from Z = 0.0004 to 0.03. *Astronomy and Astrophysics Supplement Series*, 141(3):371–383.

Graczyk, D., Konorski, P., Pietrzyński, G., Gieren, W., Storm, J., Nardetto, N., Gallenne, A., Maxted, P. F. L., Kervella, P., and Kołaczkowski, Z. (2017). The Surface Brightness-color Relations Based on Eclipsing Binary Stars: Toward Precision Better than 1% in Angular Diameter Predictions. *The Astrophysical Journal*, 837(1):7.

Günther, M. N. and Daylan, T. (2021). Allesfitter: Flexible Star and Exoplanet Inference from Photometry and Radial Velocity. *The Astrophysical Journal Supplement Series*, 254(1):13.

Hao, C. J., Xu, Y., Hou, L. G., Bian, S. B., Li, J. J., Wu, Z. Y., He, Z. H., Li, Y. J., and Liu, D. J. (2021). Evolution of the local spiral structure of the Milky Way revealed by open clusters. *Astronomy & Astrophysics*, 652:A102.

Hao, J., Koester, B. P., McKay, T. A., Rykoff, E. S., Rozo, E., Evrard, A., Annis, J., Becker, M., Busha, M., Gerdes, D., Johnston, D. E., Sheldon, E., and Wechsler, R. H. (2009). Precision measurements of the cluster red sequence using an error-corrected gaussian mixture model. *Astrophysical Journal*, 702(1):745–758.

He, M. Y. and Petrovich, C. (2017). On the stability and collisions in triple stellar systems. 12(October):1–12.

Hillenbrand, L. A. and Hartmann, L. W. (1998). A Preliminary Study of the Orion Nebula Cluster Structure and Dynamics. *The Astrophysical Journal*, 492(2):540–553.

Hippke, M., David, T. J., Mulders, G. D., and Heller, R. (2019). Wōtan : Comprehensive Time-series Detrending in Python . *The Astronomical Journal*, 158(4):143.

Hoffman, M. D. and Gelman, A. (2014). The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, 15(2008):1593–1623.

Holl, B., Sozzetti, A., Sahlmann, J., Giacobbe, P., Ségransan, D., Unger, N., Delisle, J. B., Barbato, D., Lattanzi, M. G., Morbidelli, R., and Sosnowska, D. (2022). Gaia DR3 astrometric orbit determination with Markov Chain Monte Carlo and Genetic Algorithms. Systems with stellar, substellar, and planetary mass companions. pages 1–31.

Holoien, T. W. S., Marshall, P. J., and Wechsler, R. H. (2016). EmpiriciSN: Re-sampling Observed Supernova/Host Galaxy Populations using an XD Gaussian Mixture Model. *eprint arXiv:1611.00363*, 153(6):249.

Hong, K., Woo Lee, J., Rittipruk, P., Park, J.-H., Kim, H.-Y., and Han, C. (2022). BD+47 378: An Eclipsing Binary Containing a δ Sct Pulsating Star. *The Astronomical Journal*, 164(4):121.

Huber, P. J. (1980). *Robust statistics / Peter J. Huber*. Wiley series in probability and mathematical statistics. Wiley, New York.

Hunt, E. L. and Reffert, S. (2020). Improving the open cluster census. I. Comparison of clustering algorithms applied to Gaia DR2 data. (39341).

Hunt, E. L. and Reffert, S. (2023). Improving the open cluster census. *Astronomy & Astrophysics*, 673:A114.

Hwang, H.-C., Ting, Y.-S., and Zakamska, N. L. (2022). The eccentricity distribution of wide binaries and their individual measurements. *Monthly Notices of the Royal Astronomical Society*, 512(3):3383–3399.

Johnson, H. L. and Knuckles, C. F. (1955). The Hyades and Coma Berenices Star Clusters. *The Astrophysical Journal*, 122(2):209–222.

Kharchenko, N. V., Piskunov, A. E., Schilbach, E., Röser, S., and Scholz, R. D. (2013). Global survey of star clusters in the Milky Way: II. the catalogue of basic parameters. *Astronomy and Astrophysics*, 558:1–8.

Kipping, D. M. (2013). Efficient, uninformative sampling of limb darkening coefficients for two-parameter laws. *Monthly Notices of the Royal Astronomical Society*, 435(3):2152–2160.

Kirk, B., Conroy, K., Prša, A., Abdul-Masih, M., Kochoska, A., Matijevič, G., Hambleton, K., Barclay, T., Bloemen, S., Boyajian, T., Doyle, L. R., Fulton, B. J., Hoekstra, A. J., Jek, K., Kane, S. R., Kostov, V., Latham, D., Mazeh, T., Orosz, J. A., Pepper, J., Quarles, B., Ragozzine, D., Shporer, A., Southworth, J., Stassun, K., Thompson, S. E., Welsh, W. F., Agol, E., Derekas, A., Devor, J., Fischer, D., Green, G., Gropp, J., Jacobs, T., Johnston, C., LaCourse, D. M., Saetre, K., Schwengeler, H., Toczyski, J., Werner, G., Garrett, M., Gore, J., Martinez, A. O., Spitzer, I., Stevick, J., Thomadis, P. C., Vrijmoet, E. H., Yenawine, M., Batalha, N., and Borucki, W. (2016). KEPLER ECLIPSING BINARY STARS. VII. THE CATALOG OF ECLIPSING BINARIES FOUND IN THE ENTIRE KEPLER DATA SET. *The Astronomical Journal*, 151(3):68.

Kos, J., de Silva, G., Buder, S., Bland-Hawthorn, J., Sharma, S., Asplund, M., D'Orazi, V., Duong, L., Freeman, K., Lewis, G. F., Lin, J., Lind, K., Martell, S. L., Schlesinger, K. J., Simpson, J. D., Zucker, D. B., Zwitter, T., Bedding, T. R., Čotar, K., Horner, J., Nordlander, T., Stello, D., Ting, Y. S., and Traven, G. (2018). The GALAH survey and Gaia DR2: (non-)existence of five sparse high-latitude open clusters. *Monthly Notices of the Royal Astronomical Society*, 480(4):5242–5259.

Kounkel, M., Covey, K., and Stassun, K. G. (2020). Untangling the Galaxy. II. Structure within 3 kpc. *arXiv*.

Kraus, A. L. and Hillenbrand, L. A. (2009). The Coevality of Young Binary Systems. *The Astrophysical Journal*, 704(1):531–547.

Kraus, A. L., Tucker, R. A., Thompson, M. I., Craine, E. R., and Hillenbrand, L. A. (2011). The Mass-Radius(-Rotation?) Relation for Low-mass Stars. , 728(1):48.

