



## UvA-DARE (Digital Academic Repository)

### Maximum Entropy Estimation of the Galactic Bulge Morphology via the VVV Red Clump

Coleman, B.; Paterson, D.; Gordon, C.; Macias, O.; Ploeg, H.

**DOI**

[10.1093/mnras/staa1281](https://doi.org/10.1093/mnras/staa1281)

**Publication date**

2020

**Document Version**

Final published version

**Published in**

Monthly Notices of the Royal Astronomical Society

[Link to publication](#)

**Citation for published version (APA):**

Coleman, B., Paterson, D., Gordon, C., Macias, O., & Ploeg, H. (2020). Maximum Entropy Estimation of the Galactic Bulge Morphology via the VVV Red Clump. *Monthly Notices of the Royal Astronomical Society*, 495(3), 3350–3372. <https://doi.org/10.1093/mnras/staa1281>

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

*UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)*

# Maximum entropy estimation of the Galactic bulge morphology via the VVV Red Clump

B. Coleman,<sup>1★</sup> D. Paterson,<sup>1★</sup> C. Gordon,<sup>1★</sup> O. Macias<sup>2,3</sup> and H. Ploeg<sup>1</sup>

<sup>1</sup>*School of Physical and Chemical Sciences, University of Canterbury, Christchurch, 8140, New Zealand*

<sup>2</sup>*Kavli Institute for the Physics and Mathematics of the Universe (WPI), University of Tokyo, Kashiwa, Chiba 277-8583, Japan*

<sup>3</sup>*GRAPPA Institute, University of Amsterdam, Amsterdam NL-1098 XH, the Netherlands*

Accepted 2020 April 14. Received 2020 April 14; in original form 2019 November 19

## ABSTRACT

The abundance and narrow magnitude dispersion of Red Clump (RC) stars make them a popular candidate for mapping the morphology of the bulge region of the Milky Way. Using an estimate of the RC’s intrinsic luminosity function, we extracted the three-dimensional density distribution of the RC from deep photometric catalogues of the VISTA Variables in the Via Lactea (VVV) survey. We used maximum entropy-based deconvolution to extract the spatial distribution of the bulge from  $K_s$ -band star counts. We obtained our extrapolated non-parametric model of the bulge over the inner  $40^\circ \times 40^\circ$  region of the Galactic centre. Our reconstruction also naturally matches on to a parametric fit to the bulge outside the VVV region and inpaints overcrowded and high extinction regions. We found a range of bulge properties consistent with other recent investigations based on the VVV data. In particular, we estimated the bulge mass to be in the range  $[1.3, 1.7] \times 10^{10} M_\odot$ , the X-component to be between 18 per cent and 25 per cent of the bulge mass, and the bulge angle with respect to the Sun–Galactic centre line to be between  $18^\circ$  and  $32^\circ$ . Studies of the *Fermi* Large Area Telescope (LAT) gamma-ray Galactic centre excess suggest that the excess may be traced by Galactic bulge distributed sources. We applied our deconvolved density in a template fitting analysis of this *Fermi*–LAT GeV excess and found an improvement in the fit compared to previous parametric-based templates.

**Key words:** Galaxy: bulge – Galaxy: centre – Galaxy: structure – gamma-rays: galaxies – infrared: galaxies.

## 1 INTRODUCTION

Since the advent of near-infrared surveys, we have begun to view the Milky Way centre behind dust reddening obscuration (Bland-Hawthorn & Gerhard 2016). Through the *COBE*/DIRBE survey the presence of a Galactic bulge/bar was established (Binney et al. 1991; Weiland et al. 1994). Models fitted to the DIRBE data typically found a triaxial bar with its major axis rotated at an angle in the range between  $10^\circ$  and  $45^\circ$  to the Sun–Galactic centre line (Dwek et al. 1995; Bissantz et al. 1997; Freudenreich 1998; Bissantz & Gerhard 2002). Subsequently, surveys such as OGLE, 2MASS, and VVV have provided us with increasingly sensitive observations of the stellar distribution in the Galactic centre. The main observational data set of interest to this study is the VISTA Variables in the Via Lactea (VVV) survey (Minichiello et al. 2010), in particular, the stars occupying the Red Clump (RC) region of the colour–magnitude diagram (CMD).

The narrow dispersion of the RC (Chan & Bovy 2019; Hall et al. 2019) combined with the photometric star catalogues in the near-infrared regime enables estimates of the distance to stars based on their apparent magnitudes, though this comes with some caveats (see Girardi 2016). The RC has been the focus of several studies characterizing the three-dimensional density structure of the Galactic bulge. Many studies have exploited this property of the RC to fit triaxial models to the bulge (Stanek et al. 1997; Rattenbury et al. 2007; Cao et al. 2013; Simion et al. 2017). Non-parametric methods have also been used in viewing the RC distribution, initially with an assumed constant intrinsic RC magnitude Saito et al. (2011), then later accounting for its dispersion in works such as Wegg & Gerhard (2013; from here on WG13). The Galactic RC magnitude distribution was found to produce a double peak by Nataf et al. (2010) using OGLE-III data and McWilliam & Zoccali (2010) using 2MASS. This has been interpreted as being the result of an X-shaped structure that is characteristic of the boxy/peanut-like morphology seen in extragalactic studies of barred galaxies (e.g. Laurikainen et al. 2014; Ciambur & Graham 2016) and N-body simulations (e.g. Gardner et al. 2014). However, some works have disputed the

\* Email: [bjc174@uclive.ac.nz](mailto:bjc174@uclive.ac.nz) (BC); [dnp16@uclive.ac.nz](mailto:dnp16@uclive.ac.nz) (DP); [chris.gordon@canterbury.ac.nz](mailto:chris.gordon@canterbury.ac.nz) (CG)

physical separation of the RC, positing population effects in the luminosity function account for the photometric split in the RC peaks (López-Corredoira 2016; Joo, Lee & Chung 2017; Lee et al. 2018). However, the cross-matching of VVV RC stars with *Gaia* in Sanders et al. (2019) and Clarke et al. (2019) found proper motions of the VVV RC stars that indicate a spatial separation in the split RC peak.

Triaxial symmetry has often been assumed in morphological studies, such as the analytic models used by Simion et al. (2017; from here on S17). The models used by S17 represent only a subset of the broader class of triaxial bulge models (Dwek et al. 1995). Triaxial symmetry has also been enforced for non-parametric studies such as that of WG13 (hereafter, eight-fold symmetry for this context) to overcome gaps in the data and improve signal to noise when producing their final model. In this article, we use maximum entropy and smoothness regularization (Jaynes 1957; Storm, Weniger & Calore 2017) to help estimate the bulge morphology. This allows us to make fewer symmetry assumptions and it also provides a natural way of inpainting masked regions and matching on to parametric fits outside the region of interest covered by the data.

Paterson et al. (2019), hereafter P19, modelled the VVV data without any symmetry requirements, which exposed features adjacent to the bulge. In this paper, we made a mirror symmetry assumption about the Galactic plane to enable a constrained extension of the non-parametric RC bulge model to the inner  $40^\circ \times 40^\circ$  region, which is important for our intended application described next. In addition, we absorb into our background known features outside the bulge that may otherwise be picked up by the deconvolution. We also performed systematic checks of this bulge analysis pipeline.

Knowledge of the Galactic bulge density distribution can provide useful information when modelling the *Fermi* Galactic Centre Excess (GCE; Ackermann et al. 2017) observed in the *Fermi* Large Area Telescope (LAT) data (Atwood et al. 2009). The GCE was identified early on as a possible dark matter self-annihilation signal (Goodenough & Hooper 2009; Abazajian & Kaplinghat 2012; Gordon & Macias 2013) due to its apparent diffuse spherical nature, and soon after as possibly due to a millisecond pulsar (MSP) population in the Galactic centre (Abazajian 2011). More recently, the non-spherical nature of the GCE came to the foreground in importance, interpreted as strongly in favour of gamma-ray emission tracing a Galactic bulge morphology rather than the more spherically distributed dark matter self-annihilation case (Bartels et al. 2018a; Macias et al. 2018). However, there is some debate about whether the resolved MSPs are consistent with the needed bulge population (Cholis, Hooper & Linden 2015; Hooper & Mohlabeng 2016; Ploeg et al. 2017; Bartels, Edwards & Weniger 2018b). In this work, we employ the same template fitting procedure described in Macias et al. (2019) and compare our non-parametrically deconvolved bulge model to the bulge models of past works.

Our article is arranged as follows: In Section 2, we provide an overview of our VVV data set preparation and our non-parametric deconvolution method for inverting stellar statistics to recover the three-dimensional RC density distribution. We also motivate our choice of parametric model as a prior distribution and as a simple geometric model of the bulge with a peanut/X-shape morphology. In Section 3, we test our deconvolution pipeline against simulations. We present our results and discuss them in Sections 4 and 5. In Section 6, we estimate various properties of the bulge and analyse the impact of our non-parametric model on the GCE template fitting.

## 2 METHOD

### 2.1 VVV data preparation

This paper employs the MW–BULGE–PSFPHOT ultra deep photometric catalogue of Surot et al. (2019), corrected for known calibration issues discussed by Hajdu et al. (2019) through cross-matching sources with 2MASS (Skrutskie et al. 2006). Note that the  $K_s$  and  $J$  apparent magnitudes in the catalogue have been extinction corrected (Surot et al. 2019). A standard colour cut of  $0.4 < J - K_s < 1.0$  was applied to restrict sources to the predominantly RC region of the CMD and exclude any bluer foreground stars. We bin the extinction-corrected stellar catalogue with resolution ( $0.05 \text{ mag} \times 0.2^\circ \times 0.2^\circ$ ) in magnitude ( $K_s$ ), Galactic latitude ( $l$ ), and Galactic longitude ( $b$ ). The grid was bounded by the ranges:  $11 < K_s < 15$ ,  $-10^\circ < l < 10^\circ$ , and  $-10^\circ < b < 5^\circ$ . This binned data set was corrected for completeness by dividing by the mean of the completeness values for stars in each voxel, utilizing the effective completeness value assigned to each star in the catalogue of Surot et al. (2019). Due to crowding and known photometric error effects, we argue for a mask of our gridded line-of-sight data based on the mean  $K_s$  uncertainty ( $\sigma$ ) of the binned stars in the catalogue rather than a colour excess based mask. A boundary of  $\sigma = 0.06$  was chosen. This value causes the new mask to approximately match the  $E(J - K) = 0.9$  boundary in the less crowded regions of  $|l| > 5^\circ$ . A systematic check of this method is investigated in Section 5. The data preparation is discussed in further detail in P19.

### 2.2 Luminosity function

We utilize the semi-analytic luminosity function constructed in P19 using the PARSEC + COLIBRI isochrone sets of Marigo et al. (2017) and a Chabrier lognormal Initial Mass Function (IMF; Chabrier (2003)). Using the evolutionary stage flags in the isochrones, the semi-analytic luminosity function is divided into three components: a red giant branch, an RC, and an asymptotic giant branch. An exponential function was fitted to the red giant branch, excluding the absolute magnitude range  $-1.75 < M_{K_s} < -0.75$ , to extract the red giant branch bump (RGBB) component. We assumed a bulge age of 10 Gyr and a metal content normally distributed with solar mean metallicity  $\mu_{[\text{Fe}/\text{H}]} = 0.0$  and dispersion  $\sigma_{[\text{Fe}/\text{H}]} = 0.4$  (Zoccali et al. 2008).

### 2.3 Deconvolution procedure

The RC + RGBB stellar density ( $\rho$ ) of the Galactic bulge can be reconstructed by inverting the equation of stellar statistics:

$$N(K_s, l, b) = B(K_s, l, b) + \Delta\Omega \Delta K_s \int_{4 \text{ kpc}}^{13 \text{ kpc}} \rho(s, l, b) \Phi(K_s - 5 \log s - 10) s^2 ds, \quad (1)$$

where  $N$  is the predicted number of stars in a voxel centred at  $(K_s, l, b)$  and  $B$  is the number of smooth background stars in the voxel that are neither RC or RGBB stars. The  $\Delta\Omega$  denotes the solid angle subtended by the line of sight,  $\Delta K_s$  is the width of the  $K_s$  magnitude bin, and  $s$  (measured in kpc) is the distance from the Sun. The luminosity function  $\Phi$  is the sum of the bulge RC and the bulge RGBB luminosity function components. Note that as the RGBB is a much smaller component than the RC, we sometimes refer to our obtained density in terms of the RC only, but more precisely it does contain both the RC and RGBB. As the Galactic bulge density tends to become negligible beyond several kpc, we

only integrate the range  $4 \text{ kpc} \leq s \leq 13 \text{ kpc}$  when computing the bulge contribution in modelling stellar counts.

As in P19, our analysis uses penalized likelihoods with penalties which come in two general forms: the first is maximum entropy regularization, inspired by its application in Storm et al. (2017), which is defined for a three-dimensional grid of numbers  $q$ :

$$-2 \ln \mathcal{L}_{\text{MEM}} = 2\lambda \sum_{i,j,k} (1 - q_{i,j,k} + q_{i,j,k} \ln q_{i,j,k}), \quad (2)$$

where  $i, j$ , and  $k$  are the grid points for  $K_s$ ,  $b$ , and  $l$ , respectively. The maximum entropy regularization has a minimum at  $q_i = 1$ , so for our application we will use a parametrization where  $q$  is the ratio between a modelled quantity of interest and a smooth prior estimation of the quantity. As shown in appendix A of P19, the prior relative standard deviation of the reconstructed density from the prior density is of order  $1/\sqrt{\lambda}$ . So, the larger the value of  $\lambda$  chosen, the smaller the prior uncertainty assumed and so the more regularization of the solution is applied.

The second form of likelihood penalty we use is the  $\ell_2$ -norm regularization of the second derivative of the logarithm of some quantity (also inspired by its application in Storm et al. 2017). For a three-dimensional grid of numbers,  $F$ , which varies over one dimension, we use the second-order central difference equation approximation of curvature:

$$-2 \ln \mathcal{L}_{\text{smooth}} = \eta \sum_i (\ln F_{i-1} + \ln F_{i+1} - 2 \ln F_i)^2. \quad (3)$$

This penalty has a minimum when  $F$  is the exponential of a linear function of grid coordinates. As shown in appendix A of P19, the prior relative standard deviation from an exponential of a linear function is approximately  $1/\sqrt{6\eta}$ . So, the larger the value chosen for  $\eta$ , the more smoothness regularization is applied.

## 2.4 Background

We modelled the background ( $B$ ) non-parametrically as a free parameter for each  $(K_s, l, b)$  voxel. Without regularization, we would have a Poisson likelihood for data  $n_{i,j,k}$  with expected counts  $B_{i,j,k}$  where  $i, j, k$  are the grid points for  $(K_s, l, b)$ , respectively. With maximum entropy and smoothness regularization, we have the following formula for the natural log of the penalized likelihood ( $\mathcal{L}$ ):

$$\begin{aligned} \ln \mathcal{L} = & \sum_{(i,j,k) \in \{K_s, l, b\}} \left[ n_{i,j,k} \ln B_{i,j,k} - B_{i,j,k} \right. \\ & - \lambda (1 - q_{i,j,k} + q_{i,j,k} \ln q_{i,j,k}) \\ & - \eta_{K_s} (\ln B_{i-1,j,k} + \ln B_{i+1,j,k} - 2 \ln B_{i,j,k})^2 / 2 \\ & - \eta_l (\ln B_{i,j-1,k} + \ln B_{i,j+1,k} - 2 \ln B_{i,j,k})^2 / 2 \\ & \left. - \eta_b (\ln B_{i,j,k-1} + \ln B_{i,j,k+1} - 2 \ln B_{i,j,k})^2 / 2 \right], \quad (4) \end{aligned}$$

where  $q$  is the ratio between our background model and a smooth prior estimation of the background:

$$q \equiv \frac{B}{B_{\text{prior}}}. \quad (5)$$

The first line on the RHS of equation (4) is from the usual Poisson likelihood distribution. The second line is an entropy regularization of the form of equation (2) and the third, fourth, and fifth lines are smoothness regularizations of the form given in equation (3) for  $K_s$ ,  $l$ , and  $b$ , respectively. The regularization parameter values we used are listed in Table 1, and we discuss their choice in Section 3. We

**Table 1.** Regularization parameters used when fitting to the simulated population and the VVV sample.

	$\lambda$	$\eta_s$	$\eta_l$	$\eta_b$
Background	1.0	1000.0	100.0	100.0
Three-dimensional deconvolution	0.01	400.0	200.0	100.0

maximized equation (4) using the magnitude ranges  $11 < K_s < 11.7$  and  $14.3 < K_s < 15$ , see Section 3 for more details. This means the behaviour in  $11.7 \leq K_s \leq 14.3$  is determined entirely by the prior, maximum entropy, and smoothness regularization.

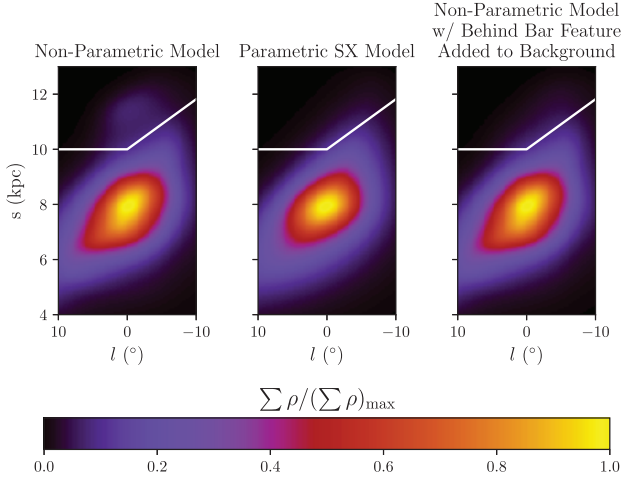
The background is mainly composed of red giant stars in the bulge and foreground disc stars, so for the prior background ( $B_{\text{prior}}$ ) we used the S-model+discs fitted by S17 with the RC and RGBB components subtracted. Only the asymptotic giant branch and red giant branch (excluding the RGBB) components of the semi-analytic luminosity function are used for the bulge component in determining the background. Included in the S-model+discs are thin and thick disc components of the Besançon galaxy model of Robin et al. (2003), where we have used the updated thin disc parameters from Robin et al. (2012) and the updated thick disc parameters from Robin et al. (2014). The S-model+discs of S17 was fitted to aperture photometry of the VVV DR2 data in the range  $12 < K_s < 14$ , so the background was underestimated for some lines of sight. To compensate for this, we multiplied each pixel (line of sight) of the prior background by a constant, so that its mean matched the mean of our data in the range  $11 < K_s < 11.5 \text{ mag}$ .

