WP3.11 Cross Matching at LSST Depths

Deliverable 3.11.1

Project Acronym	LUSC-B
Project Title	UK Involvement in the Legacy Survey of Space and Time
Document Number	LUSC-B-08

Submission date	14 August 2020
Version	1.1
Status	For review
Author(s) inc. institutional affiliation	Tim Naylor (Exeter), Tom J. Wilson (Exeter)
Reviewer(s)	Bob Mann (Edinburgh) Manda Banerji (Southampton)

Dissemination level
Public

Version History

Version	Date	Comments, Changes, Status	Authors, Contributors, Reviewers
1.0	14/Aug/20	Original version	Naylor, Wilson
1.1	14/Aug/20	Missing figure added, typo corrected.	Naylor, Wilson

Table of Contents

List of Figures

1 Executive Summary

LSST catalogues will be so deep and crowded that traditional "error circle" catalogue cross-matching techniques will fail. This work package will match the LSST to external datasets using modern cross-matching algorithms providing match tables and software to the community. There are three distinct threads to the work required. First is to use the full power of the astrometric uncertainties of the LSST by using a Bayesian match which assumes a Gaussian distribution of counterpart separations. A large fraction of LSST fields are sufficently crowded, however, that such matching will produce many false negatives because faint, unresolved contaminants in the point spread functions of stars can move their center-of-light by amounts far larger than their formal astrometric uncertainties. So, our second thread will model and correct for these non-Gaussian effects. This in turn can have the effect of weakening the ability to decide between counterpart and field stars. So, our final thread is to use a newly developed algorithm to include the photometric information from the two catalogues, which typically improves the Bayes factors of true counterparts by a factor of 10.

As unresolved contaminants much fainter than the limiting magnitude of the survey can have a significant effect on the astrometry, a model is required for the distribution in magnitude of these stars. Existing galactic models were not thought to be deep enough for LSST, and so deliverable D3.11.1 for this work package was to find or create a model for these stars. Instead we have shown that such a model is not required, by considering the intrinsic noise in the centroid of the stellar image. If we limit ourselves to modelling contaminants which produce a shift larger than 5% of the centroiding noise, existing galactic models will suffice. In solving this problem we have also taken the opportunity to revisit our model for exactly how contaminating stars affect the centroids, specifically the difference between surveys which use centroiding algorithms and those that use point-spread-function fitting. We already had a model for centroiding, and now have one for point-spread-function fitting which we have validated against WISE.

This report therefore fulfils deliverable D3.11.1 in that we now know we have models deep enough to be able to model the effect of faint stars at LSST depths. In addition we also give an outline of the science drivers and the overall plan for this work package, and our progress so far. Further technical details are supplied in an additional journal-style report.

2 Introduction

A significant fraction of LSST science will be based on being able to find reliable cross-matches between objects in the LSST and catalogues in similar or other wavebands. The two main uses of such cross matching will be the provision of multi-wavelength spectral energy distributions and identifying outburst progenitors. Traditionally astronomers have used simple nearest-neighbour matching to find counterparts, choosing the closest star within a given radius, or if there is no star within that radius assigning it as having no counterpart. The fraction of false matches that will be obtained if we persist in using this approach for the LSST is shown by considering a field at l=30°, b=−20°, i.e. significantly removed from the Galactic Plane. Even here the galactic models show that at i=26.7 (LSST full survey depth) roughly 25 percent of objects will randomly have a star within 2" separation. Lower galactic latitudes are worse, and at higher galactic latitudes the density of galaxies in the Hubble Deep Field North is similar to that of stars at l=30°, b=−20°. So, for very large swathes of the sky spurious counterparts will be a serious issue. Yet, without this package simple error-circle matching between LSST positions and the external catalogues is the default position for all programmes we know of.

2.1 Simple Bayesian cross matching.

Simple error-circle matching throws away key information; the closer together stars are, the more likely they are to be counterparts of each other. As a result, modern methods are based on using the astrometric uncertainties and relative densities of counterparts and field stars (at a given magnitude) to assess how likely a given pairing is. This is always done within a Bayesian framework, based on the method discovered by Sutherland & Saunders [1]. The advantage of such methods in crowded fields is that they often show that an association made on the basis of an error circle has a separation which is simply too large to be a credible pairing given the precision of modern astrometry.

2.2 The Astrometric Tug.

Necessary as these Bayesian methods are in crowded fields, they are vulnerable to underestimates of the true astrometric uncertainties. Hence the importance of a paper by the work-package PDRA and PI [2], whose essential message is that undetected faint stars in the PSFs of brighter stars will "tug" the astrometric position of the brighter star, so that the difference between its true and measured positions can be far larger than the nominal astrometric uncertainties. This is illustrated in Figure 1, where the dotted lines show the expected 1D (radial) distribution of counterpart separations for WISE-Gaia matches assuming the counterparts are distributed as a 2D Gaussian with the uncertainties given by the WISE catalogue. The error bars show the true distribution, which has a long tail to large separations. As a result, using the formal Gaussian astrometric uncertainties in the Galactic Plane loses about 50% of the true counterparts. The contaminating stars causing the problem can be very faint; so faint that they would lie below the survey completeness limit even if there were not a bright star there.

Crowding in the Galactic Plane (in terms of stars per PSF) is very similar for LSST single visits and WISE, so astrometric tug will be a serious problem for LSST. We can estimate how widespread in the following way, given the caveat that it depends on what fraction of stars have counterparts, and how source counts extrapolate to faint magnitudes, which we cannot know until we carry out the work proposed here. If we examine a field at at l=30°, b=−5° we are just inside the area of the plane covered at the shallower depth in the current scheduling models, i.e. in one of the lower density areas of the plane survey. Here we estimate from TRILEGAL [6] that

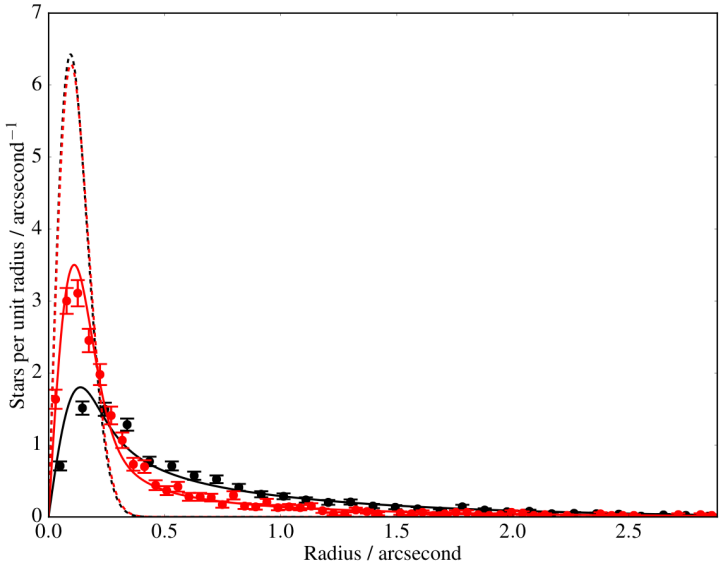

Figure 1: The underestimation of the astrometric uncertainty due to crowding. This figure compares the quoted WISE uncertainties (black dotted line) with the actual distribution of counterparts as a function of distance between star and counterpart in a crowded region (black error bars). The solid black line is a model of the astrometric uncertainties which includes the effect of contamination, which produces the long wing to the distribution. The red coloured data are from a somewhat less-crowded region, and demonstrate how the true astrometric uncertainties in less crowded regions tend towards the uncrowded model.

roughly a third of counterparts will be missed by conventional Bayesian matching (i.e. using Gaussian uncertainties); the rest of the plane and some regions of the Magellanic Clouds will probably be worse. Move to l=30°, b=−10° and the stellar density decreases, but the survey depth increases by a magnitude, yielding 5 to 10 percent missed matches. I.e. significant areas outside the plane will also be badly affected by astrometric tugging. Finally, whilst the focus of this proposal is solving the astrometric tug problem for cross matching, our solution will carry over to both proper motions and parallaxes for faint stars.

2.3 Using the photometric information.

Were we to implement the simple Bayesian matching outlined above, i.e. correctly model the effects of astrometric tugging but take no further action, the long tails in Figure 1 would dramatically increase the number of false positives, limiting the astrophysics we could undertake. Fortunately, there is still more information we can use; the magnitudes of the stars in both catalogues. For example, an object of r=15 is very unlikely to be fainter than K=15 (see also Figure 2). In a second paper [3] we demonstrated how to use this information in a Bayesian framework, and show that typically it improves the Bayes factors for true matches by a factor 10. Unlike an earlier attempt [7], this does not rely on having a physical model for the object, but uses the distribution of magnitudes of counterparts and field stars, and so a single technique will be widely applicable.

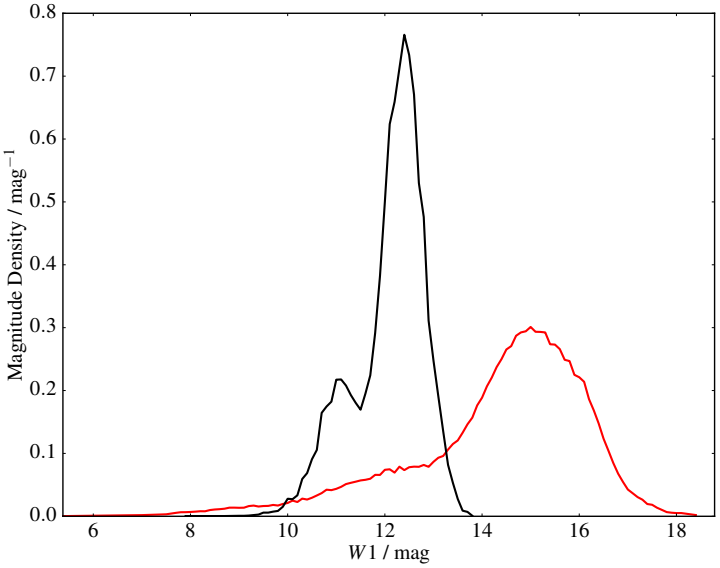

Figure 2: The distribution of magnitudes in the WISE W1 filter for stars that do not have counterparts in Gaia (red line) and stars that have Gaia counterparts between G=14 and 15 (black line). The region of sky used is $131°<l<138°$ and $-3°<b<3°$. It is clear that a star of $W1$=15 close to a star of G=14.5 is (on photometric grounds at least) far less likely to be a counterpart than one of $W1$=12. Using this information, along with the sky density of counterparts allows us to eliminate faint accidental correlations as plausible counterparts. (The two peaks of the counterpart distribution correspond to dwarfs and giants.)

