Simultaneous calibration of spectro-photometric distances and the Gaia DR2 parallax zero-point offset with deep learning

Henry W. Leung 1* and Jo Bovy 1,2†

1Department of Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada
2Dunlap Institute for Astronomy and Astrophysics, University of Toronto, 50 St. George Street, Toronto, Ontario M5S 3H4, Canada

Accepted 2019 August 9. Received 2019 August 8; in original form 2019 February 15

ABSTRACT
Gaia measures the five astrometric parameters for stars in the Milky Way, but only four of them (positions and proper motion, but not distance) are well measured beyond a few kpc from the Sun. Modern spectroscopic surveys such as APOGEE cover a large area of the Milky Way disc and we can use the relation between spectra and luminosity to determine distances to stars beyond Gaia’s parallax reach. Here, we design a deep neural network trained on stars in common between Gaia and APOGEE that determines spectro-photometric distances to APOGEE stars, while including a flexible model to calibrate parallax zero-point biases in Gaia DR2. We determine the zero-point offset to be $-52.3 \pm 2.0$ μas when modelling it as a global constant, but also train a multivariate zero-point offset model that depends on $G$, $G_{BP} - G_{RP}$ colour, and $T_{\text{eff}}$ and that can be applied to all $\approx 58$ million stars in Gaia DR2 within APOGEE’s colour–magnitude range and within APOGEE’s sky footprint. Our spectro-photometric distances are more precise than Gaia at distances $\gtrsim 2$ kpc from the Sun. We release a catalogue of spectro-photometric distances for the entire APOGEE DR14 data set which covers Galactocentric radii $2$ kpc $\lesssim R \lesssim 19$ kpc; $\approx 150,000$ stars have $<10$ per cent uncertainty, making this a powerful sample to study the chemo-dynamical structure of the disc. We use this sample to map the mean [Fe/H] and 15 abundance ratios [X/Fe] from the Galactic Centre to the edge of the disc. Among many interesting trends, we find that the bulge and bar region at $R \lesssim 5$ kpc clearly stands out in [Fe/H] and most abundance ratios.

Key words: methods: data analysis – techniques: spectroscopic – astrometry – stars: distances – stars: fundamental parameters – Galaxy: structure.

1 INTRODUCTION

The Milky Way provides a unique opportunity for the study of galaxy formation and evolution, because we can determine the three-dimensional position and velocity (using astrometry and spectroscopy), high-quality stellar parameters and elemental abundances (from high-resolution spectroscopy), and ages for large samples of individual stars (e.g. Freeman & Bland-Hawthorn 2002; Rix & Bovy 2013). Recently, great advances have been made in getting precise stellar parameters (e.g. Holtzman et al. 2015; Casey et al. 2016; Fabbro et al. 2018; Leung & Bovy 2019) and ages (e.g. Martig et al. 2016; Ness et al. 2016; Mackereth et al. 2019) for hundreds of thousands of stars across the Milky Way from high-resolution spectroscopic surveys such as APOGEE (Majewski et al. 2017).

At the same time, astrometric data from the Gaia satellite (Gaia Collaboration 2016) are providing an unprecedented view of the spatial and kinematic structure of the extended solar neighbourhood (e.g. Antoja et al. 2018; Gaia Collaboration 2018a; Bennett & Bovy 2019). However, even with Gaia’s exquisite astrometric precision, it currently only provides precise distances (and, thus, tangential motions) within about 2 kpc, even for relatively bright giants. Thus, to take full advantage of the wide disc coverage of APOGEE, we require a method for obtaining precise distances and space velocities for all stars in APOGEE. The wealth of data from Gaia allows spectro-photometric distance methods to be calibrated using the large, nearby set of Gaia parallaxes and then be applied to the full APOGEE data set. This is what we set out to do in this paper, using the modern machine-learning technique deep learning.

Machine-learning techniques for spectro-photometric distances like deep learning are powerful, because they can be trained on stars with both high-resolution spectra from APOGEE and parallaxes.
from *Gaia* to produce spectro-photometric distances. But applied blindly, such techniques propagate biases that are present in the training set to the model and the subsequently inferred distances. In particular, the *Gaia* DR2 parallaxes are known to suffer from a zero-point offset (Lindegren et al. 2018; Zinn et al. 2018) that is known to be multivariate and may have a very complex dependence on magnitude, colour, sky position, or other quantities. Simply training on the parallax data without any correction or using an inappropriate correction will result in a biased model. This would significantly bias distances obtained for distant stars such as those in the APOGEE sample, because these have small parallaxes where even a small (tens of μas) zero-point offset has a large effect.

Determinations of the *Gaia* DR2 parallax zero-point offset have either been performed using quasars (Lindegren et al. 2018), which should have no measurable parallax on average, or various types of stars. The determination using quasars is precise, but likely not directly applicable to most stars of interest, which are both brighter and redder than quasars and thus probably have a different zero-point offset due to its multivariate dependencies. Determinations using stars depend on trusting semi-empirical stellar-evolution models (Zinn et al. 2018) or use rigid (i.e. few parameter) models for standardizable candles (e.g. Sesar et al. 2017; Riess et al. 2018; Stassun & Torres 2018). These give precise determinations of the zero-point, but have to assume that the zero-point is a well-behaved (e.g. constant) function of observables such as magnitude, parallax, and colour. In this paper, we determine the parallax zero-point offset using a diverse sample of main-sequence and red-giant stars with a method that only rests on the assumption that continuum-normalized stellar spectra allow the intrinsic luminosity to be determined, but not the distance (or apparent magnitude). We represent the zero-point’s multivariate dependence on magnitude, colour, and temperature using a flexible (i.e. many parameter) neural network model using deep learning.

Deep learning is a subset of machine learning and the term refers to the usage of multilayer (‘deep’) artificial neural networks (NNs) to do various kinds of machine learning tasks in supervised and unsupervised learning, image recognition, natural language processing, etc. NNs exist in various architectures that mimic biological brains in order to represent high-dimensional mappings in a versatile manner. On top of the versatility of the NN, we also employ in our work a robust way of Bayesian deep learning that (a) takes data uncertainties in the training data into account and (b) estimates uncertainty in predictions made with the NN using an approximation to variational inference, drop out variational inference. This methodology is mostly based on our previous work on deep learning of stellar abundances with APOGEE spectra using *astroNN*1 (Leung & Bovy 2019). Supervised learning requires trusted, labelled training data for the model to learn. Because of the *Gaia* DR2 zero-point offset, we do not fully trust the training data in the current application and in training an NN to determine spectro-photometric distances from APOGEE spectra, we need to simultaneously learn how to correct the *Gaia* parallaxes for the zero-point offset. We do this by using a form of adversarial training that optimizes two NNs at the same time, one for the spectro-photometric distances and one for the zero-point calibration, in such a way that the residual between the spectro-photometric parallax and the calibrated *Gaia* parallax has no information about the parallax itself.

The outline of this paper is as follows. Section 2 describes our methodology, with Section 2.1 focusing on a general discussion of the method and assumptions and Section 2.2 discussing the specifics of the model and its implementation in more detail. Section 3 provides information on the data selection and processing from APOGEE DR14 and *Gaia* DR2 to construct training and test sets. Sections 4 and 5 discuss our *Gaia* DR2 zero-point offset findings and the precision of the spectro-photometric distances, respectively. To illustrate the power of the derived data set of APOGEE stars with precise distances, Section 6 shows maps of the elemental abundances across a wide area of the Milky Way. Section 7 discusses how our method compares to other methods for inferring the parallax zero-point and for determining spectro-photometric distances, and we look forward to future applications of this methodology. Section 8 gives a brief overview of our conclusions. Appendix A describes how to perform variational NN inference on arbitrary APOGEE spectra to determine their distances using the model used in this work.

Code to reproduce all of the plots in this paper as well as the FITS2 data file containing our neural network’s distance for the entire APOGEE DR14 data set is available at https://github.com/henrysky/astroNN_gaia_dr2_paper. The data model for this FITS file is described in Table 1 at the end of this paper.

### 2 METHODOLOGY

#### 2.1 Basic idea and assumptions

Our goal is to simultaneously calibrate spectro-photometric distances and the *Gaia* DR2 parallax zero-point offset by training a model to predict the *Gaia* DR2 catalogue value of the parallax from the near-infrared APOGEE spectra. To do this, our supervised deep-learning algorithm requires training labels for a set of stars observed by both *Gaia* and APOGEE and training proceeds by minimizing the prediction error for the training data set by adjusting the model parameters. In this application, these model parameters are the strengths of neural-network connections and the optimization is performed using a gradient-descent optimizer.

Before explaining our method in detail, we briefly summarize the main ideas and assumptions that allow us to calibrate both the spectro-photometric distance model and the *Gaia* DR2 zero-point offset at the same time.