Initial tests of our deconvolution method on the VVV data showed that our method was finding a feature in the density consistent with the structure behind the bar reported in Gonzalez et al. (2018) and P19. As we are trying to determine the bulge component, we decided to add this feature to our background, by first estimating our density using our maximum entropy background, then adding the star counts associated with any density significantly greater than our prior parametric density (see SX model of Section 2.6) to the maximum entropy background. We considered any density that was beyond the limits

$$\begin{aligned} s &> 10 \text{ kpc} & l &\geq 0^\circ \\ s &> (10 - 0.1818l) \text{ kpc} & l &< 0^\circ \end{aligned} \quad (6)$$

and at least  $2.6 \times 10^{-5} \text{ stars pc}^{-3} \text{ sr}^{-1}$  above the parametric model density to be part of the structure behind bar. In Fig. 1, we display the density summed over  $|b| < 10^\circ$ , where the feature behind the bar is visible in the model fitted using our maximum entropy method. The contribution of the feature behind the bar to the background is visible in the bottom panel of Fig. 2 as a bump in the fitted background at  $K_s \sim 13.8 \text{ mag}$ . When using the updated background, the feature behind the bar is no longer present in the density, as seen in the right-hand panel of Fig. 1.

Shown in the top panel of Fig. 2 is the fitted background for a  $1^\circ \times 1^\circ$  box around  $(l, b) = (0.9^\circ, -6.1^\circ)$ , where we can see that the fitted background is only slightly deviating from the prior background. In the bottom panel, the background fitted in a  $1^\circ \times 1^\circ$  box around  $(l, b) = (0.9^\circ, 3.1^\circ)$  fits the data well in the shaded regions. However, the background needs to deviate significantly from the prior background at  $K_s > 14.7 \text{ mag}$ , where the data may have residual extinction and completeness issues. In the unshaded region, apart from the added feature behind the bar, the background closely follows the shape of the prior solution. The background also smoothly trends back to passing through the data in the shaded regions.



**Figure 1.** Apparent structure behind the bar in the VVV data, visible in the left-hand panel, was added to the background of our model. We remove any density that is significantly greater than the fitted parametric model (middle panel) and at distances greater than indicated by the white line. In these figures, the density has been summed in the range  $|b| < 10^\circ$ .

## 2.5 Maximum entropy deconvolution

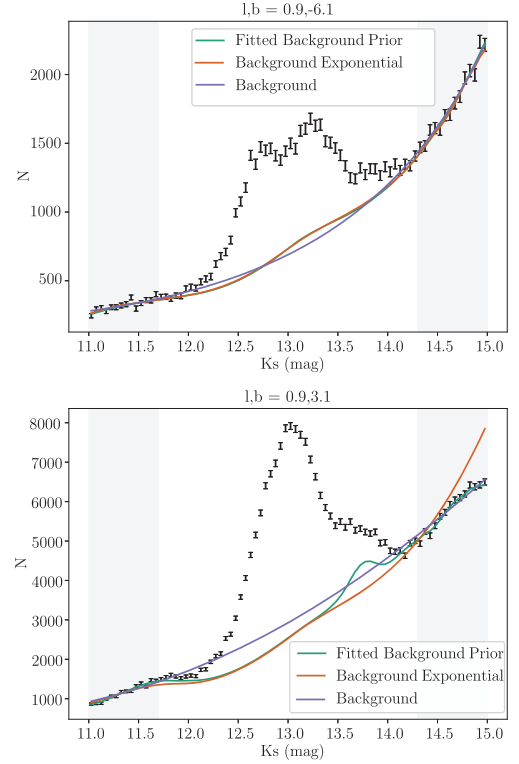
Our maximum entropy method provides a non-parametric estimate of the stellar density that predicts the binned star counts of a stellar catalogue. It maximizes the same  $\ln \mathcal{L}$  as equation (4), except that  $B$  is replaced with the total expected star counts ( $N$ ) and  $q$  is replaced by  $\kappa$ , which is the ratio between the bulge density model and a prior estimation of the density such as a parametric bulge model like that of Section 2.6:

$$\kappa \equiv \frac{\rho}{\rho_{\text{prior}}}. \quad (7)$$

Also, as we are estimating  $\rho$  on a grid of  $(s, l, b)$ , we need a separate sum for the regularization terms in contrast to equation (4) where we could use one sum as we estimated the background ( $B$ ) on a  $(K_s, l, b)$  grid. This gives

$$\begin{aligned} \ln \mathcal{L} = & \sum_{\{i,j,k\} \in \{K_s, l, b\}} (n_{i,j,k} \ln N_{i,j,k} - N_{i,j,k}) \\ & - \sum_{\{h,j,k\} \in \{s, l, b\}} \left[ \lambda (1 - \kappa_{h,j,k} + \kappa_{h,j,k} \ln \kappa_{h,j,k}) \right. \\ & + \eta_s (\ln \rho_{h-1,j,k} + \ln \rho_{h+1,j,k} - 2 \ln \rho_{h,j,k})^2 / 2 \\ & + \eta_l (\ln \rho_{h,j-1,k} + \ln \rho_{h,j+1,k} - 2 \ln \rho_{h,j,k})^2 / 2 \\ & \left. + \eta_b (\ln \rho_{h,j,k-1} + \ln \rho_{h,j,k+1} - 2 \ln \rho_{h,j,k})^2 / 2 \right]. \quad (8) \end{aligned}$$

Including the maximum entropy term in the likelihood discourages the modelled density from overfitting to regions of the data that are dominated by noise, where it will instead favour the smooth prior density. In practice, this is important in the regions where the background makes up a significant part of the model ( $K_s$  near 12.0 and 14.0), where the density should be tending towards zero. Addition of the smoothness terms discourages spurious high-frequency variations in the modelled density by minimizing curvature in the logarithm of the density. The smoothness term also has the added benefit of inpainting the density in lines of sight that have been masked out. For equation (8), we set  $\lambda = 0$  in masked regions so as they are only affected by the smoothness term and the values of the model at the edge of the mask.



**Figure 2.** Demonstration of the maximum entropy background fitting in two  $1^\circ \times 1^\circ$  regions. The background has been fitted in the grey-shaded regions using the maximum entropy method. The prior background was calculated using the S17 S-model+discs, which has been scaled to match the VVV observations between  $11.0 < K_s < 11.5$  mag. The bump in the bottom panel ‘fitted background’ at  $K_s \sim 13.8$  mag is from a feature behind the bar, see text in Section 2.4 for more details. The exponential background is described in Section A1.

## 2.6 Parametric model of the X-bulge

In light of the X-shape apparent in the eight-fold symmetrized WG13 style deconvolution, we consider a closed form parametric base case that allows for a X-bulge perturbation. We characterize its potential pathologies in fitting to data and simulations. The parametric density models fitted in this section are used as prior estimates for the density ( $\rho_{\text{prior}}$ ) with the maximum entropy deconvolution in Section 4. Our base case parametric-model fit was subsequently applied in a template fitting analysis of the *Fermi* GCE for comparison with our base non-parametric model result (see Section 6.2).

Triaxial models of the bulge have been investigated by Dwek et al. (1995), Athanassoula et al. (1990), and Freudenreich (1998). We selected the S-model, which proved successful for bulge modelling in Freudenreich (1998) and S17, as our base distribution. Inspired by the X-bulge parametric form of López-Corredoira (2016), we perturb the S-model with a X-like shape. We use a right-handed, Galactic Centre origin, Cartesian grid  $(X, Y, Z)$  aligned with the bulge axes of symmetry. The coordinates are chosen so that the X-axis lies along the major axis of the bulge and the Z-axis points towards the north Galactic pole. We refer to the arms of the X-bulge as the X-arms but these are not necessarily aligned with our X coordinate. The perturbation shape was freed in X and Y to accommodate non-circular X-arm shapes. The X-arms in this model part linearly along the bar-aligned Z-axis with gradient C. We also allowed the density of the X-arms to trail off as an exponential of a

power-law with exponent  $n$  rather than assuming an exponential or Gaussian distribution. We label this parametric form the SX model, with its components defined as follows:

$$\begin{aligned} \rho_{\text{SX}}(X, Y, Z) &= \rho_0 \text{sech}^2(r_1) \\ &\times [1 + A \times (\exp(-r_2^n) + \exp(-r_3^n))], \\ r_1^{c_{\parallel}} &= \left[ \left( \frac{|X|}{x_0} \right)^{c_{\perp}} + \left( \frac{|Y|}{y_0} \right)^{c_{\perp}} \right]^{\frac{c_{\parallel}}{c_{\perp}}} + \left( \frac{|Z|}{z_0} \right)^{c_{\parallel}}, \\ r_2 &= \left[ \left( \frac{|X - CZ|}{x_1} \right)^2 + \left( \frac{|Y|}{y_1} \right)^2 \right]^{\frac{1}{2}}, \\ r_3 &= \left[ \left( \frac{|X + CZ|}{x_1} \right)^2 + \left( \frac{|Y|}{y_1} \right)^2 \right]^{\frac{1}{2}} \end{aligned} \quad (9)$$

using a generalized ellipsoid distribution for the bulge and a simple ellipsoidal X-shape aligned with the bulge that tapers off with the same Z distribution. The parameters  $\mathfrak{P} = (\rho_0, A, n, x_0, y_0, z_0, c_{\perp}, c_{\parallel}, C, x_1, y_1)$  all need to be fit to the data. We used this parametric fit as a prior ( $\rho_{\text{prior}}$ ) for the maximum entropy non-parametric fit that did not enforce eight-fold symmetry. Equation (9) will provide us with an intermediary model between the S and non-parametric models in the *Fermi* template fitting analysis to gauge the correlation between an improved VVV fit and an improved gamma-ray distribution fit. If the GCE is tracing a bulge and there are no additional unexpected features, we might expect that a model that increasingly traces the morphological features of the bulge will improve the fit.

Investigating the parting rate of the X-arms by fitting a power law rather than the simple  $X \pm CZ$  form, we found the split was still well approximated as a linear function. To avoid convergence issues from excessive parameters, the RC split was left in the linear form.

A tapering of the density at cylindrical radii greater than a cut-off radius,  $R_c$ , was applied to the density distribution via  $\exp(-2(R - R_c)^2)$  with  $R_c$  fixed to 4.5 kpc in all fits, following the preferred choice in S17. We also fit the deviation from an 8 kpc distance from the Sun to the Galactic centre so that the new distance is 8 kpc +  $\Delta R_0$ . Additionally, we fitted  $\alpha$ , which is the angle between the bulge major axis and the line connecting the Sun to the Galactic centre.

We optimize our parametric models for parameter set  $\mathfrak{P}$  using the SCIPY BFGS routine,<sup>1</sup> minimizing the Poisson likelihood statistic:

$$\ln \mathcal{L} = \sum_{\{i,j,k\} \in \{K_s, l, b\}} (n_{i,j,k} \ln N_{i,j,k} - N_{i,j,k}) + \text{constant}, \quad (10)$$

where  $N_i$  is the corresponding model, obtained by integrating the equation of stellar statistics (equation 1) for parametric density  $\rho_{\text{SX}}$ . Our best-fitting likelihoods and uncertainties are listed in our tables of results (Tables A1, A2, and A3). The uncertainties are derived from the corresponding square root of diagonal elements of the inverse Hessian matrix produced by this routine. The SX model fit was initialized by randomly picking a starting point somewhere between qualitatively different boundaries that produce physically possible densities for the X perturbation parameters and choosing the initial S parameters from within 10 percent of the best-fitting values from the S-model.

**Table 2.** Density distribution parameters for the bulge component used for our simulation. The second row gives the total number of stars in the unmasked regions of the simulation in the range  $12 < K_s < 14$ . In cylindrical coordinates, centred at the maximum density of the bulge, the Sun is located at  $(R_{\odot}, Z_{\odot}) = (8.0 \text{ kpc}, 15.0 \text{ pc})$ .

$x_0$ (kpc)	$y_0$ (kpc)	$z_0$ (kpc)	$\alpha$ ( $^{\circ}$ )	$c_{\parallel}$	$c_{\perp}$
1.61	0.69	0.48	19.16	2.50	1.86
$N_{\text{thin}}(\times 10^6)$		$N_{\text{thick}}(\times 10^6)$		$N_{\text{bulge}}(\times 10^6)$	
1.35		1.87		17.04	

### 3 TESTING THE DECONVOLUTION AGAINST A SIMULATION

We constructed a simulated Milky Way population comprised of a thin disc, thick disc, and a bulge, as is modelled in S17. To generate the synthetic population, we used

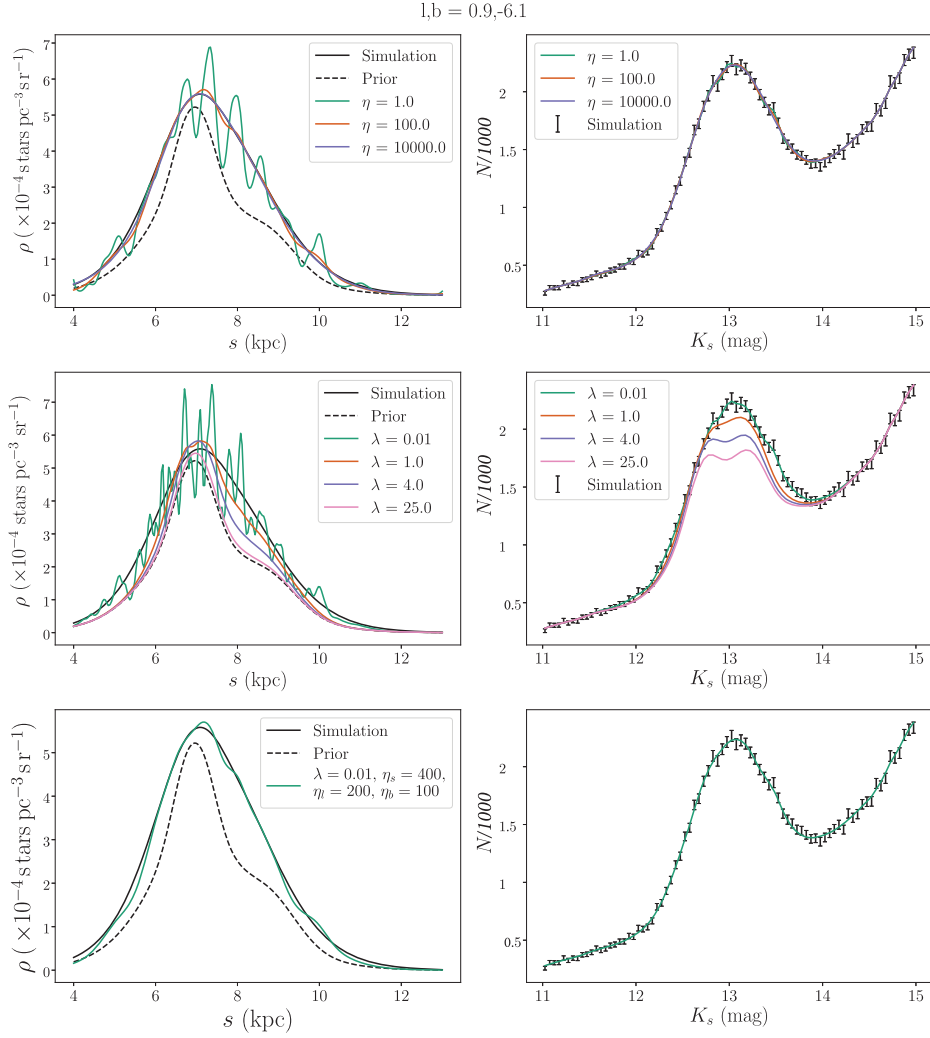
$$\begin{aligned} N(K_s, l, b) &= \Delta \Omega \Delta K_s \\ &\times \sum_i \int_0^{\infty} \rho_i(s, l, b) \Phi_i(K_s - 5 \log s - 10) s^2 ds, \end{aligned} \quad (11)$$

where  $\rho$  is the density and  $\Phi$  is the luminosity function and the sum is over the three model components, to predict the combined star counts in each  $(K_s, l, b)$  voxel. We then simulated a population of stars by drawing a Poisson random value from the binned simulation model. As in P19, the thin and thick discs were generated from the updated Besançon model parameters of Robin et al. (2012, 2014), respectively. The S-bulge model is given by equation (9) with  $A = 0$ . The simulation parameters used for this model are listed in Table 2.

The normalizations we used for each of the three components have been multiplied by the same constant chosen so that the total number of stars in the unmasked region and in  $12 < K_s < 14$  matches the number of stars in the VVV PSF catalogue. The luminosity function we used for the bulge in the simulation is the same as the one we used in our fitting procedure to the VVV data.

To choose the values of the regularization parameters, we tested a range of choices in a  $1^{\circ} \times 1^{\circ}$  region centred on  $(l, b) = (0.9, -6.1)$ . The test region was subdivided into our usual voxel size of  $(0.05 \text{ mag} \times 0.2^{\circ} \times 0.2^{\circ})$ . For this test, we did not want to use a prior that was too close to the true value, so we used the base SX (equation 9) model that had been fitted to the VVV data. We first fixed the maximum entropy regularization parameter,  $\lambda$  from equation (8), to zero and applied our maximum entropy deconvolution method with a range of smoothing regularizations,  $\eta$ . We repeated this for  $\eta = 0$  and a range of  $\lambda$  values. In Fig. 3, the deconvolved density for all choices of  $\eta$  follow the general shape of the true density. Small values of  $\eta$  give spurious oscillatory deviations from the true density, which decrease in amplitude as  $\eta$  increases. There is not a significant difference in the predicted star counts between the choices of  $\eta$ . For  $\lambda \geq 1.0$ , the predicted star counts deviate significantly from the simulation, which is also seen in the deconvolved density where it overestimates at distances less than 6 kpc, and underestimates from 6 to 8 kpc. This is because the prior density is not a good estimate of the true density for the current case. When  $\lambda = 0.01$ , the deconvolved density is scattered around the simulated density, and the predicted star counts are overfitting. The results of this test suggested that a small value of  $\lambda$  and a large value of  $\eta$  would give the most accurate density deconvolution. Therefore, we used a value of  $\lambda = 0.01$  and  $\eta = 100-1000$ . For the background modelling, a simulation is not needed to determine an optimal set of regularization parameters, as the effectiveness can be determined by directly comparing to the data. Also, the

<sup>1</sup><https://www.scipy.org/>

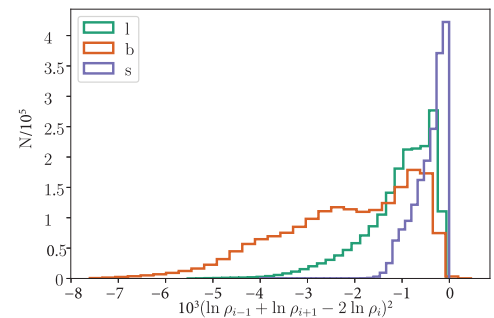


**Figure 3.** Testing the choice of regularization parameters. We perform our maximum entropy deconvolution to a  $1^\circ \times 1^\circ$  region of our simulated population, centred on  $(l, b) = (0.9, -6.1)$ . In the top panels, the maximum entropy regularization is set to zero, and a range spatial smoothness parameter values are tested. The middle panels have the spatial smoothness regularization set to zero, and a range of maximum entropy regularization values are tested. The bottom panels have the regularization parameters used in our final analysis. The left-hand panels show the deconvolved density compared to the true density in the simulation, the right show the model star counts ( $N$ ) compared to the simulated population ( $n$ ). For small values of  $\eta$ , the deconvolved density has many spurious features, which get smaller in amplitude as  $\eta$  is increased. The predicted star counts is not significantly sensitive to the choice of  $\eta$  in the range tested here. For all  $\lambda \geq 1.0$ , the predicted star counts do not match the simulation, where it is clear that the prior density distribution is not a good estimate of the true density.