2.4 The final products.

We intend to match to VISTA, VPHAS, WISE and Spitzer, with the ability to extend to EUCLID and other surveys if there are good science cases. For each survey the primary outputs from this work package will be cross-match tables which give, for each LSST source, the best-matching source in the other catalogue, with the probability that it is a match. For each survey we will also give a corresponding table which lists each source and its corresponding LSST best-match. Where there is no good match, the tables will give the probability that the source has no counterpart in the other catalogue.

3 Deliverable 3.11.1

3.1 The astrometric uncertainty function.

One of the key concepts for this work package is the Astrometric Uncertainty Function (AUF). Given the measured position of a star the AUF is the probability density function describing the possible positions of a star. Naively one may think this is a simple two-dimensional Gaussian, but as discussed above it has very broad wings which are caused by the faint, undetected stars which subtly shift the measured position.

3.2 The problem

As one moves to fainter magnitudes, the number of perturbers becomes larger, though each makes a smaller contribution. Initially one may think this means one has to understand the magnitude distribution of stars to the very faintest limits (far below the limiting magnitude of the catalogue). Actually, once one reaches a magnitude where the number of contaminating stars is larger than one, then the effects begin to average out. When this occurs depends on the size of the point-spread function relative to the density of stars, but based on Wilson & Naylor [4] we anticipated this would occur approximately 10 magnitudes below the LSST limit of 27th magnitude. This implied we need a model which reaches 37th magnitude, far below that available from TRILEGAL [6]. Deliverable 3.11.1 was to solve this problem.

3.3 The solution

We realised that although the faint perturbers do change the position of the star, this becomes largely irrelevant once the perturbation is less than the noise in the measurement of the position of the star. We developed a simple analytical model for this effect, which showed that if we wanted un-accounted perturbers to make a no more that 5% contribution to the astrometric noise budget, then for the faintest stars we needed to have a galactic model which reached only five magnitudes below the completeness limit, which TRILEGAL (just) does. The same model then allowed us to show that the situation is less critical for the brighter stars.

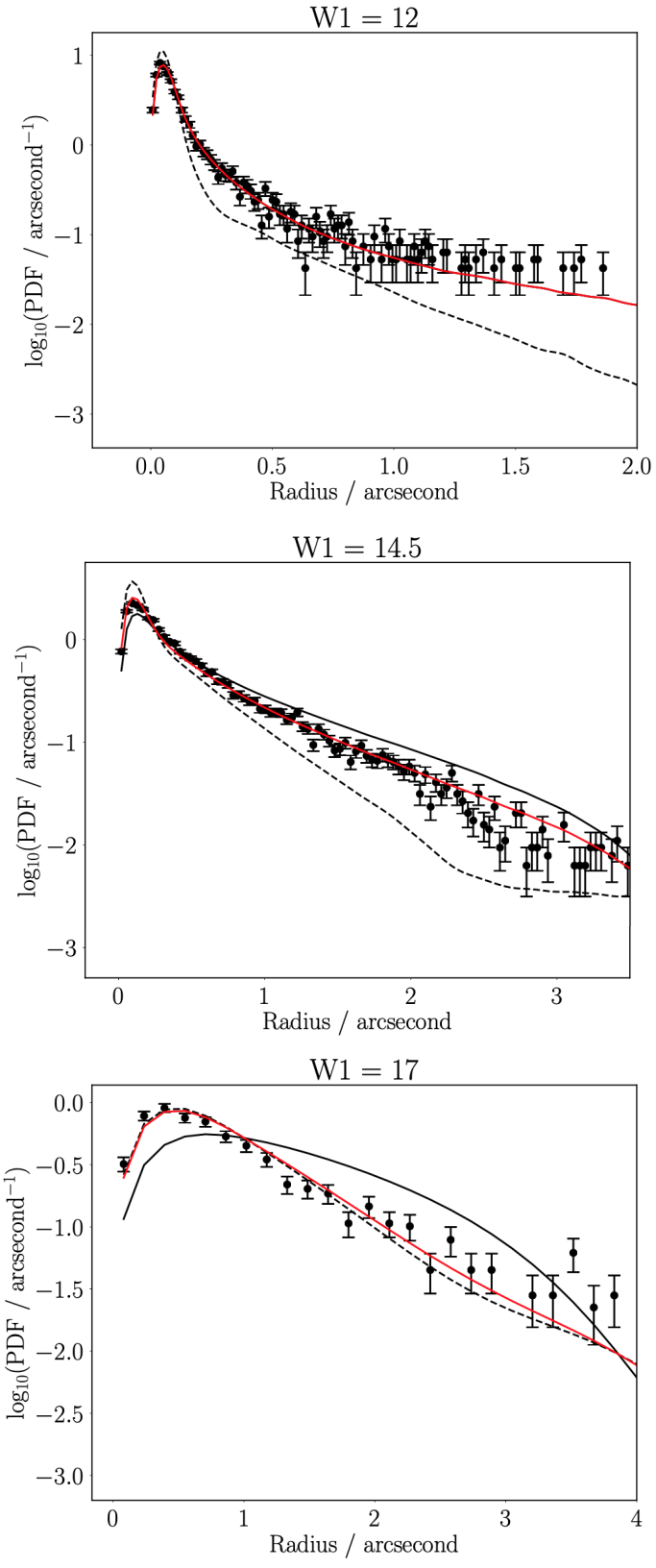

Figure 3: The new model AUFs (red lines) compared with the distribution of positions of *Gaia-WISE* cross-matches for $W1 \simeq 12$, 14.5 and 17 (shown as black error bars). The new AUF is a weighted average of the original flux-weighted centroid calculation of [4] (shown as a solid black line, which is hidden under the red line in the top panel) and the new background-dominated PSF derived centroid (the dashed black line). This figure shows how the final AUF changes from the photon-dominated model of [4] at high fluxes to the background dominated model at low fluxes. It also shows that the original model worked less well at faint magnitudes, but is significantly improved by the addition of the background dominated component.

3.4 Modelling the astrometric uncertainty

One of the inputs into the astrometric uncertainty function is an understanding of how the positions are measured in the original survey, so that the effect of faint perturbing stars can be calculated. In principle one should do this by injecting synthetic sources into the raw images, and then re-running the entire data reduction process for the survey in question. This is not a practical option, and in practice the effect of the data reduction pipeline is best simulated by a simplified model. There are two ways which surveys use to measure positions, either profile fitting or straightforward centroiding. Our understanding of the LSST pipeline is that it will use a combination of both of these, whereas the original version of our software assumed centroiding. This is not a bad assumption even when profile fitting has been used, because the distribution of posterior probabilities are such that changing the threshold at which a star is considered a counterpart from 0.9 to 0.95 (i.e. halving the threshold in the not-counterpart space) rarely changes the sample significantly. That said, having a model for profile fitting would improve our results, not only for the LSST catalogues but also because many of the catalogues we will match to also use profile fitting.

Recalculating the effect of the faint stars on the AUF gave us the opportunity to revisit this issue, and we had the means to test the results as WISE uses profile fitting. We therefore devised a semi-analytical model of the effects of faint perturbing stars on positions measured by profile fitting. This model has two extremes, faint stars where the background light ensures the pixels all have approximately equal uncertainties (the "background limited case" described in, for example [5]) and bright stars where the uncertainty is dominated by photon noise from the star. For the latter it turns out we can use our original aperture model from [4]. The final model is a mixture of these two depending on the brightness of star.

We can test the verisimilitude of the model by measuring the difference in position between WISE sources and their Gaia counterparts for many WISE sources, and then plotting the resulting data as density of counterparts per arcsecond in radius as a function of radius. Figures 3 show the result of comparing such data with the old and new AUF models. The improvement is significant, and will allow us the flexibility to deal with the LSST data whichever position-finding algorithm it uses.

3.5 Modelling the flux contamination

The astrometric perturbation also gives an indication of the flux contamination, since objects distant from their counterparts probably have significant flux within their PSFs from the perturbing star. As is the case for the astrometric perturbation, this photometric perturbation will depend on how the flux is measured. We have undertaken some work which shows that it is likely we can build a model for PSF fitting as well as our current estimates for aperture photometry, and that we will not need a galactic model any deeper than required for the astrometry. However, the photometric work is lower priority, and so we have not proceeded further with this so we can stay on track for delivering D3.11.2 and D3.11.3 on time.

References

[1] Sutherland & Saunders 1992. MNRAS 259 413.

[2] Wilson & Naylor, 2017. MNRAS 468 2517.

[3] Wilson & Naylor, 2018. MNRAS 473 5570.

[4] Wilson & Naylor, 2018. MNRAS, 481, 2148.

[5] Naylor 1998. MNRAS 296 339.

[6] Girardi et al 2012. ASSP 26 165.

[7] Budavári & Szalay, 2008. ApJ 679 301.

MNRAS **000**, 1–10 (2020) Preprint 13 August 2020 Compiled using MNRAS LATEX style file v3.0

Report for D3.11.1. Improvements to the AUF: updated algorithms for background-dominated sources and improved derived flux contaminations.

