



# Discovery and characterization of 3000+ main-sequence binaries from APOGEE spectra

Kareem El-Badry,<sup>1,2\*</sup> Yuan-Sen Ting,<sup>3,4,5,6</sup> Hans-Walter Rix,<sup>2</sup> Eliot Quataert,<sup>1</sup> Daniel R. Weisz,<sup>1</sup> Phillip Cargile,<sup>7</sup> Charlie Conroy,<sup>7</sup> David W. Hogg,<sup>2,8,9,10</sup> Maria Bergemann<sup>2</sup> and Chao Liu<sup>9,11</sup>

<sup>1</sup>Department of Astronomy and Theoretical Astrophysics Center, University of California Berkeley, Berkeley, CA 94720, USA

<sup>2</sup>Max Planck Institute for Astronomy, D-69117 Heidelberg, Germany

<sup>3</sup>Research School of Astronomy and Astrophysics, The Australian National University, Canberra ACT 0200, Australia

<sup>4</sup>Institute for Advanced Study, Princeton, NJ 08540, USA

<sup>5</sup>Observatories of the Carnegie Institution of Washington, Pasadena, CA 91101, USA

<sup>6</sup>Department of Astrophysical Sciences, Princeton University, Princeton, NJ 08544, USA

<sup>7</sup>Department of Astronomy, Harvard University, Cambridge, MA, 02138, USA

<sup>8</sup>Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA

<sup>9</sup>Center for Cosmology and Particle Physics, Department of Physics, New York University, New York, NY 10003, USA

<sup>10</sup>Center for Data Science, New York University, New York, NY 10011, USA

<sup>11</sup>Kay Lab of Optical Astronomy, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

Accepted 2018 January 25. Received 2018 January 25; in original form 2017 November 20

## ABSTRACT

We develop a data-driven spectral model for identifying and characterizing spatially unresolved multiple-star systems and apply it to APOGEE DR13 spectra of main-sequence stars. Binaries and triples are identified as targets whose spectra can be significantly better fit by a superposition of two or three model spectra, drawn from the same isochrone, than any single-star model. From an initial sample of  $\sim 20\,000$  main-sequence targets, we identify  $\sim 2500$  binaries in which both the primary and secondary stars contribute detectably to the spectrum, simultaneously fitting for the velocities and stellar parameters of both components. We additionally identify and fit  $\sim 200$  triple systems, as well as  $\sim 700$  velocity-variable systems in which the secondary does not contribute detectably to the spectrum. Our model simplifies the process of simultaneously fitting single- or multi-epoch spectra with composite models and does not depend on a velocity offset between the two components of a binary, making it sensitive to traditionally undetectable systems with periods of hundreds or thousands of years. In agreement with conventional expectations, almost all the spectrally identified binaries with measured parallaxes fall above the main sequence in the colour–magnitude diagram. We find excellent agreement between spectrally and dynamically inferred mass ratios for the  $\sim 600$  binaries in which a dynamical mass ratio can be measured from multi-epoch radial velocities. We obtain full orbital solutions for 64 systems, including 14 close binaries within hierarchical triples. We make available catalogues of stellar parameters, abundances, mass ratios, and orbital parameters.

**Key words:** methods: data analysis – binaries: spectroscopic – Galaxy: stellar content.

## 1 INTRODUCTION

About half of solar-type stars are in binary or higher order multiple-star systems (Raghavan et al. 2010; Moe & Di Stefano 2017). Beyond the Solar neighbourhood, most binaries are too close on the

sky to be spatially resolved; they appear as single photometric point sources, and both components of binary systems contribute to the spectra observed by spectroscopic surveys.

Spectroscopically identifying such unresolved binaries is straightforward only if the period is relatively short ( $P \lesssim 5$  yr). In this case, spectra exhibit split or ‘double’ lines if the two components have comparable luminosities (so-called SB2 systems), and two peaks can be identified in the cross-correlation function

\*E-mail: [kelbadry@berkeley.edu](mailto:kelbadry@berkeley.edu)

(Pourbaix et al. 2004; Fernandez et al. 2017; Merle et al. 2017). Even if the secondary is faint and does not contribute significantly to the spectrum, short-period binaries can be identified from radial velocity (RV) variability when multi-epoch spectra are available ('SB1' systems; Minor 2013; Troup et al. 2016; Price-Whelan et al. 2017; Badenes et al. 2017).

However, about half of solar-type binaries have periods exceeding 200 yr (Duquennoy & Mayor 1991; Duchêne & Kraus 2013). The typical line-of-sight velocity separation between the two stars in such systems is of order  $1 \text{ km s}^{-1}$ , while the typical change in the stars' individual velocities over a one-year baseline is of order  $0.01 \text{ km s}^{-1}$ . Such systems will be missed by binary-detection methods based on the Doppler shift.

Unresolved binarity in main-sequence stars presents both a nuisance and an opportunity for spectroscopic surveys of the Milky Way. Because spectral morphology is a strong function of effective temperature, contamination from a cooler secondary star<sup>1</sup> makes the observable spectrum of an unresolved binary different from that of the primary, and in many cases, different from that of any single star. This means that, if binarity is ignored and all spectra are simply fitted with single-star models, biases can be introduced in the stellar parameters and abundances inferred for unrecognized binaries (El-Badry et al. 2018).

On the other hand, binarity-induced features in stellar spectra can be exploited to detect binaries that could not be detected based on velocity shifts alone: binaries can be identified as systems whose spectrum can be significantly better fit by a binary spectral model (i.e. a sum of two single-star models) than any single-star model. This approach, if it can successfully be applied to large spectroscopic surveys, will make possible systematic study of the Galactic binary population on an unprecedented scale.

El-Badry et al. (2018, hereafter E18) recently demonstrated that fitting a binary model to *synthetic* APOGEE-like spectra makes it possible to spectroscopically identify many binaries and to simultaneously recover the atmospheric parameters and abundances of both component stars. In this paper, we apply the method described in E18 to real spectra from DR13 of the APOGEE survey (Majewski et al. 2017). We focus on main-sequence stars, for which the effects of unresolved binarity on the spectrum are typically larger than in giants. We demonstrate that, although the spectral signatures of binarity are strongest in close systems with a large velocity offset between the two stars, binaries with mass ratios  $0.4 \leq q \leq 0.8$  can be detected with high fidelity even in the absence of any detectable velocity offset (where  $q = m_2/m_1$ ).

This paper is organized as follows. We describe our spectral model for single and binary stars in Section 2 and its application to the combined APOGEE spectra in Section 3. In Section 2.4, we extend the model to fit multi-epoch spectra of close binaries with detectable velocity changes between visits, calculating dynamical mass ratios from the relative velocities of the two components. We identify and derive parameters for close binaries, triples, and systems with unusual velocity shifts in Section 3.2 and derive orbital solutions for the subset of binaries with sufficient visits and phase coverage in Section 3.4. We discuss our results and conclude in Section 4.

<sup>1</sup>We adopt the convention that the secondary is the less massive of the two stars (e.g. Duchêne & Kraus 2013). For the equal-age, equal-composition main-sequence binaries that we model, the secondary is always cooler and less luminous.

We provide many of the underlying model details in the appendices. Specifically, in Appendix A, we describe the spectral model; model selection and tests with semi-empirical synthetic binary spectra are described in Appendix B; shortcomings of the model and false positives are discussed in Appendix C, and diagnostics of orbit-fitting convergence are presented in Appendix D. Available catalogues are described in Appendix E.

## 2 METHODS

Our binary spectral model depends on two steps: (i) creating a data-driven generative model for single-star spectra (Section 2.1), and (ii) combining the spectra of two single-star models, with a suitable velocity offset (Section 2.2). To find candidate binaries, we fit spectra with both single-star and binary models (Section 2.3) and identify systems that can be significantly better fit by a binary model (Appendix B).

In this work, we only attempt to fit main-sequence stars; i.e. targets with  $\log g \geq 4$ . We do not attempt to identify binaries in which one star is a giant because in most giant-dwarf binaries, the dwarf secondary will contribute a negligible fraction of the total light, while in giant-giant binaries, two components with the same age will necessarily have similar masses, and thus, quite similar spectra. We note that short-period binaries containing giants can be straightforwardly detected from RV variability (Troup et al. 2016; Badenes et al. 2017), and some giant–subgiant binaries can likely be detected spectroscopically (Section 4.1).

### 2.1 Single-star spectral model

We model APOGEE spectra of single stars using a data-driven<sup>2</sup> generative model to predict the rest-frame normalized flux density at a given wavelength as a function of a set of 'labels,'  $\vec{\ell}$ , which determine the spectrum. Our approach is very similar to that employed by *The Cannon* (Ness et al. 2015): the spectral model is a fitting function that maps labels to normalized spectra, and the free parameters of this fitting function are determined by optimization on a training set, whose labels are obtained separately or known a priori, e.g. from ab-initio fitting.

The primary difference between our method and existing implementations of *The Cannon* is that, as in Ting et al. (2017) and E18, we model the normalized flux density at a particular wavelength pixel using an artificial neural network rather than a polynomial function. We find a neural network model to be more flexible than a polynomial and to typically produce smaller errors in model spectra during cross-validation; this formalism, which we refer to as *The Payne*, is described further in Appendix A and will be explained in detail in Ting et al. (in preparation). The full spectral model then consists of all the individual neural networks for all wavelength pixels stitched together.

We predict rest-frame spectra with a single-star model that depends on five labels,

$$\vec{\ell} = (T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], v_{\text{macro}}). \quad (1)$$

We use  $[\text{Mg}/\text{Fe}]$  as a proxy for all 'α-elements.' We experimented with including more elemental abundances as labels, including C,

<sup>2</sup>We also experimented with using synthetic, ab-initio spectral models, but we found them ill suited for identifying binaries because systematic shortcomings in synthetic models cause almost *all* spectra to be significantly better fit by a sum of two models than a single model.

N, O, and Si. We found that this did not substantially change our identification of likely binary targets or their inferred mass ratios, so we opted to use a relatively simple model in the interests of reduced complexity.  $v_{\text{macro}}$  primarily accounts for the effects of stellar rotation, and is small ( $< 10 \text{ km s}^{-1}$ ) for most stars with  $T_{\text{eff}} \lesssim 6000 \text{ K}$ . In practice, spectra are not observed in the rest frame, so an additional label  $v_{\text{Helio}}$  also determines the model spectrum and must be included in fitting. However, our neural network model always predicts spectra in the rest frame; Doppler shifts are applied subsequently.

An ideal training set would contain only stars known to be single a priori. Unfortunately, it is nearly impossible to conclusively rule out the possibility that an unresolved system is a binary.<sup>3</sup> We therefore construct a training set by beginning with a random sample of main-sequence APOGEE stars and then iteratively removing stars whose spectra can be significantly better fit by the binary model described in Section 2.2. The ASPCAP pipeline does not derive reliably calibrated abundances for dwarfs. ‘Ground truth’ labels for stars in the training set were derived from ab-initio fitting with single-star models, following a procedure similar to that used by Ting et al. (2017); see Ting et al. (in preparation) for details. For the initial training set, we randomly selected 2000 targets distributed throughout the region of label space within which a spectral model was desired, namely  $4200 \text{ K} < T_{\text{eff}} < 7000 \text{ K}$ ,  $4.0 < \log g < 5.0$ ,  $-1 < [\text{Fe}/\text{H}] < 0.5$ ,  $-0.4 < [\text{Mg}/\text{Fe}] < 0.6$ , and  $0 \text{ km s}^{-1} < v_{\text{macro}} < 45 \text{ km s}^{-1}$ . We only attempt to fit targets for which the labels determined from ab-initio fitting lie within this region of parameter space, as (i) we are only interested in main-sequence stars, and (ii) the labels determined from ab-initio fitting are less reliable outside this range (Ting et al., in preparation).

There is of course no guarantee that the targets in our initial training set are actually single stars. After training the initial model, we therefore fit all spectra in the training set both with the initial single-star and binary models (as described in Section 2.2) based on this single-star model. We then removed from the training set the  $\sim 300$  targets that could be significantly better fit by a binary model than a single-star model<sup>4</sup> and retrained the single-star model on the resulting ‘cleaned’ training set. We repeated this cleaning and retraining procedure until none of the targets in the training set could be significantly better fit by a binary model. This approach converges quickly: after the second iteration, fewer than 10 targets in the cleaned training set could be significantly better fit by a binary model; after the third iteration, no additional targets in the training set could be significantly better fit by a binary model.

This iterative cleaning procedure likely does not remove all unresolved binaries from the training set: only binaries whose combined spectrum is significantly different from any single-star star spectrum can be identified. For APOGEE-like spectra of solar-type stars with negligible velocity offsets, the range of mass ratios over which binarity is detectable is  $0.4 \lesssim q \lesssim 0.85$  (E18). Binaries in

<sup>3</sup>The only exception is in the immediate Solar neighbourhood ( $d \lesssim 8 \text{ pc}$ ), where a combination of direct imaging and speckle interferometry can resolve nearly all systems where a velocity offset is not detectable (Simons, Henry & Kirkpatrick 1996; Reid & Gizis 1997). However, there are only  $\sim 66$  stars in the Solar neighbourhood for which binarity can be ruled out with high confidence; of these, only the Sun and Arcturus have been observed by APOGEE.

<sup>4</sup>Here, we quantified ‘significantly better fit’ as having  $\chi^2_{\text{single star}} - \chi^2_{\text{binary}} > 1000$ . We develop a more detailed threshold for model selection in Appendix B.

the training set with mass ratios outside this range will not contaminate the spectral model, since their spectra are not significantly different from the spectrum of a single star with the labels of the primary.

Our approach would likely not work if binaries dominated the training set, or if the functional form of the spectral model were sufficiently complex to incorporate spectral features due to binarity in the single-star model. Because binaries with spectra that are significantly better fit by a binary model constitute only  $\sim 15$  per cent of the initial training set and the spectral model is not very complex (we use a small neural network with only one hidden layer of five neurons), detectable binary spectra are essentially treated as outliers and removed during iterative cleaning, preventing the model from overfitting the signature of unresolved binarity into the single-star model.

## 2.2 Binary spectral model

We assume that both components of a binary system have the same age and composition. Fitting a binary model thus adds three free parameters compared to the single-star model: the mass ratio,  $q = m_2/m_1$ , which determines  $T_{\text{eff}}$  and  $\log g$  of the secondary, and  $v_{\text{macro}}$  and  $v_{\text{Helio}}$  of the secondary. To model the normalized spectrum of a binary with a particular mass ratio, we estimate  $T_{\text{eff}}$  and  $\log g$  of the secondary using MIST isochrones (Choi et al. 2016),<sup>5</sup> predict the single-star spectra of the primary and secondary in un-normalized space, apply a Doppler shift, add the two spectra, and finally pseudo-continuum normalize the total spectrum; see E18 for details.

Since the data-driven model for single stars operates on normalized spectra, predicting un-normalized spectra for the primary and secondary requires a model for the pseudo-continuum by which the normalized spectra can be multiplied. We obtain the pseudo-continuum for a single star at a particular point in label space by applying our pseudo-continuum fitting procedure (see Section 2.3) to a spectrum produced by a synthetic spectral model trained on Kurucz spectra (Kurucz 1970, 1979, 1993). Synthetic spectra are first produced with units of surface flux density and are then multiplied by the surface area of the star in question, using radii estimated from MIST isochrones. The un-normalized flux density of an unresolved binary system viewed from a distance  $D$  is then given by

$$f_{\lambda, \text{binary}} = \frac{1}{D^2} (R_1^2 f_{\lambda,1} + R_2^2 f_{\lambda,2}), \quad (2)$$

where  $R_1$  and  $R_2$  represent the radii of the primary and secondary star, and  $f_{\lambda,1}$  and  $f_{\lambda,2}$  represent their individual flux densities. Because we subsequently normalize  $f_{\lambda, \text{binary}}$  prior to fitting, the distance  $D$  is an arbitrary scaling factor and does not enter our analysis. In practice,  $R_1$  and  $R_2$  are estimated from  $T_{\text{eff}}$ ,  $\log g$ , and  $[\text{Fe}/\text{H}]$  using a neural network trained on a large grid of MIST isochrones.

Our results are not sensitive to the choice of synthetic model spectra, which sets only the relative flux contribution of the primary and the secondary, because the total binary spectrum is again normalized prior to fitting. We have verified that we obtain similar

<sup>5</sup>In practice, we predict  $T_{\text{eff}}$  and  $\log g$  of the secondary from  $T_{\text{eff}}$  and  $\log g$  of the primary,  $[\text{Fe}/\text{H}]$ , and  $q$  using a neural network trained on a large grid of binary isochrones with  $0.01 \leq (\text{age/Gyr}) \leq 13.5$  and  $-1 \leq [\text{Fe}/\text{H}] \leq 0.5$ . We have verified through cross-validation that typical errors in the thus-estimated parameters of the secondary are small ( $\sim 20 \text{ K}$  in  $T_{\text{eff}}$  and  $\sim 0.01 \text{ dex}$  in  $\log g$ ).

results by simply defining a continuum for each star as a blackbody with appropriate  $T_{\text{eff}}$  scaled by the surface area of the star.

For long-period systems with negligible velocity shifts, our model cannot detect binaries with mass ratios  $q \lesssim 0.4$ , because the secondary contributes a negligible fraction of the total light, or  $q \gtrsim 0.85$ , because the spectra of the primary and secondary are too similar. In practice, another, often more stringent limit on the lowest detectable mass ratio is set by our spectral model's minimum  $T_{\text{eff}}$  of 4200 K. For systems with a hot primary star ( $T_{\text{eff}} \gtrsim 6500$  K), this limit is not important, since a secondary with  $T_{\text{eff}} < 4200$  K would be too faint to contribute significantly to the spectrum anyway. However, the model's minimum  $T_{\text{eff}}$  reduces the range of detectable mass ratios for systems with cooler primaries: for a primary with  $T_{\text{eff}} = 5800$  K, the effective minimum  $q$  that can be modelled is  $q_{\min} \approx 0.62$ , while for a primary with  $T_{\text{eff}} = 5000$  K and  $q_{\min} \approx 0.75$ . We discuss this further in Appendix B1.

### 2.3 Model fitting

Best-fitting labels for binary and single-star models are determined through full-spectrum fitting of normalized spectra in vacuum wavelengths. Pseudo-continuum normalization is carried out using the *Cannon*-type normalization routine from the APOGEE package (Bovy 2016), which fits a fourth-order Chebyshev polynomial to pixels in which the gradient of the data-driven spectral model with respect to the labels is small. Bad pixels and pixels with poor sky subtraction, as flagged in the bitmasks produced by the APRED pipeline (Nidever et al. 2015), are masked during normalization and fitting.

Fitting is carried out using the `SCIPY curve_fit` routine, which implements the ‘trust region reflective’ algorithm (Branch, Coleman & Li 1999) for  $\chi^2$  minimization. When fitting a single spectrum with a single-star model, we find that the optimization essentially always converges on the true global minimum, irrespective of the location in label space where it is initialized. However, for the binary model there is an obvious degeneracy: the normalized spectrum of a  $q = 1$  binary model is identical to that of a  $q = 0$  model in the limit of no velocity offset. Hence, the posterior for a binary model is often bimodal in  $q$ , and minimization can sometimes converge on a false local minimum. We therefore initialize  $\sim 10$  separate optimizers with different initial values of  $q$  when fitting a binary model. If these do not all converge to the same model, we take as the best model the one that reaches the lowest global  $\chi^2$ . We have verified by fitting semi-empirical synthetic binary spectra that this approach converges on the true global minimum in  $\sim 99$  per cent of all cases (see Appendix B1).

Most APOGEE targets are observed more than once, with time baselines between individual visits ranging from  $\sim 1$  h to  $\sim 1200$  d.<sup>6</sup> Spectra from individual visits are shifted to rest frame and co-added to produce a single combined spectrum with higher signal-to-noise ratio (S/N) than the individual visit spectra by the APSTAR pipeline (Nidever et al. 2015). It is these combined spectra that are fit by the ASPCAP pipeline to derive the stellar parameters and abundances

<sup>6</sup>The APOGEE observing strategy aims to observe most targets three times, over a minimum time baseline of 1 month. Some targets, primarily faint stars, are visited more often to accumulate S/N; some targets in unfavourable locations, such as the Galactic Bulge, are visited only once (Zasowski et al. 2013). Most targets with baselines longer than 1 yr, as well as those with multiple visits within one night, are targets which were observed initially during the survey commissioning period and again during the main survey.

published for the main survey (Holtzman et al. 2015; García Pérez et al. 2016), but the reduced spectra from individual visits are also made publicly available.

Combined spectra are easier to work with than individual visit spectra both because they have higher S/N and because stars are often observed with a different fibre and with a different barycentric velocity at each visit, so that the combined spectrum is less affected by bad pixels, poor sky subtraction, and telluric absorption than the individual visit spectra. We therefore fit the combined spectra rather than spectra from individual visits when possible. However, if a system is an unresolved close binary, the orbital configuration and relative RVs of the primary and secondary will change between visits, so that the morphology of the total binary spectrum is different in each visit. In such cases, the combined spectrum does not represent any real physical system, and fitting it can yield biased labels.

For this reason, we attempt to fit all targets that may be close binaries using the individual visit spectra rather than the combined spectrum. We identify potential close binaries as targets for which (i) the best-fitting model to the combined spectrum is a binary model in which the line-of-sight velocity separation of the two components,  $\Delta v_{\text{los}}$ , is greater than  $10 \text{ km s}^{-1}$ , or (ii) the  $V_{\text{scatter}}$  term calculated from the RVs determined by the APSTAR pipeline (Nidever et al. 2015) is greater than  $1 \text{ km s}^{-1}$ , indicating potential RV variability. Some of these targets, particularly stars with high  $T_{\text{eff}}$  or low S/N, are single stars with poorly constrained RVs, but many are close binary systems. Fitting individual visit spectra for targets with  $V_{\text{scatter}} > 1 \text{ km s}^{-1}$  also protects against the possibility of a single star erroneously appearing to be a binary if the RVs are calculated incorrectly, while creating the combined spectrum; otherwise, co-adding two visit spectra with different Doppler shifts could produce a combined spectrum bearing erroneous signatures binarity with  $q = 1$ .