Krone-Martins, A. and Moitinho, A. (2014). UPMASK: Unsupervised photometric membership assignment in stellar clusters. *Astronomy and Astrophysics*, 561.

Kroupa, P. (2001). The Birth, Evolution and Death of Star Clusters. *From Darkness to Light: Origin and Evolution of Young Stellar Clusters*, 243(1):1–12.

Kuhn, M. A., Feigelson, E. D., Getman, K. V., Baddeley, A. J., Broos, P. S., Sills, A., Bate, M. R., Povich, M. S., Luhman, K. L., Busk, H. A., Naylor, T., and King, R. R. (2014). The spatial structure of young stellar clusters. I. Subclusters. *Astrophysical Journal*, 787(2).

Kuhn, M. A., Medina, N., Getman, K. V., Feigelson, E. D., Gromadzki, M., Borissova, J., and Kurtev, R. (2017). THE STRUCTURE OF THE YOUNG STAR CLUSTER NGC 6231. I. STELLAR POPULATION.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766.

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., Gully-Santiago, M., Saunders, N., Cody, A. M., Barclay, T., Hall, O., Sagear, S., Turtelboom, E., Zhang, J., Tzanidakis, A., Mighell, K., Coughlin, J., Bell, K., Berta-Thompson, Z., Williams, P., Dotson, J., and Barentsen, G. (2018). Lightkurve: Kepler and TESS time series analysis in Python. Astrophysics Source Code Library.

Lindegren, L. (2019). The Gaia reference frame for bright sources examined using VLBI observations of radio stars. *arXiv*, (36161).

Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U., Ramos-Lerate, M., De Torres, A., Steidelmöller, H., Stephenson, C., Hobbs, D., Lammers, U., Biermann, M., Geyer, R., Hilger, T., Michalik, D., Stampa, U., McMillan, P. J., Castañeda, J., Clotet, M., Comoretto, G., Davidson, M., Fabricius, C., Gracia, G., Hambly, N. C., Hutton, A., Mora, A., Portell, J., Van Leeuwen, F., Abbas, U., Abreu, A., Altmann, M., Andrei, A., Anglada, E., Balaguer-Núñez, L., Barache, C., Becciani, U., Bertone, S., Bianchi, L., Bouquillon, S., Bourda, G., Brösemeister, T., Bucciarelli, B., Busonero, D., Buzzi, R., Cancelliere, R., Carlucci, T., Charlot, P., Cheek, N., Crosta, M., Crowley, C., De Bruijne, J., De Felice, F., Drimmel, R., Esquej, P., Fienga, A., Fraile, E., Gai, M., Garralda, N., González-Vidal, J. J., Guerra, R., Hauser, M., Hofmann, W., Holl, B., Jordan, S., Lattanzi, M. G., Lenhardt, H., Liao, S., Licata, E., Lister, T., Löffler, W., Marchant, J., Martin-Fleitas, J. M., Messineo, R., Mignard, F., Morbidelli, R., Poggio, E., Riva, A., Rowell, N., Salguero, E., Sarasso, M., Sciacca, E., Siddiqui, H., Smart, R. L., Spagna, A., Steele, I., Taris, F., Torra, J., Van Elteren, A., Van Reeven, W., and Vecchiato, A. (2018). Gaia Data Release 2: The astrometric solution. *Astronomy and Astrophysics*, 616:1–35.

Lindegren, L., Klioner, S. A., Hernández, J., Bombrun, A., Ramos-Lerate, M., Steidelmüller, H., Bastian, U., Biermann, M., de Torres, A., Gerlach, E., Geyer, R., Hilger, T., Hobbs, D., Lammers, U., McMillan, P. J., Stephenson, C. A., Castañeda, J., Davidson, M., Fabricius, C., Gracia-Abril, G., Portell, J., Rowell, N., Teyssier, D., Torra, F., Bartolomé, S., Clotet, M., Garralda, N., González-Vidal, J. J., Torra, J., Abbas, U., Altmann, M., Anglada Varela, E., Balaguer-Núñez, L., Balog, Z., Barache, C., Becciani, U., Bernet, M., Bertone, S., Bianchi, L., Bouquillon, S., Brown, A. G., Bucciarelli, B., Busonero, D., Butkevich, A. G., Buzzi, R., Cancelliere, R., Carlucci, T., Charlot, P., Cioni, M. R., Crosta, M., Crowley, C., del Peloso, E. F., del Pozo, E., Drimmel, R., Esquej, P., Fienga, A., Fraile, E., Gai, M., Garcia-Reinaldos, M., Guerra, R., Hambly, N. C., Hauser, M., Janßen, K., Jordan, S., Kostrzewa-Rutkowska, Z., Lattanzi, M. G., Liao, S., Licata, E., Lister, T. A., Löffler, W., Marchant, J. M., Masip, A., Mignard, F., Mints, A., Molina, D., Mora, A., Morbidelli, R., Murphy, C. P., Pagani, C., Panuzzo, P., Peñalosa Esteller, X., Poggio, E., Re Fiorentin, P., Riva, A., Sagristà Sellés, A., Sanchez Gimenez, V., Sarasso, M., Sciacca, E., Siddiqui, H. I., Smart, R. L., Souami, D., Spagna, A., Steele, I. A., Taris, F., Utrilla, E., van Reeven, W., and Vecchiato, A. (2020). Gaia early data release 3: The astrometric solution. *arXiv*, 649(A17):A2.

Liu, C. (2019). Smoking Gun of the Dynamical Processing of the Solar-type Field Binary Stars. 17(July):1–17.

Liu, L. and Pang, X. (2019). A catalog of newly identified star clusters in GAIA DR2.

Luri, X., Brown, A. G., Sarro, L. M., Arenou, F., Bailer-Jones, C. A., Castro-Ginard, A., Bruijne, J. D., Prusti, T., Babusiaux, C., and Delgado, H. E. (2018). Gaia Data Release 2: Using Gaia parallaxes. *Astronomy and Astrophysics*, 616(32964):1–20.

Lyttle, A. J., Davies, G. R., Li, T., Carboneau, L. M., Leung, H. H., Westwood, H., Chaplin, W. J., Hall, O. J., Huber, D., Nielsen, M. B., Basu, S., and Garciá, R. A. (2021). Hierarchically modelling Kepler dwarfs and subgiants to improve inference of stellar properties with asteroseismology. *Monthly Notices of the Royal Astronomical Society*, 505(2):2427–2446.

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Prieto, C. A., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., Pérez, A.

Leys, C., Ley, C., Klein, O., Bernard, P., and Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4):764–766.