**Spectro-photometric luminosity features:** Our spectro-photometric distance model works by mapping continuum-normalized spectra to luminosity (or absolute magnitude) that is subsequently converted to parallax using the observed apparent magnitude and the extinction. To perform this mapping from spectra to luminosity without relying on stellar evolution models, we have to assume that the continuum-normalized spectrum contains features that are indicative of the star’s luminosity. This is a plausible assumption for the red giants observed by APOGEE, as the mass dependence of internal mixing of carbon and nitrogen in red giants allows the stellar mass to be determined from APOGEE spectra (e.g. Masseron & Gilmore 2015) and mass combined with the effective temperature and surface gravity that are eminently measurable from stellar spectra allow the luminosity to be determined. It is less

1https://github.com/henrysky/astroNN

2https://github.com/henrysky/astroNN_gaia_dr2_paper/raw/master/apogee_dr24_dist.fits
clear that this is plausible for the sub-giant, turn-off, and main-sequence stars in our sample, but ultimately the success (or lack thereof) of the method validates this assumption and we will see that we are able to determine luminosities for these types of stars as well.

Continuous spectral flux–luminosity relation: We assume that the value of the luminosity is a continuous, smooth function of the continuum-normalized spectral flux values. This is a general limitation of the type of NN that we use. Thus, similar spectra are assumed to have similar luminosities. This is a generalization of the concept that ‘spectral twins’ have the same intrinsic luminosity, which has been successfully used to obtain high-precision distances to stars (Jofrè et al. 2015, 2017). However, we do not require that spectra with similar luminosities have similar spectra.

Unpredictability of the distance from continuum-normalized spectra: This is the basic assumption that allows the calibration of the Gaia DR2 zero-point offset. We assume that a star’s distance cannot be learned from the continuum-normalized spectra alone. If this assumption did not hold, then the simultaneous determination of the luminosity and distance from the continuum-normalized spectrum would allow the parallax and apparent magnitude to be determined directly and the spectro-photometric model on its own could match the Gaia catalogue parallax values at all magnitudes and colours, thus absorbing the erroneous effect of the zero-point offset. That the continuum-normalized spectrum contains no direct distance information is plausible, because in the absence of interstellar absorption or emission features, the continuum-normalized spectrum is an intrinsic, distance-independent property of the star. APOGEE spectra do contain interstellar absorption features (e.g. a strong diffuse interstellar band; Zasowski et al. 2015), which are correlated with extinction and thus indirectly with distance and this therefore weakly breaks our assumption. However, as long as the correlation with distance is not perfect, the effect of this on our method is small. We similarly assume that the continuum-normalized spectra have no features that are directly related with the zero-point offset. That a spectrum would have features directly related to the zero-point offset is highly unlikely, given that the zero-point offset is a Gaia-specific instrumental effect and the spectra are taken using an entirely different instrument. The unpredictability of the distance from continuum-normalized spectra leads therefore to the unpredictability of the Gaia zero-point offset from spectra.

Gaussian observational errors of Gaia parallaxes: We assume that the uncertainty value provided by the Gaia data reduction pipeline is the standard deviation of a Gaussian distribution of parallax error. Below, we preserve the Gaussian nature of the uncertainty by not performing any non-linear (e.g. inverse or logarithmic) operation on the parallax, but instead adopting the luminosity–parallax relation from Anderson et al. (2018)

\[
L_{\text{pseudo}} = \sigma \times 10^{\frac{m_{\text{apparent}} - M_\odot}{5M_{\odot}}} + 3.
\]

where \(L_{\text{pseudo}}\) is an alternative scaling of luminosity, a pseudo-luminosity. If the observational errors in the parallax are Gaussian and the observational errors in apparent magnitude are negligible, the propagated observational errors in \(L_{\text{pseudo}}\) are Gaussian as well.

Known extinction: The conversion between intrinsic luminosity and parallax requires the extinction-free apparent magnitude. In this work, we assume that the extinction is known and that there is no uncertainty in the extinction correction. This is a good assumption for APOGEE stars, for which the \(K_i\) extinctions are determined

---

Table 1. Data model of apogee_dr14_nn_dist.fits which contains 277 371 APOGEE DR14 stars. Most labels contain −9999 for unknown or bad data.

| Label                      | Units    | Sources            | Descriptions                                          |
|----------------------------|----------|--------------------|-------------------------------------------------------|
| apogee_id                  | n/a      | APOGEE DR14        | APOGEE ID                                             |
| location_id                | n/a      | APOGEE DR14        | Location ID of APOGEE                                  |
| ra_apogee                  | deg      | APOGEE DR14        | Right ascension (J2000)                                |
| dec_apogee                 | deg      | APOGEE DR14        | Declination (J2000)                                   |
| ra                         | deg      | Gaia DR2           | Right ascension (ICRS)                                 |
| dec                        | deg      | Gaia DR2           | Declination (ICRS)                                    |
| pmra                       | mas yr\(^{-1}\) | Gaia DR2   | Proper motion in RA                                    |
| pmdec                      | mas yr\(^{-1}\) | Gaia DR2   | Proper motion in Dec.                                  |
| pmra_error                 | mas yr\(^{-1}\) | Gaia DR2   | Proper motion uncertainty in RA                       |
| pmdec_error                | mas yr\(^{-1}\) | Gaia DR2   | Proper motion uncertainty in Dec.                     |
| phot_g_mean_mag            | mag      | Gaia DR2           | G-band mean magnitude                                 |
| bp_rp                      | mag      | Gaia DR2           | \(G_{\text{BP}} - G_{\text{RP}}\) colour             |
| fakemag                    | mag      | This work          | NN pseudo-luminosity \(L_{\text{pseudo}}\) (see equation 1) |
| fakemag_error              | mag      | This work          | NN pseudo-luminosity \(L_{\text{pseudo}}\) uncertainty (see equation 1) |
| nn_parallax\(^a\)          | mas      | This work          | NN parallax (see equation 1)                          |
| nn_parallax_error          | mas      | This work          | NN parallax model uncertainty (see equation 1)        |
| dist\(^a\)                 | parsec   | This work          | NN inverse parallax                                   |
| dist_model_error           | parsec   | This work          | NN inverse parallax model uncertainty                 |
| dist_error                 | parsec   | This work          | NN inverse parallax total uncertainty                 |
| weighted_dist\(^b\)        | parsec   | This work and Gaia DR2 | Inverse variance weighted combined distance from NN and Gaia |
| weighted_dist_error        | parsec   | This work and Gaia DR2 | Uncertainty in the weighted distance (NN model uncertainty is adapted) |

\(^a\)Missing values have −9999, due to missing spectroscopic data or photometry.

\(^b\)Calculated based on the NN distance and \(m_{\text{Gaia}} + 52\mu\text{as}\). When one of the distance sources is −9999, the other one is used directly. When both are −9999, the resulting weighted \(\text{dist}\) is −9999.
from near-infrared and mid-infrared photometry (see below). Note
that this is not a crucial assumption of the method, because we could
simultaneously infer extinction from the spectra and associated
multiband photometric data (e.g. Hogg, Eilers & Rix 2018), but
because accurate extinctions are available for APOGEE stars (see
below), it is a convenient additional assumption for the present
application.

We exploit the unpredictability of the Gaia zero-point offset
from continuum-normalized spectra to train a model that we name
the offset calibration model. This model calibrates the zero-point
offset during training while another part of the model, the
spectro-luminosity model learns how to map continuum-normalized
spectra to luminosity. Since the goal of the training is to minimize the
prediction error for the training data – in our case obtaining a
predicted Gaia parallax as close to the actual Gaia parallax as
possible given the error in the observed parallax – any offset
that is unpredictable to the spectro-luminosity model will result a
higher prediction error. The offset calibration model then effectively
provides one or more degrees of freedom in parallax space to
improve the parallax prediction. Given our assumptions, the
spectro-luminosity model returns the true parallax and the offset calibration
model therefore provides the zero-point offset.

2.2 Modelling specifics

The NN that we train and test in this work is composed of two basic
components: the offset calibration model and the spectro-luminosity
model. Both of these are NNs and they are trained simultaneously
using a single objective function. Once these two models are trained
they are used separately to either determine the luminosity (and
uncertainty, but this is easily converted to
L
)

as ApogeeDR14GaiaDR2BCNN() in astroNN and is shown in
Fig. 1. This combined network takes the continuum-normalized
APOGEE spectrum for a star
i

and maps it to the
L

pseudo

which is converted to the true parallax
ϖ

i

using the observed apparent magnitude and the known extinction. The offset
calibration model provides the zero-point offset
ϖ
offset,i

i

(2)

The objective function that is minimized during training is then the
sum of the following objective functions for each individual star
i:

\[
J(σ_i^G, ϖ_i^G) = \frac{1}{2} (ϖ_i^G - \hat{ϖ}_i^G)^2 e^{-t_i} + \frac{1}{2} s_i, 
\]

where
σ_i^G

is the Gaia-reported parallax. This objective function is minus the log likelihood if the uncertainty in the difference
is Gaussian, which we assume here. The quantity
s_i

represents
the natural logarithm of the uncertainty variance in the difference
between the predicted and the Gaia parallax. It has two contributions
that are summed in quadrature

\[
s_i = \ln[σ_{\text{err},i}^2 + σ_{\text{pred},i}^2].
\]

These two contributions are the parallax uncertainty
σ_{\text{err},i}^2,

reported by Gaia and the predictive uncertainty
σ_{\text{pred},i}^2,

in the true parallax returned by the spectro-luminosity model. This predictive
uncertainty is an additional output from the spectro-luminosity model and
the optimal mapping of input continuum-normalized spectrum to uncertainty is also simultaneously optimized during training.

