Report Jan 2020 – Attempts at parameterising an empirical AUF

Tom J. Wilson and Tim Naylor

ABSTRACT

Given the depths that LSST will reach, we possibly require a different approach than previously undertaken to compute the perturbation component of the Astrometric Uncertainty Function (AUF) for probabilistic photometric catalogue cross-matching. We document here an attempt to derive this AUF – along with any other non-statistical AUF component, such as proper motions or binary reflex motion – from the data themselves, creating in-situ AUFs which do not rely on simulation. Conclusions currently suggest this is not a good approach to take, due to the degeneracies of convolution and the unphysicality of the results.

1 INTRODUCTION

The Astrometric Uncertainty Function (AUF) is the probability density function (PDF) describing the probability of a source being detected at some position given its true location. The probability of two sources being the same intrinsic object detected twice given their respective locations – or the distance between their given positions – is the convolution of their respective AUFs. Each AUF is, itself, a convolution of the individual components describing its offset from its "true" location. The first of these, and typically the only component used, is its statistical component, in which Poisson noise in a photometric image causes imperfect centroiding of a source.

A second critical component in crowded regions, or in photometric surveys which probe significant depths relative to the length scale of the survey PSF (to which LSST and *WISE* are both subject), is the perturbation from blended sources. These objects, while hidden beneath the flux from a brighter source, contaminant the central object by both adding additional flux, and by tugging on the center-of-light of the composite object. Thus, in this simple two-component model, the AUF is

$$h(\mathbf{x}) = (f_{\mathrm{stat}} * f_{\mathrm{blend}})(\mathbf{x}) \tag{1}$$

where the notation $(f * g)(x)$ denotes a convolution.

Previously, Wilson & Naylor (2018) produced f_{blend} by a Monte Carlo simulation, via Galactic stellar densities from a TRILE-GAL simulation (Girardi et al. 2005). However, the limiting magnitude of these simulations – via the public API – is 32nd magnitude; this was fine for *WISE* where we still had access to stars 10 magnitudes fainter than the limit of the survey (necessary for probing the faintest flux ratios and sufficiently small flux weighted astrometric perturbations), being brighter than $W1 = 32$. However, the limiting magnitude of LSST is about 27th magnitude, and thus the magnitude limit is out of the range of the TRILEGAL simulations.

2 IN SITU AUF

Thus the need for deriving empirical AUFs was born. Instead of modelling the perturbations from each component individually –

simulating blended source position shifts, unknown proper motions, etc. – we could extract from the data themselves the AUF describing *all* components, minus the statistical noise component, and use the data-driven AUF subsequently.

The idea here was that we can reverse the equation above, and compute f_{blend} directly from knowing the presciption for the statistical AUF and the total PDF describing the probability of two detections being the same source detected twice given their separation. To this end we matched the *WISE* data, thought to be well understood and accurately described by its quoted uncertainties, to *Gaia* data, using the high precision on the *Gaia* dataset to effectively reduce the convolution of the two datasets to that of just the *WISE* data (see Wilson & Naylor 2017, e.g.), allowing for a one-sided probe of the test.

We are therefore looking for the deconvolution of f_{stat} and f_{deblend}, for which we can turn to Fourier transforms. Where $F = \mathcal{F}[f(x)](\rho)$ denotes the Fourier transform of f, we have

$$h = f * g \tag{2}$$

and, if we take the Fourier transform,

$$H = F \times G \tag{3}$$

where in Fourier space the convolution is described by a simple multiplication. Thus we can solve for G – or f_{blend} – as

$$G = H/F \tag{4}$$

and inverse Fourier transform, giving

$$f_{\mathrm{blend}}(\mathbf{x}) = \mathcal{F}^{-1}\left[\mathcal{F}[h(\mathbf{x})](\rho)/\mathcal{F}[f_{\mathrm{stat}}(\mathbf{x})](\rho)\right](\mathbf{x}). \tag{5}$$

2.1   Computing the Fourier Transforms

To reduce the computational expense of a two-dimensional convolution by simplifying it to a 1D integral, we use the Hankel transform,

$$\mathcal{F}[f(r,\theta)](\rho) = \mathcal{F}[f(r)](\rho) = 2\pi \int_0^\infty r f(r) J_0(2\pi r \rho) \, \mathrm{d}r \tag{6}$$

where we assume circular symmetry (i.e., the Hankel transform, or Fourier-Bessel transform, is a two-dimensional Fourier transform under the assumption of a radially symmetric integral kernel) and J_0 is the zeroth order Bessel function of first kind. This is computed via a series of small Δr bins with each function evaluated at small r (or ρ for an inverse transform) offsets, for a given ρ (or r, for inverse transforms).