Lightkurve Collaboration, Cardoso, J. V. d. M., Hedges, C., Gully-Santiago, M., Saunders, N., Cody, A. M., Barclay, T., Hall, O., Sagear, S., Turtelboom, E., Zhang, J., Tzanidakis, A., Mighell, K., Coughlin, J., Bell, K., Berta-Thompson, Z., Williams, P., Dotson, J., and Barentsen, G. (2018). Lightkurve: Kepler and TESS time series analysis in Python. Astrophysics Source Code Library.

Lindegren, L. (2019). The Gaia reference frame for bright sources examined using VLBI observations of radio stars. *arXiv*, (36161).

Lindegren, L., Hernández, J., Bombrun, A., Klioner, S., Bastian, U., Ramos-Lerate, M., De Torres, A., Steidelmöller, H., Stephenson, C., Hobbs, D., Lammers, U., Biermann, M., Geyer, R., Hilger, T., Michalik, D., Stampa, U., McMillan, P. J., Castañeda, J., Clotet, M., Comoretto, G., Davidson, M., Fabricius, C., Gracia, G., Hambly, N. C., Hutton, A., Mora, A., Portell, J., Van Leeuwen, F., Abbas, U., Abreu, A., Altmann, M., Andrei, A., Anglada, E., Balaguer-Núñez, L., Barache, C., Becciani, U., Bertone, S., Bianchi, L., Bouquillon, S., Bourda, G., Brösemeister, T., Bucciarelli, B., Busonero, D., Buzzi, R., Cancelliere, R., Carlucci, T., Charlot, P., Cheek, N., Crosta, M., Crowley, C., De Bruijne, J., De Felice, F., Drimmel, R., Esquej, P., Fienga, A., Fraile, E., Gai, M., Garralda, N., González-Vidal, J. J., Guerra, R., Hauser, M., Hofmann, W., Holl, B., Jordan, S., Lattanzi, M. G., Lenhardt, H., Liao, S., Licata, E., Lister, T., Löffler, W., Marchant, J., Martin-Fleitas, J. M., Messineo, R., Mignard, F., Morbidelli, R., Poggio, E., Riva, A., Rowell, N., Salguero, E., Sarasso, M., Sciacca, E., Siddiqui, H., Smart, R. L., Spagna, A., Steele, I., Taris, F., Torra, J., Van Elteren, A., Van Reeven, W., and Vecchiato, A. (2018). Gaia Data Release 2: The astrometric solution. *Astronomy and Astrophysics*, 616:1–35.

Lindegren, L., Klioner, S. A., Hernández, J., Bombrun, A., Ramos-Lerate, M., Steidelmüller, H., Bastian, U., Biermann, M., de Torres, A., Gerlach, E., Geyer, R., Hilger, T., Hobbs, D., Lammers, U., McMillan, P. J., Stephenson, C. A., Castañeda, J., Davidson, M., Fabricius, C., Gracia-Abril, G., Portell, J., Rowell, N., Teyssier, D., Torra, F., Bartolomé, S., Clotet, M., Garralda, N., González-Vidal, J. J., Torra, J., Abbas, U., Altmann, M., Anglada Varela, E., Balaguer-Núñez, L., Balog, Z., Barache, C., Becciani, U., Bernet, M., Bertone, S., Bianchi, L., Bouquillon, S., Brown, A. G., Bucciarelli, B., Busonero, D., Butkevich, A. G., Buzzi, R., Cancelliere, R., Carlucci, T., Charlot, P., Cioni, M. R., Crosta, M., Crowley, C., del Peloso, E. F., del Pozo, E., Drimmel, R., Esquej, P., Fienga, A., Fraile, E., Gai, M., Garcia-Reinaldos, M., Guerra, R., Hambly, N. C., Hauser, M., Janßen, K., Jordan, S., Kostrzewa-Rutkowska, Z., Lattanzi, M. G., Liao, S., Licata, E., Lister, T. A., Löffler, W., Marchant, J. M., Masip, A., Mignard, F., Mints, A., Molina, D., Mora, A., Morbidelli, R., Murphy, C. P., Pagani, C., Panuzzo, P., Peñalosa Esteller, X., Poggio, E., Re Fiorentin, P., Riva, A., Sagristà Sellés, A., Sanchez Gimenez, V., Sarasso, M., Sciacca, E., Siddiqui, H. I., Smart, R. L., Souami, D., Spagna, A., Steele, I. A., Taris, F., Utrilla, E., van Reeven, W., and Vecchiato, A. (2020). Gaia early data release 3: The astrometric solution. *arXiv*, 649(A17):A2.

Liu, C. (2019). Smoking Gun of the Dynamical Processing of the Solar-type Field Binary Stars. 17(July):1–17.

Liu, L. and Pang, X. (2019). A catalog of newly identified star clusters in GAIA DR2.

Luri, X., Brown, A. G., Sarro, L. M., Arenou, F., Bailer-Jones, C. A., Castro-Ginard, A., Bruijne, J. D., Prusti, T., Babusiaux, C., and Delgado, H. E. (2018). Gaia Data Release 2: Using Gaia parallaxes. *Astronomy and Astrophysics*, 616(32964):1–20.

Lyttle, A. J., Davies, G. R., Li, T., Carboneau, L. M., Leung, H. H., Westwood, H., Chaplin, W. J., Hall, O. J., Huber, D., Nielsen, M. B., Basu, S., and Garciá, R. A. (2021). Hierarchically modelling Kepler dwarfs and subgiants to improve inference of stellar properties with asteroseismology. *Monthly Notices of the Royal Astronomical Society*, 505(2):2427–2446.

Majewski, S. R., Schiavon, R. P., Frinchaboy, P. M., Prieto, C. A., Barkhouser, R., Bizyaev, D., Blank, B., Brunner, S., Burton, A., Carrera, R., Chojnowski, S. D., Cunha, K., Epstein, C., Fitzgerald, G., Pérez, A.

E. G., Hearty, F. R., Henderson, C., Holtzman, J. A., Johnson, J. A., Lam, C. R., Lawler, J. E., Maseman, P., Mészáros, S., Nelson, M., Nguyen, D. C., Nidever, D. L., Pinsonneault, M., Shetrone, M., Smee, S., Smith, V. V., Stolberg, T., Skrutskie, M. F., Walker, E., Wilson, J. C., Zasowski, G., Anders, F., Basu, S., Beland, S., Blanton, M. R., Bovy, J., Brownstein, J. R., Carlberg, J., Chaplin, W., Chiappini, C., Eisenstein, D. J., Elsworth, Y., Feuillet, D., Fleming, S. W., Galbraith-Frew, J., García, R. A., García-Hernández, D. A., Gillespie, B. A., Girardi, L., Gunn, J. E., Hasselquist, S., Hayden, M. R., Hekker, S., Ivans, I., Kinemuchi, K., Klaene, M., Mahadevan, S., Mathur, S., Mosser, B., Muna, D., Munn, J. A., Nichol, R. C., O'Connell, R. W., Parejko, J. K., Robin, A. C., Rocha-Pinto, H., Schultheis, M., Serenelli, A. M., Shane, N., Aguirre, V. S., Sobeck, J. S., Thompson, B., Troup, N. W., Weinberg, D. H., and Zamora, O. (2017). The Apache Point Observatory Galactic Evolution Experiment (APOGEE). *The Astronomical Journal*, 154(3):94.

