Three-dimensional interstellar dust reddening maps of the Galactic plane

B.-Q. Chen,1,⁎ Y. Huang,1 X.-W. Liu,1,⁎ H.-B. Yuan,2 C. Wang,3 D.-W. Fan,4
M.-S. Xiang,4 H.-W. Zhang,3 and Z.-J. Tian3

1South-Western Institute for Astronomy Research, Yunnan University, Kunming, Yunnan 650091, P. R. China
2Department of Astronomy, Beijing Normal University, Beijing 100875, P. R. China
3Department of Astronomy, Beijing Normal University, Beijing 100875, P. R. China
4National Astronomy Observatories, Chinese Academy of Sciences, Beijing 100101, P. R. China

Accepted ???. Received ???; in original form ???

ABSTRACT

We present new three-dimensional interstellar dust reddening maps of the Galactic plane in three colours, E(G − Ks), E(BP − RP) and E(H − Ks). The maps have a spatial angular resolution of 6 arcmin and covers over 7000 deg² of the Galactic plane for Galactic longitude 0° < l < 360° and latitude |b| < 10°. The maps are constructed from robust parallax estimates from the Gaia Data Release 2 (Gaia DR2) combined with the high-quality optical photometry from the Gaia DR2 and the infrared photometry from the 2MASS and WISE surveys. We estimate the colour excesses, E(G − Ks), E(BP − RP) and E(H − Ks) simultaneously, toward over 44 million stars with the machine learning algorithm Random Forest regression, using a training data set constructed from the large-scale spectroscopic surveys LAMOST, SEGUE and APOGEE. The results reveal the large-scale dust distribution in the Galactic disk, showing a number of features consistent with the earlier studies. The Galactic dust disk is clearly warped and show complex structures associated spatially with the Sagittarius, Local and Perseus arms. We find that the boundaries between the Local arm and the other two arms are not clearly defined, there are several clouds connecting the three arms. We also provide the empirical extinction coefficients for the Gaia photometry that can be used to convert the colour excesses presented here to the line-of-sight extinction values in the Gaia photometric bands.

Key words: ISM: dust, extinction – ISM: structure – Galaxy: structure

1 INTRODUCTION

The interstellar dust grains dim and redden stellar light from the ultraviolet (UV) to the infrared (IR) (Draine 2003). They show an inhomogeneous clumpy distribution and increase sharply towards the Galactic plane (Schlegel et al. 1998; Planck Collaboration et al. 2014), and thus pose as a serious obstacle for the study of the structure, formation and evolution of our Galaxy, especially for the low Galactic latitude regions, specifically the Galactic disk (Chen et al. 2013). For any stellar study near the Galactic plane, one needs to correct for the effects of dust extinction and reddening to interpret the observations properly.

Traditional two-dimensional (2D) extinction maps give the total amount of extinction in a given direction integrated along the line-of-sight. Consequently, the value represents an upper limit of the real one for a local disc star in that direction. In the past decades, thanks to a number of large-scale photometric and spectroscopic surveys, such as the Sloan Digital Sky Survey (SDSS; York et al. 2000), the Two Micron All Sky Survey (2MASS; Skrutskie et al. 2006), the LAMOST Experiment for Galactic Understanding and Exploration (LEGUE; Deng et al. 2012; Zhao et al. 2012) and the Pan-STARRS 1 Survey (PS1; Chambers et al. 2016), three-dimensional (3D) extinction maps constructed based on estimates of the distances and extinction values to millions of individual stars have become available.

Based on the 2MASS data, Marshall et al. (2006) present a three-dimensional (3D) extinction model of the inner Galaxy (|l| < 100° and |b| < 10°) by comparing the observed colours of giant stars for each line of sight with the synthetic values from the Besançon Galactic model (Robin et al. 2003). Using a similar method, Chen et al. (2013) and Schultheis et al. (2014) present the colour excess maps for several colours toward the Galactic Bulge (|l| < 10° and −10° < b < 5°) based on data from the ESO Public Survey, VISTA Variables in the Via Lactea (VVV; Minniti et al. 2010) and the Galactic Legacy Infrared Mid-Plane Survey Experiments (GLIMPSE; Benjamin et al. 2008) and the 2MASS survey (Skrutskie et al. 2006). A new three-dimensional extinction map has been presented by Chen et al. (2013) and Schultheis et al. (2014) from the ESO DR2 Parallax Catalogue using the Gaia DR2 catalogue (Gaia Collaboration et al. 2018).

⁎ E-mail: bchen@ynu.edu.cn (BQC); x.liu@ynu.edu.cn (XWL).
B.Q. Chen et al.