Tom J. Wilson and Tim Naylor

ABSTRACT

If we are to assess whether or not the positions of two images of stars observed in different surveys are actually the same star we need to understand the uncertainties in their positions. In addition to the uncertainty due to photon noise there is an additional error due to faint stars blended into the image which perturb centroiding. The combination of these sources of uncertainty we refer to as the Astrometric Uncertainty Function (AUF). In principle to estimate this one needs to understand the distribution of contaminating stars to very faint magnitudes, roughly 10 magnitudes fainter than the survey limit, i.e. $g = 37$ for LSST (Wilson & Naylor 2018). This is deeper than the standard Galactic models, and so Deliverable 3.11.1 was to build a contamination model which worked to these depths.

Here we show that we can sidestep this requirement, at least for LSST, by consideration of the noise levels of a given observation. Simple noise considerations show that at the faint end of a survey it is only necessary to consider perturbers up to 5 magnitudes fainter than the central source, for a very conservative estimate of including sources which contribute errors in the centroid equivalent of up to 5% the noise of the primary object in the PSF. Hence we deliver D3.11.1 not as an improved model of the contamination, but as an understanding of at what magnitude difference between a star and its perturber the photon noise in the centroid comes to dominate noise due to perturbers. In addition, we consider both astrometric limits – as the sole consideration of previous efforts – but also the photometric limits, ensuring that simulated results give consistent flux contaminations at all brightnesses. This considers the cumulative expectation number of sources within a given PSF circle down to some Δm limit, and the probability that zero sources are drawn. Limiting this to a sufficiently small fraction we can derive an acceptable perturber relative flux ratio, and show that this, along with the astrometric limit, is within the 5 magnitude faint limit constraint laid out previously.

Additionally, we derive a more accurate model to describe the perturbations of sources in the background-dominated case, crucial for ensuring the AUFs of the very crowded, faintest LSST objects are as well-modelled as possible. This model, and the AUF previously derived for bright objects whose source noise is dominant, are then parameterised as a weighted average, allowing for the smooth transition from the source-dominated to the background-dominated AUF regime.

Finally, we test the application of the new, more accurate model for astrometric perturbations to the issue of photometric contamination. Along with probabilities of cross-matches, modelled using astrometric perturbations, the composite photometric catalogues provided by the LSST:UK DAC should also provided information on the probability of relative flux brightening of sources. We therefore perform tests on the extraction and recovery of blended objects, showing that, at least qualitatively, it is reasonable to construct a completely analagous model to astrometric perturbation when describing photometric contamination.

1 INTRODUCTION

One of the key concepts for this work package is the Astrometric Uncertainty Function (AUF). Given the measured position of a star the AUF is the probability density function describing the possible positions of a star. Naively one may think this is a simple two-dimensional Gaussian, but in practice it can have very broad wings which are caused by faint, undetected stars which subtly shift the measured position. For the faintest stars in any survey, assessing the effect of even fainter stars requires knowledge of the number of stars as a function of magnitude far below the survey limit. In Section 2 we discuss how this led to the first deliverable for this work package, and how we side-stepped the need for it by considering the effects of noise.

Whilst carrying out this work, however, we discovered that the

model of the AUF we had anticipated using was inadequate at faint magnitudes, and so in Section 3 we describe an improved AUF for background-dominated sources, implicitly encoding in its derived quantities the effects of any blended objects unresolved by the data reduction pipeline of the given survey.

Another extension to WP3.11.1 we discovered during this work was the need to ensure that the *photometric* component of the Monte Carlo simulations used to model the effects of crowding were also as accurate as possible. This would then allow us to report robust statistical flux contaminations for the cross-matches reported, which would likely be of interest to some users of the resultant merged catalogue of objects. We additionally report here these photometric considerations. Firstly, we consider the Δm limit down to which to model simulated blended objects within our Monte Carlo PSFs to ensure that we capture, in all realisations, all appropriate flux brightening within each object at the brightness of the central object in Section 4. Additionally, in Section 5 we test the reporting of these flux brightenings for objects whose properties were derived with PSF-based least-squares fitting, but are not in the background-dominated regime.

2 INCLUDING THE EFFECTS OF NOISE IN THE SIMULATING OF PERTURBATION AUF COMPONENTS

A pragmatic consideration we must make when considering the cross-matching of LSST catalogues to external catalogues is that of dynamic range. LSST is expected to reach 5-sigma depths of 27th magnitude, which leads to a potential issue with Galactic source count simulations. Previously, Wilson & Naylor (2018) considered the TRILEGAL (Girardi et al. 2005) Galactic simulations for star counts; however, the faintest stars modelled in these simulations – at least, those publically available – is 32nd magnitude. Wilson & Naylor (2018) also recommended simulating faint perturbers up to 10 magnitudes fainter than the primary object, for source astrometric precision arguments. Sources sufficiently faint are not available through the TRILEGAL simulations, so another approach is needed. Previous PSF simulations do not compare the noise in the astrometric positions with the noise due to faint perturbers, an exercise we undertake in the following section.

2.1 Noise effects on faint object astrometric perturbations

For these purposes we consider a simple system: a primary object, of flux F_p , and a secondary object some distance d offset with flux F_s (or fF_p, where f is the ratio of secondary-to-primary fluxes). When calculating the perturbation component of the AUF, sources are randomly simulated within a given PSF down to some flux ratio and the perturbation position recorded in some way. However, this was previously done in a noiseless environment, with flux-weighted averages considered; the faint limit cutoff was therefore slightly arbitrary. "Real" PSF fitting, however, is done in an environment that contains noise – both from the physical sky objects, and the background sky itself. Thus we can now consider the limit case where the perturbing object cannot be seen below the noise in the primary – or composite – object,

$$F_s \geq B\sqrt{F_p + S + fF_pQ}, \tag{1}$$

where B is a factor, less than one, which dictates how many sigma we need to consider a faint object to, S is the sky flux or source counts within the PSF area in question, and Q is the fraction of the flux from the secondary within the PSF area of the primary (typically computed as the integral of a PSF centered on the secondary over the primary PSF region). For our faint source cases $f \ll 1$ so we can make the assumption that all of the noise comes from the primary object or sky, and thus

$$F_s \geq B\sqrt{F_p + S} = B\sigma_{F_p}. \tag{2}$$

Considering $f \equiv F_s/F_p$ we can divide both sides by F_p to obtain

$$f \geq B\frac{\sigma_{F_p}}{F_p} = B\,\mathrm{SNR}_p^{-1}. \tag{3}$$

As most photometric surveys are defined by their 5-sigma completeness limits, we need to focus on the signal-to-noise ratio (SNR) of the primary object, requiring $\mathrm{SNR}_p \geq 5$. Also considering a swap from relative flux ratio to magnitude offset, $\Delta m = -2.5\log_{10}(f)$, we can also select $B = 0.05$ – or sources at minimum a twentieth the noise of the primary – as this gives $\Delta m = -2.5\log_{10}(0.05/5) = 5$. Since these sources are at the faint limit of the survey, this implies we only need to model the magnitude distribution 5 magnitudes below the completeness limit, which our currently chosen Galatic model allows.

A star 5 magnitudes brighter than the faint limit, assuming it is still sky limited, will have $\mathrm{SNR}_p = 500$, for which we "only" need to consider sources down to $\Delta m = 10$, or again 5 magnitudes below the faint limit. Thus this $B = 0.05$ provides a reasonable range of faint perturbation flux limits as a function of primary flux. It is these Δm values, considering sources up to 5% the noise of the central object when creating perturbation AUFs, that we use in the following section. This value of B is incredibly conservative, and could therefore be increased slightly without significant astrometric effect, for future, even fainter surveys, such as *WFIRST* or any "deep" LSST surveys with completeness limits beyond even 27th magnitude.

We note here – but see Section 4 for more discussion – that this is only the *astrometric* flux ratio limit, and we must also consider the *photometric* limits on simulated contaminants.

3 DERIVING A MORE ACCURATE AUF IN THE SKY BACKGROUND LIMITED CASE

Previously, the algorithm used by Wilson & Naylor (2018) to create fake Monte Carlo PSFs from which the average quantities affecting the systematic offsets of the recorded positions of a given astrophysical object used a relatively simple and naive flux-weighted average scheme. The astrometric *perturbation* of a central source, caused by fainter unresolved objects within a Rayleigh (1880) criterion of 1.185 full-width at half maximums (FWHMs)[1] was simply given by $x_{\mathrm{pertub}} = (\sum_i x_i f_i)/(1 + \sum_i f_i)$ where f is the relative flux between the central object at (0, 0), defined as $f = 1$, and the faint perturber. The photometric *contamination* was quoted as being $\sum_i f_i$ – i.e., it was assumed that if there was 10% additional flux inside the system, either as two sources with 5% relative flux or a single object a tenth the brightness, the resultant photometric catalogue object was 1.1 times too bright.

This simple model should be true for the case of aperture photometry, where any extra flux *is* simply added to the total brightness

[1] Although some surveys, due to non-Gaussian wings in the telescope PSF, may have slightly larger unresolvable regions; *WISE*, for instance, cannot resolve sources within 1.3FWHM – see Cutri et al. (2012), section 4.4c.

of the central object when counting the flux within the aperture. The first moment – used to derive the position of the source – will result in essentially a flux-weighted average position, taking into consideration edge cases where a bright object appears just inside the aperture radius. However, if a more complicated model is used, and sources are fit with a PSF to derive their flux and centroid position then this assumption may not hold. Indeed, the case considered by Wilson & Naylor (2018) is that of *WISE*, which has a robust PSF fitting routine for its derived positions and flux measurements.