Maxted, P. F. (2016). Ellc: A fast, flexible light curve model for detached eclipsing binary stars and transiting exoplanets. *Astronomy and Astrophysics*, 591.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics*, 21(6):1087–1092.

Miller, N. J., Maxted, P. F., and Smalley, B. (2020). Fundamental effective temperature measurements for eclipsing binary stars-I. Development of the method and application to AI Phoenicis. *Monthly Notices of the Royal Astronomical Society*, 497(3):2899–2909.

Morton, T. D. (2015). isochrones: Stellar model grid package. Astrophysics Source Code Library, record ascl:1503.010.

Neal, R. M. (2012). MCMC using Hamiltonian dynamics. *Handbook of Markov Chain Monte Carlo*, pages 113–162.

Olivares, J., Bouy, H., Sarro, L. M., Miret-Roig, N., Berihuete, A., Bertin, E., Barrado, D., Huélamo, N., Tamura, M., Allen, L., Beletsky, Y., Serre, S., and Cuillandre, J. C. (2019). Ruprecht 147 DANCe. *Astronomy and Astrophysics*, 625:A115.

Pearson, K. (1893). Contributions to the mathematical theory of evolution. *Bulletin of Mathematical Biology*, 185(71).

Perets, H. B. (2015). C11 - The Multiple Origin of Blue Straggler Stars: Theory vs. Observations. pages 251–275.

Price-Whelan, A., Breivik, K., D'Orazio, D., Hogg, D. W., Johnson, L. C., Moe, M., Morton, T. D., and Tayar, J. (2019). Stellar multiplicity: an interdisciplinary nexus. , 51(3):206.

Price-Whelan, A. M., Hogg, D. W., Foreman-Mackey, D., and Rix, H.-W. (2017). The Joker: A Custom Monte Carlo Sampler for Binary-star and Exoplanet Radial Velocity Data. *The Astrophysical Journal*, 837(1):20.

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., Beaton, R. L., Lewis, H. M., Nidever, D. L., Almeida, A., Badenes, C., Barba, R., Beers, T. C., Carlberg, J. K., De Lee, N., Fernández-Trincado, J. G., Frinchaboy, P. M., García-Hernández, D. A., Green, P. J., Hasselquist, S., Longa-Peña, P., Majewski, S. R., Nitschelm, C., Sobeck, J., Stassun, K. G., Stringfellow, G. S., and Troup, N. W. (2020). Close Binary Companions to APOGEE DR16 Stars: 20,000 Binary-star Systems Across the Color–Magnitude Diagram. *The Astrophysical Journal*, 895(1):2.

Price-Whelan, A. M., Hogg, D. W., Rix, H.-W., De Lee, N., Majewski, S. R., Nidever, D. L., Troup, N., Fernández-Trincado, J. G., García-Hernández, D. A., Longa-Peña, P., Nitschelm, C., Sobeck, J., and Zamora, O. (2018). Binary Companions of Evolved Stars in APOGEE DR14: Search Method and Catalog of ∼5000 Companions. , 156(1):18.

Price-Whelan, A. M., Nidever, D. L., Choi, Y., Schlafly, E. F., Morton, T., Koposov, S. E., and Belokurov, V. (2019). Discovery of a Disrupting Open Cluster Far into the Milky Way Halo: A Recent Star Formation Event in the Leading Arm of the Magellanic Stream? *The Astrophysical Journal*, 887(1):19.

Prša, A., Batalha, N., Slawson, R. W., Doyle, L. R., Welsh, W. F., Orosz, J. A., Seager, S., Rucker, M., Mjaseth, K., Engle, S. G., Conroy, K., Jenkins, J., Caldwell, D., Koch, D., and Borucki, W. (2011). KEPLER ECLIPSING BINARY STARS. I. CATALOG AND PRINCIPAL CHARACTERIZATION OF 1879 ECLIPSING BINARIES IN THE FIRST DATA RELEASE. *The Astronomical Journal*, 141(3):83.

Prša, A., Kochoska, A., Conroy, K. E., Eisner, N., Hey, D. R., IJspeert, L., Kruse, E., Fleming, S. W., Johnston, C., Kristiansen, M. H., LaCourse, D., Mortensen, D., Pepper, J., Stassun, K. G., Torres, G., Abdul-Masih, M., Chakraborty, J., Gagliano, R., Guo, Z., Hambleton, K., Hong, K., Jacobs, T., Jones, D., Kostov, V., Lee, J. W., Omohundro, M., Orosz, J. A., Page, E. J., Powell, B. P., Rappaport, S., Reed, P., Schnittman, J., Schwengeler, H. M., Shporer, A., Terentev, I. A., Vanderburg, A., Welsh, W. F., Caldwell, D. A., Doty, J. P., Jenkins, J. M., Latham, D. W., Ricker, G. R., Seager, S., Schlieder, J. E., Shiao, B., Vanderspek, R., and Winn, J. N. (2022). TESS Eclipsing Binary Stars. I. Short-cadence Observations of 4584 Eclipsing Binaries in Sectors 1–26. *The Astrophysical Journal Supplement Series*, 258(1):16.