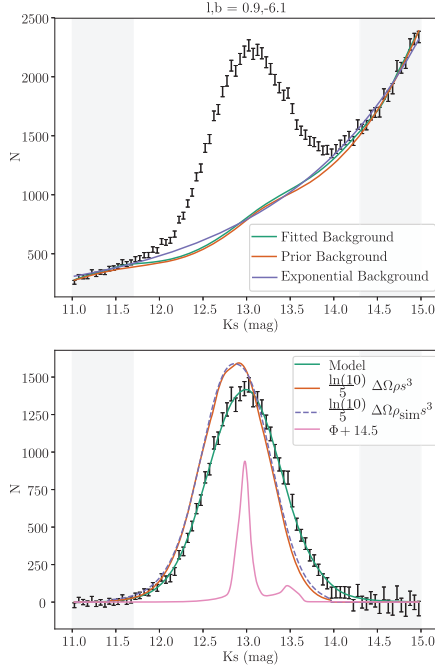
prior background from the S17 model gives a good description of the background. This means we expect less deviation from the prior and so a larger value of  $\lambda$  can be used. The regularization parameters used for the background determination are presented in Table 1.

The distribution of curvature in log-density (equation 3) for the simulated bulge in Fig. 4 is strictly negative. It is broadest in  $b$ , second broadest in  $l$ , and narrowest in  $s$ . The  $\ell_2$ -norm regularization gives a minimum penalty to the likelihood when the log of the fitted density has zero curvature. We chose  $\eta_s$ ,  $\eta_l$ , and  $\eta_b$  such that the overall curvature penalty term in equation (8) was of similar magnitude. From the distributions of the curvature term in Fig. 4, we chose the regularization parameters used for fitting the simulated population as listed in Table 1.

We applied the maximum entropy deconvolution process to the simulated star counts, first by fitting the background including the feature behind the bar, then by fitting a parametric density model to determine a prior density estimation for the full three-dimensional



**Figure 4.** Distributions of the curvature in log-density (equation 3) along respective density model coordinates of the simulated bulge.

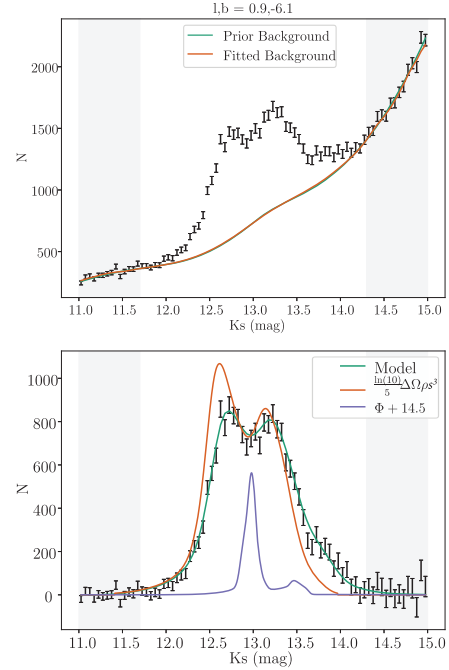


**Figure 5.** Demonstration of the maximum entropy deconvolution method to a simulated population for a  $1^\circ \times 1^\circ$  region. *Top*: The background has been fitted in the grey-shaded regions using the maximum entropy method. The prior background is the background of the model used to generate the simulation. *Bottom*: Maximum entropy deconvolution of the line-of-sight star count distribution. Shown in green is the predicted number of RC star counts from the convolution of the fitted density (orange) and assumed luminosity function (pink). The density used to produce the simulation is shown as a dashed purple line. The luminosity function has been scaled and shifted for display, where 14.5 is the distance modulus added to the absolute magnitudes,  $M_{K_s}$ .

density deconvolution. The parameters of the fitted prior density are presented in Table A3, labelled case A. The maximization of the  $\ln \mathcal{L}$  in equation (8) and  $\ln \mathcal{L}$  in equation (4) were both performed using the PYTHON implementation PYLBFGS<sup>2</sup> of the Limited Memory Broyden–Fletcher–Goldfarb–Shanno algorithm.

The density was modelled non-parametrically on a (257, 100, 50) grid of  $(s, l, b)$ , in the range  $4 < (s/\text{kpc}) < 13$ ,  $-10^\circ < b < 0^\circ$  and  $-10^\circ < l < 10^\circ$ , for a total of  $1.285 \times 10^6$  free parameters. The grid spacing is  $(\Delta s, \Delta l, \Delta b) = (35 \text{ pc}, 0.2^\circ, 0.2^\circ)$ . To make the optimization of so many parameters feasible, we evaluated the gradients of  $\ln \mathcal{L}$  in equation (8) and  $\ln \mathcal{L}$  in equation (4) analytically, see appendix A of P19 for more details. We assumed symmetry about the Galactic mid-plane so that we could reliably extend our non-parametric density model to latitudes  $b > 5^\circ$ , where there are no observations in the VVV sample. Making the mirror symmetry assumption forced us to position the Sun in the Galactic mid-plane ( $Z_\odot = 0 \text{ kpc}$ ). We fixed the reconstructed density just outside the region of interest to the prior density by setting  $\lambda = 1$  in those regions. This meant that the smoothness regularization forced the reconstructed density to smoothly transition to the parametric prior density at  $|l| > 10^\circ$  and  $|b| > 10^\circ$ .

Shown in the top panel of Fig. 5 is the background fitted to the simulation. From the deconvolution of the VVV data shown in Fig. 5, we can see the simulated population lacks a splitting of the



**Figure 6.** Demonstration of the maximum entropy deconvolution method in a  $1^\circ \times 1^\circ$  ( $5 \times 5$  pixels) region. *Top*: The background has been fitted in the grey-shaded regions using the maximum entropy method. The prior background was calculated using the S17 S-model+discs, which has been scaled to match the VVV observations between  $11.0 < K_s < 11.5$ . *Bottom*: Maximum entropy deconvolution of the line-of-sight background-subtracted star count distribution. Shown in green is the predicted number of RC star counts from the convolution of the fitted density (orange) and assumed luminosity function (purple). The luminosity function has been scaled for display.

RC peak that is present in the VVV observations case shown in Fig. 6. In Fig. 7, we compare the three-dimensional deconvolved density to the density used in simulating the population. The deconvolved density using the maximum entropy method compares well to the density used in our simulation, even inside of the masked regions where there is no data influencing the deconvolution. However, the reconstruction displays some discrepancy at around  $s = 4 \text{ kpc}$ . Note that this is due to the low star counts in the bulge at this radius that makes an accurate reconstruction difficult. Note that Fig. 7 correctly does not show the X-bulge morphology that is seen in the VVV data, which is displayed in Fig. 8.

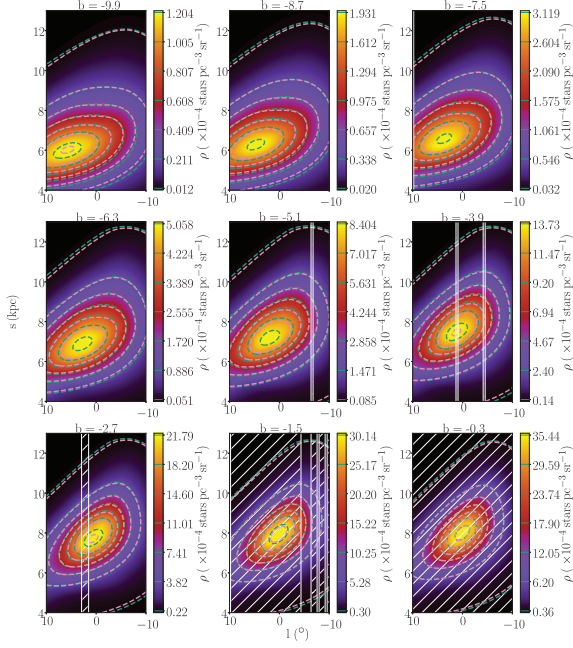
#### 4 DECONVOLUTION OF VVV

In this section, we discuss how we applied our maximum entropy deconvolution method to the VVV data sample for our base model, which we label as case A. We used a fit of the parametric SX model as the prior density distribution and the values for the regularization parameters in Table 1. The background was fitted using the maximum entropy method of Section 2.4. In Fig. 6, we present a breakdown of the maximum entropy deconvolution model components along a single line of sight through the region the photometric split clump has been observed.

Displayed in Fig. 9 is a comparison between the predicted star counts by our maximum entropy deconvolution, the fitted parametric model we used as the prior, and the VVV data. For compactness,

<sup>2</sup><https://github.com/dedupeio/pylbfgs>





**Figure 7.** Deconvolved RC + RGBB star density using the maximum entropy method for a simulated 10 Gyr S-model. The white hatched regions were masked during the analysis, and were inpainted naturally as part of the deconvolution. The green-dashed contours show the true density used in simulating the S-bulge. Pink show the parametric SX model used as the prior density.

we show every 10th magnitude bin. At  $K_s < 12$  and  $K_s > 14$ , the RC + RGBB stars contribute negligibly to the total star counts, so both the parametric model and maximum entropy deconvolution are dominated by the background. By construction, these regions are well described by the background model, though perhaps there is slight overfitting in the  $K_s = 14.975$  bin. The non-parametric model reproduced the data well and has smaller deviations in comparison to the parametric model, especially notable in the  $K_s = 12.525$  bin at  $l = 5^\circ$  where the X-bulge is prominent. The assumption of symmetry about the Galactic mid-plane seems to be reasonable, as there is no visible bias in fitting to the mirrored contours above and below the plane.

The deconvolved density and the fitted parametric density, for fixed latitude bins, are shown in Fig. 8. For compactness, only 9 of the 50 bins are displayed and only for  $b < 0^\circ$ , as the density is symmetric about  $b$ . Unlike the simulated bulge shown in Fig. 7, the density from deconvolution of the VVV data shows the arms of the X-bulge, first noticeable at  $b = -8.7^\circ$  for  $(l, s) = (4.7^\circ, 6.6 \text{ kpc})$  and  $(l, s) = (-3.3^\circ, 9 \text{ kpc})$ . As latitude decreases, the arms get closer until they merge at  $b = -2.7^\circ$ . The maximum density at  $b = -2.7^\circ$ , where the arms merge, is at longitude  $l = -0.7^\circ$ . The maximum density of the X-bulge arms in the parametric model do not align with the maximum density in the non-parametric model, which is also evident in the star counts. Cartesian versions of the reconstructed bulge from the VVV data and the simulation are shown in the first columns of Figs 10 and 11, respectively.

## 5 SYSTEMATIC TESTS

In order to gain a better understanding of the robustness of our results, we test systematics based on the following:

- (i) Adding the feature behind the bar to the background (case B).
- (ii) The VVV data mask (case J).
- (iii) The determination of the background component (case C).
- (iv) The semi-analytic luminosity function (case D and I).
- (v) The metallicity distribution (case E).
- (vi) The position of the Sun (case F, G, H, I).
- (vii) The deconvolution method used (Appendix A1).

We tested the significance of these assumptions by systematically changing one, then repeating the maximum entropy deconvolution, including the background fitting and parametric prior density model fitting. We also repeated the deconvolution with the new assumptions on the simulated population.

The results of fitting the SX model to data and simulations are listed in Tables A2 and A3, and are plotted in Figs 12, 13, and 14. Except where specified, the parametric model has been fitted twice, following the prescription of the deconvolution method in Section 2, in which the feature behind the bar is subsumed into the background. By fitting to the S-model simulation generated by the parameters in Table 2, we hoped to gauge the impact on the likelihood of different background and parametric model cases used in bulge modelling. Note that in the simulation, we chose  $Z_\odot = 15 \text{ pc}$ . As can be seen in Fig. 13, the range of fitted model parameters is much greater than the error bars in Table A3. This indicates the main cause of the variation is due to model assumptions rather than statistical error. We used the following test statistic (TS) to compare the different cases:

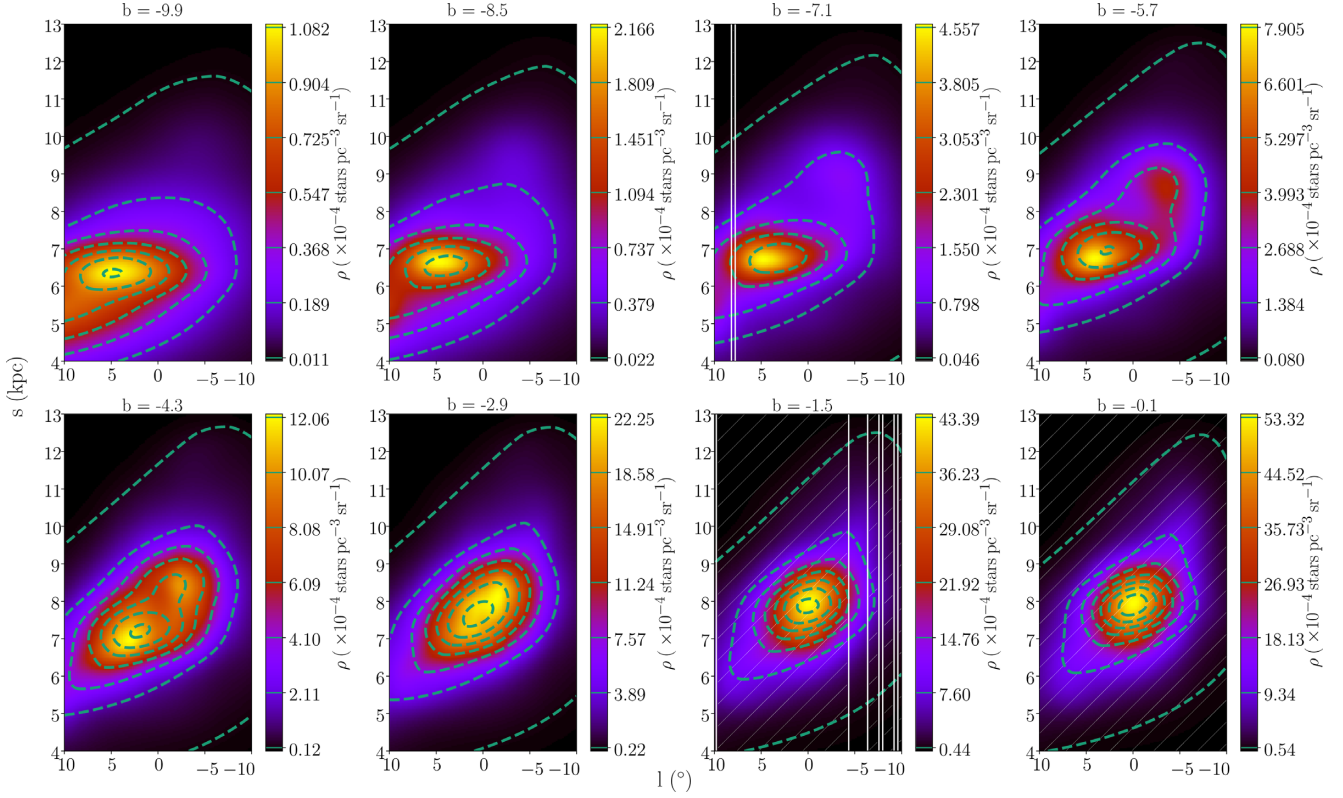
$$TS \equiv -2 \ln(\mathcal{L}/\mathcal{L}_{\text{base}}). \quad (12)$$

As most of the variation between cases was due to systematic error rather than statistical error, we did not use Wilks' theorem (Wilks 1938), which is also only suited for nested models. This means we cannot associate the TS value with a p-value in the usual way. We can get a rough measure of what a significant TS value is by comparing to the corresponding TS values seen in simulations. The median value of the simulation TSs for the combined top and bottom panels of Fig. 12 was  $TS \approx 10^4$ . We take this as our threshold above which the TS value is regarded to be significant.

### 5.1 Feature behind the bar

As can be seen both in the top and bottom panels of Fig. 12, the simulation has a negligible TS when testing against case B that does not account for a feature behind the bar. This is to be expected as this feature was not present in the simulation. In contrast, for the parametric fit (top panel), the data has a high TS for case B. This indicates that the feature behind the bulge is significant. In P19, we analysed this case in more detail. Also, in that paper we used a parametric background that then also revealed a feature in front of the bulge. The non-parametric background in this paper has absorbed the feature in front of the bar.

However, in the non-parametric case (bottom panel) we do not find a significant change in our penalized likelihood when not removing the feature behind the bar. This can be seen in the bottom panel of Fig. 12 where case B has a TS very close to zero for both the data and simulation. This is to be expected as the flexibility of the non-parametric method can easily incorporate the feature behind the bar as being part of the bulge as seen by comparing columns 1 and 2 in Fig. 10. While for the simulation, where there should be no feature behind the bar, the corresponding columns are virtually indistinguishable as seen in Fig. 11.



**Figure 8.** Deconvolved RC + RGBB star density using the maximum entropy method. The white hatched regions were masked during the analysis, and were inpainted as part of the deconvolution. The prior density model is shown in green-dashed contours.

## 5.2 Background systematics

We changed the background in case C to one that is common in the literature, a second-order polynomial in  $\log(N)$ , described in Section A1. We have already displayed this background for a couple of lines of sight in Fig. 2. At high latitudes (top panel), this background tends to estimate higher counts than the maximum entropy background for  $12 < K_s < 12.5$  and estimate fewer counts at  $13 < K_s < 14$ . At lower latitudes, this background tends to overestimate at all  $K_s$ , especially at around  $K_s = 12.0$ . On the simulation, the exponential background significantly overestimates in the range,  $11.7 < K_s < 13.0$ , as shown in the top panel of Fig. 5. As a result, the density is underestimated on the near side ( $x < 0$ ) of the bulge at low latitudes when using the exponential background rather than the maximum entropy background in both the VVV data (Fig. 10) and simulated population (Fig. 11).