3.1 The log-likelihood PSF fitting method

3.1.1 Log-likelihood maximisation

Taking the slightly simplifying assumption that the PSF in question is described by a Gaussian, we can state the fitting process as a minimisation problem, which can equally be thought of as a likelihood maximisation problem. This problem, using a similar notation to that given by Plewa & Sari (2018), equation 1, is

$$\log \mathcal{L} = -\frac{1}{2} \times L \int_{-\infty}^{\infty} [\phi(\mathbf{r}) + f\phi(\mathbf{r}-\mathbf{d}) - (1+\Delta f)\phi(\mathbf{r}-\mathbf{\Delta d})]^2 \, \mathrm{d}^2 r, \tag{4}$$

where we are minimising the square of the differences between two models, one with unity flux (up to a scaling factor of L) at the origin and another with relative flux f at position vector $\mathbf{d}$, and a single model with brightening Δf at perturbation vector $\mathbf{r} - \mathbf{\Delta d}$. Here ϕ is the equation describing the circularly symmetric Gaussian PSF,

$$\phi(\mathbf{r}) = \phi(\mathbf{r}, \sigma_\phi) = \frac{1}{2\pi\sigma_\phi^2} \exp\left(-\frac{1}{2}\frac{\mathbf{r}^2}{\sigma_\phi^2}\right). \tag{5}$$

We can extend this easily to more than one perturber, now fitting for $N+1$ sources with a single composite PSF, given by

$$\log \mathcal{L} = -\frac{1}{2} \times L \int_{-\infty}^{\infty} \left[\phi(\mathbf{r}) + \sum_i f_i \phi(\mathbf{r}-\mathbf{d}_i) - (1+\Delta f)\phi(\mathbf{r}-\mathbf{\Delta d})\right]^2 \mathrm{d}^2 r. \tag{6}$$

Expanding the brackets results in six integrand terms, each of which has a multiplicitive factor in front of the multiple of two ϕ terms, with relative offsets (e.g., the second term would be $-\frac{1}{2} L \sum_i f_i \phi(\mathbf{r})\phi(\mathbf{r}-\mathbf{d})$). These terms, integrated over all space, are the convolution of the two Gaussians, offset by a given vector, and thus each of the six integrals can be analytically computed as six convolutions. As each ϕ term has the same uncertainty σ_ϕ, the resultant Gaussian has an uncertainty of $\sqrt{2}\sigma_\phi$, as Gaussian convolutions result in a Gaussian with a variance the sum of the two composite variances. We therefore, for notation's sake, define a new term

$$\begin{aligned} \psi(\mathbf{r}) = \psi(\mathbf{r}, \sigma_\psi) &= \frac{1}{2\pi\sigma_\psi^2} \exp\left(-\frac{1}{2}\frac{|\mathbf{r}|^2}{\sigma_\psi^2}\right) \\ &\equiv \phi(\mathbf{r}, \sqrt{2}\sigma_\phi) = \frac{1}{4\pi\sigma_\phi^2} \exp\left(-\frac{1}{4}\frac{|\mathbf{r}|^2}{\sigma_\phi^2}\right). \end{aligned} \tag{7}$$

Only three of the resulting convolutions contain terms with Δf or $\mathbf{\Delta d}$ in them – for example, the first term $L\phi(\mathbf{r})$ multiplied by itself, after evaluating the convolution through the integral, becomes $-\frac{1}{2}L^2\psi(\mathbf{0})$. The three terms which are multiplications of two different terms within the square brackets of equation 6 appear twice and thus cancel the factor of a half in the definition of the log-likelihood. Combining all terms, evaluating all convolutions, and dropping constant terms, we therefore have

$$\log \mathcal{L} = L\left[(1+\Delta f)\psi(\mathbf{\Delta d}) + (1+\Delta f)\sum_i f_i \psi(\mathbf{d}_i - \mathbf{\Delta d}) - \frac{1}{2}(1+\Delta f)^2 \psi(\mathbf{0})\right]. \tag{8}$$

For small perturbations by very faint perturbers, we can derive analytic expressions for Δf and Δx (and Δy, to view sky coordinates in a small enough area to consider the region as cartesian) by taking derivatives with respect to those same values and setting to zero. For Δf we have

$$\frac{\partial \log \mathcal{L}}{\partial \Delta f} = L\left[\psi(\mathbf{\Delta d}) + \sum_i f_i \psi(\mathbf{d}_i - \mathbf{\Delta d}) - (1+\Delta f)\psi(\mathbf{0})\right] = 0. \tag{9}$$

Defining $\psi'(\mathbf{x}) \equiv \psi(\mathbf{x})/\psi(\mathbf{0})$, Δf can be solved for as

$$\begin{aligned} &\psi'(\mathbf{\Delta d}) + \sum_i f_i \psi'(\mathbf{d}_i - \mathbf{\Delta d}) - (1+\Delta f) = 0 \\ &\Delta f = \psi'(\mathbf{\Delta d}) - 1 + \sum_i f_i \psi'(\mathbf{d}_i - \mathbf{\Delta d}). \end{aligned} \tag{10}$$

In the limit that $|\mathbf{\Delta d}| \ll 1$, $\psi'(\mathbf{\Delta d}) \to 1$ and $\psi'(\mathbf{d}_i - \mathbf{\Delta d}) \to \psi'(\mathbf{d}_i)$ and thus

$$\Delta f \simeq \sum_i f_i \psi'(\mathbf{d}_i) = \sum_i f_i \exp\left(-\frac{1}{4}\frac{|\mathbf{d}_i|^2}{\sigma_\psi^2}\right), \tag{11}$$

as quoted by Plewa & Sari (2018), equation 3. Similarly we can differentiate with respect to one or other cartesian coordinate,

$$\frac{\partial \log \mathcal{L}}{\partial \Delta x} = L\left[(1+\Delta f)\frac{\Delta x}{2\sigma_\phi^2}\psi(\mathbf{\Delta d}) - (1+\Delta f)\sum_i f_i \frac{x_i - \Delta x}{2\sigma_\phi^2}\psi(\mathbf{d}_i - \mathbf{\Delta d})\right] = 0. \tag{12}$$

Rearranging and using the previous definition of ψ' we get

$$\begin{aligned} (1+\Delta f)\frac{\Delta x}{2\sigma_\phi^2}\psi'(\mathbf{\Delta d}) &= (1+\Delta f)\sum_i f_i \frac{x_i - \Delta x}{2\sigma_\phi^2}\psi'(\mathbf{d}_i - \mathbf{\Delta d}) \\ \Delta x\, \psi'(\mathbf{\Delta d}) &= \sum_i f_i (x_i - \Delta x)\psi'(\mathbf{d}_i - \mathbf{\Delta d}). \end{aligned} \tag{13}$$

Again assuming $|\mathbf{\Delta d}| \ll 1$ and thus $\psi'(\mathbf{\Delta d}) \to 1$, $x_i - \Delta x \to x_i$, and $\psi'(\mathbf{d}_i - \mathbf{\Delta d}) \to \psi'(\mathbf{d}_i)$, we have

$$\Delta x \simeq \sum_i f_i \Delta x_i\, \psi'(\mathbf{d}_i) = \sum_i f_i x_i \exp\left(-\frac{1}{4}\frac{|\mathbf{d}_i|^2}{\sigma_\psi^2}\right). \tag{14}$$

While these approximations of position perturbaton are valid for small f, they are not applicable to cases of approximately equal-brightness contaminants; we therefore require an approximation for the perturbation of these bright objects in the sky background-limited case. The general case for the increase in model flux, equation 10, *can* be used, however, once these perturbation coordinates are computed. We therefore focus now on the positional effects.

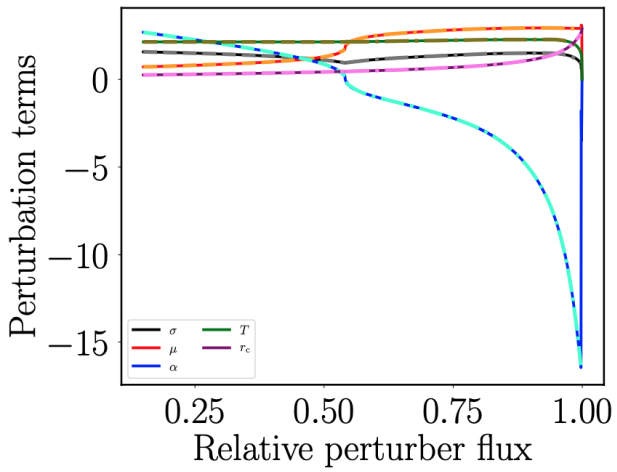

Figure 1. Parameterisation of composite flux-weighted and skew normal model of single object perturbation, Δx. Individual least-squares minimisation parameters for skew normal, and critical cutoff for composite model change (cf. equation 16), are shown in solid lines with both 15th order polynomials overlaid in dash-dot lines. Polynomials do not fit $f > 0.9975$ as parameterisation is poor at these fluxes.

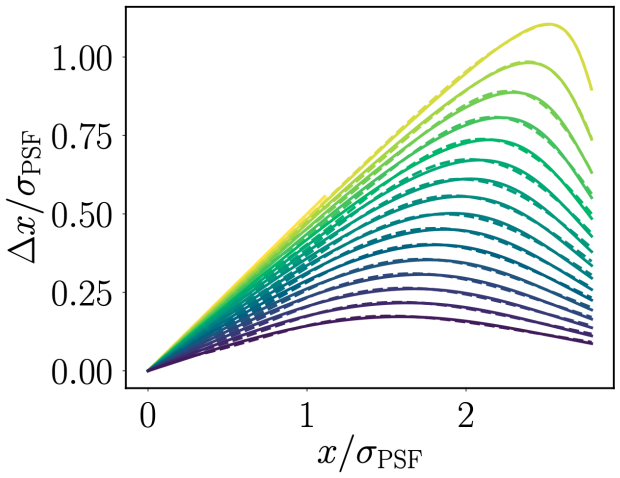

Figure 2. Comparison between empirical and parameterised perturbations, as a function of source offset and flux – represented by the colorscale, in arbitrary scaling and steps from $f \simeq 0.15$ to $f \simeq 0.1$. Empirical perturbations are shown in solid lines, with parameterisation shown as dashed lines.