The XSTPS-GAC survey area is an extinction map that covers the entire Xuyi Schmidt Telescope Photometric Survey of the Galactic Anticentre (XSTPS-GAC) survey area of over 6000 deg$^2$ (140$^\circ$ < l < 220$^\circ$ and b < 40$^\circ$), based on the XSTPS-GAC optical and the 2MASS and Wide-Field Infrared Survey Explorer (WISE; Wright et al. 2010) IR photometry. Using a hierarchical Bayesian model, Sale et al. (2014) derive a 3D extinction map in the Northern Galactic Plane (30$^\circ$ < l < 215$^\circ$ and |b| < 5$^\circ$) from the optical photometry of the INT/WFC Photometric Hα Survey (IPHAS; Drew et al. 2005), Green et al. (2015) and Green et al. (2018) apply the Bayesian approach to the PS1 optical and the 2MASS IR photometry to produce 3D maps of dust reddening over 2MASS Point Source Catalog (PSC; Skrutskie et al. 2006) provides photometry for over 500 million objects. The uncertainties of 2MASS photometric measurements are estimated to be smaller than 0.03 mag. The WISE surveys the entire sky in four IR bands, $G$, $G_{BP}$ and $G_{RP}$ magnitudes and 1.3 billion sources have parallax and proper motion measurements. The $G$ band covers the whole optical wavelength range from 330 to 1050 nm. The $G_{BP}$ and $G_{RP}$ magnitudes are derived from the low resolution spectrophotometric measurements integrated over the wavelength ranges 330 - 680 nm and 630 - 1050 nm, respectively. The internal validation of the Gaia DR2 shows that the calibration uncertainties for $G$, $G_{BP}$ and $G_{RP}$ are 2, 5 and 3 mmag, respectively. To break the degeneracy of effective temperature (or intrinsic colours) and extinction for the individual stars, we combine the Gaia optical photometry with the IR photometry of 2MASS and WISE. The 2MASS surveys the entire sky in three near-IR bands, $J$, $H$ and $K_s$, centered at 1.25, 1.65 and 2.16 $\mu$m, respectively. The 2MASS Point Source Catalog (PSC) provides photometry for over 500 million objects. The uncertainties of 2MASS photometric measurements are estimated to be smaller than 0.03 mag. The WISE surveys the entire sky in four IR bands, $W_1$ to $W_4$, centered at 3.4, 4.6, 12 and 22 $\mu$m, respectively. The AllWISE Source Catalog provides four band magnitudes and variability statistics for over 747 million objects. For the WISE, we use only the data of band $W_1$, as the longer wavelength measurements have lower sensitivities and poorer angular resolutions. Including the latter in the analysis do not improve the parameter estimation.

We select stars from the Gaia DR2 with Galactic latitude |b| < 10$^\circ$ and then cross-match them with the 2MASS PSC and the AllWISE Source catalogues using the Centre de Donnes astronomiques de Strasbourg (CDS) XMatch Service. The matching radius is set to 1.5 arcsecond. The fraction of multiple matches is less than 0.01 percent and the matches of the closest positions are adopted. To select sample stars from the combined Gaia/2MASS/WISE catalogue, we require that the sources must be detected in all bands, i.e., Gaia $G$, $G_{BP}$, and $G_{RP}$, 2MASS $J$, $H$, and $K_s$ and WISE $W_1$. We further require that the sources have photometric errors less than 0.08 mag in all bands and 2MASS quality of ‘AAA’. The cuts lead to a total 44,218,735 stars in the combined catalogue.

3 METHOD

In the SED fitting or Bayesian approach, one uses gridding (Berry et al. 2012; Chen et al. 2014) or the Markov chain Monte Carlo (MCMC) methods (Green et al. 2014; Sale et al. 2014) to sample the parameter space in order to derive the extinction (or colour excess) values of the individual stars. In the current work, we build a model that returns the colour excess values of stars given the data described in Section 2. The model is based on a machine-learning approach (GLIMPSE; Churchwell et al. 2009). By combining the SDSS optical and the 2MASS near-IR photometry, Berry et al. (2012) simultaneously estimate distances and values of extinction of stars by fitting the observed optical (and IR) spectral energy distributions (SEDs). Analogously, Chen et al. (2014) present a 3D extinction map that covers the entire Xuyi Schmidt Telescope Photometric Survey (XSTPS-GAC) survey area of over 6000 deg$^2$ (140$^\circ$ < l < 220$^\circ$ and b < 40$^\circ$), based on the XSTPS-GAC optical and the 2MASS and Wide-Field Infrared Survey Explorer (WISE; Wright et al. 2010) IR photometry. Using a hierarchical Bayesian model, Sale et al. (2014) derive a 3D extinction map in the Northern Galactic Plane (30$^\circ$ < l < 215$^\circ$ and |b| < 5$^\circ$) from the optical photometry of the INT/WFC Photometric Hα Survey (IPHAS; Drew et al. 2005), Green et al. (2015) and Green et al. (2018) apply the Bayesian approach to the PS1 optical and the 2MASS IR photometry to produce 3D maps of dust reddening over 2MASS Point Source Catalog (PSC; Skrutskie et al. 2006) provides photometry for over 500 million objects. The uncertainties of 2MASS photometric measurements are estimated to be smaller than 0.03 mag. The WISE surveys the entire sky in four IR bands, $G$, $G_{BP}$ and $G_{RP}$ magnitudes and 1.3 billion sources have parallax and proper motion measurements. The $G$ band covers the whole optical wavelength range from 330 to 1050 nm. The $G_{BP}$ and $G_{RP}$ magnitudes are derived from the low resolution spectrophotometric measurements integrated over the wavelength ranges 330 - 680 nm and 630 - 1050 nm, respectively. The internal validation of the Gaia DR2 shows that the calibration uncertainties for $G$, $G_{BP}$ and $G_{RP}$ are 2, 5 and 3 mmag, respectively. To break the degeneracy of effective temperature (or intrinsic colours) and extinction for the individual stars, we combine the Gaia optical photometry with the IR photometry of 2MASS and WISE. The 2MASS surveys the entire sky in three near-IR bands, $J$, $H$ and $K_s$, centered at 1.25, 1.65 and 2.16 $\mu$m, respectively. The 2MASS Point Source Catalog (PSC) provides photometry for over 500 million objects. The uncertainties of 2MASS photometric measurements are estimated to be smaller than 0.03 mag. The WISE surveys the entire sky in four IR bands, $W_1$ to $W_4$, centered at 3.4, 4.6, 12 and 22 $\mu$m, respectively. The AllWISE Source Catalog provides four band magnitudes and variability statistics for over 747 million objects. For the WISE, we use only the data of band $W_1$, as the longer wavelength measurements have lower sensitivities and poorer angular resolutions. Including the latter in the analysis do not improve the parameter estimation.