In Fig. 12, for the parametric fit (top panel), the exponential background (case C) has the worst TS both for the data and simulation, out of all of the cases considered in that panel. The TS was also high for both the data and simulation in the non-parametric case as shown in the bottom panel of Fig. 12. This provides further evidence that the maximum entropy method is providing a better background than exponential model approach.

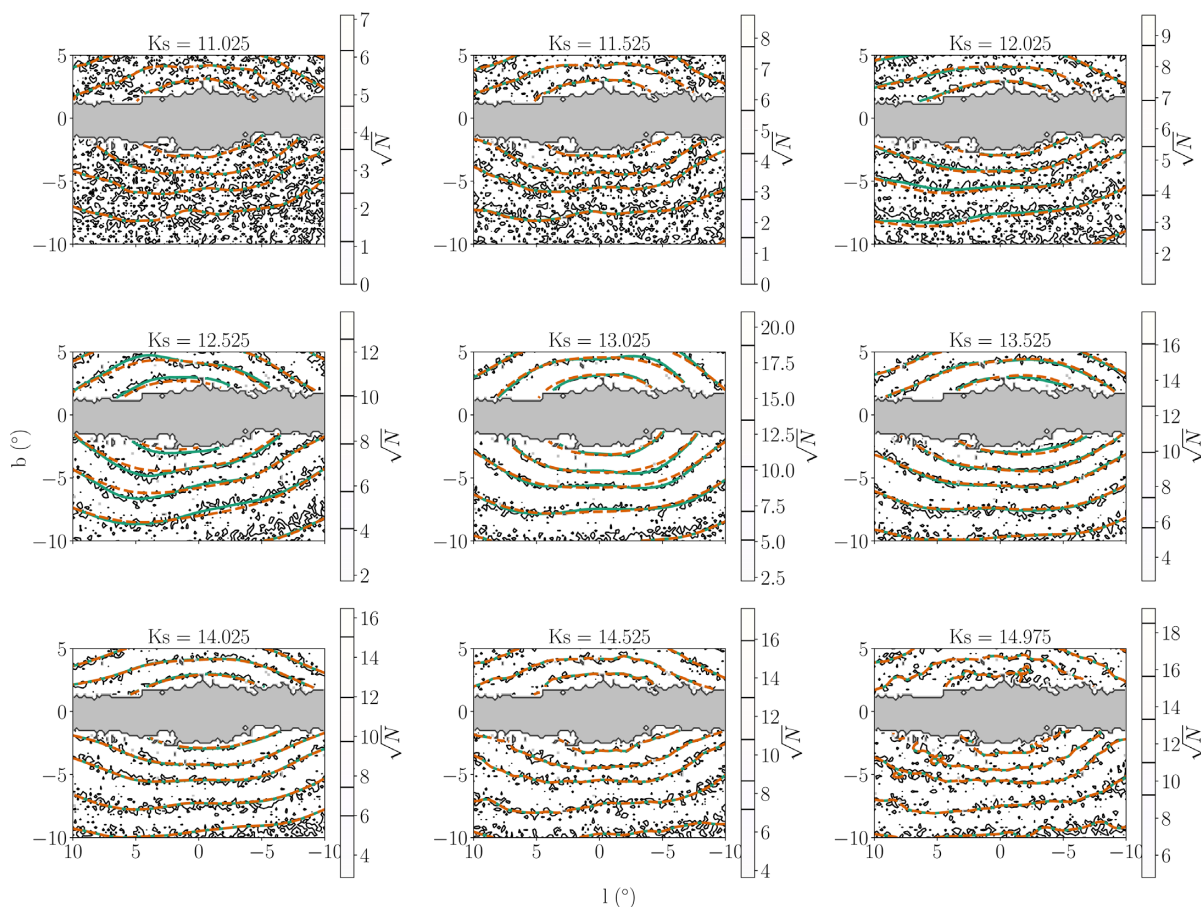
## 5.3 Luminosity function systematics

S17 found that the best-fitting luminosity function was significantly broader than the luminosity function they had simulated with GALAXIA (Sharma et al. 2011), using the same isochrones we have used in our analysis. We also tried a similarly broad luminosity function, by convolving our luminosity function (of approximate

Gaussian width 0.06 mag) with an additional Gaussian with a standard deviation of 0.24 mag. The density slices in the ‘Broad  $\Phi$ ’ column of Figs 10 and 11 are consistent with the broadened luminosity function requiring a narrower and more angled bulge. A similar relationship can be seen in Fig. 16 of S17. In the top panel of Fig. 12, the SX parametric model with broadened luminosity function (case D) had a slightly improved TS for the data, while it was disfavoured for the simulation. However, this broader luminosity function is not consistent with recent measured intrinsic RC magnitude dispersions in the  $K_s$  band of 0.03–0.09 mag (Chan & Bovy 2019; Hall et al. 2019). Also, in Fig. 14, the X-shape parameters,  $n$  and  $x_1$ , are anomalous for case D. The consequence of this was that the broader luminosity function fit resulted in unnaturally narrow X-arms as depicted in Fig. 15.

As can be seen in the non-parametric results shown in the bottom panel of Fig. 12, the broader luminosity function (case D) provided a high TS for the simulations indicating a bad fit. This is to be expected as the simulations were based on our standard narrower luminosity function. The TS for the data was so high for the broad luminosity function that we could not accommodate it in Fig. 12 without making the range of the plot too great to see any of the other details. This was because the non-parametric model was being heavily penalized for deviating greatly from the prior SX model, which had converged to a physically unnatural solution, shown in the top panel of Fig. 15.

Since our prior for the maximum entropy deconvolution was unnatural for the broad luminosity function, we wanted to check if a different prior gave similar results. So we repeated the test, but instead we used an S-model as the prior density, shown in the bottom panel of Fig. 15. As can be seen in the top panel of Fig. 12, this



**Figure 9.** Predicted star counts for our maximum entropy deconvolution method. The black contours show the VVV star counts, where the levels of the contours are indicated by the black lines on the colour bar. The green contours show the star counts predicted by the non-parametric model, where the levels match the black contours. The orange-dashed line is the parametric model used as the prior. Contours are produced using the same resolution as the data, i.e.  $0.2^\circ \times 0.2^\circ$  degree grid with  $K_s$  bins of width 0.5 mag.

S-model with a broad luminosity function (case I) was disfavoured by both the data and the simulation for the parametric case. Also, as presented in the bottom panel of Fig. 12, case I did have a significant TS for the non-parametric fit in the case of the data. This indicates that from a TS perspective, our non-parametric results disfavour a broad luminosity function.

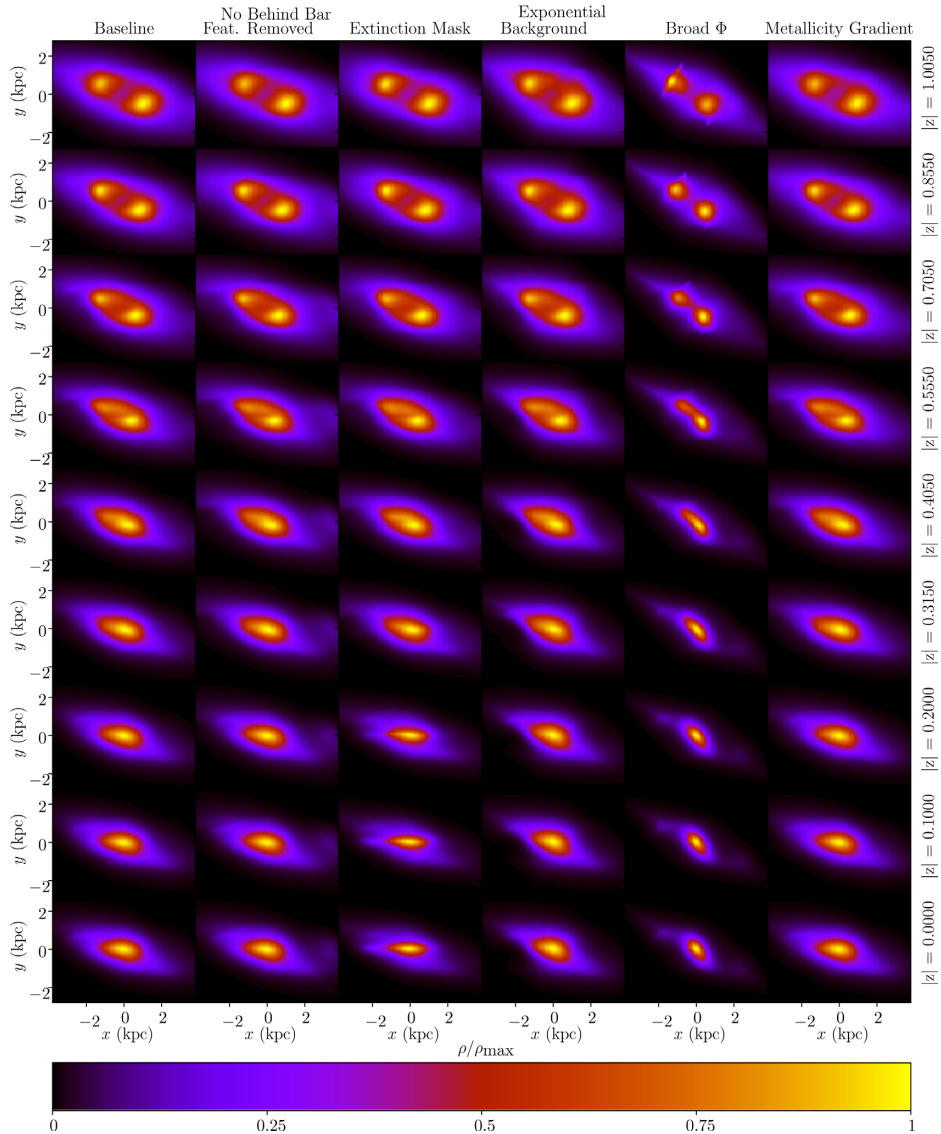
#### 5.4 Metallicity distribution systematics

Our base case assumed that the metallicity distribution is constant throughout the bulge. Several spectroscopic studies, e.g. Zoccali et al. (2017) and García Pérez et al. (2018), have observed a vertical metallicity gradient in the bulge, where stars near the Galactic mid-plane are on average more metal rich than stars on the periphery of the bulge. We used the photometric metallicity map generated by the BEAM-II calculator (Gonzalez et al. 2018) to allow the metallicity distribution function in the computation of our semi-analytic luminosity function to have a different mean metallicity for every line of sight. The metallicity dispersion was kept fixed at 0.4 for this test. Shown in Fig. 16 (top panel) is the metallicity map of Gonzalez et al. (2018), where we have filled the missing values with  $[\text{Fe}/\text{H}] = 0.0$ . From the luminosity functions in bottom panel of Fig. 16, it is clear that the lower metallicity line of sight has a fainter RC, and is naturally broader, though the difference in brightness is only 0.03 mag between  $b = -9.1$  and  $b = -3.1$ . Some

part of the broadness is from the overlapping of the RC and RGBB, since the RGBB is brighter at lower metallicities. Qualitatively, the density that was fitted using the metallicity gradient is nearly identical to the base case as seen in the last column of Fig. 10. The insensitivity to the metallicity gradient can be seen in case E for the bottom and top panel of Fig. 12. The TS changes for the metallicity cases are negligible in comparison to the TS changes associated with the other systematics. The E case does appear to have an anomalous  $x_1$  in Fig. 14. However, as  $A \approx 0$  for the E case, its X-component is negligible. We conclude from this test that the inclusion of a simple unimodal metallicity gradient does not significantly affect our results. A more sophisticated double population model, consistent with spectroscopic observations, is necessary to properly include a metallicity gradient.

#### 5.5 Sun position systematic

Our simulated population of stars had the Sun located at  $Z_\odot = 15$  pc, which is different to the  $Z_\odot = 0$  pc assumed in our base model. We tested the significance of this assumption by fitting an S-model with the Sun in the same position as in our simulation (case F). We still assumed symmetry in the maximum entropy density about  $b = 0^\circ$ . The top panel of Fig. 12 shows how parametric case F provided an improved fit to the simulation. This is to be expected as it corresponds with the model used to generate the simulation. In



**Figure 10.** Cartesian projections of the deconstructed density of the VVV RC stars in the bulge, for several systematic test cases. The  $x$ -axis is aligned with the Sun–Galactic centre line and the  $z$ -axis is perpendicular to the Galactic plane and measured in kpc. The Galactic centre is located at the maximum bulge density. The significance of each test case is discussed in the text in Section 5.

the case of the VVV data, it is harder to interpret the case F result in Fig. 12 as we have changed both the position of the Sun and the parametric form of the prior density. The difference between case F and case H is the position of the Sun, where both differ from the base case by having an S-model parametric form. The VVV data TS of case F was significantly larger than case H in the parametric case, however, there was less of a difference when fitting the parametric model to the simulation.

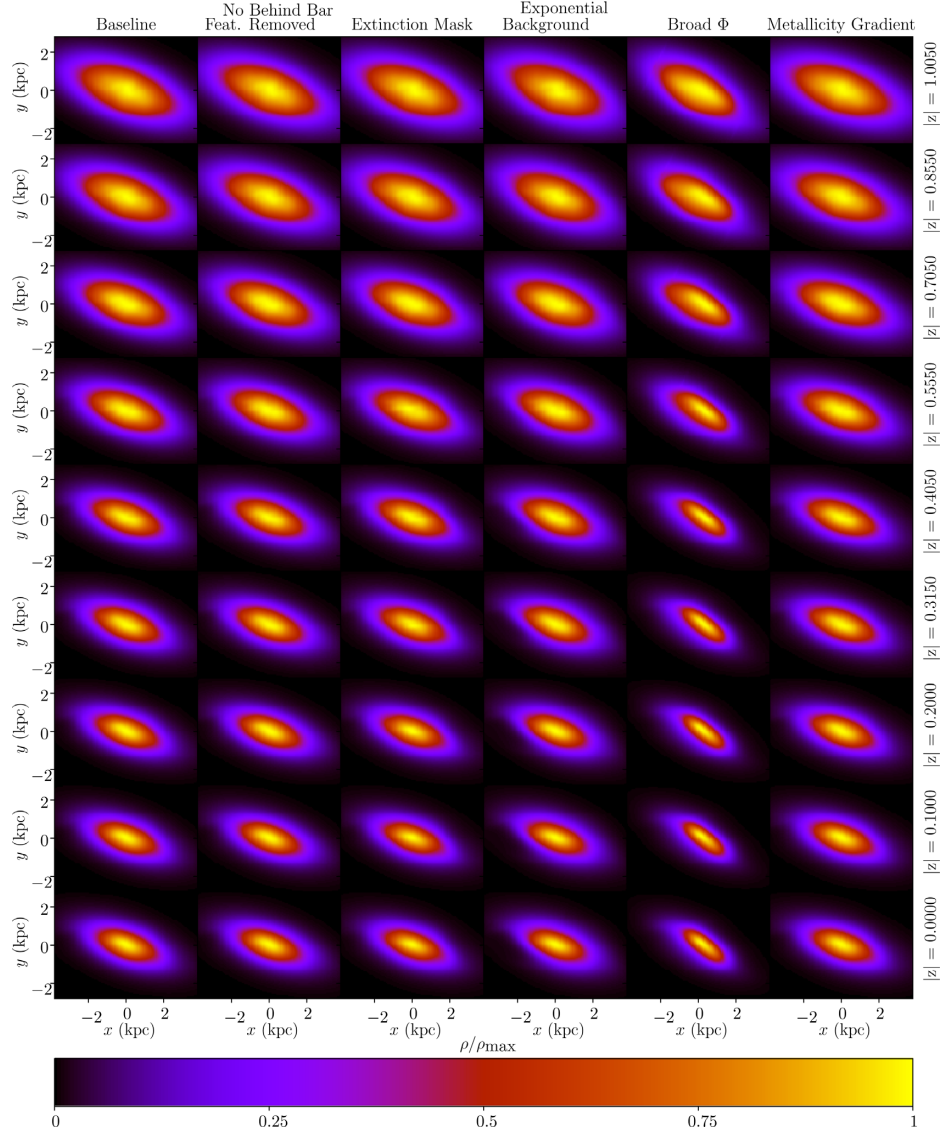
This confirms that the VVV data prefers  $Z_{\odot} = 0$  pc when fitting the parametric S-model as seen in the top panel of Fig. 12. When comparing the same cases, F and H, for the non-parametric method, case F had a significantly larger TS than case H for both the simulated population and the VVV data. It is hard to interpret this result for the non-parametric model, given that it had an assumed symmetry around the  $Z_{\odot} = 0$  pc plane. However, we relaxed this assumption in P19 without issue.

Case H is an S-model with  $Z_{\odot} = 0$  pc. As can be seen from the top panel of Fig. 12, for the parametric fit, the data significantly prefer

the SX model. Also, for the parametric fit, the F case is very slightly favoured over the SX model for the simulation. This follows in that the F case is of the same form as the model used to generate the simulation. However, case F is even more disfavoured by the data than case H. From this we conclude that, for the parametric fit, the data favours the SX model over the S-model and this conclusion is not affected by reasonable changes in  $Z_{\odot}$ .

## 5.6 Mask systematic

We changed the region in which the data are excluded, from the combined extinction and  $K_s$ -band uncertainty boundary case ( $\sigma > 0.06$ ), to a colour excess mask  $E(J - K) > 0.9$ . This systematic test changes the amount of data used in the analysis, so the likelihood is not comparable to the base case. In Fig. 10, the density that is reconstructed with an extinction only mask has a prominent bar-like feature at  $|z| < 0.2$  kpc, that is pointed nearly directly towards the Sun. Note that this feature is not seen in the corresponding



**Figure 11.** Cartesian projections of the deconstructed density of the simulated bulge population, for several systematic test cases. The  $x$ -axis is aligned with the Sun–Galactic centre line and the  $z$ -axis is perpendicular to the Galactic plane and measured in kpc. Nearly all of the cases give a qualitatively similar density to the base case. However, the exponential background gives densities that are too low at  $(x, y) = (-2.5\text{kpc}, 0\text{kpc})$ , especially at low latitudes. Also, the broadened luminosity function gives a larger bar angle than the base case. The two exceptions noted here are also seen in the VVV data (Fig. 10).

simulation result of Fig. 11. We extracted this feature by subtracting the baseline case. Plotted in Fig. 17 is the sum of the density difference for all density with  $|z| < 1$  kpc. At first glance, this apparent overdensity looks similar in structure to the younger, secondary population of bulge stars in S17 (E component of the S + E model). The green star indicates the maximum density of the difference and is located at  $(x, y) = (120\text{ pc}, 90\text{ pc})$ . This is 150 pc behind the centre of the bulge ( $(x, y) = (0\text{ pc}, 0\text{ pc})$ ). This suggests that the stars are unlikely to be from a significantly younger or more metal-rich population than the rest of the stars in our bulge model, as they would have a brighter RC in the luminosity function than we have modelled. A 5 Gyr old population with a similar metallicity distribution to our fiducial case has an RC that is 0.1 mag brighter, which corresponds to a difference of 400 pc closer at 8 kpc, indicated by the cyan triangle in Fig. 17.

