

# Report Mar 2020. Improving the perturbation AUF for faint sources

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 (this is 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.

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.

## 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.

## 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,

$$F_s \geq B\sqrt{F_p + S + fF_pQ}, \quad (1)$$

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

$$F_s \geq B\sqrt{F_p + S} = B\sigma_{F_p}. \quad (2)$$

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

$$f \geq B\frac{\sigma_{F_p}}{F_p} = B\text{SNR}_p^{-1}. \quad (3)$$

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  $\text{SNR}_p \geq 5$ . 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 Galactic model allows.

A star 5 magnitudes brighter than the faint limit, assuming it is still sky limited, will have  $\text{SNR}_p = 500$ , 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  $\Delta m$  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, but we do not currently model these when considering the astrometric functions for the rest of this report.

### 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)<sup>1</sup> was simply given by  $x_{\text{pertub}} = (\sum_i x_i f_i) / (1 + \sum_i f_i)$  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 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

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

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  $\Delta f$  at perturbation vector  $\mathbf{r} - \Delta \mathbf{d}$ . Here  $\phi$  is the equation describing the circularly symmetric Gaussian PSF,

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

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

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

Expanding the brackets results in six integrand terms, each

<sup>1</sup> 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 which has a multiplicative factor in front of the multiple of two  $\phi$  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  $\phi$  term has the same uncertainty  $\sigma_\phi$ , 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

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

Only three of the resulting convolutions contain terms with  $\Delta f$  or  $\Delta \mathbf{d}$  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

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

For small perturbations by very faint perturbers, we can derive analytic expressions for  $\Delta f$  and  $\Delta x$  (and  $\Delta y$ , 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  $\Delta f$  we have

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

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

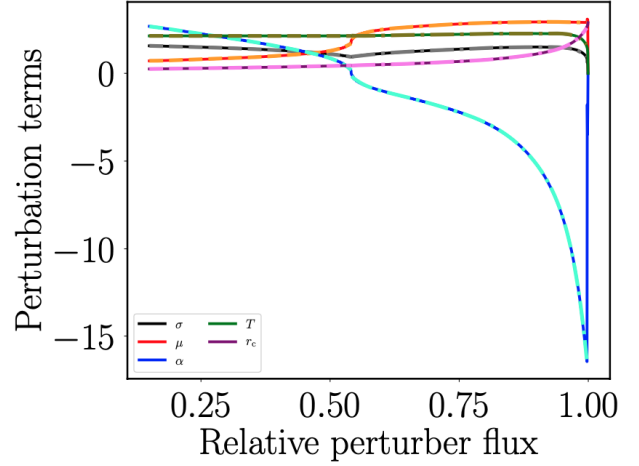
$$\begin{aligned} \psi'(\Delta \mathbf{d}) + \sum_i f_i \psi'(\mathbf{d}_i - \Delta \mathbf{d}) - (1 + \Delta f) &= 0 \\ \Delta f = \psi'(\Delta \mathbf{d}) - 1 + \sum_i f_i \psi'(\mathbf{d}_i - \Delta \mathbf{d}). \end{aligned} \quad (10)$$

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

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

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

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



**Figure 1.** Parameterisation of composite flux-weighted and skew normal model of single object perturbation,  $\Delta x$ . 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.