The spectro-luminosity model returns
L_{\text{pseudo}}

and its (Gaussian)
uncertainty, but this is easily converted to
σ_{\text{pred},i}^2

using equation (1). The purpose of including the predictive variance is to capture the uncertainty associated with the inability of the model to perfectly map spectra on to absolute magnitudes; it also captures any possible underestimation of the catalogue Gaia parallax uncertainty, for which evidence was found by Arenou et al. (2018). We do not include a predictive uncertainty in the offset calibration model, because for predictions we are primarily interested in spectro-
photometric distances. The final loss for the stochastic gradient
descent optimizer in this work (ADAM optimizer; Kingma & Ba
2014) is calculated from a mini-batch partition of the data of size
N

\[
J = \frac{1}{N} \sum_{i=1}^{N} J(σ_i^G, ϖ_i^G).
\]

We use dropout with a dropout fraction of 30 per cent in all layers
during training to prevent overfitting (Hinton et al. 2012).

During inference, we use dropout for uncertainty estimation
which is known as dropout variational inference (Kendall & Gal
2017). To obtain
L_{\text{pseudo}}

predictions for input spectra, we run every spectrum through
N

forward passes of the spectro-luminosity model with dropout turned on. Since dropout drops weights randomly, the
spectro-luminosity model becomes probabilistic and has different predictions in every forward pass through the network. The mean
value of these
N
predictions is the final prediction and the standard

\footnote{In this section, we denote predicted values of quantities using the hat operator (e.g. \( \hat{\sigma} \) for the predicted parallax) to distinguish them from the provided training data. In later sections we drop this notation and simply denote predicted parallaxes and parallax offsets without a hat.}

\[ \text{MNRAS 489, 2079–2096 (2019)} \]
deviation of the predictions is the model uncertainty. In addition to this model uncertainty from dropout variation inference, the spectro-luminosity model also gives the predictive uncertainty for each star discussed above. The total prediction uncertainty is the sum of model and predictive uncertainty in quadrature (Kendall & Gal 2017).

The reason that we use the Softplus activation in the last layer of the spectro-luminosity model is that the pseudo-luminosity \( L_{pseudo} \) from equation (1) cannot be negative, because a negative pseudo-luminosity would translate to a negative true parallax. The Softplus activation – defined as \( y = \log(1 + e^x) \), where \( x \) is the input and \( y \) is the output – is a smooth approximation of the standard rectified linear unit (ReLU) activation function that we use in all but the output layer. Although both Softplus and ReLU map all real inputs to non-negative real numbers, using the Softplus activation as the last layer for predicting stellar luminosity is better, because it never produces zero luminosity and thus never leads to zero parallax. That is, unlike ReLU, Softplus maps all real numbers to non-zero positive real numbers, while ReLU maps all negative numbers to zero.

While the true parallax returned by our model is never negative or zero, our model’s prediction for the Gaia parallax can be negative due to the zero-point correction. This in addition to taking the uncertainty in the parallax and its prediction into account in the objective function of equation (3) allows us to use negative Gaia parallaxes in our training set. That is, we do not need to artificially remove negative parallaxes in the Gaia catalogue, which may result from random noise or from zero-point biases. Because our objective function takes the Gaia parallax uncertainty into account, we are also not limited to only using high-precision Gaia parallaxes in the training and we do not do any cut on parallax SNR in the training step.

3 DATA

3.1 Spectroscopic data from APOGEE

The spectroscopic data in this work come from Data Release 14 (DR14; Abolfathi et al. 2018; Holtzman et al. 2018; Jónsson et al. 2018) of the APO Galactic Evolution Experiment (APOGEE; Majewski et al. 2017). APOGEE spectra are obtained with a 300-fiber spectrograph (Wilson et al. 2019) attached to the Sloan Foundation 2.5 m telescope at Apache Point Observatory (Gunn et al. 2006). APOGEE is an infrared (1.5–1.7 \( \mu \)m), high resolution \((R \sim 22 500)\), high signal-to-noise ratio (SNR) (typical SNR > 100) spectroscopic survey mainly targeting red-giants (Zasowski et al. 2013, 2017). As in our previous work (Leung & Bovy 2019), we perform our own continuum-normalization that uses a pre-defined set of continuum pixels (Bovy 2016; Casey et al. 2016) starting from the combined, rest-frame spectra in the APOGEE ‘apStar’ files. After continuum normalization, we set the flux value of pixels that contain the following bits in the APOGEE pixel-level mask bitset to 0 (the expected continuum), because they are likely bad: 0: bad pixel, 1: cosmic ray, 2: saturated, 3: unfixable, 4: bad from dark, 5: bad from flat, 6: high error, 7: no sky info, 12: overlaps a significant sky line. Spectroscopic parameters such as \( T_{eff} \) and elemental abundances that we use come from our own re-analysis of the DR14 data in Leung & Bovy (2019).

Apparent \( K_s \) magnitudes for all stars in the APOGEE catalogue are taken from the 2MASS catalogue (Skrutskie et al. 2006) with uncertainty in \( K_s < 0.1 \) (Zasowski et al. 2017). We correct these for extinction using the AK_TARG extinction listed in the APOGEE catalogue, which is the extinction adopted by APOGEE for targeting. The value of AK_TARG is derived on an individual-
star basis by the Rayleigh Jeans Colour Excess method (RJCE; Majewski, Zasowski & Nidever 2011). The RJCE method calculates extinctions using a combination of near- and mid-infrared photometry as
\[
A_K = 0.918(H - [4.5\mu]) - (H - [4.5\mu])_0),
\]
where \(H - [4.5\mu]\) is the measured colour and the method assumes that \((H - [4.5\mu])_0 = 0.08\) for a wide range of spectral types. \(H\)-band photometry in this equation comes from 2MASS, while the 4.5 \(\mu\) photometric data are either from Spitzer–IRAC data (Churchwell et al. 2009) or from the WISE survey (Wright et al. 2010). The typical uncertainty in \(A_K\) is \(0.05\) mag (Bovy et al. 2014) and we set all \(\text{AK}_{\text{TARG}} < -1\) equal to 0, that is, zero extinction.

3.2 Parallax data from Gaia

We use data from the second data release (DR2; Gaia Collaboration 2018b) from the European Space Agency’s Gaia mission Gaia (Gaia Collaboration 2016) to train the model and calibrate the Gaia DR2 zero-point offset. Gaia DR2 contains 1.3 billion source with 5 astrometric parameters (positions, proper motions, parallaxes). We match Gaia sources to APOGEE stars using a celestial position cross-match with a tolerance of 2 arcsec. Out of 277 371 stars in APOGEE DR14, 265 761 have both a Gaia DR2 parallax and a 2MASS \(K_s\) band apparent magnitude; the median parallax SNR of these stars is 15.4. We use parallax and Gaia pipeline-reported parallax uncertainty as well as the \(G\)-band magnitude and \(G_{BP} - G_{RP}\) colour.

3.3 Training, validation, and test data sets from APOGEE and Gaia

We create one training set and one test set from the APOGEE DR14 spectra. Both data sets consist of continuum-normalized APOGEE spectra and Gaia DR2 parallaxes. The training set contains 35 112 stars, while the test set has 33 468 stars. The main difference between the training and test set is in the SNR. The training is constructed using only high-SNR spectra with SNR > 200, whereas the test set consists only of low-SNR spectra with SNR < 200. 90 per cent of the training set is randomly selected to train the NN – that is, these stars are used to compute the gradients of the objective function in the training steps – and the remaining 10 per cent constitutes a separate validation set that is used to validate the performance of the NN during the training process.

On top of the spectral SNR cuts, we perform cuts on the quality of the Gaia parallaxes and on the quality of the APOGEE spectra. This is necessary, because all of the knowledge learned by the NN is solely driven by the training data. Therefore, we need to make sure that the training inputs and labels are as accurate as possible, because any systematic inaccuracy such as bias will be captured by the NN and propagated to new data. For this reason, we exclude spectra flagged with the \(\text{STARFLAG}\) flags and spectra with a radial velocity scatter larger than 1 \(\text{km s}^{-1}\), because these represent potential issues with the spectra, and potential binary stars, respectively. We require that the Gaia parallaxes have \(\sigma_\mu < 0.1 \text{mas}\) and \(\text{visibility}\_\text{periods\_used} > 11\) to ensure stars are astrometrically well-observed by Gaia with at least 11 gaps of at least 4 d with small uncertainty (Lindegren et al. 2018). Furthermore, we exclude stars with \(\log(\text{L}/\text{L}_{\odot} | L_{\text{Gaia}}/L_{\odot}) \leq 0\) where \(L\) refers to the use of \(L_{\text{Gaia}}\) in equation (1) and we only apply this cut to stars with positive \(\sigma_\mu\), because we are mainly interested in brighter stars that can be seen to large distances to map the Milky Way. These cuts ensure the quality of the training and test sets. Fig. 2 displays the distribution of the training and test sets in a few key quantities.

Despite all of our quality cuts, our training and test sets include negative parallaxes and parallaxes with large percentage uncertainties (low-parallax SNR). As described in Section 2.2, our model can handle negative parallaxes in a physically sensible way during the training process and as discussed in Section 2.1, our training process does not involve inverse parallax, which is a strongly biased estimate of the distance for low-parallax SNR (Bailer-Jones 2015). Furthermore, our robust objective function used during training takes parallax uncertainty into account. Therefore, we do not remove parallaxes with low SNR from the training or test set and are therefore not affected by any biases that would result from making such a cut.