There is, however, a small instability in this function computed numerically — no matter how finely you evaluate the functions, J_0, effectively a combination of a $1/\sqrt{(x)}$ function and a $\cos(x)$ function, cannot be reasonably computed with sufficient precision at high ρ in the Fourier transform. At large ρ, the $\cos(x)$ component of J_0 becomes very rapid, cyling completely every $1/r$ steps (remember it is $J_0(2\pi r \rho)$ and thus goes as $\cos(2\pi r \rho)$). This numerical instability is acceptable for a *convolution* (being of order 10^{-5}), as a convolution *multiplies* Fourier tranforms, and thus our large scale Fourier transform values, which should be effectively zero, are truncated at 10^{-10}, and thus do not contribute much when inverse Fourier transformed. Unfortunately, for a *deconvolution*, the functions are *divided*, resulting in an inverse Fourier transform which, for large ρ, has a constant *but significant* value – tending to of order unity. We therefore cannot compute the deconvolution numerically.

2.2   Mixture of Gaussian Model

Instead of attempting the deconvolution head-on, however, we can sidestep the issue of fourier transform stability altogether. The convolution of two Gaussian functions is analytically known – the new variance is simply the sum of the variances of the original two functions. Thus, if we were able to describe the total AUF as a sum of N Gaussians, we could trivially remove the statistical noise component of the AUF. To do so we fit a Mixture of Gaussian model (MoG),

$$f(x) = \sum_i \frac{a_i}{2\pi\sigma_i^2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma_i^2}\right) \tag{7}$$

where $\sum_i a_i = 1$ to ensure proper normalisation. We fit this via the scipy package module basinhopping (and minimize), although as we fit to the separation *counts* we fit a Rayleigh distribution, as opposed to a Gaussian distribution, and thus fit

$$f(x) = \sum_i \frac{a_i x}{\sigma_i^2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma_i^2}\right) \tag{8}$$

again assuming circular symmetry. We also limited all $\sigma_i > \sigma_{\text{stat}}$ to ensure a non-imaginary uncertainty in the deconvolution AUF, computed as $\sigma_{\text{blend},i} = \sqrt{\sigma_i^2 - \sigma_{\text{stat}}^2}$, giving

$$f_{\text{blend}}(x) = \sum_i \frac{a_i}{2\pi\sigma_{\text{blend},i}^2} \exp\left(-\frac{1}{2}\frac{x^2}{\sigma_{\text{blend},i}^2}\right). \tag{9}$$
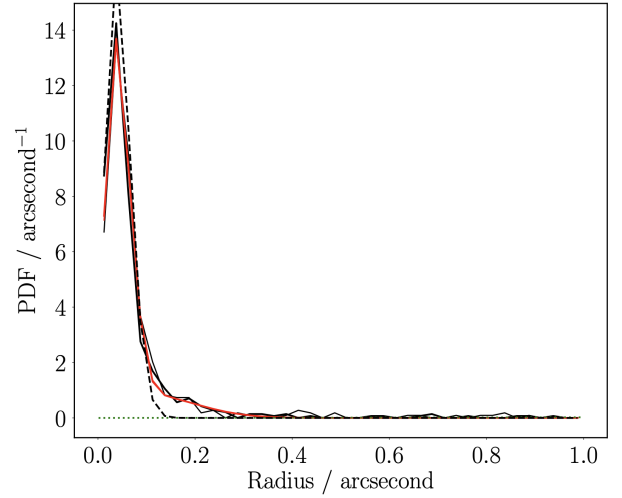

Figure 1. Fit of a Mixture of Gaussian model to *Gaia-WISE* nearest neighbour matches for $W1 = 9$. Red line is MoG model, solid black histogram shows cross-match countrate, and dashed black line is the instrinsic astrometric uncertainty.

This process was repeated for steps of one magnitude $9 \leq W1 \leq 17$ (plus $W1 = 16.5$, as $W1 = 17$ suffered from potential small-number statistics effects), for small steps in normalising density ($\Delta N = 0.01$, fitting total star counts $W1 \lesssim 15$ to Nz^m; see Wilson & Naylor 2018 for more details), magnitude ($\Delta W1 \leq 0.1$, except $\Delta W1 = 0.3$ for $W1 = 17$), and quoted astrometric uncertainty ($\Delta \sigma \leq 0.01$, except $\Delta \sigma = 0.02$ for $W1 = 17$). All results have $\chi_\nu^2 \leq 2$ (except $W1 = 11$ where $\chi_\nu^2 = 3.5$) and thus for the most part suggest good fits to the data, an example of which is shown in Figure 1.

