



Probabilistic Parameter Determination of Stars and Galaxies

Research Thesis

Author: Billy Harrison¹

Supervisor: Dr. Ralph Schönrich^{1,2}

MSc Astrophysics

September 7, 2025

¹University College London, Department of Physics & Astronomy, London, WC1E 6BT, UK

²Mullard Space Science Laboratory, University College London, Holmbury St. Mary, Dorking,
Surrey, RH5 6NT, UK

Abstract

Decoding the age and metallicity distributions of stars in the Solar neighbourhood is central to tracing the kinematic evolution of the Galactic disk, however direct age determinations are infamously uncertain. Traditional fundamental stellar parameter estimation techniques are hindered by large observational uncertainties and degeneracies from isochrones. Adopting the Bayesian methodology provides a pathway to a natural solution, enabling the integration of prior information and the probabilistic treatment of objective error calculations. In this work, we employ a Bayesian framework with isochrone fitting and apply it to *Gaia* stars with a given parallax cut. The incorporation of observational likelihoods that encompass relevant parameter space and their constraining power for age and metallicity determination is explored, with their posterior probability distribution functions (PDFs) acting as catalysts for statistical inference. We test our ages and compositions and find an age-velocity dispersion relation (AVR) consistent with the literature, with exponent $\beta = 0.29 \pm 0.01$, and expected higher vertical heating compared to in-plane. A negative metallicity-dispersion relation with kinematically hotter stars at intermediate metallicities was observed, as well as mostly consistent metallicity distribution functions given our examination of a parallax-limited sample. The results validate Bayesian isochrone fitting as a credible tool for parameter estimation, reproducing key Solar neighbourhood trends of disk heating and chemical evolution, but limitations such as large age uncertainty and over-prediction of old, metal-rich stars remain. Future improvements like prior refinement and combined photometric/spectroscopic likelihood constraints will refine the method and deepen our understanding of disk structure.

Word count: 19,115

Contents

Abstract	1
1 Introduction	4
1.1 Bayesian Inference	4
1.2 Previous Applications	5
1.2.1 Historical Origins	5
1.2.2 Recent Applications	5
1.2.3 Justifications	5
1.3 Paper Overview	6
2 Methodology Outline	7
2.1 Bayesian Formalism	7
2.2 Parameter Space & Observations	8
3 Practical Application of Algorithm	8
3.1 Priors	8
3.1.1 Choosing a Prior	8
3.1.2 Accounting for Flatness	9
3.1.3 Reparameterisation	9
3.1.4 Age and Metallicity Priors	10
3.1.5 Establishing our Choice	12
3.2 Isochrones	12
3.2.1 What are Isochrones?	12
3.2.2 Choice of Grid	14
3.3 Stellar Model PDFs	14
3.4 Constraining Metallicity	15
3.4.1 Stellar Theory and Observational Constraints	15
3.4.2 Application and Example	16
3.5 Observational Likelihoods	17
3.5.1 Context within Bayes' Theorem	17
3.5.2 Application to the <i>Gaia</i> Survey	18
3.5.3 Parameter Conversion	19
3.6 Marginalisation	19
3.6.1 Calculating Projections	19
3.6.2 Moments and Uncertainties	20
3.6.3 Confidence Intervals	22
3.7 Analysing Posteriors	23
3.7.1 Overview of Posterior Construction	23
3.7.2 Case Studies	23
4 Application: <i>Gaia</i> Survey	25
4.1 Stellar Kinematics	25
4.1.1 Velocities	25
4.1.2 Known Dispersion Forms	26
4.2 σ vs τ	26
4.2.1 Expected Relation	26
4.2.2 Ages and Uncertainties	27

4.2.3	Calculating Dispersion	28
4.2.4	Initial Fit	29
4.3	Selection Cuts	29
4.3.1	Velocity and Parallax	29
4.3.2	Age Uncertainty Filter	30
4.3.3	Observed Flattening	31
4.3.4	HRD Cuts	31
4.3.5	Final AVR	32
4.4	Population Boundaries	33
4.4.1	Definitions	33
4.4.2	Velocity, Age, Structure	34
4.4.3	Chemical Separation	34
4.5	σ vs [M/H]	35
4.5.1	Metallicity	35
4.5.2	More Selection Cuts	36
4.5.3	Results	37
4.5.4	Metallicity Gradient	38
4.6	Metallicity Distribution	38
4.6.1	Age-Metallicity Relation	38
4.6.2	Metallicity Distribution Functions	40
4.6.3	3-D Relationships	41
4.6.4	Context within the Solar Neighbourhood	41
5	Discussion	42
5.1	Future Improvements	43
5.1.1	Photometric Likelihoods	43
5.1.2	Spectroscopic Likelihoods	43
5.1.3	Astrometric Likelihoods	44
5.1.4	Degeneracies and Bias	45
5.1.5	Additional Advancements	45
5.2	Mapping the Solar Neighbourhood	46
5.2.1	Refined Population Separation	46
5.2.2	Disk Formation History	47
5.2.3	Discovering Substructure	47
6	Conclusion	48
7	Bibliography	49

1. Introduction

1.1. Bayesian Inference

Throughout all subfields of astrophysics, interstellar observations accumulate into the grand reservoir of knowledge that equips us with the tools to tackle the greatest questions and mysteries of the universe. Such observations and their uncertainties have fuelled the past few centuries of research, utilising several statistical techniques that have become canonised within standard approaches in the literature (Feigelson & Babu, 2012; Hogg et al., 2010), and which are now routinely used to study stellar population parameters en masse within the Milky Way (MW) (e.g. Rix & Bovy (2013)), and recently other galaxies as well. We can fill such a reservoir by exploiting the spectroscopic (anything involving the spectra of the electromagnetic radiation of stars), astrometric (the precise determination of the positions and movements of stars using parallax, proper motion, radial velocities) and photometric (capturing the flux of light with photometers to measure brightness) measurements of said populations under interest. Once we have a substantial supply of data through these observations, it is relatively straightforward to infer information on stellar samples, particularly stellar parameters such as chemical composition, mass, age, gravity, and kinematic knowledge that accounts for distances and characteristics such as velocity dispersion (e.g. Majewski et al. (2017)).

Historically, however, these parameter determinations have been filtered through a pipeline of best-fit estimates (Valle, G. et al., 2015) and simple averages between different contrasting methods, which may compromise juxtaposing results given identical initial assumptions. Averaging results from disparate evolutionary models may obscure or blur real model disagreements (Brown, 2014), or fail to reflect true propagated uncertainties, without accounting for parameter degeneracy or observational noise. This may be a problem as areas of interest such as age may depend strongly non-linearly on other parameter spaces. These methods also do not incorporate any prior historical astrophysical knowledge (e.g. star formation history, metallicity spread, initial mass function) into their frameworks, essentially going in blind with reference to the relationship between any expected distributions' coupled parameters, and how they will continue to evolve.

The advent of the recent increase in abundance and complexity of data sets from satellite missions such as *Gaia* (Lindegren et al., 2018) and *Hipparcos* (Esa, 1997) and their related follow-up surveys with the simultaneous exponential growth potential of computational power has gifted astrophysicists with an appropriate set of tools prime for the application of Bayesian methods. Hence, it is necessary to adopt fully integrated automation of data analysis in the context of stellar parameter determination (Ness et al., 2015), as manual exploration of stellar analysis is limited to the constraints of subjectivity and greatly reduced sample sizes incompatible with these types of large survey. The nature of this newfound data structure is no longer limited to un-idealised error estimates and gives astrophysicists the resolution to determine not just point estimates on stellar parameters, but distributions on the probabilistic likelihoods of these parameters (Silva Aguirre et al., 2015). Intelligence that encodes such a spread gives us valuable information on error distributions, e.g. variance and confidence intervals, and is key to making meaningful conclusions or links between various models of Galactic evolution and stellar structure. Therefore, the approach we need to adopt must be objective, easily programmable, and mouldable, combine every quantitative aspect of the observational data, and be applicable to stellar surveys as they become increasingly expansive. The best mathematical/statistical framework for adopting all desired qualities is Bayesian theory, hence it is a perfect fit in an astrophysical modelling context.

1.2. Previous Applications

1.2.1 Historical Origins

Although being advocated for long ago by [Jeffreys \(1939\)](#), the pragmatism and conceptual underpinnings of the Bayesian method were largely ignored in the field of astrophysics for many years. It wasn't until the late 80s/early 90s when the realisation that this school of thought would be well equipped to tackle astrophysical problems began to rear its head within the community of contemporary researchers. One of the progenitor advocates was [Loredo \(1992\)](#), who comprehensibly laid forward the case for its implementation, citing the observed interest growth in applied statistics and economics, and "look[ed] forward to a similar revolution occurring in astrophysics". He pioneered many studies in the field, including general event arrival time periodic signal detection ([Gregory & Loredo, 1992](#)), various studies on Supernovae ([Drell et al., 2000](#); [Loredo & Lamb, 1989](#)), and γ -ray burst distributions sources ([Loredo & Wasserman, 1998](#)). Other notable studies of the time include [Bailyn et al. \(1998\)](#)'s Bayesian analysis of stellar black hole mass distributions, [Hernandez et al. \(2000b\)](#)'s inference of star formation histories for local group spheroidal dwarf galaxies and the *Hipparcos* Solar neighbourhood (Snhd) ([Hernandez et al., 2000a](#)), and predictions of cosmological cold dark matter models by [Bunn & Sugiyama \(1995\)](#).

1.2.2 Recent Applications

Focusing on more contemporary stellar and Galactic physics applications, [Pont & Eyer \(2004\)](#) and [Jørgensen & Lindegren \(2005\)](#) were some of the first to present Bayesian methods of stellar age estimates from theoretical isochrones, and divulge into the statistical inference from the resulting age PDFs. [Nordström et al. \(2004\)](#) briefly outline similar approaches to age estimates in their landmark study of the *Geneva-Copenhagen-Survey* (GCS) of the Snhd. [Casagrande et al. \(2011\)](#) re-analyse the GCS to derive stellar and metallicity parameters where they provide details on the Bayesian scheme adopted for dealing with the observational errors in T_{eff} , metallicities and absolute magnitudes. [Bailer-Jones \(2011\)](#) adopts a Bayesian method for inferring intrinsic stellar parameters, using parallax and photometric data in unison to estimate non-parametric posterior probability distribution functions (PDFs). [Burnett & Binney \(2010\)](#) employ related techniques in their stellar distance inference determination. [Schönrich & Bergemann \(2014\)](#) similarly present a Bayesian framework/algorithm for stellar parameter derivation that combines all available observable and prior information. [Shkedy et al. \(2007\)](#) were among the first, in an astrophysical context, to account for both systematic and statistical measurement errors using Bayesian principles in their study of fitting theoretical spectra to observed values. In the context of the *Gaia* astrometry and spectroscopy which will be used in this work as a case study, [Liu et al. \(2012\)](#) test the performance of Bayesian algorithms, namely Aeneas, and how well they constrain effective temperature, line-of-sight extinction, metallicity, and surface gravity of A-M type stars. [Serenelli et al. \(2013\)](#) also derive masses, ages, and distances using Bayesian schemes and stellar evolutionary tracks in tandem. Additional Bayesian parameter determination efforts include open-cluster ages ([Jeffery et al., 2016](#)), line-of-sight velocities, T_{eff} , $\log g$, and [M/H] ([Walker et al., 2015](#)), and magnetic-field properties ([Petit & Wade, 2012](#)).

1.2.3 Justifications

It is clear from the abundance of previous literature, and the exponential increase in research as computational power has caught up to theory, that Bayesian inference is a major way forward. It provides a natural framework that neatly combines disparate information sources - stellar evolutionary theory encoded by the prior and contemporary observational data - into statements of probability.

Information accounts in this form are especially advantageous in an astrophysical context where data are often sparse, incomplete, or have problems with precision due to random error, or with accuracy due to systematic error.

The innovation of such inference has also been supplemented by the development of efficient algorithms that sample the posterior, such as Markov Chain Monte Carlo (MCMC - [Metropolis et al. \(1953\)](#)) methods ([Mann et al., 2015; Jullo et al., 2007](#)), nested sampling, model selection metrics ([Liddle, 2007; Schwarz, 1978](#)), and Metropolis-Hastings algorithms ([McMillan, 2011](#)), allowing higher-dimensional posterior parameter spaces to be explored with more rigour and completeness. This transition could also be argued to represent a greater philosophical shift from the use of best-fit models and deterministic principles towards model parameter distributions being probabilistically characterised. As we can see, this framework is not only now accepted, but is seen as the norm or default approach to parameter inference within astrophysics, marking a significant shift in how uncertainty and inference are viewed within the observational sciences.

1.3. Paper Overview

The paper is structured as follows. In Section 2 we provide a methodological outline of the Bayesian framework adopted throughout this work, specifically discussing the choice of parameter space and observational constraints. In Section 3 we will present and test a Bayesian algorithm for estimating stellar age and metallicity, beginning with a discussion on priors and our choice justification for these parameters. We explore the use of stellar evolutionary models called isochrones as proxies for mapping our physical parameters onto desired observables in the absence of any universally reliable age-estimation methods. We then describe the construction of stellar model PDFs that compound in conjunction with our priors in constraining our chosen parameters, as well as the incorporation of metallicity constraints via additional magnitude and colour observables. These are then linked to the formulation of observational likelihoods using *Gaia* astrometry and spectroscopy, and how they contribute to the construction of our posteriors. The section concludes with a discussion on posterior marginalisation and case-study analysis given a selection of test stars at physically interesting points on the colour-magnitude diagram (CMD).

In Section 4, we apply the above Bayesian method to a parallax-limited *Gaia* stellar sample in the Snhd. We first examine stellar kinematics within the context of our derived parameters, constructing age-velocity dispersion relations, before exploring metallicity-dispersion trends and the overall metallicity distribution of the population. Particular attention is paid to the effects of sample selection cuts and to distinguishing thin- and thick-disk populations in age, chemical, and kinematic space. Section 5 discusses the broader astrophysical implications of our results, highlighting current limitations of the Bayesian approach, and suggesting future pathways for method refinement, including additional observational constraints and age validation against other stellar-dating techniques.

The central motivation of this work is to establish whether our Bayesian inference methodology yields credible age and metallicity estimates when applied to *Gaia* stars in the Snhd. The age- and metallicity-dispersion trends provide vigorous tests as both are well-established diagnostics for disk-heating and chemical evolution. In attempting to reproduce their expected forms given our derived observational uncertainties, we probe the kinematic and chemical history of the disk, and validate the informing power of our Bayesian framework for general stellar parameter determination.

2. Methodology Outline

2.1. Bayesian Formalism

The majority of historical astrophysical research takes a more frequentist approach to parameter determination which treats the definitions of “probability” and “frequency” interchangeably, i.e. probability is encoded by the limit of the long-run frequencies estimated via sampling of a large number of repeatable trials. This approach does not consider exterior information or beliefs about the particular system being designed as a statistical model. Instead, we want full probability distributions given the available observations and prior knowledge, to explore similarities to theoretical models of astronomical phenomena, and to constrain parameter boundary space, which is given by a Bayesian formalism. This school of inference increases efficiency when dealing with problems related to the conceptual conclusions drawn from a set of premises in parameter space and the presence of uncertainty in said data. We must be very careful in our employment of these measurements, however, as systematic biases such as inadequate modelling, noisy or highly uncertain data, selection and temporal biases and thoughtless prior and posterior dependencies can steer us heavily in the wrong direction, and lead us to draw unfounded conclusions about how our universe behaves.

In our context, we express the probability of a set of stellar parameters $\mathbf{X} = X_1, X_2, \dots, X_n$ dependent on a set of observations $\mathbf{Y} = Y_1, Y_2, \dots, Y_n$, by the posterior probability $P(\mathbf{X} | \mathbf{Y})$ ([Schönrich & Bergemann, 2014](#)). This is the conditional probability of parameter set \mathbf{X} being true given an observation set \mathbf{Y} . The approach from here hinges on the result of a simple formalism known as Bayes’ theorem, given from the consequences of probability theory and presuming the Bayesian idea of how probability itself is defined - “probability is a measure of the degree of belief about a proposition” ([Trotta, 2008](#)) - and how it can be manipulated. The theorem is itself a direct consequence of the axiomatic foundations of probability theory as laid out by Andrey Kolmogorov, who noted that ”Every axiomatic (abstract) theory admits, as is well known, an unlimited number of concrete interpretations besides those from which it was derived” ([Kolmogorov, 1933](#)), underscoring the adaptability of Bayesian probability. Although itself a mathematical description of assigning probabilities, the theorem can be thought of qualitatively by exploiting our intuition to update our beliefs about the world around us in light of new evidence. In other words, it is a formal way to reason about belief revision: this is what the theorem tells us how to accomplish in the context of data described with mathematical PDFs. Bayes’ theorem shows that the posterior probability takes the form of

$$P(\mathbf{X} | \mathbf{Y}) = \frac{P(\mathbf{Y} | \mathbf{X}) \cdot P(\mathbf{X})}{P(\mathbf{Y})} \quad (1)$$

where $P(\mathbf{Y} | \mathbf{X})$ is the observed likelihood - the conditional probability of making a set of observations \mathbf{Y} given the known parameters \mathbf{X} and I, $P(\mathbf{X})$ is the prior probability, or what was known on \mathbf{X} *a priori* which intends to update our beliefs, and $P(\mathbf{Y})$ is a normalising factor independent of \mathbf{X} encoding the probability the set of observations \mathbf{Y} is made, which we can readily set to unity ([Pont & Eyer, 2004](#)). This can therefore be simplified to

$$P(\mathbf{X} | \mathbf{Y}) = P(\mathbf{Y} | \mathbf{X}) \cdot P(\mathbf{X}) \quad (2)$$

Usually, the observations being made are conditionally independent of the given parameters - i.e. the observations are uninformative on one another provided the parameters are perfectly known - meaning we can disentangle the observations and represent our posterior probabilities as a sum over our independent parameters. Qualitatively, the importance and great inferring power of this theorem lies in the fact that the posterior probability that a hypothesis is true given the data can be linked to the

probability that one would observe the data given the hypothesis being true - a term that has a much greater chance of assignment ([Sivia & Skilling, 2006](#)).

2.2. Parameter Space & Observations

In an astrophysical context, the parameter set describes all relevant parameters under investigation in our particular study. This could include a star's initial mass, age, chemical composition, position in sky, effective temperature, etc. Since stars are point sources and evidently describe single objects (as opposed to distant galaxies or clusters, which may appear as point sources), all of their parameters are intrinsically linked in one way or another. Hence, if we want to determine Bayesian estimates on stars that can only otherwise be inferred by empirical stellar models, we should choose a parameter space that can disentangle any dependencies observations may have on one another, ensuring all information is accounted for.

The parameter space chosen for this work will be a combination of $\log g$ (surface gravity), M_V (absolute visual magnitude), colour index (e.g. $B - V$, $g - r$, $G_{\text{bp}} - G_{\text{rp}}$), $\log T_{\text{eff}}$ (effective temperature), and $\log(L/L_{\odot})$ (luminosity). It is important to choose parameters that are sensitive to different evolutionary stages; e.g. $\log g$ and $\log(L/L_{\odot})$ can distinguish between lower main sequence (MS) and more evolved stars, or can capture the colour-temperature relation that informs stellar age and metallicity. Some parameters like luminosity can be degenerate with mass, e.g. an intermediate mass MS and a solar-like subgiant may both have $\log(L/L_{\odot}) = \sim 0.5$, depending on their age. Therefore, utilising the full boundaries of our parameter space to triangulate where the star truly lies on the Hertzsprung-Russell Diagram (HRD) by combining with colour or T_{eff} is useful in mitigating these degeneracies. For more discussion on this, see Section 3.4. Potential multi-modal Bayesian posterior PDFs also capture this spread in parameter likelihood areas and can be employed to break any degeneracies (see Figure 5). However, it is more likely that these multi-modal distributions are borne out of larger observational errors, especially in colour indices or gravity, so these broader or multi-peaked PDFs will actually reflect genuine uncertainty rather than discrete outcomes that non-Bayesian approaches will not retain and propagate, but instead collapsing them to potentially misleading single estimates. Hence, preserving the full posterior shape allows for ultimately transparent representations of regions of degeneracy.

Other parameters are ones of interest that we try to build posteriors, such as age and metallicity, and are constrained by evolutionary stellar models such as isochrones. Others, like parallax and distance, can inform on luminosity and magnitude and are constrained by astrometric observations or comparisons to photometric models. Our chosen parameter space is motivated by physical interpretations of movement along the HRD and accompanying stellar models; therefore, the posterior itself will probe a distinct, complementary aspect of stellar structure. It will constrain and maximise sensitivity on our interest parameters while minimising any degeneracies and potential uncertainty; hence a high-dimensional parameter space is suited for posterior inference justified by the use of a Bayesian framework.

3. Practical Application of Algorithm

3.1. Priors

3.1.1 Choosing a Prior

As discussed in Section 2.1, the choice of prior encodes any *a priori* knowledge of the stellar population distribution in the target parameter space. When data is weakly constrained, i.e. the preliminary

posterior is broad and the observed data do not sharply influence the test parameter, an appropriate choice of prior is needed to avoid any bias.

A strong or overly specific conjugate prior can impose unfounded assumptions or biased inferences that pull the posterior into mimicking the prior more so than the data (Burnett & Binney, 2010). In the case that the prior and data conflict with one another, the prior will have an unbounded influence on the posterior. Such an incompatibility between sources of information communicated via Bayes' theorem is counteracted by considering the tail-behaviour of both likelihood/prior “since conflicts are characterised by events occurring far into the tails. Thus, the speed at which the tails of the probability density decay to zero determines the importance of the source of information in the model.” (Andrade & Oney, 2013). Hence, the prior structure at high σ values should also be considered to avoid any bias.

A self-informed prior may be used in contexts where our initial prior is too constraining; the posterior can derive an updated prior and diagnose whether the prior is overly restrictive, preventing the Bayesian model from sampling the ‘true’ posterior. If there exists a large discrepancy between the original and self-informed posteriors, the original prior is too informative.

3.1.2 Accounting for Flatness

When setting priors in a multi-dimensional parameter space, it is common and good practice to use a flat prior for the main parameters under consideration. For example, Jørgensen & Lindegren (2005) assume a flat prior in age and metallicity when disentangling their information on these two parameters due to the assumption of no *a priori* correlation between them; their goal of determining the existence of an age-metallicity relationship should not assume one exists intrinsically, otherwise they will be proving their hypothesis with circular logic.

Additionally, on a shown quantity any third party can easily multiply on their own prior, in contrast to secondary parameters: it would vice versa be wrong to assume a flat metallicity prior when just providing an age posterior – e.g. for very high observed metallicity ([Fe/H] ~ 0.3 or similar), the slope of the metallicity distribution will mean that the true expectation for metallicity is significantly lower, thus a flat prior would bias the age estimate. Furthermore, when encountered with data points on the stellar grid that pass through multiple separate isochrones, Jørgensen & Lindegren (2005) examine the choice presented in the absence of any additional information to preferentially opt for the youngest of the isochrone selection as the most reasonable solution. They posit that for a field star on that particular point (e.g. $\log T_{\text{eff}} = 3.8$, $M_V = 3.0$) via comparison with exterior information, younger age stars will likely still reside in the slow-evolving core hydrogen burning stage; if higher ages are adopted, there would be an implied assumption that the star is more evolutionarily advanced, which is less probable for your random star. Here, this conscious choice of information could be quantified by the prior distribution rather than being applied *a posteriori* for each individual point in parameter space.

3.1.3 Reparameterisation

One must be careful when dealing with flat priors as they are not invariant under transformation, i.e. what is flat for one parameterisation may be informative in another. A uniform prior is only so in the chosen coordinate system; reparameterise and the prior can turn subtly informative which is unideal. Take the astrophysical scenario of stellar parallax estimation. A realistic uniform prior on parallax that is truncated between minimum and maximum values ϖ_{\min} and ϖ_{\max} could be considered. Equivalently, a uniform distance prior could be chosen where the bounds are coupled via the distance-parallax relationship $d_{\min} = \frac{1}{\varpi_{\max}}$ and $d_{\max} = \frac{1}{\varpi_{\min}}$. Although both seem to encode similar information, they behave very differently when expressed as a function of d , as shown in Figure 1.