The number of free parameters to be optimized increases substantially when we fit spectra from many visits simultaneously, since the RVs at each visit are all free parameters. This can make the fit more susceptible to convergence on an erroneous local minimum in  $\chi^2$ ; we discuss the measures taken to ensure global convergence in this case in Section 2.4.

For both single-visit and combined spectra, we inflate the uncertainties of pixels with  $S/N > 200$ – $0.5$  per cent (i.e. S/N of 200) during fitting because empirical S/N diagnostics based on the variation in a given pixel across visits show that the noise model underestimates uncertainties for bright stars and is likely limited by systematics at this level (Nidever et al. 2015). We also find that our fitting approach often performs poorly at low S/N, primarily due to poor continuum normalization. We therefore do not attempt to fit any visit spectra with median  $S/N < 30 \text{ pixel}^{-1}$ . Since most APOGEE targets are bright, this restriction excludes less than 20 per cent of the targets in our sample; for these targets, we report labels obtained by fitting the combined spectrum, which has higher S/N, but we caution that results for targets with large  $V_{\text{scatter}}$  and low S/N are likely less reliable.

We do not report uncertainties on labels for individual targets. Formal fitting uncertainties based on the concavity of the likelihood function in the vicinity of the global maximum can be computed with `curve_fit` (e.g. Ness et al. 2015; Ho et al. 2017), and comparable uncertainties can be obtained by MCMC sampling. However, the thus-obtained uncertainties are typically unrealistically small for high-S/N spectra (e.g.  $\sigma(T_{\text{eff}}) < 10 \text{ K}$  for typical APOGEE spectra) because they do not properly account for systematic errors in the spectral model. Systematic errors can arise if (i) the spectral model

is not sufficiently complex to account for all the variance in the data set, (ii) there are unaccounted-for errors in the labels assigned to the training set, or (iii) the adopted set of labels does not fully characterize all the variance in the data set. We investigate the typical precision of our best-fitting labels in Appendix B1.1.

## 2.4 Fitting multi-epoch spectra

We attempt to fit the individual visit spectra rather than the combined spectra of all stars that were visited more than once and are flagged as potential close binaries. In order to fully exploit the information contained in the spectra, we fit all single-visit spectra for each system *simultaneously*, requiring the physical parameters of the component stars to be the same at all epochs. Because we fit all visit spectra with the same spectral model, we implicitly treat the instrumental line spread function as constant across all fibres and visits. For the single-star model, we also require the line-of-sight velocity to be the same at each epoch; in this case, the model is no more complex than when fitting a single combined spectrum.

For an isolated binary system, the line-of-sight velocities of the two components are not independent: in the centre-of-mass frame, conservation of linear momentum requires that the RV of the primary along any line of sight,  $v_1$ , and that of the secondary,  $v_2$ , are related by  $v_2 = -v_1/q_{\text{dyn}}$ , where  $q_{\text{dyn}}$  is the dynamical mass ratio of the system. If the centre-of-mass heliocentric velocity of the binary is  $\gamma$ , then

$$v_{\text{Helio},2} = \gamma + (v - v_{\text{Helio},1}) / q_{\text{dyn}}. \quad (3)$$

Here,  $v_{\text{Helio}}$  denotes a velocity at a single epoch, measured in the frame of the centre of mass of the Solar system.

For true, isolated binary systems containing two main-sequence stars,  $q_{\text{dyn}}$  should be equal to the *spectral* mass ratio  $q$ , which determines the contribution of the secondary star to the binary spectrum. We will use  $q$  and  $q_{\text{spec}}$  interchangeably in the rest of this paper. However, we fit  $q_{\text{dyn}}$  and  $q_{\text{spec}}$  separately to allow for the possibility of companions whose contribution to the spectrum is different from what is predicted by the dynamical mass ratio. This could occur, for example, if there are biases in the isochrones used in the spectral model, if the secondary falls near the edge of the APOGEE fibre and only a fraction of its flux contributes to the spectrum, or if a third object is present in the system. Comparing the best-fitting  $q_{\text{dyn}}$  and  $q_{\text{spec}}$  provides a useful diagnostic of the accuracy of our spectral model.

Our basic ‘SB2’ binary model does not allow the velocities of both stars to vary freely, but instead enforces the restriction that the velocities at all epochs follow equation (3) when two or more visit spectra are fit simultaneously. In most cases, this leads to best-fitting velocities that are similar (within  $\sim 200 \text{ m s}^{-1}$  on average, and nearly always within a few  $\text{km s}^{-1}$ ) to those obtained when equation (3) is not enforced. However, there are some targets for which the best-fitting velocities are very different – and produce a much better fit – when equation (3) is not enforced than when it is. Such systems have velocities inconsistent with being a simple two-body system and likely contain a third component. To avoid mischaracterizing these systems, we also fit all targets with a binary model in which the velocities of both components are allowed to vary freely; systems that are significantly better fit by this model are classified as SB2s with an unseen third component (see Section 2.4.1 for details).

We also find systems in which there is a clear RV shift in the spectrum between different visits but no individual visit spectrum is better fit by a binary model; i.e. the existence of a companion

can be inferred from its gravitational effects on the primary, but the companion does not significantly contribute to the observed spectrum. Most of these single-line binary (‘SB1’) systems are probably ordinary main-sequence binaries with low mass ratios and relatively short periods; some are likely binaries in which the companion is a stellar remnant. To distinguish between SB1s and SB2s, we fit all potential close binary systems with an SB1 model, which is identical to the single-star model, except that the RV is allowed to vary between visits. We designate systems as SB1s if the SB1 and SB2 models converge on essentially the same fit; i.e. if there is no detectable contribution to the spectrum from the secondary.

Finally, we find some systems whose visit spectra cannot be well fit by any single star or binary model: the binary model provides a better fit than the single-star model, but many lines are poorly fit or are missing entirely from the best-fitting binary model. We find that many of these systems can be much better fit by a *triple* model: i.e. three stars with independent velocities and masses, restricted to lie on the same isochrone.

### 2.4.1 Summary of models fit to visit spectra

We simultaneously fit the  $N$  visit spectra for each object in the ‘potential close binary’ subsample with a total of five different models, which we summarize here. We classify systems based on the total  $\chi^2$  of each model, preferring the least complex model when different models have similar  $\chi^2$ .

(i) *Single star*: the single-star model has six free parameters, regardless of the number of visit spectra:

$$\vec{\ell}_{\text{single star}} = (T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], v_{\text{macro}}, v_{\text{Helio}}). \quad (4)$$

In particular, this model forces the heliocentric velocity of the star to be the same in all visits.

(ii) *SB1*: the SB1 model is identical the single-star model, except that the heliocentric velocity is allowed to vary between the  $N$  visits. The  $5 + N$  free parameters are:

$$\vec{\ell}_{\text{SB1}} = (T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], v_{\text{macro}}, v_{\text{Helio},i}), \quad (5)$$

where  $i$  enumerates the visits.

(iii) *SB2*: the SB2 model fits two stars, with different velocities at each visit, but with the restriction that the velocity satisfy equation (3). The  $9 + N$  free parameters are

$$\vec{\ell}_{\text{SB2}} = \left( T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], q, v_{\text{macro}1}, v_{\text{macro}2}, q_{\text{dyn}}, \gamma, v_{\text{Helio},i} \right). \quad (6)$$

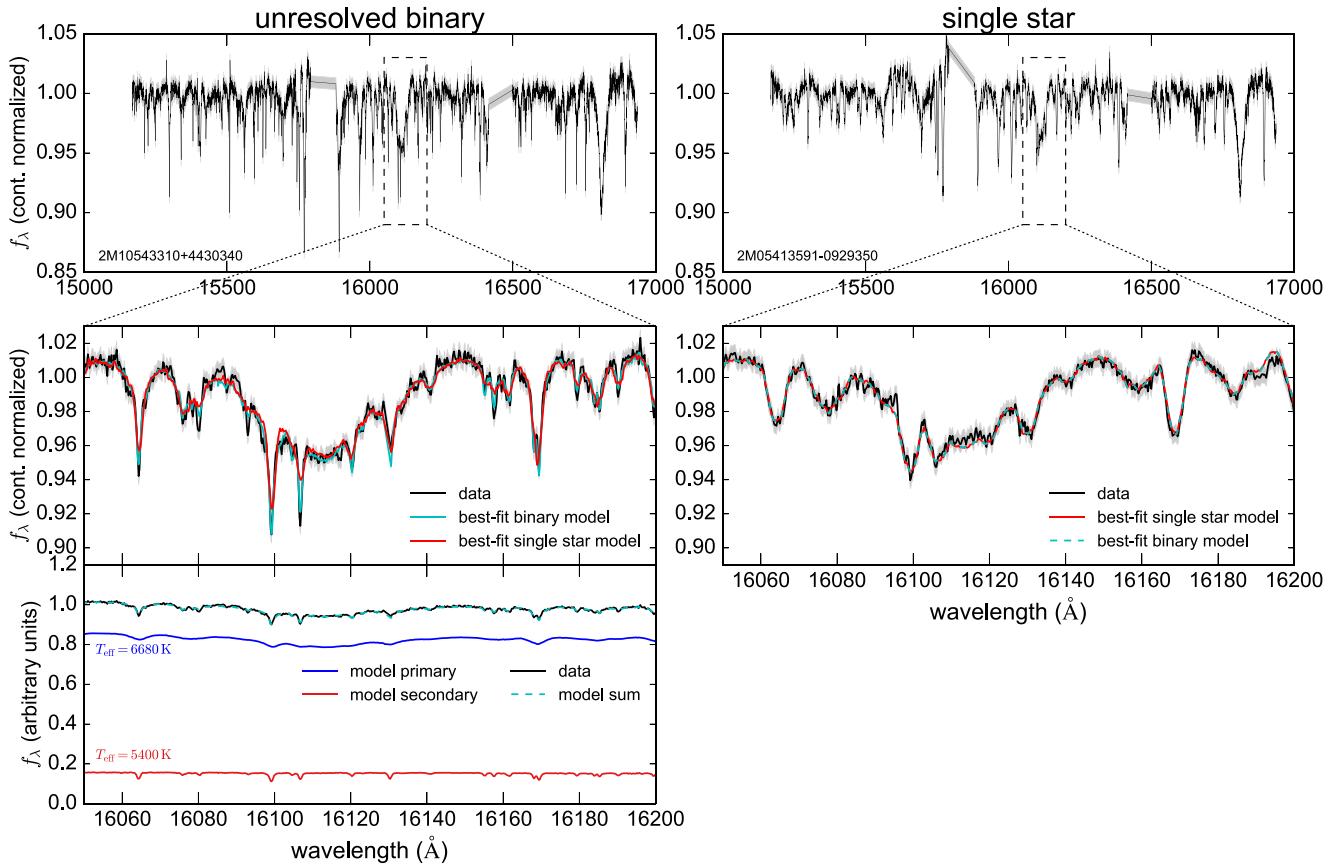
(iv) *SB2 with unseen third object*: this model fits two stars but allows their velocities to vary freely, without enforcing equation (3). If it provides a significantly better fit than the SB2 model, the relative RV shifts are inconsistent with being a simple Keplerian two-body system. The  $7 + 2N$  free parameters are:

$$\vec{\ell}_{\text{SB2, unseen 3rd object}} = \left( T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], q, v_{\text{macro}1}, v_{\text{macro}2}, v_{\text{Helio}1,i}, v_{\text{Helio}2,i} \right). \quad (7)$$

(v) *SB3*: the SB3 model fits three stars and imposes no restrictions on their velocities. The  $9 + 3N$  free parameters are:

$$\vec{\ell}_{\text{SB3}} = \left( T_{\text{eff}}, \log g, [\text{Fe}/\text{H}], [\text{Mg}/\text{Fe}], q_2, q_3, v_{\text{macro}1}, v_{\text{macro}2}, v_{\text{macro}3}, v_{\text{Helio}1,i}, v_{\text{Helio}2,i}, v_{\text{Helio}3,i} \right), \quad (8)$$

where  $q_2 = m_2/m_1$  and  $q_3 = m_3/m_1$ .



**Figure 1.** Left: spectrum of an unresolved main-sequence binary with  $q = m_2/m_1 \approx 0.7$  as observed by APOGEE. Top panel shows the full normalized spectrum. Middle panel shows the spectrum and best-fitting binary and single-star models, zoomed-in on a narrow wavelength range enclosing a hydrogen Brackett line. The binary model fits the data significantly better than the single-star model. Bottom panel shows the two components of the best-fitting binary model. The spectrum’s broad features are due primarily to the hotter star, which contributes > 80 per cent of the total light, but has no strong narrow lines; the shape of the sharp line profiles is primarily due to the cooler star. Our method makes it possible to identify many long-period binaries like this one, in which the velocity offset between the two stars is negligible. Right: spectrum of a presumed single star with similar parameters to the primary in the system shown in the left-hand panels. In this case, the best-fitting binary and single-star models are identical.

We note that the SB2 models are in principle identical to the SB1 model (and the SB3 model to the SB2 models) in the limit where  $q = 0$ . We keep these models separate in practice because our model does not transition smoothly from the minimum possible  $q$  that can be modelled (corresponding to  $T_{\text{eff}} = 4200$  K) to  $q = 0$ .

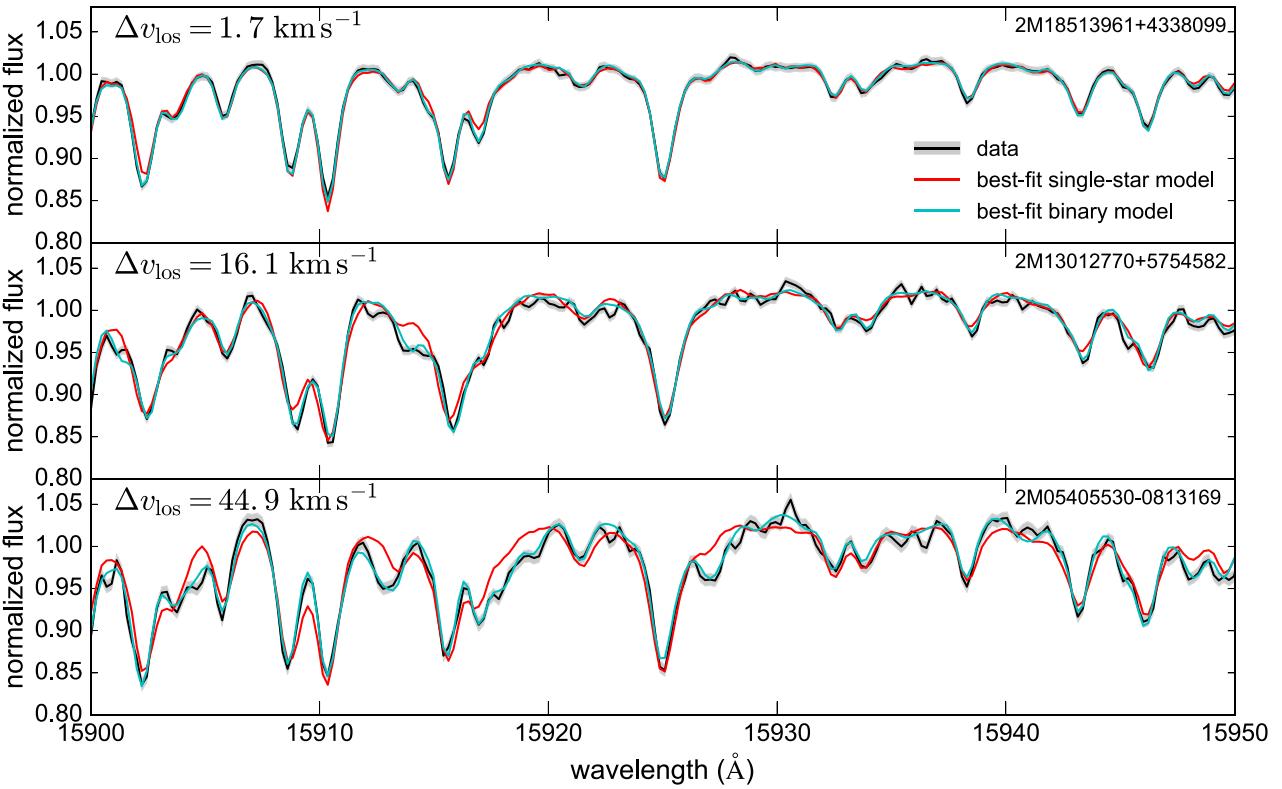
Fitting many visits simultaneously increases the number of labels to be fit, increasing the risk of the optimizer’s convergence on a local minimum. For example, for a target with 10 visits, fitting the SB2 (SB3) model requires optimization of the likelihood in 19 (39) dimensions, which is computationally demanding. In tests with synthetic binary spectra, we find that convergence on the global best fit is nearly always achieved *as long as the optimizer is initialized reasonably close to the global minimum*; i.e. with all velocities within  $\pm \sim 20 \text{ km s}^{-1}$  of their true values at all epochs. We therefore first fit individual visit spectra one at a time to estimate the velocity of each component at each epoch, and then use the resulting best-fitting labels to initialize the global optimizer during simultaneous fitting. Because the velocity offsets at each epoch are nearly uncorrelated with those in other epochs – i.e. changing  $v_{\text{Helio}, 1, i}$  only shifts the spectrum predicted for the  $i$ th visit – the optimization remains tractable in many dimensions.

### 3 RESULTS

We fit the spectra of 20,142 targets from APOGEE DR13 that ab-initio fitting with single-star models (Ting et al., in preparation) found to (i) lie on the main sequence ( $\log g > 4$ ), (ii) fall within the region of label space where the synthetic spectral model is reliable ( $4200 \text{ K} \leq T_{\text{eff}} \leq 7000 \text{ K}$  and  $-1 \leq [\text{Fe}/\text{H}] \leq 0.5$ ), and (iii) be acceptably fit, in a  $\chi^2$  sense, by synthetic spectral models. From this initial sample, we identify 2645 targets in which more than one star contributes significantly to the spectrum and an additional 663 targets with time-variable RVs but no detectable spectral contribution from the secondary. Catalogues of targets classified as single stars, binaries, and triples are presented in Appendix E.

Fig. 1 illustrates how our model identifies systems that are likely binaries but show no significant RV variability or split lines due to a velocity offset between the two components. Panels on the left show the spectrum of a target that can be significantly better fit by a binary model than a single-star model; those on the right show one that cannot.

We fit the full spectrum simultaneously, but we zoom-in on a small region to show the qualitative signatures of binarity. The



**Figure 2.** Examples of single-star and binary models fit to binary systems with a negligible (top), intermediate (middle), and large (bottom) line-of-sight velocity offset between the two stars. All three systems have a mass ratio  $q = m_2/m_1 \sim 0.7$ , a primary star with  $T_{\text{eff}} \sim 5400$  K and  $\log g \sim 4.5$ , and  $[\text{Fe}/\text{H}] \sim 0$ . Detecting binarity in systems with a large velocity offset ( $\Delta v_{\text{los}} \gtrsim 15 \text{ km s}^{-1}$ ) is straightforward, because the two stars' lines become separated in velocity space. However, binarity can also be detected in many systems, where the line-of-sight velocity offset is negligible, as in the top panel, because the two-component stars have different temperatures and ionization states, so their combined spectrum cannot be well fit by any single-star model.

spectrum in the left-hand panels contains features of both hot and cool stars: wide hydrogen lines and rotationally broadened line profiles at the wing of all lines, and deeper, narrow line cores that do not show rotational broadening. No single-star model can achieve a good fit: the absorption lines in the best-fitting single-star model are too shallow, and some lines in the data spectrum are blended in the best-fitting single-star model or are missing altogether. On the other hand, the binary model can provide a good fit and reproduces the line profiles of the observed spectrum. The decomposition of the binary model spectrum in the bottom panel shows that the broad features are all due to the hot primary star, while the sharper features originate in the spectrum of the cooler secondary.

In the right-hand panels of Fig. 1, we show the spectrum of a typical single star with stellar parameters and abundances similar to the primary in the left-hand panels; as expected, it is similar to the spectrum in the left-hand panels with the sharper, narrow lines removed. In this case, the binary and single-star models converge on what is essentially the same spectrum, so there is no reason to prefer the binary model.

The binary spectrum in the left-hand panels of Fig. 1 illustrates why it is often possible to spectrally identify binaries even when one star is much brighter than the other: although the secondary star in the binary system contributes less than 20 per cent of the total light, it contributes a large fraction of the total absorption because lines in hotter stars are often intrinsically weaker than those in cool stars. For many binaries containing a hot primary and cool secondary, the spectrum and binary model exhibit lines that are completely absent

from the spectrum of the primary because the relevant species are ionized at its higher  $T_{\text{eff}}$ .

### 3.1 Effect of a velocity offset

Although a line-of-sight velocity difference between the primary and secondary stars is in many cases not required to identify binaries with our model, a velocity offset makes the signatures of unresolved binarity more obvious and extends the range of detectable mass ratios. This is illustrated in Fig. 2, which compares the spectra of three binary systems with similar stellar parameters, abundances, and mass ratios, but a range of velocity offsets between the primary and secondary components. The system shown in the top panel has a small line-of-sight velocity offset, similar to the system in the left-hand panels of Fig. 1. In this case, the effects of binarity are quite subtle, and binarity can likely only be detected with detailed spectral modelling. As the velocity offset increases (middle and bottom panels), binarity-induced changes to the spectrum become more obvious. In all three panels of Fig. 2, we plot the APSTAR combined spectrum, not the spectra from individual visits. However, the target in the bottom panel, which is the only target of the three for which we might expect a large velocity change between visits, was only visited once.