Prusti, T., de Bruijne, J. H. J., Brown, A. G. A., Vallenari, A., Babusiaux, C., Bailer-Jones, C. A. L., Bastian, U., Biermann, M., Evans, D. W., Eyer, L., Jansen, F., Jordi, C., Klioner, S. A., Lammers, U., Lindegren, L., Luri, X., Mignard, F., Milligan, D. J., Panem, C., Poinsignon, V., Pourbaix, D., Randich, S., Sarri, G., Sartoretti, P., Siddiqui, H. I., Soubiran, C., Valette, V., van Leeuwen, F., Walton, N. A., Aerts, C., Arenou, F., Cropper, M., Drimmel, R., Høg, E., Katz, D., Lattanzi, M. G., O'Mullane, W., Grebel, E. K., Holland, A. D., Huc, C., Passot, X., Bramante, L., Cacciari, C., Castañeda, J., Chaoul, L., Cheek, N., De Angeli, F., Fabricius, C., Guerra, R., Hernández, J., Jean-Antoine-Piccolo, A., Masana, E., Messineo, R., Mowlavi, N., Nienartowicz, K., Ordóñez-Blanco, D., Panuzzo, P., Portell, J., Richards, P. J., Riello, M., Seabroke, G. M., Tanga, P., Thévenin, F., Torra, J., Els, S. G., Gracia-Abril, G., Comoretto, G., Garcia-Reinaldos, M., Lock, T., Mercier, E., Altmann, M., Andrae, R., Astraatmadja, T. L., Bellas-Velidis, I., Benson, K., Berthier, J., Blomme, R., Busso, G., Carry, B., Cellino, A., Clementini, G., Cowell, S., Creevey, O., Cuypers, J., Davidson, M., De Ridder, J., de Torres, A., Delchambre, L., Dell'Oro, A., Ducourant, C., Frémat, Y., García-Torres, M., Gosset, E., Halbwachs, J.-L., Hambly, N. C., Harrison, D. L., Hauser, M., Hestroffer, D., Hodgkin, S. T., Huckle, H. E., Hutton, A., Jasniewicz, G., Jordan, S., Kontizas, M., Korn, A. J., Lanzafame, A. C., Manteiga, M., Moitinho, A., Muinonen, K., Osinde, J., Pancino, E., Pauwels, T., Petit, J.-M., Recio-Blanco, A., Robin, A. C., Sarro, L. M., Siopis, C., Smith, M., Smith, K. W., Sozzetti, A., Thuillot, W., van Reeven, W., Viala, Y., Abbas, U., Abreu Aramburu, A., Accart, S., Aguado, J. J., Allan, P. M., Allasia, W., Altavilla, G., Álvarez, M. A., Alves, J., Anderson, R. I., Andrei, A. H., Anglada Varela, E., Antiche, E., Antoja, T., Antón, S., Arcay, B., Atzei, A., Ayache, L., Bach, N., Baker, S. G., Balaguer-Núñez, L., Barache, C., Barata, C., Barbier, A., Barblan, F., Baroni, M., Barrado y Navascués, D., Barros, M., Barstow, M. A., Becciani, U., Bellazzini, M., Bellei, G., Bello García, A., Belokurov, V., Bendjoya, P., Berihuete, A., Bianchi, L., Bienaymé, O., Billebaud, F., Blagorodnova, N., Blanco-Cuaresma, S., Boch, T., Bombrun, A., Borrachero, R., Bouquillon, S., Bourda, G., Bouy, H., Bragaglia, A., Breddels, M. A., Brouillet, N., Brüsemeister, T., Bucciarelli, B., Budnik, F., Burgess, P., Burgon, R., Burlacu, A., Busonero, D., Buzzi, R., Caffau, E., Cambras, J., Campbell, H., Cancelliere, R., Cantat-Gaudin, T., Carlucci, T., Carrasco, J. M., Castellani, M., Charlot, P., Charnas, J., Charvet, P., Chassat, F., Chiavassa, A., Clotet, M., Cocozza, G., Collins, R. S., Collins, P., Costigan, G., Crifo, F., Cross, N. J. G., Crosta, M., Crowley, C., Dafonte, C., Damerdji, Y., Dapergolas, A., David, P., David, M., De Cat, P., de Felice, F., de Laverny, P., De Luise, F., De March, R., de Martino, D., de Souza, R., Debosscher, J., del Pozo, E., Delbo, M., Delgado, A., Delgado, H. E., di Marco, F., Di Matteo, P., Diakite, S., Distefano, E., Dolding, C., Dos Anjos, S., Drazinos, P., Durán, J., Dzigan, Y., Ecale, E., Edvardsson, B., Enke, H., Erdmann, M., Escolar, D., Espina, M., Evans, N. W., Eynard Bontemps, G., Fabre, C., Fabrizio, M., Faigler, S., Falcão, A. J., Farràs Casas, M., Faye, F., Federici, L., Fedorets, G., Fernández-Hernández, J., Fernique, P., Fienga, A., Figueras, F., Filippi, F., Findeisen, K., Fonti, A., Fouesneau, M., Fraile, E., Fraser, M., Fuchs, J., Furnell, R., Gai, M., Galleti, S., Galluccio, L., Garabato, D., García-Sedano, F., Garé, P., Garofalo, A., Garralda, N., Gavras, P., Gerssen, J., Geyer, R., Gilmore, G., Girona, S., Giuffrida, G., Gomes, M., González-Marcos, A., González-Núñez, J., González-Vidal, J. J., Granvik, M., Guerrier, A., Guillout, P., Guiraud, J., Gúrpide, A., Gutiérrez-Sánchez, R., Guy, L. P., Haigron, R., Hatzidimitriou, D., Haywood, M., Heiter, U., Helmi, A., Hobbs, D., Hofmann, W., Holl, B., Holland, G., Hunt, J. A. S., Hypki, A., Icardi, V., Irwin, M., Jevardat de Fombelle, G., Jofré, P., Jonker, P. G., Jorissen, A., Julbe, F., Karampelas, A., Kochoska, A., Kohley, R., Kolenberg, K., Kontizas, E., Koposov, S. E., Kordopatis, G., Koubsky, P., Kowalczyk, A., Krone-Martins,