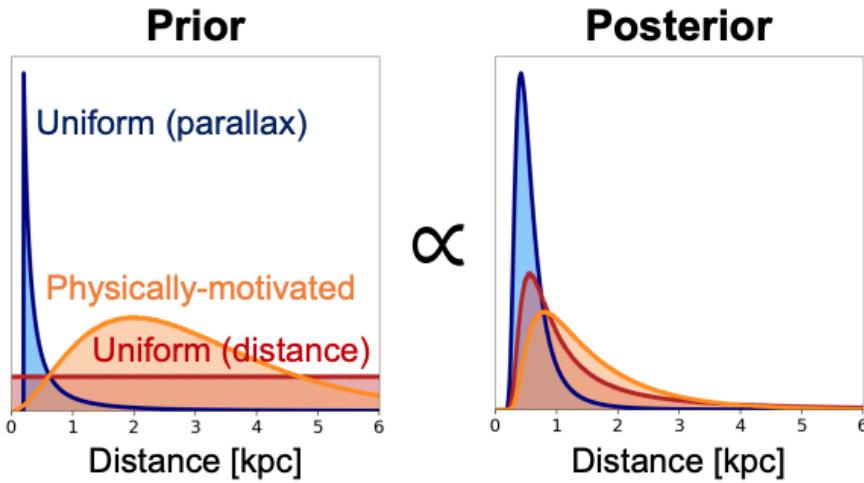


Figure 1: *Left:* Three example priors, one uniform in distance d (red), another corresponding to the reparameterisation in ϖ following the reciprocal relationship (blue), and a third more physically-motivated prior determined by [Bailer-Jones et al. \(2018\)](#) (orange), which is derived from a theoretical volume dV swept by a solid angle $d\Omega$, and implies a d^2e^{-d} relationship on the assumption of an exponential decrease in stellar density as a function of distance as one extends radially outwards in the Galactic disk. *Right:* The respective corresponding posterior PDFs given Gaussian likelihoods in d . Source of figure: [Eadie et al. \(2023\)](#).

This highlights how the choice of parameterisation influences the behaviour of non-informative parameters (strictly speaking, these are weakly informative because the physical knowledge is reflected in the chosen bounds). For an objective prior that is invariant under any parameterisation, see Jeffreys prior ([Jeffreys, 1946](#)).

3.1.4 Age and Metallicity Priors

The age prior sets the expected age distribution of the stars that are born in the data sample, this is, in other words, the Star Formation Rate (SFR). This is largely unknown and hard to capture especially for the galactic disk. A flat age prior (not necessarily flat in the intended parameter space, see below) may be a reasonable simplification. The avoidance of identifying a possible age-metallicity relation as discussed above may also be reason enough to deter us from seeking an informative prior. A decreasing prior adopted in [Schönrich & Bergemann \(2014\)](#) with upper and lower age constraints reflecting the expected age distribution in each population can be used

$$P(\tau, [\text{M/H}]) = N \cdot P([\text{M/H}]) \cdot \begin{cases} 0, & \text{if } \tau > 14 \text{ Gyr} \\ 1, & \text{if } 11 \text{ Gyr} \leq \tau \leq 14 \text{ Gyr} \\ \exp\left(\frac{\tau - 11 \text{ Gyr}}{\sigma_\tau}\right), & \text{if } \tau < 11 \text{ Gyr} \end{cases} \quad (3)$$

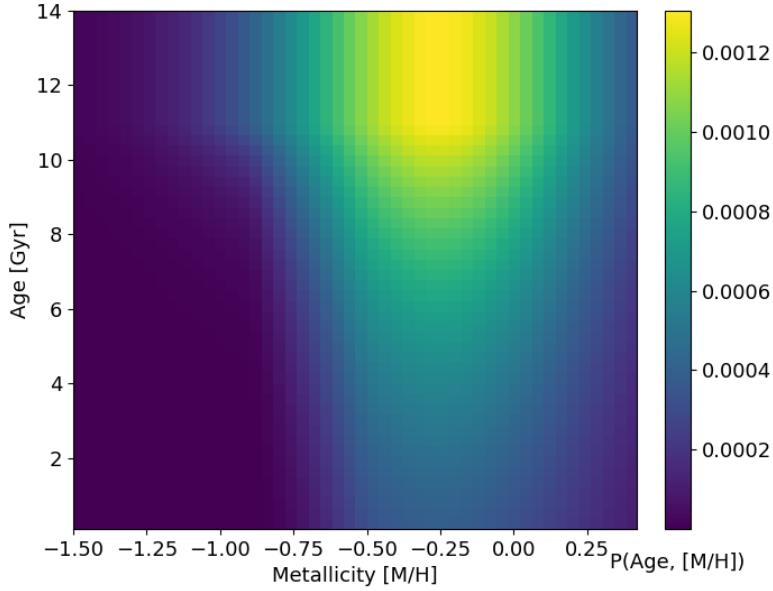


Figure 2: Combined prior for age and metallicity discussed in Section 3.1.4, binned over our chosen isochrone grid. The probability colour scale is normalised to unity.

where N represents normalisation and σ_τ encodes an age likelihood boost at intermediate ages if the star is metal-poor:

$$\sigma_\tau = \begin{cases} 1.5 \text{ Gyr}, & \text{if } [\text{M}/\text{H}] < -0.9 \\ 1.5 + 7.5 \cdot \frac{0.9 + [\text{M}/\text{H}]}{0.4} \text{ Gyr}, & \text{if } -0.9 \leq [\text{M}/\text{H}] \leq -0.5 \\ 9 \text{ Gyr}, & \text{else} \end{cases} \quad (4)$$

Here, density is constant at older ages, reflecting that older stars have had more time to populate the disk, and decays exponentially below the chosen 11Gyr. Although, as stated above, the age prior is hard to physically capture at a population-wide level, for simplicity this steepening function of age crudely reflects the significant decline in SFR with time measured for stars in the Galactic disk (Schönrich & Binney (2009); Aumer & Binney (2009)). As age decreases, the likelihood of a star existing given the initial conditions in the star-forming regions also drops off due to gentle decrease of inflow, and activity in the interstellar medium (ISM). Thus, stars with measured young ages are less likely to be observed, which should be reflected in the prior. Additional knowledge on age distributions of local stars (age of the Snhd is predicted at ~ 11 Gyr (Binney et al., 2000) is also consistent with our bias towards older stars. Previous observations also derive a coupling between older ages and weaker metallicities, which justifies this decreasing age towards weaker compositions. The adopted prior is visualised in Figure 2.

For a singular metallicity prior, the simplest option would be to consider the volume-limited sample of our data survey encoded in the likelihood function and project the measured metallicities into a 1-D PDF. Although a flat prior could avoid bias in our metallicity estimates, the metallicity distribution of nearby field stars is known to be strongly peaked at around -0.1 ± 0.3 dex; therefore, a skew in our predicted $[\text{M}/\text{H}]$ distribution is appropriate. This skew is roughly reflected in our *Gaia* data sample, which peaks at $[\text{M}/\text{H}] \sim -0.25$ dex, likely reflecting the above Snhd distribution and metallicity skewness imported by radial migration (Sellwood & Binney, 2002), and also in our combined prior,

which favours metal-poor populations. We do not assume a strict age-metallicity relation as [M/H] has remained relatively flat over the past ~ 8 Gyrs. Therefore, we can be justified in the application of the combined prior adopted in [Schönrich & Bergemann \(2014\)](#).

3.1.5 Establishing our Choice

As above, we must be confident that the true metallicities are compatible with this assumption, lest we bias the age estimate. If we are not confident in this assumption or we have adequate metallicity intelligence on the probed population, we may prefer to opt for a flat prior in order to rid ourselves of any prejudice. While our parallax cut on the data sample selects nearby stars, the prior is intended to reflect the overall Galactic population, which inherently involves a spread of older stars that have had time to distribute along the loci of the disk, as well as younger stars in lesser numbers. The age-metallicity relationship encoded within our combined prior justifies this, as older disk stars are generally metal-poor, aligning with the adopted prior.

This is evidently a physically motivated prior that reflects stellar formation history assumptions and knowledge on the age-metallicity relation (for another example of a physically motivated prior, see the distance prior in Figure 1). This can be considered weakly informative in the sense that it captures broad empirical trends at the population level, suitable for modelling age-metallicity trends without over-constraining our inference.

As discussed in Section 3.1.3, we make sure not to reparameterise from, for example, τ into $\log(\tau)$ space, as our chosen prior form is not invariant under all transformations. If we define a new variable for modelling convenience later in the methodology, $v = \log_{10}(\tau / \text{Gyr})$, then our new prior in log-age space becomes $P(v) = P(\tau(v)) \cdot \frac{d\tau}{dv} = P(\tau(v)) \cdot \ln(10) \cdot 10^v$. This transformation unintentionally emphasises younger stars unless we adjust the shape accordingly, potentially biasing the posteriors. For this reason, we remain in linear age space for the remainder of this work.

3.2. Isochrones

3.2.1 What are Isochrones?

To cover the requisite areas of parameter space and to apply appropriate physical constraints on the stellar catalogue, dense grids of theoretical isochrones - paths traced on the CMD representing stellar populations with equivalent ages and metallicities - can be used to portray stellar models and obtain corresponding PDFs. See Fig 3 for a visual representation of an isochrone grid. They effectively describe initial stellar conditions and can be used to determine stellar age constraints on large scales. Early pioneers, such as [Sandage \(1953\)](#) and [Eggen et al. \(1962\)](#), enabled by contemporary advances in theories of stellar structure and nuclear astrophysics, applied the models for the estimation of globular cluster ages - particularly exploiting the position of the MS turn-off. Subsequent developments of more sophisticated theoretical isochrones ([Iben & Renzini \(1984\)](#), and later [Bertelli et al. \(1994\)](#)) were essential in transforming CMD morphology into individual age estimates. Isochrone fitting then became the go-to method of determining the oldest populations, providing limits on the age of the Universe itself; for example, [Chaboyer et al. \(1996\)](#) demonstrated that the oldest globular clusters agreed with ages of $\sim 12\text{-}14$ Gyr compared to cosmological constraints, and later the CMB.

Sufficiently dense grids allow for maximal coverage for various combinations of stellar parameters, e.g. our chosen parameter space of $\log T_{\text{eff}}$, $\log(L/L_{\odot})$, $\log g$, colour, and absolute magnitude. The isochrone library comparison method involves PDF derivations provided by grids of desired stellar parameters for a given star ([Casagrande et al. \(2011\)](#); [Schönrich & Bergemann \(2014\)](#)), calculated by summing over these theoretical isochrones.

For our determination of ages and metallicities we must compare observational data with their

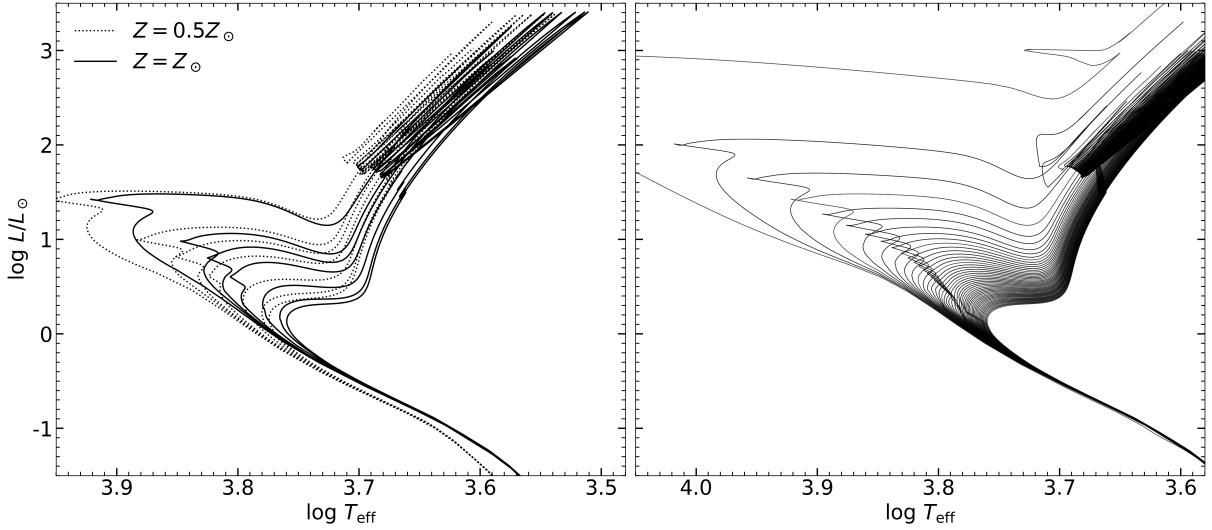


Figure 3: Physical colour-magnitude or Hertzsprung-Russell Diagram, shown by BaSTI isochrones from Pietrinferni et al. (2004); Pietrinferni et al. (2006, 2009) for *left*: two metallicities Z_\odot and $0.5Z_\odot$ at 1, 2, 3, 4.49 (approximate solar age), 8, 12, and 14 Gyr, and for *right*: all ages from 0.1 to 14 Gyr in increments of 0.1 Gyr at constant solar metallicity. These diagrams act as a reference point on the methods and issues of how we derive physical parameters (age, abundance) for stars based on observables: namely the age and abundance sensitivity of L and $\log T_{\text{eff}}$ near the turn-off and along the subgiant branch making for robust parameter estimates, whereas isochrone clumping on the MS leads to age independence of the luminosity and hence uncertain parameter estimates, etc.

corresponding theoretically computed values as denoted by the isochrones. From hereafter, the observed data \mathbf{Y} for a given star will be represented by the vector $\mathbf{q}(\mathbf{p})$ with accompanying uncertainties. Since we are utilising isochrone fitting, convenient quantities will include effective temperature, luminosity, colour and magnitude because of the association with the axes of the HRD; isochrones trace stellar evolution over differing initial mass visualised as a function of these variables. We will now also refer to the given model parameters \mathbf{X} by the parameter vector \mathbf{p} , in our case $\mathbf{p} = (\tau, [\text{M}/\text{H}])$ - the variables we are ultimately trying to map over all boundaries of \mathbf{q} . The stellar evolutionary models (theoretical isochrones) provide a mapping from the parameter space to the data space. For each point on the HRD (for example at a given T_{eff} and luminosity) the isochrone grid will provide estimates on age and metallicity given their local proximity. We must note that this is possible only if $\dim(\mathbf{q}) \geq \dim(\mathbf{p})$ - we need at least as many observable parameters as unknown parameters - and the mapping is non-degenerate - meaning that every point in the data space corresponds to only one point in the parameter space, i.e. it is informative. This is important as different model parameter combinations will otherwise yield the same observables, and inferred results will not be unique. Again, we can refer to the age-metallicity degeneracy as a well-known case of this problem - we want to avoid scenarios where different stellar ages or compositions procure data that is indistinguishable.

Through this baseline we can thus simplify our understanding of the problem of age and metallicity determination as inverting the function $\mathbf{q}(\mathbf{p})$ by probabilistic Bayesian methods given the appropriate forms of priors and observational likelihoods.

3.2.2 Choice of Grid

For our initial model where we test preliminary PDFs we choose a small subset of BaSTI (a Bag of Stellar Tracks and Isochrones, [Pietrinferni et al. \(2004\)](#); [Pietrinferni et al. \(2006, 2009\)](#)) isochrones which only cover 11 selected values in $[M/H]$ between $-2.267 < [M/H] < 0.395$, but with a denser resolution in age which spans from $0.1 < \tau < 14$ Gyr. For our purposes, where we want to balance computational efficiency by employing a sparser isochrone grid with reasonable output resolution for our PDF model tests, this is an acceptable balance. However, for our full posterior calculations it is important that our binned age/metallicity likelihood grid is sufficiently filled. A sparse metallicity grid can introduce artificial gaps in the structure of the PDFs, especially for stars that populate near bin edges. This will result in stars with $[M/H]$ values roughly equidistant between sparse nearest isochrones to be pushed towards the nearest available model; this will potentially bias the age-metallicity relation and skew inferred ages, particularly in the metal-poor tail or for those near solar composition. To add to this, isochrone morphology does not scale linearly with metallicity (see Figure 3) so without sufficient $[M/H]$ resolution, age constraints from other parameters may become inaccurate. Extensive early testing of the effects of such gaps on metallicity posterior shaping suggested the need for a denser grid. Hence, we consolidate the data structure with additional BaSTI isochrones which supplement the gaps in metallicity. We choose isochrones with corresponding *Gaia*-band colours and magnitudes that span 48 values of metallicity from $-1.5 < [M/H] < 0.42$ and a similar age range of $0.1 < \tau < 14.0$.

It is possible to refine the resolution of the isogrid through interpolation to ensure proper parameter spacing, used for example by [Jørgensen & Lindegren \(2005\)](#), where a continuous interpolation parameter u is introduced along each isochrone to calculate the general parameter function. Characterised by physically significant sections of the isochrone, u encodes equal evolutionary intervals and is taken to be proportional to the curve length along the track using arbitrary line elements. The linear interpolation ensures that the general morphology of the isochrones is preserved to disentangle separate tracks in parameter space from intersecting where they should not.

Such techniques offer a powerful solution to refining the sparseness of grids while preserving the natural structure across important evolutionary phases (core H exhaustion, Red Giant Branch (RGB) turn-off), but they also introduce complexity when considering large n-dimensional grids spanning age and metallicity. In our current study, such interpolation should not be needed to produce smooth posterior surfaces. Nevertheless, it is important to highlight how carefully grid sampling should be defined across tracks, and this could be used *a posteriori* in future works requiring finer resolution between rapid evolutionary transitions, i.e. between isochrones.

3.3. Stellar Model PDFs

To supplement the use of our priors in the Bayesian scheme, pre-computed stellar models that describe the evolution of various stellar properties can be utilised in tandem. These models force constraints on the full-dimensional parameter space by incorporating the physics of stellar evolution, essentially behaving as a quasi-prior that combines with the actual prior to shape the posterior. The PDF is calculated using a dense grid of isochrones from the BaSTI database. Population parameters can be well inferred with a sufficiently dense grid that spans the maximal coverage of stellar parameters ($\log T_{\text{eff}}$, $\log(L/L_{\odot})$, $\log g$) in comparison to utilising other models such as stellar evolutionary tracks, which require additional assumptions and detailed physical evolutionary constraints. Additionally, we use isochrones, as they are generally less computationally intensive than stellar tracks, and particularly Bayesian methods, which deal with complex posterior inference and high parameter space dimensionality. The PDF calculation is performed by summing over the isochrones across the grid, first assigning a weight at each stellar point proportional to the local parameter space volume it

represents

$$W_i = N_W \Delta[\text{M}/\text{H}] \cdot \Delta\tau \cdot \Delta M_{\text{init}} \quad (5)$$

where ΔM_{init} is the average distance in initial mass to its neighbouring points, given by $\Delta M_{\text{init}} = 0.5(M_{\text{init}, i+1} - M_{\text{init}, i-1})$, $\Delta[\text{M}/\text{H}]$ is the average distance in metallicity space an isochrone assigns to that stellar point compared with neighbouring isochrones inhabiting parameter space on the HRD, and $\Delta\tau$ is the average age difference calculated the same way.

These comparisons are calculated by parsing through every isochrone and minimising the line element $ds^2 = d\log T_{\text{eff}}^2 + d\log(L/L_{\odot})^2$ - the same axes used to encode an isochrone's trajectory on a traditional HRD. N_W is a normalisation constant that ensures that all calculated probabilities across the desired interval sum to unity. The model PDF is then updated by multiplying each weighted point with the associated uncertainty, here represented by an n-dimensional Gaussian kernel g with auxiliary dispersion vector σ spanning all relevant space - no additional uncertainties in other unwanted dimensions - and summing over the isochrone:

$$P_{\text{mod}}(\mathbf{X}) = \sum_i W_i g((\mathbf{X} - \mathbf{X}_i), \sigma) \quad (6)$$

where \mathbf{X}_i is the parameter space vector ascertained from the model grid and \mathbf{X} represents the parameter space itself (Schönrich & Bergemann, 2014). For our initial models with sparse ranges of metallicity in terms of the isochrone grid, age and metallicity are binned over ranges given by the first set of BaSTI isochrones, and a joint 2-dimensional PDF is calculated for each bin using the above method, the discrepancies in how the Gaussian kernel distributes the residual values of $\log T_{\text{eff}}$ and $\log(L/L_{\odot})$ differentiated from the isochrone grid serving as the main contributor.

3.4. Constraining Metallicity

3.4.1 Stellar Theory and Observational Constraints

So far, the isochrone grid stellar PDF has singularly implemented the constraining parameters of luminosity and effective temperature to place the predicted star in age and metallicity probability space. This choice of parameter space for placing the interested stars upon the HRD facilitates known stellar evolution mechanisms and patterns which follow well-defined trajectories, helping to predict behaviour in the distributions. For example, bright hot stars ($\log T_{\text{eff}} > 3.8$, $\log(L/L_{\odot}) > \sim 1.3$) halfway through the MS are typically young, massive ($> 2-3 M_{\odot}$), and rapidly evolving. Their positions above the Zero-Age Main Sequence (ZAMS) indicate the ignition of their hydrogen fuel reserve, a direct constraint on the speed of evolution and subsequent age at the MS turn-off. Thus, by leveraging the coupling between the chosen parameters and the stellar evolutionary state, we can provide a good view on minimum age as the bright-blue end of the MS is populated by the youngest populations. Conversely, dim-red stars are usually low-mass and evolve slowly, implying that their continued presence on the MS may indicate an older population or sustained star formation. Effective temperature helps break the luminosity degeneracy for stars by decoupling their metallicities. Metal-poor stars will appear bluer as low opacity lets energy escape, and metal-rich stars redder as high opacity impacts how the energy traverses the extended convective envelope. This is observed in the red clump and globular clusters with comparatively blue metal-poor stars.

Although these parameters are effective at predicting our desired probability distributions, we can try disentangling any age-[M/H] degeneracies (old metal-poor populations overlapping with young metal-rich ones) by introducing additional observables like surface gravity or detailed spectra. In this case, we look at absolute V-band magnitude and $B - V$ band Johnson colour indices. Although these new observables are not entirely independent of L and T_{eff} (they project intrinsic stellar measurements

into a new observation space), they do add extra information and allow us to constrain the stellar catalogue more tightly with higher dimensions. Magnitudes encode parallax/distance and luminosity, and colours depend primarily on metallicity and temperature. This is recorded using Equation 6 in which we provide probabilities by summing the Gaussian over every point on the isochrone, but now we compare residuals $\mathbf{X} - \mathbf{X}_i$ of provided $B - V$ and M_V values in addition to the aforementioned $\log(L/L_\odot)$ and $\log T_{\text{eff}}$ ones. Facilitating extra constraints clips out areas of parameter space that potentially contribute to biasing the PDF into skewed age and [M/H] distributions that would otherwise have gone unimpeded under only the influence of luminosity and effective temperature.

Additionally, our choice of colour indices will influence our ability to disentangle temperature-magnitude degeneracies. Bands blueward of the Balmer jump (Johnson U -bands, UV passbands) are sensitive to temperature due to flux drops at near-UV for cooler stars (Howes, Louise M. et al., 2019). Supplementing our $B - V$ data with these colours can enhance T_{eff} estimates and separate overlapping hot, metal-poor and cool, metal-rich sources. On a similar page, narrow-band systems such as Strömgren $uvby\beta$ photometry provide filters specially made to isolate metallicity-/temperature-sensitive line features - the m_1 index specifically probing metal-line blanketing, pushing the star redder as metallicity increases (Calamida et al., 2007). Hence, it would be ideal to constrain our inference with accessory temperature-sensitive colour bands alongside $B - V$ by tightening the temperature aspect that manifests in the posterior.