3.1.2 *Parameterising the background-dominated AUF perturbation*

Unfortunately it is not possible to analytically solve for equation 13. We therefore here lay out an attempt to parameterise the solution as accurately as possible, first splitting the composite solution into one of the vector sum of single-component perturbation solutions, cf. equation 14. For a single perturber, the offset Δx, now a scalar parameter, is a function of two components: first, the relative flux f, and second, the perturber position x (and y). x is also considered a scalar here – and always positive – but its implementation in the computing of AUFs simply takes the absolute value of any given x – positive or negative – to fit the parameterisation and matches the sign of the resulting Δx to that of x.

The paramaterisation of Δx takes two forms: first, at small relative fluxes, we keep the approximate form given by equation 14; second, at larger relative fluxes, the perturbations are modelled as a composite of flux-weighted averages at small source offsets and a skew normal distribution at larger source offsets:

$$\Delta x(x, y, f) = \begin{cases} f\, x \exp\left(-\frac{1}{4}\frac{x^2+y^2}{\sigma_\psi^2}\right) & f < 0.15 \\ \Omega(x, f) & f \geq 0.15, \end{cases} \tag{15}$$

where

$$\Omega(x, f) = \Omega(x, f, \sigma, \mu, \alpha, T, r_c) = \begin{cases} xf/(1+f) & x < r_c \text{ or } f > 0.9975 \\ 2f\frac{T}{\sigma}\lambda\left(\frac{x-\mu}{\sigma}\right)\Lambda\left(\alpha\frac{x-\mu}{\sigma}\right) & x > r_c \text{ and } f \leq 0.9975, \end{cases} \tag{16}$$

and $\lambda(x) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right)$, $\Lambda(x) = \int_{-\infty}^{x}\lambda(t)\,\mathrm{d}t = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{x}}\right)\right]$, r_c is the critical offset at which perturbation stops being modelled as flux-weighted, and σ, μ, α, and T are the scale, location, shape and scaling parameters of the skew normal, respectively. The skew normal distribution parameters and critical cutoff are themselves parameterised as $\sigma(f)$, etc., and fit as two polynomials (of 15th order; a tradeoff between accuracy and computation cost), split by the break in σ at $f = 0.54$. The parameterisation of skew normal parameters is given in Figure 1, and a comparison between empirical and parameterised perturbations is given in Figure 2. Note that extremely large relative fluxes give poorly constrained skew normal parameters and thus we treat what are essentially equal-brightness binary objects as being modelled as flux-weighted averages at all offsets (as shown in equation 16 and Figure 1). This gives maximum relative perturbations from the empirical Δx of $< 8\%$ (for the smallest relative fluxes not fit by the approximate solution), tapering to maximal shifts of $< 1\%$ for the brightest contaminants. In all cases the maximum *absolute* deviation is less than $0.02\sigma_\phi$ – that is, the offsets are always good to, at worst, of order a percent of a PSF sigma.

To test the validity of the vector sum composite model at intermediate fluxes, we then ran simulations of 300,000 realisations of a central object plus an appropriate drawing of contaminants – based on the TRILEGAL number densities in the Galactic plane at roughly $l = 130$, $b = 0$. The "proper" perturbation – the composite fitting, equation 13 – was compared to the vector sum approximation, for steps of *WISE* magnitude $W1$ 11th through 17th. An example of the improvement the parameterisation makes over the flux-weighted centroid regime – where this background-dominated assumption can be assumed to hold – can be seen in Figure 3, for $W1 \simeq 15.5$. It can be seen that there is a much tighter distribution overall between the new vector sum parameterisation and the full perturbation calculation, as compared with the flux-weighted centroid previously used. Additionally, we can compare the Δf computed using several algorithms, as seen in Figure 4. The full Δf, from equation 10, as fit via least-squares minimisation centroids agrees reasonably well with both the vector sum background-dominated perturbation parameterisation and the flux-weighted centroid shifts – neither offset, in Figure 3, was *too* wrong. However, Figure 4 shows how incorrect the naive aperture sum flux increase is as a metric for background-dominated PSF fit flux increases – the reported flux can be wrong by as much as 50-100% of the original central source flux, for heavily contaminated objects, such as those in *WISE* at 15.5th magnitude. It is therefore crucial to understand the algorithms used to create photometric catalogues and correctly apply the appropriate modelling to ensure physically motivated reporting of any modelling of

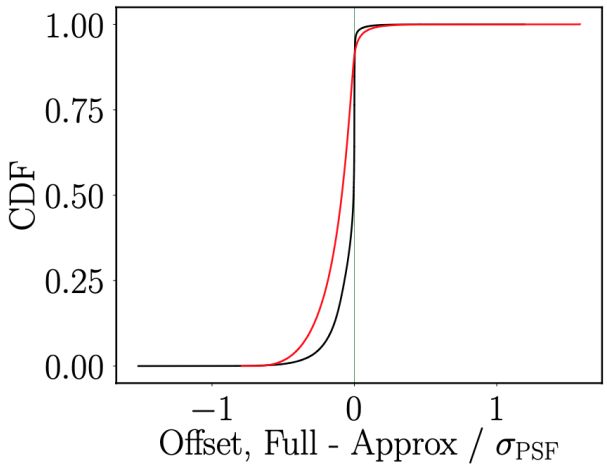

Figure 3. Comparison between the full background-dominated PSF fit centroid shift computation, cf. equation 13, and the vector sum skew normal parameterisation (equation 15), shown in black, and the flux-weighted centroid algorithm previously used by Wilson & Naylor (2018), shown in red.

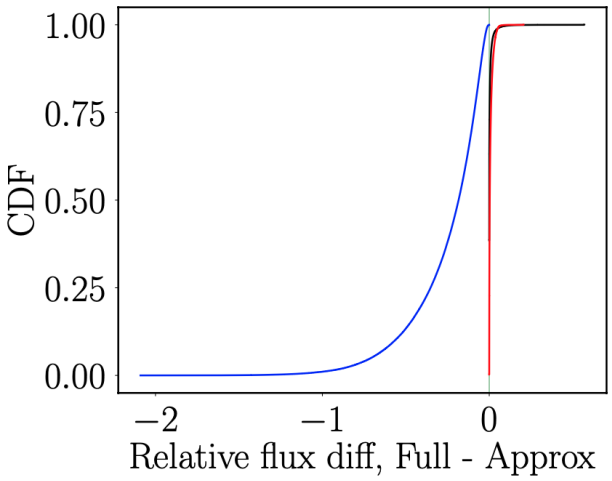

Figure 4. Comparison between the full background-dominated PSF fit flux increase computation and various other algorithms. The difference between the full PSF fit flux increase derived by equation 10 and the increase computed using the perturbations calculated using the approximation given in equation 15 is shown in black; the background-dominated case PSF flux increase using the flux-weighted perturbation centroids is shown in red; and the naive aperture sum flux increase ($\sum_i f_i$) is shown in blue.

flux contamination due to blended objects, such as that performed by Wilson & Naylor (2018).

3.2 Modelling the AUF at all signal-to-noise ratios

3.2.1 Outline solution

Now that we have a model for fitting the centroid perturbation shifts in the limit of background-dominated, approximately constant noise PSF fit objects, we can apply a further parameterisation to model the AUF at all signal-to-noise ratios. To create a probability density function describing the likelihood that two objects, detected in two photometric catalogues, are detections of the same object when one catalogue is subject to crowding, one first needs to derive the likely systematic perturbations of the first object due to said crowding. To achieve this, one must – as described in more detail by Wilson & Naylor (2018) – simulate a PSF, drawing – based on the density of sources as a function of magnitude fainter than the central object – realisations of the perturbers within the bright object, and deriving the centroid shift of the object caused by these blending objects. This is then repeated a large number of times to create a sample of systematic centroid perturbations and corresponding flux contamination levels for a source of given Galactic coordinates (affecting the overall density of sources and extinction reddening, etc.), local source density (to "smooth out" the simulated TRILEGAL source densities), and central object flux. This perturbation AUF component is then convolved with the *statistical* AUF component, the original – and often assumed only – source of position uncertainty caused by noise in the detection image.

3.2.2 Testing with real data

A good dataset to test how well this will work for faint *LSST* sources is *WISE*, because although *WISE* is a relatively shallow survey, its large point-spread function means that it suffers similar crowding to *LSST*. Wilson & Naylor (2018) previously used a flux-weighted centroid algorithm for computing the AUFs of sources, motivated primarily by relatively bright objects ($W1 \simeq 13$) in the *WISE* survey. However, this model is clearly superceded by the physically motivated fitting algorithm laid out in Section 3.1.1; this improvement can be seen in Figure 5, where the background-dominated PSF centroiding algorithm (dashed line) fits the *Gaia*-*WISE* cross-matches of $W1 \simeq 17$ much better than the flux-weighted centroiding (solid line)[2]. The exact opposite is true at $W1 \simeq 12$ in Figure 6, where the flux-weighted centroid calculated AUF is now the better fit.

We parameterise the "hand-off" between the two noise regimes – star dominated and background sky dominated – with a simple linear slope. At each *WISE* magnitude, 11th through 17th, we fit the least-squares minimisation of the weighted average of the two AUFs to the data, as

$$y = H\,\mathrm{AUF}_{\rm fw} + (1 - H)\mathrm{AUF}_{\rm bd} \tag{17}$$

for "flux-weighted" and "background-dominated" AUFs. We then parameterise H as a function of magnitude as a linear slope, capped at $0 \leq H \leq 1$. The resulting composite AUFs are seen in Figures 5 and 6 as solid red lines, but also an "intermediate" SNR regime at $W1 \simeq 14.5$ in Figure 7, where it can be seen that neither AUF can explain the data by itself, but a simple weighted average of the two explains the cross-matches with high accuracy.