For APOGEE spectra with  $R = 22\,500$ , one resolution element corresponds to a RV difference of  $\delta v \sim c/R \sim 13.5 \text{ km s}^{-1}$ . The traditional method of identifying binaries as systems in which the cross-correlation function of an observed spectrum with a synthetic

template exhibits two peaks can only reliably detect binaries in which the line-of-sight velocity offset is of order 1–3 resolution elements; such systems are usually referred to as ‘SB2’ systems. For example, Fernandez et al. (2017) found that binaries could only be reliably detected in APOGEE spectra when the maximum line-of-sight velocity separation exceeded  $\Delta v_{\text{los}} = 30 \text{ km s}^{-1}$ .<sup>7</sup> Fig. 2 shows that even a small velocity offset can substantially strengthen the signatures of binarity. How much a velocity offset improves detectability for our method depends on the stellar parameters and abundances of the primary, because it is easier to detect velocity offsets in stars with many deep, narrow lines. For most stars with  $T_{\text{eff}} \lesssim 6500 \text{ K}$ , a velocity offset of  $\Delta v_{\text{los}} \gtrsim (5–10) \text{ km s}^{-1}$  makes it possible to identify binaries from single-epoch spectra even when the mass ratio is close to  $q = 1$ ; such systems are not otherwise detectable with our method (see Appendix B1).

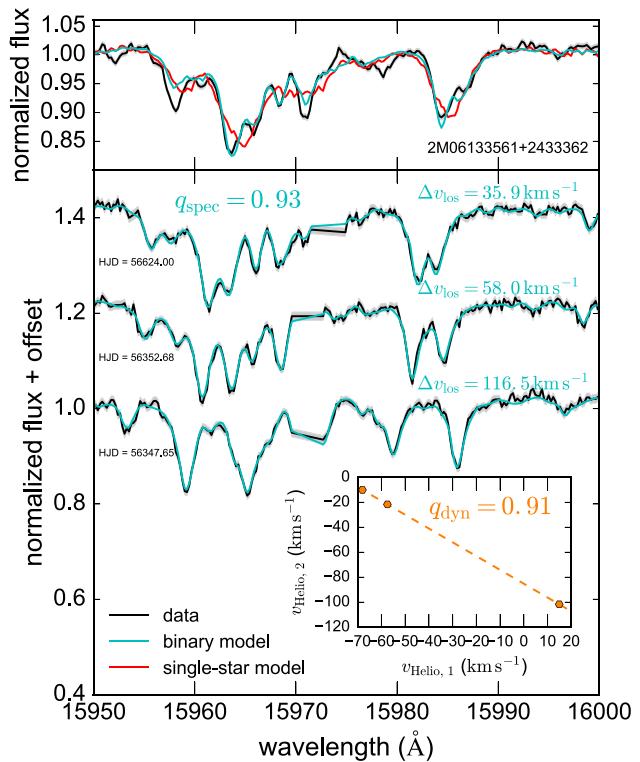
### 3.2 Results for multi-epoch spectra with velocity variability

Examples of targets whose spectra are best fit by SB2, SB1, and SB2 with an unseen third object, and SB3 models are shown in Figs 3, 5, 8, and 6.

Fig. 3 shows a system that is best fit by the SB2 model (i.e. case (iii) from Section 2.4.1) and exhibits spectra that change substantially from one epoch to the next. In the upper panel, we plot the combined spectrum and the best-fitting binary and single-star models obtained by fitting it. Although the binary model is a better fit (and our initial fit to the combined spectrum did flag the system as a likely binary), the fit is not very good: some features in the combined spectrum cannot be accommodated by either the single-star or the binary models. In the lower panel, we show the spectra obtained in the three individual visits, which are co-added to produce the combined spectrum, and the binary model obtained by simultaneously fitting them. The fit to the individual visit spectra is good. The poor fit to the combined spectrum is a consequence of the fact that the components’ velocities change between visits, meaning that the combined spectrum is an unphysical superposition of different spectra.

The inset in Fig. 3 shows the heliocentric velocities of the primary and secondary stars at each visit for the best-fitting SB2 model. The slope and intercept of the line on which these velocities fall can be used to calculate the dynamical mass ratio,  $q_{\text{dyn}}$ , and the centre-of-mass velocity,  $\gamma$ . For binary systems in which the velocities of the two stars change significantly between visits, it is therefore possible to obtain a constraint on the mass ratio that is independent of the spectral label  $q$ . Such constraints will of course not be reliable if the orbital configuration does not change significantly between visits: in this case, all measurements of  $v_{\text{Helio},1}$  and  $v_{\text{Helio},2}$  will be clustered around one point, and the slope of the line is ill constrained. We also emphasize that linear momentum conservation requires that the slope of the line on which  $v_{\text{Helio},2}$  and  $v_{\text{Helio},1}$  fall must be *negative* for a true binary system. Fernandez et al. (2017) attempted to infer  $q_{\text{dyn}}$  also from systems in which the slope of this line is positive or zero (e.g. their fig. 6), but mass ratios inferred in this way have no physical interpretation and indicate either inaccurate RV measurements or the presence of a third, unseen component.

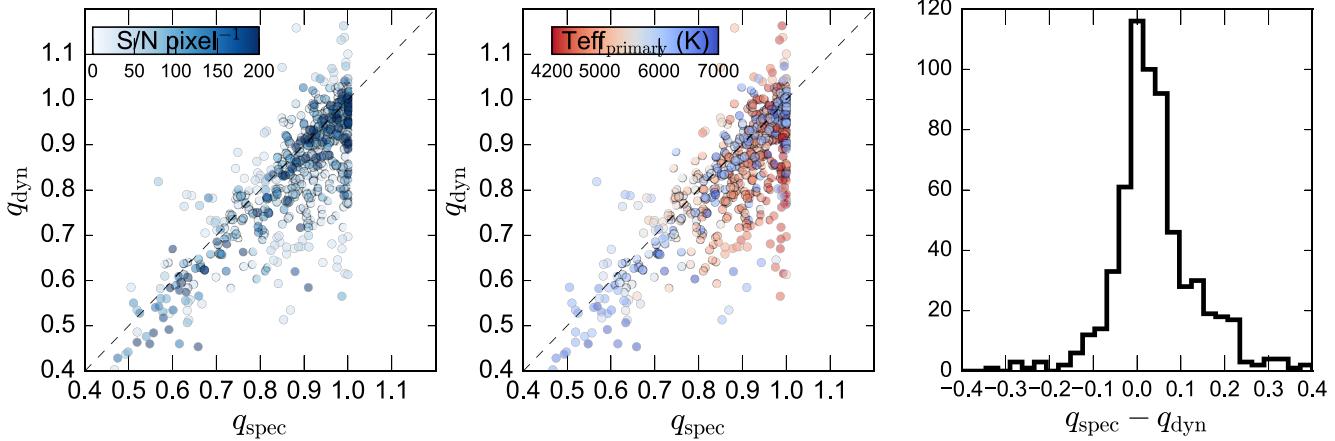
<sup>7</sup>Other surveys find similar sensitivity to spectroscopic binaries; e.g. Merle et al. (2017) found that binaries could be detected down to  $\Delta v_{\text{los}} = 15 \text{ km s}^{-1}$  in UVES spectra ( $R = 47\,000$ ) from the Gaia–ESO survey, and Matijević et al. (2010) found a minimum  $\Delta v_{\text{los}}$  for reliable detection of  $50 \text{ km s}^{-1}$  in the RAVE survey ( $R = 7500$ ).



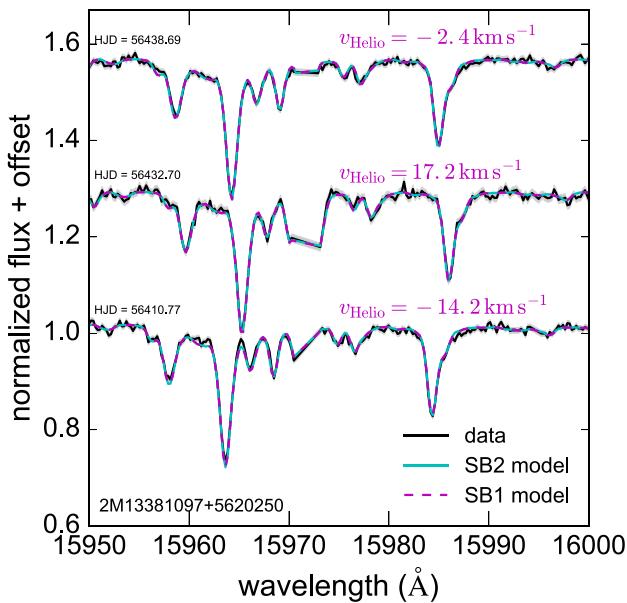
**Figure 3.** A binary system in which the stars’ velocities change between visits. Top panel shows a small portion of the *combined* spectrum (black), which is produced by co-adding spectra from different visits, best-fitting single-star model (red), and best-fitting binary model (cyan). The binary model provides a better fit than the single-star model, but it cannot fully reproduce the combined spectrum. Bottom panel shows the individual visit and the best-fitting SB2 model, which produces an excellent match to all the individual visit spectra. Inset shows heliocentric velocities of the primary and secondary stars at each epoch; momentum conservation requires that these lie on a line with slope  $-1/q_{\text{dyn}}$ , where  $q_{\text{dyn}}$  is the dynamical mass ratio. The spectrally inferred mass ratio,  $q_{\text{spec}} = 0.93$ , is in good agreement with the dynamical mass ratio,  $q_{\text{dyn}} = 0.91$ .

In Fig. 4, we compare the best-fitting values of  $q_{\text{spec}}$  and  $q_{\text{dyn}}$  obtained for SB2 systems in which  $q_{\text{dyn}}$  can be reliably measured; we identify such systems as those in which the range of  $v_{\text{Helio}}$  spanned across visits is at least  $10 \text{ km s}^{-1}$  for both stars, corresponding to a velocity shift of slightly less than one resolution element. We colour points by the median of S/N per pixel as reported in the allVisit catalogue, where the median is over all visit spectra used in the fit.

The agreement between  $q_{\text{spec}}$  and  $q_{\text{dyn}}$  is in general quite good, with a median absolute difference between  $q_{\text{spec}}$  and  $q_{\text{dyn}}$  of  $\text{med}(|q_{\text{spec}} - q_{\text{dyn}}|) = 0.048$  and a corresponding middle 68 per cent range of (0.012–0.14). The agreement is on average better for targets whose spectra have higher S/N; most systems with significantly different  $q_{\text{spec}}$  and  $q_{\text{dyn}}$  have  $S/N \lesssim 50$ . Particularly at lower mass ratios,  $q_{\text{dyn}}$  is on average slightly lower than  $q_{\text{spec}}$ ; i.e. assuming  $q_{\text{dyn}}$  is usually more accurate than  $q_{\text{spec}}$ , the latter is biased to slightly higher  $q$ . This can be understood as a consequence of the minimum  $T_{\text{eff}}$  of our spectral model, which sets an effective minimum  $q_{\text{spec}}$ . If a cool primary has a companion with  $T_{\text{eff}}$  cooler than 4200 K that cannot be fully accommodated by the spectral model, a better fit can often still be achieved by a binary model with  $T_{\text{eff}} = 4200 \text{ K}$  and a too-high  $q_{\text{spec}}$  than a single-star model which ignores the



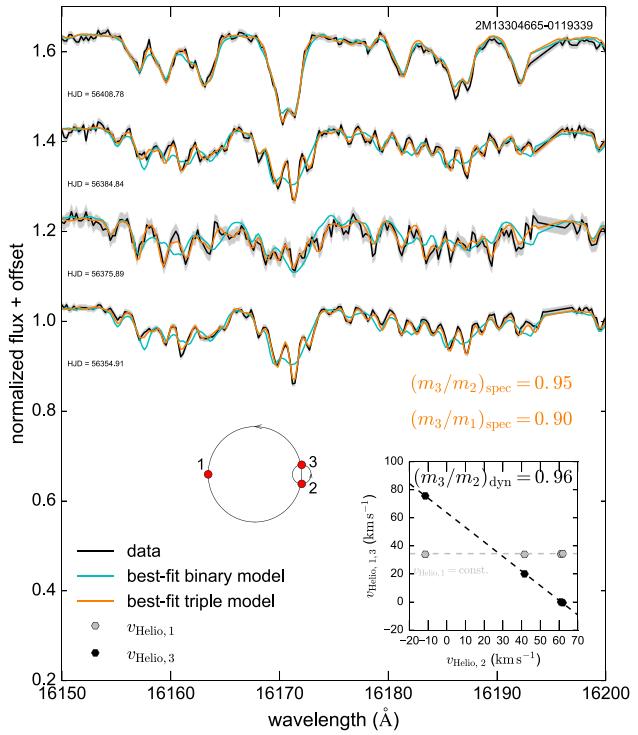
**Figure 4.** Comparison of spectroscopically and dynamically inferred mass ratios for ‘SB2’ binary systems in which a dynamical mass ratio can be measured.  $q_{\text{spec}}$  is measured from the relative contribution of each star to the spectrum, and  $q_{\text{dyn}}$ , from the relative changes of the RVs of the primary and secondary across multiple epochs (see Fig. 3). The designation of primary and secondary components is based on their relative contribution to the spectrum:  $q_{\text{spec}}$  is bounded by 1, but  $q_{\text{dyn}}$  is not. 623 systems have sufficiently short periods to allow measurement of  $q_{\text{dyn}}$ . Most systems for which the disagreement between  $q_{\text{spec}}$  and  $q_{\text{dyn}}$  is large have low S/N (left) and cool primaries (middle). Due to the spectral model’s minimum  $T_{\text{eff}}$  of 4200 K, low mass ratio systems can only be detected if the primary is hot, and mass ratios are less accurate for cooler systems. The median absolute difference between  $q_{\text{dyn}}$  and  $q_{\text{spec}}$  is 0.048.



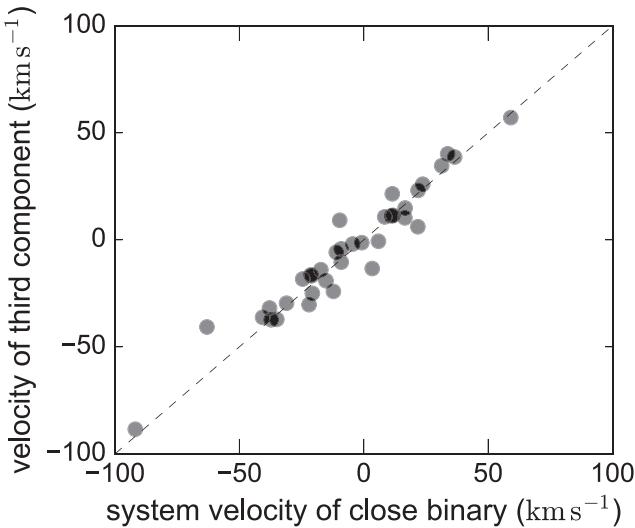
**Figure 5.** Visit spectra and best-fitting models for an SB1 system. The SB1 model contains only a single star contributing to the spectrum, but its RV can vary across visits. The SB2 model includes the possibility of a second star contributing to the spectrum. In this case, the best-fitting SB1 and SB2 models are identical, indicating that there is no detectable contribution to the spectrum from the secondary. However, RV variability of the primary clearly indicates that a companion is present. Assuming that the companion is a main-sequence star, an upper-limit of  $q \lesssim 0.45$  can be derived; if  $q$  were larger, the secondary would contribute detectably to the spectrum, and the SB2 model would provide a better fit.

secondary entirely. This in part explains the substantial number of cool systems with  $q_{\text{spec}}$  near 1 and lower  $q_{\text{dyn}}$ , though we note that most cool systems also have lower S/N.

If the secondary is very faint compared to the primary, its contribution to the spectrum may be completely undetectable, in which case binary and single-star models will converge to the same model



**Figure 6.** Visit spectra of a target identified as a triple (SB3) system. The three components have different line-of-sight velocities, so many lines can be seen in triple, and an SB2 model cannot provide a good fit. Inset shows the line-of-sight velocities of each component at each epoch. The heliocentric velocity of the primary is consistent with being constant at  $v_{\text{Helio},1} \approx 34.5 \text{ km s}^{-1}$ , so no dynamical mass ratio can be estimated for  $m_2/m_1$  or  $m_3/m_1$ . However,  $v_{\text{Helio},2}$  and  $v_{\text{Helio},3}$  fall on a line with an implied mass ratio  $m_3/m_2$  consistent with the spectrally inferred one. This implies that the system is a hierarchical triple, as is illustrated in the orbit schematic.



**Figure 7.** Line-of-sight velocities for hierarchical triples containing a close binary and third component with a much larger separation (see schematic in Fig. 6). The tight correlation between the centre-of-mass velocity of the close sub-binary and the velocity of the long-period component indicates that most systems are bona-fide gravitationally bound triples, not chance alignments between a close binary and a background or foreground star.

spectrum as long as the velocities are allowed to vary between visits. Such systems can be distinguished from isolated single stars by the fact that the ‘SB1’ model provides a better fit than the single-star model, which requires a target’s velocity to be the same at all visits. Fig. 5 shows an example of such a system. Our model makes it possible to set an upper limit on the mass ratio, under the assumption that the companion is a main-sequence star: in this case, the SB2 model would provide a better fit than the SB1 model if the secondary had  $T_{\text{eff}} \gtrsim 4200$  K. This limit is likely conservative in practice for main-sequence secondaries, as discussed above. However, it will not apply for binaries in which the companion is a stellar remnant.

We note that most SB1s and some close SB2s can be qualitatively identified as unlikely to be single based on the scatter across visits in the RVs measured by the APSTAR pipeline (e.g. Badenes et al. 2017). However, we find a nontrivial number of SB2 systems ( $\sim 100$  systems out of the  $\sim 20000$  targets studied in this work) that show clearly time-variable spectra, with changes in the velocities of both components larger than  $30 \text{ km s}^{-1}$ , for which the APSTAR-derived  $v_{\text{Helio}}$  measurements change at the  $< 1 \text{ km s}^{-1}$  level. This indicates that APOGEE RV measurements are likely problematic for these systems, and studies that flag short-period binaries based on velocity variability will miss some SB2 systems.

Fig. 6 shows an example of a spectrum classified as a triple. The SB2 model (cyan) clearly cannot provide a good fit to the observed spectra, which simply have too many lines; on the other hand, the triple model is a good fit to all visits. The inset shows the velocities of the three components at each epoch; note that these are all allowed to vary freely and are not restricted to follow any equivalent of equation (3). One component, the spectral primary, has effectively constant velocity (within  $\pm 0.5 \text{ km s}^{-1}$ ) across all visits. On the other hand, the velocities of the secondary and tertiary components vary a great deal between visits and fall on a line with negative slope, just as in case of close binaries (Fig. 3). The most straightforward explanation for these kinematics is that the system is a hierarchical triple (e.g. Ford, Kozinsky & Rasio 2000; Toonen, Hamers & Portegies Zwart

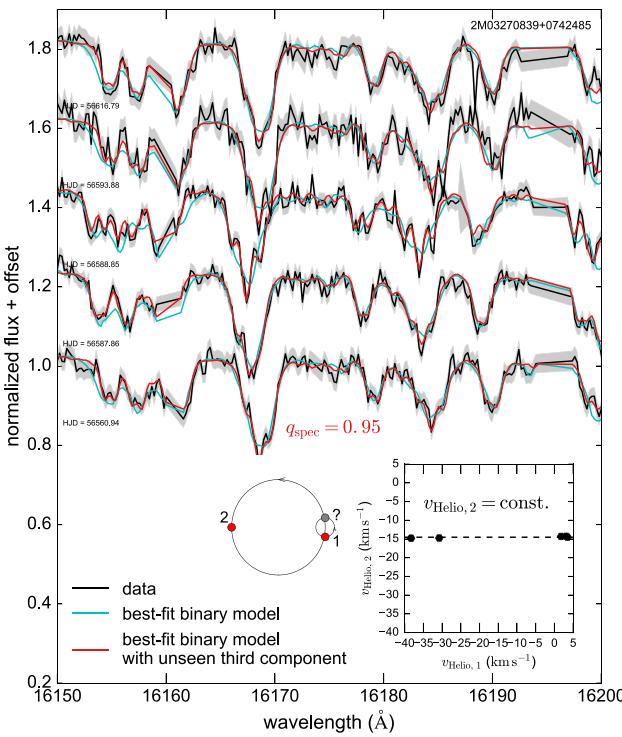
2016) consisting of a close binary orbiting a third system with a period much longer than that of the close binary, so that the velocity of the primary and the centre-of-mass velocity of the close binary do not change significantly over the temporal baseline between visits (which is  $\sim 54$  d for this target). This type of hierarchical orbital is illustrated schematically in Fig. 6. Consistent with this interpretation, the spectrally inferred mass ratio between the two components of the close binary is similar to the dynamical mass ratio inferred from the slope of the line on which their velocities fall.

We find 114 triple systems, most of which have the same qualitative velocity configuration as the system in Fig. 6: they contain one component with effectively constant velocity over all visits and two components with variable velocities that fall on a line as expected by a close binary. This is not surprising, as hierarchical configurations are the natural stable end state of the dynamical evolution of (otherwise chaotic) triple systems (Naoz et al. 2013). We also find systems in which the velocity of the third (long-period) component is not constant but changes approximately linearly with time; this is expected if the system’s outer period is long compared to the observation baseline but not so long that no change can be observed. In such cases, the heliocentric velocities of the other two components do not fall on a straight line but exhibit some intrinsic scatter; this scatter can be reduced if a constant multiple of the linear trend of the lone star is subtracted from the velocities of the other two stars.