4 RESULTS: THE GAIA DR2 ZERO-POINT OFFSET

We train models to calibrate the Gaia DR2 zero-point offset in three different variations while training the spectro-photometric NN to infer luminosity from spectra. We train a first model on unmodified Gaia DR2 parallaxes without any offset calibration model to investigate the spectro-photometric distances that we obtain without calibrating or correcting the zero-point offset. We train a second model on unmodified parallaxes with a constant zero-point offset calibration (see Section 2.2 for implementation details). The third model that we fit is trained on unmodified parallax with a zero-point offset calibration model that depends on \(G, G_{BP} - G_{RP}\), and \(T_{\text{eff}}\).

Comparing the first model that does not calibrate or correct the zero-point offset with the constant-offset second model demonstrates clearly that there is indeed a zero-point offset in Gaia DR2. This is evident from Fig. 3, where we compare the parallax obtained from the spectro-photometric NN with no zero-point offset to the parallax obtained when fitting a constant zero-point offset. Because there is a zero-point offset, the spectro-photometric NN parallax trained without accounting for the offset is unable to match the Gaia parallax over the entire range of parallaxes. Because the training sample is dominated by distant giants, the NN optimizer matches the NN parallaxes to the Gaia parallaxes at small parallaxes, but fails to do so for similar stars at larger parallaxes, leading to an increasing offset between NN and Gaia parallaxes at larger parallaxes. The model that includes a fitted constant zero-point offset is much better at matching the Gaia parallax over a wide range of parallaxes; the fact that it does not do so perfectly is because the zero-point offset is not constant, as we discuss further below.

When fitting for a constant zero-point offset, we get a zero-point offset of \(-52.3 \pm 2.0\) \(\mu\)as; these values are obtained by inputting featureless vectors of ones in the offset calibration model and sampling the posterior of the offset calibration model by dropout to get the result and uncertainty (see Section 2.2). This value is similar to that found by Zinn et al. (2018) and Arenou et al. (2018) also using red giant stars. Our result deviates from the \(\approx -30\) \(\mu\)as determined with quasars (Lindegren et al. 2018); this indicates that the zero-point may be different for different types of stars and for objects with different spectral energy distributions such as quasars (as indicated by their colour or effective temperature).

To investigate whether the zero-point offset depends on other properties, we fit the multivariate offset model that depends on \(G\),
Spectro-photometric distances and Gaia DR2 zero-point

Figure 2. Number of stars in the training and test sets as a function of four different properties and the sky distribution of the training and test sets. Upper left: APOGEE spectral SNR with the median SNR $\bar{\text{SNR}}$ shown in the legend. Upper right: observed Gaia parallax, without any zero-point offset correction. Middle left: Gaia parallax SNR as $|\varpi|/\sigma_{\varpi}$ (we do not cut on parallax SNR to construct our training and test sets). Middle right: logarithmic luminosity in solar units, derived from unmodified Gaia parallax and extinction-corrected $K_s$ apparent magnitude; we exclude stars below solar luminosity. Bottom left/right: sky distribution of the training (left) and test set (right); both follow the full APOGEE DR14 footprint.

This model therefore determines an individual zero-point offset for each star and we determine the zero-point and its uncertainty for each star by sampling the posterior of the offset calibration model using dropout variational inference. Taking the mean of the zero-point offsets in our test sample, we obtain $\approx -50\text{\,\muas}$ in good agreement with our fit of a constant zero-point offset above. The full dependence on different properties in the test sample of the zero-point offset is displayed in Fig. 4. The left column shows the zero-point as a function of one of the properties ($G$, $G_{BP} - G_{RP}$, and $T_{eff}$) at a time and the right column displays the zero-point offset as a function of $G$ and $G_{BP} - G_{RP}$ and of $G$ and $T_{eff}$. The points in the left column are colour coded by the uncertainty in the zero-point offset and it is interesting to notice that a horizontal band of low uncertainty is located at $\approx 50\text{\,\muas}$, which is similar to the result we have estimated with the constant offset model. The majority of the stars in the APOGEE red-giant sample have magnitudes, colours, and temperatures such that they fall in this $\approx 50\text{\,\muas}$ low-uncertainty regime.

From the trends in the left column of Fig. 4, the offset seems to be increasing in magnitude with $G_{BP} - G_{RP}$ and $G$, i.e. apparently redder and fainter stars seem to have more negative offset. The offset is also smaller with increasing surface temperature, consistent with the colour trend. This behaviour might explain why Zinn et al. (2018) report that the offset seems to have a dependence on parallax that is such that the zero-point offset is larger when the parallax is smaller: in a sample of giants like the APOKASC sample used by Zinn et al. (2018), stars farther away are generally fainter, and thus have a larger offset due to the strong dependency of the zero-point offset on $G$ that we find. For RC stars, Zinn et al. (2018) report $-50.2 \pm 2.5\text{(stat.)} \pm 1\text{(syst.)}\text{\muas}$ while our inverse-variance weighted mean of offset for the same RC stars is $\approx -48\text{\,\muas}$ with negligible uncertainty. Zinn et al. (2018) attribute...
the dispersion around the mean quadratic fits ranges from 4.7 to 5.5 μas.

These are obtained using quadratic regression and we find (we do not
polynomial models of the univariate trends in the left column panels.

offset found using quasars is difficult.

offset ϖ of quasars is far from the range covered in this work with giants
in this volume and the offset model overpredicts their luminosity, because the training set is dominated by giants. Overall, the model with constant zero-point offset is better able to match the Gaia parallaxes over a wide range of parallax values than the model without a zero-point correction, demonstrating that a zero-point offset is present in the Gaia DR2 parallaxes.

the different offset they find for RGB and RC stars to a systematic
in the asteroseismic radius scale, but given that we confirm this
difference without relying on the radius scale, it appears that this
is instead a true difference in parallax zero-point offset for RC and
RGB stars.

Similarly, the trends that we find might also explain the much
smaller offset found by Lindegren et al. (2018) using quasars,
because quasars are much bluer and effectively hotter and the colour
and $T_{\text{eff}}$ trends that we find are such that these properties should lead
to a smaller offset. However, quasars are also generally fainter than
the stars in our sample and the $G$ trend that we find would point
towards a larger offset. The colour–magnitude–temperature range
of quasars is far from the range covered in this work with giants
and so a direct comparison between our zero-point offsets and those
found using quasars is difficult.

We summarize the results from Fig. 4 by providing simply
polynomial models of the univariate trends in the left column panels.
These are obtained using quadratic regression and we find (we do not
provide the statistical uncertainty in these fits as it is negligible,
the dispersion around the mean quadratic fits ranges from 4.7 to
5.4 μas):

\[
\sigma_{\text{offset}}/\mu\text{as} = 1.18 (G - 16)^2 - 0.92 (G - 16) \\
- 55.49 \quad (11 \lesssim G \lesssim 17). \tag{7}
\]

\[
\sigma_{\text{offset}}/\mu\text{as} = 6.90 (G_{\text{BP}} - G_{\text{RP}})^2 - 34.63 (G_{\text{BP}} - G_{\text{RP}}) \\
- 11.67 \quad (0.8 \lesssim G_{\text{BP}} - G_{\text{RP}} \lesssim 3), \tag{8}
\]

\[
\sigma_{\text{offset}}/\mu\text{as} = 6 \times 10^{-6} (T_{\text{eff}}/K - 4500)^2 + 0.0135 (T_{\text{eff}}/K - 4500) \\
- 53.17 \quad (4000 K \lesssim T_{\text{eff}} \lesssim 5750 K). \tag{9}
\]

Our zero-point estimates can be applied to all 58,157,607 stars
in Gaia DR2 within the colour–magnitude range and sky footprint
covered by APOGEE [cuts from equation 9 with $\ell < 250$ deg and
$(-15 < b < 15)$ deg as shown in Fig. 2]. Due to the significant
variation of the zero-point offset with position on the sky (Arenou et al. 2018; Lindegren et al. 2018) on small and large scales, one
should be cautious about applying our zero-point offset outside of
the area of the sky covered by the APOGEE survey or in small
patches of the sky within the APOGEE footprint.

5 RESULTS: PRECISION AND ACCURACY OF
SPECTRO-PHOTOMETRIC DISTANCES

5.1 Comparison to Gaia

After we train the spectro-photometric NN on high-SNR APOGEE
spectra, we test the model on the low-SNR spectra test set. To
make our results easier to compare to other approaches, we use
the constant zero-point offset model for this comparison – the
discussion in Section 4 demonstrates that for the bulk of the
APOGEE stars the constant zero-point offset model works about
as well as the more complicated ($G, G_{\text{BP}} - G_{\text{RP}}, T_{\text{eff}}$)-dependent
offset model. Fig. 5 shows a comparison between the distances
obtained using the constant and the multivariate offset models for
stars in the test set. Both models agree well within 10 kpc from
the Sun, but beyond that even a slight difference in the zero-point
of a few μas in parallax results in a noticable offset and a greater
dispersion. Our multivariate zero-point model suggests that fainter
stars have bigger offsets (in absolute value) and, therefore, stars
beyond 10 kpc are generally closer when using the multivariate
offset model than when using the constant zero-point model. The
bulk of the APOGEE stars are within 10 kpc and we therefore use
the constant zero-point model when evaluating how well our NN
distances perform.