A., Kudryashova, M., Kull, I., Bachchan, R. K., Lacoste-Seris, F., Lanza, A. F., Lavigne, J.-b., Le Poncin-Lafitte, C., Lebreton, Y., Lebzelter, T., Leccia, S., Leclerc, N., Lecoeur-Taibi, I., Lemaitre, V., Lenhardt, H., Leroux, F., Liao, S., Licata, E., Lindstrøm, H. E. P., Lister, T. A., Livanou, E., Lobel, A., Löffler, W., López, M., Lopez-Lozano, A., Lorenz, D., Loureiro, T., MacDonald, I., Magalhães Fernandes, T., Managau, S., Mann, R. G., Mantelet, G., Marchal, O., Marchant, J. M., Marconi, M., Marie, J., Marinoni, S., Marrese, P. M., Marschalkó, G., Marshall, D. J., Martín-Fleitas, J. M., Martino, M., Mary, N., Matijevič, G., Mazeh, T., McMillan, P. J., Messina, S., Mestre, A., Michalik, D., Millar, N. R., Miranda, B. M. H., Molina, D., Molinaro, R., Molinaro, M., Molnár, L., Moniez, M., Montegriffo, P., Monteiro, D., Mor, R., Mora, A., Morbidelli, R., Morel, T., Morgenthaler, S., Morley, T., Morris, D., Mulone, A. F., Muraveva, T., Musella, I., Narbonne, J., Nelemans, G., Nicastro, L., Noval, L., Ordénovic, C., Ordieres-Meré, J., Osborne, P., Pagani, C., Pagano, I., Pailler, F., Palacin, H., Palaversa, L., Parsons, P., Paulsen, T., Pecoraro, M., Pedrosa, R., Pentikäinen, H., Pereira, J., Pichon, B., Piersimoni, A. M., Pineau, F.-x., Plachy, E., Plum, G., Poujoulet, E., Prša, A., Pulone, L., Ragaini, S., Rago, S., Rambaux, N., Ramos-Lerate, M., Ranalli, P., Rauw, G., Read, A., Regibo, S., Renk, F., Reylé, C., Ribeiro, R. A., Rimoldini, L., Ripepi, V., Riva, A., Rixon, G., Roelens, M., Romero-Gómez, M., Rowell, N., Royer, F., Rudolph, A., Ruiz-Dern, L., Sadowski, G., Sagristà Sellés, T., Sahlmann, J., Salgado, J., Salguero, E., Sarasso, M., Savietto, H., Schnorhk, A., Schultheis, M., Sciacca, E., Segol, M., Segovia, J. C., Segransan, D., Serpell, E., Shih, I.-c., Smareglia, R., Smart, R. L., Smith, C., Solano, E., Solitro, F., Sordo, R., Soria Nieto, S., Souchay, J., Spagna, A., Spoto, F., Stampa, U., Steele, I. A., Steidelmüller, H., Stephenson, C. A., Stoev, H., Suess, F. F., Süveges, M., Surdej, J., Szabados, L., Szegedi-Elek, E., Tapiador, D., Taris, F., Tauran, G., Taylor, M. B., Teixeira, R., Terrett, D., Tingley, B., Trager, S. C., Turon, C., Ulla, A., Utrilla, E., Valentini, G., van Elteren, A., Van Hemelryck, E., van Leeuwen, M., Varadi, M., Vecchiato, A., Veljanoski, J., Via, T., Vicente, D., Vogt, S., Voss, H., Votruba, V., Voutsinas, S., Walmsley, G., Weiler, M., Weingrill, K., Werner, D., Wevers, T., Whitehead, G., Wyrzykowski, , Yoldas, A., Žerjal, M., Zucker, S., Zurbach, C., Zwitter, T., Alecu, A., Allen, M., Allende Prieto, C., Amorim, A., Anglada-Escudé, G., Arsenijevic, V., Azaz, S., Balm, P., Beck, M., Bernstein, H.-h., Bigot, L., Bijaoui, A., Blasco, C., Bonfigli, M., Bono, G., Boudreault, S., Bressan, A., Brown, S., Brunet, P.-m., Bunclark, P., Buonanno, R., Butkevich, A. G., Carret, C., Carrion, C., Chemin, L., Chéreau, F., Corcione, L., Darmigny, E., de Boer, K. S., de Teodoro, P., de Zeeuw, P. T., Delle Luche, C., Domingues, C. D., Dubath, P., Fodor, F., Frézouls, B., Fries, A., Fustes, D., Fyfe, D., Gallardo, E., Gallegos, J., Gardiol, D., Gebran, M., Gomboc, A., Gómez, A., Grux, E., Gueguen, A., Heyrovsky, A., Hoar, J., Iannicola, G., Isasi Parache, Y., Janotto, A.-m., Joliet, E., Jonckheere, A., Keil, R., Kim, D.-w., Klagyivik, P., Klar, J., Knude, J., Kochukhov, O., Kolka, I., Kos, J., Kutka, A., Lainey, V., LeBouquin, D., Liu, C., Loreggia, D., Makarov, V. V., Marseille, M. G., Martayan, C., Martinez-Rubi, O., Massart, B., Meynadier, F., Mignot, S., Munari, U., Nguyen, A.-t., Nordlander, T., Ocvirk, P., O'Flaherty, K. S., Olias Sanz, A., Ortiz, P., Osorio, J., Oszkiewicz, D., Ouzounis, A., Palmer, M., Park, P., Pasquato, E., Peltzer, C., Peralta, J., Péturaud, F., Pieniluoma, T., Pigozzi, E., Poels, J., Prat, G., Prod'homme, T., Raison, F., Rebordao, J. M., Risquez, D., Rocca-Volmerange, B., Rosen, S., Ruiz-Fuertes, M. I., Russo, F., Sembay, S., Serraller Vizcaino, I., Short, A., Siebert, A., Silva, H., Sinachopoulos, D., Slezak, E., Soffel, M., Sosnowska, D., Straizys, V., ter Linden, M., Terrell, D., Theil, S., Tiede, C., Troisi, L., Tsalmantza, P., Tur, D., Vaccari, M., Vachier, F., Valles, P., Van Hamme, W., Veltz, L., Virtanen, J., Wallut, J.-M., Wichmann, R., Wilkinson, M. I., Ziaeepour, H., and Zschocke, S. (2016). The Gaia mission. *Astronomy & Astrophysics*, 595:A1.

Queiroz, A. B., Anders, F., Santiago, B. X., Chiappini, C., Steinmetz, M., Dal Ponte, M., Stassun, K. G., da Costa, L. N., Maia, M. A., Crestani, J., Beers, T. C., Fernández-Trincado, J. G., García-Hernández, D. A., Roman-Lopes, A., and Zamora, O. (2018). StarHorse: A Bayesian tool for determining stellar masses, ages, distances, and extinctions for field stars. *Monthly Notices of the Royal Astronomical Society*, 476(2):2556–2583.

Ricker, G. R., Winn, J. N., Vanderspek, R., Latham, D. W., Bakos, G. , Bean, J. L., Berta-Thompson, Z. K., Brown, T. M., Buchhave, L., Butler, N. R., Butler, R. P., Chaplin, W. J., Charbonneau, D., Christensen-Dalsgaard, J., Clampin, M., Deming, D., Doty, J., De Lee, N., Dressing, C., Dunham, E. W., Endl, M., Fressin, F., Ge, J., Henning, T., Holman, M. J., Howard, A. W., Ida, S., Jenkins, J. M., Jernigan, G., Johnson, J. A., Kaltenegger, L., Kawai, N., Kjeldsen, H., Laughlin, G., Levine, A. M., Lin, D., Lissauer, J. J., MacQueen, P., Marcy, G., McCullough, P. R., Morton, T. D., Narita, N., Paegert, M., Palle, E., Pepe,

F., Pepper, J., Quirrenbach, A., Rinehart, S. A., Sasselov, D., Sato, B., Seager, S., Sozzetti, A., Stassun, K. G., Sullivan, P., Szentgyorgyi, A., Torres, G., Udry, S., and Villasenor, J. (2014). Transiting Exoplanet Survey Satellite. *Journal of Astronomical Telescopes, Instruments, and Systems*, 1(1):014003.