We initially explored the possibility that the mis-match between the physically motivated AUFs and the cross-matches was due to the combining of multiple "groups" of data into a single dataset. However, in all cases the *WISE* data were drawn from a consistent SNR regime, whether measured by pure SNR, background flux counts, or star-to-background flux ratio. Thus, unable

[2] The construction of these diagrams is described in detail by Wilson & Naylor (2018). One can compare the *WISE* measured position of a source with its true position by cross-matching it with *Gaia* because the *Gaia* uncertainties are so small. Hence for a large sample of *Gaia*-*WISE* cross-matches the distribution of actual *WISE* positions with respect to the measured one is given by the density of *Gaia* counterparts at a given radius. This is the error bars marked in these diagrams.

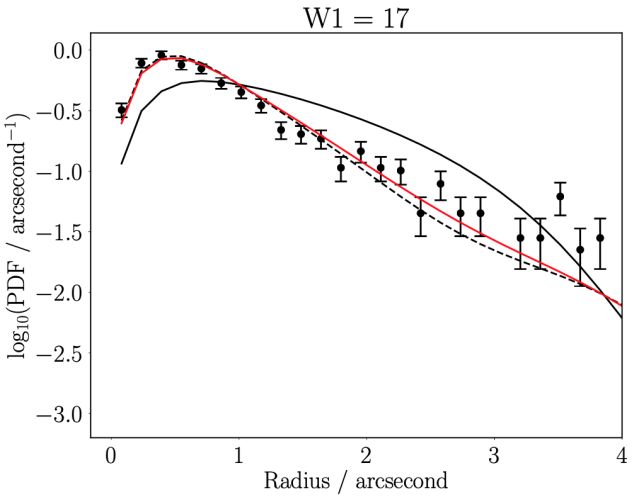

Figure 5. Model AUFs compared with *Gaia-WISE* cross-matches of $W1 \simeq 17$ (shown as black errorbars). Solid black line is AUF as derived using the Wilson & Naylor (2018) flux-weighted centroid calculation, and the dashed black line is the new, background-dominated PSF derived centroid shift laid out in Section 3.1.1. Red line is parameterisation of weighted average of two PDFs.

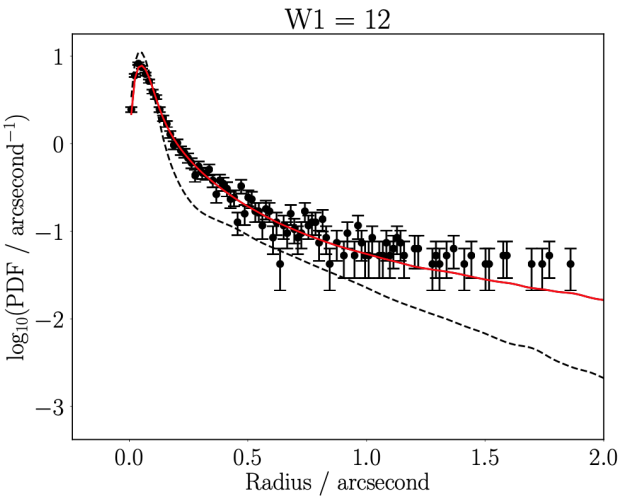

Figure 6. Model AUFs compared with *Gaia-WISE* cross-matches of $W1 \simeq 12$ (shown as black errorbars). Solid black line is AUF as derived using the Wilson & Naylor (2018) flux-weighted centroid calculation, and the dashed black line is the new, background-dominated PSF derived centroid shift laid out in Section 3.1.1. Red line is parameterisation of weighted average of two PDFs. Note that the black line in this figure is entirely beneath the red line as the weighted average AUF is entirely weighted towards the flux-weighted AUF ($H = 1$, equation 17).

to explain the composite nature of the cross-match distributions as physically different groups of objects incorrectly merged into a single distribution, the most likely explanation is simply that those sources at intermediate brightnesses, $W1 \simeq 14.5$ or thereabouts, are in a transition from star- to background-dominated noise. This hypothesis bears out under examination of the star-to-sky flux ratio, which is of order 10 at $W1 = 12$, of order 2-3 at $W1 = 14.5$, and less than 1 by $W1 = 17$.

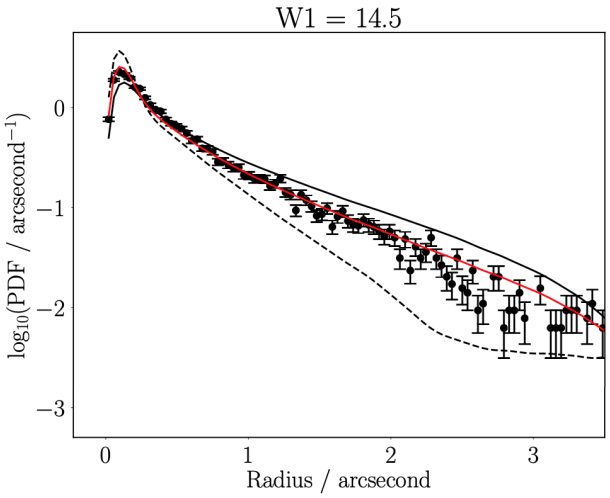

Figure 7. Model AUFs compared with *Gaia-WISE* cross-matches of $W1 \simeq 12$ (shown as black errorbars). Solid black line is AUF as derived using the Wilson & Naylor (2018) flux-weighted centroid calculation, and the dashed black line is the new, background-dominated PSF derived centroid shift laid out in Section 3.1.1. Red line is parameterisation of weighted average of two PDFs.

4 THE PHOTOMETRIC LIMIT OF THE PERTURBATION AUF

The original algorithm for deriving the additional components of the AUF – aside from the original, intrinsic, noise-based Gaussian component – used by Wilson & Naylor (2018) called for a $\Delta m = 10$ magnitude limit for the perturbation-specific aspect. In this case, this limit was a simple one, based roughly on perturbations using a flux-weighted centroid, being that of the largest "small" offset caused by this faint object. For $\Delta m = 10$ we have $f = 0.0001 - f$ being the relative flux ratio between perturber and central source – and flux-weighted perturbations at most of order 0.001", an order of magnitude below the centroiding precision of bright *WISE* objects.

However, as discussed in Section 2, we can now use a magnitude-based Δm cut, based on the *individual* SNR of an object, providing a dynamic Δm limit. As this calculation only took into account the effects on the astrometry of the bright, central object we additionally need to consider whether the Δm limit would result in a complete evaluation of the flux contamination of the objects. For this, we can turn to the number of objects simulated within a given Monte Carlo realisation of the PSF.

4.1 PSF Realisation Derivation

As described by Wilson & Naylor (2018), the simulations for creating the statistical distribution of perturbations of a bright source's position – and contamination of its flux – involve the drawing of objects and placing them randomly near to the primary source. We then assume that, on a statistical level, all of the flux that will affect the contamination level quoted has been simulated if we define our Δm such that essentially all simulations contain at least one perturber. To compute this, we again assume our simulations are modelled as per Wilson & Naylor (2018), in which small magnitude offsets are stepped through for secondary perturbers. In each small bin m_i to $m_i + \mathrm{d}m$ there is a given number density of objects (either a galaxy source rate or TRILEGAL Galaxy star count rate, e.g., cf. figure

5 of Wilson & Naylor 2018), converted to an expected number of source within an area of the PSF.

From this expectation count λ_i – for the ith bin – a given source number is drawn from a Poissonian distribution. We can therefore compute the expected number of sources, and cumulative distribution function (CDF), for the total number of sources down to a given Δm, given by the convolution of each individual Poissonian distribution for each small magnitude bin. Using the notation $\sum_{i=1}^{n} X_i = Y$, where X_i and Y are n independent random variables and the resultant convolution distribution respectively, we get

$$\sum_{i=1}^{n} \mathrm{Poisson}(\lambda_i) = \mathrm{Poisson}\left(\sum_{i=1}^{n} \lambda_i\right) \equiv \mathrm{Poisson}(\lambda), \tag{18}$$

where

$$\mathrm{Poisson}(\lambda) = P(X = k; \lambda) = \frac{\lambda^k \exp(-\lambda)}{k!} \tag{19}$$

with k the number of objects drawn. Thus the sum of n values, drawn from Poissonian distributions is, effectively, itself a drawing from a Poissonian distribution with the expectation value the sum of each individual expectation values.

We wish to find the Δm to simulate down to – or, now, the number of small magnitude bins we need to drawn from – at which we get some small fraction of realisations with zero extra sources drawn. Now we can use a Poissonian distribution with its expectation value the sum of each individual expectation value, and draw the CDF for no objects,

$$P(X \leq k; \lambda) = \frac{\Gamma(\lfloor k+1 \rfloor, \lambda)}{\lfloor k \rfloor !} \equiv \exp(-\lambda) \sum_{i=0}^{\lfloor k \rfloor} \frac{\lambda^i}{i!}, \tag{20}$$

where $\lfloor k \rfloor$ is the "floor" of k (i.e., the largest integer no larger than k), and thus our given probability, for zero objects, is

$$P(X \leq 0; \lambda) = \Gamma(1, \lambda) \equiv \exp(-\lambda). \tag{21}$$

4.2 *WISE* Simulations

We need to calculate $P(X \leq 0; \lambda) = y$, where y is some small fraction, here given as $y = 0.01$ – a 1% chance of realising a PSF containing no objects. Thus, for a given source density of potential contaminating sources, we can compute the *photometric* contamination limit, shown, along with the original astrometric limit, in Figure 8. This is simply the sum of all individual magnitude bin λ_i values (the number density of objects times the bin width times the PSF area) below the m of the bright, central object until $P(X \leq 0; \lambda) \leq 0.01$. For $k = 0$ and $y = 0.01$ this simplifies slightly to asking the Δm bin at which $\lambda \geq -\ln(0.01) \simeq 4.6$.