Rearranging and using the previous definition of  $\psi'$  we get

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

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

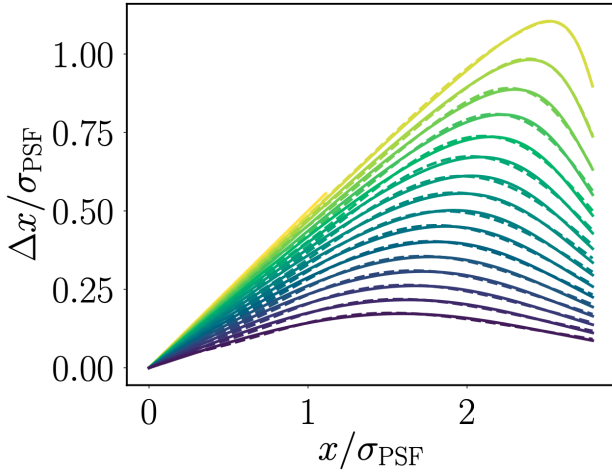
$$\Delta x \simeq \sum_i f_i \Delta x_i \psi'(\mathbf{d}_i) = \sum_i f_i x_i \exp\left(-\frac{1}{4} \frac{|\mathbf{d}_i|^2}{\sigma_\psi^2}\right). \quad (14)$$

While these approximations of position perturbation 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.

### 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  $\Delta x$ , 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  $\Delta x$  to that of  $x$ .

The parameterisation of  $\Delta x$  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



**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  $f \approx 0.15$  to  $f \approx 0.1$ . Empirical perturbations are shown in solid lines, with parameterisation shown as dashed lines.

a composite of flux-weighted averages at small source offsets and a skew normal distribution at larger source offsets:

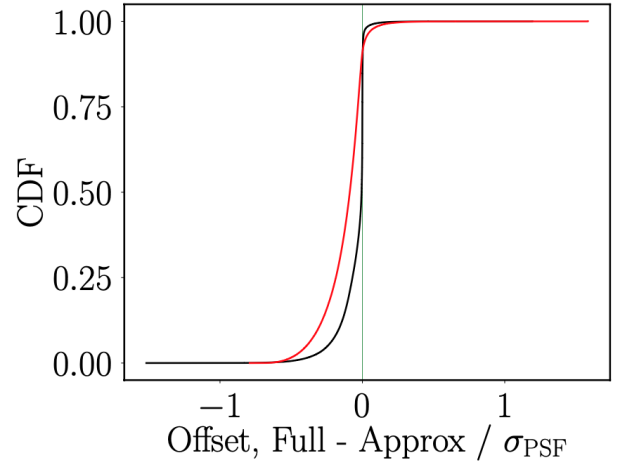
$$\Delta x(x, y, f) = \begin{cases} f x \exp\left(-\frac{1}{4} \frac{x^2 + y^2}{\sigma_\psi^2}\right) & f < 0.15 \\ \Omega(x, f) & f \geq 0.15, \end{cases} \quad (15)$$

where

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

and  $\lambda(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right)$ ,  $\Lambda(x) = \int_{-\infty}^x \lambda(t) dt = \frac{1}{2} \left[ 1 + \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right) \right]$ ,  $r_c$  is the critical offset at which perturbation stops being modelled as flux-weighted, and  $\sigma$ ,  $\mu$ ,  $\alpha$ , 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  $\sigma$  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  $\Delta x$  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,



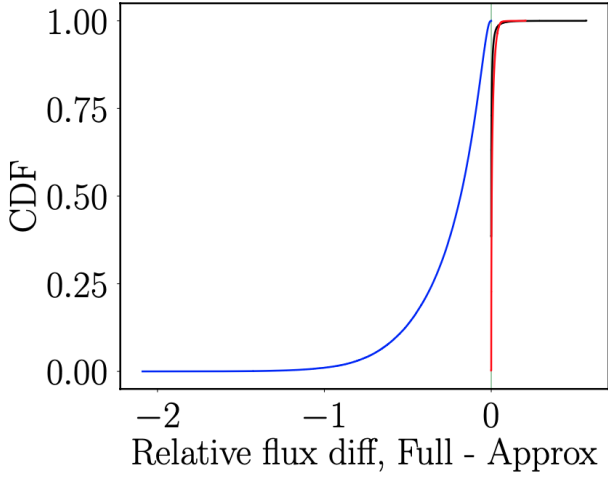
**Figure 3.** Comparison between the full background-dominated PSF fit centroid shift computation, cf. equation 13, and the vector sum skew normal parameterisation (equation 15), shown in black, and the flux-weighted centroid algorithm previously used by Wilson & Naylor (2018), shown in red.