Riello, M., De Angeli, F., Evans, D. W., Busso, G., Hambly, N. C., Davidson, M., Burgess, P. W., Montegriffo, P., Osborne, P. J., Kewley, A., Carrasco, J. M., Fabricius, C., Jordi, C., Cacciari, C., Van Leeuwen, F., and Holland, G. (2018). Gaia Data Release 2: Processing of the photometric data. *Astronomy and Astrophysics*, 616:1–19.

Riello, M., de Angeli, F., Evans, D. W., Montegriffo, P., Carrasco, J. M., Busso, G., Palaversa, L., Burgess, P. W., Diener, C., Davidson, M., Rowell, N., Fabricius, C., Jordi, C., Bellazzini, M., Pancino, E., Harrison, D. L., Cacciari, C., van Leeuwen, F., Hambly, N. C., Hodgkin, S. T., Osborne, P. J., Altavilla, G., Barstow, M. A., Brown, A. G., Castellani, M., Cowell, S., de Luise, F., Gilmore, G., Giuffrida, G., Hidalgo, S., Holland, G., Marinoni, S., Pagani, C., Piersimoni, A. M., Pulone, L., Ragaini, S., Rainer, M., Richards, P. J., Sanna, N., Walton, N. A., Weiler, M., and Yoldas, A. (2020). Gaia early data release 3: Photometric content and validation. *arXiv*, 649(2021):A3.

Rix, H.-W., Ting, Y.-S., Sippel, A., Beown, A., Anderson, S., Kollmeier, J., and Kupfer, T. (2019). Binaries Matter Everywhere: from Precision Calibrations to Re-Ionization and Gravitational Waves. , 51(3):104.

Salvatier, J., Wiecki, T. V., and Fonnesbeck, C. (2016). Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, 2(4):e55.

Santana, F. A., Beaton, R. L., Covey, K. R., O'Connell, J. E., Longa-Peña, P., Cohen, R., Fernández-Trincado, J. G., Hayes, C. R., Zasowski, G., Sobeck, J. S., Majewski, S. R., Chojnowski, S. D., De Lee, N., Oelkers, R. J., Stringfellow, G. S., Almeida, A., Anguiano, B., Donor, J., Frinchaboy, P. M., Hasselquist, S., Johnson, J. A., Kollmeier, J. A., Nidever, D. L., Price-Whelan, A. M., Rojas-Arriagada, A., Schultheis, M., Shetrone, M., Simon, J. D., Aerts, C., Borissova, J., Drout, M. R., Geisler, D., Law, C. Y., Medina, N., Minniti, D., Monachesi, A., Muñoz, R. R., Poleski, R., Roman-Lopes, A., Schlaufman, K. C., Stutz, A. M., Teske, J., Tkachenko, A., Van Saders, J. L., Weinberger, A. J., and Zoccali, M. (2021). Final Targeting Strategy for the SDSS-IV APOGEE-2S Survey. , 162(6):303.

Sebastian, D., Swayne, M. I., Maxted, P. F. L., Triaud, A. H. M. J., Sousa, S. G., Olofsson, G., Beck, M., Billot, N., Hoyer, S., Gill, S., Heidari, N., Martin, D. V., Persson, C. M., Standing, M. R., Alibert, Y., Alonso, R., Anglada, G., Asquier, J., Bárczy, T., Barrado, D., Barros, S. C. C., Battley, M. P., Baumjohann, W., Beck, T., Benz, W., Bergomi, M., Boisse, I., Bonfils, X., Brandeker, A., Broeg, C., Cabrera, J., Charnoz, S., Cameron, A. C., Csizmadia, S., Davies, M. B., Deleuil, M., Delrez, L., Demangeon, O. D. S., Demory, B. O., Dransfield, G., Ehrenreich, D., Erikson, A., Fortier, A., Fossati, L., Fridlund, M., Gandolfi, D., Gillon, M., Güdel, M., Hasiba, J., Hébrard, G., Heng, K., Isaak, K. G., Kiss, L. L., Kopp, E., Kunovac, V., Laskar, J., Etangs, A. L. d., Lendl, M., Lovis, C., Magrin, D., McCormac, J., Miller, N. J., Nascimbeni, V., Ottensamer, R., Pagano, I., Pallé, E., Pepe, F. A., Peter, G., Piotto, G., Pollacco, D., Queloz, D., Ragazzoni, R., Rando, N., Rauer, H., Ribas, I., Lalitha, S., Santerne, A., Santos, N. C., Scandariato, G., Ségransan, D., Simon, A. E., Smith, A. M. S., Steller, M., Szabó, G. M., Thomas, N., Udry, S., Van Grootel, V., and Walton, N. A. (2022). The EBLM project – IX. Five fully convective M-dwarfs, precisely measured with CHEOPS and TESS light curves. 19(September):1–19.

Serenelli, A., Weiss, A., Aerts, C., Angelou, G. C., Baroch, D., Bastian, N., Beck, P. G., Bergemann, M., Bestenlehner, J. M., Czekala, I., Elias-Rosa, N., Escorza, A., Van Eylen, V., Feuillet, D. K., Gandolfi, D., Gieles, M., Girardi, L., Lebreton, Y., Lodieu, N., Martig, M., Miller Bertolami, M. M., Mombarg, J. S. G., Morales, J. C., Moya, A., Nsamba, B., Pavlovski, K., Pedersen, M. G., Ribas, I., Schneider, F. R. N., Silva Aguirre, V., Stassun, K. G., Tolstoy, E., Tremblay, P.-E., and Zwintz, K. (2021). Weighing stars from birth to death: mass determination methods across the HRD. *The Astronomy and Astrophysics Review*, 29(1):4.

Sim, G., Lee, S. H., Ann, H. B., and Kim, S. (2019). 207 new open star clusters within 1 kpc from gaia data release 2. *Journal of the Korean Astronomical Society*, 52(5):145–158.

Skrutskie, M. F., Cutri, R. M., Stiening, R., Weinberg, M. D., Schneider, S., Carpenter, J. M., Beichman, C., Capps, R., Chester, T., Elias, J., Huchra, J., Liebert, J., Lonsdale, C., Monet, D. G., Price, S., Seitzer, P., Jarrett, T., Kirkpatrick, J. D., Gizis, J. E., Howard, E., Evans, T., Fowler, J., Fullmer, L., Hurt, R., Light, R., Kopan, E. L., Marsh, K. A., McCallon, H. L., Tam, R., Van Dyk, S., and Wheelock, S. (2006). The Two Micron All Sky Survey (2MASS). *The Astronomical Journal*, 131(2):1163–1183.