We test the parallaxes returned by the spectro-photometric NN
in combination with the extinction-corrected $K_{s}$ magnitudes by

Figure 3. Comparison of the neural network spectro-photometric parallax results obtained by training with unmodified Gaia parallaxes and to those obtained using the constant offset calibration model for the test set. The x-axis is the Gaia DR2 catalogue value for the parallax and the y-axis is the difference between the spectro-photometric parallax and the Gaia DR2 parallax, either the catalogue value for the model without zero-point correction or the zero-point-corrected value $\sigma_{\text{Gaia}} + 52 \mu\text{as}$ for the model trained with a constant offset calibration model. The curves show the median of the parallax difference values in bins in $\sigma_{\text{Gaia}}$. The purple line starting at $\sigma_{\text{Gaia}} = 0$ is the line $\sigma_{\text{NN}} = 0$, which is the smallest the NN parallax can be and therefore no points can lie below this line. That the curves follow this line is because the NN returns a parallax close to zero for most stars for which Gaia determined a negative parallax. The region within a 1 kpc from the Sun is not shown as the APOGEE stars are dominated by dwarfs in this volume and the model overpredicts their luminosity, because the training set is dominated by giants. Overall, the model with constant zero-point offset works better able to match the Gaia parallaxes over a wide range of parallax values than the model without a zero-point correction, demonstrating that a zero-point offset is present in the Gaia DR2 parallaxes.
Figure 4. The offset determined using the offset calibration model shown in Fig. 1, in which the zero-point offset depends on $G$, $G_{BP} - G_{RP}$, and $T_{eff}$. The left-hand panels show the multivariate offset being projected as a function of one of $G$, $G_{BP} - G_{RP}$, and $T_{eff}$; the points are colour coded by the zero-point uncertainty which is estimated by dropout variational inference. There is a clear horizontal band of low zero-point uncertainty located at $\approx 50$ μas which is similar to the zero-point offset that we determine with the constant offset model. The right-hand panel displays two-dimensional projections of the three-dimensional offset model. The rough location of quasars (QSO) used by Lindegren et al. (2018), Cepheids used by Riess et al. (2018), and eclipsing binaries used by Graczyk et al. (2019) are represented by ellipses with sizes that correspond to the typical range of the data used in these works, which all find an offset of $\approx -30$ μas. The Gaia DR2 zero-point offset becomes bigger in magnitude for fainter, redder, and cooler stars.

comparing them to the Gaia parallaxes, correcting the latter as $\sigma_{Gaia} + 52$ μas to account for the zero-point offset. For APOGEE DR14 as a whole, Fig. 6 compares the SNR of the distance determined using the NN to that determined by Gaia as a function of Galactic $x$ and $y$ coordinates. The blue scatter points, where Gaia performs better than the NN, are concentrated at $\lesssim 2$ kpc from the Sun. Beyond 2 kpc, the NN distances are better than the Gaia distances. The bulk of the APOGEE giants are at distances greater than 2 kpc, so the NN distances outperform Gaia for most APOGEE giants.

A summary of the comparison to Gaia parallaxes with uncertainty $<20$ per cent in the test set is displayed in Fig. 7. The top panels show the percentage uncertainty returned by the NN – this is the combination of the model and predictive uncertainty – on the left and the absolute percentage error between $\sigma_{NN}$ and $\sigma_{Gaia} + 52$ μas on the right. The absolute percentage error is generally larger for predictions with a more uncertain prediction, showing that the uncertainty in the NN parallax determined by the model is reasonable. The bottom panels show the median of the absolute percentage deviation (MAD per cent) as a function of $T_{eff}$, log $g$, spectral SNR, and $\sigma_{Gaia}$ SNR. The prediction is generally accurate to $<8$ per cent all the way from dwarfs with solar luminosity to giants brighter than the red clump. For the most luminous giants with $T_{eff} < 4250$ K or log $g < 1.6$, the parallax precision is worse at around 20 per cent or higher. The error bars indicate the typical uncertainty in the NN parallax prediction and the fact that the error bars typically reach zero and not much further demonstrates that the uncertainty estimates returned by the NN are a reasonable description of the precision in the NN parallax.

The two remaining lower panels of Fig. 7 show the $\sigma_{MAD}$ per cent absolute deviation as a function of the SNR in the APOGEE spectra or in the Gaia parallaxes used in the comparison. The NN parallax precision is constant for all but the lowest Gaia parallax SNR values, indicating that the magnitude of the $\sigma_{NN} - \sigma_{Gaia}$ deviation is driven by noise in the NN parallax rather than in the Gaia comparison parallax for these high-SNR Gaia parallaxes. For spectra with spectral SNR $> 100$ the NN parallax precision is roughly constant at $\approx 6$ per cent, which shows that the parallax precision is not limited
surveys in the $H$ spectral SNR $100$, the precision in the NN parallax degrades and especially so for scatter in the spectra–luminosity relation. Below spectral SNR of by noise in the spectra at high-spectral SNR, but rather by the scatter in the spectra–luminosity relation. Below spectral SNR of 100, the precision in the NN parallaxes degrades and especially so for spectral SNR $< 50$. Thus, if future high-resolution spectroscopic surveys in the $H$ band like SDSS-V (Kollmeier et al. 2017) want high-precision spectro-photometric distances for luminous giants, these surveys should aim for SNR at least 50 and ideally 100. That the intrinsic scatter in the spectra–luminosity relation is only $\approx 6$ per cent in distance also implies that any possible Malmquist bias in the luminosity calibration is $< 1$ per cent in implied distance.

To further validate the accuracy of the NN distances, Fig. 8 compares $\sigma_{\text{NN}}$ and $\sigma_{\text{Gaia}} + 52 \mu$as. The $\sigma_{\text{MAD}}$, is used a robust measurement of the scatter which is based on the median absolute deviation (MAD): $\sigma_{\text{MAD}} = 1.4826 \text{MAD}$, where the factor is such that for a Gaussian distribution $\sigma_{\text{MAD}}$ equals the Gaussian standard deviation. Thus, for a set of percentage residual $R$ per cent : $[R_1 \text{ per cent}, R_2 \text{ per cent}, \ldots, R_n \text{ per cent}]$ which is 100 per cent times residual $R$ over ground truth, $\sigma_{\text{MAD}}$ per cent is

$$\sigma_{\text{MAD}} \text{ per cent} = 1.4826 \text{median}(R / \text{per cent}) - \text{median}(R / \text{per cent}).$$

To focus the comparison on high-quality Gaia inverse parallaxes we only display stars with $\sigma_{\text{Gaia}} + 52 \mu$as uncertainty that is lower than 5 per cent in the test set, the $\sigma_{\text{MAD}}$ is $\approx 8$ per cent when considering all giants (top panel) and only slightly larger at $\approx 9$ per cent for the most luminous giants (lower panel). The median difference between the NN and the Gaia distance is only $\approx 2.5$ per cent for all giants and about twice as large for the most luminous giants. Thus, our NN distances are highly accurate.

### 5.2 Comparison to other spectro-photometric distances

To further test the performance of the NN parallaxes, we compare our results to previous spectro-photometric distance estimates for subsets of the APOGEE data: those for RC stars from the APOGEE-RC catalogue (Bovy et al. 2014), distances from the APOGEE-DR14 Brazilian Participation Group (BPG) Distance Estimation Catalogue (Santiago et al. 2016), and the distances for luminous red giants from Hogg et al. (2018).

The DR14 version of the APOGEE-RC catalogue (Bovy et al. 2014) contains 29 502 stars with approximately 95 per cent purity and with distances precise to 5–10 per cent. RC stars are stars in the core Helium-burning stage in the stellar evolution of low-mass stars and because they have a narrow luminosity distribution they act close to a standard candle. The distances in the RC catalogue are determined using predictions of their absolute magnitude from stellar evolution models. The overall distance scale in the RC catalogue was calibrated against parallaxes ($\leq 100$ pc) in Hipparcos and the Hipparcos zero-point offset is 0.1 mas at worst (Arenou et al. 1995). Therefore the RC distance scale can not have the same zero-point offset issue as the Gaia DR2 data and is accurate to 1 per cent. Thus, the RC catalogue can provide precise and accurate distances to test both the NN precision as well as the zero-point offset correction. We remove 1408 RC stars from consideration because they are likely contaminants: these stars have a difference between the log $g$ value determined by the APOGEE pipeline and the data-driven NN value of $0.2$ dex. Because the APOGEE RC catalogue selection is based on the pipeline value of log $g$ and we believe the NN log $g$ value to be more accurate, these 1408 stars are likely mislabelled. The distances in the APOGEE-RC catalogue are determined using the same extinction values that we use, so any systematics due to incorrectly determined extinction are the same (and cancel out in the comparison).