We select stars from the Gaia DR2 with Galactic latitude |b| < 10$^\circ$ and then cross-match them with the 2MASS PSC and the AllWISE Source catalogues using the Centre de Donnes astronomiques de Strasbourg (CDS) XMatch Service. The matching radius is set to 1.5 arcsecond. The fraction of multiple matches is less than 0.01 percent and the matches of the closest positions are adopted. To select sample stars from the combined Gaia/2MASS/WISE catalogue, we require that the sources must be detected in all bands, i.e., Gaia $G$, $G_{BP}$, and $G_{RP}$, 2MASS $J$, $H$, and $K_s$ and WISE $W_1$. We further require that the sources have photometric errors less than 0.08 mag in all bands and 2MASS quality of ‘AAA’. The cuts lead to a total 44,218,735 stars in the combined catalogue.

The paper is structured as following. In Section 2, we present the relevant Gaia DR2, 2MASS and WISE data. Section 3 describes the methods used to derive values of colour excess and to construct the 3D colour excess maps. In Section 4, we present our main results which are discussed in Section 5. We summarize in Section 6.
algorithm, the Random Forest regression, one of the most effective machine learning models for predictive analytics. A Random Forest regression is a meta estimator that fits a number of decision trees to various sub-samples of the dataset and uses average values to improve the predictive accuracy while controls over-fitting. We employ the Random Forest regression to estimate the colour excesses, $E(G - K_S)$, $E(G_{BP} - G_{RP})$ and $E(H - K_S)$ for the individual stars. We have also tried other machine learning algorithms, such as the Support Vector Machine and Extra-Tree regressions and found very similar results.

We first create an empirical training set using data from the spectroscopic surveys. The second release of value-added catalogues of the LAMOST Spectroscopic Survey of the Galactic Anticentre (LSS-GAC DR2; [Xiang et al. 2017]) provides stellar atmospheric parameters deduced from 1.8 million spectra. The catalogues for internal usage include all the observations collected by 2016 June and contains robust stellar parameters estimated from over 5 million spectra [Xiang et al. 2017; Chen et al. 2018]. In addition to the basic stellar atmospheric parameters, the catalogues also provide values of the interstellar dust reddening $E(B - V)$ toward the individual stars estimated with three independent techniques, the star-pair method, comparison with the theoretical isochrones, and SED fitting with multi-band photometry. $E(B - V)$ values yielded by the star-pair method are adopted as the recommended ones as the method is model free and achieves a better precision (about 0.02 mag; [Yuan et al. 2015]).

The recommended values of $E(B - V)$ released in the LSS-GAC DR2 catalogues are weighted means of colour excess estimates in several colours, $g - r$, $r - i$, $i - J$, $J - H$ and $H - K_S$, using the extinction law of [Yuan et al. 2013]. On the other hand, it has been known that the extinction law varies with environment [Chen et al. 2013; Schlafly et al. 2016]. In this work, we will simultaneously estimate the interstellar dust reddening values in different colours, i.e. the colour excesses, without assuming a universal extinction law. We calculate values of $E(G - K_S)$, $E(G_{BP} - G_{RP})$ and $E(H - K_S)$ for stars in the LSS-GAC DR2 using the same star-pair algorithm of [Yuan et al. 2015]. We select stars from the LSS-GAC DR2 with the criteria: LAMOST spectral S/N(4650Å) per pixel > 10, ‘objtype’ = ‘STAR’, ‘moondis’ > 30’, ‘BADFIBER’ = 0, ‘SATFIBER’ = 0, ‘BRIGHTFIBER’ = 0, ‘vr_flag’ ≤ 6 and photometric errors err($G$, $G_{BP}$, $G_{RP}$, $J$, $H$, $K_S$, W1) < 0.1 mag. The requirements lead to 3,485,460 stars for the LAMOST sample. A LAMOST reference sample, required by the star-pair method for colour excess estimation, is further defined by requiring S/N(4650Å) per pixel > 50, the Schlegel et al. (1998, SFD) $E(B - V) < 0.015$ mag and err($G$, $G_{BP}$, $G_{RP}$, $J$, $H$, $K_S$, W1) < 0.03 mag. This reference sample contains 48,243 stars.

We also use data from the SDSS DR14 that provides reliable stellar parameters for 0.4 million stars from the Sloan Extension for Galactic Understanding and Exploration (SEGUE; [Yanny et al. 2009]) and 0.2 million stars from the Apache Point Observatory Galactic Evolution Experiment (APOGEE; [Majewski et al. 2010]). Since the LAMOST surveys mainly the outer disc of the Galaxy, thus the SEGUE and APOGEE data sets complement that of LAMOST. To minimize potential systematics between stellar parameters yielded by the different surveys, colour excesses $E(G - K_S)$, $E(G_{BP} - G_{RP})$ and $E(H - K_S)$ are calculated separately for the SEGUE and APOGEE samples, using the same aforementioned star-pair technique. We select stars from the SEGUE and APOGEE catalogues with the criteria: S/N > 10, $T_{eff}$ > 0, log$g$ > 0, [Fe/H]...
or $[\text{M/H}] > -3$ and $\text{err}(G, G_{\text{BP}}, G_{\text{RP}}, J, H, K_S, W1) < 0.1$ mag. The requirements yield 116,006 and 142,994 stars for the SEGUE and APOGEE samples, respectively. The SEGUE and APOGEE reference samples are defined by requiring S/N $> 50$, SFD $E(B - V) < 0.015$ mag and $\text{err}(G, G_{\text{BP}}, G_{\text{RP}}, J, H, K_S, W1) < 0.08$ mag. The cuts yield 6,386 and 6,579 stars in the SEGUE and APOGEE reference samples, respectively.