3.4.2 Application and Example

This can be seen in, for example, a star with $\log(L/L_\odot) = 0.40$, $\log T_{\text{eff}} = 3.75$. (Figure 4, bottom left). Applying these two limits in our algorithm predicts a long metallicity tail below $[M/H] = \sim -0.5$ at high ages, and a subtle age gradient towards ~ 2 Gyr, a large spread from the ideal concentration. However, forcing on our additional constraints by pinpointing the star's corresponding colour and magnitude, accounting for bolometric correction and empirical calibrations at the exact temperatures, we see different results (Figure 4, bottom right). The tail towards low metallicity is almost completely cut off with the distribution still favouring its place in probability space of $[\text{age}, M/H] \sim [8\text{Gyr}, 0.3\text{dex}]$. Colour is sensitive to metallicity as opacity is directly correlated with metal abundance. Adding colour specifiers results in higher resolution distinction between multiple stars of slightly different [M/H] that would have otherwise overlapped on the HRD. With colour data, only a narrow range of metallicities are consistent with said stars.

We observe that age constraints are also improved, although not to the same extent. Metallicity is a secondary parameter to age in terms of determining stellar position on the HRD; altering [M/H] shifts the locus of stellar evolution tracks more in colour/magnitude space than in luminosity/temperature space. Throughout most of the HRD, especially for low-mass stars, stellar colour does not trace age fluctuations as strongly due to the MS turn-off and subsequent giant branch evolution already being captured well by L and T_{eff} in combination. MS stars of different ages but equal mass will have similar colours, and although giants and turn-off stars will better constrain age, precise photometry is also needed. Age effects are more subtle on the lower MS where all generations of stars sit very closely in colour/magnitude space, and location becomes a pure metallicity-mass indicator due to stars' distance from the turn-off. Similar sharpened metallicity distributions are observed for another example, $\log T_{\text{eff}} = 3.72$, $\log(L/L_\odot) = 0.8$, (Figure 4, top) in which [M/H] is almost identically quenched below $\sim -0.3\text{dex}$ with age being less affected as it is harder to tighten unless we have very precise turn-off data. Overall, it is clear that adding classical parameters that map the CMD onto already established HRD signifiers helps to break metallicity degeneracy for specific stellar populations. These tighter metallicity constraints but shallower age estimates are what should be expected with the choice of $[M_V, B - V]$ parameter space.

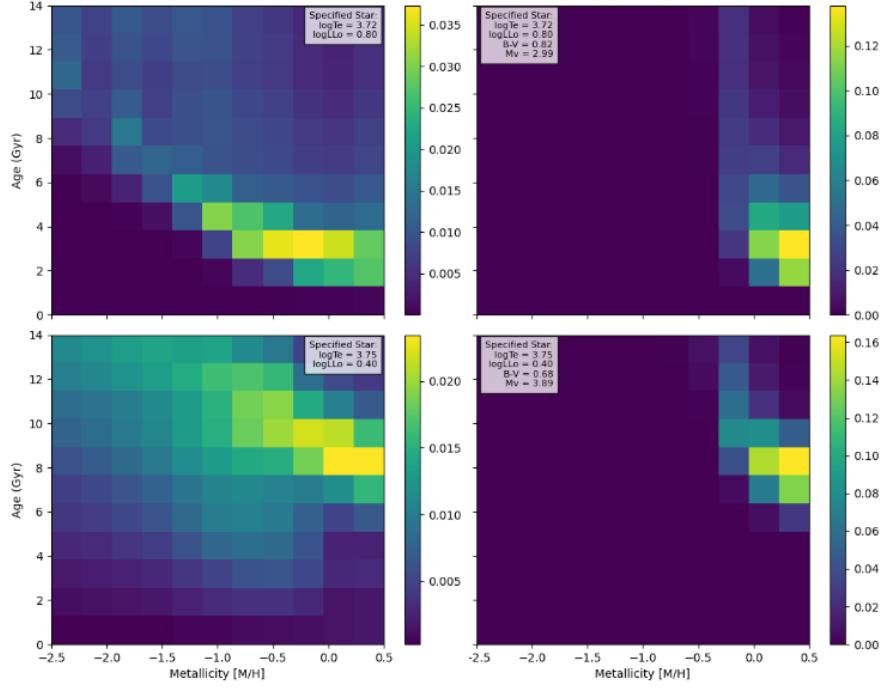


Figure 4: Stellar model PDFs $P_{mod}(\mathbf{X})$ in age and metallicity for two test stars computed from our sparse initial BaSTI isochrone grid, with model constraints of luminosity and effective temperature (*left*), and additional constraints of magnitude and colour (*right*). The addition of photometric constraints demonstrates the improved discriminating power of combined spectro-photometric isochrone based inference on stellar parameters.

We note above that the newly introduced variables are not entirely independent of the originals. Therefore, although the metallicity and age constraints are improved, slight correlations in resulting posterior parameter spaces may be considered throughout this work due to projection overlaps; changes in marginal distributions could potentially propagate through to the posterior. If we want to build star formation histories or chemical evolution models using our iso-grid priors and subsequent Bayesian scheme, a more consolidated metallicity distribution will offer much higher clarity inference into chemical enrichment, star formation rates, etc. Therefore, this choice to increase the dimensionality of parameter space to select stars of interest can be justified in pursuit of constraining output predictive and eventual posterior distributions.

3.5. Observational Likelihoods

3.5.1 Context within Bayes’ Theorem

We want to express the probability of the observed data in the form of a distribution. We can recall the observational likelihood $P(\mathbf{Y} | \mathbf{X})$, which represents the probability of observing data \mathbf{Y} given stellar model parameters \mathbf{X} . As before, we will denote the observational data points as the vector \mathbf{q} and the parameter vector that uniquely label the isochrone models, and what we are trying to infer, as \mathbf{p} .

This observational likelihood can be referred to as the probability of observing a particular dataset given a specific accompanying statistical model, in this case the BaSTI isochrones which encode how stars of given age and metallicity behave, and its parameters. The likelihood can be framed by asking the question: assuming one knows the parameters exactly, what is the distribution of the data? For a given data/parameter pair, it registers how “likely” the data is. It must not be misconstrued that the

likelihood is a function of the data, or is a PDF rather than just being proportional to a probability.

3.5.2 Application to the *Gaia* Survey

We choose the Global Astrometric Interferometer for Astrophysics (*Gaia*) Data Release 3 (DR3) as our dataset due to its precision and incorporation of astrometric, photometric and spectroscopic observable coverage across the entire sky. The catalogue includes extensive parallax inventories, its own broad optical band for magnitudes and colours, and selected astrophysical parameters for a subset of stars. We can test and calibrate our Bayesian algorithm by first probing the local Snhd within the thin-disk. A sample of 168695 stars corresponding to a parallax cut of 10mas (or an upper distance limit of 100pc) is selected, indexing a grand mix of main sequence, sub-giant, and a small amount of white dwarf stars, but omitting more distant stars that reside in the thick disk, galactic centre, or greater halo. A wide range of highly resolved and complete stellar populations is ideal for the facilitation of detailed comparisons with isochrone models which fully cover parameter space. The nearby nature of the sample minimises extinction variables and emphasises completeness of volume - it also allows for the employ of a more well-understood selection function if we should so choose. The convenient parameter space we use for this calculation given the coupling between *Gaia* DR3 and the BaSTI isochrones is taken as $[\log T_{\text{eff}}, \log(L/L_{\odot}), V - I, M_V, \log g]$; therefore, we can write the likelihood $P(\mathbf{Y}|\mathbf{X})$ as $\mathcal{L}(\mathbf{q} | \mathbf{p}) = \mathcal{L}(\log T_{\text{eff}}, \log(L/L_{\odot}), V - I, M_V, \log g | \tau, [\text{M}/\text{H}])$.

To obtain the observational likelihood \mathcal{L} for each star in the sample we assume Gaussian errors and evaluate the goodness-of-fit statistic χ^2 at every isochrone point k up to a total n for every parameter q_i where $i = \dim(\mathbf{q}) = N$

$$\mathcal{L}_j = \prod_{i=1}^{N=5} \frac{1}{(2\pi)^{N/2}\sigma_i} \exp\left(-\frac{\chi^2}{2}\right) \quad (7)$$

where

$$\chi^2 = \sum_{k=1}^n \left(\frac{q_i^{\text{obs}} - q_{k,i}(\tau, [\text{M}/\text{H}])}{\sigma_i} \right)^2 \quad (8)$$

and sum over j isochrones to get an overall total likelihood \mathcal{L} . q_i^{obs} is the observed data value for our star in our chosen parameter space given by *Gaia* DR3, $q_{k,i}$ is the corresponding isochrone data point given the constraints in parameter space, and σ_i is the weighted observational uncertainty on our parameters. A baseline σ value of 0.1 is added in quadrature as an error floor to compensate for the systematic underestimation of uncertainty and to avoid overconfidence in sharply peaked likelihood PDFs that may be observed as artificially narrow (Lindegren et al., 2018). The global parallax 0-point offset is not fully captured in the parallax error from DR3; therefore, such small uncertainties that are expected for close-distance stars ($\varpi > 10\text{mas}$) need to be accounted for. We should be safe however, because at this limit the systematic errors in parallax are a factor ~ 200 smaller or so.

Least-Squares (LS), or similarly χ^2 statistics, are commonly used procedures in data analysis and can be justified through the Bayesian approach (Sivia & Skilling, 2006). In the context where the observer is largely ignorant of the stellar mechanism background, the prior PDF is taken to be uniform; thus the resulting posterior PDF transforms to coincide with the likelihood function. Hence the ML estimate value \mathbf{p}_0 of the parameter set \mathbf{p} can be inferred from the maximum value of the posterior distribution curve. The solution invariance under monotonic transformations of PDFs or other functions, e.g. the luminosity function $\phi(L) \rightarrow \phi(\log L)$, is one of the many positive features of ML (Bailer-Jones, 2017). In the extensional case where the observer assumes the individual data are independent, further approximations can be made relating to the likelihood function; the overall joint PDF is calculated by the product of probabilities, or the likelihood function, of the individual measurements.

The maximum observed likelihood will occur when χ^2 is minimised and the corresponding \mathbf{p}_0 is known as the LS estimate. χ^2 statistics are only applicable for Normal or Gaussian distributions where sufficiently large data samples are being analysed.

3.5.3 Parameter Conversion

Before evaluating the test statistics, i.e. the likelihoods for each individual star, parametric conversions must be made in order for the comparison to keep its spatial consistency. $\log g$ is not explicitly referred to over the isochrones so must be estimated from corresponding mass M , radius R , luminosity L , and T_{eff} values. We can take our surface gravity $g = \frac{GM}{R^2}$ and eliminate R by rearranging the Stefan-Boltzmann law in terms of radius, $R^2 = \frac{L}{4\pi\sigma T_{\text{eff}}^4}$, and then form a ratio so the stellar constants cancel and the 4π is absorbed. Then, it is as simple as taking base-10 logarithms for convenience:

$$\log \frac{g}{g_{\odot}} = \log \frac{M}{M_{\odot}} + 4 \log \frac{T_{\text{eff}}}{T_{\odot}} - \log \frac{L}{L_{\odot}} \quad (9)$$

where $\log g_{\odot} = 4.438$ is the solar gravitational acceleration, $\log T_{\odot} = \log_{10}(5777) = 3.76$ is the solar temperature, and $\log M_{\odot}$ and $\log L_{\odot}$ are the solar mass and luminosity, respectively. *Gaia* photometric colour bands for $G_{\text{bp}} - G_{\text{rp}}$ must be redefined using standard Johnson-Cousins colour bands employed by BaSTI, $V - I$:

$$G_{\text{bp}} - G_{\text{rp}} = -0.0162 + 1.274(V - I) - 0.08143(V - I)^2 \quad (10)$$

(European Space Agency (2024), Table 5.6 of specification section 5.5.1: relationships with other photometric systems). Similarly, these redefined colour bands are used in converting the *Gaia* specific G-band photometric absolute magnitudes into the more universal visible M_V using a second-order polynomial fit:

$$M_V = M_G - 0.018 - 0.007(G_{\text{bp}} - G_{\text{rp}}) - 0.173(G_{\text{bp}} - G_{\text{rp}})^2 \quad (11)$$

(Table 5.8 of specification section 5.3.7). The absolute G-band magnitudes are also first calculated from the given DR3 photometric mean apparent G-band magnitude and astrometric parallax ϖ :

$$M_G = G - 5 \log_{10} \left(\frac{1000}{\varpi} \right) + 5 \quad (12)$$

with distance being simply inversely proportional to parallax and with no extinction correction included. Once these individual likelihoods/probabilities have been assigned to each star in the population-constrained dataset, the values are binned in age and metallicity, ensuring the same resolution as seen in the stellar model and prior PDFs. This results in a population-level likelihood surface used for inference over the SFR and chemical enrichment patterns, and where they are most consistent with our isochrone models.

3.6. Marginalisation

3.6.1 Calculating Projections

Instead of studying how a model should be interpreted based on the introduction of new data and prior knowledge, if we want to isolate a derivation of the probabilities of new parameters given

existing probabilities, we can introduce Bayesian marginalisation:

$$P(X_1, \dots, X_j, X_{j+2}, \dots, X_n \mid \mathbf{Y}) = \int P(X_1, \dots, X_n \mid \mathbf{Y}) dx_{j+1} \quad (13)$$

which excludes the "nuisance parameter" x_{j+1} . This is a powerful device for integrating out unwanted "nuisance parameters", such as background noise signals, quantities which systematically enter the analysis space but are of no interest. Here, for our derived age-metallicity posteriors, we can easily numerically integrate out either parameter to draw 1-D PDFs when desired. This flattening projection will offer inference on age regardless of metallicity, and vice versa.

We use a straightforward rectangle rule utilising weighted summation of the infinitesimals calculated by the NumPy gradient function (this also allows for non-uniformly spaced grids if we should so choose) to perform the integration. In other words, the integration process is discretised as

$$\int P(a, Z) da \approx \sum P(a_i, Z_j) \Delta a_i \quad (14)$$

where i and j indicate our bin indices, and a and Z are stand-ins for age and metallicity. For cases like ours where the posterior is defined on a structured grid with low dimensionality, this rectangle method will give sufficient accuracy. Proper normalisation is also ensured so that the area under the 1-D PDFs evaluates to unity. For other applications with larger core parameter spaces, higher-order integration methods like Simpson's rule may be used if the trade-off between higher-grain resolution and computational efficiency can be consolidated.

3.6.2 Moments and Uncertainties

After we have obtained conditional PDFs in lower dimensions by marginalising, we can obtain moments of the distribution in each of our variables. Taking the first moment as in

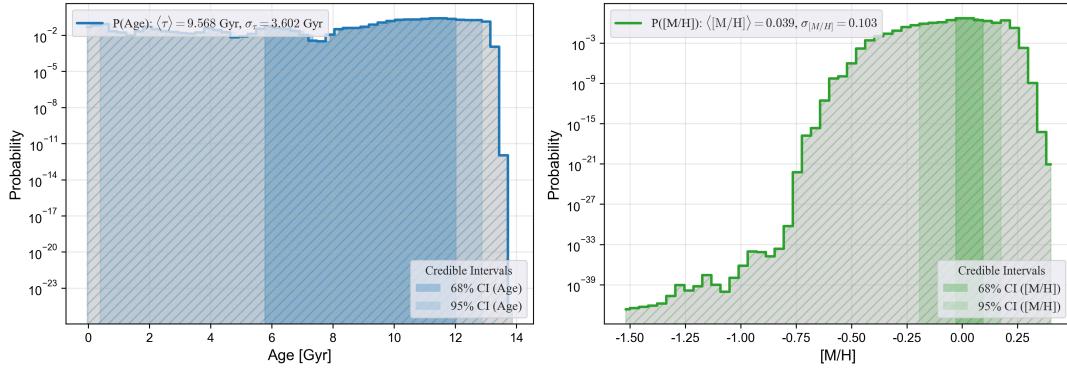
$$\langle X_j \rangle = \int x_j P(X_j \mid \mathbf{Y}) dx_j \quad (15)$$

gives us the expectation value of the parameter X_j ([Schönrich & Bergemann, 2014](#)), which generalises the weighted average of said parameter. Since we are working with a continuous source of possible outcomes, it is expected that this is calculated using integration, but like our integration process over a binned parameter range, we calculate it by summing over the bins multiplied by the PDF value and step size to achieve a discretised analogue. As this value is derived from the full Bayesian posterior of a single star which is commonly singularly peaked and mono-modal, it is valuable in inferring a summary estimate of the full shape of the distribution. However, if the posterior is highly skewed, asymmetric, or multi-modal, we may prefer to quote the mode or other confidence intervals, as the mean may lie in a low probability region. As an extension, we can calculate the second moment of the posterior PDF with

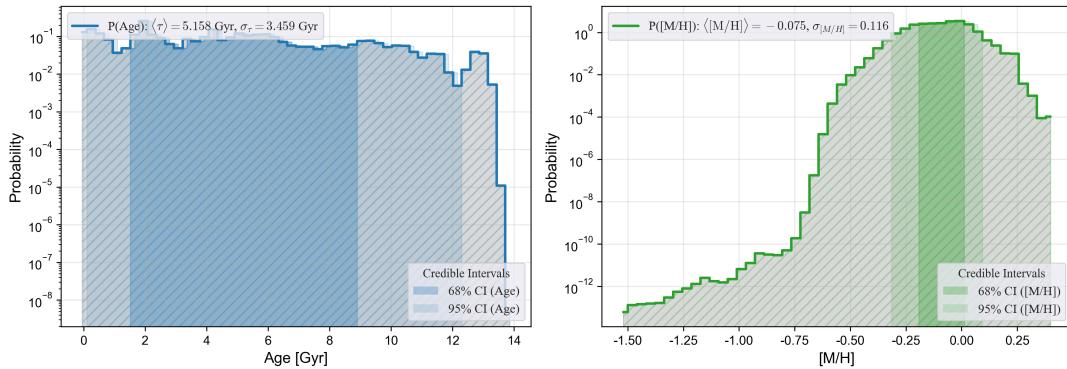
$$\langle X_j^2 \rangle = \int x_j^2 P(X_j \mid \mathbf{Y}) dx_j \quad (16)$$

which then offers us the standard deviation through $\sigma_{X_j} = \sqrt{\langle X_j^2 \rangle - \langle X_j \rangle^2}$. This is computed as described above. Rather than pinning an estimate on the central or most likely value of our variable, the standard deviation offers insight into the quantitative measure of spread across the posterior. Hence, it helps us measure the extent to which our model constrains certain measurements and helps encode model degeneracies such as age-metallicity.

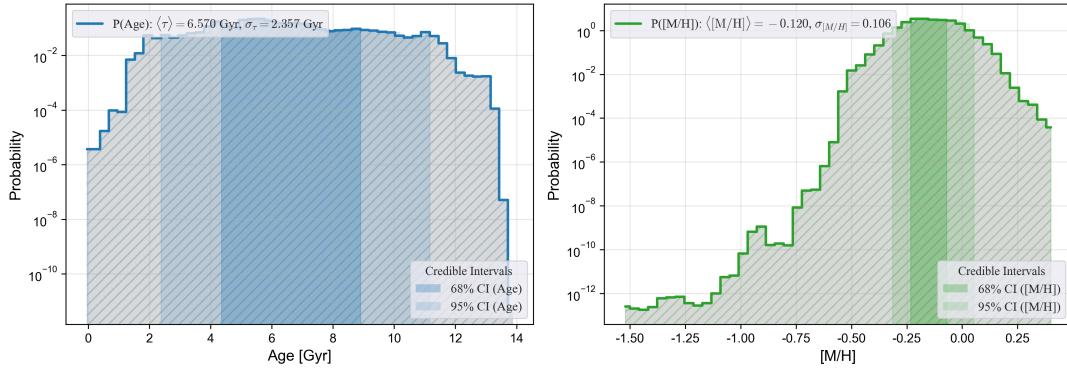
Marginalised Age and [M/H] PDFs for a star with $\log T_{\text{eff}} = 3.66$, $\log L/L_{\odot} = -0.69$, $G_{\text{BP}} - G_{\text{RP}} = 1.31$, $M_G = 10.48$



Marginalised Age and [M/H] PDFs for a star with $T_{\text{eff}} = 3.74$, $\log L/L_{\odot} = -0.13$, $G_{\text{BP}} - G_{\text{RP}} = 0.93$, $M_G = 9.55$



Marginalised Age and [M/H] PDFs for a star with $\log T_{\text{eff}} = 3.77$, $\log L/L_{\odot} = 0.13$, $G_{\text{BP}} - G_{\text{RP}} = 0.74$, $M_G = 8.98$



Marginalised Age and [M/H] PDFs for a star with $\log T_{\text{eff}} = 3.81$, $\log L/L_{\odot} = 0.64$, $G_{\text{BP}} - G_{\text{RP}} = 0.58$, $M_G = 8.06$

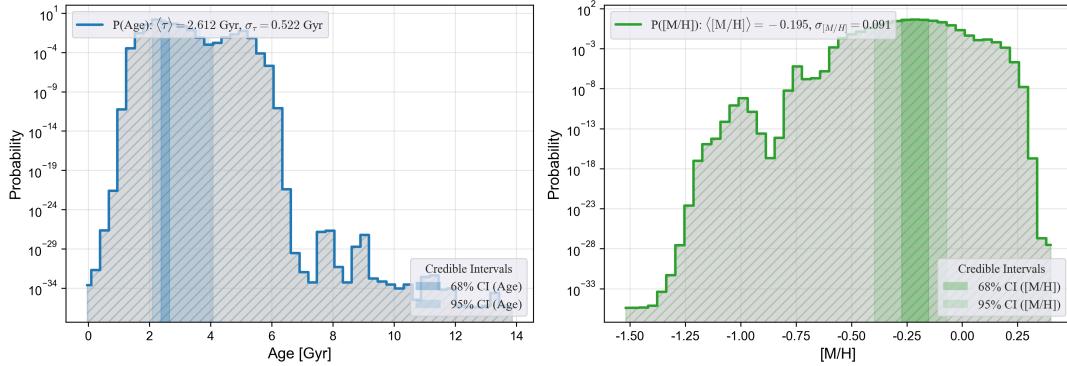


Figure 5: Marginalised posteriors for a selection of four test stars fully explored in Section 3.7. Included are the expectation values, standard deviations, and both 68% and 95% confidence intervals for both age and metallicity parameters.

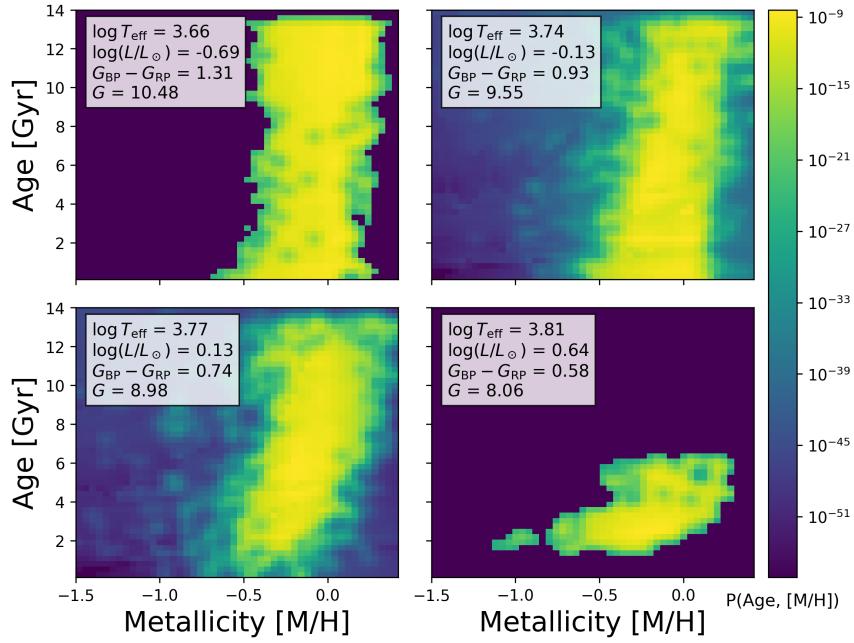


Figure 6: Age and metallicity joint posteriors for a selection of four test stars with constraints on our given parameter set of $[\log T_{\text{eff}}, \log(L/L_{\odot}), G_{\text{bp}} - G_{\text{rp}}, M_{\text{G}}]$. Our particular set of stars are depicted by red squares on the Hertzsprung-Russell Diagram populated by our *Gaia* Survey with predicted Bayesian ages and compositions (Figure 8). Note that here and for Figure 5, the PDFs are in log-scale so the observed spread in age and $[\text{M}/\text{H}]$ is not as severe as one would expect on first glance.