We argue based on the reconstructed distance from the Sun that the apparently overdense region is not consistent with a

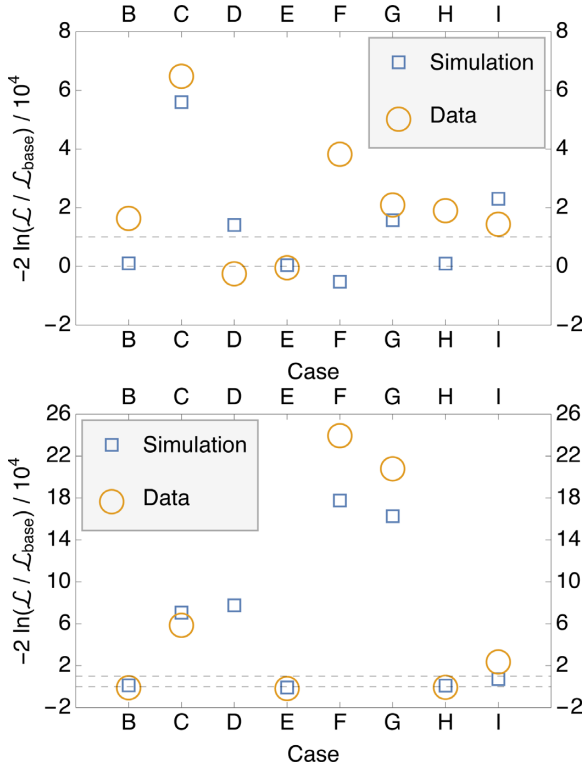
different population of stars. Its orientation, which is suspiciously pointed directly towards the Sun, and is distinctly different from the majority of the bulge population also makes it inconsistent with main population of the bulge stars. This was one of our motivations in using the crowding + extinction based mask over the extinction-only-based mask. A combination of significant crowding and residual extinction deteriorates the quality of the star count catalogues, including the photometric zero-point.

## 6 APPLICATIONS

### 6.1 Properties of the bulge

#### 6.1.1 Mass of the bulge

From the fitted density and IMF, we can estimate the total mass of the bulge. Integrating the RC + RGBB stellar density over the



**Figure 12.** The parametric (top panel) and non-parametric (bottom panel) likelihood ( $\mathcal{L}$ ) for the different cases considered. The base case’s non-parametric likelihood is  $\mathcal{L}_{\text{base}}$ . Results are shown for both the simulations and the data. In the simulation case, the base case and labelled case are both fit to the simulated data. In the bottom panel, Case D’s data symbol is not shown due to its very low likelihood value not being in the range of the plot for the non-parametric fit. See Table A1 for numerical values including the one for Case D in the non-parametric fit. The dashed lines are for  $\text{TS} = 0$  and the median of the simulations for both the parametric and non-parametric case, which is  $\text{TS} \approx 10^4$ . The cases considered are no behind-the-bar feature subtraction (B), exponential background (C), broad luminosity function (D), metallicity gradient (E), S-model prior with  $Z_{\odot} = 15$  pc (F), S-model prior and broad luminosity function with  $Z_{\odot} = 15$  pc (G), S-model prior with  $Z_{\odot} = 0$  pc (H), S-model prior with  $Z_{\odot} = 0$  pc with a broad luminosity function (I).

entire bulge region gives us a total of  $19.1 \times 10^6$  (RC + RGBB) stars. Based on our luminosity function, 0.062 per cent of all stars are in either the RC or RGBB, so the total number of stars in the bulge is  $N_{\text{total}} = 30.7 \times 10^9$ . Stars in the bulge with a mass  $> 1 M_{\odot}$  have evolved into stellar remnants, so the normalization of the IMF is then given by

$$\xi_0 = \frac{N_{\text{total}}}{\int_{0.15 M_{\odot}}^{1 M_{\odot}} \xi(m) dm}, \quad (13)$$

where  $\xi$  is the IMF and  $\xi_0$  is the normalization of the IMF. We use the Chabrier IMF, which was also used to generate our luminosity function. With the IMF correctly normalized, the mass of the bulge is then calculated by integrating the IMF multiplied by the final mass of the star, over the range  $0.15 M_{\odot} < m < 150 M_{\odot}$ . Stars with an initial mass  $< 1 M_{\odot}$  have not yet evolved into remnants, so the final mass is equal to the initial mass. Stars with initial mass  $1 M_{\odot} < m < 8 M_{\odot}$  have evolved into white dwarfs, where the final mass is related to the initial mass by  $m_f = 0.48 + 0.077m_i$  (Maraston 1998). To determine the final mass stars with initial mass  $> 8 M_{\odot}$ , which have evolved into neutron stars or black holes, we

use the results of the numerical population synthesis code SEVN (Spera, Mapelli & Bressan 2015). Therefore, the total stellar mass of the bulge (assuming a Chabrier lognormal IMF) is  $M_{\text{bulge}} = 1.64 \times 10^{10} M_{\odot}$ . This includes the mass of the stellar remnants, which make up 30.1 per cent of the total mass.

Parametric modelling of VVV bulge stars in S17 found a total stellar mass of the bulge assuming a Chabrier IMF of  $2.36 \times 10^{10} M_{\odot}$ , with the stellar remnants making up 49 per cent of the total mass. Both the total mass and remnant fraction of S17 are larger than we are reporting. However, if we were to have the same remnant fraction as S17, then our total mass would be  $2.24 \times 10^{10} M_{\odot}$  which would be consistent with S17 once our systematic uncertainties have been incorporated.

A dynamical estimate of the bulge mass by combining the VVV bulge stellar distribution of WG13 with kinematic information from BRAVA in Portail et al. (2015) found a bulge stellar mass of  $1.3\text{--}1.7 \times 10^{10} M_{\odot}$ , which is consistent with our estimated mass. They also provide a mass-to-clump ratio, which is used to estimate the total stellar mass of the bulge from the number of RC + RGBB stars. For a Chabrier IMF, there are approximately  $905 M_{\odot}$  of bulge mass for each RC + RGBB star. So, for our estimated  $19.1 \times 10^6$  (RC + RGBB) stars the estimated mass was  $1.73 \times 10^{10} M_{\odot}$ . This is remarkably similar to our value, considering Portail et al. (2015) used different isochrones, metallicity distribution, and treatment of the compact remnants to those used in our estimation. Additionally, we list the bulge mass estimates for all of our systematic test cases in Table 3. As can be seen, the mass estimates of the simulated data encompass the mass of the model used for the simulation with a spread of a few per cent. As the systematic error is much greater than the statistical error, we use the range of best-fitting bulge mass estimates for our different cases to get an estimate of the uncertainty in our mass estimate. The mass estimates for the bulge from the VVV data are in the range  $1.33\text{--}1.71 \times 10^{10} M_{\odot}$ , which is in agreement with the results of Portail et al. (2015).

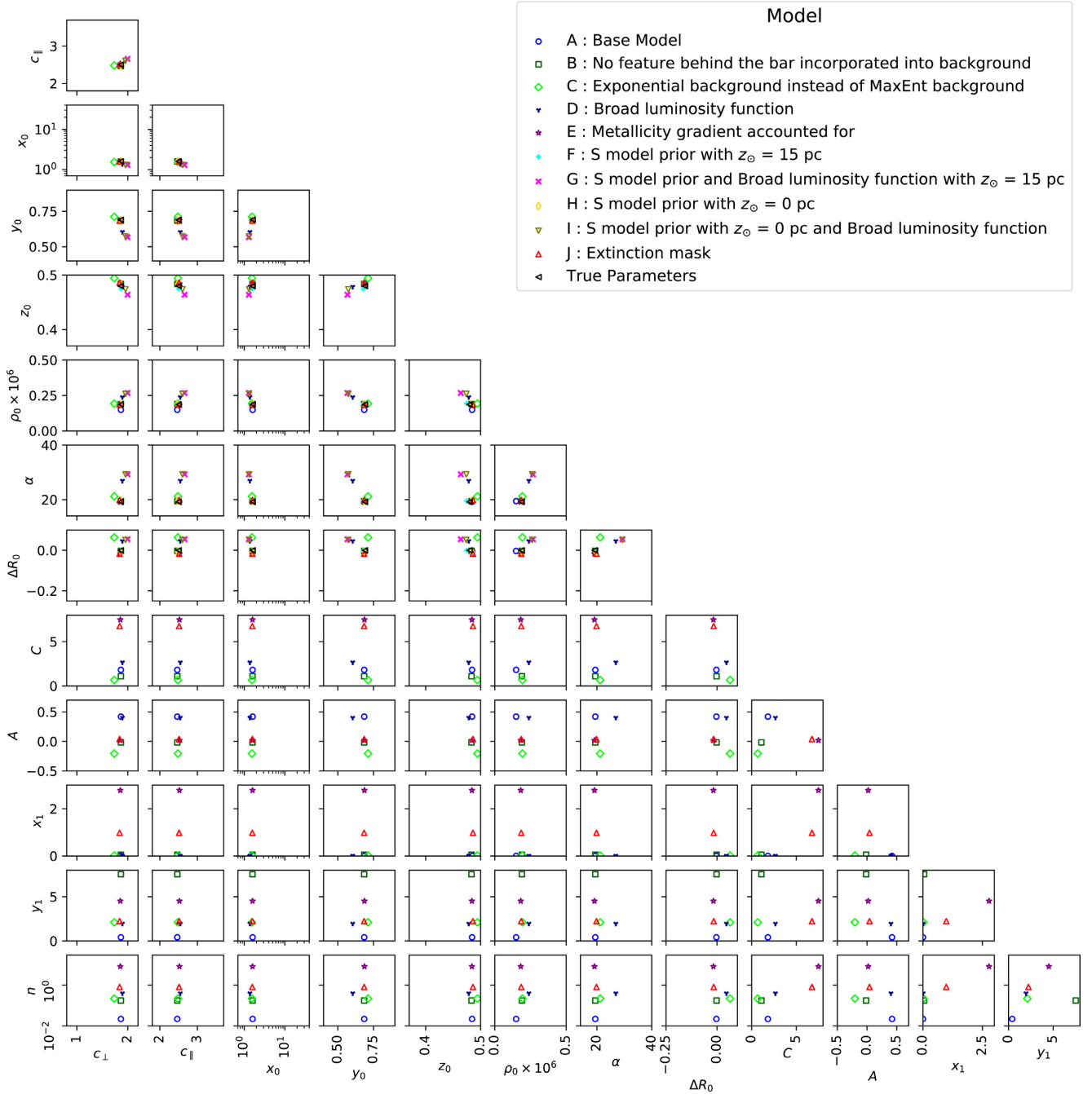
### 6.1.2 Distance to the Galactic centre

As mentioned previously, we associate the Galactic centre with the location of the maximum density of the bulge. In all cases we examined, this maximum bulge density was in the same location for the parametric and non-parametric fit. According to our base non-parametric model, the distance from the Sun to the Galactic centre is 7.9 kpc, where the assumed mean absolute magnitude of the RC is  $\mu_{M_{K_s}, \text{RC}} = -1.53$ . WG13 found the main effect of changing  $\mu_{M_{K_s}, \text{RC}}$  was to change the distance to the Galactic centre. If we had instead used the observed local RC mean magnitude of  $\mu_{M_{K_s}, \text{RC}} = 1.62$  (Chan & Bovy 2019; Hall et al. 2019), then all distances would be increased by a factor of 1.04. With the brighter RC, the distance to the Galactic centre would then be 8.24 kpc, which is consistent with the recent measurement of  $8.18 \pm 0.04$  kpc calculated using parallax observations of Sgr A\* (Gravity Collaboration 2019).

### 6.1.3 Estimating the X-component proportion

The X-component was obtained by setting the 1 in  $(1 + A)$  from the SX model definition in equation (9) to 0. The X-component proportion was then computed by integrating the X-component and SX model over all coordinates and then taking the ratio of them. These ratios are listed in Table 4.

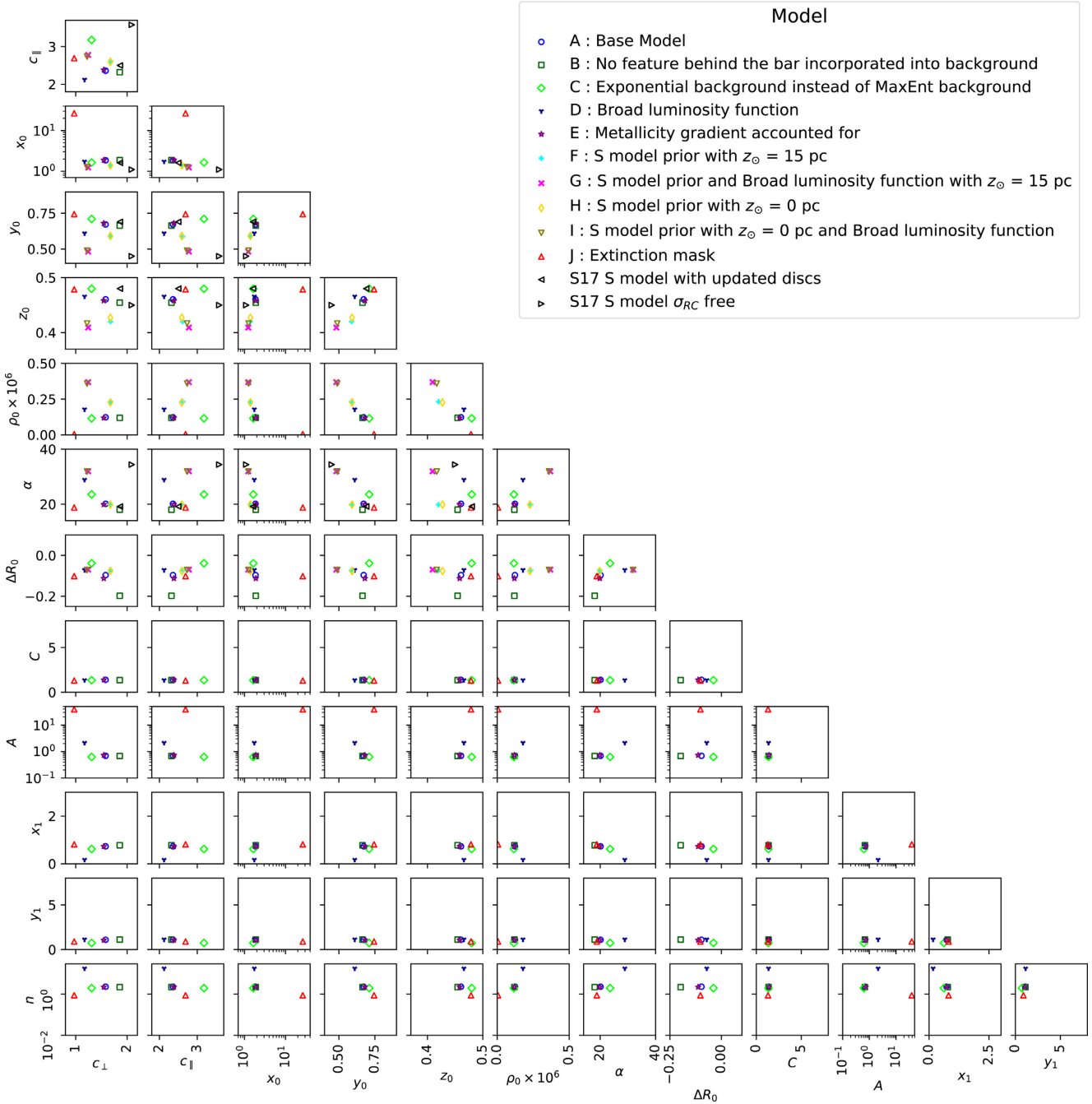
A partial degeneracy in the SX model, due to allowing the X-arm power-law exponent ( $n$ ) to vary, turns up in our extinction mask



**Figure 13.** Pair plot of parametric model parameters fitted for the base case and systematics to simulations. See Table A3. The  $n$  parameter has been plotted on a logarithmic scale.

parametric fit (case J) to the data. The additional density unveiled by the extinction mask depicted in Fig. 17 may be the main driving factor in this behaviour that only showed up in that model case. The result of this is visible in Fig. 14, where the J case is an outlier in the  $A$  and  $n$  parameters. With an exponent,  $n$ , less than 1, the X-arms become very broad. This case is not shown in Fig. 12 because it involves a different amount of data, so the change of likelihood will be on a different scale to that in the other cases. Another case of  $A$  and  $n$  replacing the bulk of the S-component of the SX model is in parametric case A on the simulations. A slice near the edge of the Galactic plane data mask, at 310 pc, is displayed in Fig. 18.

As the parameter  $n$  approaches 0, the perturbation tends towards a constant with a cusp at the X-arm origins from the exponential term. Although this model can appear to have a strong X-component, the fact we have  $n \ll 1$  tells us that this component is near constant, so it is effectively adding to the normalization of the S-component rather than giving an X-shaped perturbation. This result could in principle have come out for any of the simulation cases, so this behaviour is not particular to the A model, just the random model initialization that resulted in a convergence to a model that has the X-component trace the bulge rather than, for example, fall below the mask by having a large X-arm parting factor  $C$ .



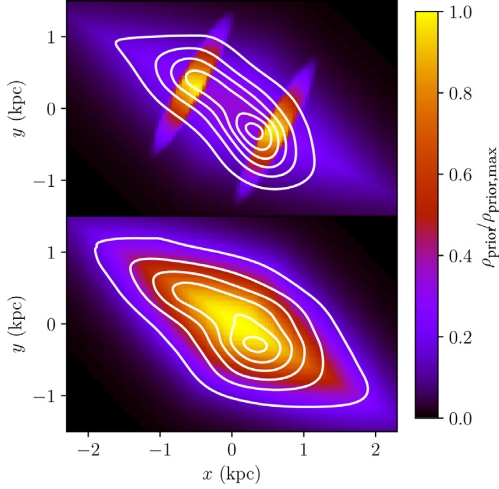
**Figure 14.** Pair plot of parametric model parameters fitted for the base case and systematics on the VVV data. Note that the axis scaling for parameters  $x_0$ ,  $A$ , and  $n$  are logarithmic. See Table A2.

Based on the above arguments, we discard the A case parametric estimate for the simulation and the J case parametric result for the data in Table 4. It follows that our simulation results are consistent with a negligible X-component, which is correct as the model used to generate the simulation had no X-component. Additionally, we can conclude that our parametric fit to the data has the X-component contributing a range of 18–25 percent to the bulge mass. This estimate of the X-bulge component contribution is consistent with that found for the WG13 model by Portail et al. (2015), which was 24 per cent.