The importance of this calculation is highlighted in Figures 9 and 10. These figures show the cumulative distribution of the derived flux contaminations of a sample of *WISE* objects at two magnitudes – one faint, one bright. Plotted are the $\Delta m = 10$ previous limit, for reference, along with a $\Delta m = 15$ case, representing a "complete" limit at all brightnesses. Also plotted are the astrometric Δm derived previously, and the new photometric limit.

It can be seen in Figure 9 that neither the astrometric limit, nor the original Δm, capture the tail in the distribution of fluxes from sources of $W1 = 9$. These objects are bright, and thus not subject to significant crowding at relative flux ratios that contribute to the overall quoted flux; however, the astrometric limit causes two thirds of cases to be quoted as having zero additional flux. While the individual sources are at $\Delta f \simeq 0.001$, we can ensure that

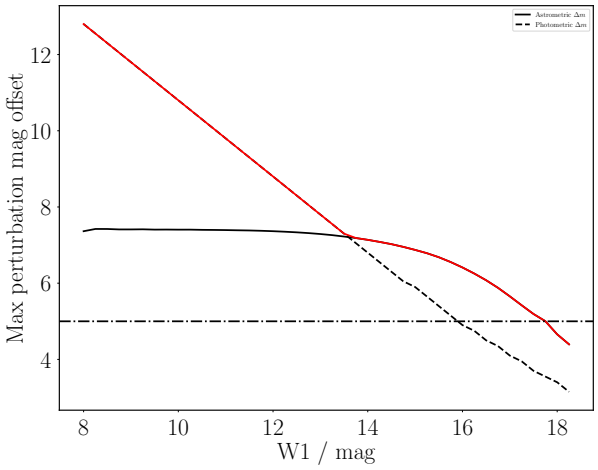

Figure 8. Magnitude offsets for considering blended objects to, for both astrometric and photometric considerations, for *WISE* sources at $l = 130$, $b = 0$. The Δm for astrometric completeness – the limit at which the secondary object is 5% the noise of the primary object – is shown in solid black. The photometric magnitude offset – the magnitude offset down to which sources must be drawn for 99% of realisations to contain at least one source – is shown in dashed black. The red line shows the maximum of the two; and the dash-dot black line shows $\Delta m = 5$, important for LSST using TRILEGAL simulations.

our distribution of flux brightenings is robust across all statistics by increasing Δm to 10 or even 12, with little computational cost. This then ensures we track our Δf flux brightenings down to 10^{-5}, providing the full statistical distribution of potential contamination levels, which could be of use.

Figure 10 shows a different story, however. At these fainter magnitudes, the photometric limit – defined as the point at which we draw zero sources no more than 1% of the time – is achieved much quicker, as these sources are much more subject to crowding. Here, the astrometric limit begins to dominate, and we can stop worrying about the photometry, as it can be assumed to be complete at the astrometric limit. Thus, as Figure 8 shows, the photometric limit matters more at bright magnitudes than the astrometric limit, as shown by the handoff between the two at $W1 \simeq 14$.

We can therefore now derive a Δm which ensures that both the astrometric perturbations and photometric contamination of simulated objects are as accurate as possible. This then allows for robust secondary parameters, such as flux brightening of individual sources subject to crowding in a given photometric catalogue, to be quoted along with likely cross-matches to secondary catalogues. In addition, this method can adjust the acceptable fraction of PSF realisations with zero perturbers y, analagous to the $B = 0.05$ result for astrometric SNR considerations, allowing for flexibility in the derivation of the photometric Δm.

5 FLUX CONTAMINATION COMPUTATIONS AT ALL SIGNAL-TO-NOISE RATIOS

Previously, we derived a new method for computing the astrometric perturbations of sources in the background-dominated, PSF-fit regime, applicable to faint LSST objects. This new algorithm was combined with the previously used flux-weighted centroids in a magnitude-based weighted average scheme (see Figure 7, e.g.).

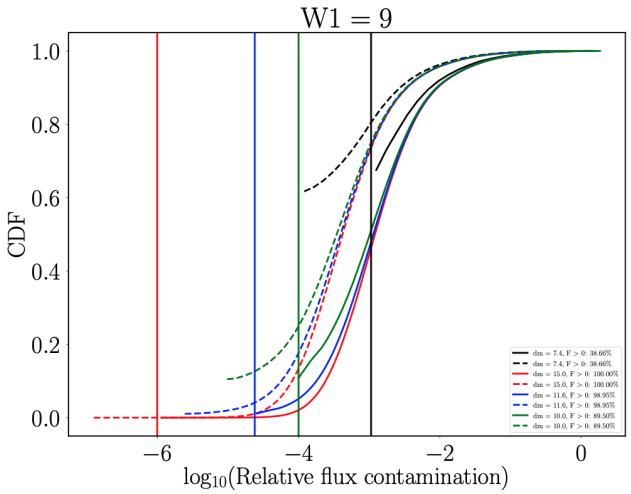

Figure 9. Cumulative distributions of derived flux contaminations from Monte Carlo simulations of blended sources within PSFs, assuming a *WISE* source of $W1 = 9$. Various Δm limits are shown: the previously computed astrometric limit (black line), 15 (red line), the photometric limit (blue line), and 10 (green line). Additionally, the flux contaminations from a simple flux-weighted centroid (solid lines) and the log-likelihood maximisation method (dashed lines) are shown. Vertical lines show the lower limit on f, based on $f = -2.5 \log \Delta m$. CDFs include sources for which zero objects were realised, which is why some CDFs do not begin at zero (i.e., if 40% of objects are "pure", the CDF would start at 0.4).

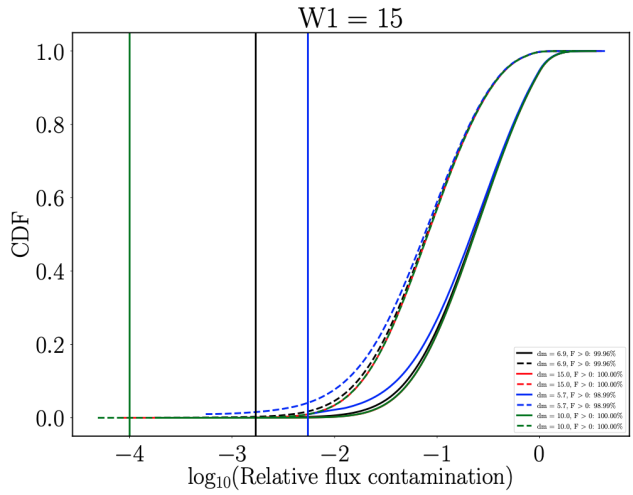

Figure 10. Cumulative distributions of derived flux contaminations from Monte Carlo simulations of blended sources within PSFs, assuming a *WISE* source of $W1 = 15$. Colours and line styles are the same as in Figure 9.

However, this only considered the astrometric modelling, and there is an outstanding question of how to handle the flux contamination of PSF-fit sources at bright magnitudes. We detail here a test into this question, using simulated test image cutouts with realistic *WISE* counts and noise.

5.1 Simulating Crowded PSFs

For a series of *W1* magnitudes, we create small representative images of a single PSF, and add extra sources, representing simulated perturbers. The PSF used is the *WISE* PSF[3], integrated over a pixel, with image cutouts being 13x13 pixels (roughly two $3\sigma_\phi$ radii, to catch the entirety of a perturber placed at its maximum offset). This is implemented using a PHOTUTILS EPSFMODEL, fit via ASTROPY'S LEVMARLSQFITTER. We calculate the "flux," or, more specifically, the DN counts, of an object by $N = 10^{-(m-m_0)/2.5}$, where $m_0 = 20.5$ is the instrumental zero point of the $W1$ filter. The background is chosen to be representative of the same region simulated previously – $l = 130$, $b = 0$ – as the average of the quoted DNs of *WISE* objects in the region, $B = 23$ (again, in DN). It is subtracted from the image once the noise has been simulated. Noise in the image is simulated by drawing each pixel from a Poissonian distribution, with a multiplicative factor of $g = 3.2e^-/\text{DN}$, the gain of the system; this is subsequently removed again once the noise is simulated. The *uncertainty* of each pixel is a combination of several sources: the Poissonian noise from the sources and background (correcting for gain), read noise of 3.1DN, the PSF uncertainty array (as provided with the *WISE* PSF), a flatfield uncertainty of 0.15% flux, and a typical quoted "confusion error" of 0.3DN. These uncertainties are combined quadratically, and the inverse standard deviation is provided as a weight to LEVMARLSQFITTER. The main source is also randomly placed within the central pixel, to simulate various pixel phases, and then realisations of the given number density of objects, from the TRILEGAL simulation used previously, are drawn and placed.

The resulting composite, blended object is then fit with a single PSF, from random starting position and flux; the build up of 50000 realisations is then flipped into a probability density function (PDF) of perturbations and flux contaminations. We also simulate just the central object, with zero contaminants, and fit the resulting distribution of recorded offsets – in both position and flux – to derive statistical, "pure" Gaussian uncertainties. We then analysed the resultant distribution of perturbed source positional offsets and flux brightenings.