The catalogue of APOGEE BPG distances that is included in APOGEE DR14 contains 211 243 stars. Distances to these stars are determined using a Bayesian method applied to spectroscopic parameters and photometry that makes use of stellar evolutionary models and a Galactic density prior based on mass, age, and metallicity. Distances have uncertainties of $\approx 10$ per cent for dwarf stars and $\approx 20$ per cent for giants.
Figure 7. Precision and accuracy of the spectro-photometric NN parallaxes in the test set. The top two panels show the percentage uncertainty of the NN parallaxes $\varpi_{\text{NN}}$ (top left) and the absolute percentage error between $\varpi_{\text{NN}}$ and $\varpi_{\text{Gaia}} + 52$ μas (top right) as a function of temperature and surface gravity determined by the NN of Leung & Bovy (2019). The percentage uncertainty returned by the NN model is correlated with the actual error, i.e. the absolute percentage error is generally larger for uncertain predictions. The four bottom panels display the median absolute percentage deviation $\sigma_{\text{MAD}}$ (see equation 10) as a function of $T_{\text{eff,logg}}$, spectral SNR, and SNR of the Gaia parallax measurement; the error bars are the median uncertainty for the NN parallax predictions in each bin (the green error bar is the model uncertainty component, while the orange error bar is the total uncertainty). The prediction is generally accurate to $<8$ per cent, except for the region where $T_{\text{eff}} < 4250$ K or $\log g < 2$, because training data are sparse in this region of luminous giants. The NN parallax precision starts to significantly degrade for spectral SNR $<50$. The NN parallax precision is constant with Gaia parallax SNR and, at spectral SNR $>100$, with spectral SNR, demonstrating that the NN parallax uncertainty is due to scatter in the relation between spectra and luminosity, which fundamentally limits the distance precision that can be obtained from APOGEE spectra.

The distances for luminous red giants from Hogg et al. (2018) are the only ones among the three comparison distances that are determined using purely data-driven techniques that are similar to those used in this work. Hogg et al. (2018) employ a linear (in exponential space) model for how parallax depends on the APOGEE spectrum and broad-band photometry. Their model has significantly fewer trainable parameters than the NN used in this work and they also independently infer the extinction to each star. Hogg et al. (2018) only provide distances for the APOGEE luminous red-giants with $\log g < 2.2$, $(J - K) < (0.4 \text{ mag}) + 0.45(G_{\text{BP}} - G_{\text{RP}})$ and $(H - W_2) > (-0.05 \text{ mag})$ where $J, K, H, W_2$ are 2MASS and WISE photometry. Hogg et al. (2018) report 10 per cent distance estimates.

The distances determined by the NN from this work are compared to those for the same stars in the three comparison catalogues in Fig. 9. The median offset and the scatter in the offset are indicated in the legend of each panel. Overall, there is almost no bias between the NN distances and those in the APOGEE-RC catalogue (median offset is $<1$ per cent); the scatter in the offset is $\approx6$ per cent, similar to the precision quoted for RC stars by Bovy et al. (2014). The typical NN uncertainty for RC stars is $\approx10$ per cent. Thus, the NN distances for RC stars are very similar to those in the RC catalogue, even though the distance estimates are obtained using very different methods.

The middle panel of Fig. 9 compares our NN distances to those in the APOGEE BPG distance catalogue. Overall, there is about a 5 per cent offset in the sense that BPG distances are larger than those returned by the NN. This trend is driven by stars located towards the centre of the Galaxy and is likely caused by the Galactic-density prior used by the BPG method. This prior contains a massive bulge that has the effect of making it more likely that a star is deep within the bulge. For stars located in the direction of the outer Galaxy, the NN and BPG distances agree much better. Aside from this overall shift towards larger distances in the BPG catalogue, the robust scatter between the NN and BPG distances is just over 10 per cent, but note that there are many stars in the tails of the difference distribution.

The right-hand panel of Fig. 9 compares our results to the distances from Hogg et al. (2018). Overall, there is only a small $\approx3$ per cent offset between these two distance estimates with scatter $\approx10$ per cent and few outliers. The good agreement
between our and the Hogg et al. (2018) distances hold over the entire ≈ 10 kpc distance range contained within the luminous red giant sample. The uncertainties in the distances reported by Hogg et al. (2018) are likewise similar to our NN model uncertainty. This is a remarkable agreement between these two different methods.

6 ABUNDANCES ACROSS THE MILKY WAY

To illustrate the power of the NN distances determined in this work, we make maps over a large fraction of the Milky Way disc of elemental abundance ratios for the elements measured by APOGEE. To do this, we combine the NN distances from this work with the NN-determined stellar parameters and abundances from Leung & Bovy (2019). We select stars from APOGEE DR14 with NN distance uncertainty less than 20 per cent, log g uncertainty less than 0.2 dex (as explained in Leung & Bovy 2019), this cut removes dwarfs for which the APOGEE and NN abundance measurements are unreliable, and [Fe/H] uncertainty less than 0.05 dex. We convert three-dimensional coordinates from the heliocentric to the Galactocentric coordinate frame by adopting the distance to the Galactic Centre of 8.125 kpc from Gravity Collaboration (2018) and the Sun’s height above the plane of 20.8 pc from Bennett & Bovy (2019). To focus on trends in the abundance ratios within the disc, we further require that stars be within 300 pc from the Galactic mid-plane; 52,476 stars pass all of these cuts. When we consider abundance ratios relative to iron, we additionally only use stars with [X/H] uncertainty less than 0.1 dex for the considered element, but the actual typical uncertainty is generally still less than 0.05 dex except [Na/H] having ≈ 0.09 dex uncertainty; this creates samples ranging in size from 15,833 (for Na) to 52,476 (for Si and Ni).

Figs 10 and 11 show the median elemental abundance ratios in bins of ≈ (500 pc)² size across the Milky Way disc. It is clear that [Fe/H] peaks at a Galactocentric radius of ≈ 5 kpc and that [Fe/H] decreases monotonically inwards and outwards from this peak. This is markedly different from the behaviour found in the inner galaxy by Hayden et al. (2014) using APOGEE DR10 data, which showed a flattening of the metallicity gradient within ≈ 6 kpc, but with a slight increase of the metallicity towards the centre all the way to R ≈ 2 kpc. This difference is most likely due to the adopted distances. Hayden et al. (2014) determined distances using a Bayesian methodology using a Galactic density prior similar to that of the BPG distances that we compared to in Section 5.2 and the effect of this prior is to place stars towards the bulge at larger distances. Because of the good agreement between our distances and the Gaia parallaxes for bright, luminous giants on the one hand and the alternative distance determination from Hogg et al. (2018) on the other hand, we believe that our distances for stars towards the bulge are more reliable. Thus, it appears that [Fe/H] in the Milky Way peaks at ≈ 5 kpc from the centre and declines to ≈ −0.3 dex in the centre. That the bulge/bar region is more metal-poor than the innermost reaches of the disc is in agreement with other analyses of ARGOS and APOGEE data (Ness et al. 2013, 2016), but our larger sample and precise distances allow the spatial metallicity trend in the inner Milky Way to be mapped in much greater detail.

Fig. 11 displays the spatial trend in elemental abundance ratios with respect to iron over a wide area of the Galactic disc. Abundance ratios of alpha elements ([O/Fe], [Mg/Fe], [Si/Fe], [S/Fe], [Ca/Fe], and [Ti/Fe]) largely trace each other in all parts of the Galactic disc, although notably [Mg/Fe] stays relatively constant towards larger Galactocentric radii, while other alpha elements like [O/Fe], [Si/Fe], and [Ti/Fe] display a clear upturn towards the outer disc. Odd-Z elements [AI/Fe] and [K/Fe] trace the alpha element [Mg/Fe] everywhere. Among iron-peak elements, [Ni/Fe] is almost constant everywhere, as expected, while [V/Fe], [Mn/Fe], and [Co/Fe] show a peak at ≈ 5 kpc similar to that in the [Fe/H] map. [C/Fe] and [N/Fe] are approximately constant outside the solar circle, but inside the solar circle and especially inside the bulge/bar, [C/Fe] strongly increases and is anticorrelated with [N/Fe], which decreases. The spatial abundance ratio trends place strong constraints on nucleosynthetic yields from different processes and on the history of chemical enrichment in different parts of the disc that we will pursue in future work.

We use the Ti abundance determined only from the Ti ii line in the APOGEE spectral region, due to issues with the analysis of neutral Ti lines in the APOGEE analysis (e.g. Hawkins et al. 2016)
Figure 9. Comparison of the NN distances from this work to three different spectro-photometric distances: APOGEE red-clump giants (Bovy et al. 2014; left-hand panel), APOGEE BPG distances (Santiago et al. 2016; middle panel), and distances for luminous red giants (Hogg et al. 2018; right-hand panel). We cut the sample to those stars with <20 per cent uncertainty in both distances that are compared in each panel.

Figure 10. [Fe/H] as a function of position in the Galactic disc for stars within 300 pc from the disc’s mid-plane. The map displays the median [Fe/H] in bins ≈(500 pc)$^2$ in size across the Milky Way disc using NN abundances and distances. Each bin contains at least 10 stars. GC denotes the Galactic Centre. The two circles centred on GC represent 5 and 8.125 kpc (the adopted Solar circle). The ellipse centred at GC illustrates a 5 kpc bar with a 2:1 axial ratio oriented at 25 deg. [Fe/H] in the Milky Way peak at R$\approx$5 kpc, likely due to the transition to the barred region, and decline strongly from this peak towards the Galactic Centre and towards the outer disc.