In Fig. 2 we compare estimates of $E(G - K_S)$, $E(G_{\text{BP}} - G_{\text{RP}})$ and $E(H - K_S)$ for common stars of the LAMOST and the SEGUE and APOGEE samples. We find no systematics amongst the different samples. For the LAMOST and SEGUE samples, the dispersions of differences are respectively $\sim 0.06$, $0.03$ and $0.008$ mag for $E(G - K_S)$, $E(G_{\text{BP}} - G_{\text{RP}})$ and $E(H - K_S)$ differences. The corresponding values are $\sim 0.08$, $0.04$ and $0.006$ mag between the LAMOST and APOGEE samples.

The resulted LAMOST, SEGUE and APOGEE samples are then combined. Abnormal colour excess estimates are excluded by plotting the intrinsic colours, $(G - K_S)_0$, $(G_{\text{BP}} - G_{\text{RP}})_0$ and $(H - K_S)_0$, versus $T_{\text{eff}}$ (Fig. 2), taking the advantage that the intrinsic colours of stars correlate well with effective temperature.

We reject stars that deviate more than 0.2, 0.15 and 0.05 mag from the best-fit $(G - K_S)_0$, $(G_{\text{BP}} - G_{\text{RP}})_0$ and $(H - K_S)_0$ versus $T_{\text{eff}}$ relations, respectively. This leads us to a spectroscopic sample consisting 3,224,373 stars. In Fig. 3 we show the distribution of those stars (in the training sample, see below) in the Galactic coordinates and in the $E(G - K_S)$ versus $T_{\text{eff}}$ plane. The sample covers almost all the Galactic latitudes and about two thirds of the Galactic longitudes. The effective temperatures range between 3000 and 10500 K. The estimated $E(G - K_S)$ values can be as high as $\sim 6$ mag, corresponding to $E(B - V) \sim 3$ mag (based on the extinction coefficients presented in Sect. 5.1). We note that the derived colour excesses for stars in the spectroscopic sample can be smaller than 0 but larger than $\sim -0.05$ mag, as a result of the photometric errors.

The stars in the final spectroscopic sample are randomly divided into two sub-samples, a training sample consisting of 80 per cent of stars and a test sample containing the remaining 20 per cent stars. The training sample is used to generate the Random Forest models, while the test sample is used to validate the generated relations. We build three separate Random Forest models for $E(G - K_S)$, $E(G_{\text{BP}} - G_{\text{RP}})$ and $E(H - K_S)$, respectively. In all cases, the input parameters are $G - J$, $G_{\text{BP}} - K_S$, $R_{\text{BP}} - J$, $J - H$, $H - K_S$ and $K_S - W1$. We use the scikit-learn package for Python (Pedregosa et al. 2011) to build the models. We test with different parameters to optimize the models. Finally we set the number of trees in the forest to be $n_{\text{estimators}}=200$, the number of features to

**Figure 2.** Intrinsic colours versus $T_{\text{eff}}$ diagrams for stars in the LAMOST/SEGUE/APOGEE spectroscopic sample. Blue pluses and associated error bars represent median values and standard deviations deduced by binning the data points in bins of 50 K. The blue lines are fits to the median values using a fifth-order polynomial. The red lines mark the boundaries used to reject the outliers that fall respectively more than 0.2, 0.15 and 0.05 mag from the fits.

**Figure 3.** Distributions of stars in the final spectroscopic sample in the Galactic coordinates (bottom panel) and in the $E(G - K_S)$ versus $T_{\text{eff}}$ plane (upper panel). Black, blue and red dots represent stars selected from the LAMOST, SEGUE and APOGEE spectroscopic surveys, respectively.
Fig. 4. Comparison of colour excess values yielded by the Random Forest models and those derived with the star-pair technique for, from left to right, $E(G - K_S)$, $E(BP - RP)$ and $E(H - K_S)$, respectively, for the test sample stars. The mean and standard deviation of the differences, are marked in each plot.

consider when looking for the best split to be max_features='auto', the minimum number of samples required to split an internal node to be min_samples_split=2 and the minimum number of samples needed at a leaf node to be min_samples_leaf=1. The 16th and 84th percentiles of the Random Forest ensemble are taken as the uncertainties. Fig. 4 compares the values of $E(G - K_S)$, $E(BP - RP)$ and $E(H - K_S)$ yielded by the Random Forest models and those given by the star-pair technique for the test sample stars. The Figure shows good agreements for all three cases. There is no systematics in the residuals. The standard deviations of the differences are 0.07, 0.04 and 0.005 for $E(G - K_S)$, $E(BP - RP)$ and $E(H - K_S)$, respectively, comparable to what expected from the typical uncertainties of colour excess of the spectroscopic sample.

For distances of the stars, we adopt the values from Bailer-Jones et al. (2018) who calculate distances of 1.3 billion stars from the Gaia measurements, imposing a prior based on the expected distribution of all stars in the Gaia catalogue. In the current work, we accept the distance estimates only for stars with Gaia parallax uncertainties smaller than 20 per cent. To map the 3D dust distribution of all stars in the Gaia catalogue. In the current work, the Gaia measurements, imposing a prior based on the expected uncertainties of colour excess of the spectroscopic sample.