#### 6.1.4 Bulge angle

As can be seen from Table A3, our bulge angles with respect to the Sun–Galactic centre line ( $\alpha$ ) for the simulation ranged from  $19.1^\circ$  to  $29.3^\circ$  that encompasses the simulated value of  $\alpha = 19.2^\circ$ . As can be seen from Table A2, our parametric fit of the VVV data had bulge angles in the range of  $18^\circ$ – $32^\circ$ . This is consistent with previous estimates. E.g. WG13 obtained a best fit of  $27^\circ$  and S17 obtained a best fit of  $20^\circ$ . The dependence of the viewing angle on the intrinsic RC luminosity dispersion for triaxial features was





**Figure 15.** SX (top) and S (bottom) parametric density models at  $z = 0.495$  kpc, fitted to the VVV data using the 10 Gyr bulge Parsec derived luminosity function that has been convolved with a Gaussian with  $\sigma = 0.24$  (case D). They are used as the prior models for a non-parametric fit. The broadened luminosity function has driven the X-component to be unnaturally contrasting to the S-component, which necessitates the non-parametric model (the white contours) modulate significantly from the prior density. By contrast, the S-model is still largely visible in the non-parametric solution, with the modulated X-bulge arms visible at  $x = \pm 0.5$ .

observed by Stanek et al. (1997) and S17. As  $\sigma_{RC}$  broadens, the depth of the bar needs to decrease along each line of sight. For a triaxial density, an increase in angle relative to the Sun–Galactic centre position will directly lead to a smaller depth through the bar for each line of sight.

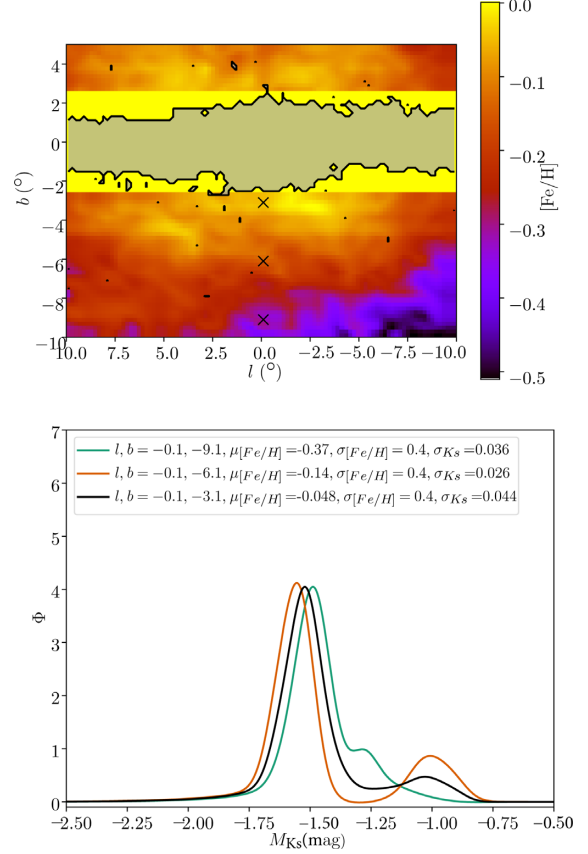
## 6.2 Gamma-ray Galactic centre excess

The work of Macias et al. (2019) found the S-bulge model (denoted by F98S hereafter) from Freudenreich (1998) provided the best fit to the *Fermi* GCE in a template fitting analysis. We created a template from our base parametric model and our non-parametric model fitted to the VVV data for comparison with the quality of the F98S template fit. We assumed that the density of MSPs is spatially correlated with the RC stellar density. The template ( $T$ ) for the *Fermi*–LAT analysis needs to be proportional to the expected flux of the MSPs, so it was constructed using

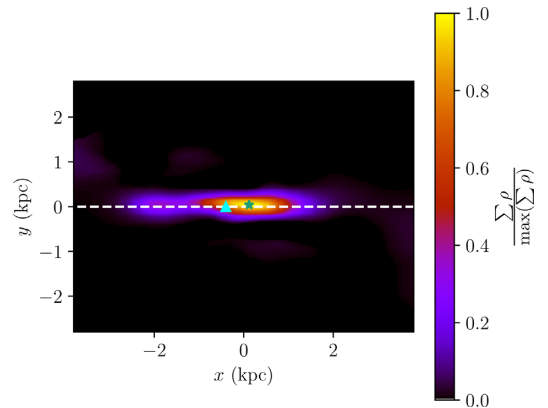
$$T(l, b) = \int_s \rho(s, l, b) ds \quad (14)$$

where  $\rho$  is, as before, the RC + RGBB stellar density of the bulge. Note that an extra factor of  $s^2$  is not necessary as this is the flux, so whilst the number density is increasing as  $s^2$  the observed flux is falling as  $s^2$ . We show a comparison between the F98S template and templates generated from our parametric and non-parametric fits in Fig. 19. Our non-parametric template has a noticeable ‘peanut’-like morphology. This may at first seem in contrast to the X-shaped morphology apparent from Fig. 10, for example. However, in that figure each slice in  $z$  is normalized by the maximum density in that slice. As is well known, when no such normalization is done the bulge has a more peanut-like morphology as can be seen from the third panel of the cross-sections in Fig. 20.

In fitting to the *Fermi*–LAT data, we followed the same method as Macias et al. (2019). The bulge template was fitted simultaneously with the resolved point sources, gas correlated templates, inverse Compton templates (ICS-F98SA50; Porter, Johannesson &



**Figure 16.** *Top:* Mean photometric metallicity map,  $[Fe/H]$  of Gonzalez et al. (2018). Where the map does not have coverage at  $|b| < 2.6$  we assume the fiducial value  $[Fe/H] = 0.0$ . The black crosses indicate the locations of the three luminosity functions plotted in the bottom panel. *Bottom:* The RC + RGBB luminosity functions for a range of fields of view, assuming a metallicity distribution as in the above panel. They have been convolved with a Gaussian with dispersion  $\sigma$ , the photometric uncertainty. In order of increasing metallicity, the mean absolute magnitude of the RC is  $-1.49$ ,  $-1.51$ , and  $-1.52$  mag.



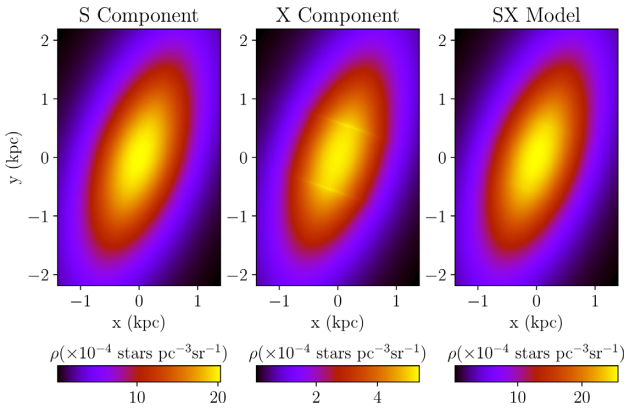
**Figure 17.** Difference between the deconvolved density using a crowding + extinction based mask and a extinction only mask in Cartesian coordinates where  $x$  is aligned with the Sun–Galactic centre line. The density difference has been summed over  $|z| < 1$  kpc]. The white dashed line indicates  $l = 0^\circ$ . The maximum density of the difference (indicated by a green star) is 150 pc behind the maximum density location of the crowding + extinction based mask reconstructed bulge. The cyan triangle is at the expected maximum density location for a population that would have an RC 0.1 mag brighter than our PARSEC derived semi-analytic luminosity function, such as a 5 Gyr old population or a more metal-rich population.

**Table 3.** Total stellar mass estimate for the Galactic bulge for all test cases. A Chabrier IMF was assumed, which gave a remnant fraction of 30.1 per cent. The cases considered are base (A), no behind-the-bar feature subtraction (B), exponential background (C), broad luminosity function (D), metallicity gradient (E), S-model prior with  $Z_{\odot} = 15$  pc (F), S-model prior and broad luminosity function with  $Z_{\odot} = 15$  pc (G), S-model prior with  $Z_{\odot} = 0$  pc (H), S-model prior with  $Z_{\odot} = 0$  pc with a broad luminosity function (I), and extinction mask (J). The mass of the simulated stellar population is  $\text{Mass}_{\text{Bulge}}^{\text{Sim}} = 1.92 \times 10^{10} M_{\odot}$ .

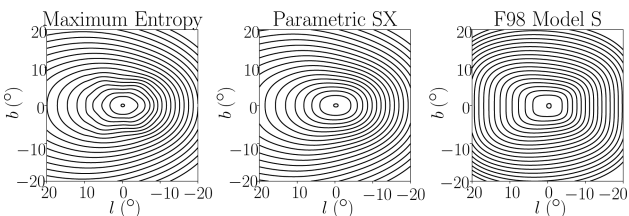
Case	$\text{Mass}_{\text{Bulge}}^{\text{VVV}} (\times 10^{10} M_{\odot})$	$\text{Mass}_{\text{Bulge}}^{\text{Sim}} (\times 10^{10} M_{\odot})$
A	1.64	1.89
B	1.70	1.92
C	1.33	1.84
D	1.61	1.90
E	1.63	1.89
F	1.52	1.91
G	1.58	1.93
H	1.53	1.92
I	1.57	1.93
J	1.71	1.90

**Table 4.** Ratios given by the X-component of each corresponding model integrated in all directions down to a scalar divided by overall integrated SX model, for data and simulation fits.

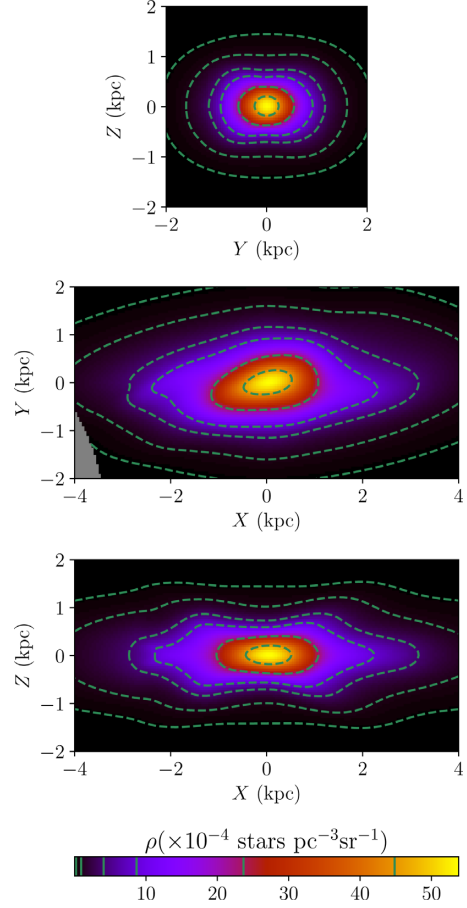
	A	B	C	D	E	J
Data	0.23	0.23	0.18	0.25	0.24	0.92
Simulations	0.20	-0.0062	-0.048	0.012	0.018	0.016



**Figure 18.** Sample slice at  $z = 310$  pc of the parametric model in case A, fitted to simulations. A simple ratio of the X-component to the full SX model can imply there is a significant X-arm component when there is not one. Due to the very small exponent  $n \sim 0.02$ , the X-component has effectively the same shape as the S-component only with small cusps at the origins of the exponential functions.



**Figure 19.** Integrated density,  $T(l, b) = \int \rho(s, l, b) ds$  for the maximum entropy deconvolution, the parametric SX prior density for the deconvolution and the parametric S-model of F98.



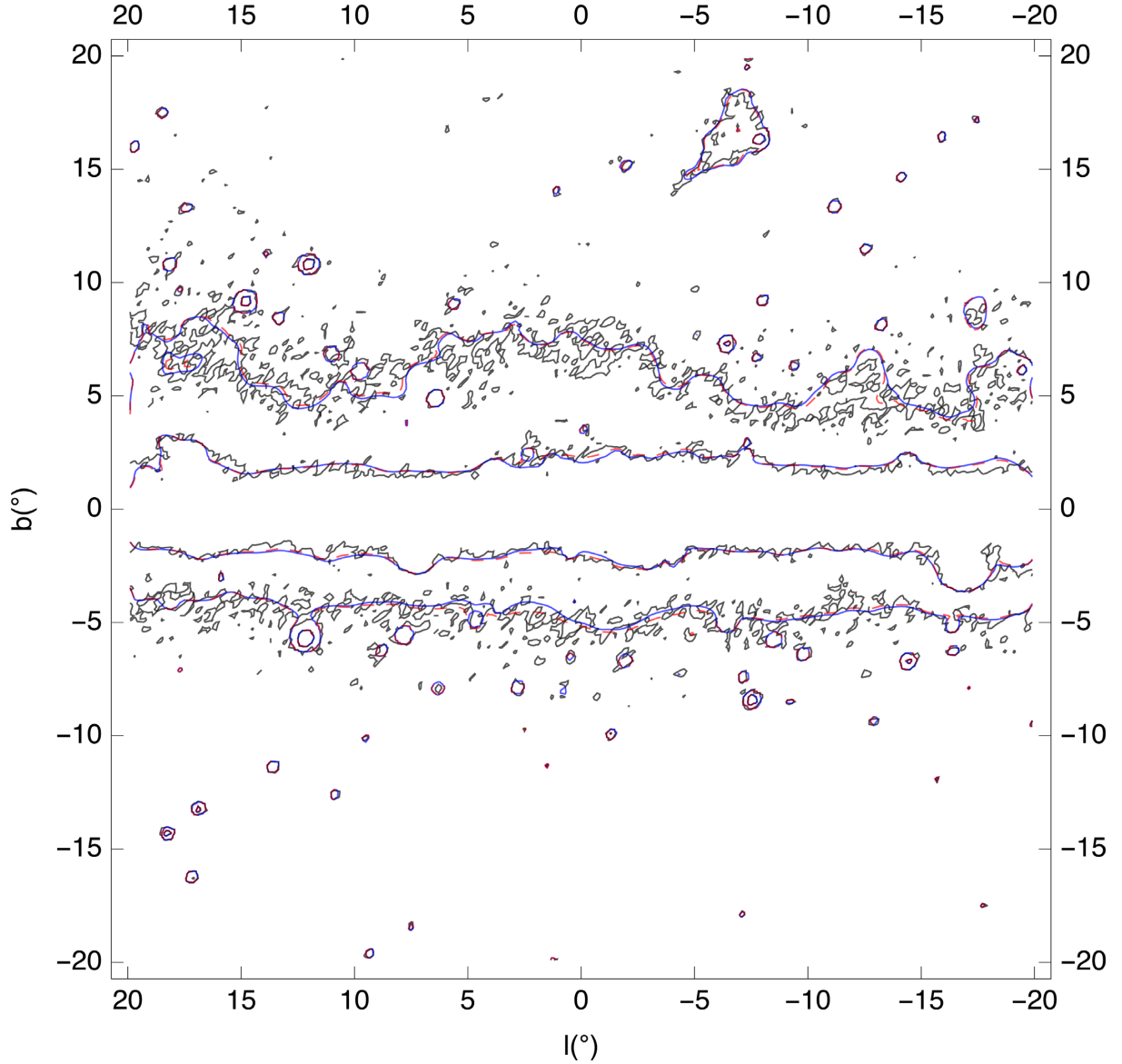
**Figure 20.** Slices at the Galactic centre of the stellar density across different axis slices for our base non-parametric model. The three perpendicular axes are aligned along the bulge angle and centre using  $\alpha$  and  $\Delta R_0$  from our best-fitting parametric model for the base case. Where  $X$  is along the main axis of the bar and  $Z$  is perpendicular to the Galactic plane.

**Table 5.** A comparison of the different bulge templates ability to explain the *Fermi*-LAT GCE. Where for model  $i$ , we list  $\Delta \text{TS}_{\text{Fermi}} = 2 \ln \mathcal{L}_{\text{nonparam}} / \mathcal{L}_i$ .

Model	$\Delta \text{TS}_{\text{Fermi}}$
Non-parametric bulge	0
SX bulge	65
S-bulge	177

Moskalenko 2017), *Fermi*-bubbles templates, and Sun/Moon templates. The unresolved MSP Galactic disc component has been found to have an undetectable contribution (Bartels et al. 2018a), and so we did not include it. The energy range of the photons used in the *Fermi*-LAT analysis was 667 MeV to 158 GeV, distributed over 15 logarithmically spaced energy bins. A  $40^\circ \times 40^\circ$  region around the Galactic centre was used with  $0.5^\circ \times 0.5^\circ$  pixels. This large region of interest was necessary to be able to constrain the background components. Also, no mask was used in the *Fermi*-LAT analysis. This made our non-parametric method of estimating the bulge from the VVV data particularly suitable as it allowed us to obtain an estimate of the bulge morphology over a  $40^\circ \times 40^\circ$  area with no masked regions.

We evaluated the improvement to the fit to the *Fermi*-LAT data by working out  $\text{TS}_{\text{Fermi}} = 2 \log \mathcal{L}_{\text{null}} - 2 \log \mathcal{L}_{\text{bulge}}$  where  $\mathcal{L}_{\text{null}}$  is



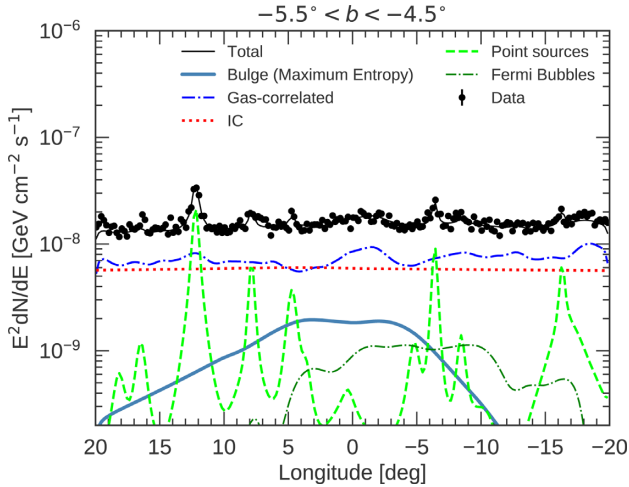
**Figure 21.** Contours of the *Fermi*–LAT data (black), a model without a Galactic bulge (blue), and model with our non-parametric Galactic bulge (red, dashed). The energy range is 1.1–2.8 GeV and the contour levels are 750 and 2000 in units of photons per square degree.

the maximum likelihood with all the above-mentioned templates’ normalizations treated as free parameters in each of the 15 energy bands.  $\mathcal{L}_{\text{bulge}}$  is the maximum likelihood estimate using all the above-mentioned templates and the bulge template where the template normalizations were all fitted simultaneously. As discussed by Macias et al. (2019), a  $\text{TS}_{\text{Fermi}} \geq 34.8$  corresponds to a  $4\sigma$  detection of a new extended source. In Table 5, we list the change in  $\text{TS}_{\text{Fermi}}$  for the different bulge templates<sup>3</sup> we considered. The non-parametric template was preferred by the *Fermi*–LAT data, with  $\Delta\text{TS}_{\text{Fermi}} = 177$  compared to the previous best-fitting template, F98S. A similar value was obtained when using a S-model fitted to the VVV data instead of F98S. Compared to our parametric SX template, our non-parametric template had  $\Delta\text{TS}_{\text{Fermi}} = 65$ . Our  $40^\circ \times 40^\circ$  templates were significantly larger than the area

covered by the VVV data. The extrapolated regions of the templates accounted for around half of the magnitude of the TS values listed in Table 5. Each successive enhancement in our bulge model, from S to SX to non-parametric, resulted in a steady improvement in the quality of fit to the *Fermi* data. This provides further evidence that the GCE traces the stellar content of the Galactic bulge. We found that the inferred gamma-ray energy spectra of the bulge was not very sensitive to the bulge morphology and was similar to previous analysis (Macias et al. 2019).