Such systems are almost certainly gravitationally bound triples, since the velocities of all three components are correlated. However, for triples in which the velocity of one component is consistent with being constant over the time baseline spanned by observations, there is no guarantee that the three stars are actually gravitationally bound: the observed velocities could also be explained by a chance alignment between a close binary system and a background or foreground star. Whether such chance alignments constitute a substantial fraction of the targets we identify as hierarchical binaries can be diagnosed by comparing the centre-of-mass velocity of the close binaries to the velocity of the third component. For gravitationally bound triples, these should be reasonably similar, with offsets of order the orbital velocity of the long-period component. The typical offsets should be larger (at minimum, of order  $30 \text{ km s}^{-1}$ , the velocity dispersion of the Milky Way’s stellar disc) for chance alignments.

We investigate this explicitly in Fig. 7. Here, we only plot systems that are consistent with the velocity of the long-period component being fixed over all epochs; we identify such cases as systems in which the change in the velocity of the long-period component across epochs is less than  $2 \text{ km s}^{-1}$  when all velocities are allowed to vary freely. Consistent with the expectation for bound triples, the system velocity of the close binary is in most cases within  $10 \text{ km s}^{-1}$  of that of the third component. There are five systems in which the offset is larger, but due to the relatively short observational baselines, we find that none of these velocity offsets are large enough to rule out the possibility that all three stars are gravitationally bound. We discuss the possibility of contamination due to chance alignments of stars further in Section 3.5.

Along with SB1s, SB2s, and SB3s, we also identify a class of systems in which the presence of a third component can be deduced from RV measurements, but only two stars contribute significantly to the spectrum. Fig. 8 shows an example of such a target. The standard SB2 model, which enforces equation (3) with  $q_{\text{dyn}} \geq 0.2$ , cannot satisfactorily fit the spectrum. However, the ‘SB2 with unseen third component’ model, which allows the velocities of both components to vary freely, provides a good fit, converging on a



**Figure 8.** Visit spectra of a triple system in which the third component does not contribute significantly to the spectrum but can be detected gravitationally. Cyan line shows best-fitting SB2 model with the restriction that  $v_{\text{Helio},1}$  and  $v_{\text{Helio},2}$  fall on a line with negative slope [equation (3)]. Red line shows the best-fitting binary model in which the velocities of the primary and secondary are allowed to vary freely. Inset shows the line-of-sight velocities corresponding to the red model. The velocity of the secondary is consistent with being constant at  $v_{\text{Helio},2} \approx -14.5 \text{ km s}^{-1}$ , while that of the primary varies substantially. This implies that the system is a hierarchical triple in which one component of the close binary does not contribute to the spectrum (i.e. it is a stellar remnant or faint M dwarf); this is shown schematically in the orbital diagram.

solution in which the velocity of one component is consistent with being fixed across epochs while that of the other varies.

As illustrated in the orbital schematic in Fig. 8, such a RV pattern can be explained straightforwardly if the system is a hierarchical triple in which the close binary is an SB1; i.e. one component of the close binary does not contribute to the spectrum, either because its mass is low or because it is a compact remnant. No dynamical mass ratio can be inferred for these systems, because the acceleration of the variable-velocity component is due primarily to the unseen component. We identify 108 SB2 systems in which the presence of a third component can be inferred dynamically; the majority of these systems have velocity configurations similar to Fig. 8, with one component's velocity essentially constant.

### 3.3 Colour-magnitude diagram

A straightforward diagnostic to verify that the targets we spectroscopically identify as binaries are primarily true binaries, as opposed to single stars whose spectra contain unusual features that are not well accounted for in the spectral model, is to examine their distribution in a colour-magnitude diagram (CMD). True binaries are expected to lie above the single-star main sequence of a CMD (Hurley & Tout 1998; Li, de Grijs & Deng 2013): binaries with  $q \sim 1$  will have the same colour as would either single star but will

be twice as luminous, while binaries with  $0.5 \lesssim q \lesssim 0.9$  will be both brighter and redder than a single star with the parameters of the primary.

Accurate measurements of absolute magnitude (and hence distance) are required to construct the CMD. To identify stars in our sample with accurate distance measurements, we cross-matched it with the Tycho–Gaia astrometric solution catalogue (Michalik, Lindegren & Hobbs 2015) using the `gaia_tools.xmatch` routine written by Jo Bovy. This revealed 1925 stars in our sample with parallax errors of 10 per cent or better,<sup>8</sup> 217 of which were spectroscopically identified as multiple-star systems in which at least two components contribute detectably to the spectrum.

We plot the CMD for these objects, based on 2MASS photometry, in the left-hand panel of Fig. 9. As expected, the majority of spectroscopically identified binaries are scattered above the main sequence. We stress that our model for identifying binaries operates exclusively on normalized spectra and does not rely whatsoever on photometry; the fact that nearly all of the spectroscopically identified binaries populate the expected locus of the CMD above the main sequence is therefore a robust confirmation that our model is finding real binaries.

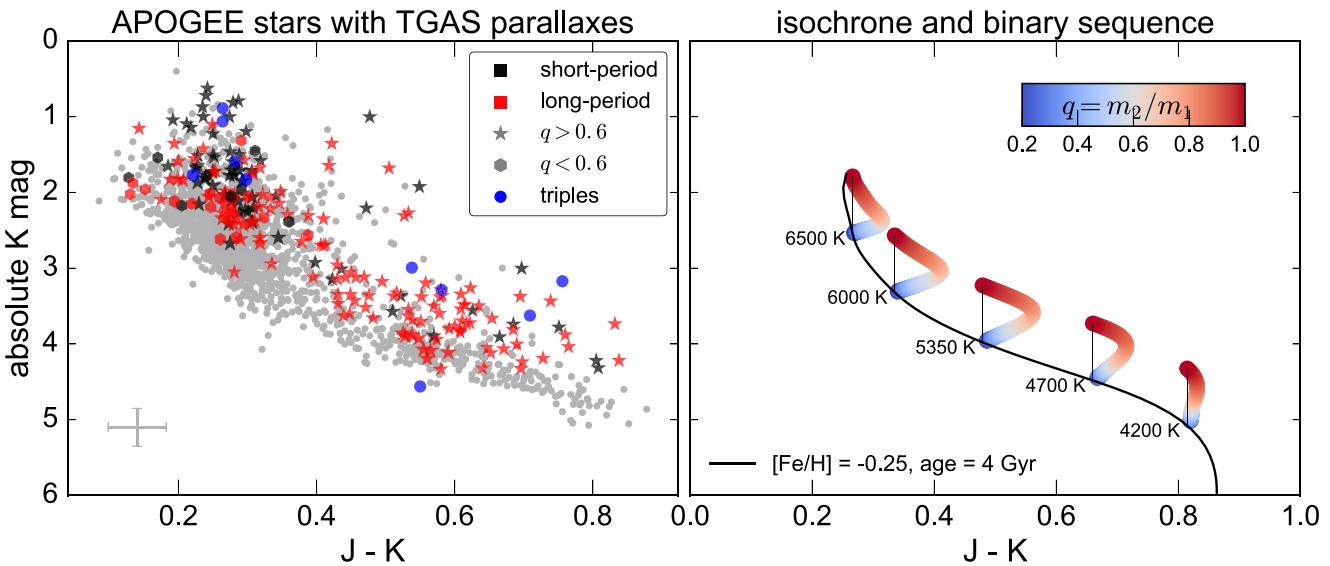
In the right-hand panel of Fig. 9, we show schematically how the presence of an unresolved companion is expected to change a star's position on the CMD. For a single stellar population, binaries with mass ratios  $0.6 \lesssim q \lesssim 1$  form a coherent second sequence  $\sim 0.75$  mag above the main sequence; i.e. it is *not* the case over this range of mass ratios that binaries with higher  $q$  fall farther above the main sequence. This occurs because as  $q$  is increased from 0.6 to 1 and  $T_{\text{eff}}$  of the secondary increases, unresolved binaries move blueward parallel to the main sequence. On the other hand, in binaries with  $q \lesssim 0.4$ , the secondary contribute so little light that the change in the unresolved system's location on the CMD is negligible.

The lowest mass ratio to which we are sensitive is  $q \sim 0.4$ , so the majority of the binaries we identify should scatter above the main sequence for their respective isochrone. In the left-hand panel of Fig. 9, we plot separately binaries with  $q \leq 0.6$  (which are only detectable around primaries with  $T_{\text{eff}} \gtrsim 5800 \text{ K}$ ; see Appendix B1) and those with higher mass ratios. As expected, the lower mass ratio systems are on average below the higher mass ratio systems on the CMD. With one exception,<sup>9</sup> systems identified as triples (SB3) fall above the binary main sequence. We do not mark SB1s in Fig. 9; their distribution on the CMD is similar to that of single stars, likely because most have low mass ratios.

The sample of stars for which accurate parallaxes are available spans a wide range of metallicities and ages, so significant intrinsic scatter is expected in the distribution of both single stars and binaries on the CMD. We note that there are some stars that our model does not identify as binaries but which still scatter well above the main sequence. We suspect that most of these systems are binaries with

<sup>8</sup>This corresponds to a magnitude error of  $\pm 0.22$  mag, plus typical 2MASS photometric errors of  $\pm 0.03$  mag. We do not attempt to correct for extinction or reddening, which is expected to be modest in the near-infrared at the distances of the stars with accurate parallaxes (which have a median distance of 200 pc).

<sup>9</sup>We have investigated the spectra of this target (2M07212735+2342096) in detail and find it to be an unambiguous triple, with clear changes in spectral morphology between visits. Why it falls below the main sequence is not clear; one possibility is that marginally resolved multiplicity led to an overestimate of its parallax.



**Figure 9.** Left: CMD of all stars in our sample with parallax errors of 10 per cent or less (grey circles). Black and red symbols represent spectroscopically identified binaries, with maximum line-of-sight velocity offsets between the two components greater (short period) and less than (long period)  $30 \text{ km s}^{-1}$ . Binaries with mass ratios  $q < 0.6$  are marked with hexagons; those with  $q > 0.6$ , with stars. Triples are marked with blue points. Most spectroscopically identified binaries lie above the main sequence. Traditional SB2 identification methods can only identify close binaries with large velocity offsets (black symbols); our method can identify many more long-period binaries with negligible velocity offsets (red symbols). Right: schematic effect of unresolved binarity on the CMD. Black line shows a MIST isochrone for single stars with a single age and metallicity; coloured loci show where binaries with different mass ratios and primary  $T_{\text{eff}}$  fall when they are spatially unresolved.

$q \sim 1$  and small velocity offsets; these are not detectable in our current framework.

We also divide suspected binaries into subsamples with large and small velocity offsets, corresponding approximately to systems which could and could not be detected with traditional Cross-correlation function (CCF)-based binary-detection methods (Fernandez et al. 2017). Only  $\sim 30$  per cent of systems have velocity offsets that are large enough to be detected with traditional methods. This highlights one of the primary advantages of the method introduced in this work: it is sensitive to a substantially larger fraction of the binary population than methods based on RV separation or variability alone.

### 3.4 Deriving orbital parameters

We derive full orbital solutions for 64 binary systems which have sufficient visits and phase coverage to constrain the orbit. Our criteria for determining whether the available velocity data are sufficient to constrain a system’s orbit are discussed in Appendix D. We only attempt to derive orbital solutions for systems in which at least two stars contribute to the spectrum; orbital solutions for SB1 systems in APOGEE can be found in Troup et al. (2016).

Velocities for both components are returned as labels for the best-fitting spectral model. An initial estimate of the RV uncertainty at each visit is obtained through bootstrapping: Gaussian noise proportional to the best-fitting model residual at each pixel is added to the data spectrum and the fit is repeated; the uncertainty on each RV is taken to be the standard deviation of the best-fitting velocity at each epoch when this procedure is repeated many times.

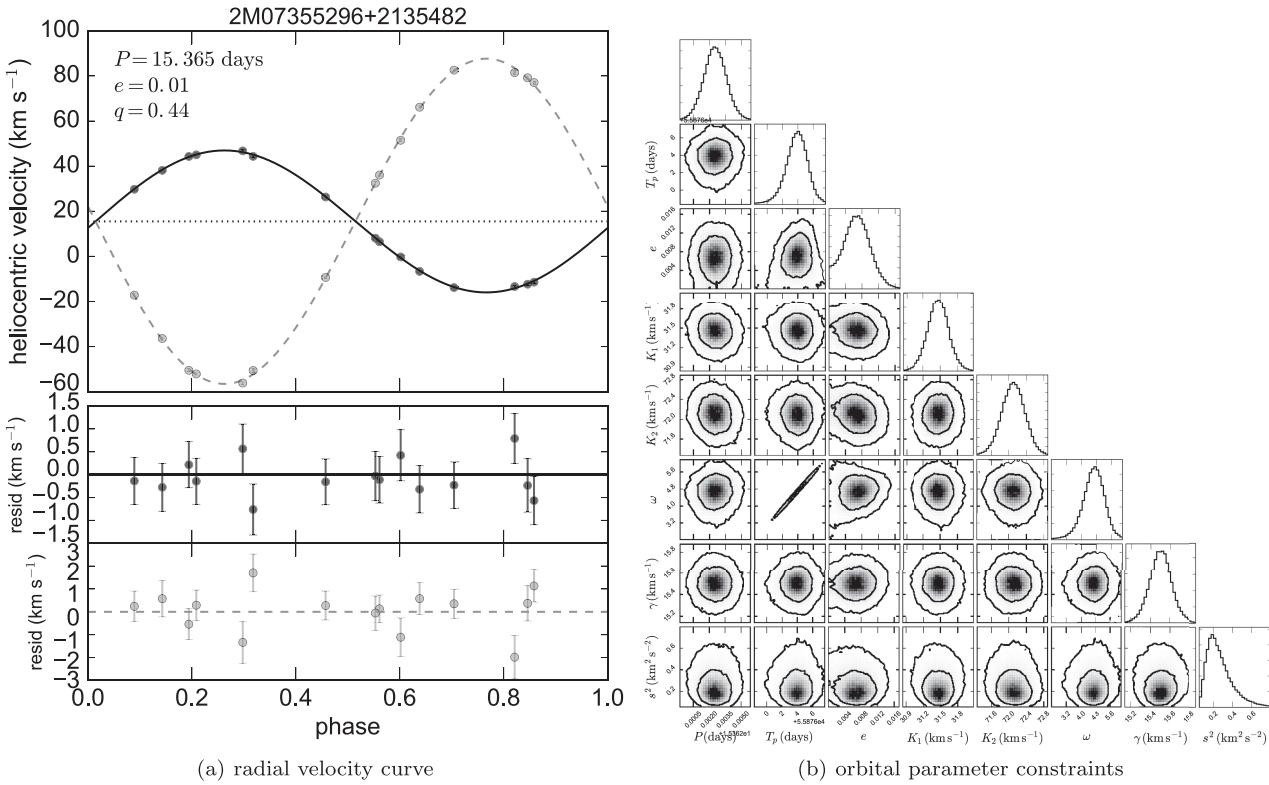
Fitting a Keplerian orbit amounts to simultaneously maximizing the likelihood of the RV curves of the primary and secondary, where the model RV for a given set of orbital parameters is obtained by

solving the two-body problem (Murray & Correia 2010). We use a custom PYTHON implementation of the adaptive simulated annealing algorithm (Iglesias-Marzoa, López-Morales & Jesús Arévalo Morales 2015) to obtain an initial maximum-likelihood estimate of the best-fitting orbital parameters and then sample the parameter space in the vicinity of the maximum likelihood with emcee (Foreman-Mackey et al. 2013) to estimate parameter uncertainties. We use non-informative, flat priors throughout. In addition to the seven standard Keplerian orbital parameters,<sup>10</sup> we fit a ‘jitter’ term,  $s^2$ , to allow for the possibility of intrinsic scatter in the RVs due to e.g. stellar pulsation or underestimated RV uncertainties (see e.g. Baluev 2009; Price-Whelan et al. 2017). The effective total uncertainties in the RVs used in fitting are then  $\sigma_{\text{tot},i}^2 = \sigma_i^2 + s^2$ , where  $\sigma_i$  are the RV uncertainties at each epoch found from bootstrapping. Explicitly, the log-likelihood function is

$$\ln \mathcal{L} = -\frac{1}{2} \sum_i^N \left\{ \frac{[v_r(t_i, \vec{\theta}_1) - v_{\text{Helio}1,i}]^2}{\sigma_{1,i}^2 + s^2} + \ln[2\pi(\sigma_{1,i}^2 + s^2)] \right. \\ \left. + \frac{[v_r(t_i, \vec{\theta}_2) - v_{\text{Helio}2,i}]^2}{\sigma_{2,i}^2 + s^2} + \ln[2\pi(\sigma_{2,i}^2 + s^2)] \right\}. \quad (9)$$

Here, the sum is over  $N$  epochs at times  $t_i$ ,  $v_r(t, \vec{\theta})$  represents the predicted RV at time  $t$  for a system with orbital parameters  $\vec{\theta}$  (Murray & Correia 2010), and  $\vec{\theta}_1 = (P, T_p, e, \omega, K_1, \gamma)$  and  $\vec{\theta}_2 = (P, T_p, e, \omega + \pi, K_2, \gamma)$  are the orbital parameters for each component. For most systems, the best-fitting jitter is small ( $s^2 \sim 0.1 \text{ km s}^{-2}$ ), indicating that our estimates of  $\sigma_i$  are reasonably accurate. However, for stars with large  $v_{\text{macro}}$ , which

<sup>10</sup> These include the period,  $P$ , periastron time,  $T_p$ , eccentricity,  $e$ , argument of periastron,  $\omega$ , centre-of-mass velocity,  $\gamma$ , and the velocity semi-amplitudes,  $K_1$  and  $K_2$ .



**Figure 10.** Left: Orbit fit to an SB2 systems with 15 epochs. In this relatively low mass ratio system, the secondary contributes only  $\sim 3$  per cent of the light, but we can still determine its velocity to  $\pm \sim 1 \text{ km s}^{-1}$ . Black and grey lines show the heliocentric velocity of the primary and secondary star. Right: 68 and 95 per cent marginalized probability regions for the system’s orbital parameters. Because the eccentricity is very nearly 0,  $T_p$ , and  $\omega$  are highly degenerate and are individually not well constrained, reflecting the fact that these quantities are meaningless for a circular orbit. We derive similar orbital solutions, shown in Table 1, for 64 systems.

indicates significant rotation, jitter is sometimes of order  $1 \text{ km s}^{-1}$ . This suggests that our velocity measurements are less accurate for rapidly rotating stars, which is also supported by our experiments with semi-empirical binary spectra (Appendix B1).

For some systems with mass ratios near 1, we found that it was initially impossible to obtain a good fit to the measured RVs because the velocity assignments of the two stars were switched in the fits for one or more visits. We attempted to fit these systems by allowing the fitting algorithm to switch the assigned velocities for individual visits if doing so would improve the fit. In most cases, this solved the problem. A few systems ( $\sim 5$  per cent of those with sufficient coverage) remained with RV curves that could not be well fitted by a Keplerian orbit, even with the possibility of switching the assigned velocities; these systems may have poorly measured RVs or contain an unseen component.

Fig. 10(a) shows an example orbital solution for a system with typical phase coverage, RV errors, and number of epochs. This system has the lowest mass ratio,  $q = 0.44$ , of the systems for which we derive orbital solutions. Because of the system’s low mass ratio, the secondary contributes only a small fraction ( $\sim 3$  per cent) of the total light in the spectrum; it is not obvious from visual inspection that more than one star contributes to the spectrum. However, the secondary is detected unambiguously by our model, and the fact that the primary and secondary velocities all fall on a Keplerian orbit confirms the validity of the detection.

Marginalized probabilities for the orbital parameters of this system are shown in Fig. 10(b). Most orbital parameters are well constrained for this system, without strong parameter covariances.

However, the periastron time,  $T_p$ , and argument of periastron,  $\omega$ , are highly degenerate, because the orbit is nearly circular (the eccentricity,  $e$ , is consistent with 0); the orbit has no well-defined periastron, and hence  $T_p$  and  $\omega$  are undefined. All systems with low eccentricities therefore have large uncertainties in  $\omega$  and in  $T_p$ , even when the meaningful parameters of the orbit are well constrained.

In Table 1, we provide best-fitting orbital parameters and marginalized uncertainties for 64 systems for which an orbital solution could be obtained. Most of these systems are ordinary double-lined binaries (SB2s), similar to the system shown in Fig. 3. However, we also include solutions for several close SB2s within hierarchical triples (similar to Fig. 6), as well as SB1s within hierarchical triples with hidden third components (similar to Fig. 8). These are fit very similarly to pure SB2 systems, with the only difference being that  $v_{\text{Heli}}$  measurements for individual stars at each visit are obtained from the ‘SB3’ and ‘SB2 with an unseen third object’ models. We only attempt to fit such systems if they are consistent with the third component having constant velocity over the observed baseline. For ‘SB2 with an unseen third object’ systems, the fit is to a single RV curve, so  $K_2$  is not measurable. The rms velocity residual for all orbital solutions ranges between  $0.04$  and  $1 \text{ km s}^{-1}$ . The systems with the largest velocity errors have (i) lower average S/N and (ii) higher  $T_{\text{eff}}$  and  $v_{\text{macro}}$ , both of which make it more difficult to accurately measure RVs.

The statistics  $U_N$  and  $V_N$  in the last column of Table 1 quantify the uniformity of coverage in orbital phase and velocity by the measured RV data. We calculate these statistics following Troup

**Table 1.** Orbital solutions for double-line spectroscopic binaries. We report the median and middle 68 per cent of the marginalized posterior samples for each parameter.  $U_N$  and  $V_N$  quantify the phase and velocity coverage of the observations (see Appendix D); systems with  $U_N V_N \lesssim 0.5\text{--}0.6$  are susceptible to erroneous bad fits.