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  $\approx 15.5$ . 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  $\Delta f$  computed using several algorithms, as seen in Figure 4. The full  $\Delta f$ , 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 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 –



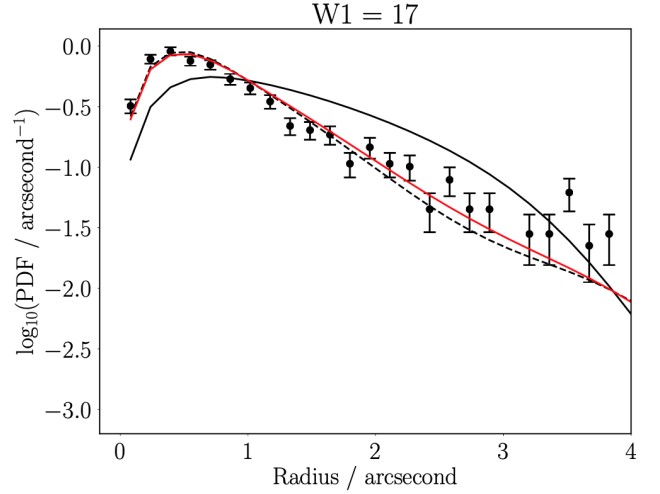
**Figure 4.** Comparison between the full background-dominated PSF fit flux increase computation and various other algorithms. The difference between the full PSF fit flux increase derived by equation 10 and the increase computed using the perturbations calculated using the approximation given in equation 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.

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 ( $W1 \approx 13$ ) in the *WISE* survey. However, this model is clearly superseded 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  $W1 \approx 17$  much better than the flux-weighted centroiding (solid line)<sup>2</sup>. The exact opposite is true at  $W1 \approx 12$  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



**Figure 5.** Model AUFs compared with *Gaia-WISE* cross-matches of  $W1 \approx 17$  (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.

– 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

$$y = H \text{AUF}_{\text{fw}} + (1 - H) \text{AUF}_{\text{bd}} \quad (17)$$

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  $W1 \approx 14.5$  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.

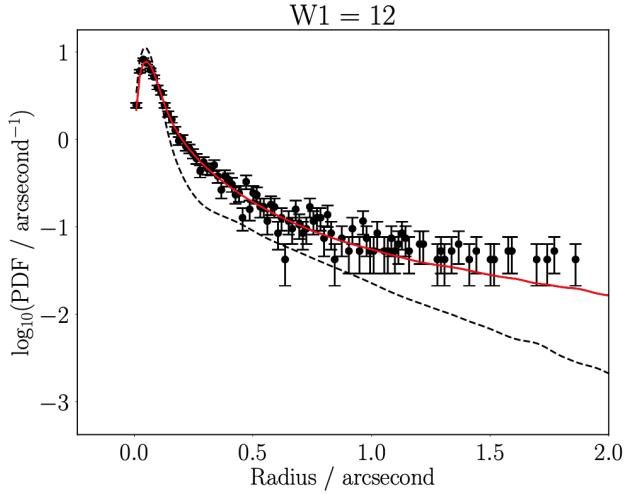
We initially explored the possibility that the mis-match between the physically motivated AUFs and the cross-matches was due to the combining of multiple “groups” of data into a single dataset. However, in all cases the *WISE* data were drawn from a consistent SNR regime, whether measured by pure SNR, background flux counts, or star-to-background flux ratio. Thus, unable 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,  $W1 \approx 14.5$  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  $W1 = 17$ .

## 4 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

<sup>2</sup> 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.