3.6.3 Confidence Intervals

Furthermore, after using marginalisation to project our parameter PDFs into 1-D to acquire single estimates, we can also provide a confidence interval (CI) at specified levels (68% proportional to $\pm 1\sigma$ or 95% ($\pm 2\sigma$) is standard). Although more at home in the frequentist realm, one can use CIs to estimate ranges in which population parameters are likely to fall. One interpretation of this is the reliability in which a parameter \mathbf{X} can be inferred based on the posterior curve area between two points, say X_1 and X_2 , equal to a percentage of the curve that is covered. Since this area is proportional to the extent in the belief that \mathbf{X} falls within the given range, this can be considered a sensible uncertainty estimation protocol. Assuming the PDF is normalised to have unit area, the $\pm 2\sigma$ confidence interval is defined as follows to determine X_1 and X_2 :

$$\int_{X_1}^{X_2} P(\mathbf{X}|\mathbf{Y}) d\mathbf{X} \approx 0.95. \quad (17)$$

Another possible interpretation: if one were to sample the posterior and construct intervals for each, one would expect a certain percentage to contain the 'true' population parameter. Marginalised PDFs in age and metallicity are shown for our selection of test stars in Figure 5 derived from posteriors in Figure 6, with confidence intervals, expectation values, and standard deviations depicted.

3.7. Analysing Posteriors

3.7.1 Overview of Posterior Construction

We now have the necessary tools to build posterior joint PDFs in age and metallicity with the final equation: $P(\mathbf{X} | \mathbf{Y}) = P(\mathbf{Y} | \mathbf{X}) \cdot P_{\text{mod}}(\mathbf{X}) \cdot P(\mathbf{X})$. This final distribution encodes how likely you are to find a specific star in a specific stellar population in age and metallicity probability space. The posterior considers evaluations from:

- i) the likelihoods $P(\mathbf{Y} | \mathbf{X})$ at each grid point comparing observed parameters to isochrone predictions
- ii) prior distributions in age and metallicity for considered populations of our *Gaia* survey $P(\mathbf{X})$
- iii) the isochrone-based stellar models $P_{\text{mod}}(\mathbf{X})$ that effectively describes a 4-dimensional constraint on our parameter space based on the same observed parameters used in i).

3.7.2 Case Studies

Here we present test stars from our parallax-limited *Gaia* sample for four different populations in the HRD, and compare the resulting joint age-[M/H] predictive posterior PDFs (and their respective marginalised 1-D posteriors) with known physics from MS and post-MS stellar evolution. The posteriors are shown in Figure 6 with their marginalised counterparts in Figure 5.

1. For our core parameter space of $[\log T_{\text{eff}}, \log(L/L_{\odot}), G_{\text{bp}} - G_{\text{rp}}, M_{\text{G}}]$ that will be used henceforth while looking at our posteriors, we select a test star at $[3.66, -0.69, 1.31, 10.48]$. We test a very faint, cool and red star placed somewhere on the mid MS. We can infer a less-than solar mass star actively burning core hydrogen around the K- or early M-dwarf regime with inert time-varying evolutionary properties. At this point on the HRD, most isochrones are tightly packed regardless of age or [M/H]; MS star positions do not change dramatically until later stages. The slowly evolving nature of such stars may inform the age uncertainty, which is manifested in the roughly equal spread of age probability, as their positions change only subtly over the order of Gyrs. Hence there is a wide range of ages consistent with this point in parameter space, the observed significant spread is expected, and thus the posterior is uninformative in terms of age. Although age predictions are roughly flat, the model predicts a peak at ~ 9.5 Gyr with an uncertainty of ~ 3.6 Gyr, a reflection of the limited age sensitivity of MS stars in this area. On the other hand, metallicity is more constrained by a star's temperature, luminosity and colour parameters (as discussed in Section 3.4); therefore, the posterior picks up metal-poor stars' tendency to appear bluer than observed, or metal-rich stars' inclination to appear redder or fainter than our test star. Since our source is moderately red and faint, the metallicity posterior is better constrained, finding a likely [M/H] value of $\sim 0.04 \pm 0.1$. This prediction balances between metal-poor isochrones which would colour the star too blue, and metal-rich isochrones which would make it too red for its given position on the HRD.
2. For a star at $[3.74, -0.13, 0.93, 9.55]$ we observe a yellowish colour (mid to late G-type) with quite a faint absolute G magnitude. The parameters likely place us as an intermediate-mass star actively burning core hydrogen, located on the upper MS - before the turn-off into the sub-giant/giant phases. Again, in this $T_{\text{eff}}-L$ range, stars evolve very slowly, taking up to ~ 15 Gyr for a $0.8-0.9 M_{\odot}$ star to leave the main sequence (Harwit, 1988), and changes in observables remain subtle. Although the age uncertainty is large here (~ 3.4 Gyr), the determined probable age falls within the intermediate range of ~ 5 Gyr, consistent with other ages predicted in this

area of the HRD where the MS begins to diverge, which is densely populated by a cluster of stars spanning $\sim 400\text{K}$ and $0.3L/L_\odot$. These appear to get older with increasing brightness and decreasing effective temperature, or towards the “top” of the MS, mimicking the evolutionary pathway of our isochrones (Figure 3). The isochrones here are less packed than for stars lower on the MS, offering us slightly more evolutionary leverage to constrain our uncertainty. On the other hand, our metallicity estimate is a reasonable subsolar -0.075dex with a modest uncertainty of $\pm 0.12\text{dex}$. Here metallicity plays a more pronounced role in adjusting colour and magnitude, a quality that is physically consistent with the source’s observed photometry and given typical Snhd thin-disk chemical compositions.

3. At [3.77, 0.13, 0.74, 8.98] on the HRD, our star is sun-like, roughly solar mass and effective temperature, but slightly higher in luminosity. We can estimate another late MS/turnoff star, or a very early subgiant depending on the exact mass. Although we are coming off the MS here and isochrones begin to diverge as they grow older, there is still a significant overlap at $\sim 1.1M_\odot$ as this region is not as evolutionarily predictable. The age-metallicity degeneracy is also worth noting here as chemically rich stars evolve with moderate haste, so older metal-poor and younger metal-rich stars may appear similar on the HRD, hence the consistently broad age posterior. Our estimation of $\tau = \sim 6.5\text{Gyr}$ is consistent with theoretical models which predict MS lifetimes of 9-10Gyr at $1M_\odot$ and 5-7Gyr at $1.1M_\odot$ and $Z = \sim 0.014$ or $[\text{M}/\text{H}] = \sim 0$ (Figure 11 on [Mowlavi et al. \(2012\)](#)). Our slightly subsolar metallicity guess is typical for $\sim 6\text{-}8\text{Gyr}$ stars in the thin-disk, so our constraint is reasonable. $[\text{M}/\text{H}]$ influences colour and the structure of MS/subgiant tracks; our colour (and temperature) place our star with intermediate metallicity. Isochrone fitting here can rule out the extremes, leaving a reasonably well constrained $[\text{M}/\text{H}]$ with symmetric uncertainty, even though age isn’t necessarily nailed down.
4. At [3.81, 0.64, 0.58, 8.06] our *Gaia* star is hotter and brighter than the sun, potentially straddling the range of a late F-type or early G-type dwarf. Although not evident on our HRD plot given our parallax limited sample, the luminosity suggests this sits near the main sequence turn-off or the beginning of the subgiant arm. At this point in parameter space, stars are characterised by more rapid evolution, with shifts in small increments of $\log T_{\text{eff}}$ or luminosity breeding large variations in age, and vice versa, as the isochrones diverge more dramatically. This results in a sharper peaked age posterior, giving us a credible estimate of $\sim 2.6\pm 0.5\text{Gyr}$ - fitting well within expectations for a star that is likely just reaching the Terminal Age Main Sequence (TAMS), a phase that occurs roughly between 2-3.5Gyr for such intermediate-mass, slightly sub-solar composition stars. Conversely to the above main sequence predictions, the metallicity constraint is weaker in this regime due to isochrone degeneracies at similar $[\text{M}/\text{H}]$, leading to a broader posterior seen in Figure 5, peaking at -0.2dex. Such an intermediate metallicity estimate fits the observed thin-disk parameters - metal-poor ($\sim -0.5\text{dex}$) isochrones would likely overshoot the temperature or luminosity, whereas metal-rich ($\sim 0.2\text{dex}$) ones would place the star in an overly cool and dim pocket for its given age.

These test stars illustrate the practical behaviour of the joint age–metallicity posteriors across different regimes of the HRD and confirm that the methodology produces results relatively consistent with expectations set forth by stellar evolutionary theory and analysis.

Having established the validity of the execution of the posterior framework on individual examples, we now extend the analysis to the parallax-limited *Gaia* sample. By comparing the inferred age and metallicity distributions to known trends and structures in the MW - such as the age–metallicity relation, velocity-dispersion trends, and galactic thick/thin disk populations - we begin to assess the consistency and comprehensibility of the derived stellar parameters across the survey.

4. Application: *Gaia* Survey

4.1. Stellar Kinematics

4.1.1 Velocities

Although the posterior distributions in age and metallicity provide useful insight into the intrinsic properties of individual stars and their evolutionary histories, they do not capture the full dynamical context in which these stars reside, particularly as members of the local Snhd. To explore how kinematics correlates with stellar populations - and potentially disentangle different structural components of the Galaxy - we now turn our attention to the velocity and proper motion information provided by the *Gaia* survey, and how it behaves in relation to our postulated Bayesian-derived stellar parameters.

Through mechanisms such as astronomical spectroscopy via observed spectral line broadening and processing of consequent radial velocities and astrometric signals, the velocity dispersion of a set of stellar objects can be determined easily. The velocity dispersion σ can be defined as the statistical dispersion about a common mean velocity for a group of astronomical objects, or gravitationally bound interacting particles, such as stars in a confined neighbourhood, open clusters, or galaxy clusters. Mathematically, it is the root mean square of the N-dimensional stellar velocity vector \mathbf{v} , which is composed of one-dimensional velocity components U, V, W measured against the local average motion. See Figure 7 for visual definitions. Relationships or empirical correlations between dispersion and the derived mass of groups have been found using the virial theorem which balances the total heating, or kinetic energy, of a system (proportional to σ^2) to its associated potential energy, and hence mass.

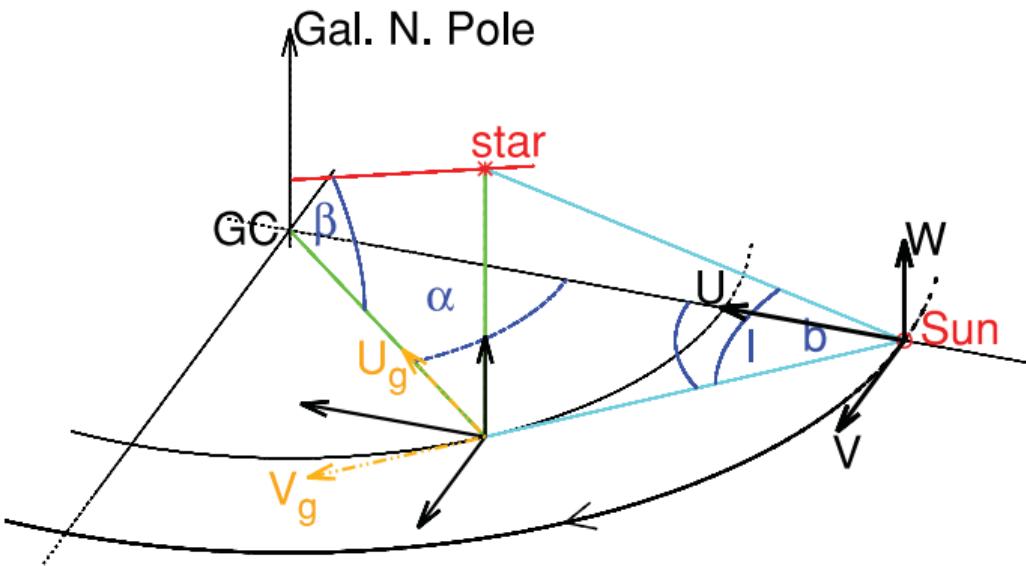


Figure 7: Visual definition of kinematic and geometric quantities. The radial U-axis denotes the direction to the galactic centre (GC), the azimuthal V-axis is orthogonal to this and encodes the direction of galactic rotation, and the vertical W-axis points towards the north galactic pole. Source of figure: [Schönrich \(2012\)](#).

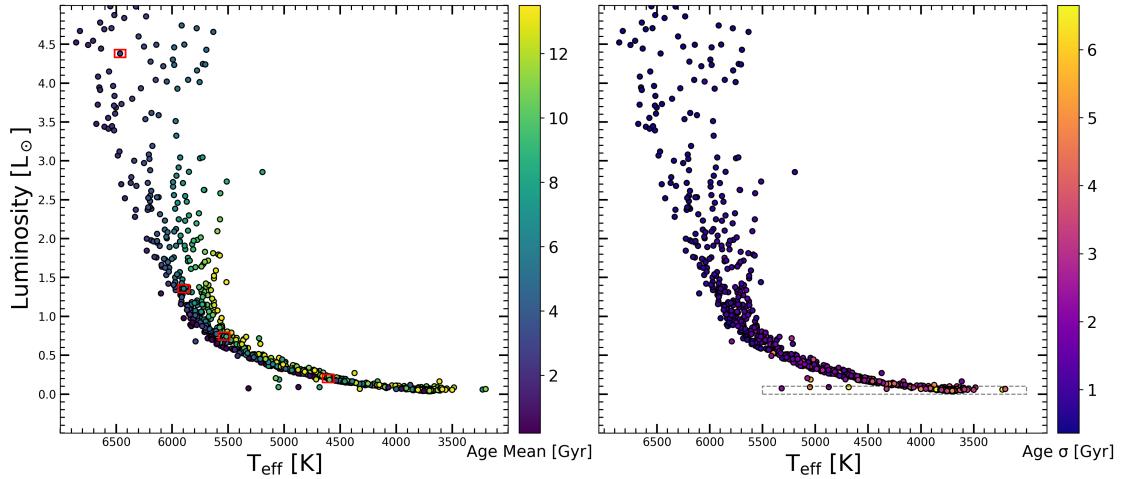


Figure 8: Hertzsprung-Russell Diagram for our parallax limited *Gaia* sample with Bayesian estimated ages (*left*) and age uncertainties (*right*). The four test stars analysed in Section 3.7 are highlighted with red squares, and our lower MS luminosity and temperature age uncertainty selection cut described in Section 4.3 is denoted by the grey-dotted rectangle.

4.1.2 Known Dispersion Forms

These dependences can take many forms, e.g. the $M - \sigma$ relation for the galactic bulge and the central supermassive black hole (Ferrarese & Merritt, 2000), or the Tully-Fisher relation for spiral galaxies (Tully & Fisher, 1977). It has also been shown that the dispersion of stellar velocities generally increases as a function of group age (Wielen, 1977), particularly when binned within the Snhd (Aumer et al., 2016). This relationship, although not properly constrained - particularly by stellar age uncertainties for MS stars and degenerate isochrone fitting in this region, as well as a bias towards younger populations for samples with provided ages (Nordström et al., 2004) - may offer insight into the mechanisms that propel stochastic heating in the local disk population. Namely, it suggests how an increase in σ over stellar time implies an increase in dynamical interactions with other gravitationally bound bodies, for example, giant molecular cloud (GMC) scattering or minor mergers. Heating mechanisms needed to explain the observed relation appear to diverge between different galactic components, simulations performed by Aumer et al. (2016) suggest that combined GMC and spiral-arm heating are sufficient for the thin disk, whereas the thick disk requires additional heating sources. Additionally, establishing a strong empirical relationship between σ and τ can allow velocity dispersion to act as a proxy for stellar age, particularly useful for slowly evolving MS stars, where isochrone fitting naturally comes with a high level of uncertainty (Soderblom, 2010).

4.2. σ vs τ

4.2.1 Expected Relation

The velocity dispersion, σ , as a function of stellar age, τ (age-velocity relation henceforth known as AVR), can be described as a simple power law with an exponent β typically found in the region of ~ 0.3 , although this is not a universally accepted quantity and has been shown to deviate. Early estimates include Lacey (1984) who derived a proto-estimate value of 0.25 for β , which turned out to be a lower value than the data indicated. Studies including Aumer & Binney (2009); Carlberg & Sellwood (1985), and Casagrande et al. (2011) suggest a flattening or plateauing of dispersion at greater ages, the latter particularly showcasing this for all stars - not just those with 'good' ages -

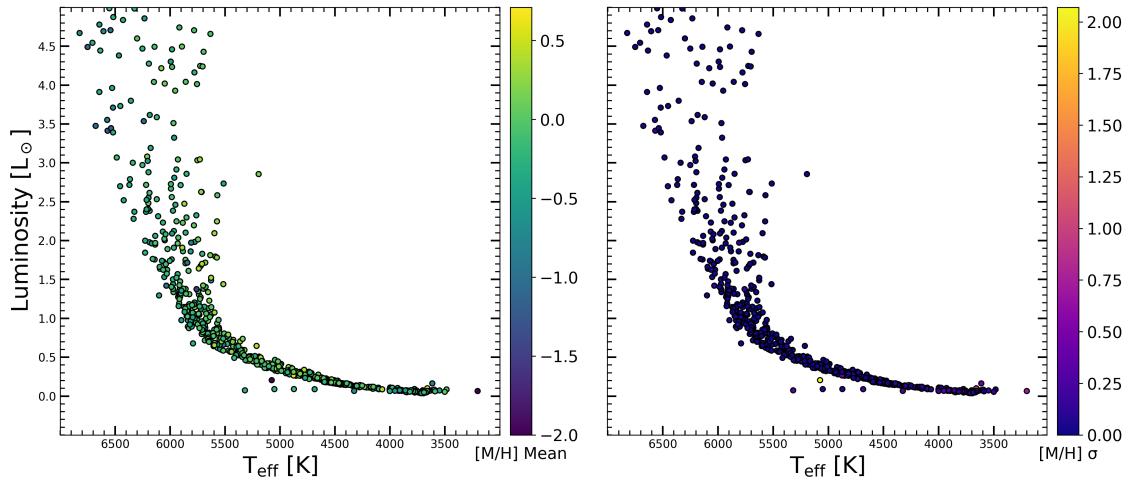


Figure 9: Hertzsprung-Russell Diagram for our parallax limited *Gaia* sample with Bayesian estimated metallicities (*left*) and [M/H] uncertainties (*right*).

derived from BaSTI isochrones with a metallicity cut-off of $[Fe/H] > -0.8$, similar to the stellar sample we derive. [Sellwood & Carlberg \(1984\)](#) predict a lower rise of σ^2 with time leading to fainter spiral heating effect patterns at later times as σ rises secularly, quenching the gravitational effects of star organisation; their simulations thus recreate such AVR as seen in [Wielen \(1974\)](#) and [Mayor \(1974\)](#). [Ting & Rix \(2019\)](#)'s use of *Gaia* DR2, and [Seabroke & Gilmore \(2007\)](#)'s study of the GCS find σ_W component has the steepest increase, commonly interpreted as a stronger vertical heating over time; but this interpretation may be oversimplified. [Schönrich & Binney \(2009\)](#) and [Aumer et al. \(2016\)](#) show that high- σ_W populations can be imported from the inner Galaxy via radial migration, not just created locally. This process naturally steepens the σ_W -age relation in the Snhd by adding older stars that were born in dynamically hotter environments. In contrast, the other components are affected more by in-plane dynamics such as non-axisymmetric potential perturbations from long-term spiral arm scattering at resonance, or radial migration and asymmetric drift ([Sellwood & Binney, 2002](#)). These observations are substantiated by [Tarricq, Y. et al. \(2021\)](#) who observe a similar preference for a strong vertical increase and also evidence for flattening out at old ages (>8 Gyr) as seen in [Casagrande et al. \(2011\)](#). These features likely reflect a combination of both local secular heating and the cumulative contribution of migrated stars from dynamically hotter inner-disk regions.

4.2.2 Ages and Uncertainties

The inherent uncertainties and degeneracies in derived stellar age estimations - particularly for MS stars where age-sensitive parameters evolve slowly - mean that traditional methods that might require high-resolution photometry or astrometry (such as empirical calibration to well-studied clusters, or through astroseismology, which are often inapplicable or unavailable for volume-limited samples) may be discarded in favour of our Bayesian framework. Our *Gaia* sample provides enough constraints in our chosen parameter space vector \mathbf{q} to feed into our Bayesian algorithm star by star, and compute posterior PDFs in $\mathbf{p} = [\tau, [M/H]]$ for each. Then it is simple to marginalise out any unwanted parameters and calculate the first and second moments to provide estimates on the mean and associated uncertainty as discussed in section 3.6. Metallicity is marginalised over to account for age degeneracy inherent with isochrone fitting and because the AVR is currently of interest here rather than a historical chemical evolution trend. The resulting HRDs populated by stars determined with our Bayesian algorithm coloured by both age and metallicity are shown in Figures 8 and 9 respectively.

Once age estimates have been calculated on a star-by-star basis, they must then be binned at a significant enough resolution in which there are enough data points to track and mathematically model meaningful trends in the dispersion relationship, while also balancing to ensure the bins aren't too small such that there are enough stellar velocities to calculate said statistical dispersion. After all, this tends towards a "true" value the greater the number of data points used, and the uncertainty scales as $\sigma/\sqrt{2N}$ where N is sample size (Nissen, 1995).