2MASS ID	$N_{\text{epochs}}$	$P$ (d)	$T_p$ (BMJD)	$e$	$\omega$ (radians)	$K_1$ ( $\text{km s}^{-1}$ )	$K_2$ ( $\text{km s}^{-1}$ )	$\gamma$ ( $\text{km s}^{-1}$ )	$U_N V_N$
06212323+1701485	8	42.24 $\pm 0.19$ $\pm 0.17$	56657.15 $\pm 0.21$ $\pm 0.23$	0.3166 $\pm 0.0091$ $\pm 0.0092$	0.684 $\pm 0.029$ $\pm 0.034$	35.27 $\pm 0.31$ $\pm 0.29$	41.40 $\pm 0.32$ $\pm 0.31$	1.37 $\pm 0.19$ $\pm 0.18$	0.79
08544465+1130053	21	39.23855 $\pm 0.00097$ $\pm 0.00093$	55933.620 $\pm 0.016$ $\pm 0.016$	0.6932 $\pm 0.0014$ $\pm 0.0014$	1.0931 $\pm 0.0035$ $\pm 0.0035$	59.79 $\pm 0.20$ $\pm 0.19$	62.93 $\pm 0.20$ $\pm 0.21$	-6.818 $\pm 0.069$ $\pm 0.069$	0.92
04030722+5150045	9	69.973 $\pm 0.076$ $\pm 0.093$	55906.20 $\pm 0.38$ $\pm 0.33$	0.569 $\pm 0.045$ $\pm 0.036$	4.026 $\pm 0.024$ $\pm 0.026$	31.5 $\pm 2.9$ $\pm 1.9$	33.4 $\pm 3.1$ $\pm 2.1$	-7.176 $\pm 0.089$ $\pm 0.091$	0.76
21313924+1307507	41	1.5567964 $\pm 0.0000016$ $\pm 0.0000015$	55731.18 $\pm 1.10$ $\pm 0.19$	0.0016 $\pm 0.0016$ $\pm 0.0012$	1.03 $\pm 4.30$ $\pm 0.76$	59.51 $\pm 0.12$ $\pm 0.12$	71.84 $\pm 0.13$ $\pm 0.13$	-52.199 $\pm 0.069$ $\pm 0.061$	0.92
18470667-0226077 <sup>a</sup>	32	7.52676 $\pm 0.00015$ $\pm 0.00014$	55823.81 $\pm 0.39$ $\pm 0.37$	0.0136 $\pm 0.0046$ $\pm 0.0050$	5.38 $\pm 0.32$ $\pm 0.30$	45.86 $\pm 0.27$ $\pm 0.26$	55.85 $\pm 0.26$ $\pm 0.26$	15.99 $\pm 0.15$ $\pm 0.15$	0.93
07355296+2135482	15	15.3645 $\pm 0.0010$ $\pm 0.0011$	55879.8 $\pm 1.5$ $\pm 1.5$	0.0065 $\pm 0.0033$ $\pm 0.0034$	4.62 $\pm 0.60$ $\pm 0.62$	31.47 $\pm 0.17$ $\pm 0.18$	72.11 $\pm 0.27$ $\pm 0.27$	15.50 $\pm 0.11$ $\pm 0.10$	0.87
15010903+3702218	7	17.5079 $\pm 0.0011$	56090.854 $\pm 0.046$ $\pm 0.045$	0.2996 $\pm 0.0014$ $\pm 0.0013$	2.5573 $\pm 0.0092$ $\pm 0.0088$	41.962 $\pm 0.100$ $\pm 0.092$	57.13 $\pm 0.14$ $\pm 0.14$	-47.938 $\pm 0.040$ $\pm 0.036$	0.61
08541894+1239291	22	1.3046792 $\pm 0.0000087$ $\pm 0.0000089$	55904.366 $\pm 0.054$ $\pm 0.052$	0.042 $\pm 0.010$ $\pm 0.010$	2.96 $\pm 0.25$ $\pm 0.24$	130.2 $\pm 1.7$ $\pm 1.7$	130.2 $\pm 1.8$ $\pm 1.8$	7.87 $\pm 0.86$ $\pm 0.80$	0.91
19303146+4210508 <sup>b</sup>	24	5.55412 $\pm 0.00013$ $\pm 0.00014$	56444.01 $\pm 0.25$ $\pm 0.22$	0.0120 $\pm 0.0029$ $\pm 0.0030$	6.05 $\pm 0.28$ $\pm 0.24$	43.20 $\pm 0.13$ $\pm 0.13$	- $\pm 0.24$	-58.76 $\pm 0.092$ $\pm 0.100$	-
08464223+1205302	17	15.0232 $\pm 0.0023$ $\pm 0.0024$	56654.146 $\pm 0.023$ $\pm 0.023$	0.2980 $\pm 0.0034$ $\pm 0.0035$	0.290 $\pm 0.010$ $\pm 0.011$	24.76 $\pm 0.13$ $\pm 0.12$	26.54 $\pm 0.14$ $\pm 0.14$	-6.493 $\pm 0.058$ $\pm 0.055$	0.93
...	...	...	...	...	...	...	...	...	...

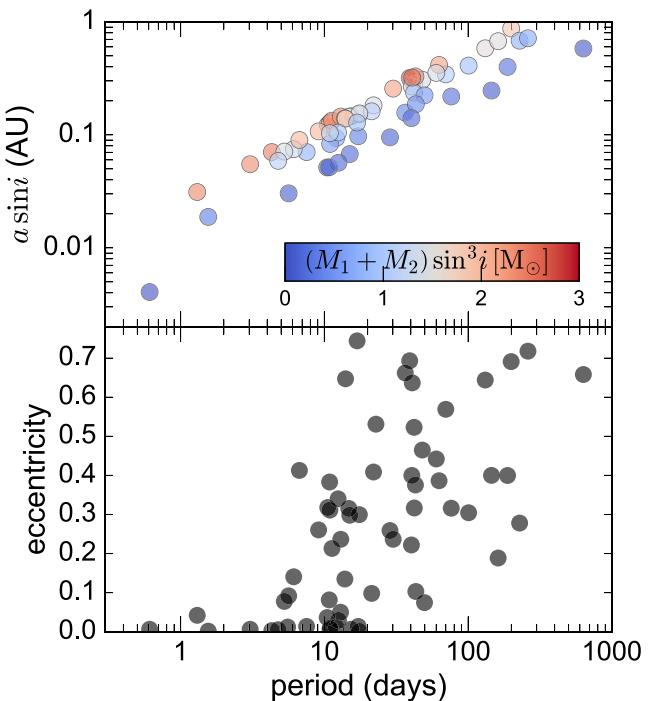
Notes. <sup>a</sup>System is an SB2 within a hierarchical triple (10 systems). <sup>b</sup>System is an SB1 within a hierarchical triple (three systems). This table is available in its entirety (with orbital solutions for 64 systems) in machine-readable form.

et al. (2016, their equations 22 and 23); both  $U_N$  and  $V_N$  are bounded between 0 and 1, with values near 1 corresponding to uniform phase and velocity coverage. Troup et al. (2016) estimated that orbital parameters are unreliable for SB1s if  $U_N V_N < 0.5$ ; of course, the probability of recovering the correct orbit also depends on the number of RV measurements and their uncertainties. In Appendix D, we carry out tests with synthetic RV data to determine the number of epochs and phase + velocity coverage required for reliable orbit recovery of SB2s with RV data similar to that obtained for real binaries.

In the top panel of Fig. 11, we plot constraints on the semimajor axes and component masses derived from the orbital parameters of all systems for which we present an orbital solution.<sup>11</sup> Our sample contains systems with periods ranging from 0.6 d (short enough that the two stars are nearly touching, with  $a \sin i \sim 0.8 R_\odot$ ) to  $\sim 600$  d. Dynamical constraints on the absolute masses of stars in individual binaries are weak due to the degeneracy with  $\sin i$ , but the highest lower limit on the mass of an individual component is  $\sim 1.5 M_\odot$ . This corresponds to  $T_{\text{eff}} \gtrsim 6600$  K for a solar-metallicity star on our adopted MIST isochrones; reassuringly, none of our dynamical mass constraints imply  $T_{\text{eff}} > 7000$  K, which is the upper limit adopted for our spectral model.

We plot the periods and eccentricities of binary systems in the bottom panel of Fig. 11. Most of the short-period ( $P \lesssim 10$  d)

<sup>11</sup>We calculate  $a \sin i$ ,  $M_1 \sin^3 i$ , and  $M_2 \sin^3 i$  using the standard formulae from Cox (2000). It is not possible to measure  $a$  or  $M$  directly from RV data alone; we note that future astrometric constraints can break the degeneracy between these quantities and orbital inclination for nearby systems (Halbwachs et al. 2017a).



**Figure 11.** Top: distribution of periods, semimajor axes, and masses for the 64 double-lined binary systems for which we derive an orbital solution. Due to APOGEE's relatively rapid cadence (most targets have maximum baselines of a few months), these systems are heavily biased towards short periods. Bottom: period-eccentricity distribution. Most systems with  $P \lesssim 10$  d have low eccentricity due to tidal circularization.

systems have nearly circular orbits, likely due tidal dissipation processes (e.g. Koch & Hrvnak 1981). However, a few systems with short periods do have eccentricity constraints that are inconsistent with zero. Some of these systems are short-period binaries within a hierarchical triple; such systems are known to be susceptible to eccentricity boosts via three-body interactions (Kozai 1962).<sup>12</sup>

### 3.5 Are binaries gravitationally bound?

Our method finds binaries and triples by identifying targets in which more than one star falls within a single APOGEE fibre and contributes to the observed spectrum. In all cases where the velocities of the components of a suspected binary or triple system are not observed to vary in a correlated way, there is no guarantee that all the components are gravitationally bound: chance alignments of stars at different distances that fall within the same fibre can produce spectra consistent with binarity.

To estimate the false-positive rate due to such ‘optical binaries’ that are not gravitationally bound, we analyse mock photometric catalogues created with GALAXIA (Sharma et al. 2011). GALAXIA implements the Besançon model of stellar population synthesis (Robin et al. 2003) to populate the Galactic distribution function and produce realistic mock surveys along arbitrary lines of sight. Using a Galactic dust extinction map computed by `mwdust` (Bovy et al. 2016), we produced mock catalogues complete to  $J = 14$  mag along three lines of sight representative of the range of stellar densities spanned by different APOGEE fields: one towards the bulge with  $(\ell, b) = (0 \text{ deg}, 5 \text{ deg})$ , one towards the Galactic anticentre with  $(\ell, b) = (180 \text{ deg}, 0 \text{ deg})$ , and one at high latitude with  $(\ell, b) = (0 \text{ deg}, 60 \text{ deg})$ , where  $\ell$  and  $b$  represent Galactic longitude and latitude. We then checked, for each star in a mock catalogue, whether any other stars fall within a circular aperture of diameter 2 arcsec centred on that star, representing a single APOGEE fibre. If more than one star was found in a given aperture (including both dwarfs and giants), we classified all stars in that aperture as a single optical binary.

Towards the Galactic Bulge, we find a 0.2 per cent probability that a star is an optical binary. The same probability is 0.05 per cent towards the Galactic anticentre and 0.005 per cent at high latitude. As these probabilities are all much smaller than our detected binary fraction of ~15 per cent, we conclude that optical binaries are unlikely to be a large source of false positives, though a small fraction of the systems we detected in fields towards the Bulge may be chance alignments masquerading as binaries.

Our model requires all components of a multiple-star system to have identical distances and abundances and fall on a single isochrone. This is likely a reasonable assumption for true, gravitationally bound binaries (Desidera et al. 2004; Andrews, Chanamé & Agüeros 2017), but it is unlikely to hold for chance alignments. One could thus distinguish between true binaries and chance alignments by allowing the stellar parameters and abundances, and/or relative distance, of the secondary to vary freely and identifying cases where

<sup>12</sup>The only short-period system which is distinctly non-circular and is not best fit by a triple model is 2M21320320+1107560, with  $P = 6.70$  d,  $e = 0.41$ , and 41 epochs. It may well also be part of a hierarchical triple in which the third (long-period) component is too faint to appreciably contribute to the spectrum. Such a system would not be identifiable as having an unseen companion, as only systems in which the unseen companion is in the short-period sub-binary have velocities inconsistent with being an isolated binary.

the best-fitting model assigns significantly different abundances or distances to the different components. We defer such analysis to future work.

## 4 DISCUSSION AND CONCLUSIONS

### 4.1 Comparison to previous work

Chojnowski et al. (2015) compiled a catalogue of double-lined spectroscopic binaries and triples in APOGEE by identifying targets whose cross-correlation function exhibited multiple peaks.<sup>13</sup> Of the 610 targets in their catalogue that were also in our initial sample, 574 were also classified as multiple systems by our pipeline, 5 were classified as SB1s, and 31 were classified as consistent with being single stars. Of the 574 stars classified as multiple systems by both pipelines, 514 are in our ‘potential close binary’ subsample, which contains variable-velocity targets and binaries with large velocity offsets between the two components (Section 2.4). This ~95 per cent agreement rate is encouraging, given the very different approaches of the two pipelines. A primary advantage of the method developed in this work is its increased sensitivity to long-period systems with negligible velocity offsets.

Recently, Badenes et al. (2017) studied the occurrence rate of short-period, velocity-variable binaries in APOGEE. They found the multiplicity fraction for main-sequence stars to be a factor of ~2 higher in the lowest metallicity tercile of their sample than in the highest metallicity tercile. We find a similar result: for short-period systems (those with velocity shifts of at least  $10 \text{ km s}^{-1}$  between epochs), the multiplicity fraction is ~60 per cent higher for the lowest metallicity tercile of our sample ( $[\text{Fe}/\text{H}] < -0.21$ ) than for the highest metallicity tercile ( $[\text{Fe}/\text{H}] > -0.02$ ). Badenes et al. (2017) studied systems with metallicities as low as  $[\text{Fe}/\text{H}] = -2.5$ ; given the smaller range of metallicities in our sample ( $[\text{Fe}/\text{H}] > -1$ ), these results are likely consistent. For long-period systems, we find the binary fraction to be consistent with being constant with metallicity. Some theoretical models (e.g. Machida 2008) predict that low-metallicity clouds should preferentially form short-period binaries, consistent with this result. However, we caution against overinterpreting this finding, as we have not attempted to quantify or correct for changes in the completeness of our method at lower metallicity.

Modelling approaches similar to the method developed in this work have previously been used on a case-by-case basis to fit ‘composite spectrum binaries’, a term that refers specifically to binaries containing a cool giant primary and a hot subgiant or main-sequence secondary (e.g. González & Levato 2006; Griffin & Griffin 2010). Similar techniques have also been employed to spectroscopically detect and characterize unresolved binaries composed of very low-mass stars or brown dwarfs with different spectral types (Burgasser 2007; Burgasser et al. 2008). Although this work focuses on modelling the spectra of binaries in which both components are main-sequence stars, the method we develop is flexible and can be straightforwardly extended to identify other flavours of binaries. The primary requirement is a robust training set spanning the range of single-star spectral types found in the data set of interest.

For systems known to be double-lined binaries, a wide variety of techniques have been developed to disentangle the spectra of the two-component stars in order to measure their individual velocities

<sup>13</sup>Their catalogue is available at <http://astronomy.nmsu.edu/drewski/apogee-sb2/apSB2.html>.

and stellar labels (e.g. Bagnuolo & Gies 1991; Simon & Sturm 1994; Hadrava 1995; Pavlovski & Hensberge 2010; Czekala et al. 2017). These techniques can reliably separate the spectra of the individual components of a binary even when lines are blended, but they generally require multi-epoch spectroscopy that captures the combined binary spectrum at several orbital configurations.

If only single-epoch spectroscopy is available or the binary is sufficiently wide that the orbital velocities of the two components do not change much between visits, the most common approach for measuring RVs is cross-correlation with a composite template spectrum (Zucker & Mazeh 1994; Halbwachs et al. 2017b); this requires first estimating the labels of the individual stars. In such cases, most previous works have attempted to first model the primary star with a synthetic or empirical template, then subtract this template from the composite spectrum, and finally fit a model for the second star to the residual spectrum. However, it is difficult to ensure with this approach that the optimal binary model has been found, as the single-star model spectrum that best fits the combined binary spectrum does not in general correspond to the true best-fitting parameters of the primary star (see E18). The method introduced in this work, which fits for the stellar parameters and velocities of both components simultaneously, avoids these complications.

Recently, a few works have shown that double-lined binaries can also be detected non-parametrically by identifying systems with peculiar spectra that are clustered outliers in a high-dimensional space of arbitrary summary statistics computed for all spectra collected by a survey (Traven et al. 2017; Reis et al. 2017). While such methods thus far primarily identify binaries with large velocity offsets, we note that non-parametric methods can likely be further optimized for binary identification by searching for targets which are precisely the kind of outliers expected to result from binarity; for example, one could identify systems that cannot be well described by a single combination of spectral PCA components but can be well described by two sums of components with different velocities.

## 4.2 Future prospects

### 4.2.1 Improving the model

A straightforward way to make our model sensitive to a larger fraction of the binary population is to extend the single-star model to cooler temperatures. As discussed in Appendix B1, the lower limit of  $T_{\text{eff}} = 4200$  K, which is due to shortcomings of ab-initio spectral models for main-sequence stars at lower temperatures, limits the model to only detecting binaries with mass ratios near  $q = 1$  at low temperatures and prevents us from fitting spectra of the coolest stars altogether. Fitting cooler stars does not required any modification of our general approach, only a robust training set at lower temperatures, which currently does not readily exist. Due to the increased importance of molecular opacity from many species at lower  $T_{\text{eff}}$ , it may be helpful to include more abundances in the spectral model for cooler stars.

Our model could also be improved by fitting for the projected rotation velocity  $v \sin i$  explicitly instead of subsuming it under the Gaussian broadening of  $v_{\text{macro}}$ , since the single-star model currently performs worst for rapidly rotating stars. This is in principle simple to accomplish: rotation velocities for stars in the training set can be obtained straightforwardly in post-processing (e.g. Díaz et al. 2011), and the inferred  $v \sin i$  can then be added as an additional label to the model. Rotation is currently not an important problem for most of the targets in our sample because stars with  $T_{\text{eff}} \lesssim 6500$  K typically lose most of their angular momentum to magnetic

braking and do not rotate rapidly on the main sequence (Glebocki, Gnacinski & Stawikowski 2000; Schatzman 1962), and stars with  $T_{\text{eff}} \geq 6500$  K represent less than 4 per cent of our data set. However, an improved treatment of rotation would make it possible to better model hot stars and would likely decrease the false-positive rate (see Appendix C). This is particularly true for young stars, which can rotate significantly even at cooler temperatures (e.g. Terrien et al. 2014).

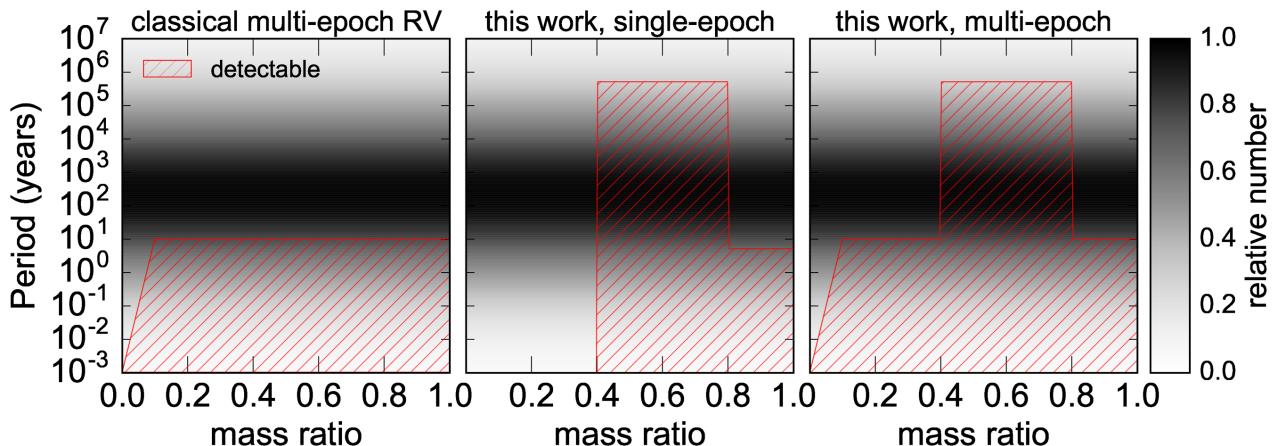
### 4.2.2 Hierarchical modelling

Beyond the Solar neighbourhood, previous spectroscopic studies of the Galactic binary population have been limited to studying the short-period tail of the binary population. Because the model presented in this work does not depend on RV variability or a line-of-sight velocity offset to detect binaries, it has the potential to substantially improve on existing constraints on the binary population of the Milky Way and/or its satellites when combined with a model for detection completeness and the survey selection function.

Existing RV surveys of the Milky Way and nearby dwarf galaxies are sensitive to binaries with periods less than  $\sim(1\text{--}10)$  yr (e.g. Matijević et al. 2011; Minor 2013; Hettinger et al. 2015; Gao et al. 2017; Badenes et al. 2017). For the lognormal period distribution for solar-type stars found in the Solar neighbourhood (Duchêne & Kraus 2013),  $\sim$ 73 per cent percent of binaries have  $P > 10$  yr; most of these systems will be missed by such surveys. The most probable period for solar-type binaries is  $\sim$ 300 yr; assuming random orbit orientations, the typical line-of-sight velocity separation for such systems is  $\Delta v_{\text{los}} \sim 2 \text{ km s}^{-1}$ , and the average RV change over a one-year baseline is  $\sim 0.02 \text{ km s}^{-1}$ . This is an order of magnitude below the detectability thresholds of existing large spectroscopic surveys, though such weak RV trends in SB1s may be marginally detectable with high-dispersion spectrographs typically used to study exoplanets (Konacki 2005; Katoh et al. 2013).