4 DUST REDDENING MAPS

Our resulted 3D colour excess maps are available in electronic form in the online version of this manuscript. Table 1 describes the data format. Rows of the Table contains information of one subfield (pixel): Galactic coordinates, measured values of the integrated $E(G - K_S)$, $E(BP - RP)$ and $E(H - K_S)$ and the associated uncertainties at the individual distance moduli. In addition, the colour excesses, $E(G - K_S)$, $E(BP - RP)$ and $E(H - K_S)$ and the associated errors of the 44,218,735 individual stars in the combined Gaia/2MASS/WISE sample are available upon request by email. Fig. 5 plots the dust corrected colour and absolute magnitude diagram of stars in the combined Gaia/2MASS/WISE sample. The colour scale represents the number of stars per colour-magnitude-WISE sample. The colour scale represents the number of stars per colour-magnitude-WISE sample from 68 per cent probability intervals of the marginalized probability distribution functions (PDFs) of each parameter, given by the accepted values after post-burn period in the MCMC chain.
Table 1. Colour excess tabulated as a function of Galactic longitude, latitude and distance module.

| l (deg) | b (deg) | $E(G - K_S)_{4.25}$ | $E(G - K_S)_{4.75}$ | $E(G - K_S)_{12.75}$ | $E(G - K_S)_{13.25}$ |
|---------|---------|----------------------|----------------------|----------------------|----------------------|
| $\sigma E(G - K_S)_{4.25}$ | $\sigma E(G - K_S)_{4.75}$ | $\sigma E(G - K_S)_{12.75}$ | $\sigma E(G - K_S)_{13.25}$ |
| $E(G_{BP} - G_{RP})_{4.25}$ | $E(G_{BP} - G_{RP})_{4.75}$ | $E(G_{BP} - G_{RP})_{12.75}$ | $E(G_{BP} - G_{RP})_{13.25}$ |
| $\sigma E(G_{BP} - G_{RP})_{4.25}$ | $\sigma E(G_{BP} - G_{RP})_{4.75}$ | $\sigma E(G_{BP} - G_{RP})_{12.75}$ | $\sigma E(G_{BP} - G_{RP})_{13.25}$ |
| $E(H - K_S)_{4.25}$ | $E(H - K_S)_{4.75}$ | $E(H - K_S)_{12.75}$ | $E(H - K_S)_{13.25}$ |
| $\sigma E(H - K_S)_{4.25}$ | $\sigma E(H - K_S)_{4.75}$ | $\sigma E(H - K_S)_{12.75}$ | $\sigma E(H - K_S)_{13.25}$ |

Notes. For each position, we provide the integrated colour excesses $E(G - K_S)$, $E(G_{BP} - G_{RP})$ and $E(H - K_S)$ and the associated errors as functions of distance module $\mu$ ranging from 4.25 to 13.25 mag with a step of 0.5 mag.

**Figure 6.** 3D colour excess $E(G_{BP} - G_{RP})$ maps of the Galactic plane, integrated to distances, from bottom to second to top, 501, 1000, 1585, 1995, 2512 and 3981 pc. The top panel shows the 2D map from Planck Collaboration et al. (2014) for comparison.

**Figure 7.** 3D maps of differential colour excess $\delta E(G_{BP} - G_{RP})$ of the Galactic plane, in units of mag kpc$^{-1}$. The six panels, from bottom to top, refer to ranges of distance from the Sun, 0 - 501 pc, 501 - 1000 pc, 1000 - 1585 pc, 1585 - 1995 pc, 1995 - 2512 pc and 2512 - 3981 pc, respectively.
agram for 28,271,752 stars with Gaia parallax errors smaller than 20 per cent. Values of absolute magnitude $M_G$ of the individual stars are estimated using the standard relation, $M_G = G - A_G - 5 \log d + 5$, where $d$ is distance estimate from [Bailer-Jones et al. (2018)] and $A_G$, the $G$-band line-of-sight extinction. As the Gaia $G$ band covers a wide range of wavelength, we calculate $A_G$ from colour excesses $E(G - K_S)$ and $E(H - K_S)$ using the near-IR extinction law of [Yuan et al. (2013)].

$$A_G = E(G - K_S) + 1.987E(H - K_S)$$ (3)

Fig. 5 presents a nice Hertzsprung-Russell diagram (HRD) very similar to the HRD presented in Fig. 5 of [Gaia Collaboration et al. (2018a)], constructed using low-extinction stars. The main sequence is quite sharp, the red clump really a clump and the red giant branch clearly visible. The well-defined HRD presented in Fig. 5 suggests the robustness of colour excess values derived in the current work.

We plot in Fig. 5 2D maps of colour excess $E(G_{\text{BP}} - G_{\text{RP}})$ in the Galactic plane, integrated respectively to selected distances, 501, 1000, 1585, 1995, 2512 and 3981 pc from the Sun. In general, the colour excess increases with distance for all pixels, but the growth rate varies from pixel to pixel, showing various structures. At close distances, we see the local dust clouds that extend to the high latitudes. At large distance, we begin to see the tilt of dust lane in the Galactic disk. Also plotted in the Figure for comparison is the Planck 2D colour excess map deduced from the dust far-IR thermal emission [Planck Collaboration et al. 2014]. The Planck map, representing the colour excess integrated along the line-of-sight to infinite, is comparable with ours integrated to 4 kpc in the direction of outer disk ($150^\circ < l < 250^\circ$). But in the direction toward the Galaxy centre, the Planck map yields systematically much larger colour excess values than ours, suggesting that there are still large amounts of dust in that direction beyond 4 kpc. Nevertheless, both maps show very similar dust features.