Soderblom, D. R. (2010). The Ages of Stars.

Speagle, J. S. (2020). DYNESTY: a dynamic nested sampling package for estimating Bayesian posteriors and evidences. , 493(3):3132–3158.

Spina, L., Ting, Y.-S., De Silva, G. M., Frankel, N., Sharma, S., Cantat-Gaudin, T., Joyce, M., Stello, D., Karakas, A. I., Asplund, M. B., Nordlander, T., Casagrande, L., D'Orazi, V., Casey, A. R., Cottrell, P., Tepper-García, T., Baratella, M., Kos, J., Čotar, K., Bland-Hawthorn, J., Buder, S., Freeman, K. C., Hayden, M. R., Lewis, G. F., Lin, J., Lind, K., Martell, S. L., Schlesinger, K. J., Simpson, J. D., Zucker, D. B., and Zwitter, T. (2021). The GALAH survey: tracing the Galactic disc with open clusters. *Monthly Notices of the Royal Astronomical Society*, 503(3):3279–3296.

Stacy, A., Greif, T. H., and Bromm, V. (2009). The first stars: formation of binaries and small multiple systems. *Monthly Notices of the Royal Astronomical Society*, 403(1):45–60.

Stassun, K. G., Oelkers, R. J., Paegert, M., Torres, G., Pepper, J., Lee, N. D., Collins, K., Latham, D. W., Muirhead, P. S., Chittidi, J., Rojas-Ayala, B., Fleming, S. W., Rose, M. E., Tenenbaum, P., Ting, E. B., Kane, S. R., Barclay, T., Bean, J. L., Brassuer, C. E., Charbonneau, D., Ge, J., Lissauer, J. J., Mann, A. W., McLean, B., Mullally, S., Narita, N., Plavchan, P., Ricker, G. R., Sasselov, D., Seager, S., Sharma, S., Shiao, B., Sozzetti, A., Stello, D., Vanderspek, R., Wallace, G., and Winn, J. N. (2019). The Revised TESS Input Catalog and Candidate Target List . *The Astronomical Journal*, 158(4):138.

Stassun, K. G., Oelkers, R. J., Pepper, J., Paegert, M., DeLee, N., Torres, G., Latham, D. W., Charpinet, S., Dressing, C. D., Huber, D., Kane, S. R., Lepine, S., Mann, A., Muirhead, P. S., Rojas-Ayala, B., Silvotti, R., Fleming, S. W., Levine, A., Plavchan, P., and Group, t. T. T. S. W. (2017). The TESS Input Catalog and Candidate Target List. *The Astronomical Journal*, 156(3):102.

Stassun, K. G., Oelkers, R. J., Pepper, J., Paegert, M., Lee, N. D., Torres, G., Latham, D. W., Charpinet, S., Dressing, C. D., Huber, D., Kane, S. R., Lépine, S., Mann, A., Muirhead, P. S., Rojas-Ayala, B., Silvotti, R., Fleming, S. W., Levine, A., and Plavchan, P. (2018). The TESS Input Catalog and Candidate Target List. *The Astronomical Journal*, 156(3):102.

Sullivan, P. W., Winn, J. N., Berta-Thompson, Z. K., Charbonneau, D., Deming, D., Dressing, C. D., Latham, D. W., Levine, A. M., McCullough, P. R., Morton, T., Ricker, G. R., Vanderspek, R., and Woods, D. (2015). THE TRANSITING EXOPLANET SURVEY SATELLITE: SIMULATIONS OF PLANET DETECTIONS AND ASTROPHYSICAL FALSE POSITIVES. *The Astrophysical Journal*, 809(1):77.

Tarricq, Y., Soubiran, C., Casamiquela, L., Castro-Ginard, A., Olivares, J., Miret-Roig, N., and Galli, P. A. B. (2022). Structural parameters of 389 local open clusters. *Astronomy & Astrophysics*, 659:A59.

Toonen, S., Portegies Zwart, S., Hamers, A. S., and Bandopadhyay, D. (2020). The evolution of stellar triples: The most common evolutionary pathways. *arXiv*, 16.

Torres, G. (2013). Fundamental properties of lower main-sequence stars. *Astronomische Nachrichten*, 334(1-2):4–9.

Torres, G., Andersen, J., and Giménez, A. (2010a). *Accurate masses and radii of normal stars: Modern results and applications*, volume 18.

Torres, G., Andersen, J., and Giménez, A. (2010b). Accurate masses and radii of normal stars: Modern results and applications. *Astronomy and Astrophysics Review*, 18(1-2):67–126.

Troup, N. W., Nidever, D. L., De Lee, N., Carlberg, J., Majewski, S. R., Fernandez, M., Covey, K., Chojnowski, S. D., Pepper, J., Nguyen, D. T., Stassun, K., Nguyen, D. C., Wisniewski, J. P., Fleming, S. W., Bizyaev, D., Frinchaboy, P. M., García-Hernández, D. A., Ge, J., Hearty, F., Meszaros, S., Pan, K., Allende Prieto, C., Schneider, D. P., Shetrone, M. D., Skrutskie, M. F., Wilson, J., and Zamora, O. (2016). Companions to APOGEE Stars. I. A Milky Way-spanning Catalog of Stellar and Substellar Companion Candidates and Their Diverse Hosts. , 151(3):85.

Utsunomiya, M., Hashimoto, M., and Kajitani, H. (1996). Development of a magnetic head suspension system for high-Speed seeking performance. *NEC Research and Development*, 37(3):369–381.

Vallenari, A., Brown, A., and Prusti, T. (2022). Gaia Data Release 3. Summary of the content and survey properties. *Astronomy & Astrophysics*, pages 1–23.

Vasiliev, E. (2019). Systematic errors in Gaia DR2 astrometry and their impact on measurements of internal kinematics of star clusters. *Monthly Notices of the Royal Astronomical Society*, 489(1):623–640.

Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., and Bürkner, P.-C. (2019). Rank-normalization, folding, and localization: An improved $\widehat{R}$ for assessing convergence of MCMC. *Bayesian Analysis*, 16(2):667–718.

Wenger, M., Ochsenbein, F., Egret, D., Dubois, P., Bonnarel, F., Borde, S., Genova, F., Jasniewicz, G., Laloe, S., Lesteven, S., and Monier, R. (2000). The SIMBAD astronomical database. 22:9–22.

Winn, J. N. (2010). Transits and Occultations.