4.2.3 Calculating Dispersion

The *Gaia* sample provides measurements on individual stars' parallax, proper motion, and radial velocity, hence we must also compute the U,V,W velocities and their corresponding σ_U , σ_V , σ_W dispersions manually. The following quantities and their uncertainties are used: $\varpi \pm \sigma_\varpi$ parallax in arcsec, $v_r \pm \sigma_{v_r}$ radial velocity in km s^{-1} , right ascension α in degrees, declination δ in degrees, $\mu_\alpha \pm \sigma_{\mu_\alpha}$ proper motion in right ascension in arscec yr^{-1} , and $\mu_\delta \pm \sigma_{\mu_\delta}$ proper motion in declination in arscec yr^{-1} . Distance d can be found in kpc taking the reciprocal of the parallax and the tangential velocity components are calculated using dimensional analysis: $V_\alpha = k \cdot \mu_\alpha^* \cdot d$, $V_\delta = k \cdot \mu_\delta \cdot d$ where k is a numerical conversion factor from mas yr^{-1} and kpc to km s^{-1} . We then need a transformation from our current equatorial coordinate system to our final galactic coordinate framework through an intermediate representation of Cartesian velocity:

$$\begin{bmatrix} V_X \\ V_Y \\ V_Z \end{bmatrix} = \begin{bmatrix} \cos \alpha \cos \delta & -\sin \alpha & -\cos \alpha \sin \delta \\ \sin \alpha \cos \delta & \cos \alpha & -\sin \alpha \sin \delta \\ \sin \delta & 0 & \cos \delta \end{bmatrix} \begin{bmatrix} V_r \\ V_\alpha \\ V_\delta \end{bmatrix} \quad (18)$$

and finally

$$\begin{bmatrix} U \\ V \\ W \end{bmatrix} = \mathbf{R} \begin{bmatrix} V_X \\ V_Y \\ V_Z \end{bmatrix} \quad (19)$$

where \mathbf{R} is the typical J2000 rotation matrix (Johnson & Soderblom, 1987). It then follows that the velocity dispersion is calculated given a set of velocities

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (v_i - \bar{v})^2} \quad (20)$$

where N is the number of stars in our chosen subset - in this case the age bins, v_i are the individual velocities (e.g. U_i , V_i or W_i) and \bar{v} is the mean component velocity of the bin population. Each velocity component v_i is subject to an uncertainty which we calculate using the equation for the variance of a function of several variables as seen in Section IV of Johnson & Soderblom (1987). To compute the subsequent uncertainty on the sampled velocity dispersion, we must consider the statistical variance of the sample given the nature of a dispersion, alongside the variance introduced by said individual measurement errors in velocity. Thus, we propagate the individual measurement uncertainties under the assumption that the intrinsic distribution and the measured errors are Gaussian and uncorrelated, just as is required for the equation of variance above to hold true. Then the uncertainty in dispersion is approximated by

$$\delta\sigma \approx \left[\frac{1}{2(N-1)} \frac{1}{N} \sum_{i=1}^N ((v_i - \bar{v})^2 + e_i^2) \right]^{1/2} \quad (21)$$

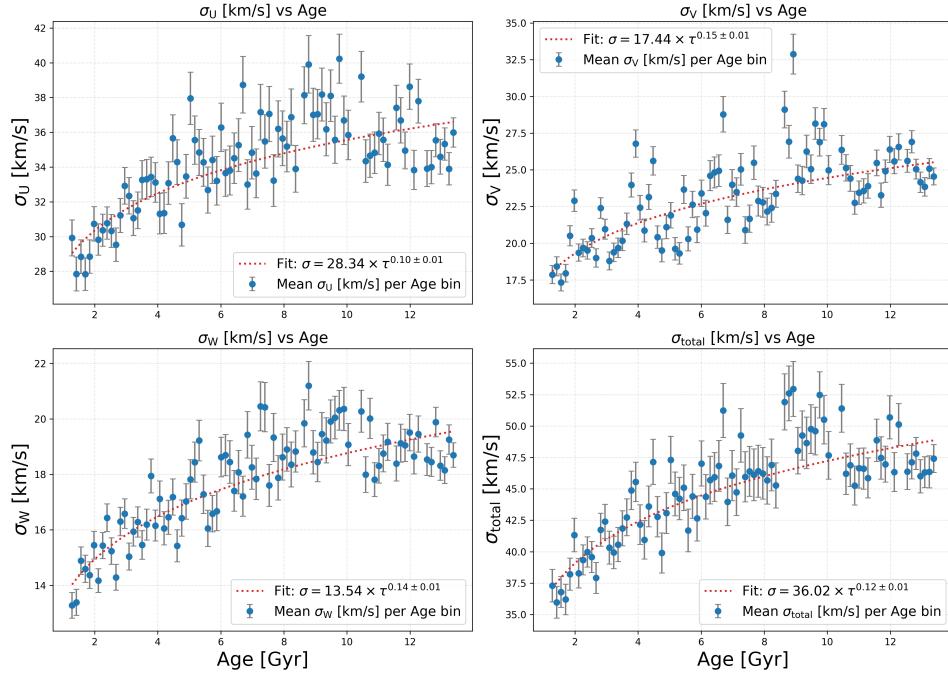


Figure 10: Total velocity dispersion σ (*bottom right*) and individual U,V,W dispersions vs age τ for our *Gaia* parallax limited sample with determined Bayesian ages. Error bars consider both uncertainty from underpopulated bins and contributions from individual errors in v_i as shown in Equation 21.

where the dispersion segment captures the spread in measured velocity and the e_i^2 segment is the mean squared measurement error across the sample from individual velocities. We prefer this frequentist approach to uncertainty as the it is analytic and computationally efficient compared to a Bayesian approach where sampling the posterior with an MCMC algorithm or constructing a PDF mean or median form where marginalising over the unobserved true velocities would be required. The used form follows from standard error propagation and sampling distribution methodology seen for example in Bevington & Robinson (2003).

4.2.4 Initial Fit

The raw data that encompass every star in our *Gaia* sample with processed age and metallicity maximum likelihood values, and are distributed across 100 age bins from $0.1 < \tau < 14.0$ Gyr, produces an AVR shown in Figure 10. The dotted red line follows from a defined power-law where σ_0 normalises the function, indexing the magnitude of dispersion for stellar groups at $\tau = 0$, and β denotes the power-law index we are trying to observe. This power-law relationship uses the known uncertainties in σ discussed above to appropriately weight the fit and provides error estimates on said model parameters, leaving us with a preliminary guess of $\sigma = (36.12 \pm 0.40) \cdot \tau^{0.12 \pm 0.01}$.

4.3. Selection Cuts

4.3.1 Velocity and Parallax

Due to the clear discrepancy with the 'recognised' literature value of $\beta = \sim 0.3$, we must model the impact of potential selection functions on the survey, and hence the AVR. We can start by defining a signal-to-noise ratio (SNR) for stellar parallax and the provided uncertainties and discarding all stars

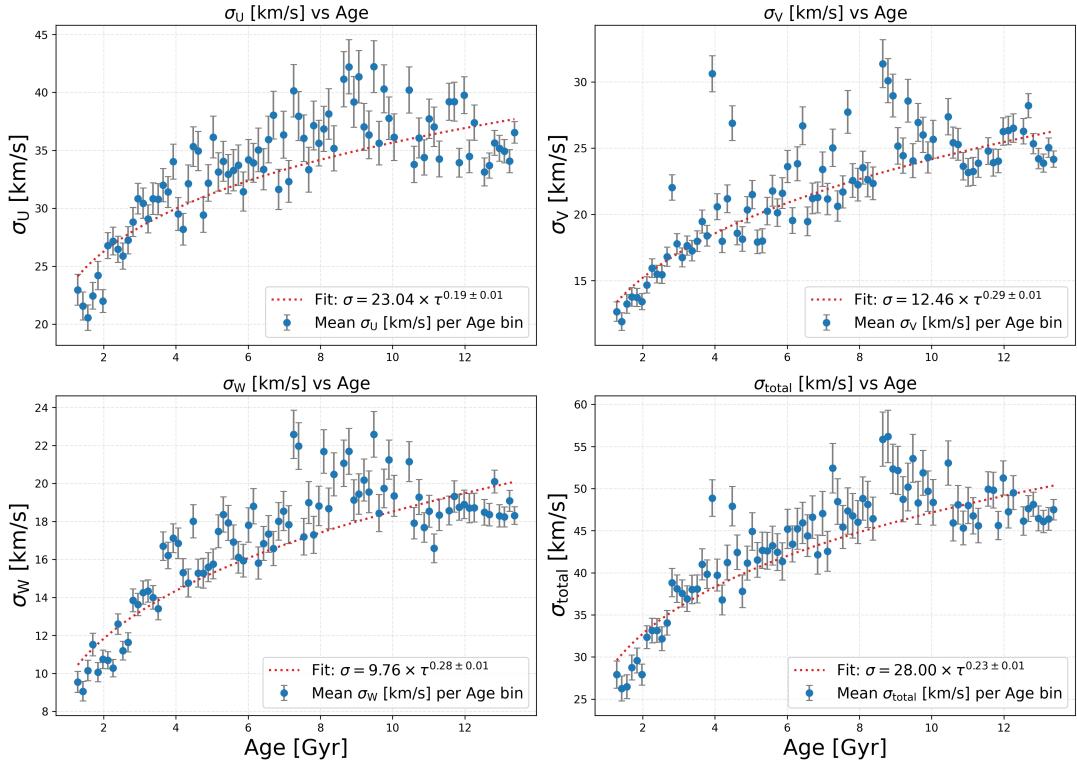


Figure 11: Same set of AVR as Figure 10 but with the following selection cuts: $\varpi/\sigma_\varpi > 100$, $\epsilon_\sigma > 2 \text{ km s}^{-1}$, $\sigma_\tau > 30\%$, $V_{tan} > 100 \text{ km s}^{-1}$. The observed slight peak at $\sim 8\text{-}10 \text{ Gyr}$ is likely due to errors piling up from old stars with 'bad' ages being scattered downward.

that overcome a certain value for relative uncertainty; *Gaia* provides notoriously accurate astrometric data for every source, so our cut-off for ϖ/σ_ϖ needs to be exceptionally high, e.g. magnitude of ~ 100 , to filter out even 5% of catalogue sources. Applying this cut reduces distance uncertainty due to its inverse coupling with parallax, limiting bias in tangential velocity calculation (large parallax uncertainty can artificially inflate distances and hence lead to unrepresentatively inflated [U,V,W] components). Extending from this, introducing cuts in [U,V,W] component errors prevents unwanted inflation in dispersion from large single measurement errors, for example from young thick-disk populations having intrinsic dispersions of $< 10 \text{ km s}^{-1}$ and associated errors at a similar magnitude, which could dominate the signal. As most of our calculated velocity component uncertainty is on the order of $\sim 1 \text{ km s}^{-1}$, we can set a cut-off for those with say $\epsilon_{U,V,W} > 2 \text{ km s}^{-1}$. A lower limit on tangential velocity V_{tan} of 100 km s^{-1} is also applied.

4.3.2 Age Uncertainty Filter

Finally, our Bayesian age and metallicity posterior algorithm lends itself to predict large age uncertainty relatively often, those of which span several Gyr and hence multiple bins, depending on the resolution. Introducing a cut to exclude such poorly constrained stellar ages, such as $\sigma_\tau > 30\%$ (uncertainty, not age dispersion) will rid our sample of erroneous and imprecise sources relative to the observed distribution. For our calculated ages, this threshold removes $\sim 40\%$ of the sample, which will directly introduce sources of bias. Less well-constrained sources including young ($< 1 \text{ Gyr}$) stars, which evolve slowly on the HRD and have tightly packed isochrones, and older ($> 10 \text{ Gyr}$) stars which may experience age/metallicity degeneracy during isochrone fitting, may flatten our observed AVR as sources are systematically removed from both extremes.

4.3.3 Observed Flattening

This flattening, especially at greater ages, can be seen in Figure 11 where we plot all individual dispersion components as well as the total dispersion as a function of age, with all discussed selection cuts on astrometric SNR, velocity uncertainty and age errors applied. The overall value for β has increased to 0.23 ± 0.01 , a step toward our intended target of ~ 0.3 , the expected literature value. These refinements deal with poorly constrained kinematic data and ages which may otherwise synthetically flatten the AVR by inflating calculated dispersion within young bins or smoothing out gradients in old ones. The remaining discrepancy might reflect observational bias and probable physical deviation from an empirical assumed power law model - particularly at old ages where vertical heating is driven by internal heating mechanisms that become less efficient at scattering hot orbits, or where thick-disk sources begin to populate the age bins. However, for our volume-limited sample with the $\varpi > 10\text{mas}$ cut, there is an inherent bias against giant stars in the galactic disk which have already evolved off the MS. These are preferentially sampled at larger distances (and smaller parallaxes) due to their higher luminosity and will therefore be under-represented in our *Gaia* sample. This is important as older stars, especially thick disk, show higher dispersions due to dynamical heating over time, or because they likely formed in more active early galactic environments. This manifests in evolved younger stars inheriting these accelerated heated motions, and hence leading to an under-representation of high- τ high- σ sources, further flattening the observed slope of the AVR.

We also see an apparent downturn in omnidirectional dispersion at high age ($>\sim 10\text{Gyr}$) on a bin-average basis towards the end of the observed plateau. This could be due to uncertainties derived from our isochrone models which distribute stars normally on the HRD, which may scatter low ages towards high ages due to small photometric errors, making stars with disk-like kinematics appear older. Another thought is that at high τ , there is a sharp upper limit on the ‘true’ ages ($\sim 13.8\text{Gyr}$, the estimated age of the universe), so there may also be a bias of stars near this limit being much more likely to scatter asymmetrically downward than slightly younger stars which may be pushed in both directions, diluting younger bins with higher, more spurious velocity dispersions. Additionally, stars at the end of the plateau may not necessarily be old, but instead reflect skewed age PDFs. Stars with poorly constrained ages often present long-tailed PDFs against the 14Gyr limit, causing an expectation value pile-up at high ages.

4.3.4 HRD Cuts

When we refer to the HRD visualisation of our selected stars in Figure 8 we see the expected lack of a Red Giant Branch (RGB) or Asymptotic Giant Branch (AGB) component that should populate parameter space in the region of lower T_{eff} ($\sim 3000\text{-}5000\text{K}$) and increased brightness ($>\sim 3L_{\odot}$). But we also notice that for stars that have survived the initial age-uncertainty cut-off, the greatest magnitude of error occurs on the MS, particularly the lower main sequence (LMS), and this increases as T_{eff} falls off. This is a well-known limitation of this subpopulation within astrophysical research (Valle, G. et al., 2015); (Babusiaux et al., 2018)) and is consistent with the slow, nearly featureless evolution of stars on this track during the hydrogen core burning phase where both hydrostatic and thermal equilibrium are stably maintained. Since isochrone-based age estimates rely on evolutionary changes, and HRD positions for low-mass lower MS stars vary little over the timescales of Gyr, the lack of movement makes age in this region particularly degenerate. This is manifested in evolutionary models, see Figure 3, as isochrones are tightly packed on the LMS - even small uncertainties in colour and magnitude can push a low-mass star from one isochrone to an adjacent one, or even across multiple isochrones. Hence, we observe large uncertainties and potentially bimodal solutions in our marginalised PDFs (see Figure 5). Again we can refer to the so-called age-metallicity degeneracy, where cold faint stars populating this area may be young and metal-rich, or old and metal-poor. Strictly speaking, we can

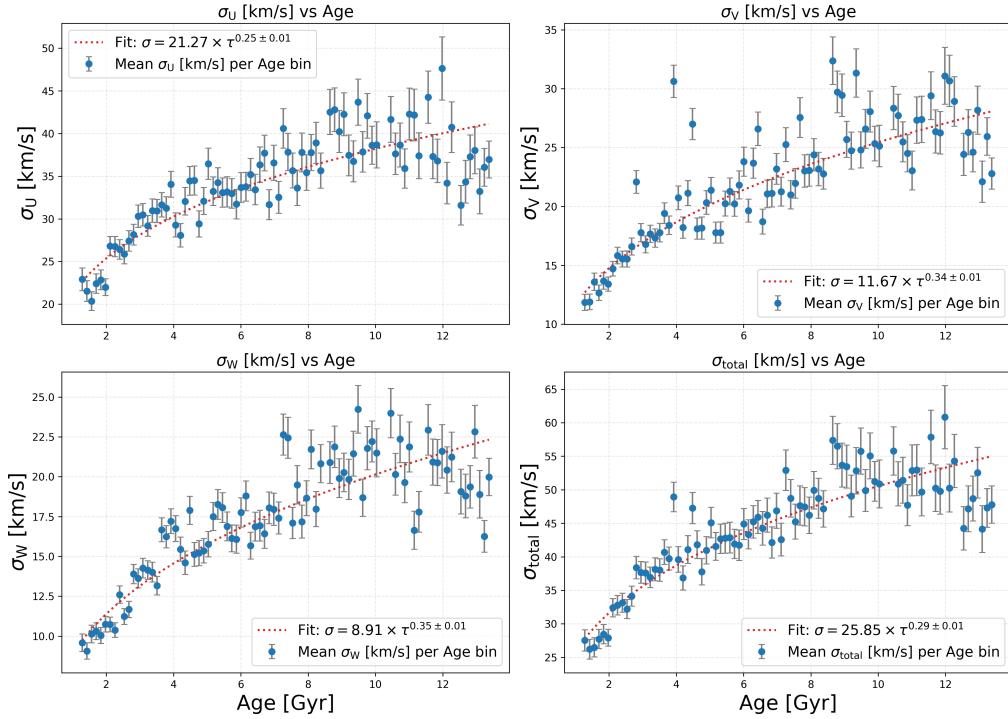


Figure 12: Same set of AVR as Figure 11 but with the following magnitude selection cuts: $T_{\text{eff}} > 5500\text{K}$ & $L/L_{\odot} > 0.1$.

try to disentangle this by precisely denoting the metal content of each star - which we have achieved in our PDF calculations - but in this case it will be easier to just not account for them in our dispersions at all.

To counter these further discrepancies in age, we can introduce further selection cuts towards the lower main sequence. We select a cut-off for sources with $T_{\text{eff}} < 5500\text{K}$ as well as those with $L/L_{\odot} < 0.1$ which will roughly exclude stars later than the early K-types, and remove stars deep in the LMS. The combination of these two parameter cuts effectively excludes the majority of stars with masses of $< 0.8\text{-}0.9 M_{\odot}$ whilst keeping the solar-type and heavier residual MS stars which have more reliable age determinations. Quantitatively, stars in our sample that fall within our manufactured lower main sequence boundary conditions (shown within the dashed box in Figure 8) have an average age uncertainty of $\sim 4\text{Gyr}$, whereas the remainder have an average of only $\sim 1.6\text{Gyr}$. Thus, this supplementary diagnostic on population location further cleans the *Gaia* sample of unwanted stars with questionable age determinations and allows us to focus on our calculation of our exponent β .

4.3.5 Final AVR

Doing so results in the relationships seen in Figure 12, where β has increased from 0.22 ± 0.01 to 0.29 ± 0.01 , well within the limits of the expected value given the literature. For further examples of accepted β ranges, (Nordström et al., 2004) find an exponent of 0.34 ± 0.05 for the total dispersion-age relation power law using Snhd stars with the best determined ages within the GCS - both in relatively close agreement with Holmberg et al. (2009) ($\beta = 0.40$) and Binney et al. (2000) ($\beta = 0.33 \pm 0.03$).

This result further validates our treatment of age uncertainty in the stellar population as the biggest hurdle against recovering this expected slope, and reinforces our interpretation of dynamical heating in the thin-disk (Aumer et al., 2016). In doing so, we have struck an agreeable balance between sample completeness and the statistical soundness of our inferred σ - τ trends to a justifiable level.

The recovered AVR s also justify our implementation of a Bayesian framework in which predicted stellar ages are internally consistent and dynamically meaningful compared to previous methodology. In a broader sense, it highlights the extent to which selective but physically motivated HRD cuts substantially improve the accuracy of population-wide kinematic inferences, proving that not all stellar sources are equally informative for deducing all astrophysical relationships, and not all regions of parameter space constrain said relationships to an equal extent. The overall incremental divergence towards a "true" estimate of the σ - τ relationship due to selection and population cuts is shown in Figure 13.

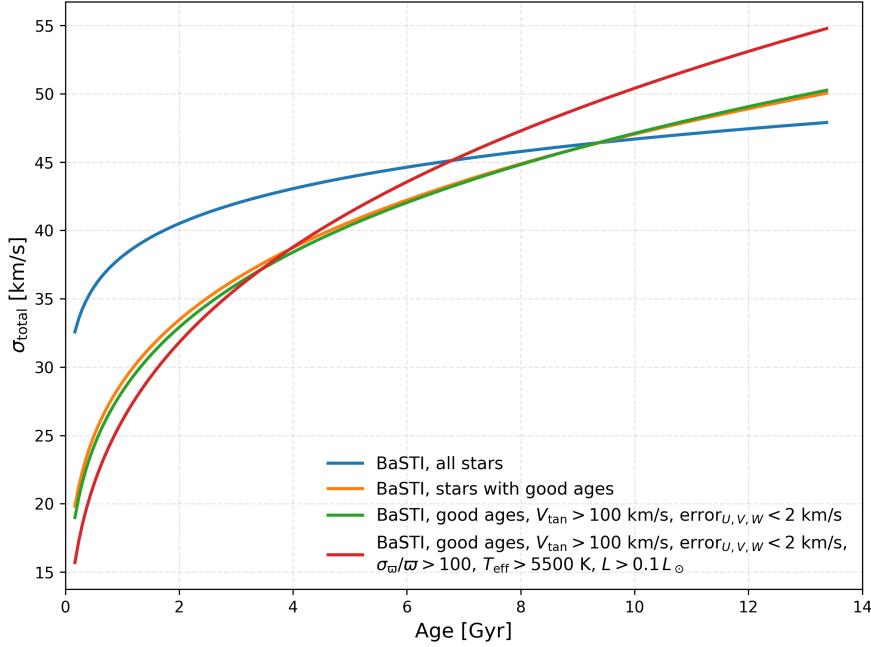


Figure 13: Total velocity dispersion $\sigma = \sqrt{\sigma_U^2 + \sigma_V^2 + \sigma_W^2}$ as a function of age τ for derived Bayesian ages and provided *Gaia* kinematics. Our power-law best fit exponent estimates are $\beta = 0.12 \pm 0.01$, 0.22 ± 0.01 , 0.23 ± 0.01 & 0.29 ± 0.01 for the blue, orange, green, and red lines respectively, which represent increasingly constraining selection cuts.

4.4. Population Boundaries

4.4.1 Definitions

It should be noted that the term 'stellar population' used commonly in this work so far is in fact quite nebulous, with no clear-cut definition provided by the previous literature. It is commonly viewed as an ensemble of stars that share some observable properties (Baade, 1944); (McCarthy, 1959). However, a more subtle definition should allow us to link the observables to properties that will uniquely define the formation conditions of such populations (Haywood, Misha et al., 2013). Ideally, our chosen proxies of age, metallicity, and kinematic distributions should offer us a solid foundation to distinguish between populations and relate them to our desired insights. In practice, it is difficult to establish such distinctions as there is no agreed-upon methodology of unambiguously defining single populations.

4.4.2 Velocity, Age, Structure

For example, initial classifications were based on the physical structure of the galaxy itself, with the thin disk defined as having shorter scale heights than the thick disk ([Majewski \(1993\)](#); [Buser et al. \(1999\)](#)), whereas the thick disk has longer scale lengths than the thin disk ([Ojha et al., 1996](#)). In terms of kinematics, thick-disk stars have often been thought to have larger velocity dispersions and lag the net rotation of the disk ([Soubiran, C. et al., 2003](#)), meaning that the difference between the star's radial velocity V_R and that of the disk at that radius is positive. V_{tot} (based on U,V,W velocities) has been used as a simple metric to distinguish between the three, with thin-disk stars populating total 3-D velocities of typically $\sim V_{tot} < 50 \text{ km s}^{-1}$, thick-disk with between $\sim 50\text{-}70 \text{ km s}^{-1}$ and $150\text{-}220 \text{ km s}^{-1}$, and halo stars adopting higher total velocities, roughly $>\sim 180 \text{ km s}^{-1}$ on average ([Feltzing, S. et al. \(2003\)](#); [Venn et al. \(2004\)](#); [Bonaca et al. \(2017\)](#); [Hawkins et al. \(2015\)](#); [Chen et al. \(2021\)](#)). See Figure 14 for a simple diagram showing these regimes for our selected sample. [Bensby et al. \(2003\)](#) adopts a probabilistic approach to separating the disks, defining ratios encoding the fraction of stars belonging to a given component, and a probability distribution function $f(U,V,W)$ that takes into account both the stellar velocities and the component velocity dispersions. Additionally, age has been used as a qualifier, with thick disk stars usually being older relative to their thin-disk counterparts ([Chiba & Beers \(2000\)](#); [Bochanski et al. \(2007\)](#)) - a common cut-off threshold age between the two disks is around $\sim 8 \text{ Gyr}$ ([Fuhrmann \(1998\)](#); [Hou et al. \(2012\)](#); [Haywood, Misha et al. \(2013\)](#)). Additionally, in terms of physical location, [Ivezic et al. \(2008\)](#) and [Loebman et al. \(2011\)](#) bring attention to a particular region in the local Snhd in their MW simulations, determining the thin-thick "transition zone" resides somewhere inside the cylindrical volume defined by $|z| = 0.5\text{-}1.0 \text{ kpc}$ and $R = 7\text{-}9 \text{ kpc}$. Within this volume a roughly equal number of thin- and thick-disk stars is expected.