Finally, it is interesting to note that the shape of constant abundance-ratio contours in the inner Galaxy does not appear to be Galactocentric circles, but that they appear to be elliptical and aligned with the bar in the inner Milky Way (illustrated with the red ellipse in Figs 10 and 11). Thus, we tentatively detect the bar in the spatial behaviour of elemental abundance ratios. Previous determinations of the bar’s shape are based on star counts only (e.g. Blitz & Spergel 1991; Wegg & Gerhard 2013; Wegg, Gerhard & Portail 2015) and the bar is also detected clearly in the kinematics of stars and gas in the inner Milky Way (e.g. Binney et al. 1991; Portail et al. 2017). But our chemical Milky Way maps provide the first direct evidence from abundance ratios that the bar region extends out to 5 kpc, consistent with a fast-rotating, long bar as found by Portail et al. (2017). Future data releases from APOGEE will include data from its southern extension, APOGEE-South, which will test this picture further by filling in the region at negative $y$ (210$^\circ$ $\lesssim l \lesssim$ 360$^\circ$) in the inner Milky Way.

7 DISCUSSION

7.1 Comparison to other Gaia DR2 zero-point offset determinations

With the release of the Gaia DR2 data, Lindegren et al. (2018) provided a determination of the parallax zero-point offset of $-29 \pm 13$ μas using quasars. Because the zero-point offset likely depends on magnitude, colour, and sky location, subsequently a number of work attempted to estimate the zero-point offset. Zinn et al. (2018) used stars in the APOKASC sample of stars with asteroseismic and spectroscopic observations to estimate the offset as $-52.8 \pm 2.4$(stat.) $\pm 1$(syst.) μas for red-giant branch stars and $-50.2 \pm 2.5$(stat.) $\pm 1$(syst.) μas for RC stars, which as we discussed above is in good agreement with our determination of the zero-point offset. Hogg et al. (2018) similarly used luminous red giants in APOGEE to estimate the offset as $-48 \pm 2$ μas. Riess et al. (2018) used Milky Way Cepheid variable stars using a previous joint determination of the period–luminosity relation and the distance ladder in the local Universe and estimated the offset to be $-46 \pm 13$ μas. Recently, Graczyk et al. (2019) employed Milky Way detached eclipsing binary stars to obtain $-31 \pm 11$ μas.

The discrepancies between these different determinations of the zero-point offset are likely caused by the multivariate nature of the zero-point offset and we have determined the multivariate zero-point dependence using a flexible model (see Fig. 4). Zinn et al. (2018) and Hogg et al. (2018) both estimate similar offsets as we have found for the bulk of the red-giants stars used in this work, because these works use similar populations of red giant stars in APOGEE as the majority of our sample. Both Lindegren et al. (2018) and Graczyk et al. (2019) favour a smaller offset compared to our work, because (a) the quasars in Lindegren et al. (2018) are generally much bluer and dimmer compared to the giants in this work and (b) the eclipsing binaries used in Graczyk et al. (2019) are bluer and brighter than the stars we consider. The quasars are so far outside of the magnitude and colour range of stars in our sample and have such different spectral energy distributions that it is difficult to extrapolate our trends in Fig. 4 to quasars. The $(G, G_{BP} - G_{RP})$ panel in Fig. 4 demonstrates that we find that stars that are at the
Figure 11. Maps of [X/Fe] abundance ratios for 15 different elements measured from APOGEE spectra across the Galactic disc. As in Fig. 10, the map displays the median [X/Fe] in bins \((500 \text{ pc})^2\) in size for bins with more than 10 stars. The circles, ellipse, and other markings are as in Fig. 10.
blue, bright end of our sample have offsets of $\gtrsim -30 \mu$as, similar to Graczyk et al. (2019). The uncertainty in the zero-point offset determined using Cepheids from Riess et al. (2018) is large enough to be consistent with all other determinations of the zero-point, including ours.

Compared to the methods employed by these other works, our technique does not rely on any physical or stellar model, but is instead entirely data-driven and our technique has far more freedom than the other methods. The determination of the parallax zero-point offset is crucial to almost all applications of the Gaia DR2 data and it has significant implications for such questions as what the Hubble constant is. The $(G, BP - G, T_{\text{eff}})$ dependence that we find in Fig. 4 points towards a zero-point offset for the bright, long-period Cepheids used in the determination of the Hubble constant that is smaller in absolute value and close to the value derived from quasars $\approx -30 \mu$as. This would place Cepheids at somewhat larger distances than the distances used by Riess et al. (2018) and would therefore lead to a slight reduction in the inferred Hubble constant, reducing the tension with the value determined using cosmology and the inverse distance ladder (Aubourg et al. 2015; Planck Collaboration VI 2018).

### 7.2 Comparison to other spectro-photometric distances for red giants

We compared the NN distances that we determined to three other methods for determining spectro-photometric distances in Section 5.2. Here we further discuss the similarities and differences between our method and these other techniques:

(i) Both the NN used in this work and the Hogg et al. (2018) technique map (APOGEE) spectra to luminosity in purely data-driven way, by training on the overlap between APOGEE and Gaia data. The BPG distance determinations rely on stellar models and a Galaxy density prior during inference. That we do not use a density prior or stellar models is both an advantage and a disadvantage. Our and the Hogg et al. (2018) distances are therefore solely informed by the data and they do not depend on imperfect stellar models and imperfect knowledge of the Galaxy’s density distribution and how the APOGEE/Gaia selection function affects the observed number counts. On the other hand, we cannot easily extrapolate our results to stellar types outside of those included in our training set. But because we train directly on Gaia parallaxes (and simultaneously infer the Gaia zero-point offset), there is no doubt that our NN distances have higher precision and accuracy than those determined using stellar models and density priors, such as the BPG distances, for stars within the bounds of the training sample.

(ii) Hogg et al. (2018) distances are determined using a much simpler model than the NN used in this work. They use a linear model for the logarithm of the parallax, whereas we use a multilayer artificial NN as a universal function approximation. Because their model is so simple, Hogg et al. (2018) only train on bright red giants. The model in this work is complex and therefore it can infer luminosity for a wider range of stars. Our model returns precise distances, even at the edges of the training sample, and it also returns realistic uncertainties that can be used to identify stars for which the NN distance is unreliable (e.g. stars outside of the bounds of the training set). Fears that the complexity allowed by the NN leads to bad extrapolation are therefore unfounded for our model. We are therefore able to provide a single distance-estimation technique that can be applied from solar-luminosity dwarfs to the most luminous red giants.

(iii) Hogg et al. (2018) distance methods contain a data-driven extinction model that makes use of additional broad-band photometry, while in our work we rely on external extinction data. This is not a great limitation for the application to APOGEE data, because extinctions determined using the RJCE method are highly reliable. But when applying our type of modelling to other data sets being able to include a data-driven extinction estimation would be highly desirable. It would be possible to construct a simpler NN to determine the extinction to each star that is similar to our existing offset calibration model, where in the extinction case the inputs would be multiband broad-band photometric data similar to Hogg et al. (2018).

(iv) Different from Hogg et al. (2018) and essentially all other spectro-photometric distance methods, we calibrate the Gaia zero-point offset simultaneously with the training of the spectro-photometric model. We are even able to infer how the zero-point offset depends on stellar properties $G, G_{BP} - G_{RP}$, and $T_{\text{eff}}$, with a flexible NN model for the zero-point offset. Thus, we have a consistent model for the Gaia DR2 zero-point offset and spectro-photometric distances.

### 7.3 Future applications and challenges

The method described in this work can be easily applied to future Gaia data release with the high-level API of the astropy NN package. The amount of data to train the algorithm will increase in future APOGEE and Gaia data releases. Other spectroscopic surveys such as GALAH (De Silva et al. 2015) and SDSS-V (Kollmeier et al. 2017) will provide even larger training sets that also contain a wider range of stellar types. The flexibility of the NN technique will allow these data sets to be incorporated into a single method. Furthermore, future Gaia data releases will include low-resolution optical spectra for all Gaia sources. Training on these data will provide a way to improve the distance estimates of all Gaia stars, although it remains to be determined by how much.

While the Gaia processing of the satellite measurements to determine parallaxes will improve and likely lead to smaller overall parallax zero-point offsets, it is likely that future data release will still suffer from unknown zero-point issues. We have shown that the zero-point offset in Gaia DR2 is a function of $G, G_{BP} - G_{RP}$, and $T_{\text{eff}}$ and the zero-point may further depend on sky location (e.g. Lindegren et al. 2018) and other properties. We have shown that with the $\approx 35\,000$ member APOGEE training set, we can determine the zero-point offset to $\approx 2\mu$as precision, even if we allow the zero-point to depend on $G, G_{BP} - G_{RP}$, and $T_{\text{eff}}$. Thus, simultaneously calibrating spectro-photometric distances and the multivariate nature of the parallax zero-point is a way to obtain a high-precision zero-point calibration. However, determining the zero-point offset’s multivariate dependencies is challenging, because the NN is complex. If we give the NN more degrees of freedom, the NN will may start deviating from physically plausible trends, as the NN only minimizes the objective function. For example, the zero-point offset is known to have spatial covariance and a quasi-regular triangular pattern with a period of about one degree (see the LMC test in Lindegren et al. 2018). We believe that it will be difficult for an NN, especially the Bayesian NN we employ here which is highly regularized, to capture such high-frequency periodic variations in equatorial coordinates, although it may be possible to recast the spatial dependence of the zero-point into a more well-behaved set of properties.