Contour plots of the data and two alternative models are shown in Fig. 21. The improvement of the fit when the Galactic bulge component is included is particularly noticeable around  $(l, b) = (5^\circ, -5^\circ)$ . The contribution of the Galactic bulge to the *Fermi*–LAT model fit is shown in Fig. 22. The peanut nature for the bulge shape is evident in this figure, even after accounting for the PSF smoothing of the *Fermi*–LAT instrument. Around the  $l = 5^\circ$  region there is a larger ratio of bulge to total signal than in other longitudes

<sup>3</sup>[https://github.com/chrisgordon1/galactic\\_bulge\\_templates](https://github.com/chrisgordon1/galactic_bulge_templates)



**Figure 22.** Spatial distribution of the main model components included in the *Fermi*-LAT fit. The flux profiles in the energy range [1.1, 2.8] GeV are displayed. The black dots represent the data and the continuous black line the total best-fitting model. Other components not shown here (e.g. isotropic, the Sun, Moon, and Loop I) are  $\sim \mathcal{O}(1)$  less bright in the region used to construct the profile.

displayed. This helps in explaining why that area has one of the most noticeable improvements in fitting to the gamma-ray data presented in Fig. 21. Also, this figure shows how typically the bulge component is an order of magnitude smaller than the overall signal. This makes it hard to assign a statistical significance to the difference in  $\Delta TS_{\text{Fermi}}$  values seen in Table 5, as small errors in the larger components could cause one template to be preferred over the other. One alternative method to account for this complication may be to use a maximum entropy non-parametric approach to modulate the larger components as handled by the SkyFACT method (Storm et al. 2017), which also found a preference for a boxy bulge model of the GCE in the *Fermi*-LAT data (Bartels et al. 2018a).

## 7 CONCLUSIONS

We have used a non-parametric method incorporating maximum entropy and smoothness regularization to deconvolve the density distribution of bulge stars in the VVV MW-BULGE-PSFPHOT catalogue. We have also proposed a maximum entropy method for determining the background non-RC + RGBB stars, based on prior estimates using parametric models. Reasonable values for the regularization parameters were found by testing the deconvolution method on a simulated stellar population of the galaxy made of a 10 Gyr old eight-fold symmetric bulge, thin disc, and thick disc. Testing our maximum entropy deconvolution and background fitting method on a simulated population, we were able to nearly perfectly reconstruct the density even in the heavily extinguished and crowded regions that had been masked in the analysis.

Applying the deconvolution method to the VVV data, we found many of the features previously observed in the literature, including the X-shaped bulge from the split RC peak, the dependence of the viewing angle on the intrinsic RC luminosity dispersion, and the feature behind the bar. The  $R_0$  gradient was not clearly seen in the MW-BULGE-PSFPHOT star counts when using the modified Richardson-Lucy deconvolution method assuming eight-fold symmetry.

We performed extensive systematic tests of the maximum entropy deconvolution method to test our assumptions regarding the choice

of background model, metallicity distribution, intrinsic dispersion of the RC, position of the Sun above the Galactic mid-plane, and the deconvolution method itself.

The maximum entropy background was significantly preferred over the widely used exponential background by both the parametric models we fitted and the maximum entropy deconvolution method. Future studies of bulge star counts should be wary using the exponential background, as we have shown it has a tendency to over estimate the background star counts at the bright end of the luminosity function, causing the density of stars to be significantly underestimated at nearby distances.

A broad, unimodal metallicity distribution with spatially varying mean metallicity did not significantly effect the bulge stellar density. A bi-modal metallicity distribution is likely needed, which will become possible as the coverage of bulge spectroscopic surveys grows.

Qualitatively, our results were broadly consistent with the modified Richardson-Lucy deconvolution of WG13. However, we were able to obtain less noisy and higher resolution reconstructions with our maximum entropy method when using the narrow RC dispersion, which recent observations with *Gaia* have favoured (Hall et al. 2019; Chan & Bovy 2019). This resulted in somewhat less dense X-arms. Our method inpainted regions where the data was masked. This meant that we did not need to assume eight-fold symmetry to obtain a reconstruction of the whole bulge area.

From our fits to several different model cases, we found our bulge angle was in the range [18°, 32°], our bulge mass was in the range [ $1.3 \times 10^{10}$ ,  $1.7 \times 10^{10}$ ]  $M_{\odot}$ , and our X-bulge contribution to the bulge was in the range [18, 25] per cent. These are all compatible with other recent bulge estimates using the VVV data.

Our non-parametric method allowed us to inpaint masked regions and smoothly join on to a parametric model outside the region of the VVV data. This made it suitable for providing a template to be used in fitting the *Fermi*-LAT GeV GCE. We found our non-parametric template provided a better fit than the previously implemented parametric S-model (F98S) and our parametric fits to the VVV data. This further supports the unresolved population of MSPs interpretation of the GeV GCE, traced by the Galactic bulge stellar population.

## ACKNOWLEDGEMENTS

We thank Iulia Simion for helpful discussions related to this work. Also, we are grateful to Elena Valenti for giving us early access to the MW-BULGE-PSFPHOT VVV catalogue. OM was supported by World Premier International Research Center Initiative, the Ministry of Education, Culture, Sports, Science and Technology (MEXT) the Government of Japan and by the Japan Society for the Promotion of Science (JSPS) under Grant Numbers KAKENHI-JP17H04836, -JP18H04340, and -JP18H04578. This work was made possible by the use of the Research Compute Cluster facilities at the University of Canterbury. The following software packages were used in this work: ASTROPY, FERMI SCIENCE TOOLS, MATPLOTLIB, NUMPY, PYLBFPS, and SCIPY.

## REFERENCES

- Abazajian K. N., 2011, *J. Cosmol. Astropart. Phys.*, 2011, 010
- Abazajian K. N., Kaplinghat M., 2012, *Phys. Rev. D*, 86, 083511
- Ackermann M. et al., 2017, *ApJ*, 840, 43
- Athanassoula E., Morin S., Wozniak H., Puy D., Pierce M. J., Lombard J., Bosma A., 1990, *MNRAS*, 245, 130

Atwood W. B. et al., 2009, *ApJ*, 697, 1071  
 Bartels R., Storm E., Weniger C., Calore F., 2018a, *Nat. Astron.*, 2, 819  
 Bartels R. T., Edwards T. D. P., Weniger C., 2018b, *MNRAS*, 481, 3966  
 Binney J., Gerhard O. E., Stark A. A., Bally J., Uchida K. I., 1991, *MNRAS*, 252, 210  
 Bissantz N., Gerhard O., 2002, *MNRAS*, 330, 591  
 Bissantz N., Englmaier P., Binney J., Gerhard O., 1997, *MNRAS*, 289, 651  
 Bland-Hawthorn J., Gerhard O., 2016, *ARA&A*, 54, 529  
 Cao L., Mao S., Nataf D., Rattenbury N. J., Gould A., 2013, *MNRAS*, 434, 595  
 Chabrier G., 2003, *PASP*, 115, 763  
 Chan V. C., Bovy J., 2019, *MNRAS*, 493, 4367  
 Cholis I., Hooper D., Linden T., 2015, *J. Cosmol. Astropart. Phys.*, 1506, 043  
 Ciambur B. C., Graham A. W., 2016, *MNRAS*, 459, 1276  
 Clarke J. P., Wegg C., Gerhard O., Smith L. C., Lucas P. W., Wylie S. M., 2019, *MNRAS*, 489, 3519  
 Dwek E. et al., 1995, *ApJ*, 445, 716  
 Freudenreich H. T., 1998, *ApJ*, 492, 495  
 García Pérez A. E. et al., 2018, *ApJ*, 852, 91  
 Gardner E., Debattista V. P., Robin A. C., Vásquez S., Zoccali M., 2014, *MNRAS*, 438, 3275  
 Girardi L., 2016, *ARA&A*, 54, 95  
 Gonzalez O. A. et al., 2018, *MNRAS*, 481, L130  
 Goodenough L., Hooper D., 2009, preprint (arXiv:0910.2998)  
 Gordon C., Macias O., 2013, *Phys. Rev.*, D88, 083521  
 Gravity Collaboration, 2019, *A&A*, 625, L10  
 Hajdu G., Dékány I., Catelan M., Grebel E. K., 2019, preprint (arXiv:1908.06160)  
 Hall O. J. et al., 2019, *MNRAS*, 486, 3569  
 Hooper D., Mohlabeng G., 2016, *J. Cosmol. Astropart. Phys.*, 1603, 049  
 Jaynes E. T., 1957, *Phys. Rev.*, 106, 620  
 Joo S.-J., Lee Y.-W., Chung C., 2017, *ApJ*, 840, 38  
 Laurikainen E., Salo H., Athanassoula E., Bosma A., Herrera-Endoqui M., 2014, *MNRAS*, 444, L80  
 Lee Y.-W., Hong S., Lim D., Chung C., Jang S., Kim J. J., Joo S.-J., 2018, *ApJ*, 862, L8  
 López-Corredoira M., 2016, *A&A*, 593, A66  
 Macias O., Gordon C., Crocker R. M., Coleman B., Paterson D., Horiuchi S., Pohl M., 2018, *Nat. Astron.*, 2, 387  
 Macias O., Horiuchi S., Kaplinghat M., Gordon C., Crocker R. M., Nataf D. M., 2019, *J. Cosmol. Astropart. Phys.*, 2019, 042  
 Maraston C., 1998, *MNRAS*, 300, 872  
 Marigo P. et al., 2017, *ApJ*, 835, 77  
 McWilliam A., Zoccali M., 2010, *ApJ*, 724, 1491  
 Minniti D. et al., 2010, *New Astron.*, 15, 433  
 Nataf D. M., Udalski A., Gould A., Fouqué P., Stanek K. Z., 2010, *ApJ*, 721, L28  
 Paterson D., Coleman B., Gordon C., 2019, preprint (arXiv:1911.04716)  
 Pietrinferni A., Cassisi S., Salaris M., Castelli F., 2004, *ApJ*, 612, 168  
 Ploeg H., Gordon C., Crocker R., Macias O., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 015  
 Portail M., Wegg C., Gerhard O., Martinez-Valpuesta I., 2015, *MNRAS*, 448, 713  
 Porter T. A., Johannesson G., Moskalenko I. V., 2017, *ApJ*, 846, 67  
 Rattenbury N. J., Mao S., Sumi T., Smith M. C., 2007, *MNRAS*, 378, 1064  
 Robin A. C., Reylé C., Derrière S., Picaud S., 2003, *A&A*, 409, 523  
 Robin A. C., Marshall D. J., Schultheis M., Reylé C., 2012, *A&A*, 538, A106  
 Robin A. C., Reylé C., Fliri J., Czekaj M., Robert C. P., Martins A. M. M., 2014, *A&A*, 569, A13  
 Saito R. K., Zoccali M., McWilliam A., Minniti D., Gonzalez O. A., Hill V., 2011, *AJ*, 142, 76

Sanders J. L., Smith L., Evans N. W., Lucas P., 2019, *MNRAS*, 487, 5188  
 Sharma S., Bland-Hawthorn J., Johnston K. V., Binney J., 2011, *ApJ*, 730, 3  
 Simion I. T., Belokurov V., Irwin M., Kuposov S. E., Gonzalez-Fernandez C., Robin A. C., Shen J., Li Z. Y., 2017, *MNRAS*, 471, 4323  
 Skrutskie M. F. et al., 2006, *AJ*, 131, 1163  
 Spera M., Mapelli M., Bressan A., 2015, *MNRAS*, 451, 4086  
 Stanek K. Z., Udalski A., Szymański M., Kałużny J., Kubiak Z. M., Mateo M., Krzemiński W., 1997, *ApJ*, 477, 163  
 Storm E., Weniger C., Calore F., 2017, *J. Cosmol. Astropart. Phys.*, 2017, 022  
 Surot F. et al., 2019, *A&A*, 629, A1  
 Wegg C., Gerhard O., 2013, *MNRAS*, 435, 1874  
 Weiland J. L. et al., 1994, *ApJ*, 425, L81  
 Wilks S. S., 1938, *Ann. Math. Stat.*, 9, 60  
 Zoccali M., Hill V., Lecureur A., Barbuy B., Renzini A., Minniti D., Gómez A., Ortolani S., 2008, *A&A*, 486, 177  
 Zoccali M. et al., 2017, *A&A*, 599, A12

## APPENDIX A: RESULTS TABLES

The best-fitting likelihood values we obtained for our parametric and non-parametric fits are listed in Table A1. The best-fitting parameter values are listed in Tables A2 and A3.

### A1 Deconvolution method systematic

Since our data differ from previous three-dimensional RC bulge studies in its photometry and completeness, we investigated how these changes are reflected in past methods applied to view the VVV RC. Given our semi-analytic formulation of a  $K_s$ -band luminosity function, we compare the results of past methods using different luminosity functions and backgrounds to our maximum entropy non-parametric density model. We continued to use the semi-analytic luminosity function derived in P19 (abbreviated here as the PARSEC luminosity function). We also used the parametric function fitted to Monte Carlo simulations of WG13 (abbreviated as the BaSTI luminosity function). The WG13 luminosity func-

**Table A1.** Minimum values of  $-2\ln\mathcal{L}$  for the parametric and non-parametric models. The base case (A) values of  $(-1.36968, -2.35631, -1.4102015, -2.32068) \times 10^8$  have been subtracted from columns one to four, respectively. The non-base cases considered are no behind-the-bar feature subtraction (B), exponential background (C), broad luminosity function (D), metallicity gradient (E), S-model prior with  $Z_\odot = 15$  pc (F), S-model prior and broad luminosity function with  $Z_\odot = 15$  pc (G), S-model prior with  $Z_\odot = 0$  pc (H), S-model prior with  $Z_\odot = 0$  pc with a broad luminosity function (I), extinction mask (J). Note that as case J has a different amount of data, its  $-2\ln\mathcal{L}$  value cannot be compared directly with the other cases.

Case	VVV Data		Simulation	
	Param.	Non-param.	Param.	Non-Param.
A	0	0	0	0
B	17 086	974	733	307
C	65 507	60 554	55 654	69 758
D	-1793	2917 614	13 797	76 778
E	266	184	109	-1641
F	38 934	241 421	-5523	176 708
G	21 665	209 841	15 475	161 736
H	19723	1361	640	95
I	15107	25589	22740	6252
J	$-2 \times 10^7$	$-4 \times 10^7$	$-2 \times 10^7$	$-3 \times 10^7$

**Table A2.** Parametric SX and S-models fitted to VVV data used as priors in Table A1. The best fits and 68 per cent errors are given for each case on alternating lines.

Label	$c_{\perp}$	$c_{\parallel}$	$x_0$	$y_0$	$z_0$	$\rho_0 \times 10^6$	$\alpha$	$\Delta R_0$	$C$	$A$	$x_1$	$y_1$	$n$
(A) Base case	1.581	2.359	1.853	0.672	0.4605	0.123	20.12	-0.0968	1.386	0.69	0.731	1.090	2.31
	0.008	0.009	0.006	0.001	0.0004	0.002	0.03	0.0009	0.005	0.02	0.004	0.005	0.09
(B) No feature behind the bar incorporated into background	1.856	2.319	1.88	0.664	0.4544	0.119	18.0	-0.198	1.359	0.68	0.781	1.11	2.2
	0.007	0.008	0.02	0.002	0.0007	0.003	0.2	0.001	0.004	0.05	0.007	0.02	0.2
(C) Exponential background instead of MaxEnt background	1.309	3.177	1.641	0.7105	0.4798	0.1158	23.55	-0.0386	1.346	0.6246	0.621	0.734	1.981
	0.001	0.002	0.001	0.0007	0.0003	0.0001	0.002	0.0005	0.002	0.0009	0.001	0.001	0.001
(D) Broad luminosity function	1.172	2.124	1.735	0.610	0.4658	0.1788	28.88	-0.0711	1.356	2.13	0.170	1.135	18.0
	0.007	0.009	0.008	0.002	0.0007	0.0009	0.06	0.0009	0.003	0.04	0.003	0.008	0.4
(E) Metallicity gradient accounted for	1.546	2.383	1.884	0.6802	0.4582	0.1193	19.863	-0.1127	1.389	0.727	0.729	1.057	2.244
	0.002	0.002	0.002	0.0003	0.0002	0.0002	0.001	0.0007	0.001	0.001	0.002	0.001	0.002
(F) S-model prior with $Z_{\odot} = 15$ pc	1.677	2.616	1.3812	0.58753	0.42	0.2322	19.7886	-0.0724	-	-	-	-	-
	0.0003	0.0002	0.0002	0.00012	0.0003	0.0004	0.0003	0.0003	-	-	-	-	-
(G) S-model prior and broad luminosity function with $Z_{\odot} = 15$ pc	1.242	2.779	1.2332	0.4819	0.40921	0.3687	31.945	-0.0698	-	-	-	-	-
	0.001	0.003	0.0013	0.0004	0.00018	0.0005	0.005	0.0008	-	-	-	-	-
(H) S-model prior with $Z_{\odot} = 0$ pc	1.6734	2.592	1.3921	0.5915	0.4271	0.2269	19.8241	-0.0767	-	-	-	-	-
	0.0008	0.003	0.0009	0.0004	0.0002	0.0003	0.0003	0.0008	-	-	-	-	-
(I) S-model prior with $Z_{\odot} = 0$ pc and broad luminosity function	1.221	2.733	1.253	0.4884	0.41672	0.3596	31.851	-0.0712	-	-	-	-	-
	0.003	0.004	0.0012	0.0004	0.00016	0.0004	0.006	0.0006	-	-	-	-	-
(J) Extinction mask	0.970	2.691	26.442	0.7440	0.4786	0.004990	18.768	-0.1018	1.302	38.903	0.815	0.891	0.8855
	0.002	0.001	0.002	0.0007	0.0002	0.000005	0.002	0.0006	0.001	0.002	0.001	0.001	0.0009

**Table A3.** Parametric SX and S-models, fitted to an S-model simulation. The best fits and 68 per cent errors are given for each case on alternating lines.