Irrespective of RV variability, long-period binaries with favourable mass ratios can be detected with our model as long as both components fall within a single spectroscopic fibre. At a distance of 1 kpc, more than 80 per cent of solar-type binaries will have projected separations of less than 1 arcsec, so that both stars would fall with a single 2-arcsec fibre; this fraction increases at larger distances. On the other hand, for long-period systems, the binary spectral model is sensitive only to intermediate mass ratio systems ( $0.4 \lesssim q \lesssim 0.8$ ), in which the primary and secondary have qualitatively different spectral types, but the secondary still contributes a non-negligible fraction of the total light (see Appendix B1 and E18). The distribution of mass ratios for solar-type binaries is approximately flat down to  $q = 0.1$  (Duchêne & Kraus 2013), so the binary model will miss many high and low mass ratio systems with long periods.

We summarize the sensitivity of our method, as well as standard binary-detection methods based on velocity variability, to systems with different periods and mass ratios in Fig. 12. RV variability can probe essentially all mass ratios, but only for the short-period tail of the binary population. On the other hand, fitting a binary spectral model to single-epoch observations can probe most of the period distribution, but only for a restricted subset of mass ratios. We thus expect that the method developed here can be fruitfully combined with existing multi-epoch RV measurements from SB1s, such as the APOGEE constraints on the short-period binary fraction presented in Badenes et al. (2017) and measurements of the binary fractions of nearby dwarf galaxies presented by Minor (2013). This would enable a full hierarchical model for binary populations that is sensitive to an unprecedented range of periods and mass ratios.



**Figure 12.** Schematic illustration of the range of binary periods and mass ratios that can be detected with different methods. Grey shading (identical in all panels) shows the distribution of periods and mass ratios for solar-type stars; hatches show the regions of parameter space that can be probed by RV variability (left) and fitting a binary model to single- and multi-epoch spectra (middle, right). Conventional multi-epoch RV surveys are sensitive to essentially all mass ratios, but only for short-period binaries, which represent roughly a third of the observed lognormal period distribution for solar-type stars. The binary spectral model introduced in this work is sensitive to all but the longest periods (as long as both stars fall within one spectroscopic fibre) with single-epoch observations, but only for intermediate mass ratio systems. When multi-epoch spectra are available, fitting a binary model can also detect all systems with variable RVs as SB1s.

An immediate advantage of our method is that it is sensitive to a large fraction of the binary population even when only single-epoch observations are available. With multi-epoch observations, our model can detect short-period systems as SB1s, with similar sensitivity to traditional methods. Our modelling approach can also be straightforwardly applied to spectra from other surveys. The precise range of mass ratios to which it is sensitive will vary with wavelength coverage: surveys at optical wavelengths will be more sensitive to binaries with higher mass ratios ( $0.8 \lesssim q \lesssim 0.9$ ; see E18) due to the increased spectral information content at shorter wavelengths, but they will be less sensitive at low  $q$  because a cooler secondary star contributes a greater fraction of a binary system's total light in the near-infrared than at optical wavelengths.

In this work, we fit normalized spectra and only used the CMD to assess the reliability of our spectral model. A promising avenue for future work is to fit spectra and photometry simultaneously, or to place a photometric prior on  $q$ . This would make it possible to detect systems with  $q \sim 1$  and negligible velocity offsets, which are twice as luminous as they would be if they were a single star. Particularly with improved parallaxes from future *Gaia* data releases, photometric constraints could substantially extend the fraction of the binary population to which our method is sensitive.

#### 4.3 Summary

We have developed a flexible data-driven method for identifying and fitting the spectra of multiple-star systems and have applied it to  $\sim 20\,000$  main-sequence targets from the APOGEE survey. Unlike most previous work, our model performs well even for long-period systems in which the line-of-sight velocity offset between components is negligible, substantially expanding the fraction of the binary population that can be probed by observations. Our method is mostly automated and can be straightforwardly applied to other spectroscopic surveys with modest adjustments. Our main results are as follows:

(i) *Spectral identification of long- and short-period binaries:* unresolved binaries can be identified as systems whose spectrum can be better fit by a sum of two single-star model spectra falling on a

single isochrone than any single-star model (Fig. 1). For systems with mass ratios  $0.4 \lesssim q \lesssim 0.8$ , in which the two stars have different spectral types, binaries can be identified spectroscopically even in the limit of no velocity offset and with only single-epoch observations. Spectral signatures of binarity are strengthened in the presence of a velocity offset of order one resolution element or greater (Fig. 2); thus, close binaries can be detected even in the limit of  $q \sim 1$ .

(ii) *Photometric test of the model:* nearly all spectroscopically identified binaries with accurate distance measurements fall above the main sequence on the CMD, as is predicted for true binaries, and triple systems fall above most binaries (Fig. 9). Photometry does not enter our binary identification procedure, so this agreement with theoretical predictions provides independent validation of our spectral model.

(iii) *Dynamical mass ratios:* for short-period binaries in which the velocities of the two components change substantially between visits, it is possible to obtain a dynamical measurement of the mass ratio from the relative changes in the stars' RVs between visits (Fig. 3). This provides a constraint on the mass ratio that is independent of the spectral mass ratio, which determines the contribution of the secondary star to the spectrum. We find good agreement between spectral and dynamical mass ratios, with a median difference of 0.048 and even better agreement for systems with high S/N spectra (Fig. 4).

(iv) *Triple systems:* we identify 114 systems in which the contributions of three stars can be identified in the spectrum (Fig. 6) and an additional 108 in which only two stars contribute significantly to the spectrum, but the presence of a third component can be inferred from its gravitational effects (Fig. 8). Most identified triples are hierarchical, consisting of a close binary orbited by a third component with a much longer period; we have verified that these systems are all likely gravitationally bound (Fig. 7).

(v) *Orbital solutions:* for double-lined systems with a sufficient number of epochs and well-sampled RV curves, we derive full Keplerian orbital solutions (Fig. 10b); some of these systems are close binaries within hierarchical triples. We derive orbital solutions for 64 binaries with periods ranging from  $\sim 0.6$  d to  $\sim 2$  yr and semi-major axes ranging from  $\sim R_\odot$  to  $\sim 1$  au. Consistent with previous

studies, we find that most binaries with  $P \lesssim 10$  d have eccentricity consistent with 0 due to tidal circularization processes (Fig. 11).

We make catalogues of best-fitting labels for all identified multiple-star systems publicly available; these are described in Appendix E. A public version of the code used for fitting spectra in this work is available at <https://github.com/kareemelbadry/binspec>.

## ACKNOWLEDGEMENTS

We are grateful to the anonymous referee for a constructive report. We thank Gaspard Duchêne, Keith Hawkins, Jessica Lu, Hans-Günter Ludwig, Adrian Price-Whelan, and Silvia Toonen for helpful conversations. We are grateful to Jan Rybizki for assistance in creating mock catalogues with GALAXIA, and to Anna Ho for help with *the Cannon*. This project was developed in part at the 2017 Heidelberg Gaia Sprint, hosted by the Max-Planck-Institut für Astronomie, Heidelberg. KE acknowledges support from the SFB 881 program (A3), a Berkeley Fellowship, a Hellman award for graduate study, and an NSF graduate research fellowship. HWR received support from the European Research Council under the European Union’s Seventh Framework Programme (FP 7) ERC Grant Agreement no. [321035]. YST is supported by the Australian Research Council Discovery Program DP160103747, the Carnegie-Princeton Fellowship and the Martin A. and Helen Chooljian Membership from the Institute for Advanced Study at Princeton. EQ is supported in part by a Simons Investigator Award from the Simons Foundation. DRW is supported by a fellowship from the Alfred P. Sloan Foundation. CC acknowledges support from NASA grant NNX15AK14G, NSF grant AST-1313280, and the Packard Foundation. The analysis in this paper relied on the PYTHON packages NUMPY (Van Der Walt, Colbert & Varoquaux 2011), MATPLOTLIB (Hunter 2007), and ASTROPY (Astropy Collaboration et al. 2013).

## REFERENCES

- Andrews J. J., Chanamé J., Agüeros M. A., 2017, *MNRAS*, 473, 5393  
 Astropy Collaboration et al., 2013, *A&A*, 558, A33  
 Badenes C. et al., 2017, *ApJ*, preprint ([arXiv:1711.00660](https://arxiv.org/abs/1711.00660))  
 Baguñolo W. G., Jr., Gies D. R., 1991, *ApJ*, 376, 266  
 Baluev R. V., 2009, *MNRAS*, 393, 969  
 Bovy J., 2016, *ApJ*, 817, 49  
 Bovy J., Rix H.-W., Green G. M., Schlafly E. F., Finkbeiner D. P., 2016, *ApJ*, 818, 130  
 Branch M. A., Coleman T. F., Li Y., 1999, *SIAM J. Sci. Comput.*, 21, 1  
 Burgasser A. J., 2007, *AJ*, 134, 1330  
 Burgasser A. J., Liu M. C., Ireland M. J., Cruz K. L., Dupuy T. J., 2008, *ApJ*, 681, 579  
 Choi J., Dotter A., Conroy C., Cantiello M., Paxton B., Johnson B. D., 2016, *ApJ*, 823, 102  
 Chojnowski S. D. et al., 2015, in American Astronomical Society Meeting Abstracts, 340.05  
 Cox A. N., 2000, Allen’s Astrophysical Quantities, 4th edn. Springer-Verlag, New York  
 Czekala I., Mandel K. S., Andrews S. M., Dittmann J. A., Ghosh S. K., Montet B. T., Newton E. R., 2017, *ApJ*, 840, 49  
 Desidera S. et al., 2004, *A&A*, 420, 683  
 Díaz C. G., González J. F., Levato H., Grosso M., 2011, *A&A*, 531, A143  
 Duchêne G., Kraus A., 2013, *ARA&A*, 51, 269  
 Duquennoy A., Mayor M., 1991, *A&A*, 248, 485  
 El-Badry K., Rix H.-W., Ting Y.-S., Weisz D. R., Bergemann M., Cargile P., Conroy C., Eilers A.-C., 2018, *MNRAS*, 473, 5043  
 Fernandez M. A. et al., 2017, *PASP*, 129, 084201  
 Ford E. B., Kozinsky B., Rasio F. A., 2000, *ApJ*, 535, 385  
 Foreman-Mackey D., Hogg D. W., Lang D., Goodman J., 2013, *PASP*, 125, 306  
 Gao S., Zhao H., Yang H., Gao R., 2017, *MNRAS*, 469, L68  
 García Pérez A. E. et al., 2016, *AJ*, 151, 144  
 Glebocki R., Gnacinski P., Stawikowski A., 2000, *Acta Astron.*, 50, 509  
 González J. F., Levato H., 2006, *A&A*, 448, 283  
 Griffin R. E. M., Griffin R. F., 2010, *MNRAS*, 402, 1675  
 Hadrava P., 1995, *A&AS*, 114, 393  
 Halbwachs J.-L. et al., 2017a, preprint, ([arXiv:1710.02017](https://arxiv.org/abs/1710.02017))  
 Halbwachs J.-L., Kiefer F., Arenou F., Famaey B., Guillout P., Ibata R., Mazeh T., Pourbaix D., 2017b, in Reyl C.é, Di Matteo P., Herpin F., Lagadec E., Lançon A., Meliani Z., Royer F., eds, SF2A-2017: Proceedings of the Annual meeting of the French Society of Astronomy and Astrophysics, Paris, p. 79  
 Hettinger T., Badenes C., Strader J., Bickerton S. J., Beers T. C., 2015, *ApJ*, 806, L2  
 Ho A. Y. Q. et al., 2017, *ApJ*, 836, 5  
 Holtzman J. A. et al., 2015, *AJ*, 150, 148  
 Hunter J. D., 2007, *Comput. Sci. Eng.*, 9, 90  
 Hurley J., Tout C. A., 1998, *MNRAS*, 300, 977  
 Iglesias-Marzoa R., López-Morales M., Jesús Arévalo Morales M., 2015, *PASP*, 127, 567  
 Katoh N., Itoh Y., Toyota E., Sato B., 2013, *AJ*, 145, 41  
 Koch R. H., Hrivnak B. J., 1981, *AJ*, 86, 438  
 Konacki M., 2005, *ApJ*, 626, 431  
 Kozi Y., 1962, *AJ*, 67, 591  
 Kurucz R. L., 1970, Smithsonian Astrophysical Observatory Spec. Rep., 309  
 Kurucz R. L., 1979, *ApJS*, 40, 1  
 Kurucz R. L., 1993, Kurucz CD-ROM No. 18. SYNTHE Spectrum Synthesis Programs and Line Data. Cambridge, Mass  
 Li C., de Grijs R., Deng L., 2013, *MNRAS*, 436, 1497  
 Machida M. N., 2008, *ApJ*, 682, L1  
 Majewski S. R. et al., 2017, *AJ*, 154, 94  
 Matijević G. et al., 2010, *AJ*, 140, 184  
 Matijević G. et al., 2011, *AJ*, 141, 200  
 Merle T. et al., 2017, *A&A*, 608, A95  
 Michalik D., Lindegren L., Hobbs D., 2015, *A&A*, 574, A115  
 Minor Q. E., 2013, *ApJ*, 779, 116  
 Moe M., Di Stefano R., 2017, *ApJS*, 230, 15  
 Murray C. D., Correia A. C. M., 2010, Keplerian Orbits and Dynamics of Exoplanets. Univ. Arizona Press, Tucson, AZ, p. 15  
 Naoz S., Farr W. M., Lithwick Y., Rasio F. A., Teyssandier J., 2013, *MNRAS*, 431, 2155  
 Ness M., Hogg D. W., Rix H.-W., Ho A. Y. Q., Zasowski G., 2015, *ApJ*, 808, 16  
 Nidever D. L. et al., 2015, *AJ*, 150, 173  
 Pavlovski K., Hensberge H., 2010, in Prša A., Zejda M., eds, ASP Conf. Ser. Vol. 435, Binaries - Key to Comprehension of the Universe. Astron. Soc. Pac., San Francisco, p. 207  
 Pourbaix D. et al., 2004, *A&A*, 424, 727  
 Price-Whelan A. M., Hogg D. W., Foreman-Mackey D., Rix H.-W., 2017, *ApJ*, 837, 20  
 Raghavan D. et al., 2010, *ApJS*, 190, 1  
 Reid I. N., Gizis J. E., 1997, *AJ*, 113, 2246  
 Reis I., Poznanski D., Baron D., Zasowski G., Shahaf S., 2017, *MNRAS*, preprint ([arXiv:1711.00022](https://arxiv.org/abs/1711.00022))  
 Robin A. C., Reylé C., Derrière S., Picaud S., 2003, *A&A*, 409, 523  
 Schatzman E., 1962, *Ann. Astrophys.*, 25, 18  
 Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, *ApJ*, 730, 3  
 Simon K. P., Sturm E., 1994, *A&A*, 281, 286  
 Simons D. A., Henry T. J., Kirkpatrick J. D., 1996, *AJ*, 112, 2238  
 Terrien R. C. et al., 2014, *ApJ*, 782, 61  
 Ting Y.-S., Rix H.-W., Conroy C., Ho A. Y. Q., Lin J., 2017, *ApJ*, 849, L9  
 Toonen S., Hamers A., Portegies Zwart S., 2016, *Comput. Astrophys. Cosmol.*, 3, 6

- Traven G. et al., 2017, *ApJS*, 228, 24  
 Troup N. W. et al., 2016, *AJ*, 151, 85  
 Van Der Walt S., Colbert S. C., Varoquaux G., 2011, *Comput. Sci. Eng.*, preprint ([arXiv:1102.1523](https://arxiv.org/abs/1102.1523))  
 Zasowski G. et al., 2013, *AJ*, 146, 81  
 Zucker S., Mazeh T., 1994, *ApJ*, 420, 806

## SUPPORTING INFORMATION

Supplementary data are available at [MNRAS](https://mnras.org) online.

[Table\\_E5\\_all\\_SB3\\_labels.csv](#)

[Table\\_E4\\_all\\_SB2s\\_hidden\\_third\\_component\\_labels.csv](#)

[Table\\_E3\\_all\\_binary\\_star\\_labels.csv](#)

[Table\\_E2\\_all\\_SB1\\_labels.csv](#)

[Table\\_E1\\_all\\_single\\_star\\_ids.csv](#)

[Table\\_1\\_orbital\\_solutions.csv](#)

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

## APPENDIX A: NEURAL NETWORK SPECTRAL MODEL

As mentioned in Section 2.1, we use a neural network to predict the normalized flux density at a given wavelength pixel as a function of stellar labels. As applied in this work, a neural network is essentially a flexible function produced through the composition of simple functions. It takes as its argument a vector of labels  $[\vec{\ell};$  equation (1)] and returns the normalized flux density predicted at a particular wavelength pixel.

For the neural network used in this work, which contains a single hidden layer with five neurons, the normalized flux density at wavelength pixel  $\lambda$  is given by

$$\hat{f}_\lambda = \tilde{w}_\lambda^i \sigma(w_{\lambda,i}^k \hat{\ell}_k + b_{\lambda,i}) + \tilde{b}_\lambda \quad (\text{A1})$$

with implied summation over  $k = 1, \dots, N_{\text{labels}}$  and  $i = 1, \dots, N_{\text{neurons}}$ . Here,  $\hat{\ell} = (\vec{\ell} - \vec{\ell}_{\min}) / (\vec{\ell}_{\max} - \vec{\ell}_{\min}) - 0.5$  is a scaled label vector,  $\vec{\ell}_{\max}$  and  $\vec{\ell}_{\min}$  are vectors of the maximum and minimum values of each label in the training set, and  $\sigma(z) = 1/(1 + e^{-z})$  is the ‘sigmoid’ activation function. The *weights*,  $w$  and  $\tilde{w}$ , and *biases*,  $b$  and  $\tilde{b}$ , parametrize the neural network; these are the free parameters that are adjusted during training.

In order to treat spectra with different line-of-sight velocities, all spectra are shifted to rest frame and linearly interpolated on to a common wavelength grid. Training the model consists of minimizing a *loss function*, comparable to the  $\chi^2$  statistic, that quantifies how well the model can fit the training set. We use an L1 loss function, which minimizes the total absolute difference between fluxes predicted by the neural network and those in the training set. We expect this to perform better than e.g. the  $\chi^2$  statistic during the iterative cleaning of the training set and retraining, since it is less sensitive to outliers. During training, we mask all pixels with  $\text{S/N} < 50$ , bad or missing pixels, and pixels with poor sky subtraction.

We implement and train the neural network using the PYTHON package PyTorch. We tested a wide range of network architectures, varying the network depth, width, and activation function, with both data-driven and synthetic spectral models. We find that using a small neural network and a large training set is the most straightforward way to prevent overfitting; using a substantially larger network with more neurons or hidden layers causes the model to reach lower losses (i.e. fit the training set better), but perform

worse in cross-validation. We verified that our spectral model performs equally well on the training and cross-validation sets at fixed S/N, so it does not overfit the training set.

An advantage of using a neural network spectral model is that the neural network’s flexibility makes it possible to model a wide range of stellar parameters in a single model, rather than stitching together multiple models covering different regions of label space. However, our basic approach of constructing a binary spectral model does not depend critically on use of a neural network; a comparable binary model could likely be built from other forms of single-star model (e.g. *The Cannon*).

## APPENDIX B: MODEL SELECTION

Because the single-star model is a special case of the binary model, it is *always* possible to obtain a binary model that fits a data spectrum at least as well as does the best-fitting single-star model. As the binary model is more complex than the single-star model, with three additional free parameters, one might expect to find a better fit, in a  $\chi^2$  sense, with the binary model even for targets which are true single stars. It is therefore necessary to formulate a heuristic to determine ‘how much’ better a fit with a binary model is required to constitute reliable evidence in favour of the binary model.

The primary statistic used for model selection is the  $\chi^2$  difference,  $\Delta\chi^2 = \chi^2_{\text{single}} - \chi^2_{\text{binary}}$ , which simply quantifies how much better a fit is obtained by the binary model. We also calculate a second statistic, the ‘improvement fraction’  $f_{\text{imp}}$ , to quantify how much better a fit the binary model achieves relative to how different it is from the single-star model. The basic idea here is that if a binary model spectrum is very different from the single-star model, but only achieves a slightly better fit to the data, this constitutes weaker evidence in favour of the binary model than a case with comparable  $\Delta\chi^2$  in which most of the difference between the best-fitting binary and single-star model goes towards improving the fit. The improvement fraction is defined as

$$f_{\text{imp}} = \frac{\sum \{ (|\hat{f}_{\lambda,\text{single}} - \hat{f}_{\lambda}| - |\hat{f}_{\lambda,\text{binary}} - \hat{f}_{\lambda}|) / \hat{\sigma}_{\lambda} \}}{\sum \{ |\hat{f}_{\lambda,\text{single}} - \hat{f}_{\lambda,\text{binary}}| / \hat{\sigma}_{\lambda} \}}, \quad (\text{B1})$$

where  $\hat{f}_{\lambda}$  and  $\hat{\sigma}_{\lambda}$  are the normalized flux density and corresponding uncertainty,  $\hat{f}_{\lambda,\text{single}}$  and  $\hat{f}_{\lambda,\text{binary}}$  are the best-fitting normalized single-star and binary model spectra, and the sum is over all wavelength pixels.

Our full acceptance criterion for preferring the binary model is given in Table B1. These thresholds were motivated in part by the  $\Delta\chi^2$  and  $f_{\text{imp}}$  values calculated for semi-empirical binaries as described below, and in part by validation with the CMD (Fig. 9). The adopted thresholds are conservative, and prevent us from identifying some binaries whose spectra can only be marginally better fit by a binary model; however, setting them to substantially lower values causes the model to begin categorizing more targets near the main sequence of the CMD as binaries, indicating a non-negligible false-positive rate. We have not tuned the minimum  $f_{\text{imp}}$  value for each  $\Delta\chi^2$ ; we set the intuitive requirement that a higher  $f_{\text{imp}}$  should be required for systems with a lower  $\Delta\chi^2$ .