All data-driven methods for determining the zero-point essentially suffer from the issue that the offset calibration might
not be physical. Using standard(zable)-candle stars such as red clump stars or variable stars as well as using extragalactic sources like quasars can only provide the offset calibration in the colour/position/magnitude space covered by each sample. It is difficult to use an ensemble of them to average out the bias or inaccuracy introduced by each one. Because our method is flexible enough to allow for a wide range of stellar types to be incorporated simultaneously, it can determine the zero-point’s dependence on colour/position/magnitude consistently over a wide range in these properties.

8 CONCLUSIONS

Milky Way stars are being surveyed astrometrically, photometrically, and spectroscopically in an unprecedented manner thanks to survey like Gaia and APOGEE. The distance to each star is one of its most crucial properties and while Gaia is revolutionizing the measurement of stellar distances using its parallax measurements, Gaia can only determine precise distances within a small volume compared to the Galaxy as a whole. Existing survey like APOGEE – which will soon also release data from its southern extension, APOGEE-S – cover a wide volume of the Milky Way. Machine learning provides an opportunity to use the precise Gaia distances for nearby APOGEE stars to train a spectro-luminosity model and to use this model to provide precise spectro-photometric distances for a sample that covers a large part of the Milky Way disc.

In this paper, we have simultaneously calibrated spectro-photometric distances and the Gaia DR2 parallax zero-point offset with deep learning using NNs. We have determined spectro-photometric distances for the entire APOGEE DR14 sample of 277,371 stellar spectra which cover Galactocentric radii 2 kpc ≤ R ≤ 19 kpc. About 150,000 of these has <10 per cent distance uncertainties. Our spectro-photometric distances have higher SNR than Gaia beyond ≈2 kpc. We determined the Gaia DR2 zero-point offset to be 52.3 ± 2.0 μas when fitting a constant offset model to mainly APOGEE giants. When allowing the Gaia offset to depend on G, GBP − G_RP, and T_eff, the result is shown in equation (9) and Fig. 4; the mean of this model is ≈50 μas with reasonable dependencies of the offset on G, GBP − G_RP, and T_eff that explain the differences between various zero-point determinations in the literature using different types of stars. Our zero-point determinations can be applied directly to ≈58 million stars in Gaia DR2 within the colour–magnitude and sky footprint covered by APOGEE. We release the catalogue of spectro-photometric distances for APOGEE DR14 stars, with the data model given in Table 1 (see Section 1 for the data-file link). Moreover, all trained NN models from this work are publicly available (see Appendix A), including code to determine the zero-point offset for any Gaia star.

We compared our spectro-photometric distances to a subset of highly accurate Gaia parallaxes and to several alternative spectro-photometric distance catalogues. Our distances show excellent agreement with highly precise Gaia parallaxes, with distances to APOGEE RC stars, and with the distances to luminous red giants from Hogg et al. (2018). But our distances give precise distances for a much wider range of stellar types and distances than any other existing method. We used the distances for the entire APOGEE DR14 sample and employed chemical abundances from our previous work (Leung & Bovy 2019) to analyse spatial trends in 15 elemental abundance ratios across the Milky Way and find interesting trends across the Galaxy. In particular, we demonstrate that a bar-like trend is clearly apparent in many of the abundance ratios, consistent with an ≈5 kpc half-length of the Galactic bar.

This method can be applied to future Gaia data releases and to future spectroscopic surveys. Improving the determination of the true complex multivariate Gaia zero-point offset will be challenging and likely require inputting a good, physical model for the more complex dependencies of the zero-point (like its spatial dependence). But the Gaia zero-point offset will likely decrease in magnitude in future Gaia data releases and even with the current zero-point model, our method demonstrates that obtaining <10 per cent spectro-photometric distances is achievable for a wide range of main-sequence and giant stellar types.

ACKNOWLEDGEMENTS

It is a pleasure to thank Frédéric Arenou for helpful comments. HL and JB received support from the Natural Sciences and Engineering Research Council of Canada (NSERC; funding reference number RGPIN-2015-05235) and from an Ontario Early Researcher Award (ER16-12-061). JB also received partial support from an Alfred P. Sloan Fellowship.

We gratefully acknowledge the support of NVIDIA Corporation with the donation of a Titan Xp GPU used in this research.

Funding for the Sloan Digital Sky Survey IV has been provided by the Alfred P. Sloan Foundation, the U.S. Department of Energy Office of Science, and the Participating Institutions. SDSS-IV acknowledges support and resources from the Centre for High-Performance Computing at the University of Utah. The SDSS web site is www.sdss.org.

This work has made use of data from the European Space Agency (ESA) mission Gaia (https://www.cosmos.esa.int/gaia), processed by the Gaia Data Processing and Analysis Consortium (DPAC, https://www.cosmos.esa.int/web/gaia/dpac/consortium). Funding for the DPAC has been provided by national institutions, in particular the institutions participating in the Gaia Multilateral Agreement.

REFERENCES

Abolfathi B. et al., 2018, ApJS, 235, 42
Anderson L., Hogg D. W., Leistedt B., Price-Whelan A. M., Bovy J., 2018, AJ, 156, 145
Antoja T. et al., 2018, Nature, 561, 360
Arenou F., Lindegren L., Froecklisch M., Gomez A. E., Turon C., Perryman M. A. C., Wielen R., 1995, A&A, 304, 52
Arenou F. et al., 2018, A&A, 616, A17
Aubourg ´E. et al., 2015, Phys. Rev. D, 92, 123516
Bailer-Jones C. A. L., 2015, PASP, 127, 994
Bennett M., Bovy J., 2019, MNRAS, 482, 1417
Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, MNRAS, 252, 210
Blitz L., Spiegel D. N., 1991, ApJ, 379, 631
Bovy J., 2016, ApJ, 817, 49
Bovy J. et al., 2014, ApJ, 790, 127
Casey A. R., Hogg D. W., Ness M., Rix H.-W., Ho A. Q. Y., Gilmore G., 2016, preprint (arXiv:1603.03040)
Churchwell E. et al., 2009, PASP, 121, 213
De Silva G. M. et al., 2015, MNRAS, 449, 2604
Fabbro S., Venn K. A., O’Briain T., Bialek S., Kielty C. L., Jahandar F., Monty S., 2018, MNRAS, 475, 2978
Freeman K., Bland-Hawthorn J., 2002, ARA&A, 40, 487
Gaia Collaboration, 2016, A&A, 595, A1
Gaia Collaboration, 2018a, A&A, 616, A11
Gaia Collaboration, 2018b, A&A, 616, A1
Graczyk D. et al., 2019, ApJ, 872, 85
APPENDIX A: EXAMPLE OF USING NEURAL NET TO INFERENCE LUMINOSITY AND CONVERT TO DISTANCE ON ARBITRARY APOGEE SPECTRA

Besides providing general tools for deep learning in astronomy in the astronn package, we also share the actual networks trained and discussed in this paper in a separate GitHub repository associated with this paper. Here we give an example of how to use the spectro-luminosity model network for determining stellar luminosity and convert to distance for a given APOGEE spectrum. Moreover, code to use the Offset Calibration Model separately to get the calibrated offset is not included in astronn, but it is provided in the Github repository for this work. First follow the following instructions:

(i) Install astronn by following instructions from https://astronn.readthedocs.io/en/v1.1.0/quick_start.html.
(ii) Obtain the repository containing the code to reproduce all figures in this paper at https://github.com/henrysky/astroNN_gaia_dr2_paper.
(iii) Open a python terminal under the repository folder but outside the neural network model folder.
(iv) Copy and paste the following code to do inference with the neural net in this paper on the star 2M19060637 + 4717296.

Listing 1: Example of using Neural Net to infer distance on APOGEE spectra

```python
from astropy.io import fits
from astroNN.apogee import visit
from astroNN.gaia import extinction_correction, fakemag_to_pc
from astroNN.models import load_folder

# arbitrary spectrum
f = fits.open('visit_spectra(dr=14, apogee = ’2M19060637 + 4717296’))
spectrum = f[1].data
spectrum_err = f[2].data
spectrum_bitmask = f[3].data

# using default continuum and bitmask values to continuum normalize
norm_spec, norm_spec_err = apogee_continuum(spectrum, spectrum_err, bitmask=spectrum_bitmask, dr=14)

# load neural net, it is recommend to use model ends with _reduced
# for example, using astroNN_constant_model_reduced instead of astroNN_constant_model
neuralnet = load_folder(‘astroNN_constant_model_reduced’)

# inference, if there are multiple visits, then you should use the globally weighted combined spectra (i.e. the second row)
pred, pred_err = neuralnet.test(norm_spec)

# correct for extinction
K = extinction_correction(f[0].header[‘K’], f[0].header[‘AKTARG’])

# convert prediction in fakemag to distance
pc, pc_error = fakemag_to_pc(pred[:, 0], K, pred_err[‘total’][:, 0])
print(f’Distance: {pc} +/− {pc_error}’)
```

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