To highlight dust features in different distance slices, we plot in Fig. 7 the distribution of the differential colour excess $\delta E(G_{\text{BP}} - G_{\text{RP}})$ (in units of mag kpc$^{-1}$) produced by local dust grains in distance slices 0 - 501 pc, 501 - 1000 pc, 1000 - 1585 pc, 1585 - 1995 pc, 1995 - 2512 pc and 2512 - 3981 pc, respectively. Only the map of $\delta E(G_{\text{RP}} - G_{\text{BP}})$ is shown, as the maps of $\delta E(G - K_S)$ and $\delta E(H - K_S)$ are very similar. The Figure shows fine structures of dust distribution at various distances bins. The local dust clouds are clearly visible in the two close distance slices ($d < 1000$ pc). For example, in the closest distance bin, $0 < d < 501$ pc, one sees the Ophiuchus, Aquila Rift, and Hercules in the inner disk, Polaris Flare, Cepheus Flare and Perseus-Taurus-Auriga complex in the direction of the Galactic anti-centre. Beyond this ($501 < d < 1000$ pc), one sees the Monoceros and Orion clouds in the Galactic anti-centre region and the Gum Nebula at $l \sim 260^\circ$. Beyond 1000 pc, one sees mainly the dust features in the Galactic thin disc, concentrated in a narrow range of latitude ($|b| < 5^\circ$), and the warp is clearly visible. Following [Marshall et al. (2006)], we fit the dust warp by the equation,

$$z_{\text{warp}} = \gamma(R - R_0) \cos(\theta - \theta_0),$$ (4)

where $z_{\text{warp}}$ is the vertical distance between the mid-plane of the dust disk and the plane defined by $b = 0^\circ$, $\gamma$ the slope of the amplitude, $R_0$ the Galactocentric radius where the warp starts and $\theta_0$ the node angle. Based on the distribution of dust grains between distances from 2 to 4 kpc (Fig. 8), we find the values of position $z$ of maximum colour excess in the individual $(R, b)$ bins. Parameters $\gamma$, $R_0$ and $\theta_0$ are then derived by fitting the results. The best-fit values are $\gamma = 0.19$, $R_0 = 7.6$ kpc and $\theta_0 = 94^\circ$. The fit is shown in Fig. 8 with the result of [Marshall et al. (2006)] over-plotted for comparison. Our fit is in good agreement with that of Marshall et al. (2006) in the common region that is also covered by the dust map of Marshall et al. ($-100^\circ < l < 100^\circ$). However, in other regions such as $170^\circ < l < 250^\circ$, there is significant deviation between the warp model of Marshall et al. and ours.

In Fig. 7 we show the differential colour-excess $\delta E(G_{\text{BP}} - G_{\text{RP}})$ (in units of mag kpc$^{-1}$) in the Galactic disc plane of $|b| < 0.1^\circ$. The Sun is located at the center of the plot at $X = -8$ kpc and $Y = 0$ kpc. The overall morphology of the dust structure inside 2 kpc matches well with those of Fig. 1 in [Lallement et al. 2014] and of Fig. 7 in [Green et al. 2018]. At distances beyond 4 kpc, not a lot of dust is seen $|\delta E(G_{\text{BP}} - G_{\text{RP}})| < 1$ mag kpc$^{-1}$. This is mainly due to the selection bias of the map. Distant stars suffering from larger dust extinction have larger distance uncertainties compared to those suffering from smaller dust extinction. Those highly red- dined distant stars are therefore discriminated against when constructing the 3D colour excess maps. One can easily locate the various dust clouds, such as the Aquila Rift complex, Cygnus Rift, and the Orion clouds in the Figure. On large scales, those clouds spatially coincide with the Galactic spiral arms delineated by log-spiral fits to the high-mass star forming regions of [Reid et al. 2014]. The Sagittarius, Local and Perseus arms appear clearly on our map. The Scutum and Outer arms locate respectively near the right and left
Figure 9. Distribution of local dust in the Galactic plane ($|b| < 0.1\degree$). The Sun, assumed to be at 8.0 kpc from the Galactic center, is located at the centre of the plot. Brown circles are placed every 2 kpc from the Sun. The directions of $l = 0\degree, 90\degree, 180\degree$ and $270\degree$ are also marked in the plot. The yellow solid and dashed curved lines denote the center and $\pm 1\sigma$ widths of spiral arm models of, from left to right, the Outer, Perseus, Local, Sagittarius and Scutum arms, taken from Reid et al. (2014). The red lines are the same model of Sagittarius but have Galactocentric azimuth $\beta$. The edges of the map and are not readily identifiable. The Perseus arm can be traced by several discrete dust clouds located at $\sim 2$ kpc from the Sun in the outer disk. The Sagittarius arm can be traced by clouds located at $R$ about 6 - 7 kpc. In addition, the clouds in directions between $l \sim 310\degree$ and $360\degree$ have extended the Sagittarius in the fourth quadrant, which is consistent with the most recently work of Xu et al. (2018), who have studied the spiral structure in the solar neighborhood using a sample of over 5000 O-B2 stars. The Local arm can be traced by several discrete clouds at $R$ about 7 - 8 kpc. For regions of $Y < 0$ kpc, the Local arm seems to split into two parts. One goes through the dust clouds at $l$ between $190\degree$ and $250\degree$ and distance $d$ from 1 to 2 kpc. It is connecting the Perseus arm. The other one goes through the dust clouds at $l$ between $270\degree$ and $290\degree$ and distance $d$ from 1 to 3 kpc. It is connecting the Sagittarius arm. In addition, in the first and second quadrants, there are also several clouds connecting the Perseus and the Sagittarius arms. For example, the clouds at $l$ between $30\degree$ and $60\degree$ is connecting the Sagittarius arm and the clouds at $l$ between $145\degree$ and $160\degree$ the Perseus arm. The boundaries between the Local arm and other two arms are not clearly defined. The morphology is quite consistent with the “bridges” and “spurs” found by previous works using other tracers (e.g. Lallement et al. 2018 and Xu et al. 2018).