4.4.3 Chemical Separation

Chemically, thick disk stars are typically described as more metal-poor than thin disk stars, with greater alpha element enhancement $[\alpha/\text{Fe}]$ (higher ratios of "alpha elements", i.e. oxygen, magnesium, calcium, compared to the typical solar value) at a given iron abundance compared to the thin disk ([Fuhrmann \(1998\)](#); [Prochaska & Wolfe \(2000\)](#); [Bensby et al. \(2003\)](#)). Approaches defining population boundaries on the basis of age and chemical properties suggest distinct epochs of formation. The above approaches suggest a thick-disk phase characterised by rapid chemical enrichment, whereas the thin-disk formed later (hence younger) under greatly different conditions. Boundary approaches inherently breed overlaps in population parameter space depending on whose definition you choose to demarcate between them. For example, stars classified kinematically as thick-disk may be younger or more chemically enhanced than thin-disk stars, or vice versa, contradicting previously established conditions we would ideally hold over every tracer we use ([Minchev et al. \(2014\)](#); [Nordström et al. \(2004\)](#)). The kinematic separation criteria alone may lead to population separation, even in simulations where different populations may not necessarily be accounted for or intended, suggesting that mechanisms such as radial migration may play a role in observed differences([Loebman et al. \(2011\)](#)). The complexity increases if one considers substructures within the galactic halo, which are usually identified kinematically and may suggest distinct accretion events that contribute to the halo composition (e.g. [Koppelman et al. \(2018\)](#)).

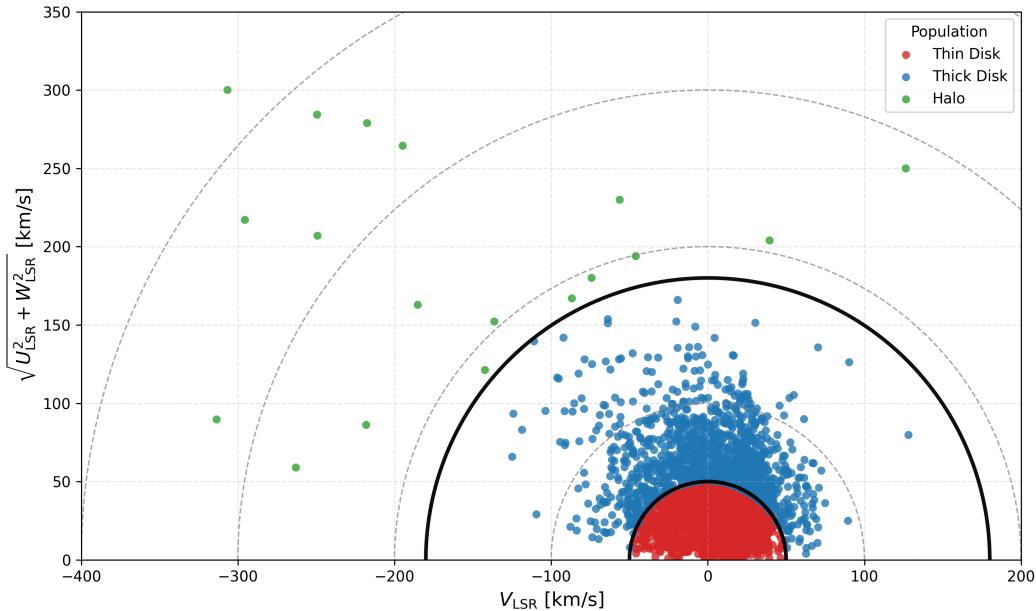


Figure 14: Toomre diagram for our *Gaia* sample distinguishing different Galactic components in terms of total velocity. Dashed lines show constant values of the total Galactic velocity $V_{tot} = 100, 200, 300$, and 400 km s^{-1} . Solid black lines separate the populations at 50 and 180 km s^{-1} . Here we work in terms of the Local Standard of Rest (LSR) - i.e. velocities relative to the average motion of the Sun within its immediate neighbourhood, to centre our distribution around $V_{LSR} = 0 \text{ km s}^{-1}$. This is performed by correcting our provided U, V, W velocities by the solar motion - the velocity of the Sun relative to the Galactic center - $(U_\odot, V_\odot, W_\odot) = (11.1, 12.24, 7.25) \text{ km s}^{-1}$ (Schönrich et al., 2010), as well as the circular velocity of the sun $V_{circ} = \sim 220 \text{ km s}^{-1}$ (Bovy et al., 2012).

4.5. σ vs [M/H]

4.5.1 Metallicity

To fully understand the relationship between the kinematics of the local Snhd, their age distributions, and any inferred insights into their chemodynamical histories, we should also consider the coupled metallicities. When we accredit this combination, ideally with supplementary photometry, we can utilise the information as a tracer of Galactic archaeology and help disentangle the origins and evolutionary histories of stellar populations. Metallicity acts as a guide for the chemical enrichment of the ISM at the time and place where a star is formed. Older stars formed when the ISM was not as enriched from e.g. supernovae and are hence more metal-poor, whereas young stars exploit gas enriched by previous generations and are more metal-rich. Our galactic components formed at different times, rates, and under varying initial conditions, and hence their population stars observe different [M/H] distributions; metallicity correlates with stellar origin. Therefore, age and kinematics are not enough individually to constrain the population. Our age estimates are highly uncertain (especially for LMS), and there exists a degeneracy between old thick/thin stars. Our kinematic observations may also be tainted with dynamical heating which can scatter velocity differences between components, and mechanisms such as radial migration - metal-rich stars transitioning radially outward with thin-disk-like velocities.

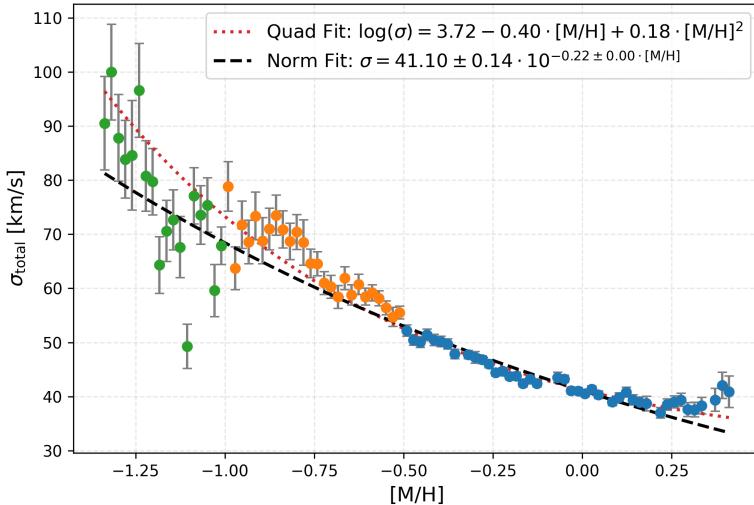


Figure 15: Total velocity dispersion $\sigma = \sqrt{\sigma_U^2 + \sigma_V^2 + \sigma_W^2}$ as a function of [M/H] for derived Bayesian metallicities and provided *Gaia* kinematics. Provided are two best fit approximations, log-norm (black-dashed) and quadratic (red-dotted).

4.5.2 More Selection Cuts

We can start to infer the populations of our sample on the basis of kinematics, age, and composition. Having already found a reasonable relationship between τ and σ in Sections 4.2, 4.3, and 4.3.4, we now need to find a similar relationship for [M/H] to constrain our populations. Using the same methodology, we bin our sample in terms of our Bayesian metallicities and uncertainties from $-1.5 < [\text{M}/\text{H}] < \sim 0.42$ across 100 segments and calculate the dispersion per bin in terms of [U,V,W] velocities. The produced relationship (Figure 15) presents both a quadratic and a log-normal power law fit showing a general decrease in velocity as stars get more enriched. Also included is a crude initial metallicity-based population identifier to show the decrease in dispersion as we transition from halo to thick-disk to thin-disk, with arbitrary cut-offs between the three at $[\text{M}/\text{H}] = -1.0$ and -0.5 . An additional halo signifier of $V_{\text{tot},LSR} > 150 \text{ km s}^{-1}$, as marked in the Toomre diagram for our sample in Figure 14, is used to differentiate further and apply future data cuts. For our relationships of σ_{total} and σ_W (dispersion along the galactic pole axis), we observe large uncertainties concentrated within halo-like metallicity bins due to their sparse population and the parallax bias within our survey, which discriminates against such dim-likely halo stars. Hence we choose two competing model fits to try to quantify the relationship slope at either ends of the metallicity spectrum where uncertainties are boosted and the gradient strays from roughly linear. Our initial guess to quantify the relationship between velocity and metallicity shows an exponent of -0.22 ± 0.005 and -0.31 ± 0.01 for σ_{total} and σ_W respectively. As before with age, slicing our dataset with selection functions and data cuts will help constrain this relation. Again we utilise the same parallax error, tangential velocity, and $V_{\text{tot}} > 150 \text{ km s}^{-1}$ cut-offs to filter out stars - this time with a MS box cut approximating the *Gaia* CMD trend: $1.5 > G_{\text{bp}} - G_{\text{rp}} > 0.5$ and $7.0 > M_G > 4.0$. Within this region stars are more uniform in their τ -M/H relations and have well-behaved kinematics. The cut also removes WDs, subgiants, and giants, which may contaminate dispersion trends due to alternate evolutionary stages. Doing so produces a fit with the log-normal power exponent increasing to -0.19 ± 0.01 and -0.29 ± 0.02 for σ_{total} and σ_W . The squared component coefficient on our quadratic fit also decreases dramatically for both, suggesting our selection cuts lead to refinement towards a more linear relationship. The results for both are shown in Figure 16.

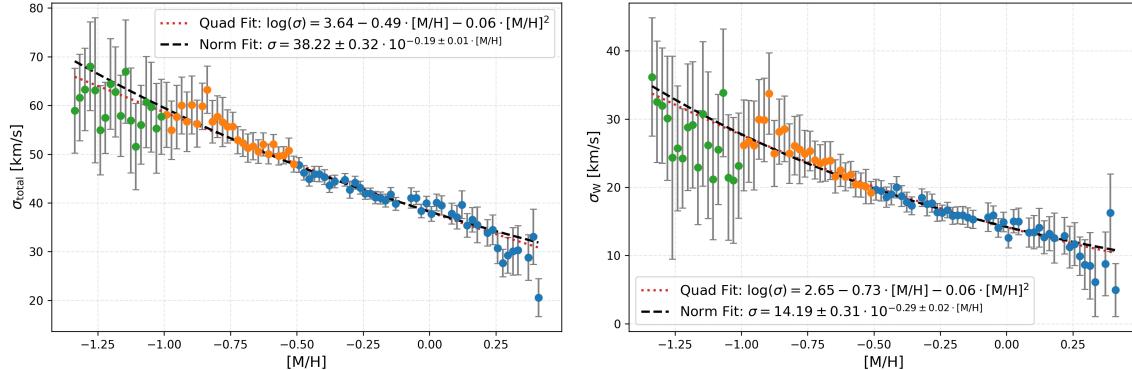


Figure 16: Total (*left*) and vertical (*right*) velocity dispersion as a function of Bayesian determined metallicity for our *Gaia* sample with $\varpi/\sigma_\varpi > 100$, $V_{\text{tot},LSR} > 150 \text{ km s}^{-1}$, $1.5 > G_{\text{bp}} - G_{\text{rp}} > 0.5$ and $7.0 > M_G > 4.0$ cuts applied. The observed increase in uncertainty at either metallicity extreme reflects the above cuts intended tendency to discriminate against metal-poor thick-disk and halo stars, as well as metal-rich spiral plane stars (Khoperskov, S. et al., 2023) (maybe less likely to contribute to σ_w , although spiral arms can also induce small amounts of vertical velocity change in stars due to the non-axisymmetric potential (Faure et al., 2014)). Along the loci of our established thin-disk population, the observed trends are maximumly constrained as the associated bins are populated more heavily compared to the edges, as is expected of this sample.

4.5.3 Results

A stronger $[\text{M}/\text{H}]$ dependence in σ_w compared to σ_{total} is fairly consistent with theoretical expectations; vertical heating mechanisms, e.g. scattering via GMCs, produce systematic relationships between these parameters. The vertical component is less contaminated by rotation effects as spirals do not directly increase σ_w significantly (Sellwood (2014); Martinez-Medina et al. (2015)) and GMCs are not the only source of a non-axisymmetric component to the gravitational field (Aumer et al., 2016) - W generally has the lowest average dispersion (Schönrich, 2012), and is least affected by streaming motions which are usually confined to the U-V velocity plane (Dehnen, 1998). Hence, the vertical component is valuable for understanding structure and for separating stellar populations, particularly for showing that the “vertical velocity dispersion progressively decreased in the thick disk, from about 50 km s^{-1} to 25 km s^{-1} , suggesting star formation proceeded in progressively thinner layers” (Haywood, Misha et al., 2013).

As a further step, we can observe the effects on the metallicity relation of removing old stars and those with ‘bad’ ages from the sample. In doing this, the σ - $[\text{M}/\text{H}]$ trend becomes roughly \cap -shaped instead of negatively correlated (see Figure 17 versus Figure 16); dispersions are lowest at the metal-poor and -rich tails, peaking around $[\text{M}/\text{H}] \simeq -0.6 \text{ dex}$. This follows naturally from our imposed selection cuts; hot, α -enhanced thick-disk stars drive the higher dispersions at the low end (Feuillet et al., 2019), while the super-solar tail is dominated by young, metal-rich, thin-disk stars and inner-disk migrators (which are numerous given our $\varpi > 10 \text{ mas}$ sample cut) that remain kinematically cooled. These enriched inner-disk stars are born in regions of rapid chemical evolution and may migrate outward via churning mechanisms (e.g., corotation resonance, Sellwood & Binney (2002)), often without adopting substantial random motions. The intermediate metallicities signify a transitional population which retains a mix of ranges up to 8Gyr and secularly heated thin-disk stars, resulting in a relatively higher dispersion plateau. However, this explanation is likely not the full picture, as the role of migrators and their kinematic imprint on metallicity are not fully explored and warrant further investigation.

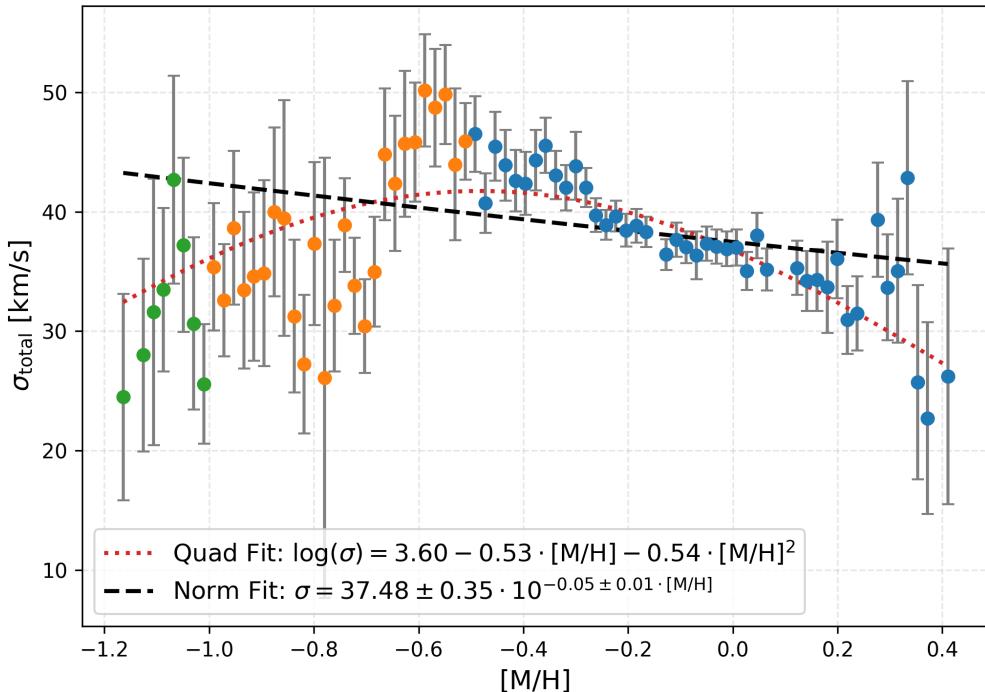


Figure 17: Metallicity-dispersion relation for our Bayesian-derived sample with 'good' ages ($\sigma_\tau < 30\%$) only, and an imposed age cut of 8Gyr.

4.5.4 Metallicity Gradient

The negative exponents on our relationship is the expected signature of an older, dynamically heated population - this also tracks with our positive τ - σ relationship as metal-poor stars are typically older. This can be tied to observed trends in metallicity gradient declining with increasing scale height $|z|$ (MW simulations give a rate of $\sim 0.18 \text{dex kpc}^{-1}$ - [Loebman et al. \(2011\)](#)), which indicates that more metal-poor stars systematically populate higher disk altitudes, in agreement with our physical understanding of the disk boundaries. Our recovered metallicity gradient for our parallax-limited sample (translating to max scale heights on the order of 0.15kpc given trace inflated parallax errors that scatter sources above the expected 0.1kpc maximum) calculated by binning $[M/H]$ and determining vertical distance from galactic plane with longitude conversions from proper motion is found as $\sim 0.343 \text{dex kpc}^{-1}$. This is consistent with a majority thin-disk population, which is expected given our sample, since older metal-poor stars usually occupy larger $|z|$ values, halo/thick-disk contaminants which may steepen or flatten the gradient are filtered out. Again, this is in quantitative agreement with previous measurements of the MW gradient with SDSS spectra - $\sim 0.3 \text{dex kpc}^{-1}$ ([Ivezic et al., 2008](#)) - and around 0.2-0.4dex kpc $^{-1}$ from RAVE mock samples ([Boeche, C. et al., 2014](#)).

4.6. Metallicity Distribution

4.6.1 Age-Metallicity Relation

We now have a good combination of σ - τ - $[M/H]$ relationships to provide a powerful diagnostic to distinguish the stellar population and see if the outcome is expected given our data set. We can also take steps to determine whether our Bayesian ages and metallicities can be considered 'good' or reliable given what we know about stellar parameter distributions in the local Snhd.

Figure 18 (left) shows our predicted trend of age and metallicity, broken up into distinct age

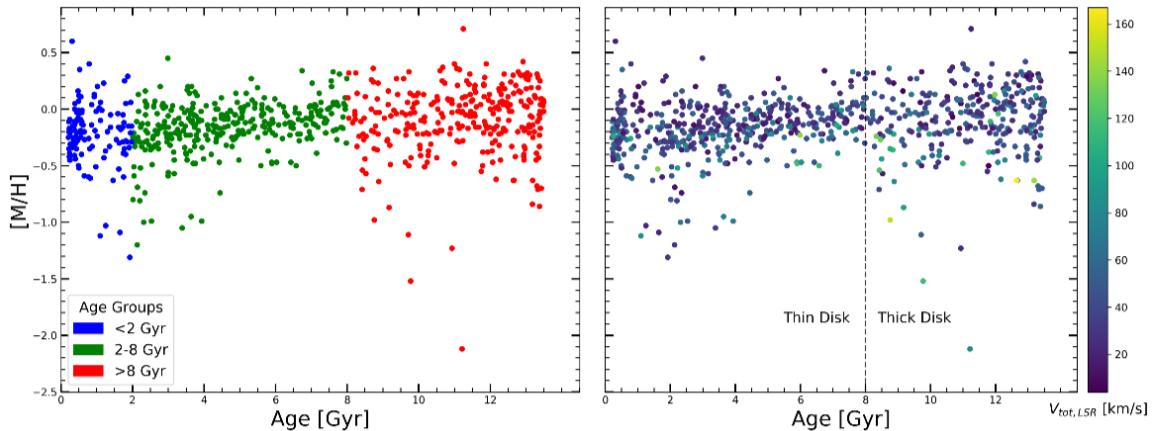


Figure 18: *Left:* Age-metallicity relation for our Bayesian-derived sample. Stars are divided into distinct age groups denoting young ($0 < \tau < 2$ Gyr), intermediate ($2 < \tau < 8$ Gyr), and old (> 8 Gyr) populations. *Right:* Same age-metallicity relation but stars are coloured by V_{tot} in the LSR. The black dotted line demarcates our arbitrary age boundary separating the thick- and thin-disks discussed in Section 4.4.

groups that one could establish as young, intermediate, and old populations. There exists a slight increased $[M/H]$ spread with increasing age, particularly for older stars - reflecting the thick-disk and halo population contamination - which is in line with literature predictions for reliable age estimations (see Casagrande et al. (2011), Figure 16 where a similar spread is shown for stars with well-determined ages). Although we observe a small contingent of old, metal-rich ($\tau > 10$ Gyr, $[M/H] > 0.1$ or so) stars, the spread in metallicity and the lack of dense age clustering suggest the avoidance of a so-called ‘terminal age bias’. This bias refers to the systematic overestimation of stellar ages near the region of the HRD for which the $\log(L/L_\odot)$ - $\log T_{\text{eff}}$ relationship changes slope sharply at the end (or terminus) age of the main sequence (TAMS)(Pont & Eyer, 2004). Small errors in supplementary photometric or spectroscopic parameters may cause a star to be mapped to an isochrone of much greater age because of how tightly packed they are in this area, see Figure 3, and hence the likelihood dominates to a skewed age posterior. Consequently strong excesses of old metal-rich stars (that are actually mis-assigned young metal-rich stars) and artificially higher densities near the TAMS should be observed if our Bayesian estimates were biased this way; this is not manifested in any significant way for our observed age-metallicity relation.

Figure 18 (*right*) shows the same age-metallicity relationship but coloured with total LSR velocities. These velocities seem largely consistent across the plot, which isn’t unexpected as individual kinematics are inherently more stochastic compared to the population-wide dispersions, which we have seen to increase with age and decrease with metallicity. We caution that individual Bayesian-derived uncertainties, particularly in age, are substantial (particularly on the LMS before the TAMS, again see Figure 8) and often asymmetric - potentially scattering stars across age-[M/H] bins and diluting clarity in source-by-source kinematic trends. While the $\sigma-\tau$ -[M/H] relations emerge clearly in aggregate statistics, arbitrary scatter in individual $V_{tot,LSR}$ values across the age-metallicity plane is expected due to the width and shape of posterior distributions. This is not necessarily a pitfall of the Bayesian scheme in this context, as the intrinsic age-[M/H] trends it predicts are palatable, and the overall binned kinematic relationships, which are more of interest, show reasonable conformity with the known physics. Individual stellar velocities, however, are only really useful for segregating stellar populations on a case-by-case level, e.g. determining halo stars with $V_{tot,LSR} > 150 \text{ km s}^{-1}$, which evidently do not occur often given our parallax cut.

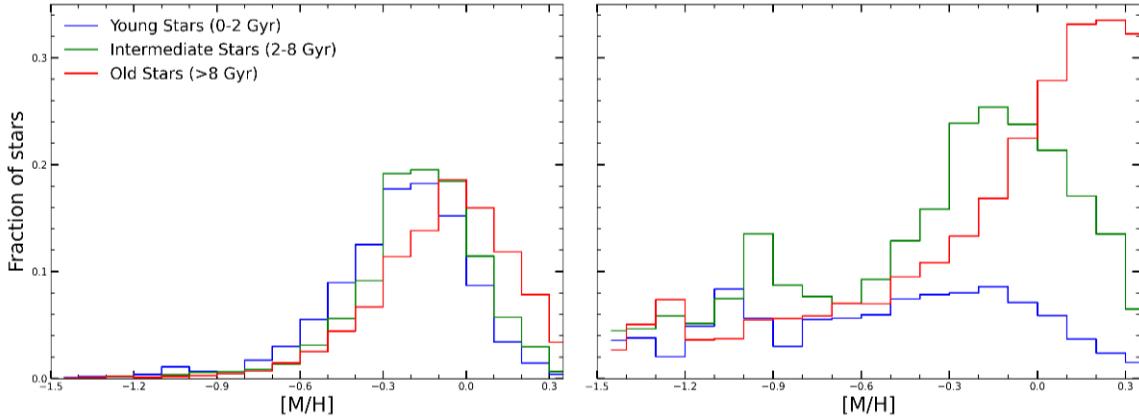


Figure 19: *Left:* Separately normalised MDF for stars belonging to our *Gaia* sample within our three age groups, *right:* MDF relative to each age group, i.e. what fraction of all sample stars at a given metallicity bin come have a particular age.