Different  $\Delta\chi^2$  thresholds are required for systems identified as potential close binaries, because (i) for these systems, we fit multiple visit spectra simultaneously, and (ii) we fit a total of five different models (see Section 2.4.1). For these systems, we begin with a fiducial threshold of  $\Delta\chi^2 = 300 \times N_{\text{epochs}}$  for each increase in model complexity, where  $N_{\text{epochs}}$  is the number of visit spectra fit simultaneously; i.e. we require  $\chi^2_{\text{single star}} - \chi^2_{\text{SB1}} > 300N_{\text{epochs}}$  for a system

**Table B1.** Minimum  $\Delta\chi^2$  and improvement fraction  $f_{\text{imp}}$  (equation B1) for a target to be classified as a binary. All systems with  $\Delta\chi^2 < 300$ , and all systems falling below the minimum  $f_{\text{imp}}$  for a given  $\Delta\chi^2$ , are classified as inconclusive; i.e. showing no strong evidence of binarity.

$\Delta\chi^2 = \chi^2_{\text{single}} - \chi^2_{\text{binary}}$	Minimum $f_{\text{imp}}$
$\Delta\chi^2 \geq 3000$	0
$2500 \leq \Delta\chi^2 < 3000$	0.05
$2000 \leq \Delta\chi^2 < 2500$	0.075
$1500 \leq \Delta\chi^2 < 2000$	0.1
$1000 \leq \Delta\chi^2 < 1500$	0.125
$750 \leq \Delta\chi^2 < 1000$	0.15
$600 \leq \Delta\chi^2 < 750$	0.175
$450 \leq \Delta\chi^2 < 600$	0.2
$300 \leq \Delta\chi^2 < 450$	0.225

to be initially classified as an SB1;  $\chi^2_{\text{SB1}} - \chi^2_{\text{SB2}} > 300N_{\text{epochs}}$  for a system to be initially classified as an SB2, etc. We then inspected the spectra of these targets individually and reclassified suspected false positives (see Appendix C).

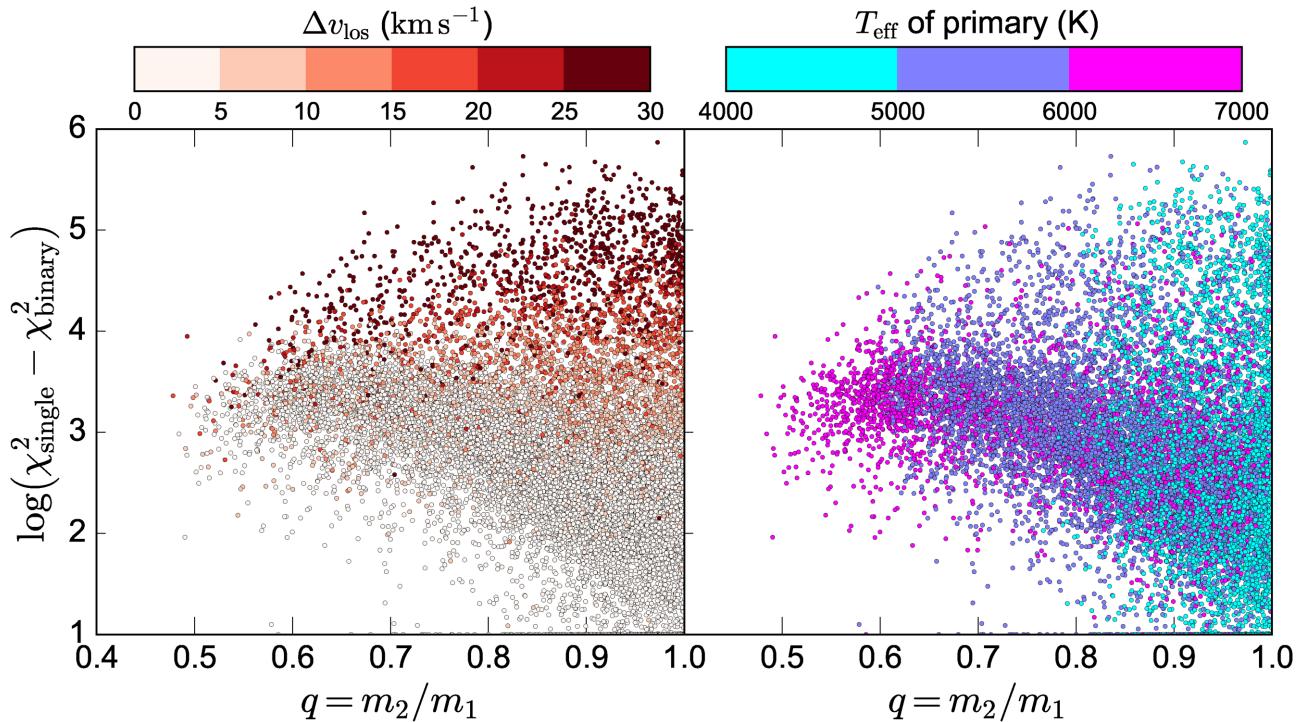
We experimented with generating single-star spectra directly from the spectral model, adding noise, and fitting them with both binary and single-star models. This produces typical  $\Delta\chi^2$  values of order unity, which is smaller than we find for the majority of APOGEE targets. This occurs because spectra generated in this way can necessarily be perfectly fit by the single-star spectral model, which is not necessarily the case for real spectra. We quantify the  $\Delta\chi^2$  values expected for real binary spectra in the next section.

## B1 Tests with semi-empirical synthetic binary spectra

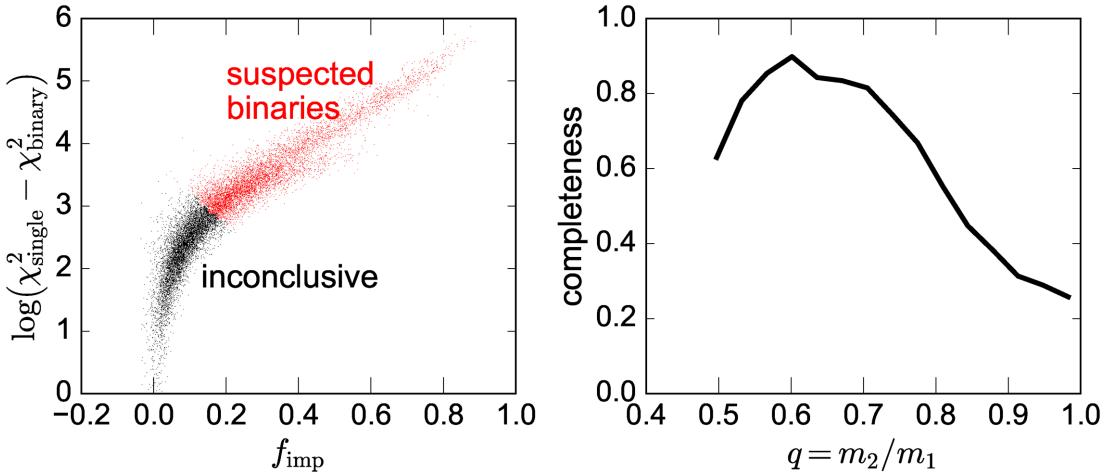
To assess the accuracy and potential systematics of our method, and to measure the expected  $\Delta\chi^2$  values for true binaries with a particular mass ratio, we construct a library of  $\sim 15\,000$  ‘semi-empirical’ synthetic binary spectra. These are created by combining randomly chosen pairs of APOGEE spectra in un-normalized space, following the method outlined in Section 2.2. We then normalize and fit these spectra following the same procedure used to fit real spectra. An advantage of constructing synthetic binary spectra by combining real spectra (as opposed to simply generating binary spectra from the data-driven model) is that this accounts for the possibility that the model does not capture all the variance in the real spectra; this is likely the case for our model, which only contains five labels.

We require that the two stars used to construct a semi-empirical binary spectrum have similar abundances (within 0.05 dex in [Fe/H] and  $[\alpha/\text{Fe}]$ ) and fall within 0.03 dex in  $\log g$  of a single isochrone. We only combine spectra consistent with being single stars; i.e. those which cannot be significantly better fit by a binary model than a single-star model according to the thresholds in Table B1. We assign realistic orbital parameters to each system following E18, drawing orbital periods from the lognormal period distribution for solar-type stars from Duchêne & Kraus (2013) and assuming random orbit orientations and phases. Results from fitting these semi-empirical binary spectra are shown in Figs B1–B4.

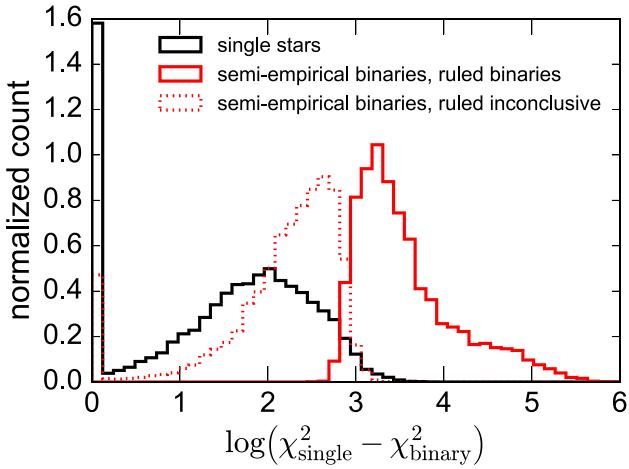
In Fig. B1, we show the  $\chi^2$  difference in favour of the binary model,  $\Delta\chi^2 = \chi^2_{\text{single}} - \chi^2_{\text{binary}}$ , as a function of the mass ratio  $q$ , line-of-sight velocity offset  $\Delta v_{\text{los}}$ , and  $T_{\text{eff}}$  of the primary star. As expected,  $\Delta\chi^2$  is a strong function of  $q$ : most binaries with  $q \lesssim 0.75$



**Figure B1.** Results of fitting semi-empirical binary spectra with single-star and binary models. Semi-empirical binary spectra are created by adding together flux-calibrated APOGEE spectra of two stars with similar abundances. At  $q \lesssim 0.8$ , most semi-empirical binary spectra can be significantly better fitted with a binary model, with  $\Delta\chi^2 \gtrsim 1000$ . The  $\chi^2$  difference is nearly always larger for systems with large velocity offsets; for systems with  $q \sim 1$ , only binaries with  $\Delta v_{\text{los}} \gtrsim 10 \text{ km s}^{-1}$  have  $\Delta\chi^2 \gtrsim 1000$ . At fixed  $q$  and  $\Delta v_{\text{los}}$ , the typical  $\Delta\chi^2$  is larger for systems with cooler primaries. Due to our single-star spectral model’s minimum  $T_{\text{eff}}$  of 4200 K, low mass ratio systems can only be modelled for hot primaries.



**Figure B2.** Results for fitting semi-empirical synthetic binary spectra. Left-hand panel shows the  $\chi^2$  difference and improvement fraction [equation (B1)]; systems passing the adopted acceptance criterion for a binary candidate to be considered legitimate are plotted in red. Right-hand panel shows the resulting completeness function; i.e. the fraction of semi-empirical binary systems at a given  $q$  that are successfully identified as binaries.



**Figure B3.**  $\chi^2$  differences between best-fitting single and binary models. Black histogram shows suspected single stars from which semi-empirical binary spectra are constructed. Solid (dotted) red histogram shows semi-empirical binary spectra which pass (fail) the  $\Delta\chi^2$  and  $f_{\text{imp}}$  binary acceptance criterion.

have  $\Delta\chi^2 > 1000$ , while most systems with  $q \sim 1$  have much lower  $\Delta\chi^2$ . This is expected, because the two stars in binaries with  $q \sim 1$  will have similar spectra, making the combined binary spectrum indistinguishable from that of either single star unless there is a sizable velocity offset between the two stars. The left-hand panel of Fig. B1 shows that  $\Delta\chi^2$  is also a strong function of the velocity offset  $\Delta v_{\text{los}}$ : at fixed  $q$ , systems in which the velocity offset is larger nearly always have larger  $\Delta\chi^2$ . In particular, most binaries with  $\Delta v_{\text{los}} \gtrsim 10 \text{ km s}^{-1}$  have  $\Delta\chi^2 > 1000$ , even at  $q \sim 1$ . Indeed, among binaries with large velocity offsets, the typical  $\Delta\chi^2$  is largest for systems with  $q \sim 1$ ; in such systems, both stars contribute significantly to the spectrum, and absorption lines are obviously split.

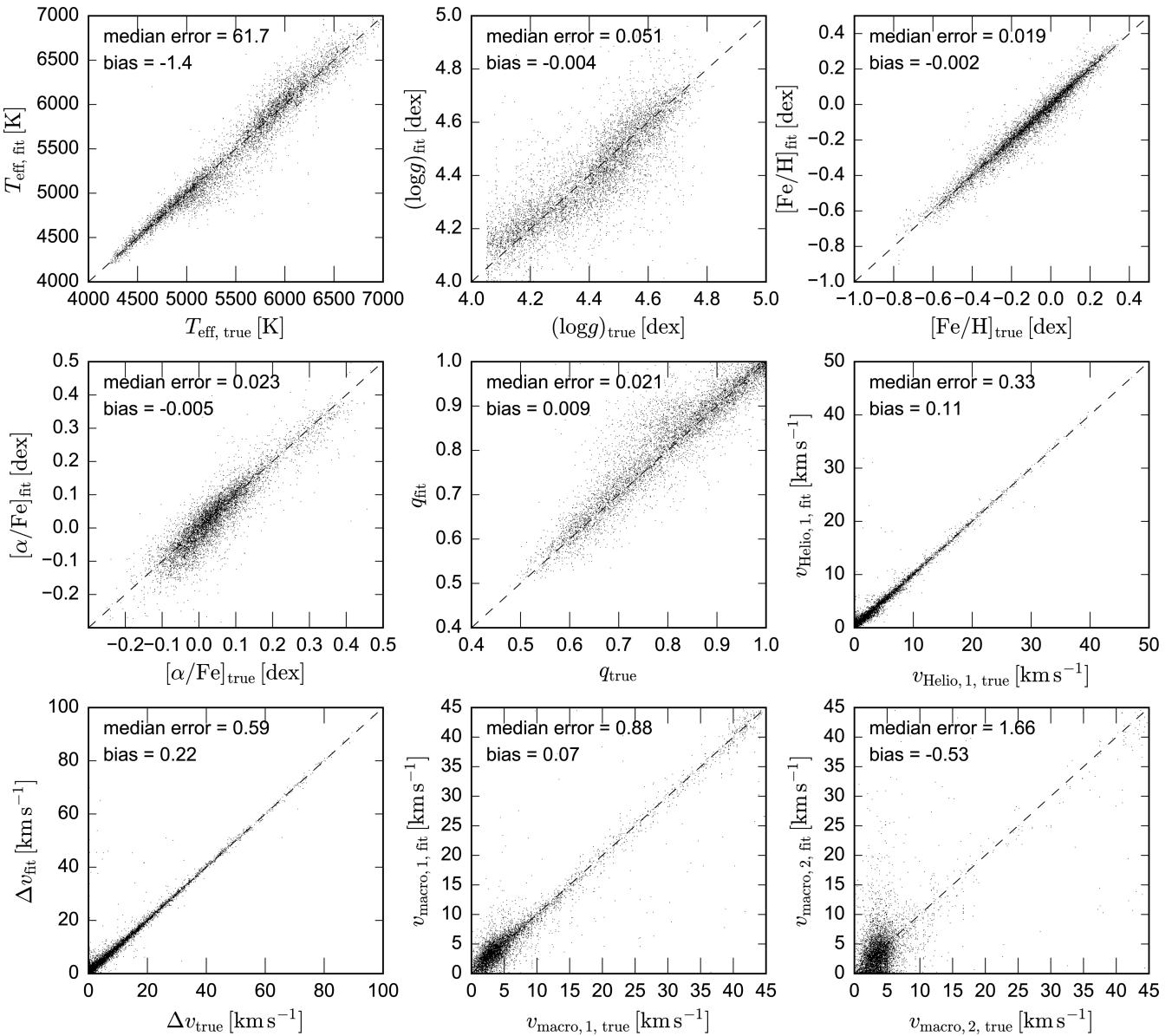
The right-hand panel of Fig. B1 shows the dependence of  $\Delta\chi^2$ , and the range of mass ratios to which our method is sensitive, on  $T_{\text{eff}}$  of the primary. At fixed mass ratio and  $\Delta v$ , the median  $\Delta\chi^2$  is slightly lower for hot stars ( $T_{\text{eff}} > 6000 \text{ K}$ ), particularly for systems

with  $q \sim 1$  and large  $\Delta v$ . This occurs because lines are on average weaker and more rotationally broadened in hot stars, reducing the information content of the spectrum. This panel also shows that, due to the minimum  $T_{\text{eff}}$  of 4200 K of the single-star spectral model, the minimum mass ratio that can be modelled varies with  $T_{\text{eff}}$  of the primary. This means that our completeness is higher for hot stars than for cool stars.

To determine whether our binary model fit has converged on the true globally optimal model rather than a local  $\chi^2$  minimum, we check whether the  $\chi^2$  value of the best-fitting binary model is at least as low as that corresponding to the binary model with the true labels of the system. We find that our fit converges on the globally optimal solution for  $\sim 99$  per cent of all semi-empirical binaries. About half of the systems in which the fit converges on a local minimum are binaries with  $q \sim 1$  in which the velocity assignments of the primary and secondary stars are switched; in these cases, the derived stellar labels are still reasonably accurate.

In Fig. B2, we show how our adopted model selection threshold translates to the range of mass ratios to which the model is sensitive. In the left-hand panel, we plot the distribution of semi-empirical binaries in  $\Delta\chi^2 - f_{\text{imp}}$  space.  $\Delta\chi^2$  and  $f_{\text{imp}}$  are correlated: most systems whose spectrum can be significantly better fit by a binary model (high  $\Delta\chi^2$ ) also have high  $f_{\text{imp}}$ . In the right-hand panel, we plot the fraction of semi-empirical binaries at a given mass ratio that pass our adopted model selection criteria to be considered reliable binary candidates. As expected, this ‘completeness’ function is a strong function of mass ratio. Most binaries with  $0.55 \lesssim q \lesssim 0.75$  have spectra that are sufficiently different from any single-star model that they can be unambiguously identified as binaries; at higher mass ratios, the spectra of the two-component stars become similar, so only the  $\sim 20$  per cent of binaries with  $\Delta v_{\text{los}} \gtrsim 10 \text{ km s}^{-1}$  can be detected. At sufficiently low mass ratios, the secondary contributes a negligible fraction of the total light.

We emphasize that this completeness function does *not* represent our global completeness function for all binaries, for two reasons. First, the population of binaries in our semi-empirical library is not statistically representative of the Galactic binary population: because our single-star spectral model cannot model stars with  $T_{\text{eff}} < 4200 \text{ K}$ , we cannot currently model low mass ratio binaries in which the primary is cool, as the secondary will have



**Figure B4.** Label recovery diagnostic for semi-empirical binary spectra that pass our  $\Delta\chi^2$  and  $f_{\text{imp}}$  threshold to be considered real detections. For each label  $\ell_i$ , inset text indicates the median error,  $\text{med}(|\ell_i, \text{fit} - \ell_i, \text{true}|)$ , and bias,  $\text{med}(\ell_i, \text{fit} - \ell_i, \text{true})$ .

$T_{\text{eff}} < 4200$  K (see Section 2.2). Second, we have made no attempt to model the APOGEE selection function, which would complicate the distribution of  $T_{\text{eff}}$  at a given  $q$ . For the particular set of semi-empirical binaries analysed here,  $\sim 50$  per cent of all binary systems pass the model selection threshold to be characterized as binaries. However, this is not representative of the global sensitivity of our model, since we make no attempt to use a realistic distribution of mass ratios in the semi-empirical library: the distribution of mass ratios in our semi-empirical binary library is skewed towards  $q = 1$ , which is precisely the regime in which the model performs poorly.

In Fig. B3, we compare the distribution of  $\Delta\chi^2$  values for the suspected single stars used in constructing the semi-empirical binary library to those for semi-empirical synthetic binary spectra. Systems for which  $0 < \Delta\chi^2 < 1$  are assigned  $\log \Delta\chi^2 = 0$  on this plot; each histogram is normalized separately. The median  $\Delta\chi^2$  is  $\sim 50$  for single stars,  $\sim 200$  for semi-empirical binaries that fail the detection

criteria to be considered reliable binaries, and  $\sim 2600$  for semi-empirical binaries that pass the threshold to be considered reliable. There is some overlap in the distribution of  $\chi^2$  values for single stars and semi-empirical binaries, since for systems with small  $\Delta v_{\text{los}}$ , the binary model in principle transitions smoothly to the single-star model both as  $q \rightarrow 1$  and  $q \rightarrow 0$ . However, for binaries with favourable mass ratios, the typical  $\Delta\chi^2$  is more than an order of magnitude greater than for single stars.

### B1.1 Cross-validation

In Fig. B4, we compare the best-fitting labels for all semi-empirical binaries that pass the binary detection threshold to the true labels used in constructing the semi-empirical binary spectra. We set the ‘true’ abundance for each binary as the average of the abundances of the two single stars, which we required to be within 0.05 dex for  $[\text{Fe}/\text{H}]$  and  $[\alpha/\text{Fe}]$ .  $T_{\text{eff}}$  and  $\log g$  refer to the primary. These

semi-empirical binaries were not used to train the spectral model, so this experiment constitutes cross-validation of the binary spectral model. In each panel, we indicate the median signed error (bias) and absolute error (scatter) in the best-fitting label. Overall, this experiment reveals that labels inferred from fitting our binary model are reasonably precise: the true and best-fitting labels fall near the one-to-one line, with small scatter. The only label for which this is not obviously true is  $v_{\text{macro},2}$ ; this occurs primarily because only hot stars have  $v_{\text{macro}} \gtrsim 10 \text{ km s}^{-1}$ , so only the few binaries in which both stars are hot have non-negligible  $v_{\text{macro},2}$ .