5 DISCUSSION

5.1 Extinction coefficients for the Gaia photometry

We provide 3D maps for $E(G - K_S)$, $E(G_{RP} - G_{BP})$ and $E(H - K_S)$ separately. In order to derive line-of-sight extinction or colour excess in other passbands, one must assume certain relation between the extinction or colour excess in different bands, namely the ex-
Table 2. Colour excess ratios and extinction coefficients for Gaia photometry

| $E(G - K_S)$ | $E(G_{RP} - J)$ | $E(H - K_S)$ | $R_G$ | $R_{G_{BP}}$ | $R_{G_{RP}}$ |
|---------------|----------------|--------------|-------|--------------|--------------|
| $1.61 \pm 0.02$ | $0.81 \pm 0.02$ | $0.13 \pm 0.01$ | $2.50 \pm 0.03$ | $3.24 \pm 0.02$ | $1.91 \pm 0.02$ |

![Figure 10](image.png)

**Figure 10.** Colour excesses $E(G - K_S)$, $E(G_{RP} - J)$ and $E(H - K_S)$ versus $E(G_{BP} - G_{RP})$. Red pluses and error bars represent medians and dispersions in bins of size 0.1 mag in abscissa. Blue lines are linear regressions passing through the origin to the red pluses.

3D Reddening maps in disk

5.2 Comparison with previous work

To examine the accuracy of colour excesses derived in the current work with a machine learning algorithm, we compare our results with measurements from a number of independent studies, including:

(i) Values of $r$-band extinction of over 13 million stars in the Galactic anti-centre from Chen et al. (2014), obtained by SED fitting to photometric measurements from the optical to the near-IR ($g$, $r$, and $i$ from the XSTPS-GAC, $J$, $H$, and $K_S$ from the 2MASS and $W_1$ and $W_2$ from the WISE).

(ii) Values of $K_S$-band extinction of over 0.1 million stars observed by the APOGEE survey from Wang et al. (2016), derived with a Bayesian approach by taking into account spectroscopic constraints from the APOGEE stellar parameters and photometric constraints from the 2MASS, as well as prior knowledge of the Milky Way.

(iii) Values of monochromatic extinction at $5495\AA$, $A_0$, of over 38 million stars in the Northern Galactic plane observed by IPHAS from Sale et al. (2014), derived with a method based on a hierarchical Bayesian model using the IPHAS photometry.

Fig. 11 compares our results with those from previous studies mentioned above. We cross-match our Gaia/2MASS/WISE sample with those of Chen et al. (2014), Wang et al. (2016) and Sale et al. (2014) with a search radius of 1 arcsec. For consistency, all the colour excess or extinction values have been re-scaled to $E(B-V)$ using appropriate extinction laws. We convert our current estimate of colour excess $E(G_{BP} - G_{RP})$ to $E(B-V)$ using the extinction coefficients presented in Sect. 5.1, and this yields $E(B-V) = 0.75 E(G_{BP} - G_{RP})$. The values of $A_0$ in Chen et al. (2014) are converted to those of $E(B-V)$ using the extinction law of Yuan et al. (2013), which gives $E(B-V) = 0.43 A_0$. For $A_{G_{BP}}$ of Wang et al. (2016), we use the Cardelli et al. (1989) extinction law and have $E(B-V) = 2.77 A_{G_{BP}}$. For $A_0$ from Sale et al. (2014), we use the relation, $A_0 = 1.003 A_{G_{BP}}$, from Sale et al. (2014) and assume $R_V = 3.1$. This gives $E(B-V) = 0.32 A_0$. Fig. 11 shows good agreement for all comparisons. Our measurements, compared to those of Chen et al. (2014), have an average difference of only $-0.001$ mag, along with an rms scatter
Figure 11. Comparisons of $E(B-V)$ values derived in the current work and those deduced by Chen et al. (2014) (left), by Wang et al. (2016) (middle) and by Sale et al. (2014) (right). Red pluses and the error bars represent medians and standard deviations in the individual bins. Red straight lines denoting complete equality are also overplotted to guide the eyes. Means and standard deviations of the differences (ours minus those from the literature), are marked in the individual panels.

Figure 12. Comparison of our current 3D map with that of Green et al. (2018). The panels show the differences of $E(B-V)$ (ours minus those of Green et al.), integrated to, from bottom to top, 501, 1000, 1585, 1995, 2512 and 3981 pc. Means and standard deviations of the differences are marked in the individual panels.

of 0.11 mag. The typical $r$-band extinction uncertainties for stars of photometric errors ~ 0.05 mag in Chen et al. (2014) are about 0.16 mag, corresponding to ~ 0.07 mag in $E(B-V)$. We can conclude that our current measurements and those of Chen et al. contribute equally to the aforementioned dispersion, suggesting that the current machine-learning method achieves an accuracy similar to the traditional SED fitting method. There are a few stars have large $E(B-V)$ values from the current work (~ 1 – 1.5 mag), but small ones in Chen et al. (2014). Those stars are found to be very cool stars (spectral type M5-M6) that have intrinsic colour $(g-i)_0$ ~ 3 mag as found by Chen et al. (2014). Due to the lack of very cool stars in our current spectroscopic training sample, we are not able to recover the colour excesses of those stars correctly. However, those stars contribute only 0.1 per cent of the entire sample. They do not have a significant effect on the 3D colour excess maps presented here.

Our results are well correlated with those of Wang et al. (2016). The mean difference is only 0.015 mag and the dispersion, 0.078 mag, is the smallest among the comparisons. This is probably because the extinction values of Wang et al. (2016) are derived from spectroscopic data that have the smallest uncertainties. The
dispersion is mainly contributed by the uncertainties of our work, which is $\sim 0.07$ mag in $E(B-V)$.

Compared to Sale et al. (2014), our results are systematically larger, with a mean difference of $-0.076$ mag, and a relatively large dispersion of 0.159 mag. The offset and dispersion may partly be caused by the different data sets and methods used, as well as by the variations of extinction coefficients in the Galactic plane.