4.6.2 Metallicity Distribution Functions

Keeping our stellar sample segregated between crude estimates of young, intermediate, and old populations, we can begin to explore the metallicity distribution function (MDF) to assess how these age groups form the chemical structure of the local Snhd, particularly the thin disk. We can take both individual normalised MDFs for each age population along with the fractional contribution of each age group at a given $[M/H]$ as diagnostics for the age-metallicity relation, as well as the plots in Figure 18, to further test the consistency of our Bayesian estimates. The increased chemical broadening with age implies that old stars - although sparse within our parallax-limited local volume - are important in interpreting the wings of the MDF, particularly individual age group contributions to chemical makeup. Figure 19 presents the MDFs for the total stellar sample, both in general and also relative to each age grouping. The first impression is that of a relative overdensity of older stars with metal-rich compositions compared to younger stars at the same $[M/H]$ - a concept directly in contention with the known makeup of local stellar parameter distributions. This apparent excess is likely influenced by $\tau \cdot [M/H]$ degeneracies from the isochrone fitting scheme, where stars with weakly constrained parameters may be misclassified as *both* older and more metal-rich than their true values. Such biases can artificially boost the metal-rich tail of this older population. Another natural explanation would be that this is due to the radial migration of stars (Sellwood & Binney, 2002) where the Snhd (which our *Gaia* sample is encompassed by) is considered to be assembled from not only regional stars, but also stars originating from the inner (metal-rich) and outer (metal-poor) Galactic disk that have migrated radially inwards due to gains/losses in angular momentum from gravitational interaction with non-axisymmetric structure (Grand et al., 2015) at different timescales (Roškar et al., 2008; Schönrich & Binney, 2009). Observational evidence (e.g. Casagrande et al. (2011); Haywood, Misha et al. (2013)) shows that older stars with super-solar metallicities are common in the Snhd despite the local ISM generally housing near-solar stars over all age populations. Simulations from e.g. Minchev, I. et al. (2013); Frankel et al. (2020) reproduce such trends pertaining to local old, metal-rich stars arriving via "churning" (changes in guiding radius without heating) as the primary mechanism. As migration is a cumulative process with time, i.e. older stars naturally have more time to migrate, an up-tick in age at the metal-rich tail could potentially be expected. Through this logic one could also expect a similar increase of old stars at the metal-poor wing if the stars originated at the outer disk, but evidence of inner-disk inflow dominating due to higher stellar density (Casagrande et al., 2011; Grand et al., 2012) biases the exchange towards an excess of metal-rich migrators to some degree.

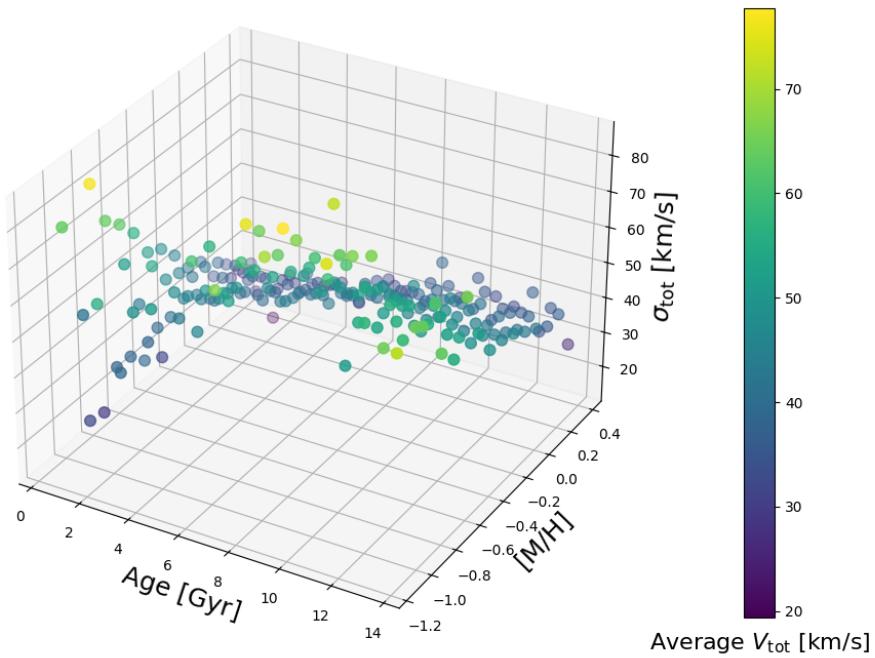


Figure 20: 3-Dimensional scatter plot of binned age vs metallicity [M/H] vs total velocity dispersion σ_{total} , coloured by the average 3-D total velocity vector for stars within that bin.

4.6.3 3-D Relationships

To further disentangle trends in age-[M/H]-dispersion space, we examine the joint distribution, which offers an independent kinematic interpretation of the MDF asymmetries revealed by our Bayesian age posteriors. Our interpretation of this space is further informed by the age-[M/H]-dispersion trend in Figure 20 where individual bins in 3-D space are coloured by mean total velocity. As expected given previous trends, our parallax-limited sample contains very few genuinely old, metal-poor stars - at least not enough to influence the bins; individual outliers are averaged away. The bulk of the metal-poor population appears comparatively young with relatively low kinematic heating. In contrast, metal-rich stars seem to span the full age range, including a family of dynamically hot, old stars which are consistent with the prerequisites of inner-disk migrators.

The absence of the old, metal-poor tail in these averaged trends may reflect on the potential scarcity of outer-disk migrators in this local volume and the survey selection effects that oppose the kinematically hottest, metal-poor thick-disk and halo stars. Bayesian posterior estimates are important here, as they facilitate the tracking of population gradients in both chemistry and kinematics whilst still propagating age uncertainties. This could reveal that the chemical-kinematic coupling is highly asymmetric between the metal-poor and -rich tails of our MDF.

4.6.4 Context within the Solar Neighbourhood

Although we can justify this old-metal-rich skewed result to some extent with the above hypotheses, a relative peak in super-solar metallicities is unexpected. Ages may not be well constrained for the older population; many stars classified as old may be misclassified intermediate-age stars scattered up due to wide posterior tails. The aforementioned isochrone degeneracies can also lead to misinformation, saturating Bayesian metallicities to the higher end. These factors can artificially broaden the age–metallicity distribution, making the wings of the MDF appear more populated than they truly are.

Therefore, while radial migration provides a plausible explanation for the observed overabundance of metal-rich old stars, we cannot exclude the fact that these features arise partly from methodological limitations and selection effects.

Separately, the age-relative MDFs for the other populations broadly follow the expected forms. A mostly flat MDF for the young population conveys a roughly equal metallicity representation up to about the solar value. This is a reasonable result as we know the recent ISM is chemically homogenous around the Snhd, with only mild scatter. Young stars with extreme metallicities either side of the solar value are rarer (shown by the Gaussian-like normalised distribution peaking just sub-solar, Figure 19, left) as stars forming in the last 2 Gyrs should be born within this chemical nursery. The drop off above this range is also manageable because, as above, super-metal-rich stars of any age are rare in the Snhd - those that do reside here likely being older and products of radial migration. The flat MDF further suggests that young stars are primarily drawn from the same chemically well-mixed solar-[M/H] ISM which aligns with observational studies with APOGEE data (Hayden et al., 2015).

When normalised within each age group, the MDFs show young and intermediate-age stars peaking near -0.2dex , with intermediate stars slightly skewed toward solar metallicity, while the oldest age bin peaks at approximately solar metallicity with a slightly flatter observed curve. This differs from the canonical expectation that the oldest populations should dominate at sub-solar metallicities (e.g., $\sim-0.5\text{dex}$) because our strict parallax cut ($\varpi > 10\text{mas}$) restricts the sample to a very local, thin-disk-dominated volume. Consequently, metal-poor old stars, which are more common at larger $|z|$ and Galactocentric radii R , are severely under-represented, also due to said parallax cut. These observed peaks per age group are roughly in line with what is found by Casagrande et al. (2011)'s analysis of GCS metallicity distributions, along with the feature of younger stars exhibiting a narrower distribution than older stars. It may also be relevant that our parallax-limited *Gaia* sample likely contains a large fraction of low-mass dwarfs (e.g., M-dwarfs), which dominate the local structure and have long lifetimes. Observational studies of M-dwarf MDFs (e.g., Casagrande et al. (2008); Bonfils et al. (2005)) consistently show a strong peak around solar [M/H] with a relatively narrow spread, reflecting the chemical enrichment of the local disk rather than the broader Galactic metallicity distribution. This naturally biases our normalised MDFs toward solar metallicity for the oldest stars, even when those stars represent long evolutionary timescales, which may help explain the strange overabundance of old stars scattered towards super-solar metallicities that is discussed above.

5. Discussion

Our current Bayesian algorithm described in Section 3 is a simplified version of the method proposed by Schönrich & Bergemann (2014), which does not account for full photometric, spectroscopic, and astrometric observation space coverage, with likelihood calculations inspired by Pont & Eyer (2004). Despite this, we have managed to recover reasonable age and metallicity distributions for our given stellar sample, and their associated kinematic and distance relationships, without these further constraints and contributions to the posterior. The main drawbacks of our chosen approach to this method are evidently manifested as (i) large tails in age uncertainty observed in the posterior, especially for lower MS stars where isochrones are degenerate, and (ii) spurious metallicity predictions, particularly a metal-rich abundance at high ages that we struggle to properly rationalise within the current adopted framework.

In the following section, we recognise the potential of future modifications or extensions to our Bayesian algorithm for inferring posteriors for individual stellar parameters, aiming to address these limitations and to incorporate a more complete observational set of constraints. In particular, we address the treatment of observational uncertainties and likelihood structure in building PDFs with

the goal of expanding our core observational parameter space and the consequent inferences we can draw from them.

5.1. Future Improvements

5.1.1 Photometric Likelihoods

If we want to predict stellar parameters from observations, our Bayesian likelihood function needs to compare model-predicted observables with actual data, as discussed in Section 3.5. Thus far we have only considered isochrones covering age, metallicity, and stellar mass, but we can expand our parameter space using Bayesian photometric grids. Synthetic stellar-model-predicted magnitude grids in, e.g. SDSS bands, spanning model grid points i at different filters k can be defined as $C_{i,k}$ and directly compared to observational magnitudes derived directly in each band k . These are subsidised by estimates weighing in parallax ϖ or stellar distance s , as well as reddening parameter r . Assuming Gaussian errors, we can represent this as something like:

$$P(O_k | C_{i,k}, s, r) = \frac{1}{\sqrt{2\pi} \sigma_k} \exp \left[-\frac{(O_k - [C_{i,k} + \mu(s) + r\hat{e}_k])^2}{2\sigma_k^2} \right] \quad (22)$$

where O_k is the photometric magnitude observation in band k , $\mu(s) = 5\log_{10}(s/10pc)$ is the distance modulus (a magnitude), and σ_k encodes the total uncertainty due to observations, model artefacts, and reddening. Multiplying across all filters gets a total photometric likelihood which represents how multiple photometric bands align to constrain certain stellar parameters. This would supplement our current likelihood techniques, as instead of building a likelihood solely from HRD space, we can compute it directly from observed magnitudes, avoiding projection biases. This would simplify assumptions made in the conversion process (e.g. assuming a blackbody, un-uniform stellar atmospheres) leading to inaccurate luminosities when transforming derived quantities [Pacifici et al. \(2023\)](#). Seamlessly incorporating distance and reddening as additional free parameters that simultaneously influence all bands also improves constraints. This works with whatever survey you choose, i.e. *Gaia* with two bands (BP + RP) or SDSS with multiple bands (ugriz + infrared).

In the context of our parameter estimation, adopting photometry across multiple bands constrains our stars' T_{eff} , luminosity, and colour, which indirectly form constraints on allowed age-metallicity combinations consistent with our other data. For example, broad/narrow-band colours may help break degeneracies between hot metal-poor and cold metal-rich stars, and colour matching MS turn-off and giant branch stars on the CMD will help determine ages at a fixed luminosity. For our local Snhd sample, reddening r is mostly negligible but it is important to include if we expand to larger surveys. Also, the CMD position changes only subtly over time so this new likelihood alone cannot constrain age in this region, mainly informing metallicity, thus reiterating the need to employ this method in conjunction with already established likelihoods.

5.1.2 Spectroscopic Likelihoods

We already employ spectroscopic data throughout this work, mainly to constrain metallicity, but we are far from obtaining the idealised end goal of incorporating *all* available spectroscopic information. We mainly employ T_{eff} , $\log g$, and luminosity, but largely ignore intelligence such as the abundance of specific elements beyond just [M/H], spectral type, and rotation. Hence it is clear that a fully fleshed out spectrum gives the observer the highest potential in constraining a star's physical properties, complementary to its photometry and astrometry. Creating simulated spectral grids as surrogates for our evolutionary stellar models or isochrones to compare with our observed spectra may

also act as a likelihood function to constrain parameters further. Obtaining a spectroscopic likelihood at every point in parameter space would involve evaluating the same goodness-of-fit statistic χ^2 used in Equation 8 at each pixel i of the observed spectrum, comparing the template and observed spectra, respectively.

Extracting this information accurately is limited, however, as modelling spectra is complex due to unphysical approximations and incomplete data. Atmospheric spectra models are routinely interpreted using 1-D model atmospheres which assume plane-parallel or spherical symmetry, which is inexact at best. Physical stars display 3-D convective motions and atmospheric inhomogeneities that scatter line formation in the emission spectra. The simplified 1-D codes simulating large spectral grids simplify these problems but cannot capture the full complexity of stellar behaviour, propagating systematic errors in line shaping into our inferred parameters. 3-D non-LTE (local thermal equilibrium; the level atom and ion populations in our stellar atmospheres are set by "local" densities and thermal gradients only) codes have been established to produce such grid spectra (Bergemann et al., 2012), but the high dimensionality involved with integrating over every opacity and wavelength makes computing separate grids for separate compositions difficult. The non-LTE constraints of these grids are also limiting as non-local processes within the radiative and convective fields quickly divert stellar atmospheres from LTE, especially in metal-poor stars, leading to biased abundance determinations and stellar parameters if left ignored. Due to the nature of these model spectra approximations, this spectroscopic likelihood is thus an imperfect representation of a true observation probability.

Furthermore, although code for calculating synthetic spectra has been adopted with success in contemporary work (SIU code (Reetz, 1999), Turbospectrum (Plez, 2012), and SYNTHE (Kurucz, 2005; Munari et al., 2005)), the complexity of combining all types of spectral information with others in this Bayesian framework is beyond the scope of this work.

5.1.3 Astrometric Likelihoods

Similarly, we can use the full astrometric coverage in our likelihoods to constrain parameters even further. An advantage of the *Gaia* survey is the unprecedented precision and quantity of parallax data; thus, it is natural to consider these measurements within the observational likelihood function. Parallax measurements only affect distance s and distance modulus μ - i.e., there are no biases introduced with respect to contaminating other parameter measurements - so it is relatively easy to combine this new likelihood with already established photometric and spectroscopic information. *Gaia* reports a measured parallax ϖ_0 with associated uncertainty σ_ϖ . A principled way to inherit this information into Bayesian inference would be to again assume Gaussian error and write the likelihood in parallax space:

$$P(\varpi_0 | \varpi) = \frac{1}{\sqrt{2\pi} \sigma_\varpi} \exp \left[-\frac{(\varpi_0 - \varpi)^2}{2\sigma_\varpi^2} \right] \quad (23)$$

where ϖ is the *true* parallax implied by a model incorporating its distance. We could also parameterise the model by distance s in pc, but we must then inform the posterior with a distance prior $P(s)$ (Bailer-Jones, 2015), as we cannot infer s exactly due to the noise involved. We cannot simply invert ϖ_0 to get distance because the reciprocal is a biased non-Gaussian estimator (probability intervals are equally sized in $1/s$ because the Gaussian is symmetric, but not in s) and because negative or low signal-to-noise parallax measurements still encode information that must be folded into the posterior. As the Gaussian has infinite support for all parallax values and uncertainties, i.e., $\varpi > 0$ everywhere, there will always be some finite portion of its probability mass arbitrarily close to 0, corresponding to undefined or large distances. Hence, the observational likelihood in parallax space will always assign some probability to regions that map to physically non-sensical distances. Thus, we must be careful with reparameterisation - see Section 3.1.3 - and choose necessary physically motivated distance priors

(Figure 1). For more, see [Luri et al. \(2018\)](#); [Bailer-Jones et al. \(2018\)](#).

The full *Gaia* astrometric solution to the likelihood should take into account proper motion, thus becoming multivariate. Establishing an observed astrometric vector $[\varpi_0, \mu_{\alpha,0}, \mu_{\delta,0}]$ and comparing with a predicted model vector helps us expand into a fully realised likelihood that covers the entire astrometric spectrum. Supplementing our current methods with this extra information improves our constraints on multiple parameters. Astrometry generally constrains absolute magnitude when combined with the full photometry, and proper motions, when coupled with distance, constrain tangential velocities. When constructing a total manifestation of Bayes' theorem evaluated in observable space, each likelihood term should keep to its appropriate data type. Combining parallax in the same framework would couple distance to luminosity and consequently give better age constraints (via magnitude and isochrone position) than photometry alone, especially near the turn-off.

5.1.4 Degeneracies and Bias

Throughout this work, we have discussed how and where degeneracies occur when competing combinations of variables produce similar observables (we have seen e.g. age-metallicity, LMS, MS turn-off). Biases that artificially boost inference towards specific areas of parameter space are also intrinsically introduced every time we build our sample with selection cuts, or through imperfect likelihood modelling assumptions. In our Bayesian framework, this can distort the posterior PDFs, so we turn our attention to potential improvements in which we can model them explicitly.

An obvious degeneracy is observed in our treatment of intrinsically multi-modal age PDFs (see Figure 5) within certain evolutionary phases such as the subgiant "hook" or red clump [Haywood, Misha et al. \(2013\)](#). Currently, our single-Gaussian posterior summaries mask this structure by collapsing the more pronounced peak to a single average age. A potential solution would be to fit mixture distributions ([Ghojogh et al., 2020](#)) (probably two Gaussian mixtures in our case) to the posteriors, which would model the weighted contribution of each peak and the heterogeneity of the data set while retaining the multimodality for uncertainty propagations. Previous astrophysical applications for reference would be [Jones & Heavens \(2019\)](#) application to distinguishing multi-modal redshift likelihoods, and the Bayesian probabilistic assignment of pulsar candidates using Gaussian mixture model algorithms by [Lee et al. \(2012\)](#).

A source of bias not yet considered is unresolved binaries that contaminate the data set. These will shift observed colours and magnitudes, manufacturing bluer/redder stars, biasing age and metallicity estimates. Again, we could apply mixture models that separate these population distributions into singles and binaries, with binary fraction or mass ratio initialised as hyperparameters. Additionally, in our familiar Bayesian form, we could marginalise over possible binary contributions determined by a nuisance parameter flag in our observational likelihood equation (see Equation 13) rather than just ignore them. [Pont & Eyer \(2004\)](#) is an early example of accounting for binaries in the Bayesian age determination context.

Survey selection function biases (e.g. CMD or HRD cuts, Section 4.3) can also distort inferred parameter posterior distributions. Forward modelling these effects, i.e. accounting for them in the likelihood function, will ensure the posterior reflects the observed sample rather than an idealised parameter-space-volume limited population. This may be as simple as applying the selection function or sample cut-offs to the simulated stars from our isochrones before likelihood calculation so that they do not weigh into posterior inference with unwanted regions of parameter space.

5.1.5 Additional Advancements

We have mentioned expanding the PDF calculations to account for the full photometric (including reddening), astrometric, and spectroscopic solutions; we will now briefly outline some additional

advancements that should be considered for future expansions.

It would be valuable to test the Bayesian framework on well-studied stellar clusters that provide mono-age and -metallicity populations. Open/globular clusters serve to validate age and metallicity posterior recovery under controlled conditions, free from the broader spread and inherent degeneracies experienced by field stars seen throughout this work. Extensive comparison against high-quality *Gaia* benchmark clusters ([Bossini, D. et al., 2019](#)) or spectroscopically calibrated clusters (e.g. APOGEE [Donor et al. \(2020\)](#), or *Gaia*-ESO [Randich et al. \(2018\)](#)) would offer important checks on accuracy, precision, and potential systematics of our Bayesian method.

Similarly, we could check our inferred parameter distributions calibrated against other stellar surveys, e.g. APOGEE ([Jönsson et al., 2020](#)), instead of just our chosen *Gaia* sample. APOGEE provides high-resolution NIR spectra across Galactic locations, including the inner disk and bulge, making it ideal for testing how our method generalises beyond the local Snhd. Comparing the resulting age-metallicity and kinematic trends across these surveys could reveal whether systematic differences arise from selection effects, chosen wavelength spectrum, or model assumptions, or if the observed age-metallicity degeneracy and predicted overdensity of old, metal-rich stars is intrinsic to our sample.

In addition to our AVR and metallicity-kinematic relations, further diagnostics to verify our distributions against known trends can be introduced. For example, the shape/orientation of the full velocity ellipsoid instead of just compartmentalised singular dispersions could be analysed. Observing anisotropies within tilt angles or axis-ratios compared with known dynamics and expectations from thick/thin disk behaviour could signify incorrect age/metallicity estimates. Different action (J_R , J_z , L_z) distributions for all age/metallicity bins could also be compared; inconsistent results like young stars with halo-like orbits could flag problematic estimates as well.

Finally, future work could also explore the implementation of self-informed priors (as discussed in Section 3.1.1), where initial results are used to iteratively refine our prior $P(\tau, [\text{M}/\text{H}])$. Although we are confident that our chosen prior reflects the population under investigation, such methodology can help reduce bias and increase posterior accuracy in regions of parameter space where degeneracy is problematic, or where they are weakly constrained.

5.2. Mapping the Solar Neighbourhood

With the above methodological improvements accounted for, our algorithm is in good shape to map the local Snhd. Bayesian parameter determination opens doors to a richer, more precise drawing of the stellar catalogue we observe around us. In addition to tracing the kinematic relationships of sources at different inferred ages and metallicities, there are numerous ways that we can refine our study.

5.2.1 Refined Population Separation

As we have seen in Section 4.4, many Galactic studies using traditional methods use hard cuts in kinematics or chemistry (e.g. [Bensby et al. \(2003\)](#); [Adibekyan, V. Zh. et al. \(2011\)](#)) to separate thin/thick/halo stars. In overlapping regions of parameter space, this may lead to an up-tick in star stellar misclassification. An advantage of using Bayesian posteriors is that they provide continuous parameter probabilities for each population, absorbing uncertainties, and allowing stars to be superposed across multiple components ([Rix & Bovy, 2013](#)).

Calculating age-metallicity- α/Fe -kinematic joint posteriors to assign probabilities of population-family would be the first step in streamlining this inference. This would allow for the avoidance of harsh/steep artificial boundaries in $[\alpha/\text{Fe}]$ - $[\text{Fe}/\text{H}]$ space ([Hayden et al. \(2015\)](#)), as well as the mapping of "transition stars" in various parameter space iterations that may assist in helping to understand the

MW Galactic archaeology. Creating spatial maps in the context of the above posteriors weighted by the population membership probabilities ([Bland-Hawthorn & Gerhard, 2016](#)) would similarly reveal smoother transitions rather than sharp breaks like those we have adopted in this work. An additional comparison with APOGEE or SDSS MDFs and abundance gradients (e.g, [Anders, F. et al. \(2014\)](#)) would be needed to supplement this work.

We identify stellar populations in Section 4.6.2 whose kinematics are inconsistent with their postulated chemistries, e.g. old metal-rich stars on cold thin-disk orbits within the Snhd. Classifying the probabilities that these stars originate from other parts of the disk via radial migration would be critical to differentiate populations further. The methodology of this could follow from e.g. [Schönrich & Binney \(2009\)](#)'s chemodynamical radial migration model, or [Wojno et al. \(2016\)](#)'s identification of potential migrators from RAVE data, and the extent to which they influence dynamical trends within the Galactic disks. Using Bayesian age estimations may help to put migration timescales on these stars given their chemical abundance observational data is high enough resolution.