5.2 Bright, PSF-Fit Source Contaminations

The astrometric uncertainty functions (AUFs) follow a very similar shape to those seen in the *Gaia-WISE* cross-matches used above and by Wilson & Naylor (2018). We cannot directly compare the simulation to the AUFs seen in the *Gaia-WISE* data, unfortunately, as we cannot simulate several systematic astrometric effects, such as pixel-to-coordinate transformations. We can compare, however, the ensemble *WISE* astrometric uncertainties at each magnitude slice to a quadrature sum of the derived "pure" Gaussian uncertainty from the simulations and a systematic uncertainty. This gives good agreement, with a systematic uncertainty of 0.035", which is a reasonable "missing" uncertainty, suggesting our simulation is robust as an experiment. A comparison of the statistical distribution of perturbations (the perturbation components of the AUF), convolved with this "representative" *WISE* astrometric uncertainty, shows good agreement with the cross-match offsets in the extreme limits of flux. Such an example can be seen in Figure 11, comparing the "flux-weighted" AUF with the *Gaia-WISE* cross-matches shown. This suggests our realisations of perturbing objects are accurate, and thus useful for evaluating the photometric contamination.

The overall trend of H with magnitude for the simulated perturbed positions agrees qualitatively with that seen in the data-driven model; H is larger at bright fluxes, decreasing to zero in the

[3] Cutri et al. (2012), 4.4c

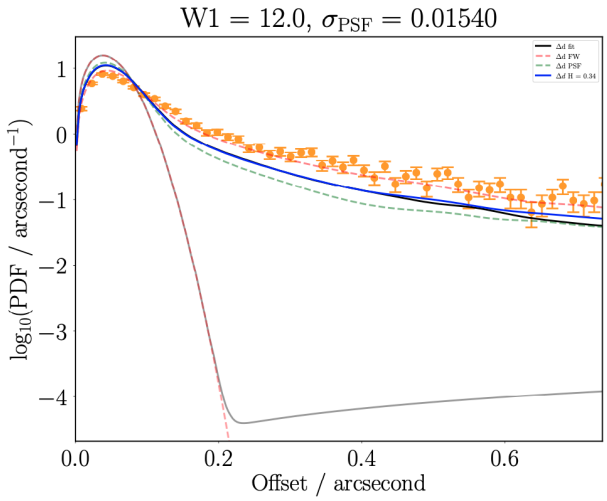

Figure 11. The distribution of perturbations of central object by blended sources. Shown in black are the offsets from "true" of simulated sources, fit via PSF photometry; overplotted in red and green dashed lines are the flux-weighted centroid and background-dominated AUFs, convolved with the intrinsic Gaussian AUF component. Faint black and red lines show simulated "pure" positional offsets, used to derive the intrinsic Gaussian uncertainty, and the intrinsic Gaussian, respectively. The blue line shows the H-weighted AUF; and the orange errorbars show typical *Gaia-WISE* cross-matches for comparison.

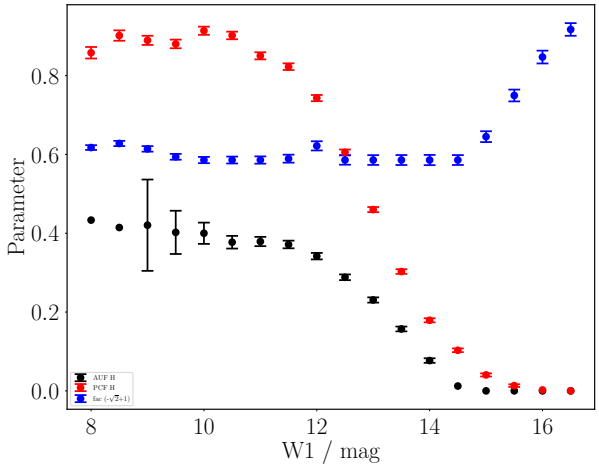

Figure 12. Fit parameters as a function of magnitude for test $W1$ PSF simulations. Shown are the weighting H between flux-weighted centroid (or aperture photometry flux) and background-dominated perturbations (contaminations) for the AUF (and PCF) in black (and red). Also shown, in blue, is the constant of proportionality between astrometry and photometry, for deriving the relative photometric uncertainty of a given source, offset by $\sqrt{2} - 1$.

background-dominated case. However, we find that H does not tend to one (shown in Figure 11), as it does with the full *Gaia-WISE* cross-match AUF fits, suggesting some discrepancy in our fits versus the "full" data – an obvious issue being the lack of multi-band fitting. On the other hand, the "H" for the photometric contamination function (PCF; the statistical distribution of Δf for a given theoretical source) *does* tend roughly to one at bright magnitudes; the discrepancy here, as with the simulated AUF H, is the magnitude at which H starts to drop. Both H trends for the full magnitude range are shown in Figure 12; see Figure 13 for an example of a PCF fit.

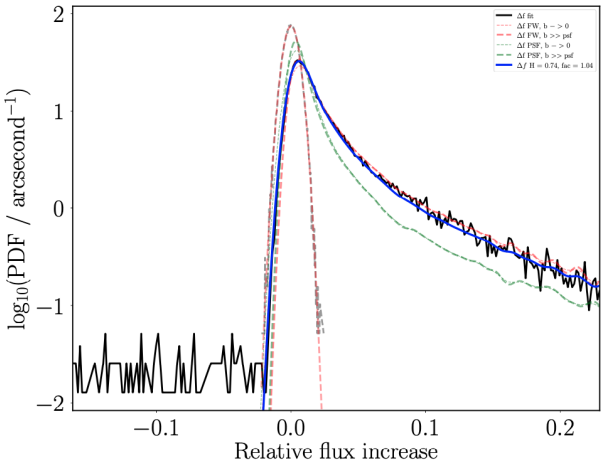

Figure 13. Distribution of photometric contaminations of central object by blended sources. The black line shows the relative flux brightenings of simulated sources, fit via PSF photometry. Faint black and red lines show simulated the "pure" Δf distribution, and its corresponding best fit Gaussian, respectively. The red and green lines show the aperture photometry and background-dominated, PSF-fit flux brightening models, convolved with the "pure" Gaussian noise component, respectively. The dashed and dotted lines (for the red/green lines) show the limits where "fac", the constant of proportionality between relative astrometric and photometric uncertainties, are 1 (no background) and $\sqrt{2}$ (background-dominated), respectively. The blue line shows the H-weighted PCF, with $1 \leq \text{fac} \leq \sqrt{2}$.

Overall then, there is *reasonable* evidence from these simulations that we can derive a completely analagous function for photometry to that used to compute the astrometric component, the PCF (the photometric "AUF"). We should be able to parameterise the contamination component of the PCF as a piece-meal weighted-average of an aperture photometry model at high SNRs and the model previously derived for background-dominated PSF-fit sources. This can then be convolved with a (now one-dimensional) statistical uncertainty component – albeit with the extra proportionality constant linking astrometric and photometric ratios (see Figure 12, and King 1983 for more details) – to provide a full distribution of potential Δf brightenings.

6 EXTENSIONS

While the above parameterisation of the magnitude offsets necessary for inclusion in the AUF model (Section 2) and the improved model for calculating centroid offsets for PSF fit sources in background-dominated noise regimes (Section 3) are vital for the correct modeling of the effects of blended sources on the astrometry of LSST sources, and the improvements to the photometric modelling side in Sections 4 and 5 are also important, this does not represent an exhaustive modelling of all sources of uncertainty in the derived parameters.

A significant unknown at present is how to model the effects of *deblending* of objects – active deblending is where additional sources are included simultaneously in the least-squares minimisation of a PSF fit source. The effects of these additional sources are not included in the AUF – where they would have perturbed the

primary by some Δx they are themselves extracted together with the primary object so no longer do so – but they also have more subtle effects not currently taken into account. Indeed, in the case of bright *WISE* objects, we find that the cross-matches with *Gaia* sources results in a larger average offset for sources which have had active deblending applied than those for which there is no attempt to deblend perturbers. However, only of order 15% of *WISE* sources have active deblending applied to them, and this additional systematic offset from the more precise *Gaia* objects is negligible by $W1 \simeq 14.5$. Hence we ignore the effects of active deblending at present.

We also stress here that this work is not quite finished, as we have prioritised WP Deliverable 3.11.2 (and 3.11.3; see the six month plan April-October 2020 for more details) due to begin May/June 2020 over finishing these minor aspects of WP deliverable 3.11.1. This report provides the qualitative conclusions to the case of the parameterisation of the flux contamination suffered by an object subject to blending at all SNRs, but will require further follow up to smooth over a few loose threads and create the eventual implementation for the final deliverable of the project. We currently assume that this parameterisation of the smooth transition from high to low SNRs through an intermediate weighted average AUF will also be applied to the PCF with the same weighting. A more accurate model for the flux brightening an object is subject to in its crowded field would be a bonus, but as this should always be treated with caution, regardless of the algorithm used, is of lower importance to the project.

7 CONCLUSION

In this report we have laid out an updated algorithm for the description of background-dominated sources for which blended object perturbations need to be accounted for. This new AUF, when combined with a previous algorithm applicable to the case of bright objects, better describes the separations of *Gaia* and *WISE* matches in the Galactic plane than previous attempts alone. When combined with a faint source cutoff motivated by the signal-to-noise ratio of the star, the new modelling gives a more accurate understanding of the cross-matches of sources at the faint end of the LSST dynamic range This is crucial as crowding becomes increasingly important at such depths, while also allowing us to use previously selected Galactic simulations, avoiding issues with simulation faint magnitude limits. We also detailed two improvements to the photometric description of blended sources: the photometric simulation limit, and the photometric contamination model, which will allow us to report more detailed, accurate, and precise information on the levels of flux contamination suffered by objects subject to blending than previously available.

REFERENCES

Cutri R. M., et al., 2012, Technical report, Explanatory Supplement to the WISE All-Sky Data Release Products
Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, A&A, 436, 895
King I. R., 1983, PASP, 95, 163
Plewa P. M., Sari R., 2018, MNRAS, 476, 4372
Rayleigh L., 1880, MNRAS, 40, 254
Wilson T. J., Naylor T., 2018, MNRAS, 481, 2148

This paper has been typeset from a TEX/LATEX file prepared by the author.