Label	$c_{\perp}$	$c_{\parallel}$	$x_0$	$y_0$	$z_0$	$\rho_0 \times 10^6$	$\alpha$	$\Delta R_0$	$C$	$A$	$x_1$	$y_1$	$n$
(A) Base case	1.864	2.464	1.608	0.6851	0.4845	0.1492	19.414	-0.0031	1.8136	0.42	0.0003	0.409	0.022
	0.004	0.003	0.001	0.0006	0.0002	0.0007	0.006	0.0003	0.0006	0.01	0.0002	0.005	0.001
(B) No feature behind the bar incorporated into background	1.864	2.467	1.600	0.6846	0.4835	0.1897	19.405	-0.0023	1.092	-0.016	0.050	7.538	0.178
	0.003	0.004	0.001	0.0004	0.0002	0.0003	0.003	0.0006	0.003	0.003	0.001	0.005	0.002
(C) Exponential background instead of MaxEnt background	1.733	2.481	1.545	0.7116	0.4943	0.1932	21.17	0.0638	0.6724	-0.205	0.020	2.10	0.222
	0.004	0.005	0.002	0.0006	0.0003	0.0007	0.02	0.0004	0.0006	0.006	0.002	0.04	0.007
(D) Broad luminosity function	1.893	2.545	1.377	0.6043	0.4785	0.2386	26.90	0.0460	2.659	0.402	0.011	1.954	0.40
	0.008	0.007	0.002	0.0006	0.0004	0.0005	0.03	0.0007	0.002	0.001	0.001	0.007	0.02
(E) Metallicity gradient accounted for	1.852	2.523	1.601	0.6864	0.4843	0.1817	19.10	-0.0178	7.483	0.019	2.779	4.51	8.308
	0.004	0.005	0.001	0.0005	0.0003	0.0003	0.01	0.0008	0.005	0.001	0.008	0.01	0.006
(F) S-model prior with $Z_{\odot} = 15$ pc	1.868	2.506	1.586	0.6790	0.4746	0.1930	19.49	-0.0003	-	-	-	-	-
	0.003	0.004	0.001	0.0004	0.0002	0.0002	0.04	0.0007	-	-	-	-	-
(G) S-model prior and broad luminosity function with $Z_{\odot} = 15$ pc	1.9941	2.6591	1.30221	0.56743	0.4640	0.2677	29.2638	0.0548	-	-	-	-	-
	0.0002	0.0002	0.00008	0.00005	0.0001	0.0002	0.0001	0.0003	-	-	-	-	-
(H) S-model prior with $Z_{\odot} = 0$ pc	1.861	2.476	1.599	0.6841	0.4840	0.1886	19.552	-0.0065	-	-	-	-	-
	0.003	0.003	0.001	0.0005	0.0002	0.0002	0.004	0.0006	-	-	-	-	-
(I) S-model prior with $Z_{\odot} = 0$ pc and Broad luminosity function	1.954	2.604	1.3187	0.5733	0.4740	0.2616	29.2719	0.0514	-	-	-	-	-
	0.001	0.002	0.0006	0.0003	0.0002	0.0002	0.0009	0.0006	-	-	-	-	-
(J) Extinction mask	1.839	2.513	1.582	0.6844	0.4861	0.1851	19.84	-0.0164	6.76	0.041	0.98	2.23	0.82
	0.005	0.006	0.002	0.0006	0.0004	0.0004	0.02	0.0007	0.05	0.003	0.01	0.07	0.03

tion construction involved random draws of star masses from a Salpeter IMF and metallicity from the Baade's window metallicity distribution measured by Zoccali et al. (2008). Then, the  $K_s$  absolute magnitude was obtained from interpolated  $\alpha$  enhanced BaSTI isochrones (Pietrinferni et al. 2004) assuming an age of 10 Gyr. The parametrization of the WG13 BaSTI based luminosity function takes the form of the sum of two Gaussians corresponding to the RC and RGBB with parameters  $\mu_{M_{K_s,RC}} = -1.72$ ,  $\sigma_{RC} = 0.18$ ,  $\mu_{M_{K_s,RGBB}} = -0.91$ ,  $\sigma_{RGBB} = 0.19$ , and relative fraction  $f_{RGBB} = 0.20$  ( $\mu$  and  $\sigma$  taking their typical meanings in a Gaussian distribution). A notable difference here is that the RC dispersion is three times the width of our semi-analytic form, which is approximately 0.06 when fitting a Gaussian to the RC component.

As in WG13, we fitted a background of the form

$$B(K_s) = \exp(a + b(K_s - 13) + c(K_s - 13)^2) \quad (\text{A1})$$

to the magnitude ranges  $11 \leq K_s \leq 11.9$  and  $14.3 \leq K_s \leq 15$  mag for each line of sight. Several adjustments they recommended were

retained for this background fit. Higher extinction and crowding in fields with  $|b| < 2^\circ$  were accommodated by setting the second-order coefficient,  $c$ , to 0 and restricting the upper fitted magnitude range to 14.5 mag. The bright latitude end magnitude range for regions where  $l \geq 5.5^\circ$  was reduced down to  $11 \leq K_s \leq 11.7$ . The star count model for each field of view takes the form of equation (1), converted to the form of a background plus a linear convolution via the transform of line-of-sight distance ( $s$ ) to distance modulus ( $\mu$ ). The luminosity function was convolved with the mean combined photometric and systematic uncertainty for each  $K_s$  along each line of sight to account for their effects. The VVV data was re-discretized into  $\sim 1.5^\circ \times \sim 0.5^\circ$  spatial bins over 0.05 mag  $K_s$  bins. For each line of sight, the density distribution was initialized to a Hann window function over a distance modulus of 11.2–17, renormalized to the observed counts. We then applied the modified Richardson–Lucy procedure of WG13, retaining their stopping criteria, for both the BaSTI and PARSEC luminosity functions. This produced an estimate of the bulge density that depended on  $\mu$ , which we mapped on to a density that depends on  $s$ . We then reprojected the bulge density to Cartesian

form using linear interpolation. For the low-resolution data, step sizes of  $(\Delta x \times \Delta y \times \Delta z) = (0.15 \times 0.1 \times 0.075)$  kpc were used. This simple reprojection only produced a noisy unsymmetrized view of the density model. For a view of the deconvolved bulge density assuming eight-fold symmetry, the appropriate frame needs to be found.

We applied a process of finding the maximally eight-fold symmetric frame following WG13. For each slice in the  $z$ -direction, we carried out a simple grid search over distance to the Galactic centre  $R_0$  and bulge angle  $\alpha$ , in steps of 0.02 kpc and  $0.5^\circ$ . For each  $\alpha$  fixed, we shifted the bulge centre to some value of  $R_0$  and computed the symmetrized density:

$$\bar{\rho}(x, y, z) = \frac{1}{N} [\rho(x, y, z) + \rho(-x, y, z) + 6 \text{ other octants}], \quad (\text{A2})$$

where octant positions without matching densities in the  $(l, b, s)$  projection were ignored from the computation. Parameter  $N$  is the number of octants with non-masked densities. Then, the quantity

$$\frac{1}{N_z} \sum_{z=0.4\text{kpc}}^{1\text{kpc}} \frac{\langle \rho_{\text{rms}} \rangle_z}{\langle \rho \rangle_z} \quad (\text{A3})$$

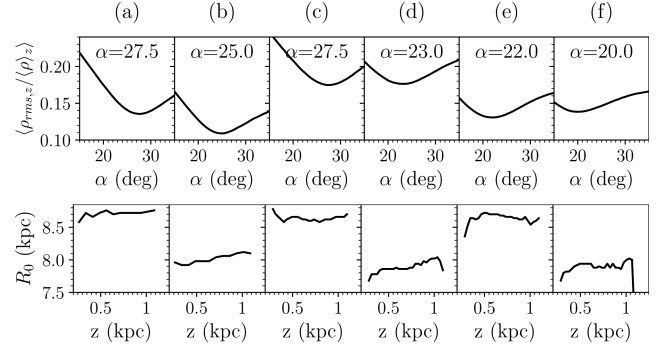
was minimized, where  $N_z$  is the number of slices between 0.4 and 0.8 kpc in the chosen Cartesian grid, so the quantity is comparable between resolutions. The parameter  $\rho_{\text{rms}}$  denotes the root-mean-square deviation between each octant's density in the symmetrization and the average density,  $\bar{\rho}$ , of those points, which was then averaged across all points in each  $z$ -slice.

Rather than minimizing equation (A3) directly,  $\langle \rho_{\text{rms}} \rangle_z / \langle \rho \rangle_z$  was minimized over individual slices of  $z$  for our  $R_0$  grid search. This was an intermediary step in the bulge angle selection process to account for potential magnitude shifts in the model resulting from factors such as metallicity gradients, on top of the required shift in finding the maximally eight-fold symmetric frame.

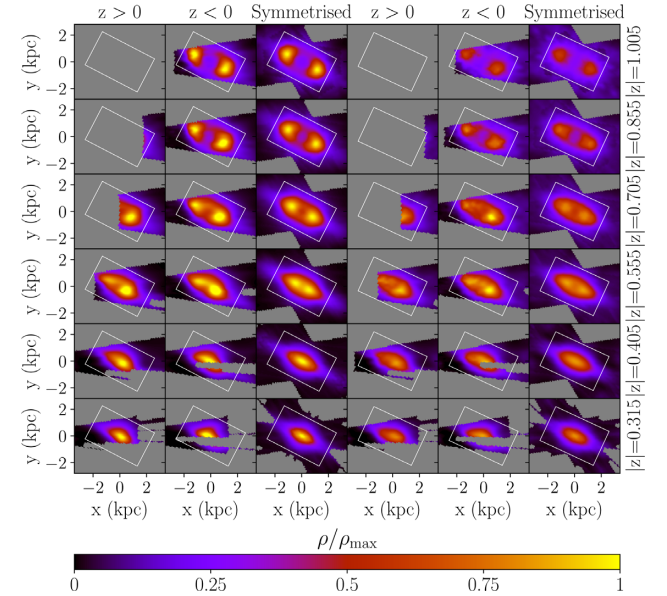
This process was then repeated for  $0.2^\circ \times 0.2^\circ$  spatial bins using our maximum entropy derived background, described in Section 2.4, and Cartesian grid spacing adjusted to  $(\Delta x \times \Delta y \times \Delta z) = (0.04 \times 0.04 \times 0.03)$  kpc, to accommodate the finer data resolution.

In Fig. A1, we recovered the relation observed in S17, in which the broader BaSTI luminosity function results in a larger bulge angle in comparison to the narrower PARSEC luminosity function. We note how the shift in  $R_0$  for each slice to maximize eight-fold symmetry is nearly flat with a constant shift in the BaSTI cases and a much shallower gradient than found by WG13 in our semi-analytic PARSEC luminosity function cases. Figs A2 and A3 show our density deconvolutions on the data using the BaSTI and PARSEC luminosity functions across the two different resolutions we considered. The region used in the maximization of eight-fold symmetry, compatible with WG13, is bounded by a white rectangle. The X-bulge structure and features seen in WG13, such as the near-far RC density asymmetry, are visibly recovered. The  $K_s$ - and  $K_s$ -band RC magnitude widths being observed using *Gaia* DR2 of 0.03–0.09 mag (Hall et al. 2019; Chan & Bovy 2019) are consistent with the PARSEC luminosity function, which is narrower than the BaSTI luminosity function.

In Fig. A4, we show a comparison between the modified Richardson–Lucy deconvolution and our non-parametric method. As can be seen from the profile plot in the rightmost panel, the modified Richardson–Lucy deconvolution with the BaSTI luminosity function has significantly denser X-arms at high  $|z|$ . However, this is primarily due to the use of the BaSTI luminosity function rather

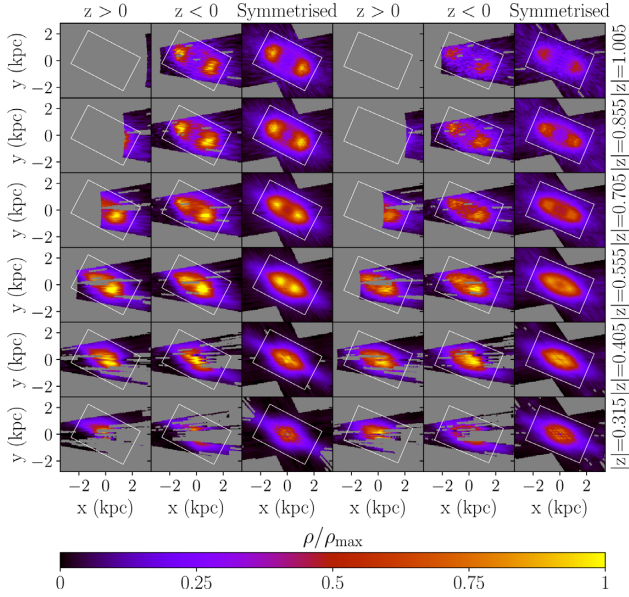


**Figure A1.** Maximally eight-fold symmetric angle (top) and  $R_0$  (bottom) orientation of modified Richardson–Lucy deprojected data. From left to right: (a) BaSTI luminosity function on low-resolution data, (b) PARSEC luminosity function on low-resolution data, (c) BaSTI luminosity function on high-resolution data, (d) PARSEC luminosity function on high-resolution data, (e) BaSTI luminosity function on simulated data, and (f) PARSEC luminosity function on simulated data.

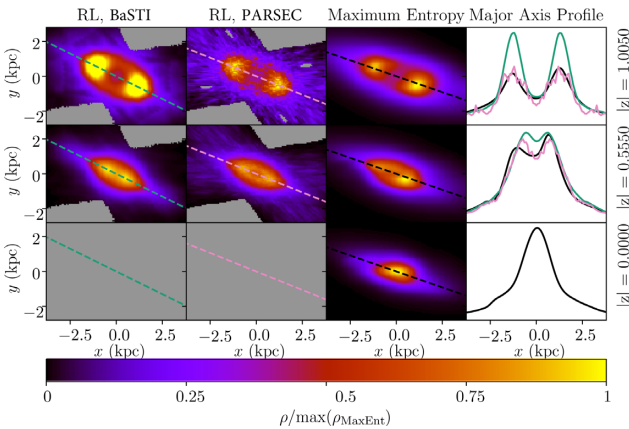


**Figure A2.** Three-dimensional reconstruction of low-resolution VVV data. Columns 1–3 using BaSTI luminosity function and 4–6 using PARSEC luminosity function. Slices of  $|z|$  (measured in kpc) normalized by the maximum of the BaSTI symmetrized model.

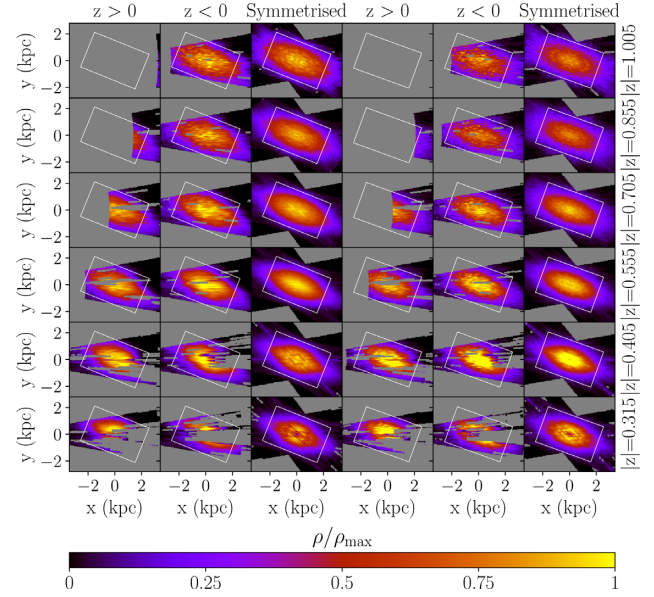
than the PARSEC luminosity function. If the PARSEC luminosity function is used with the modified Richardson–Lucy deconvolution (as in the second column), then the peaks are similar to our non-parametric deconvolution. But, as can be seen from the second column, of the figure, when the PARSEC luminosity function is used with the modified Richardson–Lucy deconvolution, a much noisier reconstruction is obtained even though the low-resolution case is being used. The PARSEC luminosity function has an intrinsic RC dispersion that is more consistent with observations (as mentioned above). It is distinct advantage that our non-parametric model can give non-noisy reconstructions with the narrower PARSEC luminosity function at higher resolution. We checked the method against simulations for the finer resolution to examine possible shortcomings in that regime independently of the actual data.



**Figure A3.** Three-dimensional reconstruction of high-resolution VVV data. Columns 1–3 using BaSTI luminosity function and 4–6 using PARSEC luminosity function. Slices of  $|z|$  are measured in kpc.



**Figure A4.** Comparison between the modified Richardson–Lucy deconvolution and maximum entropy deconvolution. The left column implements the same method and resolution as WG13 except on our updated data set. The middle column is constructed in the same way as the left column except that the narrower PARSEC luminosity function is used instead of the BaSTI luminosity function used by WG13. Density slices have been normalized to the maximum value in the corresponding maximum entropy slice. The green, pink, and black profile plots in the fourth column are along the lines shown in column one, two, and three, respectively. Slices of  $|z|$  are measured in kpc.



**Figure A5.** Three-dimensional reconstruction of S-model simulations. Columns 1–3 using BaSTI luminosity function and 4–6 using the PARSEC luminosity function. Slices of  $|z|$  are measured in kpc.

In Fig. A5, we show the results of the deconvolution and symmetrization of the simulated data with our standard  $0.2^\circ \times 0.2^\circ$  resolution. The bulge angle was effectively recovered using  $0.5^\circ$  steps in a grid search for the PARSEC luminosity function case and a larger angle using the broader BaSTI luminosity function as seen in our earlier results and also by S17. The shift in  $R_0$  is mostly flat across  $z$  slices in both cases with a slight negative gradient in the BaSTI case. Comparing to the gradient in the data fits, it is not apparent whether or not these comparably shallow gradients are spurious. The  $R_0$  eight-fold symmetric maximization on the data results in a very flat shift in  $R_0$  across  $z$  slices between 400 and 800 pc. Above 800 pc, the counts are very low at this resolution, causing excessively noisy features and below 400 pc our mask starts interfering substantially with the symmetrization procedure. We find a negligible gradient using the broader BaSTI-derived luminosity functions. It is not clear within this method how one might interpret the apparent magnitude-shift gradient depending on the broadness of the luminosity function here and how much of it is an artefact of the symmetrization, when there is a persistent asymmetry at odds with the assumption of eight-fold symmetry. Our metallicity distribution systematic in Section 5.4 for comparison, found unimodal corrections driven by observation were negligible.

This paper has been typeset from a  $\text{\TeX}/\text{\LaTeX}$  file prepared by the author.