Again, we confirm Bayesian parameter estimation as a reliable approach to combine all available observational constraints into a unified framework that ensures properly classified populations. The approach not only reduces bias from arbitrary selection thresholds, but also preserves information about ambiguous transitional objects that are fundamental for testing Galactic evolution models.

5.2.2 Disk Formation History

Disk formation models predict radial metallicity gradients that evolve over time, which are steep for younger stars and plateau for older stars ([Nordström et al., 2004](#); [Minchev, I. et al., 2013](#); [Frankel et al., 2020](#)). Similarly, vertical metallicity gradients are linked to the disk transition, with gradients steepening out at greater scale heights $|z|$, linked to early turbulence or heating events in the disk ([Haywood, Misha et al., 2013](#); [Bland-Hawthorn & Gerhard, 2016](#)). Bayesian estimates on age, metallicity, and distance would avoid bias from hard binning and supply a full uncertainty propagation when inferring said gradients. Using our marginalised posteriors would handle asymmetric errors and degeneracies, especially where age uncertainties are large, thus justifying the Bayesian methodology. Our current parallax cut on our *Gaia* sample means we are sampling mainly nearby disk stars, truncating the observed spatial coordinate parameters, adding an artificial flattener to the gradients. Despite this bias, we can still measure the slope of the local gradients and how they act as a function of age; posterior-weighted age bins could potentially reveal intermediate-age populations with flatter gradients than young stars, thereby flagging early migration.

Extending our *Gaia* astrometry with more reliable spectroscopic surveys, e.g. APOGEE or LAMOST, and increasing the range of our survey outside our $\sim 100\text{pc}$ range, we can expand the range in R and z , improving metallicity precision and allowing for a joint posterior gradient fit that covers the full disk, not just the very local neighbourhood.

5.2.3 Discovering Substructure

The age-metallicity-kinematic combination is the most discriminating space for isolating moving groups, streams, and accreted debris within the MW. The *Gaia* DR2 demonstrated how chemodynamical information can reveal such rich substructures, for example the *Gaia*-Enceladus debris object ([Helmi et al., 2018](#)), or resonant features in the local velocity distribution showing a young perturbed disk ([Antoja et al., 2018](#)). Adopting Bayesian parameter posterior samples rather than point estimates to constrain patterns that emerge from this structure preserves measurement uncertainties that would otherwise erase real structure signifiers.

Probabilistic clustering methods, such as HDBSCAN ([McInnes & Healy, 2017](#)) or previously mentioned GMMs ([Bovy et al., 2011](#)), are useful in incorporating these uncertainties, returning prob-

abilistic distributions rather than point assignments. Integral-of-motion/action-space clustering, e.g. Trick (2019), is a further application that enhances the clarity of dynamical group contrast.

Such approaches could also be applied to our observed population of old ($>8\text{Gyr}$), metal-rich ($[\text{M}/\text{H}] > \sim 0.2\text{dex}$) stars in the Snhd. Posterior-based chemical, age, and orbital properties may be compared to chemodynamical models to distinguish between the competing scenarios (Frankel et al., 2018) presented in Section 4.6.2 and the rest of this work. Hence, this Bayesian framework again presents a potential application outside the scope of this work - a principled avenue to uncover and physically interpret substructure in the MW.

6. Conclusion

In this research thesis, we have analysed the applications and major contributions of Bayesian statistics in the field of astrophysical stellar parameter determination, and conducted a modest physical implementation for a local volume of stars quantified by the *Gaia* astrometry mission (Gaia Collaboration et al., 2023), successfully showing how one could operate such mathematical methods to constrain interest parameters on large scales. We specifically focused on the algorithms proposed by Schönrich & Bergemann (2014), subsidised by isochrone-based age determination methods and likelihood function construction from Pont & Eyer (2004) and Jørgensen & Lindegren (2005), in order to build our own models that predict posterior probability distribution functions of age and metallicity for stars within the local Snhd. Using these models, we examined these PDFs with an emphasis on how the prior distribution, stellar model constraints, and likelihood function forms influence the shape of the posterior, and how biases, degeneracies, and regions of unconstrained parameter inference manifest in their uncertainty tails.

The sample of stars with computed Bayesian ages and metallicities using the above methods allowed us to explore various combinations of known parameter distributions of age, chemistry and provided *Gaia* kinematics and astrometry. Recreating such distributions and measuring them against the consensus forms given in the literature gave us confirmation diagnostics on the validity and accuracy of our adopted Bayesian method of constraining age and metallicity spreads, at least within the context of local stellar populations. We successfully recreated a reasonable age-velocity dispersion relation (AVR) which takes the form of a power-law relation $\sigma = \sigma_0 \cdot \tau^\beta$, with $\beta = 0.29 \pm 0.01$. This value is in accordance with historically predicted estimates by e.g. Lacey (1984) and Binney et al. (2000), and falls within an accepted range found in more contemporary observational studies (Holmberg et al., 2009). A β of ~ 0.3 is consistent with the gradual secular omnidirectional heating of the disk by spiral arms and other stochastic perturbations. We observed a systematic steeper power-law increase in the vertical dispersion σ_W , compared to radial or azimuthal components, consistent with the idea that vertical heating is dominated by GMCs and acts more efficiently over time than in-plane heating sources Aumer et al. (2016). We find a flattening of the AVR at the oldest ages, reflecting a saturation of these heating processes, as verified in Casagrande et al. (2011) for example. This is likely compounded by the transition to the kinematically hotter thick-disk population, whose stars were born hot rather than acquiring their velocities through long-term heating. The various proposed methods (and their justifications) to distinguish population stars between the thick/thin-disk and halo components were also explored, using our Bayesian-determined parameters as our baseline to some degree of success.

We also examined dispersion as a function of our metallicities, where the kinematic evolution with age is mirrored. We found negative exponents of -0.29 ± 0.02 and -0.19 ± 0.01 , for total and vertical dispersion, respectively. These trends are expected because metal-poor stars belong predominantly to older, dynamically hotter populations, and vice versa. The shallower vertical slope suggests that

heating is less strongly coupled to metallicity in this direction than in-plane heating. In terms of metallicity, we also recover suitable MDFs and age-metallicity relationships. The latter shows an expected spread of chemical compositions at older ages, which reflects potential thick-disk and halo contamination, as well as radial migrators that populate the old, metal-rich corner of parameter space, travelling from the inner disk. The former is in agreement with this, and studies such as Casagrande et al. (2011), where the metallicity peaks of all age populations skew slightly sub-solar, with the older generation confirming this metal-rich trend. The fact that our Bayesian metallicity estimates reproduce these established correlations lends confidence to both the interpretability of our inference method and its ability to recover physically meaningful Galactic trends around the local Snhd.

While our Bayesian framework offers sensible constraints on age and metallicity distributions, with uncertainty propagation and prior knowledge incorporation, its practical application is not without caveats. The major flags for significant limitations raised in this work lie in the large observed age uncertainties for stars (particularly on the lower MS) where isochrones of vastly different ages converge in $T_{\text{eff}}-L$ space, resulting in inherently broader age posteriors (Soderblom, 2010). This degeneracy limits our ability to maintain reliable age estimates of cooler dwarfs and dilutes population-level trends. On top of that, our model tends to overpredict the fraction of old, metal-rich stars compared to previous observational trends (e.g. Casagrande et al. (2011); Haywood, Misha et al. (2013)). Although the existence of this population in the Snhd is well documented, their prevalence in our sample is slightly puzzling and, as discussed above, has been mainly put down to radial-migration trends.

In a future study, addressing these limitations will require further refinement of prior assignments, incorporation of additional independent extension of astrometric, spectroscopic, and photometric constraints on the likelihood, and careful individual treatment of stars in poorly constrained regions of parameter space. Furthermore, cross-validation with independent age-dating techniques, such as asteroseismology for evolved stars (Chaplin & Miglio, 2013; Aguirre et al., 2017), or gyrochronology for dwarfs (Barnes, 2007), would provide external checks on the quality of the posteriors. Incorporating α -abundances into the likelihood could also help disentangle thick- and thin-disk populations and further clarify age-metallicity-kinematic correlations. In addition, testing our model against mock catalogues made in chemodynamical simulations (e.g. Minchev, I. et al. (2013); Frankel et al. (2020)) would help pinpoint bias introduction within the inference pipeline and quantify how efficiently different populations in the Snhd are recovered. This would not only strengthen the credibility of our derived results, but would also enable more precise connections to be made between observed trends and the physical mechanisms shaping the disk.

7. Bibliography

- Adibekyan, V. Zh. Santos, N. C. Sousa, S. G. Israelian, G. 2011, *A&A*, 535, L11
 Aguirre V. S., et al., 2017, *The Astrophysical Journal*, 835, 173
 Anders, F. et al., 2014, *A&A*, 564, A115
 Andrade J., Omey E., 2013, *Annals of the Institute of Statistical Mathematics*, 65
 Antoja T., et al., 2018, , 561, 360
 Aumer M., Binney J. J., 2009, , 397, 1286
 Aumer M., Binney J., Schönrich R., 2016, *Monthly Notices of the Royal Astronomical Society*, 462, 1697–1713
 Baade W., 1944, , 100, 137
 Babusiaux C., et al., 2018, *Astronomy & Astrophysics*, 616, A10
 Bailer-Jones C. A. L., 2011, , 411, 435
 Bailer-Jones C. A. L., 2015, , 127, 994
 Bailer-Jones C. A. L., 2017, Practical Bayesian Inference: A Primer for Physical Scientists. Cambridge University Press

- Bailer-Jones C. A. L., Rybizki J., Fouesneau M., Mantelet G., Andrae R., 2018, , [156, 58](#)
- Bailyn C. D., Jain R. K., Coppi P., Orosz J. A., 1998, , [499, 367](#)
- Barnes S. A., 2007, , [669, 1167](#)
- Bensby T., Feltzing S., Lundström I., 2003, , [410, 527](#)
- Bergemann M., Lind K., Collet R., Magic Z., Asplund M., 2012, , [427, 27](#)
- Bertelli G., Bressan A., Chiosi C., Fagotto F., Nasi E., 1994, , [106, 275](#)
- Bevington P. R., Robinson D. K., 2003, Data reduction and error analysis for the physical sciences
- Binney J., Dehnen W., Bertelli G., 2000, *Monthly Notices of the Royal Astronomical Society*, 318, 658
- Bland-Hawthorn J., Gerhard O., 2016, , [54, 529](#)
- Bochanski J. J., West A. A., Hawley S. L., Covey K. R., 2007, *The Astronomical Journal*, 133, 531
- Boeche, C. et al., 2014, [A&A](#), 568, A71
- Bonaca A., Conroy C., Wetzel A., Hopkins P. F., Kereš D., 2017, *The Astrophysical Journal*, 845, 101
- Bonfils X., Delfosse X., Udry S., Santos N. C., Forveille T., Ségransan D., 2005, , [442, 635](#)
- Bossini, D. et al., 2019, [A&A](#), 623, A108
- Bovy J., et al., 2011, , [729, 141](#)
- Bovy J., et al., 2012, *The Astrophysical Journal*, 759, 131
- Brown D. J. A., 2014, *Monthly Notices of the Royal Astronomical Society*, 442, 1844–1862
- Bunn E. F., Sugiyama N., 1995, , [446, 49](#)
- Burnett B., Binney J., 2010, , [407, 339](#)
- Buser R., Rong J., Karaali S., 1999, , [348, 98](#)
- Calamida A., et al., 2007, , [670, 400](#)
- Carlberg R. G., Sellwood J. A., 1985, , [292, 79](#)
- Casagrande L., Flynn C., Bessell M., 2008, , [389, 585](#)
- Casagrande L., Schönrich R., Asplund M., Cassisi S., Ramírez I., Meléndez J., Bensby T., Feltzing S., 2011, , [530, A138](#)
- Chaboyer B., Demarque P., Kernan P. J., 1996, *Science*, 271
- Chaplin W. J., Miglio A., 2013, , [51, 353](#)
- Chen D.-C., et al., 2021, *The Astrophysical Journal*, 909, 115
- Chiba M., Beers T. C., 2000, , [119, 2843](#)
- Dehnen W., 1998, , [115, 2384](#)
- Donor J., et al., 2020, , [159, 199](#)
- Drell P. S., Loredo T. J., Wasserman I., 2000, , [530, 593](#)
- Eadie G. M., Speagle J. S., Cisewski-Kehe J., Foreman-Mackey D., Huppenkothen D., Jones D. E., Springford A., Tak H., 2023, [arXiv e-prints](#), p. [arXiv:2302.04703](#)
- Eggen O. J., Lynden-Bell D., Sandage A. R., 1962, , [136, 748](#)
- Esa ., 1997, VizieR Online Data Catalog: The Hipparcos and Tycho Catalogues (ESA 1997), VizieR On-line Data Catalog: I/239. Originally published in: 1997HIP...C.....0E
- European Space Agency 2024, Gaia Mission: ESA Science Technology, <https://sci.esa.int/web/gaia>
- Faure C., Siebert A., Famaey B., 2014, *Monthly Notices of the Royal Astronomical Society*, 440, 2564
- Feigelson E. D., Babu G. J., 2012, Statistical Methods for Astronomy ([arXiv:1205.2064](#)), <https://arxiv.org/abs/1205.2064>
- Feltzing, S. Bensby, T. Lundström, I. 2003, [A&A](#), 397, L1
- Ferrarese L., Merritt D., 2000, , [539, L9](#)
- Feuillet D. K., Frankel N., Lind K., Frinchaboy P. M., García-Hernández D. A., Lane R. R., Nitschelm C., Roman-Lopes A., 2019, *Monthly Notices of the Royal Astronomical Society*, 489, 1742
- Frankel N., Rix H.-W., Ting Y.-S., Ness M., Hogg D. W., 2018, , [865, 96](#)
- Frankel N., Sanders J., Ting () Y.-S., Rix H.-W., 2020, *The Astrophysical Journal*, 896, 15
- Fuhrmann K., 1998, , [338, 161](#)
- Gaia Collaboration et al., 2023, , [674, A1](#)
- Ghojogh B., Ghojogh A., Crowley M., Karray F., 2020, Fitting A Mixture Distribution to Data: Tutorial ([arXiv:1901.06708](#)), <https://arxiv.org/abs/1901.06708>

- Grand R. J. J., Kawata D., Cropper M., 2012, , [426, 167](#)
- Grand R. J. J., Kawata D., Cropper M., 2015, , [447, 4018](#)
- Gregory P. C., Loredo T. J., 1992, , [398, 146](#)
- Harwit M., 1988, *Astrophysical Concepts*. Astronomy and Astrophysics Library, Springer New York, <https://books.google.co.uk/books?id=wtFUAAAAYAAJ>
- Hawkins K., Jofré P., Masseron T., Gilmore G., 2015, *Monthly Notices of the Royal Astronomical Society*, 453, 758
- Hayden M. R., et al., 2015, , [808, 132](#)
- Haywood, Misha Di Matteo, Paola Lehnert, Matthew D. Katz, David Gómez, Ana 2013, [A&A, 560, A109](#)
- Helmi A., Babusiaux C., Koppelman H. H., Massari D., Veljanoski J., Brown A. G. A., 2018, , [563, 85](#)
- Hernandez X., Valls-Gabaud D., Gilmore G., 2000a, , [316, 605](#)
- Hernandez X., Gilmore G., Valls-Gabaud D., 2000b, , [317, 831](#)
- Hogg D. W., Bovy J., Lang D., 2010, Data analysis recipes: Fitting a model to data ([arXiv:1008.4686](https://arxiv.org/abs/1008.4686)), <https://arxiv.org/abs/1008.4686>
- Holmberg J., Nordström B., Andersen J., 2009, , [501, 941](#)
- Hou J. L., Zhong J., Chen L., Yu J. C., Liu C., Deng L. C., 2012, *Proceedings of the International Astronomical Union*, 8, 105–105
- Howes, Louise M. Lindegren, Lennart Feltzing, Sofia Church, Ross P. Bensby, Thomas 2019, [A&A, 622, A27](#)
- Iben I., Renzini A., 1984, , [105, 329](#)
- Ivezic Ž., et al., 2008, , [684, 287](#)
- Jeffery E. J., Hippel T. v., Dyk D. A. v., Stenning D. C., Robinson E., Stein N., Jefferys W. H., 2016, *The Astrophysical Journal*, 828, 79
- Jeffreys H., 1939, *Theory of Probability*. Clarendon Press, Oxford, England
- Jeffreys H., 1946, *Proceedings of the Royal Society of London Series A*, 186, 453
- Johnson D. R. H., Soderblom D. R., 1987, , [93, 864](#)
- Jones D. M., Heavens A. F., 2019, *Monthly Notices of the Royal Astronomical Society*, 490, 3966–3986
- Jönsson H., et al., 2020, , [160, 120](#)
- Jørgensen B. R., Lindegren L., 2005, , [436, 127](#)
- Jullo E., Kneib J. P., Limousin M., Elíasdóttir Á., Marshall P. J., Verdugo T., 2007, *New Journal of Physics*, 9, 447
- Khoperskov, S. Sivkova, E. Saburova, A. Vasiliev, E. Shustov, B. Minchev, I. Walcher, C. J. 2023, [A&A, 671, A56](#)
- Kolmogorov A. N., 1933, Foundations of the Theory of Probability
- Koppelman H., Helmi A., Veljanoski J., 2018, *The Astrophysical Journal Letters*, 860, L11
- Kurucz R. L., 2005, *Memorie della Societa Astronomica Italiana Supplementi*, 8, 14
- Lacey C. G., 1984, , [208, 687](#)
- Lee K. J., Guillemot L., Yue Y. L., Kramer M., Champion D. J., 2012, , [424, 2832](#)
- Liddle A. R., 2007, , [377, L74](#)
- Lindegren L., et al., 2018, , [616, A2](#)
- Liu C., Bailer-Jones C. A. L., Sordo R., Vallenari A., Borrachero R., Luri X., Sartoretti P., 2012, *Monthly Notices of the Royal Astronomical Society*, 426, 2463
- Loebman S. R., Roškar R., Debattista V. P., Ivezic Quinn T. R., Wadsley J., 2011, *The Astrophysical Journal*, 737, 8
- Loredo T. J., 1992, *Promise of Bayesian Inference for Astrophysics*. Springer New York, New York, NY, pp 275–297, doi:[10.1007/978-1-4613-9290-3_31](https://doi.org/10.1007/978-1-4613-9290-3_31), https://doi.org/10.1007/978-1-4613-9290-3_31
- Loredo T. J., Lamb D. Q., 1989, *Annals of the New York Academy of Sciences*, 571, 601
- Loredo T. J., Wasserman I. M., 1998, , [502, 75](#)
- Luri X., et al., 2018, *Astronomy & Astrophysics*, 616, A9
- Majewski S. R., 1993, , [31, 575](#)
- Majewski S. R., et al., 2017, , [154, 94](#)
- Mann A. W., Feiden G. A., Gaidos E., Boyajian T., von Braun K., 2015, *The Astrophysical Journal*, 804, 64

- Martinez-Medina L. A., Pichardo B., Pérez-Villegas A., Moreno E., 2015, , 802, 109
- Mayor M., 1974, , 32, 321
- McCarthy M. F., 1959, , 53, 211
- McInnes L., Healy J., 2017, in 2017 IEEE International Conference on Data Mining Workshops (ICDMW). IEEE, p. 33–42, doi:10.1109/icdmw.2017.12, <http://dx.doi.org/10.1109/ICDMW.2017.12>
- McMillan P. J., 2011, , 414, 2446
- Metropolis N., Rosenbluth A. W., Rosenbluth M. N., Teller A. H., Teller E., 1953, , 21, 1087
- Minchev, I. Chiappini, C. Martig, M. 2013, *A&A*, 558, A9
- Minchev I., Chiappini C., Martig M., 2014, , 572, A92
- Mowlavi N., Eggenberger P., Meynet G., Ekström S., Georgy C., Maeder A., Charbonnel C., Eyer L., 2012, , 541, A41
- Munari U., Sordo R., Castelli F., Zwitter T., 2005, , 442, 1127
- Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, *The Astrophysical Journal*, 808, 16
- Nissen P. E., 1995, in van der Kruit P. C., Gilmore G., eds, IAU Symposium Vol. 164, Stellar Populations. p. 109
- Nordström B., et al., 2004, , 418, 989
- Ojha D. K., Bienaymé O., Robin A. C., Creze M., Mohan V., 1996, , 311, 456
- Pacifci C., et al., 2023, *The Astrophysical Journal*, 944, 141
- Petit V., Wade G. A., 2012, *Monthly Notices of the Royal Astronomical Society*, 420, 773
- Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2004, , 612, 168
- Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2006, *The Astrophysical Journal*, 642, 797
- Pietrinferni A., Cassisi S., Salaris M., Percival S., Ferguson J. W., 2009, *The Astrophysical Journal*, 697, 275
- Plez B., 2012, Turbospectrum: Code for spectral synthesis, *Astrophysics Source Code Library*, record ascl:1205.004
- Pont F., Eyer L., 2004, , 351, 487
- Prochaska J. X., Wolfe A. M., 2000, *The Astrophysical Journal*, 533, L5–L8
- Randich S., et al., 2018, , 612, A99
- Reetz J. K., 1999, PhD thesis, Ludwig-Maximilians University of Munich, Germany
- Rix H.-W., Bovy J., 2013, , 21, 61
- Roškar R., Debattista V. P., Quinn T. R., Stinson G. S., Wadsley J., 2008, , 684, L79
- Sandage A. R., 1953, , 58, 61
- Schönrich R., 2012, , 427, 274
- Schönrich R., Bergemann M., 2014, , 443, 698
- Schönrich R., Binney J., 2009, , 396, 203
- Schwarz G., 1978, *Annals of Statistics*, 6, 461
- Schönrich R., Binney J., Dehnen W., 2010, *Monthly Notices of the Royal Astronomical Society*, 403, 1829–1833
- Seabroke G. M., Gilmore G., 2007, , 380, 1348
- Sellwood J. A., 2014, *Reviews of Modern Physics*, 86, 1–46
- Sellwood J. A., Binney J. J., 2002, , 336, 785
- Sellwood J. A., Carlberg R. G., 1984, , 282, 61
- Serenelli A. M., Bergemann M., Ruchti G., Casagrande L., 2013, *Monthly Notices of the Royal Astronomical Society*, 429, 3645
- Shkedy Z., Decin L., Molenberghs G., Aerts C., 2007, *Monthly Notices of the Royal Astronomical Society*, 377, 120
- Silva Aguirre V., et al., 2015, , 452, 2127
- Sivia D. S., Skilling J., 2006. <https://api.semanticscholar.org/CorpusID:59722824>
- Soderblom D. R., 2010, *Annual Review of Astronomy and Astrophysics*, 48, 581
- Soubiran, C. Bienaym  , O. Siebert, A. 2003, *A&A*, 398, 141
- Tarricq, Y. et al., 2021, *A&A*, 647, A19
- Ting Y.-S., Rix H.-W., 2019, , 878, 21
- Trick W., 2019, in KITP Conference: In the Balance: Stasis and Disequilibrium in the Milky Way. p. 2

- Trotta R., 2008, [Contemporary Physics](#), **49**, 71
- Tully R. B., Fisher J. R., 1977, , **54**, 661
- Valle, G. Dell’Omodarme, M. Prada Moroni, P. G. Degl’Innocenti, S. 2015, [A&A](#), 575, A12
- Venn K. A., Irwin M., Shetrone M. D., Tout C. A., Hill V., Tolstoy E., 2004, , **128**, 1177
- Walker M. G., Olszewski E. W., Mateo M., 2015, [Monthly Notices of the Royal Astronomical Society](#), 448, 2717–2732
- Wielen R., 1974, [Publications of the Astronomical Society of the Pacific](#), 86, 341
- Wielen R., 1977, , **60**, 263
- Wojno J., et al., 2016, [Monthly Notices of the Royal Astronomical Society](#), 461, 4246