Finally we compare our 3D colour excess map to the most recent 3D map of Green et al. (2018). The latter covers the whole Northern sky ($\delta \geq -30^\circ$) out to a distance of several kpcs. The two maps overlap in the Galactic longitude range $0^\circ < l < 120^\circ$. They have similar distance resolution. However, the map of Green et al. uses HEALPix pixelization scheme (Gorski et al. 2005). Depending on the regions, the pixel scale varies from 3.4 to 55 arcmin. For each distance bin, we convert their map to the angular resolution of our map (6 to 6 arcmin) by 2D linear interpolations. We convert the maps of Green et al. to $E(B-V)$ using their relation, $E(B-V) = 0.996 \times (\text{Bayestar17})$. Fig. 12 shows the differences of cumulative values of $E(B-V)$ between our map and that of Green et al., integrated to distances 501, 1000, 1585, 1995, 2512 and 3981 pc, respectively. Overall, the two results agree well, with small average differences of $\sim -0.04$ mag, and small dispersions between 0.09 and 0.27 mag.

6 SUMMARY

The Gaia DR2 has provided us a great opportunity to study the 3D distribution of dust grains in the Galactic disk. By combing the high quality optical photometry from the Gaia DR2 and those from the 2MASS and WISE in the IR, we have simultaneously derived values of colour excess $E(G-K_s)$, $E(G_{BP} - G_{RP})$ and $E(H-K_s)$ for over 44 million stars, using the Random Forest regression, a machine learning algorithm. In doing so, we have built an empirical training sample of stars selected from several large-scale spectroscopic surveys, including the LAMOST, SEGUE and APOGEE. We derive values of colour excess $E(G-K_s)$, $E(G_{BP} - G_{RP})$ and $E(H-K_s)$ for over 3 million stars in the spectroscopic sample with the star-pair technique. A comparison with results in the literature shows good agreement and that our current results have an accuracy comparable to those derived from the SED fitting method or Bayesian approaches, with typical uncertainties of about 0.07 mag in $E(B-V)$. However, the machine learning technique adopted in the current work is much faster than these traditional methods. Values of $E(G-K_s)$, $E(G_{BP} - G_{RP})$ and $E(H-K_s)$ for the 44 million stars are available upon request by email.

By combining our colour excess values and the distances of Bailer-Jones et al. (2018) estimated from the Gaia parallaxes, we have constructed high-quality 3D colour excess map for the entire Galactic plane ($0^\circ < l < 120^\circ$ and $|b| < 10^\circ$). The map covers over 7000 deg$^2$ at an angular resolution of 6 arcmin, out to a distance of about 5 kpc from the Sun. The newly built map is in good agreement with those in the literature. The map reveals large-scale distribution of dust grains in the Galactic plane such as warping and spiral arms, as well as many fine features. The map will be public available, and should be quite useful for follow up studies of the Milky Way disk.

Finally, using the spectroscopic sample we have calculated colour excess ratios and the extinction coefficients for the Gaia DR2 photometric bands. Empirically, we have $R_G = 2.50 \pm 0.03$, $R_{G_{BP}} = 3.24 \pm 0.02$ and $R_{G_{RP}} = 1.91 \pm 0.02$. The extinction coefficients can be used to convert our colour excesses to line-of-sight extinction in the Gaia DR2 bands (Eqs. 3, 5 and 6).

ACKNOWLEDGEMENTS

This work is partially supported by National Key Basic Research Program of China 2014CB845700 and National Natural Science Foundation of China U1531244. HBY is supported by NSFC grant No. 11603002 and Beijing Normal University grant No. 310232102. This research made use of the cross-match service provided by CDS, Strasbourg.

This work has made use of data products from the Guoshoujing Telescope (the Large Sky Area Multi-Object Fibre Spectroscopic Telescope, LAMOST). LAMOST is a National Major Scientific Project built by the Chinese Academy of Sciences. Funding for the project has been provided by the National Development and Reform Commission. LAMOST is operated and managed by the National Astronomical Observatories, Chinese Academy of Sciences.

Funding for the Sloan Digital Sky Survey (SDSS) has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Aeronautics and Space Administration, the National Science Foundation, the U.S. Department of Energy, the Japanese Monbukagakusho, and the Max Planck Society. The SDSS Web site is http://www.sdss.org/.

The SDSS is managed by the Astrophysical Research Consortium (ARC) for the Participating Institutions. The Participating Institutions are The University of Chicago, Fermilab, the Institute for Advanced Study, the Japan Participation Group, The Johns Hopkins University, Los Alamos National Laboratory, the Max-Planck-Institute for Astronomy (MPIA), the Max-Planck-Institute for Astrophysics (MPA), New Mexico State University, University of Pittsburgh, Princeton University, the United States Naval Observatory, and the University of Washington.

This work presents results from the European Space Agency (ESA) space mission Gaia. Gaia data are being processed by the Gaia Data Processing and Analysis Consortium (DPAC). Funding for the DPAC is provided by national institutions, in particular the institutions participating in the Gaia MultiLateral Agreement (MLA). The Gaia mission website is https://www.cosmos.esa.int/gaia. The Gaia archive website is https://archives.esac.esa.int/gaia.

This publication makes use of data products from the Two Micron All Sky Survey, which is a joint project of the University of Massachusetts and the Infrared Processing and Analysis Center/California Institute of Technology, funded by the National Aeronautics and Space Administration and the National Science Foundation.

This publication makes use of data products from the Wide-field Infrared Survey Explorer, which is a joint project of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration.

REFERENCES

Andrae, R. et al. 2018, ArXiv e-prints
Bailer-Jones, C. A. L., Rybizki, J., Fouesneau, M., Mantelet, G., & Andrae, R. 2018, ArXiv e-prints
Berry, M., et al. 2012, ApJ, 757, 166