The median error in the best fit  $q$  for the semi-empirical binaries is 0.021, which is smaller than the median difference of 0.048 between  $q_{\text{dyn}}$  and  $q_{\text{spec}}$  found for real binaries in Fig. 4. This is not unexpected, because (i)  $q_{\text{dyn}}$  also has non-zero uncertainty, and (ii) ‘ $q_{\text{true}}$ ’ for the semi-empirical binaries is calculated with the same isochrones used in the model from which  $q_{\text{fit}}$  is obtained. That is, the median difference of 0.021 does not account for uncertainties in the isochrones; the larger difference of 0.048 does, because  $q_{\text{dyn}}$  is independent of isochrones.

We emphasize that while Fig. B4 shows our derived stellar labels to be *precise*, this does not guarantee that they are *accurate*. The reason for this is that the uncertainties in stellar parameters obtained from spectral fitting are often dominated by systematic uncertainties in the model, which enter primarily from errors in the ‘ground truth’ labels of the training and validation sets, and are not accounted for in cross-validation. The cross-validation errors therefore are reasonable estimates of the precision of the model, but represent lower limits on the absolute uncertainties.

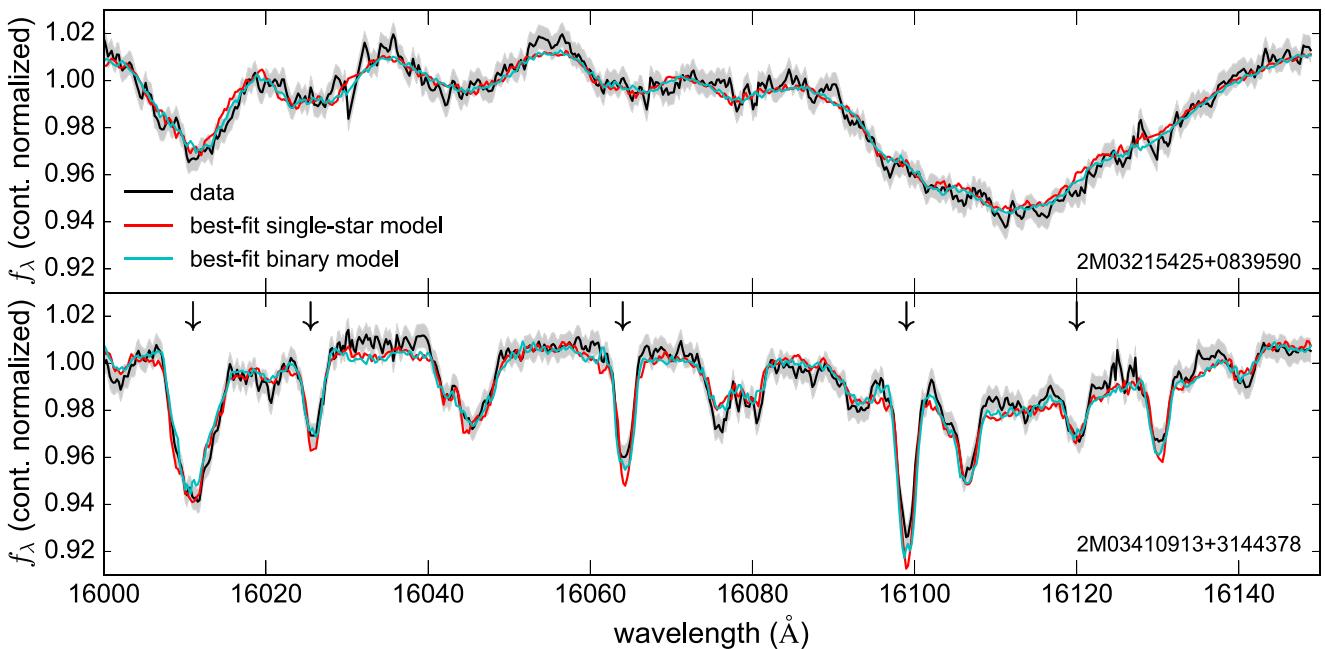
## APPENDIX C: FALSE POSITIVES

As discussed above, there are some targets for which our fitting and model selection formally prefers a binary model but visual

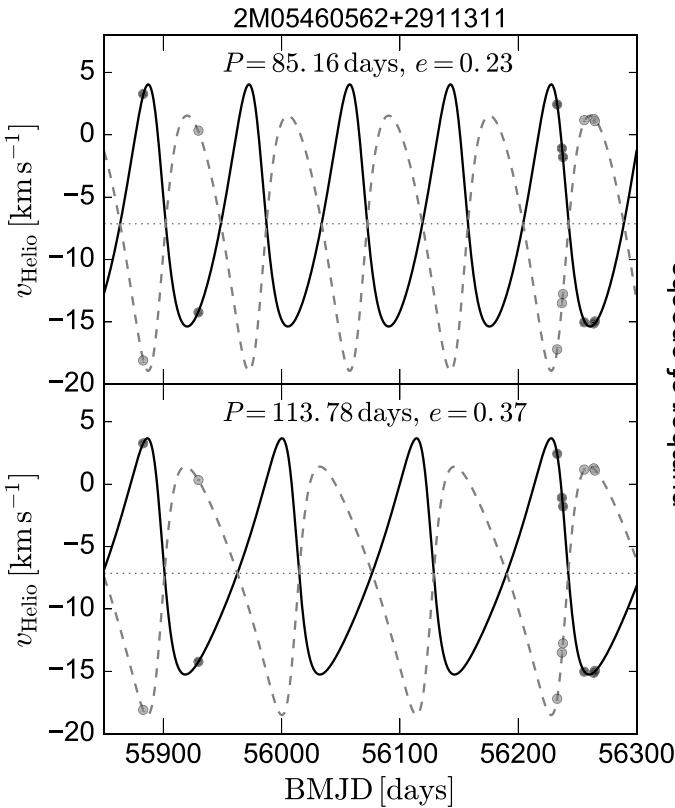
inspection of the spectrum reveals that the evidence in favour of the binary model is weak and a single-star model should likely be preferred. After visually inspecting all  $\sim 3300$  targets for which our model selection thresholds preferred a model other than a single star, we flagged  $\sim 300$  targets as probable false positives. The vast majority of such cases are hot stars ( $T_{\text{eff}} \gtrsim 6500 \text{ K}$ ) exhibiting significant rotational broadening.

In Fig. C1, we show example spectra of two stars flagged as false positives. The top panel shows the spectrum of a hot, rotating star. Such spectra are less informative than spectra of cooler stars with less rotational broadening; most metallic lines are intrinsically weaker, since more species are ionized at higher  $T_{\text{eff}}$ , and the profiles of individual lines are smeared out due to rotation. Although the binary model is formally preferred, with  $\Delta\chi^2 \sim 1000$ , it does not obviously fit the profiles of individual lines better than the single-star model. Indeed, the main difference between the binary and single-star models is that the binary model produces slightly wider blended lines.

Most of the stars we flag as false positives have spectra similar to this target and large  $v_{\text{macro}}$ . There is in principle no reason to expect the method model to generically fail for such targets; the most likely explanation is that our five-label single-star spectral model is not sufficiently complex to fully characterize the results of rotation. More than half of these targets are APOGEE telluric standard stars, which are selected to be used for telluric correction because they have nearly featureless spectra. Because hot, rotationally broadened stars constitute only a small fraction of our initial sample, we simply visually inspected and removed questionable targets by hand. A straightforward method for automating this procedure in future work would be to remove all telluric standards and/or targets for which a Fourier transform of the spectrum reveals little power at high frequencies.



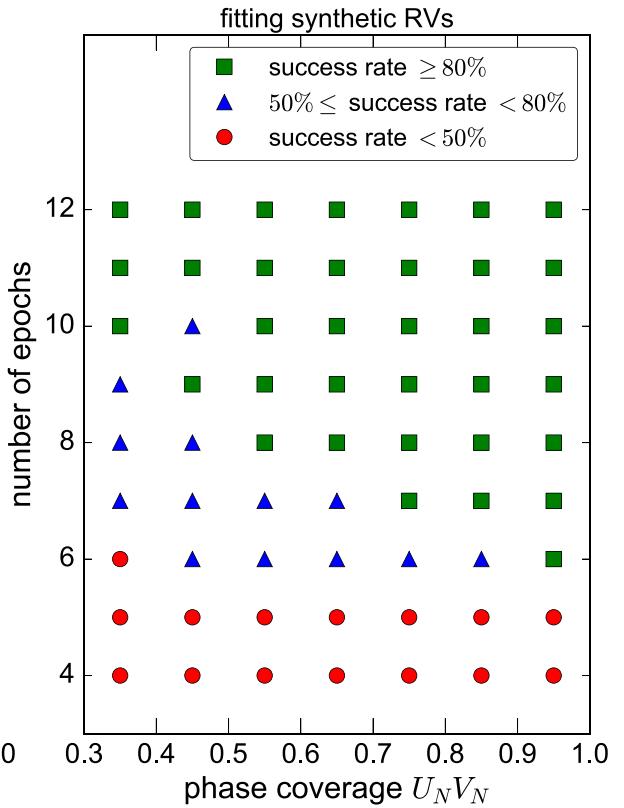
**Figure C1.** Spectra of two stars flagged as candidate binaries in our initial fitting that are likely false positives. Top: nearly featureless spectrum of a hot ( $T_{\text{eff}} \approx 7000 \text{ K}$ ), rotationally broadened star. The main difference between the single-star and binary models (which has  $q = 1$  and  $\Delta v_{\text{los}} = 40 \text{ km s}^{-1}$ ) is that the binary model produces slightly broader lines. This target is an APOGEE telluric standard. Bottom: although the binary model (with  $q = 1$  and  $\Delta v_{\text{los}} = 25 \text{ km s}^{-1}$ ) is formally preferred, it has many marginally split lines, marked with arrows, which are not present in the data spectrum. This target is a young star in an embedded cluster.



**Figure D1.** Left: example APOGEE binary system with eight epochs in which the available velocity data are not sufficient to fully constrain the orbit. Top and bottom panels show two qualitatively different orbital solutions which can both fit the measured velocity data (identical in the two panels) well. Because the measured velocities are sparse, the posterior is multimodal. We do not provide orbital solutions for such systems. Right: results of fitting synthetic RV data for an SB2 population with realistic orbital parameters and survey cadence. The ‘success rate’ indicates the fraction of systems in each cell of  $N_{\text{epochs}} - U_N V_N$  space for which our fitting procedure converged on the true orbit from which RV measurements were generated. Orbits can usually be reliably constrained for  $N_{\text{epochs}} \gtrsim 7$ , though systems with poor phase coverage can still converge on incorrect orbital solutions. We only attempt to fit orbits for systems with  $N_{\text{epochs}} \geq 7$  and  $U_N V_N > 0.5$ .

The bottom panel of Fig. C1 shows the spectrum of a false-positive candidate that is not featureless. The binary spectral model achieves a fit that is formally better than the single-star model ( $\Delta\chi^2 \sim 1500$ ), but unlike the single-star model, the binary model spectrum has noticeably split lines as a result of the velocity offset between the two components. Because the data spectrum does not show such split line profiles, the binary model fit is likely erroneous and formally preferred only because it produces wider and shallower lines than the single-star model can accommodate. False positives like this one are rare; only a few dozen spectra are identified in which the binary model produces a less realistic line profile than the single-star model despite achieving a formally better fit. A significant fraction of these are targets in young embedded clusters. Young stars often exhibit spectral features that are uncommon in older stars, such as chromospheric emission and increased rotation in cooler stars. Our method is more susceptible to incorrectly preferring a binary model if the spectrum cannot be well accommodated by the single-star model; this is likely the case for these targets. In future studies, such false positives can potentially be eliminated by using a more complex spectral model for single stars and ensuring that different varieties of ‘unusual’ spectra are represented in the training set.

Besides reclassifying stars flagged as false positives upon visual inspection to be single stars, we also reclassified some of the potential close binary targets (Section 2.4) from one multiple-star class



to another. For example, we reclassified SB2 systems for which the best-fitting velocities of the primary and secondary fall on a one-to-one line with *positive* slope as SB1s, and SB3s with broad lines that were not obviously better fit by the SB3 model than the SB2 model as SB2s. We attempted to be conservative in classifying systems as triples; i.e. some triple systems are likely miscategorized as binaries, but all systems classified as triples are unambiguously better fit by the SB3 model than the SB2 model.

#### APPENDIX D: ORBIT-FITTING CONVERGENCE

Sparse RV data can often be well fitted by several families of qualitatively different orbits, particularly when there are few RV measurements and/or phase coverage is poor (e.g. Price-Whelan et al. 2017). Due to the complex multimodal structure of the posterior in these cases, standard MCMC techniques fail to fully explore the orbital parameter space in finite time, meaning that there is no guarantee that our orbit-fitting procedure will converge on the true solution or that the orbital parameter uncertainties found by EMCEE are reliable. In the left-hand panel of Fig. D1, we show an example APOGEE binary in which the RV data are not sufficiently constraining to yield an unambiguous orbital solution: (at least) two qualitatively different orbital solutions can fit the measured RVs.

To assess the number of epochs and phase + velocity coverage required for reliable orbit constraints, we generated synthetic RV measurements for a population of synthetic binaries (as described in E18) with  $4 < N_{\text{epochs}} < 12$  and phase coverage  $0.3 < U_N V_N < 1$  and fit them using the same procedure described in Section 3.4. To ensure realistic survey cadence, we drew observation times from observations of real APOGEE stars. We added Gaussian noise to the synthetic data with  $\sigma_{\text{RV}} = 0.2 \text{ km s}^{-1}$ , which is typical for our observations.

The results of this experiment are shown in Fig. D1. We label each fit successful if the true orbital parameters fall within the marginalized 90 percent credibility regions returned by MCMC fitting. Although it is in principle possible to constrain the orbit of an SB2 with as few as five RV epochs,<sup>14</sup> reliable constraints for can only be obtained from realistic data with  $\gtrsim 7$  epochs. At fixed number of epochs, constraints are more reliable for systems with larger  $U_N V_N$ , as expected. Based on the results of this experiment, we only attempt to fit orbits to systems with  $N_{\text{epochs}} \geq 7$  and  $U_N V_N \geq 0.5$ . All but two of the targets for which we provide an orbital solution have  $U_N V_N > 0.6$ ; we caution that orbital parameters for systems with lower  $U_N V_N$  may be less reliable.

We stress that the orbit of an SB2 can in general be constrained with fewer RV measurements than that of an SB1. The basic reason for this is that even with only a few epochs, having velocity measurements for both stars pins down the system velocity  $\gamma$  exactly. Therefore, many of the families of orbits with different combinations of  $P$ ,  $e$ , and  $\gamma$  that would be permitted if RV measurements were only available for the primary can be excluded when the secondary is detected.

Other works have used more conservative limits to determine whether the available velocity data were sufficient to constrain a binary orbit. For example, Halbwachs et al. (2017a) require  $N_{\text{epochs}} \geq 11$ . The true probability of convergence on a local minimum depends on the RV uncertainties, number of epochs, and the

uniformity of observational coverage. This experiment indicates that  $N_{\text{epochs}} \geq 7$  is usually sufficient for RV data similar to what we obtain in this work, but the probability of convergence on an erroneous orbital solution is, of course, lower for systems with a larger number of epochs.

## APPENDIX E: DATA PRODUCTS

Here, we make available the best-fitting labels for all targets identified as multiple-star systems. We also provide a list of targets consistent with being single stars in order to make it possible to reconstruct our initial sample of 20 142 targets.

In Table E1, we list all targets consistent with being single stars; labels for these systems will be released by Ting et al. (in preparation). Labels for targets identified as SB1s are listed in Table E2; these are obtained by simultaneously fitting visit spectra. Labels for targets classified as binaries are listed in Table E3. As described in Section 2.3, these are obtained by simultaneously fitting visit spectra for potential close binaries, which are primarily RV-variable systems, and by fitting the combined spectrum for all other systems. For close binaries in which the orbital configuration changes substantially between visits, we list dynamical mass ratios and centre-of-mass velocities. Labels for targets classified as SB2s in which the gravitational effects with an unseen third component can be detected are listed in Table E4; these are obtained by simultaneously fitting visit spectra. Labels for targets identified as triples in which all three components contribute to the spectrum are listed in Table E5; these are obtained by simultaneously fitting visit spectra.

**Table E1.** List of targets identified as single stars. This table is available in its entirety (with 16 834 rows) in machine-readable form.

APOGEE ID
2M00000233+1452324
2M00001701+7052395
2M00003475+5723259
2M00004578+5654428
...

<sup>14</sup>Although seven parameters are required to parametrize a two-body orbit, the system velocity  $\gamma$  and mass ratio  $q = K_1/K_2$  can be obtained ‘for free’ if the RVs of the primary and secondary fall on a line  $v_{\text{Helio},2} = \alpha v_{\text{Helio},1} + \beta$ , the system velocity is  $\gamma = \beta/(1 - \alpha)$  and the mass ratio is  $q = -1/\alpha$ .

**Table E2.** Best-fitting labels for targets identified as SB1s.  $\Delta v_{\max}$  is the maximum change in RV between visits. This table is available in its entirety (with parameters for 663 systems) in machine-readable form.

APOGEE ID	$T_{\text{eff}}$ (K)	$\log g$ (dex)	[Fe/H] (dex)	[Mg/Fe] (dex)	$v_{\text{macro}}$ (km s $^{-1}$ )	$\Delta v_{\max}$ (km s $^{-1}$ )
2M00010204+0049037	6008	4.26	-0.41	0.00	6.48	2.93
2M00031962-0017109	6157	4.23	-0.23	-0.06	42.95	65.17
2M00041803+1519505	5941	4.92	0.15	0.17	2.21	0.78
2M00041859+7104111	4921	4.36	0.13	-0.01	5.64	20.26
...	...	...	...	...	...	...

**Table E3.** Best-fitting labels for targets identified as binaries in which both components contribute to the spectrum.  $T_{\text{eff}}$  and  $\log g$  refer to the primary. For the 623 targets in which the orbital configuration changes substantially between visits, we provide the dynamical mass ratios,  $q_{\text{dyn}}$ , and centre-of-mass velocity,  $\gamma$ . This table is available in its entirety (with parameters for 2423 systems) in machine-readable form.

APOGEE ID	$T_{\text{eff}}$ (K)	$\log g$ (dex)	[Fe/H] (dex)	[Mg/Fe] (dex)	$q_{\text{spec}}$	$v_{\text{macro}, 1}$ (km s $^{-1}$ )	$v_{\text{macro}, 2}$ (km s $^{-1}$ )	$q_{\text{dyn}}$	$\gamma$ (km s $^{-1}$ )
2M00003968+5722329	4517	4.62	0.05	-0.04	0.88	3.09	8.19	—	—
2M00012717+0128193	5048	4.51	0.07	0.06	0.75	0.00	14.57	—	—
2M00023179+1521164	4589	4.37	-0.20	0.10	0.99	19.29	16.15	0.84	-0.92
2M00024073+6354560	5664	4.26	0.18	-0.01	0.67	1.42	3.34	—	—
...	...	...	...	...	...	...	...	...	...

**Table E4.** Best-fitting labels for targets in which two components contribute to the spectrum but the gravitational effects of a third component can be detected (e.g. Fig. 8).  $T_{\text{eff}}$  and  $\log g$  refer to the primary. This table is available in its entirety (with parameters for 108 systems) in machine-readable form.

APOGEE ID	$T_{\text{eff}}$ (K)	$\log g$ (dex)	[Fe/H] (dex)	[Mg/Fe] (dex)	$q_{\text{spec}}$	$v_{\text{macro}, 1}$ (km s $^{-1}$ )	$v_{\text{macro}, 2}$ (km s $^{-1}$ )
2M00103470+0043200	4200	4.50	-0.21	0.21	1.00	19.98	5.40
2M00265252+6359169	6416	4.56	-0.04	-0.29	0.85	10.34	14.38
2M00310678+8508494	5742	4.44	0.01	-0.07	0.62	4.91	3.11
2M01194897+8532293	5272	4.54	-0.08	0.03	0.90	2.68	4.80
...	...	...	...	...	...	...	...

**Table E5.** Best-fitting labels for SB3s, targets in which three components contribute to the spectrum (e.g. Fig. 6).  $T_{\text{eff}}$  and  $\log g$  refer to the primary. This table is available in its entirety (with parameters for 114 systems) in machine-readable form.

APOGEE ID	$T_{\text{eff}}$ (K)	$\log g$ (dex)	[Fe/H] (dex)	[Mg/Fe] (dex)	$q_2$	$q_3$	$v_{\text{macro}, 1}$ (km s $^{-1}$ )	$v_{\text{macro}, 2}$ (km s $^{-1}$ )	$v_{\text{macro}, 3}$ (km s $^{-1}$ )
2M00182859+6207248	5398	4.33	0.22	-0.20	1.00	1.00	8.69	4.33	0.19
2M00285967+5931138	4655	4.20	-0.22	0.08	0.94	0.94	1.39	30.07	36.15
2M00470197+1751448	6386	4.07	-0.47	0.03	0.89	0.88	19.03	3.50	4.46
2M01103850+6655525	4756	4.60	0.07	-0.00	0.92	0.81	3.71	1.34	16.35
...	...	...	...	...	...	...	...	...	...

This paper has been typeset from a TeX/LaTeX file prepared by the author.