However, Figure 2 shows the resulting deconvolved AUF, with the statistical component removed. Also shown is a Monte Carlo simulation (as per Wilson & Naylor 2018), with little agreement between the two. We currently cannot explain the physicality of the MoG deconvolved model, as technically there are many g distributions which, when convolved with f, can reproduce h.

While this *may* be satisfactory, one of the strengths of the original Wilson & Naylor (2017) algorithm was its roots in the physical process, and thus we are currently loathed to persue this line of research any further at present. If, in the future, we cannot resolve the magnitude limit issue for LSST images and catalogues we may be forced to adopt an unphysical, but data-driven, model to describe faint crowded regions in LSST data.

3   AUF PARAMETER REDUCTION

During this process we also attempted to explore the possibility of reducing the number of free parameters for the description of a perturbation AUF. Currently, based on the trilegal simulations, we require knowledge of the Galactic coordinates (l and b, e.g.), central source brightness, and normalising density N (used to scale the trilegal simulated star counts). However, where star counts are described as a power law, each step in magnitude results in a factor z increase in *relative* star counts. Thus, perhaps, we could remove N as a factor and instead use N_{eff}, this "relative" density.

To test this, we took the previously computed perturbation AUFs – whether physical or not, their convolution with a statistical
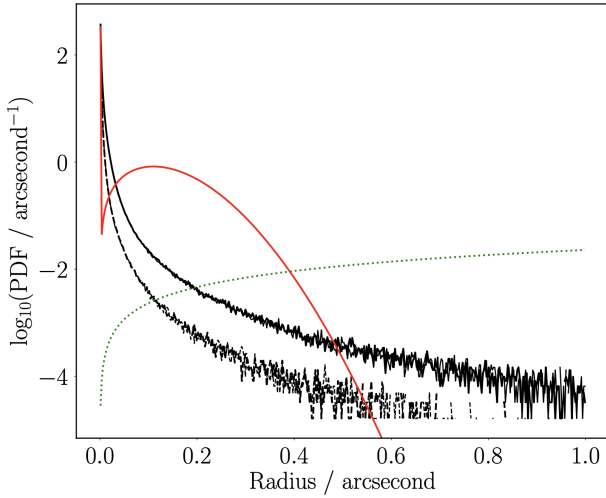
Figure 2. Deconvolution of MoG model to *Gaia-WISE* sources for $W1 = 9$. Solid black histogram is the Monte Carlo simulation distribution of perturbation offsets, with the MoG model deconvolved shown in red.

AUF component *does* describe the total AUF of the data well – and convolved it with the statistical uncertainty of a new set of sources, with magnitude $m + dm$. Again we forced the width of ΔN, Δm, and $\Delta \sigma$ to be small. We then re-selected sources based on the mode σ_{stat} and a new normalising density, a factor z^{dm} smaller, to give the same resulting N_{eff}.

While the brightness of the central source is bright, and the power-law distribution of perturbing source counts holds (for most sources, anyway), this distribution is well fit by the new AUF. However, when the central source is fainter, and most perturbing sources are in the turnover of source counts (see Figure 5, Wilson & Naylor 2018), we overestimate the number of these perturbing objects, and thus have an AUF which is too broad, and a poor fit to the data. Unfortunately, we conclude that – perhaps obvious with hindsight – due to the complex nature of stellar brightness counts within the Galaxy, the naive power law approximation of star counts does not hold and thus cannot be used to simplify our parameterisation in this case, just as it could not in previous Monte Carlo simulations of star perturbations.

4 CONCLUSIONS

We have explored some attempts to move the full perturbation-inclusive AUF to a new regime with LSST, in which we would have a greater ability to use image frame-level information to better capture the effects of blended sources. Unfortunately, preliminary investigations are inconclusive and ongoing, with a MoG fit to the data which does not currently have a physical explanation. If we can explain the effects and fully parameterise the distributions with MoG models, and thus analytically convolve statistical noise separations with further systematic offsets, we can look further to using LSST images – detected source residuals or "empty" sky otherwise – to characterise the perturbations of sources below the detection limit, extending the data-driven AUF to fainter magnitudes still.

REFERENCES

Girardi L., Groenewegen M. A. T., Hatziminaoglou E., da Costa L., 2005, A&A, 436, 895
Wilson T. J., Naylor T., 2017, MNRAS, 468, 2517
Wilson T. J., Naylor T., 2018, MNRAS, 481, 2148

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