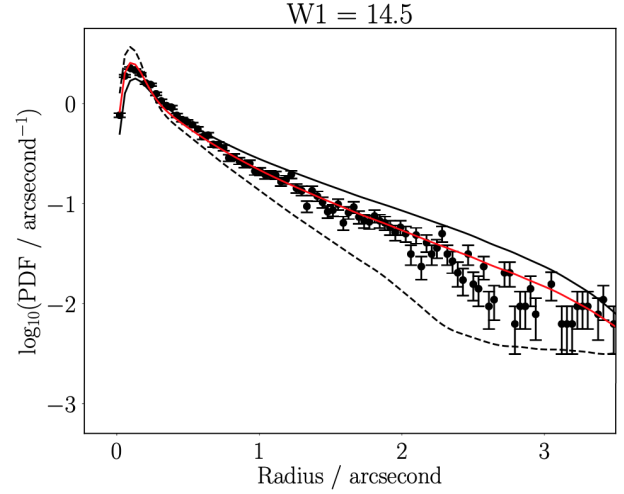


**Figure 6.** Model AUFs compared with *Gaia*-*WISE* cross-matches of  $W1 \approx 12$  (shown as black errorbars). Solid black line is AUF as derived using the Wilson & Naylor (2018) flux-weighted centroid calculation, and the dashed black line is the new, background-dominated PSF derived centroid shift 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).

correct modeling of the effects of blended sources on the astrometry and photometry of LSST sources, 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  $\Delta x$  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  $W1 \approx 14.5$ . Hence we ignore the effects of active deblending at present.

Another question which must be resolved is that of the photometric flux limit to simulate down to, as discussed briefly in Section 2. For all simulated AUFs discussed in this report we only used the  $\Delta m$  limit derived for astrometric limits, using the SNR of the primary source to set a limit at which objects can no longer perturb the astrometric position significantly compared with the effect on the primary position of the photon noise. However, in addition to measuring the true uncertainty in star positions, a secondary aim of this work package is to use the astrometric shifts to model the level to which the reported flux of the source is too high because of contaminating stars ( $\Delta f$ ). Hence we need to ensure that our models include all sources that will have an influence on the brightness of the central object. We have attempted to model this second flux limit, and are partway through this effort, but if we are to remain on



**Figure 7.** Model AUFs compared with *Gaia*-*WISE* cross-matches of  $W1 \approx 12$  (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.

schedule must now move on to the next phase. Practically, even if we were to find that we required deeper models to model the change in photometric flux these models are not available, and so we would only be able to estimate the effects of this very faint perturbers.

Finally, further to the point above, we also have to finalise the  $\Delta f$  parameterisation at all SNRs. We have shown that the aperture sum flux contamination calculation for background-dominated, PSF fit sources (cf. Figure 4) results in the belief that source photometry, as quoted in a photometric catalogue, is significantly fainter than it is in reality, as the  $\Delta f$  can be wrong by a large fraction of the source flux. We can turn to the log-likelihood derived  $\Delta f$  for these faint sources; however, we do not know at present whether this flux brightening can hold for *bright* sources that have had PSF photometry applied to them. We will therefore also have to model the effects of contaminant sources on bright objects that are fit by PSF photometry, and verify whether this algorithm holds for such objects, or if the flux-weighted algorithm previously used – which does hold for bright objects in astrometry – also applies to the photometry in this SNR regime as well. This will allow us to as accurately as possible model both aperture and PSF photometry-based catalogues, ensuring as robust as possible understanding of the effects of crowding across a wide range of datasets. However, as with the photometric AUF limit consideration above, the quoted statistical flux contamination levels are a secondary aim – the primary focus being on the astrometric effects of crowded sources – and thus we are less concerned with this goal at present. 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.

## 5 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, crucial as crowding becomes increasingly important at such depths. There remain problems with estimating the effects of crowding on the magnitudes of stars, but this is a secondary aim, and we must now move on to implementing the cross matching in an environment which mimics the DAC.

## 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  
 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  $\text{\LaTeX}$  file prepared by the author.